Wave Motion 88 (2019) 226–256

Contents lists available at ScienceDirect

Wave Motion

journal homepage: www.elsevier.com/locate/wamot

A causality-preserving Fourier method for gravity waves in a viscous, thermally diffusive, and vertically varying atmosphere

Harold K. Knight [a,*], Dave Broutman [a], Stephen D. Eckermann [b]

[a] Computational Physics, Inc, Springfield, VA, 22151, USA
[b] Space Science Division, U.S. Naval Research Laboratory, Washington DC, 20375, USA

HIGHLIGHTS

- A general causality-preserving multilayer Fourier method for linear PDEs is described.
- Numerical swamping is avoided using the Kennett algorithm from seismology.
- The method is applied to atmospheric gravity waves with viscous and thermal damping.
- Imaginary frequency shifts allow identification of upgoing and downgoing modes.
- Dissipative modes must be included in order for causality to be preserved.

ARTICLE INFO

Article history:
Received 28 June 2018
Received in revised form 24 April 2019
Accepted 5 June 2019
Available online 10 June 2019

Keywords:
Multilayer method
Boundary value problem
Dissipative modes
Scattering matrix
Kennett algorithm

ABSTRACT

A multilayer Fourier method is developed for modeling gravity waves above some altitude at which the vertical velocity is given as a boundary condition. The method is described in a general context in which the altitude and time variation of Fourier components for fixed horizontal wavenumbers is specified by a linear homogeneous partial differential equation (PDE) of any order, with coefficients varying with altitude. The coefficients are required to meet certain conditions so that causality is preserved and upgoing and downgoing modes can be defined. It is shown that causality is not preserved unless dissipative modes are included. An imaginary frequency shifting technique is introduced to allow upgoing and downgoing modes to be defined in certain situations that would otherwise be problematic. The pervasive problem of numerical swamping, endemic to multilayer approaches to viscous and thermally diffusive gravity-wave problems, is solved using a scattering matrix method, related to a method from seismology, which differs fundamentally from earlier gravity-wave solution methods cited herein. The method is applied to gravity waves in the linear, nonhydrostatic, anelastic, viscous, and thermally diffusive case with vertically varying background winds, in addition to certain limiting cases. The PDE in altitude and time is obtained from a dispersion relation that includes odd powers of the vertical wavenumber, making necessary the aforementioned imaginary frequency shifting technique.

1. Introduction

Atmospheric gravity waves generated in the lower atmosphere can propagate deep into the upper atmosphere (100–500 km altitude), where they interact with the ionospheric plasma in ways that disturb the ionosphere, affect radiowave

* Corresponding author.
E-mail address: knight@cpi.com (H.K. Knight).

https://doi.org/10.1016/j.wavemoti.2019.06.001

signal propagation, and degrade communication and geolocation applications. This importance arises in large part due to the exponential decrease in atmospheric density ρ with altitude z. In the lower atmosphere, processes such as deep convection in weather systems, jet-stream imbalances, and flow over mountains generate a spectrum of gravity waves, also known as buoyancy waves or internal waves (see, e.g., the review of [1]). As these gravity waves propagate upward, energy conservation requires that the wave amplitude solution have a growth term that is approximately proportional to $\rho^{-1/2}(z)$. This growth in wave amplitude with height can lead to instability and wavebreaking, especially in the middle atmosphere (10–100 km altitude), where such wavebreaking is an important source of turbulent mixing and plays a significant role in forcing the global weather and climate across scales [1]. Gravity waves that reach the upper atmosphere are subject to damping by molecular viscosity and thermal diffusion, which increases exponentially in proportion to $\rho^{-1}(z)$.

From the early days of atmospheric gravity-wave research (e.g., [2]), it was recognized that viscous and thermal damping must have a substantial influence on gravity-wave dynamics in the upper atmosphere, but the modeling of viscous and thermally diffusive gravity-wave dynamics has proven to have many challenging aspects. Among them is the increase in the order of the gravity-wave dispersion relation from second order in vertical wavenumber m to sixth order in m, resulting from the inclusion of viscous and thermal damping terms [3].

If kinematic viscosity ν and thermal diffusivity $\alpha \equiv \nu/Pr$ (where Pr is the Prandtl number, which is constant for a given atmospheric composition) are assumed constant in the governing equation derivations, as in Pitteway and Hines [4] and subsequent studies such as Midgley and Liemohn [5], Volland [6,7], Francis [8], Yeh and Liu [9], Klostermeyer [10,11], and ion drag and Coriolis force are omitted [12], the dispersion relation has only even powers of m and clearly separates into three pairs of upgoing and downgoing gravity waves, viscosity waves, and thermal-diffusion waves (the latter more commonly known as heat or thermal-conduction waves, e.g., [7,12]). The viscosity-wave and thermal-diffusion-wave modes are also referred to as dissipative modes. The main distinction between the two pairs of dissipative modes (i.e., viscosity and thermal-conduction waves) and the pair of gravity-wave modes is that the latter have smaller imaginary parts, which means that upgoing gravity-wave modes do not decrease in amplitude as rapidly with increasing altitude as the dissipative modes. Letting ν and α decrease towards zero, the gravity wave m roots approach the inviscid m roots, while both the real and imaginary parts of m for the dissipative modes approach infinity ([13] eq. (20b)). (Henceforward, the term "inviscid" is understood to refer to situations in which there is neither viscosity nor thermal diffusion.)

The relative significance of the dissipative modes for numerical solutions remains unclear. Francis [8, Appendix], citing the results of extensive unpublished numerical calculations, asserted that the dissipative modes could be left out of atmospheric gravity-wave calculations without a significant effect on the accuracy of the results, but it should be noted that this was based on the unrealistic assumption that ν could be treated as a constant in the governing equations. In the present work, we establish that all modes are in fact needed in order for the modeled wavefield to respond to forcing terms in a mathematically causal way.

The earlier studies just described (in particular, [5,6,8,10,11]) tended to rely on multilayer methods, involving sequences of many thin, altitude-delimited, adjacent layers within which the background parameters are assumed constant. As will be discussed in Section 4, the inclusion of dissipative modes in such methods gives rise to the problem of numerical swamping [14], which occurs when dissipative modes with large Im(m) are extended in the opposite direction to that in which they are damped. A number of work-arounds for this pervasive problem have been proposed, such as the iterative approach of Midgley and Liemohn [5], which breaks down at altitudes above approximately 200 km [6], or simply omitting viscosity and thermal diffusion from the governing equations below 150 km [6,10].

As was pointed out by Pitteway and Hines [4] and Vadas and Fritts [3], the assumption of constant $\mu = \rho\nu$ in the governing equations is more physically justifiable than constant ν, even though the latter leads to a more mathematically tractable problem. When the altitude dependence of ν is incorporated into the governing equations, assuming constant μ as in [3], the resulting sixth order dispersion relation contains odd powers in m (or a related variable), which in some cases complicate the assignment of the roots of the dispersion relation to distinct sets of upgoing and downgoing modes (see Section 5). Vadas and Fritts [3] avoid complications associated with describing the roots in m of their sixth order dispersion relation by assuming that m is real and solving the resulting simpler equations for complex intrinsic frequency $\widehat{\omega}$. This approach results in gravity-wave group trajectories (ray paths) with viscous reflection altitudes that have no counterpart in the inviscid theory, with vertical wavelengths decreasing near the reflecting altitude. Vadas [15] introduces a condition for determining whether the [3] ray approximation is valid approaching a viscous reflection altitude. At the same time, Vadas [15] reaffirms the finding in [3] that vertical wavelength decreases as the wave dissipates, given uniform temperature. The reader is referred to Walterscheid and Hickey [16], Vadas and Nicolls [17], and Heale et al. [18] for additional discussion related to this.

Finite-difference methods have successfully been applied to the problem of numerically modeling atmospheric gravity waves, as described in, e.g., Lindzen and Kuo [19],Hickey et al. [20,21]. In the finite-difference approach, real frequencies and horizontal wavenumbers are generally assumed, but Vadas and Nicolls [17] point out that it is unclear whether real frequency solutions will always yield accurate solutions for transient gravity-wave packets. One drawback of finite-difference methods versus the multilayer approach is that for the former the needed vertical grid resolution depends on the vertical wavelength being modeled. Lindzen and Kuo [19] assert that whenever a continuous solution for the governing equations and boundary conditions exists, their finite-difference method is expected to provide it (in the infinitesimal grid interval limit), but the ambiguity as to how to define upgoing and downgoing modes described in the previous paragraph raises the question of whether the existence of a solution can safely be assumed in all relevant cases.

Given the fundamental importance of both understanding and accurately modeling gravity-wave dynamics in viscous atmospheres, as well as unresolved questions about existing numerical methods, this study offers an appropriate new detailed theoretical attack on the problem. An overarching theme that guides our theoretical development is preservation of causality. The subject of causality has only rarely appeared explicitly in the literature on atmospheric gravity waves, although it is implicit in the concepts of the radiation condition and the direction of group propagation (e.g., [22]). In the gravity-wave context here, causality is the requirement for physical consistency that a solution to any problem in which all forcing terms (e.g., a nonzero boundary condition) first appear at or after time $t = 0$ must continue to vanish at $t < 0$ as the solution is extended from the location of the forcing term. We find that the multilayer concept facilitates the resolution of questions related to causality while providing a practical and accurate technique for deriving numerical solutions to upper atmospheric gravity-wave problems with viscous and thermal damping.

In Section 2 we establish basic mathematical causality conditions within the Fourier space routinely used to derive gravity-wave solutions. In Section 3 we derive basic causality criteria from these conditions for partial differential equation (PDE) boundary value problems of relevance to Fourier solution methods, assuming initially constant coefficients typical of a homogeneous background atmosphere. This purely mathematical result is later seen in Section 5.2 to resolve one of the aforementioned outstanding questions in the literature on multilayer viscous gravity-wave solutions by establishing that all modes must be included for causality to hold. Namely, given a specification of a wavefield incident at a lower boundary that first appears at or after $t = 0$ and a radiation condition at the upper boundary, there are no $t < 0$ effects in the model if and only if all upgoing modes are included: a similar statement applies when the roles of the upper and lower boundaries are reversed.

This work also highlights the importance for causality of identifying branch points in root functions giving m as a function of complex frequency. We find that the problem described above of identifying upgoing and downgoing modes when odd powers of m are present in the dispersion relation is related to the branch point issue and can be overcome using a technique involving the addition of a constant pure imaginary term to frequencies (see Section 3.4). Unlike Vadas and Fritts [3], however, we do not assume that m is real.

In Section 4 we generalize these results to coefficients that vary in some dimension. By reformulating the problem in terms of scattering matrices (e.g., [23]), which are closely related to reflection and transmission matrices appearing in seismology, as in [24, Section 7.2.4], we derive a general form of causality-preserving multilayer solution that avoids the aforementioned problems of numerical swamping and limited applicability while obviating earlier critiques of the validity of multilayer solutions [25]. Our multilayer solution can be applied to a general PDE in two variables, t and z, with causality holding in the time-like variable t and the PDE coefficients varying in the height-like z variable, provided that the PDE coefficients meet certain conditions described in Section 4.4. Our technique gives an exact solution to the PDE in the limit of vanishing layer thickness. Partial reflection, i.e., coupling between upgoing and downgoing modes, is included, but we also find that a simplified version of our technique, without partial reflection, is equivalent to the WKB approximation in the second-order case, again, in the limit of vanishing layer thickness (end of Section 4.3).

In Section 5 we apply the method and insights developed in earlier sections to the linearized, viscous, and thermally diffusive gravity-wave equations, derived by Vadas and Fritts [3] for constant μ and assuming a horizontally uniform but vertically varying background. We begin with the anelastic dispersion relation derived by Vadas and Fritts [3, eq. (23)], which constrains m and intrinsic frequency $\widehat{\omega}$ in terms of the background parameters and horizontal wavenumbers. This gives rise (and directly corresponds) to what we call a dispersion relation PDE, to which our method is applied. Relevant properties of root functions of the dispersion relation just described, as well as Boussinesq and non-thermally diffusive (but viscous) approximations thereof, are examined in Sections 5.2, 5.3, and 5.4 to establish that the conditions required for causality are met, and it is found that an imaginary frequency term, alluded to above, is needed in deriving solutions except in the Boussinesq approximation.

In Section 6 we apply our method to two types of transient gravity-wave packet examples in which reflection and/or strong viscous and thermal damping occurs. The first example, in Section 6.1, involves viscous and thermally diffusive anelastic gravity-wave packets for which the central frequencies and vertical wavenumbers correspond to examples from Vadas and Fritts [3, Figs. 1 and 2]. The second example, in Section 6.2, involves an inviscid Boussinesq gravity-wave packet reflected at a turning point in a case for which an exact analytic solution exists. The close agreement between the two methods illustrates an additional application of our method, which is to describe the behavior of inviscid atmospheric gravity waves in the vicinity of turning points without the need for special functions. Finally, in Section 7, we briefly summarize our results and major conclusions and suggest possible areas for future work that might build upon these results.

2. Fourier method causality

The Fourier transform and its inverse are defined by

$$\tilde{f}(\omega) = \mathcal{F}(f)(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\exp(i\omega t)\,dt, \tag{2.1}$$

$$f(t) = \mathcal{F}^{-1}\left(\tilde{f}\right)(t) = \int_{-\infty}^{\infty} \tilde{f}(\omega)\exp(-i\omega t)\,d\omega. \tag{2.2}$$

Here, t and ω are nominally considered time and frequency variables, respectively, with no further physical context. For functions of multiple variables, if a Fourier transform or inverse Fourier transform is only applied in one dimension, this will be denoted by specifying the variable as a subscript, e.g., $\mathcal{F}_{\omega}^{-1}$.

The frequency ω is real in the above formulas, but it will be seen to be useful in the mathematical theory that follows to consider extensions of a frequency domain function $\tilde{f}$ to complex frequencies. We define the open upper-half complex frequency (UCF) plane to be all values $\omega_r + i\omega_i$ with $\omega_i > 0$, where ω_r and ω_i are real values, and the closed UCF plane to be the same, except with the condition $\omega_i \geq 0$. An essential result on Fourier method causality is as follows:

Titchmarsh's Theorem (*e.g., [26], Section 1.6c*)**.** *If $\tilde{f}$ is square integrable on the real line (i.e., the subset of the complex frequency plane with $\omega_i = 0$) then the following two statements are equivalent: (i) $f(t) = 0$ for $t < 0$ and (ii) $\tilde{f}$ has an analytic extension to the open UCF plane, $\lim_{\omega_i \to 0} \tilde{f}(\omega_r + i\omega_i) = \tilde{f}(\omega_r)$ for almost all values of ω_r (to be referred to as the* ***continuity property****), and there is a fixed finite value, C, such that for any $\omega_i > 0$,*

$$\int_{-\infty}^{\infty} \left| \tilde{f}(\omega_r + i\omega_i) \right|^2 d\omega_r < C. \tag{2.3}$$

It follows immediately from Parseval's Theorem (e.g., [26], eq. 1.6.2) moreover, that the condition that f is square integrable can be substituted for the condition that $\tilde{f}$ is square integrable in Titchmarsh's Theorem.

We define a *Fourier method* to be a method that, given some function $\beta(t)$, provides a function $w(t)$ that has some desired property relative to β. A *Fourier method function* is a function $\widetilde{\psi}(\omega)$ that defines a Fourier method via the formula

$$w(t) = \mathcal{F}^{-1}\left(\widetilde{\psi}\widetilde{\beta}\right)(t). \tag{2.4}$$

A function $\beta(t)$ is said to have *compact support* if it is zero outside of some finite t interval, and it is said to be *smooth* if all of its derivatives are defined.

Lemma 1 (*Frequency Domain Decay Result*)**.** *If β is smooth and has compact support, and $\beta(t) = 0$ for $t < 0$, then it follows that $\widetilde{\beta}$ has a bounded (in absolute value) analytic extension to the open UCF plane with the continuity property and that for any real $r \geq 0$*

$$\left|\widetilde{\beta}(\omega)\right| = O\left(|\omega|^{-r}\right) \text{ as } \omega \to \infty \ (\mathrm{Im}(\omega) \geq 0). \tag{2.5}$$

Proof. The analytic extension of $\widetilde{\beta}$ to the open UCF plane can be expressed in terms of an integral like (2.1) with an imaginary part added to ω. Integration by parts, as in Theorem 1 of Erdelyi [27], gives (2.5). □

We say that a Fourier method or Fourier method function $\widetilde{\psi}$ *preserves causality* if for any smooth function β with compact support that has the property $\beta(t) = 0$ for $t < 0$, the function w provided by the method also has the property that $w(t) = 0$ for $t < 0$.

Lemma 2 (*Sufficient Conditions for Preserving Causality*)**.** *If $\widetilde{\psi}$ is analytic on the open UCF plane and has the continuity property, and there are real constants C, S and $p \geq 0$ such that*

$$\left|\widetilde{\psi}(\omega)\right| \leq C + S|\omega|^p \quad (\mathrm{Im}(\omega) > 0), \tag{2.6}$$

then $\widetilde{\psi}$ preserves causality.

Proof. Since r in Eq. (2.5) is arbitrary, the decay in $\widetilde{\beta}$ must overwhelm the growth in $\widetilde{\psi}$ as $\omega \to \infty$, so the requirement in Eq. (2.3) of Titchmarsh's Theorem follows. □

Typically, a multidimensional version of (2.4) will be used, with a collection of Fourier method functions $\widetilde{\psi}_{q=1\ldots n}(\omega)$ determining $w(t)$ given $\beta_{q=1\ldots n}(t)$ as follows

$$w(t) = \mathcal{F}^{-1}\left(\sum_{q=1}^{n} \widetilde{\psi}_q \widetilde{\beta}_q\right)(t), \tag{2.7}$$

with the definition of preserving causality being extended in the obvious way.

Lemma 3 (*Multidimensional Causality Conditions*)**.** *If each of the n Fourier method functions $\widetilde{\psi}_{q=1\ldots n}(\omega)$ meets the conditions described in Lemma 2, then the multidimensional Fourier method preserves causality.*

Proof. This follows immediately from Lemma 2. □

3. Method for constant coefficients

Although some of the results described here are straightforward, it proves convenient to present them for later reference. This section gives causality results for Fourier methods that solve a single-function linear homogeneous PDE, of any order, in two variables, z and t, with constant coefficients, where t is still a nominal time variable and z is nominally considered a height variable. This can be written as

$$\sum_{p=0}^{J'}\sum_{q=0}^{J} a_{p,q}\partial_t^p\partial_z^q f\,(z,t)=0, \tag{3.1}$$

where, e.g., ∂_z^q denotes the qth partial derivative with respect to z. A boundary condition of the form

$$\partial_z^{q-1} f\,(z,t)_{z=z_0}=\beta_q\,(t)\quad (q=1\ldots n)\,, \tag{3.2}$$

is to be met, where $n\leq J$. If $n<J$, then additional conditions are needed in order for f to be uniquely determined.

3.1. Boundary value problem

The Fourier transform (2.1) is applied to (3.1) in the time dimension. (Later, when gravity waves are considered, it will also be applied in the horizontal dimensions.) In the frequency domain, the type of PDE just described is equivalent to an ordinary differential equation (ODE) of the form

$$\sum_{q=0}^{J} b_q\,(\omega)\,\partial_z^q\tilde{f}\,(z,\omega)=0, \tag{3.3}$$

where each b_q is a polynomial in ω with constant coefficients given by

$$b_q\,(\omega)=\sum_{p=0}^{J'}(-i)^p\,a_{p,q}\omega^p. \tag{3.4}$$

The boundary condition will be of the form

$$\partial_z^{q-1}\tilde{f}\,(z,\omega)_{z=z_0}=\widetilde{\beta}_q\,(\omega)\quad (q=1\ldots n)\,, \tag{3.5}$$

where $n\leq J$. Define

$$c_{q=0\ldots J}\,(\omega)=i^q b_q\,(\omega)\,, \tag{3.6}$$

and let a polynomial g in a complex variable m with coefficients that are themselves polynomials in ω be given by

$$g\,(m,\omega)=\sum_{q=0}^{J} c_q\,(\omega)\,m^q, \tag{3.7}$$

which will be referred to as the *wavenumber polynomial* associated with the ODE in (3.3). Let the real and/or complex roots of g in m (i.e., for fixed values of ω) be denoted by $m_j\,(\omega)$, with $j=1\ldots J$. It will be assumed that the roots of g in m are distinct for general values of ω.

Let n roots be selected and denote them in row vector form by $\mathbf{m}\,(\omega)=[m_j\,(\omega)]_{j=1\ldots n}$, where it is assumed for notational convenience that the first n roots are selected. (In this paper, all vectors are considered row vectors, which means that matrices must operate on vectors by multiplication from the right.) Similarly, let the boundary condition (3.5) be denoted in row vector form by $\widetilde{\boldsymbol{\beta}}\,(\omega)$. Given some number Λ of values $\boldsymbol{\xi}=(\xi_\lambda)_{\lambda=1\ldots\Lambda}$, let $\mathbf{V}\,(\boldsymbol{\xi})$ denote the square Vandermonde matrix $\left(V_{\lambda,q}\right)_{\lambda=1\ldots\Lambda,\ q=1\ldots\Lambda}$ with $V_{\lambda,q}\equiv\xi_\lambda^{q-1}$ (e.g., [28]).

Lemma 4 (*Solution for Constant Coefficients*)**.** *Let a row vector of n functions of ω be given by*

$$\tilde{\boldsymbol{h}}\equiv\left[\tilde{h}_j\,(\omega)\right]_{j=1\ldots n}\equiv\widetilde{\boldsymbol{\beta}}\,(\omega)\,(\mathbf{V}\,[i\mathbf{m}\,(\omega)])^{-1}\,. \tag{3.8}$$

Then $\tilde{f}\,(z,\omega)$ given by

$$\tilde{f}\,(z,\omega)\equiv\sum_{j=1}^{n}\tilde{h}_j\,(\omega)\exp\left[i\,(z-z_0)\,m_j\,(\omega)\right] \tag{3.9}$$

solves (3.3) and (3.5) and therefore also solves Eqs. (3.1) and (3.2).

Proof. This follows immediately from the definitions. The issue of possible singularities or branch points in the roots is discussed in Sections 3.2 and 3.3. □

Let the exponential terms in (3.9) be denoted in row vector form by

$$\mathbf{e}\left(z,\omega\right)=\left(\exp\left[i\left(z-z_0\right)m_j\left(\omega\right)\right]\right)_{j=1\ldots n} \tag{3.10}$$

and Fourier method functions (see Section 2) be given in row vector form by

$$\left[\widetilde{\psi}_j\left(z,\omega\right)\right]_{j=1\ldots n}^{t}\equiv\widetilde{\boldsymbol{\psi}}\left(z,\omega\right)=\left(\mathbf{V}\left[i\mathbf{m}\left(\omega\right)\right]\right)^{-1}\mathbf{e}^{t}\left(z,\omega\right), \tag{3.11}$$

where the superscript t denotes transposition. Now the solution $\tilde{f}\left(z,\omega\right)$ can be expressed as a dot product:

$$\tilde{f}\left(z,\omega\right)=\widetilde{\boldsymbol{\beta}}\left(\omega\right)\cdot\widetilde{\boldsymbol{\psi}}\left(z,\omega\right). \tag{3.12}$$

Here, the Fourier method function concept has been extended slightly to include dependence on the variable z, but Lemma 3 (multidimensional causality conditions) can still be applied if the assumptions for that lemma hold for any particular z value.

3.2. Root functions

Since the Fourier method functions in Section 3.1 are described in terms of roots $m_{j=1\ldots n}\left(\omega\right)$ of the wavenumber polynomial $g\left(m,\omega\right)$ associated with the ODE in (3.3), the question of whether they are analytic in the open UCF plane is closely related to the question of how the roots, viewed as functions of ω, can fail to be analytic. The relation associating complex ω to roots m of $g\left(m,\omega\right)$ will necessarily be a multiple-valued function [29] in general. Continuous single-valued functions giving roots can only be defined on subsets of the complex plane, and such a single-valued function is sometimes referred to as a *branch* of the multiple-valued function [29].

As described in the section on algebraic functions in Ahlfors [29, Section 8.2], for all but a finite number of complex values of ω, letting complex ω be fixed, g will have J roots in m. We call the finite set of ω values for which there are less than J roots the set of *excluded points*. The discriminant [29] of g is a polynomial $D\left(\omega\right)$ with the property that the complex roots ω' for which $D\left(\omega'\right)=0$ are exactly the values for which there is an m' such that $g\left(m',\omega'\right)=0$ and $\partial_m g\left(m',\omega'\right)=0$: the reader is referred to, e.g., [29] for an explanation of how the discriminant is constructed. A complex ω at which the discriminant is zero is a *branch point* of one or more root functions, meaning that the function must have a jump discontinuity on any small closed curve in the complex plane going around ω. As explained in [29], the excluded points can be described as the combined set of branch points of g and zeros of $c_J\left(\omega\right)$. Let the J root functions that can be defined locally around any non-excluded point be denoted by $\mu_{j=1\ldots J}\left(\omega\right)$.

For a fixed complex value of ω, g will only have nondistinct roots in m if ω is a branch point, and at a branch point ω_0, a root m_0 will only be a multiple root if $\partial_m g\left(m_0,\omega_0\right)=0$. For any such ω_0, let j' be the minimum integer such that $\partial_m^{j'} g\left(m_0,\omega_0\right)\neq 0$. Then there will be exactly j' root functions that approach m_0 as $\omega\rightarrow\omega_0$. Analytic continuation in a small closed path around ω_0 interchanges elements within this set of root functions that approach m_0. In general, the j' root functions having this property will not be analytic in any neighborhood of ω_0 and can only be defined as j' distinct roots if branch cuts are introduced: that is, curves across which there are discontinuities.

Several definitions are needed. First, suppose there is a set of $n\leq J$ root functions, denoted by $\mu_{j=1\ldots n}$ for convenience, defined on a region $\Omega\subset\mathbb{C}$ in which the leading coefficient c_J has no zeros. If the branch points in the region Ω all either do not affect $\mu_{j=1\ldots n}$ or interchange root functions within that set then we say that the branch points *preserve* that set of root functions on Ω. Next, a function of some number of variables is said to be *symmetric* if it is invariant under all permutations of its arguments (i.e., variables). Finally, we apply the term *analytic* to a function of n variables in some region of $\mathbb{C}^n$ if the function is holomorphic in the sense defined by Krantz [30, Section 0.2].

Lemma 5 (*Removal of Branch Points by Symmetry*)**.** *Suppose that a function F of n variables is analytic in a region of $\mathbb{C}^n$ containing the image of some region $\Omega\subset\mathbb{C}$ under the mapping $\left(\mu_1,\ldots,\mu_n\right)$. If the branch points preserve $\mu_{j=1\ldots n}$ on Ω and F is symmetric, then the function defined on Ω by*

$$F\left[\mu_1\left(\omega\right),\ldots,\mu_n\left(\omega\right)\right] \tag{3.13}$$

is analytic as a function of one complex variable, ω. Similarly, if the branch points preserve subsets of the roots and F is invariant under any permutation that preserves the subsets, then (3.13) is, again, analytic as a function of ω.

Proof. Intuitively, it is clear that this result is correct, since the branch points interchange root functions within the set and the symmetry assumption therefore prevents the function defined by (3.13) from being multiple valued at any of the branch points. Details are left to the reader. □

3.3. Causality for PDE solutions

Sufficient conditions will now be given for the boundary value and PDE solution described in Section 3.1 to be causal. Instead of μ (used in Section 3.2), the variable m (with subscripts) will now be used to denote root functions.

Theorem 1 (*Constant-coefficient PDE Causality*). *Let $m_{j=1\ldots n}(\omega)$ be n roots as in Lemma 4 (solution for constant coefficients). Suppose the following conditions hold:*

1. *All branch points for the root functions in the open UCF plane preserve the set of root functions $m_{j=1\ldots n}(\omega)$ in the sense defined in Section 3.2.*
2. *The leading coefficient, $c_J(\omega)$, has no roots in the closed UCF plane.*
3. *For some finite real B_u or B_d, one of the following two inequalities holds for all ω in the open UCF plane and all of the root functions $m_{j=1\ldots n}(\omega)$:*

$$\mathrm{Im}\left[m_j(\omega)\right] > B_u \tag{3.14}$$

$$\mathrm{Im}\left[m_j(\omega)\right] < B_d. \tag{3.15}$$

Then the solution given by Lemma 4 is continuous and preserves causality relative to the boundary condition for $z \geq z_0$ if (3.14) holds and for $z \leq z_0$ if (3.15) holds. Conversely, if condition #1 does not hold, and a branch point in the open UCF plane interchanges one or more of the roots $m_{j=1\ldots n}$ with a root not in that set, then the Fourier method given by Lemma 4 does not preserve causality.

Condition #1 above will be referred to as the *root preservation property*. The conditions described by (3.14) and (3.15) will be referred to as the *upgoing* and *downgoing* properties (respectively) for solution modes, using a wave-propagation analogy that becomes salient physically in Section 5. A proof of Theorem 1 is given in Appendix A. The proof makes use of the results given up to this point, including Lemma 5 (removal of branch points by symmetry).

3.4. Imaginary frequency shifts

It will be seen later that there can be cases of interest in which one or more of the three conditions from Theorem 1 (constant-coefficient PDE causality) are not met. In such cases, it can help to apply an imaginary shift to frequency domain functions so that problematic points are no longer included in the open UCF plane, as will be described in this section. This method will not work for every PDE, but it will be seen in later sections to help with certain gravity wave problems of geophysical relevance.

Given some real value δ, which in general will be positive, and given some frequency domain function $\tilde{f}$, let $\tilde{f}^\delta$ denote the function given by the formula $\tilde{f}^\delta(\omega_r + i\omega_i) \equiv \tilde{f}[\omega_r + i(\omega_i + \delta)]$, where it is assumed that values in the open UCF plane can be obtained by means of analytic continuation from the real line $\mathbb{R}$. If $\tilde{f}$ is a function of multiple variables, one of which corresponds to frequency, then $\tilde{f}^\delta$ is defined similarly, applying the shift to the frequency variable.

If a boundary condition function β is smooth and has compact support, then it follows immediately that $\widetilde{\beta}^\delta$ exists for any δ and can be expressed as

$$\widetilde{\beta}^\delta(\omega) = \widetilde{\beta}(\omega + i\delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \beta(t)\exp\left[i(\omega + i\delta)t\right] dt = \left(\mathcal{F}\left[\exp(-\delta t)\beta(t)\right]\right)(\omega). \tag{3.16}$$

Given a wavenumber polynomial g, it can easily be seen that g^δ also exists for any δ. The notation m_j^δ and $\tilde{f}^\delta$ that appears in the following theorem is not entirely consistent with the definition of frequency shifting given above, since these functions will not be obtained directly as analytic continuations from the real axis, but it will become clear shortly that this use of the notation makes sense:

Theorem 2 (*Causal PDE Solution with Imaginary Frequency Shift*). *Suppose that either the upgoing or downgoing version of the conditions of Theorem 1 are met for the roots $m^\delta_{j=1\ldots n}(\omega)$ of g^δ. Let $\tilde{f}^\delta$ be the solution obtained by applying Eq. (3.12) to $\widetilde{\beta}^\delta$ and $m^\delta_{j=1\ldots n}(\omega)$. Then the Fourier method that provides f via the formula*

$$f(z,t) = \exp(\delta t)\left[\mathcal{F}_\omega^{-1}\left(\tilde{f}^\delta\right)\right](z,t) \tag{3.17}$$

preserves causality and solves the PDE (3.1) and boundary condition (3.2).

Proof. Assume that $\beta_q(t) = 0$ $(t < 0,\ q = 1\ldots n)$ and that $z \geq z_0$ or $z \leq z_0$ depending on whether the upgoing or downgoing condition, respectively, holds. Causality will be established first. Eqs. (2.5) and (2.6) still hold (see Appendix A) for the shifted boundary condition and Fourier method functions, so Lemma 3 implies that

$$\left[\mathcal{F}_\omega^{-1}\left(\tilde{f}^\delta\right)\right](z,t) = 0 \quad \text{for } t < 0, \tag{3.18}$$

Given this, it is clear from (3.17) that $f(z,t) = 0$ for $t < 0$, so the Fourier method preserves causality.

Now it will be shown that (3.17) solves (3.1) and (3.2). Eqs. (3.3) and (3.5) give

$$\sum_{q=0}^{J} b_q\left(\omega + i\delta\right) \partial_z^q \tilde{f}^{\delta}\left(z, \omega\right) = 0, \tag{3.19}$$

$$\partial_z^{q-1} \tilde{f}^{\delta}\left(z, \omega\right)_{z=z_0} = \widetilde{\beta}_q^{\delta}\left(\omega\right) \quad (q = 1 \ldots n)\,. \tag{3.20}$$

Applying the inverse Fourier transform, (3.16), and the definition of f in (3.17) to (3.20) gives

$$\exp\left(-\delta t\right) \partial_z^{q-1} f\left(z, t\right)_{z=z_0} = \exp\left(-\delta t\right) \beta_q\left(t\right) \quad (q = 1 \ldots n)\,, \tag{3.21}$$

so (3.2) holds. Next, substitute the right side of (3.17) into the left side of (3.1), apply the definition of the inverse Fourier transform, combine the exponential terms, and move the partial derivatives and the summation symbols inside the integral:

$$\int_{-\infty}^{\infty} \sum_{q=0}^{J} \sum_{p=0}^{J'} a_{p,q} \partial_t^p \exp\left[\left(-i\omega + \delta\right) t\right] \partial_z^q \tilde{f}^{\delta}\left(z, \omega\right) d\omega. \tag{3.22}$$

Then, noting that $-i\omega + \delta = -i\left(\omega + i\delta\right)$, Eqs. (3.4) and (3.19) give (3.1). □

Because of the $\exp\left(\delta t\right)$ term in (3.17), it cannot be guaranteed that $\tilde{f}$ will be defined for $z \neq z_0$, but assuming that $\tilde{f}$ is defined, it will follow that $\tilde{f}$ solves the corresponding Fourier domain equations, (3.3) and (3.5). The question of under what conditions the Fourier transform can be applied to f as given by (3.17) will not be considered here.

The function given by (3.17) is independent of δ under certain conditions:

Lemma 6 (*Shift-invariance of Solution*)**.** *Suppose that the Fourier method functions given by (3.11) are analytic on the region defined by* $0 < \mathrm{Im}\left(\omega\right) < \delta$*. Then the solution* f *obtained from (3.17) is identical to the one obtained by taking the inverse Fourier transform of (3.12) without an imaginary frequency shift.*

Proof. This follows from Cauchy's Integral Theorem. Details are left to the reader. □

It follows from the reasoning used to prove Theorem 1 (see Appendix A) that to establish the condition stated in Lemma 6 it is sufficient to show that any branch points in the specified region preserve the set of root functions and that the leading coefficient c_J has no zeros in the region.

Another application of imaginary frequency shifting is as follows. It will be seen in Section 5 that condition #2 from Theorem 1 on the leading coefficient $c_J\left(\omega\right)$ does not hold for inviscid and viscous, non-thermally diffusive gravity waves, since $c_J\left(\omega\right)$ will have a root on the real frequency axis. In this case, any shift $\delta > 0$ will allow condition #2 to hold, and then, letting $\delta \to 0$, it is seen that causality holds.

4. Solution method for variable coefficients

The problem to be solved here is similar to the one described in Eqs. (3.1) and (3.2), but the coefficients $a_{p,q}$ are now allowed to vary with z, which will cause coupling between upgoing and downgoing modes, making it necessary to consider modes in both directions simultaneously.

4.1. ODE solution from discrete jump approximation

It will later be seen that it helps to be able to view the solution of a linear homogeneous ODE with z-dependent coefficients as a limit of approximations based on coefficients that are piecewise constant in z with "jump" discontinuities at the interfaces. This piecewise approach has been used to model gravity waves and is known as the multilayer method as discussed in Section 1. In this section and Sections 4.2 and 4.3, ω will be fixed at any particular value and will be omitted from all formulas for convenience, except for the following definition of a new function $\mathfrak{f}$ in terms of $\tilde{f}$:

$$\mathfrak{f}\left(z\right) = \tilde{f}\left(z, \omega\right). \tag{4.1}$$

The ODE to be solved is of the form

$$\sum_{q=0}^{J} b_q\left(z\right) \partial_z^q \mathfrak{f}\left(z\right) = 0, \tag{4.2}$$

Adapting the notation from (3.6), let

$$c_q\left(z\right) = i^q b_q\left(z\right) \tag{4.3}$$

and let a polynomial in m with coefficients depending on z be given by

$$g(m,z) = \sum_{q=0}^{J} c_q(z)\, m^q. \tag{4.4}$$

The value of $\mathfrak{f}$ and its first $J-1$ derivatives will be represented in vector form by $\mathfrak{f} \equiv \left(\mathfrak{f}_q\right)_{q=1\ldots J} \equiv \partial_z^{q-1}\mathfrak{f}$ $(q = 1\ldots J)$. It will be assumed that the roots are distinct for each z, so $\mathfrak{f}$ determines J components denoted in vector form by $\mathbf{h} \equiv \left(h_j\right)_{j=1\ldots J}$ through the relation

$$\mathbf{h} = \mathfrak{f}\, \mathbf{V}^{-1}(i\mathbf{m}), \tag{4.5}$$

as in Eq. (3.8).

For the moment, assume that the coefficients $b_{q=0\ldots J}(z)$ are constant on a sequence of layers $\left(z_p, z_{p+1}\right)$ $(p = 0\ldots P-1)$ and have jump discontinuities at the layer interfaces $z_{p=1\ldots P-1}$. For $p = 0\ldots P-1$, let the root functions of $g(m,z)$ in the layer $\left(z_p, z_{p+1}\right)$ be denoted by $m_{j,p}$, with $j = 1\ldots J$. It will be assumed that the roots are distinct for each interval, so any solution on an interval $\left(z_p, z_{p+1}\right)$ must be of the form

$$\mathfrak{f}\left(z_p < z < z_{p+1}\right) = \sum_{j=1}^{J} h_{j,p} \exp\left[i m_{j,p}\left(z - z_p\right)\right], \tag{4.6}$$

where $h_{j,p}$ are constants. Suppose a boundary condition of the form

$$\partial_z^{q-1}\mathfrak{f}(z)_{z=z_0} = \tilde{\beta}_q \quad (q = 1\ldots J) \tag{4.7}$$

is provided, analogous to Eq. (3.5), except that here we are omitting ω for now. Then Lemma 4 (solution for constant coefficients) implies that (4.7) uniquely determines $h_{j,0}$ $(j = 1\ldots J)$ in (4.6). For each successive $p = 1\ldots P-1$, let $h_{j,p}$ $(j = 1\ldots J)$ be determined from $h_{j,p-1}$ $(j = 1\ldots J)$ by a *smoothness condition*, which is that $\mathfrak{f}$ and its first $J-1$ derivatives are continuous at z_p, i.e.,

$$\partial_z^{q-1}\mathfrak{f}(z)_{z=z_p^+} = \partial_z^{q-1}\mathfrak{f}(z)_{z=z_p^-} \quad (q = 1\ldots J), \tag{4.8}$$

where the $+$ and $-$ superscripts for z_p mean that the derivative is evaluated by taking the limit from above and below, respectively.

Now let $b_q(z)$ $(q = 0\ldots J)$ be continuous functions, and assume that the corresponding polynomial $g(m,z)$ has J distinct roots for every z on an interval $[z_b, z_a]$, and that a boundary condition as in (4.7) has been specified. Choose some P, as well as an increasing set of values $z_{p=0\ldots P}$ with $z_0 = z_b$ and $z_P = z_a$. For each $p = 0\ldots P-1$, choose some intermediate point $\check{z}_p$ with $z_p < \check{z}_p < z_{p+1}$. Define $\check{b}_{q=0\ldots J}(z)$ and $\check{c}_{q=0\ldots J}(z)$ as the functions that are constant on each interval $\left(z_p, z_{p+1}\right)$ and equal to $b_{q=0\ldots J}$ and $c_{q=0\ldots J}$, respectively, at each $\check{z}_p$. ODEs and polynomials corresponding to approximate coefficients such as these will be referred to as *piecewise* ODEs and polynomials, respectively, and the ODE given in (4.2) will be referred to as the *original* ODE.

Lemma 7 (*Multilayer Limit for Single ODE*). *Let $\check{\mathfrak{f}}(z)$ be determined from Eqs. (4.7) and (4.8) as described above for Eq. (4.6), but using the piecewise coefficient functions $\check{b}_{q=0\ldots J}(z)$. Then in the limit as $P \to \infty$ and the interval widths approach zero, $\check{\mathfrak{f}}$ approaches a solution to Eqs. (4.2) and (4.7) for the continuous coefficients $b_{q=0\ldots J}(z)$.*

Proof. Convert to a linear homogeneous system of J first order ODEs, using the standard procedure in which the single function $\mathfrak{f}$ is replaced with the vector $\mathfrak{f}$ of J functions, defined as before. The first $J-1$ equations of the system of ODEs are of the form $\partial_z \mathfrak{f}_q = \mathfrak{f}_{q-1}$ $(q = 2\ldots J)$, and the Jth equation is given by (4.2). The result then follows from Lemma 8, to be given shortly. □

We only consider systems of first order ODEs briefly for the purpose of indicating how the proof of Lemma 7 can be finished. Let a system of first order ODEs be written as

$$\partial_z \mathfrak{g} = \mathfrak{g}\, \mathbf{B}(z), \tag{4.9}$$

where $\mathfrak{g}$ denotes J functions in row vector form and $\mathbf{B}$ denotes a matrix of ODE coefficients, which are continuous functions of z. Let a boundary condition similar to (4.7) also be specified. As before, define a corresponding piecewise system of ODEs with a piecewise constant approximation $\check{\mathbf{B}}(z)$ of $\mathbf{B}(z)$. The smoothness condition is now that each of the J functions of $\mathfrak{g}$ is continuous at each z_p.

Lemma 8 (*Multilayer Limit for System of First Order ODEs*). *Letting $P \to \infty$ and the layer widths go to zero, solutions $\check{\mathfrak{g}}$ of the piecewise system of ODEs approaches a solution of (4.9).*

We will not give a detailed proof of Lemma 8 here, although the proof is not difficult. In the context of atmospheric gravity waves, a proof was first given for a special case by Pierce [31]. Klostermeyer [10] then gave a more general proof, based mainly on cited derivations in [32]. Later, an explicit proof was given in Klostermeyer [11, Appendix]. The proof given in [11] follows from simply noting that the limit as interval widths go to zero of the solution to the piecewise ODE equals the limit as the step size goes to zero of Euler's method applied to (4.9), and thus equals the solution of the ODE (since the limit of Euler's method is known to give the correct solution).

Early multilayer solution methods for atmospheric gravity waves [5,6,10,11] use a matrix of coefficients, as in (4.9), where each element of $\mathbf{f}$ is a state variable and each row of the matrix corresponds to a governing equation. By contrast, here we solve a higher order ODE for just one function. As the above discussion illustrates, the two approaches are mathematically interchangeable, but it is convenient for us to work with a single higher order ODE because later we directly obtain vertical structure equations for atmospheric gravity waves from dispersion relations.

The critique of multilayer techniques made by Hines [25] focuses on whether the equations describing the state variables within a layer are physically realistic. That concern is not relevant here, since it is sufficient in this purely mathematical context that, in the infinitesimally thin layer limit, the method converges to a correct solution of the ODE. The coefficients of the ODE are regarded as functions varying with z, and the question of what they mean, physically, is immaterial. Hines [25] refers to this purely mathematical multilayer solution technique as *nonstandard* and cites [10] as the first instance of such an approach. Hines [25] questions the practical usefulness of the nonstandard approach, suggesting that it would be more efficient to solve an equation such as (4.9) (c.f. [25] eq. (1)) using an iterative numerical integration procedure. We do not take such an approach here, however, and find that the piecewise constant coefficient approximation provides a conceptual framework in which the radiation condition (to be defined in Section 4.2) can readily be met (see Section 4.3) and in which causality can also be established and preserved (see Section 4.4).

Given $J+1$ ODE coefficient functions, $b_{q=0\ldots J}$, any component vector $\mathbf{h}(z)$ at one point z uniquely determines the component vector $\mathbf{h}(z')$ at any other point, z'. This linear relationship can be described in terms of what we call the *transfer matrix* $\mathbf{H}(z,z')$ as follows:

$$\mathbf{h}(z') = \mathbf{h}(z)\,\mathbf{H}(z,z')\,. \tag{4.10}$$

This is similar to the transfer matrix defined in Pérez-Álvarez and García-Moliner [23, Section 2.1]. It is also similar, but not identical, to the transfer matrix defined in [11]. Multiplication of transfer matrices for adjacent intervals clearly gives the transfer matrix for the combined interval, as follows:

$$\mathbf{H}(z,z'') = \mathbf{H}(z,z')\;\mathbf{H}(z',z'')\,. \tag{4.11}$$

Reversing the arguments of a transfer matrix yields the inverse relation:

$$\mathbf{H}(z',z) = \mathbf{H}^{-1}(z,z')\,. \tag{4.12}$$

The transfer matrices can be defined for a piecewise ODE, using the smoothness condition at the jump discontinuities. For each p, let $\mathbf{E}_p(z)$ denote the diagonal matrix with diagonal elements $\exp\left(izm_{j,p}\right)$ $(j=1\ldots J)$. Also, let $\mathbf{m}_p$ denote the vector with elements $\left(m_{j,p}\right)_{j=1\ldots J}$. The piecewise ODE transfer matrix for two adjacent intermediate points, $\check{z}_p$, $\check{z}_{p+1}$, is given by

$$\mathbf{H}\left(\check{z}_p,\check{z}_{p+1}\right) = \mathbf{E}_p\left(z_{p+1}-\check{z}_p\right)\mathbf{V}\left(i\mathbf{m}_p\right)\mathbf{V}^{-1}\left(i\mathbf{m}_{p+1}\right)\mathbf{E}_{p+1}\left(\check{z}_{p+1}-z_{p+1}\right). \tag{4.13}$$

This will be called an *elementary piecewise transfer matrix*. Lemma 7 implies that a transfer matrix for the piecewise ODE is an approximation of the transfer matrix for the original ODE. To reduce ambiguity, we could put a check accent over transfer matrices for a piecewise ODE, but this would complicate the notation, and it should be clear from the context to follow which type of ODE is meant. Two additional elementary piecewise transfer matrices are needed (in order to cover the whole altitude range from z_0 to z_P), which are

$$\mathbf{H}\left(z_0,\check{z}_0\right) = \mathbf{E}_0\left(\check{z}_0-z_0\right);\quad \mathbf{H}\left(\check{z}_{P-1},z_P\right) = \mathbf{E}_{P-1}\left(z_P-\check{z}_{P-1}\right). \tag{4.14}$$

Now one can obtain piecewise ODE transfer matrices $\mathbf{H}\left(z_0,\check{z}_p\right)$ and $\mathbf{H}\left(\check{z}_p,z_p\right)$ for any p by applying the composition rule in Eqs. (4.11) to (4.13) and (4.14) for adjacent intermediate points.

4.2. Two-altitude boundary condition

We are interested in the situation in which upgoing and downgoing components are specified as boundary conditions at z_b and z_a, the altitudes bounding the altitude range being modeled at the bottom and top of the range, respectively, where b and a are abbreviations for "below" and "above". In order to state the two-altitude boundary condition, it suffices to require that $J=2n$ and that sets of roots $[m_j(z)]_{j=1\ldots n}$ and $[m_j(z)]_{j=n+1\ldots 2n}$ have been designated "upgoing" and "downgoing" (respectively) at $z=z_b$ and $z=z_a$. Later, in Section 4.4, it will become important to characterize the upgoing and downgoing roots in terms of modified versions of Eqs. (3.14) and (3.15), but in this section it suffices to treat them merely as two disjoint subsets of n roots each.

The two-altitude boundary condition is given by the following two equations:

$$\partial_z^{q-1}(h_1 + \cdots + h_n)_{z=z_b} = \widetilde{\beta}_q \ \ (q = 1 \ldots n)\,, \tag{4.15}$$

$$h_j(z_a) = 0 \ \ (j = n+1 \ \ldots \ 2n)\,. \tag{4.16}$$

Eq. (4.15) is referred to as the *lower-boundary condition*, and Eq. (4.16) referred to as the *upper-boundary radiation condition*. Let u and d subscripts for a vector indexed by root (e.g., $\mathbf{m}$ and $\mathbf{h}$) indicate a vector in the subspace corresponding to either upgoing or downgoing roots, respectively. Eqs. (4.15) and (4.16), respectively, may now be stated as

$$\mathbf{h}_u\,(z_b) = \widetilde{\boldsymbol{\beta}}\ \mathbf{V}^{-1}\,(i\mathbf{m}_u)\,, \tag{4.17}$$

$$\mathbf{h}_d\,(z_a) = \mathbf{0}, \tag{4.18}$$

where $\mathbf{0}$ is the zero vector.

Let $\mathbf{S}_u$ and $\mathbf{S}_d$ be defined as the surjective (referring to a function that maps from one space onto another space) linear maps going from the vector space with basis elements indexed by all roots (in the sense described above) to the vector space spanned by only the first n and last n basis elements, respectively. In other words, these maps send $(\xi_1, \ldots, \xi_{2n})$ to $(\xi_1, \ldots, \xi_n)$ and $(\xi_{n+1}, \ldots, \xi_{2n})$, respectively, going from J-dimensional space to n-dimensional space. Similarly, let $\mathbf{I}_u$ and $\mathbf{I}_d$ be the injective (referring to a one-to-one function) linear maps embedding the n-dimensional vector spaces with basis elements indexed by upgoing and downgoing roots (respectively) into the full J–dimensional space, mapping the vector $(\xi_1, \ldots, \xi_n)$ to $(\xi_1, \ldots, \xi_n, 0, \ldots, 0)$ and $(0, \ldots, 0, \xi_1, \ldots, \xi_n)$, respectively.

Lemma 9 (*Solution from Transfer Matrices*). *If* $\mathbf{I}_d\ \mathbf{H}\,(z_b, z_a)\ \mathbf{S}_d$ *is nonsingular then the solution to (4.2) and the two-altitude boundary condition specified by (4.17) and (4.18) can be expressed in terms of transfer matrices by the formulas*

$$\mathbf{h}_d\,(z_b) = -\mathbf{h}_u\,(z_b)\,\mathbf{I}_u\ \mathbf{H}\,(z_b, z_a)\,\mathbf{S}_d\,[\mathbf{I}_d\ \mathbf{H}\,(z_b, z_a)\,\mathbf{S}_d]^{-1}\,, \tag{4.19}$$

$$\mathbf{h}\,(z) = (\mathbf{h}_u\,(z_b)\,\mathbf{I}_u + \mathbf{h}_d\,(z_b)\,\mathbf{I}_d)\ \ \mathbf{H}\,(z_b, z) \quad (z_b < z \le z_a). \tag{4.20}$$

Proof. It follows immediately from the definitions that Eq. (4.20) gives a solution to (4.2) and the lower boundary condition (4.17). Eq. (4.18) is then verified by plugging (4.19) into (4.20), setting $z = z_a$, multiplying both sides from the right by $\mathbf{S}_d$, and applying matrix algebra. □

Eqs. (4.19) and (4.20), combined with the piecewise approximation of the transfer matrices described at the end of Section 4.1, provide a numerical solution method for (4.2) and the two-altitude boundary condition that is accurate in the limit as layer thickness goes to zero. Care must be taken in using this approach, however, because of the problem of *numerical swamping* [14,33,34], in which certain terms grow rapidly and overwhelm other terms, rendering the resulting solution meaningless. For models of gravity waves damped by viscosity, numerical swamping is most serious when the kinematic viscosity ν is near zero, since the roots of $g\,(m, z)$ corresponding to viscosity modes have imaginary parts with leading asymptotic terms proportional to $\nu^{-1/2}$: see, e.g., [13, eq. 20b]. In fact, terms can grow so rapidly that they quickly exceed the maximum allowed double precision value on a computer. This issue makes it challenging to derive accurate solutions for gravity waves propagating from a region where viscous and thermal damping can be neglected into a region where they becomes significant.

Various methods have been used to reduce numerical swamping of viscous gravity-wave solutions. Midgley and Liemohn [5] use an iterative method that is described as Gauss–Seidel group iteration in [11]. One drawback of their approach is that it requires just one dominant mode in each direction, which, as pointed out by Volland [6], is not a valid assumption for viscous and thermally diffusive gravity wave solutions at or above approximately 200 km altitude. Volland [6] works around this problem using a three region approach, with abrupt changes in the physical assumptions at the interfaces between the regions and an assumption that certain background parameters are constant in two of the regions. The method we describe in Section 4.3 does not rely on abrupt changes like this. One thing in common between our approach to follow and [6] is the use of scattering matrices, but unlike us, Volland [6] uses just one stand-alone scattering matrix instead of a sequence of composed scattering matrices. Klostermeyer [11, Section 4] uses what appears to be a fundamentally different method based on transfer matrices, which we do not address here.

4.3. Scattering matrix method

In addition to numerical swamping, a second problem with the transfer matrix approach to solving for a two-altitude boundary condition is that modes that are causal in upgoing and downgoing directions are combined and extended in the same direction together, making it difficult to establish that the overall solution responds in a causal way to the forcing terms at the lower boundary. Both problems can be avoided if upgoing and downgoing modes are only extended upwards and downwards, respectively, which is the motivation behind the scattering matrix method described in this section.

A scattering matrix specifies the relationship between solution components at any two different heights within the modeled range $[z_b, z_a]$ in terms of reflection and transmission submatrices, which are defined in such a way that damping and not growth occurs in the transformations associated with the submatrices. The statement of this approach requires that n upgoing and n downgoing roots be designated at *every* altitude within the modeled range. Extending the notation from Section 4.2, $[m_j(z)]_{j=1\ldots n}$ and $[m_j(z)]_{j=n+1\ldots 2n}$ are to be referred to as the "upgoing" and "downgoing" roots for every z with $z_b \leq z \leq z_a$ now. The need for the conditions on upgoing and downgoing roots given by Eqs. (3.14) and (3.15), respectively will not be discussed until Section 4.4.

Given a pair of altitudes z and z', with $z < z'$, the scattering matrix associated with an ODE described in Section 4.1 (either original or piecewise) contains four submatrices, denoted by $\mathbf{T}_u$, $\mathbf{R}_u$, $\mathbf{T}_d$, and $\mathbf{R}_d$. Here, $\mathbf{T}$ and $\mathbf{R}$ stand for transmission and reflection, respectively, and the u and d subscripts stand for upgoing and downgoing, respectively, as before. The four $n \times n$ submatrices combined form a $J \times J$ matrix as in [23, eq. (2.66)], but here we focus on the submatrices rather than the combined matrix. The four submatrices describe the response to incident components at either z or z', where the incident components are upgoing at z (with the u subscript) and downgoing at z' (with the d subscript). The four submatrices are implicitly defined by the following two equations:

$$\left[\mathbf{I}_u + \mathbf{R}_u\left(z, z'\right)\mathbf{I}_d\right]\mathbf{H}\left(z, z'\right) = \mathbf{T}_u\left(z, z'\right)\mathbf{I}_u, \tag{4.21}$$

$$\left[\mathbf{I}_d + \mathbf{R}_d\left(z, z'\right)\mathbf{I}_u\right]\mathbf{H}\left(z', z\right) = \mathbf{T}_d\left(z, z'\right)\mathbf{I}_d. \tag{4.22}$$

Eq. (4.21) means that when the transfer matrix $\mathbf{H}\left(z, z'\right)$ is applied to the component vector $\mathbf{h}(z) = [\mathbf{h}_u(z), \mathbf{h}_u(z)\,\mathbf{R}_u\left(z, z'\right)]$, the resulting component vector at z' has upgoing components equal to $\mathbf{h}_u\mathbf{T}_u\left(z, z'\right)$ and downgoing components equal to zero: a similar interpretation applies to (4.22). While the arguments for the transfer matrix (i.e., the z and z' values) can be reversed, we will keep the arguments for the transmission and reflection matrices in increasing order.

The following formulas,

$$\mathbf{S}_u\,\mathbf{I}_u + \mathbf{S}_d\,\mathbf{I}_d = \mathbf{I}(J), \tag{4.23}$$

$$\mathbf{I}_u\,\mathbf{S}_u = \mathbf{I}(n); \quad \mathbf{I}_d\,\mathbf{S}_d = \mathbf{I}(n), \tag{4.24}$$

$$\mathbf{I}_u\,\mathbf{S}_d = \mathbf{0}; \quad \mathbf{I}_d\,\mathbf{S}_u = \mathbf{0}, \tag{4.25}$$

give elementary properties of the surjective and injective linear maps defined in Section 4.2, where, e.g., $\mathbf{I}(j)$ denotes a $j \times j$ identity matrix.

Explicit formulas for the reflection submatrices are easily obtained by multiplying both sides of Eqs. (4.21) and (4.22) from the right by $\mathbf{S}_d$ and $\mathbf{S}_u$, respectively:

$$\mathbf{R}_u\left(z, z'\right) = -\left[\mathbf{I}_u\,\mathbf{H}\left(z, z'\right)\mathbf{S}_d\right]\left[\mathbf{I}_d\,\mathbf{H}\left(z, z'\right)\mathbf{S}_d\right]^{-1}, \tag{4.26}$$

$$\mathbf{R}_d\left(z, z'\right) = -\left[\mathbf{I}_d\,\mathbf{H}\left(z', z\right)\mathbf{S}_u\right]\left[\mathbf{I}_u\,\mathbf{H}\left(z', z\right)\mathbf{S}_u\right]^{-1}. \tag{4.27}$$

If the two inverted terms in the above two equations are nonsingular, then the reflection submatrices will be said to be *well-defined*. Now, explicit formulas for the transmission submatrices may be obtained in terms of the formulas for the reflection submatrices by multiplying both sides of Eqs. (4.21) and (4.22) on the right by $\mathbf{S}_u$ and $\mathbf{S}_d$, respectively:

$$\mathbf{T}_u\left(z, z'\right) = \mathbf{I}_u\,\mathbf{H}\left(z, z'\right)\mathbf{S}_u + \mathbf{R}_u\left(z, z'\right)\mathbf{I}_d\,\mathbf{H}\left(z, z'\right)\mathbf{S}_u, \tag{4.28}$$

$$\mathbf{T}_d\left(z, z'\right) = \mathbf{I}_d\,\mathbf{H}\left(z', z\right)\mathbf{S}_d + \mathbf{R}_d\left(z, z'\right)\mathbf{I}_u\,\mathbf{H}\left(z', z\right)\mathbf{S}_d. \tag{4.29}$$

It can readily be shown that Eqs. (4.26), (4.27), (4.28), and (4.29) also imply Eqs. (4.21) and (4.22).

Lemma 10 (*Scattering Matrix Composition*)**.** *Given three heights, $z < z' < z''$, suppose that (i) the reflection matrices given by (4.26) and (4.27) are well-defined for the pairs (z, z') and (z', z''), (ii) the four transmission submatrices (i.e., two for each of the two pairs of heights just specified) are all nonsingular, and (iii) the matrices $\mathbf{I}(n) - \mathbf{R}_u\left(z', z''\right)\mathbf{R}_d\left(z, z'\right)$ and $\mathbf{I}(n) - \mathbf{R}_d\left(z, z'\right)\mathbf{R}_u\left(z', z''\right)$ are nonsingular. Then it follows that $\mathbf{R}_u\left(z, z''\right)$ and $\mathbf{R}_d\left(z, z''\right)$ are well-defined and $\mathbf{T}_u\left(z, z''\right)$, $\mathbf{T}_d\left(z, z''\right)$, and $\mathbf{H}\left(z, z''\right)$ are nonsingular, with*

$$\mathbf{R}_u\left(z, z''\right) = \mathbf{R}_u\left(z, z'\right) + \mathbf{T}_u\left(z, z'\right)\left[\mathbf{I}(n) - \mathbf{R}_u\left(z', z''\right)\mathbf{R}_d\left(z, z'\right)\right]^{-1}\mathbf{R}_u\left(z', z''\right)\mathbf{T}_d\left(z, z'\right) \tag{4.30}$$

$$\mathbf{R}_d\left(z, z''\right) = \mathbf{R}_d\left(z', z''\right) + \mathbf{T}_d\left(z', z''\right)\left[\mathbf{I}(n) - \mathbf{R}_d\left(z, z'\right)\mathbf{R}_u\left(z', z''\right)\right]^{-1}\mathbf{R}_d\left(z, z'\right)\mathbf{T}_u\left(z', z''\right) \tag{4.31}$$

$$\mathbf{T}_u\left(z,z''\right)=\mathbf{T}_u\left(z,z'\right)\ \left[\mathbf{I}\left(n\right)-\mathbf{R}_u\left(z',z''\right)\mathbf{R}_d\left(z,z'\right)\right]^{-1}\mathbf{T}_u\left(z',z''\right) \tag{4.32}$$

$$\mathbf{T}_d\left(z,z''\right)=\mathbf{T}_d\left(z',z''\right)\ \left[\mathbf{I}\left(n\right)-\mathbf{R}_d\left(z,z'\right)\mathbf{R}_u\left(z',z''\right)\right]^{-1}\mathbf{T}_d\left(z,z'\right). \tag{4.33}$$

Indication of proof. Our proof, the details of which are omitted for brevity, involves first expressing $\mathbf{H}\left(z,z'\right)$ and $\mathbf{H}\left(z',z''\right)$ in terms of the scattering submatrices (including inverse transmission matrices) for the two subintervals, using the injective and surjective linear maps, then applying the transfer matrix composition formula (4.11) to obtain a formula for $\mathbf{H}\left(z,z''\right)$, and finally applying the scattering submatrix definitions and Eqs. (4.23), (4.24), and (4.25). □

The equations stated in Lemma 10 are equivalent to Pérez-Álvarez and García-Moliner [23, eq. 2.72]. They are also closely related to certain equations from the Kennett algorithm in seismology [24, eqs. 7.2.56-57]. The matrix $\left[\mathbf{I}\left(n\right)-\mathbf{R}_u\left(z',z''\right)\mathbf{R}_d\left(z,z'\right)\right]^{-1}$ and the other similar matrix with terms reversed appearing in Lemma 10 will be referred to as *reverberation matrices*, following [24], and the matrices that are being inverted, e.g., $\mathbf{I}\left(n\right)-\mathbf{R}_u\left(z',z''\right)\mathbf{R}_d\left(z,z'\right)$ will be referred to as the *inverse reverberation matrices*.

The following lemma will be seen in the theorem that follows it to make it possible to describe components at z' in terms of the scattering submatrices for the two subintervals, given that either the downgoing components at z'' or the upgoing components at z are zero:

Lemma 11 (*Internal Components*)**.** *If all of the assumptions of Lemma 10 hold then*

$$\mathbf{T}_u\left(z,z''\right)\ \mathbf{I}_u\ \mathbf{H}\left(z'',z'\right)=\mathbf{T}_u\left(z,z'\right)\ \left[\mathbf{I}\left(n\right)-\mathbf{R}_u\left(z',z''\right)\mathbf{R}_d\left(z,z'\right)\right]^{-1}\ \left[\mathbf{I}_u+\mathbf{R}_u\left(z',z''\right)\mathbf{I}_d\right] \tag{4.34}$$

$$\mathbf{T}_d\left(z,z''\right)\ \mathbf{I}_d\ \mathbf{H}\left(z,z'\right)=\mathbf{T}_d\left(z',z''\right)\ \left[\mathbf{I}\left(n\right)-\mathbf{R}_d\left(z,z'\right)\mathbf{R}_u\left(z',z''\right)\right]^{-1}\ \left[\mathbf{I}_d+\mathbf{R}_d\left(z,z'\right)\mathbf{I}_u\right]. \tag{4.35}$$

Indication of proof. The proof, which is omitted for brevity, uses the same general concepts and techniques as the one indicated for Lemma 10. □

Eqs. (4.34) and (4.35) are closely related to Chapman [24, eq. 7.2.70], which describes seismic waves reaching a receiver from a source, although [24, eq. 7.2.70] includes extra terms because the source is considered to be within the medium, while our boundary condition is on the edge of the medium. Also, in comparing with the theory in [24] it should be kept in mind that our formulas are for row vectors, while [24] uses column vectors.

Theorem 3 (*Solution from Scattering Submatrices*)**.** *(a) Suppose that $\mathbf{R}_u\left(z_b,z_a\right)$ is well-defined. Then the assumption of Lemma 9 (solution from transfer matrices) holds, and Eq. (4.19) from Lemma 9 is equivalent to*

$$\mathbf{h}_d\left(z_b\right)=\mathbf{h}_u\left(z_b\right)\mathbf{R}_u\left(z_b,z_a\right). \tag{4.36}$$

(b) Suppose that the assumptions of Lemma 10 hold, substituting z_b for z, z for z', and z_a for z''. Then the assumption of Lemma 9 holds (once again), and Eq. (4.20) from Lemma 9 is equivalent to

$$\mathbf{h}\left(z\right)=\mathbf{h}_u\left(z_b\right)\mathbf{T}_u\left(z_b,z\right)\ \left[\mathbf{I}\left(n\right)-\mathbf{R}_u\left(z,z_a\right)\mathbf{R}_d\left(z_b,z\right)\right]^{-1}\ \left[\mathbf{I}_u+\mathbf{R}_u\left(z,z_a\right)\mathbf{I}_d\right], \tag{4.37}$$

which at $z=z_a$ becomes

$$\mathbf{h}\left(z_a\right)=\mathbf{h}_u\left(z_b\right)\mathbf{T}_u\left(z_b,z_a\right)\ \mathbf{I}_u. \tag{4.38}$$

Indication of roof. For part (a), the equivalence follows immediately from Eq. (4.26). Part (b) follows readily from the results given up to this point, in particular Eq. (4.34), and further details are left to the reader. □

As just indicated, (4.37) follows from (4.34). A complementary result, following from (4.35), but not explicitly given here, can be obtained for the reverse situation of the two-altitude boundary condition problem, in which a nonzero downgoing boundary condition is specified at z_a and the upgoing components are assumed to be zero at z_b.

A numerical solution method based on scattering matrices for the two-altitude boundary condition problem can now be obtained by applying Eqs. (4.30), (4.31), (4.32), (4.33), (4.36), and (4.37) to elementary piecewise scattering submatrices obtained from the elementary piecewise transfer matrices given by Eqs. (4.13) and (4.14). The question of whether the conditions needed for Theorem 3 can always be met in the numerical solution method just described is left for Section 4.4.

This scattering matrix method is mathematically consistent with the transfer matrix approach described in Section 4.2 but eliminates the problem of numerical swamping described earlier. While the transfer matrices combine both upgoing and downgoing terms, the transmission and reflection matrices are defined in such a way that they only include terms that are specific to a particular direction.

The method just described does not allow for a reflective lower boundary, so downgoing modes that cross the lower boundary have no further effect. This can be understood as a radiation condition at the lower boundary. It would be relatively straightforward to modify the method to include reflection at the lower boundary.

We now mention a simpler but less accurate alternative to method described above, in which an approximate solution of the two-altitude boundary condition problem is obtained from elementary piecewise transmission matrices as follows:

$$\mathbf{h}\left(\check{z}_p\right) \approx \mathbf{h}_u\left(z_0\right) \mathbf{T}_u\left(z_0, \check{z}_0\right) \mathbf{T}_u\left(\check{z}_0, \check{z}_1\right) \cdots \mathbf{T}_u\left(\check{z}_{p-1}, \check{z}_p\right) \mathbf{I}_u. \tag{4.39}$$

Since the downgoing components are omitted, this approximation cannot converge to a solution of (4.2) in the continuous limit unless the ODE coefficients are constant. If $J = 2$ it converges in the continuous limit to the WKB approximation [35, e.g.] applied to (4.2) and (4.15). A proof of a similar result can be found in Berry and Mount [36, Section 2]. The approximation given by (4.39) will be referred to as the *transmission-only* approximation.

4.4. Causality with variable coefficients

Now, ω dependence is introduced into the terms and results of Sections 4.1, 4.2, and 4.3. A PDE is to be solved, which is like (3.1) except that now z dependence is included in the coefficients:

$$\sum_{p=0}^{J'} \sum_{q=0}^{J} a_{p,q}\left(z\right) \partial_t^p \partial_z^q f\left(z, t\right) = 0. \tag{4.40}$$

The two-altitude boundary condition is still used, now with ω dependence given to $\widetilde{\boldsymbol{\beta}}$. The function $\mathfrak{f}\left(z\right)$ defined in Section 4.1 is replaced with $\tilde{f}\left(z, \omega\right)$, giving an ODE that is like (3.3), except now with z dependence in the coefficients:

$$\sum_{q=0}^{J} b_q\left(z, \omega\right) \partial_z^q \tilde{f}\left(z, \omega\right) = 0. \tag{4.41}$$

The wavenumber polynomial, $g\left(m, \omega, z\right)$, is defined as before, but now having coefficients that vary with z. The solution to (4.40) and the two-altitude boundary condition is obtained as the inverse Fourier transform of

$$\tilde{f}(z, \omega) = \sum_{j=1}^{J} h_j(z, \omega), \tag{4.42}$$

where $\left[h_j(z, \omega)\right]_{j=1\ldots J}$ are provided by Theorem 3.

The reverberation matrices can be interpreted as formal matrix power series:

$$\left[\mathbf{I}\left(n\right) - \mathbf{R}_u\left(z', z''\right) \mathbf{R}_d\left(z, z'\right)\right]^{-1} = \sum_{j=0}^{\infty} \left[\mathbf{R}_u\left(z', z''\right) \mathbf{R}_d\left(z, z'\right)\right]^j, \tag{4.43}$$

$$\left[\mathbf{I}\left(n\right) - \mathbf{R}_d\left(z, z'\right) \mathbf{R}_u\left(z', z''\right)\right]^{-1} = \sum_{j=0}^{\infty} \left[\mathbf{R}_d\left(z, z'\right) \mathbf{R}_u\left(z', z''\right)\right]^j. \tag{4.44}$$

If the power series are convergent then it follows that the inverse reverberation matrices are nonsingular.

Lemma 12 (*Causality for One Jump Discontinuity*)**.** *Suppose that the PDE coefficients in (4.40) are constant as functions of z except for one z value at which the coefficients change. Suppose that the polynomials $g(m, \omega, z)$ have distinct roots in m for every real ω, both above and below the jump discontinuity. Suppose that there are values B_u and B_d, with*

$$\mathrm{Im}(m_{j=1\ldots n}) > B_u > B_d > \mathrm{Im}(m_{j=n+1\ldots 2n}), \tag{4.45}$$

such that the conditions of Theorem 1, substituting $(m_j)_{j=n+1\ldots 2n}$ for the downgoing roots, hold for both the upgoing and downgoing roots for all ω and both above and below the jump discontinuity. Then the reflection matrices are well-defined and the transmission matrices are nonsingular for the interval (z_b, z_a) and the solution to the two-altitude boundary condition problem for this PDE preserves causality.

The proof is given in Appendix B.

Theorem 4 (*Piecewise Constant PDE Causality*)**.** *Given a PDE as in (4.40) with coefficients that are piecewise constant in z as in Section 4.1, suppose the following conditions all hold:*

1. *There are fixed real values B_u and B_d satisfying (4.45) such that the conditions of Theorem 1, substituting $(m_j)_{j=n+1\ldots 2n}$ for the downgoing roots, hold for those values for the upgoing and downgoing roots, respectively, for all ω and all z with $z_b \le z \le z_a$.*

2. *The polynomial* $g(m, \omega, z)$ *has distinct roots in* m *for every* ω *and* z.
3. *The eigenvalues of* $\mathbf{R}_u\left(\check{z}_p, \check{z}_{p+1}\right)\mathbf{R}_d\left(z_0, \check{z}_p\right)$ *and* $\mathbf{R}_d\left(z_0, \check{z}_p\right)\mathbf{R}_u\left(\check{z}_p, \check{z}_{p+1}\right)$ *have absolute value less than unity for all* p *with* $0 < p < P-1$,
4. *The eigenvalues of* $\mathbf{R}_u\left(\check{z}_{p+1}, z_P\right)\mathbf{R}_d\left(\check{z}_p, \check{z}_{p+1}\right)$ *and* $\mathbf{R}_d\left(\check{z}_p, \check{z}_{p+1}\right)\mathbf{R}_u\left(\check{z}_{p+1}, z_P\right)$ *have absolute value less than unity for all* p *with* $0 \le p < P-2$,
5. *The eigenvalues of* $\mathbf{R}_u\left(\check{z}_p, z_P\right)\mathbf{R}_d\left(z_0, \check{z}_p\right)$ *and* $\mathbf{R}_d\left(z_0, \check{z}_p\right)\mathbf{R}_u\left(\check{z}_p, z_P\right)$ *have absolute value less than unity for all* p *with* $0 < p < P-1$.

Then the solution to the two-altitude boundary condition problem for the piecewise constant PDE obtained by applying Theorem 3 for every real ω *preserves causality at every* z *with* $z_b \le z \le z_a$.

Indication of proof. Apply the composition rules of Lemma 10 recursively, giving the reflection and transmission matrices for all intervals $(z_0, \check{z}_p)$ with $0 < p \le P-1$ and all intervals $(\check{z}_p, z_P)$ with $0 \le p < P-1$. Then apply Theorem 3 to $z = \check{z}_p$, $0 < p < P-1$, using condition #5. Lemma 12 gives causality for elementary piecewise reflection and transmission matrices, and the assumptions on the eigenvalues imply convergence for all of the power series that arise in the algorithm just described for generating the solution at $z = \check{z}_p$, $0 < p < P-1$. Each summand in each power series is causal, so the solution that is generated is also causal. The solution is extended to the other z values using Theorem 1, only extending upgoing components upwards and downgoing components downwards. Details are left to the reader. □

Letting the piecewise-constant PDE be an approximation of a continuous PDE, and applying Lemma 7 (multilayer limit for single ODE) for each ω, it is evident that Theorem 4 and the steps outlined in its proof provide an algorithm for numerically solving the original PDE and that the solution will preserve causality if all of the conditions are met. If the condition given by (4.45) cannot be met, then the imaginary frequency shifting method can be used to work around this problem in essentially the same way as was described for the constant coefficient case in Section 3.4. We find that the eigenvalue conditions #3 and #4 can readily be made to hold by letting the layer widths be sufficiently small.

Although condition #5 has held in all of the cases we have studied up to now, it appears that caution is advised with this condition, since it involves global (i.e., for all z) interactions between locally-created reflection and transmission. The condition (4.45) can be expected to help in this regard. A systematic investigation of the possibility of condition #5 not holding in scientifically relevant situations is left for future work.

A similar causality result to Theorem 4 can be obtained for the transmission-only approximation, defined by Eq. (4.39), but it should be kept in mind that the transmission-only approximation does not converge to a solution of the PDE in the infinitesimal layer width limit (because it lacks reflection terms), while the method described by Theorem 4 does converge to a solution of the PDE.

5. Application to atmospheric gravity waves

5.1. Vadas–Fritts relations

We consider the linearized anelastic gravity-wave relations derived by Vadas and Fritts [3] for a z-dependent constant-μ atmosphere (see Section 1). We solve for vertical velocity perturbations $w'(x, y, z, t)$, where x and y are the horizontal spatial coordinates, z is height, and t is time, with other wavefield parameters given by polarization relations (e.g., Appendix B of [3]). Following standard procedures (e.g., [3,4]), we define

$$w(x, y, z, t) = \left[\frac{\bar{\rho}(z)}{\bar{\rho}_0}\right]^{\frac{1}{2}} w'(x, y, z, t), \tag{5.1}$$

where $\bar{\rho}(z)$ is background density and $\bar{\rho}_0$ is the value at some reference height. For convenience, we introduce the expressions *unphysical scaling* and *physical scaling*, where w has the unphysical scaling and w' has the physical scaling. Generalizing (2.1)–(2.2) to a three-dimensional Fourier transform in horizontal wavenumber (k, l) and ground-relative frequency ω, we define

$$\widetilde{w}(k, l, z, \omega) = \frac{1}{8\pi^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(x, y, z, t)\exp(-ikx - ily + i\omega t)\, dxdydt, \tag{5.2}$$

and the inverse transform

$$w(x, y, z, t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{w}(k, l, z, \omega)\exp(ikx + ily - i\omega t)\, dkdld\omega. \tag{5.3}$$

We seek linear numerical solutions to $\widetilde{w}(k, l, z, \omega)$ then derive wavefield solutions using (5.3).

Two brief observations will be made in passing. First, an alternative to the rescaling approach in (5.1) is commonly used, which is to instead subtract an imaginary growth term $-i/2H$ from the vertical wavenumber, as in, e.g., [7]. Second, we note that Myers and Yanowitch [37] and works cited therein also study solutions to constant-μ governing equations. Unlike us, they include a small-ν assumption and give asymptotic formulas for wavefield solutions in terms of a variable related to ν. We have nothing further to say about the [37] work for the present.

Given a background atmosphere with z-dependent wind velocity (U, V), buoyancy frequency N, and density scale height $H = -\bar{\rho}/\bar{\rho}_z$, the anelastic viscous and thermally diffusive dispersion relation for vertical wavenumber m is [3, eq. (23)]

$$\left(K^2 + m^2 + \frac{1}{4H^2}\right)\left[\widehat{\omega} - i\nu\left(-K^2 - m^2 + \frac{1}{4H^2} + \frac{im}{H}\right)\right] \left[\widehat{\omega} - \frac{i\nu}{Pr}\left(-K^2 - m^2 + \frac{1}{4H^2} + \frac{im}{H}\right)\right] - K^2N^2 = 0. \tag{5.4}$$

where $K^2 = k^2 + l^2$, $\widehat{\omega}(z) = \omega - kU(z) - lV(z)$ is the intrinsic (wind-relative) wave frequency, ν is kinematic viscosity, introduced in Section 1, and Pr is the Prandtl number, giving the thermal diffusivity α (also introduced in Section 1) as

$$\alpha = \nu/Pr \tag{5.5}$$

where $Pr \sim 0.71$ in the Earth's atmosphere. We now slightly modify the definition of the term "UCF plane" to mean the upper half complex *intrinsic* frequency plane unless otherwise noted. The causality and multilayer method results from the earlier sections carry over to intrinsic frequency with no difficulty, since the horizontal wavenumbers k and l are set to fixed values when those results are applied.

We consider complex $\widehat{\omega}$ and m in solving (5.4), in general, and it is clear that if $\widehat{\omega}$ and m satisfy (5.4), then the negative complex conjugates $-\widehat{\omega}^*$ and $-m^*$ do also. Thus, in examining the roots in m of (5.4) as functions of $\widehat{\omega}$, it suffices to focus on the right side of the complex $\widehat{\omega}$ plane, since from it, the properties of roots on the left side follow, taking the negative complex conjugate. This property also implies that the Fourier solution method described in Section 4 gives real solutions for real boundary conditions.

As described in Section 1, dispersion relations derived from governing equations in which ν is assumed constant (e.g., [4,5,7]) contain only even powers of m (or a related variable involving the imaginary growth term $-i/2H$ as mentioned above), yielding roots that group naturally into upgoing and downgoing pairs. By contrast, odd powers of m appear in the constant-μ dispersion relation (5.4), which complicates identification of upgoing and downgoing modes, as will be seen in Section 5.3. Determination of whether a mode is upgoing or downgoing is typically based on the sign of the imaginary part of the wavenumber (e.g., [6,13]), to give attenuation (rather than unbounded growth) of mode amplitude. Our definition of upgoing and downgoing modes in earlier sections is more general, in that it permits negative (positive) imaginary m for upgoing (downgoing) modes, provided that condition (4.45) is met.

Making the substitution $m \to -i\partial_z$ in (5.4) yields a sixth-order linear homogeneous ODE with z-dependent coefficients. We will refer to any ODE obtained in this way as a *dispersion relation ODE*. (To obtain the dispersion relation ODE, we first completely distribute the terms in (5.4), then reorder factors so that m is to the right of all factors, and then make the substitution $m \to -i\partial_z$. This way the z derivative is not applied to background parameters in the ODE.) These ODEs describe the behavior of $\widetilde{w}(k, l, z, \omega)$ as a function of z for fixed (k, l, ω). One can extend this definition by making analogous substitutions for k, l, and ω (e.g., $k \to -i\partial_x$) and obtain a *dispersion relation PDE*. If the substitution is only made for ω, then a PDE in two variables as in Eq. (4.40) is obtained. Note that there is nothing limiting this to constant background winds.

More generally, let an ODE that describes the variation in z of a state variable be called a vertical structure equation. We consider the dispersion relation ODE to be a vertical structure equation, but it is not the most accurate type of vertical structure equation that can be derived from the linearized governing equations given by Vadas and Fritts [3, eqs. (13)–(16)]. For any particular state variable, e.g., w', one could derive an ODE describing that state variable from the aforementioned governing equations, eliminating the other state variables and allowing for z variation in the background parameters. The resulting ODE would include vertical gradient terms for the background parameters. To illustrate this, consider the Taylor-Goldstein equation given by Baines [38, eq. (4.2.6)] for the inviscid, Boussinesq case. It includes background vertical gradient terms, e.g., U_{zz}. (See also Zhou and Morton [39], who give a vertical structure equation for the inviscid, compressible case.) We omit the background vertical gradient terms from our vertical structure equations, as do Vadas and Fritts [3], implicitly. The application of our multilayer method to vertical structure equations that include background vertical gradient terms is left for future work. (Note that omitting vertical gradient terms, e.g., U_{zz}, from the vertical structure equation does not mean that the background wind is not allowed to vary with altitude. The background parameters, e.g., $U(z)$, *are* allowed to vary in our method, as mentioned above.)

5.2. Viscous Boussinesq case without thermal diffusion

Setting $Pr \to \infty$ [so that thermal diffusivity α vanishes via (5.5)] and $H \to \infty$ in (5.4) yields the viscous Boussinesq dispersion relation of Vadas and Fritts [3, eq. (29)], which is equivalent to McHugh and Grimshaw [13, eq. (12b)] and Hooke and Jones [40, eq. (4)]:

$$\widehat{\omega}\nu m^4 + \left(2K^2\nu\widehat{\omega} - i\widehat{\omega}^2\right)m^2 + \widehat{\omega}K^4\nu - i\widehat{\omega}^2K^2 + iK^2N^2 = 0. \tag{5.6}$$

Since (5.6) is quadratic in m^2, roots occur in $\pm\sqrt{m^2}$ pairs, each identifiable as an upgoing and downgoing mode. The four roots are

$$m = s_o\sqrt{(2\nu)^{-1}i\widehat{\omega}\left(1 + s_i\sqrt{1 + i4\nu\widehat{\omega}^{-3}K^2N^2}\right) - K^2}, \tag{5.7}$$

where $s_o = \pm 1$, $s_i = \pm 1$, and the o and i subscripts represent the "outer" and "inner" square roots, respectively.

Next we make use of the definitions introduced in Section 3.2, including the concept of a branch of a multiple-valued function. The square root function can only be defined continuously on a subset of the complex plane that is missing the points on a curve, i.e., a branch cut, extending from the point 0 to infinity. Given any complex $\widehat{\omega}$, for the inner square root in (5.7) we use the branch that halves the argument (i.e., the angle in the complex plane) between $-\pi$ and π. For the outer square root we use the branch that halves the argument between 0 and 2π. The outer square root always gives a value in the closed (i.e., including the real line) upper-half complex $\widehat{\omega}$ plane, so that the imaginary part of m cannot be negative for $s_o = 1$ or positive for $s_o = -1$. Thus $s_o = \pm 1$ roots identify upgoing and downgoing mode pairs. It can be seen that the imaginary part of (5.7) is always smaller (or equal) for $s_i = -1$ and $\widehat{\omega}$ in the open UCF plane, so the roots obtained for $s_i = -1$ will be called gravity-wave roots, while the roots obtained for $s_i = 1$ will be called viscosity-wave roots (e.g., [4,40]), hereafter identified by "gra" and "vis" subscripts, respectively.

With these definitions, it can be seen that (5.7) is analytic over the entire open UCF plane except on a branch cut located on the line between 0 and $i\left(4\nu K^2N^2\right)^{1/3}$. Analytic continuation of m_{gra} past the branch cut changes m_{gra} into m_{vis} (preserving s_o) for any positive purely imaginary $\widehat{\omega}$ for which $|\widehat{\omega}| < \left(4\nu K^2N^2\right)^{1/3}$, so that this branch point preserves the set of upgoing roots, $\left\{m_{gra}, m_{vis}\right\}$, and downgoing roots (in this case, $\left\{-m_{gra}, -m_{vis}\right\}$).

There are five values of $\widehat{\omega}$ at which branch points occur:

$$\widehat{\omega}_\mathfrak{b} = i\left(4\nu K^2N^2\right)^{1/3}\exp\left[(\mathfrak{b} - 1)\,\frac{2\pi i}{3}\right] \quad (\mathfrak{b} = 1, 2, 3) \tag{5.8}$$

$$\widehat{\omega}_\mathfrak{b} = \frac{i}{2}\left(-\nu K^2 \pm \sqrt{\nu^2K^4 - 4N^2}\right) \quad (\mathfrak{b} = 4, 5)\,. \tag{5.9}$$

(taking a pure imaginary square root in (5.9) if $\nu^2K^4 < 4N^2$). Of these, only the first ($\mathfrak{b} = 1$), which is the same branch point just described in terms of the formula (5.7), lies in the UCF plane. It would not be expected that any of the other branch points $\omega_\mathfrak{b}$ ($\mathfrak{b} = 2, \ldots, 5$) would lie in the UCF plane, since no other discontinuities were seen in the root functions in the open UCF plane except on the branch cut.

Since the $\mathfrak{b} = 1$ branch point interchanges the two upgoing roots, it follows from the converse result given in Theorem 1 that both upgoing roots must be used in order for causality to be preserved in the solution given by Lemma 4. It also can be seen that if the dispersion relation coefficients vary with z then all four roots must be included in order for causality to be preserved in the solution to the two-altitude boundary condition problem.

We now consider the causality condition $B_u > B_d$ from (4.45). The upgoing and downgoing roots have nonnegative and nonpositive imaginary parts, respectively, but it can be seen that for general values of ν the imaginary parts of the roots cannot be bounded away from zero with strict inequality. This means that an argument based on imaginary frequency shifting, as described at the end of Section 3.4, is needed to establish causality. Since there are no solutions to (5.6) with real m for $\widehat{\omega}$ in the open UCF plane, any small frequency shift $i\delta$ with $\delta > 0$ will allow (4.45) to hold. In terms of the polynomial nomenclature of Eq. (3.7), since the leading m^4 coefficient in (5.6), $c_J\left(\widehat{\omega}\right) = \nu\widehat{\omega}$ (where $J = 4$), has a root on the real axis, Theorem 1 condition #2 is violated. This problem also can be circumvented by introducing a small imaginary frequency shift $\delta > 0$ and letting $\delta \to 0$.

Next we establish causality in the inviscid limit. For real $\widehat{\omega}$, the inviscid limits of m_{gra} are

$$\lim_{\nu\to 0} m_{gra} = -s_o\widehat{\omega}^{-1}K\sqrt{N^2 - \widehat{\omega}^2} \quad (|\widehat{\omega}| < N)\,, \tag{5.10}$$

$$\lim_{\nu\to 0} m_{gra} = s_o i|\widehat{\omega}|^{-1}K\sqrt{\widehat{\omega}^2 - N^2} \quad (|\widehat{\omega}| > N)\,. \tag{5.11}$$

Eq. (5.10) follows the usual derivation of upward- or downward-propagating free-propagating ($|\widehat{\omega}| < N$) roots based on the sign of the vertical group velocity (e.g., [22]). The nonpropagating forms in (5.11) for evanescent waves ($|\widehat{\omega}| > N$) yield purely imaginary upward- or downward-decaying roots.

Let these inviscid root functions be extended to the entire closed UCF plane by

$$m_{inv} = s_oK\sqrt{N^2\widehat{\omega}^{-2} - 1} \quad (\mathrm{Im}\left(\widehat{\omega}\right) \neq 0 \text{ or } |\mathrm{Re}\left(\widehat{\omega}\right)| \geq N) \tag{5.12}$$

$$m_{inv} = -s_o\mathrm{sgn}\left(\widehat{\omega}\right)K\sqrt{N^2\widehat{\omega}^{-2} - 1} \quad (\mathrm{Im}\left(\widehat{\omega}\right) = 0 \text{ and } |\mathrm{Re}\left(\widehat{\omega}\right)| < N) \tag{5.13}$$

where the square root in Eq. (5.12) is defined on the complex plane so as to halve the argument between 0 and 2π. (It is understood that m_{inv} can be either upgoing or downgoing, depending on whether $s_o = 1$ or $s_o = -1$, respectively.) It can

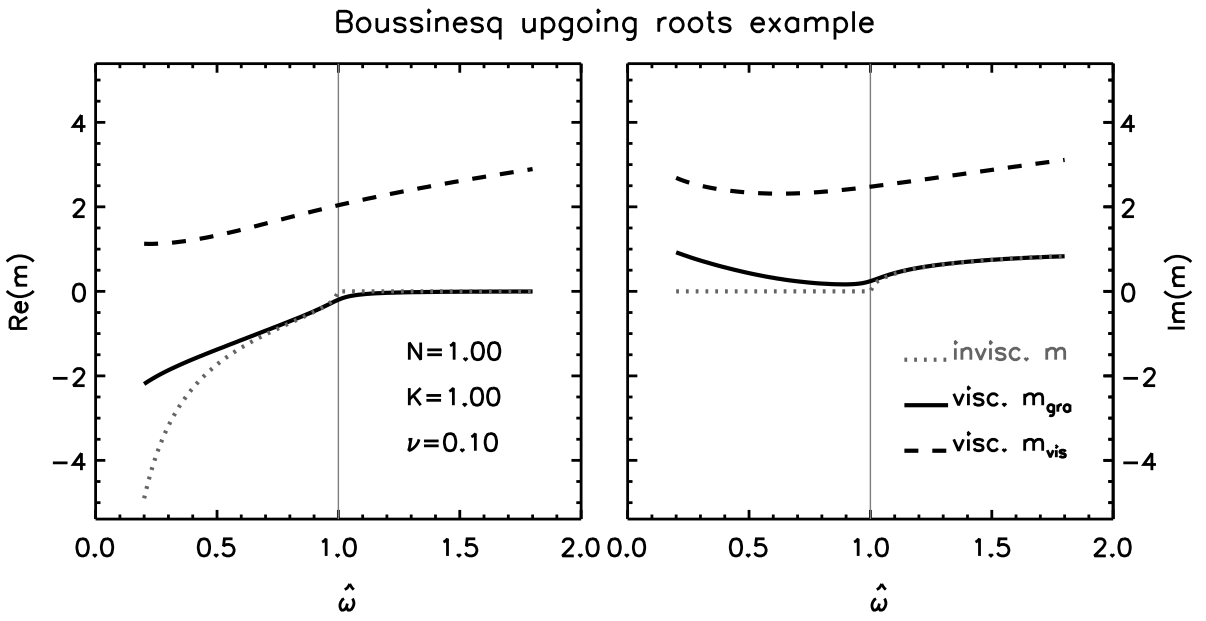

Fig. 1. Inviscid (gray dotted lines), and viscous (non-thermally diffusive) gravity-wave (thick, solid) and viscosity-wave mode (dashed) Boussinesq upgoing roots for the buoyancy frequency N, $K = \sqrt{k^2 + l^2}$ wavenumber, and viscosity ν values shown in the left panel. The thin gray vertical lines in the two panels are aligned with the value of N along the x axis.

be seen that (5.13) is consistent with the limit of (5.12) approaching real $\widehat{\omega}$ values from the open UCF plane. It can also be seen that (5.12) equals the $\nu \to 0$ limit of (5.7) with $s_i = -1$ (i.e., the gravity-wave root). For real values of $\widehat{\omega}$, (5.12) and (5.13) are consistent with the right sides of (5.10) and (5.11). There is no branch cut in the open UCF plane with this definition of the square root, so Eq. (5.12) is analytic in the open UCF plane.

Letting $\nu \to 0$, the branch points in (5.9) approach the inviscid turning-point frequency limits

$$\widehat{\omega}_{itp} = \pm N \tag{5.14}$$

where $m_{gra} \to 0$. These are the only remaining branch points in the inviscid limit since (5.8) vanishes, tending to the uniform $\widehat{\omega} = 0$ critical-level limit where $m_{gra} \to \infty$. As before, Theorem 1 condition #2 and Eq. (4.45) will not hold unless some small positive imaginary frequency shift δ is introduced here, but δ can then be made arbitrarily small, verifying causality for the full parameter space of the inviscid solutions.

Returning to the viscous case, Fig. 1 plots examples of the root functions given by Eq. (5.7) evaluated for positive real intrinsic frequencies in a nondimensional form, in which $N = 1$ and $K = 1$. The real (left panel) and imaginary (right panel) components of the roots m are plotted versus intrinsic frequency. Dotted curves show the inviscid gravity-wave roots, with imaginary components absent in the free-propagating range $|\widehat{\omega}| < 1$ and real components absent where $|\widehat{\omega}| > 1$. Solid curves show viscous m_{gra} solutions for a viscosity value of $\nu = 0.1$. They reveal a reduction relative to the inviscid roots in $|\mathrm{Re}(m_{gra})|$ for free-propagating waves due to viscous damping, particularly at lower frequencies. Dashed curves show the viscosity-wave roots, with imaginary parts larger than those of the gravity-wave roots. (The foregoing example is equivalent to a dimensional example in which $\nu K^2 N^{-1} = 0.1$, where, e.g., space and time variables are made nondimensional by multiplication by K and N, respectively.)

5.3. Viscous anelastic case without thermal diffusion

Letting $Pr \to \infty$ (so that thermal diffusivity α once again vanishes) but retaining H in (5.4) yields the viscous anelastic dispersion relation

$$\begin{aligned}\widehat{\omega}\nu m^4 - \frac{i}{H}\widehat{\omega}\nu m^3 + \left(2K^2\nu\widehat{\omega} - i\widehat{\omega}^2\right) m^2 - \frac{i}{H}\widehat{\omega}\left(K^2\nu + \frac{1}{4H^2}\nu\right) m \\ + K^4\nu\widehat{\omega} - \frac{1}{16H^4}\nu\widehat{\omega} - i\widehat{\omega}^2K^2 - i\widehat{\omega}^2\frac{1}{4H^2} + iK^2N^2 = 0.\end{aligned} \tag{5.15}$$

In the inviscid limit, (5.15) takes the form

$$\widehat{\omega}^2 m^2 + \widehat{\omega}^2 K^2 + \widehat{\omega}^2 \frac{1}{4H^2} - K^2N^2 = 0. \tag{5.16}$$

Using the nomenclature of (5.12) and (5.13), the roots of (5.16) are

$$m_{inv} = s_o K \sqrt{N^2 \widehat{\omega}^{-2} - 1 - \frac{1}{4H^2K^2}} \quad \left(\mathrm{Im}\,(\widehat{\omega}) \neq 0 \text{ or } |\mathrm{Re}\,(\widehat{\omega})|^2 \geq N^2 \left(1 + \frac{1}{4H^2K^2}\right)^{-1} \right) \tag{5.17}$$

$$m_{inv} = -s_o \mathrm{sgn}\,(\widehat{\omega})\, K \sqrt{N^2 \widehat{\omega}^{-2} - 1 - \frac{1}{4H^2K^2}} \quad \left(\mathrm{Im}\,(\widehat{\omega}) = 0 \text{ and } |\mathrm{Re}\,(\widehat{\omega})|^2 < N^2 \left(1 + \frac{1}{4H^2K^2}\right)^{-1} \right) \tag{5.18}$$

where as before the square root halves the argument between 0 and 2π and the root functions (i.e., for upgoing and downgoing waves) are analytic in the UCF plane. Given (5.17) and (5.18), a causality-preserving method can be defined for the constant-background inviscid anelastic case in a similar manner to the inviscid Boussinesq case in Section 5.2, yielding an anelastic turning-point frequency separating inviscid propagating and evanescent modes of the form

$$\widehat{\omega}_{itp} = \pm N \left(1 + \frac{1}{4H^2K^2}\right)^{-\frac{1}{2}}, \tag{5.19}$$

which reproduces the anelastic cutoff frequency for inviscid gravity waves at turning points derived by Marks and Eckermann [41, eq. (2)].

In the general viscous case, since (5.15) is a fourth-order polynomial in m, closed-form expressions for the roots exist, but are not presented since they are too complicated to be useful here. More insight is gained by presenting results for a specific case. In the example that follows, frequencies are again nondimensionalized by setting $N = 1$, but now wavenumbers are nondimensionalized by setting $H = 1$ rather than $K = 1$. Fig. 2a and b plot the real and imaginary components, respectively, of vertical wavenumber m for all four roots of the dispersion relation (5.15) for $\mathrm{Im}(\widehat{\omega}) = 0$ as a function of intrinsic frequency for a nondimensional (as just described) case with $K = 0.2$ and $\nu = 0.2$. The thin, gray, solid line marks the inviscid turning point frequency $|\widehat{\omega}_{itp}| \sim 0.37$ corresponding to $K = 0.2$ in (5.19).

Unlike the viscous Boussinesq and inviscid anelastic cases, odd powers of m appear in (5.15) and thus roots are not symmetric under multiplication by -1, making identification of the upgoing and downgoing roots more difficult. For any particular complex $\widehat{\omega}$, if one arranges the four roots for m in increasing order of $\mathrm{Im}(m)$ then those may be designated downgoing viscosity-wave, downgoing gravity-wave, upgoing gravity-wave, and upgoing viscosity-wave modes, respectively. (This definition is reasonable because it ensures that viscosity-wave modes will dissipate more rapidly than gravity-wave modes in the appropriate altitude direction. Beyond this, the issue of how to define gravity-wave and viscosity-wave modes is not as clear-cut as one might expect, except in certain asymptotic limits, e.g., as $\nu \to 0$, and when closed-form expressions are available, possibly allowing one to "see" the different types of roots in the expressions.) If the four root functions are continuously extended to other $\widehat{\omega}$ values then the order just defined changes. In Fig. 2, the identification of root type is based on the order of $\mathrm{Im}(m)$ at $\mathrm{Re}(\widehat{\omega}) = 0.05$. The four roots are shown by thick curves (both solid and dashed). The upgoing and downgoing modes are shaded black and gray, respectively, and the gravity-wave and viscosity-wave modes are solid and dashed, respectively. This provides an example of the behavior just mentioned, in which the order of $\mathrm{Im}(m)$ changes under continuous extension.

The imaginary parts of the gravity-wave roots (solid black and gray curves in Fig. 2b) change signs at $\widehat{\omega} \sim 0.25$ and ~ 0.33 and both roots have positive $\mathrm{Im}\,(m)$ within the interval $0.25 < \widehat{\omega} < 0.33$. Fig. 2c and d plot roots [versus $\mathrm{Re}(\widehat{\omega})$] derived for the same problem after introducing a small imaginary frequency shift of $\mathrm{Im}(\widehat{\omega}) = \delta = 0.05$. (Note that it is just a coincidence that the value 0.05 also occurs above in the context of identifying root types at $\mathrm{Re}(\widehat{\omega}) = 0.05$.) Inspection of Fig. 2d reveals no curve crossings and preservation of both the sign of $\mathrm{Im}(m)$ for each root and root ordering based on $\mathrm{Im}(m)$ values for all $\mathrm{Re}(\widehat{\omega})$. (Ma [42] described the behavior of roots of a gravity wave dispersion relation in certain cases as a function of altitude for a realistic model atmosphere with viscosity and thermal diffusion. There was insufficient time for us to study that work to determine the connection, if any, with the behavior we are describing.)

Since the curve crossings in Fig. 2b are associated with sign changes in $\mathrm{Im}(m)$ within a narrow frequency interval, the solid curves in Fig. 3 plot the locus of points on the complex $\widehat{\omega}$ plane for which there are roots m with $\mathrm{Im}(m) = 0$ for this problem. Diamonds in Fig. 3 show the branch points, analogous to those derived and presented for the Boussinesq problem in Section 3.2. The curves in Fig. 3 show that real solutions [i.e., $\mathrm{Im}(\widehat{\omega}) = 0$] at the $\mathrm{Re}(\widehat{\omega})$ values of 0.25 and 0.33, noted above, appear to be connected through a nearby branch point having a small positive $\mathrm{Im}(\widehat{\omega})$ value. The inset panel magnifies this branch-point region by a factor of 300, revealing that the branch point actually lies just inside this curve. (The same applies to its symmetric counterpart on the left.) At these branch points, the gravity-wave roots (black and gray solid curves in Fig. 2a, b) converge to the same "double root". (The methods used to establish this are omitted for brevity.)

The crossing of the black and gray solid curves in Fig. 2b means that condition (4.45), required for causality and our multilayer method, cannot be met with real intrinsic frequencies. Imaginary frequency shifts (see Section 3.4) can be used to avoid this problem. Consider a frequency shift $i\delta$ greater than the maximum imaginary value on the solid curves in Fig. 3. At or above the line $\mathbb{R} + i\delta$ in the complex plane of Fig. 3, the imaginary part of a root function cannot change sign, since sign changes can only occur on the curve of frequencies corresponding to real m values. Moreover, it can also be

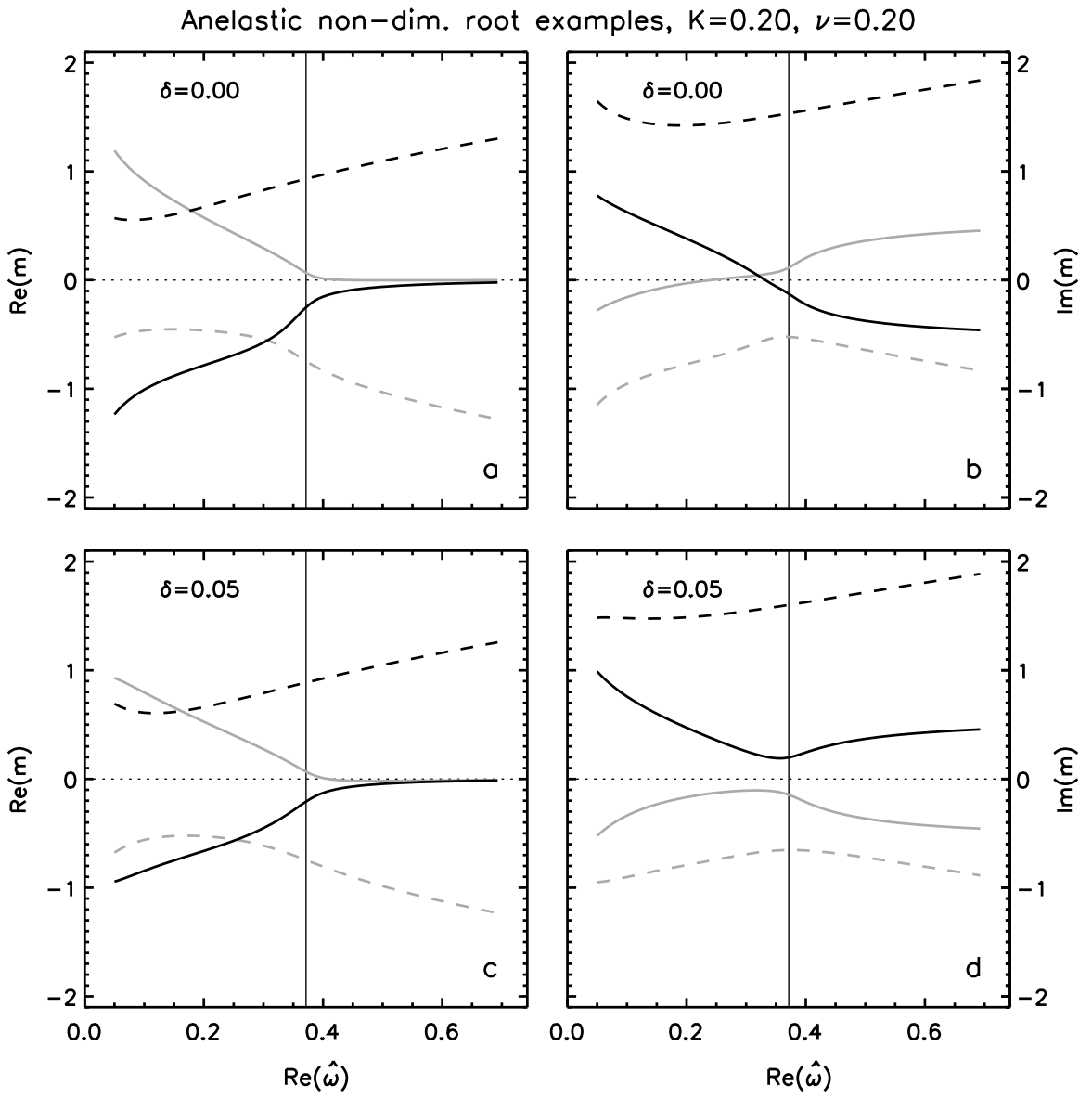

Fig. 2. Roots of the dispersion relation in the nondimensionalized viscous non-diffusive anelastic case with $H = 1$, $N = 1$, $K = 0.2$, and $\nu = 0.2$. Panels a and b show values for real intrinsic frequencies, $\widehat{\omega}$, while panels c and d show values for complex $\widehat{\omega}$ with $\mathrm{Im}(\widehat{\omega}) \equiv \delta = 0.05$. The thick, solid curves are gravity-wave roots, and the thick, dashed curves are viscosity-wave roots. See the text for an explanation of the black and gray shades for the thick curves. The thin, solid, vertical, lines show the inviscid turning point frequency, and the thin, dotted, horizontal lines show y-axis value $= 0$ in each panel.

shown using large $|\widehat{\omega}|$ asymptotics that, for sufficiently large $|\widehat{\omega}|$, there are exactly two roots m with positive imaginary parts and two with negative imaginary parts, such that

$$|\mathrm{Im}(m)| \geq \left(K^2 + \frac{1}{4H^2}\right)^{1/2} + O_{|\widehat{\omega}|\to\infty}\left(|\widehat{\omega}|^{-1}\right). \tag{5.20}$$

Analytic continuation using earlier arguments can then be used to show that there must be exactly two root functions with positive imaginary part and two root functions with negative imaginary part at or above the line $\mathbb{R} + i\delta$. In fact, it can be shown using (5.20) and a continuity argument that for $\mathrm{Im}(\widehat{\omega}) \geq \delta$, the minimum of $\mathrm{Im}[m(\widehat{\omega})]$ [i.e., B_u in (4.45)] for the two upgoing roots must be strictly less than the maximum of $\mathrm{Im}[m(\widehat{\omega})]$ [i.e., B_d in (4.45)] for the two downgoing roots, with $B_d < 0 < B_u$. At any branch point above the line $\mathbb{R} + i\delta$, either the two upgoing or the two downgoing roots become one double root, so all of the branch points above the line preserve upgoing and downgoing sets. The leading

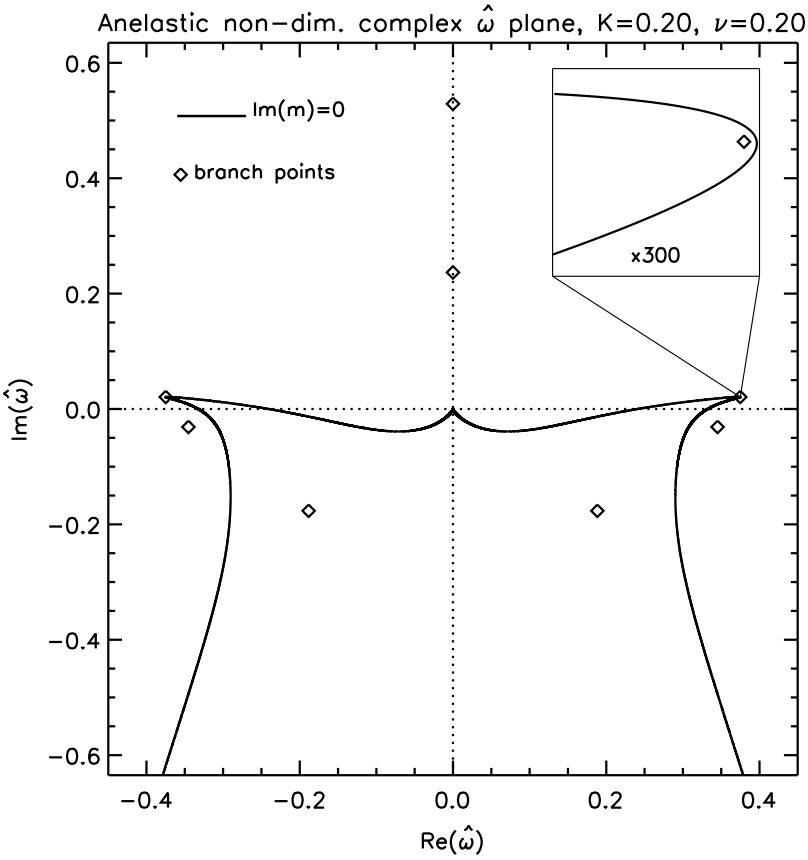

Fig. 3. Locations in the complex intrinsic frequency plane (i.e., $\widehat{\omega}$) of branch points (diamonds, see text) and complex $\widehat{\omega}$ that allow $\mathrm{Im}\,(m\,(\widehat{\omega})) = 0$ (thick, solid curves) for the nondimensionalized viscous non-diffusive anelastic case with $H = 1$, $N = 1$, $K = 0.2$, and $\nu = 0.2$.

coefficient in (5.15), $c_J\,(\widehat{\omega}) = \nu\widehat{\omega}$ (with $J = 4$ in this case), has no roots in the region at or above $\mathbb{R} + i\delta$. Thus all of the conditions needed for causality are met if a δ with the properties described above is introduced.

This raises the question of how to choose an appropriate δ for a given problem. It can be shown that (5.15) satisfies the following limits:

$$\max\,(\widehat{\omega}_i) \;= 0 \;\; \text{(for real } m \text{ and } K \geq 1/2H) \tag{5.21}$$

$$\max\,(\widehat{\omega}_i) \;\leq \nu\left(1/4H^2 \;- K^2\right) \;\; \text{(for real } m \text{ and } K < 1/2H). \tag{5.22}$$

This means that an imaginary frequency shift of magnitude

$$\delta > \nu\left(1/4H^2 \;- K^2\right) \tag{5.23}$$

fulfills the required properties. Fig. 2c and d show the effect of applying an imaginary shift $i\delta$ of $\delta = 0.05$, a value slightly larger than the upper limit of (5.23), which has a dimensionless value of 0.042 in this particular example. As expected, the imaginary parts of the roots in Fig. 2d no longer change sign.

Any δ value greater than the right side of Eq. (5.23) will suffice mathematically, but sometimes smaller values also suffice and are necessary, in practical terms, due to limits on numerical precision and grid resolution. This results from the factor of $\exp(\delta t)$ in Eq. (3.17) introducing a numerical instability when δ is much larger than it needs to be. When $K < 1/2H$ and $\nu \gtrapprox H^2N$, for example, it can be shown that a smaller δ than what is given by the right side of Eq. (5.23) suffices by allowing $B_d > 0$ or $B_u < 0$ in (4.45), whereupon the causality criterion (4.45) can be established for a δ found by considering curves in the complex frequency plane like those shown in Fig. 3, but solving instead for the locus of points where $\mathrm{Im}(m) \equiv \sigma$ for some $\sigma \neq 0$ and

$$|\sigma| < \sqrt{K^2 + \frac{1}{4H^2}}. \tag{5.24}$$

Given Eq. (5.20), analytic continuation leads to the result that, on the line $\mathbb{R} + i\delta$ in the complex $\widehat{\omega}$ plane, two of the vertical wavenumber roots will have $\mathrm{Im}(m) > \sigma$ and two will have $\mathrm{Im}(m) < \sigma$. Using analysis of this type, the details of which are omitted, we have determined that a frequency shift of

$$\delta = 0.065 \times N \tag{5.25}$$

is always sufficient, regardless of the value of ν or the other parameters (although a much smaller shift usually suffices). This issue will be revisited in Sections 5.4 and 6.1.

5.4. Viscous anelastic case with thermal diffusion

We now briefly return to the original sixth-order dispersion relation (5.4) with all original terms retained. The leading coefficient is $c_J(\widehat{\omega}) = -\nu^2/Pr$ (with $J = 6$ in this case). The same problem of a sign change in the imaginary parts of m occurs as in Section 5.3, so that imaginary frequency shifting is again needed and judicious choices for δ are required. In this case, the limit

$$\delta \geq \max\left(1, Pr^{-1}\right) \nu \left(1/4H^2 \ - K^2\right). \tag{5.26}$$

takes the place of (5.23) in placing an upper bound on the maximum δ for which solutions with real m are possible. An argument based on large $\widehat{\omega}$ asymptotics, similar to the one described in Section 5.3, shows that, for sufficiently large $\widehat{\omega}$, there will in general be three roots m with positive imaginary components and three with negative imaginary components. For any particular complex $\widehat{\omega}$, if the six individual root functions $m(\widehat{\omega})$ are sorted in increasing order of $\mathrm{Im}(m)$, the first two and last two root are the downgoing and upgoing dissipative mode roots, respectively (i.e. the viscosity-wave and thermal conduction-wave modes), and the middle two roots are downgoing and upgoing gravity-wave modes, respectively. Similar reasoning to that used in Section 5.3 can be applied when curve crossing of imaginary components of the gravity-wave roots occurs.

As in Section 5.3, a smaller shift than (5.26) is needed in some numerical applications. Eq. (5.26) was obtained by deriving a bound for $\mathrm{Im}(\widehat{\omega})$ on curves in the complex $\widehat{\omega}$ plane for which $\mathrm{Im}(m) \equiv 0$, but it is possible to obtain a smaller shift value by generalizing this approach to curves for which $\mathrm{Im}(m) \equiv \sigma$ for some real σ, not necessarily equal to zero. Analysis similar to that described at the end of Section 5.3 shows that the shift value given by Eq. (5.25) suffices in the viscous and thermally diffusive case for any Pr.

6. Numerical gravity-wave packet examples

6.1. Reflection due to vertical increases in kinematic viscosity

Here we apply our method to viscous and thermally diffusive gravity-wave examples that originally appeared in Vadas and Fritts [3, Section 6]. The atmosphere is idealized: an isothermal 250 K temperature profile is adopted, yielding constant $H = 7$ km and $N = 0.02$ s^{-1}, and no winds are imposed ($U = V = 0$) so that $\widehat{\omega} \equiv \omega$ here. Atmospheric densities $\bar{\rho}(z)$ decrease exponentially with z at the density-scale height H from a value of $\bar{\rho}_0 = 1$ kg m^{-3} at the reference altitude of $z = 0$. These densities combined with a constant $\mu = 1.7\ 10^{-5}$ kg m^{-1} s^{-1} and $Pr = 1$ lead to exponential increases in ν and α with height, as discussed in Section 1. As in [3], we assume that waves propagate in the x-direction, so that $k = K$ and $l = 0$, although this detail is moot given the isotropic atmosphere and our neglect of any small metric terms due to spherical geometry. Vadas and Fritts [3] applied spatial ray methods derived for a dispersion relation equivalent to (5.4) and assuming real m. Since ν varies with z, our constant coefficient method from Section 3 is not sufficient, and we apply our multilayer method from Section 4.

Using the space–time and Fourier domain forms for vertical velocity, w and $\widetilde{w}$, respectively, as defined in (5.1)–(5.3), and the notation of Section 4, our numerical method provides w wavefield solutions as the sum of those for the individual roots:

$$w = \sum_{j=1}^{J} h_j. \tag{6.1}$$

We also define w_{up} and w_{dn} as the partial sums over just the upgoing and downgoing components, respectively, such that

$$w = w_{up} + w_{dn}. \tag{6.2}$$

Three different lower boundary conditions are considered, denoted by a subscript $\mathfrak{l}$, to be defined below. The lower boundary condition at $z_b = 0$ launches an upward-propagating packet of central wavenumber $k_{\mathfrak{l}}$, and frequency $\omega_{\mathfrak{l}}$ with Gaussian amplitude forcing in time, viz

$$w_{up}(x, z_b, t) = A \exp\left[ik_{\mathfrak{l}}x - i\omega_{\mathfrak{l}}t - \left(\frac{\omega_{\mathfrak{l}}t}{5\pi}\right)^2\right], \tag{6.3}$$

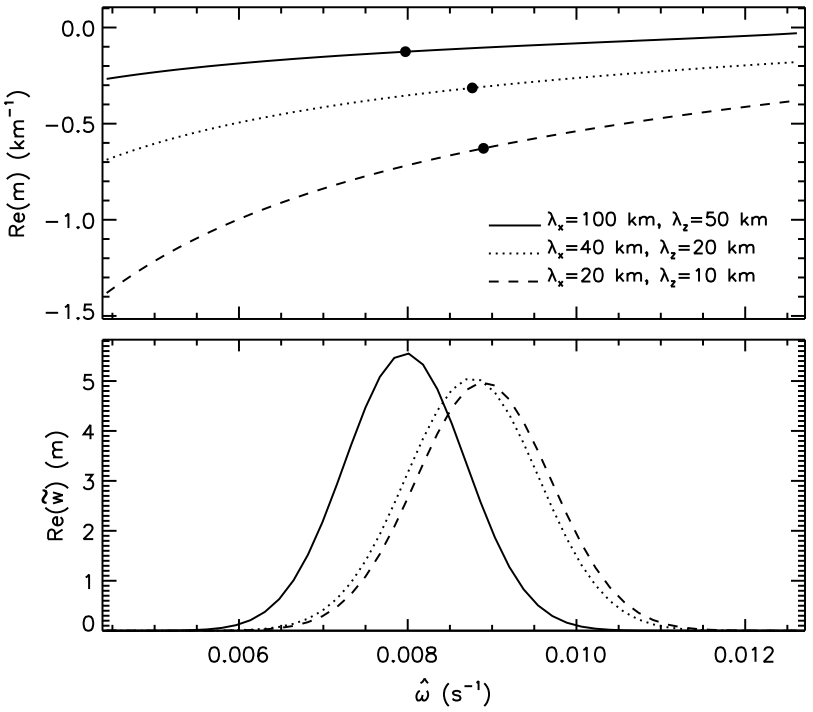

Fig. 4. The upper panel shows the local variation of Re(m), i.e., the real part of the vertical wavenumber, as a function of intrinsic frequency, $\widehat{\omega}$, for three different wave packets at the lower boundary (i.e., $z = 0$) for an example described in the text. The large dots mark the m values at the centers of the respective wave packets. The lower panel shows the three incident wave packets in the frequency domain at $z = 0$.

which takes the equivalent Fourier form of

$$\widetilde{w}_{up}\,(k, z_b, \omega) = \frac{5}{2\,\omega_{\mathfrak{l}}}\sqrt{\pi}A\;\Delta(k - k_{\mathfrak{l}})\exp\left\{-\left[\frac{5\pi\;(\omega - \omega_{\mathfrak{l}})}{2\omega_{\mathfrak{l}}}\right]^2\right\}, \tag{6.4}$$

where Δ denotes the Dirac delta function. The amplitude factor A is set to 0.01 m s^{-1}.

For comparison with three of the cases in [3, Fig. 1] (as well as related examples in later work), we performed three experiments in which $\lambda_z = 2\pi/|m|$ was initialized at z_b to be 50 km, 20 km, and 10 km. We set the index label $\mathfrak{l}$ in (6.3) based on these initial λ_z values: i.e., the $\mathfrak{l} = 50$, $\mathfrak{l} = 20$, and $\mathfrak{l} = 10$ cases. Horizontal wavelengths $\lambda_h = 2\pi/k_{\mathfrak{l}}$ are set to $2\lambda_z$ at z_b in each case. The real and imaginary parts of $\omega_{\mathfrak{l}}$ are initialized at $z_b = 0$ using Eqs. (58) and (59) of [3], respectively, although the imaginary part is tiny since their eq. (59) scales with ν, which is very small at the surface. The upper panel of Fig. 4 plots these initial $\omega_{\mathfrak{l}}$ with filled circles, as well as Re(m) for the upgoing gravity-wave root of the dispersion relation (5.4). The lower panel shows the Fourier form of the lower-boundary condition of Eq. (6.4), revealing a broad frequency distribution around the central frequency $\omega_{\mathfrak{l}}$ in each case, and highlighting the broad range of viscous wavenumber solutions (upper panel) that must be evaluated to accompany these packet frequencies when deriving solutions for all (z, t). The vertical wavenumber m for each Fourier component varies with height due to the z-dependence of ν in the dispersion relation (5.4).

We derive numerical solutions for $\widetilde{w}\,(k_{\mathfrak{l}}, z, \omega)$ using z, t, and ω grid resolutions of 1 km, 36 s, and $2\pi\,(1024 \cdot 36)^{-1}$ s^{-1}, respectively, using a domain of 1024 time/frequency grid points and 300 vertical layers. The upper-boundary radiation condition (4.18) is applied at $z = 300$ km.

In the $\mathfrak{l} = 10$ and $\mathfrak{l} = 20$ cases, k_{10} and k_{20} are both greater than $1/2H$ so that, according to Eq. (5.21), $\delta = 0$ should suffice. However, we adopted a nonzero $\delta = 1 \times 10^{-5}$ s^{-1} on the assumption that it suppresses potential for numerical instability within the solutions. In the third case, k_{50} violates the condition specified in (5.21) and so a nonzero δ was required. However, as $z \to 300$ km the requisite δ required to satisfy the inequality (5.26) becomes sufficiently large that numerical problems arise, of the type discussed in Sections 5.3 and 5.4. We used the type of analysis described at the end of Section 5.3 (the details of which are omitted) to show that $\delta = 8 \times 10^{-4}$ s^{-1} suffices in this case.

As discussed in Section 5.1, our method gives purely real and purely imaginary solutions for purely real and purely imaginary lower boundary conditions, respectively, so by taking the real part of the solution, for example, one obtains the solution corresponding to the real part of the boundary condition. The figures to be presented and discussed next plot only the real parts of the wavefield, and it will be understood that these are the solutions corresponding to the real parts of the lower boundary conditions described earlier.

Fig. 5 shows the full wavefield solutions $w = w_{up} + w_{dn}$ at $x = 0$, with the unphysical scaling described by (5.1), in time-height form for all three cases. All six modes are included, i.e., the two gravity-wave modes and the four dissipative modes. To assess the sensitivity of our new scattering matrix-based algorithms to vertical layer resolution, we repeated the calculations using a 0.5 km vertical resolution, and found absolute differences between the two w solutions of less than 10^{-5} m s^{-1} throughout the domain shown in Fig. 5. (This works out to a maximum difference of 0.1% of the maximum amplitude.) These solutions reveal the basic near-linear group trajectory of each packet, with different slopes revealing different vertical group velocities of each packet and showing broad similarities to the single ray group trajectories in [3, Fig. 1]. A similar figure (not included here) plotting only w_{dn} shows the wavefield of the downward-reflected trajectory, which does reach $z_b = 0$ but with much smaller amplitudes than the wavefield of the upward trajectory.

Fig. 6 presents these same solutions in the unscaled, physical form given by w' (see Eq. (5.1)). Here the w' solutions are separated into their w'_{up} and w'_{dn} components, with all three upgoing modes included in w'_{up} and all three downgoing modes included in w'_{dn}. The right panels of Fig. 6 all reveal the appearance of downgoing packet energy at altitudes where the primary upgoing packet is undergoing significant local amplitude attenuation due to viscous dissipation. On inspecting the color scales in Fig. 6, we see that the ratios of typical downgoing amplitudes to typical upgoing amplitudes are largest for initial $\lambda_z = 50$ km and smallest for initial $\lambda_z = 10$ km. Combining the upgoing and downgoing values in Fig. 6, one obtains full wavefield values (not shown) that are very similar to the upgoing values shown in Fig. 6a, c, e. The maximum amplitudes for the full solutions with the physical scaling are within 5% of those for the upgoing solutions, and the phase structure (indicated by the contours) is similar.

Substantial coupling and destructive interference occurs between the gravity-wave and dissipative modes when all modes are included in the multilayer solution, and the dissipative modes have a significant contribution to the combined solution. That said, one should bear in mind that the result of removing the dissipative modes from the solution after it has been generated is not the same as using an approximation that includes only the gravity-wave modes from the beginning. In the latter case, it may be possible to obtain a good approximation using only the gravity-mode waves, depending on the situation, but we are not considering such an approximation here. We believe that the approach of Francis [8, Appendix] mentioned in Section 1 was more like the latter type of approximation, i.e., including only gravity-wave modes from the beginning. We plan to discuss the accuracy of approximations of this type for a realistic case in a forthcoming paper, but we are leaving this issue aside here.

The results in Fig. 5c for the wave packet with initial $\lambda_z = 10$ km can be somewhat directly compared to Liu et al. [43, Fig. 3] and Heale et al. [18, Fig. 12a]. Liu et al. [43] applied a numerical solution of nonlinear governing equations including viscosity and thermal diffusion, using low amplitudes so as to make the behavior approximately linear. The wave packet results illustrated in [43, Fig. 3] are based on the same background parameters as those of the $\lambda_z = 10$ case in [3, Fig. 1] and show horizontal perturbation velocity with what we call the unphysical scaling. The boundary condition is defined differently in [43], in that the [43] wave packet begins with an initial condition in time, while our examples begin at a lower altitude boundary. Nonetheless, it is reasonable to expect the packets to behave similarly, since they have similar widths in altitude. It is also reasonable to expect the variations in phase and relative amplitude to be similar for vertical and horizontal velocity, and we do, in fact, see such similarity between Fig. 5c and [43, Fig. 3]. Heale et al. [18, Fig. 12a] is based on the same wave packet example just described from [43], again using a nonlinear model with small amplitudes. The similarities with Fig. 5c are clearer in [18, Fig. 12a], since they plot versus time and altitude.

Liu et al. [43, Fig. 10] shows horizontal velocity with the physical scaling for the initial $\lambda_z = 10$ km and $\lambda_z = 20$ km cases. Since, as mentioned above, the full solutions (i.e., with combined upgoing and downgoing values) with the physical scaling are very similar to the upgoing values in Fig. 6, comparisons can be made between Fig. 6c, e and [43, Fig. 10a, b] respectively. We have made rough comparisons of this type. For both the $\lambda_z = 10$ km and $\lambda_z = 20$ km cases, the variations in phase and relative amplitude with time and altitude appear to be consistent, for the most part. It would be helpful to be able to make this type of comparison for initial $\lambda_z = 50$ km, but we are not aware of any such published results. Liu et al. [43, Fig. 5e, f] give momentum flux and vertical wavelength (respectively) power spectra for their $\lambda_z = 50$ km model results, but we are not considering momentum flux or including power spectra analysis in the present work.

6.2. Inviscid turning point

Here, our method is applied to a nondimensional, inviscid, Boussinesq turning-point problem, with background parameters selected in such a way that the Fourier domain functions of altitude, $\widetilde{w}(k, l, z, \omega)$ are exactly given by Airy functions, making it possible to compare our method's results to exact analytic expressions. The z parameter varies from $z_b = -2$ to $z_a = 0.95$.

Two key parameters of our example are $k_0 > 0$ and $\omega_0 > 0$, defining the horizontal wavenumber and central frequency, respectively. We make the problem nondimensional by setting $k_0 = 1$ and $\omega_0 = 1$, which is equivalent to dividing all wavenumbers and frequencies by k_0 and ω_0, respectively, and multiplying all space and time variables by k_0 and ω_0, respectively.

We confine wavenumbers to the (k, m) plane ($l = 0$), so that the y and l dimensions are dropped hereafter. There is no background wind ($U = V = 0$). The nondimensional buoyancy frequency varies with nondimensional height z as

$$N(z) = \sqrt{1 - z}, \tag{6.5}$$

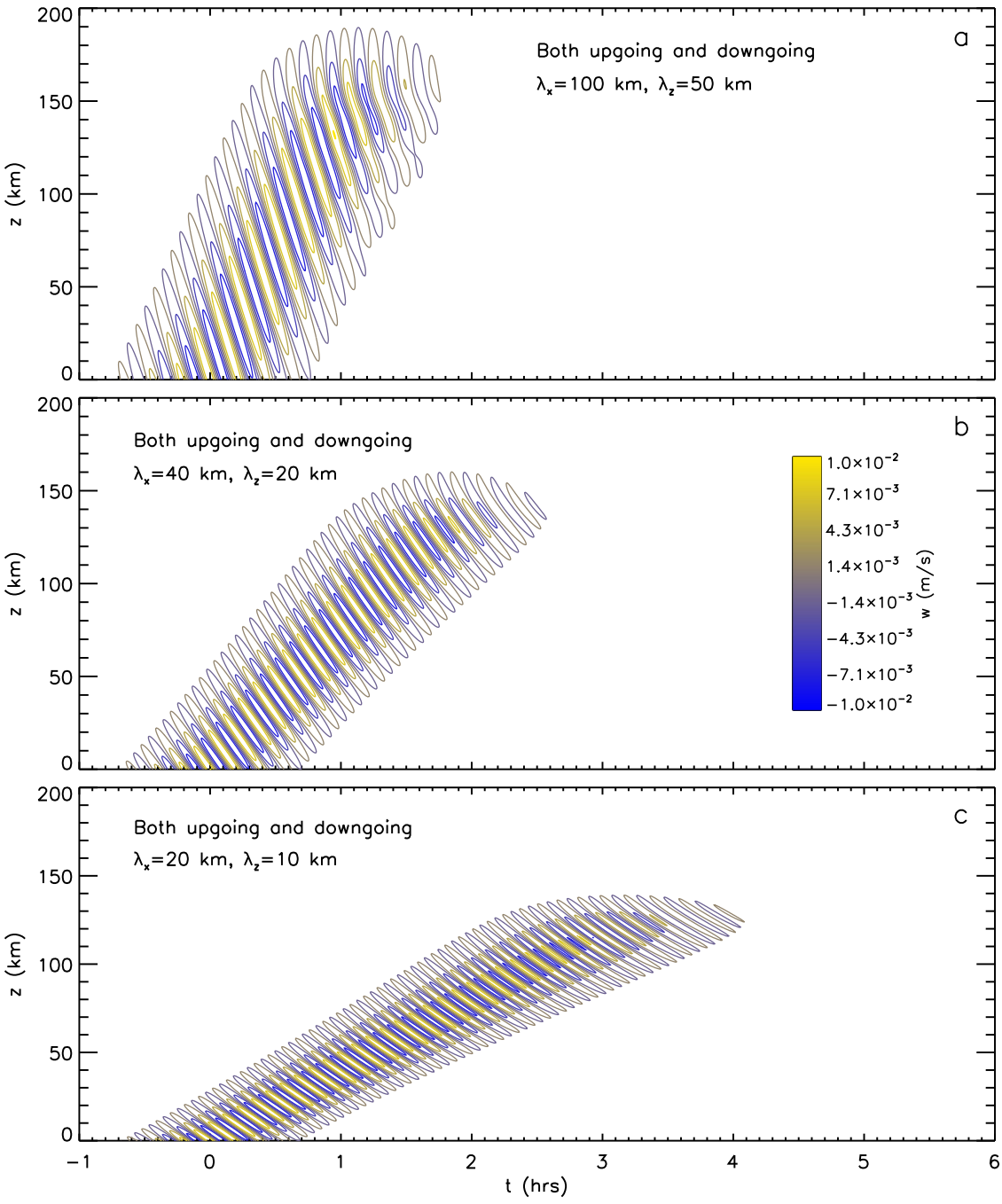

Fig. 5. Vertical velocity contour plots for the example with viscosity and thermal diffusion described in the text. The velocities are scaled in proportion to the square root of background density, as in Eq. (5.1). The three panels correspond to the three upgoing wave packet boundary conditions described in the text, as indicated by the labels showing λ_x and λ_z. The three panels show the full solutions, i.e., $w = w_{up} + w_{dn}$, including all solution modes, at $x = 0$.

vanishing to yield neutral stability at $z = 1$ and having the property $N(0) = \omega_0$, so that from (5.14) we anticipate an inviscid turning point for the central frequency at $z = 0$. The lower boundary condition at $z_b = -2$ for vertical velocity takes the temporal wave-packet form

$$w_{up}(x, z_b, t) = \exp\left[ik_0 x - i\omega_0 t - \left(\frac{\omega_0 t}{5\pi}\right)^2\right] \tag{6.6}$$

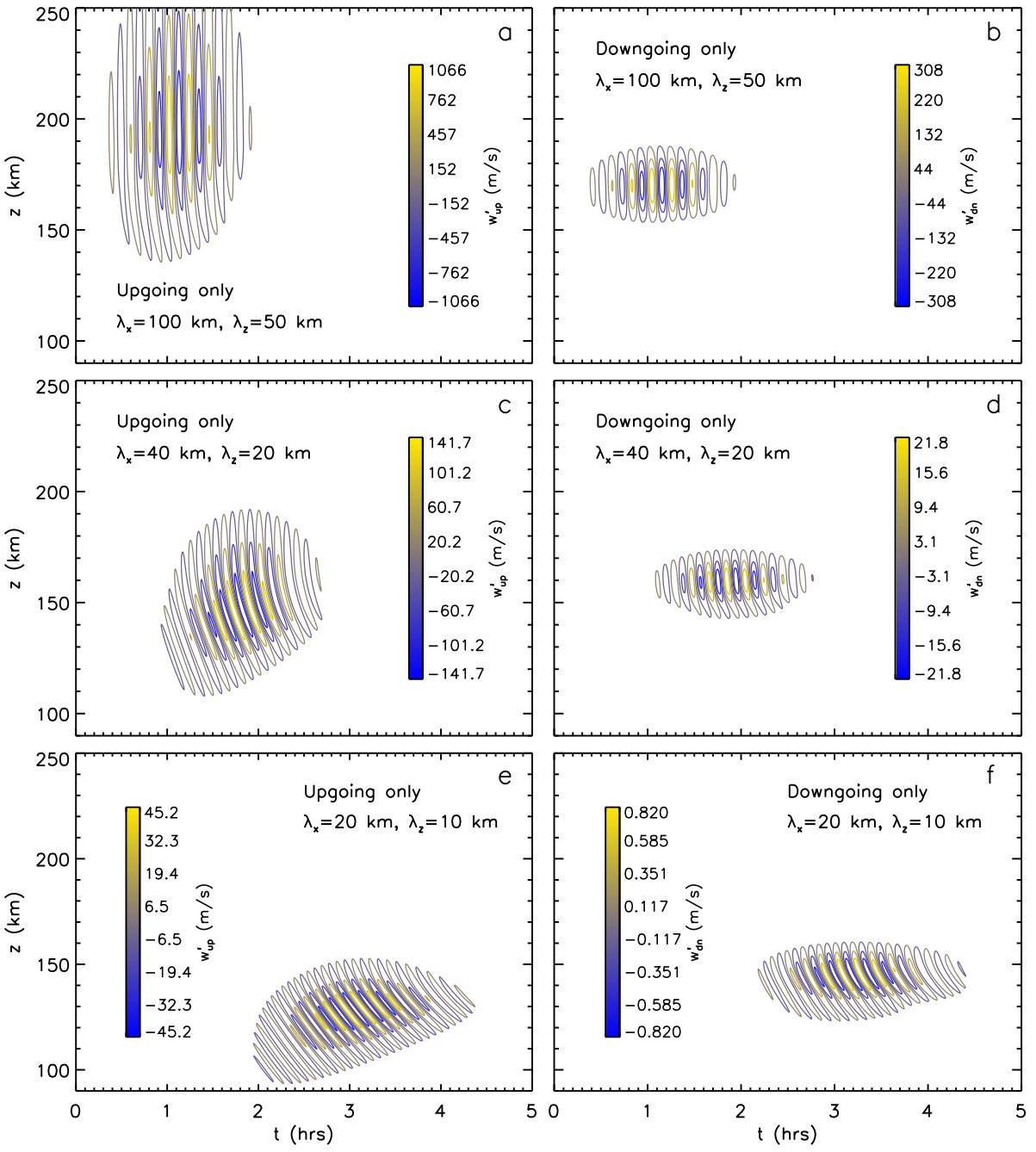

Fig. 6. Vertical velocity contour plots for the same example as in Fig. 5, but now showing the actual vertical velocities, without scaling in proportion to the square root of background density. Another difference is that here the upgoing and downgoing contributions are shown separately, with w'_{up} in panels a, c, and e and w'_{dn} in panels b, d, and f. Unlike in Fig. 5, the contours have a different scale in each panel.

and the equivalent Fourier form,

$$\widetilde{w}_{up}\,(k_0, z_b, \omega) = \frac{5}{2}\sqrt{\pi}\;\Delta(k-k_0)\;\exp\left\{-\left[\frac{5\pi\;(\omega-\omega_0)}{2}\right]^2\right\}, \tag{6.7}$$

where Δ is again the Dirac delta function.

Numerical solutions were derived for the inviscid Boussinesq dispersion relation [with root functions given by (5.12)–(5.13)] in a domain with nondimensional altitude, time, and ω resolutions of 0.01, 0.3, and $2\pi/(1024\cdot 0.3)$, respectively,

with 1024 time/frequency grid points and 295 altitude grid points. The upper-boundary radiation condition (4.18) is applied at $z_a = 0.95$. We included an imaginary frequency shift of $\delta = 0.1$ to avoid numerical instability near turning points at $z \approx 0$. As in Section 6.1, it is to be understood that plotted results to follow show the real part of the wavefield, corresponding to the real part of the lower boundary condition.

Fig. 7a shows numerical solutions at $x = 0$ to this problem using our scattering matrix method, revealing a trapped-wave response that grows and decays in time with the transient forcing. As discussed towards the end of Section 4.3, upward reflection of downgoing waves returning to the lower boundary, z_b, is not currently included in our method.

Fig. 7b shows the "transmission-only" approximation to w, while Fig. 7c shows the upgoing solution w_{up}. The two differ in that the former is only a WKB approximation, including only products of upward transmission matrices, as in Eq. (4.39), while the latter gives the upgoing portion of the full solution, as explained in Section 4.3. Fig. 7c and d show the w_{up} and w_{dn} contributions to the full w solution in Fig. 7a. One can see that at $z = z_b$ the upgoing packet amplitude peaks near $t = 0$ during the time of peak forcing in (6.6), whereas the downgoing packet amplitude peaks at $t \approx 9$, illustrating the group-propagation delay between the incident packet propagating from $z_b = -2$ to $z \approx 0$, where it reflects and then propagates downward back to $z_b = -2$. Both w_{up} and w_{dn} show structure centered near the $z = 0$ turning point that extends to much later times but disappears as the two destructively interfere when added to form the full w solution in Fig. 7a. This demonstrates that w_{up} and w_{dn} must be interpreted in combination and not separately. These canceling features in w_{up} and w_{dn} have frequencies close to the local buoyancy frequency and seem to be modal byproducts of the evanescent behavior near the turning point.

The dispersion relation ODE for this particular problem is

$$\frac{d^2}{dz^2}\widetilde{w}\,(k_n, z, \omega) = \left[(z-1)\,\omega^{-2} + 1\right]\widetilde{w}. \tag{6.8}$$

The Airy functions [44, Section 10.4] Ai (z) and Bi (z) solve the differential equation $f'' - zf = 0$, so solutions to (6.8) can be obtained by applying the Airy functions to linear transformations of the altitude variable. We used this approach to derive corresponding solutions for this problem based on Airy functions and found that the resulting values (not shown) were visually indistinguishable from the plots shown in Fig. 7a, with absolute differences in w of $< 1 \times 10^{-4}$ across the (z, t) domain. (This works out to a maximum difference of 0.01% of the maximum amplitude at the lower boundary.) The ease with which our method provides an accurate numerical approximation in this case suggests that our multilayer approach can provide a viable alternative to Fourier-ray solution methods that rely on special functions in the vicinity of turning points (see, e.g., [45]).

7. Conclusion

We have established general causality results for linear homogeneous PDEs in which one of the two variables is time, allowing for coefficients that vary in the other variable, which is treated here as altitude, and in which boundary conditions are given at the upper and lower edges of an altitude range. The main condition needed in order for our causality results to hold is given by Eq. (4.45), which requires the imaginary parts of the upgoing and downgoing vertical wavenumber roots of a polynomial associated with the PDE, constraining the vertical wavenumbers and the frequencies, to be separated from each other by a finite interval of nonzero width. The main mathematical underpinning for our theory is Titchmarsh's theorem. We also reviewed the mathematics of root functions and branch points and applied the concept of analytic continuation in establishing our results.

The causality results were readily applied to atmospheric gravity-wave cases with a horizontally uniform but vertically varying background, and causality was seen to be a useful organizing principle for understanding these types of waves. Our method allows a definition of upgoing and downgoing modes that extends earlier definitions based on group velocity in the inviscid case (e.g., [22]) and the sign of the imaginary part of vertical wavenumbers when ν is assumed constant in the governing equations (e.g., [6]). We found that difficulties associated with odd powers of vertical wavenumber in the viscous and thermally diffusive anelastic dispersion relation of Vadas and Fritts [3] can be resolved by means of imaginary frequency shifts, which also allow numerical difficulties associated with critical layers and turning points in the inviscid case to be avoided. In the course of establishing causality for the case where PDE coefficients vary with the altitude variable, we also outlined and demonstrated a new method, based on scattering matrix concepts, for circumventing the pervasive problem of numerical swamping, encountered routinely in earlier studies of atmospheric gravity waves in the presence of viscosity and thermal diffusion. There are similarities between our scattering matrix technique and a technique that has appeared in the field of seismology (e.g., [24] Section 7.2.4), but time has not permitted a systematic study of these similarities as of the writing of this paper.

One implication of our application of our method to atmospheric gravity waves is that all modes must be included in order for atmospheric gravity waves generated by transient forcing terms to accurate, since otherwise there would have to be non-causal (and therefore unphysical) effects. In other words, it is insufficient to only include the gravity-wave modes: the dissipative modes must also be included. Because of time and space limitations, we did not attempt to indicate the magnitude of resulting non-causal effects here. In Section 6.1 we report that the dissipative modes have a significant contribution to the overall wavefield in two of the examples considered in Section 6.1.

The question of under what assumptions it can rigorously be proved that the power series in Eq. (4.43) and (4.44) converge is left aside (especially as this relates to condition #5 in Theorem 4), but convergence was found for the examples

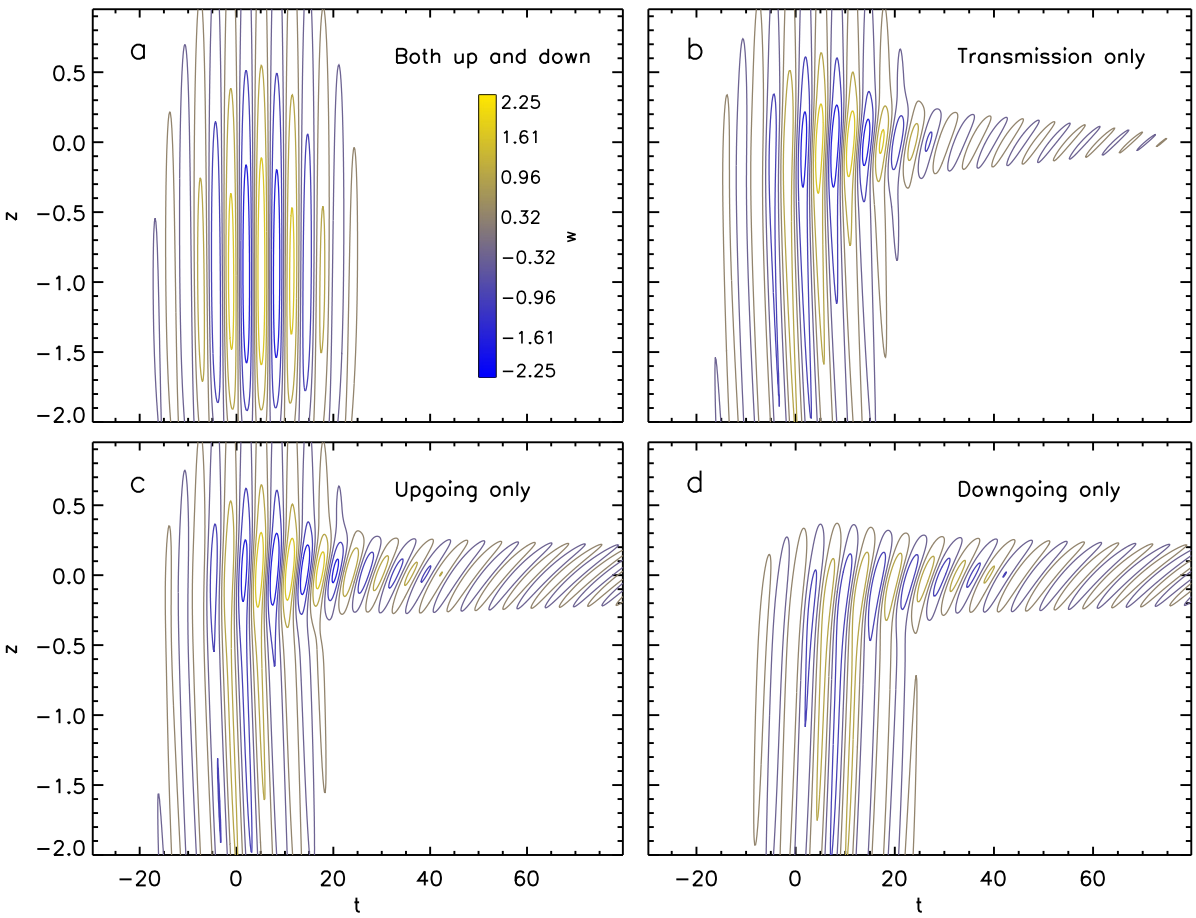

Fig. 7. An inviscid Boussinesq turning point example at $x = 0$, with an incident wave packet and background parameters as described in the text. Time and distance are nondimensional, with $N(0) = 1$, $k = 1$, and $l = 0$. Panel a shows the full multilayer solution for $\mathrm{Re}\,(w)$. Panel b shows the transmission-only approximation. Panels c and d show the upgoing and downgoing contributions to the values in panel a. There is substantial cancellation between the upgoing and downgoing parts near $z = 0$ towards the right sides of the panels, as discussed in the text.

described in Section 6. Also left for future studies is a systematic investigation of how best to determine the minimum necessary imaginary frequency shift δ for any particular case being modeled (in cases where there are odd powers of the vertical wavenumber in the dispersion relation). Using a frequency shift that is larger than necessary does not affect the method in a mathematical sense, but large frequency shifts can present a problem for numerical applications.

Future development of the method could include extending the results of Section 5 to include ion drag, another source of upper-atmospheric damping. It would also be interesting to see how readily the method can be applied to more realistic PDEs for atmospheric gravity waves that include vertical derivatives of background parameters, as discussed in Section 5.1. Additionally, it would be worthwhile to extend these results to linear inhomogeneous PDEs so that more general source terms could be included.

We plan to apply the method to realistic cases of gravity waves generated by deep tropospheric convection that subsequently propagate into the thermosphere. Although we have omitted consideration of nonlinear effects such as wavebreaking from the present discussion, such effects are important and will need to be approximated (e.g., using a parameterized model) in applications such as the one just mentioned. Other areas of application could be in the study of processes associated with gravity waves in the viscous upper atmospheres of other planets (e.g., [46]) and the effect of background turbulence on gravity waves in the lower and middle atmosphere [40,47].

Acknowledgments

This work was supported by the Chief of Naval Research via the NRL base 6.2 program and by the Office of Naval Research through the Basic Research Challenge “Understanding the Phase-Resolved Bottomside Ionosphere (BSION)” (BAA N00014-16-S-BA10).

Appendix A. Proof of causality result for PDE with constant coefficients

Here, Theorem 1 will be proved using Lemma 3 (multidimensional causality conditions). For any fixed z, the Fourier method functions are $\widetilde{\psi}_{j=1\ldots n}$, where each $\widetilde{\psi}_j(\omega)$ is the jth vector elements of the expression $(\mathbf{V}[i\mathbf{m}(\omega)])^{-1}\mathbf{e}^t(z,\omega)$ in (3.11). It is necessary to show that each $\widetilde{\psi}_j$ is analytic in the open UCF plane, has the continuity property, and is bounded as in Eq. (2.6).

Each $\widetilde{\psi}_j(\omega)$ be understood as $F_j[m_1(\omega),\ldots,m_n(\omega)]$ for a function F_j of n variables implicitly defined by (3.11). Since it is assumed in the statement of the theorem that the branch points preserve the roots, Lemma 5 (removal of branch points by symmetry) will imply that $\widetilde{\psi}_j$ is analytic if it can be shown that $F_{j=1\ldots n}$ is both symmetric and analytic on a region of $\mathbb{C}^n$ containing the image of Ω (i.e., the open UCF plane) under the mapping $[m_1(\omega),\ldots,m_n(\omega)]$. It is not difficult to show that each F_j is symmetric, and this is left to the reader.

The Vandermonde determinant formula ([28], eq. (3.4)) gives

$$\det[\mathbf{V}(\boldsymbol{\xi})] = \prod_{1\le\lambda<\lambda'\le n}(\xi_{\lambda'}-\xi_\lambda). \tag{A.1}$$

As is well known, the entries of the inverse of a square matrix can be expressed as the ratios of the determinants of submatrices to the determinant of the overall matrix. Also, the determinant of an $n'\times n'$ matrix can be expressed as a multinomial of order n' in the matrix entries. This means that the Fourier method functions can all be expressed as ratios for which the numerators are sums of products of exponential terms $\exp\left[i(z-z_0)m_j(\omega)\right]$ with the multinomials just described and the denominators are all identical and given by (A.1) applied to $\mathbf{V}[i\mathbf{m}(\omega)]$. This shows that a singularity can only occur when two root functions are equal, which can only be at a branch point that affects both root functions, in the sense described in Section 3.2. The matrix determinant has the property that if two rows are interchanged then the determinant is multiplied by -1, which means that if two roots are interchanged then this also has the effect of multiplying the Vandermonde determinant by -1. Since it has already been established that the elements of (3.11) are symmetric, this means that the numerators must also have the property that interchanging two root functions has the effect of multiplying the numerator value by -1. This means that the numerator must equal zero at a branch point where two root functions, say, $m_j(\omega)$ and $m_{j'}(\omega)$ with $1\le j,j'\le n$ become equal, so the singularity that would result from the factor $\left[m_j(\omega)-m_{j'}(\omega)\right]$ in the denominator is canceled out, and there is no singularity at such a branch point.

The continuity property follows from the definition of the root functions. All that remains to be shown is the bounding condition given by Eq. (2.6), which follows from the definition of the root functions and the assumptions given by Eqs. (3.14) and (3.15), so the theorem has been proved.

The converse statement (i.e., that if a branch point in the open UCF plane interchanges one or more of the roots $m_{j=1\ldots n}$ with a root not in that set then the Fourier method given by Lemma 4 does not preserve causality) follows from Titchmarsh's Theorem and Lemma 1.

Appendix B. Proof of causality result for one jump discontinuity

Here, a proof will be given for Lemma 12. The PDE coefficients $a_{p,q}(z)$ are assumed constant on $[z_b,z_a]$ except for jump discontinuities at a single point, $z_{\mathcal{J}}$, where the $\mathcal{J}$ subscript stands for "jump". The subscripts $-$ and $+$ will be added to terms to indicate values taken on below and above $z_{\mathcal{J}}$, respectively.

Let the notation $\mathbf{E}_{\ldots}(z,\omega)$ have the same meaning as in Section 4.1, denoting a diagonal matrix with exponential terms, but with ω dependence added and the subscript for $\mathbf{E}$ indicating more generally which set of roots is used, e.g., with a $-$ subscript indicating the set of roots below $z_{\mathcal{J}}$. The transfer matrix (which is now a function of ω) going from z_b to z_a may be obtained as in (4.13):

$$\mathbf{H}(z_b,z_a,\omega) = \mathbf{E}_-(z_{\mathcal{J}}-z_b,\omega)\,\mathbf{V}[i\mathbf{m}_-(\omega)]\,\mathbf{V}^{-1}[i\mathbf{m}_+(\omega)]\,\mathbf{E}_+(z_a-z_{\mathcal{J}},\omega). \tag{B.1}$$

The points z_b and z_a will temporarily be brought infinitesimally close to $z_{\mathcal{J}}$, so now z_b is written as $z_{\mathcal{J}-}$, z_a is written as $z_{\mathcal{J}+}$, and Eq. (B.1) becomes

$$\mathbf{H}(z_{\mathcal{J}+},z_{\mathcal{J}-},\omega) = \mathbf{V}[i\mathbf{m}_-(\omega)]\,\mathbf{V}^{-1}[i\mathbf{m}_+(\omega)]. \tag{B.2}$$

In this case, Eq. (4.21) can be rewritten as

$$\mathbf{I}_u\mathbf{V}(i\mathbf{m}_-(\omega)) = [\mathbf{T}_u(z_{\mathcal{J}-},z_{\mathcal{J}+},\omega),\ -\mathbf{R}_u(z_{\mathcal{J}-},z_{\mathcal{J}+},\omega)]\,\mathbf{V}\left(\left[i\mathbf{m}_{u,+}(\omega),\ i\mathbf{m}_{d,-}(\omega)\right]\right), \tag{B.3}$$

where the matrices $\mathbf{T}_u$ and $-\mathbf{R}_u$ have been adjoined to give an $n\times J$ matrix and the vectors $i\mathbf{m}_{u,+}$ and $i\mathbf{m}_{d,-}$ have been adjoined to give a J dimensional vector. This immediately gives the following formulas for the reflection and transmission matrices:

$$\mathbf{R}_u(z_{\mathcal{J}-},z_{\mathcal{J}+},\omega) = -\mathbf{I}_u\mathbf{V}[i\mathbf{m}_-(\omega)]\,\mathbf{V}^{-1}\left(\left[i\mathbf{m}_{u,+}(\omega),\ i\mathbf{m}_{d,-}(\omega)\right]\right)\mathbf{S}_d, \tag{B.4}$$

$$\mathbf{T}_u(z_{\mathcal{J}-},z_{\mathcal{J}+},\omega) = \mathbf{I}_u\mathbf{V}[i\mathbf{m}_-(\omega)]\,\mathbf{V}^{-1}\left(\left[i\mathbf{m}_{u,+}(\omega),\ i\mathbf{m}_{d,-}(\omega)\right]\right)\mathbf{S}_u. \tag{B.5}$$

It follows from the distinctness assumption for the roots and Eq. (4.45) that the Vandermonde matrix that is inverted in Eqs. (B.4) and (B.5) is nonsingular.

First it will be shown that $\mathbf{R}_u\left(z_{\mathcal{J}-}, z_{\mathcal{J}+}, \omega\right)$ is well-defined. Suppose that $\mathbf{I}_d\mathbf{H}\left(z_{\mathcal{J}-}, z_{\mathcal{J}+}, \omega\right)\mathbf{S}_d$ is singular. It follows from the assumption on the distinctness of the roots and Eq. (B.2) that $\mathbf{H}\left(z_{\mathcal{J}-}, z_{\mathcal{J}+}, \omega\right)$ is nonsingular. It must be the case that (B.4) and (B.5) solve (4.21). Multiplying both sides of Eq. (4.21) from the right by $\mathbf{S}_d$, one obtains a contradiction to the aforementioned nonsingularity of $\mathbf{H}\left(z_{\mathcal{J}-}, z_{\mathcal{J}+}, \omega\right)$, so $\mathbf{I}_d\mathbf{H}\left(z_{\mathcal{J}-}, z_{\mathcal{J}+}, \omega\right)\mathbf{S}_d$ must be nonsingular and $\mathbf{R}_u\left(z_{\mathcal{J}-}, z_{\mathcal{J}+}, \omega\right)$ is well-defined. This reasoning can readily be extended to the case in which $z_b < z_{\mathcal{J}} < z_a$. The proof that $\mathbf{T}_u\left(z_b, z_a, \omega\right)$ is nonsingular is similar and omitted.

The remainder of the proof concerns causality. Let $\mathbf{s}\left(n\right)$ denote the row vector with n elements all equal to 1. Let the upgoing $\widetilde{\boldsymbol{\psi}}_u = \left(\widetilde{\psi}_{u,j}\right)_{j=1\ldots n}$ and downgoing $\widetilde{\boldsymbol{\psi}}_d = \left(\widetilde{\psi}_{d,j}\right)_{j=1\ldots n}$ method functions be defined as the functions giving the upgoing and downgoing portions, respectively, of the solution in terms of the lower boundary condition, as in the following two expressions involving dot products of row vectors:

$$\mathbf{s}\left(n\right)\cdot\mathbf{h}_u(z,\omega) = \widetilde{\boldsymbol{\beta}}\cdot\widetilde{\boldsymbol{\psi}}_u(z,\omega) \tag{B.6}$$

$$\mathbf{s}\left(n\right)\cdot\mathbf{h}_d(z,\omega) = \widetilde{\boldsymbol{\beta}}\cdot\widetilde{\boldsymbol{\psi}}_d(z,\omega). \tag{B.7}$$

It follows from (4.36) and (4.38) that the downgoing method functions at $z_{\mathcal{J}-}$ and $z_{\mathcal{J}+}$, respectively, are given by

$$\widetilde{\boldsymbol{\psi}}_d^t\left(z_{\mathcal{J}-},\omega\right) = -\mathbf{V}^{-1}\left[i\mathbf{m}_{u,-}\left(\omega\right)\right]\mathbf{I}_u\mathbf{V}\left[i\mathbf{m}_{-}\left(\omega\right)\right]\mathbf{V}^{-1}\left(\left[i\mathbf{m}_{u,+}\left(\omega\right),\ i\mathbf{m}_{d,-}\left(\omega\right)\right]\right)\mathbf{S}_d\ \mathbf{s}^t\left(n\right), \tag{B.8}$$

$$\widetilde{\boldsymbol{\psi}}_u^t\left(z_{\mathcal{J}+},\omega\right) = \mathbf{V}^{-1}\left[i\mathbf{m}_{u,-}\left(\omega\right)\right]\mathbf{I}_u\mathbf{V}\left[i\mathbf{m}_{-}\left(\omega\right)\right]\mathbf{V}^{-1}\left(\left[i\mathbf{m}_{u,+}\left(\omega\right),\ i\mathbf{m}_{d,-}\left(\omega\right)\right]\right)\mathbf{S}_u\ \mathbf{s}^t\left(n\right). \tag{B.9}$$

Lemma 3 is to be applied to show that the method functions preserve causality. The continuity property and the existence of a bound as in Eq. (2.6) follow as in Appendix A, so the only condition for Lemma 3 left to be shown is that each $\widetilde{\psi}_{u,j}$ and $\widetilde{\psi}_{d,j}$ is analytic in the open UCF plane. Each method function defined by Eqs. (B.8) and (B.9) is a function of three sets of roots: $F(\mathbf{m}_{u,-}, \mathbf{m}_{d,-}, \mathbf{m}_{u,+})$. The conditions of the theorem being proved imply that all of the branch points in the open UCF plane preserve the three sets of roots. Lemma 5 (removal of branch points by symmetry) will imply that each method function is analytic on the open UCF if it can be shown that each F is analytic on a region of $\mathbb{C}^{3n}$ containing the image of Ω (i.e., the open UCF plane) and invariant under any permutation preserving the three sets of roots.

The $\mathbf{m}_{d,-}$ terms drop out of $\mathbf{V}^{-1}\left[i\mathbf{m}_{u,-}\left(\omega\right)\right]\mathbf{I}_u\mathbf{V}\left[i\mathbf{m}_{-}\left(\omega\right)\right]$ (occurring in both (B.8) and (B.9)) because of the presence of $\mathbf{I}_u$. A permutation of the $\mathbf{m}_{u,-}$ roots permutes the columns of $\mathbf{V}^{-1}\left[i\mathbf{m}_{u,-}\left(\omega\right)\right]$ and the rows of $\mathbf{I}_u\mathbf{V}\left[i\mathbf{m}_{-}\left(\omega\right)\right]$ the same way, so the product of these two matrices remains unaffected by such a permutation. A permutation of either the $\mathbf{m}_{u,+}$ or $\mathbf{m}_{d,-}$ roots permutes the columns of $\mathbf{V}^{-1}\left(\left[i\mathbf{m}_{u,+}\left(\omega\right),\ i\mathbf{m}_{d,-}\left(\omega\right)\right]\right)$. The presence of the $\mathbf{S}_d\ \mathbf{s}^t\left(n\right)$ and $\mathbf{S}_u\ \mathbf{s}^t\left(n\right)$ terms in (B.8) and (B.9), respectively, again means that these types of permutations (i.e., preserving the $\mathbf{m}_{u,+}$ or $\mathbf{m}_{d,-}$ roots) do not affect the two formulas, so that the symmetry property has been established for the method functions.

As in Appendix A, it follows from the Vandermonde determinant formula, (A.1), that a singularity can only occur at a complex frequency for which two of the roots from either $\mathbf{m}_{u,-}$ or $\left[\mathbf{m}_{u,+},\ \mathbf{m}_{d,-}\right]$ become equal. If it is a pair of roots within one of the three sets, $\mathbf{m}_{u,-}$, $\mathbf{m}_{u,+}$, or $\mathbf{m}_{d,-}$, then the same argument as in Appendix A applies here. A root from $\mathbf{m}_{u,+}$ and a root from $\mathbf{m}_{d,-}$ cannot become equal in the open UCF plane because of the condition (4.45). Thus it has been established that method functions are analytic in the open UCF plane, so transmitted and reflected components have a causal relationship to the incident components at $z_{\mathcal{J}}$.

Now the result can readily be extended to the more general jump discontinuity case in which $z_b < z_{\mathcal{J}} < z_a$. The details of this are left to the reader, but this step involves simply multiplying the incident (at z_b), transmitted, and reflected components by the appropriate exponential terms. It follows from Theorem 1 that causality is preserved for the reflected and transmitted components.

References

[1] D.F. Fritts, M.J. Alexander, Gravity wave dynamics and effects in the middle atmosphere, Rev. Geophys. 41 (2003).
[2] C.O. Hines, Internal atmospheric gravity waves at ionospheric heights, Can. J. Phys. 38 (1960) 1441–1481.
[3] S.L. Vadas, D.C. Fritts, Thermospheric responses to gravity waves: Influences of increasing viscosity and thermal diffusivity, J. Geophys. Res. 110 (2005) D15103.
[4] M.L.V. Pitteway, C.O. Hines, The viscous damping of atmospheric gravity waves, Can. J. Phys. 41 (12) (1963) 1935–1948.
[5] J.E. Midgley, H.B. Liemohn, Gravity waves in a realistic atmosphere, J. Geophys. Res. 71 (15) (1966) 3729–3748.
[6] H. Volland, Full wave Calculations of gravity wave propagation through the thermosphere, J. Geophys. Res. Sp. Phys. 74 (7) (1969) 1786–1795.
[7] H. Volland, The upper atmosphere as a multiply refractive medium for neutral air motions, J. Atmos. Terr. Phys. 31 (1969) 491–514.
[8] S.H. Francis, Acoustic-gravity modes and large-scale traveling ionospheric disturbances of a realistic, dissipative atmosphere, J. Geophys. Res. 76 (13) (1973) 2278–2301.
[9] K.C. Yeh, C.H. Liu, Acoustic-gravity waves in the upper atmosphere, Rev. Geophys. Sp. Phys. 12 (2) (1974) 193–216.
[10] J. Klostermeyer, Numerical calculation of gravity wave propagation in a realistic thermosphere, J. Atmos. Terr. Phys. 34 (1972) 765–774.
[11] J. Klostermeyer, Computation of acoustic-gravity waves, kelvin-helmholtz instabilities, and wave-induced eddy transport in realistic atmospheric models, J. Geophys. Res. 85 (C5) (1980) 2829–2839.
[12] M.P. Hickey, K.D. Cole, A quartic dispersion equation for internal gravity waves in the thermosphere, J. Atmos. Terr. Phys. 49 (9) (1987) 889–899.

[13] J. McHugh, R. Grimshaw, Reflection of an internal wave at an interface representing a rapid increase in visocity, Geophys. Astrophys. Fluid Dyn. 107 (2013) 603–613.
[14] S. Maeda, Numerical solutions of the coupled equations for acoustic-gravity waves in the upper thermosphere, J. Atmos. Terr. Phys. 47 (1985) 965–972.
[15] S.L. Vadas, Horizontal and vertical propagation and dissipation of gravity waves in the thermosphere from lower atmospheric and thermospheric sources, J. Geophys. Res. 112 (2007) A06305.
[16] R.L. Walterscheid, M.P. Hickey, Group velocity and energy flux in the thermosphere: Limits on the validity of group velocity in a viscous atmosphere, J. Geophys. Res. 116 (2011) D12101.
[17] S.L. Vadas, M.J. Nicolls, The phases and amplitudes of gravity waves propagating and dissipating in the thermosphere: Theory, J. Geophys. Res. 117 (A05322) (2012).
[18] C.J. Heale, J.B. Snively, M.P. Hickey, C.J. Ali, Thermospheric dissipation of upward propagating gravity wave packets, J. Geophys. Res. 119 (2014) 3857–3872.
[19] R.S. Lindzen, H.L. Kuo, A reliable method for numerical integration of a large class of ordinary and partial differential equations, Mon. Weather Rev. 97 (1969) 732–734.
[20] M.P. Hickey, M.J. Taylor, C.S. Gardner, C.R. Gibbons, Full-wave modeling of small-scale gravity waves using Airborne Lidar and Observations of the Hawaiian Airglow (aloha-93) O (1s) images and coincident Na wind/temperature lidar measurements, J. Geophys. Res. 103 (D6) (1998) 6439–6453.
[21] M.P. Hickey, G. Schubert, R. Walterscheid, Propagation of tsunami-driven gravity waves into the thermosphere and ionosphere, J. Geophys. Res. 114 (2009) A08304.
[22] J. Lighthill, Waves in Fluids, Cambridge University Press, Cambridge, UK, 1978.
[23] R. Pérez-Álvarez, F. García-Moliner, Transfer Matrix, Green Function, and Related Techniques: Tools for the Study of Multilayer Heterostructures, Universitat Jaume I, Castelló de la Plana, Spain, 2004.
[24] C.H. Chapman, Fundamentals of Seismic Wave Propagation, Cambridge University Press, Cambridge, UK, 2004.
[25] C.O. Hines, A critique of multilayer analyses in application to the propagation of acoustic-gravity waves, J. Geophys. Res. 78 (1) (1973) 265–273.
[26] H.M. Nussenzveig, Causality and Dispersion Relations, Academic Press, New York, N.Y., USA, 1972.
[27] A. Erdelyi, Asymptotic representations of Fourier integrals and the method of stationary phase, J. Soc. Ind. Appl. Math. 3 (1) (1955) 17–27.
[28] K.E. Atkinson, An Introduction to Numerical Analysis, John Wiley & Sons, New York, N.Y., USA, 1978.
[29] L. Ahlfors, Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable, second ed., McGraw-Hill Book Company, New York, N.Y., USA, 1966.
[30] S.G. Krantz, Function Theory of Several Complex Variables, John Wiley & Sons, New York, N.Y., USA, 1982.
[31] A.D. Pierce, Justification of the use of multiple isothermal layers as an approximation to the real atmosphere for acoustic-gravity wave propagation, Radiat. Sci. 1 (3) (1966) 265–267.
[32] R. Zurmühl, Praktische Mathematik, Springer-Verlag, Berlin, 1965.
[33] M.L.V. Pitteway, The numerical calculation of wavefields, reflection coefficients, and polarization for long radio waves in the lower ionosphere, Phiol. Trans. Roy. Soc. Lond. A257 (1965) 219–241.
[34] K.G. Budden, The Propagation of Radio Waves, Cambridge University Press, Cambridge, UK, 1988.
[35] C.M. Bender, S.A. Orszag, Advanced Mathematical Methods for Scientists and Engineers, McGraw-Hill Book Company, New York, N.Y., USA, 1978.
[36] M.V. Berry, K.I. Mount, Semiclassical approximations in wave mechanics, Rep. Progr. Phys. 35 (1972) 315–397.
[37] R. Myers, M. Yanowitch, Small oscillations of a viscous isothermal atmosphere, J. Comput. Phys. 8 (1971) 241–257.
[38] P. Baines, Topographic Effects in Stratified Fluids, Cambridge University Press, Cambridge, UK, 1995.
[39] Q. Zhou, Y. Morton, Gravity wave propagation in a nonisothermal atmosphere with height varying background wind, Geophys. Res. Lett. 34 (2007) L23803.
[40] W.H. Hooke, R.M. Jones, Dissipative waves excited by gravity-wave encounters with the stably stratified planetary boundary layers, J. Atmos. Sci. 43 (1986) 2048–2060.
[41] C. Marks, S. Eckermann, A three-dimensional nonhydrostatic ray-tracing model for gravity waves: formulation and preliminary results for the middle atmosphere, J. Atmos. Sci. 52 (1995) 1959–1984.
[42] J. Ma, Atmospheric layers in response to the propagation of gravity waves under nonisothermal, wind-shear, and dissipative conditions, J. Mar. Sci. Eng. 4 (25) (2016).
[43] X. Liu, J. Xu, J. Yue, S.L. Vadas, Numerical modeling study of the momentum deposition of small amplitude gravity waves in the thermosphere, Ann. Geophys. 31 (2013) 1–14.
[44] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, N.Y., USA, 1972.
[45] D. Broutman, J.W. Rottman, S.D. Eckermann, A simplified Fourier method for nonhydrostatic mountain waves, J. Atmos. Sci. 60 (2003) 2686–2696.
[46] S.D. Eckermann, J. Ma, X. Zhu, Scale-dependent infrared radiative damping rates on Mars and their role in the deposition of gravity-wave momentum flux, Icarus 211 (2011) 429–442.
[47] W.K. Hocking, S. Fukao, M. Yamamoto, T. Tsuda, S. Kato, Viscosity waves and thermal-conduction waves as a cause of "specular" reflectors in radar studies of the atmosphere, Radiat. Sci. 26 (1991) 1281–1303.