Fully nonlinear simple internal waves over subcritical slopes in continuously stratified fluids: Theoretical development

Shimizu,Kenji

(Citation)
Physics of Fluids,31(1):016601-016601

(Issue Date)
2019-01

(Resource Type)
journal article

(Version)
Version of Record

Physics of Fluids

Fully nonlinear simple internal waves over subcritical slopes in continuously stratified fluids: Theoretical development

Cite as: Phys. Fluids **31**, 016601 (2019); https://doi.org/10.1063/1.5074095
Submitted: 21 October 2018 . Accepted: 26 December 2018 . Published Online: 22 January 2019

Kenji Shimizu

Physics of Fluids ARTICLE scitation.org/journal/phf

Fully nonlinear simple internal waves over subcritical slopes in continuously stratified fluids: Theoretical development

Cite as: Phys. Fluids 31, 016601 (2019); doi: 10.1063/1.5074095
Submitted: 21 October 2018 • Accepted: 26 December 2018 •
Published Online: 22 January 2019

Kenji Shimizu[a)]

AFFILIATIONS

Kobe University, 1-1 Rokkodai-Cho, Nada-Ku, Kobe-Shi, Kobe 657-0013, Japan

[a)]kenji.shimizu.rc@gmail.com

ABSTRACT

The dynamics of fully nonlinear internal waves in continuously stratified fluids is investigated using a new simple-wave (or Riemann-wave) theory with 2nd-order topographic effects. Unlike previous fully nonlinear internal-wave theories, the proposed theory is applicable to large water-depth change over a long distance, provided that the local bottom slopes are mild. An essential step in the theoretical development is the separation of "adiabatic" and "diabatic" topographic effects. First, adiabatic topographic effects are incorporated into the previous simple-wave theory, in such a way that one-directional, fully nonlinear, nondispersive waves become conservative over topography. Since the Riemann variable (or invariant) remains constant along the characteristic, this extends the applicability of the solution method based on characteristic curves from flat bottom to mild slopes. Then, 2nd-order diabatic topographic effects are added to the adiabatic simple-wave solution by a perturbation approach. The application of the proposed theory to Wunsch's subcritical wedge problem suggests that the proposed theory is applicable to subcritical slopes, although the error unavoidably increases as the slope angle approaches the propagation angle of internal wave rays (i.e., the critical angle). An important finding is the relatively rapid growths of wave-induced isopycnal setup and mean flow under the combined effects of strong nonlinearity and topography, in contrast to temporally stationary setup and mean flow under weak topographic effects. This implies, for example, that large-amplitude internal waves on continental shelves have an inherent tendency to modify the "background" stratification and currents without mixing and that it could occur within a spring-neap tidal cycle.

I. INTRODUCTION

Large-amplitude internal waves are commonly observed on continental shelves,[1–3] in marginal seas,[4–6] and in stratified lakes.[7–9] They may steepen and form bores or degenerate into a packet of internal solitary waves (ISWs), which are considered to be important for mixing,[1,10,11] horizontal transport of water mass,[12–14] sediment transport,[15–17] and offshore structures and operations.[4,18] Nonlinear effects are essential in these deformation processes,[19,20] but topography can also play important roles.[21] Although weakly nonlinear theories with 1st-order topographic effects have been used to investigate these processes,[21–23] a theory with strong nonlinear and topographic effects is preferable,[24] for example, to understand the dynamics of large-amplitude internal waves on continental shelves.[2,25] This study develops a fully nonlinear simple-wave theory with 2nd-order topographic effects and illustrates some of the processes due to combined effects of strong nonlinearity and topography.

Several fully nonlinear theories have been developed for internal gravity waves over a flat bottom.[26] Some of them are based on the theory of simple (or Riemann) waves, which is used in various fields of physics to analyze fully nonlinear, nondispersive waves propagating in one direction (see from Sec. 2 to Sec. 6 and Sec. 13 in Ref. 27 or from Sec. 2.8 to Sec. 2.14 in Ref. 28). It assumes that only one of the possible

modes of propagation has nonzero amplitude. This assumption reduces the degrees of freedom of the wave field to one and hence enormously simplifies the description of fully nonlinear waves. Simple waves follow a simple evolutionary equation

$$\left(\frac{\partial}{\partial t} + c^{N}\frac{\partial}{\partial x}\right)R = 0, \tag{1}$$

where t is the time, x is the horizontal coordinate, $c^N(R)$ is the fully nonlinear propagation speed, and $R(x, t)$ is a variable called Riemann invariant. Since R remains constant along the characteristic $dx/dt = c^N$, the problem can be solved by constructing the solution using characteristic curves (see Sec. 2.1 in Ref. 27), which is sometimes called the Riemann method (see Sec. 2.8 in Ref. 28). Hereafter, this method is called the characteristic method. For internal waves over a flat bottom, Ostrovsky and Grue[29] developed a simple-wave theory for two-layer stratified fluids, and Shimizu[30] recently developed the counterpart for continuously stratified fluids. There are also a Miyata-Choi-Camassa (MCC)-type theory[31,32] that allows waves propagating in opposing directions and a fully nonlinear, fully nonhydrostatic theory to calculate internal waves of permanent form using the Dubreil-Jacotin-Long (DJL) equation.[33,34] This study focuses on simple waves without nonhydrostatic effects for simplicity.

The incorporation of topographic effects appears to be a bottleneck in the application of the fully nonlinear internal-wave theories to oceans and lakes. Although it is relatively straightforward to add the effects of small water-depth change to (1) by the perturbation method, water-depth variation in the field is often too large to assume small depth change. For example, the transformation of internal tides into bores and ISWs on continental shelves typically occurs in a depth range of ≈200–≲100 m.[3,22] In general, it appears that there are very limited studies that incorporate the effects of spatially varying composition and cross section into a fully nonlinear simple-wave theory (e.g., see Sec. 2.8 in Ref. 28), although a comprehensive literature survey is difficult because there are vast fields of physics in which the simple-wave theory can be applied. In general physical context, Sec. 2.13 in Ref. 28 showed that the simple-wave theory can be extended to gradually varying composition and cross section assuming small wave amplitude. For sound waves, Varley and Cumberbatch[35] extended the weakly nonlinear theory by Whitham[36] to fully nonlinear pulses in a gradually varying medium; however, their approach based on wave groups is not very useful for internal waves in oceans and lakes, where the resolution of individual waves is usually required.

Korteweg-de Vries (KdV)-type theories with topographic effects[22,37–39] shed some light on the incorporation of large water-depth change into the fully nonlinear simple-wave theory. Including 1st-order nonlinear and topographic effects, the equations (hereafter topography-modified KdV-type equations) can be written as

$$\left(\frac{\partial}{\partial t} + (c + \alpha\eta)\frac{\partial}{\partial x}\right)\sqrt{\frac{c^3}{h}}\eta = S, \tag{2}$$

where $\eta(x, t)$ represents some isopycnal-displacement amplitude, $c(x)$ is the propagation speed of long linear gravity waves, $\alpha(x)$ is the 1st-order nonlinear coefficient, and $h(x)$ is a function determined by the definition of η (water depth if applied to barotropic gravity waves). The "source" term $S(\eta, x)$ could represent the weak effects of not only nonhydrostatic wave dispersion[21,37] but also two-dimensional wave spreading,[38] damping,[22] and rotation.[39] The factor $\sqrt{h^{-1}c^3}$ takes the so-called "adiabatic" topographic adjustment of η into account so that $\sqrt{h^{-1}c^3}\eta$ remains constant along the characteristic if $S = 0$. Interestingly, this is true even if the variations of c, α, and h on the left-hand-side (LHS) are not small. When seen in the frame of reference moving with the nonlinear propagation speed $c + \alpha\eta$, an appropriate scaling of the equation (shown later) suggests that the local bottom slopes need to be mild for the equation to be valid, but no restriction is imposed on the variations of c, α, and h. This suggests a possibility to incorporate adiabatic topographic effects into (1), in such a way that allows large water-depth change over a long distance provided that the local bottom slopes are mild. This requires obtaining an adiabatic fully nonlinear simple-wave solution that follows (1) in the presence of topography. Once this is done, relatively small "diabatic" topographic effects can be added to the right-hand-side (RHS) of (1) by a perturbation approach.

This study aims to extend the fully nonlinear simple-wave theory for continuously stratified fluids[30] to incorporate 2nd-order topographic effects and to illustrate some of the processes due to combined effects of strong nonlinearity and topography. Throughout this study, the bottom slope is assumed to be subcritical, or the slope angle is less than the propagation angle of internal wave rays.[40,41] As a first step in developing a fully nonlinear theory of simple waves with topographic effects, we assume that waves neither break nor form shocks, for simplicity. Although obtaining fully nonlinear solution requires a numerical approach, the focus of this study is primarily on the development of theory rather than a numerical model.

This paper is organized as follows. In Sec. II, we develop a theory of adiabatic, fully nonlinear simple waves over topography, using surface gravity waves in homogeneous fluids as an example. The aim is to illustrate essential ideas and methods in a simpler case, before dealing with more complicated cases in continuously stratified fluids. In Sec. III, we develop a theory of fully nonlinear simple internal waves over subcritical slopes in continuously stratified fluids. The governing equations are first transformed into such a form that the method developed in Sec. II can be directly applied to calculate adiabatic simple internal waves. Then, 2nd-order diabatic topographic effects are added to the adiabatic simple-wave solution by a perturbation approach. In Sec. IV, example applications are used to illustrate some important features of fully nonlinear simple internal waves over subcritical slopes. Implications of the results of this study are discussed in Sec. V, followed by brief conclusions in Sec. VI.

II. BASIC IDEAS AND MATHEMATICAL METHODS FOR CALCULATION OF ADIABATIC SIMPLE WAVES

In this section, we extend the fully nonlinear simple-wave theory from a flat bottom to sloping bottom, using surface gravity waves in homogeneous fluids as an example. Only adiabatic topographic effects are considered in this section. Equations in this section are written in forms that are applicable to continuously stratified fluids in Secs. III and IV, as much as possible.

The simple-wave theory for surface gravity waves is based on the shallow water equations. Assuming background water depths of $H(x)$ and no background current, the equations for the wave-induced deviations of surface elevation $\eta(x, t)$ and horizontal velocity $u(x, t)$ can be written in a matrix form as

$$\frac{\partial \mathbf{p}}{\partial t} = -\mathbf{K}\frac{\partial \mathbf{p}}{\partial x} - \mathbf{G}\mathbf{p}, \tag{3a}$$

$$\mathbf{p} = \begin{bmatrix} \eta \\ u \end{bmatrix}, \tag{3b}$$

$$\mathbf{K} = \begin{bmatrix} u & H+\eta \\ c^2 H^{-1} & u \end{bmatrix}, \tag{3c}$$

$$\mathbf{G} = \begin{bmatrix} 0 & \partial_x H \\ \partial_x (c^2 H^{-1}) & 0 \end{bmatrix}, \tag{3d}$$

where $c(x) = \sqrt{gH}$ is the propagation speed of long linear surface waves, g is the acceleration due to gravity, and ∂ represents the derivative with respect to the variable shown in the subscript. Although $\partial_x(c^2H^{-1}) = \partial_x g = 0$ for homogeneous fluids, we retain this term so that subsequent equations in this section are applicable to continuously stratified fluids.

A. Revisiting simple waves over a flat bottom

For a flat bottom, the simple-wave analysis proceeds in the following four steps.[27,28,30] In the first step, we calculate the eigenvalues and eigenvectors of $\mathbf{K}$. The eigenvalue problem

$$c^{\mathrm{N}}\tilde{\mathbf{l}}^{H} = \tilde{\mathbf{l}}^{H}\mathbf{K} \tag{4}$$

yields

$$c^{\mathrm{N}}_{\pm} = u \pm c\sqrt{1+\eta/H}, \tag{5a}$$

$$\tilde{\mathbf{l}}_{\pm} = \begin{bmatrix} \pm\dfrac{c}{H\sqrt{1+\eta/H}} \\ 1 \end{bmatrix}, \tag{5b}$$

where $c^{\mathrm{N}}_{\pm}$ are the propagation speeds of fully nonlinear simple waves, $\tilde{\mathbf{l}}_{\pm}$ are the left-eigenvectors, and the subscripts + and – denote the right- and left-propagating modes, respectively. Hereafter, the superscript H denotes conjugate transpose, and signs are vertically ordered. In this study, the modes associated with (4) are called "Riemann modes" to distinguish them from vertical modes introduced later. Note that a left-eigenvector can be multiplied by an arbitrary scalar factor. The tilde is used to denote a variable whose magnitude depends on the normalization of left-eigenvectors throughout this paper. There are also right-eigenvectors of $\mathbf{K}$,

$$\tilde{\mathbf{r}}_{\pm} = \frac{1}{2}\begin{bmatrix} \pm\dfrac{H\sqrt{1+\eta/H}}{c} \\ 1 \end{bmatrix}. \tag{6}$$

The left- and right-eigenvectors satisfy the bi-orthogonality

$$\tilde{\mathbf{l}}^{H}_{r}\tilde{\mathbf{r}}_{s} = \delta_{r,s}, \tag{7}$$

where $r, s = +, -$ are the indices for Riemann modes and $\delta_{r,s}$ is the Kronecker delta (=1 if $r = s$, but 0 otherwise).

In the second step, we derive the evolutionary equations of Riemann invariants $\tilde{Q}_r(x, t)$, which are defined through the differential relationships (see Sec. 5.3 in Ref. 27)

$$d\tilde{Q}_r = \tilde{\mathbf{l}}^{H}_{r}\, d\mathbf{p}. \tag{8}$$

Multiplying (3a) without the $\mathbf{G}$ term by $\tilde{\mathbf{l}}^{H}_{r}$ from the left and using the above definition, we get

$$\left(\frac{\partial}{\partial t} + c^{\mathrm{N}}_{r}\frac{\partial}{\partial x}\right)\tilde{Q}_r = 0. \tag{9}$$

These equations mean that $\tilde{Q}_r$ remain constant along the characteristics $dx/dt = c^{\mathrm{N}}_{r}$, but note that c^{N}_{r} depend on both $\tilde{Q}_+$ and $\tilde{Q}_-$.

In the third step, we reduce the degrees of freedom of the wave field by assuming that all the Riemann invariants have zero amplitude except for a mode of interest. To present equations that are applicable to continuous stratification, we hereafter use the subscripts $+p$ and q to denote the Riemann mode of interest (hereafter primary Riemann mode, assumed to be right-propagating) and the rest of the modes (nonprimary Riemann modes), respectively. Assuming waves propagating into initially quiescent water, we get $\tilde{Q}_q(= \tilde{Q}_-) = 0$ along the characteristics for left-propagating mode (see an example in Fig. 1). So, the integration of (8) for $r = -$ yields

$$u = \int_0^{\eta} \frac{c}{H\sqrt{1+\eta'/H}}\, d\eta' = 2c\left(\sqrt{1+\frac{\eta}{H}} - 1\right). \tag{10}$$

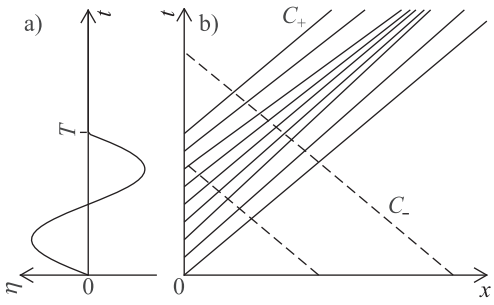

FIG. 1. Schematics of characteristics in a signalling problem. (a) Boundary forcing and (b) characteristics. In (b), solid and dashed lines show right- and left-propagating characteristics, respectively.

Then, (8) for $r = +$ yields

$$\tilde{Q}_{+p} = 2u = 4c\left(\sqrt{1+\frac{\eta}{H}} - 1\right). \tag{11}$$

Alternatively, $\boldsymbol{p}$ can be expressed as a function of $\tilde{Q}_{+p}$ as

$$\boldsymbol{p} = \begin{bmatrix} H\left(1+\tilde{Q}_{+p}/(4c)\right)^2 - H \\ \tilde{Q}_{+p}/2 \end{bmatrix}. \tag{12}$$

Note that the degrees of freedom in terms of the wave field are reduced from 2 (η and u) in (3a) to 1 because η, u, and $\tilde{Q}_{+p}$ are now related through (11).

In the fourth step, initial and boundary conditions are expressed in terms of $\tilde{Q}_{+p}$ using (11), and the evolution of $\tilde{Q}_{+p}$ is followed by (9) for $r = +p$ or the characteristic method.[27,28] The results in terms of η and u are obtained through (12).

B. Adiabatic simple waves over a sloping bottom

We now seek to extend the theory to incorporate adiabatic topographic effects. For a sloping bottom, the simple-wave analysis as presented above is inapplicable because the **G** term in (3a) causes changes in $\tilde{Q}_{+p}$ along the characteristics, as well as wave reflection, or the interaction between the Riemann modes. The starting point for the extension is to modify the definition of Riemann invariants as

$$d\tilde{R}_r = \frac{\tilde{\lambda}_r}{c_r^N}\tilde{\boldsymbol{l}}_r^H(\mathbf{K}d\boldsymbol{p} + \mathbf{G}\boldsymbol{p}dx), \tag{13}$$

where $\tilde{\lambda}_r$ are the weight functions introduced for the reason explained shortly. Note that c_r^N, $\tilde{\boldsymbol{l}}_r$, and $\tilde{R}_r$ depend on c and H rather than x through (4), but it is more convenient to reduce the number of independent variables by using the fact that c and H depend only on x. Multiplying (3a) by $\tilde{\boldsymbol{l}}_r^H$ from the left and using the above definition, we get

$$\left(\frac{\partial}{\partial t} + c_r^N\frac{\partial}{\partial x}\right)\tilde{R}_r = 0, \tag{14}$$

which appear to be the sought-after equations. However, the problem is not so simple as the above equations appear, and we still need to solve the following two problems.

The first remaining problem is to make $\tilde{R}_r$ integrable or to make $\tilde{R}_r$ independent of integration paths. To illustrate the necessity, we proceed by focusing on the $+p$th mode and tentatively assuming that $\tilde{R}_q = \tilde{Q}_q = 0$. Since (8) for $r = +$ implies

$$\boldsymbol{p}(\tilde{Q}_{+p}, x) = \int_0^{\tilde{Q}_{+p}} \tilde{\boldsymbol{r}}_{+p}(\tilde{Q}'_{+p}, x)d\tilde{Q}'_{+p}, \tag{15}$$

we get

$$d\boldsymbol{p} = \left(\frac{\partial \boldsymbol{p}}{\partial \tilde{Q}_{+p}}\right)_x d\tilde{Q}_{+p} + \left(\frac{\partial \boldsymbol{p}}{\partial x}\right)_{\tilde{Q}_{+p}} dx = \tilde{\boldsymbol{r}}_{+p}d\tilde{Q}_{+p} + \left(\frac{\partial \boldsymbol{p}}{\partial x}\right)_{\tilde{Q}_{+p}} dx. \tag{16}$$

Substituting this into (13) for $r = +p$ yields

$$d\tilde{R}_{+p} = \tilde{\lambda}_{+p}d\tilde{Q}_{+p} + \frac{\tilde{\lambda}_{+p}}{c_{+p}^N}\tilde{G}_{+p}dx, \tag{17}$$

where (4) and (7) are used and

$$\tilde{G}_r = \tilde{\boldsymbol{l}}_r^H\left(c_r^N\left(\frac{\partial \boldsymbol{p}}{\partial x}\right)_{\tilde{Q}_{+p}} + \mathbf{G}\boldsymbol{p}\right). \tag{18}$$

Compared to (13), the simple-wave assumption reduced the number of independent variables to 2 (i.e., $\tilde{Q}_{+p}$ and x). If we consider two different integration paths as in Fig. 2(a), the difference in $\tilde{R}_{+p}$ for the infinitesimal paths O-A and O-B is

$$\iint\left(\left(\frac{\partial \tilde{\lambda}_{+p}}{\partial x}\right)_{\tilde{Q}_{+p}} - \left(\frac{\partial}{\partial \tilde{Q}_{+p}}\left(\frac{\tilde{\lambda}_{+p}}{c_{+p}^N}\tilde{G}_{+p}\right)\right)_x\right)d\tilde{Q}_{+p}dx. \tag{19}$$

This shows that $\tilde{R}_{+p}$ is path-dependent in general, but an appropriate choice of $\tilde{\lambda}_{+p}$ makes $\tilde{R}_{+p}$ path-independent, as in Fig. 2(b). This is the reason why $\tilde{\lambda}_r$ are introduced in (13). Mathematically, the requirement

$$\left(\frac{\partial \tilde{\lambda}_{+p}}{\partial x}\right)_{\tilde{Q}_{+p}} = \left(\frac{\partial}{\partial \tilde{Q}_{+p}}\left(\frac{\tilde{\lambda}_{+p}}{c_{+p}^N}\tilde{G}_{+p}\right)\right)_x \tag{20}$$

is the condition for $\tilde{R}_{+p}$ to be an exact differential or the integrability condition of Pfaff's differential equation (17). An example of such a condition in physics is the Maxwell relations in thermodynamics. Once $\tilde{R}_{+p}$ is made integrable, the assumption of $\tilde{R}_{+p} = 0$ at $\tilde{Q}_{+p} = 0$ leads to

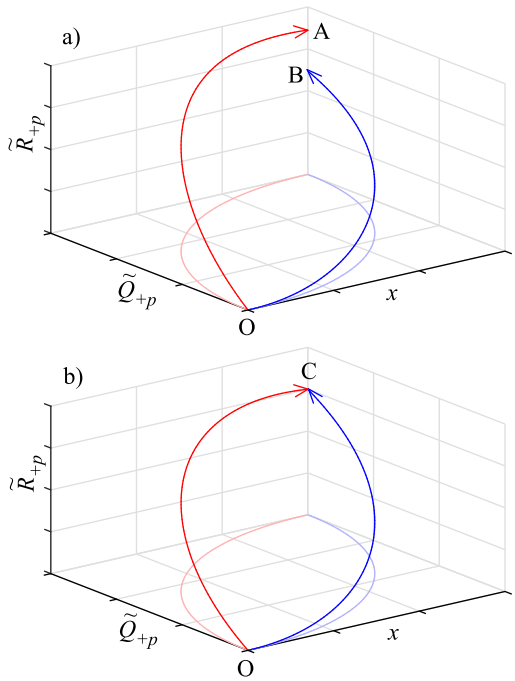

FIG. 2. Schematics of integration of (17) along two paths, shown in red and blue. (a) Not integrable or $\tilde{R}_{+p}$ is path-dependent. (b) Integrable or $\tilde{R}_{+p}$ is path-independent so that the integrals along the two paths are given by $\tilde{R}_{+p}(C) - \tilde{R}_{+p}(O)$.

$$\tilde{R}_{+p} = \int_0^{\tilde{Q}_{+p}} \tilde{\lambda}_{+p}(\tilde{Q}'_{+p}, x) d\tilde{Q}'_{+p}. \tag{21}$$

The topographic factor $\tilde{\lambda}_{+p}$ can also be seen as a normalization factor of $\tilde{\mathbf{l}}_{+p}$ because $\tilde{R}_{+p} = \tilde{Q}_{+p}$ if $\tilde{\mathbf{l}}_{+p}$ is replaced by $\tilde{\lambda}_{+p}\tilde{\mathbf{l}}_{+p}$ from the beginning of the derivation. In a flat-bottom case, the normalization of $\tilde{\mathbf{l}}_{+p}$ is arbitrary, but the same normalization is applied for all x. In a sloping-bottom case, (20) implies that the normalization is arbitrary at a reference location (i.e., a boundary condition for the integration in x), but not arbitrary at the other locations for $\tilde{R}_{+p}$ to be integrable.

The linear limit of the above integrability problem deserves special attention. In this limit, (17) becomes

$$d\tilde{R}_{+p} \sim \tilde{\lambda}_{+p} d\tilde{Q}_{+p} + \frac{\tilde{\lambda}_{+p}}{\sqrt{cH}} \frac{\partial \sqrt{cH}}{\partial x} \tilde{Q}_{+p} dx, \tag{22}$$

where (5)–(7) and (12) are used to evaluate the 2nd term on the RHS. Then, the integrability condition (20) at $\tilde{Q}_{+p} = 0$ yields

$$\tilde{\lambda}_{+p} = A\sqrt{cH}, \tag{23}$$

where A is an integration constant. Since $\eta \sim (2c)^{-1} H \tilde{Q}_{+p} = (2c\tilde{\lambda}_{+p})^{-1} H \tilde{R}_{+p}$ and $c = \sqrt{gH}$, we see that η varies as $\sqrt{H/c^3} \propto H^{-1/4}$ holding $\tilde{R}_{+p}$ fixed (along the characteristic). This agrees with the Green's formula often used in coastal engineering.[42] Note also that (20) involves nonzero $\partial\tilde{\lambda}_{+p}/\partial\tilde{Q}_{+p}$ in general, but not in the linear limit. Therefore, (23) provides the initial condition for the integration of (20) in $\tilde{Q}_{+p}$.

The second remaining problem is how to determine $\tilde{Q}_q$, which is assumed to be 0 in the above analysis. In a flat-bottom case, $\tilde{Q}_q$ is determined by the initial condition $\tilde{Q}_q = 0$, and it remains constant along the characteristic. So, $\tilde{Q}_q = 0$ everywhere for waves propagating into initially quiescent water (e.g., Fig. 1). In a sloping-bottom case, such a characteristic approach is impossible in general because a sloping bottom causes reflection or conversion of the primary Riemann mode to a nonprimary mode. Even for a mild slope, the characteristic approach is difficult because it requires $\tilde{R}_q$ to be integrable, but it appears impossible to satisfy this requirement in one-directional simple-wave analysis.

To determine $\tilde{Q}_q$, we assume a mild slope and take a perturbation approach commonly used in KdV-type theories, which does not require integration along the characteristics. We introduce a frame of reference moving with the nonlinear propagation speed of the primary mode and define

$$\tau = \int \frac{1}{c^N_{+p}(x)} dx - t, \tag{24a}$$

$$\xi = \delta x. \tag{24b}$$

Here, δ is a typical ratio of a horizontal length scale of the topography to the wavelength, assumed to be small. So ξ is a slowly varying horizontal coordinate. The difference from KdV-type theories is that nonlinear propagation speed is used, instead of linear propagation speed. We also assume that (8) is applied with respect to τ, but not ξ. To emphasize this, we replace (8) by

$$\left(\frac{\partial \tilde{Q}_r}{\partial \tau}\right)_\xi = \tilde{\mathbf{l}}^H_r \left(\frac{\partial \mathbf{p}}{\partial \tau}\right)_\xi. \tag{25}$$

Then, noting that the order of relative magnitude of the $\mathbf{G}$ term in (3a) is δ after appropriate scaling, we multiply (3a) by $\tilde{\mathbf{l}}^H_r$ from the left and use (25), yielding

$$\left(-1 + \frac{c^N_r}{c^N_{+p}}\right) \frac{\partial \tilde{Q}_r}{\partial \tau} = -\delta \tilde{\mathbf{l}}_r \left(\mathbf{K} \frac{\partial \mathbf{p}}{\partial \xi} + \mathbf{G}\mathbf{p}\right). \tag{26}$$

To the leading order in δ, we get

$$\left(-1 + \frac{c^N_r}{c^N_{+p}}\right) \frac{\partial \tilde{Q}_r}{\partial \tau} = 0. \tag{27}$$

Assuming waves propagating into initially quiescent water ($\tilde{Q}_q = 0$ at $\tau = 0$), the above equation for $r = q$ yields $\tilde{Q}_q = 0$. $\tilde{Q}_{+p}$ is undetermined at this order, but to the 1st order in δ, we get

$$0 = -\tilde{\lambda}_{+p} \tilde{\mathbf{l}}_{+p} \left(\mathbf{K} \frac{\partial \mathbf{p}}{\partial \xi} + \mathbf{G}\mathbf{p}\right) = -c^N_{+p} \frac{\partial \tilde{R}_{+p}}{\partial \xi}, \tag{28}$$

where $\tilde{\lambda}_{+p}$ is multiplied to convert $\tilde{Q}_{+p}$ to $\tilde{R}_{+p}$ and (13) but $x \to \xi$ is used to get the 2nd expression. In (x, t) coordinates, this becomes (14) for $r = +p$. The above analysis shows that (14) is applicable only to the primary Riemann mode provided that $\tilde{R}_{+p}$ is made integrable and bottom slopes are mild. The evolution of the nonprimary mode needs to be determined from another relation that does not require the integrability of $\tilde{R}_q$, i.e., (27) for $r = q$.

At this point, it is worth noting the conservation properties of the simple-wave solution developed above. The simple-wave solution satisfies the conservation of volume and momentum in (3a), as well as the conservation of total (i.e., available potential plus kinetic) energy associated with the equations

$$\frac{\partial E}{\partial t} = -\frac{\partial J_E}{\partial x}, \tag{29a}$$

$$E = \frac{\rho}{2}\left(\frac{c^2}{H}\eta^2 + (H+\eta)u^2\right), \tag{29b}$$

$$J_E = \rho\left(\frac{c^2}{H}\eta(H+\eta)u + \frac{1}{2}(H+\eta)u^3\right), \tag{29c}$$

where ρ is the density, E is the total energy per unit surface area, and J_E is the vertically integrated energy flux. Since only one Riemann mode has nonzero amplitude, the simple wave solution (12) satisfies the above energy conservation equation. Therefore, the simple-wave solution can be regarded as an adiabatic (or conservative) solution. This adiabatic nature is important to extend the above theory to continuously stratified fluids in Sec. III.

In summary, we have extended the fully nonlinear simple-wave theory from a flat bottom to sloping bottom. The

theory requires that bottom slopes are locally mild ($\delta \ll 1$), but otherwise, no restriction is imposed on the variation of H, c_{+p}^{N}, and $\tilde{\lambda}_{+p}$. Since higher-order topographic effects and continuous stratification cause changes of $\tilde{R}_{+p}$ along the characteristic as we shall see below, we hereafter call $\tilde{R}_{+p}$ the (complete) Riemann variable instead of the Riemann invariant, following Ref. 27 (Sec. 5.3). Also, we call $\tilde{Q}_r$ as incomplete Riemann variables because, unlike $\tilde{R}_{+p}$, $\tilde{Q}_r$ are defined as incomplete (or path-dependent) integrals. In this study, $\tilde{Q}_r$, $\tilde{R}_{+p}$, and $\tilde{\lambda}_{+p}$ are assumed to be dimensionless variables.

III. THEORY OF SIMPLE WAVES OVER MILD SLOPES IN CONTINUOUSLY STRATIFIED FLUIDS

In this section, we extend the theory developed in Sec. II B to continuously stratified fluids. The strategy here is to transform the governing equations into such a form that the theory in Sec. II B is directly applicable. To do so, we write the equations in an isopycnal coordinate and project the equations onto vertical modes in the presence of topography. The adiabatic part of the equations allows us to calculate adiabatic simple waves over topography. Diabatic topographic effects are added later to the adiabatic simple-wave solution by a perturbation approach. Unfortunately, intermediate equations in this section are rather complicated because of the existence of many vertical modes in continuously stratified fluids and because of the need to separate adiabatic and diabatic topographic effects. However, the end results are relatively simple equations, such as the propagation equation for the Riemann variable (14) but with diabatic "source" terms on the RHS.

A. Governing equations in the isopycnal coordinate

We consider a stably stratified, inviscid, incompressible fluid bounded by a free surface at the top and variable topography at the bottom in a nonrotating frame of reference. We assume no background current and a mild bottom slope. We also assume horizontally uniform stratification and hydrostaticity for brevity in this study, although these assumptions can be relaxed.

Following the previous theory of simple waves in continuously stratified fluids,[30] we use an isopycnal vertical coordinate s that is a monotonic function of the background density $\rho_{\mathrm{ref}}(z)$, where z is the upward-positive vertical coordinate with the origin at the equilibrium water surface. For horizontally uniform stratification considered in this study, a particularly useful choice of s is the undisturbed height coordinate Z, which is used in conjugate flow calculation[43] and in association with the DJL equation.[33,34] The governing equations and boundary conditions in the s coordinate are derived from those in the height (z) coordinate using the standard formulae for vertical coordinate transformation.[44] It makes z a prognostic variable and introduces the Montgomery potential $M = p + \rho g z$, where p is the pressure (assumed to be zero at the surface). Then, we separate wave-induced deviations from the background state as

$$z(x,s,t) = Z(s) + \eta(x,s,t), \tag{30a}$$

$$M(x,s,t) = M_{\mathrm{ref}}(s) + m(x,s,t), \tag{30b}$$

$$s^t(x,t) = S^t + \sigma^t(x,t), \tag{30c}$$

$$s^b(x,t) = S^b(x) + \sigma^b(x,t), \tag{30d}$$

where s^t and s^b represent the temporally evolving boundary densities at the top and bottom, respectively. Hereafter, the superscripts t and b are used to indicate values at the top and bottom boundaries, respectively. Note that S^t is assumed be constant because we have assumed horizontally uniform background stratification.

The governing equations and boundary conditions in the s coordinate are as follows. The background condition satisfies the hydrostatic balance

$$0 = -\frac{\partial M_{\mathrm{ref}}}{\partial s} + g\frac{d\rho}{ds}Z \tag{31}$$

and the boundary conditions

$$M_{\mathrm{ref}} = \rho g Z \qquad \text{at } s = S^t, \tag{32a}$$

$$Z = -H(x) \qquad \text{at } s = S^b. \tag{32b}$$

By subtracting (31) from the original, we get the following governing equations for the wave-induced deviations:

$$\frac{\partial}{\partial t}\left(\frac{\partial \eta}{\partial s}\right) = -\frac{\partial}{\partial x}\left(\frac{\partial (Z+\eta)}{\partial s}u\right), \tag{33a}$$

$$\rho\frac{Du}{Dt} = -\frac{\partial m}{\partial x}, \tag{33b}$$

$$0 = -\frac{\partial m}{\partial s} + g\frac{d\rho}{ds}\eta, \tag{33c}$$

where $D/Dt \equiv \partial/\partial t + u\partial/\partial x$. (Note that the diapycnal term $\dot{s}\partial/\partial s$ is neglected because we neglect the diffusion of heat and constituents, such as salt, in this study.) Since s is used as the vertical coordinate, x- and t-derivatives in this paper are assumed to be taken with holding s fixed unless otherwise stated. The kinematic boundary conditions are given by

$$\frac{D\sigma^t}{Dt} = 0 \qquad \text{at } s = S^t + \sigma^t, \tag{34a}$$

$$\frac{D\sigma^b}{Dt} = -u\frac{\partial S^b}{\partial x} \qquad \text{at } s = S^b + \sigma^b. \tag{34b}$$

Since σ^t follows an advection equation without a source term and σ^t is assumed to be the wave-induced deviation, we hereafter set $\sigma^t = 0$. Using (33a), (34b) can be written in a "conservative" form as

$$\frac{\partial}{\partial t}\left(\frac{\partial (Z+\eta)}{\partial s}\sigma^b\right) + \frac{\partial}{\partial x}\left(u\frac{\partial (Z+\eta)}{\partial s}\sigma^b\right) = -u\frac{\partial (Z+\eta)}{\partial s}\frac{\partial S^b}{\partial x} \qquad \text{at } s = S^b + \sigma^b. \tag{35}$$

The above kinematic boundary conditions are used together with

$$m = \rho g \eta \qquad \text{at } s = S^t, \tag{36a}$$

$$Z + \eta = -H \qquad \text{at } s = S^b + \sigma^b, \tag{36b}$$

where (32a) has been subtracted from the dynamic boundary condition at the free surface.

For small σ^b, we may impose approximate bottom boundary conditions at the equilibrium boundary position. This is a standard approach to include the effects of mild slopes[41,45,46] and apparent variation of wave-induced boundary-layer thickness.[47–49] In this study, it is convenient to expand (35) and (36b) in Taylor series around $s = S^b$ because it is consistent with the perturbation expansion employed later. Retaining only terms required in the later derivation, (36b) can be expanded as

$$\eta \sim -\eta_B = -(\eta_B^{(0)} + \eta_B^{(1)}) \quad \text{at } s = S^b, \tag{37}$$

where

$$\eta_B^{(0)} = \frac{\partial (Z+\eta)}{\partial s}\sigma^b, \tag{38a}$$

$$\eta_B^{(1)} = \frac{1}{2}\frac{\partial^2 (Z+\eta)}{\partial s^2}(\sigma^b)^2. \tag{38b}$$

The straightforward Taylor-series expansion of (35) yields an equation for $\eta_B^{(0)} + 2\eta_B^{(1)}$, but we prefer obtaining an equation for $\eta_B = \eta_B^{(0)} + \eta_B^{(1)}$ to be consistent with (37). Fortunately, we may derive an equation for η_B because the expanded equations for the first two orders (for $\eta_B^{(0)}$ and $\eta_B^{(1)}$) are decoupled. To see this, note that $\partial u/\partial s$ is negligible to the first two orders (see Appendix A), which linearizes (35) in terms of $\eta_B^{(0)}$ and $\eta_B^{(1)}$. The resulting equation for η_B is

$$\frac{\partial \eta_B}{\partial t} + \frac{\partial}{\partial x}(u\eta_B) \sim -u\theta \frac{\partial (Z+\eta)/\partial s}{\partial Z/\partial s} - \frac{1}{2}u\theta \frac{\partial^2 (Z+\eta)/\partial s^2}{\partial Z/\partial s (\partial (Z+\eta)/\partial s)}\eta_B \quad \text{at } s = S^b, \tag{39}$$

where $\theta \equiv (dZ/ds)(dS^b/dx)$ is the bottom slope. The scaling used to evaluate the order of magnitude of the terms in (37) and (39) is introduced in Subsection III B, and the details are explained in Appendix A.

B. Projection of governing equations onto vertical modes

In order to use vertical modes under topographic effects, we introduce vertical modes that vary horizontally. Previous studies used horizontally varying vertical modes to include the effects of mild slopes on weakly nonlinear waves[45,46] and those of steep slopes on linear internal waves.[50–52]

Following Shimizu,[30] we introduce vertical modes $\hat{\phi}(x,s)$ and $\hat{\pi}(x,s)$ that satisfy

$$\hat{h}\frac{\partial \hat{\phi}}{\partial s} = \frac{dZ}{ds}\hat{\pi}, \tag{40a}$$

$$\frac{c^2}{\hat{h}}\frac{\partial (\rho\hat{\pi})}{\partial s} = g\frac{d\rho}{ds}\hat{\phi}, \tag{40b}$$

$$c^2\hat{h}^{-1}\hat{\pi} = g\hat{\phi} \qquad \text{at } s = S^t, \tag{40c}$$

$$\hat{\phi} = 0 \qquad \text{at } s = S^b(x), \tag{40d}$$

where $\hat{h}$ is a separation constant. These equations are assumed to be satisfied locally for the boundary position $S^b(x)$; hence, $\hat{\phi}$, $\hat{\pi}$, c, and $\hat{h}$ vary with x. By combining these equations, we get an eigenvalue problem

$$c^2\frac{\partial}{\partial s}\left(\frac{1}{gd\rho/ds}\frac{\partial (\rho\hat{\pi})}{\partial s}\right) = \frac{dZ}{ds}\hat{\pi}, \tag{41a}$$

$$\frac{1}{gd\rho/ds}\frac{\partial (\rho\hat{\pi})}{\partial s} = \hat{\pi} \qquad \text{at } s = S^t, \tag{41b}$$

$$\frac{1}{gd\rho/ds}\frac{\partial (\rho\hat{\pi})}{\partial s} = 0 \qquad \text{at } s = S^b(x). \tag{41c}$$

This problem yields an infinite number of propagation speeds of long linear waves c_n and the associated vertical modes $\hat{\pi}_n$, and the use of (40) yields $\hat{\phi}_n$. Hereafter, the subscripts l, m, n are used as indices for vertical modes, which is 0 for the barotropic vertical mode (hereafter VM0), 1 for the 1st baroclinic vertical mode (VM1), 2 for the 2nd baroclinic vertical mode (VM2), etc. The vertical modes $\hat{\pi}_n$ satisfy the orthogonality relationships

$$\int_{S^b}^{S^t} \hat{\pi}_n \rho \frac{dZ}{ds}\hat{\pi}_m ds = \hat{\rho}\hat{h}_n\delta_{n,m}, \tag{42a}$$

$$\left(\hat{\pi}_n\frac{\rho}{g}\hat{\pi}_m\right)^t - \int_{S^b}^{S^t}\frac{\partial (\rho\hat{\pi}_n)}{\partial s}\frac{1}{gd\rho/ds}\frac{\partial (\rho\hat{\pi}_m)}{\partial s}ds = \frac{\hat{\rho}\hat{h}_n}{c_n^2}\delta_{n,m}. \tag{42b}$$

Hereafter, $(\,)^t$ and $(\,)^b$ are used to indicate that all variables within the parentheses are evaluated at $s = S^t$ and $s = S^b$, respectively. Note that (42a) for $m = n$ defines the norm of the vertical modes and that $\hat{\rho}$ and $\hat{h}_n(x)$ are arbitrary normalization factors with the units of density and water depth, respectively. The caret is used to denote a modal variable whose magnitude depends on the choice of these normalization factors. Although it is customary to set the maximum magnitude of $\hat{\phi}_n$ to be 1, we choose $\hat{h}_n = H$ in this study because the analytic relationship $d\hat{h}_n/dx = -\theta$ is convenient to calculate analytic and numerical solutions. In addition, we choose the sign of $\hat{\pi}_n$ to be positive at the bottom.

We project the governing equation (33) onto the vertical modes by assuming generalized Fourier series of the form

$$\eta(x,s,t) = \hat{\phi}_n(x,s)\hat{\eta}_n(x,t), \tag{43a}$$

$$m(x,s,t) = \frac{\rho(s)}{\hat{\rho}}\hat{\pi}_n(x,s)\hat{m}_n(x,t), \tag{43b}$$

$$u(x,s,t) = \hat{\pi}_n(x,s)\hat{u}_n(x,t). \tag{43c}$$

In this study, summation over all the available (Riemann or vertical) modes applies to repeated dummy subscript indices unless otherwise stated. (If a dummy index appears more than twice within a term, it is understood that the sum of the whole term is taken on that index, instead of the sum for each pair.) Note that we define $\hat{\phi}_n$ and $\hat{\pi}_n$ as nondimensional functions so that a modal amplitude ($\hat{\eta}_n$, $\hat{m}_n$, or $\hat{u}_n$) carries the unit of the original variable. To project (33a)–(33c) onto the vertical modes, we multiply the equations, respectively, by $\rho\hat{\pi}_n$, $(dZ/ds)\hat{\pi}_n$, and $\hat{\phi}_n$, integrate vertically from $s = S^b$ to S^t, apply integration by parts where necessary [in particular, the LHS of (33a) and in the m term in (33c)], substitute the boundary conditions (36a) and (37), substitute then the modal expansion (43), and use the orthogonality relationships (42). The kinematic bottom boundary condition (39) is also required to close the problem. After eliminating $\hat{m}_n$ and applying the scaling explained shortly, the resulting equations are

$$\frac{\partial\hat{\eta}_n}{\partial t} = -\frac{\partial}{\partial x}\left((\hat{h}_n\delta_{n,m} + \epsilon\hat{N}_{mln}\hat{\eta}_l)\hat{u}_m\right) + \delta\left(\hat{L}_{mn}\hat{h}_m + \epsilon\hat{M}_{mln}\hat{\eta}_l\right)\hat{u}_m - \delta\left(\hat{B}^L_{mn} + \epsilon\hat{B}^M_{mln}\frac{\hat{\eta}_l}{\hat{h}_l}\right)\hat{u}_m - \delta\frac{\rho^b\hat{\pi}^b_n}{\hat{\rho}}\frac{\partial\eta_B}{\partial t} + O(\epsilon^2\delta^3), \tag{44a}$$

$$\frac{\partial\hat{u}_n}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{c_n^2}{\hat{h}_n}\hat{\eta}_n\right) - \epsilon\hat{N}_{lnm}\hat{u}_l\frac{\partial\hat{u}_m}{\partial x} - \delta\left(\hat{L}_{nm}\frac{c_m^2}{\hat{h}_m}\hat{\eta}_m + \epsilon\hat{M}_{lnm}\hat{u}_l\hat{u}_m\right), \tag{44b}$$

$$\frac{\partial\eta_B}{\partial t} = -\left(1 + \epsilon\hat{\pi}^b_l\frac{\hat{\eta}_l}{\hat{h}_l}\right)\hat{\pi}^b_m\hat{u}_m\theta - \epsilon\frac{\partial}{\partial x}\left(\hat{\pi}^b_m\hat{u}_m\eta_B\right) - \epsilon\delta\frac{1}{2}\left(\frac{d^2Z/ds^2}{(dZ/ds)^2}\right)^b\hat{\pi}^b_m\hat{u}_m\theta\eta_B + O\left(\epsilon\delta^2, \epsilon\delta\frac{C^2}{gD}\right). \tag{44c}$$

Note that no sum is taken on n because it is not a dummy index. In the above equations, the following variables are modal interaction coefficients defined as

$$\hat{N}_{nlm} = \frac{1}{\hat{\rho}\hat{h}_l}\int_{S^b}^{S^t}\hat{\pi}_n\rho\frac{dZ}{ds}\hat{\pi}_l\hat{\pi}_m ds, \tag{45a}$$

$$\hat{L}_{nm} = \frac{1}{\hat{\rho}\hat{h}_n}\int_{S^b}^{S^t}\hat{\pi}_n\rho\frac{dZ}{ds}\frac{\partial\hat{\pi}_m}{\partial x}ds, \tag{45b}$$

$$\hat{M}_{nlm} = \frac{1}{\hat{\rho}\hat{h}_l}\int_{S^b}^{S^t}\hat{\pi}_n\rho\frac{dZ}{ds}\hat{\pi}_l\frac{\partial\hat{\pi}_m}{\partial x}ds, \tag{45c}$$

$$\hat{B}^L_{nm} = \frac{1}{\hat{\rho}}(\hat{\pi}_n\rho\hat{\pi}_m)^b\theta, \tag{45d}$$

$$\hat{B}^M_{nlm} = \frac{1}{\hat{\rho}}(\hat{\pi}_n\rho\hat{\pi}_l\hat{\pi}_m)^b\theta. \tag{45e}$$

From the top, they represent interactions due to the advection of horizontal momentum and isopycnal heaving, linear and nonlinear topographic effects, and linear and nonlinear effects of varying bottom boundary density, respectively. Note that $\hat{N}_{nlm}$, $\hat{B}^L_{nm}$, and $\hat{B}^M_{nlm}$ are dimensionless, but $\hat{L}_{nm}$ and $\hat{M}_{nlm}$ have the unit of the inverse of horizontal length. It is useful to note the rules for index swapping and the identities among the interaction coefficients.[30] Additional relationships from Shimizu[30] are $\hat{M}_{lnm} = \hat{h}_n^{-1}\hat{h}_l\hat{M}_{nlm}$, $\hat{B}^L_{mn} = \hat{B}^L_{nm}$, $\hat{B}^M_{mln} = \hat{B}^M_{lnm} = \hat{B}^M_{nlm}$, $\hat{L}_{mn}c_n^2/\hat{h}_n = -\hat{L}_{nm}c_m^2/\hat{h}_m$ for $(n \neq m)$ (see Appendix B for proof), and

$$\hat{M}_{nlm} = \hat{N}_{nlj}\hat{L}_{jm}, \tag{46}$$

where the sum on j is taken. [To get (46), note that $\partial\hat{\pi}_m/\partial x = \hat{L}_{jm}\hat{\pi}_j$ from (45b), substitute it into (45c), and use (45a).]

The following scaling is applied to (44) to evaluate the order of magnitude of the terms. Since we have chosen $\hat{h}_n = H$, we get $\hat{h}_n = O(D)$ and $\hat{\pi} = O(1)$, where D is the typical water depth. This makes $\hat{\phi}_n$, $\hat{N}_{nlm}$, $\hat{B}^L_{nm}$, and $\hat{B}^M_{nlm}$ variables of O(1). Then, we scale the other variables in (44) based on D, typical bottom slope Θ, wavelength L, wave propagation speed C, and wave-induced current speed U. These result in

$$x = Lx^\dagger,\ t = (L/C)t^\dagger, \tag{47a}$$

$$c = Cc^\dagger, \hat{h} = D\hat{h}^\dagger, \theta = \Theta\theta^\dagger, \tag{47b}$$

$$\hat{\eta} = (UD/C)\hat{\eta}^\dagger, \hat{m} = (\hat{\rho}UC)\hat{m}^\dagger, \hat{u} = U\hat{u}^\dagger, \tag{47c}$$

$$\eta_B = (U\Theta L/C)\eta_B^\dagger, \tag{47d}$$

$$\hat{L}_{nm} = (\Theta/D)\hat{L}^\dagger_{nm},\ \hat{M}_{nlm} = (\Theta/D)\hat{M}^\dagger_{nlm}, \tag{47e}$$

$$\hat{B}^L_{nm} = \Theta\hat{B}^{L\dagger}_{nm},\ \hat{B}^M_{nlm} = \Theta\hat{B}^{M\dagger}_{nlm}, \tag{47f}$$

where † denotes a scaled variable. After the above scaling, we get (44) in nondimensional form with $\epsilon = U/C$ and $\delta = \Theta L/H$, setting $\hat{\rho} = 1$ and omitting † for brevity. The dimensional form of the equations is recovered by setting $\epsilon = \delta = 1$. Note that ϵ is unnecessary for fully nonlinear cases, but it is kept above for convenience in reducing the equations to linear or weakly nonlinear cases.

As we see in (44a) and (44b), the evolutionary equations of modal amplitudes have many more terms compared to the original governing equations (33). For example, the volume flux term in (33a) becomes the linear and nonlinear terms without topographic effects, linear topographic terms, and mixed nonlinear-topographic terms in (44a). It is certainly a disadvantage to deal with more terms, but (44) has an advantage that the effects of different physical processes are separated. In Subsection III C, we shall see that the separation of physical processes is essential to calculate adiabatic simple waves in continuously stratified fluids.

C. Adiabatic simple waves in continuously stratified fluids

We now apply the theory developed in Sec. II to continuously stratified fluids. To do so, it is convenient to write the

evolutionary equations of modal amplitudes (44a) and (44b) in a matrix form. We consider terms up to 2nd order in δ, which are necessary in the perturbation analysis in Subsection III D. We also focus on internal waves and assume that $C^2/(gH) \lesssim \delta$. Then, we write (44a) and (44b) as

$$\frac{\partial \hat{\mathbf{p}}}{\partial t} \sim -\hat{\mathbf{K}}\frac{\partial \hat{\mathbf{p}}}{\partial x} - \delta\hat{\mathbf{G}}\hat{\mathbf{p}} - \delta\left(\hat{\mathbf{B}}\hat{\mathbf{p}} + \hat{\mathbf{b}}\frac{\partial \eta_{\mathrm{B}}}{\partial t}\right), \tag{48}$$

where

$$\hat{\mathbf{G}} = \hat{\mathbf{G}}_{\mathrm{all}} + \hat{\mathbf{B}}_{\mathrm{self}}, \tag{49a}$$

$$\hat{\mathbf{B}} = \hat{\mathbf{B}}_{\mathrm{int}} - \hat{\mathbf{B}}_{\mathrm{self}} + \hat{\mathbf{B}}_{\mathrm{bnd}} \tag{49b}$$

and η_{B} is determined by the kinematic bottom boundary condition (44c). Here, the vectors and matrices are schematically defined as

$$\hat{\mathbf{p}} = \begin{bmatrix} \hat{\eta}_n \\ \hat{u}_n \end{bmatrix}, \tag{50a}$$

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{\rho}^{-1}\rho^b\hat{\pi}_n^b \\ 0 \end{bmatrix}, \tag{50b}$$

$$\hat{\mathbf{K}}(\hat{\mathbf{p}}) = \begin{bmatrix} \epsilon\hat{\mathrm{N}}_{lmn}\hat{u}_l & \hat{h}_n\delta_{n,m} + \epsilon\hat{\mathrm{N}}_{mln}\hat{\eta}_l \\ c_n^2\hat{h}_n^{-1}\delta_{n,m} & \epsilon\hat{\mathrm{N}}_{lnm}\hat{u}_l \end{bmatrix}, \tag{50c}$$

$$\hat{\mathbf{G}}_{\mathrm{all}}(\hat{\mathbf{p}}) = \begin{bmatrix} 0 & \partial_x(\hat{h}_n)\delta_{n,m} + \epsilon\partial_x(\hat{\mathrm{N}}_{mln})\hat{\eta}_l \\ \partial_x(c_n^2\hat{h}_n^{-1})\delta_{n,m} & 0 \end{bmatrix}, \tag{50d}$$

$$\hat{\mathbf{B}}_{\mathrm{int}}(\hat{\mathbf{p}}) = \begin{bmatrix} -\epsilon\hat{\mathrm{M}}_{lmn}\hat{u}_l & -\hat{\mathrm{L}}_{mn}\hat{h}_m \\ \hat{\mathrm{L}}_{nm}c_m^2\hat{h}_m^{-1} & \epsilon\hat{\mathrm{M}}_{lnm}\hat{u}_l \end{bmatrix}, \tag{50e}$$

$$\hat{\mathbf{B}}_{\mathrm{self}}(\hat{\mathbf{p}}) = \begin{bmatrix} -\epsilon\hat{\mathrm{N}}_{lmn}\hat{\mathrm{L}}_{nn}\hat{u}_l & -\hat{\mathrm{L}}_{nn}\hat{h}_n\delta_{n,m} \\ \hat{\mathrm{L}}_{nn}c_n^2\hat{h}_n^{-1}\delta_{n,m} & \epsilon\hat{\mathrm{N}}_{lnm}\hat{\mathrm{L}}_{mm}\hat{u}_l \end{bmatrix}, \tag{50f}$$

$$\hat{\mathbf{B}}_{\mathrm{bnd}}(\hat{\mathbf{p}}) = \begin{bmatrix} 0 & \hat{\mathrm{B}}^{\mathrm{L}}_{mn} + \epsilon\hat{\mathrm{B}}^{\mathrm{M}}_{mln}\hat{\eta}_l\hat{h}_l^{-1} \\ 0 & 0 \end{bmatrix}. \tag{50g}$$

The argument $\hat{\mathbf{p}}$ of the above matrices corresponds to $\hat{\eta}_n$ and $\hat{u}_n$ used to evaluate the nonlinear terms. Although the matrices and vectors are written for the nth mode, (48) can also be interpreted as matrix-form equations for all the vertical modes. For example, if we include a n_{M} number of vertical modes, $\hat{\mathbf{p}}$ can be seen as a $2n_{\mathrm{M}} \times 1$ column vector and $\hat{\mathbf{K}}$ can be seen as a $2n_{\mathrm{M}} \times 2n_{\mathrm{M}}$ matrix. The latter interpretation is more convenient to deal with eigenvalue problems associated with $\hat{\mathbf{K}}$.

To calculate adiabatic simple waves, we consider only the following terms in (48),

$$\frac{\partial \hat{\mathbf{p}}}{\partial t} \sim -\hat{\mathbf{K}}\frac{\partial \hat{\mathbf{p}}}{\partial x} - \delta\hat{\mathbf{G}}\hat{\mathbf{p}}, \tag{51}$$

and replace $\boldsymbol{p}$, $\mathbf{K}$, and $\mathbf{G}$ in Sec. II by $\hat{\mathbf{p}}$, $\hat{\mathbf{K}}$, and $\hat{\mathbf{G}}$, respectively, in this section. [Note that the above equation reduces to (3a) for homogeneous fluids. To see this, we consider only the barotropic mode ($n = 0$), choose $\hat{h}_0 = H$ and $\hat{\rho} = \rho$, and note that the bottom boundary density does not change ($\eta_{\mathrm{B}} = 0$). Then, we get $\hat{\pi}_0(\mathrm{s}) \sim 1$ and $\hat{\mathrm{N}}_{000} \sim 1$.] More specifically, simple-wave solution can be obtained by solving the eigenvalue problem for Riemann modes (4), focusing on a particular right-propagating (+pth) Riemann mode, assuming that $\tilde{\mathrm{Q}}_q = 0$, expressing the prognostic variable vector $\hat{\mathbf{p}}$ as a function of $\tilde{\mathrm{Q}}_{+p}$ using (15), calculating $\tilde{\lambda}_{+p}$ from (20), and calculating $\tilde{\mathrm{R}}_{+p}$ from (21). Then, the evolution of simple waves is described by a simple propagation equation of the Riemann variable $\tilde{\mathrm{R}}_{+p}$, (14) for $r = +p$. Although (48) has terms that are not used in the simple-wave calculation, the effects of those small terms of order δ can be added to the adiabatic simple-wave solution by a perturbation approach, as shown in Subsection III D. Before doing so, however, we need to look at some properties of the simple-wave solution.

Although (51) has a form similar to (3a), there is a difference in that $\hat{\mathbf{G}}$ in (51) includes topographic self-interaction terms, represented by $\hat{\mathbf{B}}_{\mathrm{self}}$ in (49a). These terms are included in $\hat{\mathbf{G}}$ in order to make the solution to (51) independent of the normalization of vertical modes, except for a constant multiplication factor. The details are provided in Appendix C.

The above simple-wave solution is adiabatic or satisfies the conservation of volume, momentum, and total (i.e., available potential plus kinetic) energy, as for surface waves in homogeneous fluids (Sec. II). However, the energy conservation property of (51) is not so obvious because the additional $\hat{\mathrm{N}}_{nlm}$ factors break an analogy with (3a). The derivation of the total energy equation associated with (51) is shown in Appendix D. An important point from the appendix is that only terms that contribute to horizontal energy flux are retained in (51).

D. Perturbation expansion for diabatic topographic effects

In this subsection, we add diabatic topographic effects to the adiabatic simple-wave solution derived in Subsection III C. To do so, we follow essentially the perturbation approach used in Sec. II, but consider up to 2nd-order topographic effects.

To add the diabatic topographic effects, we need to consider nonprimary Riemann modes. It is convenient to separate the following variables into the parts due to the primary mode (or $\tilde{\mathrm{R}}_{+p}$) and nonprimary modes (or $\tilde{\mathrm{Q}}_q$):

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}^{\mathrm{SW}} + \Delta\hat{\mathbf{p}}, \tag{52a}$$

$$c_r^{\mathrm{N}} = c_r^{\mathrm{SW}} + \Delta c_r, \tag{52b}$$

$$\tilde{\mathbf{l}}_r = \tilde{\mathbf{l}}_r^{\mathrm{SW}} + \Delta\tilde{\mathbf{l}}_r, \tag{52c}$$

$$\eta_{\mathrm{B}} = \eta_{\mathrm{B}}^{\mathrm{SW}} + \Delta\eta_{\mathrm{B}}, \tag{52d}$$

where (52a) applies to its components, $\hat{\eta}_n$ and $\hat{u}_n$. The variables with the superscript SW are due to the primary Riemann mode. Note that $\hat{\mathbf{p}}^{\mathrm{SW}}$, c_{+p}^{SW}, and $\tilde{\mathbf{l}}_r^{\mathrm{SW}}$ are calculated from the relationships for the adiabatic simple-wave solution as functions of $(\tilde{\mathrm{R}}_{+p}, x)$; however, these variables are not the adiabatic parts

because $\tilde{R}_{+p}$ is modified by diabatic topographic effects, as shown below. This is analogous to non-equilibrium thermodynamics in which, for example, adiabatically defined Gibbs free energy is calculated using temperature affected by diabatic mixing. The symbol Δ denotes the part due to nonprimary modes, which are due to diabatic topographic effects.

Although we follow essentially the perturbation approach used in Sec. II, the existence of nonprimary modes requires some changes. First, we define (τ, ξ) coordinates using c_{+p}^{SW} instead of c_{+p}^{N} as

$$\tau = \int \frac{1}{c_{+p}^{SW}(x)} dx - t, \quad \xi = \delta x. \tag{24'}$$

Second, the definition of $\tilde{R}_{+p}$ is now written as

$$d\tilde{R}_{+p} = \tilde{\lambda}_{+p} d\tilde{Q}_{+p} + \frac{\tilde{\lambda}_{+p}}{c_{+p}^{SW}} \tilde{G}_{+p}(\hat{\mathbf{p}}^{SW}) d\xi, \tag{17'}$$

where the definition of $\tilde{G}_r$ is modified as

$$\tilde{G}_r(\hat{\mathbf{p}}) = \tilde{\mathbf{l}}_r(\hat{\mathbf{p}})^H \left(c_r^N(\hat{\mathbf{p}}) \left(\frac{\partial \hat{\mathbf{p}}}{\partial \xi} \right)_{\tilde{Q}_{+p}, \eta_B} + \hat{\mathbf{G}}(\hat{\mathbf{p}}) \hat{\mathbf{p}} \right). \tag{18'}$$

Third, the integrability condition becomes

$$\left(\frac{\partial \tilde{\lambda}_{+p}}{\partial \xi} \right)_{\tilde{Q}_{+p}, \eta_B} = \left(\frac{\partial}{\partial \tilde{Q}_{+p}} \left(\frac{\tilde{\lambda}_{+p}}{c_{+p}^{SW}} \tilde{G}_{+p}(\hat{\mathbf{p}}^{SW}) \right) \right)_{\xi}. \tag{20'}$$

Note that the τ-derivative does not appear in the last two equations because only $\tilde{Q}_{+p}$ and η_B have direct τ dependence, but these variables are held fixed in the x-derivatives in (18) and (20). Finally, using the above definitions, the evolutionary equations of $\tilde{Q}_r$ and η_B in (τ, ξ) coordinates are obtained by multiplying (48) by $\tilde{\mathbf{l}}_r^H$ from the left and applying the coordinate transformation (24′) and the definitions of $\tilde{Q}_r$ (25). This yields

$$\left(-1 + \frac{c_r^N}{c_{+p}^{SW}} \right) \frac{\partial \tilde{Q}_r}{\partial \tau} \sim -\delta \left(c_r^N \tilde{\mathbf{l}}_r^H \frac{\partial \hat{\mathbf{p}}}{\partial \xi} + \tilde{\mathbf{l}}_r^H \hat{\mathbf{G}}(\hat{\mathbf{p}}) \hat{\mathbf{p}} + \tilde{B}_r(\hat{\mathbf{p}}) \right), \tag{53a}$$

$$- \frac{\partial \eta_B}{\partial \tau} + \epsilon \frac{1}{c_{+p}^{SW}} \frac{\partial}{\partial \tau} \left(u^b \eta_B \right) \sim -\left(1 + \epsilon \hat{\pi}_l^b \frac{\hat{\eta}_l}{\hat{h}_l} \right) u^b \theta$$
$$- \epsilon \delta \frac{\partial}{\partial \xi} \left(u^b \eta_B \right) - \epsilon \delta \frac{1}{2} \left(\frac{d^2 Z/ds^2}{(dZ/ds)^2} u \right)^b \theta \eta_B, \tag{53b}$$

where $u^b = \hat{\pi}_m^b \hat{u}_m$ is the horizontal component of velocity at the bottom and

$$\tilde{B}_r(\hat{\mathbf{p}}) = \tilde{\mathbf{l}}_r(\hat{\mathbf{p}})^H \left(\hat{\mathbf{B}}(\hat{\mathbf{p}}) \hat{\mathbf{p}} - \hat{\mathbf{b}} \frac{\partial \eta_B(\hat{\mathbf{p}})}{\partial \tau} \right). \tag{54}$$

In the definition of $\tilde{G}_r(\hat{\mathbf{p}})$ in (18′) and $\tilde{B}_r(\hat{\mathbf{p}})$ above, the arguments of c_r^N, $\tilde{\mathbf{l}}_r$, and η_B are used to indicate the state vector used to calculate them; in other words, $c_r^N(\hat{\mathbf{p}}^{SW}) = c_r^{SW}$, $c_r^N(\hat{\mathbf{p}}^{SW} + \Delta\hat{\mathbf{p}}) = c_r^{SW} + \Delta c_r$, $\tilde{\mathbf{l}}_r(\hat{\mathbf{p}}^{SW}) = \tilde{\mathbf{l}}_r^{SW}$, $\tilde{\mathbf{l}}_r(\hat{\mathbf{p}}^{SW} + \Delta\hat{\mathbf{p}}) = \tilde{\mathbf{l}}_r^{SW} + \Delta\tilde{\mathbf{l}}_r$, and $\eta_B(\hat{\mathbf{p}}^{SW}) = \eta_B^{SW}$.

For perturbation analysis, we expand the relevant variables in (4), (7), (17′), (20′), (25), and (53) as

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}^{(0)} + \delta \hat{\mathbf{p}}^{(1)} + \delta^2 \hat{\mathbf{p}}^{(2)} + \cdots, \tag{55a}$$

$$\eta_B = \eta_B^{(0)} + \epsilon\delta \eta_B^{(1)} + \cdots, \tag{55b}$$

$$c_r^N = c_r^{(0)} + \delta c_r^{(1)} + \delta^2 c_r^{(2)} + \cdots, \tag{55c}$$

$$\tilde{\mathbf{l}}_r = \tilde{\mathbf{l}}_r^{(0)} + \delta \tilde{\mathbf{l}}_r^{(1)} + \cdots, \tag{55d}$$

$$\hat{\mathbf{K}} = \hat{\mathbf{K}}^{(0)} + \delta \hat{\mathbf{K}}^{(1)} + \delta^2 \hat{\mathbf{K}}^{(2)} + \cdots, \tag{55e}$$

$$\hat{\mathbf{G}} = \hat{\mathbf{G}}^{(0)} + \delta \hat{\mathbf{G}}^{(1)} + \cdots, \tag{55f}$$

$$\hat{\mathbf{B}} = \hat{\mathbf{B}}^{(0)} + \delta \hat{\mathbf{B}}^{(1)} + \cdots, \tag{55g}$$

$$\tilde{Q}_r = \tilde{Q}_r^{(0)} + \delta \tilde{Q}_r^{(1)} + \delta^2 \tilde{Q}_r^{(2)} + \cdots, \tag{55h}$$

$$\tilde{R}_{+p} = \tilde{R}_{+p}^{(0)} + \delta \tilde{R}_{+p}^{(1)} + \delta^2 \tilde{R}_{+p}^{(2)} + \cdots, \tag{55i}$$

$$\tilde{\lambda}_{+p} = \tilde{\lambda}_{+p}^{(0)} + \delta \tilde{\lambda}_{+p}^{(1)} + \cdots, \tag{55j}$$

where only the terms required in the derivation are shown. Note that (52) applies to the 1st- and 2nd-order terms of the relevant variables and that the expansion of η_B is irregular because it comes from the Taylor expansion of the kinematic bottom boundary condition (35). Note also that the perturbation of $\hat{\mathbf{p}}$ causes the perturbations of $\tilde{\lambda}_{+p}$, $\hat{\mathbf{K}}$, $\hat{\mathbf{G}}$, and $\hat{\mathbf{B}}$ and that $\hat{\mathbf{K}}^{(k)}$, $\hat{\mathbf{G}}^{(k)}$, and $\hat{\mathbf{B}}^{(k)}$ depend only on $\hat{\mathbf{p}}^{(k)}$ because these matrices are linear in $\hat{\mathbf{p}}$. Substituting the above expansion into (4), (7), (17′), (25), and (53) and collecting like terms in δ, we get the governing equations for $\tilde{R}_{+p}$, $\tilde{Q}_q$, and η_B to each order in δ. To express the results in dimensional form, the resulting equations and variables are combined to a required order and put back to (x, t) coordinates.

To 1st order in δ, the resulting equations are

$$\left(\frac{\partial}{\partial t} + c_{+p}^N \frac{\partial}{\partial x} \right) \tilde{R}_{+p} = -\tilde{\lambda}_{+p} \tilde{S}_{+p}(\hat{\mathbf{p}}), \tag{56a}$$

$$\frac{\partial \eta_B}{\partial t} - \epsilon \frac{1}{c_{+p}^N} \frac{\partial}{\partial t} \left(u^b \eta_B \right) = -\left(1 + \epsilon \hat{\pi}_l^b \frac{\hat{\eta}_l}{\hat{h}_l} \right) u^b \theta, \tag{56b}$$

where

$$\tilde{S}_{+p}(\hat{\mathbf{p}}) = \delta \Delta \tilde{G}_{+p} + \tilde{B}_{+p}(\hat{\mathbf{p}}), \tag{57a}$$

$$\Delta \tilde{G}_{+p} = \tilde{G}_{+p}(\hat{\mathbf{p}}) - \tilde{G}_{+p}(\hat{\mathbf{p}}^{SW}). \tag{57b}$$

Note that $\tilde{Q}_q = 0$, $\hat{\mathbf{p}} = \hat{\mathbf{p}}^{SW}$, $c_{+p}^N = c_{+p}^{SW}$, and $\eta_B = \eta_B^{SW}$ to this order, and $\hat{\mathbf{p}}^{SW}$ is given by

$$\hat{\mathbf{p}}^{SW} = \int_0^t \tilde{\mathbf{r}}_{+p} \left(\tilde{Q}_{+p}(x, t'), x \right) \frac{\partial \tilde{Q}_{+p}}{\partial t'} dt', \tag{58}$$

where no sum is taken on $+p$.

Now, the problem is essentially reduced to a forced one-directional propagation equation for $\tilde{R}_{+p}$, (56a). This is because η_B is fully determined by the initial condition $\eta_B = 0$ and

$\tilde{R}_{+p}(x,t)$ and hence not an independent variable. This is clearer for an initial-value problem without boundaries, in which η_B becomes a diagnostic variable using $\partial/\partial t \sim -c^N_{+p}(\partial/\partial x)$ in (56b); however, we prefer keeping $\partial/\partial t$ because we consider boundary-value problems in example applications in Sec. III E. Equation (56a) is the continuous-stratification-version of (14) for $r = +p$ and has additional diabatic terms on the RHS. Note that the diabatic term disappears in the limit of weak nonlinearity, which recovers the weakly nonlinear simple-wave theory in Sec. 2.13 of Ref. 28. [To see this, note that the terms with ϵ can be neglected in (56b) under the assumption that $\epsilon = \delta$ and that $\partial/\partial\tau = -\partial/\partial t$ in (54).] Although it is unclear in the 1st-order theory, the diabatic term represents the conversion of the primary Riemann mode to nonprimary modes, which becomes clear in the 2nd-order theory below.

To 2nd order in δ, the resulting equations are

$$\left(\frac{\partial}{\partial t} + c^N_{+p}\frac{\partial}{\partial x}\right)\tilde{R}_{+p} = -\tilde{\lambda}_{+p}\tilde{S}_{+p}(\hat{\boldsymbol{p}}), \tag{59a}$$

$$\left(1 - \frac{c^{SW}_q}{c^{SW}_{+p}}\right)\frac{\partial \tilde{Q}_q}{\partial t} = -\tilde{S}_q(\hat{\boldsymbol{p}}^{SW}), \tag{59b}$$

$$\frac{\partial \eta_B}{\partial t} + \epsilon\frac{\partial}{\partial x}\left(u^b\eta_B\right) = -S_{\eta_B}(\hat{\boldsymbol{p}}), \tag{59c}$$

where $\hat{\boldsymbol{p}}^{SW}$ is defined by (58), $\tilde{S}_{+p}$ is defined by (57a), $\Delta\hat{\boldsymbol{p}}$ is defined by

$$\Delta\hat{\boldsymbol{p}} = \int_0^t \tilde{\boldsymbol{r}}_q\left(\tilde{Q}_{+p}(x,t'),x\right)\frac{\partial \tilde{Q}_q}{\partial t'}dt', \tag{60}$$

and $\tilde{S}_q$ and S_{η_B} are defined by

$$\tilde{S}_q(\hat{\boldsymbol{p}}) = \tilde{G}_q(\hat{\boldsymbol{p}}) + \tilde{B}_q(\hat{\boldsymbol{p}}), \tag{61a}$$

$$S_{\eta_B}(\hat{\boldsymbol{p}}) = \left(1 + \epsilon\hat{\pi}^b_l\frac{\hat{\eta}_l}{\hat{h}_l}\right)u^b\theta + \epsilon\delta\frac{1}{2}\left(\frac{d^2Z/ds^2}{(dZ/ds)^2}u^{SW}\right)^b\theta\eta^{SW}_B. \tag{61b}$$

Also, Δc_{+p} to the 2nd order and $\Delta\tilde{\boldsymbol{l}}_{+p}$ to the 1st order are required in the calculation. They are calculated from

$$\Delta c_{+p} = (\tilde{\boldsymbol{l}}^{SW}_{+p})^H\Delta\hat{\boldsymbol{K}}\tilde{\boldsymbol{r}}^{SW}_{+p}, \tag{62a}$$

$$\Delta\tilde{\boldsymbol{l}}_{+p} = \tilde{a}_{+p,r}\tilde{\boldsymbol{l}}^{SW}_r, \tag{62b}$$

$$\tilde{a}_{r,s} = \frac{1}{c^{SW}_r - c^{SW}_s}(\tilde{\boldsymbol{l}}^{SW}_r)^H\Delta\hat{\boldsymbol{K}}\tilde{\boldsymbol{r}}^{SW}_s, \tag{62c}$$

$$\Delta\hat{\boldsymbol{K}} = \epsilon\delta\begin{bmatrix}\hat{N}_{lmn}\Delta\hat{u}_l & \hat{N}_{mln}\Delta\hat{\eta}_l \\ 0 & \hat{N}_{lnm}\Delta\hat{u}_l\end{bmatrix}, \tag{62d}$$

where $\Delta\hat{\eta}_l$ and $\Delta\hat{u}_l$ are calculated as the components of $\Delta\hat{\boldsymbol{p}}$. In principle, these relationships are obtained by applying the perturbation expansion (55) to (4); however, the expansion is irregular because only $\hat{\boldsymbol{p}}^{(0)}$ and $\hat{\boldsymbol{p}}^{SW(1)}$ are available to the 1st order and up to $\hat{\boldsymbol{p}}^{SW(2)}$ and $\Delta\hat{\boldsymbol{p}}^{(1)}$ to the 2nd order. The above equations show that topographic effects excite all the Riemann modes to the 2nd order because $\tilde{Q}_q \neq 0$ for all $q \neq +p$. Note that there is some arbitrariness in converting the perturbation equations to dimensional form because variables in the highest-order terms can be either only the primary-mode parts (e.g., $\hat{\boldsymbol{p}}^{SW}$ and η^{SW}_B) or the full expressions (e.g., $\hat{\boldsymbol{p}}$ and η_B). We choose the former in this study. With this choice, $\tilde{Q}_q$ are not independent (in a way diagnostic) variables because they are fully determined by the initial condition $\tilde{Q}_q = 0$ and $\tilde{R}_{+p}(x,t)$. For this reason, the integrability issue does not arise for non-zero $\tilde{Q}_q$, and a simple-wave approach is applicable with 2nd-order topographic effects. In contrast to $\tilde{Q}_q$, η_B becomes independent from $\tilde{R}_{+p}$ to this order. So the degrees of freedom in terms of the wave field increases from 1 (i.e., $\tilde{R}_{+p}$) in the 1st-order theory to 2 (i.e., $\tilde{R}_{+p}$ and η_B) in the 2nd-order theory. To the 2nd order, the problem essentially consists of a forced one-directional propagation equation for $\tilde{R}_{+p}$, (59a), and a forced advection equation for η_B, (59c).

In this study, solutions to (56) or (59) are called topography-modified simple waves, in order to distinguish them from adiabatic simple waves.

E. Reduction to weakly nonlinear theory

Reducing the above theory to a weakly nonlinear theory is useful for comparisons with KdV-type theories and the derivation of analytic solutions. To do so, we assume that $\epsilon = \delta^2$ in (56) and (59) in this subsection. For a weakly nonlinear problem, it is more convenient to change indices for Riemann modes from r and s to $\pm n$ and $\pm m$, where n and m are the indices for vertical modes and + and – signs indicate the right- and left-propagating Riemann modes, respectively.

The base adiabatic solution is the following weakly nonlinear solution. To 0th order in δ, each vertical mode follows equations equivalent to the shallow water equations (3a), but c, H, η, and u are replaced by c_n, $\hat{h}_n$, $\hat{\eta}_n$, and $\hat{u}_n$, respectively. The eigenvalues and eigenvectors of $\hat{\boldsymbol{K}}$ are given by (5) and (6) with $\eta = u = 0$ for each vertical mode, and $\tilde{\lambda}_{+p}$ is given by by (23). We also need $O(\epsilon = \delta^2)$ nonlinear correction to the propagation speed, which is available from the power-series expansion of simple-wave solution for a flat bottom.[30] Then, the base adiabatic solution is

$$c^{SW}_{+p} = c_p + \frac{\hat{\alpha}_p\hat{h}_p}{2c_p}\tilde{Q}_p, \tag{63a}$$

$$(\hat{\boldsymbol{p}}^{SW})_m = \frac{1}{2}\begin{bmatrix}\hat{h}_p/c_p \\ 1\end{bmatrix}\tilde{Q}_{+p}\delta_{m,p}, \tag{63b}$$

$$\tilde{\lambda}_{+p} = A\sqrt{c_p\hat{h}_p}, \tag{63c}$$

$$\tilde{R}_{+p} = \tilde{\lambda}_{+p}\tilde{Q}_{+p}, \tag{63d}$$

$$\hat{\alpha}_p = \frac{3c_p}{2\hat{h}_p}\hat{N}_{ppp}. \tag{63e}$$

Note that no sum is taken on p because the pth mode is the single vertical mode of interest. The parentheses with the subscript m are used to indicate mth and $(m + n_M)$th components of $\hat{\boldsymbol{p}}^{SW}$ (a $2n_M \times 1$ column vector), where n_M is the number of

vertical modes used in the calculation. Substituting this into (56b) with $\epsilon = 0$, we get

$$\frac{\partial \eta_B^{SW}}{\partial t} = -\frac{1}{2}\hat{\pi}_p^b \theta \tilde{Q}_{+p}. \tag{64}$$

The problems up to 2nd order in δ are solved as follows. To evaluate $\Delta\hat{\mathbf{p}}$ to the 1st order, we assume that $\eta_B = \eta_B^{SW}$ and use the 0th-order solution to calculate $\tilde{G}_{\pm n}$ and $\tilde{B}_{\pm n}$, yielding

$$\tilde{G}_{\pm n} = \frac{1}{2}\left(\frac{dc_p}{dx} \pm \frac{c_p}{\hat{h}_p}\frac{d\hat{h}_p}{dx} + (1 \mp 1)\hat{L}_{pp}c_p\right)\tilde{Q}_{+p}\delta_{n,p}, \tag{65a}$$

$$\tilde{B}_{\pm n} = \frac{1}{2}\left(\mp\hat{L}_{pn}\frac{\hat{h}_p}{\hat{h}_n}c_n + \hat{L}_{np}c_p - (1 \mp 1)\hat{L}_{pp}c_p\delta_{n,p}\right)\tilde{Q}_{+p}. \tag{65b}$$

Then, using (B4), (59b) yields

$$\frac{\partial \tilde{Q}_{\pm n}}{\partial t} = \mp\frac{c_p^2}{2c_n}\frac{(c_p \pm c_n)^2}{c_p^2 - c_n^2}\hat{L}_{np}\tilde{Q}_{+p} \qquad \text{for } n \neq p, \tag{66a}$$

$$\frac{\partial \tilde{Q}_{-p}}{\partial t} = \frac{1}{4}\frac{\partial c_p}{\partial x}\tilde{Q}_{+p}. \tag{66b}$$

Note that we need $\tilde{Q}_{-p}$ to evaluate $\Delta\hat{\mathbf{p}}$. Using $\Delta\hat{\mathbf{p}}$ from (60), (59c) to 1st order in δ becomes

$$\frac{\partial \eta_B}{\partial t} = -\frac{1}{2}\hat{\pi}_p^b\theta\tilde{Q}_{+p} - \left(\frac{1}{8}\hat{\pi}_p^b\frac{dc_p}{dx} - \sum_{n\neq p}\hat{\pi}_n^b\frac{c_p^2}{c_p^2 - c_n^2}\hat{L}_{np}c_p\right)\theta\int\tilde{Q}_{+p}dt, \tag{67}$$

and $\Delta\tilde{G}_{+p}$ and $\tilde{B}_{+p}$ to the 1st order become

$$\Delta\tilde{G}_{+p} = \frac{1}{8}\frac{dc_p}{dx}\frac{dc_p}{dx}\int\tilde{Q}_{+p}dt, \tag{68a}$$

$$\tilde{B}_{+p} = -\frac{c_p^2}{2}\sum_{n\neq p}\frac{c_p^2 + 3c_n^2}{c_p^2 - c_n^2}\hat{L}_{pn}\hat{L}_{np}\int\tilde{Q}_{+p}dt. \tag{68b}$$

Since η_B is determined by $\tilde{Q}_{+p}$ and the initial condition $\eta_B = 0$, it is not an independent variable in this weakly nonlinear problem. Also, note that diabatic corrections to the propagation speed Δc_{+p} and the left-eigenvector $\Delta\tilde{\mathbf{l}}_r$ are $O(\delta^3)$ and hence negligible. Therefore, to 2nd order in δ, we get the following single equation for $\tilde{R}_{+p}$ from (59a):

$$\left(\frac{\partial}{\partial t} + \left(c_p + \frac{\hat{\alpha}_p\hat{h}_p}{2\tilde{\lambda}_{+p}c_p}\tilde{R}_{+p}\right)\frac{\partial}{\partial x}\right)\tilde{R}_{+p} = -c_p\hat{\nu}_p\int\tilde{R}_{+p}dt, \tag{69a}$$

$$\hat{\nu}_p = \frac{1}{8c_p}\frac{dc_p}{dx}\frac{dc_p}{dx} - \frac{c_p}{2}\sum_{n\neq p}\frac{c_p^2 + 3c_n^2}{c_p^2 - c_n^2}\hat{L}_{pn}\hat{L}_{np}, \tag{69b}$$

where (B6) is used to simplify the equation.

The above equation is a topography-modified KdV-type equation. To see this, we use $\tilde{R}_{+p} = 2\tilde{\lambda}_{+p}c_p\hat{h}_p^{-1}\hat{\eta}_p + O(\delta^2)$ and $\partial/\partial t = -c_p\partial/\partial x + O(\delta^2)$ and assume that $\epsilon \sim \delta^2 \sim \mu \equiv D^2/L^2$. Note that we can simply add the linear nonhydrostatic dispersion term of the KdV equation because the term is of $O(\mu)$ and determined independent of nonlinear and topographic effects (ϵ and δ) to the leading order. After these manipulations, we get

$$\left(\frac{\partial}{\partial t} + \left(c_p + \hat{\alpha}_p\hat{\eta}_p\right)\frac{\partial}{\partial x} + \hat{\beta}_p\frac{\partial^3}{dx^3}\right)\sqrt{\frac{c_p^3}{\hat{h}_p}}\hat{\eta}_p = \hat{\nu}_p\int\sqrt{\frac{c_p^3}{\hat{h}_p}}\hat{\eta}_p dx, \tag{70}$$

where $\hat{\beta}_p$ is the 1st-order nonhydrostatic dispersion coefficient. The basic topography-modified KdV equation,[21–23,37] which is 1st order in ϵ, δ, and μ, is recovered by neglecting the RHS term, which provides the 2nd-order topographic correction.

IV. EXAMPLE APPLICATIONS

To illustrate important features of topography-modified, fully nonlinear simple waves, we focus on the fastest right-propagating internal waves ($p = 1$), use the undisturbed height coordinate $s = Z$, and consider two example applications in this section. The first example is the nonlinear steepening of sinusoidal internal waves over a linear slope, which illustrates similarities and differences compared to the process over a flat bottom. The second example is Wunsch's subcritical wedge problem.[53] This problem is important because analytic linear solution is available for arbitrary slopes, although its practical use is limited to subcritical slopes.[54] We use the analytic solution to estimate the error in the proposed 2nd-order theory and then illustrate additional features due to combined effects of nonlinearity and topography.

Since the focus of this study is on the development of theory rather than a numerical model, numerical methods used in this study were kept rather basic (see Appendix E). Although the theory presented in Sec. III is fully nonlinear, internal-wave amplitudes in the examples are restricted by two factors. First, we do not consider such large amplitudes that lead to shock formation within the model domain. This is because shock conditions in continuously stratified fluids remain to be investigated, and shocks tend to disappear if wave dispersion is added to the theory (e.g., in order to develop an ISW model). Second, we assume that isopycnal "layers" have nonzero thicknesses [i.e., $(d\rho/ds)^{-1}(\partial(Z + \eta)/\partial s) < 0$]. Some isopycnal layer thicknesses could reach zero under full nonlinearity, and it appears possible to continue numerical integration beyond this point. However, we leave the handling of zero layer thickness to future studies because it requires details that are important for numerical modeling but not for the theoretical development in this study.

A. Background conditions and boundary forcing

Bathymetry and coordinates used in example applications in this section are shown in Fig. 3. We use a uniform bottom slope with the gradient θ. Water depths at the incoming and outgoing boundaries are H and $H/2$, respectively. Considering typical water-depth ranges where shoaling deformation of internal waves occurs on continental shelves,[3,22,25] we use $H = 160$ m and θ in the range of 10^{-3}–10^{-2}. The origin of the horizontal coordinate is located at the incoming boundary, and horizontal length of the domain $H/(2\theta)$ is 8–80 km. Note that

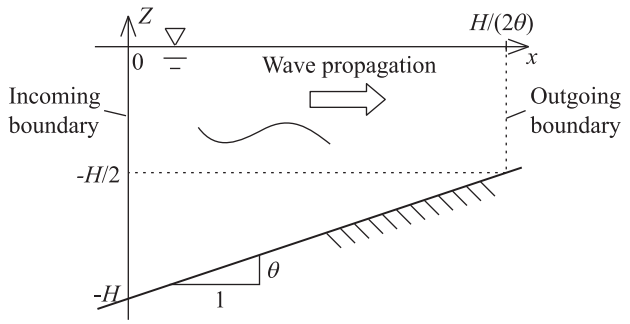

FIG. 3. Bathymetry and coordinates used in example applications. H = 160 m is used throughout this paper.

the bottom slope is relatively mild, but the total depth change (80 m) is large.

We consider the hyperbolic-tangent and linear stratification. Figure 4 shows the vertical profiles of density deviation from the background and the associated vertical modes. The pycnocline depth and width of the hyperbolic tangent stratification [h and d in (25) in Ref. 30] are 54 and 17 m, respectively. Note that as the vertical modes vary horizontally, so do parameters associated with the vertical modes (Fig. 5). For example, the celerities c_n decrease with decreasing water depth [Figs. 5(a) and 5(e)], and shapes of the vertical modes change depending on the relative location of pycnocline within the water column [Figs. 4(b), 4(c), 5(c), and 5(d)]. The conversion factor between the modal displacement amplitude, $\hat{\eta}_1$,

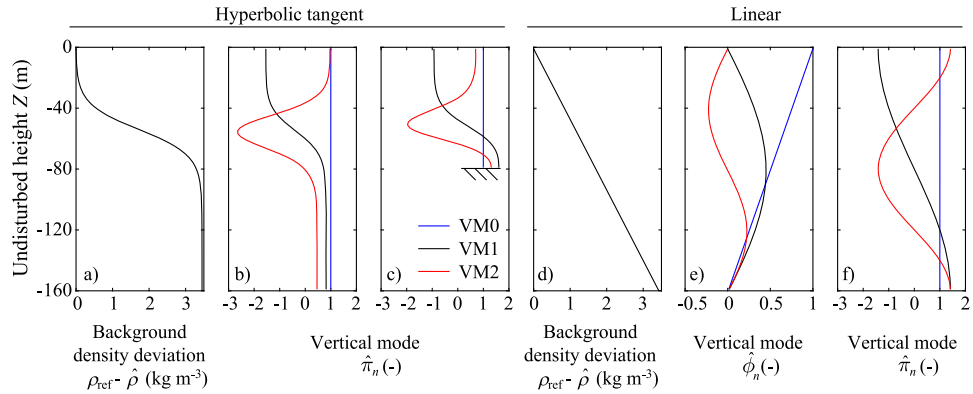

FIG. 4. Stratification and vertical modes used in example applications. [(a) and (d)] Deviation of background density from the reference value of $\hat{\rho}$ = 1000 kg m^{-3}, [(b) and (f)] vertical modes $\hat{\pi}_n$ at the incoming boundary (x = 0), (c) $\hat{\pi}_n$ at the outgoing boundary (x = $H/(2\theta)$), and (e) vertical modes $\hat{\phi}_n$ at the incoming boundary. The density difference between the surface and bottom is the same for (a) and (d) and corresponds to the temperature difference of ≈13 K for seawater with a salinity of 0.035 kg m^{-3}. For linear stratification, vertical modes are stretched or compressed vertically depending on water depth, but otherwise the shapes remain the same as in [(e) and (f)].

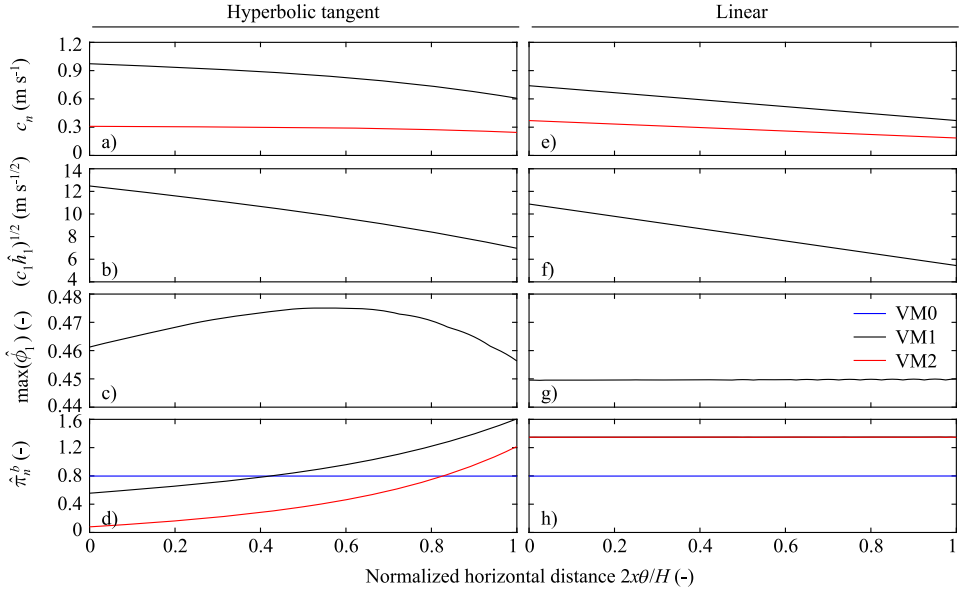

FIG. 5. Horizontal variation of selected parameters related to vertical modes. [(a) and (e)] Celerities of VM1 and VM2, [(b) and (f)] topographic factor $\tilde{\lambda}_{+1}$ in the linear limit, [(c) and (g)] maximum of $\hat{\phi}_1$ within the water column, and [(d) and (h)] $\hat{\pi}_n$ for VM0, VM1, and VM2 at the bottom. Note that the black line in (h) is overlaid by a red line.

and the maximum (physical) isopycnal displacement within the water column, $\max(|\eta|)$, is given by $\max(\hat{\phi}_1)$. It is essentially constant for the linear stratification, but not constant for the hyperbolic-tangent stratification under the vertical-mode normalization used in this study [Figs. 5(c) and 5(g)].

Parameters related to adiabatic simple waves are calculated based on water depth, stratification, and the associated vertical modes, but without specifying forcing (Fig. 6). Simple-wave solutions without topographic effects are calculated from (4) and (8) but using $\hat{\mathbf{p}}$ and $\hat{\mathbf{K}}$ defined in (50). The solutions provide modal amplitudes of isopycnal displacements $\hat{\eta}_n$, horizontal velocities $\hat{u}_n$, and the associated nonlinear propagation speed c_r^{SW} for all the Riemann modes as a function of the incomplete Riemann variable for the primary mode $\tilde{Q}_{+1}$. Note that $\hat{\eta}_1$ and $\hat{u}_1$ have monotonic relationships with $\tilde{Q}_{+1}$ [Figs. 6(a) and 6(b)]. c_{+1}^{SW} is almost independent of $\tilde{Q}_{+1}$ for the linear stratification, but shows maxima for the hyperbolic-tangent stratification [Figs. 6(c) and 6(d)]. Since the maxima occur approximately when the pycnocline is at the mid-depth, the maxima occur at positive $\tilde{Q}_{+1}$ (i.e., elevated isopycnals) for water depth shallower than 108 m, but negative $\tilde{Q}_{+1}$ deeper than 108 m. Then, using simple-wave solutions without topographic effects, (20′) is solved to determine the topographic factor $\tilde{\lambda}_{+1}$ [Figs. 6(e) and 6(f)], which is needed to convert $\tilde{Q}_{+1}$ to the (complete) Riemann variable $\tilde{R}_{+1}$. In the numerical solutions, left Riemann modes $\tilde{\mathbf{l}}_r$ are normalized to account for the variation of $\tilde{\lambda}_{+1}$ in the linear limit $\sqrt{c_p\hat{h}_p}/\sqrt{(c_p\hat{h}_p)_{\mathrm{ref}}}$, using the incoming boundary as a reference location (see Appendix E). Therefore, $\tilde{\lambda}_{+1} = 1$ at the incoming boundary ($x = 0$) and deviates from 1 with decreasing water depth due to combined effects of nonlinearity and topography.

Boundary conditions are set as follows. Incident waves at $x = 0$ are assumed to be sinusoidal with a fixed period of 12 h to represent semi-diurnal internal tides. In numerical solutions, incident-wave amplitude is gradually increased from 0 within the first 12 h. No outgoing boundary condition is required for $\tilde{R}_{+1}$ because (56a) and (59a) are one-directional wave-propagation equations. To 2nd order in δ, we do need boundary conditions for η_B because of the advection term in (59b), but the theory does not appear to dictate how to set them. In this study, we exploit the fact that, under oscillatory flows considered in this study, the boundary conditions tend

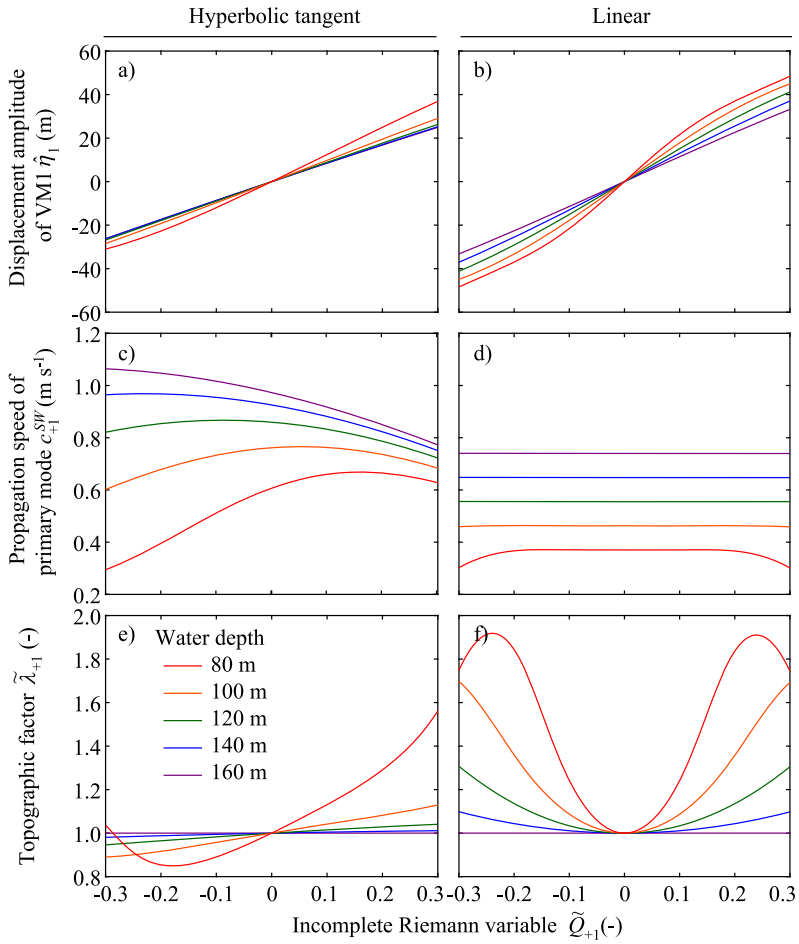

FIG. 6. Selected parameters related to fastest, right-propagating adiabatic simple internal waves. [(a) and (b)] Displacement amplitude of VM1 $\hat{\eta}_1$, [(c) and (d)] propagation speed of primary Riemann mode c_{+1}^{SW}, and [(e) and (f)] topographic factor $\tilde{\lambda}_{+1}$.

to affect only the regions within one wave-induced horizontal particle excursion from the boundaries, typically 1-6 km in the following applications. Therefore, we use buffer zones to avoid artifacts due to the boundary conditions for η_B.

B. Nonlinear steepening of sinusoidal waves

In this example, we use the hyperbolic-tangent stratification in Fig. 4(a) and a mild bottom slope of $\theta = 10^{-3}$. Incident-wave (semi-) amplitude is $\hat{\eta}_1 = 14$ m [or 6.5 m in maximum isopycnal displacement within the water column, using the conversion factor $\max(\hat{\phi}_1)$ in Fig. 5(c)]. The amplitude is chosen so that shocks are not formed within the model domain.

To compare the results with the corresponding flat-bottom case, we first look at the adiabatic solution, which is a solution to (59a) without the RHS term (Fig. 7). Since $\tilde{R}_{+1}$ remains constant along the characteristic, the problem can be solved by the characteristic method;[27,28] however, note that the wave propagation speed c_{+1}^{SW} changes with water depth (compare Figs. 1 and 7). Wave steepening occurs behind the crests for $x < 50$ km because the wave troughs propagate faster there, but in front of the crests for $x > 50$ km where the crests propagate faster. Although $\tilde{R}_{+1}$ remains constant, the amplitude in $\hat{\eta}_1$ increases adiabatically as the waves shoal [dashed lines in Fig. 7(a)].

The adiabatic solution provides a good approximation near the incoming boundary, but diabatic topographic effects become important as the waves shoal. Figures 8 and 9 show a numerical solution for fully nonlinear simple waves with 2nd-order topographic effects. In Fig. 8, $\tilde{R}_{+1}$ is shown together with the characteristics for adiabatic simple waves from Fig. 7. The adiabatic solution provides a good approximation for $x \lessapprox 50$ km, but diabatic effects are apparent in shallower water. Figure 9 shows a time slice of the solution. The propagation speed due to the primary Riemann mode (c_{+1}^{SW}) shows substantial nonlinear fluctuation [compare with the black line in Fig. 5(a)]. The adjustment due to the nonprimary modes (Δc_{+1}) is overall small. As the waves shoal, the modal velocity of VM1 deviates from the adiabatic solution, and the contributions from VM0 and higher baroclinic modes increase [Fig. 9(b)]. As a result, the two-dimensional velocity field deviates from the simple VM1 structure for $x \gtrapprox 50$ km [Fig. 9(d)].

An interesting feature of the fully nonlinear solution with 2nd-order topographic effects is the relatively rapid development of mean components. Figure 10 shows time-series of the modal amplitudes at $x = 60$ km and the corresponding mean vertical profiles over a wave period. Note that $\hat{\eta}_n^{SW}$ and $\hat{u}_n^{SW}$ for

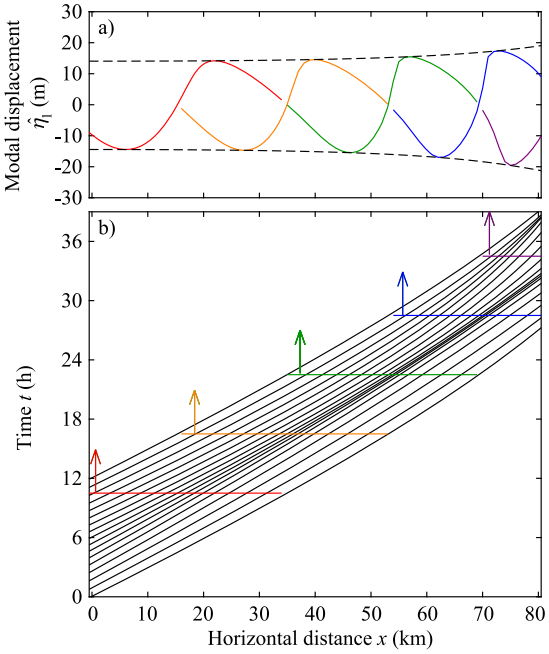

FIG. 7. Adiabatic simple internal waves in the signaling problem under hyperbolic-tangent stratification [Fig. 4(a)] and bottom slope $\theta = 10^{-3}$. Boundary forcing at $x = 0$ is sinusoidal with an amplitude of $\hat{\eta}_1 = 14$ m. Panel (b) shows characteristics, and panel (a) shows spatial slices of $\hat{\eta}_1$ at times indicated by colored lines in (b). Dashed lines in (a) indicate variations of maximum and minimum $\hat{\eta}_1$ calculated from the adiabatic relationship between $\tilde{R}_{+1}$ and $\hat{\eta}_1$. To convert $\hat{\eta}_1$ to maximum isopycnal displacements within the water column, use the factor in Fig. 5(c).

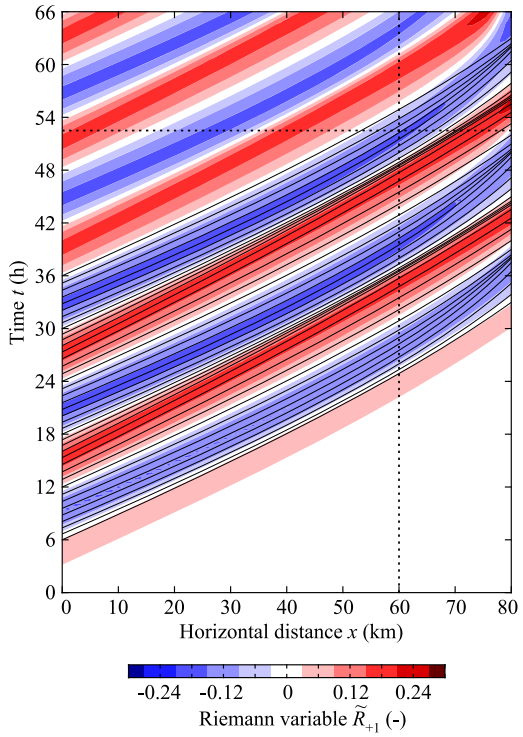

FIG. 8. Comparison of adiabatic and topography-modified simple-wave solutions in the signaling problem shown in Fig. 7. Solid lines show characteristics of adiabatic simple waves from Fig. 7(b), and color shows the (complete) Riemann variable $\tilde{R}_{+1}$ from numerical solution for simple waves with 2nd-order topographic effects. Note that incident-wave amplitude increases gradually within the first 12 h in the numerical solution. Dotted lines indicate the time and location shown in Figs. 9 and 10, respectively.

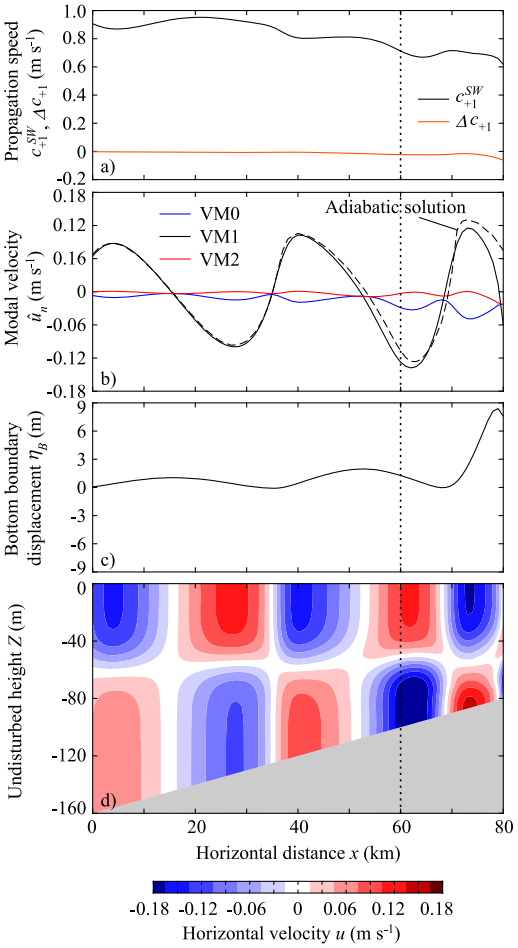

FIG. 9. Fully nonlinear simple internal waves with 2nd-order topographic effects for the signaling problem shown in Figs. 7 and 8. (a) Wave propagation speed due to primary mode c_{+1}^{SW} and that due to nonprimary modes Δc_{+1}, (b) modal velocities $\hat{u}_n$ of the lowest 3 vertical modes, (c) bottom boundary displacement η_B, and (d) horizontal velocity field u at t = 52.5 h. The dashed line in (b) shows adiabatic solution for $\hat{u}_1$. Vertical dotted lines indicate the location shown in Fig. 10.

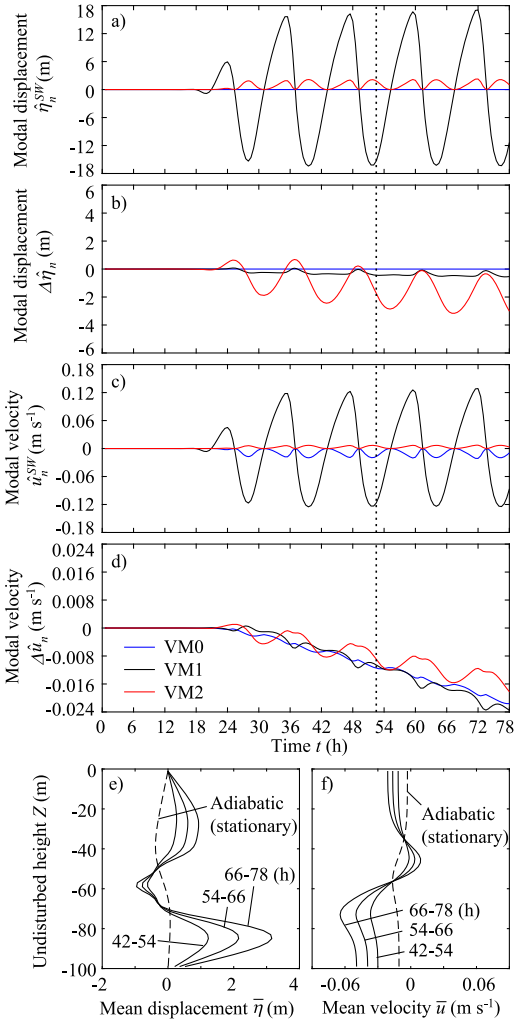

FIG. 10. Evolution of isopycnal displacements and velocities at 100-m water depth in the signaling problem shown in Figs. 7–9. (a) Modal displacements of the lowest 3 vertical modes due to primary Riemann mode $\hat{\eta}_n^{SW}$, (b) those due to the sum of nonprimary modes $\Delta\hat{\eta}_n$, (c) modal velocities of the lowest 3 vertical modes due to primary Riemann mode $\hat{u}_n^{SW}$, (d) those due to the sum of nonprimary modes $\Delta\hat{u}_n$, and [(e) and (f)] vertical profiles of isopycnal displacements and velocities averaged over a wave period. Dotted lines in (a)–(d) indicate the time shown in Fig. 9. Dashed lines in (e) and (f) show results for adiabatic simple waves. They are stationary in time.

VM0 and VM2 tend to be either positive or negative with half the period compared to VM1 because they are excited by nonlinearity in the adiabatic solution [Figs. 10(a) and 10(c)]. This indicates the isopycnal setup (or set-down) and mean flow. In addition, diabatic effects cause slow growths of $\hat{\eta}_n^{SW}$ and $\hat{u}_n^{SW}$. By contrast, $\Delta\hat{\eta}_n$ and $\Delta\hat{u}_n$ primarily show growths, except for $\Delta\hat{\eta}_2$ [Figs. 10(b) and 10(d)]. The development of mean components is the reason why the deviation from the adiabatic solution increases with time for $x \gtrapprox 50$ km in Fig. 8. These results suggest that diabatic topographic effects tend to strengthen wave-induced isopycnal setup and mean flow.

C. Wunsch's subcritical wedge problem

In this example, we use the linear stratification in Fig. 4(d) and vary the bottom slope from the relatively mild value of $\theta = 10^{-3}$ used in the previous example, to the critical slope $\theta = 10^{-2}$. We first derive an analytic linear solution based on the proposed theory and compare the results with the analytic solution by Wunsch.[53] Then, we calculate fully nonlinear solutions for two cases. The first case is for a relatively steep subcritical slope of $\theta = 0.0066$ and incident-wave (semi-) amplitude of $\hat{\eta}_1 = 2.2$ m [or 1.0 m in maximum isopycnal displacement within the water column, using the conversion factor $\max(\hat{\phi}_1)$ in Fig. 5(g)]. This rather small incoming-wave amplitude was necessary to keep isopycnal-layer thickness

nonzero. The second case is for the bottom slope and incident wave amplitude used in the example with hyperbolic-tangent stratification, i.e., $\theta = 10^{-3}$ and incident-wave (semi-) amplitude of $\hat{\eta}_1 = 14$ m.

Assuming linear stratification with small vertical density variation (i.e., the Boussinesq approximation), we can derive a linear analytic solution for Wunsch's subcritical wedge problem with 2nd-order topographic effects, as follows. In the derivation, we consider a linear slope in Fig. 3, but assume that the origin of the x coordinate is located at the virtual apex where water depth is 0, for brevity. Following Wunsch,[53] we also assume a rigid lid and neglect the barotropic mode. Then, water depth, vertical modes, and the associated topographic interaction coefficients and 1st-order nonlinear coefficient are

$$H(x) = -\theta x, \tag{71a}$$

$$c^{SW}_{\pm n}(x) = \mp\frac{N\theta}{n\pi}x \qquad (n \neq 0), \tag{71b}$$

$$\hat{\pi}_n(x,s) = (-1)^n\sqrt{2}\cos\left(-n\pi\frac{Z}{\theta x}\right) \quad (n \neq 0), \tag{71c}$$

$$\hat{L}_{nm}(x) = -\frac{2n^{-2}}{n^{-2}-m^{-2}}\frac{1}{x} \quad (n \neq m;\ n,m \neq 0), \tag{71d}$$

$$\hat{\alpha}_p(x) = 0, \tag{71e}$$

where $\hat{h}_n = H(x)$ is assumed. The solution to (69) has the form

$$\tilde{R}_{+p} = (\theta x)^{i\gamma}\exp(-i\omega t). \tag{72}$$

Substituting this into (69), we get

$$\gamma = -\frac{2p\pi}{\delta}\left(1 - B\left(\frac{\delta}{2p\pi}\right)^2\right), \tag{73a}$$

$$\delta = \frac{2N\theta}{\omega}, \tag{73b}$$

$$B = \frac{1}{8} + 2\sum_{n\neq 0,p}\frac{p^{-2}+3n^{-2}}{(p^{-2}-n^{-2})^3}p^{-2}n^{-2}. \tag{73c}$$

Although wavelengths in the solution continuously change in x, we may follow Wunsch[53] and introduce a local horizontal wavenumber $k_x = \gamma x^{-1}$. Assuming the local horizontal length scale of $L = 2\pi k_x^{-1}$, the above δ is equivalent to δ used in the scaling in Sec. III for VM1. Note that $\delta = 2$ for the critical slope. Setting $B = 0$ yields the corresponding solution with 1st-order topographic effects. To calculate the wave field associated with this solution, we need to calculate $\tilde{Q}_q$ from (66), $\hat{\boldsymbol{p}}$ from $\tilde{R}_{+p}$ and $\tilde{Q}_q$ using (58), (60), and (63d), and then η, m, and u from (43).

Comparisons of the above 1st- and 2nd-order solutions with Wunsch's analytic solution suggest that the proposed theory is applicable to subcritical slopes (Fig. 11). The behavior of the primary Riemann mode, which is characterized by a single parameter γ in the above analytic solution, is captured reasonably well by the 1st-order solution [Fig. 11(d)]. The main role of the 2nd-order topograhic effects is the diabatic corrections to the wave fields by the nonprimary

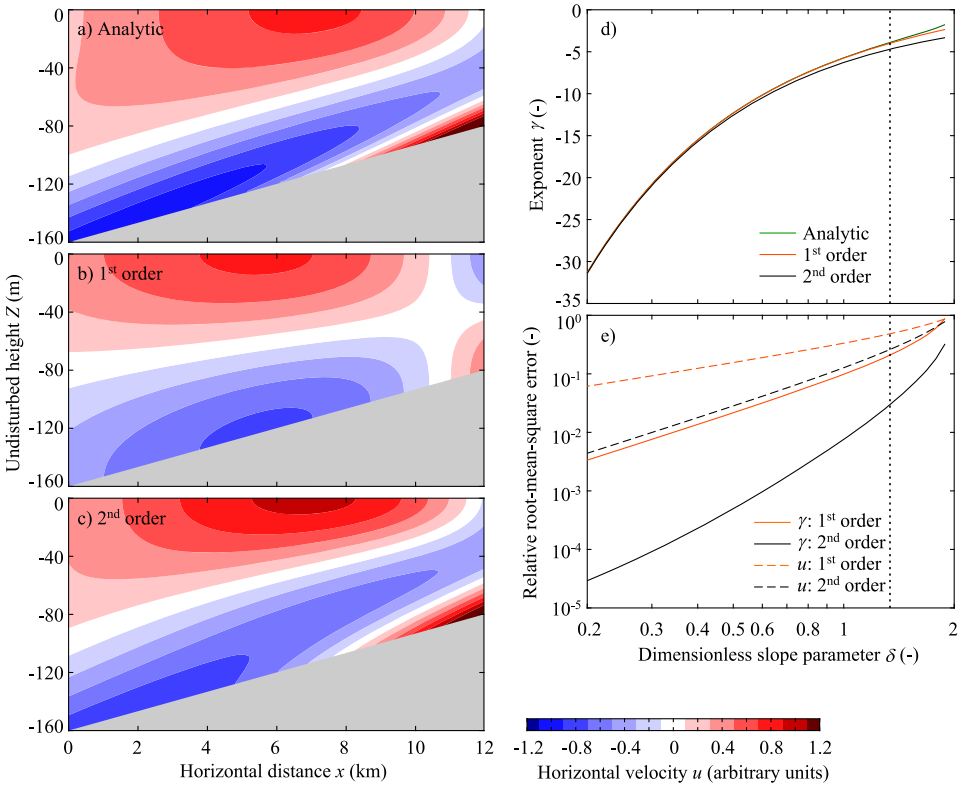

FIG. 11. Comparisons of analytic solution and linear-wave solutions with 1st- and 2nd-order topographic effects for Wunsch's subcritical wedge problem. Stratification is linear [Fig. 4(d)], and the bottom slope is $\theta = 0.0066$ (or dimensionless slope $\delta = 4/3$). (a) Horizontal velocity field from analytic solution and [(b) and (c)] those from linear-wave solutions with 1st- and 2nd-order topographic effects, respectively. (d) Comparisons of exponent γ in (72) and (e) relative root-mean-square errors (RMSEs) of γ and horizontal velocity field u. Relative RMSEs of γ are calculated using the analytic solution as reference, and those of u are calculated as the integral of squared error over the area shown in (a)–(c) normalized by the corresponding integral of squared horizontal velocity from the analytic solution. Dotted lines in (d) and (e) indicate δ used in (a)–(c).

modes. When the bottom slope is close to the critical, the 1st-order solution cannot capture a ray-like feature in the wave field because the vertical structure is given by a single vertical mode [Figs. 11(a) and 11(b)]. The 2nd-order solution captures the main features of Wunsch's analytic solution for δ as large as 1.3 [Fig. 11(c)]. Unavoidably, the error increases as the bottom slope approaches the critical slope [δ = 2 in Fig. 11(e)]. This suggests that the proposed theory based on the 2nd-order perturbation expansion is applicable to subcritical slopes despite the mild slope assumption ($\delta \ll 1$) in the derivation. The comparisons also confirm that the proposed theory is applicable to large depth change.

For the case shown in Fig. 12, the nonlinear solution is qualitatively similar to the linear solution in the oscillating components, but different in the mean components (Figs. 12 and 13). Similarity in the oscillating components originates from the fact that the incident-wave amplitude is rather small and that the wave propagation speed remains almost independent of the wave amplitude [Fig. 6(d)]. This is in contrast to the previous case with hyperbolic-tangent stratification, in which the amplitude dependence causes nonlinear steepening. Since the bottom slope considered in this example is much steeper than the previous example in Fig. 9, topographic adjustments from the adiabatic solution are quantitatively large [Figs. 12(b) and 12(d)]. Wave-induced isopycnal setup and mean flow grow

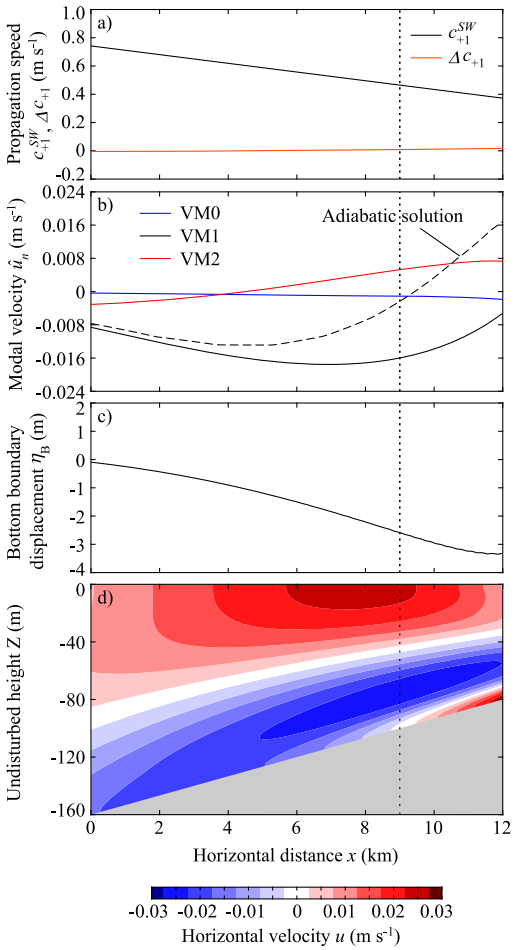

FIG. 12. Fully nonlinear simple-wave solution with 2nd-order topographic effects for Wunsch's subcritical wedge problem. Input parameters are the same as those used in Fig. 11 but with an incident-wave (semi-) amplitude of $\hat{\eta}_1$ = 2.2 m. (a) Wave propagation speed due to primary Riemann mode c_{+1}^{SW} and that due to nonprimary modes Δc_{+1}, (b) modal velocities $\hat{u}_n$ of the lowest 3 vertical modes, (c) bottom boundary displacement η_B, and (d) horizontal velocity field u at t = 34.5 h. Dashed line in (b) shows adiabatic solution for $\hat{u}_1$. Vertical dotted lines indicate the location shown in Fig. 13.

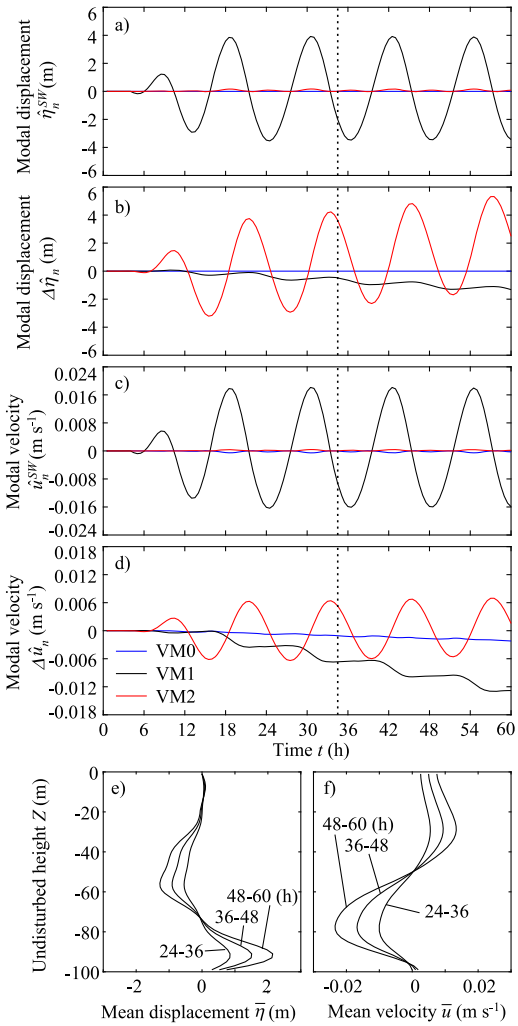

FIG. 13. Evolution of isopycnal displacements and velocities at 100-m water depth in Wunsch's subcritical wedge problem shown in Fig. 12. (a) Modal displacements of the lowest 3 vertical modes due to primary Riemann mode $\hat{\eta}_n^{SW}$, (b) those due to the sum of nonprimary modes $\Delta\hat{\eta}_n$, (c) modal velocities of the lowest 3 vertical modes due to primary Riemann mode $\hat{u}_n^{SW}$, (d) those due to the sum of nonprimary modes $\Delta\hat{u}_n$, and [(e) and (f)] vertical profiles of isopycnal displacements and velocities averaged over a wave period. Dotted lines in (a)–(d) indicate the time shown in Fig. 12.

much faster than in the previous example and obtain amplitudes comparable to the oscillatory components only after 3-4 wave periods. The change of mean stratification by this setup is the main reason why relatively small incident-wave amplitude was needed to avoid zero isopycnal-layer thickness.

Combined effects of nonlinearity and topography are important not only for wave-induced isopycnal setup and mean flow but also for the wave fields. This is much clearer in Fig. 14 than the previous two examples in Figs. 9 and 12. Compared to the linear solution with 2nd-order topographic effects, the fully nonlinear solution shows more localized upslope currents both in width and height [Figs. 14(a) and 14(b)]. This tendency is already clear near the incoming boundary and strengthened as the waves shoal. The flow field near the outgoing boundary shows resemblance to the so-called boluses,[55,56] which are ISWs of elevation with a trapped core on a sloping bottom [Figs. 14(b) and 14(c)], although the wavelength is too long for a bolus in the figure.

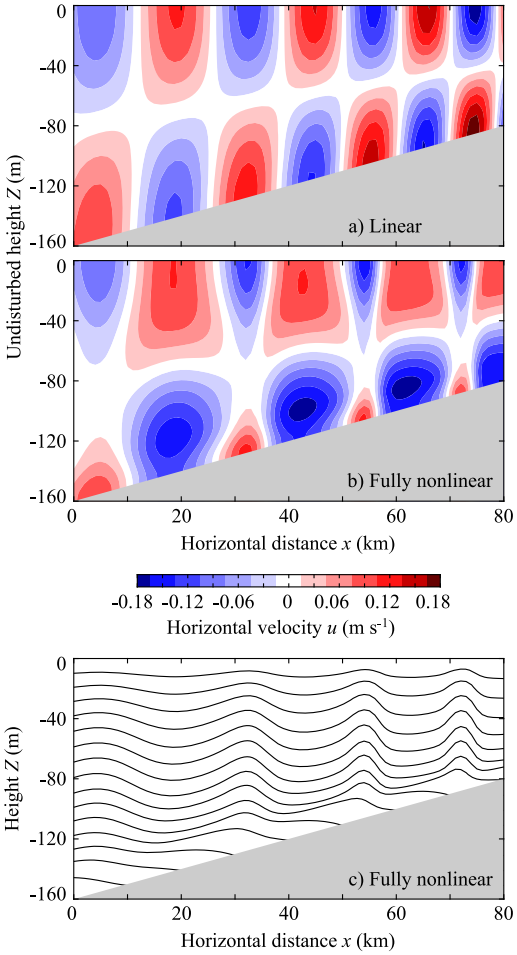

FIG. 14. Comparison of linear and fully nonlinear simple-wave solution with 2nd-order topographic effects for Wunsch's subcritical wedge problem. Stratification is linear [Fig. 4(d)], the bottom slope is $\theta = 10^{-3}$ (or dimensionless slope $\delta = 0.2$), and incident-wave (semi-) amplitude is $\hat{\eta}_1 = 14$ m. (a) Horizontal velocity field u from linear solution, (b) that from fully nonlinear solution, and (c) isopycnal-height field from fully nonlinear solution at $t = 64.5$ h.

Comparisons of wave-induced isopycnal setup and mean flow with the previous cases provide some indication of the sensitivity to background stratification, incident-wave amplitude, and the bottom slope. The wave-induced isopycnal setup and mean flow in this case develop roughly 2-3 times faster than the case with the same bottom slope and a similar incident-wave amplitude but with hyperbolic-tangent stratification (Fig. 10). Compared to the previous case with linear stratification (Fig. 13), the wave-induced isopycnal setup and mean flow in this case develop roughly 3-4 and 6-7 times faster, respectively. Note that the effects of nonlinear and topographic effects are partly compensated because the incident-wave amplitude is 6.5 times larger, but the bottom slope is 6.6 times milder than the previous case.

V. DISCUSSION

This study developed a theory of fully nonlinear simple internal waves over subcritical slopes in continuously stratified fluids. To my knowledge, this study is the first to incorporate topographic effects in a fully nonlinear simple-internal-wave theory. More generally, the theory might be novel as a theory of fully nonlinear simple waves in a gradually varying cross section, although a comprehensive literature survey is difficult because there are vast fields of physics in which the simple-wave theory can be applied. The proposed theory can be seen as a generalization of the weakly nonlinear simple-wave theory with gradually varying composition and cross section by Lighthill.[28] It can also be seen as an extension of the topography-modified KdV-type theory[22,37–39] to full nonlinearity. [Although nonhydrostatic effects are not considered in this study, it is straightforward to add a linear dispersion term to (56a) and (59a), as done in (70).] For example, (2) without the RHS term and (69) can be seen as the power-series expansion of (1) and (59a) for weak nonlinear and topographic effects, respectively. In the KdV theory, it has been known that the quantity $\sqrt{\hat{h}_p^{-1} c_p^3}\hat{\eta}_p$ remains constant under adiabatic topographic effects.[22,39,57] This study appears to be the first to identify the quantity as the Riemann variable $\tilde{R}_{+p}$ in the limit of weak nonlinear and topographic effects.

An important feature of the proposed theory is the applicability to large water-depth change over a long distance, provided that the local bottom slopes are mild. This is a significant advantage over the more straightforward perturbation expansion of the flat-bottom simple-wave solution[29,30] for small depth change. It was enabled by separating adiabatic and diabatic topographic effects and using the adiabatic simple-wave solution as a base solution for the perturbation expansion with respect to diabatic topographic effects. Although the evolutionary equations of modal amplitudes (44) unfortunately have many more terms than the original governing equations (33), the projection of (33) onto vertical modes was essential to separate adiabatic and diabatic topographic effects. As a

by-product of this approach, the applicability of the characteristic method[27,28] has been extended from flat bottom to mild slopes (e.g., Fig. 7). Although the method is applicable only to the adiabatic part of the solution, it still provides a useful first-order approximation for shoaling internal waves under large depth change if the local bottom slopes are far from critical [Fig. 8; also compare Figs. 9(b) and 12(b)].

The development of wave-induced isopycnal setup (or set-down) and mean flow (Figs. 10 and 13) is one of the processes in which the combined effects of strong nonlinearity and topography are important. For both surface waves[42,58] and internal waves,[59–63] wave-induced surface/isopycnal setup and mean flow are often studied based on the Stokes drift and radiation stress, which are first derived for a flat bottom assuming weak nonlinearity but could then be applied to slowly varying topography. Since the adiabatic simple-wave solution in this study is fully nonlinear, the effects of the Stokes drift and radiation stress (more precisely their counterparts in the isopycnal coordinate) are included in the solution, which can be seen by expanding the solution in power series around zero amplitude. Adiabatic simple internal waves do induce isopycnal setup and mean flow, but they remain stationary in time, even under the effects of subcritical slopes. By contrast, the setup and mean flow due to nonlinear and diabatic topographic effects grow in time, which makes the combined effects of strong nonlinearity and topography more important [compare solid and dashed lines in Figs. 10(e) and 10(f)].

It should be emphasized here that the isopycnal setup and mean flow shown in this study are only the parts associated with topography-modified simple internal waves. To investigate the total wave-induced setup and mean flow on continental shelves in the future, it would be necessary to develop a formulation for the mean (or slowly varying) components under the apparent forcing from the fluctuating (internal-wave) components, which is essentially given by the diabatic forcing terms on the RHS of (56a) or (59a). This is because the simple-wave theory focuses only on one Riemann mode propagating in one direction, and other modes and waves (or signals) propagating in the other direction are necessary to establish a mean state. For example, the mean flow in Fig. 10 has a barotropic component, which needs to be compensated to satisfy the zero volume flux condition at the coast. Despite this caveat, the relatively rapid growths of isopycnal setup and mean flow (Fig. 13) imply, for example, that large-amplitude internal waves on continental shelves have an inherent tendency to modify the "background" stratification and currents without mixing and that it could occur within a spring-neap tidal cycle. This could be one of the causes for the relatively rapid modulation of the bore or ISW activity.[64]

Since the proposed theory uses a perturbation approach to include diabatic topographic effects, the applicable slope range needs to be checked using some reference solutions. In this study, this was done using the analytic linear solution to Wunsch's subcritical wedge problem (Fig. 11). The comparisons suggest that the proposed 2nd-order theory fortunately provides a reasonable approximation even for relatively steep subcritical slopes of $\delta = O(1)$. For slopes very close to critical ($\delta \lesssim 2$), the proposed theory of course breaks down, but any inviscid solution, including Wunsch's analytic solution, is not very useful because mixing becomes significant.[54] Physically, the applicability of one-directional (including simple) wave theory to subcritical slopes is unsurprising because internal wave rays reflected on a subcritical slope travel in the forward direction (i.e., the same horizontal propagation direction as the incident waves).[40,65] These facts suggest that the proposed theory is applicable to subcritical slopes but not near-critical slopes. However, it should be noted that the comparisons in this study are limited to linear cases, and it is preferable to check the applicable slope range using fully nonlinear solutions, such as the results of fully nonlinear numerical simulations, in the future.

The 2nd-order topographic effects included in the proposed theory are necessary for understanding the dynamics of large-amplitude internal waves on continental shelves. Theoretical models for nonlinear internal waves, such as topography-modified KdV-type models, usually include only 1st-order topographic effects.[22,37–39] This is probably because the models are mostly applied to ISWs, whose short wavelengths tend to make higher-order topographic effects unimportant. However, internal tides and ISWs need to be modeled together on continental shelves because ISWs are often generated from internal tides within the shelf[3,22,25] and because combined effects of topography and nonlinearity are important for internal tides (Fig. 14). To get a rough estimate of topographic effects on internal tides, we may use the dimensionless slope parameter $\delta = \Theta L/H$ and typical numbers for semi-diurnal internal tides on continental shelves. We consider only internal tides propagating shoreward past the generation zone because the proposed theory assumes one-directional waves. Two-dimensional effects, such as multiple generation zones[66] and oblique interaction,[67,68] are outside of the scope of this study. Using typical numbers from the Australian North West Shelf,[2,22] inner Californian shelf,[64] New Jersey coast,[69] Oregon shelf,[70] and Scotian Shelf[25] ($L \approx 5$–30 km, $H \approx 30$–150 m, and $\Theta \approx 1$–5×10^{-3}), for example, we get $\delta \approx 0.4$–0.9. These values are within the range where 2nd-order topographic effects are necessary to capture topographic modification of the wave fields as well as the wave-induced isopycnal setup and mean flow. Note that the above scaling does not apply to bores and ISWs that have shorter horizontal scales, but does apply to the overall (mean) wave form over a semi-diurnal tidal cycle (e.g., triangular for an idealized bore over a flat bottom) in the presence of bores and ISWs.

The original motivation of this study was to develop a fully nonlinear, weakly nonhydrostatic internal wave model applicable to continental shelves. Having developed a fully nonlinear simple-wave model with 2nd-order topographic effects, it is now straightforward to extend the model to include weak nonhydrostaticity. The result would be a one-directional version of the MCC-type theory for continuously stratified fluids with 2nd-order topographic effects. Such an extension,

application of the model to large-amplitude internal waves on continental shelves, and investigation into their effects on isopycnal setup and mean flow are in progress.

VI. CONCLUSIONS

This study proposed a theory of fully nonlinear simple (or Riemann) internal waves with 2nd-order topographic effects. An important feature of the theory is the applicability to large water-depth change, provided that the local bottom slopes are mild. The incorporation of adiabatic topographic effects into the previous simple-wave theory[30] made the Riemann variable $\hat{R}_{+p}$ constant along the characteristic over topography and hence extended the applicability of the characteristic method[27,28] from a flat bottom to mild slopes. Comparisons with the analytic solution to Wunsch's subcritical wedge problem suggested that the proposed theory with 2nd-order diabatic topographic effects is applicable to subcritical slopes; however, the error unavoidably increases as the bottom slope approaches the critical slope. An important finding is the relatively rapid development of wave-induced isopycnal setup and mean flow, induced by the combined effects of strong nonlinearity and topography. Since previous fully nonlinear internal-wave theories are inapplicable to large depth change, the proposed theory is a significant step forward in understanding the dynamics of fully nonlinear internal waves over topography, such as large-amplitude internal waves on continental shelves.

ACKNOWLEDGMENTS

I thank two anonymous reviewers for their constructive comments and Steve Buchan for proof reading.

APPENDIX A: SCALING OF BOTTOM BOUNDARY CONDITIONS

This appendix explains the scaling applied to the bottom boundary conditions (37) and (39). Applying the scaling (47) to (37), we get

$$\eta = -\delta(\eta_B^{(0)} + \epsilon\delta\eta_B^{(1)}) + O(\epsilon^2\delta^3) \quad \text{at} \quad s = S^b. \tag{A1}$$

To scale (39), we need to evaluate $\partial u/\partial s$ and $\partial^2\eta/\partial s^2$ at $s = S^b$. Applying the scaling (47) and the above equation, (33a) and (33c) at $s = S^b$ are

$$\frac{\partial u}{\partial t} + \epsilon\frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right) = -\frac{\partial}{\partial x}\left(\frac{m}{\rho}\right), \tag{A2a}$$

$$0 = -\frac{\partial m}{\partial s} - \delta g\frac{d\rho}{ds}(\eta_B^{(0)} + \epsilon\delta\eta_B^{(1)}) + O(\epsilon^2\delta^3), \tag{A2b}$$

where $\rho^{-1}g(d\rho/ds)/(dZ/ds) = O(C^2D^{-2})$ is used. Taking the s-derivative of the upper equation and then substituting the lower equation, we get

$$\frac{\partial}{\partial t}\left(\frac{\partial u}{\partial s}\right) + \epsilon\frac{\partial}{\partial x}\left(u\frac{\partial u}{\partial s}\right) = \delta\frac{g}{\rho}\frac{d\rho}{ds}\frac{\partial}{\partial x}\left(\eta_B^{(0)} + \epsilon\delta\eta_B^{(1)}\right) + \left(\frac{C^2}{gD}\right)\frac{1}{\rho^2}\frac{d\rho}{ds}\frac{\partial m}{\partial x} + O(\epsilon^2\delta^3). \tag{A3}$$

Since we assume waves propagating into quiescent water, $\partial u/\partial s = 0$ at $s = S^b$ initially, and the order of magnitude of $\partial u/\partial s$ is determined by the "forcing" terms on the RHS. Therefore,

$$\frac{\partial u}{\partial s} = O\left(\delta, \frac{C^2}{gD}\right). \tag{A4}$$

To evaluate $\partial^2\eta/\partial s^2$ at $s = S^b$, we substitute (43a), use (40), and apply the scaling (47), yielding

$$\frac{\partial^2\eta}{\partial s^2} = \frac{d^2Z}{ds^2}\hat{\pi}_l\frac{\hat{\eta}_l}{\hat{h}_l} + O\left(\frac{C^2}{gD}\right). \tag{A5}$$

The Taylor expansion of (35) and the above scaling lead to the order of magnitude of the remaining terms in (44c).

APPENDIX B: ANALYTIC EXPRESSIONS OF $\hat{L}_{nm}$

The analytic expressions of $\hat{L}_{nm}$ in terms of the vertical modes $\hat{\pi}_n$ and the bottom slope θ are derived in this appendix. Although the resulting expressions are essentially the same as those for multilayer stratification,[51] the derivation in the isopycnal s-coordinate is presented here because the derivation is not straightforward. Note that the expressions in this appendix are applicable to steep slopes.

To derive analytic expressions of $\hat{L}_{nm}$, we begin with the expression

$$\int_{S_b}^{S_t}\frac{\partial\hat{\pi}_n}{\partial x}\rho\frac{dZ}{ds}\hat{\pi}_m ds = \int_{S_b}^{S_t}\frac{\partial}{\partial x}\left(c_n^2\frac{\partial}{\partial s}\left(\frac{1}{gd\rho/ds}\frac{\partial(\rho\hat{\pi}_n)}{\partial s}\right)\right)\rho\frac{dZ}{ds}\hat{\pi}_m ds, \tag{B1}$$

which comes from (41a). We apply integration by parts with respect to s to the RHS twice, noting that the x-derivative of ρ is 0 with s fixed. When evaluating the boundary values, it is essential to distinguish x-derivatives at constant s and along the boundary.[44] To calculate the derivative along the boundary, it is useful to consider the boundary following coordinate σ, which is constant along the boundaries. The standard rule for vertical coordinate transformation[44] requires

$$\left(\frac{\partial}{\partial x}\right)_s = \left(\frac{\partial}{\partial x}\right)_\sigma - \left(\frac{\partial s}{\partial x}\right)_\sigma\frac{\partial}{\partial s}. \tag{B2}$$

Note that the bottom boundary condition (41c) can be substituted into the x-derivative along the bottom (fixed σ), but not that with s fixed. After applying the above relationship to the x-derivatives at the bottom and applying the boundary conditions (41b) and (41c), we get

$$\frac{c_n^2 - c_m^2}{c_m^2c_n^2}\hat{L}_{nm}\frac{c_m^2}{\hat{h}_m} = \frac{1}{\hat{h}_m\hat{h}_n}\left(\frac{1}{c_n^2}\frac{dc_n^2}{dx}\hat{h}_n\delta_{n,m} + \frac{1}{\hat{\rho}}(\hat{\pi}_n\rho\hat{\pi}_m)^b\theta\right), \tag{B3}$$

where $\theta = (dZ/ds)(\partial s/\partial x)_{\sigma=\text{bottom}} = (dZ/ds)(dS^b/dx)$. This yields the rule for index swapping,

$$\hat{L}_{mn}\frac{c_n^2}{\hat{h}_n} = -\hat{L}_{nm}\frac{c_m^2}{\hat{h}_m}(n \neq m), \tag{B4}$$

as well as the analytic expression for dc_n/dx,

$$\frac{dc_n}{dx} = -\frac{c_n}{2\hat{\rho}\hat{h}_n}(\hat{\pi}_n\rho\hat{\pi}_n)^b\theta. \tag{B5}$$

To get an expression for $\hat{L}_{nn}$, we take the x-derivative of (42a) and substitute the above relationship, yielding

$$\hat{L}_{nn} = -\frac{\hat{h}_n}{2c_n^2}\frac{d}{dx}\left(\frac{c_n^2}{\hat{h}_n}\right). \tag{B6}$$

Note that $\hat{L}_{nn}$ vanishes when $\hat{h}_n$ is proportional to c_n^2. Note also that, in general, $d\hat{h}_n/dx$ is required to evaluate $\hat{L}_{nn}$. The general analytic expression for $d\hat{h}_n/dx$ appears unavailable, but we may set $\hat{h}_n$ using variables whose x-derivatives are known. In this study, we use $\hat{h}_n = H$ so that analytic evaluation of $\hat{L}_{nn}$ is possible.

APPENDIX C: THE ROLE OF TOPOGRAPHIC SELF-INTERACTION TERMS IN CALCULATION OF ADIABATIC SIMPLE WAVES

This appendix explains how the topographic self-interaction terms [$\hat{\mathbf{B}}_{\rm self}$ in (49)] make an adiabatic simple-wave solution independent of vertical-mode normalization, except for a constant multiplication factor. Substituting (B6) into $\hat{L}_{nn}$ in $\hat{\mathbf{B}}_{\rm self}$, (51) becomes

$$\frac{\partial}{\partial t}\begin{bmatrix}\eta_n' \\ u_n'\end{bmatrix} = -\frac{\partial}{\partial x}\begin{bmatrix}c_n^2\hat{u}_n' + \mathrm{N}'_{mln}\eta_l'u_m' \\ \eta_n'\end{bmatrix} - \begin{bmatrix}0 \\ \mathrm{N}'_{lnm}u_l'\end{bmatrix}\frac{\partial u_m'}{\partial x}, \tag{C1}$$

where

$$\eta_n' = \frac{c_n}{\sqrt{\hat{h}_n}}\hat{\eta}_n, \tag{C2a}$$

$$u_n' = \frac{\sqrt{\hat{h}_n}}{c_n}\hat{u}_n, \tag{C2b}$$

$$\mathrm{N}'_{nlm} = \frac{c_n c_m}{c_l}\sqrt{\frac{\hat{h}_l}{\hat{h}_n\hat{h}_m}}\hat{\mathrm{N}}_{nlm}. \tag{C2c}$$

To understand what this means, we need to consider how modal variables vary when the normalization factors $\hat{h}_n$ are changed. Using the terminology in tensor analysis,[71] (40) and (42a) show that $\hat{\phi}_n$ and $\hat{\pi}_n$ are contra- and co-variant with respect to the "metric" $\hat{g}_{nn} \equiv \hat{h}_n$, respectively. Since η and u are physical variables independent of vertical-mode normalization, this means that $\hat{\eta}_n$ and $\hat{u}_n$ are co- and contra-variant, respectively. Using the notation in tensor analysis (i.e., super- and subscripts for contra- and co-variant variables, respectively), the above relations become

$$\eta'(n) = \frac{c(n)}{\sqrt{\hat{g}_{nn}}}\hat{\eta}_n, \tag{C3a}$$

$$u'(n) = \frac{\sqrt{\hat{g}_{nn}}}{c(n)}\hat{u}^n, \tag{C3b}$$

$$\mathrm{N}'(nlm) = \frac{c(n)c(m)}{c(l)}\sqrt{\frac{\hat{g}_{ll}}{\hat{g}_{nn}\hat{g}_{mm}}}\hat{\mathrm{N}}^l_{nm}. \tag{C3c}$$

Here, indices in parentheses are used to denote physical components in tensor analysis, which do not vary with the metric $\hat{g}_{nn}$. This shows that (C1) are written in physical components. It also shows that if vertical modes are normalized such that $c_n^2/\hat{h}_n$ is constant independent of x, (51) becomes (C1) because $\hat{L}_{nn} = 0$ from (B6). From these, we see that the self-interaction terms in $\hat{\mathbf{B}}_{\rm self}$ are required to make (51) independent of vertical-mode normalization. Although this normalization independence is convenient, the equations in physical components are less intuitive due to the loss of similarity to the shallow water equations (3a). For example, dimensions of the prognostic variables are changed, and the variables corresponding to water depth and wave propagation speed in (C1) are c_n^2 and 1, respectively. For this reason, we use (51) rather than (C1) in the main body of this paper.

APPENDIX D: TOTAL ENERGY EQUATION ASSOCIATED WITH (51)

To derive the total energy equation associated with (51), it is convenient to assume that $\hat{\mathbf{B}}_{\rm self} = \mathbf{O}$ (by choosing $c_n^2/\hat{h}_n$ = const. or applying appropriate normalization in Appendix C) and write the equations as

$$\frac{\partial}{\partial t}\begin{bmatrix}\hat{\boldsymbol{\eta}} \\ \hat{\mathbf{u}}\end{bmatrix} = -\frac{\partial}{\partial x}\begin{bmatrix}(\hat{\mathbf{H}}+\hat{\mathbf{Y}})\hat{\mathbf{u}} \\ \mathbf{C}^2\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\eta}}\end{bmatrix} - \begin{bmatrix}\mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\mathbf{U}}\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}\hat{\boldsymbol{\eta}} \\ \hat{\mathbf{u}}\end{bmatrix}, \tag{D1}$$

where the vectors and matrices are defined as

$$(\hat{\boldsymbol{\eta}})_n = \hat{\eta}_n, \tag{D2a}$$

$$(\hat{\mathbf{u}})_n = \hat{u}_n, \tag{D2b}$$

$$(\mathbf{C})_{nm} = c_n\delta_{n,m}, \tag{D2c}$$

$$(\hat{\mathbf{H}})_{nm} = \hat{h}_n\delta_{n,m}, \tag{D2d}$$

$$(\hat{\mathbf{Y}})_{nm} = \hat{\mathrm{N}}_{mln}\hat{\eta}_l, \tag{D2e}$$

$$(\hat{\mathbf{U}})_{nm} = \hat{\mathrm{N}}_{lnm}\hat{u}_l \tag{D2f}$$

and parentheses with subscripts are used to denote matrix components. It is also convenient to introduce the following equations equivalent to (D1):

$$\frac{\partial}{\partial t}\begin{bmatrix}\mathbf{C}^2\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\eta}} \\ \hat{\mathbf{q}}\end{bmatrix} = -\begin{bmatrix}\mathbf{O} & \mathbf{C}^2\hat{\mathbf{H}}^{-1} \\ \hat{\mathbf{H}}+\hat{\mathbf{Y}} & \mathbf{O}\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}\mathbf{C}^2\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\eta}} \\ \hat{\mathbf{q}}\end{bmatrix} - \frac{\partial}{\partial x}\begin{bmatrix}\mathbf{O} \\ \hat{\mathbf{U}}^{\mathrm{T}}\hat{\mathbf{q}}\end{bmatrix}, \tag{D3}$$

where $\hat{\mathbf{q}} = (\hat{\mathbf{H}}+\hat{\mathbf{Y}})\hat{\mathbf{u}}$ is the modal amplitude of "layer" transport, $q = (d\rho/ds)^{-1}(d(Z+\eta)/ds)\rho u$. Note that if we define $\hat{\mathbf{p}} = [\hat{\boldsymbol{\eta}}\ \ \hat{\mathbf{u}}]^{\mathrm{T}}$ and $\hat{\mathbf{p}}_{\mathrm{D}} = [\mathbf{C}^2\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\eta}}\ \ \hat{\mathbf{q}}]^{\mathrm{T}}$, the total energy is $\hat{\rho}\hat{\mathbf{p}}_{\mathrm{D}}^{\mathrm{T}}\hat{\mathbf{p}}/2$, where the superscript T denotes matrix transpose. So the total energy equation can be derived by multiplying (D1) by $\hat{\rho}\hat{\mathbf{p}}_{\mathrm{D}}^{\mathrm{T}}$ and (D3) by $\hat{\rho}\hat{\mathbf{p}}^{\mathrm{T}}$ from the left and adding the equations, yielding

$$\frac{\partial \mathrm{E}}{\partial t} = -\frac{\partial \mathrm{J_E}}{\partial x}, \tag{D4a}$$

$$\mathrm{E} = \frac{\hat{\rho}}{2}\left(\hat{\boldsymbol{\eta}}^{\mathrm{T}}\mathbf{C}^2\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\eta}} + \hat{\mathbf{q}}^{\mathrm{T}}\hat{\mathbf{u}}\right), \tag{D4b}$$

$$\mathrm{J_E} = \hat{\rho}\hat{\mathbf{q}}^{\mathrm{T}}\mathbf{C}^2\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\eta}} + \frac{\hat{\rho}}{2}\hat{\mathbf{q}}^{\mathrm{T}}\hat{\mathbf{U}}\hat{\mathbf{u}}. \tag{D4c}$$

The advantage of using the above "dual" formulation is that, unlike (29), the nonlinear energy flux term remains 3rd order with respect to the prognostic variables (note that $\hat{\mathbf{U}}$ is linear in $\hat{u}_l$), which simplifies the mathematics in the presence of inter-mode interactions. This idea was originally based on the energetics calculation in electromagnetism[72] and applied to multi-layer stratified fluids.[73,74] The derivation above shows that only the terms that contribute to horizontal energy flux are retained in (51). Since only one Riemann mode has nonzero amplitude in the simple-wave solution, the substitution of $\hat{\boldsymbol{\eta}}$, $\hat{\mathbf{u}}$, and $\hat{\mathbf{q}}$ from (15) or (58) into the above equation shows the energy conservation property of adiabatic simple waves.

APPENDIX E: NUMERICAL METHODS

Numerical methods used to calculate topography-modified simple waves in this study are briefly explained in this appendix. Since the focus of this paper is the development of theory rather than a numerical model, the explanation here is brief. The details will be presented elsewhere.

The topography-modified simple-wave solutions were calculated in four steps. First, (40) was discretized using the control volume (or finite volume) method.[75] It effectively yielded the equations for multilayer stratification, and the equations were solved following Shimizu.[51] Second, simple-wave solutions were calculated as Ref. 30, but normalizing left Riemann modes $\tilde{\mathbf{l}}_r$ such that the component corresponding to maximam modal velocity is $\sqrt{c_p\hat{h}_p}/\sqrt{(c_p\hat{h}_p)_{\mathrm{ref}}}$ instead of 1 Ref. 30, using the incoming boundary as a reference location. This normalization removed the variation of $\tilde{\lambda}_{+p}$ due to linear topographic effects and made numerical calculation of $\tilde{\lambda}_{+p}$ easier. Third, $\tilde{\lambda}_{+p}$ was calculated from (20′) but in (x, t) coordinates. I first attempted to solve the equation by simple finite-difference discretization in $(\tilde{Q}_{+p}, x)$ space, but the problem appeared very sensitive to small errors in the initial condition. So (20′) was expanded into power-series in the $\tilde{Q}_{+p}$ direction, using the power-series solution for simple waves in terms of $\tilde{Q}_{+p}$ Ref. 30. The equation was numerically integrated in the x direction and then extrapolated to larger or smaller $\tilde{Q}_{+p}$ using the power series, starting from $\tilde{Q}_{+p} = 0$. Since the higher-order power-series coefficients of $\tilde{\lambda}_{+p}$ were unreliable, the process of the numerical integration in x and the extrapolation in $\tilde{Q}_{+p}$ with a low-order (up to 3rd) power series were repeated with a sufficiently small step in $\tilde{Q}_{+p}$. Finally, (56) or (59) were discretized using the control volume method[75] with explicit time stepping. The advection terms were solved using the QUICKEST scheme.[76] The number of grid points in (x, Z) was (80, 85) regardless of the bottom slope, and the number of vertical modes was 10. Buffer zones of 10 extra grid points in the x direction were used to allow η_B to adjust to model internal dynamics from potentially inconsistent boundary conditions (see Sec. IV A).

REFERENCES

[1] H. Sandstrom and J. A. Elliott, "Internal tide and solitons on the Scotian Shelf: A nutrient pump at work," J. Geophys. Res. **89**, 6415–6426, https://doi.org/10.1029/jc089ic04p06415 (1984).
[2] N. F. Smyth and P. E. Holloway, "Hydraulic jump and undular bore formation on a shelf break," J. Phys. Oceanogr. **18**, 947–962 (1988).
[3] E. L. Shroyer, J. N. Moum, and J. D. Nash, "Energy transformations and dissipation of nonlinear internal waves over New Jersey's continental shelf," Nonlinear Processes Geophys. **17**, 345–360 (2010).
[4] A. R. Osborne and T. L. Burch, "Internal solitons in the Andaman Sea," Science **208**, 451–460 (1980).
[5] J. R. Apel, J. R. Holbrook, A. K. Liu, and J. J. Tsai, "The Sulu Sea internal soliton experiment," J. Phys. Oceanogr. **15**, 1625–1651 (1985).
[6] J. M. Klymak, R. Pinkel, C.-T. Liu, A. K. Liu, and L. David, "Prototypical solitons in the South China Sea," Geophys. Res. Lett. **33**, L11607, https://doi.org/10.1029/2006gl025932 (2006).
[7] S. A. Thorpe, "Asymmetry of internal seiche in Loch Ness," Nature **231**, 306–308 (1971).
[8] L. Boegman, J. Imberger, G. N. Ivey, and J. P. Antenucci, "High-frequency internal waves in large stratified lakes," Limnol. Oceanogr. **48**, 895–919 (2003).
[9] J. Appt, J. Imberger, and H. Kobus, "Basin-scale motion in stratified Upper Lake Constance," Limnol. Oceanogr. **49**, 919–933 (2004).
[10] A. Wüest and A. Lorke, "Small-scale hydrodynamics in lakes," Annu. Rev. Fluid Mech. **35**, 373–412 (2003).
[11] K. G. Lamb, "Internal wave breaking and dissipation mechanisms on the continental slope/shelf," Annu. Rev. Fluid Mech. **46**, 231–254 (2014).
[12] K. G. Lamb, "A numerical investigation of solitary internal waves with trapped cores formed via shoaling," J. Fluid Mech. **451**, 109–144 (2002).
[13] A. Brandt and K. R. Shipley, "Laboratory experiments on mass transport by large amplitude mode-2 internal solitary waves," Phys. Fluids **26**, 046601 (2014).
[14] D. Deepwell and M. Stastna, "Mass transport by mode-2 internal solitary-like waves," Phys. Fluids **28**, 056606 (2016).
[15] D. Bogucki, T. Dickey, and L. G. Redekopp, "Sediment resuspension and mixing by resonantly generated internal solitary waves," J. Phys. Oceanogr. **27**, 1181–1196 (1997).
[16] J. M. Klymak and J. N. Moum, "Internal solitary waves of elevation advancing on a shoaling shelf," Geophys. Res. Lett. **30**, 2045, https://doi.org/10.1029/2003gl017706 (2003).
[17] D. Bourgault, M. Morsilli, C. Richards, U. Neumeier, and D. E. Kelley, "Sediment resuspension and nepheloid layers induced by long internal solitary waves shoaling orthogonally on uniform slopes," Cont. Shelf Res. **72**, 21–33 (2014).
[18] J. Grue and J. K. Sveen, "A scaling law of internal run-up duration," Ocean Dyn. **60**, 993–1006 (2010).
[19] S. I. Badulin, V. M. Vasilenko, and N. N. Golenko, "Transformation of internal waves in the equatorial Lomonosov current," Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. **26**, 110–117, (1990).
[20] K. D. Sabinin and A. N. Serebryanyĭ, "'Hot spots' in the field of internal waves in the ocean," Acoust. Phys. **53**, 357–380 (2007).
[21] R. Grimshaw, E. Pelinovsky, T. Talipova, and A. Kurkin, "Simulation of the transformation of internal solitary waves on oceanic shelves," J. Phys. Oceanogr. **34**, 2774–2791 (2004).
[22] P. E. Holloway, E. Pelinovsky, T. Talipova, and B. Barnes, "A nonlinear model of internal tide transformation on the Australian North West Shelf," J. Phys. Oceanogr. **27**, 871–896 (1997).
[23] K. G. Lamb and W. Xiao, "Internal solitary waves shoaling onto a shelf: Comparisons of weakly-nonlinear and fully nonlinear models for hyperbolic-tangent stratifications," Ocean Modell. **78**, 17–34 (2014).
[24] J. Grue, "Nonlinear interfacial wave formation in three dimensions," J. Fluid Mech. **767**, 735–762 (2015).

[25]H. Sandstrom, J. A. Elliott, and N. A. Cochrane, "Observing groups of solitary internal waves and turbulence with BATFISH and echo-sounder," J. Phys. Oceanogr. **19**, 987–997 (1989).
[26]K. R. Helfrich and W. K. Melville, "Long nonlinear internal waves," Annu. Rev. Fluid Mech. **38**, 395–425 (2006).
[27]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, 1974).
[28]J. Lighthill, *Waves in Fluids* (Cambridge University Press, 1978).
[29]L. A. Ostrovsky and J. Grue, "Evolution equations for strongly nonlinear internal waves," Phys. Fluids **15**, 2934–2948 (2003).
[30]K. Shimizu, "Physical uniqueness of higher-order Korteweg-de Vries theory for continuously stratified fluids without background shear," Phys. Fluids **29**, 106604 (2017).
[31]M. Miyata, "An internal solitary wave of large amplitude," La Mer **23**, 43–48 (1985).
[32]W. Choi and R. Camassa, "Fully nonlinear internal waves in a two-fluid system," J. Fluid Mech. **396**, 1–36 (1999).
[33]R. R. Long, "Some aspects of the flow of stratified fluids. Part I. A theoretical investigation," Tellus **5**, 42–58 (1953).
[34]M. Stastna and K. G. Lamb, "Large fully nonlinear internal solitary waves: The effect of background current," Phys. Fluids **14**, 2987 (2002).
[35]E. Varley and E. Cumberbatch, "Large amplitude waves in stratified media: Acoustic pulses," J. Fluid Mech. **43**, 513–537 (1970).
[36]G. B. Whitham, "On the propagation of weak shock waves," J. Fluid Mech. **1**, 290–318 (1956).
[37]A. K. Liu, "Analysis of nonlinear internal waves in the New York Bight," J. Geophys. Res. **93**, 12317–12329, https://doi.org/10.1029/jc093ic10p12317 (1988).
[38]J. Small, "A nonlinear model of the shoaling and refraction of internal solitary waves in the ocean. Part I: Development of the model and investigation of the shoaling effect," J. Phys. Oceanogr. **31**, 3163–3183 (2001).
[39]R. Grimshaw, C. Guo, K. Helfrich, and V. Vlasenko, "Combined effect of rotation and topography on shoaling oceanic internal solitary waves," J. Phys. Oceanogr. **44**, 1116–1132 (2014).
[40]O. M. Phillips, *The Dynamics of the Upper Ocean*, 2nd ed. (Cambridge University Press, 1977).
[41]A. E. Gill, *Atmosphere-Ocean Dynamics* (Academic Press, 1982).
[42]K. Horikawa, *Coastal Engineering: An Introduction to Ocean Engineering* (Wiley, 1978).
[43]K. G. Lamb and B. Wan, "Conjugate flows and flat solitary waves for a continuously stratified fluid," Phys. Fluids **10**, 2061–2079 (1998).
[44]A. Kasahara, "Various vertical coordinate systems used for numerical weather prediction," Mon. Weather Rev. **102**, 509–522 (1974).
[45]S. G. Llewellyn-Smith and W. R. Young, "Conversion of the barotropic tide," J. Phys. Oceanogr. **32**, 1554–1566 (2002).
[46]R. Maugé and T. Gerkema, "Generation of weakly nonlinear nonhydrostatic internal tides over large topography: A multi-modal approach," Nonlinear Processes Geophys. **15**, 233–244 (2008).
[47]B. Johns, "A boundary layer method for the determination of the viscous damping of small amplitude gravity waves," Q. J. Mech. Appl. Math. **21**, 93–103 (1968).
[48]K. Shimizu and J. Imberger, "Damping mechanisms of internal waves in continuously stratified rotating basins," J. Fluid Mech. **637**, 137–172 (2009).
[49]K. Shimizu, "An analytical model of capped turbulent oscillatory bottom boundary layers," J. Geophys. Res. **115**, C03011, https://doi.org/10.1029/2009jc005548 (2010).
[50]S. D. Griffiths and R. H. J. Grimshaw, "Internal tide generation at the continental shelf modeled using a modal decomposition: Two-dimensional results," J. Phys. Oceanogr. **37**, 428–451 (2007).
[51]K. Shimizu, "A theory of vertical modes in multilayer stratified fluids," J. Phys. Oceanogr. **41**, 1694–1707 (2011).
[52]S. M. Kelly, J. D. Nash, K. I. Martini, M. H. Alford, and E. Kunze, "The cascade of tidal energy from low to high modes on a continental slope," J. Phys. Oceanogr. **42**, 1217–1232 (2012).
[53]C. Wunsch, "Progressive internal waves on slopes," J. Fluid Mech. **35**, 131–144 (1969).
[54]D. Cacchione and C. Wunsch, "Experimental study of internal waves over a slope," J. Fluid Mech. **66**, 223–239 (1974).
[55]B. C. Wallace and D. L. Wilkinson, "Run-up of internal waves on a gentle slope in a two-layered system," J. Fluid Mech. **191**, 419–442 (1988).
[56]K. R. Helfrich, "Internal solitary wave breaking and run-up on a uniform slope," J. Fluid Mech. **243**, 133–154 (1992).
[57]K. R. Helfrich, W. K. Melville, and J. W. Miles, "On interfacial solitary waves over slowly varying topography," J. Fluid Mech. **149**, 305–317 (1984).
[58]M. S. Longuet-Higgins and R. W. Stewart, "Radiation stress and mass transport in gravity waves, with application to 'surf beats,'" J. Fluid Mech. **13**, 481–504 (1962).
[59]S. A. Thorpe, "On the shape of progressive internal waves," Philos. Trans. R. Soc., A **263**, 563–614 (1968).
[60]E. Støylen and J. E. H. Weber, "Mass transport induced by internal Kelvin waves beneath shore-fast ice," J. Geophys. Res. **115**, C03022, https://doi.org/10.1029/2009jc005298 (2010).
[61]T. S. van den Bremer and B. R. Sutherland, "The mean flow and long waves induced by two-dimensional internal gravity wavepackets," Phys. Fluids **26**, 106601 (2014).
[62]J. E. H. Weber, K. H. Christensen, and G. Broström, "Stokes drift in internal equatorial Kelvin waves: Continuous stratification versus two-layer models," J. Phys. Oceanogr. **44**, 591–599 (2014).
[63]B. Semin, G. Facchini, F. Pétrélis, and S. Fauve, "Generation of a mean flow by an internal wave," Phys. Fluids **28**, 096601 (2016).
[64]J. A. Colosi, N. Kumar, S. H. Suanda, T. M. Freismuth, and J. H. MacMahan, "Statistics of internal tide bores and internal solitary waves observed on the inner continental shelf off Point Sal, California," J. Phys. Oceanogr. **48**, 123–143 (2018).
[65]C. C. Eriksen, "Observations of internal wave reflection off sloping bottoms," J. Geophys. Res. **87**, 525–538, https://doi.org/10.1029/jc087ic01p00525 (1982).
[66]V. Vlasenko, N. Stashchuk, M. E. Inall, and J. E. Hopkins, "Tidal energy conversion in a global hot spot: On the 3-D dynamics of baroclinic tides at the Celtic Sea shelf break," J. Geophys. Res.: Oceans **119**, 3249–3265, https://doi.org/10.1002/2013jc009708 (2014).
[67]C. Wang and R. Pawlowicz, "Oblique wave-wave interactions of nonlinear near-surface internal waves in the Strait of Georgia," J. Geophys. Res. **117**, C06031, https://doi.org/10.1029/2012jc008022 (2012).
[68]K. Shimizu and K. Nakayama, "Effects of topography and Earth's rotation on the oblique interaction of internal solitary-like waves in the Andaman Sea," J. Geophys. Res.: Oceans **122**, 7449–7465, https://doi.org/10.1002/2017jc012888 (2017).
[69]E. L. Shroyer, J. N. Moum, and J. D. Nash, "Observations of polarity reversal in shoaling nonlinear internal waves," J. Phys. Oceanogr. **39**, 691–701 (2009).
[70]J. N. Moum, D. M. Farmer, E. L. Shroyer, W. D. Smyth, and L. Armi, "Dissipative losses in nonlinear internal waves propagating across the continental shelf," J. Phys. Oceanogr. **37**, 1989–1995 (2007).
[71]R. Aris, *Vectors, Tensors, and the Basic Equations of Fluid Mechanics* (Dover, 1962).
[72]I. Imai, "On the definition of macroscopic electromagnetic quantities," J. Phys. Soc. Jpn. **60**, 4100–4118 (1991).
[73]K. Shimizu, "Application of modal analysis to strongly stratified lakes," Ph.D. thesis, University of Western Australia, Crawley, Western Australia, Australia, 2009.
[74]A. de la Fuente, K. Shimizu, Y. Niño, and J. Imberger, "Nonlinear and weakly nonhydrostatic inviscid evolution of internal gravitational basin-scale waves in a large, deep lake: Lake Constance," J. Geophys. Res. **115**, C12045, https://doi.org/10.1029/2009jc005839 (2010).
[75]S. V. Patankar, *Numerical Heat Transfer and Fluid Flow* (Taylor and Francis, 1980).
[76]B. P. Leonard, "A stable and accurate convective modelling procedure based on quadratic upstream interpolation," Comput. Methods Appl. Mech. Eng. **19**, 59–98 (1979).