I. INTRODUCTION

Holmboe instability in stratified shear flows appears in a variety of physical contexts such as in astrophysics, the Earth’s atmosphere, and oceanography.1–9 Although the typical growth rate is smaller than that of Kelvin-Helmholtz instability, it is present for arbitrarily large values of the global Richardson number making Holmboe instability a good candidate for the generation of turbulence and mixing in many physical scenarios.

What distinguishes Holmboe from the Kelvin-Helmholtz instability is that unlike the latter instability, the Holmboe unstable modes have nonzero phase velocity that depends on the wavenumber (i.e., traveling dispersive modes). It was first identified by Holmboe10 in a simplified model of a continuous piecewise linear velocity profile and a step-function density profile. Several authors have expanded Holmboe’s theoretical work11–14 by considering different stratification and velocity profiles that do not include the simplifying symmetries Holmboe used in his model. Hazel15 and more recently Smyth and Peltier16 and Alexakis17 have shown that Holmboe’s results hold for smooth density and velocity profiles as long as the length scale of the density variation is sufficiently smaller than the length scale of the velocity variation. Furthermore, effects of viscosity and diffusivity,18,19 nonlinear evolution,20–23 and mixing properties24 of the Holmboe instability have also been investigated. The predictions of Holmboe have also been tested experimentally. Browand and Winant25 first performed shear flow experiments in a stratified environment under conditions for which Holmboe’s instabilities are present. Their investigation has been extended further by more recent experiments.11,26–32

Although the understanding of Holmboe instability has progressed a lot since the time of Holmboe, there are still basic theoretical questions that remain unanswered, even in the linear theory. Most of the work for the linear stage of the instability has been based on the Taylor-Goldstein (TG) equation (see Ref. 33), which describes linear normal modes of a parallel shear flow in a stratified, inviscid, nondiffusive, Boussinesq fluid:

$$\frac{d^2 \phi}{dz^2} + \left[ k^2 + \frac{U''}{U - c} - \frac{J(z)}{(U - c)^2} \right] \phi = 0, \quad (1)$$

where $\phi(z)$ is the complex amplitude of the stream function for a normal mode with real wavenumber $k$. $c$ is the complex phase velocity. $\text{Im}(c) > 0$ implies instability with growth rate given by $\zeta = k \text{Im}(c)$, $U(z)$ is the unperturbed velocity in the $x$ direction. $J(z) = -g \rho' \rho$ is the squared Brunt-Väisälä frequency where $\rho$ is the unperturbed density stratification and $g$ is the acceleration of gravity. The prime on the unperturbed quantities indicates differentiation with respect to $z$. Equation (1) together with the boundary conditions $\phi \to 0$ for $z \to \pm \infty$ forms an eigenvalue problem for the complex eigenvalue $c$.

Here, a few known results for the Taylor-Goldstein equation are presented, and some results for Holmboe’s instability are reviewed, in order to help the reader with the mathematical derivations and the discussion that follow in the next sections. If $c$ is real and in the range of $U$, there is a height $z_c$ at which $U(z_c) = c$. At this height $z_c$, called the critical height, Eq. (1) has a regular singular point. For some conditions, unstable modes exist with the real part of the phase velocity within the range of $U$. The phase speed of these modes satisfies Howard’s semicircle theorem, $|c - 1/2(\text{sup}\{U\} + \text{inf}\{U\})| < 1/2(\text{sup}\{U\} - \text{inf}\{U\})$. If these unstable modes exist, the Miles-Howard theorem34 guarantees that somewhere in the flow the local Richardson number defined by
restricted in the finite region shown in Fig. 1(a) (marked with dark gray). The boundary of this region is composed of modes with \( c = 0 \) (see Refs. \textsuperscript{15} and \textsuperscript{35}). The Holmboe unstable region forms a semi-infinite stripe in the diagram and is present for arbitrary large values of the global Richardson number \( J_0 \) (marked with light gray in Fig. 1). The unstable modes in this region are dispersive with phase velocity satisfying \( 0 < |c| < 1 \). At that time of Hazel's\textsuperscript{15} investigation, the kind of modes that determine the boundaries were not determined. The two regions on the left and on the right of the Holmboe instability stripe are stable and consist of stable gravity waves and singular modes that are part of the continuous spectrum.\textsuperscript{39,40}

Hazel observed that if \( R > 2 \), there is always a height at which the local Richardson number \( \text{Ri}(z) = J_0 \cosh(z)^4 / \cosh(Rz)^2 \) is smaller than 1/4. Based on this observation, Hazel conjectured that \( R > 2 \) is the critical value of \( R \) above which the Holmboe instability appears. Later careful numerical examination by Smyth\textsuperscript{16} found unstable Holmboe modes to appear only for values of \( R \) larger than \( R > 2.4 \). More recently, Alexakis\textsuperscript{17} showed that the instability can be found for smaller values of \( R \) up to \( R = 2.2 \), making the conjecture by Hazel still plausible. It is worth noting that the width of the Holmboe instability stripe and the growth rate of the modes decrease as the control parameter \( R \) is approaching the value 2 from above, making the detection of the Holmboe unstable modes with a numerical code difficult.

More specifically, Alexakis\textsuperscript{17} showed (numerically) for the Hazel model that the left instability boundary of the Holmboe instability region [see Fig. 1(a)] is composed of marginally unstable modes with phase velocity equal to the maximum or the minimum of the shear velocity. Such a condition has been known to hold for smooth velocity and discontinuous density profiles.\textsuperscript{36-38} Finding these marginally unstable modes corresponds to solving for the energy states \( E = -k^2 \) in a Schrödinger problem for a particle in a potential well,

\[
\frac{d^2 \phi}{dz^2} - [k^2 + V_c(z)] \phi = 0, \tag{3}
\]

where

\[
V_c(z) = \frac{U''}{U - c} - \frac{J(z)}{(U - c)^2}, \tag{4}
\]

c is taken to be \( c = U_{\text{max/min}} \) the maximum or minimum velocity of the shear layer. Note that the modes \( c = U_{\text{max/min}} \) do not always exist. For the Hazel model, for example, these modes exist only if \( R \geq 2 \) (see Ref. \textsuperscript{17}). The right boundary of the Holmboe unstable region, on the other hand [see Fig. 1(a)], is composed of singular modes with phase velocity within the range of the shear velocity. These modes can be determined by imposing the condition that the solution close to the critical height can be expanded in terms of only one of the two corresponding Frobenius solutions (see Ref. \textsuperscript{17}).

Furthermore, it was shown that for sufficiently large \( J_0 \), more than one instability stripe exists. These new instability stripes are related with the higher internal gravity modes of the unforced system, and their right instability boundaries are
determined by the higher eigenstates in the Schrödinger problem (3). Figure 1(b) shows the instability region for \( R = 3 \) and \( J_0 \) up to 80. We note that different unstable Holmboe modes have been found experimentally in Ref. 32 that were then interpreted in terms of the multilayer model of Ref. 12.

The physical picture, described in Ref. 17, for the Holmboe instability is as follows. For sufficiently small wavenumbers, the solutions of the TG equation include a discrete number of stable gravity waves with phase velocity larger than the velocity of the shear. As the wavenumber is increased, the phase speed of these modes decreases approaching the maximum value of the velocity of the shear. If the stratification length scale is small enough, there is a critical wavenumber \( k_0 \) for which the phase velocity of the waves becomes equal to the maximum wind velocity. This corresponds to the left instability boundary of the Holmboe unstable region. For wavenumbers larger than \( k_0 \), the phase velocity of the gravity waves is smaller than the maximum wind velocity and the modes become unstable. The instability persists up to another critical value of the wavenumber \( k_s \) for which the growth rate is zero but the real part of the phase velocity is within the range of \( U \). The mode with this wavenumber exhibits a singular behavior at the critical height and determines the right instability boundary of the Holmboe unstable region. For wavenumbers smaller than \( k_s \), a continuum of singular neutral modes exists.

The understanding, however, of the linear part of Holmboe instability for smooth shear and density profiles still remains conjectural and most of the results are based on numerical calculations, therefore they do not constitute proofs. This work attempts to address some of these issues. In the next section, it is proven for a general class of velocity profiles that the modes that have phase velocity equal to the maximum/minimum velocity of the shear are marginally unstable. Section III examines the case for which the parameter \( R \) is slightly larger than its critical value, and the dispersion relation inside the instability region is derived based on an asymptotic expansion. In Sec. IV, these results are tested for specific shear and density profiles. A summary of the results, their physical interpretation, and the final conclusions are in Sec. V.

II. MARGINAL WAVENUMBER

In this section, for a general family of flows the modes with phase velocity equal to the maximum/minimum velocity of the flow are examined with the aim of determining under what conditions these modes constitute a stability boundary.

Consider an infinite shear layer specified by the monotonic velocity profile \( U(y) \) that has the asymptotic values \( U(\pm \infty) = U_{\pm \infty} \). Since the system is Galilean invariant with no loss of generality, we can set \( U_{+\infty} = -U_{-\infty} = U_{0} \). More precisely, we will assume that the asymptotic behavior of \( U(y) \) for \( y \to +\infty \) is going to be given by

\[
U(y) = U_{0} - U' e^{-\nu y},
\]

The layer is stably stratified with \( J(y) > 0 \) having asymptotic behavior \( J(y) = J^* e^{-\nu y} \) for \( y \to +\infty \). In what follows, we are going to concentrate only on the modes with phase velocity close to \( c = U_{0} \); the results can easily be reproduced for the \( c = U_{+\infty} \) modes by following the same arguments.

At this stage, it is useful to rescale the variables in the TG equation using the maximum velocity \( U_{0} \) and the length scale \( \alpha^{-1} \). The new vertical coordinate now becomes \( y = \alpha z \). The wavenumber becomes \( q = k/\alpha \) and \( c \) is measured in units of \( U_{0} \). The resulting nondimensional control parameters for our system are the ratio of the two velocities, \( \sigma = U' / U_{0} \); the asymptotic Richardson number, \( J_{\infty} = J^* / (U'^2) \); and the ratio of the two length scales, \( R = \beta/\alpha \). To avoid introducing more symbols, the same symbol for the functions \( U, J, \phi, \nu_c \) is going to be used for both coordinates \( z \) and \( y \).

The physical picture, described in Ref. 17, for the Holmboe unstable region. For wavenumbers smaller than \( k_0 \), the phase velocity of the gravity waves is smaller than the maximum wind velocity and the modes become unstable. The instability persists up to another critical value of the wavenumber \( k_s \) for which the growth rate is zero but the real part of the phase velocity is within the range of \( U \). The mode with this wavenumber exhibits a singular behavior at the critical height and determines the right instability boundary of the Holmboe unstable region. For wavenumbers smaller than \( k_s \), a continuum of singular neutral modes exists.

The understanding, however, of the linear part of Holmboe instability for smooth shear and density profiles still remains conjectural and most of the results are based on numerical calculations, therefore they do not constitute proofs. This work attempts to address some of these issues. In the next section, it is proven for a general class of velocity profiles that the modes that have phase velocity equal to the maximum/minimum velocity of the shear are marginally unstable. Section III examines the case for which the parameter \( R \) is slightly larger than its critical value, and the dispersion relation inside the instability region is derived based on an asymptotic expansion. In Sec. IV, these results are tested for specific shear and density profiles. A summary of the results, their physical interpretation, and the final conclusions are in Sec. V.

Let us assume that a solution \( \phi_{0}(y) \) of the Schrödinger problem described in Eq. (3) for the wavenumber \( q_0 = k_0/\alpha \) exists. Note that for large \( y \) and for \( c = 1 \), the potential \( V_{c}(y) \) has the behavior \( V_{c}(y) = 1 - J_{c} e^{-R(y-2)} \). Clearly, if \( R > 2 \) the asymptotic behavior of \( V_{c} \) for large \( y \) is \( V_{c}(y) \approx 1 \) and \( \phi_{0}(y) \) behaves as \( \phi_{c} e^{-\lambda y} \) with \( \lambda = \sqrt{q^2 + 1} \). If, however, we have \( R = 2 \), then \( V_{c}(y) \approx 1 - J_{c} \) for \( y \to \infty \) and \( \lambda = \sqrt{q^2 + 1} - J_{c} \). For abbreviation denote for both cases

\[
\lambda = \sqrt{q^2 + 1} - \bar{J},
\]

where \( \bar{J} \) (that will be defined precisely later on) takes the values \( \bar{J} = 0 \) when \( R > 2 \) and \( \bar{J} = J_{c} \) when \( R = 2 \). As discussed in Ref. 17, no solution exists that satisfies the boundary conditions for the Schrödinger problem described in Eq. (3) if \( R < 2 \) since it corresponds to finding bounded eigenstates in an unbounded potential well.

The aim in this section is to find how \( c \) changes from the value 1 as we increase \( q \) from the value \( q_0 \). We proceed by carrying out a regular asymptotic expansion by letting \( q = q_0 + \epsilon q_{1} \) and \( c = c_{1} - \epsilon c_{2} + \cdots \) with \( 0 < \epsilon \ll 1 \) and \( c_{1} \) in general complex. However, as we deviate from the \( c = 1 \) case, the behavior of the potential \( V_{c}(y) \) drastically changes \([O(1)] \) change] in the large \( y \) region and only slightly (linearly with respect to the change in \( c \)) for \( y \sim O(1) \). Figure 2 illustrates this change for the Hazel model. This implies that two different expansions are needed, one for \( y \) being of \( O(1) \) and one for large \( y \).
A. Local solution: $y=O(1)$

Starting with the local solution and expanding $\phi$ as $\phi = \phi_0 + e\phi_1 + \cdots$ for $y=O(1)$ results at first order in the equation

$$
\frac{d^2 \phi_1}{d y^2} = \left[ q_0^2 + V_0(y) \right] \phi_1 = \left[ 2q_1q_0 - c_1V_1(y) \right] \phi_0,
$$

where $V_0 = V_\epsilon$ is given in Eq. (4) for $c=1$ and

$$
V_1(y) = \left. \frac{d^2 U}{d y^2} - \frac{J_\zeta}{U - c_1 \left( U - c \right)^2} \right|_{c=1} = \frac{J_\zeta}{(U - c)^2} - 2 J_\zeta (U - 1) - \frac{2 J_\zeta}{(U - 1)^2}.
$$

The solution of this inhomogeneous equation can be found using the Wronskian to obtain

$$
\phi_1 = \phi_0(y) \int_0^y \frac{\int_0^y \left[ 2 q_1 q_0 - c_1 V_1(y') \right] d\phi_0(y') dy'}{\phi_0^2(y')} dy',
$$

where the normalization condition $\phi_0(0) = \phi_0(0)$ is chosen. Clearly this solution satisfies the boundary condition for $y \to \infty$. For $y \to \infty$, by performing the integrations, we obtain

$$
\phi_1 = \frac{2 q_1 q_0 I_1 - c_1 I_2}{2 \lambda \varphi_\infty} e^{\lambda y} + O(e^{-\lambda y}),
$$

where $I_1 = \int_0^\infty e^{\phi_0} dy > 0$ and $I_2 = \int_0^\infty V_1 \phi_0^2 dy$. Here we need to assume that the integral $I_2$ exists and is finite. Note that if $V_1$ is not singular, $|I_2| < \infty$ for $R>2$, but $|I_2| < \infty$ only if $\lambda > 1/2$ for $R=2$. The $O(e^{-\lambda y})$ terms can be neglected when compared with $\phi_0$ but not the $O(e^{\lambda y})$ terms since for sufficiently large $y$ they can become important.

So the large $y$ behavior of $\phi$ based on the local solution is given by

$$
\phi \approx \varphi_\infty e^{\lambda y} + e^{2 q_1 q_0 I_1 - c_1 I_2} e^{\lambda y} + \cdots.
$$

B. Far away solution: $y=O(\ln[1/\epsilon])$

As discussed at the beginning of this section, the behavior of $V_\epsilon(y)$ drastically changes when $c \neq 1$ [$O(1)$ change] for large values of $y$. In particular, the TG equation (6) for large values of $y$ and for $c=1-\epsilon c_1$ reads

$$
\frac{d^2 \phi}{d y^2} - \left[ \frac{q^2 - \sigma}{\epsilon c_1 e^{y^2} - \sigma} \right] \frac{\alpha^2 J_\zeta e^{(-R-2)\bar{y}^2}}{(\epsilon c_1 e^{y^2} - \sigma)^2} \phi = 0.
$$

For $y$ of $O(1)$, this expression reduces to the case $c=1$ to first order. This is no longer true when the denominators in the above equation are close to zero (i.e., $\epsilon c_1 e^{y^2} - \sigma$). To capture, therefore, the large $y$ behavior, we need to make the change of variables $\bar{y} = y - y_\epsilon$, where $y_\epsilon = -\ln(\epsilon c_1 / \sigma)$ is the location of the singularity determined by $U(y) = c$. Note that for $c_1$ complex, $\bar{y}$ does not coincide with the real $y$ axis.)

The Taylor-Goldstein equation (1) then reads

$$
\frac{d^2 \phi}{d s^2} - \left[ q^2 - 1 \right] \frac{J_\zeta \cdot (\epsilon c_1 / \sigma)^{\delta} \cdot (e^{\bar{y}})^{\delta} \cdot \left( q^2 - 1 - s \right)^{-1} - \left( e^\bar{y} - 1 \right)^2 \right] \phi = 0,
$$

where $\delta = R-2$ and only the leading terms have been kept. Introducing the variable $s = e^{\bar{y}}$ leads to

$$
s^2 \frac{d^2 \phi}{d s^2} + \frac{d \phi}{d s} \left[ q^2 - \frac{s}{1-s} \cdot \frac{J_\zeta^2}{(1-s)^2} \right] \phi = 0.
$$

To deal with the singularities at $s=0$ and 1, we can make the substitution $\phi = s^{(1-s)\mu} h(s)$ with $\mu = 1/2 - \sqrt{1/4 - \bar{J}}$. This leads to the hypergeometric equation

$$
\begin{align*}
&\frac{d^2 h}{d s^2} h + \left[ (2q+1) - (2\mu + 2q + 1) s \right] \frac{d h}{d s} h
\end{align*}
$$

$$
+ [1 - \mu - (q + 1)] h = 0,
$$

the solution of which is the hypergeometric function $h(s) = F(a, b, d; s)$ with

$$
\begin{align*}
a &= (\mu + q) + \sqrt{q^2 + 1 - \bar{J}},
\end{align*}
$$

$$
\begin{align*}
b &= (\mu + q) - \sqrt{q^2 + 1 - \bar{J}},
\end{align*}
$$

and
Note that $q + \mu - a = -\lambda$ and $q + \mu - b = +\lambda$. Some basic properties of the hypergeometric function are given in Appendix A; here only the resulting asymptotic behavior of $\phi$ is given,

\[
\lim_{s \to 0} \phi = e^{s^b} = e^{-s^b},
\]

\[
\lim_{s \to +\infty} \phi = \frac{\Gamma(d)\Gamma(b-a)}{\Gamma(b)\Gamma(d-a)} e^{s^{b-a}} + \frac{\Gamma(d)\Gamma(a-b)}{\Gamma(a)\Gamma(d-b)} e^{s^{a-b}}.
\]

Returning to the $y$ variable and up to a normalization factor $A$, the asymptotic behavior of $\phi$ for $y < y_c$ is

\[
\phi \approx A \left[ e^{-s^b} + (-\epsilon c_1)^{2\lambda} \frac{\Gamma(a)\Gamma(b-d)\Gamma(-2\lambda)}{\Gamma(b)\Gamma(d-a)\Gamma(2\lambda)} e^{-s^b} \right].
\]

**C. Matching**

Matching the exponentially decreasing terms of the local and the faraway solutions, we obtain $A = \varphi_\infty$, and from the exponentially increasing terms we have

\[
\frac{2q_y q_\theta - c_1 l_z}{2\lambda \varphi_\infty} = \varphi_\infty (-\epsilon c_1)^{2\lambda} \frac{\Gamma(a)\Gamma(b-d)\Gamma(-2\lambda)}{\Gamma(b)\Gamma(d-a)\Gamma(2\lambda)}.
\]

The equation above can be solved iteratively by letting $\epsilon c_1 = \epsilon c_1' + e^{2b} c_2' + \cdots$. To first order, we obtain

\[
c_1' = \frac{2q_y q_\theta}{l_z},
\]

which gives the first correction to the phase speed and determines if the real part of phase speed is increasing or decreasing with the wavenumber. If, for example, $l_z > 0$ (which will be the case in the examples that follow), then $c$ is decreasing with $q$ and the correction $c_1'$ is positive for $q_1$ and negative for negative $q_1$. The opposite holds if $l_z < 0$. From now on we will assume that $l_z > 0$, which is the physically expected case (Re(e(k)) being a decreasing function of $k$; for example, a step function density profile gives $c \sim 1/\sqrt{k}$). If, however, there is a velocity and density profile such that $l_z < 0$, the same results will hold but for the opposite direction in $q$ (i.e., wavenumbers smaller than $q_0$ will be unstable and wavenumbers larger than $q_0$ will be stable). This first-order correction, however, is real, and contains no information about the growth rate. At the next order, we have

\[
\frac{c_1'' l_z}{2\lambda \varphi_\infty} = \varphi_\infty (-\epsilon c_1')^{2\lambda} \frac{\Gamma(a)\Gamma(b-d)\Gamma(-2\lambda)}{\Gamma(b)\Gamma(d-a)\Gamma(2\lambda)}.
\]

This correction is much smaller but contains the first-order correction of the imaginary part of $c$. The dispersion relation of $c$ for $q$ close to $q_0$ can then be written in terms of $q$ as

\[
c = 1 - \frac{2q_y l_1}{l_2} (q - q_0) + \frac{2\lambda \varphi_\infty^2}{l_2} \left( -2k y_1 (q - q_0) \right)^{2\lambda}
\times \left( \frac{\Gamma(a)\Gamma(b-d)\Gamma(-2\lambda)}{\Gamma(b)\Gamma(d-a)\Gamma(2\lambda)} \right).
\]

Special care is needed to interpret the term $(-c_1')^{2\lambda}$ for $c_1'$ given by Eq. (21). When $q_1 < 0$, $c_1'$ is negative and the term $(-c_1')^{2\lambda}$ is real, which corresponds to the case in which $c$ becomes larger than the shear velocity and no critical layer is formed. The Howard semicircle theorem then guarantees stability. This proves that wavenumbers slightly smaller than $q_0$ are stable. When $q_1 > 0$, $c_1'$ is positive and $(-c_1')^{2\lambda}$ becomes a complex number that can take different values depending on whether the minus sign is interpreted as $e^{i\pi}$ or $e^{-i\pi}$. The choice depends on the location of the singularity on the complex plane when we integrate the Taylor-Goldstein equation (1). If $\text{Im}(c_1) > 0$, then $(-c_1')^{2\lambda}$ should be interpreted as $|c_1|^{2\lambda} e^{i2\lambda \pi}$ because the integration is going over the singularity. If $\text{Im}(c_1) < 0$, then $(-c_1')^{2\lambda}$ should be interpreted as $|c_1|^{2\lambda} e^{-i2\lambda \pi}$ because the integration is going under the singularity. Here we arrive at an important point in the derivation: the sign of the imaginary part of $c$ based on Eq. (23) depends on the original assumption about the sign of $\text{Im}(c_1)$ when the TG equation is integrated across the singularity. Thus, in order for the matching to be successful, we need to verify that the original assumption about the sign of $\text{Im}(c_1)$ is consistent with the final result. If we assume that $\text{Im}(c_1) > 0$, then from (22) we have that

\[
0 < \text{Im}(c_1) = \sin(2\lambda \pi) |c_1|^{2\lambda} \frac{2\lambda \varphi_\infty^2}{l_2} \left( -2k y_1 (q - q_0) \right)^{2\lambda} \times \left( \frac{\Gamma(a)\Gamma(b-d)\Gamma(-2\lambda)}{\Gamma(b)\Gamma(d-a)\Gamma(2\lambda)} \right),
\]

where $\text{Im}(c_1')^{2\lambda}$ is written as $-\sin(2\lambda \pi) |c_1|^{2\lambda} e^{i2\lambda \pi}$ as previously discussed. The matching is successful only if the sign of the r.h.s. of Eq. (24) is positive as originally assumed and only then is the dispersion relation (23) valid. (We arrive at the same condition if we initially assume that $\text{Im}(c) < 0$). It is shown in Appendix B that if $R > 2$ (i.e., $\tilde{J} > 0$), the r.h.s. of (24) is always positive and the matching is successful. For the special case, however, in which $R = 2$ (i.e., $\tilde{J} = J_c$), the matching is not always successful because the product $\Gamma(b)\Gamma(d-a)$ that appears in Eq. (24) can change sign depending on the value of $\tilde{J}$. In particular, it is shown in Appendix B that if $\tilde{J} > 2q/(2q + 1)^2$, the r.h.s. of Eq. (24) is negative and thus we end up with a contradiction. Therefore, in the $R = 2$ case, we have shown instability only if

\[
\tilde{J} < 2q/(2q + 1)^2.
\]

Note that the maximum of the right-hand side of Eq. (25) is 1/4, and this condition implies that the Richardson criterion should hold at the location of the critical height $\tilde{J} = \text{Ri}(y_c) < 1/4$.

The unsuccessful matching when the condition (25) is not satisfied suggests (but does not prove) that there is no smooth solution that satisfies the TG equation in this limit. However, the TG spectrum does not consist only of smooth
modes. There is an infinity of neutral modes with a discontinuity of the first derivative at the critical height and it cannot be considered as the limit of smooth unstable solutions for \( \text{Im}(c) \to 0 \). These modes form the continuous spectrum of the Taylor-Goldstein equation and have been studied before in the literature.\(^{39,40}\) It is possible, therefore, that the reason there is no successful matching for the modes with \( q > q_0 \) is that in this region only modes of the continuous spectrum exist. It is also important to emphasize that the lack of instability at this order does not imply stability. Nonzero growth rate of smaller order can still exist and therefore the above result should be interpreted only as a sufficient condition for instability.

To summarize this section, it has been shown that if \( R > 2 \), the modes with phase velocity equal to the maximum phase velocity of the shear are marginally unstable: wavenumbers with \( q < q_0 \) are stable and wavenumbers with \( q > q_0 \) are unstable. If \( R=2 \), these modes are marginally unstable only if the condition (25) is further satisfied and stable (to the examined order) otherwise. The only assumptions that were needed for the proof is that (i) the asymptotic behavior of the velocity and density profile has the exponential behavior described at the beginning of Sec. II, (ii) modes with \( c(q)=1 \) exist, and (iii) the integral \( I_2 \) exists and is finite.

III. MARGINAL \( R \)

In the preceding section, marginal instability was shown when the wavenumber \( q \) is varied from the critical value \( q_0 \). However, the wavenumber is not a control parameter in a system. It is desirable, therefore, to examine a system for which one of the control parameters \( (J_0 \) or \( R \)) is close to the critical value for which the instability begins. Since Holmboe instability is present for arbitrarily large values of \( J_0 \), the only other control parameter left is \( R \). It is interesting, therefore, to consider a case for which \( R=2+\delta \) with \( 0 < \delta \ll 1 \) and the \( c=1 \) solution \((\phi_0,q_0)\) is known with \( q_0 \) such that \( J_\infty > 2q_0/(2q_0+1) \) so that the \( R=2 \) case gives no instability at the examined order. We make a small variation in \( q=q_0+\epsilon q_1 \) and \( c=1-\epsilon c_1 \) with the exact relation between \( \delta \) and \( \epsilon \) still undetermined. At this stage, it is assumed that \( \epsilon \) is sufficiently smaller than \( \delta \) so that the procedure in the previous section is still valid, and then the value of \( \epsilon \) gradually increases until the approximations in the previous section start to fail. As the value of \( \epsilon \) is increased, the most sensitive term (in \( \epsilon \)) that will be affected first is the term proportional to \( \bar{J} = J_\infty (\epsilon c_1/\sigma) \delta^2 \) in Eq. (15), for which \( \epsilon \) is raised to the smallest appearing power. Note that if \( \epsilon \ll \exp[-1/\delta] \), then \( \bar{J} \ll 1 \) and the results of the previous section are still valid. If, however, \( \epsilon \sim \mathcal{O}[\exp(-1/\delta)] \), then \( \bar{J} \sim \mathcal{O}(1) \). Following the same steps as in the previous section, we end up in the dispersion relation given by Eq. (23), but as in the \( R=2 \) case, \( \bar{J} \) cannot be treated as a small parameter.

The difference from the \( \delta=\mathcal{O}(1) \) case will therefore appear when we try to determine the sign of the r.h.s. of Eq. (24). To have successful matching, we need to satisfy the condition (25). Since \( \bar{J} \) is finite, the condition \( \bar{J} < 2q_1/(2q + 1)^2 \) that also appears in the \( R=2 \) case could be violated. To capture the whole unstable region, we define \( \epsilon \) such that \( J_\infty \epsilon^2 = 2q_1/(1 + 2q_0)^2 \) or

\[
\epsilon = \left[ \frac{2q_0 J_\infty}{(1 + 2q_0)^2} \right]^{1/\delta} \ll 1.
\]

Note that the term inside the brackets is always smaller than 1. For such a choice, the condition (25) for instability reads

\[
\bar{J} = J_\infty \left( \frac{\epsilon c_1}{\sigma} \right)^\delta = \left[ \frac{2q_0}{(1 + 2q_0)^2} \right] \left[ 1 + \delta \ln(c_1/\sigma) + \mathcal{O}(\delta^2) \right]
\]

or \( c_1/\sigma < 1 \). Already at this stage it can be seen that there is instability only if \( c_1 = 2q_1 q_0 I_2 / I_1 < 1 \) and therefore the instability is confined in the region of wavenumbers

\[
q_0 < q < q_0 + \Delta q,
\]

where \( \Delta q = \epsilon \sigma I_2 / 2q_0 I_1 \). Therefore, the second instability boundary for the Holmboe instability is given by \( q + \Delta q \). To get the full dispersion relation in this asymptotic limit, we need to expand in terms of \( \delta \) the product \( \Gamma(d-a)\Gamma(b) \) that appears in Eq. (23) since this is the term that can change sign depending on the value of \( \bar{J} \). This is done in Appendix B, and the resulting growth rate inside the instability region to the first nonzero order becomes

\[
\zeta = q_0 \text{Im}(c) = -\delta C_1 q_0 q_0 - q^{23} \ln \left( \frac{2(q - q_0)q_0}{\epsilon \sigma} \right),
\]

where \( C_1 > 0 \) is an \( \mathcal{O}(1) \) quantity and is given in Eq. (B1).

The maximum of the growth rate is obtained for \( q = q_0 = \epsilon^{-1/2} I_2 / (2q_0 I_1) \) with the growth rate being given by

\[
\max[\zeta] = \epsilon^{23} \left( \frac{C_1 I_0}{2} \right)^{23} \frac{I_2 \sigma}{q_0 I_1}.
\]

Therefore, the growth rate scales like \( \epsilon^{23} \) and the width of the instability region scales like \( \Delta q \sim \epsilon \). In terms of \( \delta \), these relations are given by \( \zeta \sim \epsilon^{23} \delta^{2\gamma} \) and \( \Delta q \sim \epsilon^{-2\gamma \delta} \), where \( \gamma \) is a positive constant. This very strong dependence with \( \delta \) suggests that both \( \zeta \) and \( \Delta q \) decrease very rapidly as \( \delta \) becomes smaller. This can explain the difficulty numerical codes have, when attempting to calculate growth rate for values of \( R \) very close to \( R=2 \).

IV. EXAMPLES

The previous sections presented some general results for the Holmboe unstable modes. This section examines some specific examples often used in the literature to model Holmboe’s instability.

Consider first the Hazel model that was introduced in Sec. I. Based on the definitions given in Sec. II, we have that \( \alpha = 2, \beta = 2R, U^* = 2, \) and \( J_0 = 4J_0 \). The resulting nondimensional quantities are \( J_\infty = J_0 / 4, \sigma = 2, q = k/2, \) and \( R \) has the same meaning. This model satisfies all the conditions that are stated in Sec. II, therefore for \( R > 2 \) the modes with \( c(k) = \pm 1 \) are marginally unstable. Furthermore, for the case \( R = 2 \) there is instability only if the condition (25) is satisfied, or in the units of this example if \( J_0 / 4 < k(k+1)^2 \). A simple
numerical integration shows that this is not the case for this profile (see Fig. 3). Therefore, the $R=2$ is stable (to the examined order) and is the critical value beyond which the Holmboe instability begins. The imaginary part of $c(k)$ for this profile for the case in which $R=2.1$ and $J_0=1.2$ is shown if Fig. 4, where the numerical result is compared with the asymptotic expansion of Eq. (29). Although $\delta=0.1$ is not very small, there is satisfactory agreement (a 20% difference) between the asymptotic and the numerical result. It is worth mentioning that it is very hard to find a range of values of $\delta$ in which both the asymptotic result is valid and $\text{Im}(c)$ is large enough to be captured by a numerical code. Note that decreasing the value of $\delta$ from 0.1 to 0.05 has resulted in a drop of $\text{Im}(c)$ by three orders of magnitude.

A second family of flows that is considered in this section assumes a velocity profile given by $U(y)=\text{tanh}(y)$ as in the Hazel model and the density stratification being determined by

$$-g\rho'\rho = \frac{J_0}{\cosh^{3J_0}(y)}.$$  

The advantage of this profile is that there are analytic solutions for the Kelvin-Helmholtz stability boundaries $J_0(k)$ for the cases in which $R=0, 1,$ and $2$. These stability boundaries are determined by the neutral modes that have phase velocity equal to the velocity of the flow at the inflection point, i.e., $c(k)=0$. In addition, there is an analytic solution $J_0(k)$ for the modes $k$ for which $c(k)=1$ for the $R=2$ case. The $R=0$ case was examined in Ref. 41, in which it was shown that the Kelvin-Helmholtz unstable modes satisfy $J_0<k^2(1-k^2)$. The $R=1$ case (which reduces to the $R=1$ case of the Hazel model) was investigated in Ref. 35, in which it was shown that the Kelvin-Helmholtz unstable modes satisfy $J_0<k(1-k)$. The $R=2$ case has not been investigated before (to the author’s knowledge). One can show following the same methods used for the $R=0, 1$ cases$^{35,41,42}$ that the $c(k)=0$ modes satisfy

$$J_0 = \frac{k(1-k)(2+k)(3+k)}{4(k+1)^2}$$  

with $\phi(y)=\left[1-\text{tanh}(y)\right]^{1/2}\cdot\left[\text{tanh}(y)\right]^{1/4-1/4J_0}$ and provide the Kelvin-Helmholtz instability boundary. The $c(k)=1$ modes, on the other hand, that are of interest for the Holmboe instability satisfy

$$J_0 = \frac{k(3+2k)}{(k+1)^2}$$  

for $k<1$. The stream function $\phi$ for these modes is given by $\phi=\left[1+\text{tanh}(y)\right]^{3/2}\cdot\left[1-\text{tanh}(y)\right]^{1/4-1/4J_0}$. The Kelvin-Helmholtz stability boundaries for the three cases $R=0, 1, 2$ along with the $c=1$ solutions for the $R=2$ case are shown in Fig. 5. For this example, $\sigma=2$ and $q=k/2$ and $J_\infty=2^{2R-3}J_0$. The $J_0(k)$ relation for the $c=1$ solutions does not satisfy the criterion (25), which now reads $J_0<k(k+1)^2$, thus the $R=2$ case is stable (to the examined order) and is the critical value above which the Holmboe instability begins. Because $J_\infty$ is four times bigger than in the Hazel model (for the same $J_0$), the resulting growth rate is smaller by a factor of $4^{-2/\delta H}$, which is close to $10^{-14}$ for the $\delta=0.1$ case. Figure 6 shows the growth rate based on the asymptotic expansion (29). No numerical results could be obtained for this case for values of $\delta<0.1$ that would justify a comparison with the asymptotic expansion. This example, when compared with one of Hazel, clearly demonstrates the sensitivity of the resulting growth rate to large $y$ asymptotic behavior of $J(y)$ and $U(y)$: a change by a factor of 4 in $J_\infty$ resulted in a 14 orders of magnitude difference in the growth rate.
the cases examined, Ri


FIG. 6. The Im[\epsilon] for the J(y)=J_0 \cosh^{-2\delta}(y) model for J_0=1.2 and R=2 + \delta with \delta=0.1 (solid line) and \delta=0.075 (dashed line).

V. CONCLUSIONS

In this paper, Holmboe’s instability for smooth density and velocity profiles is examined analytically. It is shown for a large family of flows that the modes with phase velocity equal to the maximum or minimum of the unperturbed velocity profile when they exist, and if the parameter R is above the critical value (R_{crit}=2) they constitute a stability boundary. This result confirms the results obtained numerically in Ref. 17, where the fact that the \eta=U_{max/min} modes are marginally unstable was only conjectured based on physical arguments and numerical results. It is also the first time shown analytically that the value of R=2 for the Hazel model is the critical value R_{crit} above which the Holmboe instability begins.

For the case in which the parameter R is only slightly larger than its critical value R_{crit}=2, the dispersion relation c(k) was obtained based on an asymptotic expansion. For this marginally unstable flow, the growth rate \eta as well as the width of the instability stripe \Delta \eta have a very strong dependence on the deviation of R from its critical value. In particular, the growth rate \eta and the width of the instability \Delta k scale as \exp[-2\Delta y/(R-R_{crit})] and \exp[-\Delta y/(R-R_{crit})], respectively (for some positive constant \gamma). For this reason, the numerical investigations performed in the past\{16,17\} were not able to capture the instability for values of R very close to R_{crit}.

The author believes also that the present results go beyond the clarification of a mathematical detail in the literature. They demonstrate the mechanisms involved in the Holmboe instability in a quantitative way, for the limit examined in this paper. In physical terms, one recognizes two regions that are important. First there is the region \gamma=O(1) that determines to first order the real part of the phase velocity of the gravity waves (i.e., the correction \eta_1 of order \epsilon). The gravity waves are coupled to the critical layer that appears at the height at which the velocity of the shear is equal to the phase velocity of the gravity waves. The location of the critical layer for an unstable gravity wave mode has to be at large enough heights so that the shear strain can overcome stratification. One would then expect that the height of the critical layer would be such that Ri(\gamma_1) < 1/4. [Note that for the cases examined, Ri(\gamma_1) \approx J_0 e^{-\beta J_0}. This restricts \gamma_1 to the range \ln(4J_0)/\delta \approx \gamma_1 and can be very large.] The growth rate of the mode will then strongly depend on the properties of the shear at this height (i.e., the correction \eta_1 of order \epsilon^2).

The coupling between the critical layer and the gravity wave gives rise to the instability. This necessary coupling between the gravity wave and the critical layer restricts the unstable wavenumbers in the following way. If the wavenumber is too small, the gravity wave travels faster than the shear velocity and there is no height \gamma that \U(\gamma) > \eta, no critical layer will be formed, and the gravity wave will be stable. If the wavenumber is too large, the gravity wave is slow and the critical layer forms in small heights such that Ri(\gamma) > 1/4, the shear strain will not be able to overcome the stratification, and as a result the gravity wave will be stable again. Therefore, unstable wavenumbers are only the ones whose phase speed is smaller than U_{max} but large enough so that the critical layer is formed in the Ri(\gamma) < 1/4 region. As the parameter R approaches from above the critical value R_{crit}=2, the smallest allowed height of the critical layer becomes larger and as a result the range of allowed wavenumbers for Holmboe’s instability becomes smaller and the growth rate is decreased.

For example, for the density and velocity profiles considered in this paper, if the phase speed of a mode close to the critical wavenumber \eta_0 can be written in terms of a Taylor expansion as \eta(\eta_0 + \Delta \eta) = 1 + a_1 \Delta \eta + \cdots (with \ a_1 = [\partial_d \eta(q)]_q=\eta_0 < 0), the unstable wavenumbers will be restricted in two ways. First, \Delta \eta must be positive (\Delta \eta > 0) so that the phase velocity is smaller than 1. Second, the condition Ri(\gamma_1) \approx J_0 e^{-\Delta \gamma} < 1/4 needs to be satisfied. The critical layer will form at a height \gamma_1 \approx -\ln(1/\Delta \eta) where the behavior \U(\gamma) \approx 1 - e^{-\gamma} is assumed again. This leads us to the estimate \Delta \eta \approx a_1^{-1/4 J_{\gamma_1}}^{-1/\delta}.

Furthermore, as the critical layer moves at larger heights, its coupling with the gravity wave becomes weaker. How weak this coupling is will depend on the amplitude of the gravity wave mode at the critical height. If we estimate the critical height by Ri(\gamma_1) \approx J_\gamma e^{-\Delta \gamma} \sim r (for some r in the range 0 < r < 1/4), and taking into account that \phi \sim e^{-\lambda \gamma} for large \gamma leads to an estimate of the growth rate \eta \sim e^{\lambda \gamma} \sim (r/J_{\gamma})^{1/\lambda \delta}.

The scaling that we get for \Delta \eta and \eta as a function of J_{\gamma} and \delta with these simple phenomenological arguments is exactly the same with the scalings that were obtained in the detailed calculation. Although these phenomenological estimates should not be trusted to a large extent, they can provide a first order of magnitude estimate of the expected unstable wavenumbers and their growth rates, in situations where the exact functional form of the density stratification and the velocity profile are not precisely known, as in experiments, and in geophysical and astrophysical flows.

Finally, it is the author’s belief that the results given in this paper can provide a basis for further numerical and analytical investigations such as an examination of the weakly nonlinear theory where a small nonlinearity is taken into account in order to examine the long time evolution of an unstable mode beyond the linear stage.

ACKNOWLEDGMENTS

Most of this work was done while the author was at ASP post-doc at the National Center for Atmospheric Research
APPENDIX A: THE HYPERGEOMETRIC EQUATION: BASIC PROPERTIES

The hypergeometric equation is

\[ z(1-z) \frac{d^2f}{dz^2} + [d - (a + b + 1)z] \frac{df}{dz} - abf = 0. \]  

(A1)

The solution that remains finite as \( z \to 0 \) is the hypergeometric function, \( f = F(a,b;c;z) \). For the normalization condition we are using, we have the following limits:

\[
\lim_{z \to 0} F(a,b,c;z) = 1/\Gamma(d),
\]

\[
\lim_{z \to 1} F(a,b,c;z) = \frac{\Gamma(d)\Gamma(d-a-b)}{\Gamma(d-a)\Gamma(d-b)}
+ \frac{\Gamma(d)\Gamma(a+b-d)}{\Gamma(a)\Gamma(b)} (1-z)^{d-a-b},
\]

\[
\lim_{z \to +\infty} F(a,b,c;z) = \frac{\Gamma(d)\Gamma(b-a)}{\Gamma(b)\Gamma(d-a)} (-z)^{-a}
- \frac{\Gamma(d)\Gamma(a-b)}{\Gamma(a)\Gamma(d-b)} (-z)^{-b}
\]

provided that \( d \neq 0,-1,-2, \ldots \) and \( a-b \) is not an integer.

APPENDIX B: THE SIGN OF THE INSTABILITY TERM

To determine whether we have successful matching, we need to find the sign of the imaginary part in the dispersion relation (23). We examine each term separately. Clearly, \( \Gamma(a) \), \( \Gamma(2\lambda) \), and \( \Gamma(d-a) \) are all positive factors since the argument of the \( \Gamma \) function is positive. The factor \( \Gamma(-2\lambda) \) is changing sign every time \( 2\lambda \) is an integer. However, its product with \( \sin(2\lambda \pi) \) always remains negative. The factors \( \Gamma(b) \) and \( \Gamma(d-a) \), however, can change sign depending on the value of \( \tilde{J} \). Using the expressions for \( a, b, d \), one can show that \(-1 \leq b \leq 0 \) if \( \tilde{J} \leq (2q)/(2q+1)^2 \) or if \( 2q \leq 1 \), and positive otherwise. Similarly, we have that \(-1 \leq d-a < 0 \) if \( \tilde{J} > (2q)/(2q+1)^2 \) and \( 2q < 1 \) and non-negative otherwise. Combining these two inequalities, we can determine the sign of the product

\[
\Gamma(b)\Gamma(d-a) \leq 0 \quad \text{if and only if} \quad \tilde{J} \leq (2q)/(2q+1)^2.
\]

The result in (24) then follows.

To find the dispersion relation for the small \( \delta \) and \( \epsilon \) given by (26), we need to find an expression for the term \( \Gamma(b)\Gamma(d-a) \). Substituting the choice of \( \epsilon \) given by (26) in the expression for \( b \) and \( d-a \) and using \( \tilde{J} = J_\infty (\epsilon \sigma \sigma')^\delta = J_\infty (\epsilon^\delta [1 + \delta \ln(c_\infty /\sigma)]) \), we have that to first order in \( \delta \)

\[
b \approx \frac{1}{2} \delta J_\infty \epsilon^\delta \ln(c_\infty /\sigma) \left[ \frac{1}{\sqrt{1 - 4J_\infty \epsilon^\delta}} + \frac{1}{\sqrt{1 + q_0 - J_\infty \epsilon^\delta}} \right]
\]

and \( b = O(1) \) if \( 2q_0 < 1 \). Similarly,

\[
d - a \approx \frac{1}{2} \delta J_\infty \epsilon^\delta \ln(c_\infty /\sigma) \left[ \frac{1}{\sqrt{1 - 4J_\infty \epsilon^\delta}} + \frac{1}{\sqrt{1 + q_0 - J_\infty \epsilon^\delta}} \right]
\]

if \( 2q < 1 \) and \( d-a = O(1) \) if \( 2q > 1 \). Using the \( \Gamma \)-function property \( \Gamma(1 + \delta) = \Gamma(1) \Gamma(1 + \delta)/\delta \), we can write the dispersion relation for \( q_0 < q < q_0 + \epsilon \sigma I_2/k_d \) as

\[
c = 1 - 2(q - q_0)q_d I_2 + \delta C_1 (q_0 - q)^{2b} \ln \left( \frac{2(q - q_0)q_d I_2}{I_2 \sigma} \right),
\]

where

\[
C_1 = \frac{J_\infty \sigma^2 \varphi^2}{I_2^2} \left( \frac{2q_d I_2}{I_2} \right)^{2b} \sin(2\lambda) \Gamma(1) \Gamma(d-a) / 2\lambda
\times \left[ \frac{1}{\sqrt{1 - 4J_\infty \epsilon^\delta}} + \frac{1}{\sqrt{1 + q_0 - J_\infty \epsilon^\delta}} \right]^{11/2}
\]

with \( w = b \) if \( 2q_0 < 1 \) and \( w = d-a \) if \( 2q_0 > 1 \).