Mon. Not. R. Astron. Soc. **260**, 323–336 (1993)

Vertical shear instability in accretion discs

S. Kumar[1] and C. S. Coleman[2]

[1]*Astronomy Centre, Physics Building, University of Sussex, Falmer, Brighton BN1 9QH*
[2]*Department of Defence, Guided Weapons Division, PO Box 1500, Salisbury, SA 5108, Australia*

Accepted 1992 July 6. Received 1992 July 6; in original form 1991 July 15

ABSTRACT

Two-dimensional shear flows in the $(R-z)$ and $(\phi-z)$ planes in thin accretion discs are studied. The vertical shear is set up by the resonant response of a disc with small tilt. It is shown that the flow is dynamically unstable, with growth rates in excess of $\sim 0.1\Omega^{-1}$, corresponding to a growth time-scale of one orbital period at a given radius. Standard analytical results for the two-dimensional flow instability are extended to take the effect of finite-density free boundaries into account. Ray theory is used to trace the path of energy transport. A small-wavelength approximation (WKBJ) is used to study the energy transport and leakage which lead to instability. The behaviour of longer wavelength modes is studied numerically, and interpreted in terms of mode coupling and wave action.

Key words: accretion, accretion discs – hydrodynamics – instabilities.

1 INTRODUCTION

Shear flows are intrinsic ingredients of accretion discs. The shear for the planar disc over the radial extent comes from the variation of the angular velocity with radius, e.g. $\Omega \propto R^{-3/2}$, in cylindrical coordinates (R, ϕ, z) for Keplerian discs. The vertical structure is usually taken to be static in the local corotating frame of reference, since the disturbances in the vertical plane reach equilibrium on the dynamical or sound-crossing time-scale. If, however, the disc is slightly tilted (where the tilt is less than the disc opening angle), which may be a generic feature of discs, then shearing motions, which are proportional to the vertical velocity, are set up in the $(R-\phi)$ and $(R-z)$ planes (Papaloizou & Pringle 1983). This disc response is inversely proportional to the viscosity (i.e. it is resonant). If the shear becomes sufficiently large, the flow can become dynamically unstable, and provide a step towards the turbulence that must characterize the flow in the accretion discs (Kumar 1988). This motivates the study of instability in compressible, plane-parallel shear flows with free boundaries, which are an important feature of accretion discs.

The perturbation problem is stated in Section 2, and standard results that characterize the instability are extended to include the free boundaries of accretion discs. These results are used in Section 3 to make order-of-magnitude estimates for the warped discs. Section 4 contains analytical results that serve as starting points for the numerical work, and also illuminate the character of the shear instability in the limit of small wavelengths. The numerical problem is studied in Section 5, and the conclusions are stated in Section 6.

2 TWO-DIMENSIONAL SHEAR FLOWS

The equations for the disturbances of the flow are derived in this section and provide the eigenvalue problem studied analytically in Section 4 and numerically in Section 5. There exist some analytical results concerning the behaviour of linearized disturbances of a stationary shear flow profile, established initially for incompressible flows (Drazin & Reid 1981). We extend some of these results in the following subsections. In particular, we wish to consider the effect of finite contributions from the free boundaries in the context of two-dimensional, plane-parallel, compressible shear flows. The fluid is barotropic, $p=p(\rho)$, and its flow is given by

$$\boldsymbol{U}=[U(z), 0, 0], \tag{1}$$

where the flow in Cartesian coordinates (x, y, z) is in the (x, z) plane, as shown in Fig. 1. This two-dimensional flow is shown in the following section to be an appropriate idealization for studying the instability of vertical shear flows in accretion discs. In particular, the disc section at a fixed radius, the $\phi-z$ cylindrical surface, can be opened out as the $x-z$ plane. We assume that there is vertical gravity, $g=-z$, in dimensionless coordinates. Its purpose is to provide the equilibrium density and pressure distributions, and it does not affect the disturbances (since there are no entropy gradients) except through the boundary conditions. The choice of this form of gravity comes from the context of

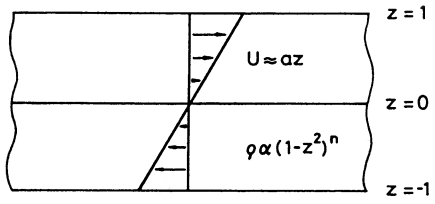

Figure 1. A schematic picture of the compressible shear flow studied. Here the fluid is polytropic, which implies boundaries at finite vertical distance. The scaling makes this unity. The rate of shear is constant, although cubic profiles could also be considered. The flows arise from the effects of tilting of accretion discs. The density and pressure at the surface are taken to vanish.

accretion discs, and differs from the laboratory fluid problems in that the gravity always points into the flow. With the enthalpy $h=\int \mathrm{d}p/\rho$, the equilibrium configuration is given by

$$\nabla(h+\Phi)=0 \tag{2}$$

(Landau & Lifshitz 1982), where Φ is the gravitational potential. Though the numerical work concentrates on linear shear flows, the results here are more general, though only in the context of accretion discs. While they are given below for barotropic flows, they can be extended to plane-parallel flows with entropy gradients (Kumar 1992). These results also serve as a check on the numerical work.

It has been shown that a stratified, incompressible shear flow is stable if the Richardson number $Ri>1/4$ everywhere over the flow (Miles 1961). This sufficient condition for stability can be used to put a bound on the instability, as performed by Howard (1961). This was extended by Chimonas (1970) for compressible fluids. However, it was assumed that the boundary was rigid and prescribed, or else the domain of flow extended to $\pm\infty$ in the z-direction. We show below that a similar calculation may be carried out for a free-boundary problem for accretion discs when the surface density is finite, but the surface pressure is zero. The linearized disturbance is characterized by the Lagrangian displacement $\boldsymbol{\xi}=(\xi_x, 0, \xi_z)$ and the Eulerian perturbations of density or enthalpy. The equations of motion, continuity and energy are

$$\frac{\mathrm{D}^2\boldsymbol{\xi}}{\mathrm{D}t^2}=-\nabla\delta h, \tag{3}$$

$$\delta\rho=-\nabla\cdot(\rho\boldsymbol{\xi}), \tag{4}$$

$$\delta h=\mathscr{H}(\rho)\,\delta\rho, \tag{5}$$

where $\mathrm{D}/\mathrm{D}t=\partial/\partial t+\boldsymbol{U}\cdot\nabla$ is the Lagrangian derivative and $\mathscr{H}(\rho)$ is some given function of ρ ($\mathscr{H}$ is always positive definite). Since the equilibrium is stationary and independent of the flow direction, we can use the normal-mode analysis in the form

$$f(\boldsymbol{x}, t)=f(z)\exp[\mathrm{i}(kx+\sigma t)], \tag{6}$$

where f is any of the perturbation variables (ξ_x, ξ_z, δh), k is the horizontal wavenumber, and σ is the frequency whose imaginary part gives the growth rate of the instability ($=-\sigma_{\mathrm{I}}$). Define $\overline{\sigma}=\sigma+kU$. Then $\Re(\overline{\sigma})=\sigma_{\mathrm{R}}+kU$ is the pattern frequency, where σ_{R} is the real part of σ. This is the local frequency of the disturbance as seen by an observer comoving with the flow. Using the variables

$$Z=\sqrt{\overline{\sigma}}\rho\xi_z \tag{7}$$

and

$$W=\delta h/\sqrt{\overline{\sigma}}, \tag{8}$$

the perturbation equations (3), (4) and (5) become, writing $\mathrm{D}=\mathrm{d}/\mathrm{d}z$,

$$\frac{\overline{\sigma}}{\rho}Z-\left(\mathrm{D}+\frac{k\mathrm{D}U}{2\overline{\sigma}}\right)W=0, \tag{9}$$

$$\left(\mathrm{D}-\frac{k\mathrm{D}U}{2\overline{\sigma}}\right)Z+\left(\frac{\overline{\sigma}}{\mathscr{H}}-\frac{k^2\rho}{\overline{\sigma}}\right)W=0. \tag{10}$$

For a global bound, we need to consider integral quantities in the perturbed variables W and Z formed by multiplying (9) by Z^* and (10) by W^*, adding and taking the integral I of the sum over z. Then, $I=0$. Taking the imaginary part to be separately equal to zero, and partially integrating, we find that the finite surface term $-(Z^*W)$ gives a positive definite contribution, and therefore does not affect the known stability criterion. The boundary condition used, appropriate for free surfaces, is that the Lagrangian pressure perturbation vanishes, $\Delta p=0$. The form used here is $\mathrm{D}/\mathrm{D}t(\Delta p)=0$, which implies the condition $\rho\overline{\sigma}W\pm Z=0$ at $z=\pm1$. This leads to the necessary condition for instability:

$$\int \mathrm{d}z(|Z|^2/\rho)[1-(\mathrm{D}U/2|\overline{\sigma}|)^2]<0, \tag{11}$$

that is, $|\overline{\sigma}|<|\mathrm{D}U/2|$ somewhere in the flow. This gives the required bound, using the inequality

$$|\overline{\sigma}_{\mathrm{I}}|\le|\overline{\sigma}|<|\mathrm{D}U/2|_{\max}. \tag{12}$$

For a linear shear flow $U=az$, $|\overline{\sigma}|_{\mathrm{I}}<a/2$, where $a>0$. Note that this global bound does not depend on the wavenumber. This inequality suggests that boundaries may play a role in destabilizing the flow, and may affect growth rates, but only up to a point. They can never overwhelm the shear destabilization; in particular, the surface mode, when it is unstable, must be mediated by shear, as is now understood to be the case (e.g. Narayan, Goldreich & Goodman 1987).

It is known that unstable modes of a linearized disturbance must have a vanishing pattern speed, $\overline{\sigma}_{\mathrm{R}}=\sigma_{\mathrm{R}}+kU=0$, somewhere in the flow (Drazin & Reid 1981). We extend this by showing that the presence of free boundaries does not affect this result. For this purpose, we use the perturbed enthalpy δh as the variable, in which case equations (9) and (10) combine to give

$$\mathrm{D}\left(\frac{\rho}{\overline{\sigma}^2}\mathrm{D}\delta h\right)+\left(\frac{\rho}{c^2}-\frac{k^2\rho}{\overline{\sigma}^2}\right)\delta h=0, \tag{13}$$

where c is the sound speed. Multiplying (13) by δh^*, and integrating over z using the boundary condition $\mathrm{D}\delta h=\pm\overline{\sigma}^2\delta h$ at $z=\pm1$, we find

$$\int \mathrm{d}z\frac{\rho\overline{\sigma}^{*2}}{|\overline{\sigma}|^4}(|\mathrm{D}\delta h|^2+k^2|\delta h|^2)=\int \mathrm{d}z\frac{\rho}{c^2}|\delta h|^2+\sum_{\text{boundary}}\rho|\delta h|^2. \tag{14}$$

Note that the right-hand side of equation (14) is positive definite. To use this result, we define $\sigma_c = \sigma^* + kU_c$, where U_c is an arbitrary but constant velocity. Then $\overline{\sigma}^* = \sigma_c + k(U - U_c)$. We then substitute in (14), which can be written in the form

$$A\sigma_c^2 + B\sigma_c + C = 0, \tag{15}$$

where A, B, C are as follows:

$$A = \int \mathrm{d}z\, \rho/|\overline{\sigma}|^4 \mathscr{F}, \tag{16}$$

$$B = \int \mathrm{d}z\, 2k\rho(U - U_c)/|\overline{\sigma}|^4 \mathscr{F}, \tag{17}$$

$$C = \int \mathrm{d}z\, k^2\rho/|\overline{\sigma}|^4 (U - U_c)^2 \mathscr{F} - \int \mathrm{d}z\, \rho/c^2 |\delta h|^2 - \sum_{\text{boundary}} \rho |\delta h|^2, \tag{18}$$

and $\mathscr{F} \equiv |\mathrm{D}\,\delta h|^2 + k^2|\delta h|^2$. For the critical layer, $\sigma_{cR} = \sigma_R + kU_c = 0$, for a certain choice of U_c. Then, from (15), by separating real and imaginary parts, and assuming $\sigma_I \neq 0$, we require $B = 0$. From (17), this is possible only if the integrand changes sign, or equivalently if $U = U_c$ somewhere in the flow. This demonstrates the existence of a critical layer (there can be more than one if U is non-monotonic) for an unstable mode within the flow. Alternatively, if we take the imaginary part of (14) we have the result that the critical layer lies in the flow.

We use the above results, in particular equations (14)–(18), to put a wavenumber-dependent bound on the shear flow instability. Again we find that the presence of free boundaries does not affect this bound, further arguing against a pivotal role for the boundary as the primary cause of instability. Since A, B and C are real, we see from (15) that $\sigma_{cI} = \sigma_I \neq 0$ only if $B^2 - 4AC < 0$. Since $A > 0$, we find that

$$\left|\frac{\sigma_I}{k}\right|^2 < \frac{\int \mathrm{d}z\, \rho(U - U_c)^2/|\overline{\sigma}|^4 \mathscr{F}}{\int \mathrm{d}z\, \rho/|\overline{\sigma}|^4 \mathscr{F}}. \tag{19}$$

Taking the maximum value of $|U - U_c|$, we obtain the required bound:

$$|\sigma_I| < k|U - U_c|_{\max}. \tag{20}$$

This value holds at either boundary for a monotonic flow profile. In any case, $|U - U_c|_{\max} \leq 2|U|_{\max}$, and we may write

$$|\sigma_I| < 2k|U|_{\max}. \tag{21}$$

We see from this relation that free-boundary effects are not destabilizing as $k \to 0$, i.e. in the limit of large wavelengths, for plane-parallel, compressible shear flows. We also see directly from (15)–(18) that $B = 0$ and $C < 0$ when $k = 0$. Now, σ_c^2 is identical to σ^2 and, from equation (15), takes a value ≥ 0.

We note for completeness that finite contributions from the free boundaries do not affect the Howard Semicircle Theorem, which states that, for a flow with given maximum and minimum velocities, the frequency and the growth rate of any unstable mode lie in a semicircle, as follows:

$$[\sigma_R/k - 1/2(U_{\max} + U_{\min})]^2 + (\sigma_I/k)^2 \leq 1/4(U_{\max} - U_{\min})^2 \tag{22}$$

(Howard 1961; Drazin & Reid 1981). A sufficient condition for stability of planar shear flows is provided by the Rayleigh criterion: for an incompressible fluid with rigid boundaries, D^2U should not change sign within the flow (Drazin & Reid 1981). This has been extended to compressible flows (Grinfeld 1984), for which the sufficient condition becomes: (i) the flow is always subsonic, and (ii) U and $\mathrm{D}(\mathrm{D}U/\rho)$ have the same sign everywhere in the flow. If the surface density drops to zero, as in the model of interest, then the flow necessarily becomes supersonic there, violating the sufficiency criterion regardless of the second condition.

3 APPLICATION TO SHEAR FLOWS IN TWISTED DISCS

It is likely that most accretion discs that occur in nature are warped, in the sense that the total angular momentum may not be aligned with the perpendicular axis. In other words, in Cartesian coordinates (x, y, z), the angular momentum components $J_x, J_y \neq 0$, though they may be small. However, for small values of the viscosity, the flows set up in the disc in the azimuthal and radial directions, here locally represented by the x- and y-coordinates, exhibit a resonant response (Papaloizou & Pringle 1983; Kumar 1988), and become large. It is then likely that they will become unstable and proceed to turbulence. Since viscosity in discs is likely to be turbulent, it is of interest to determine whether or not such flows are dynamically (i.e. generally Kelvin–Helmholtz) unstable, and what the growth rates are, in their role as precursors to shear turbulence.

In accretion disc theory, the fluid motion is Keplerian, i.e. $\Omega \propto R^{-3/2}$ in cylindrical coordinates (R, ϕ, z). The internal energy is much less than the gravitational energy, so that the vertical extent of the disc, H, is much less than the radial extent, i.e. $H/R \ll 1$. Vertically, the fluid is in hydrostatic equilibrium. However, viscosity leads to angular momentum exchange, and therefore matter infall (see Pringle 1981 and Meyer 1986 for reviews of accretion disc theory). The nature of viscosity is a major uncertainty of the theory. The viscosity is therefore modelled by various prescriptions, of which the α-disc model (due to Shakura & Sunyaev 1973) has now become the standard. The viscous stress $T_{R\phi} = \alpha p$, where $0 \leq \alpha \leq 1$ is a dimensionless quantity usually taken to be constant throughout the disc, though there have been some attempts to model it as $\propto (H/R)^\delta$, where δ is some number, often taken to be 3/2 (Meyer & Meyer-Hofmeister 1984). If α is taken to be a constant (this is not essential for the relation below), then the flow velocities set up by the twist may be written as

$$U(z) = zf(R)/\alpha^2, \qquad V(z) = zg(R)/\alpha^2, \tag{23}$$

where f and g are some functions calculated from accretion disc theory, here effectively constants. Although this motivates the study we may, for purposes of analysis in this section, take $U(z)$ and $V(z)$ to have arbitrary profiles. By the use of Squire's transformation, a standard method (see Drazin & Reid 1981), we may show that the preceding results for two-dimensional plane-parallel flows are applicable for this problem. It will be seen below that the application of this method is quite straightforward. As before, we take $p = p(\rho)$ and $\boldsymbol{U} = [U(z), V(z), 0]$, and transform the coordinates and the variables to their dimensionless equivalents. Now $\Omega = 1$, and the gravity force $\nabla\Phi = (0, 0,$

$-z$). We consider the perturbation equation that is identical to equations (3), (4) and (5), but with $\xi=(\xi_x, \xi_y, \xi_z)$. Using normal modes, with

$$f(\boldsymbol{x}, t)=f(z)\exp[\mathrm{i}(kx+ly+\sigma t)], \tag{24}$$

$$\mathrm{D}/\mathrm{D}\tau=\mathrm{i}(\sigma+kU+lV)=\mathrm{i}\bar{\sigma}, \tag{25}$$

the equations become

$$\bar{\sigma}^2\xi_x=\mathrm{i}k\delta h, \tag{26}$$

$$\bar{\sigma}^2\xi_y=\mathrm{i}l\delta h, \tag{27}$$

$$\bar{\sigma}^2\xi_z=\mathrm{D}\delta h, \tag{28}$$

$$\delta h=\mathscr{H}(\rho)[\mathrm{i}\rho(k\xi_x+l\xi_y)+\mathrm{D}(\rho\xi_z)]. \tag{29}$$

Equations (26)–(29) may now be transformed to an equivalent two-dimensional form, as follows. We choose a new set of variables:

$$\tilde{k}=\sqrt{k^2+l^2}, \tag{30}$$

$$\tilde{\xi}=(k\xi_x+l\xi_y)/\tilde{k}, \tag{31}$$

$$\tilde{U}=(kU+lV)/\tilde{k}, \tag{32}$$

$$\bar{\sigma}=\sigma+\tilde{k}\tilde{U}, \tag{33}$$

so that equations (26)–(29) become identical to (9) and (10), but with the appropriate tildes. The instability bound then becomes

$$|\sigma_\mathrm{I}|<|\mathrm{D}\tilde{U}/2|_{\max}<\tfrac{1}{2}[\sqrt{(\mathrm{D}U)^2+(\mathrm{D}V)^2}]. \tag{34}$$

We use this inequality to put a qualitative bound on the turbulent viscosity that may result from the instability. We now consider dimensional quantities. From the theory of twisted discs, we have

$$|\mathrm{D}U|\sim|\mathrm{D}V|\sim\beta\Omega/\alpha^2, \tag{35}$$

where β is the local disc tilt, and satisfies $\beta<H/R$. If the turbulent, kinematic viscosity $\nu\sim H^2\Omega|\sigma_\mathrm{I}|$, then from (35) we get $\nu=\alpha H^2\Omega\leq\beta\Omega H^2/\alpha^2$, which implies $\alpha^3\leq\beta$, or $\alpha\leq(H/R)^{1/3}$. Self-consistency of the twisted disc theory requires $\alpha>H/R$, which is equivalent here to the requirement $|\sigma_\mathrm{I}|\geq H/R$. If this is not satisfied, then $\tilde{U}\sim v_\phi$, the Keplerian velocity $=\Omega R$, we have dynamical rather than viscous motions, and the flow is not that of an accretion disc. Typically, H/R varies from 10^{-3}–10^{-1}. Growth rates larger than this for the fastest growing mode are sufficient to make the connection between instability, α-viscosity and consistency of the twisted discs theory encouraging.

For the purposes of numerical computation, we take a linear profile, $U=az$, where $a=\beta/\alpha^2$. This vorticity is bounded by $\alpha<a<a^{-1}$. Taking α to be typically between 3×10^{-2} and 1, from numerical modelling and observational estimates, the relevant range of parameters is $3\times10^{-2}<a<3\times10^{2}$, but $10^{-1}<a<10$ may be significant for any value of the wavenumber k. From length-scale arguments, we expect $k^{-1}\sim$ width of the flow. With coordinates normalized by the extent of the boundaries, i.e., $z_\mathrm{B}=\pm1$, $k\sim1$. In practice, we expect the range $10^{-1}<k<10$ for the fast-growing unstable mode. If instead we write $\alpha=\mathscr{C}|\sigma_\mathrm{I}|$, then with $|\sigma_\mathrm{I}|\sim0.2$ we find that the three conditions $\alpha^3<\beta$, $\beta<\gamma$ and $\alpha>\gamma$ are satisfied for $8\times10^{-3}\mathscr{C}^3<\beta<10^{-2}$–$10^{-1}$, or $0.2\mathscr{C}>10^{-2}$–10^{-1} for $\mathscr{C}\sim1$. From this last condition, $\mathscr{C}$ cannot be much less than 1, therefore $\alpha\sim\sigma_\mathrm{I}$ is a reasonable assumption. If $|\sigma_\mathrm{I}|\sim0.3$, then $\mathscr{C}\sim1/3$ satisfies the inequalities, but it may be larger, since we may expect a larger α-viscosity as H increases.

The application of the extension of the Rayleigh criterion (Grinfeld 1984) to shear flows in twisted discs shows that the sufficient condition for stability is satisfied for a constant-density (but compressible) fluid flow, when the boundary is taken to be rigid. Even for a variable-density but rigid-boundary flow, the conclusion is the same, unless the flow becomes supersonic somewhere. If $\rho=$ constant, the criterion is $U(z)/\mathrm{D}^2U(z)>0$ over the flow. For an isothermal disc, with $\rho\propto\mathrm{e}^{-z^2/2}$, the shear flow profile is $U(z)=c\sinh(z)$, where c is some constant, and the criterion $\rho^2U(z)/[\mathrm{D}^2U(z)-\mathrm{D}U\mathrm{D}\rho/\rho]>0$ becomes $\rho^2/[1+z\coth(z)]>0$, which is identically satisfied in the flow. From this point of view, the case considered by Narayan et al. (1987) is degenerate.

4 ANALYTICAL PRELIMINARIES

As a prelude to numerical computation, in Section 4.1 we derive useful results for the special cases of perturbed flows. The modes and their frequencies for a static compressible fluid with density distribution characteristics of an accretion disc provide the discrete eigenspectrum for any wavenumber. They are also the starting point for the picture of mode coupling leading to instability, which is discussed in Section 6. In Section 4.2, ray theory provides a picture of the distribution of the energy in the perturbed flow, again specific to the disc density profile. This refractive effect argues that global disturbances cannot easily penetrate in the radial direction; instead, they may lead to dissipation near the disc surface. In Section 4.3 the coupling of surface modes leading to instability is studied in the WKBJ approximation. This enables a determination of the relation between the growth rate, the rate of shear and the wavenumber of the disturbance, and facilitates the physical interpretation of the instability.

4.1 The spectra of static configurations

We first show that a static, two-dimensional fluid, with free boundaries, is stable. We then find the frequency spectrum for a given wavenumber, and see that the frequencies are real. These results are used to start the numerical computations. Putting $U=0$ in equation (13) for the perturbed enthalpy δh (which we shall from now on call X to simplify notation), we obtain

$$\mathrm{D}(\rho\mathrm{D}X)+(\sigma^2/c^2-k^2)\rho X=0. \tag{36}$$

The regularity condition at the free boundaries, $z_\mathrm{B}=\pm1$, in the dimensionless variables used here, is $\mathrm{D}X=\pm\sigma^2X$. Multiplying (36) by X^* and integrating over z, we find

$$\int\mathrm{d}z\rho\mathscr{T}=\sigma^2\left(\sum_{\text{boundary}}\rho|X|^2+\int\mathrm{d}z\frac{\rho}{c^2}|X|^2\right). \tag{37}$$

The left-hand side of equation (37) is real and positive definite, as are the terms in the brackets on the right-hand side. This implies that σ^2 is real and ≥0, and proves the stability of all static configurations. For such stable waves, the coefficient (σ^2/c^2-k^2) will not vanish in the domain of the fluid. Since there is no critical layer, these waves do not

extract energy from the fluid even when the fluid is isothermal, which can be seen from the explicit value of the spectrum in equation (42). This suggests that, even when the equations are formulated with the radial extent taken into account (Lin, Papaloizou & Savonije 1990), the vertically propagating waves will not dissipate locally at a given point (R, ϕ, z), apart from loss due to wave-steepening near the disc surface. Shear is a necessary ingredient for shock formation, and its consequent dissipation.

For an incompressible fluid, $c^2 \to \infty$, and if $\rho =$ constant, equation (36) reduces to $\mathrm{D}^2X - k^2X = 0$, with the boundary condition $\mathrm{D}X = \pm \sigma^2 X$ at $z = \pm 1$. We then find that

$$\sigma(k) = \pm\sqrt{k \tanh k}, \qquad \pm\sqrt{k \coth k}. \tag{38}$$

If the static fluid is isothermal then (36) may be written as

$$\mathrm{D}^2X - z\mathrm{D}X + (\sigma^2 - k^2)X = 0. \tag{39}$$

We write $\nu + 1/2 = \sigma^2 - k^2$, and substitute $X = \mathrm{e}^{z^2/4}u(z)$ to find the parabolic cylinder equation $\mathrm{D}^2u + (\nu + 1/2 - z^2/4)u = 0$. The solution is $u(z) = A\,\mathrm{D}_\nu(z) + B\mathrm{D}_{-(\nu+1)}(-\mathrm{i}z)$, on which we impose the regularity condition for $|z| \to \infty$. The appropriate condition is that $\rho|X|^2$ be finite, which is equivalent to $|u|^2$ being finite for $|z| \to \infty$. To impose this condition, we use the following asymptotic forms for the $\mathrm{D}_\nu(z)$:

$$\mathrm{D}_\nu \sim z^\nu \mathrm{e}^{-z^2/4} \qquad (z \to \infty; |\arg z| < 3\pi/4), \tag{40}$$

$$\mathrm{D}_\nu \sim z^\nu \mathrm{e}^{-z^2/4} - \frac{\sqrt{2\pi}}{\Gamma(-\nu)} \mathrm{e}^{\mathrm{i}\pi\nu} z^{-(\nu+1)} \mathrm{e}^{z^2/4} \tag{41}$$

$$(z \to \infty; \pi/4 < \arg z < 5\pi/4).$$

(Bender & Orszag 1978). We therefore require that $B = 0$ from (40), when $z \to \infty$. When $z \to -\infty$, regularity requires that $\Gamma(-\nu) = \infty$ in (41). This in turn implies that ν is a non-negative integer, and is the required condition for the frequency spectrum:

$$\sigma^2 = k^2 + m + 1/2 \qquad (m \geq 0 \text{ is integer}). \tag{42}$$

We therefore have a countable infinity of frequencies for the static, isothermal fluid. This situation resembles the quantum harmonic oscillator.

If the fluid is homogeneous but compressible, we have $\mathrm{D}^2X + (\sigma^2/c^2 - k^2)X = 0$ with the same boundary conditions. Writing $L = +\sqrt{\sigma^2/c^2 - k^2}$, the eigenvalues are given implicitly by

$$\mathrm{e}^{4L}\left(\frac{L-\sigma^2}{L+\sigma^2}\right)^2 = 1, \tag{43}$$

which must, however, be solved numerically. Note that L may a priori be imaginary, but stability for such a static configuration implies that it is always real.

A useful starting point for the computations is also provided by the static, adiabatic disc with a polytropic index $n = 1/2$. For finite wavenumbers k, the oscillation modes are the solution of a Mathieu equation. In the limit of long wavelength, $k \to 0$, there are two infinite classes of modes – the odd and the even. The frequencies are $\sigma = (2m+1)$ [the odd mode, $\propto \sin(\sigma z)$] and $\sigma = 2m$ [the even mode, $\propto \cos(\sigma z)$], where $m = (0, \pm 1, \pm 2, \ldots)$.

4.2 Ray propagation

In the short-wavelength limit, when the wavelength $\lambda \sim k^{-1} \ll H$, the local pressure scaleheight, we can find the path of energy transport of the linear oscillations, i.e. the rays. The sheared medium is anisotropic, therefore the directions of the local ray vector, the group velocity and the wave normal, i.e. the phase velocity, do not coincide. For the purposes of this section, the perturbed enthalpy equation, (13), may be written, using subscripts x, z, T to denote the partial derivatives $\partial/\partial x$, $\partial/\partial z$ and the Lagrangian derivative $\mathrm{D}/\mathrm{D}t$, as

$$\frac{1}{c^2}X_{TTT} - (X_{xx} + X_{zz})_T - \mathscr{R}(z)X_{zT} + 2\mathrm{D}UX_{zx} = 0, \tag{44}$$

where $\mathscr{R}(z) = \mathrm{D}\ln\rho$. For a static atmosphere, this reduces to

$$X_{tt} - c^2[X_{xx} + \mathscr{R}(z)X_z + X_{zz}] = 0, \tag{45}$$

where subscript t denotes $\partial/\partial t$. Now we use the geometrical optics form for X (Landau & Lifshitz 1982; Gough 1987):

$$X(t, \boldsymbol{x}) = A(t, \boldsymbol{x}) \exp[\mathrm{i}\Phi(t, \boldsymbol{x})], \tag{46}$$

where the amplitude A is a slowly varying function, and $\Phi(t, \boldsymbol{x})$ is the eikonal. We define the frequency as $\omega = \partial_t\Phi$, the wave vector as $\boldsymbol{k} = -\nabla\Phi$, and the Hamiltonian as $H(t, \boldsymbol{x}, \boldsymbol{k}) = \omega(t, \boldsymbol{x})$. Then the ray equations are $\dot{\boldsymbol{k}} = -\nabla H$ and $\dot{\boldsymbol{x}} = \nabla_k H$, where equation (48) below gives the group velocity $\boldsymbol{v}_\mathrm{g}$ of the wave. There are three wave solutions for the hyperbolic equation (45), whose phase speeds are $v_\mathrm{p} = 0, \pm c$. Then the Hamiltonian, H, becomes

$$H = v_\mathrm{p}|\boldsymbol{k}| + \boldsymbol{U}\cdot\boldsymbol{k}, \tag{47}$$

where v_p takes the three different values found from the eikonal equation

$$(\omega - \boldsymbol{k}\cdot\boldsymbol{U})(\omega - \boldsymbol{k}\cdot\boldsymbol{U} + c|\boldsymbol{k}|)(\omega - \boldsymbol{k}\cdot\boldsymbol{U} - c|\boldsymbol{k}|) = 0. \tag{48}$$

Using (46) and (47) in (48), we now find the ray trajectories for some special cases. We write $\boldsymbol{n} = \boldsymbol{k}/|\boldsymbol{k}|$ as the unit wave normal. If $U = 0$, then $H = \pm c|\boldsymbol{k}|$, k_x is a constant, and $\boldsymbol{v}_\mathrm{g} = c\boldsymbol{n}$. The equation for the vertical wavenumber k_z, found by combining (46) and (47), is

$$\mathrm{D}K_z = -\frac{(\boldsymbol{k}\cdot\boldsymbol{k})}{k_z}\mathrm{D}\ln c. \tag{49}$$

For an isothermal atmosphere $c =$ constant, so k_z is also a constant. For an adiabatic atmosphere, with $\rho \propto (1-z^2)^n$ and $c^2 = (1-z^2)/2n$,

$$k_z^2 = \frac{2m\omega^2}{(1-z^2)} - k_x^2, \tag{50}$$

and the path is represented by $\boldsymbol{x} - \boldsymbol{x}_0 = \int \mathrm{d}s\boldsymbol{m}$, where s is the path-length. The amplitude variation is given by

$$\nabla\cdot(A^2\boldsymbol{k}) = -\mathscr{R}(z)A^2k_z. \tag{51}$$

If the disc is homogeneous, the right-hand side of equation (51) $= 0$, and

$$A(z) = A_0 \exp\left(-\int_\mathscr{C} \mathrm{d}s\nabla^2\Phi/2|\boldsymbol{k}|\right) = A_0, \tag{52}$$

where the curve $\mathscr{C}$ is the ray path. For an adiabatic disc,

$$A(z)=(1-z^2)^{(1-n)/2}/\sqrt{\omega^2-k_x^2/2n(1-z^2)}. \quad (53)$$

Note that $A\to\infty$ as $z\to 1$, conserving $\rho A^2 k_z$ (the wave action here). For an isothermal disc, the amplitude $A(z)=A_0\exp(z^2-z_0^2)/4$, and increases with z to preserve the energy density of the wave. In a sheared medium, with $DU\neq 0$, by writing $\tau^2=(\boldsymbol{k}\cdot\boldsymbol{k})/k_x^2$ and using (46)–(48), τ satisfies

$$D(c\tau)+DU=0, \quad (54)$$

from which we obtain $\tau(z)=(\tau_0+M_0)\,c_0/c(z)-M(z)$, where M is the Mach number, $M(z)=U(z)/c(z)$. The angle between the wave normal and the ray, μ, is given by

$$\cos\mu=\frac{1}{\sqrt{1+M^2\pm 2M/\tau}}\,(\pm M/\tau). \quad (55)$$

These results can be used to solve for the eikonal Φ which is found, by using (54) and (55), to be

$$\Phi(t,\boldsymbol{x})=-\omega t+k_x\left[x\pm\int \mathrm{d}z\sqrt{\tau^2(z)-1}\right], \quad (56)$$

where k_x is a constant, and ω is a constant, from (48). The acoustic wave action flux $\boldsymbol{F}$ satisfies $\nabla\cdot\boldsymbol{F}=0$ away from the source, where $\boldsymbol{F}=\rho A^2/(\omega-\boldsymbol{k}\cdot\boldsymbol{U})\boldsymbol{v}_g$ (Landau & Lifshitz 1982). From (46), (47) and (54), we may find the ray-path in the $(x$–$z)$ plane, given by

$$\boldsymbol{x}-\boldsymbol{x}_0=\int \mathrm{d}s\,\boldsymbol{m}=\int \mathrm{d}z\,\frac{\sqrt{1+M^2\pm 2M/\tau}}{m_z}\,\boldsymbol{m}, \quad (57)$$

where $\boldsymbol{m}=\boldsymbol{v}_g/|\boldsymbol{v}_g|$, is now known; $\boldsymbol{v}_g$ includes the velocity of the flow and the velocity of energy transport in the local rest frame. This gives a large horizontal ray propagation distance for the initial value of τ, $\tau_0=1$, taking $n=1.5$, $a=1$ and $z_0=0$. However, if $\tau_0=2$ then this distance is $\sim H$, the disc scaleheight. Thermal stratification has a focusing effect for most of the rays, although they are usually reflected from their vertical extremes after going through length $\sim H$ only. While the traveltime for most of the waves is increased by refractive effects, most of them should not propagate long distances. This result has also been found numerically for the inward propagation of acoustic disturbances from the outer edge of an accretion disc (Lin et al. 1990).

In the supersonic regime, we may have zero-frequency rays (which are the shock characteristics), as follows. From (48), $(M^2-1)k_x^2=k_z^2$, and (47) gives

$$x-x_0=\pm\int_{z_0}^{z}\mathrm{d}z\sqrt{M^2(z)-1}. \quad (58)$$

Taking z_0 to be z_s, and the upper limit to be $z=1$, for a linear velocity profile $U(z)=az$, the horizontal shock characteristic extent is $L=x-x_0$, where L is a function of the shear rate a: $L\sim a\sqrt{n}$ for small a and $L\sim a\sqrt{n/2}$ for large a. For a warped accretion disc, from Section 3 we find the length ratio to be $L/R\sim\beta\gamma/\alpha^2<1$, so we do not expect that the shocks will either wrap around the disc at a given radius or significantly extend in the radial direction.

Here $1\geq|z|,|z_0|>|z_s|$, where z_s is the sonic layer given by $U(z)=c(z)$. The time for energy transport along the ray for the adiabatic case is given, for the same linear profile $U(z)=az$, by

$$t=\int \mathrm{d}s/c=\int_{z_0}^{z}\mathrm{d}zM/c=a/2\,\ln\left(\frac{1-z_0^2}{1-z^2}\right), \quad (59)$$

so that $t\to\infty$ as $|z|\to 1$, which indicates the deposition of energy of the disturbances near the surface. In practice, no boundary will have $\rho=0$, since the energy deposited near the boundary will heat the surface layers and blow them away or, in a steady state, create a corona.

4.3 Eigenmodes and the global dispersion relation in the WKBJ limit

It is possible to find analytically the frequency and eigenmodes of linear perturbations, e.g. from equation (13) with the accompanying boundary condition in the limit of large wavenumbers. This requires the use of the WKBJ method to join appropriate solutions in different parts of the flow, with special attention given to the change in phase. For the problem with which we are concerned, we require that the density or pressure scaleheight be much larger than the mode wavelength. The perturbed variables, e.g. the perturbed horizontal velocity v, are expanded in the form

$$v=\exp\left(\frac{1}{\varepsilon}\sum_{n=0}^{\infty}\varepsilon^n v_n\right), \quad (60)$$

where $\varepsilon=k^{-1}$, $k\to\infty$ is the expansion parameter. The first two terms, v_0 and v_1, provide the lowest order WKBJ approximation used here.

The method of joining the solutions is as follows. We write the WKBJ equations for v, the perturbed horizontal velocity. In the two boundary regions, v must satisfy the regularity conditions. The standing wave in one of the boundary regions is then asymptotically matched to the adjacent region, which is the wave-trapping region. If we start at the boundary, $z/\sqrt{2n}=1$, then A is the first turning point of the WKBJ equation. Here the mode acquires an exponentially growing part (the exponentially decaying, or subdominant, part can be neglected). There is then an additional contribution to this value in the critical layer, i.e. around the critical point $\bar{\sigma}\approx 0$. The mode then continues its exponential growth up to the next turning point, B. Here we have the second wave-trapping region, which must asymptotically match with the solution in the second boundary $(z/\sqrt{2n}=-1)$. A similar procedure may be followed for a wave from the second to the first boundary. From each such global solution, we find a dispersion relation for the real part of σ, σ_R, only. The term $\sigma_I<0$ is exponentially small, but can be estimated. It should be noted that the wave-trapping regions near the boundary mean that much of the wave energy exists here, and therefore makes the role of the boundary important. This is in contrast to the behaviour of the large-wavelength modes. These latter modes are substantially affected by the presence of the critical layer, as numerical work shows. Since these are the fastest growing modes [particularly $k\sim O(1)$], we also need to understand the behaviour of velocity profiles with special properties, e.g. no inflection point.

The WKBJ limit for the surface waves may be contrasted with instability fuelled by radiative losses, where waves

travelling in one direction, to infinity, establish the direction of the WKBJ corrections, e.g. for an infinite torus with an inner boundary, the energy and angular momentum losses in the WKBJ limit determine the computation direction from outside to the inner radius (Papaloizou & Pringle 1987). A related treatment is that of compressible, two-dimensional planar discs with one reflecting edge (Kato 1987), which show instability if the wave reflecting edge is on the same side of the corotation radius as the over-reflected wave. The problem considered in this section differs from the WKBJ treatment of the homogeneous shearing sheet (Narayan et al. 1987). They consider rigid boundaries at both ends, $z_B = \pm 1$. The density variation and free boundaries both affect the phase conditions (85) and (87) below. Moreover, their eigenequation differs in that it considers the effect of rotational shear, whereas here it is the vertical effect of gravity that is taken into account.

We now compute the global eigenfunction and find the dispersion relation. The variable used is the perturbed horizontal velocity v, which is labelled by the subscript r, as v_r ($r=1,\ldots,7$), and denotes the function in the respective region. The stationary shear flow has the profile $U=az$, and the fluid is adiabatic, with index $\gamma=1+1/n$. The length-scales are based on the factor $1/\sqrt{2n}$, so that the boundaries are $z_B=\pm 1$. This enables us to consider the limit $n\to\infty$, which is the isothermal case, and also enables an asymptotic determination of the frequencies.

Some notation is now in order for the brevity of the WKBJ approximations. $\bar{\sigma}_{B1}$, $\bar{\sigma}_{B2}$ denote the pattern frequencies at the two boundaries,

$$\bar{\sigma}_{B1}=\sigma+ka\sqrt{2n}, \qquad \bar{\sigma}_{B2}=\sigma-ka\sqrt{2n}. \tag{61}$$

The eigenfrequency $\sigma=\sigma_R+i\sigma_I$, but in the WKBJ limit σ_I is exponentially small and negative, and may be neglected compared to σ_R, so that $\sigma\approx\sigma_R$. We define constant a_1, a_2, b_1 and b_2 by

$$a_1=n\left(1+\frac{a\bar{\sigma}_{B1}}{\sqrt{2n}k}\right), \qquad a_2=n\left(1-\frac{a\bar{\sigma}_{B2}}{\sqrt{2n}k}\right), \tag{62}$$

$$b_1=n\bar{\sigma}_{B1}^2, \qquad b_2=n\bar{\sigma}_{B2}^2. \tag{63}$$

Two quantities, $\tilde{A}$ and $\tilde{B}$, are related to the turning points A and B, of the WKBJ equation,

$$D^2v=Q(z)v, \tag{64}$$

with $Q(z)=k^2-\bar{\sigma}^2/E$. The boundary condition is $Dv=\bar{\sigma}^2v/(\bar{\sigma}DU/k+z_B)$. Away from the critical layer, the perturbed vertical velocity w and pressure δp are given by

$$w\approx(-i/k)Dv, \tag{65}$$

$$\delta p\approx(-\rho\bar{\sigma}/k)v. \tag{66}$$

We write

$$\tilde{A}=1-A/\sqrt{2n}, \qquad \tilde{B}=1+B/\sqrt{2n}. \tag{67}$$

The two phase angles that arise from the asymptotic matching of the eigenfunctions near the two boundaries, and are crucial to the dispersion relation, are

$$\phi=\pi/4-a_1\pi/2-2\sqrt{b_1\tilde{A}}, \tag{68}$$

$$\psi=-\pi/4+a_2\pi/2+2\sqrt{b_2\tilde{B}}. \tag{69}$$

Regularity in the boundary region implies

$$v_1=(\sqrt{b_1\xi})^{(1-a)}J_a(2\sqrt{b_1\xi}), \tag{70}$$

where J_a is the Bessel function of order a. The asymptotic limit for matching gives

$$\lim_{\xi\to\infty} v_1=\frac{1}{\sqrt{\pi}}(b_1\xi)^{1/4}\cos(2\sqrt{b_1\xi}+a_1\pi/2-\pi/4), \tag{71}$$

from which we have, in the first wave-trapping region,

$$v_2=\frac{1}{\sqrt{\pi}}(b_1\bar{\sigma}_{B1}^2/2)^{1/4}\frac{1}{\sqrt[4]{-Q(z)}}\cos\left(\int_A^z dz\sqrt{-Q}+\phi\right). \tag{72}$$

Using the appropriate turning-point formula we find that, in the evanescent region,

$$v_3=\frac{1}{\sqrt{\pi}}(b_1\bar{\sigma}_{B1}^2/2)^{1/4}\sin(\phi+\pi/4)\frac{1}{\sqrt[4]{Q(z)}} \times\exp\left(\int_z^A dz\sqrt{Q}\right). \tag{73}$$

In the region around the critical layer $z=z_c$, with width $\sim k^{-2/3}$, we use the perturbed vertical velocity w as the appropriate variable to find the jump across z_c. Here $\bar{\sigma}\approx 0$, with a small but non-vanishing contribution from $\gamma=\Im(\sigma)$: $z_c\approx-\sigma_R/ka$. Using $\eta=z-z_c$, the WKBJ equation becomes

$$\frac{d^2v}{d\eta^2}=\left(\frac{1}{\eta^2}+k^2\right)v-\frac{1}{\eta}\frac{dv}{d\eta}. \tag{74}$$

Note that in this layer the variation of $\bar{\sigma}$ and ρ may be neglected. Writing $x=k\eta$, the solution for v becomes

$$v\approx d_1x+d_2[(\ln x/2+\gamma)x/2+1/x]. \tag{75}$$

The singularity prompts us to find a variable for which the solution near the critical point is regular, although its derivative may jump; conditions satisfied by the vertical velocity w, whose variation is governed by

$$\frac{d^2w}{d\eta^2}+q\frac{dw}{d\eta}+\left(\frac{q}{\eta}-k^2\right)w=0, \tag{76}$$

where $q=D\ln\rho$ is a constant. $\eta=0$ is a regular singular point of this equation. w is regular, but $dw/d\eta$ has a jump of magnitude $-i\pi c_2$, where c_2 is the coefficient of the $\eta\ln\eta$ part of w. The relation between v and w around the critical layer,

$$v\approx i(q\omega+Dw)/k, \tag{77}$$

gives the required condition on v for the jump as follows:

$$[v]_{z_c^+}^{z_c^-}=\pi\tilde{c}/k, \tag{78}$$

where $\tilde{c}$ is a constant related to the coefficient of the exponential term in (73). This enables us to find the subsequent value of v in the evanescent region before it approaches the next turning point B:

$$v_5=\frac{1}{\sqrt{\pi}}(b_1\bar{\sigma}_{B1}^2/2)^{1/4}\sin(\phi+\pi/4) \times(1+i\pi/k)\frac{1}{\sqrt[4]{Q(z)}}\exp\left(\int_z^A dz\sqrt{Q}\right). \tag{79}$$

We now use the second turning-point formula to find the form of v in the next wave-trapping region:

$$v_6=\frac{2}{\sqrt{\pi}}(b_1\bar{\sigma}_{B1}^2/2)^{1/4}\sin(\phi+\pi/4)(1+i\pi/k) \tag{80}$$
$$\times\frac{1}{\sqrt[4]{-Q(z)}}\exp\left(\int_B^A dz\sqrt{Q}\right)\cos\left(\int_z^B dz\sqrt{-Q}-\pi/4\right).$$

A similar procedure can start from v_7 and go to v_1. Thus we have two sets of trapped modes, and their respective slow leakages, i.e. the coupling between the modes near the boundaries. We write these as follows:

$$v_2^1=\frac{1}{\sqrt{\pi}}(b_1\bar{\sigma}_{B1}^2/2)^{1/4}\frac{1}{\sqrt[4]{-Q(z)}}\cos\left(\int_A^z dz\sqrt{-Q}+\phi\right), \tag{81}$$

$$v_2^2=\frac{2}{\sqrt{\pi}}\left(\frac{b_2\bar{\sigma}_{B2}^2}{2}\right)^{1/4}\left(1-\frac{i\pi}{k}\right)\sin\left(\frac{\pi}{4}-\psi\right) \tag{82}$$
$$\times\exp\left(\int_B^A dz\sqrt{Q}\right)\frac{1}{\sqrt[4]{-Q(z)}}\cos\left(\int_A^z dz\sqrt{-Q}-\pi/4\right),$$

for the first wave-trapping region, while for the second wave-trapping region we find

$$v_6^1=\frac{2}{\sqrt{\pi}}\left(\frac{b_1\bar{\sigma}_{B1}^2}{2}\right)^{1/4}\left(1-\frac{i\pi}{k}\right)\sin\left(\frac{\pi}{4}+\phi\right) \tag{83}$$
$$\times\exp\left(\int_B^A dz\sqrt{Q}\right)\frac{1}{\sqrt[4]{-Q(z)}}\cos\left(\int_z^B dz\sqrt{-Q}-\pi/4\right),$$

$$v_6^2=\frac{1}{\sqrt{\pi}}(b_2\bar{\sigma}_{B2}^2/2)^{1/4}\frac{1}{\sqrt[4]{-Q(z)}}\cos\left(\int_z^B dz\sqrt{-Q}-\psi\right). \tag{84}$$

The phases in (81) and in (83) should match, particularly for large k. From (61), we find the first quantum condition for σ:

$$\phi+\pi/4=2m\pi, \tag{85}$$

where m is an integer. Since this eigenvalue equation is complicated, we consider the limit $n\to\infty$, and the shear a, such that $na^2\to\infty$. We write $\tilde{\sigma}=\sigma/(\sqrt{2nka})\approx\sigma_R/(\sqrt{2nka})$. Then condition (62) gives

$$\tilde{\sigma}=\pm\frac{\pi}{4}[\sqrt{n}+(4m-1)/\sqrt{n}]-1 \tag{86}$$

for all integer values of m. A similar computation gives the eigencondition

$$\psi-\pi/4=2l\pi, \tag{87}$$

where l is any integer. With an approximation similar to that used for (86), we find that

$$\tilde{\sigma}=\pm\frac{\pi}{4}[(4l+1)/\sqrt{n}-\sqrt{n}]+1 \tag{88}$$

for all integer l values. Note, however, that $|\tilde{\sigma}|<1$ for modes that have a critical layer in the flow, and these can therefore be unstable. If the standing-wave profile is computed from the critical layer towards the boundaries, the same phase relationship and dispersion relations obtain. The growth rate, $-\sigma_I$, is of the order of the energy leakage rate of the mode, divided by the original trapped energy. We see from (61) and (62) that, for the modes that couple,

$$-\sigma_I\sim\exp\left[-2\int_B^A dz\sqrt{Q(z)}\right]\approx\exp(-k/a). \tag{89}$$

This shows that, for high wavenumber, the growth rate is exponentially small, and also that as the shear increases the growth rate increases. A qualitative extension of (89) suggests that, for comparable shear rates and wavenumbers, the growth rates can be on truly dynamical time-scales. However, the limit $a\to\infty$ cannot be reliably estimated for finite but large k.

5 NUMERICAL WORK

As a first step towards investigating the stability of a warped accretion disc, it is necessary to perform a linear analysis of an idealized model of the local flow field. In view of the results described above, the model chosen is a two-dimensional shear layer with linear velocity profile, and the vertical pressure and density structure of an adiabatic thin disc. Free surface boundary conditions are imposed at the top and bottom of the shear layer (Fig. 1). Note that, although this model neglects the effects of orbital curvature and shear due to differential rotation, both of which should be small in the outer disc region of interest here, the two-dimensional disturbances have the fastest growth rates in this approximation, as shown in Section 3 (the version here of Squire's result; see Drazin & Reid 1981). The eigenvalue problem is described in Section 5.1. The numerical results are given in Section 5.2 and Figs 2–12.

5.1 The eigenvalue problem

The flow model is expressed in terms of dimensionless variables using the disc half-height H, the inverse orbital frequency Ω^{-1}, the central density ρ_0 and the quantity $\rho_0H^2\Omega^2$ as the units of distance, time, density and pressure, respectively. The shear velocity is then given by $U=az$, density $\rho=(1-z^2)^n$ and pressure $p=(1-z^2)^{n+1}/2(n+1)$, where $n=1/(\gamma-1)$ is the polytropic index. The model is specified by two parameters, the shear parameter a which is related to the degree of disc warping, and the polytropic index n. In all numerical calculations, a value of $n=3/2$ is used, corresponding to an adiabatic index $\gamma=5/3$, as appropriate for a non-relativistic monatomic gas.

The fluid equations describing the flow model are linearized subject to perturbations of the form $f(z)\exp[i(kx+\sigma t)]$, where $f(z)$ is the z-dependence of the perturbation of any hydrodynamic variable f, k is the (real) wavenumber, and σ is the (complex) eigenfrequency to be determined. After eliminating all variables in favour of the enthalpy $X(z)$, the system reduces to a second-order ordinary differential equation:

$$D^2X-\left[\frac{2ka}{\bar{\sigma}}+\frac{2nz}{(1-z^2)}\right]DX+\left[\frac{2n\bar{\sigma}^2}{(1-z^2)}-k^2\right]X=0, \tag{90}$$

where $\bar{\sigma}=\sigma+kaz$ is the Doppler-shifted wave frequency and $DX=dX/dz$. The free surface boundary conditions require that $DX=z\bar{\sigma}X$ at $z=\pm1$, and the system is closed by the continuity of both X and DX.

Numerical solution of equation (90) is facilitated by expressing it in terms of a new variable, $P=w/X$, where $w=iDX/\bar{\sigma}$ is the perturbed z-velocity. This leads to a Ricatti equation:

$$DP=i\left[\frac{k^2}{\bar{\sigma}}-\frac{2n\bar{\sigma}}{(1-z^2)}\right]+\left[\frac{2nz}{(1-z^2)}+\frac{ka}{\bar{\sigma}}\right]P+i\bar{\sigma}P^2, \qquad (91)$$

where the free surface boundary conditions require that $P=i\bar{\sigma}z$ at $z=\pm1$, and P must be continuous.

Note from symmetry that it is sufficient to consider only positive values of k and a. Further, solutions occur as pairs, $\sigma=\pm\sigma_R+i\sigma_I$, corresponding to identical waves propagating in opposite directions. When posed as an initial-value problem, the unstable modes contribute to the disturbance as before, but the neutral and decaying modes are affected. Eigenvalues σ are obtained by numerically integrating equation (91) from each of the boundaries, and then iterating σ by linear interpolation until the two parts of the solution match at an intermediate point. Since equation (91) is singular at the boundary points, series solutions about $z=\pm1$ are used to start the integrations.

Resonance-driven instability may occur if the flow contains a critical layer, where the wave phase velocity is equal to the shear flow velocity (Drazin & Reid 1981; Craik 1988). Such a layer exists if $|\Re(\sigma)|<ka$, in which case the singularity occurs within the domain of integration. (This is referred to as the corotation singularity, by analogy with that which occurs in differentially rotating systems.)

The integration proceeds from $z=-1$ to $z=1$ along a path in the complex z-plane, chosen in accordance with the Landau prescription from the requirement that the solution be that of a valid initial-value problem (e.g. Drazin & Reid 1981). The path must pass below the corotation singularity in the complex z-plane. The integration is along the real z-axis for unstable solutions, as the singularity lies in the upper half-plane. It deviates from the real axis to pass below the singularity for the damped solutions. The unstable modes are of interest in the problem studied here.

5.2 Numerical results

In the simple case of an incompressible, static fluid layer with no shear, $n=a=0$, equation (91) can be solved to obtain two non-trivial stable modes, $\sigma^2=k\tanh k$ and $\sigma^2=k\coth k$. Numerical solutions can be sought for the generalizations of these modes for $n=3/2$ and $a\neq0$. Since sonic instabilities are expected to occur above $a\approx1$, we have the mode structure for values of the shear parameter $a=0$, 1, 2 and 5. In addition, we have investigated the behaviour of the solution at wavenumber $k=1$ as a function of the shear parameter a in the range $0\leq a\leq5$.

The real and imaginary parts of the eigenfrequencies are plotted as functions of wavenumber k for values of shear parameter $a=0$, 1, 2 and 5 in Figs 2, 3, 4 and 5 respectively. In the static case ($a=0$), the frequencies are real for all wavenumbers. In Fig. 2, the solution for $n=1.5$ has been plotted alongside the analytic result for $n=0$ given above.

For $a>0$, the mode emerging from $\sigma(0)=1$ still exists. When this mode crosses the line $\Re(\sigma)=ka$, corresponding to the entry of a critical layer in the flow, it becomes complex. The imaginary part may vary either discontinuously (as for $a=1$) or continuously (as for $a=5$) from zero at this point, and is always positive (corresponding to a damped solution) when the critical layer enters the flow. The discontinuity in the imaginary part of the frequency for $a=1$ may be a numerical artefact, but is of little relevance to the study of the instability, since it affects only the damped solutions.

The value of $\Re(\sigma)$ decreases with k until it vanishes. This point corresponds to a crossing at the two modes representing identical waves propagating in opposite directions. Both become standing waves [$\Re(\sigma)=0$], but one is strongly damped while the other is strongly unstable. At a larger wavenumber the modes again decouple, and $\Re(\sigma)$ increases with the wavenumber k without bound.

The mode emerging from $\sigma(0)=0$ in the static flow splits into a sequence of modes emerging from the line $\Re(\sigma)=ka$ when $a\neq0$. In Figs 3, 4 and 5 only the first mode in this sequence is represented, to avoid the plots becoming confused. These modes behave in a mannner identical to that emerging from $\sigma(0)=1$, in that the solution is damped when the critical layer first enters the flow, $\Re(\sigma)$ decreases with k until it vanishes, a pair of damped and unstable standing waves occurs and then $\Re(\sigma)$ increases without bound.

This simple situation is complicated by the fact that different modes undergo crossings at points well away from $\Re(\sigma)=0$. Several such crossings are represented in Figs 3, 4 and 5. In each case the value of $\Re(\sigma)$ varies only slowly with wavenumber in the crossing region, and one mode is destabilized while the other is damped. For $a=1$, which may be

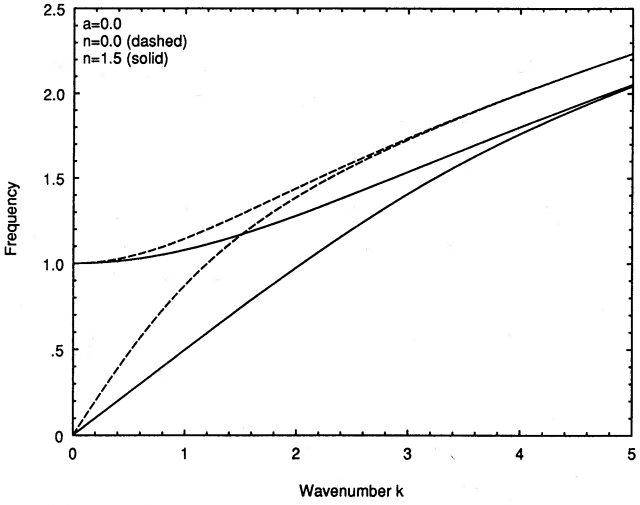

Figure 2. The eigenfrequency for a static layer of fluid for two values of the polytropic index $n=0$ (dashed) and $n=1.5$ (solid). The frequency increases with the wavenumber k. There are only four modes for $n=0$, the homogeneous, incompressible fluid. This can also be seen from equation (48) in Section 4.1. From equation (52), it can be seen that, when $n=\infty$, i.e. when the fluid is isothermal, the number of modes is infinite for each wavenumber k. Only two modes are shown for the polytropic index $n=1.5$, although there is a countable infinity of them, as indicated in Section 4.1. All positive real eigenfrequencies are shown; there are corresponding negative values as well.

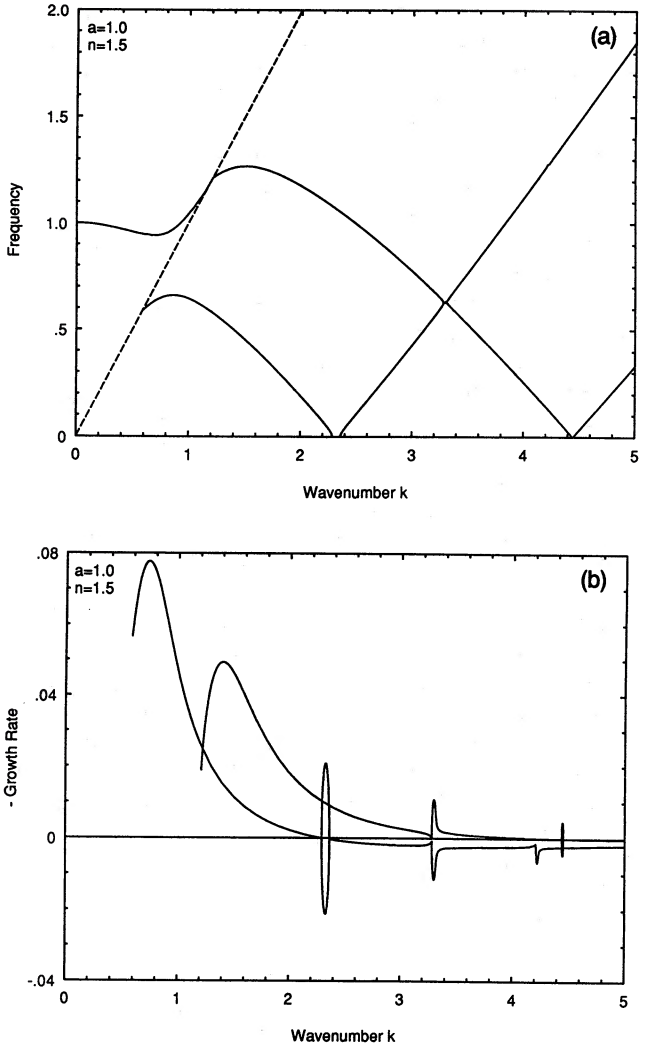

Figure 3. The real part (a) and imaginary part (b) of the eigenfrequencies are plotted for the first two modes for $a=1.0$ and $n=1.5$. Narrow instability bands occur where the eigenvalues cross.

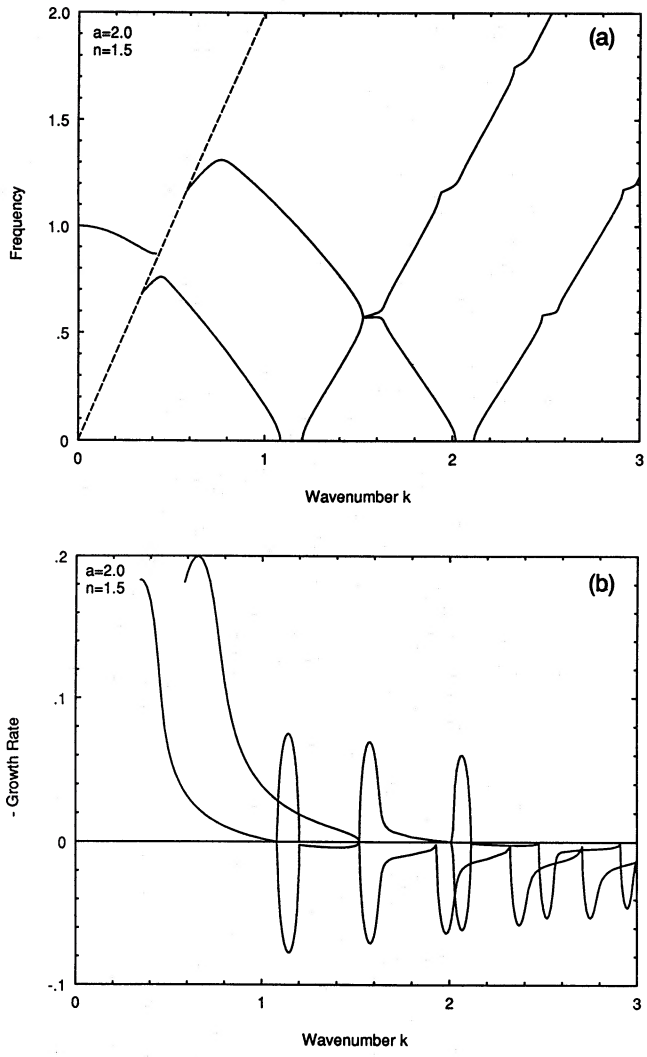

Figure 4. As Fig. 3, but for $a=2.0$. The appearance of further bands and more complicated interaction patterns can be seen.

typical of a weakly warped disc, the instability occurs at a wavenumber of $k=2.33$ in the first standing-wave band. The corresponding frequency is $\sigma=-0.021121\mathrm{i}$. Subsequent instability peaks decrease rapidly with wavenumber, so this is likely to be the global instability peak for $a=1$. This corresponds to an exponential growth time $\tau\approx 7.5T$, where $T=2\pi/\Omega$ is the orbital period. For $a=5$, a value typical of a relatively strongly warped disc, the instability peaks for each mode crossing have a similar magnitude. The lowest wavenumber peak occurs at $k=0.46$, where $\sigma=-0.112782\mathrm{i}$, corresponding to an exponential growth time $\tau\approx 1.4T$. This is clearly a highly dynamically unstable situation, and would be expected to lead to either a radical modification of the disc structure, or possibly local disruption if non-linear limiting of the disturbance does not take place.

The behaviour of the lowest frequency modes as a function of the shear parameter is presented in Fig. 6. In Fig. 6(a), a more complete picture of the mode-crossing behaviour is apparent. Referring to the imaginary part of the eigenfrequency in Fig. 6(b), it may be noted that those branches of a mode for which $\Re(\sigma)$ decreases with wavenumber generally correspond to damped solutions, while those for which $\Re(\sigma)$ increases with wavenumber correspond to unstable solutions.

The nature of the instability is understood in terms of mode coupling (Craik 1988). When the fluid is static, all modes are neutral. However, as the shear increases, the eigenvalues of the different modes are dragged around in the (a, σ_R) plane. When the two modes cross, an unstable mode begins. More precisely, the mode-coupling behaviour is stated as

$$D_1(\sigma)D_2(\sigma)=\varepsilon(\sigma), \tag{92}$$

where $D_1(\sigma_1)=0$, $D_2(\sigma_2)=0$ are the dispersion relations for each of the modes independent of the other, while ε is the weak coupling that exists between them, here due to the finite vertical extent of the shear flow. If the mode-crossing (i.e. critical) frequency is σ_c, then writing $\delta=\sigma_2-\sigma_1$ and $\Delta=\sigma-\sigma_1$, and Taylor-expanding the dispersion relation, gives the mode coupling equation

$$\Delta(\Delta-\delta)\frac{\partial D_1}{\partial\sigma}\frac{\partial D_2}{\partial\sigma}\approx\varepsilon. \tag{93}$$

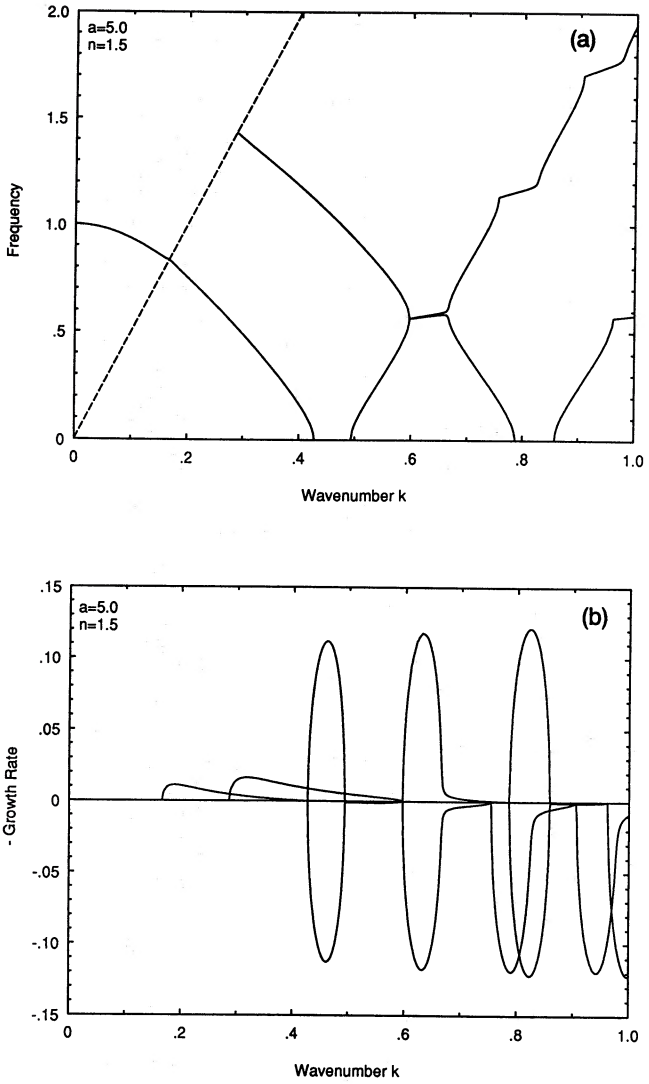

Figure 5. As Fig. 3, but for $a=5.0$. The appearance of further bands and more complicated interaction patterns can be seen.

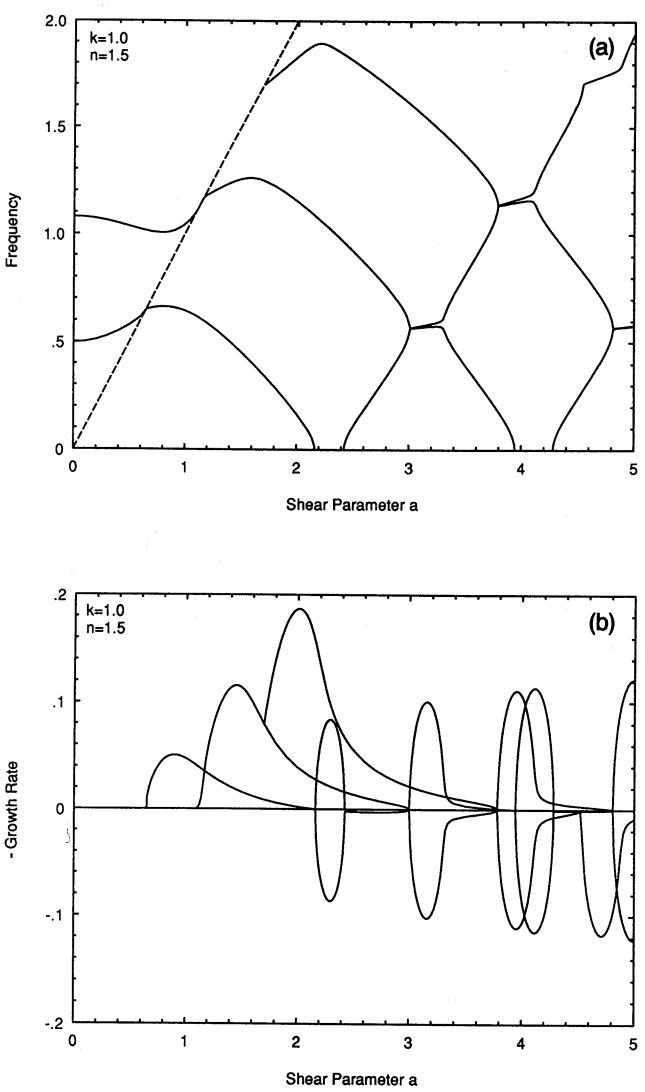

Figure 6. The eigenvalues are plotted as a function of the shear rate a for wavenumber $k=1$ for the polytrope with $n=1.5$. (a) Shows the real part and (b) the imaginary part of the eigenfrequencies as they are dragged by the shear. This leads to the crossings associated with the instability.

The value of Δ depends on the relative signs of ε, $\partial D_1/\partial\sigma$ and $\partial D_2/\partial\sigma$. The mode crossing is denoted by Δ acquiring an imaginary part, while avoided crossings [i.e. $\Im(\Delta)=0$] mean that there is no instability. In the simpler cases of incompressible flows (Craik 1988), or when the fluid is homogeneous (Glatzel 1988), it is possible to compute these relations analytically, but for the case studied here this approach is too complicated to be useful.

For completeness, the eigenfunctions for the perturbed pressure, density, the x- and the z-velocities, the wave action and its vertical flux (defined in the following section) are presented in Figs 7–12 for the parameter values $a=1.0$, $k=2.33$, $\sigma=-0.021121\mathrm{i}$: the instability peak for the $a=1$ case. The solid line represents the real part of the eigenfunction, the dashed line represents the imaginary part and the critical layer occurs at $z=0$. Note that the real part of the vertical action flux changes in sign at the critical layer, lending weight to the interpretation of this instability in terms of the creation of wave action at the critical layer.

6 CONCLUSIONS

An insight into the physical basis of the instability may be gained from earlier work on the closely related problem of the stability of a uniform shear layer (treated by, e.g., Narayan et al. 1987). In that case also, instability occurred only if the flow contained a critical layer on which the pattern speed of the disturbance vanished. The source of this instability was traced to the existence of a conserved action, the sign of which changed across the critical layer. Conservation of such an action clearly requires that a wave incident on one side of the critical layer be reflected with increased amplitude. The critical layer therefore acts as a 'corotation amplifier' and, with the addition of feedback by reflection from the fluid surface, instability is unavoidable. This mechanism operates in the regions both above and below the critical layer, each of which behaves as a pumped resonant cavity.

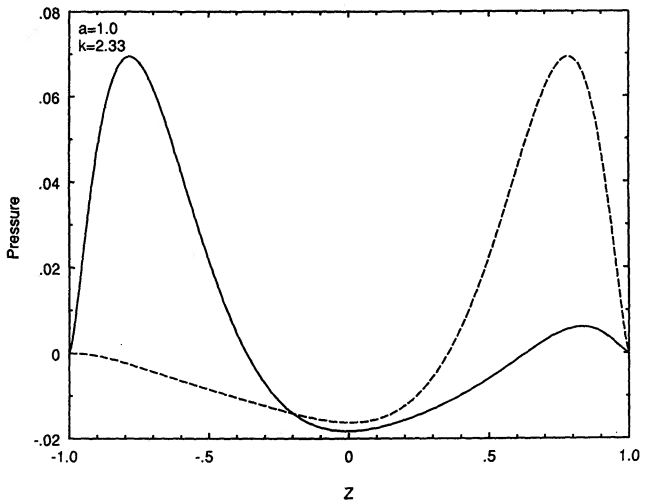

Figure 7. The perturbed pressure eigenfunction is plotted for the parameters $a=1.0$ and $k=2.33$. These are associated with the fastest growing instability for these values of the parameters. The solid line denotes the real part while the dashed line denotes the imaginary part.

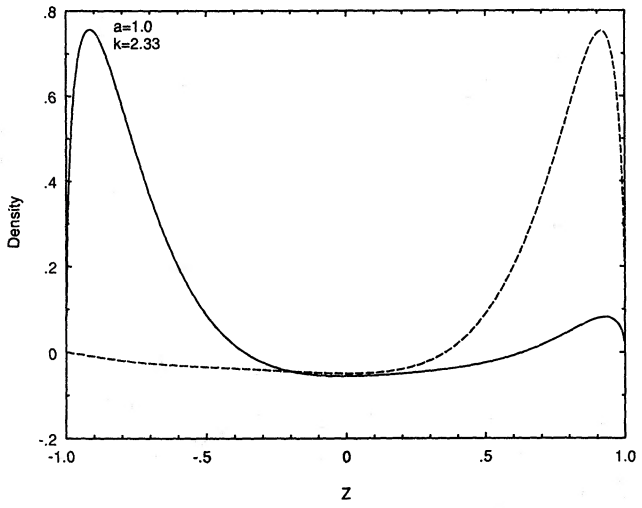

Figure 8. As Fig. 7, but for the perturbed density.

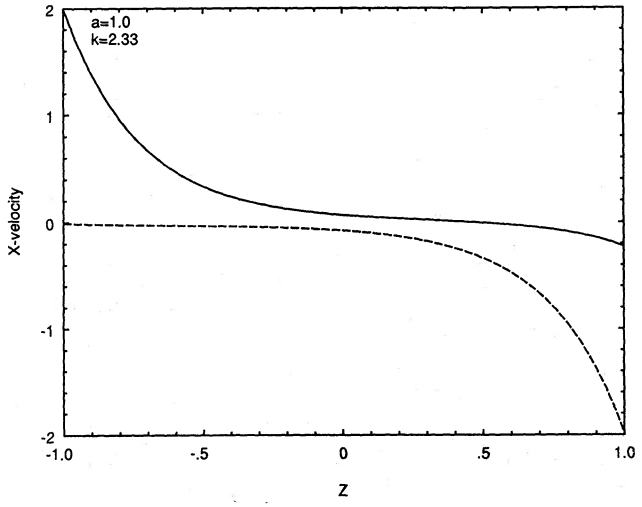

Figure 9. As Fig. 7, but for the perturbed x-velocity

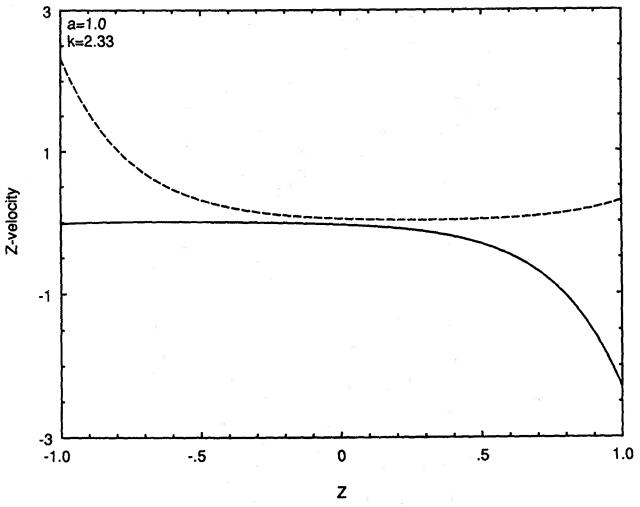

Figure 10. As Fig. 7, but for the perturbed z-velocity

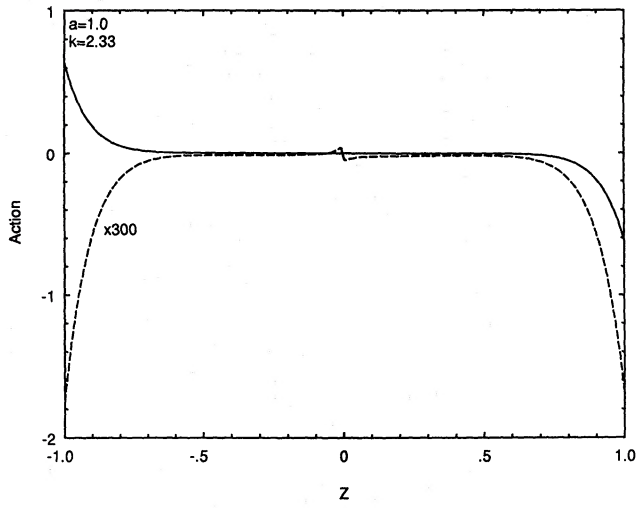

Figure 11. As Fig. 7, but for the wave action. The imaginary part of the action has been multiplied by 300 to make it visible.

In the uniform shear layer there are many modes, most of which are neutrally stable, with instability occurring only when a stringent phase condition is obeyed. In the present case, however, the density non-uniformity allows part of the (previously conserved) action to be absorbed at the critical layer (Drury 1985). For a neutral mode, the action is $A=\varepsilon/\bar{\sigma}$, where ε is the wave energy density,

$$\varepsilon=\rho\langle\partial\boldsymbol{v}^2+\delta h^2/c^2\rangle/2,$$

and the flux is $\boldsymbol{F}$ given in Section 4.2 and satisfies, in general, the conservation law

$$\frac{\partial A}{\partial t}+\nabla\cdot\boldsymbol{F}=0. \tag{94}$$

The angular brackets denote the phase-averaging over one horizontal wavelength of the disturbance (Drazin & Reid 1981), i.e., for a function α of perturbed variables,

$$\langle\alpha\rangle=\frac{2\pi}{k}\int_0^{k/2\pi}\alpha\,\mathrm{d}x. \tag{95}$$

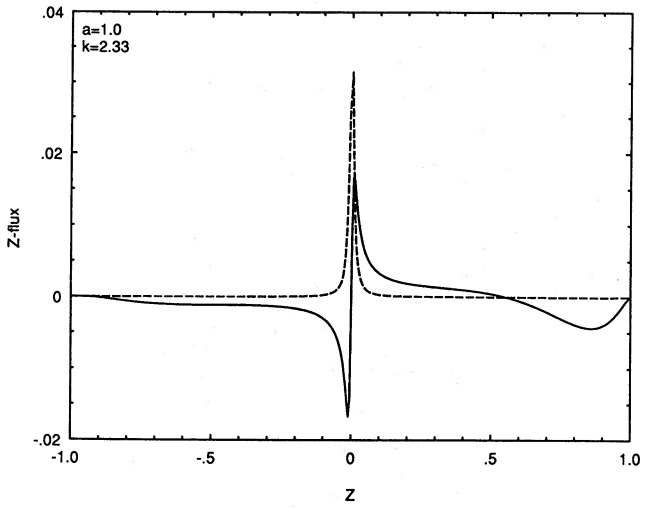

Figure 12. As Fig. 11, but for the vertical flux of the wave action. Note the change of sign at the critical layer $z=0$ for this mode. The asymmetry with respect to z reflected in the figures may be seen from the perturbation equation (90) and the boundary conditions. For $\bar{\sigma}=\sigma_I$, the eigenvalue equation is not invariant under the transformation $z\rightarrow -z$, unless $a\rightarrow -a$. If the shear flow with the horizontal velocity profile $U(z)=-az$ were studied, all the plots of the perturbed quantities with respect to z would flip accordingly.

If $\zeta=\mathrm{i}P$ in the Ricatti equation (91) then, for neutral modes, all coefficients are real, and $\mathrm{sgn}F=\mathrm{sgn}\,\Im(\zeta)$. The integration is indented below the singularity at corotation, the only place where the sign of F can change. Thus wave action is created and destroyed in this layer. For neutral modes, $\nabla\cdot\boldsymbol{F}=0$, i.e. the wave action is conserved, a result particularly useful for the rays treated in Section 4.2. The instability can only take place when the flux is directed away from the corotation layer. As indicated by Narayan et al., this has the effect of splitting neutrally stable modes into pairs of unstable and decaying modes such as those described above. Thus the non-uniform density profile of an adiabatic shear layer is an additional requirement for instability of many modes. Another interpretation of the instability is obtained by writing the energy evolution equation for the modes in terms of the change in the wave energy density, ε, and the perturbed velocities and pressure, where the angular brackets again denote the phase average:

$$\frac{\partial\varepsilon}{\partial t}=-2\sigma_I\varepsilon-\rho\mathrm{D}U\langle v\,w\rangle-\mathrm{D}\langle w\,\delta p\rangle. \tag{96}$$

The second term on the right-hand side is the coupling of the Reynolds stress to the shear of the mean flow, while the last term represents the energy transport by the pressure fluctuations. The tapping of the energy in the mean flow by the shear can lead to instability, or decay of the mode. Written in terms of the perturbed enthalpy, δh (i.e. X), it can be seen that terms proportional to $-\sigma_I$ also contribute to the right-hand side of this equation. The work from the pressure fluctuations can be integrated out in the global effect. We define E as the total energy of the disturbance, i.e. as the integral of ε over the volume of the fluid, or equivalently over z, where $1\geq z\geq -1$. Then the global energy equation becomes

$$\frac{\mathrm{d}E}{\mathrm{d}t}=-2\sigma_I E-\sigma_I f\left[\int_{-1}^{1}\mathrm{d}z\,\frac{\rho|\mathrm{D}U|^2|\mathrm{D}\,\delta h|^2}{|\bar{\sigma}|^4}+\sum_{\text{boundary}}\rho|\delta h|^2\right]$$

$$-kf\int_{-1}^{1}\mathrm{d}z\,\frac{\rho\mathrm{D}U}{|\bar{\sigma}|^2}\Im(\delta h^*\mathrm{D}\,\delta h), \tag{97}$$

where the boundary contribution is always positive apart from the sign of σ_I. The factor f takes the growth or decay of the waves into account, $f=\exp(-2\sigma_I t)$. The wave energy is conserved for the neutral modes, and its time variation is an indicator of instability. The last term on the right-hand side above vanishes for the neutral mode; otherwise, together with the global wave energy E, it provides an indicator of instability.

The principal conclusion of this work is that a two-dimensional linear shear layer, with the vertical density structure of an adiabatic thin accretion disc and free surface boundary conditions, is unstable on a dynamical time-scale. Note that the total shear velocity across the flow, Δv, is related to the disc Keplerian velocity, v_ϕ, by the expression $\Delta v/v_\phi\approx a\gamma$, where γ is the disc opening angle. For a thin disc with $\gamma\ll 1$, therefore, the models treated here represent discs with relatively weak warping.

The effect of this instability on an accretion disc is difficult to deduce from the simpler linear analysis described here, but it is clear that, in a low-viscosity disc, $\alpha\ll 1$, the instability grows rapidly for disc tilt angles $\beta>\alpha^2$. In this case, therefore, only an exceptionally well-aligned disc is stable. The non-linear development may lead to the violent disruption of the disc, or it may introduce additional dissipation as it saturates, which increases the effective viscosity parameter α. In the earlier case, there is no guarantee that the flow model used here is a good representation of the local disc flow. Since such a disc is certain to have the basic requirements of a shear flow and a non-uniform density profile, however, it is reasonable to expect that disruption may not occur if the disc is not sufficiently misaligned.

ACKNOWLEDGMENTS

We should like to thank J. E. Pringle for suggesting the problem and for a clarification, H. C. Spruit for discussion and D. W. Sciama for his kind hospitality at SISSA, Trieste, where this work was started. SK would also like to thank E. Meyer-Hofmeister, F. Meyer, R. Kippenhahn, J. Binney and R. J. Tayler for their kind hospitality at Munich, Oxford and Sussex respectively. We thank the referees for comments and a correction.

REFERENCES

Bender C., Orszag S., 1978, Advanced Mathematical Methods for Scientists and Engineers. McGraw-Hill Kogakusha, New York
Chimonas M., 1970, J. Fluid Mech., 43, 833
Craik A., 1988, Wave Interactions in Fluids. Cambridge Univ. Press, Cambridge
Drazin P., Reid W., 1981, Hydrodynamic Stability. Cambridge Univ. Press, Cambridge
Drury L. O'C., 1985, MNRAS, 217, 821
Glatzel W., 1988, MNRAS, 231, 795

Gough D., 1987, Lectures at the Les Houches 1987 Summer School on Astrophysical Fluid Dynamics
Grinfeld M. A., 1984, Geophys. Astrophys. Fluid Dyn., 28, 31
Howard L. N., 1961, J. Fluid Mech., 10, 509
Kato S., 1987, PASJ, 39, 645
Kumar S., 1988, MNRAS, 233, 33
Kumar S., 1992, Geophys. Astrophys. Fluid Dyn., 61, 235
Landau L., Lifshitz E., 1982, Fluid Mechanics. Pergamon Press, Oxford
Lin C. C., 1955, The Theory of Hydrodynamic Stability. Cambridge Univ. Press, Cambridge
Lin D. N. C., Papaloizou J., Savonije G., 1990, ApJ, 364, 326
Meyer F., 1986, in Mihalas D., Winkler K., eds, Radiation Hydrodynamics in Stars and Compact Objects. Springer-Verlag, Berlin, p. 249
Meyer F., Meyer-Hofmeister E., 1984, A&A, 128, 420
Miles J. W., 1961, J. Fluid Mech., 10, 496
Narayan R., Goldreich P., Goodman J., MNRAS, 228, 1
Papaloizou J., Pringle J. E., 1983, MNRAS, 202, 1181
Papaloizou J., Pringle J. E., 1987, MNRAS, 225, 267
Pringle J. E., 1981, ARA&A, 19, 137
Shakura N., Sunyaev R., 1973, A&A, 24, 337