J. Fluid Mech. (1994), vol. 264, pp. 165–184

Nonlinear stability of multilayer quasi-geostrophic flow

By MU MU[1], ZENG QINGCUN[1], THEODORE G. SHEPHERD[2] AND LIU YONGMING[3]

[1] LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100080, China

[2] Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada

[3] Institute of Mathematics, Anhui University, Heifei 230039, China

(Received 24 June 1993 and in revised form 24 September 1993)

New nonlinear stability theorems are derived for disturbances to steady basic flows in the context of the multilayer quasi-geostrophic equations. These theorems are analogues of Arnol'd's second stability theorem, the latter applying to the two-dimensional Euler equations. Explicit upper bounds are obtained on both the disturbance energy and disturbance potential enstrophy in terms of the initial disturbance fields. An important feature of the present analysis is that the disturbances are allowed to have non-zero circulation. While Arnol'd's stability method relies on the energy–Casimir invariant being sign-definite, the new criteria can be applied to cases where it is sign-indefinite because of the disturbance circulations. A version of Andrews' theorem is established for this problem, and uniform potential vorticity flow is shown to be nonlinearly stable. The special case of two-layer flow is treated in detail, with particular attention paid to the Phillips model of baroclinic instability. It is found that the short-wave portion of the marginal stability curve found in linear theory is precisely captured by the new nonlinear stability criteria.

1. Introduction

Arnol'd (1965, 1966) established two theorems for the nonlinear stability of steady solutions of the two-dimensional Euler equations. His method is essentially a finite-amplitude extension of the variational technique of Fjørtoft (1950), and is based on the construction of a conserved functional, usually the energy plus a suitably chosen Casimir invariant (which for the two-dimensional Euler equations consists of the spatial integral of a function of the vorticity), which is sign-definite for arbitrary perturbations. The basic state is then an extremum of the conserved functional. Arnol'd's first stability theorem corresponds to cases where the conserved functional is positive definite, while the second theorem corresponds to cases where it is negative definite. Since the method is general and can be cast in terms of Hamiltonian theory, it can be applied to any fluid system which has an underlying Hamiltonian structure (see e.g. Holm *et al.* 1985; Shepherd 1990).

Generally speaking, the establishment of an analogue of Arnol'd's second theorem is much more difficult than is the case with the first theorem. For the multilayer quasi-geostrophic equations, a widely used model for describing large-scale atmospheric and oceanic dynamics (e.g. Pedlosky 1979), an analogue of Arnol'd's first theorem was proved by Holm *et al.* (1985), and significantly generalized to incorporate momentum conservation by Zeng (1989) and Ripa (1992). Analogues of Arnol'd's second theorem have recently been established for this system by Mu (1991) and Ripa (1992). The present work continues this line of investigation and derives new nonlinear stability criteria analogous to Arnol'd's second theorem, which are superior to those derived in the aforementioned papers. The results establish rigorous upper bounds on both the energy and potential enstrophy of finite-amplitude disturbances to steady basic states, which are expressed in terms of the initial disturbance fields. These bounds hold uniformly in time, and tend to zero uniformly as the initial disturbance amplitude decreases to zero. It follows that the bounds establish nonlinear (normed) stability of the basic state. This analysis occupies §3.

The results are applied in §4 to the important case of the two-layer model. Particular attention is paid to the classical Phillips model of baroclinic instability. According to linear theory, the Phillips basic state is unstable for sufficiently large vertical wind shear provided that the disturbance wavenumber is not too large. The minimum critical shear for instability corresponds to violation of the nonlinear Charney–Stern stability criterion (Shepherd 1988). It is found here that the short-wave portion of the marginal stability curve is precisely captured by the new nonlinear stability criteria, thereby rigorously explaining the existence of a maximum wavenumber for normal-mode instability at any given shear.

While Arnol'd's stability method relies on the conserved energy–Casimir functional being sign-definite, our new criteria can be applied to cases where it is sign-indefinite because of the disturbance circulations. This fact is highlighted by the construction of an explicit example in §4.3.

Andrews (1984) showed that if the flow domain is zonally symmetric (e.g. a zonal channel), then any basic state that is stable by Arnol'd's first stability theorem must itself be zonally symmetric. There are generally two ways to prove Andrews' theorem: the first is based on the fact that an Arnol'd-stable flow is an extremum of the energy–Casimir functional (Carnevale & Shepherd 1990), while the second is based on explicit integral inequalities. The first approach is not applicable to the present case because the energy–Casimir functional need not be sign-definite. However, by modifying the argument of Andrews (1984) we prove an analogue of his result for the stability criteria obtained in §3. This analysis is presented in §5.

The case of uniform potential vorticity flow is not generally accessible to Arnol'd's theorems, although it is well known that such a flow is stable to normal-mode disturbances. It is shown in §6 that the present analysis nevertheless can be applied, and can be used to prove the nonlinear stability of uniform potential vorticity flows.

The stability criteria are summarized in §7.

2. Governing equations

We consider a stably stratified fluid of N superimposed layers of constant density $\rho_1 < \ldots < \rho_N$, with equal density jumps $\rho_{i+1} - \rho_i = \rho'$, and mean layer depths d_i. The flow is presumed to be governed by the multilayer quasi-geostrophic potential vorticity equation (e.g. Pedlosky 1979)

$$\frac{\partial P_i}{\partial t} + \partial(\Phi_i, P_i) = 0 \quad (i = 1, \ldots, N), \tag{2.1}$$

where $\Phi_i(x, y, t)$ is the stream function in layer i, and

$$P_i(x, y, t) = \nabla^2 \Phi_i + F_i \sum_{j=1}^{N} T_{ij} \Phi_j + f_i(x, y) \quad (i = 1, \ldots, N) \tag{2.2}$$

is the potential vorticity in layer i. In the above, $\partial(f,g) \equiv f_x g_y - f_y g_x$ is the two-dimensional Jacobian, x and y are respectively the eastward and northward coordinates, t is time, $F_i = f_0^2 \rho_0 / g\rho' d_i$ is an inverse stratification parameter (the so-called 'rotational Froude number'), where f_0 is a representative value of the Coriolis parameter, ρ_0 is a mean density, and g the gravitational acceleration;

$$f_i(x,y) = \delta_{iN}(f_0/d_N)\,h(x,y) + f_0 + \beta y$$

represents the combined effects of topography h in the lowest layer and the Coriolis term, where δ_{iN} is the Kronecker delta; and the matrix T_{ij} is given by

$$T_{ij} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -2 & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & -2 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & -1 \end{pmatrix}. \tag{2.3}$$

Because the density jumps have been taken to be equal, we have the condition

$$d_i F_i = \frac{f_0^2 \rho_0}{\rho' g} = \text{constant}\ \forall i. \tag{2.4}$$

The horizontal domain D under consideration is a bounded, multiply (or simply) connected domain on the beta-plane, with a smooth boundary ∂D consisting of $J+1$ simple closed curves ∂D_j. The boundary conditions are the usual ones of no normal flow and conservation of circulation in each layer, namely

$$\frac{\partial \Phi_i}{\partial s} = 0 \quad \text{on} \quad \partial D, \quad \frac{\mathrm{d}}{\mathrm{d}t}\oint_{\partial D_j} \nabla \Phi_i \cdot \hat{\boldsymbol{n}}\,\mathrm{d}s = 0 \quad \text{for} \quad j = 0, \dots, J, \tag{2.5a, b}$$

where s is arclength along the boundary ∂D, and $\hat{\boldsymbol{n}}$ the outward unit normal.

Now suppose that $(\Phi_i, P_i) = (\Psi_i, Q_i)$ is a steady solution to the system (2.1)–(2.5); it follows that $\partial(\Psi_i, Q_i) = 0$ for each i, and consequently the isolines of Ψ_i and Q_i are coincident. We further assume that there exist continuously differentiable functions $\Psi_i(\cdot)$ such that

$$\Psi_i(x,y) = \Psi_i(Q_i(x,y)) \quad \forall (x,y) \in D. \tag{2.6}$$

A finite-amplitude disturbance (ψ_i, q_i) to this steady basic state is defined according to

$$\Phi_i = \Psi_i + \psi_i, \quad P_i = Q_i + q_i, \tag{2.7}$$

with

$$q_i(x,y,t) = \nabla^2 \psi_i + F_i \sum_{j=1}^{N} T_{ij} \psi_j. \tag{2.8}$$

3. Nonlinear stability theorems

We now assume (corresponding to the hypothesis of Arnol'd's second theorem) that the functional relations (2.6) are monotonic with a negative slope, and that there exist positive constants c_{1i} and c_{2i} such that

$$0 < c_{1i} \leqslant -\frac{\mathrm{d}\Psi_i}{\mathrm{d}Q_i} \leqslant c_{2i} < \infty \quad (i = 1, \dots, N), \tag{3.1}$$

where $\mathrm{d}\Psi_i/\mathrm{d}Q_i = \nabla\Psi_i/\nabla Q_i$. The goal now is to establish upper bounds for the disturbance energy

$$E(t) = \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i|\nabla\psi_i|^2 + \sum_{i=1}^{N-1} d_i F_i(\psi_{i+1}-\psi_i)^2\right\}\mathrm{d}x\,\mathrm{d}y \tag{3.2}$$

and disturbance potential enstrophy

$$Z(t) = \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i(q_i)^2\right\}\mathrm{d}x\,\mathrm{d}y \tag{3.3}$$

in terms of the initial disturbance fields.

To do so, first define the functions $G_i(\eta) = \int^{\eta} \Psi_i(\eta)\,\mathrm{d}\eta$, using the functional relations (2.6). In the usual way (cf. Arnol'd 1966), the definition of the function $\Psi_i(\cdot)$ may be extended if necessary outside the range of Q_i in the basic state, while maintaining the property (3.1): such extension is necessary for 'non-isovortical' disturbances such as are considered in this paper. Using conservation of total energy, total potential enstrophy, and total circulation in each layer, it follows that the functional

$$\begin{aligned}\iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i|\nabla(\Psi_i+\psi_i)|^2 + \sum_{i=1}^{N-1} d_i F_i(\Psi_{i+1}+\psi_{i+1}-\Psi_i-\psi_i)^2\right\}\mathrm{d}x\,\mathrm{d}y \\ +\iint_D \left\{\sum_{i=1}^{N} d_i G_i(Q_i+q_i)\right\}\mathrm{d}x\,\mathrm{d}y - \sum_{i=1}^{N}\sum_{j=0}^{J}\oint_{\partial D_j} d_i\,\Psi_i\,\nabla(\Psi_i+\psi_i)\cdot\hat{\boldsymbol{n}}\,\mathrm{d}s\end{aligned} \tag{3.4}$$

is conserved in time. Using this result, together with the manipulation

$$\begin{aligned}&\iint_D \left\{\sum_{i=1}^{N} d_i\,\nabla\Psi_i\cdot\nabla\psi_i + \sum_{i=1}^{N-1} d_i F_i(\Psi_{i+1}-\Psi_i)(\psi_{i+1}-\psi_i)\right\}\mathrm{d}x\,\mathrm{d}y \\ &= \iint_D \left\{\sum_{i=1}^{N}\left[d_i\,\nabla\cdot(\Psi_i\,\nabla\psi_i) - d_i\,\Psi_i\nabla^2\psi_i - d_i F_i\,\Psi_i\sum_{j=1}^{N} T_{ij}\,\psi_j\right]\right\}\mathrm{d}x\,\mathrm{d}y \\ &= \sum_{i=1}^{N}\sum_{j=0}^{J}\oint_{\partial D_j} d_i\,\Psi_i\,\nabla\psi_i\cdot\hat{\boldsymbol{n}}\,\mathrm{d}s - \iint_D \left\{\sum_{i=1}^{N} d_i\,G_i'(Q_i)\,q_i\right\}\mathrm{d}x\,\mathrm{d}y,\end{aligned} \tag{3.5}$$

it is easy to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}(E(t)+A(t)) = 0, \tag{3.6}$$

where

$$A(t) = \iint_D \left\{\sum_{i=1}^{N} d_i[G_i(Q_i+q_i) - G_i(Q_i) - G_i'(Q_i)\,q_i]\right\}\mathrm{d}x\,\mathrm{d}y. \tag{3.7}$$

Note that under the hypothesis (3.1), $A(t) \leqslant 0$. $E+A$ is the energy–Casimir functional referred to in the Introduction.

We wish to use the exact, nonlinear conservation law (3.6) to obtain upper bounds on $E(t)$ and $Z(t)$. It is helpful in this regard to decompose the disturbance (ψ_i, q_i) into two parts, following Mu & Shepherd (1993). To wit, let $\psi_{0i}(x,y)$ and $q_{0i}(x,y)$ be the initial disturbance stream function and potential vorticity fields. Define

$$q_i^* = \frac{\iint_D q_{0i}\,\mathrm{d}x\,\mathrm{d}y}{\iint_D \mathrm{d}x\,\mathrm{d}y}, \quad q_i' = q_i - q_i^* \quad (i = 1,\ldots,N). \tag{3.8}$$

The reason for making this decomposition is to take advantage of the Poincaré inequality (cf. (3.29)), which applies to q_i' but not to q_i. Since $\mathrm{d}/\mathrm{d}t(\iint_D q_i \,\mathrm{d}x\,\mathrm{d}y) = 0$, it follows that

$$\iint_D q_i' \,\mathrm{d}x\,\mathrm{d}y = 0 \quad \forall t \quad [i = 1, \ldots, N]. \tag{3.9}$$

Let ψ_i' be defined by

$$\nabla^2 \psi_i' + F_i \sum_{j=1}^{N} T_{ij} \psi_j' = q_i' \quad \text{in} \quad D; \quad \left.\frac{\partial \psi_i'}{\partial s}\right|_{\partial D} = 0; \tag{3.10a}$$

$$\oint_{\partial D_j} \nabla \psi_i' \cdot \hat{\boldsymbol{n}} \,\mathrm{d}s = 0 \quad \text{for} \quad j = 0, \ldots, J. \tag{3.10b}$$

We must establish existence and uniqueness of such a solution. To do so let the matrix $\boldsymbol{K}$ be defined by

$$K_{ij} = \operatorname{diag}(F_1^{\frac{1}{2}}, \ldots, F_n^{\frac{1}{2}}) = \delta_{ij} F_i^{\frac{1}{2}}, \tag{3.11}$$

and let $(\lambda_1, \ldots, \lambda_N)$ be the eigenvalues of the matrix $-\boldsymbol{KTK}$. According to Liu & Mu (1992), there exists an orthogonal matrix $\boldsymbol{L}$ such that

$$\boldsymbol{L}^T \boldsymbol{KTKL} = -\operatorname{diag}(\lambda_1, \ldots, \lambda_N), \tag{3.12}$$

where $\boldsymbol{L}^T$ is the transpose of $\boldsymbol{L}$, $\boldsymbol{L}^T\boldsymbol{L} = \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix, and the eigenvalues are non-negative and distinct:

$$0 = \lambda_1 < \ldots < \lambda_N. \tag{3.13}$$

Multiplying (3.10) on the left by $\boldsymbol{L}^T\boldsymbol{K}^{-1}$ gives the problem

$$\nabla^2 p_i - \lambda_i p_i = b_i \quad \text{in} \quad D; \quad \left.\frac{\partial p_i}{\partial s}\right|_{\partial D} = 0; \tag{3.14a}$$

$$\oint_{\partial D_j} \nabla p_i \cdot \hat{\boldsymbol{n}} \,\mathrm{d}s = 0 \quad \text{for} \quad j = 0, \ldots, J; \tag{3.14b}$$

where

$$\boldsymbol{p} = \boldsymbol{L}^T \boldsymbol{K}^{-1} \boldsymbol{\psi}', \quad \boldsymbol{b} = \boldsymbol{L}^T \boldsymbol{K}^{-1} \boldsymbol{q}', \tag{3.15}$$

after using the property (3.12). However, according to Mu (1992, theorem A 1), for specified b_i the problem (3.14) has a unique solution p_i for $i > 1$, since $\lambda_i > 0$. For $i = 1$, we have $\lambda_1 = 0$ and $\iint_D b_1 \,\mathrm{d}x\,\mathrm{d}y = 0$ by (3.9) and (3.15); therefore p_1 is defined uniquely up to an additive function of time.

This proves the existence of solutions to (3.10). As for uniqueness, suppose that $(\tilde{\varphi}_1, \ldots, \tilde{\varphi}_N)$ and $(\hat{\varphi}_1, \ldots, \hat{\varphi}_N)$ are two solutions to (3.10). Then $\varphi_i = \tilde{\varphi}_i - \hat{\varphi}_i$ satisfies the homogeneous form of (3.10). Multiplying the ith such equation by φ_i / F_i, integrating by parts, and using the boundary conditions yields

$$\iint_D \left\{ \sum_{i=1}^{N} \frac{1}{F_i} |\nabla \varphi_i|^2 + \sum_{i=1}^{N-1} (\varphi_{i+1} - \varphi_i)^2 \right\} \mathrm{d}x\,\mathrm{d}y = 0;$$

this implies

$$\tilde{\varphi}_i - \hat{\varphi}_i = \tilde{\varphi}_{i+1} - \hat{\varphi}_{i+1} = J(t) \quad (i = 1, \ldots, N-1),$$

where $J(t)$ is a function of time alone. Hence the non-uniqueness in the definition of ψ_i' will not affect the energies $E'(t)$ and E^* (see (3.21)) associated respectively with ψ_i' and ψ_i^*, as defined below.

It follows from (3.8) and (3.10) that ψ_i^*, defined by $\psi_i^* = \psi_i - \psi_i'$, satisfies

$$\nabla^2\psi_i^* + F_i \sum_{j=1}^{N} T_{ij}\psi_j^* = q_i^* \quad \text{in} \quad D; \quad \left.\frac{\partial \psi_i^*}{\partial s}\right|_{\partial D} = 0; \tag{3.16a}$$

$$\oint_{\partial D_j} \nabla\psi_i^* \cdot \hat{\boldsymbol{n}}\, \mathrm{d}s = \oint_{\partial D_j} \nabla\psi_{0i} \cdot \hat{\boldsymbol{n}}\, \mathrm{d}s \quad \text{for} \quad j = 0, \ldots, J. \tag{3.16b}$$

Now, using (3.8) and (3.9) we have

$$\iint_D \left\{ \sum_{i=1}^{N} d_i (q_i')^2 \right\} \mathrm{d}x\, \mathrm{d}y + \iint_D \left\{ \sum_{i=1}^{N} d_i (q_i^*)^2 \right\} \mathrm{d}x\, \mathrm{d}y = \iint_D \left\{ \sum_{i=1}^{N} d_i (q_i)^2 \right\} \mathrm{d}x\, \mathrm{d}y; \tag{3.17}$$

while from (3.7), (3.1), and Taylor's remainder theorem, we have (cf. Arnol'd 1966)

$$\iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{1i} d_i (q_i)^2 \right\} \mathrm{d}x\, \mathrm{d}y \leqslant -A(t).$$

Combining the above with (3.6) then yields

$$\iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{1i} d_i (q_i)^2 \right\} \mathrm{d}x\, \mathrm{d}y \leqslant -A(t) = E(t) - E(0) - A(0). \tag{3.18}$$

Applying the inequality

$$(a+b)^2 \leqslant (1+\alpha)\, a^2 + \left(1 + \frac{1}{\alpha}\right) b^2, \tag{3.19}$$

which holds for any positive constant α, to ($a = \psi_i^*$, $b = \psi_i'$) and ($a = \nabla\psi_i^*$, $b = \nabla\psi_i'$) gives

$$E(t) \leqslant (1+\alpha)\, E^* + \left(1 + \frac{1}{\alpha}\right) E'(t), \tag{3.20}$$

where

$$E^* = \iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} d_i |\nabla\psi_i^*|^2 + \sum_{i=1}^{N-1} d_i F_i (\psi_{i+1}^* - \psi_i^*)^2 \right\} \mathrm{d}x\, \mathrm{d}y, \tag{3.21a}$$

$$E'(t) = \iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} d_i |\nabla\psi_i'|^2 + \sum_{i=1}^{N-1} d_i F_i (\psi_{i+1}' - \psi_i')^2 \right\} \mathrm{d}x\, \mathrm{d}y. \tag{3.21b}$$

Combining this with (3.17) and (3.18) then yields

$$\iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{1i} d_i (q_i')^2 \right\} \mathrm{d}x\, \mathrm{d}y \leqslant (1+\alpha)\, E^* + \left(1 + \frac{1}{\alpha}\right) E'(t) - E(0) - A(0)$$
$$- \iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{1i} d_i (q_i^*)^2 \right\} \mathrm{d}x\, \mathrm{d}y. \tag{3.22}$$

The constant α will be determined later. From (3.1) and (3.7), it follows that

$$-A(0) \leqslant \iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{2i} d_i (q_{0i})^2 \right\} \mathrm{d}x\, \mathrm{d}y = \iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{2i} d_i [(q_{0i}')^2 + (q_i^*)^2] \right\} \mathrm{d}x\, \mathrm{d}y. \tag{3.23}$$

Combining (3.23) with (3.22) then yields

$$\iint_D \frac{1}{2} \left\{ \sum_{i=1}^{N} c_{1i} d_i (q_i')^2 \right\} \mathrm{d}x\, \mathrm{d}y \leqslant \alpha E^* + H + \left(1 + \frac{1}{\alpha}\right) E'(t) - E(0)$$
$$\leqslant \alpha E^* + H + \left(1 + \frac{1}{\alpha}\right) E'(t), \tag{3.24}$$

where

$$H = E^* + \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N}(c_{2i}-c_{1i})\,d_i(q_i^*)^2\right\}\mathrm{d}x\,\mathrm{d}y + \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} c_{2i}\,d_i(q'_{0i})^2\right\}\mathrm{d}x\,\mathrm{d}y. \quad (3.25)$$

Note that H is defined solely by the initial disturbance fields, is always positive for non-zero disturbances, and is independent of the free parameter α.

Now, everything on the right-hand side of (3.24) is determined by the initial conditions, with the exception of $E'(t)$. However, as in the one-layer case (Mu & Shepherd 1993) we can bound $E'(t)$ from above by applying a Poincaré inequality, as follows. First, note that in the light of (3.10), the expression (3.21*b*) for $E'(t)$ may be rewritten, after integrating by parts, as

$$E'(t) = -\iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i\,\psi'_i\,q'_i\right\}\mathrm{d}x\,\mathrm{d}y; \quad (3.26)$$

in terms of the variables p_i and b_i defined by (3.15), this takes the form

$$E'(t) = -\iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i F_i p_i b_i\right\}\mathrm{d}x\,\mathrm{d}y$$

$$= \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i F_i[|\nabla p_i|^2 + \lambda_i(p_i)^2]\right\}\mathrm{d}x\,\mathrm{d}y, \quad (3.27)$$

the second equality following after integration by parts, using (3.14). Let λ be the least positive eigenvalue of the problem

$$\nabla^2\phi + \lambda\phi = 0 \quad \text{in} \quad D; \quad \left.\frac{\partial\phi}{\partial s}\right|_{\partial D} = 0; \quad (3.28a)$$

$$\oint_{\partial D_j} \nabla\phi\cdot\hat{\boldsymbol{n}}\,\mathrm{d}s = 0 \quad \text{for} \quad j = 0, \ldots, J. \quad (3.28b)$$

Then from (3.14) and (3.27) together with standard theory of partial differential equations, we have the Poincaré inequality

$$E'(t) \leqslant \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N}\left(\frac{1}{\lambda+\lambda_i}\right) d_i F_i(b_i)^2\right\}\mathrm{d}x\,\mathrm{d}y. \quad (3.29)$$

The expression on the right-hand side of (3.29) must now be cast in terms of q'_i. Define the diagonal matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{C}$ by

$$\Lambda_{ij} = \mathrm{diag}\left(\frac{1}{\lambda+\lambda_1}, \ldots, \frac{1}{\lambda+\lambda_N}\right), \quad C_{ij} = \mathrm{diag}(c_{1i}, \ldots, c_{1N}). \quad (3.30)$$

Then in matrix form, (3.29) can be written as

$$E'(t) \leqslant \iint_D \tfrac{1}{2}(d_i F_i)\boldsymbol{b}^T\boldsymbol{\Lambda}\boldsymbol{b}\,\mathrm{d}x\,\mathrm{d}y$$

$$= \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q}')^T\boldsymbol{K}^{-1}\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T\boldsymbol{K}^{-1}\boldsymbol{q}'\,\mathrm{d}x\,\mathrm{d}y \quad (3.31)$$

(recalling that $d_i F_i$ is a constant, by (2.4)); while the left-hand side of (3.24) can be written as

$$\iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} c_{1i}\,d_i(q'_i)^2\right\}\mathrm{d}x\,\mathrm{d}y = \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q}')^T\boldsymbol{K}^{-1}\boldsymbol{C}\boldsymbol{K}^{-1}\boldsymbol{q}'\,\mathrm{d}x\,\mathrm{d}y. \quad (3.32)$$

Combining (3.31) and (3.32) with (3.24) then yields the inequality

$$\iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q}')^T \boldsymbol{K}^{-1}\left[\boldsymbol{C}-\left(1+\frac{1}{\alpha}\right)\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T\right]\boldsymbol{K}^{-1}\boldsymbol{q}'\,\mathrm{d}x\,\mathrm{d}y \leqslant \alpha E^* + H. \tag{3.33}$$

The expression (3.33) is very nearly what we need in order to bound the disturbance potential enstrophy $Z(t)$ in terms of the initial fields. Unfortunately, it depends on the matrix $\boldsymbol{L}$, which is not known explicitly. However, it turns out that $\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T$ may be expressed in terms of the basic physical parameters of the problem. In particular,

$$\begin{aligned}\boldsymbol{L}\boldsymbol{\Lambda}^{-1}\boldsymbol{L}^T &= \boldsymbol{L}\,\mathrm{diag}\,(\lambda+\lambda_1,\ldots,\lambda+\lambda_N)\,\boldsymbol{L}^T\\ &= \lambda\boldsymbol{I}+\boldsymbol{L}\,\mathrm{diag}\,(\lambda_1,\ldots,\lambda_N)\,\boldsymbol{L}^T\\ &= \lambda\boldsymbol{I}-\boldsymbol{KTK}\end{aligned}$$

after using (3.12), whence

$$\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T = (\boldsymbol{L}\boldsymbol{\Lambda}^{-1}\boldsymbol{L}^T)^{-1} = (\lambda\boldsymbol{I}-\boldsymbol{KTK})^{-1}. \tag{3.34}$$

Now define the matrix

$$\boldsymbol{M} = \boldsymbol{C}-\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T = \boldsymbol{C}-(\lambda\boldsymbol{I}-\boldsymbol{KTK})^{-1}. \tag{3.35}$$

It is clear from (3.33) that if $\boldsymbol{M}$ is a positive definite matrix, then the left-hand side of (3.33) should be positive for sufficiently large α and stability would then follow. This is indeed the case, as is shown explicitly below. Hence we hypothesize that $\boldsymbol{M}$ is a positive definite matrix with minimum eigenvalue k_1, namely

$$\boldsymbol{\phi}^T\boldsymbol{M}\boldsymbol{\phi} \geqslant k_1|\boldsymbol{\phi}|^2 \quad \forall\boldsymbol{\phi}, \quad \text{with } k_1 > 0. \tag{3.36}$$

Since from (3.13) and (3.30) the largest eigenvalue of the matrix $\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T$ is $1/\lambda$, we have

$$\left(k_1-\frac{1}{\alpha\lambda}\right)\iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q}')^T(\boldsymbol{K}^{-1})^2\boldsymbol{q}'\,\mathrm{d}x\,\mathrm{d}y$$
$$\leqslant \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q}')^T\boldsymbol{K}^{-1}\left[\boldsymbol{C}-\left(1+\frac{1}{a}\right)\boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}^T\right]\boldsymbol{K}^{-1}\boldsymbol{q}'\,\mathrm{d}x\,\mathrm{d}y. \tag{3.37}$$

The inequality (3.37) can be combined with (3.33) to yield

$$B(\alpha)\iint_D \frac{1}{2}\left\{\sum_{i=1}^N d_i(q_i')^2\right\}\mathrm{d}x\,\mathrm{d}y \leqslant \alpha E^* + H, \tag{3.38}$$

where $B(\alpha) = k_1-(1/\alpha\lambda)$. If we choose $\alpha > 1/\lambda k_1$, so that $B(\alpha) > 0$, then (3.38) together with (3.17) gives the bound

$$Z(t) \leqslant \frac{\alpha E^*+H}{B(\alpha)} + \iint_D \frac{1}{2}\left\{\sum_{i=1}^N d_i(q_i^*)^2\right\}\mathrm{d}x\,\mathrm{d}y. \tag{3.39}$$

The task is now to choose α so as to minimize the right-hand side of (3.39), in order to obtain the sharpest possible bound on the disturbance potential enstrophy. First consider the case $E^* = 0$, for which $|\nabla\psi_i^*| = 0$ and $\psi_{i+1}^*-\psi_i^* = 0$, so that $q_i^* = 0$. In this case the right-hand side of (3.39) is minimized for the maximum value of $B(\alpha)$, which is achieved in the limit $\alpha\to\infty$ and is just k_1. This yields the inequality

$$Z(t) \leqslant H/k_1. \tag{3.40}$$

In the general case $E^* \neq 0$, the minimum of the right-hand side of (3.39) is attained at

$$\alpha_{min} = \frac{1+[1+\lambda k_1(H/E^*)]^{\frac{1}{2}}}{\lambda k_1}, \tag{3.41}$$

for which $B(\alpha_{min}) > 0$ and

$$\frac{\alpha_{min} E^* + H}{B(\alpha_{min})} = \frac{2\{E^* + [E^*(E^* + \lambda k_1 H)]^{\frac{1}{2}}\} + \lambda k_1 H}{\lambda k_1^2}. \tag{3.42}$$

Together with (3.39), this yields

$$Z(t) \leqslant \frac{2\{E^* + [E^*(E^* + \lambda k_1 H)]^{\frac{1}{2}}\} + \lambda k_1 H}{\lambda k_1^2} + \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i (q_i^*)^2\right\} \mathrm{d}x\,\mathrm{d}y. \tag{3.43}$$

In this way we have obtained an upper bound on the disturbance potential enstrophy $Z(t)$ in terms of the initial disturbance, since E^*, H and q^* depend only on the initial disturbance. It is worth noting that the special case (3.40) can be obtained from (3.43) by setting $E^* = 0$ and $q_i^* = 0$ in the latter expression.

We now proceed to determine an upper bound on the disturbance energy. From (3.31) and the fact that the largest eigenvalue of the matrix $\boldsymbol{L\Lambda L}^T$ is $1/\lambda$, we have

$$E'(t) \leqslant \frac{Z'(t)}{\lambda} \equiv \frac{1}{\lambda} \iint_D \frac{1}{2}\left\{\sum_{i=1}^{N} d_i (q_i')^2\right\} \mathrm{d}x\,\mathrm{d}y. \tag{3.44}$$

First consider the case $E^* = 0$, for which $E(t) = E'(t)$ and $Z(t) = Z'(t)$. In this case (3.40) and (3.17) may be invoked with (3.44) to yield

$$E(t) = E'(t) \leqslant H/\lambda k_1. \tag{3.45}$$

In the general case $E^* \neq 0$, on the other hand, substituting (3.44) into (3.20) yields

$$E(t) \leqslant (1+\alpha) E^* + \left(1 + \frac{1}{\alpha}\right) \frac{Z'}{\lambda}. \tag{3.46}$$

The minimum of the right-hand side of (3.46) is attained at

$$\alpha_{min} = (Z'/\lambda E^*)^{\frac{1}{2}}, \tag{3.47}$$

which when substituted back into (3.46) gives

$$E(t) \leqslant [E^{*\frac{1}{2}} + (Z'/\lambda)^{\frac{1}{2}}]^2. \tag{3.48}$$

Then using the bound on Z' that comes from (3.43) and (3.17), this yields

$$E(t) \leqslant \left(\frac{1+\lambda k_1}{\lambda k_1} E^{*\frac{1}{2}} + \frac{(E^* + \lambda k_1 H)^{\frac{1}{2}}}{\lambda k_1}\right)^2. \tag{3.49}$$

In this way we have obtained an upper bound on the disturbance energy $E(t)$ in terms of the initial disturbance. Note that the special case (3.45) can be obtained from (3.49) by setting $E^* = 0$ in the latter expression.

It is easy to see that when the initial disturbance potential enstrophy and initial disturbance circulations tend to zero, then E^*, q^* and H tend to zero also. This fact, together with the rigorous upper bounds (3.43) and (3.49), therefore demonstrates that the disturbance potential enstrophy and disturbance energy can be bounded for all time below any given positive constants, for sufficiently small initial potential enstrophy and initial circulations. We take this as the definition of nonlinear stability. Thus we can state:

CRITERION 3.1. *Suppose the basic state* (Ψ_i, Q_i) *satisfies* (2.6) *and* (3.1), *and the matrix* $\boldsymbol{M}$ *defined by* (3.35) *is positive definite. Then* (Ψ_i, Q_i) *is nonlinearly stable. In particular, upper bounds on the finite-amplitude disturbance potential enstrophy and disturbance energy are provided by* (3.43) *and* (3.49), *respectively.*

When $N \geqslant 3$, it is generally a difficult task to obtain the smallest eigenvalue k_1 of the matrix $\boldsymbol{M}$, and one may have to resort to numerical methods. On the other hand, an explicit expression for k_1 in the case $N = 2$ is obtainable, and will be given in the next section. But it is an easy matter to obtain a simpler, albeit weaker, stability criterion in the N-layer case, as follows. For any vector $\boldsymbol{\phi}$, we have the inequality

$$\boldsymbol{\phi}^T \boldsymbol{M} \boldsymbol{\phi} = \boldsymbol{\phi}^T \boldsymbol{C} \boldsymbol{\phi} - \boldsymbol{\phi}^T \boldsymbol{L} \boldsymbol{\Lambda} \boldsymbol{L}^T \boldsymbol{\phi} = \boldsymbol{\phi}^T \boldsymbol{C} \boldsymbol{\phi} - (\boldsymbol{L}^T \boldsymbol{\phi})^T \boldsymbol{\Lambda} \boldsymbol{L}^T \boldsymbol{\phi} \geqslant \left(\min_i c_{1i} - \frac{1}{\lambda}\right) |\boldsymbol{\phi}|^2. \quad (3.50)$$

Therefore the condition

$$\lambda \min_i c_{1i} > 1 \quad (3.51)$$

implies that $\boldsymbol{M}$ is a positive definite matrix. This yields:

CRITERION 3.2. *Suppose the basic state* (Ψ_i, Q_i) *satisfies* (2.6) *and* (3.1), *and* (3.51) *holds. Then* (Ψ_i, Q_i) *is nonlinearly stable in the sense described in Criterion* 3.1.

Obviously it is easier to verify Criterion 3.2 than Criterion 3.1, although the latter is better insofar as it applies to a wider class of basic states. Also, whenever (3.51) holds, it follows from (3.50) that

$$k_1 \geqslant \tilde{k}_1 \equiv \min_i c_{1i} - \frac{1}{\lambda} > 0. \quad (3.52)$$

Thus to calculate explicit upper bounds on $Z(t)$ and $E(t)$ when the basic flow is stable according to Criterion 3.2, one may use (3.43) and (3.49) with k_1 replaced by $\tilde{k}_1$.

It is well known that when the problem is zonally symmetric, conservation of zonal momentum can play an important role in the study of nonlinear stability (McIntyre & Shepherd 1987; Zeng 1989; Mu 1991; Ripa 1992). We therefore incorporate this additional information under the assumption that the topography h is independent of x, i.e. $h = h(y)$, and consider the boundaries ∂D_j to consist of lines of constant y. (A zonal channel would be the most common such geometry.) It is straightforward to show from (2.1)–(2.5) that in that case the zonal momentum (or impulse)

$$M(t) = \iint_D \left\{ \sum_{i=1}^N d_i y P_i \right\} \mathrm{d}x\, \mathrm{d}y \quad (3.53)$$

is conserved in time. We now consider a basic state (Ψ_i, Q_i) that is zonally symmetric and for which a constant α and functions Ψ_i^α exist such that

$$\Psi_i + \alpha y = \Psi_i^\alpha(Q_i) \quad (i = 1, \ldots, N). \quad (3.54)$$

(The reason for the restriction to zonally symmetric flows is that Andrews' theorem applies in this case: see § 5.) Suppose further that for such α, constants c_{1i} and c_{2i} exist such that

$$0 < c_{1i} \leqslant -\frac{\mathrm{d}\Psi_i^\alpha}{\mathrm{d}Q_i} \leqslant c_{2i} < \infty \quad (i = 1, \ldots, N). \quad (3.55)$$

Now, using the invariance of (3.53) it is easy to show that (3.6) still holds with $A(t)$ given by (3.7), provided G_i is defined by $G_i(\eta) = \int^{\eta} \Psi_i^{\alpha}(\eta)\,\mathrm{d}\eta$. The derivation leading to Criterion 3.1 above then can be followed step by step, and the estimates (3.43) and (3.49) follow directly. Hence we can state:

CRITERION 3.3. *Suppose the basic state* (Ψ_i, Q_i) *satisfies* (3.54) *and* (3.55), *and the matrix* $\boldsymbol{M}$ *defined by* (3.35) *is positive definite. Then* (Ψ_i, Q_i) *is nonlinearly stable. In particular, upper bounds on the finite-amplitude disturbance potential enstrophy and disturbance energy are provided by* (3.43) *and* (3.49), *respectively. If* (3.51) *holds, then* $\boldsymbol{M}$ *is guaranteed to be positive definite and the basic state is nonlinearly stable.*

We now compare the above criteria with Criterion 4.2 obtained by Mu (1991), which states that the basic state (Ψ_i, Q_i) is nonlinearly stable if (3.1) holds and

$$\tilde{\lambda} \min_i c_{1i} > 1, \tag{3.56}$$

where $\tilde{\lambda}$ is the smallest positive eigenvalue of a boundary value problem for an elliptic system (see the Appendix). It is shown in the Appendix that $\lambda \geqslant \tilde{\lambda}$. Therefore Criterion 3.2, and its sharper form Criterion 3.1, are stronger than Mu's (1991) Criterion 4.2. Perhaps more importantly, in order to verify applicability of the present criteria, one needs only to find the smallest positive eigenvalue λ of the two-dimensional problem (3.28), whereas $\tilde{\lambda}$ is more difficult to determine. It may also be added that while Mu (1991) only obtained implicit bounds on the finite-amplitude disturbance potential enstrophy and disturbance energy, in the present work we have derived explicit bounds on those quantities.

Ripa (1992) examined the nonlinear stability properties of the model (2.1)–(2.5), and came up with a criterion analogous to our Criterion 3.2, though with λ replaced by a different minimum eigenvalue, call it $\hat{\lambda}$. It is shown in Mu & Shepherd (1993) that $\hat{\lambda} < \lambda$ for simply connected domains, while $\hat{\lambda} = \lambda$ for a periodic channel. Moreover, Criterion 3.1 is superior to Criterion 3.2: an explicit example demonstrating this is provided by the case of the Phillips model of baroclinic instability (§4.2 below).

4. The two-layer model

4.1. *General results*

We consider the case of two layers, $N = 2$. Then

$$\boldsymbol{K} = \begin{pmatrix} F_1^{\frac{1}{2}} & 0 \\ 0 & F_2^{\frac{1}{2}} \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{4.1}$$

and so

$$-\boldsymbol{KTK} = \begin{pmatrix} F_1 & -(F_1 F_2)^{\frac{1}{2}} \\ -(F_1 F_2)^{\frac{1}{2}} & F_2 \end{pmatrix} \tag{4.2}$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2 = F_1 + F_2$. The matrix $\boldsymbol{M}$ defined by (3.35) is seen to be

$$\boldsymbol{M} = \begin{pmatrix} c_{11} - \dfrac{\lambda + F_2}{\lambda(\lambda + F_1 + F_2)} & -\dfrac{(F_1 F_2)^{\frac{1}{2}}}{\lambda(\lambda + F_1 + F_2)} \\ -\dfrac{(F_1 F_2)^{\frac{1}{2}}}{\lambda(\lambda + F_1 + F_2)} & c_{12} - \dfrac{\lambda + F_1}{\lambda(\lambda + F_1 + F_2)} \end{pmatrix}, \tag{4.3}$$

which is positive definite if and only if

$$\lambda c_{11}+\frac{F_1}{\lambda+F_1+F_2}>1, \tag{4.4a}$$

$$\lambda c_{12}+\frac{F_2}{\lambda+F_1+F_2}>1, \tag{4.4b}$$

$$\left(\lambda c_{11}+\frac{F_1}{\lambda+F_1+F_2}-1\right)\left(\lambda c_{12}+\frac{F_2}{\lambda+F_1+F_2}-1\right)>\frac{F_1F_2}{(\lambda+F_1+F_2)^2}. \tag{4.4c}$$

In particular, we can see that if

$$\lambda c_{11}>1 \quad \text{and} \quad \lambda c_{12}>1, \tag{4.5}$$

then (4.4*a–c*) are all satisfied and $\boldsymbol{M}$ is positive definite. The condition (4.5) is the two-layer version of (3.51), so we have recovered Criterion 3.2 directly. The best bounds on $E(t)$ and $Z(t)$ will however come from the use of k_1, which is given explicitly by

$$k_1=\tfrac{1}{2}(M_{11}+M_{22}-[(M_{11}-M_{22})^2+4(M_{12})^2]^{\frac{1}{2}}), \tag{4.6}$$

where M_{ij} are the entries of $\boldsymbol{M}$ as provided by (4.3).

Note that since (4.4*a, b*) imply $c_{11}>0$ and $c_{12}>0$, in proving the stability of any particular basic state it is sufficient to establish that (4.4*a–c*) hold.

4.2. *The Phillips model*

The Phillips model of baroclinic instability (e.g. Pedlosky 1979, §7.11) has the basic state consisting of a constant zonal flow in each layer,

$$\Psi_i(y)=-U_i y, \quad Q_i(y)=(-1)^{i+1}F_i U_s y+f_0+\beta y, \tag{4.7}$$

with $U_s=U_1-U_2$ and where U_1, U_2 are constants. The domain is the periodic channel

$$D=\{-\pi\leqslant x\leqslant\pi,\ -L\leqslant y\leqslant L\}, \tag{4.8}$$

for which $\lambda=(\pi/2L)^2$. Since the problem is zonally symmetric we employ Criterion 3.3. Evidently

$$\Psi_1+\alpha y=(\alpha-U_1)y, \quad \Psi_2+\alpha y=(\alpha-U_2)y, \tag{4.9}$$

and

$$Q_1=(\beta+F_1U_s)y+f_0, \quad Q_2=(\beta-F_2U_s)y+f_0. \tag{4.10}$$

There are four cases to consider, depending on the nature of the potential vorticity gradients.

Case 1. $\beta+F_1U_s=0$. We choose $\alpha=U_1$, in which case the basic state satisfies (3.55) with

$$c_{11}=c, \quad c_{12}=\frac{U_s}{F_2U_s-\beta}=\frac{1}{F_1+F_2}, \tag{4.11}$$

where c is an arbitrary positive constant. Condition (4.4*b*) is satisfied if

$$\frac{\lambda}{F_1+F_2}-\frac{\lambda+F_1}{\lambda+F_1+F_2}>0\Leftrightarrow\lambda^2>F_1(F_1+F_2), \tag{4.12}$$

while if (4.12) holds then (4.4*a*) and (4.4*c*) can be assured to hold by choosing c sufficiently large. Thus (4.12) is a sufficient condition for nonlinear stability, by Criterion 3.3.

Case 2. $\beta - F_2 U_s = 0$. We choose $\alpha = U_2$, in which case the basic state satisfies (3.55) with

$$c_{11} = \frac{U_s}{\beta + F_1 U_s} = \frac{1}{F_1 + F_2}, \quad c_{12} = c, \tag{4.13}$$

where c is an arbitrary positive constant. Following the argument in Case 1 above, we see that the basic state in this case is nonlinearly stable by Criterion 3.3 if

$$\lambda^2 > F_2(F_1 + F_2). \tag{4.14}$$

Case 3. $(\beta + F_1 U_s)(\beta - F_2 U_s) > 0$, i.e.

$$-\frac{\beta}{F_1} < U_s < \frac{\beta}{F_2}. \tag{4.15}$$

By (4.9) and (4.10), the basic state satisfies (3.55) for any $\alpha < \min(U_1, U_2)$. By taking $\alpha \to -\infty$ we can make c_{11} and c_{12} arbitrarily large, in which case (4.4*a–c*) hold and the flow is nonlinearly stable by Criterion 3.3. Of course, since $(\beta + F_1 U_s)(\beta - F_2 U_s) > 0$ then the basic-state potential vorticity gradients $\mathrm{d}Q_1/\mathrm{d}y$ and $\mathrm{d}Q_2/\mathrm{d}y$ are of the same sign, so the basic state is nonlinearly stable in any case by the finite-amplitude version of the Charney–Stern theorem (Shepherd 1988). This fact explains why no restriction on λ akin to (4.12) or (4.14) arises in this case.

Case 4. $(\beta + F_1 U_s)(\beta - F_2 U_s) < 0$, i.e.

$$U_s < -\beta/F_1 \quad \text{or} \quad U_s > \beta/F_2. \tag{4.16a, b}$$

Then the basic state satisfies (3.55) with

$$c_{11} = \frac{U_1 - \alpha}{\beta + F_1 U_s}, \quad c_{12} = \frac{U_2 - \alpha}{\beta - F_2 U_s}, \tag{4.17}$$

where we take α such that $U_1 < \alpha < U_2$ in the case (4.16*a*), and such that $U_2 < \alpha < U_1$ in the case (4.16*b*). We must determine under what conditions α may be chosen such that (4.4*a–c*) hold. (Recall that if (4.4*a,b*) hold, then c_{11} and c_{12} are necessarily positive.) By isolating α between (4.4*a*) and (4.4*b*), one finds that a solution for α satisfying (4.4*a, b*) exists if and only if

$$\lambda^2 - 2F_1 F_2 + (F_1 - F_2)\beta/U_s > 0. \tag{4.18}$$

In the case of (4.4*c*), expressing the left-hand side in terms of powers of α yields a quadratic in α, which satisfies the inequality (4.4*c*) if and only if a certain discriminant is positive. This condition boils down to

$$U_s^2 \lambda^2(\lambda^2 - 4F_1 F_2) + 2U_s \beta(F_1 - F_2)\lambda^2 + (F_1 + F_2)^2 \beta^2 > 0. \tag{4.19}$$

Therefore the existence of an α such that the matrix $\boldsymbol{M}$ is positive definite is equivalent to (4.18) and (4.19). Stated otherwise, (4.18) and (4.19) are sufficient conditions for nonlinear stability, by Criterion 3.3. Note that applying the obvious generalization of Criterion 3.2 (i.e. including α) to Case 4 gives the nonlinear stability criterion $\lambda > F_1 + F_2$, which is evidently weaker than (4.18) and (4.19): in particular, it has no dependence on U_s.

To summarize the above, the basic state (4.7) of the Phillips model of baroclinic instability has been shown to be nonlinearly stable by Criterion 3.3, provided one of the following conditions is satisfied:

(i) $U_s = -\beta/F_1$ and $\lambda^2 > F_1(F_1+F_2)$;

(ii) $U_s = \beta/F_2$ and $\lambda^2 > F_2(F_1+F_2)$;

(iii) $-\beta/F_1 < U_s < \beta/F_2$;

(iv) $U_s < -\beta/F_1$ or $U_s > \beta/F_2$, and (4.18), (4.19) hold.

Condition (iii) is not new insofar as it is obtainable using the finite-amplitude Charney–Stern theorem, but the other conditions have not been derived before in the context of nonlinear stability.

Since these conditions are sufficient for stability, their violation is necessary for instability. From Pedlosky (1979, equation (7.11.6)) it is evident that normal-mode instability occurs in the Phillips model whenever the total wavenumber κ satisfies

$$\beta^2(F_1+F_2)^2 + 2\beta U_s \kappa^4(F_1-F_2) - \kappa^4 U_s^2(4F_1 F_2 - \kappa^4) < 0. \tag{4.20}$$

Clearly, normal modes with sufficiently large κ^2 cannot satisfy (4.20): this is the well-known short-wave cut-off in the Phillips model of baroclinic instability. Note that (4.20) is exactly the opposite of (4.19) if we set $\kappa^2 = \lambda$. Since $\kappa^2 \geqslant \lambda$ necessarily, we see that satisfaction of (4.20) requires violation of (4.19), as we would expect. By the same token, the short-wave portion of the marginal curve described by (4.20) is captured precisely by the nonlinear stability condition (4.19).

In the special case $\beta = 0$, the stability conditions (i)–(iv) collapse to the single sufficient condition

$$\lambda^2 > 4F_1 F_2 \tag{4.21}$$

for non-trivial basic states $U_s \neq 0$. This nonlinear stability criterion was also obtained by Ripa (1992, equation (5.5)).

In the special case $F_1 = F_2 = F$, conditions (i)–(iv) can be consolidated into the following, any one of which is sufficient for nonlinear stability:

(*a*) $U_s^2 = \beta^2/F^2$ and $\lambda^2 > 2F^2$;

(*b*) $U_s^2 < \beta^2/F^2$;

(*c*) $U_s^2 > \beta^2/F^2$ and either

$$(c1)\,\lambda^2 > 4F^2 \quad \text{or} \quad (c2)\ 2F^2 < \lambda^2 < 4F^2 \quad \text{and} \quad U_s^2 < \frac{4F^2\beta^2}{\lambda^2(4F^2-\lambda^2)}.$$

The relative weakness of Criterion 3.2 relative to Criterion 3.1 (making the obvious generalizations to include α) is demonstrated by the fact that the former requires $\lambda^2 > 4F^2$ in cases (*a*) and (*c*) above.

4.3. *An interesting example*

The mathematical method behind Arnol'd's nonlinear stability theorems is usually presented in terms of constructing an invariant functional, known as the energy–Casimir (or more generally the energy–momentum–Casimir) functional, which is sign-definite for admissible disturbances to the specified basic state. In our case $E+A$ is the energy–Casimir functional. However, we have nowhere had to appeal to $E+A$ being sign-definite in order to prove nonlinear stability; indeed it need not be so, as the following example demonstrates.

Consider the periodic channel (4.8), for which $\lambda = (\pi/2L)^2$. Let the basic state be given by

$$\Psi_i = A_i \cos y + B_i y \quad (i = 1, 2), \tag{4.22}$$

where $A_1 > 0$, $A_2 < 0$, $A_1 + A_2 \neq 0$, and where B_1 and B_2 are chosen to satisfy

$$\frac{F_1(B_2-B_1)+\beta}{F_1(A_2-A_1)-A_1} = \frac{B_1}{A_1}, \quad \frac{F_2(B_1-B_2)+\beta}{F_2(A_1-A_2)-A_2} = \frac{B_2}{A_2}. \tag{4.23}$$

Using the presumed conditions on A_i, it is easy to show that such B_i always exist. Now, using (4.23) it can be seen that

$$\Psi_1 = \frac{A_1}{F_1(A_2-A_1)-A_1}(Q_1-f_0), \quad \Psi_2 = \frac{A_2}{F_2(A_1-A_2)-A_2}(Q_2-f_0). \tag{4.24}$$

Therefore (3.1) is satisfied with

$$c_{11} = \frac{A_1}{A_1+F_1(A_1-A_2)} > 0, \quad c_{12} = \frac{A_2}{A_2+F_2(A_2-A_1)} > 0, \tag{4.25}$$

and the basic state (4.22) is nonlinearly stable by Criterion 3.2 if

$$L < \frac{\pi}{2}\min\left(\left(\frac{A_1}{A_1+F_1(A_1-A_2)}\right)^{\frac{1}{2}}, \left(\frac{A_2}{A_2+F_2(A_2-A_1)}\right)^{\frac{1}{2}}\right). \tag{4.26}$$

For this basic state, the energy–Casimir functional $E+A$ defined by (3.2) and (3.7) takes the form

$$E+A = \iint_D \frac{1}{2}\left\{\sum_{i=1}^{2} d_i[|\nabla\psi_i|^2 - c_{1i}(q_i)^2] + (d_i F_i)(\psi_1-\psi_2)^2\right\} \mathrm{d}x\,\mathrm{d}y. \tag{4.27}$$

If the disturbance (ψ_i, q_i) is given by $\psi_1 = \psi_2 = \varepsilon \sin y$, $q_1 = q_2 = -\varepsilon \sin y$, with ε being an arbitrary positive constant, then

$$E+A = \frac{\varepsilon^2}{2}\iint_D \{(d_1+d_2)\cos^2 y - (d_1 c_{11} + d_2 c_{12})\sin^2 y\}\,\mathrm{d}x\,\mathrm{d}y. \tag{4.28}$$

On the other hand, if the disturbance is given by $\psi_1 = \psi_2 = \varepsilon \cos y$, $q_1 = q_2 = -\varepsilon \cos y$, then

$$E+A = \frac{\varepsilon^2}{2}\iint_D \{(d_1+d_2)\sin^2 y - (d_1 c_{11} + d_2 c_{12})\cos^2 y\}\,\mathrm{d}x\,\mathrm{d}y. \tag{4.29}$$

For given A_1, A_2, it is clear that we may choose L sufficiently small so that (4.28) is positive while (4.29) is negative; yet (4.26) remains valid. Therefore we have constructed an explicit example of a basic state that is nonlinearly stable by Criterion 3.2, but for which the associated energy–Casimir functional $E+A$ is not sign-definite.

5. Andrews' theorem

It has been proved by Andrews (1984) that any basic state that is nonlinearly stable by Arnol'd's first theorem in a domain with zonally symmetric boundaries must itself be zonally symmetric. We show here that this result extends to our new criteria, where stable states need not be extrema of the energy–Casimir invariant (see §4.3).

The claim is that if the basic state (Ψ_i, Q_i) (in a domain D with zonally symmetric boundaries) satisfies (3.55) and the matrix $\boldsymbol{M}$ is positive definite, then it follows that $\partial\Psi_i/\partial x = 0$ for all i and therefore the basic state, which is stable by Criterion 3.3, must be zonally symmetric.

The proof is by contradiction. Suppose that for some i, $\partial\Psi_i/\partial x \neq 0$ somewhere in D. Then there exists a constant a such that

$$\psi_i(x,y) \equiv \Psi_i(x+a,y)-\Psi_i(x,y) \tag{5.1}$$

satisfies $\partial\psi_i/\partial x \neq 0$ somewhere in D. Obviously ψ_i defined by (5.1) satisfies

$$\left.\frac{\partial\psi_i}{\partial s}\right|_{\partial D} = 0 \quad \text{and} \quad \oint_{\partial D_j} \nabla\psi_i\cdot\hat{\boldsymbol{n}}\,\mathrm{d}s = 0 \quad \text{for} \quad j=0,\dots,J \quad (i=1,\dots,N), \tag{5.2}$$

since the boundary ∂D is presumed to be zonally symmetric. It is also clear from (5.1) that $q_i^* = 0$ and so, together with (5.2), it follows that $\psi_i^* = 0$ and hence

$$\psi_i = \psi_i', \quad q_i = q_i', \quad E = E'. \tag{5.3}$$

Since the total energy, the total momentum, and the Casimir functionals

$$\iint_D d_i\, G_i(P_i)\,\mathrm{d}x\,\mathrm{d}y$$

are all unaltered by the perturbation (5.1), we have

$$E+A = 0, \tag{5.4}$$

where A is given by (3.7) with $G_i(\eta) = \int^\eta \Psi_i^\alpha(\eta)\,\mathrm{d}\eta$. From (3.31),

$$E \leqslant \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T \boldsymbol{K}^{-1}\boldsymbol{L\Lambda L}^T\boldsymbol{K}^{-1}\boldsymbol{q}\,\mathrm{d}x\,\mathrm{d}y. \tag{5.5}$$

Putting (5.5) together with (3.18) and (5.4) gives

$$\iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T \boldsymbol{K}^{-1}\boldsymbol{C}\boldsymbol{K}^{-1}\boldsymbol{q}\,\mathrm{d}x\,\mathrm{d}y \leqslant -A = E \leqslant \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T \boldsymbol{K}^{-1}\boldsymbol{L\Lambda L}^T\boldsymbol{K}^{-1}\boldsymbol{q}\,\mathrm{d}x\,\mathrm{d}y$$

$$\Leftrightarrow \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T \boldsymbol{K}^{-1}[\boldsymbol{C}-\boldsymbol{L\Lambda L}^T]\boldsymbol{K}^{-1}\boldsymbol{q}\,\mathrm{d}x\,\mathrm{d}y \leqslant 0. \tag{5.6}$$

But since $\boldsymbol{M} = \boldsymbol{C}-\boldsymbol{L\Lambda L}^T$ has been assumed to be positive definite, (5.6) implies that

$$\iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T (\boldsymbol{K}^{-1})^2\boldsymbol{q}\,\mathrm{d}x\,dy = 0. \tag{5.7}$$

Using the fact that the largest eigenvalue of the matrix $\boldsymbol{L\Lambda L}^T$ is $1/\lambda$ (cf. (3.44)), we thus obtain the chain of inequalities

$$E \leqslant \iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T \boldsymbol{K}^{-1}\boldsymbol{L\Lambda L}^T\boldsymbol{K}^{-1}\boldsymbol{q}\,\mathrm{d}x\,\mathrm{d}y \leqslant \frac{1}{\lambda}\iint_D \tfrac{1}{2}(d_i F_i)(\boldsymbol{q})^T (\boldsymbol{K}^{-1})^2\boldsymbol{q}\,\mathrm{d}x\,\mathrm{d}y = 0,$$

which implies $E = 0$. But this contradicts the fact that $\partial\psi_i/\partial x \neq 0$ somewhere in D. Hence $\partial\Psi_i/\partial x = 0$ for all i, and the basic state must be zonally symmetric.

6. Nonlinear stability of uniform potential vorticity flows

Consider the case of uniform potential vorticity flow, $\nabla Q_i = 0$. It is easy to see that Arnol'd's stability theorems are not applicable to such a situation, except in the special case of uniform velocity: the functional relation (3.54) can only exist, when $\nabla Q_i = 0$, if $V_i = \partial\Psi_i/\partial x = 0$ and $U_i = -\partial\Psi_i/\partial y = \alpha$. It is, however, well known that uniform potential vorticity flows are always linearly stable to normal-mode disturbances. We show below that such flows are in fact nonlinearly stable.

When $\nabla Q_i = 0$, the potential vorticity equation (2.1) reduces to

$$\frac{\partial q_i}{\partial t} + \partial(\Psi_i + \psi_i, q_i) = 0 \quad (i = 1, \ldots, N), \tag{6.1}$$

from which it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t} \iint_D (q_i)^2 \,\mathrm{d}x \,\mathrm{d}y = 0 \quad \forall i. \tag{6.2}$$

But (6.2) demonstrates immediately that the disturbance potential enstrophy $Z(t)$ is a conserved quantity:

$$Z(t) = Z(0). \tag{6.3}$$

To establish an upper bound on the disturbance energy, as in §3, first note that in the light of (3.17), (6.3) implies $Z'(t) = Z'(0) \equiv Z'_0$. Using this fact together with (3.20) and (3.44) implies

$$E(t) \leqslant (1+\alpha) E^* + \left(1 + \frac{1}{\alpha}\right) \frac{Z'_0}{\lambda} \tag{6.4}$$

for any positive constant α. If $E^* = 0$, then the minimum of the right-hand side of (6.4) is attained in the limit $\alpha \to \infty$, which yields (noting that $q^* = 0$ in this case) the bound

$$E(t) \leqslant \frac{1}{\lambda} Z(0). \tag{6.5}$$

If $E^* \neq 0$, then the minimum of the right-hand side of (6.4) is attained at

$$\alpha_{min} = (Z'_0/\lambda E^*)^{\frac{1}{2}}. \tag{6.6}$$

Substituting (6.6) into (6.4) then yields the bound

$$E(t) \leqslant [E^{*\frac{1}{2}} + (Z'_0/\lambda)^{\frac{1}{2}}]^2. \tag{6.7}$$

The bounds on the disturbance potential enstrophy and disturbance energy provided by (6.3), (6.5) and (6.7) demonstrate that uniform potential vorticity flows are always nonlinearly stable, in any domain D.

It is worth noting that although (6.3) and (6.7) prove nonlinear stability in the sense we have defined it, (6.7) does not prove Liapunov stability in the energy norm because the right-hand side of (6.7) cannot be bounded from above in terms of $E(0)$: that is to say, the disturbance energy cannot be bounded *a priori* in terms of the initial disturbance energy. In contrast, (6.3) proves Liapunov stability in the potential enstrophy norm. Thus, while the potential enstrophy remains constant in time, the energy can in principle amplify by an arbitrarily large amount. This is of course well known, and can be seen most visibly in the case of plane Couette flow (e.g. Shepherd 1985).

7. Summary

By using conservation of energy and potential vorticity, nonlinear stability theorems have been obtained for the multilayer quasi-geostrophic equations which are analogous to Arnol'd's second stability theorem. Like Arnol'd's theorem, these new theorems can be seen as involving two conditions: one on the basic flow, and one on the geometry of the domain. The results establish rigorous upper bounds on both the energy and potential enstrophy of finite-amplitude disturbances to steady basic states, which are expressed in terms of the initial disturbance fields. These bounds hold uniformly in time, and tend to zero uniformly as the initial disturbance amplitude decreases to zero. It follows that the bounds establish nonlinear (normed) stability of the basic state. The present stability criteria improve significantly on previous results in this area.

For non-parallel basic states, the sufficient conditions for nonlinear stability consist of (3.1) together with the matrix $\boldsymbol{M}$ defined by (3.35) being positive definite; the latter condition involves the geometry of the domain. A simpler, though less powerful, alternative to the condition on $\boldsymbol{M}$ is given by (3.51). When the problem is zonally symmetric, then incorporation of the conservation of zonal momentum leads to a more powerful stability criterion in which (3.1) is replaced by (3.55) in the above description. In the case of the two-layer model, the stability criteria are concisely described by (4.4). Applying this to the Phillips model of baroclinic instability yields four regimes in which nonlinear stability holds. An interesting result is that the short-wave cut-off found in normal-mode instability is recovered precisely by (4.4), including the detailed shape of the short-wave part of the marginal stability curve when the vertical shear exceeds the minimum critical shear.

This paper was written while M.M. was visiting the University of Toronto under the auspices of an International Scientific Exchange Award from the Natural Sciences and Engineering Research Council of Canada. M.M., Q.C.Z. and Y.M.L. are supported by the National Natural Science Foundation of China. T.G.S. is supported by NSERC as well as by the Atmospheric Environment Service of Canada.

Appendix. Proof that $\tilde{\lambda} \leqslant \lambda$

Following Mu (1991), let $\tilde{\lambda}$ be the least positive eigenvalue of the elliptic problem

$$\nabla^2\phi_i + F_i \sum_{j=1}^{N} T_{ij}\,\phi_j + \tilde{\lambda}\phi_i = 0 \quad \text{in} \quad D \quad (i = 1, \ldots, N), \tag{A 1}$$

with boundary conditions

$$\phi_1|_{\partial D_0} = 0, \quad \left.\frac{\partial \phi_1}{\partial s}\right|_{\partial D} = 0, \quad \oint_{\partial D_j} \nabla\phi_1 \cdot \hat{\boldsymbol{n}}\, \mathrm{d}s = 0 \quad \text{for} \quad j = 1, \ldots, J, \tag{A 2a}$$

$$\left.\frac{\partial \phi_i}{\partial s}\right|_{\partial D} = 0, \quad \oint_{\partial D_j} \nabla\phi_i \cdot \hat{\boldsymbol{n}}\, \mathrm{d}s = 0 \quad \text{for} \quad j = 0, \ldots, J \quad (i = 2, \ldots, N). \tag{A 2b}$$

Let v_i denote the eigenfunction corresponding to $\tilde{\lambda}$. It is easy to verify from (A 1), (A 2) and the definition of T_{ij} that

$$\tilde{\lambda} = \frac{\displaystyle\iint_D \left\{ \sum_{i=1}^{N} d_i |\nabla v_i|^2 + \sum_{i=1}^{N-1} d_i\, F_i (v_{i+1} - v_i)^2 \right\} \mathrm{d}x\, \mathrm{d}y}{\displaystyle\iint_D \left\{ \sum_{i=1}^{N} d_i (v_i)^2 \right\} \mathrm{d}x\, \mathrm{d}y}. \tag{A 3}$$

Let u be the eigenfunction corresponding to the least positive eigenvalue λ of the problem (3.28). One sees from (3.28) that

$$\iint_D u\, \mathrm{d}x\, \mathrm{d}y = 0 \tag{A 4}$$

and that
$$\lambda = \frac{\iint_D |\nabla u|^2 \,\mathrm{d}x\,\mathrm{d}y}{\iint_D u^2 \,\mathrm{d}x\,\mathrm{d}y}. \tag{A 5}$$

Now define the vector function
$$w_i = u - c_0 \quad \forall i, \tag{A 6}$$

where $c_0 = u|_{\partial D_0}$. Obviously w_i satisfies the boundary conditions (A 2), whence we have

$$\tilde{\lambda} \leqslant \frac{\iint_D \left\{ \sum_{i=1}^{N} d_i |\nabla w_i|^2 + \sum_{i=1}^{N-1} d_i F_i (w_{i+1} - w_i)^2 \right\} \mathrm{d}x\,\mathrm{d}y}{\iint_D \left\{ \sum_{i=1}^{N} d_i (w_i)^2 \right\} \mathrm{d}x\,\mathrm{d}y}$$
$$= \frac{\iint_D |\nabla u|^2 \,\mathrm{d}x\,\mathrm{d}y}{\iint_D (u^2 + c_0^2) \,\mathrm{d}x\,\mathrm{d}y} \leqslant \frac{\iint_D |\nabla u|^2 \,\mathrm{d}x\,\mathrm{d}y}{\iint_D u^2 \,\mathrm{d}x\,\mathrm{d}y} = \lambda. \tag{A 7}$$

(In obtaining (A 7), the property (A 4) has been used.) This proves that $\tilde{\lambda} \leqslant \lambda$.

REFERENCES

ANDREWS, D. G. 1984 On the existence of nonzonal flows satisfying sufficient conditions for stability. *Geophys. Astrophys. Fluid Dyn.* **28**, 243–256.

ARNOL'D, V. I. 1965 Conditions for nonlinear stability of stationary plane curvilinear flows of an ideal fluid. *Dokl. Akad. Nauk. SSSR* **162**, 975–978. (English transl. *Sov. Maths* **6**, 773–777 (1965).)

ARNOL'D, V. I. 1966 On an a priori estimate in the theory of hydrodynamical stability. *Izv. Vyssh. Uchebn. Zaved. Matematika* **54**, no. 5, 3–5. (English transl. *Am. Math. Soc. Transl., Series 2* **79**, 267–269 (1969).)

CARNEVALE, G. F. & SHEPHERD, T. G. 1990 On the interpretation of Andrews' theorem. *Geophys. Astrophys. Fluid Dyn.* **51**, 1–17.

FJØRTOFT, R. 1950 Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex. *Geofys. Publ.* **17**, no. 6, 1–52.

HOLM, D. D., MARSDEN, J. E., RATIU, T. & WEINSTEIN, A. 1985 Nonlinear stability of fluid and plasma equilibria. *Phys. Rep.* **123**, 1–116.

LIU YONGMING & MU MU 1992 A problem related to nonlinear stability criteria for multilayer quasigeostrophic flow. *Adv. Atmos. Sci.* **9**, 337–345.

MCINTYRE, M. E. & SHEPHERD, T. G. 1987 An exact local conservation theorem for finite-amplitude disturbances to non-parallel shear flows, with remarks on Hamiltonian structure and on Arnol'd's stability theorems. *J. Fluid Mech.* **181**, 527–565.

MU MU 1991 Nonlinear stability criteria for motions of multilayer quasigeostrophic flow. *Science in China* B **34**, 1516–1528.

MU MU 1992 Nonlinear stability of two-dimensional quasi-geostrophic motions. *Geophys. Astrophys. Fluid Dyn.* **65**, 57–76.

MU MU & SHEPHERD, T. G. 1993 On Arnol'd's second nonlinear stability theorem for two-dimensional quasi-geostrophic flow. *Geophys. Astrophys. Fluid Dyn.* (in press).

PEDLOSKY, J. 1979 *Geophysical Fluid Dynamics*. Springer.

RIPA, P. 1992 Wave energy-momentum and pseudoenergy-momentum conservation for the layered quasi-geostrophic instability problem. *J. Fluid Mech.* **235**, 379–398.

SHEPHERD, T. G. 1985 Time development of small disturbances to plane Couette flow. *J. Atmos Sci.* **42**, 1868–1871.

SHEPHERD, T. G. 1988 Nonlinear saturation of baroclinic instability. Part I: The two-layer model. *J. Atmos. Sci.* **45**, 2014–2025.

SHEPHERD, T. G. 1990 Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Adv. Geophys.* **32**, 287–338.

ZENG QINGCUN 1989 Variational principles of instability of atmospheric motions. *Adv. Atmos. Sci.* **6**, 137–172.