University of Reading

Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics: part 1. pseudomomentum-based theory

Article

Published Version

Kushner, P. J. and Shepherd, T. G. ORCID: https://orcid.org/0000-0002-6631-9968 (1995) Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics: part 1. pseudomomentum-based theory. Journal Of Fluid Mechanics, 290. pp. 67-104. ISSN 0022-1120 doi: 10.1017/S0022112095002424 Available at https://centaur.reading.ac.uk/32895/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Published version at: http://dx.doi.org/10.1017/S0022112095002424

To link to this article DOI: http://dx.doi.org/10.1017/S0022112095002424

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

J. Fluid Mech. (1995), *vol.* 290, *pp.* 67–104
Copyright © 1995 Cambridge University Press

Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics. Part 1. Pseudomomentum-based theory

By PAUL J. KUSHNER AND THEODORE G. SHEPHERD
Department of Physics, University of Toronto, Toronto, Canada M5S 1A7

(Received 2 February 1994 and in revised form 22 August 1994)

There exists a well-developed body of theory based on quasi-geostrophic (QG) dynamics that is central to our present understanding of large-scale atmospheric and oceanic dynamics. An important question is the extent to which this body of theory may generalize to more accurate dynamical models. As a first step in this process, we here generalize a set of theoretical results, concerning the evolution of disturbances to prescribed basic states, to semi-geostrophic (SG) dynamics. SG dynamics, like QG dynamics, is a Hamiltonian balanced model whose evolution is described by the material conservation of potential vorticity, together with an invertibility principle relating the potential vorticity to the advecting fields. SG dynamics has features that make it a good prototype for balanced models that are more accurate than QG dynamics.

In the first part of this two-part study, we derive a pseudomomentum invariant for the SG equations, and use it to obtain: (i) linear and nonlinear generalized Charney–Stern theorems for disturbances to parallel flows; (ii) a finite-amplitude local conservation law for the invariant, obeying the group-velocity property in the WKB limit; and (iii) a wave–mean-flow interaction theorem consisting of generalized Eliassen–Palm flux diagnostics, an elliptic equation for the stream-function tendency, and a non-acceleration theorem. All these results are analogous to their QG forms.

The pseudomomentum invariant – a conserved second-order disturbance quantity that is associated with zonal symmetry – is constructed using a variational principle in a similar manner to the QG calculations. Such an approach is possible when the equations of motion under the geostrophic momentum approximation are transformed to isentropic and geostrophic coordinates, in which the ageostrophic advection terms are no longer explicit. Symmetry-related wave-activity invariants such as the pseudomomentum then arise naturally from the Hamiltonian structure of the SG equations. We avoid use of the so-called 'massless layer' approach to the modelling of isentropic gradients at the lower boundary, preferring instead to incorporate explicitly those boundary contributions into the wave-activity and stability results. This makes the analogy with QG dynamics most transparent.

This paper treats the f-plane Boussinesq form of SG dynamics, and its recent extension to β-plane, compressible flow by Magnusdottir & Schubert. In the limit of small Rossby number, the results reduce to their respective QG forms. Novel features particular to SG dynamics include apparently unnoticed lateral boundary stability criteria in (i), and the necessity of including additional zonal-mean eddy correlation terms besides the zonal-mean potential vorticity fluxes in the wave–mean-flow balance in (iii).

In the companion paper, wave-activity conservation laws and stability theorems based on the SG form of the pseudoenergy are presented.

1. Introduction

A classical and important area of research in fluid mechanics concerns the evolution of disturbances to prescribed basic states. Notable examples of theoretical questions falling within this framework include the propagation of waves in inhomogeneous media, the stability or instability of basic states, and the driving of mean-flow changes by wave transience and dissipation. Within the field of large-scale atmospheric dynamics, there is a well-developed body of such theory based on quasi-geostrophic (QG) dynamics. It includes Eliassen–Palm (E–P) flux diagnostics (Andrews & McIntyre 1976, 1978); finite-amplitude wave-activity conservation laws (McIntyre & Shepherd 1987); nonlinear stability theorems for parallel and/or steady flows (Holm *et al.* 1985; McIntyre & Shepherd 1987; Shepherd 1989, hereinafter referred to as S89); and rigorous saturation bounds on baroclinic instabilities (Shepherd 1988, 1993; S89). Aspects of this theory have been successfully applied to diagnose the role of eddies in simple models, primitive-equations general circulation models, and observations (e.g. Edmon, Hoskins & McIntyre 1980; Dunkerton, Hsu & McIntyre 1981; Hoskins 1983; Held & Hoskins 1985; Randel & Held 1991).

It is nevertheless widely recognized that QG theory has serious quantitative limitations. Therefore it is natural to ask to what extent the body of theory described above generalizes to other, more complicated dynamical models. To answer this question, one must consider what is fundamental about this theory, and what is specific to the QG model. Despite its seemingly broad range of form and application, it turns out that the above-mentioned QG theory has two related, central features, within which everything may be concisely understood. First, the dynamics can be described in terms of material advection of potential vorticity (PV), together with some sort of 'invertibility' or 'balance' relation that determines the advecting velocity given the PV (see e.g. Hoskins, McIntyre & Robertson 1985); and second, there is an underlying Hamiltonian structure which provides conservation laws for so-called 'wave-activity' invariants such as the pseudomomentum and pseudoenergy (see e.g. Shepherd 1990). Both of these features are exact, finite-amplitude properties of the dynamics.

Seen from this perspective, it seems plausible that the wave-activity conservation laws and stability theorems that underlie the modern understanding of QG dynamics may generalize to other models that are more accurate, yet which contain the two ingredients of balanced dynamics and Hamiltonian structure. The derivation of such models is an active topic of current research. At the present time, there is only one such model: semi-geostrophic (SG) dynamics.

The SG model was first introduced by Hoskins (1975) in the context of the f-plane Boussinesq equations. It is based on the 'geostrophic momentum' approximation of Eliassen (1948), and has been widely used to study the dynamics of mesoscale phenomena. More recently, the SG model has been extended to the β-plane compressible equations by Magnusdottir & Schubert (1990, hereinafter referred to as MSc90).

The purpose of the present study is to generalize the well-established body of QG theory consisting of wave-activity conservation laws and stability theorems to these two SG models. While the study is motivated by analogies between QG and SG dynamics, its length is largely due to the differences between them. Both models are described by material conservation of PV, with appropriate time-evolution equations for quantities at the boundaries. In both models, an invertibility principle relates the advecting stream-function potential to the PV and boundary terms. However, the SG model differs from the QG one in its use of a coordinate transformation to express the

invertibility principle, in the nonlinear nature of the principle, and, most problematically, in its form of the boundary conditions. The coordinate transformation to isentropic geostrophic coordinates (IGC) is used to reduce the SG model to its simplest form. However, the transformation introduces complications into the dynamics at fixed geometric bounding surfaces. In the transformed space, these surfaces become dynamic quantities whose positions evolve with time. This variability has been explicitly incorporated, where feasible, into the theory.

The paper is structured as follows. We first introduce the reader to SG dynamics, establishing notation and providing some insight into its properties. In §2, the f-plane Boussinesq system of Hoskins (1975) is introduced and transformed to IGC. The set of equations (2.18)–(2.20) summarize the prognostic interior and boundary equations and the diagnostic invertibility relation. Section 3 motivates the rest of the paper by using direct methods to derive a version of the Charney–Stern theorem for disturbances to parallel flows that includes explicit boundary contributions from both vertical and lateral bounding surfaces. The theorem is expressed in terms of a conservation law (equation (3.10) or (3.12)) that turns out to be an expression of linearized pseudomomentum conservation.

We then proceed with a systematic theoretical treatment. In §4, the conservation laws for the system are derived in physical coordinates. In §5, the pseudomomentum – i.e. the wave-activity invariant for disturbances to parallel flows that is related to the zonal symmetry of the system – is derived from a variational principle, after a discussion of the nature of the basic state and the variational problem at the boundaries. In §6, a nonlinear generalization of the Charney–Stern theorem is derived, with some restrictions on the nature of the perturbations at the boundaries. In §7, a local form of the wave-activity conservation law is developed, and the group-velocity property verified for the particular form of the flux. In §8, wave–mean-flow E-P flux diagnostics are developed, including a non-acceleration theorem. In §9, the results of §§2–8 are extended to the β-plane compressible system of MSc90. The results are summarized, and their relevance and limitations discussed, in §10.

In the sequel to this paper (Kushner & Shepherd 1995), wave-activity conservation laws and stability theorems based on the SG form of the pseudoenergy are presented.

2. Semi-geostrophic dynamics

Our discussion begins with the hydrostatic Boussinesq inviscid adiabatic equations of motion under the geostrophic momentum approximation (Eliassen 1948; Hoskins 1975), expressed in physical coordinates (x, y, z), where the pseudoheight z is defined by

$$z = z_p = \left[1 - \left(\frac{p}{p_0}\right)^{\kappa}\right]\frac{H_s}{\kappa}, \tag{2.1}$$

where p is the pressure, and the scale height $H_s = p_0/(\rho_0 g)$ is defined with respect to a reference level pressure p_0 and density ρ_0 (Hoskins & Bretherton 1972). The horizontal momentum equations are

$$\frac{\mathrm{D}u_g}{\mathrm{D}t} - fv = -\phi_x, \qquad \frac{\mathrm{D}v_g}{\mathrm{D}t} + fu = -\phi_y, \tag{2.2a, b}$$

where ϕ is the geopotential, and f a constant reference value of the Coriolis parameter. The material derivative is

$$\mathrm{D}/\mathrm{D}t = \partial_t + \boldsymbol{u}\cdot\boldsymbol{\nabla}_h + w\partial_z, \tag{2.3}$$

with velocity components

$$\boldsymbol{u} = (u, v) = \left(\frac{\mathrm{D}x}{\mathrm{D}t}, \frac{\mathrm{D}y}{\mathrm{D}t}\right), \quad w = \frac{\mathrm{D}z}{\mathrm{D}t}. \tag{2.4}$$

The geostrophic velocity $\boldsymbol{u}_g$ is defined by

$$\boldsymbol{u}_g = (u_g, v_g) = (-\phi_y/f, \phi_x/f). \tag{2.5}$$

Coordinate subscripts denote partial derivatives, and $\nabla_h = (\partial_x, \partial_y)$. From (2.2), we see that the geostrophic momentum approximation amounts to a neglect of the Lagrangian acceleration of the ageostrophic velocity $\boldsymbol{u}_{ag} = \boldsymbol{u} - \boldsymbol{u}_g$. The continuity equation is

$$\nabla_h \cdot \boldsymbol{u} + w_z = 0, \tag{2.6}$$

the thermodynamic equation

$$\mathrm{D}\theta/\mathrm{D}t = 0, \tag{2.7}$$

where θ is the potential temperature, and hydrostatic balance requires

$$\phi_z = g\theta/\theta_0, \tag{2.8}$$

where θ_0 is a reference value of θ.

We transform the system to IGC

$$(X, Y, Z, T) = \left(x + \frac{v_g}{f}, y - \frac{u_g}{f}, \frac{g\theta}{f^2\theta_0}, t\right), \tag{2.9}$$

and define the Montgomery–Bernoulli potential by

$$\Psi = \phi - f^2 Z z + \tfrac{1}{2}(u_g^2 + v_g^2). \tag{2.10}$$

The sum of the first and second terms on the right-hand side of (2.10) is the Montgomery potential for the system, associated with the transformation to isentropic coordinates. The sum of the first and third terms, on the other hand, is the Bernoulli function associated with geostrophic coordinates (Hoskins 1975). Further discussion of the Legendre transformations relating these different potentials may be found in Sewell & Roulstone (1993) and Roulstone & Norbury (1994). The partial derivatives transform as

$$\begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \\ \partial_t \end{pmatrix} = \begin{pmatrix} X_x & Y_x & Z_x & 0 \\ X_y & Y_y & Z_y & 0 \\ X_z & Y_z & Z_z & 0 \\ X_t & Y_t & Z_t & 1 \end{pmatrix} \begin{pmatrix} \partial_X \\ \partial_Y \\ \partial_Z \\ \partial_T \end{pmatrix}, \tag{2.11}$$

where the partial derivatives on the left-hand side and in the matrix are evaluated in physical space, and the derivatives ∂_X etc. in IGC. From (2.8)–(2.11) we find

$$\left.\begin{aligned} (\Psi_X, \Psi_Y, \Psi_T) = (\phi_x, \phi_y, \phi_t) = (f^2(X - x), f^2(Y - y), \phi_t) \\ \Psi_Z = -f^2 z, \quad \phi_z = f^2 Z. \end{aligned}\right\} \tag{2.12}$$

In IGC, the material derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \partial_T + \frac{\mathrm{D}X}{\mathrm{D}t}\partial_X + \frac{\mathrm{D}Y}{\mathrm{D}t}\partial_Y + \frac{\mathrm{D}Z}{\mathrm{D}t}\partial_Z,$$

and using the momentum equations (2.2), as well as (2.5), (2.7), (2.9) and (2.12), we find

$$\left(\frac{DX}{Dt}, \frac{DY}{Dt}, \frac{DZ}{Dt}\right) = \left(-\frac{\Psi_Y}{f}, \frac{\Psi_X}{f}, 0\right) = (u_g, v_g, 0).$$

Thus the material derivative contains no explicit ageostrophic advection terms:

$$\frac{D}{Dt} = \partial_T + \frac{1}{f}\Psi_X \partial_Y - \frac{1}{f}\Psi_Y \partial_X = \partial_T + \frac{1}{f}\partial_{XY}(\Psi, \cdot). \tag{2.13}$$

It is a straightforward matter to show that the materially conserved potential vorticity (Hoskins 1975) is proportional to the dimensionless quantity

$$q = \frac{\partial(X, Y, Z)}{\partial(x, y, z)} = \partial_{xyz}(X, Y, Z).$$

The fact that q is the Jacobian of the transformation $(x, y, z) \to (X, Y, Z)$ is used often below. Defining the non-dimensional potential pseudodensity (the term used in MSc90) $\sigma = 1/q$, we find

$$D\sigma/Dt = 0. \tag{2.14}$$

Careful treatment of the boundary conditions in the different coordinate systems is necessary in order to prove global invariance and to develop the wave-activity-based theory. In physical (pseudoheight) space, solid horizontal boundaries are subject to the condition that the boundaries are material surfaces:

$$\boldsymbol{u} \cdot \hat{\boldsymbol{n}} = 0 \quad \text{at horizontal walls,} \tag{2.15}$$

where $\hat{\boldsymbol{n}}$ is the unit vector normal to the horizontal wall. The vertical boundary condition

$$w = Dz/Dt = 0 \quad \text{at bounding } z\text{-surfaces} \tag{2.16}$$

is an approximate boundary condition, consistent with Boussinesq scaling (Hoskins & Bretherton 1972; White 1977), and equivalent to taking the domain to be bounded by material surfaces of constant pseudoheight, or pressure. That is, (2.16) implies that $\omega \equiv Dp/Dt = 0$ at the boundaries.

Other, and in general simpler, boundary conditions that will be considered below take the flow to be periodic or unbounded in a particular direction, or steady at a given boundary.

Since the coordinates in IGC are dependent on the velocity fields, material boundaries fixed in physical coordinates become dynamic quantities under the transformation. This detracts from the apparent simplicity of the equations derived above and summarized in (2.18)–(2.20) below. Andrews (1983) and others have tried to circumvent such problems by using the so-called 'massless layer' approach of Lorenz (1955): isentropic surfaces intersecting the earth are taken to extend below ground and are assigned to have the surface pressure. By this construction, no mass is trapped between the subsurface layers, so that the 'pseudodensity' $-p_\theta$ vanishes there: the layer is massless but infinitely stable. The bounding isentropic surface at the ground is the largest value of θ remaining below ground everywhere. Taking this surface to be one of constant height, we obtain as a vertical boundary condition

$$\phi = \text{const. at bounding, below-ground } \theta \text{ surface,} \tag{2.17}$$

with additional conditions at the interface, as outlined for example by Andrews (1983).

The approach has a strong conceptual appeal in that it is consistent with the idea of the equivalence of interior PV and surface θ dynamics in the equations of motion (Bretherton 1966). However, for the problem of determining conservation laws for the system, disconcerting questions remain regarding the nature of the kinematics of the massless layer and the boundary position, both of which are difficult to determine explicitly. Andrews (1983), for example, proposes the need for additional 'body force' terms in the momentum equations below ground. Furthermore, it is unclear how the approach generalizes to IGC when, for example, treating the intersection of Y-surfaces with bounding meridional walls. We also wish the wave activity to include explicitly the dynamic effect of surface θ-distributions, in order to aid physical interpretation and to allow a close comparison with the QG theory. For these reasons, our analysis avoids the 'massless layer' concept and makes direct use of the boundary conditions (2.15) and (2.16).

It is useful at this point to summarize the equations in their IGC forms both for future reference and to express the invertibility principle for the system. We consider flow in a box $[x_1, x_2]\times[y_1, y_2]\times[z_1, z_2]$ bounded on all sides by material surfaces. The governing prognostic equations are then

$$\mathrm{D}\sigma/\mathrm{D}t = 0 \quad \text{in the interior,} \tag{2.18a}$$

$$\mathrm{D}x/\mathrm{D}t = 0 \quad \text{at} \quad x = x_1, x_2, \tag{2.18b}$$

$$\mathrm{D}y/\mathrm{D}t = 0 \quad \text{at} \quad y = y_1, y_2, \tag{2.18c}$$

$$\mathrm{D}z/\mathrm{D}t = 0 \quad \text{at} \quad z = z_1, z_2, \tag{2.18d}$$

where the material derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \partial_T + \frac{1}{f}\partial_{XY}(\Psi, \cdot). \tag{2.19}$$

The diagnostic invertibility relations are

$$(x, y, z, \sigma) = \left(X - \frac{\Psi_X}{f^2}, Y - \frac{\Psi_Y}{f^2}, -\frac{\Psi_Z}{f^2}, \partial_{XYZ}(x, y, z)\right). \tag{2.20}$$

In IGC, the parallel between the QG and SG systems is clear: in the interior, both are governed by the quasi-two-dimensional advection of a three-dimensional PV-type quantity, itself related through an inversion to a scalar potential. Quantities advected at the boundaries are also related to the potential Ψ. The principal differences between the models lie in the complexity of the inversion, and in the variability of the boundaries. The invertibility condition requires σ to be of definite sign; since σ is proportional to the inverse of the potential vorticity, we require sign-definite PV. This ensures that the transformation to IGC is globally well-defined, since σ is its Jacobian. (It is possible to have well-defined coordinate transformations in the SG system which are associated with multivalued potentials (Purser & Cullen 1987; Chynoweth & Sewell 1989; Sewell & Roulstone 1993). This is important in the treatment of discontinuous solutions such as fronts.) For small Rossby number, from (2.9) and (2.20), $\sigma \approx z_Z \approx (f/N)^2$, where N is the buoyancy frequency. We therefore take the sign of σ to be positive.

Note that the necessary information pertaining to the mass distribution between isentropic Z-surfaces (Hoskins *et al.* 1985) is also implicit in the coordinate transformation.

3. A linear Charney–Stern theorem

In this section we derive a linear Charney–Stern theorem for small disturbances to a parallel steady basic state, directly from the linearized equations of motion. The method follows that of Eliassen (1983) and MSc90, but unlike the latter includes explicit boundary criteria. The theorem is connected to the conservation of pseudomomentum developed in §5, and can be obtained from that result. Nevertheless, we present a direct derivation here for a variety of reasons. The development introduces certain mathematical techniques necessary to approach the complicated boundary conditions, and serves as an independent cross-check for the plausibility of the pseudomomentum-based result. In addition, it provides some physical insight into the stability criteria, particularly the boundary contributions, that aids in the understanding of the more formal wave-activity derivations.

The approach is straightforward. We consider a zonally periodic channel bounded with material surfaces laterally at y_1, y_2 and vertically at z_1, z_2. We linearize the motion about a zonal steady basic state $\tilde{u}_g(y,z)$ that will be discussed in more precise detail in §5.1. In the basic state $\partial_x = 0$, and so from (2.11), $\partial_X = 0$. It follows that $\tilde{v}_g = 0$, and thus $\tilde{x} = X$, where here and henceforth tildes indicate the basic state and primes the disturbance quantities. Therefore averages in x and X are the same at the basic state. We obtain in the interior

$$\tilde{\mathrm{D}}\sigma'/\mathrm{D}t + \tilde{\sigma}_Y v_g' = 0, \tag{3.1}$$

where the linearized advection operator is

$$\tilde{\mathrm{D}}/\mathrm{D}t = \partial_T + \tilde{u}_g \partial_X. \tag{3.2}$$

Note from (2.9) and (2.12) that $v_g' = -fx' = \Psi'_X/f$. At the solid boundaries

$$\tilde{\mathrm{D}}y'/\mathrm{D}t + \tilde{y}_Y v_g' = 0 \quad \text{at} \quad y = y_1, y_2, \tag{3.3}$$

and

$$\tilde{\mathrm{D}}z'/\mathrm{D}t + \tilde{z}_Y v_g' = 0 \quad \text{at} \quad z = z_1, z_2. \tag{3.4}$$

Multiplying (3.1) by $(\sigma'/\tilde{\sigma}_Y)$ and integrating in $X = \tilde{x}$, we obtain

$$\frac{\partial}{\partial T}\left(\frac{\overline{(\sigma')^2}}{2\tilde{\sigma}_Y}\right) + \overline{v_g'\sigma'} = 0, \tag{3.5}$$

where zonal integration is denoted by an overbar.

We shall see later that the term in parentheses in (3.5) is proportional to the linearized pseudomomentum. We investigate the second term, the flux of inverse PV by the meridional geostrophic flow, following MSc90. To first order in disturbance amplitude,

$$\begin{aligned}\sigma' &= \partial_{XYZ}(x', \tilde{y}, \tilde{z}) + \partial_{XYZ}(\tilde{x}, y', \tilde{z}) + \partial_{XYZ}(\tilde{x}, \tilde{y}, z') \\ &= -(1/f^2)[\Psi'_{XX}\tilde{\sigma} + \partial_{YZ}(\Psi'_Y, \tilde{z}) + \partial_{YZ}(\tilde{y}, \Psi'_Z)],\end{aligned} \tag{3.6}$$

using $\tilde{x} = X$ and $\tilde{\sigma} = \partial_{YZ}(\tilde{y}, \tilde{z})$ at the basic state, together with (2.20). Then multiplying (3.6) by $v_g' = \Psi'_X/f$ we obtain

$$\begin{aligned}v_g'\sigma' = -\frac{1}{f^3}\Bigg[&\left(\tilde{\sigma}\frac{(\Psi'_X)^2}{2}\right)_X + \partial_{YZ}(\Psi'_X\Psi'_Y, \tilde{z}) + \partial_{YZ}(\tilde{y}, \Psi'_X\Psi'_Z) \\ &- \Psi'_Y\partial_{YZ}(\Psi'_X, \tilde{z}) - \Psi'_Z\partial_{YZ}(\tilde{y}, \Psi'_X)\Bigg] \qquad (3.7a)\end{aligned}$$

$$\begin{aligned}= \frac{\partial}{\partial X}\left\{\frac{1}{2f^3}\left[-\tilde{\sigma}(\Psi'_X)^2 + \tilde{y}_Y(\Psi'_Z)^2 - 2\tilde{y}_Z\Psi'_Y\Psi'_Z + \tilde{z}_Z(\Psi'_Y)^2\right]\right\} \\ - \partial_{YZ}\left(\frac{\Psi'_X\Psi'_Y}{f^3}, \tilde{z}\right) - \partial_{YZ}\left(\tilde{y}, \frac{\Psi'_X\Psi'_Z}{f^3}\right), \qquad (3.7b)\end{aligned}$$

where in (3.7*b*) we have used thermal wind balance to take $\tilde{z}_Y = \tilde{y}_Z$ (see (2.20)). Then invoking periodicity in X, we obtain

$$\overline{v_g' \sigma'} = \overline{\partial_{YZ}(v_g' y', \tilde{z})} + \overline{\partial_{YZ}(\tilde{y}, v_g' z')}. \tag{3.8}$$

The expression (3.7*b*) is the linearized E–P flux divergence, as will be seen in §7, and as is noted by MSc90 for the compressible β-plane case. Consider now the global integral of (3.8) over the domain in IGC:

$$\begin{aligned}\int_{\tilde{D}} v_g' \sigma' \,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z &= \int_{\tilde{D}} [\partial_{XYZ}(\tilde{x}, v_g' y', \tilde{z}) + \partial_{XYZ}(\tilde{x}, \tilde{y}, v_g' z')]\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z \\ &= \int_D [\partial_{\tilde{x}\tilde{y}\tilde{z}}(\tilde{x}, v_g' y', \tilde{z}) + \partial_{\tilde{x}\tilde{y}\tilde{z}}(\tilde{x}, \tilde{y}, v_g' z')]\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z}, \end{aligned} \tag{3.9}$$

where we have changed variables in the final step. The basic-state limits of integration in IGC are symbolized by $\tilde{D}$, while the limits of integration in physical coordinates are constant and the region is denoted D (cf. §5). Note that y' and z' are fluctuations in IGC and may be sensibly defined as functions of the physical coordinates. Now

$$\partial_{\tilde{x}\tilde{y}\tilde{z}}(\tilde{x}, v_g' y', \tilde{z}) = \partial_{\tilde{y}}(v_g' y')|_{\tilde{z}, \tilde{x}},$$

and similarly for the second Jacobian in (3.9). Then (3.5) and (3.9) imply

$$\int_D \frac{1}{\tilde{\sigma}} \frac{\partial}{\partial \tilde{T}} \left(\frac{(\sigma')^2}{2\tilde{\sigma}_Y} \right) \mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z} + \left[\int_{z_1}^{z_2} \overline{v_g' y'}\,\mathrm{d}\tilde{z} \right]_{y_1}^{y_2} + \left[\int_{y_1}^{y_2} \overline{v_g' z'}\,\mathrm{d}\tilde{y} \right]_{z_1}^{z_2} = 0,$$

which after using (3.3) and (3.4) may be written

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\{ \int_D \frac{\overline{(\sigma')^2}}{2\tilde{\sigma}\tilde{\sigma}_Y} \mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z} - \left[\int_{z_1}^{z_2} \frac{\overline{(y')^2}}{2\tilde{y}_Y} \mathrm{d}\tilde{z} \right]_{y_1}^{y_2} - \left[\int_{y_1}^{y_2} \frac{\overline{(z')^2}}{2\tilde{z}_Y} \mathrm{d}\tilde{y} \right]_{z_1}^{z_2} \right\} = 0, \tag{3.10}$$

where we have used the fact that in the basic state, averages in X and $\tilde{x}$ are identical.

Equation (3.10) and its generalization to the compressible β-plane case (9.17) provides a statement of the linearized Charney–Stern theorem for this system. If $\tilde{\sigma}\tilde{\sigma}_Y$ throughout the interior, $-\tilde{y}_Y$ over $y = y_2$, $\tilde{y}_Y$ over y_1, $-\tilde{z}_Y$ over z_2, and $\tilde{z}_Y$ over z_1 all have the same sign, then the disturbance may not grow (in a mean-square sense). Following MSc90, we can restate the result in terms of particle displacements η' brought about by the meridional geostrophic wind. Defining η' by

$$v_g' = \tilde{\mathrm{D}}\eta'/\mathrm{D}t$$

we may integrate (3.1), (3.3) and (3.4) to obtain

$$\left.\begin{aligned} \sigma' + \tilde{\sigma}_Y \eta' &= 0 \quad \text{in the interior,} \\ y' + \tilde{y}_Y \eta' &= 0 \quad \text{at} \quad y = y_1, y_2, \\ z' + \tilde{z}_Y \eta' &= 0 \quad \text{at} \quad z = z_1, z_2, \end{aligned}\right\} \tag{3.11}$$

so that we may rewrite (3.10) as

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\{ \int_D \frac{\tilde{\sigma}_Y \overline{(\eta')^2}}{2\tilde{\sigma}} \mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z} - \left[\int_{z_1}^{z_2} \frac{\tilde{y}_Y \overline{(\eta')^2}}{2} \mathrm{d}\tilde{z} \right]_{y_1}^{y_2} - \left[\int_{y_1}^{y_2} \frac{\tilde{z}_Y \overline{(\eta')^2}}{2} \mathrm{d}\tilde{y} \right]_{z_1}^{z_2} \right\} = 0. \tag{3.12}$$

Equations (3.10) and (3.12) provide statements of 'formal stability' (Holm *et al.*

1985) in the sense that they hold independently of the form of the small-amplitude disturbance. In particular, the disturbance is not required to be a normal mode. The following remarks concern (3.10), (3.12), and their β-plane compressible counterparts, (9.17) and (9.18).

(i) The Charney–Stern theorem of MSc90 comprises only the conservation of the interior PV term in (9.18), corresponding to the first term in (3.12). MSc90 take the boundary variability expressed in the third term to be included in the interior integral, with the massless-layer ideas discussed above as implicit justification. The idea has some merit from a conceptual viewpoint, but it is useful to have an explicit expression for the boundary contribution. At the meridional boundaries MSc90 use an approximate – and generally incorrect – boundary condition ($v_g' = 0$ there) that suppresses the walls' contributions to the conservation law.

(ii) The theorem is consistent with the QG Charney–Stern theorem in the small Rossby number limit, where the Rossby number $\epsilon = U/fL$, with U and L being characteristic velocity and length scales. For small ϵ, $\tilde{z}_Y \approx -(f/N)^2 Z_y$ for the QG vertically stratified reference state with a prescribed buoyancy profile $N = N(z)$. In the same limit, lateral wall contributions to (3.12) disappear because they are an ageostrophic effect, and we may correctly take $v = v_g + v_{ag} \approx v_g = 0$ there. Then we obtain the classical Charney–Stern theorem: the flow is necessarily stable if the meridional PV($\propto 1/\sigma$) gradient has the opposite sign to the meridional temperature gradient at the top and the same sign as the meridional temperature gradient at the bottom.

(iii) The lateral contributions from the second term in (3.12) have implications for the effect of lateral bounding surfaces on the stability of large-scale flows running parallel to them. Unless there is a large barotropic shear near one of the boundaries, we expect that at both meridional boundaries $\tilde{y}_Y > 0$, since $\tilde{y}_Y = 1 + (\tilde{u}_g)_Y/f = 1 + O(\epsilon)$. We will see below, moreover, that requiring $y_Y > 0$ at the boundary, i.e. requiring that Y be a reasonable meridional coordinate at the boundary, is a necessary condition for the construction of the pseudomomentum. Under this condition, we conclude that *the presence of the two lateral boundaries prevents the system from satisfying the Charney–Stern stability criteria.* A possible instability mechanism arising from a violation of the stability criteria would be a resonance between counterpropagating Kelvin edge waves, analogous to the instability arising from the resonance between counterpropagating Rossby edge waves in the baroclinic Eady model (cf. Hoskins *et al.* 1985, §6). That this instability can indeed occur has been demonstrated by Kushner (in preparation) for the special case of a linear anticyclonic barotropic shear flow, although the growth rates are exceedingly weak.

(iv) Kushner (1995) has extended the conservation law (3.12) to zonal flow along zonally invariant topography.

The conservation laws (3.10) and (3.12) serve as motivation for the rest of this study. It will be shown below that (3.10) is a statement of the conservation of linearized pseudomomentum for the system. The invariance of the pseudomomentum defined for finite-amplitude disturbances is, however, a more powerful result. From it we find a stability theorem for finite-amplitude disturbances to a linearly stable basic state, as well as a local finite-amplitude wave-activity conservation law that in turn yields a theorem of wave–zonal-mean-flow interaction.

4. Global invariants

We consider closed box, channel, or doubly periodic boundary conditions. Working with the system in physical coordinates, we obtain from (2.2)–(2.8) the energy flux law

$$E_t + \nabla_h \cdot [\boldsymbol{u}(E+\phi)] + \partial_z[w(E+\phi)] = 0, \tag{4.1}$$

where

$$E = \tfrac{1}{2}(u_g^2 + v_g^2) - f^2 Zz = \Psi - \phi \tag{4.2}$$

is the energy density in (x, y, z)-space (Hoskins 1975). Under the Boussinesq approximation, variations in density do not contribute to the energy. The density is thus factored out of (4.2), so that the energy has dimensions of velocity squared. Integrating (4.1) over the domain using the boundary conditions (2.16) and either (2.15) or x or y periodicity, we obtain conservation of the total energy, $\mathscr{E}$:

$$\frac{\mathrm{d}\mathscr{E}}{\mathrm{d}t} \equiv \frac{\mathrm{d}}{\mathrm{d}t}\int E\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int E_t\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 0. \tag{4.3}$$

In a similar manner, (2.2a) can be rewritten as a local flux law for the zonal impulse M:

$$M_t + \nabla_h \cdot (\boldsymbol{u}M) + \partial_z(wM) + \phi_x = 0, \tag{4.4}$$

where

$$M = -fY \tag{4.5}$$

has dimensions of momentum divided by mass. There is a corresponding conserved global impulse $\mathscr{M}$ for flow periodic in x:

$$\frac{\mathrm{d}\mathscr{M}}{\mathrm{d}t} \equiv \frac{\mathrm{d}}{\mathrm{d}t}\int M\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 0. \tag{4.6}$$

With material conservation of σ (2.18a) and of Z (2.7), we may determine a family of globally conserved quantities. For any function $C(\sigma, Z)$ we find using the continuity equation (2.6) that

$$C_t + \nabla_h \cdot (\boldsymbol{u}C) + \partial_z(wC) = 0. \tag{4.7}$$

This yields the global conservation law

$$\frac{\mathrm{d}\mathscr{C}}{\mathrm{d}t} \equiv \frac{\mathrm{d}}{\mathrm{d}t}\int C\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 0. \tag{4.8}$$

In the variational calculations below we will use the IGC form of the fields. The functional variations will be taken on integrals over the field densities in IGC, i.e. the densities in physical space multiplied by the Jacobian of the transformation from physical coordinates to IGC, $\partial_{XYZ}(x, y, z) = \sigma$. For example the zonal impulse density in IGC is M^*, defined by

$$\mathscr{M} = \int M\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int M^*\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z, \tag{4.9}$$

i.e.

$$M^* = -fY\sigma. \tag{4.10}$$

The functionals $\mathscr{C}$ thus take the form

$$\mathscr{C} = \int C^*(\sigma, Z)\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z, \tag{4.11}$$

where $C^* = \sigma C(\sigma, Z)$ is an arbitrary function of σ and Z.

The conservation of the quantities $\mathscr{E}$, $\mathscr{M}$, and $\mathscr{C}$ is linked to the underlying Hamiltonian structure of the dynamics (Salmon 1985, 1988a; Roulstone & Norbury

1994; Kushner 1993). From this point of view, $\mathscr{E}$ is the Hamiltonian functional reflecting, through Noether's theorem, the time symmetry in the system, and $\mathscr{M}$ the corresponding zonal impulse invariant reflecting the system's zonal symmetry. The functionals $\mathscr{C}$ are the system's Casimir invariants, arising from degeneracies in the Poisson operator in the Hamiltonian formulation, and corresponding to the 'particle relabelling' symmetries (e.g. Salmon 1988*b*; Shepherd 1990). Further discussion is provided in Part 2 of this study (Kushner & Shepherd 1995).

5. Pseudomomentum

A 'wave-activity' invariant is defined to be a conserved quantity that is second order in the amplitude of a disturbance to a reference or 'basic' state with some continuous symmetry. A review of the concept of these invariants, together with a general discussion of their role within Hamiltonian theory, is provided by Shepherd (1990). The 'pseudomomentum' is the wave activity, defined with respect to zonally symmetric basic states, that is associated with the zonal symmetry of the system; it is an exact invariant of the nonlinear dynamics. In a similar way the pseudoenergy, which is considered in Part 2 of this study, is associated with the temporal symmetry of the system and is likewise an exact invariant. The 'wave action' (Whitham 1965; Bretherton & Garrett 1968) is related but distinct from these quantities; it is associated with the symmetry properties of a phase-averaged Lagrangian, rather than an exact coordinate symmetry.

5.1. *Basic state*

Consider a steady zonal basic state in physical space. As noted in § 3, $\partial_x = 0$ implies that $\partial_X = 0$ in the basic state. Because $\tilde{v}_g = \tilde{\phi}_x/f = 0$ in the basic state, the meridional boundaries are straight in IGC, i.e. constant in Y for all X, at every Z-level. However, in general the distance $Y_2 - Y_1$ may vary with Z, and the bounding z-surfaces will be curved as well.

At the basic state

$$(\Psi, \sigma) = (\tilde{\Psi}(Y, Z), \tilde{\sigma}(Y, Z)). \tag{5.1}$$

Following previous approaches (e.g. McIntyre & Shepherd 1987; S89) we define an inverse map in the interior, $\tilde{Y}_{int}(\tilde{\sigma}, Z)$: for every Z, values of σ map onto particular Y-values at the basic state. If $\tilde{\sigma}$ is not monotonic in Y, $\tilde{Y}_{int}$ will be multivalued in $\tilde{\sigma}$, and additional Lagrangian information will be needed to fully define the inverse map. See McIntyre & Shepherd (1987) for a more complete discussion of this issue.

At each of the surfaces at the basic state, there is an isentropic distribution $\tilde{Z}_s(Y)$, where s is a discrete index specifying the surface. Let us denote the surface inverse map $\tilde{Y}_s(Z)$ such that for example at z_1

$$\tilde{Y}_s(Z) = [\tilde{Z}_{z_1}(Y)]^{-1} \quad \text{at the basic state, lower } z\text{-surface.}$$

Depending on the monotonicity of $\tilde{Z}$ in Y, $\tilde{Y}_s$ may also be multivalued. The index s serves in a sense as another Lagrangian marker. There is also a basic-state surface distribution of $\tilde{\sigma}$, labelled $\tilde{\sigma}_s$, and defined by

$$\tilde{\sigma}_s(Z) = \tilde{\sigma}(\tilde{Y}_s(Z), Z). \tag{5.2}$$

Although the surface maps are defined with respect to the basic-state Z-distributions, they may be treated simply as functions of Z alone. Thus integrals over the functions are Casimirs of the form (4.11), although independent of σ. It is for this reason that $\tilde{\sigma}_s$ is defined as a function of Z rather than of Y.

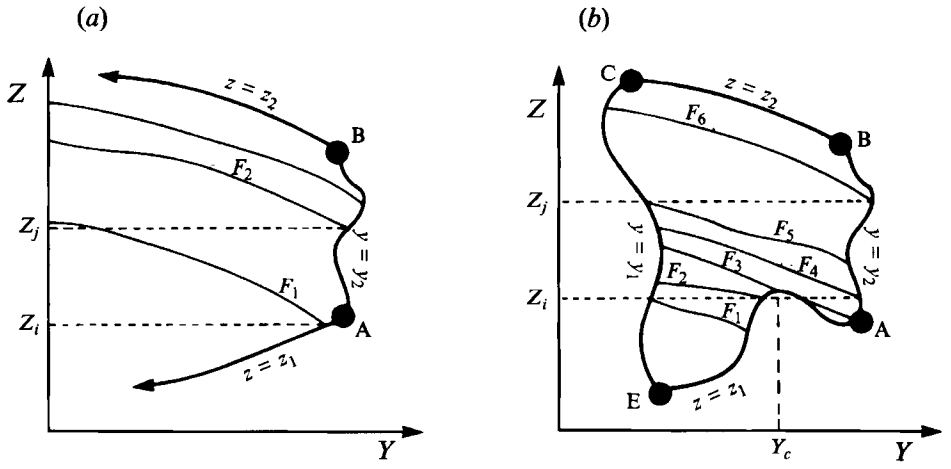

FIGURE 1. (*a*) A zonal basic-state profile sketched in IGC. The heavy solid line shows the material boundaries, with the $z = z_1$ surface unbounded to the south and ending at A to the north, the northern $y = y_2$ surface running between A and B, and the upper $z = z_2$ surface running south from B. The solid contours are curves of constant $\tilde{F}(Y, Z)$, a function monotonic in Y and Z. Dashed contours are of the surface function $\tilde{F}_s(Z)$. This basic state is linearly stable for $-\tilde{\sigma}_Y > 0$. (*b*) As in (*a*), but for a basic state breaking the stability criteria in two ways, for $-\tilde{\sigma}_Y > 0$: the presence of the southern boundary $y = y_1$, between C and E, and a surface warm anomaly at $Y = Y_c$ on the z_1-surface. Now $\tilde{F}_s(Z)$ is multivalued.

Figures 1 (*a*) and 1 (*b*) illustrate the definition of the surface maps. Figure 1 (*a*) depicts a Y, Z cross-section of a zonal basic state that would be stable according to (3.10) for positive interior PV gradients, $-\tilde{\sigma}_Y > 0$. The flow is unbounded to the south, the lower boundary $z = z_1$ ends at point A, the northern boundary y_2 runs between A and B, and the upper boundary z_2 continues south from point B. For these boundaries $-\tilde{z}_Y$ at z_1, $\tilde{z}_Y$ at z_2, and $\tilde{y}_Y$ at y_2 are all positive for $\tilde{z}_Z > 0$. Consider a function $\tilde{F}(Y, Z)$ which for purposes of this illustration is monotonic and increasing in both Y and Z. Contours of $\tilde{F}$ are sketched in the figure. At a boundary, the surface distribution $\tilde{F}_s$ is defined uniquely for all Z: we define it to be the value of $\tilde{F}$ at the point (Y_s, Z) where the surface's isentropic height is Z. Horizontal contours of $\tilde{F}_s$ are sketched with dashed lines in the figure. Hence $\tilde{F}_s(Z_i) = \tilde{F}_{z_1}(Z_i) = F_1$ and $\tilde{F}_s(Z_j) = F_{y_2}(Z_j) = F_2$.

Considering figure 1 (*b*), we see how the uniqueness of the inverse surface maps is broken under conditions that violate the linear stability criteria. We include for illustration two 'problem' regions, a 'bump' (corresponding to a 'warm anomaly') at Y_c on the z_1-surface, where $\tilde{z}_Y$ changes sign; and the presence of a southern boundary $y = y_1$ with $\tilde{y}_Y > 0$ running between points C and E. $\tilde{F}_s(Z)$ is no longer defined without additional 'point of origin' information. At Z_j, for example, the function can take either the value F_6 if $s = y_2$ or F_5 if $s = y_1$. At Z_i, the function can have one of four values: F_4 if $s = y_2$; F_3 if the region $Y > Y_c$ is specified on z_1; F_2 for $s = z_1$ and $Y < Y_c$; or F_1 for $s = y_1$. We emphasize the connection between violation of the stability criteria and the multivaluedness of $\tilde{F}_s$ whatever the monotonicity properties of $\tilde{F}$.

5.2. *Finite-amplitude pseudomomentum*

We seek a conserved quantity, $\mathscr{S}$, that is second order relative to the zonal basic state. This requires that $\mathscr{S}$ be of the form

$$\mathscr{S} = \mathscr{A}[\sigma] - \mathscr{A}[\tilde{\sigma}], \tag{5.3}$$

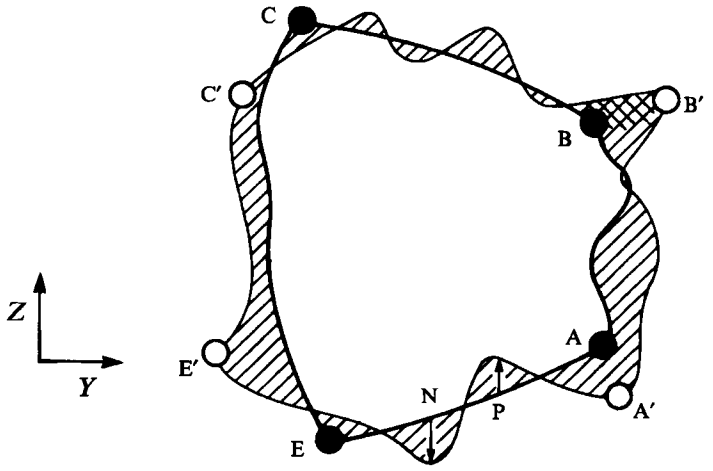

FIGURE 2. Cross-section of boundaries for a given basic state (heavy solid curve ABCE) and perturbed state (lighter solid curve A'B'C'E') at a given value of X. The hatched region is D'. The arrows at P and N represent positive and negative Z-variations of the z_1-surface. The cross-hatched region near the upper and northern edge B is of second order relative to the variations at P and N.

where σ and $\tilde{\sigma}$ symbolize all perturbed and basic-state fields, and that $\mathscr{A}$ be extremal at the basic state, i.e.

$$\delta\mathscr{A} = 0 \quad \text{at the basic state.} \tag{5.4}$$

In order to apply a variational principle of the form (5.4) we must include the variation at the boundaries explicitly. Consider an integral $\mathscr{F}$ of a density F whose local variation is denoted δF. Rather than specify limits of integration, we will denote the variable region as $\tilde{D}$ and variations to it by D'. We find

$$\delta\mathscr{F} = \int_{\tilde{D}} \delta F \,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z + \left[\int_{y_t,\tilde{D}} F\delta Y|_{y_t}\,\mathrm{d}Z\,\mathrm{d}X\right]\Bigg|_{y_1}^{y_2} + \left[\int_{z_t,\tilde{D}} F\delta Z|_{z_t}\,\mathrm{d}X\,\mathrm{d}Y\right]\Bigg|_{z_1}^{z_2}. \tag{5.5}$$

The variations at x_1 and x_2 cancel owing to periodicity. The quantity $\delta Z|_{z_1}$ is the displacement in IGC of the z_1 boundary at a given (X, Y), and $\delta Y|_{y_1}$ similarly represents displacements of y_1 in IGC for each (Z, X). These displacements are unambiguously defined as long as

$$0 < |y_Y|_{Z,X}| < \infty \text{ at } y\text{-boundaries}, \quad 0 < |z_Z|_{X,Y}| < \infty \text{ at } z\text{-boundaries}, \tag{5.6}$$

for both the perturbed and unperturbed boundaries. If we take the gradients to be positive, the conditions (5.6) require, in essence, that Z and Y act as sensible vertical and meridional coordinates in a statically stable flow with Rossby number less than unity (see remark (iii) in §3).

We note that (5.5) neglects as second-order quantities the variations in the limits of integration along the meridional edges $\{(x, y_1, z_1)\colon x_1 \leqslant x \leqslant x_2\}$, etc. The contributions at the corners cancel owing to periodicity. Hence, to leading order, the limits of integration on all the integrals are taken to be those of the basic state and there is no additional need to consider the variations D'. The subscripts $\tilde{D}$ on the surface y- and z-integrals indicate this approximation.

Figure 2 schematically illustrates the boundary variations in IGC. At a given X, the

boundary of the domain is outlined by a heavy curve for the basic state and a lighter curve for the perturbed (non-zonal) state. The region D' is the hatched region representing the perturbation of the domain. A perturbation $\delta Z|_{z_1}$ is shown to be positive at point P and negative at point N. It is supposed that under perturbation, each point on a given bounding surface maps onto another, and the variations $\delta Y|_y, \delta Z|_z$ parametrize these displacements everywhere except near the edges. Such a second-order edge contribution is represented by the cross-hatched region in the figure.

Consider, for example, variations in the zonal impulse $\mathscr{M}$. From (4.9), (4.10) and (5.5),

$$\delta\mathscr{M} = -\int_{\tilde{D}} fY\delta\sigma\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z - \left[\int_{y_t,\tilde{D}} fY\sigma\,\delta Y|_{y_t}\,\mathrm{d}Z\,\mathrm{d}X\right]\Bigg|_{y_1}^{y_2} - \left[\int_{z_t,\tilde{D}} fY\sigma\,\delta Z|_{z_t}\,\mathrm{d}X\,\mathrm{d}Y\right]\Bigg|_{z_1}^{z_2}. \quad (5.7)$$

Clearly the zonal impulse is not extremal for this basic state as its first variation is, in general, non-zero. As is standard practice for wave-activity construction (e.g. Shepherd 1990), we define the invariant

$$\mathscr{A} = \mathscr{M} + \mathscr{C}, \quad (5.8)$$

whose conservation is guaranteed by (4.6) and (4.8), and then determine $\mathscr{C}$ such that $\mathscr{A}$ satisfies (5.4). On general grounds one may always expect such a $\mathscr{C}$ to exist (e.g. Shepherd 1990).

For any Casimir $\mathscr{C}$ of the form (4.11),

$$\delta\mathscr{C} = \int_{\tilde{D}} \left.\frac{\partial C^*}{\partial\sigma}\right|_Z \delta\sigma\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z + \left[\int_{y_t,\tilde{D}} C^*\,\delta Y|_{y_t}\,\mathrm{d}Z\,\mathrm{d}X\right]\Bigg|_{y_1}^{y_2} + \left[\int_{z_t,\tilde{D}} C^*\,\delta Z|_{z_t}\,\mathrm{d}X\,\mathrm{d}Y\right]\Bigg|_{z_1}^{z_2}. \quad (5.9)$$

In order to satisfy (5.4), we require from (5.7) and (5.9) that

$$\left.\begin{aligned} \partial C^*/\partial\sigma|_Z &= fY \text{ in the interior, basic state,} \\ C^* &= fY\sigma \text{ at all boundaries, basic state.} \end{aligned}\right\} \quad (5.10)$$

A Casimir density satisfying (5.10) takes the local form

$$C^*(\sigma, Z) = \int_{\tilde{\sigma}_s(Z)}^{\sigma} f\tilde{Y}_{int}(\hat{\sigma}, Z)\,\mathrm{d}\hat{\sigma} + f\tilde{Y}_s(Z)\,\tilde{\sigma}_s(Z). \quad (5.11)$$

We now consider a finite-amplitude disturbance to the basic state. We define disturbance quantities Ψ' and σ' by

$$\Psi = \tilde{\Psi} + \Psi', \quad \sigma = \tilde{\sigma} + \sigma', \quad (5.12)$$

where we have dropped the δ-variational notation to emphasize that the disturbance can be large. We leave the limits of integration of both the $\tilde{D}$- and the D'-regions unspecified for the moment.

Note that the inverse maps $\tilde{Y}_{int}$, $\tilde{Y}_s$ and $\tilde{\sigma}_s$ are only defined over the range of values of σ and Z present in the basic state. If the disturbance (5.12) introduces values outside that range, then the function (5.11) must be extended to include such values. One can make this extension arbitrarily without compromising the fact that C^* is a Casimir

density, and is therefore conserved (cf. Arnol'd 1966). Bearing this possibility in mind, the finite-amplitude pseudomomentum can be written

$$\begin{aligned}
\mathscr{S} &= \mathscr{M}[\sigma]+\mathscr{C}[\sigma]-\mathscr{M}[\tilde{\sigma}]-\mathscr{C}[\tilde{\sigma}] \\
&= f\Big\{\int_{\tilde{D}+D'}\Big[-Y\tilde{\sigma}-Y\sigma'+\int_{\tilde{\sigma}_s}^{\sigma}\tilde{Y}_{int}(\hat{\sigma},Z)\,\mathrm{d}\hat{\sigma}+\tilde{Y}_s(Z)\,\tilde{\sigma}_s(Z)\Big]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z \\
&\quad -\int_{\tilde{D}}\Big[-Y\tilde{\sigma}+\int_{\tilde{\sigma}_s}^{\tilde{\sigma}}\tilde{Y}_{int}(\hat{\sigma},Z)\,\mathrm{d}\hat{\sigma}+\tilde{Y}_s(Z)\,\tilde{\sigma}_s(Z)\Big]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z\Big\} \\
&= f\Big\{\int_{\tilde{D}+D'}\Big[\int_{\tilde{\sigma}}^{\sigma}[\tilde{Y}_{int}(\hat{\sigma},Z)-\tilde{Y}_{int}(\tilde{\sigma},Z)]\,\mathrm{d}\hat{\sigma}\Big]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z \\
&\quad +\int_{D'}\Big[\int_{\tilde{\sigma}_s}^{\tilde{\sigma}}\tilde{Y}_{int}(\hat{\sigma},Z)\,\mathrm{d}\hat{\sigma}-Y\tilde{\sigma}+\tilde{Y}_s(Z)\,\tilde{\sigma}_s(Z)\Big]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z\Big\} \\
&= f\Big\{\int_{\tilde{D}+D'}\Big[\int_{0}^{\sigma'}[\tilde{Y}_{int}(\tilde{\sigma}+\hat{\sigma},Z)-\tilde{Y}_{int}(\tilde{\sigma},Z)]\,\mathrm{d}\hat{\sigma}\Big]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z \\
&\quad +\int_{D'}\Big[\int_{\tilde{\sigma}_s}^{\tilde{\sigma}}-\hat{\sigma}\frac{\partial\tilde{Y}_{int}(\hat{\sigma},Z)}{\partial\hat{\sigma}}\,\mathrm{d}\hat{\sigma}\Big]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z\Big\} \\
&\equiv \mathscr{S}_{int}+\mathscr{S}_s,
\end{aligned} \tag{5.13}$$

where in the derivation of (5.13) we have used the identities

$$\tilde{Y}_{int}(\tilde{\sigma},Z)=Y,\quad \tilde{Y}_{int}(\tilde{\sigma}_s(Z),Z)=\tilde{Y}_s(Z).$$

The final expressions in (5.13) formally define the pseudomomentum in terms of both surface and interior contributions.

5.3. *Comparison with QG theory and small-amplitude reduction*

It is worthwhile to compare (5.13) with the pseudomomentum invariant of QG theory. From S89,

$$\begin{aligned}
\mathscr{S}_{QG} &= \int_D\Big[-\rho\int_0^{q'}[\tilde{Y}_{int}(\tilde{q}+\hat{q};z)-\tilde{Y}_{int}(\tilde{q};z)]\,\mathrm{d}\hat{q}\Big]\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\
&\quad +\sum_{i=1}^{2}(-1)^i\int_{z_i}\frac{\rho fg}{N^2\theta_s}\Big[\int_0^{\theta'}[\tilde{Y}_{z_i}(\tilde{\theta}+\hat{\theta})-\tilde{Y}_{z_i}(\tilde{\theta})]\,\mathrm{d}\hat{\theta}\Big]\mathrm{d}x\,\mathrm{d}y.
\end{aligned} \tag{5.14}$$

In (5.14), $\rho=\rho(z)$, $\theta_s=\theta_s(z)$ and $N=N(z)$ are the prescribed reference-state density, potential temperature and buoyancy frequency profiles, q is the QG PV, and $\theta=\tilde{\theta}+\theta'$ the potential temperature deviation away from θ_s. Both q and θ are related, via the QG invertibility principle, to the stream function. The form, as written, is defined on the β-plane, with the β-term included in the definition of the PV. The $\tilde{Y}_{int}$ maps defined at every height z are similar to those we have used for the SG analysis except that there is no explicit interior-surface connection as defined in (5.2). The reason for this is that z is not a dynamical variable in the QG model.

By inspection we see that the interior contributions to $\mathscr{S}$, i.e the first integrals in (5.13) and (5.14), are analogous. At small amplitude

$$\mathscr{S}_{int} \equiv \int_{\tilde{D}+D'} \left[f \int_0^{\sigma'} [\tilde{Y}_{int}(\tilde{\sigma}+\hat{\sigma}, Z) - \tilde{Y}_{int}(\tilde{\sigma}, Z)] \, \mathrm{d}\hat{\sigma} \right] \mathrm{d}X \, \mathrm{d}Y \, \mathrm{d}Z$$

$$\approx \int_{\tilde{D}} \frac{f(\sigma')^2}{2\tilde{\sigma}_Y} \mathrm{d}X \, \mathrm{d}Y \, \mathrm{d}Z = \int_D \frac{\overline{f(\sigma')^2}}{2\tilde{\sigma}\tilde{\sigma}_Y} \mathrm{d}\tilde{y} \, \mathrm{d}\tilde{z}, \tag{5.15}$$

which has the same form, within a factor of f, as the first term in braces in (3.10). The small-amplitude form of the QG interior wave-activity density in (5.14) is

$$S_{int,QG} \approx -\frac{\rho(q')^2}{2\tilde{q}_y}. \tag{5.16}$$

The differences between (5.15) and (5.16) are due to the fact that σ is the inverse of q, and that we have removed the dimensional reference density in the Boussinesq SG system.

The appearance of the finite-amplitude surface contributions in (5.13) and (5.14), on the other hand, are quite distinct. This is not entirely surprising given the differences in construction of the QG and SG expressions. In the QG system, the zonal impulse and Casimirs have separate contributions at the vertical boundaries (cf. Shepherd 1990, §4), and all limits of integration are fixed. Conversely, in the SG system, the first equality of (5.13) has no explicit boundary contributions, but the contributions emerge as differences in the limits of integration of the basic- and disturbed-states' volume integrals.

Despite the different forms of the pseudomomentum boundary integrals, it may be shown that at small amplitude the SG surface contributions are analogous to the QG ones. To do so, we first divide the boundary contributions in (5.13) surface by surface, taking

$$\mathscr{S}_s = \sum_{i=1}^{2} (\mathscr{S}_{y_i} + \mathscr{S}_{z_i}),$$

where at small amplitude

$$\mathscr{S}_{y_i} = \int_{\tilde{D}, y_i} \left\{ (-1)^i f \int_{\tilde{Y}}^{\tilde{Y}+Y'} \left[\int_{\tilde{\sigma}_s}^{\tilde{\sigma}} -\hat{\sigma} \frac{\partial \tilde{Y}_{int}(\hat{\sigma}, Z)}{\partial \hat{\sigma}} \mathrm{d}\hat{\sigma} \right] \mathrm{d}\hat{Y} \right\} \mathrm{d}Z \, \mathrm{d}X, \tag{5.17}$$

and

$$\mathscr{S}_{z_i} = \int_{\tilde{D}, z_i} \left\{ (-1)^i f \int_{\tilde{Z}}^{\tilde{Z}+Z'} \left[\int_{\tilde{\sigma}_s}^{\tilde{\sigma}} -\hat{\sigma} \frac{\partial \tilde{Y}_{int}(\hat{\sigma}, \hat{Z})}{\partial \hat{\sigma}} \mathrm{d}\hat{\sigma} \right] \mathrm{d}\hat{Z} \right\} \mathrm{d}X \, \mathrm{d}Y. \tag{5.18}$$

The paths of integration in $\hat{Y}$ or $\hat{Z}$ are meant to represent the displacement of a material boundary in IGC. The notation $\tilde{D}$ indicates that fixed basic-state limits of integration are to be taken for the outer variables of integration, i.e. (Z, X) in (5.17) and (X, Y) in (5.18). The expressions are only valid for small amplitude because they neglect the edge contributions as discussed above. Thus the displacement of the z_i-surface is parametrized at fixed (X, Y) as a Z-displacement $\hat{Z}$. Similarly, the displacement of the y_i-surface is expressed at fixed (Z, X) as a Y-displacement $\hat{Y}$. The lower limit of integration $\tilde{Z}(X, Y)$ represents the Z-position of the z_i-surface at the basic state, for every (X, Y); a similar statement applies to $\tilde{Y}(Z, X)$. The integrands in square brackets are thus integrated for each fixed outer variable towards the perturbed position, $\tilde{Z}+Z'$ or $\tilde{Y}+Y'$.

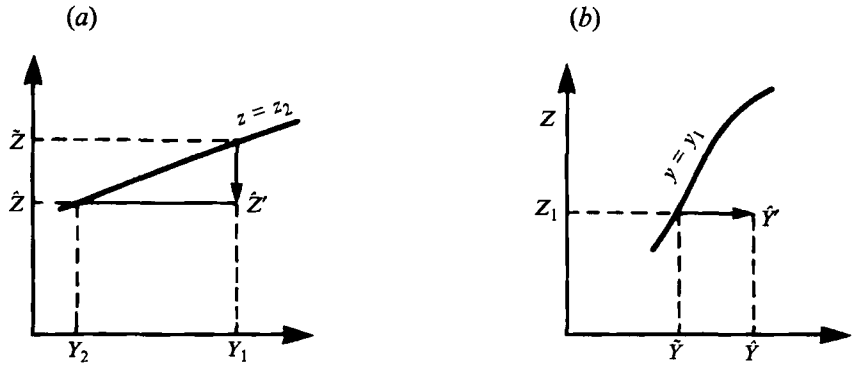

FIGURE 3. (*a*) Sketch used to evaluate (5.20), for a displacement $\hat{Z}' = \hat{Z} - \tilde{Z} < 0$ away from the surface $z = z_2$. Here $\tilde{Y}_s(\tilde{Z}) = Y_2$, and $Y_1 - \tilde{Y}_s(\tilde{Z}) = Y_1 - Y_2 \approx -Y_Z|_{z_2}\hat{Z}'$. (*b*) As in (*a*), for a displacement $\hat{Y}' = \hat{Y} - \tilde{Y} > 0$ away from the surface $y = y_1$.

The first-order contribution to the integrands in square brackets in (5.17) and (5.18) is

$$\int_{\tilde{\sigma}_s}^{\tilde{\sigma}} -\hat{\sigma}\frac{\partial \tilde{Y}_{int}(\hat{\sigma}, Z)}{\partial \hat{\sigma}}\,\mathrm{d}\hat{\sigma} \approx \left(-\tilde{\sigma}\frac{\partial \tilde{Y}_{int}}{\partial \tilde{\sigma}}\bigg|_{\tilde{\sigma}\approx\tilde{\sigma}_s}\right)(\tilde{\sigma}-\tilde{\sigma}_s). \tag{5.19}$$

From (5.2),

$$\tilde{\sigma} - \tilde{\sigma}_s = \tilde{\sigma}(Y, Z) - \tilde{\sigma}(\tilde{Y}_s(Z), Z) \approx (Y - \tilde{Y}_s(Z))\,\tilde{\sigma}_Y. \tag{5.20}$$

To evaluate (5.20), we first take a point away from a given z-surface. Figure 3(*a*) illustrates a point $(Y_1, \hat{Z})$ located near and below the basic-state z_2-surface point $(Y_1, \tilde{Z}) = (Y_1, Z_s)$, representing the lower limit of integration in (5.18), with $\hat{Z} - Z_s = \hat{Z}' < 0$. In the illustration, the surface slope $Z_Y|_{z_2} > 0$. To lowest order, $Y - \tilde{Y}_s = Y_1 - Y_2 \approx -Y_Z|_{z_2}\hat{Z}'$. The argument applies generally so that near a z_i-surface

$$\tilde{\sigma}(Y, Z) - \tilde{\sigma}_{z_i}(Z) = -\tilde{\sigma}_Y\, Y_Z|_{z_i}\hat{Z}'. \tag{5.21}$$

For displacements of the y-surface, we consider the example in figure 3(*b*), showing the point $(\hat{Y}, Z_1)$ to the right of the y_1 basic-state surface point $(\tilde{Y}, Z_1) = (Y_s, Z_1)$. Here $\hat{Y} - Y_s = \hat{Y}' > 0$, and in general

$$\tilde{\sigma}(\hat{Y}, Z) - \tilde{\sigma}_{y_i}(Z) = \tilde{\sigma}_Y\,\hat{Y}'. \tag{5.22}$$

Combining (5.18)–(5.21), and using the small-amplitude relation $z'|_Z = -\tilde{z}_Z Z'|_z$ with $\tilde{z}_Z$ satisfying the second of conditions (5.6), we obtain for the z-surface contribution

$$\mathscr{S}_{z_i} \approx (-1)^i f \int_{\tilde{D}, z_i} \tilde{\sigma} Y_Z|_{z_i}\frac{(Z')^2}{2}\,\mathrm{d}X\,\mathrm{d}Y \tag{5.23a}$$

$$= (-1)^i f \int_{z_i} \partial_{XYZ}(\tilde{x}, \tilde{y}, \tilde{z})\,\partial_{XZ\tilde{z}}(X, Y, \tilde{z})\,[\partial_{XY\tilde{z}}(X, Y, Z)]^2$$

$$\times\, \partial_{\tilde{x}\tilde{y}\tilde{z}}(X, Y, \tilde{z})\frac{(z')^2}{2}\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y} \tag{5.23b}$$

$$= -(-1)^i f \int_{y_1}^{y_2} \frac{\overline{(z')^2}}{2\tilde{z}_Y}\,\mathrm{d}\tilde{y}\bigg|_{z=z_i}. \tag{5.23c}$$

Note that since $\partial_{\tilde{x}\tilde{y}\tilde{z}}(X, Y, \tilde{z}) = \tilde{z}_Z/\tilde{\sigma}$, the change of variables made to obtain (5.23*b*) is well defined. Equation (5.23*c*) is identical (within a factor of f) to the separate z-surface contributions in the conservation equation (3.10).

The QG contribution to the wave activity at small amplitude at the z-surfaces is, from (5.14),

$$\mathscr{S}_{z_i,QG} \approx (-1)^i \int_{z_i} \frac{\rho f g(\theta')^2}{2N^2 \theta_s \tilde{\theta}_y} \mathrm{d}x\,\mathrm{d}y, \tag{5.24}$$

with fluctuations defined in the physical coordinates. The form (5.24), representing the boundary available potential energy, is analogous to (5.23c) after appropriate scalings of the fluctuations and gradients.

Similarly to the derivation of (5.23) it may also be shown that

$$\mathscr{S}_{y_i} \approx -(-1)^i f \int_{\tilde{D},y_i} \tilde{\sigma} \frac{(Y')^2}{2} \mathrm{d}Z\,\mathrm{d}X = -(-1)^i f \int_{z_1}^{z_2} \frac{\overline{(y')^2}}{2\tilde{y}_Y} \mathrm{d}\tilde{z}\Big|_{y=y_i}, \tag{5.25}$$

whose form is identical within the f-factor to the meridional boundary contributions in (3.10). In deriving (5.25) we have used the relation $y'|_Y = -\tilde{y}_Y\, Y'|_y$, requiring that $\tilde{y}_Y$ satisfy the first of conditions (5.6). As noted in §3, the contributions from the lateral boundary have no counterpart in QG dynamics, in which the standard scaling sets the normal component of the geostrophic velocity to zero on the boundary.

From (5.15), (5.23) and (5.25), we conclude that within a factor of f, $\mathscr{S}$ reduces at small amplitude to the quantity in braces in (3.10).

6. Nonlinear stability

The stability analysis presented here follows the approach of S89. In §5.3 we have shown that the sign of the small-amplitude form of the pseudomomentum (5.13) depends entirely on the signs of the meridional gradients of basic-state quantities. By conservation of the pseudomomentum these sign-dependences define a linearly stable basic state. It turns out that it is possible to define a positive-definite measure of the disturbance that is bounded for finite-amplitude disturbances to such basic states. Unlike the QG case of S89, however, the nonlinear stability theorem presented here only applies to a restricted class of perturbations, and the disturbance quantity is not, in general, a norm.

S89's approach is to bound a disturbance enstrophy-like norm from above and below by the QG pseudomomentum. Then using pseudomomentum conservation, the norm may be bounded in terms of its initial value, which is a statement of Liapunov stability. The method is first illustrated here for the simple case of a disturbance to a zonal basic state under the restriction that the disturbance vanish on the boundaries. Alternatively, the domain might be periodic or unbounded in x and y, and bounded vertically by isentropic surfaces. The boundary contributions represented by the integration over D' in (5.13) then vanish. For small-amplitude stability $\tilde{\sigma}_Y$ must be sign-definite in the interior, and for definiteness we choose it to be negative with a view to the meteorological case. We suppose further that there exist functions $K_\sigma(Z)$ and $k_\sigma(Z)$ such that in the interior

$$0 < k_\sigma(Z) \leqslant -\tilde{\sigma}_Y \leqslant K_\sigma(Z) < \infty. \tag{6.1}$$

At every point (X, Y, Z) in the interior

$$\frac{f(\sigma')^2}{2K_\sigma} \leqslant f \int_0^{\sigma'} \min\left(-\frac{\partial \tilde{Y}_{int}}{\partial \hat{\sigma}}\right) \hat{\sigma}\,\mathrm{d}\hat{\sigma} \leqslant (-S_{int}) \leqslant \frac{f(\sigma')^2}{2k_\sigma}, \tag{6.2}$$

where S_{int} is the interior wave-activity density, defined as the quantity within square brackets in (5.15).

In this case we may define a disturbance norm $\|\sigma'\|$ by

$$\|\sigma'\|^2 = \int_D \frac{f(\sigma')^2}{2K_\sigma} \mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z, \tag{6.3}$$

and we find
$$\|\sigma'\|^2(t) \leqslant -\mathscr{S}(t) = -\mathscr{S}(0) \leqslant M\,\|\sigma'\|^2(0), \tag{6.4}$$

where
$$M = \max_z (K_\sigma/k_\sigma). \tag{6.5}$$

We have used pseudomomentum conservation in (6.4). The maximum in (6.5) is taken over all Z-levels in the interior.

The inequality (6.4) is a statement of Liapunov stability similar in form to the QG case of S89, without the z-boundary contributions. A potential tightening of the bound relative to the approach of that study has been realized by considering the $(-\tilde{\sigma}_Y)_{\max}/(-\tilde{\sigma}_Y)_{\min}$ ratio at each level and seeking its maximum, rather than taking the ratio of the global maximum and minimum values. This modified approach follows Held's suggested improvement to the saturation bound of the two-layer model in Shepherd (1988) (see §5 of that paper).

We now proceed to construct a Liapunov stability theorem explicitly including boundary contributions. In order to do so we have found it necessary to consider a restricted class of perturbations, namely those that vanish along the edges of the domain. The edges are the four lines defined by $\{(x, y_1, z_1)\colon x \in [x_1, x_2]\}$, etc. Restricting the perturbations in this way makes the expressions (5.17) and (5.18) exact at finite amplitude: we may think of the displacement of the surfaces in IGC as that of a membrane fixed at its boundaries, like a drum. The restriction is artificial but appears necessary to obtain the nonlinear stability theorem.

A second problem emerges when considering the potentially destabilizing role of the meridional boundaries. From §3, remark (iii), we recall that $\mathscr{S}_{y_1}$ and $\mathscr{S}_{y_2}$ will have opposite sign, at least at small amplitude. The point is also clearly illustrated by (5.25), where we see that for $\tilde{\sigma} > 0$ at the boundaries, the northern contribution must be negative and the southern one positive. A basic state that satisfies the Charney–Stern stability criteria, then, cannot be constructed if variability is allowed at all bounding surfaces. Thus, besides the edge restrictions, we further take the perturbations to vanish at lateral boundaries, in which case $\mathscr{S}_{y_i} \equiv 0$. We expect the bounds to apply to regimes where the disturbance region is located well away from lateral boundaries. It is important to remember that the necessity of considering these restrictions for the nonlinear theory does not invalidate the linear stability theorem of §3.

Without a detailed knowledge of the particular basic state to be considered it is difficult to determine the tightest bound on the complex expression (5.18); thus we just provide a general bound that might be tightened for a specific basic state. We suppose that the basic state obeys (6.1) both in the interior and, in addition, at the boundaries. For linear stability, $-\tilde{z}_Y$ at the lower surface and $\tilde{z}_Y$ at the upper surface were required to be positive. For nonlinear stability, we further require that the quantities

$$-(-1)^i\, Y_Z|_{z_i} = (-1)^i (z_Z/z_Y)|_{z_i}$$

and $\tilde{\sigma}$ are positive and bounded away from zero and infinity at the boundaries.

We bound the surface wave-activity densities S_{z_i} (defined as the double integrals within braces in (5.18)) as follows. For example, the lower surface contribution obeys

$$-S_{z_1} \leqslant f \int_{\tilde{Z}}^{\tilde{Z}+Z'} \left[\max_\sigma \left(-\sigma \frac{\partial \tilde{Y}_{int}}{\partial \sigma}\right)\right] (\tilde{\sigma} - \tilde{\sigma}_s)\,\mathrm{d}Z \leqslant \frac{f}{k_{z_1}} \frac{(Z')^2}{2}, \tag{6.6}$$

where k_{z_i} is defined generally by

$$\frac{1}{k_{z_i}(X,Y)} = \left[\max_{Z,\sigma}\left(-\sigma\frac{\partial\tilde{Y}_{int}}{\partial\sigma}\right)\right]\left[\max_{Z}(-\tilde{\sigma}_Y)\right]\left[\max_{Z}[-(-1)^i(Y_Z|_{z_i})]\right]. \tag{6.7}$$

The maximum in σ in (6.7) is taken for $\sigma\in[\tilde{\sigma}_s,\tilde{\sigma}]$, and that in Z taken for $Z\in[\tilde{Z},\tilde{Z}+Z']$. We may also define the positive functions

$$\frac{1}{K_{z_i}(X,Y)} = \left[\min_{Z,\sigma}\left(-\sigma\frac{\partial\tilde{Y}_{int}}{\partial\sigma}\right)\right]\left[\min_{Z}(-\tilde{\sigma}_Y)\right]\left[\min_{Z}[-(-1)^i(Y_Z|_{z_i})]\right]. \tag{6.8}$$

Therefore

$$\frac{f(Z')^2}{2K_{z_i}} \leqslant -S_{z_i} \leqslant \frac{f(Z')^2}{2k_{z_i}}. \tag{6.9}$$

Following the same approach as the derivation of (6.4), we define a positive-definite measure of disturbance amplitude by

$$\|\sigma'\|^2 = \int_D \frac{f(\sigma')^2}{2K_\sigma}\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z + \sum_{i=1}^{2}\int_{\tilde{D},z_i}\frac{f(Z')^2}{2K_{z_i}}\mathrm{d}X\,\mathrm{d}Y. \tag{6.10}$$

Because the limits of integration of the first integral of (6.10) depend on Z', $\|\sigma'\|$ does not satisfy the homogeneity property $\|\lambda\sigma'\| = |\lambda|\,\|\sigma'\|$, where λ is a constant and σ' represents the vector of disturbance quantities $(\sigma', Z'|_{z_i})$. Thus (6.10) does not qualify as a norm. Nonetheless, the expression may be bounded from above and below in the same way as (6.3) and may be shown to obey

$$\|\sigma'\|^2(t) \leqslant M\|\sigma'\|^2(0), \tag{6.11}$$

with

$$M = \max\left[\max_{Z}\left(\frac{K_\sigma}{k_\sigma}\right), \max_{z_1}\left(\frac{K_{z_1}}{k_{z_1}}\right), \max_{z_2}\left(\frac{K_{z_2}}{k_{z_2}}\right)\right], \tag{6.12}$$

where the maximum over Z is taken for all interior values of Z, and that over each z-surface taken for all (X, Y) there. This proves nonlinear stability.

The bound (6.11) with (6.12) is analogous to the QG form found in S89, which also consisted of interior and vertical boundary terms. The differences lie in the restrictions on the perturbations, and in the comparatively complex form of the quantities K_{z_i} and k_{z_i}.

7. Local conservation law

We now develop a local wave-activity conservation law of the form

$$\frac{\partial S_{int}}{\partial T} + \nabla\cdot\boldsymbol{F} = 0 \tag{7.1}$$

for the pseudomomentum density. Since (7.1) is a local relation, we can analyse the interior equation alone. Taking (7.1) as a definition of $\boldsymbol{F}$ does not specify it completely: the field is only defined to within a non-divergent flux. To further specify the field's form, we require that it satisfy the group-velocity property that in the WKB limit of a small-amplitude monochromatic wave packet propagating through a slowly varying medium,

$$\langle\boldsymbol{F}\rangle = c_g\langle S_{int}\rangle, \tag{7.2}$$

where c_g is the group velocity and the angular brackets denote a phase average.

Consider the form of the interior pseudomomentum density S_{int}, the quantity within square brackets in (5.15), whose spatial integral comprises the interior contribution to the pseudomomentum (5.13). We note that from σ-conservation (2.18*a*)

$$\frac{\mathrm{D}\tilde{\sigma}}{\mathrm{D}t} = -\frac{\mathrm{D}\sigma'}{\mathrm{D}t} = \frac{\Psi'_X}{f}\tilde{\sigma}_Y. \tag{7.3}$$

Now
$$\left.\begin{aligned} \frac{\partial S_{int}}{\partial \sigma'} &= f[\tilde{Y}_{int}(\tilde{\sigma}+\sigma', Z) - \tilde{Y}_{int}(\tilde{\sigma}, Z)], \\ \frac{\partial S_{int}}{\partial \tilde{\sigma}} &= f[\tilde{Y}_{int}(\tilde{\sigma}+\sigma', Z) - \tilde{Y}_{int}(\tilde{\sigma}, Z)] - f\frac{\partial \tilde{Y}_{int}}{\partial \tilde{\sigma}}\sigma', \end{aligned}\right\} \tag{7.4}$$

whence
$$\frac{\mathrm{D}S_{int}}{\mathrm{D}t} = \left(\frac{\partial S_{int}}{\partial \tilde{\sigma}} - \frac{\partial S_{int}}{\partial \sigma'}\right)\frac{\mathrm{D}\tilde{\sigma}}{\mathrm{D}t} = -\Psi'_X\sigma' = -fv'_g\sigma'. \tag{7.5}$$

Note that to within a factor of f, the zonal average of (7.5) reduces to (3.5) in the small-amplitude limit, with the small-amplitude form of the density S_{int} given in the IGC integral in (5.15). To obtain a finite-amplitude flux law, we must express $fv'_g\sigma'$ as the divergence of a flux. This is done in Appendix A, §A.1, yielding a wave-activity conservation law of the form (7.1). In QG theory, the invertibility principle is linear, implying that the meridional PV flux has only quadratic terms at finite amplitude. From Appendix A, §A.1 we see that since the SG invertibility principle is cubic-nonlinear in Ψ, the finite-amplitude flux has terms of cubic and quartic order. The small-amplitude reduction of the flux is, from (A 3),

$$\boldsymbol{F} = \begin{bmatrix} \tilde{u}_g S_{int} - \frac{1}{2}\tilde{\sigma}(v'_g)^2 + \frac{1}{2}f^2[\tilde{z}_Z(y')^2 - 2\tilde{z}_Y y'z' + \tilde{y}_Y(z')^2] \\ f(\tilde{z}_Z v'_g y' - \tilde{y}_Z v'_g z') \\ f(-\tilde{z}_Y v'_g y' + \tilde{y}_Y v'_g z') \end{bmatrix}. \tag{7.6}$$

The zonal average of $\nabla \cdot \boldsymbol{F}$ reduces to f times (3.8).

To explore the wave properties of the system, and to verify the group-velocity property, we introduce the WKB ansatz

$$\Psi' = \mathrm{Re}\,[\Psi^*(\mu X, \mu Y, \mu Z)\,\mathrm{e}^{\mathrm{i}(\boldsymbol{K}\cdot\boldsymbol{X} - \Omega T)}], \tag{7.7}$$

where
$$\mu = \frac{\text{perturbation lengthscale}}{\text{wave envelope lengthscale}} \ll 1,$$

the dimensional wave vector $\boldsymbol{K} = (K, L, M)$, and $\Omega(\boldsymbol{K}, Y, Z)$ is the frequency. Substituting (7.7) into (7.6) and keeping the leading-order contribution in μ yields

$$\langle \boldsymbol{F} \rangle = \begin{bmatrix} \tilde{u}_g\langle S_{int}\rangle \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2}\tilde{\sigma}K^2 + \frac{1}{2}[\tilde{z}_Z L^2 - 2\tilde{z}_Y LM + \tilde{y}_Y M^2] \\ -(\tilde{z}_Z KL - \tilde{y}_Z KM) \\ -(-\tilde{z}_Y KL + \tilde{y}_Y KM) \end{bmatrix} \frac{|\Psi^*|^2}{2f^2}. \tag{7.8}$$

To verify the group-velocity property we must use (3.6), the linearized invertibility relation for small-amplitude disturbances to the basic state. From (3.6), in matrix form

$$\sigma' = h_{ij}\partial_{X_i}\partial_{X_j}\Psi', \tag{7.9}$$

where

$$h_{ij} = -\frac{1}{f^2}\begin{pmatrix} \tilde{\sigma} & 0 & 0 \\ 0 & \tilde{z}_Z & -\tilde{y}_Z \\ 0 & -\tilde{z}_Y & \tilde{y}_Y \end{pmatrix} \tag{7.10}$$

is a symmetric matrix from the thermal wind relation. In order for (7.9) to be invertible, we require h_{ij} to be sign definite. We find

$$-h_{11} = \frac{\tilde{\sigma}}{f^2}, \quad \det\begin{pmatrix} -h_{11} & -h_{12} \\ -h_{21} & -h_{22} \end{pmatrix} = \frac{\tilde{z}_Z \tilde{\sigma}}{f^4}, \quad \det(-h_{ij}) = \frac{(\tilde{\sigma})^2}{f^6}, \tag{7.11}$$

implying that h_{ij} is negative definite for a statically stable basic state ($\tilde{z}_Z > 0$) with positive PV. Substituting the WKB form (7.7) into (7.9) yields, to lowest order in μ,

$$\sigma' = -h_{ij} K_i K_j \Psi'. \tag{7.12}$$

The dispersion relation obtains upon substituting the solution into the linearized governing equation and taking the terms of lowest order in μ. We find

$$\Omega = \tilde{u}_g K + \frac{(\tilde{\sigma}_Y/f) K}{(-h_{ij} K_i K_j)}, \tag{7.13}$$

where the minus sign accompanies the denominator to keep the quadratic form positive. This is clearly analogous to the QG dispersion relation, which under the WKB approximation is

$$\omega = \tilde{u}_g k - \frac{\tilde{q}_y k}{k^2 + l^2 + (f/N)^2 m^2}, \tag{7.14}$$

where ω is the frequency, (k, l, m) the wave vector in physical space, and $\tilde{q}_y$ is the meridional QG PV gradient. The group velocity is

$$c_{g,l} = \partial_{K_l} \Omega = \left(\tilde{u}_g + \frac{\tilde{\sigma}_Y/f}{(-h_{ij} K_i K_j)}\right)\delta_{l1} + \frac{2(\tilde{\sigma}_Y/f) K}{(-h_{ij} K_i K_j)^2} h_{lm} K_m, \tag{7.15}$$

where we have used the symmetry of h_{ij}. The phase-averaged linearized pseudo-momentum is, from (5.15) and (7.12),

$$\langle S_{int} \rangle = \frac{(-h_{ij} K_i K_j)^2}{4\tilde{\sigma}_Y/f} |\Psi^*|^2, \tag{7.16}$$

and the group-velocity property (7.2) follows directly from (7.15) and (7.16) if we rewrite (7.8) as

$$\langle F_{(l)} \rangle = (\tilde{u}_g \langle S_{int} \rangle - \tfrac{1}{4} h_{ij} K_i K_j |\Psi^*|^2)\delta_{l1} + \tfrac{1}{2} K h_{lj} K_j |\Psi^*|^2. \tag{7.17}$$

8. Wave–mean-flow interaction theory

Using the results of §7 we develop diagnostic equations to determine the effects of departures from the zonal flow on the zonal flow itself. We must first distinguish the time-dependent zonal-mean state $\bar{\sigma}$, from the steady zonal basic state $\tilde{\sigma}$, with respect to which the disturbance σ' is defined (see (5.12)). We may divide σ' and the other disturbance fields into zonal and eddy parts as follows:

$$\sigma' = \overline{\sigma'}(Y, Z, T) + \sigma^*(X, Y, Z, T), \quad \text{such that } \overline{\sigma^*} = 0. \tag{8.1}$$

For small-amplitude disturbances, both $\overline{\sigma'}$ and σ^* are small compared to $\bar{\sigma}$. Using this decomposition based on the zonal-mean statistics of the flow we will be able to see clearly how the evolution of the zonal mean is influenced by a given eddy forcing.

We add a term G to the right-hand side of (7.1) to represent processes that generate or dissipate pseudomomentum locally. Then from the interior governing equation (2.18*a*), (7.1) and (7.5), we find that to lowest order in the amplitude of the disturbance

$$\frac{\partial \bar{\sigma}}{\partial T} = \frac{\partial \overline{\sigma'}}{\partial T} = -\frac{\partial}{\partial Y}\overline{v_g^* \sigma^*} = -\frac{\partial}{\partial Y}\frac{\nabla \cdot \bar{F}}{f} = \frac{\partial}{\partial Y}\left(\frac{1}{f}\frac{\partial \overline{S_{int}}}{\partial T} - \frac{\bar{G}}{f}\right). \tag{8.2}$$

We will examine (8.2) in some detail. The equation expresses the way in which eddies force the zonal mean σ (inverse PV) tendency. The components of the generalized E–P flux $\bar{F}$ are, from (7.6),

$$\bar{F} = \begin{pmatrix} f\tilde{z}_Z & -f\tilde{y}_Z \\ -f\tilde{z}_Y & f\tilde{y}_Y \end{pmatrix}\begin{pmatrix} \overline{v_g^* y^*} \\ \overline{v_g^* z^*} \end{pmatrix}. \tag{8.3}$$

To obtain the QG limit, one may use the invertibility relations (2.20) to find

$$\tilde{y}_Y \approx 1, \quad \tilde{z}_Z \approx \frac{f^2}{N^2}, \quad y' \approx \frac{1}{f}u_g', \quad z' \approx -\tilde{z}_Z Z' \approx -\frac{g\theta'}{N^2\theta_0},$$

whence

$$\lim_{\epsilon \to 0} \bar{F} = \frac{f^2}{N^2}\begin{pmatrix} 1 & 0 \\ 0 & -g/(f\theta_0) \end{pmatrix}\begin{pmatrix} \overline{v_g^* u_g^*|_z} \\ \overline{v_g^* \theta^*|_z} \end{pmatrix}, \tag{8.4}$$

where we have taken $u_g'|_Z \approx u_g'|_z$. The form (8.4) is a scaled version of the E–P flux from Edmon *et al.* (1980), with an overall change of sign due to their definition of the linearized wave activity, which is negative relative to (5.16).

Comparison of (8.3) and (8.4) suggests that the off-diagonal thermal-wind terms in the matrix play an important role in Rossby-wave propagation for large Rossby number. This can be seen from the meridional and vertical components of the group velocity (7.15),

$$\begin{pmatrix} c_{g(Y)} \\ c_{g(Z)} \end{pmatrix} = \frac{-2f^2(\tilde{\sigma}_Y/f)K}{(\tilde{\sigma}K^2 + \tilde{z}_Z L^2 - 2\tilde{y}_Z LM + \tilde{y}_Y M^2)^2}\begin{pmatrix} \tilde{z}_Z & -\tilde{y}_Z \\ -\tilde{z}_Y & \tilde{y}_Y \end{pmatrix}\begin{pmatrix} L \\ M \end{pmatrix}, \tag{8.5}$$

where the thermal-wind terms affect the relative magnitudes of the horizontal and vertical components. Since both the basic-state and perturbation quantities depend on the coordinate transformation, however, the extent to which SG dynamics modifies the QG-dynamics-based picture of Rossby-wave propagation is not yet clear. This issue is currently being examined.

The relation (8.2) provides additional diagnostic insight when it is expressed as a wave–mean-flow *closure*. The idea, as in QG theory (see e.g. Andrews, Holton & Leovy 1987), is to determine the balanced zonal-mean response to a given eddy forcing. We will see, however, that the QG and SG forms of the zonal-mean balance differ significantly. The source of this difference lies in the nonlinear invertibility relation of SG theory, and can be seen by considering the zonal mean invertibility problem. In QG dynamics, $q = \mathscr{L}_{QG}(\Psi)$, where $\mathscr{L}_{QG}$ is a linear operator and Ψ may be obtained from q subject to linear boundary conditions. The linearity of the system implies that the zonal mean disturbance, $\overline{\Psi'}$, or the zonal-mean tendency $\overline{\Psi}_T = \overline{\Psi}_T$ may be obtained given $\overline{q'}$ or $\overline{q'}_T$. The SG invertibility relation (2.20), on the other hand, may be represented symbolically by $\sigma = \mathscr{N}(\Psi, \Psi, \Psi)$. Since the system is invertible by assumption, Ψ may be obtained from σ, but a closure problem emerges when zonal-

mean quantities are desired, because $\overline{\sigma'} = \mathcal{N}(\overline{\Psi'}, \overline{\Psi^*\Psi^*}, \overline{\Psi^*\Psi^*\Psi^*}, \ldots)$, i.e. correlations between eddy quantities come into play in the balance. Thus more than the $\overline{\sigma'}$ or $\overline{\sigma'_T}$ distribution is needed to determine the corresponding $\overline{\Psi'}$ or $\overline{\Psi'_T}$.

We find that the second-order eddy correlations are significant even in the linear theory. Equation (8.2) shows that the zonal-mean PV tendency $\overline{\sigma'_T}$ is second order in disturbance amplitude a. In other words, the dynamics constrain the zonal-mean PV to evolve on a slow timescale. Thus, if $\overline{\sigma'} = 0$ initially, we have $\overline{\sigma'}/\tilde{\sigma} = O(a^2)$. From (2.20) we find to lowest order in disturbance amplitude that

$$\overline{\sigma'} = \mathscr{L}(\overline{\Psi'}) + \nabla \cdot \bar{J}, \tag{8.6}$$

where the linear operator $\mathscr{L}$ has the form

$$\mathscr{L} = -f^{-2}[\tilde{z}_Z(\cdot)_{YY} - 2\tilde{y}_Z(\cdot)_{YZ} + \tilde{y}_Y(\cdot)_{ZZ}], \tag{8.7}$$

and the divergence term consists of second-order eddy correlations:

$$\nabla \cdot \bar{J} = \overline{\partial_{XYZ}(x^*, y^*, \tilde{z})} + \overline{\partial_{YZ}(y^*, z^*)} + \overline{\partial_{XYZ}(x^*, \tilde{y}, z^*)}. \tag{8.8}$$

In (8.6), we have neglected higher-order terms such as $\partial_{YZ}(\overline{y'}, \overline{z'})$ and $\overline{\partial_{XYZ}(x^*, y^*, z^*)}$.

The zonal-mean stream-function tendency, $\overline{\Psi'_T}$, can be obtained from the following:

$$\mathscr{L}(\overline{\Psi'_T}) = -\frac{\partial}{\partial T}\nabla \cdot \bar{J} - \frac{\partial}{\partial Y}\overline{v_g^* \sigma^*} \tag{8.9a}$$

$$= -\nabla \cdot \left[\frac{\partial \bar{J}}{\partial T} + \frac{\partial}{\partial Y}\left(\frac{\bar{F}}{f}\right)\right] = \frac{\partial}{\partial T}\left[-\nabla \cdot \bar{J} + \frac{\partial}{\partial Y}\left(\frac{\overline{S_{int}}}{f}\right)\right] - \frac{\partial}{\partial Y}\left(\frac{\bar{G}}{f}\right), \tag{8.9b}$$

where we have used (8.2) and (8.6). The boundary conditions on $\overline{\Psi'_T}$ are, from (2.18c, d), (2.20), and (3.3) and (3.4),

$$\left.\begin{aligned} \overline{\Psi'_{YT}} &= -f^2\overline{y'_T} = f^2\frac{\partial}{\partial Y}\overline{v_g^* y^*} = -\frac{\partial}{\partial Y}\frac{\partial}{\partial T}\left[\frac{f^2\overline{(y')^2}}{2\tilde{y}_Y}\right] \quad \text{on} \quad y = y_1, y_2, \\ \overline{\Psi'_{ZT}} &= -f^2\overline{z'_T} = f^2\frac{\partial}{\partial Y}\overline{v_g^* z^*} = -\frac{\partial}{\partial Y}\frac{\partial}{\partial T}\left[\frac{f^2\overline{(z')^2}}{2\tilde{z}_Y}\right] \quad \text{on} \quad z = z_1, z_2. \end{aligned}\right\} \tag{8.10}$$

The quantities in square brackets on the right-hand sides of (8.10) are the scaled small-amplitude surface pseudomomentum densities in (3.10) (see also (5.23) and (5.25)).

The operator $\mathscr{L}$ is elliptic for a statically stable ($\tilde{z}_Z > 0$) basic state with $\tilde{\sigma} > 0$. The ellipticity is used in Appendix B to show that the solution $\overline{\Psi'_T}(Y, Z, T)$ to (8.9) with boundary conditions (8.10), for given interior PV flux $f\overline{v_g^* \sigma^*} = \nabla \cdot \bar{F}$, eddy correlations $(\nabla \cdot \bar{J})_T$, and boundary fluxes $(\overline{v_g^* y^*}, \overline{v_g^* z^*})$, is unique to within a spatially constant function of time. Since the Y-derivative of such a constant vanishes, and because this solution is unique, we obtain as a consequence the following *non-acceleration theorem*: in the absence of transience (namely $(\nabla \cdot \bar{J})_T = [\overline{S_{int}}]_T = [\overline{S}_{y_i}]_T = [\overline{S}_{z_i}]_T = 0$), and of pseudomomentum generation and dissipation (namely $\bar{G} = 0$), the zonally averaged flow does not accelerate, i.e. $\overline{u_{g_T}} = \overline{u'_{g_T}} = 0$ throughout the domain.

We may cast (8.9) into a form that allows us to obtain the zonal acceleration directly from the inversion. If an assumption is made that the basic-state meridional lengthscales are much larger than those of the disturbance, then $-(f^{-1}\mathscr{L}[\overline{\Psi'_T}])_Y \approx \mathscr{L}(\overline{u'_{g_T}})$, and we obtain

$$\mathscr{L}(\overline{u'_{g_T}}) = \frac{\partial^2}{\partial Y \partial T}\left(\frac{\nabla \cdot \bar{J}}{f}\right) + \frac{\partial^2}{\partial Y^2}\frac{\nabla \cdot \bar{F}}{f^2} = \frac{\partial^2}{\partial Y \partial T}\left(\frac{\nabla \cdot \bar{J}}{f}\right) - \frac{\partial^2}{\partial Y^2}\left(\frac{1}{f^2}\frac{\partial \overline{S_{int}}}{\partial T} - \frac{\bar{G}}{f^2}\right). \tag{8.11}$$

Equation (8.11) closely resembles the wave–mean-flow balance of QG theory. In the QG limit, using scaling arguments similar to those used to derive (8.4), (8.11) becomes

$$\left[\frac{\partial^2}{\partial y^2}+\frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial}{\partial z}\right)\right]\overline{u'_{g_T}}=\frac{\partial^2}{\partial y^2}\left[-\frac{\partial\overline{v_g^* u_g^*}}{\partial y}+\frac{\partial}{\partial z}\left(\frac{f\overline{v_g^*\theta^*}}{\tilde{\theta}_z}\right)\right]. \tag{8.12}$$

Equation (8.12) matches the Boussinesq inviscid limit of equation (3.5.7) of Andrews *et al.* (1987), which expresses the QG zonal-mean response to eddy forcing in log-pressure coordinates. In deriving (8.12), we have neglected the term involving $(\nabla\cdot\bar{\boldsymbol{J}})_T$ as being $O(\epsilon)$ compared to the right-hand side of the equation.

In conclusion, we have shown that while PV fluxes suffice to determine the zonal-mean PV tendency (equation (8.2)), the nonlinearity of the invertibility relation implies that the zonal-mean stream-function tendency can only be obtained if the flux divergence tendency $\nabla\cdot\bar{\boldsymbol{J}}_T$ is also specified. It is not obvious what net effect the eddy fluxes $\bar{\boldsymbol{J}}$ have on the mean flow. As a limiting case, we find for WKB conditions (7.7) that they have no effect, since $\langle(\nabla\cdot\bar{\boldsymbol{J}})_T\rangle=0$.

9. β-plane compressible flow

9.1. *Introduction*

In this section, we extend the results of §§2–8 to the β-plane compressible system of MSc90. Like the Hoskins (1975) system, the MSc90 model has the same two mathematical features that made it possible to find the pseudomomentum in §5: first, its dynamics are also governed by PV advection, with appropriate boundary conditions, and an invertibility principle relating the PV and boundary terms to the velocity; and secondly, it is a Hamiltonian system. The explicit form of the Hamiltonian structure is presented in the sequel to this paper (Kushner & Shepherd 1995).

Given the similarity in structure of the two systems, many of the derivations of the results of this section closely parallel those of the previous sections. We will thus omit the details of the developments below, referring the reader to the earlier sections.

9.2. *Semi-geostrophic dynamics for β-plane compressible flow*

MSc90 generalize the hydrostatic compressible equations to the β-plane as follows:

$$\left.\begin{aligned}\mathrm{D}u_g/\mathrm{D}t-[f(Y)v+\beta\Delta y v_g]&=-(1/\rho)p_x,\\ \mathrm{D}v_g/\mathrm{D}t+[f(Y)u+\beta\Delta y u_g]&=-(1/\rho)p_y,\end{aligned}\right\} \tag{9.1}$$

where $\Delta y=y-Y$, $f=f(Y)$ is the variable Coriolis parameter, and $\beta=\mathrm{d}f/\mathrm{d}Y$ is constant. The coordinates are (x,y,z), where in contrast to the previous sections z refers to geometric height, not pseudoheight. Equations (9.1) anticipate the transformation to IGC by explicit inclusion of the coordinate Y in the variable f and in the Δy-terms. The terms in square brackets in (9.1) represent the SG approximation to the Coriolis terms $f\hat{\boldsymbol{z}}\times\boldsymbol{u}$ in the horizontal momentum equations. These expressions do not represent the β-plane approximation *per se*, since while β is constant f still varies in Y; nor is Δy a deviation from some reference latitude. Rather (9.1) to order Rossby number ϵ accuracy expresses the behaviour of the geostrophic coordinate transformation on the β-plane. MSc90 arrive at the equations by requiring that the horizontal advection in IGC space have the canonical form found in the first two components of (9.10) below. The 'primitive' momentum equations then follow on transformation back to physical space. Equations (9.1) are part of a family of results (e.g. Magnusdottir & Schubert 1991; Lu 1993) based on a similar approach of taking

the advection to be simplified in the transformed coordinate space. Craig (1993) has shown that this form of the momentum equations also follows from a systematic asymptotic expansion of the primitive equations in a modified Rossby number.

The material derivative is (2.3), with velocity components given by (2.4), where here and henceforth z is taken to be the geometric height. The geostrophic velocity is now

$$\boldsymbol{u}_g = (u_g, v_g) = \left(-\frac{p_y}{f(Y)\rho}, \frac{p_x}{f(Y)\rho} \right), \tag{9.2}$$

where ρ is the density, the continuity equation

$$\rho_t + \boldsymbol{\nabla}_h \cdot (\rho \boldsymbol{u}) + \partial_z(\rho w) = 0, \tag{9.3}$$

the thermodynamic equation

$$\frac{\mathrm{D}\theta}{\mathrm{D}t} = \frac{\theta}{c_p T}\left[c_p \frac{\mathrm{D}T}{\mathrm{D}t} - \frac{1}{\rho}\frac{\mathrm{D}p}{\mathrm{D}t} \right] = 0, \tag{9.4}$$

and hydrostatic balance requires

$$\partial p/\partial z = -g\rho. \tag{9.5}$$

We transform the system to IGC

$$(X, Y, \Theta, T) = \left(x + \frac{v_g}{f}, y - \frac{u_g}{f}, \theta, t \right), \tag{9.6}$$

where henceforth $f = f(Y)$, and because f varies we choose not to rescale θ as in (2.9). The Montgomery–Bernoulli potential (M^* in MSc90) in IGC takes the form

$$\Psi = \phi + c_p T + \tfrac{1}{2}(u_g^2 + v_g^2) = \phi + \theta\Pi + \tfrac{1}{2}(u_g^2 + v_g^2), \tag{9.7}$$

with $\Pi = c_p(p/p_0)^\kappa$ the Exner function, and $\phi = gz$. Equations (2.10) and (9.7) differ only in their internal energy terms, which turn out to be closely related. Using (2.1), (2.9), and the ideal gas equation of state, it may be shown that $c_p T = -f^2 Z(z_p - z_a)$, where z_p is the pseudoheight, $z_a = \mathrm{H_s}/\kappa \approx 28$ km the maximum of z_p for standard atmospheric values (Hoskins & Bretherton 1972), and f is the constant reference value in (2.9). The sum of the first two terms in (9.7) is once again the Montgomery potential, and that of the first and third terms the geostrophic-coordinate potential.

The partial derivatives transform as

$$\begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \\ \partial_t \end{pmatrix} = \begin{pmatrix} X_x & Y_x & \Theta_x & 0 \\ X_y & Y_y & \Theta_y & 0 \\ X_z & Y_z & \Theta_z & 0 \\ X_t & Y_t & \Theta_t & 1 \end{pmatrix} \begin{pmatrix} \partial_X \\ \partial_Y \\ \partial_\Theta \\ \partial_T \end{pmatrix}. \tag{9.8}$$

From (9.2) and (9.5)–(9.8) we find (MSc90)

$$\Psi_X = p_x/\rho = f^2(X - x), \tag{9.9a}$$

$$\Psi_Y = \frac{p_y}{\rho} + \frac{\beta}{f\rho}[p_y(Y-y) + p_x(X-x)] = f^2(Y-y) + f\beta[(Y-y)^2 + (X-x)^2], \tag{9.9b}$$

$$\Psi_T = p_t/\rho, \quad \Psi_\Theta = \Pi, \tag{9.9c,d}$$

with $f = f(Y)$. In IGC, the material derivative is

$$\frac{\mathrm{D}}{\mathrm{D}t} = \partial_T + \frac{\mathrm{D}X}{\mathrm{D}t}\partial_X + \frac{\mathrm{D}Y}{\mathrm{D}t}\partial_Y + \frac{\mathrm{D}\Theta}{\mathrm{D}t}\partial_\Theta,$$

and using the momentum equations (9.1), the thermodynamic equation (9.4), (9.6) and (9.9), we find

$$\left(\frac{\mathrm{D}X}{\mathrm{D}t}, \frac{\mathrm{D}Y}{\mathrm{D}t}, \frac{\mathrm{D}\Theta}{\mathrm{D}t}\right) = \left(-\frac{\Psi_Y}{f}, \frac{\Psi_X}{f}, 0\right) = [u_g - \beta f^{-2}(u_g^2 + v_g^2), v_g, 0]. \tag{9.10}$$

The material derivative in IGC is then

$$\frac{\mathrm{D}}{\mathrm{D}t} = \partial_T + \frac{1}{f}\partial_{XY}(\Psi, \cdot). \tag{9.11}$$

Following MSc90 we introduce the potential pseudodensity $\sigma = f/q$ (σ^*/g in their notation), where q is the dimensional potential vorticity. The PV is conserved following the motion, so that

$$\mathrm{D}(\sigma/f)/\mathrm{D}t = 0.$$

The discussion of the boundary conditions in §2 applies here as well. Material boundaries obey the 'no normal flow' condition, equations (2.15) and (2.16). Note that because z is now a geometric coordinate, the second of these equations applies at geopotential rather than isobaric surfaces.

From MSc90, equation (4.1), we find

$$\sigma = -\partial_{XY\Theta}(x, y, p)/g. \tag{9.12}$$

To obtain an explicit invertibility principle, and to develop expressions for the local pseudomomentum flux below, it is necessary to relate the physical coordinates (x, y) and the pressure p to the potential Ψ. The zonal coordinate x is easily obtained from (9.9*a*). For the meridional coordinate, the first component of (9.10) may be written as an equation quadratic in u_g,

$$\Psi_Y = -fu_g + (\beta/f)(u_g^2 + v_g^2). \tag{9.13}$$

We use a scaling consistency argument to determine the appropriate root to solve for $u_g = f(y - Y)$. The quadratic terms in (9.13) are small relative to the linear term, for

$$\frac{O[\beta(u_g^2 + v_g^2)/f]}{O(fu_g)} = O(\beta U/f^2) = O(\beta L\epsilon/f) = O(L\epsilon/r_0),$$

where U and L characterize zonal and meridional velocity and length scales, ϵ is the Rossby number, and r_0 the Earth's radius. Thus even for planetary-scale waves we may reasonably take $O(fu_g) = O(\Psi_Y)$, forcing a choice of the smaller root of (9.13), namely

$$\begin{aligned} u_g &= (f^2/2\beta)[1 - (1 - 4\beta f^{-3}(\beta f^{-3}(\Psi_X)^2 - \Psi_Y))^{1/2}] \\ &\approx -\Psi_Y/f + \beta f^{-4}[(\Psi_X)^2 + (\Psi_Y)^2] + O[(L\epsilon/r_0)^2 U], \end{aligned} \tag{9.14}$$

because the larger root yields $O(u_g) = O(f^2/\beta) \sim 10^3$ m s^{-1}. We have thus expressed y in terms of Y, Ψ_X and Ψ_Y, and note that the relation is nonlinear.

For the pressure, from the definition of the Exner function we have

$$p = p_0(\Psi_\Theta/c_p)^{1/\kappa} = p(\Psi_\Theta).$$

To summarize, the governing prognostic equations are

$$\mathrm{D}(\sigma/f)/\mathrm{D}t = 0 \quad \text{in the interior,} \tag{9.15a}$$

$$\mathrm{D}x/\mathrm{D}t = 0 \quad \text{at} \quad x = x_1, x_2, \tag{9.15b}$$

$$\mathrm{D}y/\mathrm{D}t = 0 \quad \text{at} \quad y = y_1, y_2, \tag{9.15c}$$

$$\mathrm{D}z/\mathrm{D}t = 0 \quad \text{at} \quad z = z_1, z_2, \tag{9.15d}$$

with the material derivative given by (9.11). The diagnostic invertibility relations are

$$\begin{bmatrix} x \\ y \\ z \\ p \\ \sigma \end{bmatrix} = \begin{bmatrix} X - \Psi_X/f^2 \\ Y + (f/2\beta)[1-(1-4\beta f^{-3}(\beta f^{-3}(\Psi_X)^2 - \Psi_Y))^{1/2}] \\ (1/g)(\Psi - \frac{1}{2}f^2[(x-X)^2+(y-Y)^2] - \Theta\Psi_\Theta) \\ p_0(\Psi_\Theta/c_p)^{1/\kappa} \\ -(1/g)\partial_{XY\Theta}(x,y,p) \end{bmatrix}. \tag{9.16}$$

We note the additional nonlinearity for y and p in (9.16) as compared to the f-plane case (2.20). MSc90 characterize the nonlinearity as 'weak' in the sense that to good approximation the quadratic terms in, for example, (9.13) are small compared to the linear term, so that $u_g \approx -\Psi_Y/f(Y)$. Under a quasi-Boussinesq approximation, the pressure may be taken to be linear in Ψ_Θ.

Simpler systems may be obtained from (9.15) and (9.16). For modelling on the f-plane, these equations may be used with f constant and $\beta = 0$. For modelling Boussinesq β-plane flow, equations (2.2) may be used with the terms fv and fu replaced by the terms in brackets in (9.1).

9.3. *Linear Charney–Stern theorem*

Using the same direct methods of §3, we find that the linear Charney–Stern theorem takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\int_D \frac{\tilde{\rho}\overline{(\sigma')^2}}{2\tilde{\sigma}(\tilde{\sigma}/f)_Y}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z} - \left[\int_{z_1}^{z_2} f\tilde{\rho}\frac{\overline{(y')^2}}{2\tilde{y}_Y}\,\mathrm{d}\tilde{z}\right]_{y_1}^{y_2} - \left[\int_{y_1}^{y_2} f\tilde{\rho}\frac{\overline{(z')^2}}{2\tilde{z}_Y}\,\mathrm{d}\tilde{y}\right]_{z_1}^{z_2}\right\} = 0, \tag{9.17}$$

or in terms of the meridional geostrophic displacements η',

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\int_D \frac{f^2\tilde{\rho}(\tilde{\sigma}/f)_Y\overline{(\eta')^2}}{2\tilde{\sigma}}\,\mathrm{d}\tilde{y}\,\mathrm{d}\tilde{z} - \left[\int_{z_1}^{z_2} \frac{f\tilde{\rho}\tilde{y}_Y\overline{(\eta')^2}}{2}\,\mathrm{d}\tilde{z}\right]_{y_1}^{y_2} - \left[\int_{y_1}^{y_2} \frac{f\tilde{\rho}\tilde{z}_Y\overline{(\eta')^2}}{2}\,\mathrm{d}\tilde{y}\right]_{z_1}^{z_2}\right\} = 0. \tag{9.18}$$

Here meridional gradients in $\tilde{\sigma}/f = 1/\tilde{q}$ replace those of $\tilde{\sigma}$ in (3.10). Since f is variable, the effects of differential rotation are allowed for. The conservation of the first term alone corresponds to (5.9) of MSc90. We refer to the remarks following (3.12) for a discussion of the implications of the boundary terms.

9.4. *Global invariants*

Considering the same flow domains as in §4, we obtain the energy flux law

$$E_t + \nabla_h\cdot[\boldsymbol{u}(E+p)] + \partial_z[w(E+p)] = 0, \tag{9.19}$$

where

$$E = \rho(\tfrac{1}{2}(u_g^2+v_g^2) + \Theta\Psi_\Theta + \phi) - p \tag{9.20}$$

is the dimensional energy density in physical space. Integrating (9.19) over the domain using the boundary conditions (9.15*b–d*), or periodicity or unboundedness in some direction, we find conservation of $\mathscr{E}$ as given by (4.3).

The zonal impulse obeys

$$M_t + \nabla_h\cdot(\boldsymbol{u}M) + \partial_z(wM) + p_x = 0, \tag{9.21}$$

where

$$M = -\rho F(Y), \tag{9.22}$$

with

$$\mathrm{d}F/\mathrm{d}Y = f(Y). \tag{9.23}$$

For flow periodic in x, we have the corresponding global conservation law (4.6).

For any function $C(\sigma/f, \Theta)$ we have

$$(\rho C)_t + \nabla_h \cdot (\boldsymbol{u}\rho C) + \partial_z(w\rho C) = 0, \tag{9.24}$$

which implies the global conservation law for the Casimirs

$$\frac{\mathrm{d}\mathscr{C}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\int \rho C \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 0. \tag{9.25}$$

For the variational calculations we will again need the IGC form of the fields. The IGC zonal impulse density M^* is

$$M^* = -F(Y)\,\sigma \Rightarrow \mathscr{M} = \int M^* \,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}\Theta, \tag{9.26}$$

where we have used the relation $\rho\partial_{XY\Theta}(x, y, z) = \sigma$, which follows from the fifth component of (9.16) together with hydrostatic balance (9.5). The functionals $\mathscr{C}$ have the form

$$\mathscr{C} = \int f C^*(\sigma/f, \Theta)\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}\Theta, \tag{9.27}$$

where C^* is an arbitrary function of two arguments.

9.5. *Pseudomomentum*

The basic state is given by (5.1), and we will make use of the inverse maps

$$\tilde{F}_{int}(\sigma/f, \Theta) = F(\tilde{Y}_{int}(\sigma/f, \Theta)),$$

where F is defined by (9.23), and the boundary maps $\tilde{F}_s(\Theta) = \tilde{F}(\tilde{Y}_s(\Theta))$ and $[\tilde{\sigma}/f]_s(\Theta)$. We seek $\mathscr{S}$ of the form (5.3), involving the sum of the zonal impulse and a Casimir chosen to satisfy the extremal property (5.4). Boundary variability is treated similarly to the f-plane system. From (9.26) and (9.27), (5.10) is replaced by

$$\left.\begin{aligned} \partial C^*/\partial(\sigma/f)\,|_\Theta = F(Y) \quad &\text{in the interior, basic state,} \\ C^* = F(Y)(\sigma/f) \quad &\text{at all boundaries, basic state,} \end{aligned}\right\} \tag{9.28}$$

with solution

$$C^*(\sigma/f, \Theta) = \int_{[\tilde{\sigma}/f]_s(\Theta)}^{\sigma/f} \tilde{F}_{int}(\hat{\sigma}/f, \Theta)\,\mathrm{d}(\hat{\sigma}/f) + \tilde{F}_s(\Theta)\,[\tilde{\sigma}/f]_s(\Theta). \tag{9.29}$$

Using disturbance quantities as defined in (5.12), the β-plane compressible version of (5.13) is then

$$\begin{aligned} \mathscr{S} = &\int_{\tilde{D}+D'} f\left[\int_0^{\sigma'/f} [\tilde{F}_{int}(\tilde{\sigma}/f + \hat{\sigma}/f, \Theta) - \tilde{F}_{int}(\tilde{\sigma}/f, \Theta)]\,d(\hat{\sigma}/f)\right]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}\Theta \\ &+ \int_{D'} f\left[\int_{[\tilde{\sigma}/f]_s}^{\tilde{\sigma}/f} -(\hat{\sigma}/f)\frac{\partial \tilde{F}_{int}(\hat{\sigma}/f, \Theta)}{\partial(\hat{\sigma}/f)}\mathrm{d}(\hat{\sigma}/f)\right]\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}\Theta \\ \equiv\ &\mathscr{S}_{int} + \mathscr{S}_s. \end{aligned} \tag{9.30}$$

Following the approach of §5.3, the small-amplitude reduction of the pseudomomentum (9.30) yields the quantity within braces in (9.17). Note from (9.23) that for $f > 0$, F is monotonic and increasing in Y, and thus the $\tilde{F}_{int}$ and $\tilde{Y}_{int}$ maps play analogous roles in the stability theory.

9.6. *Stability*

The nonlinear stability theorem has the form (6.11) and (6.12) with the following substitutions:

$$\sigma \to (\sigma/f), \quad Z \to \Theta, \quad \tilde{Y}_{int} \to \tilde{F}_{int},$$

in (6.1), (6.7), (6.8), and (6.12); and with (6.10) replaced by

$$\|\sigma'\|^2 = \int_D \frac{(\sigma')^2}{2K_{(\sigma/f)}} \mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}\Theta + \sum_{i=1}^{2} \int_{\tilde{D},z_i} \frac{f(\Theta')^2}{2K_{z_i}} \mathrm{d}X\,\mathrm{d}Y. \tag{9.31}$$

9.7. *Local conservation law*

From (9.30) the interior pseudomomentum density and its small-amplitude reduction are

$$S_{int} \equiv f \int_0^{\sigma'/f} [\tilde{F}_{int}(\tilde{\sigma}/f + \hat{\sigma}/f, \Theta) - \tilde{F}_{int}(\tilde{\sigma}/f, \Theta)]\,\mathrm{d}(\hat{\sigma}/f) \approx \frac{(\sigma')^2}{2(\tilde{\sigma}/f)_Y}. \tag{9.32}$$

Similarly to the derivation of (7.5), we find

$$\frac{\partial S_{int}}{\partial T} + \frac{\partial}{\partial X}\left(-\frac{\Psi'_Y S_{int}}{f}\right) + \frac{\partial}{\partial Y}\left(\frac{\Psi'_X S_{int}}{f}\right) + \Psi'_X \sigma' = 0. \tag{9.33}$$

The term $\Psi'_X \sigma'$ is expressed as the divergence of a flux in Appendix A, §A.2. To do so it is necessary to expand y- and p-fields in powers of the disturbance field Ψ', because from (9.16) y and p are not integer powers of gradients of Ψ. The small-amplitude reduction of the E–P flux is, from (A 5) and (A 12)

$$F_{(X)} = -\frac{\tilde{\Psi}_Y}{f} S_{int} - \frac{\tilde{\sigma}}{2f^2}(\Psi'_X)^2 + \frac{1}{2g}\left[-\frac{\tilde{p}_\Theta}{f\hat{f}}(\Psi'_Y)^2 + 2\frac{\tilde{p}_Y}{f\hat{f}}\Psi'_Y \Psi'_\Theta + \tilde{\Upsilon}\tilde{y}_Y(\Psi'_\Theta)^2\right], \tag{9.34a}$$

$$F_{(Y)} = (1/g)((\tilde{p}_\Theta/f\hat{f})\,\Psi'_X \Psi'_Y + \tilde{\Upsilon}\tilde{y}_\Theta \Psi'_X \Psi'_\Theta), \tag{9.34b}$$

$$F_{(\Theta)} = (1/g)(-(\tilde{p}_Y/f\hat{f})\,\Psi'_X \Psi'_Y - \tilde{\Upsilon}\tilde{y}_Y \Psi'_X \Psi'_\Theta), \tag{9.34c}$$

where

$$\hat{f} = f(1 + 4\beta\tilde{\Psi}_Y/f^3)^{1/2}, \quad \Upsilon = \mathrm{d}p/\mathrm{d}\Psi_\Theta = \rho\Theta. \tag{9.35}$$

In deriving the above we have used the small-amplitude relation

$$y' = u'_g/f \approx -\Psi'_Y/(f\hat{f}), \tag{9.36}$$

together with thermal wind balance in the basic state

$$\tilde{\Upsilon}\tilde{y}_\Theta = -\tilde{p}_Y/(f\hat{f}), \tag{9.37}$$

both of which may be found using (9.16). Using the WKB ansatz (7.7), with Z replaced by Θ,

$$\langle F_{(X)}\rangle = -\frac{\tilde{\Psi}_Y}{f}\langle S_{int}\rangle + \left\{-\frac{\tilde{\sigma}}{2f^2}K^2 + \frac{1}{2g}\left[-\frac{\tilde{p}_\Theta}{f\hat{f}}L^2 + 2\frac{\tilde{p}_Y}{f\hat{f}}LM + \tilde{\Upsilon}\tilde{y}_Y M^2\right]\right\}\frac{|\Psi^*|^2}{2}, \tag{9.38a}$$

$$\langle F_{(Y)}\rangle = \frac{1}{g}\left(\frac{\tilde{p}_\Theta}{f\hat{f}}KL + \tilde{\Upsilon}\tilde{y}_\Theta KM\right)\frac{|\Psi^*|^2}{2}, \tag{9.38b}$$

$$\langle F_{(\Theta)}\rangle = \frac{1}{g}\left(-\frac{\tilde{p}_Y}{f\hat{f}}KL - \tilde{\Upsilon}\tilde{y}_Y KM\right)\frac{|\Psi^*|^2}{2}. \tag{9.38c}$$

To verify the group-velocity property, we have from (9.16) that to leading order

$$\sigma' = h_{ij}\,\partial_{X_i}\partial_{X_j}\Psi' + p_i\,\partial_{X_i}\Psi', \tag{9.39}$$

where $$h_{ij} = -\frac{1}{g}\begin{pmatrix} g\tilde{\sigma}/f^2 & 0 & 0 \\ 0 & -\tilde{p}_\theta/(f\hat{f}) & \tilde{p}_Y/(f\hat{f}) \\ 0 & -\tilde{Y}\tilde{y}_\theta & \tilde{Y}\tilde{y}_Y \end{pmatrix}, \quad p_i = \frac{1}{g}\begin{pmatrix} 0 \\ \partial_{Y\theta}[1/(f\hat{f}),\tilde{p}] \\ -\partial_{Y\theta}(\tilde{y},\tilde{Y}) \end{pmatrix}. \tag{9.40}$$

The matrix h_{ij} is symmetric from thermal wind balance (9.37). In order for (9.39) to be invertible, we require h_{ij} to be sign definite. As in the Boussinesq f-plane case, for a statically stable basic state with positive PV, h_{ij} is negative definite since in that case

$$-h_{11} = \frac{\tilde{\sigma}}{f^2} > 0, \quad \det\begin{pmatrix} -h_{11} & -h_{12} \\ -h_{21} & -h_{22} \end{pmatrix} = -\frac{\tilde{p}_\theta\,\tilde{\sigma}}{g f^3 \hat{f}} > 0, \quad \det(-h_{ij}) = \frac{\tilde{Y}(\tilde{\sigma}/f)^2}{g f \hat{f}} > 0. \tag{9.41}$$

Under the WKB approximation

$$\sigma' = -h_{ij}\,K_i\,K_j\,\Psi', \tag{9.42}$$

the dispersion relation is

$$\Omega = -\frac{\tilde{\Psi}_Y}{f}K + \frac{(\tilde{\sigma}/f)_Y K}{(-h_{ij}K_iK_j)}, \tag{9.43}$$

and the group velocity is

$$c_{g(l)} = \partial_{K_l}\Omega = \left(-\frac{\tilde{\Psi}_Y}{f} + \frac{(\tilde{\sigma}/f)_Y}{(-h_{ij}K_iK_j)}\right)\delta_{l1} + \frac{2(\tilde{\sigma}/f)_Y K}{(-h_{ij}K_iK_j)^2}h_{lm}K_m, \tag{9.44}$$

which follows from the symmetry of h_{ij}. The phase-averaged linearized pseudo-momentum is, from (9.32) and (9.42),

$$\langle S_{int}\rangle = \frac{(-h_{ij}K_iK_j)^2}{4(\tilde{\sigma}/f)_Y}|\Psi^*|^2, \tag{9.45}$$

and the group-velocity property follows directly from (9.44) and (9.45) if we rewrite (9.38) as

$$\langle F_l\rangle = \left(-\frac{\tilde{\Psi}_Y}{f}\langle S_{int}\rangle + \frac{-h_{ij}K_iK_j}{4}|\Psi^*|^2\right)\delta_{l1} + \frac{Kh_{lj}K_j}{2}|\Psi^*|^2. \tag{9.46}$$

9.8. *Wave–mean-flow interaction theory*

We find that (8.2) holds with f variable and S_{int} given by the small-amplitude form in (9.32). The zonally averaged E–P flux is given by

$$\bar{F} = \frac{f}{g}\begin{pmatrix} -\tilde{p}_\theta & \tilde{y}_\theta \\ \tilde{p}_Y & -\tilde{y}_Y \end{pmatrix}\begin{pmatrix} \overline{v_g^* y^*} \\ \overline{v_g^* p^*} \end{pmatrix}. \tag{9.47}$$

Equation (8.6) holds with the operator

$$\mathscr{L} = \frac{1}{g}\left[\frac{\tilde{p}_\theta}{f\hat{f}}(\cdot)_{YY} - 2\frac{\tilde{p}_Y}{f\hat{f}}(\cdot)_{Y\theta} - \tilde{Y}\tilde{y}_Y(\cdot)_{\theta\theta} + \partial_{Y\theta}(1/(f\hat{f}),\tilde{p})(\cdot)_Y - \partial_{Y\theta}(\tilde{y},\tilde{Y})(\cdot)_\theta\right], \tag{9.48}$$

from (9.40), and

$$\nabla\cdot\bar{J} = -\frac{1}{g}[\overline{\partial_{XY\theta}(x^*,y_1^*,\tilde{p})} + \overline{\partial_{Y\theta}(y_1^*,p_1^*)} + \partial_{Y\theta}(\overline{y_2'},\tilde{p}) + \partial_{Y\theta}(\tilde{y},\overline{p_2'}) + \overline{\partial_{XY\theta}(x^*,\tilde{y},p_1^*)}], \tag{9.49}$$

where the subscripts 1 and 2 indicate the first- and second-order terms in the expansions of p' and y' in (A 4) and (A 7). The linear terms in the Jacobians in (9.49) are

$$x^* = -\Psi^*_X/f^2, \quad y^*_1 = -\Psi^*_Y/(f\hat{f}), \quad p^*_1 = \tilde{\Upsilon}\Psi^*_\theta, \tag{9.50}$$

and the zonally averaged quadratic eddy correlations are, to leading order,

$$\overline{y'_2} = \frac{\beta}{f^2\hat{f}}\left[\frac{\overline{(\Psi^*_X)^2}}{f^2}+\frac{\overline{(\Psi^*_Y)^2}}{\hat{f}^2}\right], \quad \overline{p'_2} = \frac{\mathrm{d}\tilde{\Upsilon}}{\mathrm{d}\tilde{\Psi}_\theta}\overline{(\Psi^*_\theta)^2}. \tag{9.51}$$

The nonlinear relation between p, y and Ψ in (9.16) implies, bearing in mind the discussion of §8, that the boundary conditions on $\overline{\Psi'_T}$ will be more complex than (8.10). We find from (A 7) that

$$\overline{y'_T} = -\overline{\Psi'_{YT}}/(f\hat{f})+\overline{y'_{2T}}, \tag{9.52}$$

and from (9.16) that

$$\overline{z'_T} = \frac{1}{g}\left(\overline{\Psi'_T}-\Theta\overline{\Psi'_{\theta T}}-f\tilde{u}_g\overline{y'_T}-\frac{\partial\overline{(v^*_g)^2}}{\partial T}\right). \tag{9.53}$$

The boundary conditions on $\overline{\Psi'_T}$ are thus, from (9.15c, d),

$$\overline{\Psi'_{YT}} = f\hat{f}[\overline{y'_{2T}}+(\overline{v^*_g y^*})_Y] = f\hat{f}\frac{\partial}{\partial T}\left[\overline{y'_2}-\frac{\partial}{\partial Y}\frac{\overline{(y')^2}}{2\tilde{y}_Y}\right] \quad \text{on} \quad y = y_1, y_2, \tag{9.54a}$$

$$\begin{aligned}\overline{\Psi'_T}-\Theta\overline{\Psi'_{\theta T}}+\frac{\tilde{u}_g}{\hat{f}}\overline{\Psi'_{YT}} &= \frac{\partial}{\partial T}\left(f\tilde{u}_g\overline{y'_2}+\frac{\overline{(\Psi^*_X)^2}}{f^2}\right)-\frac{\partial}{\partial Y}(g\overline{v^*_g z^*}) \\ &= \frac{\partial}{\partial T}\left[f\tilde{u}_g\overline{y'_2}+\frac{\overline{(\Psi^*_X)^2}}{f^2}+\frac{\partial}{\partial Y}\frac{g\overline{(z')^2}}{2\tilde{z}_Y}\right] \quad \text{on} \quad z = z_1, z_2. \end{aligned} \tag{9.54b}$$

The ellipticity of the operator (9.48) is guaranteed for a statically stable positive-PV basic state. The proof of uniqueness of the boundary value problem, given all eddy forcings in the interior and at the boundaries, is similar to that given in Appendix B for the f-plane Boussinesq case. Given this uniqueness, the non-acceleration theorem holds for this system.

Under the appropriate WKB-like conditions, (8.11) is replaced by

$$\mathscr{L}(\overline{u'_{g_T}}) = \frac{\partial^2}{\partial Y\partial T}\left(\frac{\nabla\cdot\bar{J}}{\hat{f}}\right)+\frac{\partial^2}{\partial Y^2}\frac{\nabla\cdot\bar{F}}{f\hat{f}} = \frac{\partial^2}{\partial Y\partial T}\left(\frac{\nabla\cdot\bar{J}}{\hat{f}}\right)-\frac{\partial^2}{\partial Y^2}\left(\frac{1}{f\hat{f}}\frac{\partial\overline{S_{int}}}{\partial T}-\frac{\bar{G}}{f\hat{f}}\right). \tag{9.55}$$

10. Discussion

In the Introduction we referred to a body of theoretical work based on QG wave-activity invariants. We have derived the pseudomomentum invariant for the SG system, and have used it to extend some of this QG theory to SG dynamics: the linearized Charney–Stern stability theorem and its nonlinear generalization, the finite-amplitude local wave-activity conservation law, and the E–P flux diagnostics describing wave–mean-flow interaction. The existence of the pseudomomentum, and its similarity to the QG form, is guaranteed because SG dynamics, like QG dynamics, is a PV-advecting invertible system, with an underlying Hamiltonian structure.

At finite amplitude, the interior contributions to the SG and QG pseudomomenta are similar, involving the spatial distribution of the basic-state PV. The more

complicated form of the finite-amplitude boundary contribution to the SG pseudomomentum reflects the behaviour of the dynamic isentropic and geostrophic coordinates at the boundaries. At small amplitude, the SG and QG vertical boundary contributions are analogous. However, the ageostrophic circulations allowed in the SG form at meridional boundaries yield contributions to the pseudomomentum that are absent in QG dynamics.

Conservation of the small-amplitude form of the pseudomomentum yields the linearized Charney–Stern theorem. The QG theorem describes conditions on the signs of the meridional gradients of the interior PV and of the vertical-boundary potential-temperature distributions that must be obeyed for the flow to be necessarily stable. The terms at the meridional boundaries included in the SG pseudomomentum yield additional stability criteria. According to these criteria stable basic states for flow in a channel are difficult, if not impossible, to construct. On the other hand, the stability theorem derived in Part 2 of this study (Kushner & Shepherd 1995), which arises from pseudoenergy conservation and is analogous to Arnol'd's first theorem, does not suffer from this restriction.

The nonlinear generalization of the QG Charney–Stern theorem of S89 bounds a disturbance enstrophy-like norm, consisting of both interior and vertical boundary contributions, in terms of its initial value. For the extension to SG theory, the complex boundary dynamics require us to consider perturbations that are restricted to vanish on the domain edges, and the potentially destabilizing effect of the meridional boundaries requires further restrictions at these boundaries. Furthermore, the positive-definite disturbance quantity that is bounded in terms of its initial value is not, in general, a norm. The resulting Liapunov stability theorem, *mutatis mutandi*, is similar to that of QG theory. It is an unresolved question whether having to make these restrictions represents a mere technical difficulty or has significant physical implications for the nonlinear theory. Before constructing saturation bounds for disturbances to unstable basic states in SG theory, it will be necessary to understand this issue more fully.

The QG and SG finite-amplitude local pseudomomentum fluxes are also similar. Because of the nonlinear invertibility relation in SG dynamics, the SG meridional PV flux includes terms of higher-than-quadratic order in the disturbance.

The QG wave–mean-flow interaction theorem provides a closure relation between eddy forcing and tendencies in the zonal-mean flow. Because the SG invertibility relation is nonlinear, additional eddy correlation terms besides the zonal-mean PV fluxes are needed to obtain the mean-flow stream-function tendency. It remains to be seen how critical and realistic this effect is, for example, in the eddy-driven circulation. A useful test would be a direct comparison of the SG and QG E–P flux diagnostics for a highly baroclinic state.

One principal motivation of this study is to obtain improved diagnostics for the analysis of large-scale flows from SG theory. SG theory overcomes many of the practical limitations of QG theory, since unlike the latter it: (i) allows for large ageostrophic circulations; (ii) does not build in an assumption that the flow is linearized about a resting reference state; and (iii) makes no explicit restrictions on the shape and size of the variation of underlying topography relative to the scale height. These points suggest that SG-based analysis could be superior to that of QG in regimes, such as the lower stratosphere and the oceans, where the stratification varies strongly with latitude. Regarding the meridional boundaries as finite-amplitude jumps in the topography, we have seen how (iii) comes into play in the Charney–Stern theorem. It may be shown that the theorem also generalizes to flow parallel to finite-amplitude

ridges (Kushner 1995). In §9, we have shown that the results of §§2–8 generalize to β-plane geometry in a compressible atmosphere. The effects of β and large-scale compressibility are critical to any proper model of large-scale dynamics on the sphere. It should be noted that Magnusdottir & Schubert (1991) have extended SG dynamics to the hemisphere, and we are confident that the results presented here should also extend to that geometry.

The claim that SG theory could provide a superior diagnostic tool might be criticized given that the formal and numerical accuracy of three-dimensional SG dynamics, as well as its physical realism in simulations, have been questioned (McWilliams & Gent 1980; Barth, Allen & Newberger 1990; Snyder, Skamarock & Rotunno 1991). These objections generally focus on SG dynamics as an accurate, predictive intermediate balanced model. The emphasis in this work, however, is on the diagnostic use of the system for insight into observations and model output, in the spirit of the use of QG theory as a tool of understanding rather than of prediction. The critical question for applications of the results in this study then becomes whether or not the distinct processes captured by SG theory are reasonably well-represented. This is a very different issue than that of long-term accuracy of the model equations. Even though SG dynamics may be inaccurate in simulations because of small-error buildup over time, at any instant the SG diagnostic relationships may be quite good.

The question of whether the MSc90 system models β and compressibility effects well has not yet been fully answered: MSc90 support their system by investigating its Rossby–Haurwitz normal modes and comparing them to the modes in Laplace's tidal equations. The results of this study provide a framework for further tests, such as general circulation model or observed data analysis.

We would like to thank O. Bokhove and J. C. Bowman for helpful discussions. P.J.K. is supported by a postgraduate fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). T.G.S. is supported by NSERC as well as by the Atmospheric Environment Service of Canada.

Appendix A. Local flux law calculations

A.1. *f-plane Boussinesq system*

Equation (7.5) may be rewritten

$$\partial S_{int}/\partial T + \partial_X(u_g S_{int}) + \partial_Y(v_g' S_{int}) + f v_g' \sigma' = 0. \tag{A 1}$$

We will show that the last term, the meridional σ-flux, can be expressed as the divergence of a vector field. The finite-amplitude disturbance σ' is

$$\sigma' = \partial_{XYZ}(\tilde{x}+x', \tilde{y}+y', \tilde{z}+z') - \partial_{XYZ}(\tilde{x}, \tilde{y}, \tilde{z}),$$

whence from (2.20) we have (noting that $\tilde{x} = X$)

$$\begin{aligned} f v_g' \sigma' = & = \partial_{XYZ}[-\tfrac{1}{2}(v_g')^2, y, z] + f\partial_{YZ}(v_g' y', \tilde{z}) + f\partial_{YZ}(\tilde{y}, v_g' z') \\ & \quad - f y' \partial_{YZ}(v_g', \tilde{z}) - f z' \partial_{YZ}(\tilde{y}, v_g') \\ & \quad + \partial_{XYZ}(\Psi', y', z') - \Psi'_Y \partial_{ZX}(y', z') - \Psi'_Z \partial_{XY}(y', z') \\ & = \partial_{XYZ}[-\tfrac{1}{2}(v_g')^2, y, z] + f\partial_{YZ}(v_g' y', \tilde{z}) + f\partial_{YZ}(\tilde{y}, v_g' z') \\ & \quad + \tfrac{1}{2} f^2 \partial_X[\tilde{z}_Z (y')^2 - 2\tilde{z}_Y y' z' + \tilde{y}_Y (z')^2] \\ & \quad + \partial_{XYZ}(\Psi', y', z') + \tfrac{1}{2} f^2 [\partial_{ZX}((y')^2, z') + \partial_{XY}(y', (z')^2)]. \end{aligned} \tag{A 2}$$

From (A 1) and (A 2), a flux $\boldsymbol{F}$ satisfying (7.1) is

$$F_{(X)} = u_g S_{int} - \tfrac{1}{2}(v_g')^2 \partial_{YZ}(y,z) + \tfrac{1}{2}f^2[\tilde{z}_Z(y')^2 - 2\tilde{z}_Y y'z' + \tilde{y}_Y(z')^2] + \Psi' \partial_{YZ}(y',z') + 2f^2 y'_Z y'z', \tag{A 3a}$$

$$F_{(Y)} = v_g' S_{int} - \tfrac{1}{2}(v_g')^2 \partial_{ZX}(y,z) + f(\tilde{z}_Z v_g' y' - \tilde{y}_Z v_g' z') + \Psi' \partial_{ZX}(y',z') - f^2 z'_X y'z', \tag{A 3b}$$

$$F_{(Z)} = -\tfrac{1}{2}(v_g')^2 \partial_{XY}(y,z) + f(\tilde{y}_Y v_g' z' - \tilde{z}_Y v_g' y') + \Psi' \partial_{XY}(y',z') - f^2 y'_X y'z', \tag{A 3c}$$

where y and z are to be distinguished from $\tilde{y}$ and $\tilde{z}$.

A.2. *β-plane compressible system*

The derivation follows that of §A.1, with the important difference that we must expand the pressure and y-fields in power series as follows. From (9.16), we write

$$p' = \sum_{i=1}^{\infty} \frac{\mathrm{d}^i \tilde{p}}{\mathrm{d}(\tilde{\Psi}_\theta)^i} \frac{(\Psi'_\theta)^i}{i!} = \tilde{\Upsilon}\Psi'_\theta + \frac{\mathrm{d}\tilde{\Upsilon}}{\mathrm{d}\tilde{\Psi}_\theta}\frac{(\Psi'_\theta)^2}{2} + \ldots, \tag{A 4}$$

where Υ is given in (9.35). For y, from (9.16) we find

$$y = Y + \frac{f}{2\beta}(1-(1+a)^{1/2}), \tag{A 5}$$

where

$$a = 4\beta f^{-3}[\Psi_Y - \beta f^{-3}(\Psi_X)^2], \quad \tilde{a} = 4\beta f^{-3}\tilde{\Psi}_Y, \quad a' = 4\beta f^{-3}[\Psi'_Y - \beta f^{-3}(\Psi'_X)^2]. \tag{A 6}$$

We may now expand y as

$$y' = \sum_{j=1}^{\infty} \frac{\mathrm{d}^j \tilde{y}}{\mathrm{d}\tilde{a}^j}\frac{(a')^j}{j!} = -\frac{\Psi'_Y}{f\hat{f}} + \frac{\beta}{f^2\hat{f}}\left[\frac{(\Psi'_X)^2}{f^2} + \frac{(\Psi'_Y)^2}{\hat{f}^2}\right] + \ldots, \tag{A 7}$$

where $\hat{f}$ is also given in (9.35).

The meridional flux of σ', from (9.16), is proportional to

$$\begin{aligned} g\Psi'_X\sigma' &= -\Psi'_X \partial_{XY\theta}(x',y,p) - \Psi'_X[\partial_{Y\theta}(\tilde{y},p') + \partial_{Y\theta}(y',\tilde{p}) + \partial_{Y\theta}(y',p')] \\ &= \partial_{XY\theta}(\tfrac{1}{2}v_g^2,y,p) - \partial_{Y\theta}(\tilde{y},p'\Psi'_X) - \partial_{Y\theta}(y'\Psi'_X,\tilde{p}) \\ &\quad - \partial_{XY\theta}(\Psi',y',p') + R_1 + R_2, \end{aligned} \tag{A 8}$$

where

$$\begin{aligned} R_1 &= p'\partial_{Y\theta}(\tilde{y},\Psi'_X) + y'\partial_{Y\theta}(\Psi'_X,\tilde{p}) - \beta f^{-3}(\Psi'_X)^2[\partial_{\theta X}(\tilde{y},p') + \partial_{\theta X}(y',\tilde{p})] \\ &= -\partial_{\theta X}(\tilde{y},\beta f^{-3}(\Psi'_X)^2 p') - \partial_{\theta X}(\beta f^{-3}(\Psi'_X)^2 y',\tilde{p}) \\ &\quad + p'[\partial_{Y\theta}(\tilde{y},\Psi'_X) + \partial_{\theta X}(\tilde{y},\beta f^{-3}(\Psi'_X)^2)] + y'[\partial_{Y\theta}(\Psi'_X,\tilde{p}) + \partial_{\theta X}(\beta f^{-3}(\Psi'_X)^2,\tilde{p})] \\ &= -\partial_{\theta X}(\tilde{y},\beta f^{-3}(\Psi'_X)^2 p') - \partial_{\theta X}(\beta f^{-3}(\Psi'_X)^2 y',\tilde{p}) - \tilde{\Psi}_{Y\theta}[(\mathrm{d}\tilde{y}/\mathrm{d}\tilde{a})p'a'_X + \tilde{\Upsilon}y'\Psi'_{\theta X}] \\ &\quad + \frac{\partial}{\partial X}\left[\frac{\tilde{p}_\theta}{4\beta f^{-3}}\sum_{j=1}^{\infty}\frac{\mathrm{d}^j\tilde{y}}{\mathrm{d}\tilde{a}^j}\frac{(a')^{j+1}}{(j+1)!} + \tilde{y}_Y\sum_{i=1}^{\infty}\frac{\mathrm{d}^i\tilde{p}}{\mathrm{d}(\tilde{\Psi}_\theta)^i}\frac{(\Psi'_\theta)^{i+1}}{(i+1)!}\right] \end{aligned} \tag{A 9}$$

and

$$\begin{aligned} R_2 &= (a'/4\beta f^{-3})\partial_{\theta X}(y',p') + \Psi'_\theta \partial_{XY}(y',p') \\ &= \partial_{\theta X}\left[y', \frac{p'a'}{4\beta f^{-3}}\right] + \partial_{XY}(\Psi'_\theta y', p') + \tilde{\Psi}_{Y\theta}\left[\frac{\mathrm{d}\tilde{y}}{\mathrm{d}\tilde{a}}p'a'_X + \frac{\mathrm{d}\tilde{p}}{\mathrm{d}\tilde{\Psi}_\theta}y'\Psi'_{\theta X}\right] \\ &\quad - \partial_X(\tilde{\Psi}_{Y\theta}y'p'). \end{aligned} \tag{A 10}$$

Gathering the terms in (A 8)–(A 10),

$$\begin{aligned} g\Psi'_X\sigma' &= \partial_{XY\theta}(\tfrac{1}{2}v_g^2,y,p) - \partial_{Y\theta}(\tilde{y},p'\Psi'_X) - \partial_{Y\theta}(y'\Psi'_X,\tilde{p}) - \partial_{XY\theta}(\Psi',y',p') \\ &\quad + \frac{\partial}{\partial X}\left[\frac{\tilde{p}_\theta}{4\beta f^{-3}}\sum_{j=1}^{\infty}\frac{\mathrm{d}^j\tilde{y}}{\mathrm{d}\tilde{a}^j}\frac{(a')^{j+1}}{(j+1)!} - \tilde{\Psi}_{Y\theta}y'p' + \tilde{y}_Y\sum_{i=1}^{\infty}\frac{\mathrm{d}^i\tilde{p}}{\mathrm{d}(\tilde{\Psi}_\theta)^i}\frac{(\Psi'_\theta)^{i+1}}{(i+1)!}\right. \\ &\quad \left. + \beta f^{-3}(\Psi'_X)^2(-\tilde{y}_\theta p' + \tilde{p}_\theta y')\right] + \partial_{\theta X}[y', p'a'/(4\beta f^{-3})] + \partial_{XY}(\Psi'_\theta y', p'). \end{aligned} \tag{A 11}$$

Combining (9.33) with (A 11), a flux F satisfying (7.1) is

$$F_{(X)} = -\frac{\Psi_Y}{f} S_{int} + \frac{1}{g}\Bigg[\tfrac{1}{2}(v_g')^2 \partial_{Y\Theta}(y,p) + \frac{\tilde{p}_\Theta}{4\beta f^{-3}} \sum_{j=1}^{\infty} \frac{\mathrm{d}^j \tilde{y}}{\mathrm{d}\tilde{a}^j} \frac{(a')^{j+1}}{(j+1)!}$$
$$- \tilde{\Psi}_{Y\Theta} y' p' + \tilde{y}_Y \sum_{i=1}^{\infty} \frac{\mathrm{d}^i \tilde{p}}{\mathrm{d}(\tilde{\Psi}_\Theta)^i} \frac{(\Psi'_\Theta)^{i+1}}{(i+1)!} + \beta f^{-3} (\Psi'_X)^2 (-\tilde{y}_\Theta p' + \tilde{p}_\Theta y')$$
$$- \Psi' \partial_{Y\Theta}(y',p') + (y'_\Theta p' a'/4\beta f^{-3}) + \Psi'_\Theta y' p'_Y \Bigg], \tag{A 12a}$$

$$F_{(Y)} = \frac{\Psi'_X}{f} S_{int} + \frac{1}{g}[\tfrac{1}{2}(v_g')^2 \partial_{\Theta X}(y,p) + \Psi'_X(\tilde{y}_\Theta p' - \tilde{p}_\Theta y') - \Psi' \partial_{\Theta X}(y',p') - \Psi'_\Theta y' p'_X], \tag{A 12b}$$

$$F_{(\Theta)} = \frac{1}{g}\Bigg[\tfrac{1}{2}(v_g')^2 \partial_{XY}(y,p) + \Psi'_X(\tilde{p}_Y y' - \tilde{y}_Y p') - \Psi' \partial_{XY}(y',p') - \frac{y'_X p' a'}{4\beta f^{-3}}\Bigg]. \tag{A 12c}$$

Appendix B. Uniqueness of solutions to (8.9)

THEOREM *The solution* $\overline{\Psi'_T}(Y,Z,T)$ *to* (8.9) *with given* $\nabla\cdot\bar{F}$ *and* $\nabla\cdot\bar{J}_T$, *and boundary conditions* (8.10) *with given* $\overline{v_g^* y^*}$ *and* $\overline{v_g^* z^*}$ *at the appropriate surfaces, is unique to within a spatially constant function of time.*

Proof The proof uses the methods of §3. Suppose we have two solutions $\bar{\Psi}_{1,T}$ and $\bar{\Psi}_{2,T}$ to the problem. Then their difference $\bar{\Phi}_T = \bar{\Psi}_{1,T} - \bar{\Psi}_{2,T}$ satisfies the homogeneous problem

$$\mathscr{L}(\bar{\Phi}_T) = 0, \tag{B 1}$$

with Neumann boundary conditions

$$\bar{\Phi}_{YT} = 0 \quad \text{on} \quad y = y_1, y_2, \quad \bar{\Phi}_{ZT} = 0 \quad \text{on} \quad z = z_1, z_2. \tag{B 2}$$

We find

$$\int_{\bar{D}} \bar{\Phi}_T \mathscr{L}(\bar{\Phi}_T)\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z = \frac{1}{f^2}\int_{\bar{D}} (\bar{\Phi}_{TY}\,\bar{\Phi}_{TZ}) \begin{pmatrix} \tilde{z}_Z & -\tilde{z}_Y \\ -\tilde{y}_Z & \tilde{y}_Y \end{pmatrix} \begin{pmatrix} \bar{\Phi}_{TY} \\ \bar{\Phi}_{TZ} \end{pmatrix} \mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z,$$
$$= 0, \tag{B 3}$$

where the boundary terms are treated in a similar manner to the terms in (3.8) and (3.9), and (B 2) has been used.

The quadratic form in (B 3) is positive definite for a statically stable basic state ($\tilde{z}_Z > 0$) with positive basic-state PV. For the equality to hold, then, $\bar{\Phi}_{TY} = \bar{\Phi}_{TZ} = 0$. The theorem follows. □

REFERENCES

ANDREWS, D. G. 1983 A finite-amplitude Eliassen–Palm theorem in isentropic coordinates. *J. Atmos. Sci.* **40**, 1877–1883.

ANDREWS, D. G., HOLTON, J. R. & LEOVY, C. B. 1987 *Middle Atmosphere Dynamics.* Academic Press.

ANDREWS, D. G. & MCINTYRE, M. E. 1976 Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. *J. Atmos. Sci.* **33**, 2031–2048.

ANDREWS, D. G. & MCINTYRE, M. E. 1978 Generalized Eliassen–Palm and Charney–Drazin theorems for waves on axisymmetric mean flows in compressible atmospheres. *J. Atmos. Sci.* **35**, 175–185.

ARNOL'D, V. I. 1966 On an a priori estimate in the theory of hydrodynamical stability. *Izv. Vyssh. Uchebn. Zaved. Matematika* **54** (5), 3–5. (English transl. *Am. Math. Soc. Transl., Series 2*, **79**, 267–269 (1969).)

BARTH, J. A., ALLEN, J. S. & NEWBERGER, P. A. 1990 On intermediate models for barotropic continental shelf and slope flow fields: Part II, comparison of numerical model solutions in doubly-periodic domains. *J. Phys. Oceanogr.* **20**, 1044–1076.

BRETHERTON, F. P. 1966 Critical layer instability in baroclinic flows. *Q. J. R. Met. Soc.* **92**, 325–334.

BRETHERTON, F. P. & GARRETT, C. J. R. 1968 Wavetrains in inhomogeneous moving media. *Proc. R. Soc. Lond.* A **302**, 529–554.

CHYNOWETH, S. & SEWELL, M. J. 1989 Dual variables in semigeostrophic theory. *Proc. R. Soc. Lond.* A **424**, 155–186.

CRAIG, G. C. 1993 A scaling for the three-dimensional semigeostrophic approximation. *J. Atmos. Sci.* **50**, 3350–3355.

DUNKERTON, T. J., HSU, C.-P. F. & MCINTYRE, M. E. 1981 Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.* **38**, 819–843.

EDMON, H. J., HOSKINS, B. J. & MCINTYRE, M. E. 1980 Eliassen–Palm cross sections for the troposphere. *J. Atmos. Sci.* **37**, 2600–2616.

ELIASSEN, A. 1948 The quasi-static equations of motion. *Geofys. Publ.* **17**, No. 3.

ELIASSEN, A. 1983 The Charney–Stern theorem on barotropic–baroclinic instability. *Pure Appl. Geophys.* **121**, 563–572.

HELD, I. M. & HOSKINS, B. J. 1985 Large-scale eddies and the general circulation of the atmosphere. *Adv. Geophys.* **28A**, 3–31.

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

HOSKINS, B. J. 1975 The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.* **32**, 233–242.

HOSKINS, B. J. 1983 Modelling of the transient eddies and their feedback on the mean flow. In *Large-Scale Dynamical Processes in the Atmosphere* (ed. B. J. Hoskins & R. P. Pearce), pp. 169–199. Academic Press.

HOSKINS, B. J. & BRETHERTON, F. P. 1972 Atmospheric frontogenesis models: mathematical formulation and solution. *J. Atmos. Sci.* **29**, 11–37.

HOSKINS, B. J., MCINTYRE, M. E. & ROBERTSON, A. W. 1985 On the use and significance of isentropic potential-vorticity maps. *Q. J. R. Met. Soc.* **111**, 877–946.

KUSHNER, P. J. 1993 Nonlinear stability and wave, mean-flow interaction in semi-geostrophic theory. In *Ninth Conf. on Atmospheric and Oceanic Waves and Stability*. Preprint volume, pp. 363–366. American Meteorological Society.

KUSHNER, P. J. 1995 A generalized Charney–Stern theorem for semi-geostrophic dynamics. *Tellus* (In press).

KUSHNER, P. J. & SHEPHERD, T. G. 1995 Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics. Part 2. Pseudoenergy-based theory. *J. Fluid Mech.* **290**, 105–129.

LORENZ, E. N. 1955 Available potential energy and the maintenance of the general circulation. *Tellus* **7**, 157–167.

LU, C. D. 1993 Balanced dynamics for three dimensional curved flows. PhD dissertation, Colorado State University.

MAGNUSDOTTIR, G. & SCHUBERT, W. H. 1990 The generalization of semi-geostrophic theory to the β-plane. *J. Atmos. Sci.* **47**, 1714–1720 (referred to herein as MSc90).

MAGNUSDOTTIR, G. & SCHUBERT, W. H. 1991 Semigeostrophic theory on the hemisphere. *J. Atmos. Sci.* **48**, 1449–1456.

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.

MCWILLIAMS, J. C. & GENT, P. R. 1980 Intermediate models of planetary circulations in the atmosphere and ocean. *J. Atmos. Sci.* **37**, 1657–1678.

PURSER, R. J. & CULLEN, M. J. P. 1987 A duality principle in semi-geostrophic theory. *J. Atmos. Sci.* **44**, 3449–3468.

RANDEL, W. J. & HELD, I. M. 1991 Phase speed spectra of transient eddy fluxes and critical layer absorption. *J. Atmos. Sci.* **48**, 688–697.

ROULSTONE, I. & NORBURY, J. 1994 A Hamiltonian structure with contact geometry for the semi-geostrophic equations. *J. Fluid Mech.* **272**, 211–233.

SALMON, R. 1985 New equations for nearly geostrophic flow. *J. Fluid Mech.* **153**, 461–477.

SALMON, R. 1988*a* Semigeostrophic theory as a Dirac-bracket projection. *J. Fluid Mech.* **196**, 345–358.

SALMON, R. 1988*b* Hamiltonian fluid mechanics. *Ann. Rev. Fluid Mech.* **20**, 225–256.

SEWELL, M. J. & ROULSTONE, I. 1993 Anatomy of the canonical transformation. *Phil. Trans. R. Soc. Lond.* A **345**, 577–598.

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

SHEPHERD, T. G. 1989 Nonlinear saturation of baroclinic instability. Part II: Continuously stratified fluid. *J. Atmos. Sci.* **46**, 888–907 (referred to herein as S89).

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

SHEPHERD, T. G. 1993 Nonlinear saturation of baroclinic instability. Part III: Bounds on the energy. *J. Atmos. Sci.* **50**, 2697–2709.

SNYDER, C., SKAMAROCK, W. . & ROTUNNO, R. 1991 A comparison of primitive-equation and semigeostrophic simulations of baroclinic waves. *J. Atmos. Sci.* **48**, 2179–2194.

WHITE, A. A. 1977 Modified quasi-geostrophic equations using geometric height as vertical coordinate. *Q. J. R. Met. Soc.* **103**, 383–396.

WHITHAM, G. B. 1965 A general approach to linear and nonlinear dispersive waves using a Lagrangian. *J. Fluid Mech.* **22**, 273–283.