Are there Higher-Accuracy Analogues of Semigeostrophic Theory?*

Michael E. McIntyre
Centre for Atmospheric Science† at the
Department of Applied Mathematics and Theoretical Physics,
Silver St, Cambridge CB3 9EW,
UK,
http://www.atm.damtp.cam.ac.uk/people/mem/

Ian Roulstone
Joint Centre for Mesoscale Meteorology (Met Office),
University of Reading,
Meteorology Building, Earley Gate,
Reading RG6 6BB,
UK,
ian.roulstone@metoffice.com

As finally to press, June 2001

*Article to appear in *Large-scale Atmosphere–Ocean Dynamics: II: Geometric Methods and Models,* ed. I. Roulstone & J. Norbury. Cambridge, University Press (2001). Also Internal Report No. 128 and NWP Scientific Paper No. 61 of the Joint Centre for Mesoscale Meteorology (Met Office).

†The Centre for Atmospheric Science is a joint initiative of the Department of Chemistry and the Department of Applied Mathematics and Theoretical Physics (http://www.atm.damtp.cam.ac.uk/).

Are there Higher-Accuracy Analogues of Semigeostrophic Theory?*

Michael E. McIntyre and *Ian Roulstone*

1 Prelude

This article describes an unfinished journey and its main landmarks so far, along with elementary tutorial material to review and prepare the background. The journey takes us through territory that has long been familiar and elementary to geometers but may well seem strange to meteorologists and fluid dynamicists. We think that this justifies the tutorial emphasis.

Ever since the seminal work of Eliassen (1948, 1962), Hoskins & Bretherton (1972), and Hoskins (1975), semigeostrophic theory has had a special place in dynamical meteorology despite its limited accuracy. The reason is that semigeostrophic theory has certain useful and elegant mathematical properties, in particular Hamiltonian structure, Legendre duality, contact structure, and the convexity of certain potential functions (Chynoweth & Sewell 1989, 1991; Roulstone & Norbury 1994; Roulstone & Sewell 1997). These permit both robust numerical integration — even in cases where frontal discontinuities form — together with a rather complete knowledge of mathematical properties such as existence and uniqueness of solutions (Cullen & Purser 1984, Purser & Cullen 1987, Cullen *et al.* 1991).

Semigeostrophic theory also permits the evolution to be described in a particularly simple and elegant way (Section 4 below) in terms of potential vorticity (PV) advection and PV inversion (e.g. Hoskins *et al.* 1985). The PV is the materially conserved scalar quantity associated with the so-called particle-relabelling symmetry. Inversion means deducing the mass and velocity fields from the PV; in semigeostrophic theory this requires the solution of a Monge–Ampère equation. The simplicity and elegance is associated with the Legendre and contact structure, manifesting itself through a certain transformation to canonical coordinates originally discovered by Hoskins (1975). This transformation has turned out to be useful, also, in the practical data assimilation procedures used in numerical weather prediction (Desroziers & Lafore 1993).

Semigeostrophic theory is one example a so-called 'balanced model', meaning a model constructed so as to eliminate the fast 'inertia–gravity' waves or oscillations that can occur in numerical weather prediction or other 'primitive-equation' models. Such waves can be generated spuriously by, for instance,

*Article to appear in *Large-scale Atmosphere–Ocean Dynamics: II: Geometric Methods and Models,* ed. I.Roulstone & J.Norbury. Cambridge, University Press (2001).

errors in data assimilation and model initialization, or by the so-called 'parametrizations' of subgrid-scale processes such as cumulonimbus convection. However, the limited accuracy of semigeostrophic theory— formally correct only to leading order in the Rossby number ϵ, referred to here as 'geostrophic accuracy' — limits its practical applicability to the problems of weather forecasting and data assimilation. The Rossby number ϵ, which semigeostrophic theory requires to be small, will be referred to repeatedly in this article and for convenience will be defined simply as

$$\epsilon \;=\; \sup|\zeta/f| \;, \tag{1.1}$$

the supremum, in magnitude, of the ratio of vorticity ζ relative to the Earth and the absolute vorticity f due to the Earth's rotation. Here vertical components are understood; f vanishes at the equator and is often called the Coriolis parameter.

The limited accuracy of semigeostrophic theory raises the question of how to construct more accurate balanced models that share at least some of the same useful and elegant mathematical properties. Work at the Newton Institute Programme has led for the first time to the construction of what appears to be a class of such models, and to progress toward understanding their properties, though we are still far from a full understanding. A novel and mathematically interesting feature — very surprising to fluid dynamicists — is that *complex canonical coordinates* arise in a natural and seemingly inevitable way (McIntyre & Roulstone 1996, hereafter 'MR96'), generalizing Hoskins' real canonical coordinates. The requirement for greater accuracy leads inexorably, it seems, to the complex values. These in turn suggest a role for what is called *Kähler structure* or *Kähler geometry* (Roubstov & Roulstone 1997, 2001), a complex counterpart of the *symplectic geometry* that arises in all Hamiltonian dynamical models (e.g. McDuff & Salamon 1998).

Several lines of research have contributed to these developments. The key points are, first, that one may systematically construct Hamiltonian balanced models from their corresponding unbalanced 'parent' models, such as primitive-equation models, by applying balance conditions in the form of workless 'momentum–configuration constraints' that restrict the evolution to a submanifold $\mathcal{M}^{\mathrm{C}}$ of the parent phase space, through a method pioneered in this field by Salmon (1988a); second, an explicit recognition that the application of such constraints always 'splits' the velocity field into at least two separate velocity fields, one of them related to particle motion and the other, or others, to canonical momenta and to conserved quantities like energy, PV, and Kelvin circulation (Allen & Holm 1996, MR96); and third, the discovery of the complex canonical coordinates already mentioned (MR96). Semigeostrophic theory turns out to be a case of a 'two-way' or 'double' split: there are three velocity fields, the first related to particle motion, the second to kinetic energy, and the third to PV. Two of these three velocity fields (in fact the first and second, Section 5 below) are interrelated by a certain explicitly-invertible contact transformation, itself containing Hoskins' transformation to canonical coordinates and thereby connecting the model's vorticity and potential vorticity with the first of the three velocity fields as well as with the third. This set of interconnections seems to be part of what permits the remarkable analytical simplifications characteristic of semigeostrophic theory, the most basic of which are reviewed in Sections 4–5 below.

The plan of the article is as follows. Sections 2–4 review the standard semigeostrophic theory and its transformation properties, including the Legendre transformation involved, in the simplest 'f-plane' case with constant Coriolis parameter. Section 5 shows how these same transformation properties also reflect contact structure, a fact first pointed out by Blumen (1981), along with an associated symplectic structure that is quite distinct from that associated with the parent dynamics. Section 6, moving beyond semigeostrophic theory, shows how Salmon's method can be used to construct Hamiltonian balanced models whose accuracy is limited, in principle, only by the limitations of the concept of balance itself. Those ultimate limitations are associated with the phenomenon of spontaneous-adjustment emission or 'Lighthill radiation' of inertia–gravity waves by unsteady vortical motions. Section 7 reviews Salmon's method in detail in the simplest possible context of a toy particle-dynamics model, showing how velocity-splitting arises. Section 8 carries out the extension to shallow-water dynamics. Sections 9–10 present the complex-valued canonical coordinates, generalizing Hoskins' canonical coordinates and the associated Jacobian vorticity formulae to a far wider class of Hamiltonian balanced models, and showing that this class includes models with variable Coriolis parameter f. Section 11 points out, following recent work of Wunderer (2001), that mathematically well-behaved generalizations of semigeostrophic theory will almost certainly share with semigeostrophic theory the property of double splitting. Sections 12–14 show how Kähler geometry comes into view, along with a so-called hyper-Kähler and hence quaternionic structure. Section 15, the Postlude, briefly describes the generalization to fully stratified systems and then takes a broader view of our journey and how far it has progressed, looking toward interesting parts of the research horizon.

2 Shallow water theory

Throughout this article — except in the Postlude, Section 15 — we use the so-called shallow-water equations, also called by meteorologists the 'shallow-water primitive equations', as the parent dynamics from which Hamiltonian balanced models are to be constructed. These equations represent the simplest dynamical system for which the foregoing ideas are nontrivial, a single layer of incompressible fluid of unit mass density with a free top surface under gravity. The layer depth h is taken to be uniform in the absence of motion relative to a rotating frame of reference, the case of 'no bottom topography'. We shall need both the Lagrangian and the Eulerian descriptions of the fluid motion. In the Lagrangian description a typical particle, more precisely, fluid column, has Cartesian horizontal coordinates $\mathbf{x} = \mathbf{x}(\mathbf{a}, t)$, or

$$x = x(a, b, t), \qquad y = y(a, b, t) \ , \tag{2.1}$$

expressed as functions of the particle labels $\mathbf{a} = (a, b) \in \mathbb{R}^2$ and the time $t \in \mathbb{R}^1$. For convenience we label each particle by its position at a reference time $t = 0$, in other words define the functions in (2.1) to have the property $\mathbf{x}(\mathbf{a}, t) = \mathbf{a}$ at $t = 0$.

Incompressibility requires the layer depth h, when expressed as a function $h(\mathbf{a}, t)$ to satisfy

$$h(\mathbf{a}, t) = h(a, b, t) = \left\{ \frac{\partial(x, y)}{\partial(a, b)} \right\}^{-1} h_0 \ , \tag{2.2}$$

with the Jacobian of the mapping (2.1) appearing on the right. Consistently with the above, we have taken $h(\mathbf{a}, 0) = h_0 =$ constant. The mass element is

$$\mathrm{d}m \ = \ h_0 \, \mathrm{d}\mathbf{a} \ = \ h \, \mathrm{d}\mathbf{x} \ . \tag{2.3}$$

We assume that the mapping (2.1) is invertible, taking $\mathbf{a} \leftrightarrow \mathbf{x}$; thus the inverse

$$a = a(x, y, t), \quad b = b(x, y, t) \ , \tag{2.4}$$

when inserted into (2.2), expresses $h(\mathbf{a}, t)$ as another function

$$h(\mathbf{x}, t) = h(x, y, t) = \frac{\partial(a, b)}{\partial(x, y)} \, h_0 \ , \tag{2.5}$$

bringing in the Eulerian description. Here, for economy of notation, the same symbol h is being used as shorthand for the two different functions $h(\mathbf{x}, t)$ and $h(\mathbf{a}, t)$, emphasizing that, for given t, they have the same value for the same fluid particle. Thus $h(\mathbf{x}, t)$ is shorthand for what could have been written using a different symbol h as $\mathrm{h}(\mathbf{x}, t) = h(\mathbf{a}(\mathbf{x}, t),\ t)$ if we were using a different, 'fixed-slot' convention emphasizing the functional dependences as such. Here, however, the idea is that the single symbol $h(\cdot)$ is used to represent a single geometrical entity, the mass configuration, in one way or another — a geometrical entity basic to describing the Hamiltonian structure, and every other aspect of the fluid dynamics.

Another important quantity is the shallow-water PV defined in the rotating frame by

$$Q = \frac{1}{h} \left(f + \frac{\partial \dot{y}}{\partial x} - \frac{\partial \dot{x}}{\partial y} \right) \ . \tag{2.6}$$

This is h^{-1} times the vertical component of absolute vorticity, the Coriolis parameter f being the contribution to that vorticity from the Earth's rotation, as mentioned earlier. Rossby (1936, eq.(75)) showed that Q is an exact material invariant for frictionless shallow-water systems. The dots signify the material derivative or rate of change following the particle, i.e. partial differentiation of $\mathbf{x}(\mathbf{a},t)$ with respect to t with $\mathbf{a}$ held constant. In (2.6), $\dot{x}$ and $\dot{y}$ are each being regarded as functions of $(\mathbf{x},t)$ obtained by inserting (2.4) into the material derivatives $\dot{x}(\mathbf{a},t)$ and $\dot{y}(\mathbf{a},t)$. We take f = constant except in Section 10; this is the simplest, so-called 'f-plane', case.

The Eulerian counterpart of (2.2) is

$$\frac{\partial h}{\partial t} + \nabla\cdot(h\dot{\mathbf{x}}) \;=\; \frac{\partial h}{\partial t} + \frac{\partial(h\dot{x})}{\partial x} + \frac{\partial(h\dot{y})}{\partial y} \;=\; 0 \;, \tag{2.7}$$

and the horizontal momentum equations are

$$\ddot{x} + g\frac{\partial h}{\partial x} - \dot{y}f = 0 \;, \quad \ddot{y} + g\frac{\partial h}{\partial y} + \dot{x}f = 0 \;, \tag{2.8}$$

where g is the 'gravity acceleration', representing the combined effect of gravitational and centrifugal accelerations and here taken constant. If we take the Lagrangian view, these equations can be regarded as governing the time-evolution of the functions (2.1), obtaining $h(\mathbf{x},t)$ from (2.4) and (2.5) hence $\partial h/\partial x$ and $\partial h/\partial y$, and hence the time evolution from (2.8). Throughout this article we assume an unbounded physical domain $\mathcal{D} = \mathbb{R}^2$ with boundary conditions of sufficient evanescence at infinity; MR96 discusses boundary conditions for finite domains, which are nontrivial; see also Kushner *et al.* (1998). The shallow-water equations support unbalanced motion in the form of inertia–gravity waves; it is a straightforward exercise to verify this by linearizing about a state of relative rest with h as well as f constant and looking for solutions $\propto \exp(ikx + ily - i\sigma t)$ where k, l, and σ are real constants, giving a solution if the Poincaré dispersion relation

$$\sigma^2 = f^2 + gh(k^2 + l^2) \tag{2.9}$$

is satisfied.

3 Semigeostrophic theory

Semigeostrophic theory results from making the so-called geostrophic momentum approximation in equations (2.8) (Eliassen 1948, 1962). One replaces $(\ddot{x},\, \ddot{y})$ in those equations by the material derivative $(\dot{u}^{\mathrm{G}},\, \dot{v}^{\mathrm{G}})$ of a vector whose components are

$$u^{\mathrm{G}} = -\frac{g}{f}\frac{\partial h}{\partial y} \;, \qquad v^{\mathrm{G}} = \frac{g}{f}\frac{\partial h}{\partial x} \;, \tag{3.1}$$

or in physical (Euclidean) vector notation

$$\mathbf{u}^{\mathrm{G}} \;=\; (u^{\mathrm{G}},\, v^{\mathrm{G}}) \;=\; \frac{g}{f}\,\hat{\mathbf{z}} \times \nabla h \;, \tag{3.2}$$

where $\hat{\mathbf{z}}$ is a unit vertical vector and the cross denotes the vector product. This is called the 'geostrophic velocity' relative to the rotating frame. It is useful because under the conditions of interest in practical meteorology or oceanography it is often fairly close to the particle velocity $(\dot{x},\, \dot{y})$ in a large-scale, coarse-grain view of the atmosphere or ocean, the main exceptions being near the cores of strong atmospheric cyclones. The resulting equations are

$$\dot{u}^{\mathrm{G}} + g\frac{\partial h}{\partial x} - \dot{y}f = 0, \qquad \dot{v}^{\mathrm{G}} + g\frac{\partial h}{\partial y} + \dot{x}f = 0 \;, \tag{3.3}$$

which can again be regarded as soluble in principle for $\dot{x}(a,b,t)$ and $\dot{y}(a,b,t)$, as before, though the use of (2.1)–(2.5) becomes still more intricate than for (2.8). This system (with its evanescent boundary conditions) has the conserved energy

$$\mathsf{H}^{\mathrm{G}} = \mathsf{V} + \int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}^{\mathrm{G}}|^2 \;\mathrm{d}m \tag{3.4}$$

where $\mathcal{D}$ is the physical domain and $\mathrm{d}m$ the mass element (2.3), and where

$$\mathsf{V} = \int_{\mathcal{D}} \tfrac{1}{2}gh \;\mathrm{d}m \;, \tag{3.5}$$

the potential energy of the mass configuration. Here H^{C} and V are scalar-valued functionals of the mass configuration alone. The system also has the material invariant

$$Q_{\mathrm{S}}^{\mathrm{C}} = \frac{1}{h}\left[f + \frac{\partial v^{\mathrm{G}}}{\partial x} - \frac{\partial u^{\mathrm{G}}}{\partial y} + \frac{1}{f}\frac{\partial(u^{\mathrm{G}}, v^{\mathrm{G}})}{\partial(x,\, y)} \right] \;, \tag{3.6}$$

which is the shallow-water counterpart of Hoskins' (1975) PV. (There is a rationale for the notation, to emerge in a moment.) We shall assume that $Q_{\mathrm{S}}^{\mathrm{C}}$ is positive everywhere.

From a fluid-dynamical viewpoint, (3.6) is peculiarly different from (2.6), because of the Jacobian term. Its origin, which is now understood, will be discussed in Section 6 below; but for later reference we may note one interesting fact about (3.6), implicit in the results of Salmon (1988a) and first pointed out explicitly by Roulstone & Sewell (1996), namely that the right-hand side of (3.6) is the result of replacing $(\dot{x},\, \dot{y})$ in (2.6) not by (3.2) but by the velocity field $\mathbf{u}_{\mathrm{S}}^{\mathrm{C}}$ defined by

$$\mathbf{u}_{\mathrm{S}}^{\mathrm{C}} \;=\; (u_{\mathrm{S}}^{\mathrm{C}},\, v_{\mathrm{S}}^{\mathrm{C}}) \;=\; \mathbf{u}^{\mathrm{G}} - \;\tfrac{1}{2}\,\hat{\mathbf{z}} \times f^{-1}\mathbf{u}^{\mathrm{G}} \cdot \nabla \mathbf{u}^{\mathrm{G}} \;. \tag{3.7}$$

That is,

$$Q_{\mathrm{S}}^{\mathrm{C}} \;=\; \frac{1}{h}\left(f+\frac{\partial v_{\mathrm{S}}^{\mathrm{C}}}{\partial x}-\frac{\partial u_{\mathrm{S}}^{\mathrm{C}}}{\partial y}\right). \tag{3.8}$$

The equivalence of (3.8) to (3.6) can be verified in a few lines of manipulation.

We also note for later reference that (3.2) and (3.7) each define a velocity field whenever a mass configuration $h(x,\,y)$ is given. That is what is meant, in general, by a balance condition or 'workless momentum–configuration constraint'. The superscript C stands for 'constraint', implying restriction to a submanifold of phase space. For reasons to be explained at the end of Section 6, the constraint defined by $\mathbf{u}_{\mathrm{S}}^{\mathrm{C}}$, i.e. by the expression on the right of (3.7), will be referred to as the symplectic constraint for semigeostrophic theory, or 'Salmon's constraint'.

The balance conditions (3.2) and (3.7) both have the 'near-local' property that the velocity at a point $(x,\,y)$ in the physical domain $\mathcal{D}$ depends on the local value of h and its gradient only, and not on h values at other points in the domain. This contrasts with the highly accurate, but more complicated, balance conditions used in recent studies testing the limits of accuracy of non-Hamiltonian balanced models based on PV inversion (Norton 1988, McIntyre & Norton 2000, hereafter 'MN', McIntyre 2001, Mohebalhojeh & Dritschel 2001). Such highly accurate balance conditions, defined for instance by equations (3.6*a*–*i*) or (5.4*a*–*b*) of MN00, are necessarily nonlocal. That is, the velocity at each point of the physical domain is a functional of $h(\cdot)$, the entire mass distribution. The dependence on h values at distant points of the domain evanesces exponentially, in typical cases, over distances of order

$$L_{\mathrm{R}} = (gh_0)^{1/2}/f\ , \tag{3.9}$$

a lengthscale of the problem usually called the Rossby length, Rossby radius, or Rossby radius of deformation.

There is one more peculiarity to be noted for later reference. Under 'the conditions of interest in practical meteorology or oceanography' the Rossby number ϵ can often be considered small. Then the last term on the right of (3.7) is $O(\epsilon)$ relative to the first. The last term would look like a small correction to make the balance condition more accurate except that it has the wrong sign (minus not plus) and the wrong numerical coefficient ($\frac{1}{2}$ not 1). We return to this point in Sections 8ff., where we introduce the '$\sqrt{3}$ model'. That model, so-called for reasons to emerge in Section 9, via (9.8), is based on a variant of (3.7) that has the *right* coefficient in the correction term, i.e., has $\mathbf{u}_{\mathrm{S}}^{\mathrm{C}}$ replaced by

$$\mathbf{u}_{\sqrt{3}}^{\mathrm{C}} \;=\; \mathbf{u}^{\mathrm{G}} \,+\, \hat{\mathbf{z}}\times f^{-1}\mathbf{u}^{\mathrm{G}}\cdot\nabla\mathbf{u}^{\mathrm{G}}\ . \tag{3.10}$$

The rotational part of this velocity field is asymptotically correct. In other words, the corresponding vorticity is correctly balanced to two orders in ϵ

(e.g. Snyder *et al.* 1991, Craig 1993). This can be verified by standard scaling arguments applied to the divergence equation of the parent dynamics, i.e. to the divergence of (2.8). It is not yet known precisely what impact this has on overall accuracy, i.e., what in practical terms the overall accuracy of the $\sqrt{3}$ model might be. But it seems likely that the accuracy will be much better than that of semigeostrophic theory, especially since the remaining, divergent part of the velocity field is $O(\epsilon)$ smaller than the relative vorticity. Let us return, however, to semigeostrophic theory in order to review some of its remarkable transformation properties.

4 Transformations of the semigeostrophic equations

Hoskins (1975) discovered that (3.3) is simplified, very remarkably indeed, by the following transformation of coordinates:

$$X \;=\; x + \frac{\partial \phi}{\partial x}\;, \qquad Y \;=\; y + \frac{\partial \phi}{\partial y}\;, \tag{4.1}$$

where for convenience we have written

$$\phi(x,y,t) = \frac{g}{f^2} h(x,y,t)\;, \tag{4.2}$$

with physical dimensions of length squared, cf. ${L_\mathrm{R}}^2$. Because of (3.1), which now reads

$$u^\mathrm{G} = -f\frac{\partial \phi}{\partial y}\;, \quad v^\mathrm{G} = f\frac{\partial \phi}{\partial x}\;, \tag{4.3}$$

the transformation (4.1) is often called the geostrophic momentum transformation; and (X,Y) are often called geostrophic coordinates because, when f is constant,

$$\dot{X} = u^\mathrm{G}\;, \quad \dot{Y} = v^\mathrm{G}\;, \tag{4.4}$$

from (3.3) and (4.3). When $h(x,y,t)$, equivalently $\phi(x,y,t)$, is regarded as a known smooth function of x and y with t fixed — i.e., when we take a snapshot view of the evolving system, assumed well-behaved — then (4.1) specifies a transformation of the form

$$X = X(x,y,t), \quad Y = Y(x,y,t) \qquad (t \text{ fixed}) \tag{4.5}$$

which, at each t, is assumed to have an inverse

$$x = x(X,Y,t), \quad y = y(X,Y,t)\;, \tag{4.6}$$

requiring that

$$\frac{\partial(X,Y)}{\partial(x,y)} \neq 0,\; \infty \tag{4.7}$$

everywhere or almost everywhere (Sewell, this volume). To get the maximum simplification we can further transform in various ways. For instance, in place of $h(x,y,t)$ or $\phi(x,y,t)$ we can use the potential

$$P(x,y,t) \;=\; \phi(x,y,t) \;+\; \tfrac{1}{2}(x^2+y^2) \;, \tag{4.8}$$

or its Legendre dual

$$R(X,Y,t) = Xx + Yy - P \;. \tag{4.9}$$

The transformation

$$\{x,y,P\} \mapsto \{X,Y,R\} \tag{4.10}$$

is a Legendre transformation because, from (4.1) and (4.8),

$$X = \frac{\partial P}{\partial x} \;, \quad Y = \frac{\partial P}{\partial y} \;. \tag{4.11}$$

Its inverse is (4.9) together with

$$x = \frac{\partial R}{\partial X} \;, \quad y = \frac{\partial R}{\partial Y} \;. \tag{4.12}$$

Defining (cf. (4.8))

$$\Phi(X,Y,t) \;=\; -R(X,Y,t) \;+\; \tfrac{1}{2}(X^2+Y^2) \;, \tag{4.13}$$

we see first from (4.1), (4.12) and (4.13) that

$$\frac{\partial \Phi}{\partial X} = \frac{\partial \phi}{\partial x} = X - x \qquad \text{and} \qquad \frac{\partial \Phi}{\partial Y} = \frac{\partial \phi}{\partial y} = Y - y \;, \tag{4.14}$$

second from (4.8), (4.9) and (4.13) that

$$\Phi = \phi + \tfrac{1}{2}(X-x)^2 + \tfrac{1}{2}(Y-y)^2 \;, \tag{4.15}$$

and finally from (4.3), (4.4) and (4.14) that

$$\boxed{\dot{X} \;=\; -f\frac{\partial \Phi}{\partial Y} \;, \qquad \dot{Y} \;=\; f\frac{\partial \Phi}{\partial X} \;.} \tag{4.16}$$

Recall that the dots signify the material derivative; notice furthermore that $\Phi(X,Y,t)$ is a field to be solved for, not a prescribed function or functional. After multiplication by f (which is constant), Φ is a streamfunction or quasi-Hamiltonian for the particle motion viewed in (X,Y) space and not, of course, the dynamical Hamiltonian in the infinite-dimensional phase space of the problem — an entirely different mathematical object to be discussed in Section 8, and in this case given by (3.4), a prescribed functional of the mass configuration. We may refer to (4.16) as the 'streamfunction property' of semigeostrophic theory.

Semigeostrophic evolution can now be described in a remarkably simple way, in terms of PV advection and inversion (e.g. Hoskins *et al.* 1985). Note first that the semigeostrophic PV (3.6) can be written in the Jacobian form discovered by Hoskins (1975),

$$Q_{\rm S}^{\rm C} = \frac{g}{f\phi}\,\frac{\partial(X,Y)}{\partial(x,y)} > 0\ , \tag{4.17}$$

This can be verified by substituting (4.1) and (4.3) into (4.17) to recover (3.6). We take $Q_{\rm S}^{\rm C} > 0$, consistently with (4.7). From (4.11) and (4.17), $Q_{\rm S}^{\rm C}$ may also be written in the form noted by Cullen & Purser (1984), in terms of the Hessian determinant of P with respect to (x,y):

$$Q_{\rm S}^{\rm C} = \frac{g}{f\phi}{\rm hess}_{xy}(P) = \frac{g}{f\phi}\begin{vmatrix} \dfrac{\partial^2 P}{\partial x^2} & \dfrac{\partial^2 P}{\partial x\partial y} \\ \dfrac{\partial^2 P}{\partial y\partial x} & \dfrac{\partial^2 P}{\partial y^2}\end{vmatrix}\ . \tag{4.18}$$

We have

$$Q_{\rm S}^{\rm C} = \frac{g}{f\phi}\frac{\partial(X,Y)}{\partial(x,y)} = \frac{g}{f\phi}{\rm hess}_{xy}(P) = \frac{g}{f\phi}{\rm hess}_{xy}\{\phi + \tfrac12(x^2+y^2)\}\ . \tag{4.19}$$

Because of the Legendre duality and the properties of Jacobians, we can equally well substitute (4.12) into the reciprocal of (4.17); thus

$$\frac{1}{Q_{\rm S}^{\rm C}} \;=\; \frac{f\phi}{g}\frac{\partial(x,y)}{\partial(X,Y)} \;=\; \frac{f\phi}{g}{\rm hess}_{XY}(R) \;=\; \frac{f\phi}{g}{\rm hess}_{XY}\left(\Phi - \tfrac12(X^2+Y^2)\right)\ , \tag{4.20}$$

noting (4.13). Equivalently, from (4.14) and (4.15),

$$\frac{1}{Q_{\rm S}^{\rm C}} \;=\; \frac{f}{g}\left[\Phi - \tfrac12\left\{\left(\frac{\partial\Phi}{\partial X}\right)^2 + \left(\frac{\partial\Phi}{\partial Y}\right)^2\right\}\right]{\rm hess}_{XY}\left(\Phi - \tfrac12(X^2+Y^2)\right)\ . \tag{4.21}$$

Note further that the quantity in square brackets in (4.21), being equal to ϕ by (4.14) and (4.15), is a positive quantity, given that h is positive everywhere.

If the PV field $Q_{\rm S}^{\rm C} > 0$ is prescribed as a function of X and Y, then (4.21) is an elliptic Monge–Ampère equation for Φ in (X,Y) space, as is, also, (4.19) for ϕ in (x,y) space. For both can be seen by inspection to be special cases of the general 2-dimensional Monge–Ampère equation

$$A + B\rho + 2C\tau + D\varsigma + E(\rho\varsigma - \tau^2) = 0 \tag{4.22}$$

where in (x, y) space we define $p, q, \rho, \varsigma, \tau$ by

$$p = \frac{\partial \phi}{\partial x}, \quad q = \frac{\partial \phi}{\partial y}, \quad \rho = \frac{\partial^2 \phi}{\partial x^2}, \quad \varsigma = \frac{\partial^2 \phi}{\partial y^2}, \quad \tau = \frac{\partial^2 \phi}{\partial x \partial y} \tag{4.23}$$

and where A, B, C, D, E are given functions of (x, y, ϕ, p, q). In (X, Y) space we replace x, y and ϕ by X, Y and Φ. Ellipticity holds if $BD - C^2 - AE > 0$ (e.g. Lychagin *et al.* 1993). For (4.19), after multiplying by $g^{-1} f \phi$, we have $A = 1 - g^{-1} f \phi Q_{\mathrm{S}}^{\mathrm{C}}$, $B = D = 1$, $C = 0$, and $E = 1$, whence $BD - C^2 - AE = 1 - 1 + g^{-1} f \phi Q_{\mathrm{S}}^{\mathrm{C}} > 0$, by hypothesis, and similarly for (4.21). Hence both are elliptic.

Once Φ is obtained, (4.16) gives the particle velocities in (X, Y) space — that is, it gives the velocities of the images, in (X, Y) space, of the material particle positions in (x, y) space. We can now exploit the material conservation of $Q_{\mathrm{S}}^{\mathrm{C}}$. Continuing to regard $Q_{\mathrm{S}}^{\mathrm{C}}$ as a function of X and Y, we can, heuristically speaking, timestep either

$$\left(\frac{\partial}{\partial t} + \dot{X} \frac{\partial}{\partial X} + \dot{Y} \frac{\partial}{\partial Y} \right) Q_{\mathrm{S}}^{\mathrm{C}} = 0 \text{ or } \left(\frac{\partial}{\partial t} + \dot{X} \frac{\partial}{\partial X} + \dot{Y} \frac{\partial}{\partial Y} \right) \left(\frac{1}{Q_{\mathrm{S}}^{\mathrm{C}}} \right) = 0 \tag{4.24}$$

in order to advect $Q_{\mathrm{S}}^{\mathrm{C}}$ or $1/Q_{\mathrm{S}}^{\mathrm{C}}$. We can then invert by solving (4.21) for Φ, then use (4.16) to obtain the advection velocity, then re-advect, and so on. At any time t, the mass configuration and locations of the material particles in physical space can be obtained from (4.14).

Thus equations (4.16), (4.21) and (4.24), with (4.14), express in a remarkably simple way the standard idea of evolving the balanced motion by PV advection and PV inversion. Moreover, they do so in a form that can be shown to be mathematically well-behaved using the theory of Monge–Ampère equations (Cullen & Douglas 1998; see also Cullen in Volume 1 and Douglas in Volume 1), and generalizable to cases of non-smooth ϕ or Φ that can model atmospheric fronts (Chynoweth & Sewell 1989, 1991; see also Sewell in this volume). Note that the assumption that the PV is positive, $Q_{\mathrm{S}}^{\mathrm{C}} > 0$ (with $h > 0$, $\phi > 0$), corresponds to convexity of the graphs of $P(x, y)$ and $R(X, Y)$.

The Legendre structure associated with the intermediate sets of variables $\{x, y, P\}$ and $\{X, Y, R\}$ and expressed by (4.9)–(4.12) is special to, and an essential part of, semigeostrophic theory as the term is used here. Salmon and others have, however, used the term 'semigeostrophic theory' in a wider sense, to include a class of Hamiltonian balanced models of comparable accuracy in Rossby number ϵ but whose members may lack the Legendre structure. Purser (this volume) points out that these other models may permit the emergence of certain topological defects that can give trouble when the models are discretized for numerical integration. Conversely, he shows that the exact Legendre structure expressed by (4.9)–(4.12) makes semigeostrophic theory, as the term is used here, immune from such trouble. Sewell (this volume) reminds us that the Legendre structure also allows semigeostrophic theory to tolerate isolated zeros or infinities in the Jacobian (4.7).

5 Contact structure of semigeostrophic theory

The transformation that so remarkably simplifies (3.3) into (4.16), (4.21) and (4.24) belongs to the general class called contact transformations or contactomorphisms (see e.g. Carathéodory 1982; Sewell, this volume), related in turn to what are called canonical transformations or symplectomorphisms. For an authoritative overview of these topics and of the associated ideas of contact and symplectic structure, and of the interesting state of research on them today, the reader may consult the monograph by McDuff & Salamon (1998), hereafter 'MS98'.

Because we are interested in generalizations to models of higher accuracy than semigeostrophic theory, we digress briefly, in this section, to recall the modern definitions of contact structure and contact transformation, noting also the distinction between contact and canonical transformation (e.g. Sewell & Roulstone 1993, 1994). The occurrence of contact structure in semigeostrophic theory was first noted by Blumen (1981), and was used by Purser (1993) to generalize semigeostrophic theory to hemispherical geometry.

The clearest way to describe what is involved is to use the language of the exterior calculus. In brief — for more detail see e.g. Misner *et al.* (1973) or Schutz (1980) — the exterior calculus applies in a flat or curved space (manifold) of arbitrary dimension, possibly infinite, and has the same information content as what used to be called the calculus of antisymmetric covariant tensors and their contractions with contravariant vectors (now called simply *vectors*). All the information is metric-independent: no Christoffel symbols appear, and any coordinates $\mathbf{x}$ can be used, orthogonal or non-orthogonal. The exterior calculus has a beautiful geometric interpretation in terms of metric-independent properties such as tangency and intersection. An *exterior derivative* operator d is defined, which, when applied to a scalar field $\phi(\mathbf{x})$, produces a first-degree (first-rank) differential form or *1-form* $\mathrm{d}\phi$ having the same information content as the gradient $\nabla\phi = (\partial\phi/\partial x_1,\, \partial\phi/\partial x_2, \dots)$, a covariant vector. The information content is the same because $\mathrm{d}\phi = (\partial\phi/\partial x_1)\,\mathrm{d}x_1 + (\partial\phi/\partial x_2)\,\mathrm{d}x_2 + \cdots$ by the chain rule. This expresses the 1-form $\mathrm{d}\phi$ in terms of $\mathrm{d}x_1,\ \mathrm{d}x_2, \dots,$ the *basis 1-forms* associated with the coordinates $\mathbf{x}$. An associative, distributive, anti-commutative *exterior product* or *wedge product* is defined to have the same information content as an antisymmetrized outer or tensor product. For instance the wedge product of $\mathrm{d}\phi$ and $\mathrm{d}\psi$ is defined to be the *2-form* $\mathrm{d}\phi\wedge\mathrm{d}\psi = -\mathrm{d}\psi\wedge\mathrm{d}\phi = \mathrm{d}\phi\otimes\mathrm{d}\psi - \mathrm{d}\psi\otimes\mathrm{d}\phi$ (no factor $\frac{1}{2}$). Antisymmetrization is understood whenever d is applied to a differential form, so that, in particular, $\mathrm{d}\,\mathrm{d}\phi = 0$, generalizing $\operatorname{curl}\operatorname{grad}\phi = 0$ to arbitrary dimensions (Poincaré's lemma), and $\mathrm{d}(\phi\,\mathrm{d}\psi) = \mathrm{d}\phi\wedge\mathrm{d}\psi = -\mathrm{d}\psi\wedge\mathrm{d}\phi$. In other words, d corresponds to an antisymmetrized gradient operator, and $\mathrm{d}\phi \wedge \mathrm{d}\psi$ to the second-rank covariant tensor $\nabla\phi \otimes \nabla\psi - \nabla\psi \otimes \nabla\phi$. The same antisymmetry shows up in Jacobian determinants and their minors; for instance, if $\mathbf{x}$ is 2-dimensional with $\mathbf{x} = (x,\, y)$, then

$$\mathrm{d}\phi \wedge \mathrm{d}\psi \;=\; \frac{\partial(\phi,\psi)}{\partial(x,y)}\,\mathrm{d}x \wedge \mathrm{d}y\ . \tag{5.1}$$

Consider now the 4-dimensional space $\mathbb{R}^4$ with coordinates $\{x, y, p, q\}$, say, and the 5-dimensional space $\mathbb{R}^5$ with coordinates $\{x, y, \phi, p, q\}$, say. The latter space when equipped with the 1-form

$$\theta = \mathrm{d}\phi - p\,\mathrm{d}x - q\,\mathrm{d}y\ , \tag{5.2}$$

called the Cartan 1-form, or standard contact 1-form, is said to have *contact structure*. The structure may be visualized as a field of hyperplanes, 4-dimensional hyperplanes in our case. At each point P of the space $\mathbb{R}^5$, the 1-form θ defines an associated hyperplane, spanned by those vectors that annihilate θ, i.e. produce zero when contracted with θ. The hyperplane can be (though need not be) thought of as living in the neighbourhood of P. (More precisely, the 1-form θ lives in the 'cotangent space' at P, and the associated hyperplane lives in the 'tangent space' at P; e.g. Schutz pp. 36, 53; MS98, p. 106.) The hyperplane changes its orientation, in general, as the point P moves through the space $\mathbb{R}^5$. If we watch, for instance, the intersection of the hyperplane with the 3-dimensional subspace spanned by $\{x, \phi, p\}$, then we see an ordinary plane tilting about an axis parallel to the p-axis. The plane tilts toward the ϕ-axis whenever the point P moves in any way such that the coordinate p increases, and vice versa. The field of hyperplanes defined by θ can be shown to be *nonintegrable* or *nondegenerate* in the sense that the hyperplanes define no integral hypersurfaces: if P is imagined to move along some curve whose line elements always lie in the local hyperplane through P defined by θ, then the set of all such curves emanating from a given point P_0 traces out a quasi-helical structure that cannot be contained within any single hypersurface through P_0.

Now the map

$$\mathcal{F}_{\mathcal{C}} : \{x, y, \phi, p, q\} \mapsto \{x', y', \phi', p', q'\} \tag{5.3}$$

from $\mathbb{R}^5$ to $\mathbb{R}^5$ is called a *contact transformation* or *contactomorphism* if it preserves θ up to a (nonvanishing) multiplicative factor β, that is, if

$$\mathrm{d}\phi' - p'\,\mathrm{d}x' - q'\,\mathrm{d}y' \;=\; \beta(x, y, \phi, p, q)(\mathrm{d}\phi - p\,\mathrm{d}x - q\,\mathrm{d}y)\ . \tag{5.4}$$

Similarly, the restriction of $\mathcal{F}_{\mathcal{C}}$ to $\mathbb{R}^4$, taking $\{x, y, p, q\} \mapsto \{x', y', p', q'\}$, is called a *canonical transformation* or *symplectomorphism* if it preserves the 2-form Ω obtained by taking the exterior derivative of (5.2):

$$\Omega \;=\; \mathrm{d}\theta \;=\; -\,\mathrm{d}p \wedge \mathrm{d}x - \mathrm{d}q \wedge \mathrm{d}y \;=\; \mathrm{d}x \wedge \mathrm{d}p + \mathrm{d}y \wedge \mathrm{d}q\ . \tag{5.5}$$

That is, $\{x, y, p, q\} \mapsto \{x', y', p', q'\}$ is called canonical if

$$\mathrm{d}x' \wedge \mathrm{d}p' + \mathrm{d}y' \wedge \mathrm{d}q' = \mathrm{d}x \wedge \mathrm{d}p + \mathrm{d}y \wedge \mathrm{d}q \ . \tag{5.6}$$

Clearly (5.4) implies (5.6) if and only if $\beta = 1$. Thus, loosely speaking, a canonical transformation is a special case of a contact transformation, both because of the restriction to a smaller space and because of the restriction to $\beta = 1$. We shall see in a moment that each of the three steps in the sequence of transformations described in Section 4 is a case of (5.4) with $\beta = 1$, and therefore, in the loose sense just described, is both contact and canonical.

The 2-form Ω has a nondegeneracy property corresponding to the nondegeneracy property of θ. Nondegeneracy means that Ω gives a nonvanishing result when contracted with almost any pair of nonparallel vectors (in the older terminology, contravariant vectors). It is 'almost any pair' rather than 'any pair' of nonparallel (nonzero) vectors because there is an infinitesimally small set of special pairs of such vectors that do make the contraction vanish, as is clear by continuity and antisymmetry. Exchanging the two vectors changes the sign of the result, so rotating their plane must take the result through zero. This is related to the existence of so-called *Lagrangian submanifolds* having half the dimensions, in this case 2 dimensions, on which contractions with Ω always vanish (MS98).† Ω defines what is called a symplectic structure in $\mathbb{R}^4$, for which $\{x, y, p, q\}$ and $\{x', y', p', q'\}$ are sets of *canonical coordinates.* These are called canonical because of the simplicity of the 2-forms on the right and left of (5.6), having only two terms as compared with the six terms, with variable coefficients, that would result from applying an arbitrary coordinate transformation. (The reader is warned that, when we specialize to the case of Section 4, the symplectic structure defined by Ω will not be the same symplectic structure as that of the parent Hamiltonian dynamics, not only because of the different dimensionality, $\mathbb{R}^4$ rather than $\mathbb{R}^\infty$, but also because (p, q) will be equated to $(\partial\phi/\partial x,\ \partial\phi/\partial y)$ and not to the components of (3.7), (3.10), (4.3), or any similar expression related to the dynamical canonical momenta.)

When $\beta = 1$, the classical generating-function formalism is available (e.g. Carathéodory 1982, Sections 97–109; Sewell & Roulstone 1993; MS98 chapter 9). This guarantees that, for instance, a transformation defined by a generating function of the form $S(x, y, x', y')$, namely the transformation implicitly defined by

$$\frac{\partial S}{\partial x} = p, \quad \frac{\partial S}{\partial y} = q, \quad \frac{\partial S}{\partial x'} = -p', \quad \frac{\partial S}{\partial y'} = -q' \ , \tag{5.7}$$

†More precisely, nondegeneracy can be defined by saying that whenever the direction of one of the vectors is chosen arbitrarily, then the other vector can always be chosen to give a nonvanishing contraction, i.e. to be transverse to any Lagrangian submanifold containing the first vector. In our simple case, the Lagrangian submanifolds defined by Ω include the 2-dimensional subspaces of $\mathbb{R}^4 = \{x, y, p, q\}$ spanned by any two nonvanishing vectors of the form $(a, 0, c, 0)$ and $(0, b, 0, d)$. A less trivial example will be encountered in the paragraph below (5.12).

is canonical, as substitution into (5.6) will immediately verify. (MS98 call this $S(x, y, x', y')$ a generating function of 'type S', and discuss related global existence questions of interest in current research.) Furthermore, if we take

$$\phi' = \phi - S \tag{5.8}$$

then (5.7) with (5.8) defines a contact transformation, as can be seen at once by substitution into (5.4) with $\beta = 1$. The process of appending (5.8) to (5.7) while extending $\mathbb{R}^4$ into $\mathbb{R}^5$ is called 'lifting' a canonical transformation into a contact transformation, resulting in the 'contactification' of the symplectic structure defined by Ω. (The reader is warned that what is called 'symplectification' of a contact structure is *not* the reverse process; in our case it would give rise to a symplectic structure in $\mathbb{R}^6$ rather than $\mathbb{R}^4$.)

For present purposes we are especially interested in the case

$$S(x, y, x', y') \;=\; -\tfrac{1}{2}(x' - x)^2 \;-\; \tfrac{1}{2}(y' - y)^2 \;, \tag{5.9}$$

for which (5.7) gives

$$x' = x + p \;, \quad y' = y + q \;, \quad p' = p \;, \quad q' = q \;. \tag{5.10}$$

If we are given a (2-dimensional) hypersurface $\mathcal{G}$ within $\mathbb{R}^5$ defined, at given fixed t, by the three relations

$$\phi = \phi(x, y, t) \;, \quad p = \partial\phi/\partial x \;, \quad q = \partial\phi/\partial y \;, \tag{5.11}$$

and its image $\mathcal{G}'$ defined by (5.8) with

$$\phi' = \phi'(x', y', t) \;, \quad p' = \partial\phi'/\partial x' \;, \quad q' = \partial\phi'/\partial y' \;, \tag{5.12}$$

then the particular contact transformation (5.8)–(5.10), restricted to the graph of ϕ in the sense defined by (5.11), i.e. restricted to the hypersurface $\mathcal{G}$, is just the geostrophic momentum transformation (4.1) together with (4.14) and (4.15) (Sewell & Roulstone 1994, Theorem 11). Here we indentify (x', y') with the previous (X, Y), and $\phi'(x', y', t)$ with the previous $\Phi(X, Y, t)$ (the time t still being held fixed), and (p', q') with $(\partial\phi'/\partial x',\ \partial\phi'/\partial y')$, i.e. with the previous $(\partial\Phi/\partial X,\ \partial\Phi/\partial Y)$.

Through (5.2) and (5.11), each point on the hypersurface $\mathcal{G}$ is thus associated with a 1-form θ defining those hyperplanes tangent to $\mathcal{G}$ that are also parallel to the p and q axes. The projection of this picture into $\mathbb{R}^3 = \{x, y, \phi\}$ is the surface $\phi = \phi(x, y)$ together with its tangent planes. A contact transformation preserves this geometric structure, i.e., preserves the tangency of θ to $\mathcal{G}$. We assume here that the derivatives $\partial\phi/\partial x$ and $\partial\phi/\partial y$ exist, so that the notion of tangency makes sense. Notice also that the projection of the same picture into $\mathbb{R}^4 = \{x, y, p, q\}$ gives us another example of a Lagrangian submanifold. Because of the antisymmetry of exterior products and the symmetry of mixed partials, $\partial^2\phi/\partial x\partial y = \partial^2\phi/\partial y\partial x$, the 2-form Ω defined by (5.5) vanishes when restricted to the graph of ϕ.

If we exchange x with p, y with q, x' with p' and y' with q' in the contact transformation (5.8)–(5.10), then we obtain another contact transformation with generating function of the form $\tilde{S} = \tilde{S}(p, q, p', q')$, i.e., a transformation implicitly defined by

$$\frac{\partial\tilde{S}}{\partial p} = x, \quad \frac{\partial\tilde{S}}{\partial q} = y, \quad \frac{\partial\tilde{S}}{\partial p'} = -x', \quad \frac{\partial\tilde{S}}{\partial q'} = -y' \ . \tag{5.13}$$

This is readily verified to be just the transformation that takes ϕ into P via (4.8) when we identify P with ϕ' and (p', q') with $(\partial\phi'/\partial x',\ \partial\phi'/\partial y')$, and take $\tilde{S} = -\frac{1}{2}(p-p')^2 - \frac{1}{2}(q-q')^2$. Similarly, the transformation that takes $-R$ into Φ via (4.13) is another contact transformation. Geometrically speaking, one is simply adding a parabola and its sloping tangent planes to the original surface and its tangent planes, which trivially preserves tangency.

To summarize, then, using the notation of Section 4, the foregoing verifies (a) that the transformation taking

$$(x,\ y,\ \phi,\ \partial\phi/\partial x,\ \partial\phi/\partial y) \mapsto (X,\ Y,\ \Phi, \partial\Phi/\partial X,\ \partial\Phi/\partial Y) \tag{5.14}$$

is a contact transformation, and (b) that the first and third steps in that transformation, respectively

$$(x,\ y,\ \phi,\ \partial\phi/\partial x,\ \partial\phi/\partial y) \mapsto (x,\ y,\ P,\ \partial P/\partial x,\ \partial P/\partial y) \tag{5.15}$$

and

$$(X,\ Y,\ -R,\ -\partial R/\partial x,\ -\partial R/\partial y) \mapsto (X,\ Y,\ \Phi, \partial\Phi/\partial X,\ \partial\Phi/\partial Y)\ , \tag{5.16}$$

are also contact transformations. It follows that the intermediate step

$$(x,\ y,\ P,\ \partial P/\partial x,\ \partial P/\partial y) \mapsto (X,\ Y,\ -R,\ -\partial R/\partial X,\ -\partial R/\partial Y) \tag{5.17}$$

is itself, also, a contact transformation as well as being a Legendre transformation.

The demonstration just given that the transformation between x, y and X, Y is part of a contact transformation justifies the assertion in Section 1 that two of the velocity fields involved in semigeostrophic theory 'are interrelated by a certain contact transformation ... '. The two velocity fields are the particle velocity $\dot{\mathbf{x}} = \mathbf{u}^{\mathrm{P}}$, say, and the geostrophic velocity $\dot{\mathbf{X}} = \mathbf{u}^{\mathrm{G}}$. This last relation follows from (4.1), (4.3), (4.14), and (4.16).

Note also the explicit invertibility of the overall transformation, enhancing the usefulness of (4.16). We can transform in the direction $(X,\ Y,\ \Phi,\ \partial\Phi/\partial X,\ \partial\Phi/\partial Y) \mapsto (x,\ y,\ \phi,\ \partial\phi/\partial x,\ \partial\phi/\partial y)$ just as easily in the opposite direction $(x,\ y,\ \phi,\ \partial\phi/\partial x,\ \partial\phi/\partial y) \mapsto (X,\ Y,\ \Phi,\ \partial\Phi/\partial X,\ \partial\Phi/\partial Y)$, in part because of the symmetry and simple form of the generating function (5.9), which in the current notation, the notation of Section 4, reads

$$S(x, y, X, Y) \ = \ -\tfrac{1}{2}(X-x)^2 \ - \ \tfrac{1}{2}(Y-y)^2\ . \tag{5.18}$$

6 Constraints and velocity-splitting

We now look beyond semigeostrophic theory and its Legendre and contact structure toward the more general, and in some cases more accurate, balanced models referred to in Sections 1,3. These can be constructed systematically from a given 'parent dynamics' that admits unbalanced motion, such as the shallow-water primitive equations, or the stratified primitive equations used in numerical weather prediction, by Salmon's method of applying a balance condition or workless momentum–configuration constraint to the flow in the parent phase space (Salmon 1988a, equation (4.3)). We regard the parent dynamics as 'exact' in the sense of setting the standard against which accuracy will be judged. Therefore, a constraint that is as accurate as possible is one that interferes with the phase-space flow as little as possible. Because the constraint is a functional relation tying the velocity field to the mass configuration $h(\mathbf{x}, t)$, at any given t, as already illustrated in (3.2) and in (3.7)ff., we write

$$\mathbf{u} \;=\; \mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot)) \tag{6.1}$$

to denote the general form of the constraint (suppressing reference to t). Thus the symbol $\mathbf{u}^{\mathrm{C}}$ represents both a field, i.e. a function of $\mathbf{x}$, and also a mass-configuration functional. When one or other aspect needs emphasis we may speak of the '$\mathbf{u}^{\mathrm{C}}$ field' or '$\mathbf{u}^{\mathrm{C}}$ functional' respectively. The dependence on the mass configuration has to be nonlocal, as pointed out above (3.9), if we want the greatest possible accuracy.

Now the Hamiltonian balanced model that results from applying the constraint is always in a certain sense 'schizophrenic', in that it naturally and inevitably has not one, but two different velocity fields. In such a model there is no such thing as 'the' velocity field. This has caused confusion in the past, even though the point is perhaps obvious once recognized. We find it mnemonically useful to say that applying the constraint 'splits' the parent velocity field $\mathbf{u}$ into two different velocity fields. The first of these is $\mathbf{u}^{\mathrm{C}}$ itself; and $\mathbf{u}^{\mathrm{C}}$ cannot be ignored because it enters naturally into the balanced model's conservation relations, as will become clear shortly. For instance the formula for the balanced model's conserved PV has the same appearance as (3.8), but is valid for a completely general $\mathbf{u}^{\mathrm{C}}$ field:

$$Q^{\mathrm{C}} = \frac{1}{h}\left(f + \frac{\partial v^{\mathrm{C}}}{\partial x} - \frac{\partial u^{\mathrm{C}}}{\partial y}\right). \tag{6.2}$$

The second velocity field is the particle velocity $\dot{\mathbf{x}} = \mathbf{u}^{\mathrm{P}}$, which differs from $\mathbf{u}^{\mathrm{C}}$ for the reasons to be noted next. We denote the difference $\mathbf{u}^{\mathrm{P}} - \mathbf{u}^{\mathrm{C}}$ by $\mathbf{u}^{\mathrm{S}}$ and call it the 'velocity-split'. It is a natural generalization of the 'ageostrophic velocity' discussed in meteorological textbooks, and has the useful property of being reference-frame-independent. It is indifferent, in particular, to whether one uses a rotating frame as hitherto, or an inertial frame, as will be found convenient below.

The velocity-split $\mathbf{u}^{\mathrm{S}}$ is generically nonzero because of its relation to a certain mass-configuration functional $\mathbf{R}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$ that directly measures the constraint's interference with the parent phase-space flow. We may call $\mathbf{R}^{\mathrm{C}}$ the residual, or unbalanced, contribution to the gravitational and pressure-gradient forces per unit mass. It too is indifferent to the choice of rotation rate for the frame of reference, and is defined by

$$\mathbf{R}^{\mathrm{C}} = \mathbf{F} - \mathbf{A}^{\mathrm{C}} \ , \tag{6.3}$$

where $\mathbf{A}^{\mathrm{C}}$ will be called the absolute 'constraint acceleration' and where $\mathbf{F}$ is the net gravitational and pressure-gradient force per unit mass defined by the parent dynamics, for the given mass configuration. $\mathbf{F}$ is equal, by definition, to the absolute material acceleration that would occur if the constraint were suddenly removed.

The constraint acceleration $\mathbf{A}^{\mathrm{C}}$ is defined to be the material rate of change of $\mathbf{u}^{\mathrm{C}}$ evaluated from a fictitiously-evolving mass configuration that coincides, at the instant considered, with the given configuration, but evolves as if all particles moved with velocity $\mathbf{u}^{\mathrm{C}}$ (viewed in the inertial frame). Thus $\mathbf{R}^{\mathrm{C}}$ would be identically zero if that fiction were fact, i.e., if the constraint did not interfere with the parent dynamics. The superscript C will be used throughout to indicate a functional dependence on the mass configuration that is known when the $\mathbf{u}^{\mathrm{C}}$ functional is given. The field $\mathbf{A}^{\mathrm{C}}$ qualifies as having such functional dependence because, as a moment's reflection makes clear, $\mathbf{A}^{\mathrm{C}}$ can be expressed solely in terms of $\mathbf{u}^{\mathrm{C}}$ and its functional derivative with respect to $h(\cdot)$. This will be made explicit in (8.11) below. The functional derivative enters because of the need to represent the fictitious rate of change of the mass configuration that would occur if particles moved with velocity $\mathbf{u}^{\mathrm{C}}$.

The relation between $\mathbf{u}^{\mathrm{S}}$ and $\mathbf{R}^{\mathrm{C}}$ will be shown to take the simple generic form

$$\boxed{\sum_{j=1}^{2} \int_{\mathcal{D}} \Omega_{ij}^{\mathrm{C}}(\mathbf{x}, \mathbf{x}')\; u_j^{\mathrm{S}}(\mathbf{x}')\; \mathrm{d}m(\mathbf{x}') \;= R_i^{\mathrm{C}}(\mathbf{x}) \qquad (i = 1, 2) \ ,} \tag{6.4}$$

in suffix notation with $(x, y) = (x_1, x_2)$, etc. The kernel $\Omega_{ij}^{\mathrm{C}}(\mathbf{x}, \mathbf{x}')$ is known in terms of $\mathbf{u}^{\mathrm{C}}$ and the mass configuration, as indicated by the superscript C , and again involves the functional derivative of $\mathbf{u}^{\mathrm{C}}$ with respect to $h(\cdot)$. The kernel will be defined in Section 8. The kernel is antisymmetric in the sense that $\Omega_{ij}^{\mathrm{C}}(\mathbf{x}, \mathbf{x}') = -\Omega_{ji}^{\mathrm{C}}(\mathbf{x}', \mathbf{x})$. We call (6.4) the 'splitting equation'. It generalizes the elliptic partial differential equation found by Salmon (1985) for the ageostrophic velocity field in the case $\mathbf{u}^{\mathrm{C}} \equiv \mathbf{u}^{\mathrm{G}}$, i.e. in the case of the Hamiltonian balanced model now called L_1 dynamics. Salmon's equation is equivalent to

$$\left(\nabla^2 - K^2\right) \mathbf{U}^{\mathrm{S}} \;=\; -\frac{f}{g}\, \hat{\mathbf{z}} \times (\mathbf{u}^{\mathrm{G}} \cdot \nabla \mathbf{u}^{\mathrm{G}}) \tag{6.5}$$

under the conditions of Section 2, where $\mathbf{U}^{\mathrm{S}}(\mathbf{x}) = h\,\mathbf{u}^{\mathrm{S}} = h\,(\mathbf{u}^{\mathrm{P}} - \mathbf{u}^{\mathrm{G}})$, and $K^2 = K^2(\mathbf{x}) = fQ^{\mathrm{C}}(\mathbf{x})/g$ with $\mathbf{u}^{\mathrm{C}}$ replaced by $\mathbf{u}^{\mathrm{G}}$. The derivation of (6.5) from (6.4) is given in Appendix A. K^2 is robustly positive under the conditions of interest in practical meteorology or oceanography, with ϵ small and f dominant in (6.2); then $K^2 = {L_{\mathrm{R}}}^{-2}(1 + O(\epsilon))$; cf. (3.9).

Formally speaking, (6.4) completes the equations determining the time-evolution of the Hamiltonian balanced model defined by the $\mathbf{u}^{\mathrm{C}}$ functional. Given the mass configuration at some instant, all quantities with superscript C are known. Provided that (6.4) can be solved for $\mathbf{u}^{\mathrm{S}}$, we then know $\mathbf{u}^{\mathrm{P}} = \mathbf{u}^{\mathrm{C}} + \mathbf{u}^{\mathrm{S}}$ and can advect the mass configuration and time-march the problem. The mass configuration, described by the single scalar field $h(\mathbf{x})$, provides sufficient initial conditions because only the first time derivatives $\dot{\mathbf{x}} = \mathbf{u}^{\mathrm{P}}$ are involved, in place of the second time derivatives $\ddot{\mathbf{x}}$ in (2.8). We shall note later that the PV can be used in place of the mass field.

But why should $\mathbf{R}^{\mathrm{C}}$ be generically nonzero, no matter how cleverly we choose the $\mathbf{u}^{\mathrm{C}}$ functional? The reason is the existence of weak but generically nonzero 'spontaneous-adjustment emission', or (generalized) 'Lighthill radiation', of inertia–gravity waves by the unsteady vortical fluid motions described by the parent dynamics (Ford *et al.* 2000 & refs.). The existence of such wave emission means that there is no such thing as a constraint of the form $\mathbf{u} = \mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$ that does not interfere to some extent with the parent phase-space flow. At the very least, applying the constraint must suppress the spontaneous-adjustment emission. The phase space of the balanced model is just the constraint manifold, $\mathcal{M}^{\mathrm{C}}$, i.e. the submanifold of the parent phase space defined by $\mathbf{u} = \mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$. In this smaller phase space, essentially the parent configuration space, there is no room for degrees of freedom corresponding to inertia–gravity waves, a fact reflected in the form of (6.4), which as already remarked involves only the first time derivatives of $\mathbf{x}$ and not the second as in (2.8).

To put the same point another way — since it has been controversial in the past — the existence of spontaneous-adjustment emission implies that the parent dynamics can have no invariant 'slow manifold' in the sense originally hypothesized by Leith and Lorenz. For history and further discussion see Ford *et al.* (2000 & refs.). In place of such an invariant manifold there must exist within the parent phase space a fuzzy *'slow quasimanifold'*, a thin stochastic layer or chaotic layer of the kind familiar from low-order models of weakly coupled oscillators, such as the separatrix layer of a weakly perturbed pendulum (cf. Lynch, this volume). (Here the pendulum motion is regarded as analogous to the balanced vortical fluid motion, and the weak perturbation — due to coupling with a faster oscillator, for instance — is regarded as analogous to the coupling with inertia–gravity waves in the fluid system.) So the constraint manifold $\mathcal{M}^{\mathrm{C}}$ cannot be an invariant manifold of the parent dynamics, no matter how $\mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$ is chosen. Thin though the slow quasimanifold may be —

and the shallow-water experiments of MN00 show that it can be astonishingly thin in large regions of phase space — it is not infinitesimally thin, and no constraint manifold can ever be more than an approximation to it. There must be some ultimately irreducible error. The detailed analysis of Ford *et al.* (*op. cit.*) gives some idea of the order of magnitude of this irreducible error in certain parameter ranges; see also the further discussion in Saujani & Shepherd (2001 & refs.).

The simple form (6.4) of the splitting equation is a consequence of having just one constraint manifold $\mathcal{M}^{\mathrm{C}}$. We emphasize that semigeostrophic theory has no such simple splitting equation because, as mentioned earlier, its derivation from the parent primitive equations involves a 'double splitting' that produces *three* generically different velocity fields. There are *two* constraint manifolds within the parent phase space, $O(\epsilon)$ apart, one defined by the geostrophic constraint (3.2) and the other by Salmon's constraint (3.7). These define two of the relevant velocity fields and separately constrain, respectively, the Hamiltonian functional and the symplectic structure. Salmon's constraint shows itself in the Jacobian formula (4.17) for the conserved PV, but not in the formula 3.4 for the conserved energy. Appendix B discusses the — not quite trivial — relation between material PV conservation in physical space and symplectic-form invariance in phase space, explaining why it is Salmon's constraint (3.7) and not the geostrophic constraint (3.2) that enters into the semigeostrophic PV formulae (3.6) (3.8), and (4.17).

The constrained Hamiltonian is just the conserved energy (3.4), equal to the parent Hamiltonian evaluated with the geostrophic velocity (3.2). It is, of course, a mass-configuration functional. It can also be regarded as a result of applying Salmon's constraint (3.7) to a slightly different parent Hamiltonian in which $\frac{1}{2}|\mathbf{u}|^2$, in the standard parent Hamiltonian

$$\mathsf{H}\,(h(\cdot),\mathbf{u}(\cdot)) \;=\; \mathsf{V} + \int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}|^2 \,\mathrm{d}m \quad \text{with} \quad \mathsf{V} \;=\; \int_{\mathcal{D}} \tfrac{1}{2} g h \,\mathrm{d}m \;, \tag{6.6}$$

is replaced by $\frac{1}{2}|(\mathbf{u} - \mathbf{u}_{\mathrm{S}}^{\mathrm{C}} + \mathbf{u}^{\mathrm{G}})|^2$. The replacement is permissible because $\mathbf{u}_{\mathrm{S}}^{\mathrm{C}}$ and $\mathbf{u}^{\mathrm{G}}$ are both mass-configuration functionals. The third velocity field is of course $\dot{\mathbf{x}}$, i.e. $\mathbf{u}^{\mathrm{P}}$.

7 Ω_{ij}^{C} for particle dynamics

To see where the splitting equation's kernel Ω_{ij}^{C} comes from, it is easiest to consider first a simple textbook particle-dynamics analogue or toy problem. We consider the simplest conceivable such problem. The parent dynamics is that of a particle of unit mass moving in 2 dimensions under a potential $V(\mathbf{x})$ in an inertial reference frame, with the classical Hamiltonian function

$$H(\mathbf{x},\mathbf{u}) = \tfrac{1}{2} u_i u_i + V(\mathbf{x}) \tag{7.1}$$

where $\mathbf{x} = \{x_i\} = (x_1,\, x_2)$ denotes the particle-position coordinates defining the system configuration at a given instant (we suppress explicit reference to the time t), and $\mathbf{u} = \{u_i\}$ denotes the corresponding set of canonical momenta. Summation over over repeated indices is understood from here on. The parent phase space is $\mathbb{R}^4 = \{x_1, x_2, u_1, u_2\}$. To express the symplectic structure involved, it is again natural to use the exterior calculus. Here, however, we have Euclidean structure also, within the subspaces $\{x_1, x_2\}$ and $\{u_1, u_2\}$, representing physical distance and velocity; and in any case we want the notation to stay close to that of fluid-dynamical equations such as (2.8). So we compromise by using elementary textbook notation, with occasional reference to the more elegant exterior-calculus formulae.

The four Hamilton's equations are

$$\dot{x}_i = \frac{\partial H}{\partial u_i}\;, \qquad \dot{u}_i = -\frac{\partial H}{\partial x_i} \qquad (i = 1, 2)\;, \tag{7.2}$$

and an alternative way of expressing the same information is to introduce arbitrary variations $\delta x_i, \delta u_i$ and to note that the equations are equivalent to

$$\dot{x}_i \delta u_i - \dot{u}_i \delta x_i = \frac{\partial H}{\partial x_i}\delta x_i + \frac{\partial H}{\partial u_i}\delta u_i = \delta H\;, \text{ say.} \tag{7.3}$$

In terms of exterior calculus, the left-hand side of (7.3) can be recognized as the contraction, or inner or 'interior' product, of a pair of vectors with the 2-form $\mathrm{d}x_i \wedge \mathrm{d}u_i = \mathrm{d}x_1 \wedge \mathrm{d}u_1 + \mathrm{d}x_2 \wedge \mathrm{d}u_2$ expressing the symplectic structure of the parent phase space $\mathbb{R}^4$. The vectors are the phase-space flow vector $(\dot{\mathbf{x}}, \dot{\mathbf{u}}) = (\dot{x}_1, \dot{x}_2, \dot{u}_1, \dot{u}_2)$ and the variation vector $(\delta\mathbf{x}, \delta\mathbf{u}) = (\delta x_1, \delta x_2, \delta u_1, \delta u_2)$, expressed in the canonical coordinates (x_1, x_2, u_1, u_2). Here the symplectic structure is mathematically the same as that associated with (5.5) and (5.6), but physically with quite different associations, being a property of the parent phase space in the toy problem.

Now (7.3) has a key advantage for our purpose, namely the simplicity and directness with which a new Hamiltonian dynamical system can be constructed from (7.3) by applying a workless momentum–configuration constraint. There is no need to introduce Lagrange multipliers, nor integrals with respect to t nor caveats about their end points. To generate a constrained problem that inherits Hamiltonian structure from the parent dynamics, one need only adhere to the following basic principle:

$$\boxed{\begin{array}{c}\textit{Apply the same constraint to the variations } \delta\mathbf{x}, \delta\mathbf{u} \\ \textit{as is applied to the phase-space flow } \dot{\mathbf{x}}, \dot{\mathbf{u}}.\end{array}} \tag{7.4}$$

In the fluid problem in a finite domain this principle is crucial, incidentally, to finding boundary conditions that preserve Hamiltonian structure, as illustrated in MP96.

In the present toy problem, (7.4) means that if the constraint manifold $\mathcal{M}^{\mathrm{C}}$ is described by a set of smooth functions $x_i = x_i^{\mathrm{C}}(\mathbf{z})$, $u_i = u_i^{\mathrm{C}}(\mathbf{z})$, where $\mathbf{z} = \{z_j\}$ denotes a smaller set of variables, then in order to convert (7.3) into the new, constrained problem, preserving Hamiltonian structure, one merely substitutes the same functions $x_i^{\mathrm{C}}(\mathbf{z})$, $u_i^{\mathrm{C}}(\mathbf{z})$ into $\delta x_i, \delta u_i$ as into $\dot{x}_i, \dot{u}_i$ and $H(\mathbf{x}, \mathbf{u})$ (Salmon 1988a, equation (4.4)). Here $\mathbf{z} = \{z_j\}$ will be taken to be 2-dimensional: $j = 1, 2$. It is necessary to assume nondegeneracy in the sense that $\mathcal{M}^{\mathrm{C}}$ is not part of any Lagrangian submanifold (Section 5) of the parent phase space.

First consider the expressions $u_i \dot{x}_i$ and $-u_i \delta x_i$. Constraining these expressions produces, respectively,

$$u_i \dot{x}_i = u_i^{\mathrm{C}}(\mathbf{z}) \frac{\mathrm{d}}{\mathrm{d}t} x_i^{\mathrm{C}}(\mathbf{z}) = -\theta_j^{\mathrm{C}}(\mathbf{z}) \dot{z}_j \ , \tag{7.5}$$

say, and

$$-u_i \delta x_i = -u_i^{\mathrm{C}}(\mathbf{z}) \delta x_i^{\mathrm{C}}(\mathbf{z}) = \theta_j^{\mathrm{C}}(\mathbf{z}) \delta z_j \ , \tag{7.6}$$

where both expressions involve the same set of known functions $\theta_j^{\mathrm{C}}(\mathbf{z})$, namely

$$\theta_j^{\mathrm{C}}(\mathbf{z}) = -u_i^{\mathrm{C}}(\mathbf{z}) \frac{\partial x_i^{\mathrm{C}}(\mathbf{z})}{\partial z_j} \ . \tag{7.7}$$

Adding the first variation of (7.5) to $\mathrm{d}/\mathrm{d}t$ of (7.6) and noting that two terms $\pm u_i \delta \dot{x}_i$ cancel on the left, similarly $\pm \theta_j^{\mathrm{C}} \delta \dot{z}_j$ on the right, we see that the left-hand side of (7.3) becomes simply

$$\begin{aligned} \dot{x}_i \delta u_i - \dot{u}_i \delta x_i &= -\dot{z}_i \delta \theta_i^{\mathrm{C}} + \dot{\theta}_i^{\mathrm{C}} \delta z_i & (7.8) \\ &= \Omega_{ij}^{\mathrm{C}} \dot{z}_i \delta z_j \ , \ \text{say} & (7.9) \end{aligned}$$

where

$$\Omega_{ij}^{\mathrm{C}}(\mathbf{z}) = -\frac{\partial \theta_i^{\mathrm{C}}(\mathbf{z})}{\partial z_j} + \frac{\partial \theta_j^{\mathrm{C}}(\mathbf{z})}{\partial z_i} \ , \tag{7.10}$$

because

$$\delta \theta_j^{\mathrm{C}} = \frac{\partial \theta_j^{\mathrm{C}}}{\partial z_i} \delta z_i \qquad \text{and} \qquad \dot{\theta}_j = \frac{\partial \theta_j^{\mathrm{C}}}{\partial z_i} \dot{z}_i \ . \tag{7.11}$$

Nondegeneracy says that the constraint functions, $x_i^{\mathrm{C}}(\mathbf{z})$ and $u_i^{\mathrm{C}}(\mathbf{z})$, are such that the antisymmetric matrix Ω_{ij}^{C} is invertible. This is possible because $\mathbf{z}$ is even-dimensional. The expression (7.9) is the contraction or inner product of the two vectors $\dot{\mathbf{z}}$ and $\delta \mathbf{z}$ (in the tangent space at a point of $\mathcal{M}^{\mathrm{C}}$) with the 2-form

$$\Omega^{\mathrm{C}} = \tfrac{1}{2} \Omega_{ij}^{\mathrm{C}} \, \mathrm{d}z_i \wedge \mathrm{d}z_j \tag{7.12}$$

produced by restricting the parent 2-form $\mathrm{d}x_i \wedge \mathrm{d}u_i$ to the submanifold $\mathcal{M}^{\mathrm{C}}$. Figure 1 illustrates this process pictorially, for a special case with $\mathbf{x}^{\mathrm{C}}(\mathbf{z}) = \mathbf{z}$; see the caption for further explanation.

Herein lies the beauty of using differential forms instead of partial derivatives. Applying the constraint restricting $\mathrm{d}x_i \wedge \mathrm{d}u_i$ to $\mathcal{M}^{\mathrm{C}}$ not only preserves metric-independent geometric properties such as tangency and intersection, but is utterly straightforward, because the use of the differential form $\mathrm{d}x_i \wedge \mathrm{d}u_i$ introduces no prior prejudices about directions in phase space and therefore no need to undo such prejudices via Lagrange multipliers.

The 2-form Ω^{C} is the exterior derivative of the 1-form $\theta^{\mathrm{C}} = \theta^{\mathrm{C}}_i \mathrm{d}z_i$ (whether or not we add an extra term of the form $\mathrm{d}\phi^{\mathrm{C}}$); that is, $\Omega^{\mathrm{C}} = \mathrm{d}\theta^{\mathrm{C}}$. Note from the antisymmetry of Ω^{C}_{ij} and the symmetry of $\partial^2/\partial z_i \partial z_j$ that

$$\frac{\partial \Omega^{\mathrm{C}}_{ij}}{\partial z_k} + \frac{\partial \Omega^{\mathrm{C}}_{jk}}{\partial z_i} + \frac{\partial \Omega^{\mathrm{C}}_{ki}}{\partial z_j} = 0 \,. \tag{7.13}$$

This expresses $\mathrm{d}\Omega^{\mathrm{C}} = \mathrm{d}\,\mathrm{d}\theta^{\mathrm{C}} = 0$. (In words, Ω^{C} is 'closed', $\mathrm{d}\Omega^{\mathrm{C}} = 0$, because it is 'exact', $\Omega^{\mathrm{C}} = \mathrm{d}\theta^{\mathrm{C}}$.) From (7.3) and (7.9), then, writing $H^{\mathrm{C}}(\mathbf{z}) = H(\mathbf{x}^{\mathrm{C}}(\mathbf{z}), \mathbf{u}^{\mathrm{C}}(\mathbf{z}))$, we have

$$\Omega^{\mathrm{C}}_{ij}\, \dot{z}_i\, \delta z_j = \delta H^{\mathrm{C}} \,. \tag{7.14}$$

Removing the arbitrary factors δz_j, we see that the new problem that results from constraining (7.3) is

$$\Omega^{\mathrm{C}}_{ij}\, \dot{z}_i = \frac{\partial H^{\mathrm{C}}}{\partial z_j} \,. \tag{7.15}$$

Nondegeneracy says that the matrix Ω^{C}_{ij} has an inverse J^{C}_{ij}, with $\Omega^{\mathrm{C}}_{ij} J^{\mathrm{C}}_{jk} = \delta_{ik}$, where δ with suffixes is the Kronecker delta, giving Hamilton's equations for the new, constrained problem in noncanonical form:

$$\dot{z}_i = J^{\mathrm{C}}_{ji} \frac{\partial H^{\mathrm{C}}}{\partial z_j} \,. \tag{7.16}$$

The matrix Ω^{C}_{ij} is called the noncanonical *symplectic matrix* and J^{C}_{ij} the noncanonical *cosymplectic matrix*, though the two entities are sometimes conflated. The corresponding Poisson or cosymplectic *bracket*, $\{\!\{\cdot,\cdot\}\!\}$ acting on pairs of scalar-valued functions, is

$$\{\!\{A(\mathbf{z}), B(\mathbf{z})\}\!\} = \frac{\partial A}{\partial z_i} J^{\mathrm{C}}_{ji} \frac{\partial B}{\partial z_j} \tag{7.17}$$

and satisfies Jacobi's identity

$$\{\!\{\{\!\{A,B\}\!\},C\}\!\} + \{\!\{\{\!\{B,C\}\!\},A\}\!\} + \{\!\{\{\!\{C,A\}\!\},B\}\!\} = 0 \tag{7.18}$$

for all $A(\mathbf{z}), B(\mathbf{z}), C(\mathbf{z})$, or equivalently

$$J^{\mathrm{C}}_{il}\frac{\partial J^{\mathrm{C}}_{jk}}{\partial z_l} \;+\; J^{\mathrm{C}}_{jl}\frac{\partial J^{\mathrm{C}}_{ki}}{\partial z_l} \;+\; J^{\mathrm{C}}_{kl}\frac{\partial J^{\mathrm{C}}_{ij}}{\partial z_l} \;=\; 0\ , \tag{7.19}$$

which in turn is equivalent to (7.13) for nonsingular Ω^{C}_{ij} and J^{C}_{ij}. (This can be verified by multiplying (7.19) by $\Omega^{\mathrm{C}}_{im}\Omega^{\mathrm{C}}_{jn}\Omega^{\mathrm{C}}_{kp}$ then using $\partial/\partial z_l$ of the relation $\Omega^{\mathrm{C}}_{ij}J^{\mathrm{C}}_{jk} = \delta_{ik}$.) Each of the two versions, (7.15) with (7.13) on the one hand, and (7.16) with (7.19) on the other, are Hamiltonian systems by standard definitions (e.g. Salmon 1988b, Shepherd 1990), as expected from the basic principle (7.4).

Energy conservation for the new, constrained problem follows at once from the antisymmetry of the symplectic matrix Ω^{C}_{ij}. Replacing the variations in (7.14) by flow rates on $\mathcal{M}^{\mathrm{C}}$, we have

$$\frac{\mathrm{d}H^{\mathrm{C}}}{\mathrm{d}t} \;=\; \Omega^{\mathrm{C}}_{ij}\,\dot z_i\,\dot z_j \;=\; 0\ , \tag{7.20}$$

by antisymmetry. In the exterior-calculus language, we may use $\dot{\mathbf{z}}\rfloor\Omega^{\mathrm{C}} = \mathrm{d}H^{\mathrm{C}}$ in place of (7.14), where Ω^{C} is the 2-form (7.12) and where $\dot{\mathbf{z}}\rfloor\Omega^{\mathrm{C}}$ denotes contraction or inner or interior multiplication (analogous to the 'scalar' product of ordinary vector calculus, and also notated $\iota(\dot{\mathbf{z}})\Omega^{\mathrm{C}}$ or $\iota_{\dot{\mathbf{z}}}\Omega^{\mathrm{C}}$). Thus the derivation of (7.20) can be rewritten

$$\frac{\mathrm{d}H^{\mathrm{C}}}{\mathrm{d}t} \;=\; \dot{\mathbf{z}}\rfloor\,\mathrm{d}H^{\mathrm{C}} \;=\; \dot{\mathbf{z}}\rfloor\dot{\mathbf{z}}\rfloor\Omega^{\mathrm{C}} \;=\; 0\ , \tag{7.21}$$

by antisymmetry; note that $\dot{\mathbf{z}}\rfloor\dot{\mathbf{z}}\rfloor\Omega^{\mathrm{C}}$ means $\dot{\mathbf{z}}\rfloor\,(\,\dot{\mathbf{z}}\rfloor\Omega^{\mathrm{C}})$.

Now the constraints of interest in this article are always momentum–configuration constraints, for which

$$\mathbf{x}^{\mathrm{C}}(\mathbf{z}) = \mathbf{z}\ . \tag{7.22}$$

Then $\partial x^{\mathrm{C}}_i/\partial z_j = \delta_{ij}$, whence (7.7) gives

$$\theta^{\mathrm{C}}_j(\mathbf{z}) \;=\; -u^{\mathrm{C}}_j(\mathbf{z}) \;=\; -u^{\mathrm{C}}_j(\mathbf{x}) \qquad\qquad (\,\theta^{\mathrm{C}} = \;\mathrm{d}\phi^{\mathrm{C}} - u^{\mathrm{C}}_j\;\mathrm{d}x_j\,) \tag{7.23}$$

so that (7.8), (7.9) and (7.14) become

$$\dot x_j\delta u_j - \dot u_j\delta x_j \;=\; \dot x_j\delta u^{\mathrm{C}}_j - \dot u^{\mathrm{C}}_j\delta x_j \;=\; \Omega^{\mathrm{C}}_{ij}\dot x_i\delta x_j \;=\; \delta H^{\mathrm{C}}\ , \tag{7.24}$$

where now

$$\boxed{\Omega^{\mathrm{C}}_{ij} \;=\; \frac{\partial u^{\mathrm{C}}_i}{\partial x_j} - \frac{\partial u^{\mathrm{C}}_j}{\partial x_i}\ .} \tag{7.25}$$

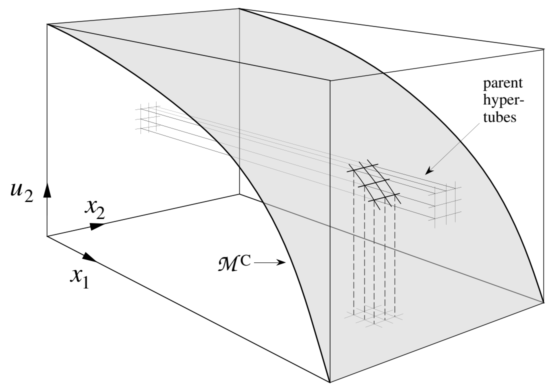

FIGURE 1. Partial visualization (seen in three out of four phase-space dimensions) of a constraint manifold $\mathcal{M}^{\mathrm{C}}$ and its intersection with the parent symplectic structure ('horizontal' tubes), for the simplest possible case, the toy problem of Section 7, a single particle in a 2-dimensional configuration space $\{x, y\} = \{x_1, x_2\}$. In this illustration, the momentum–configuration constraint defining $\mathcal{M}^{\mathrm{C}}$ has the special form $u_1 = u_1^{\mathrm{C}}$ = constant, $u_2 = u_2^{\mathrm{C}}$ = function of x_1 alone, giving nondegenerate (invertible) Ω_{ij}^{C} provided that $\partial u_2^{\mathrm{C}}/\partial x_1 \neq 0$, i.e. $\mathcal{M}^{\mathrm{C}}$ nowhere 'horizontal' in the figure. Only part of the parent symplectic structure is visible. That structure (corresponding to the 2-form $\mathrm{d}x_1 \wedge \mathrm{d}u_1 + \mathrm{d}x_2 \wedge \mathrm{d}u_2$) consists of two sets of infinitesimal oriented hypertubes (e.g. Misner *et al.* 1973, Schutz 1980), 'tubes' for short, the first set (not shown) intersecting only the x_1u_1 plane and the second (shown schematically as the 'horizontal' tubes) intersecting only the x_2u_2 plane. 'Oriented' means that the sign of the circulation around each tube is defined, positive being anticlockwise in the case shown, i.e. such that a right-handed screw would move in the positive x_1 direction. Only a few members of the second set of parent tubes are shown, as if they had finite cross-sections. A continuum limit needs to be understood. Because (x_1, x_2, u_1, u_2) are global canonical (Darboux) coordinates, the two sets of parent tubes fill phase space homogeneously. The tubes' infinitesimal cross-sectional shapes are irrelevant: instead of squares they could be parallelograms, hexagons or any other area-measurable shape. They are significant only as regards the signed total 'number of tubes' threading any specified small parallelogram A (not shown) that is arbitrarily oriented in the 4-dimensional phase space, the sign being positive if the tubes' orientation agrees with that of A. This signed total, in the continuum limit with A becoming infinitesimal, is proportional to the contraction or interior product of the 2-form with the pair of vectors defining A. Because metric concepts like 'angle' and 'orthogonality' are not used, the relevant intersection properties are inherited when the constraint is applied in accordance with (7.4), as here, with A lying in $\mathcal{M}^{\mathrm{C}}$ (corresponding to the 2-form $\frac{1}{2}\Omega_{ij}^{\mathrm{C}}\,\mathrm{d}x_i \wedge \mathrm{d}x_j$ when $\mathbf{x}$ is used, as here, to track position on $\mathcal{M}^{\mathrm{C}}$). Nondegeneracy (invertibility of Ω_{ij}^{C}) says, in this illustration, that when A lies in $\mathcal{M}^{\mathrm{C}}$ the signed total number of parent tubes threading A does not vanish — true here, as the figure is drawn, because, with u_1^{C} = constant, the first set of parent tubes does not intersect $\mathcal{M}^{\mathrm{C}}$ at all, while the second set does. We would have a degenerate case if, for instance, u_2^{C} were left as it is but u_1^{C} changed from a constant to a function defined by the indefinite integral $u_1^{\mathrm{C}} = \int(\partial u_2^{\mathrm{C}}/\partial x_1)\,\mathrm{d}x_2$, in which case the two sets of parent tubes would give mutually cancelling contributions for any A lying in $\mathcal{M}^{\mathrm{C}}$. Then $\mathcal{M}^{\mathrm{C}}$ would be a Lagrangian submanifold of $\{x_1, x_2, u_1, u_2\}$. When this picture is extended to the infinite-dimensional fluid cases of interest, under parameter conditions favouring near-geostrophic balance, $\mathcal{M}^{\mathrm{C}}$ will be, heuristically speaking, 'steeper' than the figure suggests: particles need not move far to upset near-geostrophic balance when Rossby numbers ϵ are small.

In the exterior-calculus language, the last of (7.24) reads $\delta\mathbf{x}\rfloor\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{C}} = \delta H^{\mathrm{C}}$, equivalent to $\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{C}} = \mathrm{d}\,H^{\mathrm{C}}$) by arbitrariness of $\delta\mathbf{x}$, where now (cf. (7.12))

$$\Omega^{\mathrm{C}} \;=\; \tfrac{1}{2}\Omega^{\mathrm{C}}_{ij}\,\mathrm{d}\,x_i\wedge\mathrm{d}\,x_j\;. \tag{7.26}$$

We now have

$$\delta H^{\mathrm{C}} \;=\; \delta H(\mathbf{x},\;\mathbf{u}^{\mathrm{C}}(\mathbf{x})) \;=\; \delta V + u^{\mathrm{C}}_j\delta u^{\mathrm{C}}_j\;. \tag{7.27}$$

Defining

$$\mathbf{u}^{\mathrm{S}} \;=\; \dot{\mathbf{x}} - \mathbf{u}^{\mathrm{C}}(\mathbf{x})\;, \tag{7.28}$$

we may now derive the splitting equation for the toy problem. The shortest route uses (7.24) and (7.27):

$$\begin{aligned} 0 \;&=\; -\dot{x}_j\delta u^{\mathrm{C}}_j + \dot{u}^{\mathrm{C}}_j\delta x_j + \delta V + u^{\mathrm{C}}_j\delta u^{\mathrm{C}}_j && (7.29)\\ &=\; -u^{\mathrm{S}}_j\delta u^{\mathrm{C}}_j + \dot{u}^{\mathrm{C}}_i\delta x_i + \delta V && (7.30)\\ &=\; -u^{\mathrm{S}}_j\delta u^{\mathrm{C}}_j + (\dot{u}^{\mathrm{C}}_i - A^{\mathrm{C}}_i)\delta x_i + \delta V + A^{\mathrm{C}}_i\delta x_i && (7.31)\end{aligned}$$

for any vector $\mathbf{A}^{\mathrm{C}}$. If we now choose

$$A^{\mathrm{C}}_i \;=\; u^{\mathrm{C}}_j\frac{\partial u^{\mathrm{C}}_i}{\partial x_j}, \tag{7.32}$$

which is the constraint acceleration in the sense of (6.3)ff., then

$$\dot{u}^{\mathrm{C}}_i - A^{\mathrm{C}}_i \;=\; \dot{x}_j\frac{\partial u^{\mathrm{C}}_i}{\partial x_j} - u^{\mathrm{C}}_j\frac{\partial u^{\mathrm{C}}_i}{\partial x_j} \;=\; u^{\mathrm{S}}_j\frac{\partial u^{\mathrm{C}}_i}{\partial x_j}\;. \tag{7.33}$$

Then (7.31) becomes simply

$$\Omega^{\mathrm{C}}_{ij}u^{\mathrm{S}}_j\delta x_i \;=\; R^{\mathrm{C}}_i\delta x_i, \tag{7.34}$$

where

$$R^{\mathrm{C}}_i \;=\; F_i - A^{\mathrm{C}}_i \;=\; -\frac{\partial H}{\partial x_i} - A^{\mathrm{C}}_i \;=\; -\frac{\partial V}{\partial x_i} - A^{\mathrm{C}}_i\;, \tag{7.35}$$

giving the splitting equation:

$$\boxed{\Omega^{\mathrm{C}}_{ij}u^{\mathrm{S}}_j \;=\; R^{\mathrm{C}}_i \qquad \text{or equivalently} \qquad u^{\mathrm{S}}_i \;=\; J^{\mathrm{C}}_{ij}R^{\mathrm{C}}_j\;.} \tag{7.36}$$

In words, the splitting kernel is the symplectic matrix of the constrained problem. The term $-\partial H/\partial x_i = -\partial V/\partial x_i$ in (7.35) represents, as before, the particle acceleration that would occur if the constraint were suddenly removed.

Notice from (7.27)ff. that the derivation depended on the classical form (7.1) of $H(\mathbf{x},\mathbf{u})$, with u_i occurring quadratically, so that the parent canonical momenta coincide with ordinary momenta. Of course (7.36) is no more than an alternative way of expressing Hamilton's equations (7.15), (7.16) for the constrained problem, which can now be rewritten

$$\boxed{\Omega^{\mathrm{C}}_{ij}\dot{x}_j \;=\; -\frac{\partial H^{\mathrm{C}}}{\partial x_i} \qquad \text{or equivalently} \qquad \dot{x}_i \;=\; -J^{\mathrm{C}}_{ij}\,\frac{\partial H^{\mathrm{C}}}{\partial x_j}\;,} \tag{7.37}$$

again using antisymmetry. (The minus signs arise from contracting on the right, with the summation over j, instead of on the left as in $\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{C}} = \mathrm{d}H^{\mathrm{C}}$; see MS98 Remark 3.3, p. 84, on sign conventions.)

We note in passing — though this will not be used in the sequel — that another view of the problem is, of course, obtainable through traditional Lagrange multipliers. This is related though not identical to the approach of Allen & Holm (1996).

In our version, the velocity-split $\mathbf{u}^{\mathrm{S}}$ turns out to be equal to the Lagrange multiplier $\boldsymbol{\lambda}$ of the momentum–configuration constraint $\mathbf{u}-\mathbf{u}^{\mathrm{C}} = 0$. Following the traditional recipe, we replace $H(\mathbf{x},\mathbf{u})$ by $\widehat{H}(\mathbf{x},\mathbf{u},\boldsymbol{\lambda}) = H(\mathbf{x},\mathbf{u}) + \lambda_j(\mathbf{x})\phi_j(\mathbf{x},\mathbf{u})$ where, in our version, $\phi_j = 0$ $(j = 1,2)$ is the constraint in standard notation, not to be confused with the geopotential of earlier sections: here, $\phi_j = u_j - u^{\mathrm{C}}_j(\mathbf{x})$. Then we get six equations

$$\dot{x}_i \;=\; \frac{\partial\widehat{H}}{\partial u_i}\;, \qquad \dot{u}_i \;=\; -\frac{\partial\widehat{H}}{\partial x_i}\;, \qquad \phi_j \;=\; 0\;, \tag{7.38}$$

to be solved for the six unknowns $\dot{x}_i, \dot{u}_i$ and λ_i. This system of six equations is not self-evidently Hamiltonian, but can readily be shown to be equivalent to (7.15) and hence Hamiltonian by implication, in a noncanonical description. For the first of (7.38) shows at once that

$$\boldsymbol{\lambda} = \dot{\mathbf{x}} - \mathbf{u}^{\mathrm{C}} = \mathbf{u}^{\mathrm{S}}\;; \tag{7.39}$$

and, using (7.39), the second of (7.38) can be shown in two or three lines of manipulation to be equivalent to (7.15), either directly or by using the facts that $\{\!\{\phi_j,\widehat{H}\}\!\} = 0$ and $\{\!\{\phi_j,\phi_j\}\!\} = 0$, where $\{\!\{\cdot,\cdot\}\!\}$ is the canonical Poisson or cosymplectic bracket of the parent dynamics,

$$\{\!\{A,\,B\}\!\} \;=\; \frac{\partial A}{\partial x_i}\frac{\partial B}{\partial u_i} - \frac{\partial B}{\partial x_i}\frac{\partial A}{\partial u_i}\;. \tag{7.40}$$

In this bracket, as distinct from that of (7.17), it is crucial to read all the partial differentiations as being taken in the full parent phase space — the only way they make sense — and not on the constraint manifold $\mathcal{M}^{\mathrm{C}}$. This is where the restrictiveness of partial differentiation, as compared with the freedom allowed by differential forms, makes the technicalities a little more complicated. (The Lagrange multipliers λ_j can, however, be held constant during all such differentiations, because they are always multiplied by $\phi_j = 0$.)

Notice finally that (7.24) and (7.34) can easily be generalized to include dissipation or forcing terms. Such terms can be added on the right of Hamilton's equations (7.2) and carried through the whole analysis. Of course the extra terms might interfere with balance, and the possible accuracy of the resulting balanced model is a separate question.

8 Extension to shallow-water dynamics

It is straightforward to generalize the foregoing to the shallow-water system described in Section 2. Instead of the phase space $\mathbb{R}^4 = (x_1, x_2, u_1, u_2)$ we have the Cartesian product $\mathbb{R}^4 \otimes \mathbb{R}^4 \otimes \cdots$ of an infinite number of such $\mathbb{R}^4$, one for each value of the fluid-particle label $\mathbf{a} = (a, b) = (a_1, a_2) \in \mathbb{R}^2$. Symbolically, the phase space is $\mathbb{R}^\infty = \{x_1(\mathbf{a}),\, x_2(\mathbf{a}),\, u_1(\mathbf{a}),\, u_2(\mathbf{a})\}$. The constraint manifold $\mathcal{M}^{\mathrm{C}}$ defined by a momentum–configuration constraint $\mathbf{u} = \mathbf{u}^{\mathrm{C}}$ has half the dimensions (in the Lagrangian description), corresponding to $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \cdots$ or $\mathbb{R}^\infty = \{x_1(\mathbf{a}),\, x_2(\mathbf{a})\}$. Thus the summation over $j = 1, 2$ in (7.36) has to be replaced by a similar summation together with integration over all mass elements $\mathrm{d}m = h_0\,\mathrm{d}\mathbf{a} = h\,\mathrm{d}\mathbf{x}$ in the physical domain $\mathcal{D}$, whence the form of (6.4). In the following, we again suppress explicit reference to the time t.

The formula for the kernel $\Omega^{\mathrm{C}}_{ij}(\mathbf{x}, \mathbf{x}')$ in (6.4) straightforwardly resembles its toy-problem counterpart, the symplectic matrix (7.25), provided that the Lagrangian description of fluid motion is used in an inertial frame of reference. This requires us to reinterpret $\mathbf{u}^{\mathrm{C}}$ as an absolute velocity, and to use the configuration mapping $\mathbf{a} \leftrightarrow \mathbf{x}$ to rewrite $\mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$ in its Lagrangian form $\mathbf{u}^{\mathrm{C}}(\mathbf{a}; \mathbf{x}(\cdot))$, again using the shorthand convention introduced in equation (2.5). The shorthand convention serves to emphasize that $\mathbf{u}^{\mathrm{C}}$ has the same value for the same fluid particle and the same mass configuration. Arguments will be shown explicitly, as here, whenever there might be danger of confusion. Then (7.25) is replaced by the following expression, for a given pair of fluid particles or mass elements,

$$\Omega^{\mathrm{CL}}_{ij}(\mathbf{a}, \mathbf{a}') \;=\; \frac{\delta^{\mathrm{L}} u^{\mathrm{C}}_i(\mathbf{a}; \mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')} - \frac{\delta^{\mathrm{L}} u^{\mathrm{C}}_j(\mathbf{a}'; \mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_i(\mathbf{a})}\ , \tag{8.1}$$

with Lagrangian functional derivatives in $\mathbb{R}^\infty$ replacing partial derivatives in $\mathbb{R}^2$. The functional derivatives (cf. (8.8) below) refer to the second argument in $\mathbf{u}^{\mathrm{C}}(\mathbf{a}; \mathbf{x}(\cdot))$, holding the first constant, and describe how $\mathbf{u}^{\mathrm{C}}$ for a given fluid particle varies as the mass configuration varies. The antisymmetry property mentioned below (6.4) is now evident. Exchanging the two horizontal directions i, j and the two particles or mass elements reverses the sign of the kernel.

Invertibility is now a nontrivial issue, though in the special case (6.5) there is clearly no problem because, in that 'near-local' case, the left-hand side becomes a differential operator, indeed nothing but a modified Helmholtz operator, as shown in Appendix A. So in that case at least, the splitting equation (6.4) is robustly and uniquely invertible, given the evanescent boundary conditions. For the most accurate balance conditions, in which $\mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$ has nonlocal dependence on the mass configuration, (6.4) becomes an integro–differential equation for $\mathbf{u}^{\mathrm{S}}$. Even though it is a linear equation, for a given mass configuration, there are many unanswered mathematical questions.

It is desirable to put (8.1) into Eulerian form, to make explicit the fact that everything involved in the dynamics respects the particle-relabelling symmetry, i.e. is invariant to mass-distribution-conserving ($h(\mathbf{x})$-conserving) particle rearrangements, isomorphic to particle relabellings (which permute subsets of the factors of $\mathbb{R}^4 \otimes \mathbb{R}^4 \otimes \cdots$ leaving all dynamical quantities unchanged, including the value of the Hamiltonian). The dynamics can involve only the Eulerian description $h(\mathbf{x})$ of the mass configuration.

To deal with mass-configuration functionals like $\mathbf{u}^{\mathrm{C}}$ that are also fields, one must distinguish between Lagrangian and Eulerian variations. Here the notation δ^{L} will always mean a Lagrangian variation, with the implication that $\mathbf{a}$ is held constant, as with the configuration-mapping variation $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and the functional derivatives in (8.1). The notation δ^{E} will always mean an Eulerian variation, with $\mathbf{x}$ held constant. Variations of scalar-valued mass-configuration functionals, such as the constrained Hamiltonian functional (cf. 3.4)

$$\mathsf{H}^{\mathrm{C}}(h(\cdot)) \;=\; \mathsf{V} + \int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}^{\mathrm{C}}|^2 \,\mathrm{d}m \quad \text{with} \quad \mathsf{V} \;=\; \int_{\mathcal{D}} \tfrac{1}{2} g h \,\mathrm{d}m \;, \tag{8.2}$$

will denoted by δ, thus $\delta\mathsf{H}^{\mathrm{C}}$, and similarly $\delta\mathsf{H}$ for variations of the parent Hamiltonian functional $\mathsf{H}(h(\cdot), \mathbf{u}(\cdot))$, defined in (6.6), which takes points in the parent phase space $\mathbb{R}^\infty$ into real scalars.

The Eulerian and Lagrangian variations $\delta^{\mathrm{E}}h$ and $\delta^{\mathrm{L}}h$, with $h(\cdot)$ regarded as a function of $\mathbf{x}$, are related to the mapping variations by

$$\delta^{\mathrm{L}}h(\mathbf{x}) \;=\; \delta^{\mathrm{L}}\mathbf{x}\cdot\nabla h \;+\; \delta^{\mathrm{E}}h \;=\; -h\,\nabla\cdot(\delta^{\mathrm{L}}\mathbf{x}) \;, \tag{8.3}$$

$$\delta^{\mathrm{E}}h(\mathbf{x}) \;=\; -\nabla\cdot(h\,\delta^{\mathrm{L}}\mathbf{x}) \;, \tag{8.4}$$

consistently with the general relation

$$\delta^{\mathrm{L}} \;=\; \delta^{\mathrm{L}}\mathbf{x}\cdot\nabla \;+\; \delta^{\mathrm{E}} \;. \tag{8.5}$$

This last is applicable to any function of $\mathbf{x}$. In (8.3) and (8.4), the notation $\delta^{\mathrm{L}}\mathbf{x}$ is shorthand for $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ re-expressed as a function $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{x})$ of $\mathbf{x}$ via the mapping $\mathbf{a} \leftrightarrow \mathbf{x}$. Observe that the notation $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{x})$ is consistent with (8.5), because $\delta^{\mathrm{E}}\mathbf{x}$ is trivially zero. Observe also that δ^{E} commutes with ∇ or $\partial/\partial x_i$, whereas δ^{L} does not.

The toy-problem relation (7.27) has the counterpart

$$\delta\mathsf{H}^{\mathrm{C}} \;=\; \delta\mathsf{V} + \delta\int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}^{\mathrm{C}}|^2 \,\mathrm{d}m \;=\; \delta\mathsf{V} + \int_{\mathcal{D}} u_j^{\mathrm{C}}\,\delta^{\mathrm{L}} u_j^{\mathrm{C}}\,\mathrm{d}m\;, \tag{8.6}$$

again integrating over mass elements $\mathrm{d}m$ as well as summing the index j from 1 to 2, and using the fact that the mass element $\mathrm{d}m$ has zero Lagrangian variation.

From (8.3) and (8.4) it is straightforward to show that the relevant Lagrangian and Eulerian functional derivatives are related by the appropriate form of the chain rule,

$$\frac{\delta^{\mathrm{L}} u_i^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')} \;=\; \frac{h_0}{h(\mathbf{x})}\,\frac{\partial u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\partial x_j}\,\delta(\mathbf{x}-\mathbf{x}') \;+\; h_0\,\frac{\partial}{\partial x_j'}\,\frac{\delta^{\mathrm{E}} u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')}\;, \tag{8.7}$$

where δ with argument $(\mathbf{x}-\mathbf{x}')$ denotes the 2-dimensional Dirac delta function. In the first term on the right, $\partial/\partial x_j$ connotes that the function argument $h(\cdot)$ is not varied. In the last term, $\delta^{\mathrm{E}}/\delta^{\mathrm{E}}h$ connotes that the Eulerian position argument $\mathbf{x}$ is not varied. The second position argument $\mathbf{x}'$, appearing in the denominator, is analogous to the denominator index j in the first term and also, like the $\mathbf{a}'$ on the left, indicates which dummy variable of integration to use when computing variations. Specifically, by the standard definition of a functional derivative we have, for sufficiently smooth variation fields $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{E}}h(\mathbf{x})$,

$$\delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}(\mathbf{a}) \;=\; \delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot)) \;=\; \int_{\mathcal{D}} \frac{\delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')}\,\delta^{\mathrm{L}} x_j(\mathbf{a}')\,\mathrm{d}\mathbf{a}' \tag{8.8}$$

and

$$\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}(\mathbf{x}) \;=\; \delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}(\mathbf{x};h(\cdot)) \;=\; \int_{\mathcal{D}} \frac{\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')}\,\delta^{\mathrm{E}} h(\mathbf{x}')\,\mathrm{d}\mathbf{x}'\;. \tag{8.9}$$

The definition of $\mathbf{R}^{\mathrm{C}}$ is still $\mathbf{R}^{\mathrm{C}} = \mathbf{F} - \mathbf{A}^{\mathrm{C}}$ but now with parent force per unit mass

$$F_i(\mathbf{x}) \;=\; -\frac{1}{h_0}\,\frac{\delta\mathsf{V}}{\delta^{\mathrm{L}} x_i(\mathbf{a})} \;=\; -\frac{\partial}{\partial x_i}\,\frac{\delta\mathsf{V}}{\delta^{\mathrm{E}} h(\mathbf{x})} \tag{8.10}$$

and constraint acceleration

$$\begin{aligned} A_i^{\mathrm{C}}(\mathbf{x}) \;&=\; \int_{\mathcal{D}} u_j^{\mathrm{C}}(\mathbf{a}';\mathbf{x}(\cdot))\,\frac{\delta^{\mathrm{L}} u_j^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')}\,\mathrm{d}\mathbf{a}' \\ &=\; u_j^{\mathrm{C}}\frac{\partial u_i^{\mathrm{C}}}{\partial x_j} - \int_{\mathcal{D}} \frac{\partial\{h(\mathbf{x}')u_j^{\mathrm{C}}(\mathbf{x}';h(\cdot))\}}{\partial x_j'}\,\frac{\delta^{\mathrm{E}} u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')}\,\mathrm{d}\mathbf{x}'\;, \end{aligned} \tag{8.11}$$

using (8.7) and the boundary condition of evanescence at infinity to rewrite $\mathbf{A}^{\mathrm{C}}$ in terms of Eulerian functional derivatives. Because the integral in (8.8) is over $\mathrm{d}\mathbf{a}'$, which is h_0^{-1} times the mass element, the Lagrangian form of the splitting equation is $\int \Omega_{ij}^{\mathrm{CL}} u_j^{\mathrm{S}}\,\mathrm{d}\mathbf{a}' = \int h_0^{-1}\Omega_{ij}^{\mathrm{CL}} u_j^{\mathrm{S}}\,\mathrm{d}m' = R_i^{\mathrm{C}}$, with $R_i^{\mathrm{C}} = F_i - A_i^{\mathrm{C}}$, corresponding directly to (6.4). Defining, therefore, $\Omega_{ij}^{\mathrm{C}}(\mathbf{x},\mathbf{x}') = h_0^{-1}\Omega_{ij}^{\mathrm{CL}}(\mathbf{a},\mathbf{a}')$, and again using (8.7), we arrive at the Eulerian form of the splitting equation:

$$\boxed{\int_{\mathcal{D}} \Omega_{ij}^{\mathrm{C}}(\mathbf{x},\mathbf{x}')u_j^{\mathrm{S}}(\mathbf{x}')\,\mathrm{d}m(\mathbf{x}') = (\boldsymbol{\zeta}^{\mathrm{C}}\times\mathbf{u}^{\mathrm{S}})_i + \int_{\mathcal{D}} \omega_{ij}^{\mathrm{C}}(\mathbf{x},\mathbf{x}')u_j^{\mathrm{S}}(\mathbf{x}')\,\mathrm{d}m(\mathbf{x}') = R_i^{\mathrm{C}}(\mathbf{x}),} \tag{8.12}$$

where

$$\omega_{ij}^{\mathrm{C}}(\mathbf{x},\mathbf{x}') \;=\; -\omega_{ji}^{\mathrm{C}}(\mathbf{x}',\mathbf{x}) \;=\; \frac{\partial}{\partial x_j'}\frac{\delta^{\mathrm{E}} u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')} - \frac{\partial}{\partial x_i}\frac{\delta^{\mathrm{E}} u_j^{\mathrm{C}}(\mathbf{x}';h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x})}\;, \tag{8.13}$$

which, it can be noted, is reference-frame-independent because the Eulerian variation δ^{E} of a constant, solidly-rotating velocity field is trivially zero, and where we have defined

$$\boldsymbol{\zeta}^{\mathrm{C}} \;=\; \hat{\mathbf{z}}\, h\, Q^{\mathrm{C}}\;. \tag{8.14}$$

This is just the absolute 'constraint vorticity', i.e. the curl of the (absolute, inertial-frame) $\mathbf{u}^{\mathrm{C}}$ field. It comes from the delta-function term in (8.7), when substituted into the antisymmetric expression (8.1). In addition, we may note the following five points.

1. Despite having used the inertial frame of reference in the above derivation, we now have the problem in a form that is entirely reference-frame-independent. Not only are $\mathbf{R}^{\mathrm{C}}$ and $\mathbf{u}^{\mathrm{S}}$ frame-independent, but also $\omega_{ij}^{\mathrm{C}}(\mathbf{x},\mathbf{x}')$, as just noted. The constraint vorticity $\boldsymbol{\zeta}^{\mathrm{C}}$ is an absolute vorticity, by definition, because of its relation to the conserved PV, Q^{C}.

2. The dynamical effects of rotation enter solely through the $\mathbf{u}^{\mathrm{C}}$ functional or functionals, for instance through $\boldsymbol{\zeta}^{\mathrm{C}}$ and through the way in which the $\mathbf{u}^{\mathrm{C}}$ field changes when the mass and therefore pressure field changes. This is convenient when, for instance, it comes to applying (8.12) to variable-Coriolis-parameter models (MR96). The derivation makes no use of the present assumption that the Coriolis parameter f is constant.

3. Because $\mathbf{R}^{\mathrm{C}}$ is the residual force per unit mass, the splitting equation (8.12) says that we may think of $\mathbf{u}^{\mathrm{S}}$ as a correction to $\mathbf{u}^{\mathrm{C}}$. That is, when $\mathbf{u}^{\mathrm{C}}$ falls short of being as accurate as the slow quasimanifold permits, we may expect $\mathbf{u}^{\mathrm{P}} = \mathbf{u}^{\mathrm{C}} + \mathbf{u}^{\mathrm{S}}$ to be more accurate, as judged against

the standard provided by the parent model (cf. MN00). This expectation has been confirmed for a few specific examples, two of which are noted in Appendix A and another, at a higher level of approximation, in recent work by Wunderer (2001). If the term in $\boldsymbol{\zeta}^{\mathrm{C}}$ were the only term in the middle member of (8.12), then, with ϵ small so that ζ^{C} is dominated by f, when written in a rotating frame with Coriolis parameter f, the correction would be simply a velocity increment whose Coriolis force balances the force increment $\mathbf{R}^{\mathrm{C}}$. The actual correction also involves mass rearrangement, through the ω_{ij}^{C} terms.

4. The simple way in which the constrained Hamiltonian functional $\mathsf{H}^{\mathrm{C}}(h(\cdot))$ enters the foregoing derivation is partly due to the use of Lagrangian variations in the kinetic energy term on the right of (8.6), the mass element $\mathrm{d}m$ having zero Lagrangian variation. The potential energy term, $\delta\mathsf{V}$ in (8.6), with V given by (3.5), need not be computed here because we need only the parent force $\mathbf{F}$ per unit mass, already known as part of the elementary specification of the parent dynamics. As a check, though, it is easy to compute $\mathbf{F}$ from the last expression on the right of (8.10), i.e. from $\mathbf{F} = -\nabla\left(\delta\mathsf{V}/\delta^{\mathrm{E}}h\right)$, after rewriting (3.5) as $\mathsf{V} = \int \frac{1}{2}gh^2\,\mathrm{d}\mathbf{x}$ to exploit the fact that the area element $\mathrm{d}\mathbf{x}$ has zero Eulerian variation.

5. PV as well as energy conservation follow by standard arguments (also Appendix B below), provided that the PV is defined by Q^{C} and the energy by H^{C}, as in (8.2), and provided also that, in the case of energy conservation in bounded domains, information about boundary conditions, implicit in (7.4) and (8.12), is used (MR96). Of course in an unbounded domain H^{C} is numerically infinite, even though its variations need not be, and so it is natural to go back to the rotating frame, in which evanescent boundary conditions apply.

9 Canonical coordinates and PV inversion

The Hamiltonian balanced model defined by the $\mathbf{u}^{\mathrm{C}}$ functional and the associated splitting equation (8.12) is in noncanonical form, just as was the related toy problem with its noncanonical symplectic form $\Omega^{\mathrm{C}} = \frac{1}{2}\Omega_{ij}^{\mathrm{C}}\,\mathrm{d}x_i \wedge \mathrm{d}x_j$; see (7.36)–(7.37). If canonical coordinates $(X, Y) = (X_1, X_2)$ are found, then $\Omega_{ij}^{\mathrm{CL}}(\mathbf{a}, \mathbf{a}')$ in (8.1) will simplify (cf. minus signs in (7.37)) to

$$\begin{pmatrix} 0 & \delta(\mathbf{a}-\mathbf{a}') \\ -\delta(\mathbf{a}-\mathbf{a}') & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \delta(\mathbf{a}-\mathbf{a}')\ , \tag{9.1}$$

where δ with argument $(\mathbf{a}-\mathbf{a}')$ is again the 2-dimensional Dirac delta function and where rows and columns correspond to i and j respectively. This can be thought of as a symplectic matrix in canonical form, with infinite-dimensional identity submatrices. (In the toy problem the delta functions would be replaced by 1's.) Then (4.16) will generalize to

$$\dot{X} \;=\; \frac{\delta\mathsf{H}^{\mathrm{C}}}{\delta^{\mathrm{L}}Y}\;, \qquad \dot{Y} \;=\; -\,\frac{\delta\mathsf{H}^{\mathrm{C}}}{\delta^{\mathrm{L}}X}\;; \tag{9.2}$$

cf. Salmon (1985), p. 469, remembering that here we are still in the inertial frame. The functional derivatives refer to mass-configuration rearrangements as before, X and Y being mass-configuration functionals as stated explicitly in (10.1) below.

In the case of singly-split models, there will be a corresponding formal simplification of the splitting equation. Darboux's theorem for finite-dimensional phase spaces prompts the speculation that canonical coordinates may indeed exist for a general class of $\mathbf{u}^{\mathrm{C}}$ functionals, on the infinite-dimensional constraint manifold $\mathcal{M}^{\mathrm{C}} = \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \cdots$, though not necessarily given by any simple analytical formula.

However, as was first shown in MR96, analytically simple, explicitly defined canonical coordinates $(X, Y) = (X_1, X_2)$ do exist for a certain class of $\mathbf{u}^{\mathrm{C}}$ functionals. That class includes Salmon's constraint (3.7), in which case the canonical coordinates coincide with Hoskins' geostrophic coordinates. The canonical coodinates exist also in the other cases mentioned in Section 3, which include Salmon's L_1 dynamics by implication, the '$\sqrt{3}$ model' first described in MR96 (see below), and an infinite-dimensional infinity of other Hamiltonian balanced models, to be characterized in two lemmas below. In every one of these cases the conserved PV, Q^{C}, is given by a simple Jacobian formula analogous to that discovered by Hoskins (1975) for semigeostrophic theory:

$$\boxed{\;Q^{\mathrm{C}} \;=\; \frac{f}{h}\,\frac{\partial(X,Y)}{\partial(x,y)}\;.\;} \tag{9.3}$$

That is, the transformation of (3.6) or (3.8) into (4.17) is just one special case among an infinity of others, all following the same pattern. For comparison with (4.17), recall that

$$\phi(\mathbf{x},t) \;=\; \frac{g}{f^2}\,h(\mathbf{x},t)\;. \tag{9.4}$$

An immediate consequence of (9.3) is that every one of these cases shares with semigeostrophic theory the 'streamfunction property' expressed by (4.16). In other words, the flow appears solenoidal when viewed in (X, Y) space,

$$\frac{\partial\dot{X}}{\partial X} + \frac{\partial\dot{Y}}{\partial Y} \;=\; 0\;, \tag{9.5}$$

where the dots denote material derivatives or rates of change as before. This is because (9.3) and the material conservation of Q^{C} imply that the mass element in (X, Y) space is proportional to the area element. There must therefore exist a streamfunction $f\Phi(X, Y, t)$ (which is also a mass-configuration functional) such that the functional derivatives on the right of (9.2) are equal to f times the partial derivatives of $\Phi(X, Y, t)$. That is, $\dot{X} = \delta\mathsf{H}^{\mathrm{C}}/\delta^{\mathrm{L}}Y = -f\partial\Phi/\partial Y$ and $\dot{Y} = -\delta\mathsf{H}^{\mathrm{C}}/\delta^{\mathrm{L}}X = f\partial\Phi/\partial X$.

The Jacobian form of (9.3) arises in all these cases for a simple reason. It reflects the symplectic structure inherited from the infinite-dimensional parent phase space through restriction of the parent symplectic 2-form to the constraint manifold $\mathcal{M}^{\mathrm{C}} = \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \cdots$, when $\mathbf{x}$ is replaced by the canonical coordinates $\mathbf{X}$. The remarks at the end of Section 6 are relevant here, as is the further discussion in Appendix B, along with reference to a forthcoming paper by Bridges *et al.* (2001). The latter makes use of a multi-symplectic formalism that deals directly with the projection of the symplectic structure from $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes ...$ into each individual $\mathbb{R}^2$, thereby avoiding the standard summation over mass elements.

So how are the coordinates $\mathbf{X}$ defined? We first state the result for the particular $\mathbf{u}^{\mathrm{C}}$ functionals mentioned in Section 3. It will prove convenient to stay in the inertial frame of reference and to rewrite those functionals in their generic form

$$\mathbf{u}^{\mathrm{C}} \;=\; \tfrac{1}{2} f\,\hat{\mathbf{z}} \times \mathbf{x} \;+\; \mathbf{u}^{\mathrm{G}} \;+\; \alpha\,\hat{\mathbf{z}} \times f^{-1}\mathbf{u}^{\mathrm{G}} \cdot \nabla \mathbf{u}^{\mathrm{G}}\;, \tag{9.6}$$

in which, however, $\mathbf{u}^{\mathrm{G}}$ retains its original meaning, defined by (3.2), as the geostrophic velocity relative to a frame rotating with angular velocity $(0, 0, \frac{1}{2}f)$. The constant number α can take any real value. The cases $\alpha = 0,\ -\frac{1}{2}$, and 1 correspond to (3.2), (3.7), and (3.10), rewritten in the inertial frame — respectively representing Salmon's L_1 dynamics, semigeostrophic theory, and the $\surd 3$ model. For all such $\mathbf{u}^{\mathrm{C}}$, MR96 discovered to their great surprise that the simple formula

$$\boxed{\;\mathbf{X} \;=\; \mathbf{x} \;+\; \nabla\phi \;-\; ic\,\hat{\mathbf{z}} \times \nabla\phi\;,\;} \tag{9.7}$$

generalizing Hoskins' transformation $\mathbf{X} = \mathbf{x} + \nabla\phi$, gives canonical coordinates for the balanced model provided that

$$c \;=\; \surd(2\alpha + 1)\;. \tag{9.8}$$

The derivation is reproduced in Appendix C below. We recall that (9.6) and (9.7) need have no connection with the balance condition or momentum–configuration constraint that is applied to the Hamiltonian functional. It is enough that (9.6) is the constraint applied to the parent symplectic structure. This follows from the remarks near the end of Section 6 and in Appendix B. A corollary is that semigeostrophic theory is indeed included, as the case

$\alpha = -\frac{1}{2}$, for which $c = 0$. The $\sqrt{3}$ model has $\alpha = 1$, therefore $c = \sqrt{3}$, hence the name. In all these cases the coordinates (9.7) are canonical.

The factor i in (9.7) is not a misprint, and it does mean $\sqrt{(-1)}$. For all cases of (9.7) more accurate than Salmon's constraint (3.7), the case $\alpha > -\frac{1}{2}$, the canonical coordinates $\mathbf{X}$ are complex-valued! Greater accuracy requires $\alpha > -\frac{1}{2}$, hence real c. One consequence is that both the real and the imaginary parts of (9.5) must vanish. In (9.5), $\dot{X}$ and $\dot{Y}$ must of course be read as holomorphic functions of X and Y, satisfying the appropriate Cauchy–Riemann relations. We will take $c \geqslant 0$ for definiteness.

It is easy to check by direct substitution that the Jacobian in (9.3) is real-valued, as it must be. Also real-valued (Appendix C) is the corresponding infinite-dimensional symplectic 2-form Ω^{CL} in phase space — (9.3) being essentially its 2-dimensional reflection, or projection on an individual $\mathbb{R}^2$ (see also (13.8) below). Furthermore, as pointed out by Roubtsov & Roulstone (1997), substitution of (9.7) into (9.3), cf. (4.1) into (4.17), produces a (real) Monge–Ampère equation that is elliptic, in circumstances of interest (ϵ small), as noted next. This equation can be solved for ϕ when Q^{C} is given. The fact that a Monge–Ampère equation is obtained is evident at once from the general form of the Monge–Ampère equation in 2 dimensions — recall (4.22)–(4.23) — together with the fact that (9.7) involves a linear combination of gradients of ϕ with respect to (x, y). In fact the equation obtained is (4.22) with $B = D = 1$, $C = 0$, as before, and analogously with the semigeostrophic case $A = 1 - g^{-1} f \phi Q_{\mathrm{S}}^{\mathrm{C}}$ we have $A = 1 - g^{-1} f \phi Q^{\mathrm{C}} = 1 - f^{-1} h Q^{\mathrm{C}} = 1 - f^{-1} \zeta^{\mathrm{C}}$. But $E \neq 1$; in fact, we now have $E = 1 - c^2$. That is, the Monge–Ampère equation for arbitrary c is

$$1 + \nabla^2\phi + (1 - c^2)\,\mathrm{hess}_{xy}(\phi) \;=\; \zeta^{\mathrm{C}}/f\ . \tag{9.9}$$

The ellipticity criterion is $0 < BD - AE = 1 - (1 - f^{-1}\zeta^{\mathrm{C}})(1 - c^2)$, that is,

$$(c^2 - 1)\frac{\zeta^{\mathrm{C}}}{f} \;<\; c^2\ . \tag{9.10}$$

This is always satisfied in the L_1-dynamics case, $\alpha = 0$ and $c = 1$, and is satisfied for positive ζ^{C}/f in the semigeostrophic case, $\alpha = -\frac{1}{2}$ and $c = 0$, as already noted. In the case of the $\sqrt{3}$ model, $\alpha = 1$ and $c = \sqrt{3}$, we have ellipticity whenever $\zeta^{\mathrm{C}}/f < 3/2$. Thus in all three cases (and also, in fact, in every intermediate case $0 < c < \sqrt{3}$ and beyond) the Monge–Ampère equation is elliptic over a range of circumstances much wider than the physically relevant range, small ϵ with $\zeta^{\mathrm{C}}/f = 1 + O(\epsilon)$. That is, we have ellipticity *a fortiori* whenever the constraint manifold $\mathcal{M}^{\mathrm{C}}$ defined by (9.6) approximates the actual slow quasimanifold within the parent phase space, which is possible only when ϵ is small.

The Monge–Ampère equation (9.9) gives us a second way of timestepping any of the balanced models in question after solving the splitting equation (8.12). This second way, which is likely to be the better-conditioned numerically, is to use the $\mathbf{u}^{\mathrm{P}}$ field to advect not the mass configuration h but Q^{C} instead, then invert the Monge–Ampère equation to obtain ϕ and h. The two ways are equivalent in the absence of numerical truncation errors, because of the exact material conservation of Q^{C}, on particles moving with velocity $\mathbf{u}^{\mathrm{P}}$. However, under the circumstances of interest, with ϵ small, it is likely that the second way would be less sensitive to truncation error than the first. For small ϵ, the $\mathbf{u}^{\mathrm{C}}$ functional is such that slight errors in the mass field have large, $O(\epsilon^{-1})$ effects on the velocity field relative to the rotating frame. This is analogous to large, $O(\epsilon^{-1})$ slopes of the toy problem's $\mathcal{M}^{\mathrm{C}}$ surface in Figure 1. For given ϵ there will be a tradeoff between such sensitivity and the resulting ill-conditionedness, on the one hand, and the cost of solving two elliptic problems instead of one at each timestep, on the other.

10 Generalizations including variable Coriolis parameter

To see the full generality of (9.3) and its consequences, we now ask what class of general, nonlocal $\mathbf{u}^{\mathrm{C}}$ functionals admits canonical coordinates $\mathbf{X}$. Canonical coordinates on the slow manifold $\mathcal{M}^{\mathrm{C}}$, in the most general possible sense, are not only fields but also mass-configuration functionals that respect the particle-relabelling symmetry, like $\mathbf{u}^{\mathrm{C}}$. In our shorthand notation,

$$\mathbf{X} = \mathbf{X}(\mathbf{a}; \mathbf{x}(\cdot)) = \mathbf{X}(\mathbf{x}; h(\cdot)) \ . \tag{10.1}$$

On the right, we can still read this as a transformation within $\mathcal{D} = \mathbb{R}^2$, the physical domain, provided we recognize that there is a different such transformation for each mass configuration $h(\cdot)$. This of course is already true of (9.7), because of the near-local dependence on $\nabla\phi$ and hence on ∇h; but now the dependence on $h(\cdot)$ can be fully nonlocal, as it has to be for the greatest possible accuracy.

The components $(x, y) = (x_1, x_2)$ of $\mathbf{x}$ in (10.1) do not, incidentally, need to be components referred to Cartesian axes. For instance spherical components could be used, as in the theories of Shutts (1989) and Purser (1993) extending semigeostrophic theory to the sphere. This is because the mathematical objects involved belong to the exterior calculus and are thus metric-independent. The connection between Shutts' equation and the theory of contact transformations was established by Roulstone & Sewell (1997, Section 7).

The coordinates $\mathbf{X}$ are canonical by definition if, for some scalar-valued mass-configuration functional, say $\mathsf{B}(\mathbf{x}(\cdot)) = \mathsf{B}(h(\cdot))$,

$$-\int_{\mathcal{D}} \mathrm{d}m\, u_i^{\mathrm{C}} \delta^{\mathrm{L}} x_i \ = \ \tfrac{1}{2} f \int_{\mathcal{D}} \varepsilon_{jk} X_k \delta^{\mathrm{L}} X_j \, \mathrm{d}m + \delta\mathsf{B} \ , \tag{10.2}$$

where ε_{ik} is the 2-dimensional alternating tensor ($\varepsilon_{12} = -\varepsilon_{21} = 1$, $\varepsilon_{11} = \varepsilon_{22} = 0$), and where f can still be taken to be the Coriolis parameter, in the constant-f cases considered so far, but where, more generally, $\frac{1}{2}f$ can be regarded as an arbitrary constant normalization factor introduced to give $\mathbf{X}$ the dimensions of length. The factor $\frac{1}{2}f$ will usually, but need not, be identified with some typical or average angular velocity of the spinning mass of fluid, as hitherto. (We are still in the inertial frame of reference.)

The left-hand side of (10.2) is analogous to the toy-problem expression $\theta_j^{\mathrm{C}}(\mathbf{x})\delta x_j = -u_j^{\mathrm{C}}(\mathbf{x})\delta x_j$. That is, the left-hand side of (10.2) is the shallow-water counterpart of the 1-form θ^{C} displayed in (7.23), after contraction or inner multiplication with the infinite-dimensional vector field $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$, if we omit the contribution corresponding to the arbitrary function ϕ^{C} in (7.23). Thus, before contraction, the left-hand side of (10.2) itself would correspond to a 1-form θ^{CL}, say, defined as $-\int_{\mathcal{D}} \mathrm{d}m\ u_j^{\mathrm{C}}\,\mathrm{d}^{\mathrm{L}}x_j$ where d^{L} is the exterior derivative in phase space, again omitting any contribution analogous to ϕ^{C} in (7.23) — effectively absorbing it into the B term. What is important here is that the exterior derivative of θ^{CL} is nothing but the symplectic 2-form $\Omega^{\mathrm{CL}} = \int_{\mathcal{D}} \mathrm{d}m\ \mathrm{d}^{\mathrm{L}}x_j \wedge \mathrm{d}^{\mathrm{L}}u_j^{\mathrm{C}}$ whose coefficients are given by (8.1). This transforms to $\Omega^{\mathrm{CL}} = \frac{1}{2}f\int_{\mathcal{D}} \mathrm{d}m\ \varepsilon_{jk}\,\mathrm{d}^{\mathrm{L}}X_k \wedge \mathrm{d}^{\mathrm{L}}X_j = -f\int_{\mathcal{D}} \mathrm{d}m\ \mathrm{d}^{\mathrm{L}}X \wedge \mathrm{d}^{\mathrm{L}}Y$, as can be seen by taking the exterior derivative of the 1-form $\frac{1}{2}f\int_{\mathcal{D}} \mathrm{d}m\ \varepsilon_{jk}X_k\,\mathrm{d}^{\mathrm{L}}X_j$. The minus sign is related to the presence of the vector product in the first term on the right of (9.6).

The desired general characterization of $\mathbf{u}^{\mathrm{C}}$ functionals now follows, expressed as absolute velocities, i.e. as velocities in the inertial frame:

Lemma 1 (Canonical coordinate lemma) *General canonical coordinates* $\mathbf{X}(\mathbf{x}; h(\cdot))$ *are functionally related to absolute (inertial-frame) constraint velocities* $\mathbf{u}^{\mathrm{C}}$ *by*

$$u_i^{\mathrm{C}} = -\tfrac{1}{2}f\varepsilon_{jk}X_k\frac{\partial X_j}{\partial x_i} - \frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x_i}\left\{\tfrac{1}{2}f\varepsilon_{jk}\int_{\mathcal{D}} X_k(\mathbf{x}'; h(\cdot))\frac{\delta^{\mathrm{E}}X_j(\mathbf{x}'; h(\cdot))}{\delta^{\mathrm{E}}h(\mathbf{x})}\,\mathrm{d}m(\mathbf{x}') + \frac{\delta\mathsf{B}(h(\cdot))}{\delta^{\mathrm{E}}h(\mathbf{x})}\right\} \tag{10.3}$$

where the mass-configuration functional B *may be chosen arbitrarily.*

The boldface notation $\boldsymbol{\partial}/\boldsymbol{\partial}x_i$ signifies differentiation acting on the ith component of $\mathbf{x}$ wherever it occurs, implicitly or explicitly — in this case the dependence on $\mathbf{x}$ associated with the denominators of the two functional derivatives. By contrast, the operator $\partial/\partial x_i$ in the first term acts on the ith component of the first argument of $X_j(\mathbf{x}; h(\cdot))$ only.

The proof of (10.3) is a straightforward application of the same Eulerian–Lagrangian functional differentiation machinery as before, especially (8.4), (8.5), and (8.7). Direct substitution shows, after a few lines of manipulation, that (10.3) satisfies (10.2). Arbitrariness of $\delta^{\mathrm{L}}\mathbf{x}$ takes us back to (10.3).

Our main result (9.7) now follows after specializing to the case of near-local canonical models. Both the $\mathbf{u}^{\mathrm{C}}$ functional and the canonical coordinates $\mathbf{X}$ are now taken to have near-local form. That is, $\mathbf{u}^{\mathrm{C}}$ and $\mathbf{X}$ are taken to be general pointwise functions of the layer-depth $h(\mathbf{x})$ and a finite number of its derivatives:

$$\mathbf{u}^{\mathrm{C}}(\mathbf{x}, h(\cdot)) \;=\; \mathbf{u}^{\mathrm{C}}(x_j, h, h_{,i}, h_{,ij}, \cdots) \;, \tag{10.4}$$

$$\mathbf{X}(\mathbf{x}, h(\cdot)) \;=\; \mathbf{X}(x_j, h, h_{,i}, h_{,ij}, \cdots) \;, \tag{10.5}$$

where $h_{,i} = \partial h(\mathbf{x})/\partial x_i$ etc. For consistency the mass-configuration functional B is taken in the corresponding form $\mathsf{B} = \int_{\mathcal{D}} B \; \mathrm{d}m$, where the integrand

$$B(\mathbf{x}, h(\cdot)) = B(x_j, h, h_{,i}, h_{,ij}, \cdots) \;. \tag{10.6}$$

When the number of derivatives is arbitrary but finite, these are the most general possible forms expressing near-local functional dependence. Substitution into (10.3) gives

Lemma 2 (Canonical coordinate lemma, near-local version) *Near-local canonical coordinates* $\mathbf{X}(x_j, h, h_{,i}, h_{,ij}, \cdots)$ *are functionally related to absolute (inertial-frame) constraint velocities* $\mathbf{u}^{\mathrm{C}}$ *by*

$$\begin{aligned} u_i^{\mathrm{C}} = &-\tfrac{1}{2} f \varepsilon_{jk} X_k \frac{\partial X_j}{\partial x_i} \\ &-\frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x_i}\left\{ \tfrac{1}{2} f \varepsilon_{jk}\left[hX_k \frac{\partial X_j}{\partial h} - \frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x_p}\left(hX_k \frac{\partial X_j}{\partial h_{,p}}\right) + \frac{\boldsymbol{\partial}^2}{\boldsymbol{\partial} x_p \boldsymbol{\partial} x_q}\left(hX_k \frac{\partial X_j}{\partial h_{,pq}}\right) - \cdots \right] \right. \\ &\qquad \left. + \left(B + h\frac{\partial B}{\partial h} - \frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x_j}\left(h \frac{\partial B}{\partial h_{,j}}\right) + \frac{\boldsymbol{\partial}^2}{\boldsymbol{\partial} x_j \boldsymbol{\partial} x_k}\left(h\frac{\partial B}{\partial h_{,jk}}\right) - \cdots \right) \right\} \end{aligned} \tag{10.7}$$

for near-local but otherwise arbitrary $B(x_j, h, h_{,i}, h_{,ij}, \cdots)$.

In particular, by noting that (9.7) is the simplest choice compatible with the above and compatible also with invariance to coordinate-axis rotations, and by making a correspondingly simple choice of B and substituting both into (10.7), then simplifying the resulting expression, which is lengthy — keeping in mind the distinction between ∂ and $\boldsymbol{\partial}$ — we recover (9.6) and (9.8). The details are summarized in Appendix C. We note incidentally that these lemmas answer the old question of how to find Hamiltonian balanced models with exact canonical coordinates for *variable Coriolis parameter*; cf., e.g., §3 of Salmon (1985), and Magnusdottir & Schubert (1990). For instance $\mathbf{x}$ in (9.7) can be replaced by $\nabla\Pi(\mathbf{x})$ with any function $\Pi(\mathbf{x})$ that satisfies $\mathrm{hess}_{xy}(\Pi) = \tilde{f}(\mathbf{x})/f$, where $\tilde{f}(\mathbf{x})$ is the variable Coriolis parameter and f keeps its role as a constant normalizing factor. This is straightforward to verify from lemma 2 above; details are in MR96 §10.

11 $\sqrt{3}$ models: quo vadis?

As mentioned below (3.10) in Section 3, it is the choice $\alpha = 1$ in (9.6), corresponding to $c = \sqrt{3}$ in (9.7), that gives the balance condition that is formally the most accurate of that class. This choice defines the '$\sqrt{3}$ model', or rather, as implied by the remarks at the end of Section 6 and in Appendix B, an infinite family of '$\sqrt{3}$ models'. Their mathematical properties are largely unknown, apart from the generic, purely formal properties already mentioned — which they share with semigeostrophic theory and L_1 dynamics, for which $\alpha = -\frac{1}{2}$, 0, and $c = 0$, 1, respectively. These generic properties are first the existence of complex canonical coordinates $\mathbf{X}$, second the Jacobian formula (9.3) for Q^{C}, third the Monge–Ampère equation (9.9) for ϕ given Q^{C}, and fourth the streamfunction property analogous to (4.16), arising from the solenoidality property (9.5).

The same four generic properties will be shared by all Hamiltonian balanced models in which (9.6) constrains the symplectic structure, even if not the Hamiltonian. If we constrain only the symplectic structure by (9.6), producing a doubly-split model, then we lose the splitting equation in its simplest form (8.12). But it is now emerging such loss of formal simplicity might carry with it a compensating gain in terms of good mathematical behaviour. Recent work by Wunderer (2001) has shown that in the case of the singly-split $\sqrt{3}$ model, with (9.6) constraining both the Hamiltonian and the symplectic structure, the splitting equation fails to share with Salmon's equation for L_1 dynamics, (6.5) above, the property of being robustly invertible.

All constraints that are near-local produce splitting equations in the form of linear partial differential equations, because the functional derivatives in (8.13) then reduce to delta functions and their derivatives, as illustrated in Appendix A. As shown there, the second derivatives in Salmon's equation (6.5) arise from the first derivatives in the geostrophic constraint (3.2). Similarly, the splitting equation for the $\sqrt{3}$ and other singly-split models arising solely from (9.6) all involve third derivatives, whenever $\alpha \neq 0$ and $c \neq 1$, because of the second derivatives appearing in (9.6). When the $\sqrt{3}$ model was first discovered, we thought that these third derivatives would cancel and give us another second-order elliptic equation. But Wunderer's work has shown, on the contrary, that the sum of the third-derivative terms is nonzero. Therefore the $\sqrt{3}$ model's splitting equation cannot be elliptic. In this context, therefore, the correction term with $\alpha = 1$ in (9.6) has given rise to a singular perturbation, in striking contrast with the benign effect of the same term in the case of semigeostrophic theory, with $\alpha = -\frac{1}{2}$.

One way to regularize the model without losing Hamiltonian structure might be to add a higher correction to (9.6), involving third derivatives. Then the splitting equation would involve fourth derivatives and could be elliptic, though formidably complicated. Alternatively, as already hinted, we could abandon singly-split models. We could construct regularized variants of the

$\sqrt{3}$ model by keeping (9.6) for the purpose of constraining the symplectic structure — thus keeping the simple formula (9.7) for $\mathbf{X}$ — but changing the constraint on the Hamiltonian, i.e. introducing double splitting. This is another way to produce models sharing with semigeostrophic theory the four generic properties listed above while attaining greater accuracy than semigeostrophic theory. The extra freedom gained via double splitting can be used for regularization purposes.

Indeed, it can only be thus that semigeostrophic theory itself avoids the pathology discovered by Wunderer. As we have emphasized, despite having a nonzero value of α semigeostrophic theory is a supremely regular, mathematically well-behaved theory. We have some specific ideas that should lead to well-behaved models within the (infinite) family of doubly-split $\sqrt{3}$ models, but at the time of going to press those ideas remain to be verified in detail.

12 Complex contact structure

What other properties are shared between semigeostrophic theory and the models just considered, including the doubly-split variants? For instance, is the transformation $(x, y) \mapsto (X, Y)$ defined by (9.7) part of an explicitly invertible contact transformation? The answer, frustratingly, is almost certainly not. This limits the usefulness of (9.7), because to make practical use of the canonical coordinates (X, Y) we need the transformation inverse to (9.7). The forward transformation (9.7) gives $\mathbf{X}(\mathbf{a})$ if the Lagrangian mass configuration $\mathbf{x}(\mathbf{a})$ is known in the physical, (x, y) domain, hence h and ϕ known via (2.2)–(2.5) and (9.4). But if we solve the problem in the (X, Y) domain then we need the inverse transformation to get back to the (x, y) domain. This is an important difference *vis-à-vis* semigeostrophic theory, in which the inverse transformation is given explicitly by (4.14), by virtue of the contact structure and its symmetric generating function (5.18). In the absence of such structure, and given only $\mathbf{X}(\mathbf{a})$, inversion of (9.7) would be nontrivial, requiring solution of a nonlinear partial differential equation — nonlinear because of the nonlinearities in (2.2)–(2.5) — to yield a description of the mass configuration in the $\mathbf{x}$ domain given its description in the $\mathbf{X}$ domain. It is to this part of the problem, presumably, that the pathology found by Wunderer would migrate if we were to use (9.1) to simplify the splitting equation.

However, Roubtsov & Roulstone (2001, hereafter 'RR01') point out that the 'conjugate' of the transformation $(x, y) \mapsto (X, Y)$, defined as $(x, y) \mapsto (X, \bar{Y})$ where $\bar{Y}$ is the complex conjugate of Y, is part of an explicitly invertible contact transformation $(x,\, y,\, \phi,\, \partial\phi/\partial x,\, \partial\phi/\partial y) \mapsto (X,\, \bar{Y},\, \hat{\Phi},\, \partial\hat{\Phi}/\partial X,\, \partial\hat{\Phi}/\partial\bar{Y})$ continuous with $(x,\, y,\, \phi,\, \partial\phi/\partial x,\, \partial\phi/\partial y) \mapsto (X,\, Y,\, \Phi,\, \partial\Phi/\partial X,\, \partial\Phi/\partial Y)$, the transformation found for semigeostrophic theory and specified by (4.14) and

(4.15). The generating function, $\hat{S}$ say, has an extra term proportional to ic:

$$\hat{S}(x,y,X,\bar{Y}) \;=\; \frac{-\frac{1}{2}(X-x)^2 \;-\; \frac{1}{2}(\bar{Y}-y)^2 \;+\; ic(X-x)(\bar{Y}-y)}{1+c^2}\;. \tag{12.1}$$

This is still symmetric, like (5.18), with respect to exchanging (x,y) and $(X,\bar{Y})$. The new complex potential, replacing (4.15), is

$$\hat{\Phi} \;=\; \phi \;-\; \hat{S}\;. \tag{12.2}$$

The rest of the transformation — see also below (13.5) — is only slightly less simple than (4.14):

$$\frac{\partial\hat{\Phi}}{\partial X} = \frac{\partial\phi}{\partial x} = \frac{(X-x)-ic(\bar{Y}-y)}{1+c^2} \quad \text{and} \quad \frac{\partial\hat{\Phi}}{\partial\bar{Y}} = \frac{\partial\phi}{\partial y} = \frac{(\bar{Y}-y)-ic(X-x)}{1+c^2}\;. \tag{12.3}$$

The symmetry and explicit invertibility are now evident, just as before. If we have solved for the evolution in $(X,\bar{Y})$ space and know the function $\hat{\Phi}(X,\bar{Y})$, at each time t, then we know $\partial\hat{\Phi}/\partial X$ and $\partial\hat{\Phi}/\partial\bar{Y}$. We can then regard (12.3) as a pair of linear algebraic equations for x and y. Provided that the determinant

$$\begin{vmatrix} -1 & ic \\ ic & -1 \end{vmatrix} \;=\; 1+c^2 \;\neq\; 0\;, \tag{12.4}$$

which holds in all the cases of interest, $c \geqslant 0$, we can then deduce the physical position x,y of any particle from its image in $(X,\bar{Y})$ space, just as we could from (X,Y) in semigeostrophic theory. Thus knowledge of $\mathbf{X}(\mathbf{a})$ provides us with knowledge of $x(\mathbf{a})$, and therefore of the mass configuration, in a simple and explicit way. Alternatively, and again as in semigeostrophic theory, we can stay entirely within the Eulerian description. The formulae (12.1)–(12.3) tell us at once that the new potential $\hat{\Phi}(X,\bar{Y})$ contains, in easily recoverable form and with no question of pathology, the same information as $\phi(x,y)$ and therefore $h(x,y)$. This makes it plain that the function $\hat{\Phi}(X,\bar{Y})$ completely specifies the Eulerian mass configuration, i.e. specifies the mass configuration up to particle relabelling.

It follows that, in principle, the balanced model can be formulated entirely in terms of $\hat{\Phi}(X,\bar{Y})$ and the material derivatives or rates of change of X and $\bar{Y}$, just as semigeostrophic theory can be formulated entirely in terms of its transformed potential function $\Phi(X,Y)$ and the material derivatives of Hoskins' real X and Y. However, the new potential $\hat{\Phi}(X,\bar{Y})$ cannot be identified with the complex streamfunction $\Phi(X,Y)$ implied by (9.5), and there is therefore no reason to expect $\hat{\Phi}(X,\bar{Y})$ to enter into the evolution equations with anything like the simplicity of (4.16).

13 Kähler and hyper-Kähler structure

The foregoing remarks, together with those of Sections 9–10, present us with an intriguing, tantalizing, yet frustrating situation — a kind of parting of the ways forced on us as soon as $c > 0$, as is necessary if we are to gain more accuracy. In the more accurate models under consideration, some of the properties of semigeostrophic theory are echoed in the transformation $(x, y) \mapsto (X, Y)$. Others are echoed in the transformation $(x, y) \mapsto (X, \bar{Y})$, and the two are different whenever $c \neq 0$.

There is a suggestion here that in order to gain deeper insight we must consider both transformations together, $(x, y) \mapsto (X, Y)$ and $(x, y) \mapsto (X, \bar{Y})$, implying consideration of the subspace of $\mathbb{C}^4$ spanned by the four interdependent complex variables $X, Y, \bar{X}, \bar{Y}$. That space $\{X, Y, \bar{X}, \bar{Y}\}$ has the dimensionality of $\mathbb{R}^4$, being the image of the symplectic space or manifold $\{x, y, p, q\}$ discussed in Section 5, under the linear mapping

$$\left.\begin{array}{ll} X = x + p + icq\ , & Y = y + q - icp \\ \bar{X} = x + p - icq\ , & \bar{Y} = y + q + icp \end{array}\right\} . \tag{13.1}$$

This mapping corresponds to (9.7) and its complex conjugate when (p, q) replaces $(\partial\phi/\partial x,\ \partial\phi/\partial y)$. It becomes one-to-one as soon as $c \neq 0$. As noted by RR01, its Jacobian is simply

$$\frac{\partial(X, Y, \bar{X}, \bar{Y})}{\partial(x, y, p, q)} = -4c^2 \tag{13.2}$$

and its inverse

$$x = \frac{(X + \bar{X})}{2} + \frac{(Y - \bar{Y})}{2ic}\ , \quad y = \frac{(Y + \bar{Y})}{2} - \frac{(X - \bar{X})}{2ic}\ , \tag{13.3}$$

$$p = -\frac{(Y - \bar{Y})}{2ic}\ , \qquad q = \frac{(X - \bar{X})}{2ic}\ . \tag{13.4}$$

In what seems to be a natural way, the space $\{X, Y, \bar{X}, \bar{Y}\}$, when equipped with nothing but the 2-forms $\mathrm{d}X \wedge \mathrm{d}Y$ and $\mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y}$, reflects both of the symplectic structures that have arisen in the development so far. It does so in a manner to be explained next.

The first structure is that associated with the symplectic 2-form Ω of Section 5, and thereby also with the Cartan 1-form θ of Section 5 and the contact transformations of semigeostrophic theory and of (12.1)–(12.3). The second is that associated with the symplectic 2-form Ω^{CL} of Sections 6–10 and Appendices B and C, and thereby also with the 1-form θ^{CL} corresponding to the left-hand side of (10.2), inherited from the phase space of the parent dynamics. RR01 point out that the first symplectic structure defines what is called Kähler structure, when viewed in the space $\{X, Y, \bar{X}, \bar{Y}\}$, and that the two symplectic structures together form part of what is called hyper-Kähler structure in the space $\{X, Y, \bar{X}, \bar{Y}\}$. The hyper-Kähler structure can in turn be embedded within a 6-dimensional 'twistor space'. These are well-studied geometric structures (e.g. MS98; Hitchin 1987; Atiyah & Hitchin 1988 & refs.; Freed 1999, hereafter 'F99'), which may hold important clues toward further progress on our journey.

Consider the first of the two symplectic structures. The Ω of Section 5 was defined in (x, y, p, q) space as $\Omega = \mathrm{d}x \wedge \mathrm{d}p + \mathrm{d}y \wedge \mathrm{d}q$, and the corresponding Cartan 1-form in (x, y, ϕ, p, q) space as $\theta = \mathrm{d}\phi - p\,\mathrm{d}x - q\,\mathrm{d}y$. When transformed to $(X, Y, \bar{X}, \bar{Y})$ space and multiplied by $-c$, the Ω of Section 5 becomes simply

$$-c\,\Omega \;=\; \frac{1}{2i}\,(\mathrm{d}X \wedge \mathrm{d}Y - \mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y}) \;=\; \mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y)\;, \tag{13.5}$$

as is evident by inspection of (13.1). Alternatively, we may view this result in terms of the contact transformation (12.1)–(12.3), with (p, q) replacing $(\partial\phi/\partial x,\ \partial\phi/\partial y)$ and, say, $(\hat{P}, \hat{Q})$ replacing $(\partial\hat{\Phi}/\partial X,\ \partial\hat{\Phi}/\partial\bar{Y})$. We have $\hat{P} = p$ and $\hat{Q} = q$; therefore, from (13.4), $\hat{P} = -(Y - \bar{Y})/2ic$ and $\hat{Q} = (X - \bar{X})/2ic$. The contact transformation preserves the Cartan 1-form; therefore

$$\theta \;=\; \mathrm{d}\phi - p\,\mathrm{d}x - q\,\mathrm{d}y \;=\; \mathrm{d}\hat{\Phi} - \hat{P}\,\mathrm{d}X - \hat{Q}\,\mathrm{d}\bar{Y}\;, \tag{13.6}$$

the exterior derivative of which is

$$\begin{aligned} \Omega \;=\; \mathrm{d}\theta \;&=\; \mathrm{d}X \wedge \mathrm{d}\hat{P} + \mathrm{d}\bar{Y} \wedge \mathrm{d}\hat{Q} \\ &=\; -\,\mathrm{d}X \wedge \mathrm{d}\left(\frac{Y - \bar{Y}}{2ic}\right) \;+\; \mathrm{d}\bar{Y} \wedge \mathrm{d}\left(\frac{X - \bar{X}}{2ic}\right) \\ &=\; -(2ic)^{-1}\,(\mathrm{d}X \wedge \mathrm{d}Y + \mathrm{d}\bar{Y} \wedge \mathrm{d}\bar{X})\;, \end{aligned} \tag{13.7}$$

equivalent to (13.5).

The relation (13.5) exposes an aspect of the Jacobian formula (9.3) that was previously invisible. Multiplying the Jacobian $\partial(X, Y)/\partial(x, y)$ by $\mathrm{d}x \wedge \mathrm{d}y$ (cf. (5.1)) turns it into the 2-form $\mathrm{d}X \wedge \mathrm{d}Y$ under discussion,

$$\frac{\partial(X, Y)}{\partial(x, y)}\,\mathrm{d}x \wedge \mathrm{d}y \;=\; \mathrm{d}X \wedge \mathrm{d}Y\;, \tag{13.8}$$

when X and Y are expressed as functions of x and y as in (9.7). That is, (13.8) holds under restriction to the graph of ϕ. When we view things in 4 dimensions — as distinct from 5 dimensions as in (5.11) and (5.12) — restriction to the graph of ϕ simply means restriction to the image in $(X, Y, \bar{X}, \bar{Y})$ space of the 2-dimensional surface defined in (x, y, p, q) space by $(x, y, p, q) = (x,\ y,\ \partial\phi(x, y)/\partial x,\ \partial\phi(x, y)/\partial y)$. For any smooth $\phi(x, y)$, the surface in question is a Lagrangian submanifold with respect to Ω, as was pointed out in the paragraph below (5.12). That is, Ω vanishes identically when restricted to the graph of ϕ. Consequently, $\mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y)$ becomes invisible when we restrict to the graph — $\mathrm{d}X \wedge \mathrm{d}Y$ then being real, like the Jacobian itself — even though $\mathrm{d}X \wedge \mathrm{d}Y$ is plainly complex-valued at a general point in $(X, Y, \bar{X}, \bar{Y})$ space.

Now the structure or geometry imposed on the space $\{X, Y, \bar{X}, \bar{Y}\}$ by the 2-form $(2i)^{-1}(\mathrm{d}X \wedge \mathrm{d}Y - \mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y})$, i.e. by $\mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y)$, is a simple example of what is called Kähler structure or geometry (e.g. MS98; F99). In general this can be defined using a real-valued 'Kähler potential' $\mathcal{K}(X, Y, \bar{X}, \bar{Y})$, which in our case will turn out to be related, in a peculiar way, to the complex-valued potential $\hat{\Phi}(X, \bar{Y})$ that arose in the contact transformation (12.1)–(12.3). By construction, $\hat{\Phi}(X, \bar{Y})$ is automatically a holomorphic function of each of its arguments when everything is restricted to the graph of ϕ, as pointed out in RR01. The associated Cauchy–Riemann relations state that

$$\frac{\partial \hat{\Phi}}{\partial \bar{X}} = \frac{\partial \hat{\Phi}}{\partial Y} = 0 \qquad \text{on the graph of } \phi\ . \tag{13.9}$$

The reader is warned that, in the literature on Kähler geometry, $\bar{Y}$ is notated Y and *vice versa.*

Specifically, the space $\{X, Y, \bar{X}, \bar{Y}\}$ is said to have Kähler structure when equipped with a closed *Kähler 2-form* ω_K defined by

$$\omega_K = \tfrac{1}{2} i\, \partial\, \bar{\partial} \mathcal{K}(X, Y, \bar{X}, \bar{Y})\ , \tag{13.10}$$

where ∂ and $\bar{\partial}$ are exterior derivative operators restricted, respectively, to the subspaces $\{X, \bar{Y}\}$ and $\{\bar{X}, Y\}$. Thus the first operation $\bar{\partial}$ produces (in our non-standard notation) the 1-form

$$\bar{\partial} \mathcal{K} = \frac{\partial \mathcal{K}}{\partial \bar{X}}\, \mathrm{d}\bar{X} + \frac{\partial \mathcal{K}}{\partial Y}\, \mathrm{d}Y\ , \tag{13.11}$$

with no terms in $\mathrm{d}X$ and $\mathrm{d}\bar{Y}$. The second operation ∂ converts this into the 2-form

$$\partial\, \bar{\partial} \mathcal{K} = \frac{\partial^2 \mathcal{K}}{\partial X \partial \bar{X}}\, \mathrm{d}X \wedge \mathrm{d}\bar{X} + \frac{\partial^2 \mathcal{K}}{\partial X \partial Y}\, \mathrm{d}X \wedge \mathrm{d}Y + \frac{\partial^2 \mathcal{K}}{\partial \bar{Y} \partial \bar{X}}\, \mathrm{d}\bar{Y} \wedge \mathrm{d}\bar{X} + \frac{\partial^2 \mathcal{K}}{\partial \bar{Y} \partial Y}\, \mathrm{d}\bar{Y} \wedge \mathrm{d}Y\ . \tag{13.12}$$

The first and last terms on the right are each pure imaginary (because of the antisymmetry of wedge products), as is the sum of the remaining two terms. Hence the Kähler 2-form $\omega_K = \frac{1}{2} i\, \partial\, \bar{\partial} \mathcal{K}$ is always real-valued as well as closed. Closedness follows from the relevant Poincaré lemmas $\partial\, \partial = 0$ and $\bar{\partial}\, \bar{\partial} = 0$: we have $\mathrm{d}\, \partial\, \bar{\partial} \mathcal{K} = (\partial + \bar{\partial})\, \partial\, \bar{\partial} \mathcal{K} = 0$. The foregoing needs no further qualification when, as here, the associated 'complex structure', i with $i^2 = -1$, is just the ordinary imaginary unit.

The simplest standard case (MS98 p. 130) is that in which $\mathcal{K}(X,Y,\bar{X},\bar{Y}) = X\bar{X} + Y\bar{Y}$, producing another 2-form $\frac{1}{2}i(\mathrm{d}X \wedge \mathrm{d}\bar{X} + \mathrm{d}\bar{Y} \wedge \mathrm{d}Y)$, in our non-standard notation. Our 2-form $\omega_K = \mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y)$ is obtained by choosing instead

$$\mathcal{K}(X,Y,\bar{X},\bar{Y}) \;=\; -(X-\bar{X})(Y-\bar{Y})\;, \tag{13.13}$$

or, equally well,

$$\mathcal{K}(X,Y,\bar{X},\bar{Y}) \;=\; -XY \;-\; \bar{X}\bar{Y}\;, \tag{13.14}$$

since the extra terms in (13.13) are 'harmless' insofar as they contribute nothing to (13.12). Substituting either of these into (13.12) and multiplying by $\frac{1}{2}i$, we get cross-terms only, hence

$$\omega_K \;=\; \mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y) \tag{13.15}$$

as anticipated. (The normalizations used here for ω_K and for $\mathcal{K}$ itself follow MS98, and differ by constant factors from those used in RR01 and in F99.)

Other cases with less simple $\mathcal{K}$ have Kähler 2-forms with variable coefficients describing a 'curved' Kähler structure, nontrivially different from, and more restrictive than, ordinary real symplectic structure, especially in its global aspects (MS98). This may yet prove significant for later stages on our journey. Here one may speak of 'curvature' in more than one sense, in the first place because the Kähler 2-form and associated symplectic structure may admit a 'symplectic connection' (MS98) defining parallel transport, e.g. of geometric structures like the tubes in Figure 1. Nonzero curvature means that parallel transport produces different results over different paths. In the second place, the Kähler 2-form together with the complex structure induces a symmetric bilinear form and hence a Riemannian or Minkowskian inner product and metric, which may have curvature. More precisely, when the Kähler 2-form is contracted with the pair of vectors $(\delta X, \delta Y, \delta\bar{X}, \delta\bar{Y})$ and $(i\,\delta X', -i\,\delta Y', -i\,\delta\bar{X}', i\,\delta\bar{Y}')$ (the sign pattern reflecting our non-standard notation), then the factors i convert skew-symmetry into symmetry, producing a real-valued symmetric bilinear form and metric, Riemannian or sign-definite in our case (13.15) and in the standard case mentioned above (13.13). Minkowskian or sign-indefinite cases plainly exist, as exemplified by linear combinations of the Kähler potentials and therefore 2-forms already encountered, and also by the case

$$\mathcal{K}(X,Y,\bar{X},\bar{Y}) \;=\; X\bar{X} \;-\; Y\bar{Y}\;. \tag{13.16}$$

This last is the same as the abovementioned standard case apart from the change of sign between terms, and is therefore Minkowskian since the standard case is Riemannian. Minkowskian cases are sometimes called 'pseudo-Kähler'.

In the third place, the complex structure, when generalized to mean any algebraic object whose square is minus the identity, may itself be a spatially variable field. This raises questions of 'compatibility' of the complex structure with the 2-form and metric under symplectic parallel transport (e.g. MS98, F99). Such questions do not arise, however, in the simple 'flat' case considered here and in RR01. Indeed this case is the simplest possible example of what is called 'special Kähler' or 'rigid special Kähler' structure (F99), in which, more generally, the symplectic structure may be perfectly flat (with tubes like those in Figure 1 straight, when viewed in a suitable coordinate system, and with parallel transport globally unique) even when the metric and complex structures are not flat.

There now arises a conundrum, almost a mathematical pun. F99 shows that every special Kähler structure, flat or curved, possesses not only its Kähler potential $\mathcal{K}(X, Y, \bar{X}, \bar{Y})$ but also what is called a 'holomorphic prepotential', in terms of which $\mathcal{K}$ can always be defined. In our normalization (and non-standard notation) this prepotential is a holomorphic function $\Psi(X, \bar{Y})$ such that

$$\mathcal{K}(X, Y, \bar{X}, \bar{Y}) \;=\; \frac{1}{2i}\left(\frac{\partial\Psi}{\partial X}\,\bar{X} \;-\; \frac{\partial\bar{\Psi}}{\partial Y}\,\bar{Y} \;-\; \frac{\partial\bar{\Psi}}{\partial\bar{X}}\,X \;+\; \frac{\partial\Psi}{\partial\bar{Y}}\,Y\right)\,. \qquad (13.17)$$

Inspection shows at once that this formula produces the $\mathcal{K}$ defined in (13.14) if we set $\Psi(X, \bar{Y}) = -iX\bar{Y}$. What is peculiar is that, as shown in RR01, instead of $-iX\bar{Y}$ we may take $\Psi(X, \bar{Y}) \;=\; 2c\,\hat{\Phi}(X, \bar{Y})$, where $\hat{\Phi}$ is the complex potential defined in (12.2), and substitute this into the *same formula* (13.17), to produce the $\mathcal{K}$ of (13.13) — which is equivalent to that of (13.14) as far as (13.12) is concerned, and therefore as far as the Kähler structure itself is concerned. This last result holds only, however, after restriction to the Lagrangian submanifold represented by the graph of ϕ. Away from the graph, $\hat{\Phi}$ as defined in (12.2) ceases to be holomorphic.

Still more peculiar is the very fact that one obtains such a result at all, relating $\hat{\Phi}(X, \bar{Y})$ to $\mathcal{K}(X, Y, \bar{X}, \bar{Y})$ alone, even on the graph of ϕ. This is because $\hat{\Phi}(X, \bar{Y})$ is a description of the mass configuration and therefore changes into different functions of X and $\bar{Y}$ as the dynamical system evolves. Recall that we have suppressed explicit reference to time t and therefore to the fact that $\hat{\Phi}$ is really a function $\hat{\Phi}(X, \bar{Y}, t)$. So, unlike $\mathcal{K}$ and its prepotential Ψ, the complex potential $\hat{\Phi}$ is not a static property of the space $\{X, Y, \bar{X}, \bar{Y}\}$ and its underlying structure. It is a function of the dynamics as well. Since $\hat{\Phi}$ is a holomorphic function of $(X, \bar{Y})$ on the graph of ϕ, one can imagine using analytic continuation to extend it some distance off the graph. But apart from the practical certainty that singularities would be encountered not too far away, one would still have a structure, including the locations of the singularities, that changed from moment to moment as the mass configuration changed.

Another peculiar fact is that there is actually no holomorphic prepotential at all for the $\mathcal{K}$ of (13.13). Inspection of (13.17) shows that if there were such a potential then it would have to conform to $\partial\Psi/\partial X = i(Y - \bar{Y})$ and $\partial\Psi/\partial\bar{Y} = -i(X - \bar{X})$; but the presence of the other two variables, $\bar{X}$ and Y, immediately stops us from constructing any such *holomorphic* function throughout $\{X, Y, \bar{X}, \bar{Y}\}$.‡ We are now close to the frontiers of the territory explored so far; the significance of the facts just described is far from clear.

Consider now the second of the two symplectic structures recalled at the start of this section. The Ω^{CL} of Sections 6–8 inherited from the parent phase space manifests itself, as previously noted, in the *real* part of the 2-form $\mathrm{d}X \wedge \mathrm{d}Y$. Because we are now viewing everything in the finite-dimensional space $\{X, Y, \bar{X}, \bar{Y}\}$, we again see only a 2-dimensional projection, or reflection, of the full infinite-dimensional Ω^{CL} (Bridges *et al.* 2001). The real part of $\mathrm{d}X \wedge \mathrm{d}Y$ is also a Kähler 2-form. For if we now take

$$\mathcal{K}(X, Y, \bar{X}, \bar{Y}) \;=\; -\,i\,(X - \bar{X})(Y + \bar{Y})\;, \tag{13.18}$$

or, equally well,

$$\mathcal{K}(X, Y, \bar{X}, \bar{Y}) \;=\; -\,i\,XY \;+\; i\,\bar{X}\bar{Y} \tag{13.19}$$

(again real-valued), then (13.12) produces

$$\tfrac{1}{2}i\,\partial\,\bar{\partial}\mathcal{K} \;=\; \mathrm{Re}\,(\mathrm{d}X \wedge \mathrm{d}Y)\;. \tag{13.20}$$

By taking the sum or difference of the Kähler potentials (13.14) and (13.19) we can now see, moreover, that $\mathrm{d}X \wedge \mathrm{d}Y$ and its complex conjugate $\mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y}$ are themselves Kähler 2-forms.

Transforming $\mathrm{Re}\,(\mathrm{d}X \wedge \mathrm{d}Y)$ back into (x, y, p, q) space, i.e. substituting $X = x + p + icq$ and $Y = y + q - icp$ into $\mathrm{Re}\,(\mathrm{d}X \wedge \mathrm{d}Y)$, we obtain a new 2-form

$$\begin{aligned}\omega \;&=\; \mathrm{Re}\,(\mathrm{d}X \wedge \mathrm{d}Y)\\ &=\; \mathrm{d}x \wedge \mathrm{d}y + \mathrm{d}p \wedge \mathrm{d}y + \mathrm{d}x \wedge \mathrm{d}q + (1 - c^2)\,\mathrm{d}p \wedge \mathrm{d}q\;,\end{aligned} \tag{13.21}$$

giving a noncanonical representation of the reflection in $\{X, Y, \bar{X}, \bar{Y}\}$ of the second symplectic structure. It is noncanonical, when viewed in (x, y, p, q) space, because Ω^{CL} is noncanonical. This in turn is a consequence, it will be recalled, of restricting the parent 2-form to $\mathcal{M}^{\mathrm{C}}$ by applying the constraint (9.6).

To summarize so far, the 2-form $\mathrm{d}X \wedge \mathrm{d}Y$ and its complex conjugate $\mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y}$ reflect both the symplectic structures of interest. All four 2-forms — that is, $\mathrm{d}X \wedge \mathrm{d}Y$, $\mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y}$, and their real and imaginary parts — are Kähler

‡Specifically, if $\partial\Psi/\partial X = i(Y - \bar{Y})$, then $\Psi = iX(Y - \bar{Y}) + \mathrm{func}(\bar{X}, Y, \bar{Y})$. If also $\partial\Psi/\partial\bar{Y} = -i(X - \bar{X})$, then then $\Psi = -i\bar{Y}(X - \bar{X}) + \mathrm{func}(X, \bar{X}, Y)$. Therefore $\Psi = i(XY + \bar{X}\bar{Y} - X\bar{Y}) + \hat{g}(\bar{X}, Y)$, say. But it is impossible to choose the function $\hat{g}$ so as to make $\partial\Psi/\partial\bar{X}$ and $\partial\Psi/\partial Y$ vanish as required by the Cauchy–Riemann relations for holomorphic $\Psi(X, \bar{Y})$; cf. (13.9).

2-forms on the space $\{X, Y, \bar{X}, \bar{Y}\}$. None of this is surprising, because the relation between Kähler geometry and symplectic geometry is well known and entirely straightforward in flat spaces like those dealt with here and in RR01. The transformations of the various 2-forms between the original space $\{x, y, p, q\}$ and its complex image $\{X, Y, \bar{X}, \bar{Y}\}$ under (13.1)–(13.4) can now be summarized thus:

$$\mathrm{d}X \wedge \mathrm{d}Y \;=\; \omega \,-\, ic\,\Omega \;, \qquad \mathrm{d}\bar{X} \wedge \mathrm{d}\bar{Y} \;=\; \omega \,+\, ic\,\Omega \;. \tag{13.22}$$

Again recall that this is consistent with (9.3) because of the vanishing of Ω on the graph of ϕ, making the contribution $ic\,\Omega$ invisible on the graph.

What then of hyper-Kähler structure or geometry? Here there is a hint of something entirely new. The pair of Kähler 2-forms $\mathrm{Re}\,(\mathrm{d}X \wedge \mathrm{d}Y)$ and $\mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y)$ together with the Kähler 2-form $\frac{1}{2}i\,(\mathrm{d}X \wedge \mathrm{d}\bar{X} - \mathrm{d}\bar{Y} \wedge \mathrm{d}Y)$ generated by (13.16) together make up what is called a *hyper-Kähler triplet* of closed 2-forms, as RR01 point out. These define a *hyper-Kähler geometry*[§] in the space $\{X, Y, \bar{X}, \bar{Y}\}$ and are conventionally normalized and notated as follows

$$\begin{aligned}
\omega_I &= \mathrm{Re}\,(\mathrm{d}X \wedge \mathrm{d}Y)\;, && (13.23)\\
\omega_J &= \tfrac{1}{2}i\,(\mathrm{d}X \wedge \mathrm{d}\bar{X} + \mathrm{d}Y \wedge \mathrm{d}\bar{Y})\;, && (13.24)\\
\omega_K &= \mathrm{Im}(\mathrm{d}X \wedge \mathrm{d}Y)\;. && (13.25)
\end{aligned}$$

(Here, in fact, Y and $\bar{Y}$ now appear in the standard way.) Just as Kähler geometry is underpinned by complex structure, hyper-Kähler geometry can be shown to be underpinned by quaternion or spin-matrix structure. More precisely, a hyper-Kähler triplet of 2-forms is related to a single metric through a triplet (I, J, K) of complex structures satisfying the rules of quaternion algebra. Those rules are $I^2 = J^2 = K^2 = -1$ (stating that I, J, K are indeed complex structures) together with $I = JK = -KJ$, $J = KI = -IK$, and $K = IJ = -JI$. In the present case (I, J, K) can be taken to be the triplet of 4×4 diagonal or antidiagonal matrices $I = \mathrm{antidiag}(-1,\,1,-1,\,1)$, $J = \mathrm{diag}(-i,-i,\,i,\,i)$, $K = \mathrm{antidiag}(i,-i,-i,\,i)$, where 'antidiag' is to be read from bottom left to top right. It is easy to check that the quaternion rules are satisfied. Using the shorthand $\omega_I(\delta X,\,\delta X')$ to denote the contraction of ω_I with a pair of vectors $(\delta X, \delta Y, \delta\bar{X}, \delta\bar{Y})$ and $(\delta X', \delta Y', \delta\bar{X}', \delta\bar{Y}')$, and defining also

$$\tilde{g}(\delta X,\,\delta X') \;=\; \tfrac{1}{2}\left(\delta X\delta\bar{X}' + \delta X'\delta\bar{X} + \delta Y\delta\bar{Y}' + \delta Y'\delta\bar{Y}\right)\;, \tag{13.26}$$

we can easily verify that $\omega_I(\delta X,\,\delta X') = \tilde{g}(\delta X,\;I\,\delta X')$, $\omega_J(\delta X,\,\delta X') = \tilde{g}(\delta X,\;J\,\delta X')$ and that $\omega_K(\delta X,\,\delta X') = \tilde{g}(\delta X,\;K\,\delta X')$, where $I\,\delta X'$,

[§]Hyper-Kähler geometry arises naturally in a given space or manifold whenever, for instance, the space has a Lagrangian submanifold (half the dimensions) whose cotangent bundle (e.g. Schutz 1980) can be identified with the original space (twice half the dimensions). F99 gives a clear example.

$J\,\delta X'$ and $K\,\delta X'$ are to be read as denoting 4-dimensional matrix multiplication. The symmetric, real-valued bilinear form $\tilde{g}$ defines a inner product that corresponds to a flat Riemannian metric, the single metric associated with our hyper-Kähler triplet. It follows that, as can also be directly verified, the triplet has the cyclic property $\omega_I(\delta X,\ J\,\delta X') = \omega_K(\delta X,\ \delta X')$, $\omega_J(\delta X,\ K\,\delta X') = \omega_I(\delta X,\ \delta X')$, $\omega_K(\delta X,\ I\,\delta X') = \omega_J(\delta X,\ \delta X')$.

Now we have seen from Section 5 onwards that ω_I and ω_K are mathematical objects having clear counterparts in semigeostrophic theory. As the two relations in (13.22) remind us, ω_I and ω_K transform back to ω and $-c\,\Omega$ in (x, y, p, q) space. They describe the two symplectic structures shared by all Hamiltonian balanced models arising from (9.6) through Salmon's method, including semigeostrophic theory. But the remaining member of the triplet, ω_J, is a new mathematical object altogether. It was not merely invisible, but actually nonexistent, before c became nonzero. Substituting (13.1) into (13.24), we see that ω_J transforms back to

$$\omega_J \;=\; c\,(\mathrm{d}p\wedge\mathrm{d}y + \mathrm{d}x\wedge\mathrm{d}q) \;+\; 2c\,\mathrm{d}p\wedge\mathrm{d}q\;, \tag{13.27}$$

which vanishes identically if $c = 0$. Again frustratingly, its significance for present purposes has yet to be understood; but we need to consider the possibility that its emergence is not accidental. Furthermore, this needs to be investigated alongside the connection to twistor space already mentioned. Efforts to follow these clues are underway; the reader is referred to RR01 for some further discussion.

14 Connections with Monge–Ampère operators

The parallel derivations of (9.9) and (13.21) — respectively by substitution of (9.7) into the Jacobian $\partial(X,Y)/\partial(x,y)$ and by substitution of (13.1) into the 2-form $\mathrm{d}X\wedge\mathrm{d}Y$ — remind us of the well-known fact that 2-forms like ω_I and ω_J are simply Monge–Ampère operators in disguise (e.g. Lychagin *et al.* 1993). In 2 dimensions the general case is

$$\begin{aligned}\omega_{\rm gen} \;=\; & A\,\mathrm{d}x\wedge\mathrm{d}y \;+\; B\,\mathrm{d}p\wedge\mathrm{d}y \;+\; C(\mathrm{d}x\wedge\mathrm{d}p - \mathrm{d}y\wedge\mathrm{d}q)\\ & \;+\; D\,\mathrm{d}x\wedge\mathrm{d}q \;+\; E\,\mathrm{d}p\wedge\mathrm{d}q\;,\end{aligned} \tag{14.1}$$

with ellipticity criterion

$$\frac{\omega_{\rm gen}\wedge\omega_{\rm gen}}{\Omega\wedge\Omega} \;=\; BD \;-\; C^2 \;-\; AE \;>\; 0\;. \tag{14.2}$$

In the recent literature the expression $(\omega_{\rm gen}\wedge\omega_{\rm gen})/(\Omega\wedge\Omega)$ is often called the Pfaffian of $\omega_{\rm gen}$ (Lychagin *et al.*, *op. cit.*). Restricting the 2-form $\omega_{\rm gen}$ to the graph of ϕ produces $\omega_{\rm gen} = [A + B\rho + 2C\tau + D\varsigma + E(\rho\varsigma - \tau^2)]\,\mathrm{d}x\wedge\mathrm{d}y$, as is easy to verify. This expression is just $\mathrm{d}x\wedge\mathrm{d}y$ times the general Monge–Ampère operator in (4.22). As before, ρ, ς, τ are the second partial derivatives of ϕ defined by (4.23). Notice the minus sign in the C term of (14.1), as distinct from the plus in the definition of $\Omega = \mathrm{d}x \wedge \mathrm{d}p + \mathrm{d}y \wedge \mathrm{d}q$. Because Ω vanishes on the graph of ϕ, there is some freedom in the arrangement of the C term. (Ω vanishes on any graph, i.e. vanishes for any differentiable function $\phi(x, y)$, whether or not it satisfies the Monge–Ampère equation.) Thus for instance $\mathrm{d}x \wedge \mathrm{d}p$ in the C term can be replaced by $-\mathrm{d}y \wedge \mathrm{d}q$ without changing the term $2C\tau$ in the Monge–Ampère operator itself.

The 2-form ω defined in (13.21), corresponding to ω_I in the hyper-Kähler triplet, can now be seen to correspond to the Monge–Ampère operator on the *left* of (9.9), with $A = 1$, and with the remaining parameters $B = D = 1$, $C = 0$, and $E = 1 - c^2$, as noted earlier. The new 2-form ω_J in the hyper-Kähler triplet corresponds to another elliptic Monge–Ampère operator, but not one that has come to light in any other connection.

It is not yet clear whether, or how, the foregoing will carry over to cases in which we replace $A = 1$ by the variable coefficient $A = 1 - f^{-1}\zeta^{\mathrm{C}}$, as suggested by the dynamical problem with general PV and vorticity fields $Q^{\mathrm{C}}(x, y)$, $\zeta^{\mathrm{C}}(x, y)$, involving the full Monge–Ampère operator of (9.9). The first term $\mathrm{d}x \wedge \mathrm{d}y$ on the right of (13.21) is then replaced by $(1 - f^{-1}\zeta^{\mathrm{C}})\mathrm{d}x \wedge \mathrm{d}y$. This 2-form is still closed, despite the variable coefficient, essentially because the coefficient is a function of (x, y) alone and because $\mathrm{d}x \wedge \mathrm{d}x = \mathrm{d}y \wedge \mathrm{d}y = 0$.¶ However, the new coefficient $A = 1 - f^{-1}\zeta^{\mathrm{C}}$ is not sign-definite; it is a dimensionless measure of minus the relative vorticity, and our Hamiltonian balanced models, like the real world, can have anticyclones as well as cyclones. In any case there appears to be no corresponding hyper-Kähler structure, because it is well known (Atiyah & Hitchin, 1988; G.W. Gibbons, personal communication) that such structures can be curved only in a very special way, such that the Ricci curvature tensor vanishes even if the Riemann curvature tensor does not.

15 Postlude

This article has, we hope, gone some way toward answering the questions posed in the title and in Section 1, even though not as far as we had originally hoped. We feel that it has clarified, or at least exposed, some of the issues involved — some of the main landmarks in the territory surveyed. We have been mainly concerned with issues of formalism rather than with genuinely mathematical questions, though a few of the latter have been raised as well, such as the invertibility or otherwise of the splitting equation (8.12) (Wunderer 2001; see also Theiss 1999). This last issue of invertibility has in turn prompted a new search for regularized, therefore well-behaved, members of the $\sqrt{3}$ family of Hamiltonian balanced models. We already know that such well-behaved models will, like semigeostrophic theory, exhibit double splitting.

¶For any function $k(x, y)$ we have $\mathrm{d}k(x, y)\mathrm{d}x \wedge \mathrm{d}y = \mathrm{d}k \wedge \mathrm{d}x \wedge \mathrm{d}y = (k_x \mathrm{d}x + k_y \mathrm{d}y) \wedge \mathrm{d}x \wedge \mathrm{d}y = 0$, where $k_x = \partial k/\partial x$, $k_y = \partial k/\partial y$.

Throughout this article, we have confined attention to the shallow-water equations, representing the simplest fluid-dynamical system in which all the foregoing issues are nontrivial. The extension to 3-dimensional, fully stratified flow systems is, however, straightforward if we regard such systems as layerwise-2-dimensional. More precisely, the extension is straightforward if we begin with a hydrostatic, or so-called primitive-equation parent system, consisting of a stack of shallow-water layers or the continuum limit thereof, and then apply a momentum–configuration constraint $\mathbf{u} = \mathbf{u}^{\mathrm{C}}$ to each layer. The layers are coupled together via the potential energy V, appropriately generalized. Then the parent phase space is still an infinite Cartesian product $\mathbb{R}^4 \otimes \mathbb{R}^4 \otimes \cdots$, particle by particle. The application of Salmon's method is straightforwardly as described in Sections 7–8, producing a quasi-even-dimensional phase space $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \cdots$, and there is no need to bring in, for instance, the Dirac theory of constraints. Notice that, in the analogy with the toy problem of Section 7, this corresponds to taking many particles interacting through some potential V and each moving in *two* space dimensions. It is quite unlike the generalization of the toy problem to three space dimensions, which is degenerate because Ω^{C}_{ij} is then 3×3 and antisymmetric, and therefore singular.

Specifically, to make the extension to the 3-dimensional, fully stratified fluid case, we may redefine the notation $h(\mathbf{x})$ in the following way. The stratified system is now regarded as a set of many interacting 2-dimensional layers in the physical domain $\mathcal{D}$. In the continuum limit of the layered system, the specific entropy s, or some function of it such as potential temperature, keeps track of vertical position in $\mathcal{D}$, taking advantage of the stable stratification. Thus s now has the role of a Lagrangian label that is also a physical, measurable quantity and is not, of course, subject to the particle-relabelling symmetry; and $\mathbf{x}$ is the horizontal projection of position on a given stratification surface $s =$ constant. The mass element becomes

$$\mathrm{d}m = h_0 \,\mathrm{d}\mathbf{a}\,\mathrm{d}s \;=\; h(\mathbf{x}, s)\,\mathrm{d}\mathbf{x}\,\mathrm{d}s\ , \tag{15.1}$$

where $\mathbf{a} = (a, b)$, $\mathbf{x} = (x, y)$, and where h is now equated to $(\partial z/\partial s)_{x,y}$ with z denoting geometrical altitude. Note incidentally that h could be zero at a given horizontal position $\mathbf{x}$ and a given value of s, as can happen for instance when isentropes intersect a horizontal boundary. The use of s to help identify particles within the physical domain $\mathcal{D}$ via $(\mathbf{a}, s)$ or $(\mathbf{x}, s)$ amounts to using the well known isentropic or isopycnic vertical 'coordinate' (as it is called in the meteorological literature), or more aptly 'label'.

Theiss (1999) has taken an alternative approach to the 3-dimensional stratified problem, in which Salmon's method is applied directly to the full 3-dimensional nonhydrostatic Euler equations. The constraint functional must then impose vertical, quasi-hydrostatic balance, as well as horizontal balance through some horizontally-oriented $\mathbf{u} = \mathbf{u}^{\mathrm{C}}$ as above. Because of the odd number of physical space dimensions, the configuration space of the constrained problem now has the form $\mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \cdots$, particle by particle, rather than $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \cdots$; and, as Theiss points out, naive application of a 3-dimensional constraint of the form $\mathbf{u} = \mathbf{u}^{\mathrm{C}}$ (rashly attempted in MR96) produces a splitting kernel $\Omega_{ij}^{\mathrm{CL}}(\mathbf{a}, \mathbf{a}')$ that is degenerate, with a nontrivial null space. This is analogous to the degeneracy of the 3×3 matrix Ω_{ij}^{C} in the 3-dimensional version of the toy problem of Section 7. The associated Lagrangian functional is singular, suggesting the use of the Dirac theory of constraints, a systematic method for dealing with singular Lagrangians. Theiss shows in detail how the Dirac theory can be used to overcome the degeneracy problem and to achieve a fully 3-dimensional yet nondegenerate reformulation of the splitting equation, for the singly-split models produced by a general class of constraint functionals.

So where does all this leave us? Both in the shallow-water and, by implication, in the fully stratified case, a new understanding of the complex-valued canonical coordinates (9.7) is beginning to emerge through recognition of the associated abstract mathematical and geometrical structures. When we go from semigeostrophic theory to the more accurate models considered here, the transformed streamfunction $\Phi(X, Y)$ and the transformed complex potential $\hat{\Phi}(X, \bar{Y})$ become two separate entities. In semigeostrophic theory the two are indistinguishable: we may characterize semigeostrophic theory as the unique case in which $\Phi(X, Y)$ and $\hat{\Phi}(X, \bar{Y})$ merge into a single function, and the Jacobian (13.2) collapses to zero. Conversely, the functions Φ and $\hat{\Phi}$ become distinct as soon as the parameter $c > 0$. And making $c > 0$ is inescapable, with $c = \sqrt{3}$ as the optimum, if accuracy is to be improved, within the category of near-local balanced models defined by (9.6). These are the next steps beyond semigeostrophic theory. In dealing with the space $\{X, Y, \bar{X}, \bar{Y}\}$ we encounter an overarching hyper-Kähler geometry. This is itself part of what is called twistor geometry, as studied in a well-known research programme initiated by Penrose (e.g. Atiyah & Hitchin 1988 & refs.). Efforts to exploit these facts are underway.

The importance of understanding the associated abstract structures as they apply in the cases of interest here will, in the end, turn on whether the search for regularized, doubly-split $\sqrt{3}$ models is successful and on how accurate such models prove to be, and on whether they possess tractable streamfunctions $\Phi(X, Y)$ defined in a sufficiently simple way. One obstacle to progress is the fact that although, in general, $\hat{\Phi}(X, \bar{Y})$ is simple to define explicitly, $\Phi(X, Y)$ is not.

There are, of course, many steps still further beyond semigeostrophic theory, into the wider territory of fully nonlocal balance conditions defined by fully nonlocal functionals $\mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$, where $h(\cdot)$ symbolically represents the mass configuration in a shallow-water or in a fully-stratified model, in the manner sketched above. Some incursions into that territory have already been made by Allen & Holm (1996), to two orders in ϵ, which implies nonlocality in the irrotational part of the $\mathbf{u}^{\mathrm{C}}$ field even though not in ζ^{C} and the rotational part. Further on still, hierarchies of $\mathbf{u}^{\mathrm{C}}$ functionals are known that respect more and more accurately the delicacy and subtlety of real vortical motion. They are defined by rather complicated sets of equations (e.g. MN00, Mohebalhojeh & Dritschel 2001, McIntyre 2001). The considerations of Sections 6ff. show that Salmon's method can, in principle, be used to convert any of these extremely accurate, nonlocal $\mathbf{u}^{\mathrm{C}}$ functionals into Hamiltonian balanced models, though the purely formal obstacles — even with computer-aided symbolic manipulation — are extremely daunting and to our knowledge have never been tackled.

Still less is anything known about the *mathematical* properties of these accurate $\mathbf{u}^{\mathrm{C}}$ functionals, though numerical experimentation has clearly shown that some of them are well behaved, and astonishingly accurate, over a vast parameter range including unbounded ranges of ϵ values (for instance at the equator of a hemispherical model, where $\epsilon = \infty$). It seems likely that nonlocality in $\mathbf{u}^{\mathrm{C}}$ may be another way of preventing the kind of pathology encountered by Wunderer (*op. cit.*). Accurate $\mathbf{u}^{\mathrm{C}}$ functionals can of course be used to construct non-Hamiltonian as well as Hamiltonian balanced models, as in fact was done in the work of MN00 and Mohebalhojeh & Dritschel just cited. Again, little is known about the mathematical structure of such non-Hamiltonian balanced models, which, for reasons connected with Lighthill radiation, may ultimately turn out to be the most accurate possible balanced models. This remains very much a question for the future. One recent surprise, stimulated by work begun at the Newton Institute Programme, has been the recognition that, beyond a certain accuracy (two orders in ϵ), non-Hamiltonian balanced models share with Hamiltonian balanced models the phenomenon of velocity splitting (Mohebalhojeh and McIntyre 2001). It is possible that the most powerful approach, in the end, will involve not only considerations of Hamiltonian and associated geometric structures, but also considerations of exceedingly small departures therefrom.

Acknowledgements

We thank T. Bridges, R. Ford, G.W. Gibbons, M.B. Green, V.V. Lychagin, A.R. Mohebalhojeh, P.J. Morrison, E. Neven, N. Nichols, S. Reich, S. Ren, V.N. Roubtsov, R. Salmon, N.I. Shepherd-Barron, T.G. Shepherd, J. Theiss, T. Warn, and C.F. Wunderer for helpful comments or correspondence. MEM wishes to acknowledge generous support from the UK Natural Environment Research Council through the UK Universities' Global Atmospheric Modelling Programme, from the Isaac Newton Institute for Mathematical Sciences, and from an SERC/EPSRC Senior Research Fellowship.

Appendix A: Derivation of equation (6.5)

In this Appendix we again use the inertial reference frame, except that the symbol $\mathbf{u}^{\mathrm{G}}$ still denotes the geostrophic velocity relative to a frame rotating with angular velocity $(0, 0, \frac{1}{2}f)$, as it did in (3.2) and (9.6).

Before deriving (6.5), we consider first what the splitting equation (8.12) tells us when $\mathbf{u}^{\mathrm{C}}$ is taken as solid rotation, the case known as Salmon's L_0 dynamics. Although L_0 dynamics does not yet give a balanced model capable of dynamical evolution, it does illustrate the tendency of $\mathbf{u}^{\mathrm{P}} = \mathbf{u}^{\mathrm{C}} + \mathbf{u}^{\mathrm{S}}$ to be an improved approximation to the balanced motion in comparison with $\mathbf{u}^{\mathrm{C}}$.

For L_0 dynamics we simply take

$$\mathbf{u}^{\mathrm{C}} = \tfrac{1}{2} f \hat{\mathbf{z}} \times \mathbf{x} \ , \tag{A.1}$$

describing solid rotation. Then the constraint vorticity becomes simply the Coriolis parameter of that solid rotation, $\zeta^{\mathrm{C}} = f =$ constant. The right-hand side of (8.13) vanishes, because the $\mathbf{u}^{\mathrm{C}}$ of (A.1) does not depend on the mass configuration, so that the ω^{C}_{ij} terms in the splitting equation (8.12) vanish. The residual unbalanced force $\mathbf{R}^{\mathrm{C}}$ per unit mass becomes simply the contribution to the horizontal pressure-gradient force per unit mass that is not balanced by the centrifugal acceleration of the solid rotation. Thus (8.12) becomes simply the geostrophic relation, $\mathbf{u}^{\mathrm{S}} = \mathbf{u}^{\mathrm{G}}$, with $\mathbf{u}^{\mathrm{G}}$ the relative geostrophic velocity given in terms of the mass configuration by (3.2), wherein $h = h_0 + h^{\mathrm{abs}} - h^{\mathrm{centrif}}$ with h^{abs} defined as the actual surface elevation, and h^{centrif} as the paraboloidal surface elevation for the solid rotation (A.1). The evolution described by (6.4) or (8.12) is now trivial in the sense that particles follow geostrophic streamlines, within an Eulerian mass configuration $h(\mathbf{x})$ that does not change with time when viewed in the rotating frame. This is because, as is easily checked from (3.2), $\nabla \cdot (h\mathbf{u}^{\mathrm{G}}) = 0$ so that the Eulerian mass-conservation equation (2.7) implies $\partial h/\partial t = 0$.

For L_1 dynamics we take, iteratively,

$$\mathbf{u}^{\mathrm{C}} \ = \ \tfrac{1}{2} f \hat{\mathbf{z}} \times \mathbf{x} \ + \ \mathbf{u}^{\mathrm{G}} \ . \tag{A.2}$$

The constraint vorticity ζ^{C} in (8.12) now becomes the geostrophic absolute vorticity $f + \hat{\mathbf{z}} \cdot \nabla \times \mathbf{u}^{\mathrm{G}}$. The corresponding materially conserved PV,

$$Q^{\mathrm{C}} = h^{-1} \zeta^{\mathrm{C}} \ , \tag{A.3}$$

is assumed to be positive everywhere. Also (for arbitrary bottom topography)

$$\mathsf{V} = \int_{\mathcal{D}} \left(g h^{\mathrm{abs}} - \tfrac{1}{2} g h \right) \, \mathrm{d}m = \int_{\mathcal{D}} \left(g h^{\mathrm{abs}} h - \tfrac{1}{2} g h^2 \right) \, \mathrm{d}\mathbf{x} \ , \tag{A.4}$$

implying that

$$\delta \mathsf{V} = \int_{\mathcal{D}} \left(gh \, \delta^{\mathrm{E}} h^{\mathrm{abs}} + g h^{\mathrm{abs}} \delta^{\mathrm{E}} h - gh \, \delta^{\mathrm{E}} h \right) \mathrm{d}\mathbf{x} = \int_{\mathcal{D}} g h^{\mathrm{abs}} \delta^{\mathrm{E}} h \, \mathrm{d}\mathbf{x} \tag{A.5}$$

(because with any fixed topography Eulerian variation is simplest, $\delta^{\mathrm{E}} h^{\mathrm{abs}} = \delta^{\mathrm{E}} h$), whence, by (8.4) and (8.10),

$$\delta \mathsf{V} = g \int_{\mathcal{D}} \delta^{\mathrm{L}} \mathbf{x}(\mathbf{x}) \cdot \nabla h^{\mathrm{abs}} \, \mathrm{d}m \; ; \; \Rightarrow \; \mathbf{F} = -g \nabla h^{\mathrm{abs}} \; . \tag{A.6}$$

Note incidentally — with an eye to Remarks 4 and 5 of Section 8 — how the inertial-frame description is related to the standard description relative to a rotating frame of reference. There, the Hamiltonian functional is usually defined to contain only the relative kinetic energy and to contain only the part of the potential energy associated with gh, not gh^{abs} (e.g. Salmon 1983). Here we have instead the absolute kinetic energy and the actual gravitational potential. It is straightforward to check that the two are equivalent, provided that one remembers that as well as using absolute (inertial-frame) $\mathbf{u}^{\mathrm{C}}$ values in H, one must also use absolute $\mathbf{u}^{\mathrm{P}}$ values in place of $\dot{\mathbf{x}}$ in the fluid counterpart of $-\dot{\mathbf{x}} \rfloor \Omega^{\mathrm{C}} + dH^{\mathrm{C}} = 0$, and allow for any work done by moving boundaries. Cancellations then lead to the standard rotating-frame description.

We now use the restriction to constant g and f and to zero topography (bottom boundary parallel to h^{centrif}). A short calculation from (8.11) and (A.6) gives

$$\mathbf{R}^{\mathrm{C}} = \mathbf{F} - \mathbf{A}^{\mathrm{C}} = -\mathbf{u}^{\mathrm{G}} \cdot \nabla \mathbf{u}^{\mathrm{G}} \; . \tag{A.7}$$

There is no mass-rearrangement term in this case, because now $h \nabla h = \frac{1}{2} \nabla (h^2)$, implying

$$\nabla \cdot \mathbf{u}^{\mathrm{G}} \;=\; 0 \;=\; \nabla \cdot (h \mathbf{u}^{\mathrm{G}}) \; , \tag{A.8}$$

annihilating the relative part of the last term in (8.11). If particles were to move with absolute velocity $\mathbf{u}^{\mathrm{C}}$, i.e. with relative velocity $\mathbf{u}^{\mathrm{G}}$, then $\partial h / \partial t$ evaluated in the rotating frame would be exactly zero. So in this example there is no contribution to $\mathbf{R}^{\mathrm{C}}$ from the functional dependence of $\mathbf{u}^{\mathrm{G}}$ and hence $\mathbf{u}^{\mathrm{C}}$ on the mass configuration, $h(\mathbf{x})$. The reference to rotating frame is only for computational convenience and can now be dispensed with.

It remains to evaluate ω^{C}_{ij}. The Eulerian functional derivatives of $\mathbf{u}^{\mathrm{C}}$ and $\mathbf{u}^{\mathrm{G}}$ are equal because of the reference-frame indifference of δ^{E}. In the present case we therefore have

$$\frac{\delta^{\mathrm{E}} u_i^{\mathrm{C}}(\mathbf{x}; h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')} = \frac{\delta^{\mathrm{E}} u_i^{\mathrm{G}}(\mathbf{x}; h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')} = \frac{g}{f} \varepsilon_{ik} \frac{\partial}{\partial x'_k} \delta(\mathbf{x} - \mathbf{x}') \; , \tag{A.9}$$

where ε_{ik} is defined as before by $\varepsilon_{12} = -\varepsilon_{21} = 1$, $\varepsilon_{11} = \varepsilon_{22} = 0$. Thus, from

(8.13), for any $\mathbf{x}$ interior to the physical domain $\mathcal{D}$,

$$\begin{aligned}
&\int_{\mathcal{D}} \omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')u^{\mathrm{S}}_j(\mathbf{x}')\,\mathrm{d}m(\mathbf{x}') \\
&\quad = \frac{g}{f}\int_{\mathcal{D}}\left[\left\{\varepsilon_{ik}\frac{\partial^2}{\partial x'_j\partial x'_k} - \varepsilon_{jk}\frac{\partial^2}{\partial x'_i\partial x'_k}\right\}\delta(\mathbf{x}-\mathbf{x}')\right]\, u^{\mathrm{S}}_j(\mathbf{x}')\ h(\mathbf{x}')\ \mathrm{d}\mathbf{x}' \\
&= \frac{g}{f}\int_{\mathcal{D}}\left[\varepsilon_{ij}\frac{\partial^2}{\partial x'_k\partial x'_k}\delta(\mathbf{x}-\mathbf{x}')\right]\, u^{\mathrm{S}}_j(\mathbf{x}')\ h(\mathbf{x}')\ \mathrm{d}\mathbf{x}' \ = \ \frac{g}{f}\varepsilon_{ij}\nabla^2 U^{\mathrm{S}}_j \ , \qquad \text{(A.10)}
\end{aligned}$$

where as before $\mathbf{U}^{\mathrm{S}}(\mathbf{x}) = h(\mathbf{x})\mathbf{u}^{\mathrm{S}}(\mathbf{x})$. To verify this, replace u^{S}_j in the second line by $u^{\mathrm{S}}_p\delta_{pj} = -u^{\mathrm{S}}_p\varepsilon_{pq}\varepsilon_{qj}$ then use $\varepsilon_{ik}\varepsilon_{qj} = \delta_{iq}\delta_{kj} - \delta_{ij}\delta_{kq}$, then integrate by parts twice. Rewriting the $(\boldsymbol{\zeta}^{\mathrm{C}}\times\mathbf{u}^{\mathrm{S}})_i$ term in (8.12) as $-\varepsilon_{ij}Q^{\mathrm{C}}U^{\mathrm{S}}_j$, we see that (8.12) now reduces to

$$\left(\nabla^2 - K^2(\mathbf{x})\right)\mathbf{U}^{\mathrm{S}} = -\frac{f}{g}\hat{\mathbf{z}}\times(\mathbf{u}^{\mathrm{G}}\cdot\nabla\mathbf{u}^{\mathrm{G}}) \ , \qquad \text{(A.11)}$$

for any $\mathbf{x}$ within the domain $\mathcal{D}$, which is just (6.5).

Under the parameter conditions favouring balance, ∇^2 and K^2 reinforce, and we can usually assume that K^2 is not negligible against ∇^2. Then the typical order of magnitude of $\mathbf{u}^{\mathrm{S}}$ is given immediately by comparing the right-hand side of the above equation with the second term $K^2\mathbf{U}^{\mathrm{S}}$ on the left. After cancellation of factors f/g this shows at once that $\mathbf{u}^{\mathrm{S}}$ will be one power smaller in Rossby number ϵ than the relative constraint velocity, $\mathbf{u}^{\mathrm{C(rel)}}$ say, $= \mathbf{u}^{\mathrm{G}}$ in this case.

Appendix B: Symplectic-form invariance and PV conservation

At the end of Section 6 we used the fact that double splitting does not affect material PV conservation. This is because of the well-known fact that PV conservation can be regarded as a corollary of the symplectic structure and the particle-relabelling symmetry, a case of Noether's Second Theorem (e.g. Salmon 1988b, Section 4). The result holds for any Hamiltonian flow, explaining the indifference to the choice of parent Hamiltonian functional.

However, the symplectic structure *per se* is seldom called to mind in this context, any more than a fish would think of mentioning the existence of water; but in the spirit of this volume we thought it would be interesting take a geometric viewpoint in which (material) PV conservation, in a Hamiltonian balanced model constructed by Salmon's method, is regarded as a corollary of the invariance of the symplectic 2-form $\Omega^{\mathrm{CL}} = \int_{\mathcal{D}}\mathrm{d}m\ \mathrm{d}^{\mathrm{L}}x_i\wedge\mathrm{d}^{\mathrm{L}}u^{\mathrm{C}}_i$ inherited from the parent dynamics, where d^{L} is again the exterior derivative in phase space. We thought it would be especially interesting since, contrary to folklore, the argument needs to make use of a nontrivial gambit in order to go directly from symplectic-form invariance to PV conservation. So far we have been unable to find this argument and gambit anywhere in the literature, though a different, elegant route to essentially the same result — a variation on the theme of Noether's theorem, using a multi-symplectic formalism to delete the summation over mass elements and project from phase space $\mathbb{R}^\infty = \mathbb{R}^4 \otimes \mathbb{R}^4 \otimes ...$ into $\mathbb{R}^4$ — is taken in the forthcoming paper by Bridges *et al.* (2001).

The argument presented here (and more sketchily in MR96) applies to the parent dynamics just as well as to as to any balanced model derived from the parent dynamics by Salmon's method. This is because the argument deals with an Eulerian mass configuration h that is not varied and with an Eulerian velocity field, $\mathbf{u}$ in the case of the parent dynamics and $\mathbf{u}^{\mathrm{C}}$ in the case of the balanced model, that is not varied either. That is, the variations required by the argument are such that $\delta^{\mathrm{E}} h = 0$ and, consistently, $\delta^{\mathrm{E}}\mathbf{u} = 0$ or $\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}} = 0$ as appropriate, leading respectively to material conservation of Q or Q^{C}.

The main steps in the balanced-model version of the argument are (a) that any flow advecting a geometric structure in phase space conserves intersection properties; (b) that any such flow therefore conserves the value, Q say, of the symplectic 2-form Ω^{CL} contracted with any pair of variations, i.e. pair of tangent vectors $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$, $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$, when both the vectors and the 2-form are advected by the phase-space flow (recall caption to Figure 1: the 'number of tubes threading A' is conserved); (c) that if the flow, $\dot{\mathbf{x}}(\mathbf{a})$ say, is Hamiltonian (for any Hamiltonian functional $\mathsf{H}^{\mathrm{arbitrary}}$) then Ω^{CL} itself is invariant (its Lie derivative vanishes, $\dot{\mathbf{x}}\rfloor(\mathrm{d}\,\Omega^{\mathrm{CL}}) + \mathrm{d}(\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{CL}}) = \mathrm{d}(\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{CL}}) = \mathrm{d}(\mathrm{d}\mathsf{H}^{\mathrm{arbitrary}}) = 0$) so that Ω^{CL} itself can be considered *not* to be advected, and can be considered to remain equal to the prescribed Ω^{CL} of the balanced model, the corresponding Q still being a constant of the motion provided that $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ are still advected; and (d) that when $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$, which describe material displacement fields in the physical domain $\mathcal{D}$, are chosen such that all Eulerian variations δ^{E} vanish in (8.5) and (8.7) — a choice made possible by the existence of the particle-relabelling symmetry — then Q becomes a weighted physical-domain integral of Q^{C}, namely $\mathsf{Q} = \int_{\mathcal{D}} \mathrm{d}m\, \{\delta^{\mathrm{L}}x(\mathbf{a})\delta^{\mathrm{L}}y'(\mathbf{a}) - \delta^{\mathrm{L}}y(\mathbf{a})\delta^{\mathrm{L}}x'(\mathbf{a}))\} hQ^{\mathrm{C}}$ coming from the first term on the right of (8.7) when substituted into (8.1), with Q^{C} defined by (6.2). All this applies just as well to the parent dynamics if we delete the superscript C from $\mathbf{u}^{\mathrm{C}}$ and Q^{C} in the above and in (6.2), (8.1) and (8.7), continuing to take $\delta^{\mathrm{E}} = 0$ in the last term of (8.7).

The final step (e), which uses the nontrivial gambit mentioned earlier, is to show that the constancy of Q implies the material conservation of Q^{C} or Q. This depends on two things, first that there is enough arbitrariness in the choice of the displacement fields $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ to allow deletion of the summation $\int_{\mathcal{D}} \mathrm{d}m$ over mass elements, and second that the weighting factor multiplying Q^{C} or Q in the integral is a function of mass alone and is therefore materially conserved, i.e. is invariant for each mass element. The remaining factor, Q^{C} or Q, must then be materially conserved also.

Both things can be seen to follow when we confine attention to displacement fields $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ in the form of rigid rotations of a small disk or ring of particles embedded within $\mathcal{D}$, with all the remaining particles left undisplaced. It is simplest to take a small circular ring, call it $\mathcal{R}$, and to take $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ to be rotations of the same ring $\mathcal{R}$ through different angles. The diameter of $\mathcal{R}$ is taken to be much smaller than all spatial scales of the fluid motion; then rigid rotation through any angle leaves the Eulerian mass configuration h undisturbed, as required by the condition $\delta^{\mathrm{E}}h = 0$, since h can now be regarded as locally constant to sufficient accuracy. It is essential — and this is the nontrivial gambit — to make the two angular displacements finitely different, so that for each mass element, i.e. each $\mathbf{a}, \in \mathcal{R}$ the sine of the angle between the two displacement vectors $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ is nonzero, equivalently $\delta^{\mathrm{L}}x(\mathbf{a})\delta^{\mathrm{L}}y'(\mathbf{a}) - \delta^{\mathrm{L}}y(\mathbf{a})\delta^{\mathrm{L}}x'(\mathbf{a}) \neq 0$. The gambit is permissible — we can treat these displacement fields as if they were infinitesimal despite the finite angular displacements — because the parent symplectic structure is flat and homogeneous and because $\partial u_i^{\mathrm{C}}/\partial x_j$ or $\partial u_i/\partial x_j$ can, like h, be taken to be constant over $\mathcal{R}$ to sufficient accuracy. Even though the gradients of the mapping $\mathbf{a} \mapsto \mathbf{x}$ are not, by contrast, approximately constant over $\mathcal{R}$, they are irrelevant and never appear in the argument. In the stratified case $\mathcal{R}$ must, of course, lie in a single stratification surface, so as not to violate the particle-relabelling symmetry.

Finally, the fact that $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ are advected implies that the weighting factor $h\{\delta^{\mathrm{L}}x(\mathbf{a})\delta^{\mathrm{L}}y'(\mathbf{a}) - \delta^{\mathrm{L}}y(\mathbf{a})\delta^{\mathrm{L}}x'(\mathbf{a})\}$ is materially conserved, being h times the area of an advected parallelogram, and the material conservation of Q^{C} or Q follows.

It might be thought that one could find two truly infinitesimal displacement fields $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ that would serve equally well, for instance by taking distinct but overlapping rings $\mathcal{R}$ and $\mathcal{R}'$. However, there are then two contributions to Q, from the two overlap locations, which cancel to give $\mathsf{Q} = 0$ to leading order. It can be shown that all pairs of infinitesimal displacement fields $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}'(\mathbf{a})$ give rise to essentially similar cancellations; this may be why related results in the literature (e.g. Friedman & Schutz 1978; Arnol'd & Khesin 1998) seem to give results about PV increments or gradients and not about the PV itself.

In the alternative route taken by Bridges *et al.* (2001) already mentioned, it is interesting that the multi-symplectic formalism avoids using the particle-relabelling symmetry until after an identity describing the general (nonmaterial) conservation of 'PV-substance' is obtained (Haynes & McIntyre 1990). The particle-relabelling symmetry is used only in order to deduce material conservation from general conservation.

Appendix C: The transformation (9.7)

We start from (10.7), which is the general near-local formula relating $\mathbf{u}^{\mathrm{C}}$ to the canonical coordinates $\mathbf{X}$. In our f-plane, no-topography model, physical acceptability requires invariance to coordinate-axis rotations. The simplest acceptable choices of $\mathbf{X}(\mathbf{x}, h(\cdot)) = \mathbf{X}(x_j, h, h_{,i}, h_{,ij}, \cdots)$ and $B(\mathbf{x}, h(\cdot)) = B(x_j, h, h_{,i}, h_{,ij}, \cdots)$ therefore take the following form:

$$X_i \;=\; x_i + \tilde{a}\frac{\partial h}{\partial x_i} + i\tilde{c}\,\varepsilon_{ij}\frac{\partial h}{\partial x_j}\;, \qquad B = i\tilde{b}h\;, \tag{C.1}$$

where

$$\tilde{a} = \frac{g}{f^2}\;, \quad \tilde{b} = \frac{g}{f}c\;, \quad \tilde{c} = \frac{g}{f^2}c\;. \tag{C.2}$$

The first of (C.1) is (9.7) rewritten in suffix notation. Inserting B from (C.1) into the last line of (10.7) and X_j from (C.1) into the right-hand factors of the first two lines gives (with the B contribution first)

$$\begin{aligned}\frac{1}{f}u_i^{\mathrm{C}} = -2\frac{i\tilde{b}}{f}\frac{\partial h}{\partial x_i} \quad - \tfrac{1}{2}\varepsilon_{jk}\Bigg[X_k\left(\delta_{ij} + \tilde{a}\frac{\partial^2 h}{\partial x_i \partial x_j} + i\tilde{c}\varepsilon_{jp}\frac{\partial^2 h}{\partial x_i \partial x_p}\right) \\ - \frac{\partial^2}{\partial x_i \partial x_q}\Big(hX_k\left(\tilde{a}\delta_{jq} + i\tilde{c}\varepsilon_{jp}\delta_{pq}\right)\Big)\Bigg]\;.\end{aligned} \tag{C.3}$$

Using the identity $\varepsilon_{jk}\varepsilon_{jp}\delta_{pq} = \delta_{kq}$ in (C.3) yields, after some cancellation between terms,

$$\begin{aligned}\frac{1}{f}u_i^{\mathrm{C}} = &-\overbrace{2\frac{i\tilde{b}}{f}\frac{\partial h}{\partial x_i}}^{1} - \overbrace{\tfrac{1}{2}\varepsilon_{ik}X_k}^{2,3,4} \\ &+ \overbrace{\frac{\tilde{a}}{2}\varepsilon_{qk}h\frac{\partial^2 X_k}{\partial x_i \partial x_q}}^{5} + \overbrace{\frac{\tilde{a}}{2}\varepsilon_{qk}\frac{\partial h}{\partial x_q}\frac{\partial X_k}{\partial x_i}}^{6,7,8} + \overbrace{\frac{\tilde{a}}{2}\varepsilon_{qk}\frac{\partial h}{\partial x_i}\frac{\partial X_k}{\partial x_q}}^{9} \\ &+ \overbrace{\frac{i\tilde{c}}{2}h\frac{\partial^2 X_k}{\partial x_i \partial x_k}}^{10} + \overbrace{\frac{i\tilde{c}}{2}\frac{\partial h}{\partial x_i}\frac{\partial X_k}{\partial x_k}}^{11,12} + \overbrace{\frac{i\tilde{c}}{2}\frac{\partial h}{\partial x_k}\frac{\partial X_k}{\partial x_i}}^{13,14,15}\;.\end{aligned} \tag{C.4}$$

The terms are numbered for subsequent reference. Substituting for the remaining X_k factors with (C.1) now gives

$$\frac{1}{f}u_i^{\mathrm{C}} = -\overbrace{2\frac{i\tilde{b}}{f}\frac{\partial h}{\partial x_i}}^{1} - \overbrace{\tfrac{1}{2}\varepsilon_{ik}x_k}^{2} - \overbrace{\frac{\tilde{a}}{2}\varepsilon_{ik}\frac{\partial h}{\partial x_k}}^{3} + \overbrace{\frac{i\tilde{c}}{2}\frac{\partial h}{\partial x_i}}^{4} - \overbrace{\frac{i\tilde{c}\tilde{a}}{2}h\frac{\partial}{\partial x_i}\nabla^2 h}^{5} + \overbrace{\frac{\tilde{a}}{2}\varepsilon_{qi}\frac{\partial h}{\partial x_q}}^{6}$$
$$+ \overbrace{\frac{\tilde{a}^2}{2}\varepsilon_{qk}\frac{\partial h}{\partial x_q}\frac{\partial^2 h}{\partial x_i \partial x_k}}^{7} - \overbrace{\frac{i\tilde{c}\tilde{a}}{2}\frac{\partial h}{\partial x_q}\frac{\partial^2 h}{\partial x_i \partial x_q}}^{8} - \overbrace{\frac{i\tilde{c}\tilde{a}}{2}\frac{\partial h}{\partial x_i}\nabla^2 h}^{9} + \overbrace{\frac{i\tilde{c}\tilde{a}}{2}h\frac{\partial}{\partial x_i}\nabla^2 h}^{10} + \overbrace{i\tilde{c}\frac{\partial h}{\partial x_i}}^{11}$$
$$+ \overbrace{\frac{i\tilde{c}\tilde{a}}{2}\frac{\partial h}{\partial x_i}\nabla^2 h}^{12} + \overbrace{\frac{i\tilde{c}}{2}\frac{\partial h}{\partial x_i}}^{13} + \overbrace{\frac{i\tilde{c}\tilde{a}}{2}\frac{\partial h}{\partial x_k}\frac{\partial^2 h}{\partial x_i \partial x_k}}^{14} - \overbrace{\frac{\tilde{c}^2}{2}\varepsilon_{kl}\frac{\partial h}{\partial x_k}\frac{\partial^2 h}{\partial x_i \partial x_l}}^{15}\ , \tag{C.5}$$

writing ∇^2 for $\partial^2/\partial x_i \partial x_i$. Because $\tilde{b}/f = \tilde{c}$, term 1 cancels with terms 4, 11 and 13. Furthermore, terms 5 and 10, 9 and 12, and 8 and 14 all cancel in pairs. This eliminates *all* the terms proportional to $i\tilde{c}$, and leaves us (after renaming dummy indices) with

$$\frac{1}{f}u_i^{\mathrm{C}} = -\tfrac{1}{2}\varepsilon_{ik}x_k - \tilde{a}\varepsilon_{ik}\frac{\partial h}{\partial x_k} + \frac{\tilde{a}^2 - \tilde{c}^2}{2}\varepsilon_{jk}\frac{\partial h}{\partial x_j}\frac{\partial^2 h}{\partial x_i \partial x_k}\ , \tag{C.6}$$

where, on the right, the surviving terms are 2, 3 and 6, and 7 and 15 respectively. This real-valued result is equivalent to (9.6), with (9.8) and (C.2).

A useful check on the foregoing calculation, including the multiple cancellations in (C.5), is to substitute the first of (C.1) directly into the infinite-dimensional 2-form Ω^{CL} describing the symplectic structure on $\mathcal{M}^{\mathrm{C}}$, $\Omega^{\mathrm{CL}} = \int_{\mathcal{D}} \mathrm{d}m\ \mathrm{d}^{\mathrm{L}}x_i \wedge \mathrm{d}^{\mathrm{L}}u_i^{\mathrm{C}} = \tfrac{1}{2}f\int_{\mathcal{D}} \mathrm{d}m\ \varepsilon_{jk}\ \mathrm{d}^{\mathrm{L}}X_k \wedge \mathrm{d}^{\mathrm{L}}X_j = -\int_{\mathcal{D}} \mathrm{d}m\ \mathrm{d}^{\mathrm{L}}X \wedge \mathrm{d}^{\mathrm{L}}Y$, i.e.

$$-\Omega^{\mathrm{CL}} = f\int_{\mathcal{D}} \mathrm{d}m\ \mathrm{d}^{\mathrm{L}}\left(x_1 + \tilde{a}\frac{\partial h}{\partial x_1} + i\tilde{c}\frac{\partial h}{\partial x_2}\right) \wedge \mathrm{d}^{\mathrm{L}}\left(x_2 + \tilde{a}\frac{\partial h}{\partial x_2} - i\tilde{c}\frac{\partial h}{\partial x_1}\right)\ , \tag{C.7}$$

where $\int_{\mathcal{D}} \ldots \mathrm{d}m$ represents summation over mass elements in the 2-dimensional, Euclidean, physical domain $\mathcal{D}$ or its Lagrangian label space, whereas d^{L} is the exterior derivative in the infinite-dimensional phase space. The expression $\tfrac{1}{2}f\int_{\mathcal{D}} \mathrm{d}m\ \varepsilon_{jk}\ \mathrm{d}^{\mathrm{L}}X_j \wedge \mathrm{d}^{\mathrm{L}}X_k$ is the exterior derivative of the infinite-dimensional 1-form θ^{CL} corresponding to (10.2). This 2-form (C.7) is real-valued, for complex X_j, accounting for the real values of (9.3) and (C.6). The terms proportional to $i\tilde{c}$ add to zero. To see this, note that those terms are

$$i\tilde{c}f\int_{\mathcal{D}} \mathrm{d}m\left[-\mathrm{d}^{\mathrm{L}}x_1 \wedge \mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_1}\right) - \tilde{a}\ \mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_1}\right) \wedge \mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_1}\right)\right.$$
$$\left. + \ \mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_2}\right) \wedge \mathrm{d}^{\mathrm{L}}x_2 + \tilde{a}\ \mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_2}\right) \wedge \mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_2}\right)\right]. \tag{C.8}$$

The second and fourth terms vanish by the skew-symmetry of the wedge product, leaving

$$i\tilde{c}f\int_{\mathcal{D}}\mathrm{d}m\;\mathrm{d}^{\mathrm{L}}\left(\frac{\partial h}{\partial x_i}\right)\wedge\mathrm{d}^{\mathrm{L}}x_i \;=\; i\tilde{c}f\int_{\mathcal{D}}\mathrm{d}m\left[\frac{\partial^2 h}{\partial x_i\partial x_j}\mathrm{d}^{\mathrm{L}}x_j+\mathrm{d}^{\mathrm{E}}\left(\frac{\partial h}{\partial x_i}\right)\right]\wedge\mathrm{d}^{\mathrm{L}}x_i\,, \tag{C.9}$$

where the last step uses $\mathrm{d}^{\mathrm{L}}=\mathrm{d}^{\mathrm{L}}\mathbf{x}.\nabla+\mathrm{d}^{\mathrm{E}}$, for consistency with (8.5), since the exterior derivatives d^{L}, d^{E} can be thought of as placeholders for variations δ^{L}, δ^{E}, with antisymmetrization understood. as when going from (7.12) to (7.9). The second-derivative term vanishes, again by the skew-symmetry of the wedge product. Therefore we are left with $\mathrm{d}^{\mathrm{E}}(\partial h/\partial x_i)\wedge\mathrm{d}^{\mathrm{L}}x_i$. For consistency with (8.4),

$$\mathrm{d}^{\mathrm{E}}h \;=\; -\frac{\partial}{\partial x_j}(h\;\mathrm{d}^{\mathrm{L}}x_j)\;, \tag{C.10}$$

Taking $\partial/\partial x_i$ of this, noting that d^{E} commutes with $\partial/\partial x_i$; so we have

$$i\tilde{c}f\int_{\mathcal{D}}\mathrm{d}m\;\mathrm{d}^{\mathrm{E}}\left(\frac{\partial h}{\partial x_i}\right)\wedge\mathrm{d}^{\mathrm{L}}x_i \;=\; -i\tilde{c}f\int_{\mathcal{D}}\mathrm{d}m\;\frac{\partial^2}{\partial x_i\partial x_j}\left(h\;\mathrm{d}^{\mathrm{L}}x_j\right)\wedge\mathrm{d}^{\mathrm{L}}x_i\;. \tag{C.11}$$

Integration by parts, with $\mathrm{d}m=h\,\mathrm{d}\mathbf{x}$, now gives

$$i\tilde{c}f\int_{\mathcal{D}}\mathrm{d}\mathbf{x}\left[\frac{\partial}{\partial x_j}\left(h\;\mathrm{d}^{\mathrm{L}}x_j\right)\right]\wedge\left[\frac{\partial}{\partial x_i}\left(h\;\mathrm{d}^{\mathrm{L}}x_i\right)\right]\;=\;0\;. \tag{C.12}$$

For bounded domains, it turns out that the boundary conditions implicit in (7.4) and (8.12) make the boundary terms vanish, and the above still holds. Thus the infinite-dimensional 2-form $\Omega^{\mathrm{CL}} = -\int_{\mathcal{D}}\mathrm{d}m\;\;\mathrm{d}^{\mathrm{L}}X\wedge\mathrm{d}^{\mathrm{L}}Y$ is real-valued — re-checking, incidentally, the real-valuedness of its 'reflection', the 2-dimensional Jacobian in (9.3), whose value on a single fluid particle amounts to the projection, or reflection, of Ω^{CL} into a single one of the $\mathbb{R}^2$ subspaces within the phase space $\mathbb{R}^\infty=\mathbb{R}^2\otimes\mathbb{R}^2\otimes\cdots$ of the balanced model.

References

Allen, J.S. & Holm, D.D. 1996. Extended-geostrophic Hamiltonian models for rotating shallow water motion. *Physica D* **98**, 229–248.

Arnol'd, V. I. & Khesin, B. A. 1998. *Topological Methods in Hydrodynamics.* Heidelberg, Springer, 374 pp.

Atiyah, M.F. & Hitchin, N.J. 1988. *The Geometry and Dynamics of Magnetic Monopoles.* Princeton, University Press, 134 pp. See chapters 4–5.

Blumen, W. 1981. The geostrophic coordinate transformation. *J. Atmos. Sci.* **38**, 1100–1105.

Bridges, T.J., Hydon, P.E. & Reich, S. 2001. Vorticity and symplecticity in Lagrangian fluid dynamics. *Proc. Roy. Soc. Lond.* **A**, submitted. Preprint available from Maths. Dept., U. of Surrey, Guildford, Surrey, UK.

Carathéodory, C. 1982. *Calculus of Variations and Partial Differential Equations of the First Order.* 2nd English Edition; New York: Chelsea.

Chynoweth, S. & Sewell, M.J. 1989. Dual variables in semigeostrophic theory. *Proc. Roy. Soc. Lond.* **A 424**, 155–186.

Chynoweth, S. & Sewell, M.J. 1991. A concise derivation of the semi-geostrophic equations. *Q. J. Roy. Meteorol. Soc.* **117**, 1109–1128.

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

Cullen, M.J.P. & Purser, R.J. 1984. An extended Lagrangian theory of semi-geostrophic frontogenesis. *J. Atmos. Sci.* **41**, 1477–1497.

Cullen, M.J.P. & Douglas, R. J. 1999. Applications of the Monge-Ampère equation and Monge transport problem to meteorology and oceanography. In Proc. NSF-CBMS Conference on the *Monge-Ampère Equation: Applications to Geometry and Optimization*, ed. L. A. Caffarelli & M. Milman. Providence, RI, Amer. Math. Soc., *Contemporary Mathematics* **226**, pp. 33–53.

Cullen, M.J.P., Norbury, J. & Purser, R.J. 1991. Generalized Lagrangian solutions for atmospheric and oceanic flows. *SIAM J. Appl. Math.* **51**, 20–31.

Desroziers, G. & Lafore, J.-P. 1993. A coordinate transformation for objective frontal analysis. *Mon. Wea. Rev.* **121**, 1531–1553.

Eliassen, A. 1948. The quasi-static equations of motion with pressure as an independent variable. *Geofys. Publ.* **17**(3), 1–44.

Eliassen, A. 1962. On the vertical circulation in frontal zones. *Geofys. Publ.* **24**, 147–160.

Ford, R., McIntyre, M.E. & Norton, W.A. 2000. Balance and the slow quasimanifold: some explicit results. *J. Atmos. Sci.* **57**, 1236–1254.

Freed, D.S. 1999. Special Kähler manifolds. *Commun. Math. Phys.*, **203**, 31–52.

Friedman, J.L. & Schutz, B.F. 1978. Lagrangian perturbation theory of nonrelativistic fluids. *Astrophys. J.*, **221**, 937–957.

Haynes, P.H. & McIntyre, M.E. 1990. On the conservation and impermeability theorems for potential vorticity. *J. Atmos. Sci.*, **47**, 2021–2031.

Hitchin, N.J. 1987. *Monopoles, minimal surfaces and algebraic curves.* Montreal, Les Presses de l'Université de Montréal, Sém. Math. Supérieures **105**, 94 pp.

Hoskins, B.J. 1975. The geostrophic momentum approximation and the semigeostrophic equations. *J. Atmos. Sci.* **32**, 233–242.

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. Roy. Meteorol. Soc.* **111**, 877–946. Also **113**, 402–404.

Kushner, P.J., McIntyre, M.E. & Shepherd, T.G. 1998. Coupled Kelvin-wave and mirage-wave instabilities in semi-geostrophic dynamics. *J. Phys. Oceanog.* **28**, 513–518.

Lychagin, V.V., Roubtsov, V.N. & Chekalov, I.V. 1993. A classification of Monge–Ampère equations. *Ann. Scient. École Norm. Supér.* 4e série **26**, 281–308.

Magnusdottir, Gudrun, and Schubert, W.H. 1990. The generalization of semigeostrophic theory to the β-plane. *J. Atmos. Sci.*, **47**, 1714–1720.

McDuff, D. & Salamon, D. 1998. *Introduction to Symplectic Topology*, 2nd edn. Oxford, Clarendon Press, 486 pp.

McIntyre, M.E. 2001. Balance, potential-vorticity inversion, Lighthill radiation, and the slow quasimanifold. In *Proc. IUTAM/IUGG/Royal Irish Academy Symposium on Advances in Mathematical Modelling of Atmosphere and Ocean Dynamics*, ed. P.F. Hodnett. Kluwer Academic Publishers, pp. 45–68.

McIntyre, M.E. & Norton, W.A. 2000. Potential-vorticity inversion on a hemisphere. *J. Atmos. Sci.* **57**, 1214–1235; *Corrigendum* **58**, 949, also archived at the web site http://www.atm.damtp.cam.ac.uk/people/mem/

McIntyre, M.E. & Roulstone, I. 1996. Hamiltonian balanced models: constraints, slow manifolds and velocity-splitting. *Forecasting Research Scientific Paper* **41**, *Meteorological Office, U.K.* This report contains tutorial material, now incorporated into the present article; it also contains new results to be resubmitted to *J. Fluid Mech.* The original full text and *Corrigenda* are archived at the web site http://www.atm.damtp.cam.ac.uk/people/mem/

Misner, C.W., Thorne, K.S. & Wheeler, J.A. 1973. *Gravitation.* San Francisco: W.H. Freeman.

Mohebalhojeh, A.R. & Dritschel, D.G. 2001. Hierarchies of balance conditions for the f-plane shallow water equations. *J. Atmos. Sci.*, in press.

Mohebalhojeh, A.R. & McIntyre, M.E. 2001. Local mass conservation and velocity splitting in PV-conserving balanced models. *J. Atmos. Sci.*, submitted.

Norton, W.A. 1988. *Balance and potential vorticity inversion in atmospheric dynamics.* PhD Thesis, University of Cambridge. Available from the Superintendent of Manuscripts, Cambridge University Library, West Rd, Cambridge CB3 9DR.

Purser, R.J. 1993. Contact transformations and Hamiltonian dynamics in generalized semigeostrophic theories. *J. Atmos. Sci.*, **50**, 1449–1468.

Purser, R.J. & Cullen, M.J.P. 1987. A duality principle in semigeostrophic theory. *J. Atmos. Sci.* **44**, 3449–3468.

Rossby, C.-G. 1936. Dynamics of steady ocean currents in the light of experimental fluid mechanics. *Mass. Inst. of Technol. & Woods Hole Oc. Inst. Papers in Physical Oceanography and Meteorology* **5**(1), 1–43. [See especially Eq. (75).]

Roubtsov, V.N. & Roulstone, I. 1997. Examples of quaternionic and Kähler structures in Hamiltonian models of nearly geostrophic flow. *J. Phys. A* **30**, L63–L68.

Roubtsov, V.N. & Roulstone, I. 2001. Holomorphic structures in hydrodynamical models of nearly-geostrophic flow. *Proc. Roy. Soc. Lond.* **A 457**, 1519–1531.

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

Roulstone, I. & Sewell, M.J. 1996. Potential vorticities in semi-geostrophic theory. *Q. J. Roy. Meteorol. Soc.* **122**, 983–992.

Roulstone, I. & Sewell, M.J. 1997. The mathematical structure of theories of semi-geostrophic type. *Phil. Trans. Roy. Soc. Lond.* **A 355**, 2489–2517.

Salmon, R. 1983. Practical use of Hamilton's principle. *J. Fluid Mech.* **132**, 431–444.

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

Salmon, R. 1988a. Semi-geostrophic theory as a Dirac-bracket projection. *J. Fluid Mech.* **196**, 345–358.

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

Saujani, S. & Shepherd, T.G. 2001. Comments on "Balance and the Slow Quasi-manifold: Some Explicit Results". *J. Atmos. Sci.* **58**, in press; Reply by Ford, R., McIntyre, M.E. & Norton, W.A., *ibid.*, in press.

Schutz, B.F. 1980. *Geometrical Methods of Mathematical Physics.* Cambridge University Press.

Sewell, M.J. & Roulstone, I. 1993. Anatomy of the canonical transformation. *Phil. Trans. Roy. Soc. Lond.* **A 345**, 577–598.

Sewell, M.J. & Roulstone, I. 1994. Families of lift and contact transformations. *Proc. Roy. Soc. Lond.* **A 447**, 493–512.

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

Shutts, G.J. 1989. Planetary semi-geostrophic equations derived from Hamilton's principle. *J. Fluid Mech.*, **208**, 545–573.

Snyder, C., Skamarock, W.C., & Rotunno, R. 1991. A comparison of primitive-equation and semigeostrophic simulations of baroclinic waves. *J. Atmos. Sci.*, **46**, 2179–2194.

Theiss, J. 1999. *Hamiltonian balanced models in the atmosphere.* PhD thesis, University of Cambridge. Available from the Superintendent of Manuscripts, Cambridge University Library, West Rd, Cambridge CB3 9DR.

Wunderer, C.F. 2001. *Hamiltonian theories of balanced vortical flow.* PhD thesis in preparation, DAMTP, University of Cambridge.