Hamiltonian balanced models: constraints, slow manifolds and velocity splitting

By

Michael E. McIntyre
Centre for Atmospheric Science[*] *at the*
Department of Applied Mathematics and Theoretical Physics,
Silver Street, Cambridge CB3 9EW. UK.

and

Ian Roulstone
Meteorological Office,
London Road, Bracknell RG12 2SZ. UK.

Submitted to the Journal of Fluid Mechanics

1 September 1996

Abstract

The general mathematical structure of Hamiltonian balanced models of vortical atmosphere–ocean dynamics is clarified, at arbitrary accuracy, along with its relation to the concepts of slow manifold, slow quasi-manifold and potential-vorticity (PV) inversion. Accuracy means closeness to an exact dynamics, meaning a primitive or Euler-equation Hamiltonian dynamics (with constant or variable Coriolis parameter), regarded as the exact 'parent' of the balanced model. Arbitrary means limited not by any particular expansion method or approximate formula, or fast–slow scaling assumption, but only by the irreducible, residual inaccuracy or imbalance associated with the Lighthill radiation, or spontaneous-adjustment emission, of inertia–gravity waves by unsteady vortical motions. The clarification shows (a) which features of Hamiltonian balanced models like Hoskins' semigeostrophy and Salmon's L_1 dynamics are special and which are general, and how such models can be generalized to arbitrary accuracy without losing Hamiltonian structure, (b) how such generalizations can be constructed, at least formally, by inserting into the Hamiltonian framework any of the highly accurate balance conditions used in recent studies of accurate PV inversion, and (c) how, in a certain class of models, canonical coordinates $\mathbf{X}$ analogous to Hoskins' geostrophic-momentum coordinates can be found. This class, defined by a new pair of 'canonical coordinate theorems', includes both Salmon's L_1 dynamics (in two versions, one of which filters boundary Kelvin waves) and also a new f-plane Hamiltonian balanced model accurate to a further order in Rossby number, that is, accurate to the same formal order as the standard (non-Hamiltonian) 'balance equations'. For these particular cases $\mathbf{X}$ takes the form $\mathbf{x} + \nabla\varphi - \gamma\hat{\mathbf{z}} \times \nabla\varphi$ (cf. Hoskins' $\mathbf{x} + \nabla\varphi$), where $\nabla = (\partial/\partial x, \partial/\partial y)$ and $\hat{\mathbf{z}} \times \nabla = (-\partial/\partial y, \partial/\partial x)$, φ being f^{-2} times a suitably defined geopotential anomaly or Montgomery potential anomaly, and $\mathbf{x} = (x, y)$ the horizontal coordinates for the physical domain. A remarkable and surprising feature is that the dimensionless

*The Centre for Atmospheric Science is a joint initiative of the Department of Chemistry and the Department of Applied Mathematics and Theoretical Physics.

coefficient γ must be pure imaginary, $\gamma = i$ for L_1 dynamics and $\gamma = i\sqrt{3}$ for the new, more accurate balanced model. Reality of the symplectic one-form and two-form requires only that γ^2 be real, not γ. The Legendre duality property of semigeostrophy disappears, along with the semigeostrophic Monge–Ampère equation, when $\gamma \neq 0$.

All Hamiltonian balanced models, regardless of accuracy — if derived by constraining a parent dynamics supporting inertia–gravity waves and Lighthill radiation — have a characteristic general property that may be called 'velocity splitting', recognition of which is important for a full understanding of the dynamics. Imposing any balance condition as a constraint on the parent dynamics splits the parent velocity field, into just two velocity fields in the simplest set of cases. The first, $\mathbf{u}^{\mathrm{P}}$ say, is the velocity in the ordinary sense, the velocity with which material particles move. The second, $\mathbf{u}^{\mathrm{C}}$ say, is the constraint velocity defining the balanced model's 'slow manifold', a prescribed functional of the mass configuration alone; $\mathbf{u}^{\mathrm{C}}$ is also the velocity that enters the simplest forms of the model's PV, energy and other conserved quantities, which quantities, when evaluated with $\mathbf{u}^{\mathrm{C}}$, are given by the same formulae as in the parent dynamics. The model's conserved PV, in particular, is always the Rossby–Ertel PV when expressed in terms of the 'constraint vorticity' $\boldsymbol{\zeta}^{\mathrm{C}}$, i.e. the absolute vorticity evaluated with $\mathbf{u}^{\mathrm{C}}$, and furthermore is always given by the same (real) Jacobian determinants involving $\mathbf{X}$ as in Hoskins' semigeostrophy, in cases where canonical coordinates $\mathbf{X}$ are known. Semigeostrophy itself is more complicated, from this viewpoint, exhibiting double splitting in the sense of having two slow manifolds and three distinct velocity fields $\mathbf{u}^{\mathrm{P}}$, $\mathbf{u}^{\mathrm{C(H)}}$ and $\mathbf{u}^{\mathrm{C}(\Omega)}$, of which the second enters into energy conservation via the parent Hamiltonian and the third into PV conservation via the parent symplectic structure.

There is a fundamental integral equation that governs the difference-velocity field or 'velocity split' $\mathbf{u}^{\mathrm{S}} = \mathbf{u}^{\mathrm{P}} - \mathbf{u}^{\mathrm{C}}$, and defines the balanced model dynamics including any boundary conditions additional to those used in defining the configuration space. The formulation in terms of $\mathbf{u}^{\mathrm{S}}$ gives a mathematical description that is manifestly independent of reference-frame rotation rate. The dynamical effects of rotation enter solely through the $\mathbf{u}^{\mathrm{C}}$ functional or functionals, for instance through $\boldsymbol{\zeta}^{\mathrm{C}}$. This greatly simplifies the formulation of variable-Coriolis-parameter models, and brings other conceptual and formal simplifications such as the ability to accommodate Galilean-invariant $\mathbf{u}^{\mathrm{C}}$. The integral equation takes an especially simple form for singly-split models. Solving the equation gives $\mathbf{u}^{\mathrm{S}}$ as a functional of the mass configuration alone, defining, in turn, how particles move, and showing that $\mathbf{u}^{\mathrm{S}}$ is generally nonzero though small for an accurate model. Any norm $\|\mathbf{u}^{\mathrm{S}}\|$ provides a natural intrinsic measure of model accuracy, suggesting minimization of such norms as an approach to finding 'optimal' Hamiltonian balanced models. The nonvanishing of $\|\mathbf{u}^{\mathrm{S}}\|$ is connected with the existence of Lighthill radiation and the nonexistence of a parent slow manifold.

1 Introduction

A longstanding problem in the theory of atmosphere–ocean vortex dynamics, closely relevant to today's problems of data assimilation and environmental forecasting, has been how to generalize simple balanced models like Hoskins' semigeostrophic theory, hereafter 'semigeostrophy', to higher accuracy while preserving Hamiltonian structure. For instance, Hoskins' elegant canonical coordinates $\mathbf{X}$ (Hoskins 1975) have been important in clarifying salient mathematical features of semigeostrophy, such as Legendre duality, and their use in understanding atmosphere–ocean phenomena such as fronts (e.g. Roulstone & Sewell 1996*b*, and many references therein). This paper builds on the foregoing work and on the pioneering work of Salmon (1983, 1985, 1988*a*) to further clarify what is involved, both in generalizing Hamiltonian balanced models beyond semigeostrophy and in finding canonical coordinates for them.

The crucial step is to recognize a universal general property, velocity splitting (not to be confused with regime bifurcation) that characterizes all balanced models that inherit their Hamiltonian structure from an exact Hamiltonian 'parent' dynamics. The parent dynamics is assumed to support inertia–gravity waves, and is regarded as exact in the sense that it provides the standard of accuracy against which the balanced model is to be assessed. In this paper the parent dynamics is assumed furthermore to have a particle-relabelling symmetry and a corresponding materially conserved Rossby–Ertel potential vorticity (PV).

The simplest, most straightforward and conceptually clearest way in which a Hamiltonian balanced model can be constructed from an exact parent is to impose a balance condition — usually of geostrophic or higher accuracy — that constrains the dynamics to a single 'slow' manifold $\mathcal{M}^{\mathrm{C}}$ within the full phase space of the parent dynamics. More precisely, this means imposing a suitable workless 'momentum–configuration constraint' $\mathbf{u} = \mathbf{u}^{\mathrm{C}}$ on the parent velocity field $\mathbf{u}$, regarded as a phase space coordinate, where $\mathbf{u}^{\mathrm{C}}$ is a prescribed functional of the mass configuration that respects the particle-relabelling symmetry. Thus $\mathbf{u}^{\mathrm{C}}$ is prescribable solely as a functional of the Eulerian fields $\rho(\mathbf{x})$ and $\sigma(\mathbf{x})$ of mass and entropy per unit volume for three-dimensional stratified dynamics, or of layer depth $h(\mathbf{x})$ for shallow water dynamics, where $\mathbf{x} \in \mathcal{D}$ is position in the physical domain $\mathcal{D}$. The simplest nontrivial choice is to prescribe $\mathbf{u}^{\mathrm{C}}$ as the geostrophic velocity; but some of the more accurate possible choices require a fully nonlocal functional dependence. Velocity splitting means that the actual particle velocity $\mathbf{u}^{\mathrm{P}}$ predicted by a Hamiltonian balanced model, viewed in the physical domain — what would normally be regarded as 'the' velocity — always differs from $\mathbf{u}^{\mathrm{C}}$, no matter how accurate $\mathbf{u}^{\mathrm{C}}$ may be, except in at most a tiny subset of special cases whose Lighthill radiation is exactly zero. The only known such cases are certain steady vortex and Rossby-wave flows. The significance of Lighthill radiation, the 'spontaneous-adjustment emission' of inertia–gravity waves by unsteady vortical motions, will emerge shortly. Note that the constraint velocity $\mathbf{u}^{\mathrm{C}}$ is both a field, a function of $\mathbf{x}$, and also a mass-configuration functional; and when one or other aspect needs emphasis we refer to the '$\mathbf{u}^{\mathrm{C}}$ field' or the '$\mathbf{u}^{\mathrm{C}}$ functional'. When both need emphasis, as when dealing with $\mathbf{u}^{\mathrm{C}}$ mathematically, we sometimes use explicit notations like $\mathbf{u}^{\mathrm{C}}(\mathbf{x}; \rho(\cdot), \sigma(\cdot))$ or $\mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$.

We find it mnemonically useful to say that imposing the constraint splits the parent velocity field $\mathbf{u}$ into $\mathbf{u}^{\mathrm{P}}$ and $\mathbf{u}^{\mathrm{C}}$. Such splitting is possible because of the fact that, in a classical Hamiltonian system, 'velocity field' has two separate though possibly coincident meanings, particle velocity field and phase space coordinate. It is the latter, not the former, that is constrained. A sufficient reason to pay attention to both velocity fields is that the conservation properties of the balanced model take simple forms only when expressed in terms of $\mathbf{u}^{\mathrm{C}}$, not $\mathbf{u}^{\mathrm{P}}$. In particular, the model's conserved PV is always the parent Rossby–Ertel PV evaluated with $\mathbf{u}^{\mathrm{C}}$, whose respective forms for three-dimensional stratified dynamics and shallow water dynamics are simply

$$Q^{\mathrm{C}} = \rho^{-1} \boldsymbol{\zeta}^{\mathrm{C}} . \nabla(\sigma/\rho) \qquad \text{and} \qquad Q^{\mathrm{C}} = h^{-1} \zeta^{\mathrm{C}} \; ; \tag{1}$$

$\boldsymbol{\zeta}^{\mathrm{C}}$ and ζ^{C} are respectively the three-dimensional and two-dimensional forms of the 'constraint vorticity', the absolute vorticity evaluated with $\mathbf{u}^{\mathrm{C}}$, and are natural generalizations of the standard notion of absolute 'geostrophic vorticity'. Similar statements hold for energy and momentum.

These general results were implicit in Salmon's work and in a recent study of semigeostrophy by Roulstone & Sewell (1996*a*, eqs. (50)ff.), and have been noted explicitly in the independent work of Allen & Holm to be mentioned below. It is emphasized that the purpose of calling $\mathbf{u}^{\mathrm{C}}$ a velocity, despite its not being a true velocity of material particles, is the mnemonic value of the simple rule for obtaining the conserved quantities of the balanced model. The rule is that one substitutes $\mathbf{u}^{\mathrm{C}}$ into the corresponding parent conserved quantities, as if $\mathbf{u}^{\mathrm{C}}$ were the true particle velocity. The difference field or 'velocity-split' $\mathbf{u}^{\mathrm{S}}(\mathbf{x}) = \mathbf{u}^{\mathrm{P}} - \mathbf{u}^{\mathrm{C}}$ can be thought of as a natural generalization of the 'ageostrophic velocity' that appears in simple models like semigeostrophy or Salmon's L_1 dynamics.

The fact that $\mathbf{u}^{\mathrm{S}}(\mathbf{x})$ is generally nonzero can, under reasonable assumptions, be proved rigorously. The essential points are sketched here and set out carefully in §3 below, where a relevant 'velocity-splitting theorem' is proved. This theorem makes only the mildest of assumptions about the prescribed $\mathbf{u}^{\mathrm{C}}$ functional and the associated slow manifold $\mathcal{M}^{\mathrm{C}}$. It amounts to saying that if the constraint has any effect at all on the parent dynamics, then $\mathbf{u}^{\mathrm{S}}(\mathbf{x})$ cannot be identically zero. Conversely, only a constraint having no effect at all will fail to split the parent velocity field $\mathbf{u}$. But the existence of Lighthill radiation in the parent dynamics means that — except in a tiny subset of special cases — the imposition of the constraint must, in fact, have a nonvanishing effect. It must act to suppress the Lighthill radiation. The existence of nonvanishing Lighthill radiation, from all or nearly all unsteady vortical motions, is one of the 'reasonable assumptions' we make. This has not been proved rigorously, to our knowledge, but is supported by overwhelming analytical and numerical evidence (e.g. Crighton 1981, Ford *et al.* 1996, & refs.). That evidence includes powerful heuristic analytical arguments of the kind originally used by Lighthill, for the fundamentally similar problem of aerodynamic sound generation. It also includes corroborative laboratory experiments (e.g. Webster 1970; Kambe *et al.* 1990, 1993).

The quantity measuring the effect of the constraint on the parent dynamics, for these purposes, is a field $\mathbf{R}(\mathbf{x})$ that besides depending on position $\mathbf{x}$ has functional dependence, like $\mathbf{u}^{\mathrm{C}}$, on the mass configuration. $\mathbf{R}$ may be called the residual, or unbalanced, contribution to the gravitational and pressure-gradient forces per unit mass. It shares with $\mathbf{u}^{\mathrm{S}}$ the convenient property of being indifferent to the choice of rotation rate for the frame of reference. In an inertial frame, $\mathbf{R}$ is defined by

$$\mathbf{R} = \mathbf{F} - \mathbf{A}^{\mathrm{C}} \ , \tag{2}$$

where $\mathbf{A}^{\mathrm{C}}(\mathbf{x})$ will be called the absolute 'constraint acceleration' (including what are usually counted as centrifugal and Coriolis accelerations) and where $\mathbf{F}(\mathbf{x})$ is the net gravitational and pressure-gradient force per unit mass defined by the parent dynamics for the given mass configuration. Thus $\mathbf{F}$ 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 also a functional of the mass configuration. $\mathbf{A}^{\mathrm{C}}$ is the material rate of change of $\mathbf{u}^{\mathrm{C}}$ evaluated from a fictitiously evolving mass configuration that coincides, at the moment considered, with the given configuration but changes as if all particles moved with velocity $\mathbf{u}^{\mathrm{C}}$. Thus $\mathbf{R}$ would be identically zero if this fiction were fact, i.e., if the constraint had no effect on the parent dynamics. The superscript C will be used throughout, as just illustrated, to signal a functional dependence on the mass configuration that is known as soon as the $\mathbf{u}^{\mathrm{C}}$ functional is given. The field $\mathbf{A}^{\mathrm{C}}$ qualifies as having such a 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 first functional derivatives. Such derivatives enter the reckoning when taking account of the fictitious rate of change of the configuration when particles move with velocity $\mathbf{u}^{\mathrm{C}}$. With these definitions, the abovementioned 'velocity-splitting theorem' states that *every mass configuration whose* $\mathbf{R}$ *field is not everywhere zero in the physical domain* $\mathcal{D}$ *has a* $\mathbf{u}^{\mathrm{S}}$ *field that is not everywhere zero in* $\mathcal{D}$.

The theorem can be established most easily by showing, as will be done in §3, that the $\mathbf{R}(\mathbf{x})$ and $\mathbf{u}^{\mathrm{S}}(\mathbf{x})$ fields are connected, exactly, by a certain linear integral equation, for any given $\mathbf{u}^{\mathrm{C}}$ functional. This equation is of interest for other reasons as well. It not only provides an easy proof of the theorem but also a general, but conceptually economical, way of defining the entire balanced dynamics associated with the given $\mathbf{u}^{\mathrm{C}}$, avoiding any artificial dependence on reference-frame rotation rates. It provides an easy route, for instance, to variable-Coriolis-parameter models. It explicitly shows why imposing the constraint must suppress Lighthill radiation. It takes the form

$$\sum_j \int_{\mathcal{D}} \Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')\ u^{\mathrm{S}}_j(\mathbf{x}')\ \mathrm{d}m(\mathbf{x}')\ = R_i(\mathbf{x})\ , \tag{3}$$

where the kernel $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$, to be defined in §3, is antisymmetric in the sense that $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}') = -\Omega^{\mathrm{C}}_{ji}(\mathbf{x}',\mathbf{x})$, and like $\mathbf{A}^{\mathrm{C}}$ is known in terms of $\mathbf{u}^{\mathrm{C}}$ and the mass configuration. Here $\mathrm{d}m(\mathbf{x}')$ is the mass element, $\rho(\mathbf{x}')\mathrm{d}\mathbf{x}'$ or $h(\mathbf{x}')\mathrm{d}\mathbf{x}'$, and the index j is summed over the space dimensions of the physical domain $\mathcal{D}$, 1 to 3 for stratified and 1 to 2 for shallow water flow. The theorem now follows at once, by reductio ad absurdum, for any $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ whose behaviour is good enough for the left hand side of (3) to vanish when $\mathbf{u}^{\mathrm{S}}$ is identically zero. This is hardly a significant restriction. It will become clear by example that the behaviour of Ω^{C}_{ij} is, indeed, typically more than good enough for this purpose; $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ will be seen to consist of a term equal to the constraint vorticity at $\mathbf{x}$ times a Dirac delta function $\delta(\mathbf{x}-\mathbf{x}')$ plus further terms proportional to spatial derivatives of $\delta(\mathbf{x}-\mathbf{x}')$ plus, in some cases, finite-valued, integrable contributions, corresponding to fully nonlocal contributions to $\mathbf{u}^{\mathrm{C}}$.

If, furthermore, $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ is invertible, then equation (3) can be regarded as a prognostic equation for the balanced dynamics, with the nice property of being reference-frame-independent in a natural sense. The dynamical effects of rotation enter solely through $\mathbf{u}^{\mathrm{C}}$ and the associated (absolute) constraint vorticity. Solving (3) yields $\mathbf{u}^{\mathrm{S}}$, and thence $\mathbf{u}^{\mathrm{P}}$, given the mass configuration, allowing integration forward in time t. Notice that, because $\mathbf{u}^{\mathrm{P}}$ is only the first time derivative of particle position, no dynamics of the form (3) can exhibit Lighthill radiation, simply because such dynamics cannot support inertia–gravity wave motion. Indeed it cannot support bidirectional wave propagation of any kind, because of the single time derivative. That is why imposing the constraint must always suppress such propagation, and radiation, together with the associated arrow-of-time effects, implying $\mathbf{R} \neq 0$.

A noteworthy corollary of the foregoing is that the parent dynamics, given that it supports inertia–gravity waves, cannot possess an exact, invariant slow manifold in phase space (e.g Errico 1982, Warn 1996, Warn & Ménard 1986, Ford *et al.* 1996). If there were such an exact parent slow manifold, then it would be possible to choose $\mathbf{u}^{\mathrm{C}}$ to coincide with it. This would make $\mathbf{R}$ identically zero for all time, by its definition (2), for any evolution on that manifold, contradicting the above. As Ford *et al.* (*op. cit.*) point out, this can be made sense of by assuming that the parent dynamics must possess not a single, invariant slow manifold but only what is sometimes called a 'slow quasi-manifold' $\mathcal{Q}$ in the form of a thin chaotic or 'stochastic' layer, of the kind familiar from studies of finite numbers of weakly coupled pendulums or other oscillators. $\mathcal{Q}$ is thin to the extent that Lighthill radiation is weak. The most accurate $\mathbf{u}^{\mathrm{C}}$ functionals correspond to $\mathcal{M}^{\mathrm{C}}$ lying within $\mathcal{Q}$.

We have not been able to prove any general result about the invertibility of $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ but, on the evidence so far, invertibility seems likely to hold for suitably chosen $\mathbf{u}^{\mathrm{C}}$ functionals that are close to geostrophic or gradient-wind balance. Invertibility has a geometric meaning in infinite dimensional phase space; some aspects of this are shown in figure 1 below, via the simplest finite-dimensional analogue. The technicalities in infinite dimensions are nontrivial, and include what becomes, in infinite dimensions, a rather subtle distinction between symplectic and contact structure. This is touched on in §8; a full discussion of this distinction is beyond the scope of this paper, and in any case seems not to be crucial to the invertibility of $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ in well-behaved function spaces. One choice of $\mathbf{u}^{\mathrm{C}}$ for which invertibility has been explicitly verified is the case of Salmon's L_1 dynamics, corresponding to geostrophic $\mathbf{u}^{\mathrm{C}}$. In that case invertibility not only holds but is demonstrably robust, as will be shown in §§4–7 and Appendix A for the simplest, nontopographic f-plane shallow water case. We may conjecture that similar results hold for other $\mathbf{u}^{\mathrm{C}}$ functionals corresponding to physically reasonable balance conditions, representing refinements to geostrophic accuracy; but the corresponding mathematics is again nontrivial. In the case of L_1 dynamics we shall find that (3) can be reduced to a second order elliptic partial differential equation, as originally shown by Salmon (1985), together with exactly enough information about boundary conditions to complement the configuration-space boundary conditions chosen *a priori*. For L_1 dynamics there are at least two reasonable choices for the latter, one of which leads to a new 'filtered L_1 dynamics', so called because it filters boundary Kelvin waves and is therefore fully PV-invertible, in a sense to be explained in §7.

A wider class of models that includes both versions of L_1 dynamics is that of 'near-local' Hamiltonian balanced models, meaning those defined by $\mathbf{u}^{\mathrm{C}}$ functionals that are near-local in the sense that $\mathbf{u}^{\mathrm{C}}$ depends only locally on $\rho(\mathbf{x})$, $\sigma(\mathbf{x})$ or $h(\mathbf{x})$ and on a finite number of their spatial derivatives. We say 'near-local' rather than 'local' to emphasize that a neighbourhood of $\mathbf{x}$ is involved (because for instance pressure gradients are involved), and not just the point values of

$\rho(\mathbf{x})$ etc. at $\mathbf{x}$. It is known that the accuracy of the balance condition $\mathbf{u} = \mathbf{u}^{\mathrm{C}}$, whether used in a Hamiltonian or a non-Hamiltonian model, is then limited to just two orders in Rossby number Ro, or one more than for L_1 dynamics and semigeostrophy — that is, to the same formal order of accuracy as a first approximation to gradient-wind balance, or the standard (non-Hamiltonian) 'balance equations' of Bolin and Charney (e.g. Whitaker 1993). To get further orders of accuracy it would be necessary to use a fully nonlocal functional dependence, equivalently an infinite number of derivatives at $\mathbf{x}$, because of the so called omega-equation nonlocality of certain horizontally irrotational contributions to $\mathbf{u}^{\mathrm{C}}$ ('omega' equation referring to the elliptic equation governing such contributions, not to Ω^{C}, e.g., Hoskins *et al.* (1978)). Nevertheless, two orders in Ro is already a practically useful order of accuracy, far superior to geostrophic in problems typical of the real atmosphere for instance; and the near-local Hamiltonian balanced models prove to be interesting theoretically as well as practically.

One theoretically interesting point is that we can find very simple canonical coordinates $\mathbf{X}$ like those of Hoskins (1975) for a significantly wider class of near-local models than semigeostrophy; and these canonical coordinates have a surprising twist. The discovery of these models, which for brevity we call 'near-local canonical models', was originally motivated by recent results on the application of the theory of contact transformations to Hoskins' semigeostrophy (Sewell & Roulstone 1994, Roulstone & Sewell 1996*b*). §8 derives the relevant general results as a pair of 'canonical coordinate theorems' that characterize such models and their further generalizations. The theorems establish for what class of Hamiltonian balanced models, near-local or fully nonlocal, a given $\mathbf{X}$ is canonical. It does so by characterizing the corresponding class of $\mathbf{u}^{\mathrm{C}}$ functionals, showing for instance how the special case found by Salmon (1988*a*) fits into a fully general framework. These models include a large infinity of models with variable Coriolis parameter (§10 below). Near-local canonical models, with both constant and variable Coriolis parameter, make up an important subclass.

It turns out that L_1 dynamics is a member of this subclass, as is also a new Hamiltonian balanced model accurate to two orders in Ro, which we put forward as a strong candidate for further investigation. The simplest, f-plane versions of these models correspond, in an inertial frame of reference, to

$$\mathbf{u}^{\mathrm{C}} \;=\; \tfrac{1}{2}f\hat{\mathbf{z}}\times\mathbf{x} \;+\; \mathbf{u}^{\mathrm{G(rel)}} \;-\; \alpha\,\hat{\mathbf{z}}\times f^{-1}\mathbf{u}^{\mathrm{G(rel)}}.\nabla\mathbf{u}^{\mathrm{G(rel)}} \tag{4}$$

with different choices of α, a constant parameter. Here $\frac{1}{2}f$ is some typical or average value of the rotating fluid's bulk absolute angular velocity, also taken here as a constant, $\hat{\mathbf{z}}$ is a unit vertical vector, and $\mathbf{u}^{\mathrm{G(rel)}}$ is the relative geostrophic velocity associated with the mass configuration and the parameter f, i.e.,

$$\mathbf{u}^{\mathrm{G(rel)}} = \hat{\mathbf{z}}\times\frac{1}{f}\nabla\phi = \frac{1}{f}\left(-\frac{\partial\phi}{\partial y},\frac{\partial\phi}{\partial x}\right)\ , \tag{5}$$

where ∇ denotes the horizontal gradient operator, and ϕ is the relevant surface-elevation geopotential for shallow water dynamics, or Montgomery potential for stratified dynamics in an isentropic-coordinate description. Then the particular near-local canonical models defined by (4) turn out to have canonical coordinates given by the remarkably simple formula $\mathbf{X} = \mathbf{x} + f^{-2}(\nabla\phi - \gamma\hat{\mathbf{z}} \times \nabla\phi)$, i.e.

$$X = x + \frac{1}{f^2}\left(\frac{\partial\phi}{\partial x} + \gamma\frac{\partial\phi}{\partial y}\right), \quad Y = y + \frac{1}{f^2}\left(\frac{\partial\phi}{\partial y} - \gamma\frac{\partial\phi}{\partial x}\right) , \tag{6}$$

provided that

$$\gamma = \sqrt{2\alpha - 1} \; ; \tag{7}$$

and $\alpha = \frac{1}{2}$, i.e. $\gamma = 0$, gives Hoskins' $\mathbf{X}$, for which (4) becomes the constraint discovered by Salmon (1988*a*, eq. (5.18)); for the relevance to Hoskins' semigeostrophy see §11 below. But now comes the surprising twist. To get a more accurate model of this class, α must be chosen $< \frac{1}{2}$, i.e. γ must be chosen to be pure imaginary.

In particular, the choice $\gamma = i \;\; (\alpha = 0)$ provides canonical coordinates for Salmon's L_1 dynamics — a new result in itself. But, still more interestingly, the choice $\gamma = i\sqrt{3} \;\; (\alpha = -1)$ defines a new model with canonical coordinates $\mathbf{X}$ that has one further order of accuracy in Ro . The choice $\gamma = i\sqrt{3}$ is the only such choice when $\mathbf{X}$ has the simplicity of (6).

These results suggest that, in order to gain such accuracy, it is necessary to depart from the gradient form of $\mathbf{X}$ exemplified by (6) with $\gamma = 0$, which gradient form gives rise to the Monge–Ampère equation of semigeostrophy (e.g. Purser & Cullen 1987; Roulstone & Sewell 1996*b*). The suggestion is confirmed in Appendix B below, which examines the class of near-local canonical models in which $\mathbf{X}$ has the most general possible gradient form that respects the particle-relabelling symmetry. It is found that all these models lead to the same Monge–Ampère equation as in semigeostrophy, but that none of them can be formally more accurate than semigeostrophy itself.

A point to emphasize is that the conceptual structure we are dealing with is simple. This becomes most evident from the modern geometric view of Hamiltonian dynamics, which maintains a clear distinction between the symplectic or contact structure on the one hand and the Hamiltonian functional on the other. Imposing the constraint representing the balance condition then becomes a very straightforward process and avoids, for instance, the introduction of the traditional Lagrange multipliers. Nor is there any need to consider solvability or consistency conditions, or sequences of such conditions, of the sort that arise in the Dirac–Bergmann theory of constraints (e.g. Gotay *et al.* 1978). There the problem is of a different and more complicated kind, that of discovering implicit, pre-existing constraints or 'Lagrangian submanifolds' associated with degenerate symplectic structure due to the use of too large a phase space at the outset. All that is relevant here is the nondegeneracy of the imposed constraint represented by the $\mathbf{u}^{\mathrm{C}}$ functional, as reflected by the invertibility of $\Omega^{\mathrm{C}}_{ij}(\mathbf{x}, \mathbf{x}')$. The essentials can be appreciated from simple finite-dimensional examples; the simplest such is given in the next section.

The plan of the paper is as follows. §2 presents the simplest finite-dimensional example, and §3 shows how this extends to fluid systems respecting the particle relabelling symmetry, establishing the 'velocity-splitting theorem' on the nonvanishing of $\mathbf{u}^{\mathrm{S}}$ together with the fundamental equation (3) that leads to it. The simplest system for which the balance concept is nontrivial is f-plane shallow water dynamics, and this is used for the most part when making details explicit, though the formula for $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ is exhibited also for the three-dimensional stratified case, as are the general canonical coordinate theorems of §8. §4 shows how the fundamental equation (3) quickly leads to the equations of L_1 dynamics, as first derived via a longer route by Salmon (1983). This provides a clear example in which $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ is robustly invertible. §5 describes a general procedure for extracting boundary conditions from (3), and shows that the balanced model conserves energy when those conditions are satisfied; §§6 and 7 illustrates the procedure by deriving and discussing the two sets of boundary conditions for L_1 dynamics. §7 also discusses the sense in which filtered L_1 dynamics is PV-invertible, with implications for numerical solution. §8 presents the canonical coordinate theorems, and §9 sketches the application to L_1 dynamics and to the new, higher-accurate Hamiltonian balanced model for which $\gamma = i\sqrt{3}$ in (6). §10 shows how the present approach simplifies the formulation of variable-Coriolis-parameter models. §11 puts Hoskins' semigeostrophy itself, sometimes called 'geostrophic momentum theory', into the present perspective of inheritance from a parent dynamics. From this perspective semigeostrophy is doubly split, in a sense to be explained. There may be other such models with useful analytical simplicities, though keeping track of their accuracy is conceptually more complicated. The relation between material conservation of PV and symplectic-form invariance is noted, in the version relevant here. §12 gives concluding remarks and looks at some remaining challenges.

Some of the ideas and results developed here can be found in the independent work of Allen & Holm (1996), to which the reader is referred for an alternative view of problems of this general kind, based on Hamilton's principle and Kelvin's circulation theorem as contrasted with the present approach via symplectic and contact geometry. What we have called velocity splitting is implicit in their results, for instance in the Kelvin circulation theorems for the Hamiltonian balanced models they derive, corresponding to material conservation of PV as defined by (1) above. As is clear from their presentation, the velocity entering the integrands of the Kelvin circulation integrals differs from the velocity with which the contours of integration must move in order to conserve circulation. Translated into our notation, it is $\mathbf{u}^{\mathrm{C}}$ that enters the integrands but $\mathbf{u}^{\mathrm{P}}$ that moves the contours of integration.

2 Velocity splitting in the simplest possible example

The fundamental equation (3) for singly split models stems directly from the properties of Hamiltonian flow in phase space, and the properties of nondegenerate momentum–configuration constraints. This is so basic that it seems worth demonstrating first in a simpler, finite-dimensional case, the simplest possible 'toy problem'. The theory can be presented most succinctly in the modern, abstract mathematical language of differential forms, symplectic geometry and contact geometry, but for wider readability we mostly use the standard, older textbook notation, especially as this points most directly, and in a very simple way, to the infinite-dimensional fluid-dynamical cases of interest here.

The simplest example showing the basic points is that of particle motion in two dimensions under a potential $V(\mathbf{x})$, with Hamiltonian function $H(\mathbf{x},\mathbf{p}) = (2m)^{-1}p_ip_i + V(\mathbf{x})$, where $\mathbf{x} = \{x_i\}$ denotes the particle-position coordinates defining the system configuration, at a given instant t, and $\mathbf{p} = \{p_i\}$ denotes the corresponding set of canonical momenta. Most of what follows applies to an arbitrary though finite number of particles in two dimensions, but the simplest case of a single particle of mass m is sufficient for our immediate purposes. The phase space for the Hamiltonian description is then four-dimensional, being spanned by (x_1, x_2, p_1, p_2) where the canonical momenta (p_1, p_2) coincide, in this example, with the ordinary momenta $m\dot{x}_i, m\dot{x}_2$ because of the classical form of the Hamiltonian function. The four Hamilton's equations are

$$\dot{x}_i = \frac{\partial H}{\partial p_i} \;, \qquad \dot{p}_i = -\frac{\partial H}{\partial x_i} \qquad (i = 1, 2) \;; \tag{8}$$

and the first pair gives $p_i = m\dot{x}_i$.

Now an alternative way of expressing the same information is to introduce arbitrary variations dx_i, dp_i (often notated $\delta x_i, \delta p_i$ in textbooks) and to note that Hamilton's equations are equivalent to

$$\dot{x}_i dp_i - \dot{p}_i dx_i = \frac{\partial H}{\partial x_i}dx_i + \frac{\partial H}{\partial p_i}dp_i = dH \;, \quad \text{say} \tag{9}$$

where, here and subseqently, summation over repeated indices is understood. This has great advantages for our purposes, of which the most crucial is the simplicity and directness with which a new Hamiltonian dynamical system can be derived from (9) by applying a workless constraint. There is no need to introduce Lagrange multipliers, nor integrals with respect to time 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:

$$\begin{gathered}\textit{Apply the same constraint to the arbitrary variations } d\mathbf{x}, d\mathbf{p} \\ \textit{as is applied to the phase space flow } \dot{\mathbf{x}}, \dot{\mathbf{p}}.\end{gathered} \tag{10}$$

In the fluid case this will prove crucial to finding boundary conditions, as well as equations, that express Hamiltonian structure in an appropriately general way. Geometrically, this principle says that the dynamics is constrained to a single manifold $\mathcal{M}^{\mathrm{C}}$, as suggested in figure 1.

In the present, finite-dimensional example, (10) means that, if the constraint is expressed as a set of smooth functions $x_i = x^{\mathrm{C}}(\mathbf{z})$, $p_i = p_i^{\mathrm{C}}(\mathbf{z})$ where $\mathbf{z} = \{z_j\}$ denotes a smaller set of variables, which will here be taken to be just two variables, then in order to convert (9) into the new, constrained problem one simply substitutes the same functions $x_i^{\mathrm{C}}(\mathbf{z})$, $p_i^{\mathrm{C}}(\mathbf{z})$ into dx_i, dp_i as into $\dot{x}_i, \dot{p}_i$ and $H(\mathbf{x},\mathbf{p})$. Hamiltonian structure is then automatically preserved. We assume here that certain degeneracies are not introduced, specifically that the constraint functions $x_i^{\mathrm{C}}(\mathbf{z})$ and $p_i^{\mathrm{C}}(\mathbf{z})$ are such that a certain matrix Ω_{ij}^{C}, which arises from the substitutions, is invertible. Figure 1 and the equations to follow show what is involved.

First consider the expressions $p_i \dot{x}_i$ and $-p_i dx_i$. Constraining these expressions produces, respectively,

$$p_i \dot{x}_i = \theta_j^{\mathrm{C}}(\mathbf{z}) \dot{z}_j \ , \tag{11}$$

and

$$-p_i dx_i = -\theta_j^{\mathrm{C}}(\mathbf{z}) dz_j \ , \tag{12}$$

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

$$\theta_j^{\mathrm{C}}(\mathbf{z}) = p_i^{\mathrm{C}}(\mathbf{z}) \frac{\partial x_i^{\mathrm{C}}(\mathbf{z})}{\partial z_j} \ . \tag{13}$$

Adding the first variation of (11) to d/dt of (12) and noting that two terms $\pm p_i d\dot{x}_i$ cancel on the left, similarly $\pm \theta_j^{\mathrm{C}} d\dot{z}_j$ on the right, we see that the left hand side of (9) becomes

$$\begin{aligned} \dot{x}_i dp_i - \dot{p}_i dx_i &= \dot{z}_i d\theta_i^{\mathrm{C}} - \dot{\theta}_i^{\mathrm{C}} dz_i & (14) \\ &= \Omega_{ij}^{\mathrm{C}} \dot{z}_i dz_j \ , \text{ say} & (15) \end{aligned}$$

where

$$\Omega_{ij}^{\mathrm{C}}(\mathbf{z}) = \frac{\partial \theta_i^{\mathrm{C}}(\mathbf{z})}{\partial z_j} - \frac{\partial \theta_j^{\mathrm{C}}(\mathbf{z})}{\partial z_i} \ , \tag{16}$$

because

$$d\theta_j^{\mathrm{C}} = \frac{\partial \theta_j^{\mathrm{C}}}{\partial z_i} dz_i \qquad \text{and} \qquad \dot{\theta}_j = \frac{\partial \theta_j^{\mathrm{C}}}{\partial z_i} \dot{z}_i \ . \tag{17}$$

Note from the antisymmetry of Ω_{ij}^{C} and the symmetry of $\partial^2/\partial z_j \partial z_k$ that

$$\frac{\partial \Omega_{ij}^{\mathrm{C}}}{\partial z_k} + \frac{\partial \Omega_{jk}^{\mathrm{C}}}{\partial z_i} + \frac{\partial \Omega_{ki}^{\mathrm{C}}}{\partial z_j} = 0 \ . \tag{18}$$

From (9) and (14), then, writing $\widetilde{H}(\mathbf{z}) = H(\mathbf{x}^{\mathrm{C}}(\mathbf{z}), \mathbf{p}^{\mathrm{C}}(\mathbf{z}))$, we have

$$\Omega_{ij}^{\mathrm{C}} \dot{z}_i dz_j = d\widetilde{H} \ . \tag{19}$$

Removing the arbitrary factors dz_j, we see that the new problem that results from constraining (9) is

$$\Omega_{ij}^{\mathrm{C}} \ \dot{z}_i = \frac{\partial \widetilde{H}}{\partial z_j} \ . \tag{20}$$

Notice that energy conservation for the new, constrained problem follows at once from the antisymmetry of Ω_{ij}^{C}; replacing the remaining arbitrary variations in (19) by phase space flow rates, we have

$$\frac{d\widetilde{H}}{dt} = \Omega_{ij}^{\mathrm{C}} \dot{z}_i \dot{z}_j = 0 \ . \tag{21}$$

Nondegeneracy says that the matrix Ω_{ij}^{C} has an inverse J_{ij}^{C}, with $\Omega_{ij}^{\mathrm{C}} J_{jk}^{\mathrm{C}} = \delta_{ik}$, giving Hamilton's equations for the new, constrained problem in noncanonical form:

$$\dot{z}_i = J_{ji}^{\mathrm{C}} \frac{\partial \widetilde{H}}{\partial z_j} \ . \tag{22}$$

The corresponding Poisson bracket is

$$\{A(\mathbf{z}), B(\mathbf{z})\} = \frac{\partial A}{\partial z_i} J^{\mathrm{C}}_{ji} \frac{\partial B}{\partial z_j} \ , \tag{23}$$

satisfying Jacobi's identity

$$J^{\mathrm{C}}_{il} \frac{\partial J^{\mathrm{C}}_{jk}}{\partial z_l} + J^{\mathrm{C}}_{jl} \frac{\partial J^{\mathrm{C}}_{ki}}{\partial z_l} + J^{\mathrm{C}}_{kl} \frac{\partial J^{\mathrm{C}}_{ij}}{\partial z_l} = 0 \ , \tag{24}$$

whose equivalence to (18) for nonsingular Ω^{C}_{ij} and J^{C}_{ij} can be verified by multiplying (24) by $\Omega^{\mathrm{C}}_{im}\Omega^{\mathrm{C}}_{jn}\Omega^{\mathrm{C}}_{kp}$ then using $\partial/\partial z_l$ of $\Omega^{\mathrm{C}}_{ij} J^{\mathrm{C}}_{jk} = \delta_{ik}$. Each of the three versions, (19) or (20) with (18) or (22) with (24), are Hamiltonian systems by standard definitions (e.g. Salmon 1988*b*, Shepherd 1990). The matrix Ω^{C}_{ij} is sometimes called the symplectic matrix and J^{C}_{ij} the cosymplectic matrix.

In the language of differential forms, the right-hand sides of (14) and (15), i.e., the left-hand side of (19), can be regarded as the interior product $\dot{\mathbf{z}} \rfloor \Omega^{\mathrm{C}}$ of the vector field $\dot{\mathbf{z}}$, defining the phase space flow of the constrained problem, with the symplectic two-form $\Omega^{\mathrm{C}} = -d\theta^{\mathrm{C}}_j \wedge dz_j = +dz_j \wedge d\theta^{\mathrm{C}}_j = \frac{1}{2}\Omega^{\mathrm{C}}_{ij} dz_i \wedge dz_j$. This requires reinterpreting dz_i as the exterior derivative of the i^{th} coordinate field z_i. The simple way in which (11) and (12) lead to (16) then (18) corresponds to two successive exterior differentiations of the expression (12),

$$-\theta^{\mathrm{C}}_j dz_j = \theta^{\mathrm{C}}, \quad \text{say}, \tag{25}$$

regarded as a one-form. In the standard notation, $\Omega^{\mathrm{C}} = d\theta^{\mathrm{C}}$; $d\Omega^{\mathrm{C}} = dd\theta^{\mathrm{C}} = 0$, of which the last corresponds to (18) and is called closedness of Ω^{C}. Energy conservation (21) is $\dot{\mathbf{z}} \rfloor d\widetilde{H} = \dot{\mathbf{z}} \rfloor \dot{\mathbf{z}} \rfloor \Omega^{\mathrm{C}} = 0$.

Now the constraints of interest in this paper are always momentum–configuration constraints for which the dimension of $\mathbf{z}$ equals that of $\mathbf{x}$ (2 here, and ∞ in the fluid case), and

$$\mathbf{x}^{\mathrm{C}}(\mathbf{z}) = \mathbf{z} \ , \tag{26}$$

so that the parent *configuration* space or part of it becomes the *phase* space of the new, constrained problem. Then $\partial x^{\mathrm{C}}_i / \partial z_j = \delta_{ij}$, whence (13) gives

$$\theta^{\mathrm{C}}_j(\mathbf{z}) = p^{\mathrm{C}}_j(\mathbf{z}) = m u^{\mathrm{C}}_j(\mathbf{x}) \tag{27}$$

so that (14), (15) and (19) become

$$\dot{x}_j dp_j - \dot{p}_j dx_j = m(\dot{x}_j du^{\mathrm{C}}_j - \dot{u}^{\mathrm{C}}_j dx_j) = \Omega^{\mathrm{C}}_{ij} \dot{x}_i dx_j = d\widetilde{H} \tag{28}$$

(or $\dot{\mathbf{x}} \rfloor \Omega^{\mathrm{C}} = d\widetilde{H}$) with

$$\Omega^{\mathrm{C}}_{ij}(\mathbf{x}) = m \left(\frac{\partial u^{\mathrm{C}}_i}{\partial x_j} - \frac{\partial u^{\mathrm{C}}_j}{\partial x_i} \right) \ . \tag{29}$$

Figure 1 illustrates the associated geometric structure in a simple example; the caption provides further explanation. We now have

$$d\widetilde{H}(\mathbf{x}) \ = \ dH(\mathbf{x},\, m\mathbf{u}^{\mathrm{C}}(\mathbf{x})) \ = \ dV + m u^{\mathrm{C}}_k du^{\mathrm{C}}_k \ , \tag{30}$$

or equivalently

$$\frac{\partial \widetilde{H}(\mathbf{x})}{\partial x_j} = \left(\frac{\partial H(\mathbf{x},\mathbf{p})}{\partial x_j} + m \frac{\partial H(\mathbf{x},\mathbf{p})}{\partial p_k} \frac{\partial u_k^{\mathrm{C}}(\mathbf{x})}{\partial x_j} \right)_{\mathbf{p}\,=\,m\mathbf{u}^{\mathrm{C}}(\mathbf{x})} \;=\; \frac{\partial V}{\partial x_j} + m u_k^{\mathrm{C}} \frac{\partial u_k^{\mathrm{C}}}{\partial x_j} \;. \tag{31}$$

We can now derive corresponding equations for

$$\mathbf{u}^{\mathrm{S}} = \dot{\mathbf{x}} - \mathbf{u}^{\mathrm{C}}(\mathbf{x}) \;, \tag{32}$$

directly demonstrating velocity splitting in this simplest example. The shortest route uses (28) and (30):

$$\begin{aligned} 0 \;&=\; -m\left(\dot{x}_j du_j^{\mathrm{C}} - \dot{u}_j^{\mathrm{C}} dx_j\right) + dV + m u_j^{\mathrm{C}} du_j^{\mathrm{C}} \;\; (\,= -\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{C}} + d\widetilde{H}\,) &(33)\\ &=\; -m\left(u_j^{\mathrm{S}} du_j^{\mathrm{C}} - \dot{u}_i^{\mathrm{C}} dx_i\right) + dV &(34)\\ &=\; -m\left\{u_j^{\mathrm{S}} du_j^{\mathrm{C}} - (\dot{u}_i^{\mathrm{C}} - A_i^{\mathrm{C}}) dx_i\right\} + dV + m A_i^{\mathrm{C}} dx_i &(35) \end{aligned}$$

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

$$A_i^{\mathrm{C}} = u_j^{\mathrm{C}} \frac{\partial u_i^{\mathrm{C}}}{\partial x_j}, \tag{36}$$

which is the constraint acceleration in the sense of (2)ff., then

$$\dot{u}_i^{\mathrm{C}} - A_i^{\mathrm{C}} = \dot{x}_j \frac{\partial u_i^{\mathrm{C}}}{\partial x_j} - u_j^{\mathrm{C}} \frac{\partial u_i^{\mathrm{C}}}{\partial x_j} = u_j^{\mathrm{S}} \frac{\partial u_i^{\mathrm{C}}}{\partial x_j} \;. \tag{37}$$

Then (35) becomes simply

$$\Omega_{ij}^{\mathrm{C}} u_j^{\mathrm{S}} dx_i = R_i dx_i, \tag{38}$$

where

$$R_i = F_i - m A_i^{\mathrm{C}} = -\frac{\partial H}{\partial x_i} - m A_i^{\mathrm{C}} = -\frac{\partial V}{\partial x_i} - m A_i^{\mathrm{C}} \;, \tag{39}$$

implying that

$$\Omega_{ij}^{\mathrm{C}} u_j^{\mathrm{S}} = R_i \qquad \text{and} \qquad u_i^{\mathrm{S}} = J_{ij}^{\mathrm{C}} R_j \;. \tag{40}$$

Notice that the first of (40) is the discrete-particle analogue of equation (3) of §1. When $m = 1$, (39) corresponds to the definition (2); in the fluid case it is more convenient to define $\mathbf{R}$ and $\mathbf{F}$ as force per unit mass, corresponding to the most convenient particle labelling; see (45) below. The first term $-\partial H/\partial x_i = -\partial V/\partial x_i$ on the right of (39) likewise correponds to m times the first term $\mathbf{F}$ of (2), the particle acceleration that would occur if the constraint were suddenly removed.

We note in passing — though this will not be used in the sequel — that an alternative but equivalent view of the problem is easily obtained from 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 field $\boldsymbol{\lambda}$ of the momentum–configuration constraint $\mathbf{p} - \mathbf{p}^{\mathrm{C}} = 0$, where $\mathbf{p}^{\mathrm{C}} = m\mathbf{u}^{\mathrm{C}}$. Lagrange multipliers $\boldsymbol{\lambda}$ for such constraints are velocities because they must yield energies when multiplied by $(\mathbf{p} - \mathbf{p}^{\mathrm{C}})$ and summed over particles; in the modern geometric language, $\{\mathbf{p}^{\mathrm{C}}\}$ and $\{\boldsymbol{\lambda}\}$ belong respectively to the cotangent and tangent bundles of the configuration manifold $\{\mathbf{x}\}$. The standard recipe is to retain the Poisson bracket of the parent dynamics and to replace $H(\mathbf{x},\mathbf{p})$ by $\widehat{H} = H(\mathbf{x},\mathbf{p},\boldsymbol{\lambda}) + \lambda_j(\mathbf{x})\phi_j(\mathbf{x},\mathbf{p})$ where, in our version, $\phi_j = 0 (j = 1,2)$ is the constraint in standard notation, not to be confused with geopotential: $\phi_j = p_j - p_j^{\mathrm{C}}(\mathbf{x})$. (It is simplest to avoid Salmon's (1988*a*) Dirac-bracket notation because the Dirac bracket defined there is not manifestly bilinear in its two arguments, hence is not manifestly a bracket in the usual sense.) Then we get four equations

$$\dot{x}_i = \frac{\partial \widehat{H}}{\partial p_i}\ , \qquad \dot{p}_i = -\frac{\partial \widehat{H}}{\partial x_i} \tag{41}$$

which, together with $\phi_j = 0$ constitute the equations to be solved for the six unknowns $\dot{x}_i, \dot{p}_i$ and λ_i. The entire system of equations for $\mathbf{x}$, $\mathbf{p}$ and $\boldsymbol{\lambda}$ is not self-evidently Hamiltonian, but can readily be shown to be equivalent to (20) and hence is Hamiltonian by implication, in a noncanonical description. The first of (41) shows at once that (with $m = 1$)

$$\boldsymbol{\lambda} = \dot{\mathbf{x}} - \mathbf{p}^{\mathrm{C}} = \mathbf{u}^{\mathrm{S}}\ . \tag{42}$$

Again, if the constraint has any effect, then the Lagrange multipliers, i.e., $\mathbf{u}^{\mathrm{S}}$, must differ from zero. Using (42), the second of (41) can be shown in two or three lines of manipulation to be equivalent to (20), either directly or by using the facts that that $\{\phi_j, \widehat{H}\} = 0$ and $\{\phi_j, \phi_j\} = 0$ where $\{.\ ,\ .\}$ is the canonical Poisson bracket of the parent dynamics.

In this scheme it is essential to interpret all partial differentiations as being taken in the full parent phase space 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 $\phi_j = 0$.) Once again, no Dirac–Bergmann consistency conditions enter the picture, because the problem is one of changing the dynamical system by imposing a nondegenerate constraint, requiring (29) to be nonsingular — not the subtler problem of finding a pre-existing constraint or Lagrangian submanifold implicit in a degenerate parent dynamics.

Notice finally that the variational forms (28) and (38) can easily be adapted to dynamical models that include dissipation or forcing terms. Such terms can be added on the right of (8) and carried through the analysis. Thus there is no difficulty with the formalism, as presented here, in modifying Hamiltonian dynamics by adding forcing and dissipation. Of course the extra terms can interfere with balance, and the possible accuracy of the resulting balanced model is a separate question.

3 Extension to fluid dynamics

The simplest fluid-dynamical system for which the concept of balance is nontrivial is shallow water dynamics. Here the mass configuration is completely described, in a way that is invariant to particle relabellings, by the layer depth $h(\mathbf{x})$ as a function of position $\mathbf{x}$. References to the dependence on time t are suppressed. It proves overwhelmingly advantageous to use both the Eulerian description, as with $h = h(\mathbf{x})$, and also the Lagrangian or particle-following description of the fluid motion. Use of the Lagrangian description allows the symplectic structure on the slow manifold $\mathcal{M}^{\mathrm{C}}$ to be represented in a simple way that precisely parallels equations (16)–(25) of §2. Fields like $\mathbf{u}^{\mathrm{C}}$, $\mathbf{A}^{\mathrm{C}}$, $\mathbf{F}$ and $\mathbf{R}$, on the other hand, are simplest to handle in the Eulerian description, because this is by far the simplest way of representing the pressure-gradient and other forces associated with the mass configuration and the corresponding potential-energy variations in the Hamiltonian functional, as will be illustrated in (70) below. Use of the Eulerian description is also the simplest way to guarantee that the particle-relabelling symmetry is respected. If a $\mathbf{u}^{\mathrm{C}}$ functional did not respect the particle-relabelling symmetry, then the corresponding balanced model would violate material PV conservation as well as the particle-relabelling symmetry itself. The Eulerian description is needed, furthermore, to make explicit the precise sense in which the fundamental equation (3) is reference-frame independent, thereby simplifying, for instance, routes to variable-Coriolis-parameter models.

We therefore use both descriptions, Eulerian and Lagrangian, together. This requires continual use of the configuration mapping $\mathbf{a} \leftrightarrow \mathbf{x}$ between present positions $\mathbf{x}$ and the reference positions $\mathbf{a}$ that label material fluid elements. The mapping is assumed to be well behaved, meaning invertible and sufficiently differentiable. Then, given the configuration mapping, any function of $\mathbf{a}$ can equally well be regarded as a function of $\mathbf{x}$, and vice versa. Thus the field $h(\mathbf{x})$ can equally well be regarded as a function of $\mathbf{a}$; and to reduce symbolic proliferation we adopt the notational convention that the same symbol may be used to denote both functions, so that the equality of values when $\mathbf{x}$ corresponds to $\mathbf{a}$ is written as $h(\mathbf{x}) = h(\mathbf{a})$. (Note that this is shorthand for what could have been written $h(\mathbf{x}(\mathbf{a})) = \mathrm{h}(\mathbf{a})$ by a different, 'fixed-slot' convention.) Variations $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ in the configuration mapping, naturally regarded as functions of $\mathbf{a}$ when using the Lagrangian description, can also, in the same way, be regarded as functions of $\mathbf{x}$ and written as $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{x})$. Again,

$$\Omega^{\mathrm{C}}_{ij}(\mathbf{x}, \mathbf{x}') = \Omega^{\mathrm{C}}_{ij}(\mathbf{a}, \mathbf{a}') \tag{43}$$

in (3), and so on. Arguments will be shown explicitly whenever there might be danger of confusion. The notation δ^{L} will always mean a Lagrangian variation, with the implication that $\mathbf{a}$ is held constant, as with $\delta^{\mathrm{L}}\mathbf{x}$; and δ^{E} will always mean an Eulerian variation, with $\mathbf{x}$ held constant. The distinction is crucial when dealing with functionals like $\mathbf{u}^{\mathrm{C}}$ that are also fields. Variations of ordinary functionals like the Hamiltonian, with no dependence on $\mathbf{a}$ or $\mathbf{x}$, will notated δ.

For a rigid side boundary $\partial\mathcal{D}$,

$$\delta^{\mathrm{L}}\mathbf{x} \,.\, \hat{\mathbf{n}} = 0 \quad \text{on} \quad \partial\mathcal{D} \ , \tag{44}$$

where $\hat{\mathbf{n}}$ is the outward unit normal. The mass element

$$\mathrm{d}m = h\mathrm{d}\mathbf{x} = \mathrm{d}\mathbf{a} \ , \tag{45}$$

if, as we prescribe, the reference positions are actual positions in a notional initial layer of uniform depth 1 unit, and density 1 unit, in appropriate units. Therefore, with (x_1, x_2) replacing (x, y) from now on,

$$h = \frac{\partial(a_1, a_2)}{\partial(x_1, x_2)} = \left(\frac{\partial(x_1, x_2)}{\partial(a_1, a_2)}\right)^{-1} . \tag{46}$$

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 mapping variations expressed as $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{x})$ by

$$\delta^{\mathrm{L}}h(\mathbf{x}) = \delta^{\mathrm{L}}\mathbf{x}.\nabla h + \delta^{\mathrm{E}}h = -h\nabla.(\delta^{\mathrm{L}}\mathbf{x}) , \tag{47}$$

$$\delta^{\mathrm{E}}h(\mathbf{x}) = -\nabla.(h\delta^{\mathrm{L}}\mathbf{x}) , \tag{48}$$

consistently with the general relation

$$\delta^{\mathrm{L}} = \delta^{\mathrm{L}}\mathbf{x}.\nabla + \delta^{\mathrm{E}} . \tag{49}$$

This last is applicable to any function of $\mathbf{x}$. Observe that the notation $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{x})$ is consistent with (49), because $\delta^{\mathrm{E}}\mathbf{x}$ is trivially zero. Observe also that δ^{E} commutes with ∇ or $\partial/\partial x_i$ — an important simplifying feature — whereas δ^{L} does not.

In order for the $\mathbf{u}^{\mathrm{C}}$ functional to respect the particle-relabelling symmetry it must be expressible in Eulerian form: $\mathbf{u}^{\mathrm{C}} = \mathbf{u}^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot)) = \mathbf{u}^{\mathrm{C}}(\mathbf{x}; h(\cdot))$. This says that the value of $\mathbf{u}^{\mathrm{C}}$ at any position $\mathbf{x}$ is nonlocally dependent on the mass configuration $h(\mathbf{x})$, but independent of where individual fluid elements are located within that mass distribution. The degree of nonlocality depends, as already mentioned, on the accuracy desired. From (47) and (48) it is straightforward to show that the relevant Lagrangian and Eulerian functional derivatives are related by the appropriate form of the chain rule,

$$\frac{\delta^{\mathrm{L}}u_i^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}}x_j(\mathbf{a}')} = \frac{1}{h(\mathbf{x})}\,\frac{\partial u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\partial x_j}\,\delta(\mathbf{x}-\mathbf{x}') + \frac{\partial}{\partial x'_j}\frac{\delta^{\mathrm{E}}u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}}h(\mathbf{x}')} , \tag{50}$$

where δ with argument $(\mathbf{x}-\mathbf{x}')$ denotes the two-dimensional Dirac delta function. In the first term on the right, $\partial/\partial x_j$ connotes that the function argument $h(\cdot)$ is unvaried. In the last term, $\delta^{\mathrm{E}}/\delta^{\mathrm{E}}h$ connotes that the position argument $\mathbf{x}$ is unvaried. 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, signals which dummy variable of integration to use when computing variations. Thus by definition, for sufficiently smooth $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$ and $\delta^{\mathrm{E}}h(\mathbf{x})$,

$$\delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}(\mathbf{a}) = \delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot)) = \int_{\mathcal{D}}\frac{\delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}}x_j(\mathbf{a}')}\delta^{\mathrm{L}}x_j(\mathbf{a}')\,\mathrm{d}\mathbf{a}' \tag{51}$$

and

$$\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}(\mathbf{x}) = \delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}(\mathbf{x};h(\cdot)) = \int_{\mathcal{D}}\frac{\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}}h(\mathbf{x}')}\delta^{\mathrm{E}}h(\mathbf{x}')\,\mathrm{d}\mathbf{x}' \tag{52}$$

where, as in §2, summation over repeated indices is understood. Note that $\int \mathrm{d}\mathbf{a}'\delta^{\mathrm{L}}/\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a}')$ and $\int \mathrm{d}\mathbf{x}'\delta^{\mathrm{E}}/\delta^{\mathrm{E}}h(\mathbf{x}')$ both have dimensions $(\text{length})^{-1}$, the same as $\mathbf{x}^{-1}$ and h^{-1}, and, for instance, from (46),

$$\frac{\delta^{\mathrm{L}}}{\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a}')} = \frac{1}{h(\mathbf{x}')}\frac{\delta^{\mathrm{L}}}{\delta^{\mathrm{L}}\mathbf{x}(\mathbf{x}')} , \tag{53}$$

consistent with (45). To verify (50), substitute it into (51), change variable of integration to $\mathbf{x}'$ using $\mathrm{d}\mathbf{a}' = h(\mathbf{x}')\mathrm{d}\mathbf{x}'$, integrate the δ^{E} term by parts, recalling (44), and use (48), (52) and (49). Note that the relation (50) is similar to, though simpler than, the relations needed to derive the Eulerian (Morrison & Greene 1980) formula for the parent Poisson bracket from the canonical (Lagrangian) formula.

Equation (3) now follows simply by noting that the infinite-dimensional counterpart of (29) is

$$\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}') = \Omega^{\mathrm{C}}_{ij}(\mathbf{a},\mathbf{a}') = \frac{\delta^{\mathrm{L}} u^{\mathrm{C}}_i(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')} - \frac{\delta^{\mathrm{L}} u^{\mathrm{C}}_j(\mathbf{a}';\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_i(\mathbf{a})} \;, \tag{54}$$

and that the parent Hamiltonian functional is

$$\mathsf{H} = \mathsf{V} + \int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}|^2 \, \mathrm{d}m \;, \tag{55}$$

where V is the potential energy of the mass configuration, so that

$$\widetilde{\mathsf{H}} = \mathsf{V} + \int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}^{\mathrm{C}}|^2 \, \mathrm{d}m \;; \tag{56}$$

$\widetilde{\mathsf{H}}$ and V are scalar-valued functionals of the mass configuration alone. The finite-dimensional relation (30) has the counterpart

$$\delta\widetilde{\mathsf{H}}(\mathbf{x}) \;=\; \delta\mathsf{V} + \delta\int_{\mathcal{D}} \tfrac{1}{2}|\mathbf{u}^{\mathrm{C}}|^2 \mathrm{d}m \;=\; \delta\mathsf{V} + \int_{\mathcal{D}} u^{\mathrm{C}}_k \delta^{\mathrm{L}} u^{\mathrm{C}}_k \mathrm{d}m \;, \tag{57}$$

now integrating over mass elements $\mathrm{d}m$ as well as summing the index k from 1 to 2. Note that δ^{L}, by its definition, commutes with $\int ...\mathrm{d}m$, i.e. with $\int ...\mathrm{d}\mathbf{a}$.

It remains only to carry out the steps that parallel (33) - (35), again with d replaced by δ or δ^{L} as appropriate, and applying integration over $\mathcal{D}$ with respect to the mass element $\mathrm{d}m$ as well as index summation from 1 to 2. The definition of $\mathbf{R}$ is still $\mathbf{R} = \mathbf{F} - \mathbf{A}^{\mathrm{C}}$ but now with

$$F_i(\mathbf{x}) = -\frac{\delta\mathsf{V}}{\delta^{\mathrm{L}} x_i(\mathbf{a})} = -\frac{\partial}{\partial x_i}\frac{\delta\mathsf{V}}{\delta^{\mathrm{E}} h(\mathbf{x})} \;, \tag{58}$$

and

$$\begin{aligned} A^{\mathrm{C}}_i(\mathbf{x}) \;&=\; \int_{\mathcal{D}} u^{\mathrm{C}}_j(\mathbf{a}';\mathbf{x}(\cdot)) \frac{\delta^{\mathrm{L}} u^{\mathrm{C}}_i(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')} \, \mathrm{d}\mathbf{a}' \\ &=\; u^{\mathrm{C}}_j \frac{\partial u^{\mathrm{C}}_i}{\partial x_j} - \int_{\mathcal{D}} \frac{\partial\{h(\mathbf{x}')u^{\mathrm{C}}_j(\mathbf{x}';h(\cdot))\}}{\partial x'_j} \frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_i(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')} \mathrm{d}\mathbf{x}' \;, \end{aligned} \tag{59}$$

using (50) and (44) to rewrite $\mathbf{A}^{\mathrm{C}}$ in terms of Eulerian functional derivatives. Similarly rewriting (54) we get finally

$$\boxed{\int_{\mathcal{D}} \Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}') u^{\mathrm{S}}_j(\mathbf{x}') \, \mathrm{d}m(\mathbf{x}') = (\boldsymbol{\zeta}^{\mathrm{C}} \times \mathbf{u}^{\mathrm{S}})_i + \int_{\mathcal{D}} \omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}') u^{\mathrm{S}}_j(\mathbf{x}') \mathrm{d}m(\mathbf{x}') \;=\; R_i(\mathbf{x}), \qquad (60)}$$

where

$$\omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}') = -\omega^{\mathrm{C}}_{ji}(\mathbf{x}',\mathbf{x}) = \frac{\partial}{\partial x'_j}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_i(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')} - \frac{\partial}{\partial x_i}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_j(\mathbf{x}';h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x})}\ . \tag{61}$$

Note that the $(\boldsymbol{\zeta}^{\mathrm{C}} \times \mathbf{u}^{\mathrm{S}})$ term comes from the first term in (50), when substituted into the antisymmetric expression (54). It is crucial to remember here that both $\Omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')$ and $\mathbf{A}^{\mathrm{C}}$ are absolute (inertial-frame) quantities: absolute $\mathbf{u}^{\mathrm{C}}$ values must be used in (54) and (59). Thus, for instance, $\mathbf{A}^{\mathrm{C}}$ includes what we normally call the Coriolis and centrifugal accelerations; this point is further discussed in §10 below.

For the general three-dimensional stratified, compressible case, with $\mathbf{u}^{\mathrm{C}}(\mathbf{x};h(\cdot))$ replaced by $\mathbf{u}^{\mathrm{C}}(\mathbf{x};\rho(\cdot),\sigma(\cdot))$, a precisely similar calculation, whose details will be omitted for brevity, produces (60) with

$$\begin{aligned}\omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}') = -\omega^{\mathrm{C}}_{ji}(\mathbf{x}',\mathbf{x}) &= \frac{\partial}{\partial x'_j}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_i(\mathbf{x};\rho(\cdot),\sigma(\cdot))}{\delta^{\mathrm{E}} \rho(\mathbf{x}')} - \frac{\partial}{\partial x_i}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_j(\mathbf{x}';\rho(\cdot),\sigma(\cdot))}{\delta^{\mathrm{E}} \rho(\mathbf{x})}\\ &+ \ \frac{\sigma(\mathbf{x}')}{\rho(\mathbf{x}')}\frac{\partial}{\partial x'_j}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_i(\mathbf{x};\rho(\cdot),\sigma(\cdot))}{\delta^{\mathrm{E}} \sigma(\mathbf{x}')} - \frac{\sigma(\mathbf{x})}{\rho(\mathbf{x})}\frac{\partial}{\partial x_i}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_j(\mathbf{x}';\rho(\cdot),\sigma(\cdot))}{\delta^{\mathrm{E}} \sigma(\mathbf{x})}\ .\end{aligned} \tag{62}$$

In brief, we replace $h(\mathbf{x})$ by $\rho(\mathbf{x})$ in (48), giving $dm = \rho \mathrm{d}\mathbf{x} = \mathrm{d}\mathbf{a}$ and

$$\delta^{\mathrm{E}}\rho(\mathbf{x}) = -\nabla\,.\,(\rho\delta^{\mathrm{L}}\mathbf{x}) \tag{63}$$

with ∇ now three-dimensional; and we use also

$$\delta^{\mathrm{E}}\sigma(\mathbf{x}) = -\nabla\,.\,(\sigma\delta^{\mathrm{L}}\mathbf{x}) \tag{64}$$

and the following counterpart of (50)

$$\begin{aligned}\frac{\delta^{\mathrm{L}} u^{\mathrm{C}}_i(\mathbf{a};\mathbf{x}(\cdot))}{\delta^{\mathrm{L}} x_j(\mathbf{a}')} &= \frac{1}{\rho(\mathbf{x})}\ \frac{\partial u^{\mathrm{C}}_i(\mathbf{x};\rho(\cdot);\sigma(\cdot))}{\partial x_j}\,\delta(\mathbf{x}-\mathbf{x}')\\ &+ \frac{\partial}{\partial x'_j}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_i(\mathbf{x};\rho(\cdot);\sigma(\cdot))}{\delta^{\mathrm{E}} \rho(\mathbf{x}')} + \frac{\sigma(\mathbf{x}')}{\rho(\mathbf{x}')}\frac{\partial}{\partial x'_j}\frac{\delta^{\mathrm{E}} u^{\mathrm{C}}_i(\mathbf{x};\rho(\cdot);\sigma(\cdot))}{\delta^{\mathrm{E}} \sigma(\mathbf{x}')}\ .\end{aligned} \tag{65}$$

We may speak of a 'well behaved' Hamiltonian balanced model whenever the foregoing operations make mathematical sense. More precisely:

Definition *The Hamiltonian balanced model defined by a given, functionally differentiable $\mathbf{u}^{\mathrm{C}}$ will be called 'well behaved' for some given set of mass configurations if (a) the constraint vorticity $\boldsymbol{\zeta}^{\mathrm{C}}$ is finite, (b) the $\mathbf{u}^{\mathrm{P}}$ and $\mathbf{u}^{\mathrm{S}}$ fields exist and are smooth enough for the integral $\int_{\mathcal{D}}\omega^{\mathrm{C}}_{ij}(\mathbf{x},\mathbf{x}')u^{\mathrm{S}}_j(\mathbf{x}')\mathrm{d}m(\mathbf{x}')$ to exist, and (c) the ω^{C}_{ij} fields implied by the given $\mathbf{u}^{\mathrm{C}}$ functional are such that the integral vanishes whenever $\mathbf{u}^{\mathrm{S}}(\mathbf{x}) = 0$ everywhere in the physical domain $\mathcal{D}$.*

We can now assert, as an immediate consequence of (60) by reductio ad absurdum:

Theorem 1 (Velocity-splitting theorem) *For any well behaved Hamiltonian balanced model as just defined, every mass configuration whose $\mathbf{R}$ field is not everywhere zero in $\mathcal{D}$ has a $\mathbf{u}^{\mathrm{S}}$ field that is not everywhere zero in $\mathcal{D}$.*

Remarks. Note further that:

1. Equation (60) is exact in the sense that it follows without approximation as soon as the parent dynamics and the $\mathbf{u}^{\rm C}$ functional are specified. The accuracy of $\mathbf{u}^{\rm C}$ itself, as measured by the smallness of norms $||\mathbf{R}||$ or $||\mathbf{u}^{\rm S}||$, or closeness to the parent dynamics, is a separate question.

2. The antisymmetric kernel $\omega_{ij}^{\rm C}(\mathbf{x},\mathbf{x}')$ arises solely from the way in which mass rearrangement affects the $\mathbf{u}^{\rm C}$ functional, hence its definition solely in terms of Eulerian functional derivatives with respect to the mass configuration.

3. As already mentioned, the functional dependence of $\mathbf{u}^{\rm C}$ upon the mass configuration $h(\mathbf{x})$ or $\rho(\mathbf{x})$, $\sigma(\mathbf{x})$, notated $h(\cdot)$ or $\rho(\cdot), \sigma(\cdot)$, can be fully nonlocal, as is necessary for the highest accuracies.

4. The dynamical effects of rotation enter solely through the $\mathbf{u}^{\rm C}$ functional or functionals, for instance through $\boldsymbol{\zeta}^{\rm C}$ and through the way in which the $\mathbf{u}^{\rm C}$ field changes when, for instance, the pressure field changes. This will prove very convenient when, for instance, it comes to applying (60) to variable-Coriolis-parameter models (§10). It gives a natural way of distinguishing 'centrifugal' and 'Coriolis' accelerations from 'relative' accelerations that is based entirely on the $\mathbf{u}^{\rm C}$ functional and the constraint accelerations $\mathbf{A}^{\rm C}$ derived from it.

5. The constrained Hamiltonian functional $\widetilde{\mathsf{H}}(h(\cdot))$ or $\widetilde{\mathsf{H}}(\rho(\cdot),\sigma(\cdot))$ enters the foregoing derivation very simply, owing to the use of the Lagrangian description in treating the kinetic energy term, as in the second term of (57). Some more details are given in (82)ff. The potential energy term, $\delta\mathsf{V}$ in (57), does not need to be computed here because it enters solely through its relation to the parent force $\mathbf{F} = \delta\mathsf{V}/\delta^{\rm L}\mathbf{x}$ per unit mass, which can usually be written down from elementary knowledge of the parent dynamics (cf., e.g., (71)).

6. When $\mathbf{F}$ does need to be computed from V, this is simplest via the Eulerian description, as will be illustrated in (69)ff.

7. Both PV and energy conservation follow by standard arguments, provided that the PV is given by (1) and the energy by $\widetilde{H}$, as in (21), and provided also that, in the case of energy conservation in bounded domains, the information about boundary conditions implicit in (60) is used. Details for energy are given in (82)ff, in connection with the general procedure for obtaining boundary conditions. Details for PV are given at the end of §11.

4 Equations of L_0 dynamics and L_1 dynamics

We now show that Salmon's equations for f-plane L_0 dynamics and L_1 dynamics, for f-plane shallow water parent dynamics under constant gravity g, follow very directly from the result just obtained. These are the simplest examples of Hamiltonian balanced models of the kind under discussion. For L_0 dynamics we simply take

$$\mathbf{u}^{\rm C} = \tfrac{1}{2} f \hat{\mathbf{z}} \times \mathbf{x} \ , \tag{66}$$

describing solid rotation. Then the constraint vorticity becomes simply the Coriolis parameter of that solid rotation, $\zeta^{\rm C} = f =$ constant. The right-hand side of (61) vanishes, because $\mathbf{u}^{\rm C}$ does not depend on the mass configuration, so that the $\omega_{ij}^{\rm C}$ terms in (60) vanish. The residual unbalanced force $\mathbf{R}$ 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 (60) becomes simply the geostrophic relation, $\mathbf{u}^{\rm S} = \mathbf{u}^{\rm G(rel)}$, with $\mathbf{u}^{\rm G(rel)}$ the relative geostrophic velocity given in terms of the mass configuration by (5), wherein $\phi = \phi^{\rm abs} - \phi^{\rm centrif}$ with $\phi^{\rm abs}$ defined as g times the actual surface elevation, and $\phi^{\rm centrif}$ as g times the paraboloidal surface elevation for the solid rotation (66).

Although (66) does not yet give a balanced model capable of dynamical evolution, it does illustrate the tendency of $\mathbf{u}^{\rm P} = \mathbf{u}^{\rm C} + \mathbf{u}^{\rm S}$ to be an improved approximation to the balanced motion in comparison with $\mathbf{u}^{\rm C}$.

For L_1 dynamics we take, iteratively,

$$\mathbf{u}^{\rm C} = \tfrac{1}{2} f \hat{\mathbf{z}} \times \mathbf{x} + \mathbf{u}^{\rm G(rel)} \ . \tag{67}$$

The constraint vorticity $\zeta^{\rm C}$ in (60) now becomes the geostrophic absolute vorticity $\zeta^{\rm C} = f + \hat{\mathbf{z}}\,.\nabla\times\mathbf{u}^{\rm G(rel)} = \zeta^{\rm G}$, say. The corresponding materially conserved potential vorticity (§11),

$$Q^{\rm G} = h^{-1}\zeta^{\rm G} \ , \tag{68}$$

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

$$\mathsf{V} = \int_{\mathcal{D}} \left(\phi^{\rm abs} - \tfrac{1}{2} g h\right)\, {\rm d}m = \int_{\mathcal{D}} \left(\phi^{\rm abs} h - \tfrac{1}{2} g h^2\right)\, {\rm d}\mathbf{x} \ , \tag{69}$$

implying that

$$\delta\mathsf{V} = \int_{\mathcal{D}} \left(h\delta^{\rm E}\phi^{\rm abs} + \phi^{\rm abs}\delta^{\rm E} h - g h \delta^{\rm E} h\right)\, {\rm d}\mathbf{x} = \int_{\mathcal{D}} \phi^{\rm abs}\delta^{\rm E} h\, {\rm d}\mathbf{x} \tag{70}$$

(because with any fixed topography Eulerian variation is simplest, $\delta^{\rm E}\phi^{\rm abs} = g\delta^{\rm E} h$), whence, by (44), (45), (48) and (58),

$$\delta\mathsf{V} = \int_{\mathcal{D}} \delta^{\rm L}\mathbf{x}(\mathbf{x}).\nabla\phi^{\rm abs}\, {\rm d}m \ ; \quad \Rightarrow \quad \mathbf{F} = -\nabla\phi^{\rm abs} \ . \tag{71}$$

Note incidentally — with an eye to §10 and to Remark **4** of §3 — 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 ϕ, not $\phi^{\rm abs}$ (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) $\mathbf{u}^{\rm C}$ values in H, one must also use absolute $\mathbf{u}^{\rm P}$ values in place of $\dot{\mathbf{x}}$ in the fluid counterpart of $-\dot{\mathbf{x}}\rfloor\Omega^{\rm C} + d\widetilde{H} = 0$ (more detail in (82) below), and allow for any work done by moving boundaries. Cancellations then lead to the standard rotating-frame description.

We now restrict attention, for illustrative purposes, to the simplest possible case. This is the 'no-topography' case in which the bottom of the shallow water layer coincides with a centrifugal–gravitational level surface of the basic solid rotation associated with f. Then a short calculation from (59) and (71) gives

$$\mathbf{R} = \mathbf{F} - \mathbf{A}^{\mathrm{C}} = -\mathbf{u}^{\mathrm{G(rel)}}.\nabla\mathbf{u}^{\mathrm{G(rel)}} \ . \tag{72}$$

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

$$\nabla.\mathbf{u}^{\mathrm{G(rel)}} \ = \ 0 \ = \ \nabla.(h\mathbf{u}^{\mathrm{G(rel)}}) \ , \tag{73}$$

annihilating the relative part of the last term in (59). If particles were to move with absolute velocity $\mathbf{u}^{\mathrm{C}}$, i.e. with relative velocity $\mathbf{u}^{\mathrm{G(rel)}}$, then $\partial h/\partial t$ evaluated in the rotating frame would be exactly zero. So in this example there is no contribution to $\mathbf{R}$ from the functional dependence of $\mathbf{u}^{\mathrm{G(rel)}}$ 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 be dispensed with, for instance with the help of (122) below).

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

$$\frac{\delta^{\mathrm{E}}u_i^{\mathrm{C}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}}h(\mathbf{x}')} = \frac{\delta^{\mathrm{E}}u_i^{\mathrm{G(rel)}}(\mathbf{x};h(\cdot))}{\delta^{\mathrm{E}}h(\mathbf{x}')} = \frac{g}{f}\varepsilon_{ik}\frac{\partial}{\partial x_k'}\delta(\mathbf{x}-\mathbf{x}') \ , \tag{74}$$

where ε_{ik} is the two-dimensional alternating tensor ($\varepsilon_{12} = -\varepsilon_{21} = 1$, $\varepsilon_{11} = \varepsilon_{22} = 0$). Thus, from (61), for any $\mathbf{x}$ interior to the fluid domain $\mathcal{D}$,

$$\begin{aligned}
&\int_{\mathcal{D}} \omega_{ij}^{\mathrm{C}}(\mathbf{x},\mathbf{x}')u_j^{\mathrm{S}}(\mathbf{x}')\mathrm{d}m(\mathbf{x}') = \\
&= \ \frac{g}{f}\int_{\mathcal{D}}\left[\left\{\varepsilon_{ik}\frac{\partial^2}{\partial x_j'\partial x_k'} - \varepsilon_{jk}\frac{\partial^2}{\partial x_i'\partial x_k'}\right\}\delta(\mathbf{x}-\mathbf{x}')\right] u_j^{\mathrm{S}}(\mathbf{x}')\ h(\mathbf{x}')\ \mathrm{d}\mathbf{x}' \\
&= \ \frac{g}{f}\int_{\mathcal{D}}\left[\varepsilon_{ij}\frac{\partial^2}{\partial x_k'\partial x_k'}\delta(\mathbf{x}-\mathbf{x}')\right] u_j^{\mathrm{S}}(\mathbf{x}')\ h(\mathbf{x}')\ \mathrm{d}\mathbf{x}' \ = \ \frac{g}{f}\varepsilon_{ij}\nabla^2 U_j^{\mathrm{S}} \ ,
\end{aligned} \tag{75}$$

where

$$\mathbf{U}^{\mathrm{S}}(\mathbf{x}) = h(\mathbf{x})\mathbf{u}^{\mathrm{S}}(\mathbf{x}) \ . \tag{76}$$

To verify this, replace u_j^{S} in the second line by $u_p^{\mathrm{S}}\delta_{pj} = -u_p^{\mathrm{S}}\varepsilon_{pq}\varepsilon_{qj}$ 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 (60) as $-\varepsilon_{ij}Q^{\mathrm{C}}U_j^{\mathrm{S}}$, we see that (60) reduces to

$$\left(\nabla^2 - K^2(\mathbf{x})\right)\mathbf{U}^{\mathrm{S}} = -\frac{f}{g}\hat{\mathbf{z}}\times\left(\mathbf{u}^{\mathrm{G(rel)}}.\nabla\mathbf{u}^{\mathrm{G(rel)}}\right) \ , \tag{77}$$

for any $\mathbf{x}$ within the domain $\mathcal{D}$, where $K^2 = fQ^{\mathrm{C}}/g$, the appropriate inverse square Rossby length, which is positive under the parameter conditions favouring balance. (In the three-dimensional stratified case, negative K^2 would correspond to an inertially unstable state, with any initial balance breaking down spontaneously.) The elliptic equation (77) is Salmon's (1985) equation, and is robustly invertible whenever $f\zeta^{\mathrm{C}}$ hence fQ^{C} are everywhere positive, for an unbounded domain $\mathcal{D}$ with evanescent boundary conditions.

Again 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 (77) 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 of Rossby number Ro smaller than the relative constraint velocity, $\mathbf{u}^{\mathrm{C(rel)}}$ say, $= \mathbf{u}^{\mathrm{G(rel)}}$ in this case. A similar argument applies to the relative magnitudes of $\mathbf{u}^{\mathrm{S}}$ and $\mathbf{u}^{\mathrm{C(rel)}}$ in the *general* case governed by (60), where the corresponding comparison is between $\mathbf{R}$ on the right and $\boldsymbol{\zeta}^{\mathrm{C}} \times \mathbf{u}^{\mathrm{S}}$ on the left.

5 Boundary conditions and energy conservation for rigidly bounded domains

Contrary to appearances, the fundamental equation (3) or (60), with (62) or (61), contains essential information about boundary conditions and their role in energy conservation. This involves some mathematical subtlety. We illustrate this by discussing the rigidly bounded shallow water problem, with points of greater generality indicated where appropriate. In what follows, the side boundary $\partial\mathcal{D}$ can have one or more connected parts $\partial\mathcal{D}_i$ ($\partial\mathcal{D} = \bigcup_i \partial\mathcal{D}_i$) so, that, for example, an annular domain $\mathcal{D}$ is possible. Rigidity is understood to include impermeability: no fluid particles may cross the boundary, implying that, in a reference frame co-rotating with $\partial\mathcal{D}$, the particle velocity relative to that frame, $\mathbf{u}^{\mathrm{P(rel)}}$ say, must satisfy

$$\mathbf{u}^{\mathrm{P(rel)}}.\,\hat{\mathbf{n}} = 0 \quad \text{on} \quad \partial\mathcal{D} \tag{78}$$

as well as its variational counterpart (44). This is part of the definition of the configuration space for the parent and balanced dynamics.

It turns out — and this seems to be another new result — that there are *two* natural choices of boundary conditions for rigidly bounded L_1 dynamics, each resulting in a well determined problem, and each conforming to (78) and (44). If we make one of these two choices, Kelvin waves are filtered, and if we make the other they are not.

Why should there be any freedom to choose? This is a general point. One might think that there is no reasonable alternative to defining the configuration space of the balanced model to be the same as that of the parent dynamics, then looking to the balanced dynamics to determine the remaining boundary conditions. But there is freedom in the sense that one can make the configuration space smaller by, in effect, imposing a more restrictive constraint. That such a thing is possible in principle is clear from the toy example of figure 1 (q.v.). Nondegeneracy in that example requires that the $\mathcal{M}^{\mathrm{C}}$ surface is nowhere 'horizontal'; but there is no reason why some part of it should not tend to the 'vertical'. It will still intersect the symplectic hypertubes. One could imagine, for instance, p_2 asymptoting to $-\infty$ on $\mathcal{M}^{\mathrm{C}}$ as x_1 increases to some finite value beyond the range of values depicted in the figure. The resulting dynamics plainly has a configuration space that is smaller than the configuration space for the parent dynamics.

Consider, now, the case in which the parent is shallow water dynamics. The parent configuration space is a set of sufficiently smooth functions $\{h(\mathbf{x})\}$ such that total mass is constant,

$$\int_{\mathcal{D}} h\,\mathrm{d}\mathbf{x} = \text{constant} \ . \tag{79}$$

Configuration changes must conform to (78) and (44)), but there are no boundary restrictions on the value of h. Note incidentally that, because of (79), configurations are rays, not points, in the *function* space $\{h(\mathbf{x})\}$, as with quantum-mechanical Hilbert spaces.

We can, as already suggested, define a balanced model with the same configuration space as the parent. For L_1 dynamics this is the choice that does not filter Kelvin waves (details in §6). But we can also define the balanced model to have the slightly smaller configuration space in which, in additional to the foregoing, the condition

$$\partial\phi/\partial\ell = 0 \quad \text{on} \quad \partial\mathcal{D} \tag{80}$$

is imposed (details in §7), where ℓ is a tangential coordinate along $\partial\mathcal{D}$, such as arc length, and ϕ is defined in §4. This is the choice that filters Kelvin waves, evidently so because such waves depend on undulations in ϕ at $\partial\mathcal{D}$. The imposition, at the outset, of the extra condition (80) is like making the $\mathcal{M}^{\mathrm{C}}$ of figure 1 infinitely steep in certain directions in configuration space. In the present, infinite-dimensional fluid case these directions correspond to $\mathbf{a}$ values signifying fluid particles located on $\partial\mathcal{D}$.

To find the remaining boundary conditions, once the choice of configuration space has been made, we may probe these same directions in configuration space by forming the variational counterpart of (3) and considering suitable variations at $\partial\mathcal{D}$. The following applies equally to shallow water and three-dimensional stratified cases. Boundary conditions that preserve Hamiltonian structure must satisfy the basic principle (10): their variational form must be derivable by applying the same constraint to the variations as to the phase space flow. Thus, for instance, if (80) is imposed, then we must also impose

$$\partial\delta^{\mathrm{E}}\phi/\partial\ell = 0 \quad \text{on} \quad \partial\mathcal{D} \ . \tag{81}$$

Now the variational counterpart of (3) can be exhibited in three useful versions, as follows, where, because of (57)ff., $\mathbf{u}^{\mathrm{P}}$ and $\mathbf{u}^{\mathrm{C}}$ are absolute (inertial-frame) velocities:

$$-\int_{\mathcal{D}} \left(u_i^{\mathrm{P}}\delta^{\mathrm{L}}u_i^{\mathrm{C}} - \dot{u}_i^{\mathrm{C}}\delta^{\mathrm{L}}x_i\right)\mathrm{d}m + \delta\widetilde{\mathsf{H}} = 0 \ , \qquad \text{cf. (33)}, \tag{82}$$

or equivalently, via (57),

$$-\int_{\mathcal{D}} \left(u_i^{\mathrm{S}}\delta^{\mathrm{L}}u_i^{\mathrm{C}} - \dot{u}_i^{\mathrm{C}}\delta^{\mathrm{L}}x_i\right)\mathrm{d}m + \delta\mathsf{V} = 0 \ , \qquad \text{cf. (34)}, \tag{83}$$

whose Eulerian form, via (49), is evidently

$$-\int_{\mathcal{D}} \left\{u_j^{\mathrm{S}}\delta^{\mathrm{E}}u_j^{\mathrm{C}} + \left(u_j^{\mathrm{S}}\frac{\partial u_j^{\mathrm{C}}}{\partial x_i} - \dot{u}_i^{\mathrm{C}}\right)\delta^{\mathrm{L}}x_i\right\}\mathrm{d}m + \delta\mathsf{V} = 0 \ , \tag{84}$$

with a first term that is reference-frame-independent. The symbol $\dot{\mathbf{u}}^{\mathrm{C}}$ means, as before, the material rate of change of $\mathbf{u}^{\mathrm{C}}$ when the mass configuration evolves with the balanced model dynamics, and is not to be confused with $\mathbf{A}^{\mathrm{C}}$. The second of these three equivalent versions, (83), is the version closest to the integral equation (3), and can easily be obtained from it. One need only multiply (3) scalarly by $\delta^{\mathrm{L}}\mathbf{x}$ then integrate over $\mathcal{D}$, a process analogous to going from the first of (40) back to (34).

The first version, (82), will be most convenient for deducing energy conservation, being the fluid counterpart of $-\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{C}} + d\widetilde{H} = 0$ in (33), which fact provides another route to all three versions and a useful cross-check on their correctness. The third version, (84), will be the most convenient for deducing boundary information.

The boundary information is contained in the reference-frame-independent first term of (84). The reason is that the $\mathbf{u}^{\mathrm{C}}$ functionals of physical interest, such as (67) and its more accurate counterparts, contain contributions involving gradients and possibly higher spatial derivatives of the field or fields describing the mass configuration. Therefore the term containing $\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}$ has to be integrated by parts in order to bring out a factor $\delta^{\mathrm{L}}\mathbf{x}$, making the variations removable from the integral over $\mathcal{D}$. This will generally produce boundary terms. Notice the contrast with the parent dynamics, which is described by (82) with the C's replaced by P's, and in which $\delta^{\mathrm{L}}\mathbf{u}^{\mathrm{P}}$ and $\delta^{\mathrm{L}}\mathbf{x}$ are independently variable because of the larger parent phase space, and are therefore removable as they are, in the standard way, to give just the parent Hamilton's equations and no further boundary conditions.

The reader might wonder how the relation (84) can appear to make more information available than (3), even though derivable from it as just indicated. This is unlike the finite-dimensional (34), which manifestly contains no more and no less information than (40). The answer lies in a mathematical technicality peculiar to infinite dimensions, and related to the limiting process in which the boundary $\partial\mathcal{D}$ is approached from the interior.

Note first that (60) and its original derivation in §3 has a clear meaning only for particles not *at* $\partial\mathcal{D}$, even if arbitrarily close. This is because of the Dirac delta functions in (50) and, for instance, in (74). By contrast, (84) contains no delta functions when the $\mathbf{u}^{\mathrm{C}}$ field and its variations are smooth. This allows us to handle the limiting process in a straightforward way. More precisely, (84), unlike (3), can accommodate limiting forms of the variations $\delta^{\mathrm{L}}\mathbf{x}$, and the associated $\delta^{\mathrm{E}}h = g^{-1}\delta^{\mathrm{E}}\phi$, that probe conditions at the boundary $\partial\mathcal{D}$ alone. In the limit, we may leave particle positions unvaried throughout the interior of $\mathcal{D}$, while varying h, equivalently ϕ, on $\partial\mathcal{D}$ alone. We can furthermore vary independently, on $\partial\mathcal{D}$ alone, as many of the normal derivatives of h or ϕ as we please. This is done by moving particles in an arbitrarily thin band adjacent to $\partial\mathcal{D}$, requiring, in the limit, negligible virtual mass flux to or from the interior. Such variations are invisible in the derivation of (60), which relies on (51) and (52) and therefore on $\delta^{\mathrm{L}}\mathbf{x}$ and $\delta^{\mathrm{E}}h = g^{-1}\delta^{\mathrm{E}}\phi$ being smooth functions, able to be multiplied by Dirac delta functions $\delta(\mathbf{x}-\mathbf{x}')$ and their derivatives and then integrated over $\mathbf{x}'$ when $\mathbf{x} \notin \partial\mathcal{D}$.

The upshot of the general procedure based on (84), using the freedom of boundary variations just noted, is that the boundary terms emerging from all the integrations by parts of the first term of (84) — which are line integrals over the $\partial\mathcal{D}_i$, containing arbitrary variations $\delta^{\mathrm{E}}h = g^{-1}\delta^{\mathrm{E}}\phi$ and, in general, a finite number of the normal derivatives of those variations — imply that the coefficient of each such variation, $\delta^{\mathrm{E}}\phi$ and any normal derivatives that appear, must vanish, along with the integrand multiplying $\delta^{\mathrm{L}}\mathbf{x}$ in the remaining integral over $\mathcal{D}$. The vanishing of the latter integrand gets us back to (60), as can be verified in a few lines of manipulation; but we now have the necessary complement of boundary conditions as well.

Before turning to explicit illustrations, we point out that one of the integrations by parts, namely the last of the sequence, the one that frees the factor $\delta^{\mathrm{L}}\mathbf{x}$ in the integral over $\mathcal{D}$, always fails to produce additional boundary information because it has been, so to speak, pre-empted by the boundary information imposed when defining the configuration space. The boundary term that emerges at that stage is always annihilated by the configuration-space boundary condition (44), the vanishing at $\partial\mathcal{D}$ of the normal component of $\delta^{\mathrm{L}}\mathbf{x}$. The same thing, it will have been noticed, happens with the integrations by parts needed to derive (50), the basic relation between Lagrangian and Eulerian functional derivatives. Thus (50) and its three-dimensional stratified counterpart are true without qualification; we can go freely between the Lagrangian and Eulerian forms of the functional derivatives without producing any boundary terms. All this is because of (44) and the contraction with $\delta^{\mathrm{L}}\mathbf{x}$ in (51). Thus for instance it is immaterial, for present purposes, whether we begin with (3) or with (60).

In the case of L_1 dynamics, with $\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}$ proportional to gradients of $h(\mathbf{x})$ — making (60) equivalent to the partial differential equation (77), as already shown — just one extra boundary term is produced. Details are spelt out in the next section. Higher derivatives in $\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C}}$ will produce a higher order differential operator and correspondingly more boundary terms. With a fully nonlocal $\mathbf{u}^{\mathrm{C}}$ functional, we must expect (60) to be equivalent to an infinite-order partial differential equation with an infinite number of boundary conditions, probably most conveniently treated as a integro-differential equation with a kernel free of delta functions and a finite number of boundary conditions. Details for cases of this kind remain to be worked out.

We can now see how to show that energy conservation will hold for a general Hamiltonian balanced model of this kind. Energy conservation must, of course, on physical grounds, be expected to depend on boundary conditions. Notice first that, when (82) holds true for arbitrary $\delta^{\mathrm{L}}\mathbf{x}$, it holds true, in particular, when $\delta^{\mathrm{L}}\mathbf{x}$ is replaced by $\mathbf{u}^{\mathrm{P}}$ and $\delta^{\mathrm{L}}\mathbf{u}^{\mathrm{C}}$ by $\dot{\mathbf{u}}^{\mathrm{C}}$. (In the language of exterior calculus, this is the interior product of the phase-space flow with (82) regarded as a one-form, the fluid counterpart of $-\dot{\mathbf{x}}\rfloor\dot{\mathbf{x}}\rfloor\Omega^{\mathrm{C}} + \dot{\mathbf{x}}\rfloor d\widetilde{H} = 0$ of §2.) The first term of (82) then plainly vanishes, by skew-symmetry, and the second term becomes the time derivative of $\widetilde{\mathsf{H}}$. Now (82) does hold true in general when, and only when, all the boundary conditions and interior equations, derived according to the above general procedure, are satisfied. Energy conservation then follows, as expected, for the balanced model, with $\widetilde{\mathsf{H}}$ as the conserved energy.

We now make the foregoing more concrete by presenting explicit details for L_1 dynamics.

6 Boundary conditions for L_1 dynamics

First consider the case in which, straightforwardly, the configuration space for the balanced model is defined to be the same as the configuration space for the parent dynamics. This means that the boundary conditions will include, ab initio, as part of the specification of the configuration space, only the no-normal-flow conditions (44) and (78), the latter being equivalent to

$$\mathbf{U}^{\mathrm{S}}.\hat{\mathbf{n}} = -\mathbf{u}^{\mathrm{C(rel)}}.\,\hat{\mathbf{n}} = -\mathbf{u}^{\mathrm{G(rel)}}.\,\hat{\mathbf{n}} \quad \text{on} \quad \partial\mathcal{D}\ . \tag{85}$$

Inspection of (77) reminds us that this provides only half the needed boundary information. Following now the general procedure outlined in the previous section, we substitute the geostrophic $\mathbf{u}^{\mathrm{C}}$, (67), straight into the first term of (84), then integrate by parts. The first term of (67), $f\hat{\mathbf{z}}\times\mathbf{x}$, makes no contribution because it is mass-configuration-independent and has vanishing Eulerian variation. We have simply $\delta^{\mathrm{E}}u_j^{\mathrm{C}} = \delta^{\mathrm{E}}u_j^{\mathrm{G(rel)}} = -(g/f)\varepsilon_{jk}\delta^{\mathrm{E}}(\partial h/\partial x_k)$, so that the first term of (84) is

$$-\frac{g}{f}\int_{\mathcal{D}} u_j^{\mathrm{S}}\varepsilon_{jk}\delta^{\mathrm{E}}\left(\frac{\partial h}{\partial x_k}\right)\,\mathrm{d}m = \frac{g}{f}\int_{\mathcal{D}}\varepsilon_{jk}\frac{\partial U_j^{\mathrm{S}}}{\partial x_k}\delta^{\mathrm{E}}h\,\mathrm{d}\mathbf{x} - \frac{g}{f}\oint_{\partial\mathcal{D}}\mathrm{d}\ell\;\varepsilon_{jk}n_k U_j^{\mathrm{S}}\delta^{\mathrm{E}}h\ . \tag{86}$$

One more integration by parts is required to expose the interior variation $\delta^{\mathrm{L}}\mathbf{x}$, after use of (48), but as explained above this produces no further boundary terms, because of (44). The boundary term displayed in (86) contains all the remaining boundary information in this case. That term must vanish separately, for the reasons discussed below (84), with, for the chosen configuration space, $\delta^{\mathrm{E}}h$ an arbitrary function on $\partial\mathcal{D}$. It follows that $\varepsilon_{jk}n_kU_j^{\mathrm{S}} = 0$ on $\partial\mathcal{D}$, i.e.,

$$\mathbf{U}^{\mathrm{S}}.\,\hat{\boldsymbol{\ell}} = 0 \quad \text{on} \quad \partial\mathcal{D}\ , \tag{87}$$

where $\hat{\boldsymbol{\ell}} = \hat{\mathbf{z}}\times\hat{\mathbf{n}}$ is a tangential unit vector. The remaining terms in (84), all integrals over the interior $\mathcal{D}$, recover Salmon's equation (77) after a few lines of manipulation; so this calculation also provides a check on the previous derivation of that equation direct from (60).

The boundary condition (87) was first postulated by Allen *et al.* (1990), from energetic considerations, and has been derived from Hamilton's principle in the Clebsch representation by Salmon (personal communication) and via a different route by Allen and Holm (1996).

7 Filtered L_1 dynamics

The Hamiltonian balanced model that we call filtered L_1 dynamics, because it filters boundary Kelvin waves and is PV-invertible, results from imposing $\partial h/\partial\ell = 0$ on $\partial\mathcal{D}$ at the outset, as well as (85), and the variational counterparts of both of these. As already discussed in §5, the extra conditions can be regarded as constraining the parent dynamics slighly more severely than before, resulting in a slightly smaller configuration space. Because $\mathbf{u}^{\mathrm{C}}$ itself has not changed, the same line integral, that in (86), must still vanish. So we can read off the remaining boundary conditions at once, giving the complete set

$$\mathbf{U}^{\mathrm{S}}.\,\hat{\mathbf{n}} = 0 \quad \text{on} \quad \partial\mathcal{D} \tag{88}$$

from (78), because now $\mathbf{u}^{\mathrm{G(rel)}}.\,\mathbf{n} = 0$ on $\partial\mathcal{D}$, and

$$\frac{\partial h}{\partial \ell} = 0 \qquad \text{hence} \qquad \frac{\partial^2 h}{\partial \ell \partial t} = \frac{\partial}{\partial \ell}(\nabla.\mathbf{U}^{\mathrm{S}}) = 0 \quad \text{on} \quad \partial\mathcal{D}\ ; \tag{89}$$

with

$$\oint_{\partial\mathcal{D}_i} \mathbf{U}^{\mathrm{S}}\ .\,\mathrm{d}\boldsymbol{\ell} = 0 \tag{90}$$

separately on each connected part $\partial\mathcal{D}_i$ of $\partial\mathcal{D}$, from (86). This last condition, (90), replaces (87) because of the reduced freedom in $\delta^{\mathrm{E}} h$, which in accordance with the basic principle (10) can now change only its single value on each connected part $\partial\mathcal{D}_i$ of $\partial\mathcal{D}$. As a check that the general procedure of §5 has, indeed, led to well determined boundary value problems for L_1 dynamics and filtered L_1 dynamics, uniqueness proofs for both are given for both in Appendix A. Existence, too, can be demonstrated (R. Temam, personal communication); see end of Appendix A.

It is known of course that L_1 dynamics, as defined by (77) with (85) and (87), supports Kelvin waves, indeed accurately describes them in the long-wave limit, in which the wavelength along the boundary and the boundary radius of curvature are both much greater than the Rossby length K^{-1}. This is known from, for instance, the work of Allen *et al.* (*op. cit.*). The Kelvin-wave solutions illustrate how L_1 dynamics resolves the *boundary conflict* between the no-normal-flow condition and the $\mathbf{u}^{\mathrm{C}}$ functional. Near the boundary (more precisely, within about a Rossby length K^{-1} of the boundary), the $\mathbf{u}^{\mathrm{S}}$ field increases in magnitude to be comparable to $\mathbf{u}^{\mathrm{C}}$. This is in contrast to its usual magnitude, with $\mathbf{u}^{\mathrm{S}}$ being $O(\mathrm{Ro})$ smaller than $\mathbf{u}^{\mathrm{C}}$.

Filtered L_1 dynamics, on the other hand, exemplifies Hamiltonian balanced models with no such boundary conflict, and no scaling anomaly near the boundary. Related to this is the fact that filtered L_1 dynamics is fully PV-invertible, in the following sense. The point has more than theoretical interest: it could also be useful when performing numerical integrations of such models, because it relates to well-conditionedness at small Fr and Ro.

PV inversion means using a balance condition to deduce the velocity and mass fields from the Eulerian PV field $Q(\mathbf{x})$, in the shallow water case, or from isentropic distributions $Q(x_1, x_2, s)$ and boundary $s(x_1, x_2)$ fields in the three-dimensional stratified case, where s is specific entropy σ/ρ (Hoskins *et al.* 1985). Here the relevant balance condition is that defined by the $\mathbf{u}^{\mathrm{C}}$ functional; therefore, PV inversion will yield $\mathbf{u}^{\mathrm{C}}$ and not $\mathbf{u}^{\mathrm{P}}$. The implication is that we may timestep such a model in two equivalent ways, after solving (3) or (60) for $\mathbf{u}^{\mathrm{S}}$ and hence $\mathbf{u}^{\mathrm{P}}$.

The first way is to use the $\mathbf{u}^{\mathrm{P}}$ field to advect the mass configuration directly, followed by evaluation of $\mathbf{u}^{\mathrm{C}}$. The second way, likely to be better-conditioned numerically, is to use the $\mathbf{u}^{\mathrm{P}}$ field to advect the PV (and similarly to advect the surface s in the stratified case), then invert to get the new mass configuration simultaneously with $\mathbf{u}^{\mathrm{C}}$. The two ways are equivalent, i.e. mutually consistent, in the absence of numerical truncation errors, because of Q and surface s being exactly conserved on particles moving with velocity $\mathbf{u}^{\mathrm{P}}$. However, under the usual parameter conditions, it is likely that the second way will be less sensitive to truncation error. Under typical rapidly-rotating conditions with $\mathrm{Ro} \ll 1$, the $\mathbf{u}^{\mathrm{C}}$ functional is such that slight errors in the mass field have large, $O(\mathrm{Ro}^{-1})$ effects on the relative velocity field. For a given Ro there will be a tradeoff between such sensitivity and the cost of solving two elliptic problems instead of one, at each timestep.

PV inversion is indeed possible under the boundary condition $\partial h/\partial \ell = 0$ on $\partial\mathcal{D}$, if we also specify the absolute boundary circulations Γ_i on each connected part $\partial\mathcal{D}_i$ of $\partial\mathcal{D}$,

$$\Gamma_i = \oint_{\partial\mathcal{D}_i} \mathbf{u}^{\mathrm{C}}.\,\mathrm{d}\boldsymbol{\ell}\ . \tag{91}$$

As indicated in §11 below, the Γ_i are separately constants of the motion. Hence the Γ_i (not to be confused with the line integrals (90) coming from (60)) are known from the initial conditions. Alternatively, we may specify the total mass (79), also known from initial conditions (in the stratified case, this becomes the mass under each isentropic surface), together with all but one of the Γ_i, hence none at all in the simply connected case.

It can be shown that either set of conditions is sufficient to determine the solution to the inversion equation, which, in the shallow water case with balance condition (67), can be seen from the definition (1) of the PV, Q^{C}, to be

$$\left(\nabla^2 - K^2(\mathbf{x})\right) h = -f^2/g \tag{92}$$

where, as before, $K^2 = fQ^{\mathrm{C}}/g$, bringing in the given PV field $Q^{\mathrm{C}}(\mathbf{x})$. That the first set of conditions is sufficient (specifying all the Γ_i) can be proved by using relation (145) of Appendix A with h in place of $\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}$. That the second is sufficient (specifying the total mass (79) plus all but one of the Γ_i) follows from the relation

$$\sum_i \Gamma_i = \frac{g}{f}\sum_i \oint_{\partial\mathcal{D}_i} \frac{\partial h}{\partial n}\,\mathrm{d}\ell = \int_{\mathcal{D}} \left(Q^{\mathrm{C}}h - f\right)\mathrm{d}\mathbf{x}\ , \tag{93}$$

which shows that (with the given $Q^{\mathrm{C}}(\mathbf{x}) > 0$) the single free Γ_i value increases monotonically as we add mass increments to the layer. This is because h increases monotonically *pointwise* at each $\mathbf{x}$, because of the linearity and maximum-on-boundary properties of the problem $(\nabla^2 - K^2)\eta = 0$ in $\mathcal{D}$ with $\eta = 0$ on all $\partial\mathcal{D}_i$ except one on which $\eta = \eta_0 = \text{constant} > 0$, implying $0 \le \eta \propto \eta_0$.

8 Canonical-coordinate theorems

Which Hamiltonian balanced models have canonical coordinates $\mathbf{X}$ given by explicit formulae, like those discovered by Hoskins for semigeostrophy? We can now provide, rather easily, a general formal answer to this longstanding question. The first step was taken by Salmon (1988*a*, eq. 5.18), who found a $\mathbf{u}^{\mathrm{C}}$ functional that corresponds to Hoskins' $\mathbf{X}$, the case $\alpha = \frac{1}{2}$ in (4). The full generalization emerged, during the present investigation, by analogy with recent results applying the theory of contact transformations to semigeostrophy (Sewell & Roulstone 1994, Roulstone & Sewell 1996*b*). The theory of this section applies to the three-dimensional stratified as well as to the shallow water case, in two separate senses to be explained below, depending on the choice of parent dynamics.

Canonical coordinates on the slow manifold $\mathcal{M}^{\mathrm{C}}$, in the most general sense analogous to Hoskins' $\mathbf{X}$, are both fields and also mass-configuration functionals that respect the particle-relabelling symmetry, like $\mathbf{u}^{\mathrm{C}}$:

$$\mathbf{X} = (X_1, X_2) = \mathbf{X}(\mathbf{a}; \mathbf{x}(\cdot)) = \mathbf{X}(\mathbf{x}; h(\cdot)) \ , \tag{94}$$

or $\mathbf{X}(\mathbf{x}; \rho(\cdot), \sigma(\cdot))$ for the three-dimensional stratified case. The dependence on the mass configuration can be nonlocal. $\mathbf{X}$ is canonical by definition if, for some scalar configuration functional, say $\mathsf{B}(\mathbf{x}(\cdot)) = \mathsf{B}(h(\cdot))$ or $\mathsf{B}(\mathbf{x}; \rho(\cdot), \sigma(\cdot))$, and with $\mathbf{u}^{\mathrm{C}}$ the absolute constraint velocity, i.e., defined in the inertial frame,

$$\theta^{\mathrm{C}} \equiv - \int_{\mathcal{D}} u_i^{\mathrm{C}} \delta^{\mathrm{L}} x_i \mathrm{d}m = f_0 \int_{\mathcal{D}} \tfrac{1}{2} \varepsilon_{jk} X_k \delta^L X_j \mathrm{d}m + \delta \mathsf{B} \ . \tag{95}$$

Here f_0 is a constant normalization factor, introduced to give $\mathbf{X}$ dimensions of length. It will usually, but need not, be identified with some typical or average angular velocity of the spinning mass of fluid; see §10 for the application to models with variable Coriolis parameter.

The central result can be stated as follows.

Theorem 2 (Canonical coordinate theorem, general version) *General canonical coordinates* $\mathbf{X}$ *are functionally related to the absolute constraint velocity* $\mathbf{u}^{\mathrm{C}}$ *by*

$$u_i^{\mathrm{C}} = -\tfrac{1}{2} f_0 \varepsilon_{jk} X_k \frac{\partial X_j}{\partial x_i} - \frac{\partial}{\partial x_i} \left\{ \tfrac{1}{2} f_0 \varepsilon_{jk} \int_{\mathcal{D}} X_k(\mathbf{x}'; h(\cdot)) \frac{\delta^{\mathrm{E}} X_j(\mathbf{x}'; h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x})} \mathrm{d}m(\mathbf{x}') + \frac{\delta \mathsf{B}(h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x})} \right\} \tag{96}$$

for shallow water dynamics, and by

$$\begin{aligned} u_i^{\mathrm{C}} = & -\tfrac{1}{2} f_0 \varepsilon_{jk} X_k \frac{\partial X_j}{\partial x_i} \\ & - \frac{\partial}{\partial x_i} \left\{ \tfrac{1}{2} f_0 \varepsilon_{jk} \int_{\mathcal{D}} X_k(\mathbf{x}'; \rho(\cdot); \sigma(\cdot)) \frac{\delta^{\mathrm{E}} X_j(\mathbf{x}'; \rho(\cdot); \sigma(\cdot))}{\delta^{\mathrm{E}} \rho(\mathbf{x})} \mathrm{d}m(\mathbf{x}') + \frac{\delta \mathsf{B}(\rho(\cdot); \sigma(\cdot))}{\delta^{\mathrm{E}} \rho(\mathbf{x})} \right\} \\ & - \frac{\sigma(\mathbf{x})}{\rho(\mathbf{x})} \frac{\partial}{\partial x_i} \left\{ \tfrac{1}{2} f_0 \varepsilon_{jk} \int_{\mathcal{D}} X_k(\mathbf{x}'; \rho(\cdot); \sigma(\cdot)) \frac{\delta^{\mathrm{E}} X_j(\mathbf{x}'; \rho(\cdot); \sigma(\cdot))}{\delta^{\mathrm{E}} \sigma(\mathbf{x})} \mathrm{d}m(\mathbf{x}') + \frac{\delta \mathsf{B}(\rho(\cdot); \sigma(\cdot))}{\delta^{\mathrm{E}} \sigma(\mathbf{x})} \right\} \end{aligned} \tag{97}$$

for three-dimensional stratified dynamics, for some configuration functional B.

The proof is a straightforward application of the same Eulerian–Lagrangian functional differentiation machinery[1] as used in §3, especially (49), (50) and (48) for shallow water dynamics, and

[1] Direct substitution shows that (96) and (97) satisfy (95); arbitrariness of $\delta^{\mathrm{L}}\mathbf{x}$ takes us back to (96) and (97).

their counterparts including (63), (64) and (65) for stratified dynamics. Note that, in (97), i runs from 1 to 3 but j and k still run from 1 to 2. Equations (96) and (97) can be regarded as functional equations defining the possible canonical coordinates $\mathbf{X}$, if any, for a given $\mathbf{u}^{\mathrm{C}}$ functional and any choice of the arbitrary functional B. It can also, more tractably, be regarded as defining the class of $\mathbf{u}^{\mathrm{C}}$ functionals that yield Hamiltonian balanced models with canonical coordinates of the type $\mathbf{X}$, given by simple formulae or otherwise. One need only substitute a given such formula into the right-hand side. Of course what is interesting is to choose the formula for $\mathbf{X}$ in such a way as to produce a balanced model that has good accuracy in some sense. To show that this is possible, it is sufficient to produce specific examples; the next section will present what is arguably the simplest such set of examples.

Notice that the contributions to the $\mathbf{u}^{\mathrm{C}}$ fields given by (96) and (97) are equal either to a gradient in physical space, or to such a gradient times $\sigma(\mathbf{x})/\rho(\mathbf{x})$, with one important exception, namely the leading contribution, the term $\frac{1}{2}f_0\varepsilon_{jk}X_k\partial X_j/\partial x_i$. This leading contribution is simple, involving no functional derivatives, because it arises solely from Lagrangian corrections associated with the first term $\delta^{\mathrm{L}}\mathbf{x}.\nabla$ on the right of (49). This gives corresponding simplifications in the vorticities and PVs. The shallow water ζ^{C} and the Rossby–Ertel PV in the stratified case, $\boldsymbol{\zeta}^{\mathrm{C}}.\nabla(\sigma/\rho) = \boldsymbol{\zeta}^{\mathrm{C}}.\nabla(s)$, are therefore correspondingly simple. Indeed they have exactly the simplicity of semigeostrophy, being expressible as the same Jacobian determinants found by Hoskins for semigeostrophy. The results, which are an immediate consequence of (96) and (97), and hence have the same generality, can be summarized as follows:

Theorem 3 (Jacobian vorticity/PV theorem) *If canonical coordinates $\mathbf{X}$ of the form (94) are known for a Hamiltonian balanced model generated by constraining a parent dynamics, as in §3 or §11 below, then the (absolute) constraint vorticity ζ^{C} can always be expressed in Jacobian form:*

$$\zeta^{\mathrm{C}} = f_0\frac{\partial(X_1, X_2)}{\partial(x_1, x_2)}\ . \tag{98}$$

The materially conserved potential vorticities Q^{C}, for the stratified and shallow water cases (cf. equation (1)), can likewise always be expressed in Jacobian form. Respectively, these are

$$Q^{\mathrm{C}} = \frac{f_0}{\rho}\frac{\partial(X_1, X_2, s)}{\partial(x_1, x_2, x_3)}\ , \qquad Q^{\mathrm{C}} = \frac{f_0}{h}\frac{\partial(X_1, X_2)}{\partial(x_1, x_2)}\ , \tag{99}$$

where as before s is the specific entropy σ/ρ. The second of (99) also applies to the stratified case, when described using Rossby's isentropic coordinates. With the exception of this last statement, to be justified below, the results (98) and (99) follow from (96) and (97) by straightforward manipulations, including use of the chain rule for Jacobians and the mass-element relation (46) and its three-dimensional counterpart $\rho = \partial(a_1, a_2, a_3)/\partial(x_1, x_2, x_3)$:

$$\frac{Q^{\mathrm{C}}}{f_0} = \frac{\partial(X_1, X_2, s)}{\partial(a_1, a_2, a_3)} = \frac{1}{\rho}\frac{\partial(X_1, X_2, s)}{\partial(x_1, x_2, x_3)}\ , \ \text{ and } \ \frac{Q^{\mathrm{C}}}{f_0} = \frac{\partial(X_1, X_2)}{\partial(a_1, a_2)} = \frac{1}{h}\frac{\partial(X_1, X_2)}{\partial(x_1, x_2)}\ . \tag{100}$$

The first of (99) generalizes the three-dimensional result derived in Hoskins (1975). Note incidentally that any function of s alone, potential temperature for instance, can be substituted for s

in the definition of the stratified (Rossby–Ertel) PV, provided that we make the corresponding substitution in the Jacobian formulae (99) and (100).

Anticipating the considerations of §11 — where it is pointed out that semigeostrophy uses, in effect a second slow manifold to constrain the Hamiltonian functional — we note that the foregoing results, unlike (3) and (60), have no dependence on the Hamiltonian functional. They are purely a consequence of restricting to $\mathcal{M}^{\mathrm{C}}$ the parent symplectic or contact structure[2] (see §11), and they hold for any Hamiltonian functional whatever. Thus the Jacobian formulae found in semigeostrophy are indeed special cases of the foregoing, in which $\mathcal{M}^{\mathrm{C}}$ is the particular $\mathcal{M}^{\mathrm{C}}$ given by (4) with $\alpha = \frac{1}{2}$. The result (98) shows also, for instance, how the concept of 'vorticity coordinates' introduced by Schubert & Magnusdottir (1994) fits into the present framework, together with the recent results of Roulstone & Sewell (1996*a*). Note well that the vorticity, for this purpose, is always the *constraint* vorticity, the vorticity evaluated from $\mathbf{u}^{\mathrm{C}}$.

Note further that the first of (99) is appropriate to the case in which the parent dynamics is fully nonhydrostatic, so that (96) gives a three-dimensional $\mathbf{u}^{\mathrm{C}}$. This $\mathbf{u}^{\mathrm{C}}$ generally has a nonzero vertical component u_3^{C}, in contrast with $\mathbf{X}$, which still has only two quasi-horizontal components $\mathbf{X} = (X_1, X_2)$, as it must in order for the contraction with the two-dimensional alternating tensor ε_{jk} to make sense. We may expect u_3^{C} to be small for choices that lead to an accurate balanced model, but it cannot generally be zero. Thus (96) significantly restricts the class of $\mathbf{u}^{\mathrm{C}}$ functionals that give three-dimensional Hamiltonian balanced models with two-dimensional canonical coordinates like those of Hoskins.

It remains to discuss the sense in which the second, also, of (99) applies to the three-dimensional stratified problem. It applies to a case of that problem in which the parent dynamics is taken to be the meteorological 'primitive equations', with the hydrostatic approximation already imposed. We then need to reinterpret the notation $h(\mathbf{x})$ etc. in the following way. The stratified system is now regarded as a set of many interacting two-dimensional layers in the physical domain.[3] Therefore we use a description in which specific entropy s, or some function of it such as potential temperature, keeps track of vertical position in the physical domain $\mathcal{D}$, 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 subject to the particle-relabelling symmetry; and (x_1, x_2) is the horizontal projection of position on a given stratification surface $s =$ constant. The mass element

[2] The phrase 'symplectic or contact' is used for the following reason. The distinction between symplectic and contact structure is clear in finite dimensions: the first applies only to even-dimensional manifolds, and the second only to odd. Respectively, they refer to an even-dimensional manifold equipped with a two-form like Ω^{C}, and an odd-dimensional manifold equipped with a one-form like θ^{C}, together with appropriate nondegeneracy conditions. The distinction between even and odd dimensions has no obvious meaning in infinite dimensions; but what seems to matter here is that one still has a θ^{C} whose exterior derivative is Ω^{C}, with sufficient nondegeneracy to give invertible $\Omega_{ij}^{\mathrm{C}}(\mathbf{x}, \mathbf{x}')$ in some reasonable function space.

[3] This incidentally is very unlike the generalization of the toy problem of §2 to three $\mathbf{x}$ dimensions, which is degenerate. It is more like taking many particles, interacting through the potential V, and each moving in two space dimensions.

$\mathrm{d}m$ is now written in the form

$$\mathrm{d}m = \mathrm{d}a_1\mathrm{d}a_2\mathrm{d}s = h(x_1,x_2,s)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}s\ , \tag{101}$$

where h is given by $h = \rho(\partial x_3/\partial s)_{x_1,x_2}$ with x_3 denoting geometrical height as before. Note incidentally that h could be zero at a given horizontal position $\mathbf{x}$ and a given value of s, corresponding to (x_1,x_2) being outside the image of $\{(a_1,a_2)\}$ at that s value, as can happen for instance when isentropes intersect a horizontal boundary — the counterpart of shallow water 'lens' or 'blob' models with h vanishing at a free side boundary, and zero mass outside that boundary. The use of s to help identify particles within the physical domain $\mathcal{D}$ via (a_1,a_2,s) and (x_1,x_2,s) amounts to using the well known Rossby isentropic or isopycnic 'coordinate', more aptly 'label'.

The components x_i of $\mathbf{x}$ in (94) do not, incidentally, need to be components referred to Cartesian axes. For instance spherical components could be used, as in the theory of Shutts (1989), extending semigeostrophy 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 (1996*b*, §7).

We now show how the foregoing results apply to the important case of near-local canonical models, for brevity confining attention to shallow water dynamics. Both the $\mathbf{u}^{\mathrm{C}}$ functional and the putative canonical coordinates $\mathbf{X}$ are all now assumed to take near-local form, generalizing (6) and (4). That is, $\mathbf{u}^{\mathrm{C}}$ and $\mathbf{X}$ are assumed to be completely general functions of the depth $h(\mathbf{x})$ and a finite number of its derivatives, here denoted by

$$\mathbf{u}^{\mathrm{C}}(\mathbf{x},h(\cdot)) = \mathbf{u}^{\mathrm{C}}(x_j,h,h_{,i},h_{,ij},\cdots)\ , \tag{102}$$

$$\mathbf{X}(x_j,h,h_{,i},h_{,ij},\cdots) = \mathbf{X}(x_j,h,h_{,i},h_{,ij},\cdots)\ . \tag{103}$$

For consistency the functional B is assumed to take the corresponding form $\int_{\mathcal{D}} B\,\mathrm{d}m$ where

$$B(\mathbf{x},h(\cdot)) = B(x_j,h,h_{,i},h_{,ij},\cdots)\ . \tag{104}$$

When the number of arguments is arbitrary but finite, these are the most general possible forms expressing near-local functional dependence. We now have, starting for instance from (96):

Theorem 4 (Canonical coordinate theorem, near-local shallow water version) *Near-local canonical coordinates* $\mathbf{X}$ *are functionally related to the absolute constraint velocity* $\mathbf{u}^{\mathrm{C}}$ *by*

$$\begin{aligned} u_i^{\mathrm{C}} = {} & -\tfrac{1}{2}f_0\varepsilon_{jk}X_k\frac{\partial X_j}{\partial x_i} \\ & - \frac{\partial}{\partial x_i}\left\{\tfrac{1}{2}f_0\varepsilon_{jk}\left[hX_k\frac{\partial X_j}{\partial h} - \frac{\partial}{\partial x_p}\left(hX_k\frac{\partial X_j}{\partial h_{,p}}\right) + \frac{\partial^2}{\partial x_p\partial x_q}\left(hX_k\frac{\partial X_j}{\partial h_{,pq}}\right) - \cdots\right]\right. \\ & \left. + \left(B + h\frac{\partial B}{\partial h} - \frac{\partial}{\partial x_j}\left(h\frac{\partial B}{\partial h_{,j}}\right) + \frac{\partial^2}{\partial x_j\partial x_k}\left(h\frac{\partial^2 B}{\partial h_{,jk}}\right) - \cdots\right)\right\} \end{aligned} \tag{105}$$

for some near-local $B(x_j,h,h_{,i},h_{,ij},\cdots)$ *and near-local* $\mathbf{X}(x_j,h,h_{,i},h_{,ij},\cdots)$.

These $\mathbf{u}^{\mathrm{C}}$ functionals characterize all shallow-water near-local canonical models, in the sense under discussion; the next section gives some examples.

9 Canonical coordinates for Salmon's $\mathbf{L_1}$ dynamics and for the new balanced model

The simplest near-local examples take us back again to the f-plane, no-topography shallow water case ($f =$ constant, $\phi = gh$). This allows us to narrow the search for suitable explicit $\mathbf{X}$ formulae by restricting attention to physically acceptable expressions for f-plane dynamics, requiring invariance to coordinate axis rotations. The simplest such choice is the near-local form

$$X_i = x_i + a\frac{\partial h}{\partial x_i} + c\varepsilon_{ij}\frac{\partial h}{\partial x_j}\ , \qquad B = bh\ , \tag{106}$$

for suitably chosen constants a, b, c (not to be confused with Lagrangian labels $\mathbf{a}$). We make the choice

$$a = \frac{g}{f^2},\ \ b = \frac{g}{f}\gamma,\ \ c = \frac{g}{f^2}\gamma\ , \tag{107}$$

where γ is the constant introduced in (6) and (7). With this choice of constants we set $f_0 = f$ in (105). Inserting B from (106) into the last line of (105) and X_j from (106) into the right-hand factors of the first two lines gives (with the B contribution first)

$$\begin{aligned}\frac{1}{f}u_i^{\mathrm{C}} &= -2\frac{b}{f}\frac{\partial h}{\partial x_i}\\ &\quad -\tfrac{1}{2}\varepsilon_{jk}\left[X_k\left(\delta_{ij} + a\frac{\partial^2 h}{\partial x_i\partial x_j} + c\varepsilon_{jp}\frac{\partial^2 h}{\partial x_i\partial x_p}\right) - \frac{\partial^2}{\partial x_i\partial x_q}\Big(hX_k\left(a\delta_{jq} + c\varepsilon_{jp}\delta_{pq}\right)\Big)\right].\end{aligned} \tag{108}$$

Using the identity $\varepsilon_{jk}\varepsilon_{jp}\delta_{pq} = \delta_{kq}$ in (108) yields, after some cancellation between terms,

$$\begin{aligned}\frac{1}{f}u_i^{\mathrm{C}} = &-\overbrace{2\frac{b}{f}\frac{\partial h}{\partial x_i}}^{1} - \overbrace{\tfrac{1}{2}\varepsilon_{ik}X_k}^{2,3,4}\\ &+\overbrace{\frac{a}{2}\varepsilon_{qk}h\frac{\partial^2 X_k}{\partial x_i\partial x_q}}^{5} + \overbrace{\frac{a}{2}\varepsilon_{qk}\frac{\partial h}{\partial x_q}\frac{\partial X_k}{\partial x_i}}^{6,7,8} + \overbrace{\frac{a}{2}\varepsilon_{qk}\frac{\partial h}{\partial x_i}\frac{\partial X_k}{\partial x_q}}^{9}\\ &+\overbrace{\frac{c}{2}h\frac{\partial^2 X_k}{\partial x_i\partial x_k}}^{10} + \overbrace{\frac{c}{2}\frac{\partial h}{\partial x_i}\frac{\partial X_k}{\partial x_k}}^{11,12} + \overbrace{\frac{c}{2}\frac{\partial h}{\partial x_k}\frac{\partial X_k}{\partial x_i}}^{13,14,15}.\end{aligned} \tag{109}$$

The terms are numbered for the purpose of comparison later. Substituting for the remaining X_k factors with (106) now gives

$$\begin{aligned}\frac{1}{f}u_i^{\mathrm{C}} = &-\overbrace{2\frac{b}{f}\frac{\partial h}{\partial x_i}}^{1} - \overbrace{\tfrac{1}{2}\varepsilon_{ik}x_k}^{2} - \overbrace{\frac{a}{2}\varepsilon_{ik}\frac{\partial h}{\partial x_k}}^{3} + \overbrace{\frac{c}{2}\frac{\partial h}{\partial x_i}}^{4}\\ &-\overbrace{\frac{ac}{2}h\frac{\partial}{\partial x_i}\nabla^2 h}^{5} + \overbrace{\frac{a}{2}\varepsilon_{qi}\frac{\partial h}{\partial x_q}}^{6} + \overbrace{\frac{a^2}{2}\varepsilon_{qk}\frac{\partial h}{\partial x_q}\frac{\partial^2 h}{\partial x_i\partial x_k}}^{7} - \overbrace{\frac{ac}{2}\frac{\partial h}{\partial x_q}\frac{\partial^2 h}{\partial x_i\partial x_q}}^{8} - \overbrace{\frac{ac}{2}\frac{\partial h}{\partial x_i}\nabla^2 h}^{9}\\ &+\overbrace{\frac{ac}{2}h\frac{\partial}{\partial x_i}\nabla^2 h}^{10} + \overbrace{c\frac{\partial h}{\partial x_i}}^{11} + \overbrace{\frac{ac}{2}\frac{\partial h}{\partial x_i}\nabla^2 h}^{12} + \overbrace{\frac{c}{2}\frac{\partial h}{\partial x_i}}^{13} + \overbrace{\frac{ac}{2}\frac{\partial h}{\partial x_k}\frac{\partial^2 h}{\partial x_i\partial x_k}}^{14} + \overbrace{\frac{c^2}{2}\varepsilon_{kl}\frac{\partial h}{\partial x_k}\frac{\partial^2 h}{\partial x_i\partial x_l}}^{15}\end{aligned} \tag{110}$$

where $\nabla^2 = \partial^2/\partial x_i\partial x_i$. Because $b/f = c$ term 1 cancels with 4, 11 and 13. Further, terms 5 and 10, 9 and 12, and 8 and 14 all cancel in pairs. This eliminates *all* the terms proportional to c, and therefore to γ, and leaves (relabelling dummy indices)

$$\frac{1}{f}u_i^{\mathrm{C}} = -\tfrac{1}{2}\varepsilon_{ik}x_k - a\varepsilon_{ik}\frac{\partial h}{\partial x_k} + \frac{a^2+c^2}{2}\varepsilon_{jk}\frac{\partial h}{\partial x_j}\frac{\partial^2 h}{\partial x_i \partial x_k}, \tag{111}$$

which, on the right, are terms 2, 3 and 6, and 7 and 15 respectively.

If we set $\gamma = i$ then $a^2 = -c^2$, and (111) with (107) and (5) becomes

$$u_i^{\mathrm{C}} = -\tfrac{1}{2}f\varepsilon_{ik}x_k + u_i^{\mathrm{G(rel)}} . \tag{112}$$

This is, again, precisely the constraint defining Salmon's L_1 dynamics. The canonical coordinates for L_1 dynamics are therefore, as stated earlier, given by $\gamma = i$ in

$$X_i = x_i + \frac{g}{f^2}\frac{\partial h}{\partial x_i} + \gamma\frac{g}{f^2}\varepsilon_{ij}\frac{\partial h}{\partial x_j} . \tag{113}$$

The new balanced model has $\gamma = i\sqrt{3}$ hence $\frac{1}{2}(a^2+c^2) = -g^2/f^4$ in (111), giving two orders of accuracy in Rossby number Ro , that is, one order better than L_1 dynamics or semigeostrophy. This can be verified most easily either by using (98) with the first of (106), or by taking the curl of (111), to give

$$\zeta^{\mathrm{C}} = f\frac{\partial(X_1, X_2)}{\partial(x_1, x_2)} = \zeta^{\mathrm{G}} - \frac{2}{f}\frac{\partial(u_1^{\mathrm{G(rel)}}, u_2^{\mathrm{G(rel)}})}{\partial(x_1, x_2)} , \tag{114}$$

whose last term is like that in the physical-coordinate form of Hoskins' semigeostrophic vorticity, except that its coefficient is -2 times that in Hoskins' formula. As is well known, and can easily be verified by standard scaling arguments applied to the divergence equation of the parent dynamics, (114) implies that the new balance condition is correct to two orders in Rossby number Ro (e.g. Craig 1993).

A useful check on the foregoing calculation, including the multiple cancellations in (110), is to take the same canonical coordinates $X_i = x_i + a\partial h/\partial x_i + c\varepsilon_{ij}\partial h/\partial x_j$ and insert them into the corresponding two-form (with $f_0 = f$)

$$\tfrac{1}{2}f\int_{\mathcal{D}}\mathrm{d}m\ \varepsilon_{jk}\delta^{\mathrm{L}}X_j \wedge \delta^{\mathrm{L}}X_k = f\int_{\mathcal{D}}\mathrm{d}m\ \delta^{\mathrm{L}}\left(x_1 + a\frac{\partial h}{\partial x_1} + c\frac{\partial h}{\partial x_2}\right) \wedge \delta^{\mathrm{L}}\left(x_2 + a\frac{\partial h}{\partial x_2} - c\frac{\partial h}{\partial x_1}\right) . \tag{115}$$

This is the infinite-dimensional exterior derivative of the one-form θ^{C} defined by (95). It should be real-valued when γ and hence c is pure imaginary, for consistency with (111) and (110); the terms proportional to c, and hence γ, in the two-form must vanish. Those terms are

$$fc\int_{\mathcal{D}}\mathrm{d}m\left[-\delta^{\mathrm{L}}x_1 \wedge \delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_1}\right) - a\,\delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_1}\right) \wedge \delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_1}\right) + \delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_2}\right) \wedge \delta^{\mathrm{L}}x_2 + a\,\delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_2}\right) \wedge \delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_2}\right)\right] \tag{116}$$

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

$$fc\int_{\mathcal{D}}\mathrm{d}m\ \delta^{\mathrm{L}}\left(\frac{\partial h}{\partial x_i}\right) \wedge \delta^{\mathrm{L}}x_i = fc\int_{\mathcal{D}}\mathrm{d}m\left[\frac{\partial^2 h}{\partial x_i \partial x_j}\delta^{\mathrm{L}}x_j + \delta^{\mathrm{E}}\left(\frac{\partial h}{\partial x_i}\right)\right] \wedge \delta^{\mathrm{L}}x_i , \tag{117}$$

where the last step uses $\delta^{\mathrm{L}} = \mathbf{x}.\nabla + \delta^{\mathrm{E}}$, (49) of §3. The second-derivative term vanishes, again by the skew-symmetry of the wedge product. Therefore we are left with the Eulerian part $\delta^{\mathrm{E}}(\partial h/\partial x_i)\wedge$

$\delta^{\mathrm{L}} x_i$. Recalling that

$$\delta^{\mathrm{E}} h = -\frac{\partial}{\partial x_j}(h\delta^{\mathrm{L}} x_j) \ , \tag{118}$$

from (48) of §3, and that δ^{E} commutes with $\partial/\partial x_i$, we have

$$fc\int\limits_{\mathcal{D}} \mathrm{d}m \ \delta^{\mathrm{E}}\left(\frac{\partial h}{\partial x_i}\right) \wedge \delta^{\mathrm{L}} x_i = fc\int\limits_{\mathcal{D}} \mathrm{d}m \ \frac{\partial^2}{\partial x_i \partial x_j}\left(h\delta^{\mathrm{L}} x_j\right) \wedge \delta^{\mathrm{L}} x_i \ ; \tag{119}$$

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

$$fc\int\limits_{\mathcal{D}} \mathrm{d}\mathbf{x} \left[\frac{\partial}{\partial x_j}\left(h\delta^{\mathrm{L}} x_j\right)\right] \wedge \left[\frac{\partial}{\partial x_i}\left(h\delta^{\mathrm{L}} x_i\right)\right] = 0 \ . \tag{120}$$

Note that the boundary condition (44), relevant to the rigidly bounded case, forces the expression $\hat{n}_i\delta^{\mathrm{L}} x_j \wedge \delta^{\mathrm{L}} x_i$ to vanish on $\partial\mathcal{D}$ for each value of j separately; so no boundary terms arise here. The same is true of free boundaries where h itself vanishes, as in 'lens' or 'blob' models. Therefore, in all these cases, the two-form $\frac{1}{2}\int_{\mathcal{D}}\mathrm{d}m \ \varepsilon_{jk}\delta^{\mathrm{L}} X_j \wedge \delta^{\mathrm{L}} X_k$ is real; it contains only the square of γ and not the first power, confirming (111) and (110). A corollary is that the Jacobians in (98), (99) and (100) are real.

10 Galilean-invariant balance, and variable Coriolis parameter

For constant-Coriolis-parameter models, the $\mathbf{u}^{\mathrm{C}}$ functional has the general form

$$\mathbf{u}^{\mathrm{C}} \ = \ \mathbf{r}(\mathbf{x}) \ + \ \mathbf{u}^{\mathrm{C(rel)}} \ , \tag{121}$$

as illustrated by equation (4), where $\mathbf{r}(\mathbf{x}) = \frac{1}{2}f\hat{\mathbf{z}}\times\mathbf{x}$, representing solid rotation with angular velocity $\frac{1}{2}f$. Notice again that $\mathbf{r}(\mathbf{x})$ is mass-configuration-independent. That is, its Eulerian functional derivatives vanish. By contrast, $\mathbf{u}^{\mathrm{C(rel)}}$ is mass-configuration-dependent, $\mathbf{u}^{\mathrm{C(rel)}} = \mathbf{u}^{\mathrm{C(rel)}}(\mathbf{x}; h(\cdot))$ or $\mathbf{u}^{\mathrm{C(rel)}}(\mathbf{x}; \rho(\cdot), \sigma(\cdot))$, and rotationally invariant; e.g., for shallow water dynamics:

$$-\int\limits_{\mathcal{D}} r_j(\mathbf{x}')\frac{\partial h(\mathbf{x}')}{\partial x'_j} \ \frac{\delta^{\mathrm{E}}\mathbf{u}^{\mathrm{C(rel)}}(\mathbf{x}; h(\cdot))}{\delta^{\mathrm{E}} h(\mathbf{x}')}\mathrm{d}\mathbf{x}' = -\mathbf{r}.\nabla\mathbf{u}^{\mathrm{C}} - \mathbf{u}^{\mathrm{C}}.\nabla\mathbf{r} = -\mathbf{r}.\nabla\mathbf{u}^{\mathrm{C(rel)}} - \mathbf{u}^{\mathrm{C(rel)}}.\nabla\mathbf{r} \ . \tag{122}$$

That is, rotating the $h(\mathbf{x})$ pattern with $\mathbf{r}(\mathbf{x})$ must correspondingly rotate the $\mathbf{u}^{\mathrm{C}}(\mathbf{x})$ field, i.e. must Lie-advect it with $\mathbf{r}$.

Note once again (Remark **4** of §3) that $\frac{1}{2}f$ need not be thought of as a *reference-frame* rotation rate. It can be defined in wholly physical terms, as a typical or average rotation rate of the spinning mass of fluid whose dynamics we are trying to model. As already suggested, this gives us a natural way of distinguishing 'centrifugal' and 'Coriolis' accelerations from 'relative' accelerations that is based entirely on the $\mathbf{u}^{\mathrm{C}}$ functional and the constraint acceleration $\mathbf{A}^{\mathrm{C}}$ derived from it, and not at all on arbitrary choices of reference-frame rotation rate. The question of whether or not the shape of the bottom boundary matches the centrifugal–gravitational paraboloids, spheroids or other equipotentials becomes a separate question. (If nonaxisymmetric topography or other nonaxisymmetric boundaries are present, then a particular rotating frame is, of course, distinguished physically, and shows itself through (122), which will fail unless the topography and other boundaries rotate with absolute velocity $\mathbf{r}(\mathbf{x})$.) The centrifugal acceleration associated with (121) is defined as the mass-configuration-independent contribution to $\mathbf{A}^{\mathrm{C}}$, and the Coriolis acceleration as the cross term in (59). (To verify this last statement, note that because $\nabla.\mathbf{r} = 0$ the $\mathbf{r}$ contribution to the last term of (59) reduces to the left-hand side of (122); this takes the place of the reference-frame transformation of $\partial/\partial t$ in textbook derivations of the Coriolis term, and in the corresponding argument for the parent dynamics.) One might summarize all this by saying that the formulation in terms of $\mathbf{u}^{\mathrm{C}}$ and the integral equation (3) or (60) gives us a natural way of incorporating into a balanced model the appropriate kind of Galilean invariance.

It also gives us a natural and straightforward way of defining what are usually called variable-Coriolis-parameter models. We need only take $\mathbf{r}(\mathbf{x})$ in (121) to be some function of $\mathbf{x}$ alone that departs from solid rotation. The only formal restriction is that we must still have

$$\nabla.\mathbf{r}(\mathbf{x}) \;=\; 0\;, \tag{123}$$

because we want to avoid a situation in which all mass distributions will be rearranged by the velocity field $\mathbf{r}(\mathbf{x})$. The corresponding constraint vorticity

$$\boldsymbol{\zeta}^{\mathrm{C}} = \nabla\times\mathbf{r}(\mathbf{x}) + \nabla\times\mathbf{u}^{\mathrm{C(rel)}} \tag{124}$$

has a first term, $\nabla\times\mathbf{r}(\mathbf{x}) \;=\; \hat{\mathbf{z}}f(\mathbf{x})$, say, that can be identified with the Coriolis parameter of a variable-Coriolis-parameter model. Again, 'centrifugal' and 'Coriolis effects can still be distinguished by using (122) and (123) and the fact that the Eulerian functional derivatives of $\mathbf{r}(\mathbf{x})$ vanish.

In the near-local shallow water case, the last two relations become

$$\mathbf{u}^{\mathrm{C}} = \mathbf{r}(\mathbf{x}) + \text{terms involving } h(\mathbf{x}) \text{ and its derivatives} \tag{125}$$

and

$$\zeta^{\mathrm{C}} = f(\mathbf{x}) + \quad \text{"} \qquad \text{"} \qquad \text{"} \quad . \tag{126}$$

There is no formal restriction on the choice of $f(\mathbf{x})$. This remains true even if we specialize, further, to the near-local *canonical* models that have been the main focus of interest in the literature. Hamiltonian balanced models with arbitrary $f(\mathbf{x})$ and canonical coordinates $\mathbf{X}$ can always be obtained. This is essentially because the canonical coordinate theorems (96) and (105) allow the canonical coordinates $\mathbf{X}$ to have any dependence on $\mathbf{x}$. Inspection of (105), for instance, suggests trying

$$\mathbf{X} = \boldsymbol{\chi}(\mathbf{x}) + \text{terms involving } h(\mathbf{x}) \text{ and its derivatives}\;, \tag{127}$$

for some function $\boldsymbol{\chi}(\mathbf{x})$ of $\mathbf{x}$ alone. This can indeed give any desired $f(\mathbf{x})$. The quickest way to show it is to recall that (98) is the curl of (96) or (105), so that (127) implies

$$\zeta^{\mathrm{C}} = \frac{\partial(\chi_1, \chi_2)}{\partial(x_1, x_2)} + \text{terms involving } h(\mathbf{x}) \text{ and its derivatives} . \tag{128}$$

The first term on the right can be made equal to any desired $f(\mathbf{x})$ by taking, for instance,

$$\chi_i = \frac{\partial \Phi}{\partial x_i} \tag{129}$$

where $\Phi(\mathbf{x})$ is some scalar function of $\mathbf{x}$ alone that satisfies

$$\det \mathrm{Hes}\, \Phi \equiv \det \left(\frac{\partial^2 \Phi}{\partial x_i \partial x_j} \right) = f(\mathbf{x}) \ . \tag{130}$$

This is a Monge-Ampère equation for Φ given $f(\mathbf{x})$. It is known such equations have smooth solutions; these can be used in (129) then (127). Notice that with non-solid $\mathbf{r}(\mathbf{x})$, the condition (122) no longer has a completely clear physical meaning, and the definition of 'the' Coriolis acceleration must to some extent become a matter of convention (cf. Salmon 1985).

Recalling the remark above (97) about Shutts' spherical-coordinate theory and the independence of physical-domain metrics, we may claim that the foregoing results give (we believe for the first time) a clear general answer to the question of how to construct Hamiltonian balanced models having both canonical coordinates and variable Coriolis parameter, in any physical-domain coordinate system whatever.

11 Semigeostrophy is doubly split

By a doubly split Hamiltonian balanced model we mean one whose relation to the parent dynamics is defined by two slow manifolds $\mathcal{M}^{\mathrm{C}(\Omega)}$ and $\mathcal{M}^{\mathrm{C}(\mathsf{H})}$, say, corresponding respectively to two constraint functionals

$$\mathbf{u} = \mathbf{u}^{\mathrm{C}(\Omega)} \qquad \text{and} \qquad \mathbf{u} = \mathbf{u}^{\mathrm{C}(\mathsf{H})} \ , \qquad \text{say.} \tag{131}$$

The doubly split model is defined by constraining the symplectic or contact structure to $\mathcal{M}^{\mathrm{C}(\Omega)}$ and the Hamiltonian functional to $\mathcal{M}^{\mathrm{C}(\mathsf{H})}$. That is, the constrained Hamiltonian functional will be $\mathsf{V} + \int_{\mathcal{D}} \frac{1}{2}|\mathbf{u}^{\mathrm{C}(\mathsf{H})}|^2 \mathrm{d}m$. The two slow manifolds will have be close to each other, of course, if the resulting balanced model is to be accurate. The conserved PV is now, evidently, the Rossby–Ertel PV evaluated with $\mathbf{u}^{\mathrm{C}(\Omega)}$, whereas the conserved energy and conservable physical momenta are just the parent energy and momenta evaluated with $\mathbf{u}^{\mathrm{C}}$.

Another way to think of such a model is to imagine changing the kinetic energy term $\frac{1}{2}|\mathbf{u}|^2$ in the parent Hamiltonian H (e.g., equation (55)), by making the replacement

$$\tfrac{1}{2}|\mathbf{u}|^2 \to \tfrac{1}{2}|\mathbf{u} - \mathbf{u}^{\mathrm{C}(\Omega)} + \mathbf{u}^{\mathrm{C}(\mathsf{H})}|^2 \ , \tag{132}$$

and then constraining the new Hamiltonian functional that results, call it H', say, to $\mathcal{M}^{\mathrm{C}(\Omega)}$ along with the symplectic or contact structure. Note that H' is admissible as a Hamiltonian functional, because in (132) both $\mathbf{u}^{\mathrm{C}(\Omega)}$ and $\mathbf{u}^{\mathrm{C}(\mathsf{H})}$ are functionals of the mass configuration.

It is now plain that Hoskins' semigeostrophy is exactly a case of this. Here, $\mathcal{M}^{\mathrm{C}(\Omega)}$ is the manifold defined by (4) with $\alpha = \frac{1}{2}$. In an inertial frame of reference, this is

$$\mathbf{u} = \mathbf{u}^{\mathrm{C}(\Omega)} = \tfrac{1}{2} f \hat{\mathbf{z}} \times \mathbf{x} \ + \ \mathbf{u}^{\mathrm{G(rel)}} - \tfrac{1}{2} \hat{\mathbf{z}} \times f^{-1} \mathbf{u}^{\mathrm{G(rel)}} . \nabla \mathbf{u}^{\mathrm{G(rel)}} \ , \tag{133}$$

the constraint functional discovered by Salmon (1988*a*). The other slow manifold $\mathcal{M}^{\mathrm{C(H)}}$ is defined simply by

$$\mathbf{u} = \mathbf{u}^{\mathrm{C(H)}} = \tfrac{1}{2} f \hat{\mathbf{z}} \times \mathbf{x} \ + \ \mathbf{u}^{\mathrm{G(rel)}} \ , \tag{134}$$

the geostrophic velocity viewed in an inertial frame. This is one way of making sense of the fact that Hoskins' PV, (114) with $\alpha = \frac{1}{2}$, is not simply the geostrophic PV whereas Hoskins' conserved energy is, by contrast, the geostrophic energy.

It should be noted that the simple forms (3) and (60) of the fundamental equation apply to the case of a single slow manifold only, that is, to singly-split balanced models only. This is because, as pointed out below equation (55) and in (82)ff., the simplicity depends on having the standard kinetic energy term in the parent Hamiltonian. In doubly-split cases the corresponding results are generally more complicated, though semigeostrophy itself has remarkable compensating simplicities, dependent on the gradient form of $\mathbf{X}$ and on a judicious choice of the functional $\mathbf{u}^{\mathrm{C}(\Omega)}$. Whether simplifications can be found as a tradeoff for double splitting in more accurate problems remains to be seen; we already know from Appendix B that any such simplifications will have to be of a different kind than those of semigeostrophy.

Double splitting does not affect PV conservation and the associated circulation theorems, which depend only on the symplectic or contact structure on $\mathcal{M}^{\mathrm{C}(\Omega)}$. The essential points are (a) that any flow advecting a geometric structure in phase space conserves intersection properties; (b) any such flow therefore conserves the value, Q say, of the symplectic two-form Ω^{C} on any pair of variations, i.e. pair of tangent vectors $\delta^{\mathrm{L}}\mathbf{x}^{(1)}(\mathbf{a}), \delta^{\mathrm{L}}\mathbf{x}^{(2)}(\mathbf{a})$ in phase space, when both the vectors and the two-form are advected by the phase-space flow (recall caption to figure 1), (c) that if the flow is Hamiltonian (for any Hamiltonian functional) then Ω^{C} itself is invariant (its Lie derivative vanishes), and can therefore be considered *not* to be advected and to remain equal to the prescribed Ω^{C} of the balanced model, the corresponding Q still being a constant of the motion provided that $\delta^{\mathrm{L}}\mathbf{x}^{(1)}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}^{(2)}(\mathbf{a})$ are still advected, and (d) that Q is equal to an arbitrarily weighted material domain integral of the PV, Q^{C}, when $\delta^{\mathrm{L}}\mathbf{x}^{(1)}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}^{(2)}(\mathbf{a})$ are chosen such that all Eulerian variations vanish, $\delta^{\mathrm{E}} = 0$ in (49), as the particle-relabelling symmetry makes possible. To show the arbitrariness it is enough to consider a pair $\delta^{\mathrm{L}}\mathbf{x}^{(1)}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}^{(2)}(\mathbf{a})$ in which a small disk or ring of particles within $\mathcal{D}$ undergoes small angular displacements, the whole disk in turn being much smaller in diameter than all spatial scales of the fluid flow: $\delta^{\mathrm{L}}\mathbf{x}^{(1)}(\mathbf{a})$ and $\delta^{\mathrm{L}}\mathbf{x}^{(2)}(\mathbf{a})$ are made distinct, giving nonvanishing Q, by starting the angular displacements from two finitely different angular positions. In the stratified case the disk must, of course, lie in a stratification surface $s = \sigma/\rho =$ constant, so as not to violate the particle-relabelling symmetry.

The corresponding Kelvin and Kelvin–Bjerknes circulation theorems follow by standard manipulations, for simply connected $\mathcal{D}$: for a finite closed curve in $\mathcal{D}$ moving with velocity $\mathbf{u}^{\mathrm{P}}$, i.e. a material curve, the absolute circulation

$$\Gamma = \oint \mathbf{u}^{\mathrm{C}} \,.\, \mathrm{d}\boldsymbol{\ell} = \oint \mathbf{u}^{\mathrm{C}(\Omega)} \,.\, \mathrm{d}\boldsymbol{\ell} = \text{constant} \;, \tag{135}$$

where $\mathrm{d}\boldsymbol{\ell}$ is an element of the curve. In the stratified case, as V. Bjerknes showed, the curve has to be taken to lie in a stratification surface $\sigma/\rho = $ constant. Again, this holds for evolution under any Hamiltonian, being a consequence solely of the symplectic or contact structure on $\mathcal{M}^{\mathrm{C}(\Omega)}$, so that the integrand involves $\mathbf{u}^{\mathrm{C}} = \mathbf{u}^{\mathrm{C}(\Omega)}$ and not $\mathbf{u}^{\mathrm{P}}$.

If the material curve cannot be shrunk to a point without encountering fluid boundaries, as is possible in a domain $\mathcal{D}$ that is not simply connected, then (135) is easiest to establish from the relative integral-invariant associated with the one-form $\theta^{\mathrm{C}} = -\int_{\mathcal{D}} u_i^{\mathrm{C}} \delta^{\mathrm{L}} x_i \mathrm{d}m$, replacing the small ring of particles by a finite circuit, and using a single $\delta^{\mathrm{L}}\mathbf{x}(\mathbf{a})$. An equivalent argument is to be published in a forthcoming paper by Holm (1996).

12 Concluding remarks

The theory of Hamiltonian balanced models in the form presented here has clarified several longstanding questions. One such question has been the extent to which the mathematical structure associated with semigeostrophy generalizes to a wider class of models. The new canonical coordinate theorems (96) and (105), the latter defining what we called the class of 'near-local canonical models', help to answer this question. They give insight by showing, for instance, exactly what class of $\mathbf{u}^{\mathrm{C}}$ functionals lead to Hamiltonian balanced models possessing near-local, canonical coordinates $\mathbf{X}$ like those in Hoskins' theory, in the sense of $\mathbf{X}$ being related, by explicit formulae, to the physical-space coordinates $\mathbf{x}$ and mass configuration, and being two-dimensional or layerwise two-dimensional in the physical domain $\mathcal{D}$. The same theorems show, furthermore, that for any such model the constraint vorticity and the corresponding materially conserved PV can always be expressed in the Jacobian forms (98)ff., paralleling the Jacobian forms discovered by Hoskins (1975) for semigeostrophy and their variants summarized by Chynoweth & Sewell (1991).

The new canonical coordinate theorems (96) and (105) also answer a longstanding question about the Monge–Ampère equation and Legendre-duality properties of semigeostrophy. There is a wide class of models having parallel properties (for details see Appendix B), but the results of §9 strongly suggest, and Appendix B confirms, that the necessary gradient form of $\mathbf{X}$, such as that defined in (6) with $\gamma = 0$, cannot lead to a balance condition more accurate than semigeostrophy.

There has been a deafening silence in this paper about that excellent old workhorse, quasigeostrophy à la Charney and Obukhov. The reason is that, as began to emerge from the work of Salmon (1988*a*), quasigeostrophy is an isolated case in the sense that it does not belong to the class of Hamiltonian balanced models considered here. Though it has its own Hamiltonian structure (e.g. Weinstein 1983, R. Salmon and V. Zeitlin, personal communication), that structure is not inherited from any exact parent, at least not in the simple and straightforward sense discussed here. There is certainly no simple, explicitly prescribable $\mathbf{u}^{\mathrm{C}}$ functional that, when substituted into the Rossby–Ertel formula for PV, will produce the conserved quasi-PV of quasigeostrophic theory. Consequently there is no natural systematic procedure for constructing sequences of higher approximations, with guaranteed Hamiltonian structure, that start from quasigeostrophy and approach an exact parent dynamics.

In the examples studied so far, which are all 'physically reasonable' balanced models in that they all have $\mathbf{u}^{\mathrm{C}}$ close to geostrophic, the $\boldsymbol{\zeta}^{\mathrm{C}} \times \mathbf{u}^{\mathrm{S}}$ contribution in (60) tends to reinforce the mass-rearrangement contribution involving ω_{ij}^{C}, indicating robust invertibility as already emphasized, and thence a well-defined velocity-split $\mathbf{u}^{\mathrm{S}}$ and associated dynamics. We may regard $\mathbf{u}^{\mathrm{C}}$ and $\mathbf{u}^{\mathrm{S}}$ as natural generalizations of the 'geostrophic' and 'ageostrophic' velocity fields that arise in simple balanced models. Under the usual parameter conditions, scaling analysis on (60) shows that the magnitude of $\mathbf{u}^{\mathrm{S}}$ is one order smaller in Rossby number Ro than the magnitude of $\mathbf{u}^{\mathrm{C}}$. Moreover, adding $\mathbf{u}^{\mathrm{S}}$ to $\mathbf{u}^{\mathrm{C}}$ to give $\mathbf{u}^{\mathrm{P}}$ tends to improve on the accuracy of $\mathbf{u}^{\mathrm{C}}$ though not, in general, to correct it fully to the next Rossby-number order (the case of L_0 dynamics being an exception). All this was explicitly illustrated in §§4–7.

Equation (60) implies that the dependence of $\mathbf{u}^{\mathrm{P}}$ on the mass configuration respects the particle labelling symmetry, hence qualifies $\mathbf{u}^{\mathrm{P}}$ as the $\mathbf{u}^{\mathrm{C}}$ of a further Hamiltonian balanced model. It follows that, formally at least, there is an iterative sequence of such models in which the $\mathbf{u}^{\mathrm{P}}$ of one member is taken to be the $\mathbf{u}^{\mathrm{C}}$ of the next. The first two members are Salmon's L_0 dynamics and L_1 dynamics (§4). We conjecture, from the examples looked at so far, that $||\mathbf{R}||$ and $||\mathbf{u}^{\mathrm{S}}||$ will decrease like a geometric sequence, on iteration, until the accuracy approaches the limits set by Lighthill radiation.

Alternatively, given a $\mathbf{u}^{\mathrm{C}}$ functional that already has such accuracy — and $\mathbf{u}^{\mathrm{C}}$ functionals seemingly approaching such accuracy have already been found from studies of shallow water PV inversion, (e.g. Norton 1988, McIntyre and Norton 1990*a*, *b*, 1996) — one can insert such a $\mathbf{u}^{\mathrm{C}}$ straight into (60) or (84) to obtain a Hamiltonian balanced model whose error is of the same order in Rossby number Ro, and probably smaller numerically. It is a separate question, far from being answered, whether the best accuracy by comparison with the parent dynamics, in a given norm, will be attained by Hamiltonian or by non-Hamiltonian balanced models or, if by Hamiltonian models, whether singly or doubly split (§11). It has recently been found that what is arguably the most accurate PV-conserving balanced model yet to be discovered is both non-Hamiltonian (because its particle velocity is defined to be the same as the velocity that enters the model's Rossby–Ertel PV) and also violates local mass conservation (McIntyre and Norton 1996; S. Ren, personal communication). There is an intuitive argument that rationalizes this result in terms of the dynamics of Lighthill radiation. It says that, to the extent that such radiation is associated with a continual, spontaneous local adjustment towards balance (a notion whose precise meaning is not obvious, because of the fuzziness of the parent slow quasi-manifold) one might expect a very accurate balanced model to try to mimic this situation by letting mass rearrange itself nonlocally. One might also say that velocity splitting is the way in which a Hamiltonian balanced model copes, so to speak, with the same problem. For instance one could insert into the present Hamiltonian framework the same highly accurate balance condition as in the abovementioned non-Hamiltonian model, thereby defining a model whose $\mathbf{u}^{\mathrm{P}}$ is, by construction, exactly compatible with local mass conservation, but differs by $\mathbf{u}^{\mathrm{S}}$ from the velocity $\mathbf{u}^{\mathrm{C}}$ that enters the model's Rossby–Ertel PV.

Which way of 'coping' has the greatest potential for accuracy, near the Lighthill-radiation limits, is very far from obvious, involving as it does nontrivial mathematical questions about the presumably chaotic nature of the parent slow quasi-manifold. It is not enough to point out that Hamiltonian balanced models conserve energy. Inertia–gravity waves affect the energy and momentum budgets of the parent dynamics, making it conceivable that the suppression of Lighthill radiation while maintaining the best possible accuracy might require departures from energy and momentum conservation.

For Hamiltonian balanced models, perhaps the grandest mathematical challenge of all is the problem of minimizing $||\mathbf{R}||$ or $||\mathbf{u}^{\mathrm{S}}||$ over $\mathbf{u}^{\mathrm{C}}$ functionals. Can sharp estimates of minimal $||\mathbf{R}||$ or $||\mathbf{u}^{\mathrm{S}}||$ be found? Can ingenuity with the mathematics of large function spaces find a way to home in on the actual minimum or minima, i.e., to find an optimal $\mathbf{u}^{\mathrm{C}}$ functional or functionals? One would have to restrict the set of mass configurations that are taken as the domain of $\mathbf{u}^{\mathrm{C}}$, in such a way that the $\mathbf{u}^{\mathrm{C}}$ functional belongs to the realm of physically reasonable balance conditions, for instance imposing bounds on the permitted range of ζ^{C}/f. At present, such questions, to our knowledge, are very far from being answered.

We thank O. Bokhove, M. J. P. Cullen, R. Ford, S. Kuksin, R.S. MacKay, M. T. Montgomery, P. J. Morrison, J. Norbury, M. J. Sewell, T. G. Shepherd, J. Theiss, T. Warn, and V. Zeitlin for helpful comments or correspondence, and especially J. Allen, D. Holm and R. Salmon for kindly showing us their unpublished results on closely related problems, and R. Temam for providing the existence proof referred to at the end of Appendix A. The present work was supported by the Isaac Newton Institute through a Programme on the Mathematics of Atmosphere and Ocean Dynamics, by the UK Meteorological Office, by the Engineering and Physical Sciences Research Council through a Senior Research Fellowship to MEM and by the Natural Environment Research Council through the UK Universities' Global Atmospheric Modelling Programme.

A Uniqueness theorems for L_1 dynamics and filtered L_1 dynamics

Basic lemma: Consider a rigidly bounded domain $\mathcal{D}$ in the xy plane. For any domain function

$\psi(\mathbf{x})$, define

$$P(\psi(\cdot)) = \int_{\mathcal{D}} (|\nabla\psi|^2 + K^2(\mathbf{x})\psi^2)\,\mathrm{d}\mathbf{x}\ , \tag{136}$$

and note that $P(\psi(\cdot))$ is non-negative if $K^2(\mathbf{x}) > 0$. Then, with n the outward normal coordinate and ℓ the tangential co-ordinate,

$$(\nabla^2 - K^2)\psi = 0 \qquad \Rightarrow \qquad P(\psi(\cdot)) = \oint_{\partial\mathcal{D}} \psi \frac{\partial\psi}{\partial n}\mathrm{d}\ell\ , \tag{137}$$

where $\partial\mathcal{D}$ is the boundary of $\mathcal{D}$. This will be applied to several choices of $\psi(\mathbf{x})$, and we assume throughout that ψ is smooth enough that $P(\psi(\cdot)) = 0$ implies $\psi \equiv 0$.

1. Uniqueness theorem for conventional L_1 dynamics: Given $h(\mathbf{x})$ hence $\mathbf{u}^{\mathrm{C}} = \mathbf{u}^{\mathrm{G(rel)}}$ and hence $\mathbf{R}$, we need (at each timestep) to solve (77), i.e.

$$(\nabla^2 - K^2)\mathbf{S} = \mathbf{R}\ , \tag{138}$$

for the velocity-split $\mathbf{S}$, with both $\mathbf{S}\,.\,\hat{\boldsymbol{\ell}}$ and $\mathbf{S}\,.\,\hat{\mathbf{n}}$ (tangential and normal components) prescribed on $\partial\mathcal{D}$, respectively 0 and $-\mathbf{u}^{\mathrm{G(rel)}}.\hat{\mathbf{n}}$ (known). Consider, in the standard way, the difference problem, whose solution must be shown to be identically zero for uniqueness [note, this standard technique will apply to the $\widetilde{\mathbf{U}}^{\mathrm{S}}$ problem for *any* Hamiltonian balanced model because the $\mathbf{S}$ problem is always linear, by the fundamental equation (60)]. Denoting by $\widetilde{\mathbf{U}}^{\mathrm{S}}$ the difference between solutions of this boundary-value problem, we have

$$(\nabla^2 - K^2)\widetilde{\mathbf{U}}^{\mathrm{S}} = 0\ ; \tag{139}$$

$$\widetilde{\mathbf{U}}^{\mathrm{S}}.\hat{\boldsymbol{\ell}} = \widetilde{\mathbf{U}}^{\mathrm{S}}.\hat{\mathbf{n}} = 0\ , \quad \Rightarrow \widetilde{\mathbf{U}}^{\mathrm{S}} = 0, \quad \text{on} \quad \partial\mathcal{D}\ . \tag{140}$$

Apply (137) to with ψ taken as x-component $\widetilde{U}^{\mathrm{S}}_1$ of $\widetilde{\mathbf{U}}^{\mathrm{S}}$, $\quad \Rightarrow \widetilde{U}^{\mathrm{S}}_1 \equiv 0 \quad \forall \mathbf{x} \in \mathcal{D}$. Similarly, the y-component $\widetilde{U}^{\mathrm{S}}_2 \equiv 0$.

2. Uniqueness theorem for filtered L_1 dynamics: We have the same equation (138), but the boundary conditions are now less simple:

$$\widetilde{\mathbf{U}}^{\mathrm{S}}.\hat{\mathbf{n}} = 0 \quad \text{on} \quad \partial\mathcal{D} \tag{141}$$

because $\mathbf{u}^{\mathrm{G(rel)}}.\mathbf{n} = 0$ on $\partial\mathcal{D}$, and

$$\frac{\partial}{\partial\ell}\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}} = 0 \quad \text{on} \quad \partial\mathcal{D}\ , \tag{142}$$

because

$$\frac{\partial h}{\partial\ell} = 0 \qquad \text{hence} \qquad \frac{\partial^2 h}{\partial\ell\partial t} = \frac{\partial}{\partial\ell}(\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}) = 0 \qquad \text{on} \quad \partial\mathcal{D}, \tag{143}$$

and

$$\oint_{\partial\mathcal{D}_i} \widetilde{\mathbf{U}}^{\mathrm{S}}\ .\ \mathrm{d}\boldsymbol{\ell} = 0 \tag{144}$$

separately on each connected part $\partial\mathcal{D}_i$ of $\partial\mathcal{D}$, each of which are closed curves.

Take the divergence of (139) and apply (137) with $\psi = \nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}$:

$$P(\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}(\cdot)) = \oint\limits_{\partial\mathcal{D}} \nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}\, \frac{\partial\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}}{\partial n}\, \mathrm{d}\ell = \sum_i (\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}})_i \oint\limits_{\partial\mathcal{D}_i} \frac{\partial\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}}{\partial n}\, \mathrm{d}\ell \ , \tag{145}$$

with $(\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}})_i$ the value of $\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}$ on the i^{th} connected part $\partial\mathcal{D}_i$ of $\partial\mathcal{D}$, whose ℓ-independence is implied by (142). Now the Helmholtz decomposition consistent with the boundary condition (141) enables us to write

$$\widetilde{\mathbf{U}}^{\mathrm{S}} = \nabla\Phi \ + \ \hat{\mathbf{z}} \times \nabla\Psi \tag{146}$$

for some smooth, single-valued domain functions Φ, Ψ with

$$\partial\Phi/\partial n = 0 = \partial\Psi/\partial\ell \quad \text{on} \quad \partial\mathcal{D} \ , \tag{147}$$

in terms of which (139) becomes

$$\nabla(\nabla^2 - K^2)\Phi = -\hat{\mathbf{z}} \times \nabla(\nabla^2 - K^2)\Psi \ . \tag{148}$$

Together with (141) or (147) this implies, with (n, ℓ) right-handed, that

$$\frac{\partial}{\partial n}\nabla^2\Phi = \frac{\partial}{\partial\ell}\nabla^2\Psi \quad \text{on} \quad \partial\mathcal{D} \ . \tag{149}$$

Because $\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}} = \nabla^2\Phi$, (145) now becomes

$$P(\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}}(\cdot)) = \sum_i (\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}})_i \oint\limits_{\partial\mathcal{D}_i} \frac{\partial}{\partial\ell}\nabla^2\Psi\, \mathrm{d}\ell = 0 \ ,$$

because $\nabla^2\Psi$ is single-valued. Therefore

$$\nabla.\widetilde{\mathbf{U}}^{\mathrm{S}} \equiv 0 \qquad \forall\mathbf{x} \in \mathcal{D} \ . \tag{150}$$

So we may take $\nabla\Phi = 0$ in (146), whereupon (148) implies that

$$\nabla(\nabla^2 - K^2)\Psi = 0$$

and therefore that

$$(\nabla^2 - K^2)\Psi = \ \mathrm{func}(t) = 0 \ , \tag{151}$$

if we use the freedom to replace Ψ by $(\Psi - \ \mathrm{func}(t)/K^2)$, which is permissible as long as we allow Ψ to be a (thus far unrestricted) function of time t on each connected part $\partial\mathcal{D}_i$ of $\partial\mathcal{D}$, consistent with (147). Now applying this to (137) with $\psi = \Psi$, (141) with and then using (144), we see that

$$P(\Psi(\cdot)) = \sum_i \oint\limits_{\partial\mathcal{D}_i} \Psi\frac{\partial\Psi}{\partial n}\mathrm{d}\ell = \sum_i \Psi_i \oint\limits_{\partial\mathcal{D}_i} \frac{\partial\Psi}{\partial n}\mathrm{d}\ell = 0 \ , \tag{152}$$

with Ψ_i the value of Ψ on $\partial\mathcal{D}_i$, hence that

$$\Psi \equiv 0, \qquad \widetilde{\mathbf{U}}^{\mathrm{S}} \equiv 0 \quad \forall\mathbf{x} \in \mathcal{D} \ . \tag{153}$$

This verifies that the boundary conditions (141-144) for filtered L_1 dynamics in a rigidly bounded domain, singly or multiply connected, are indeed sufficient to ensure a unique solution for $\mathbf{S}(\mathbf{x})$, given $h(\mathbf{x})$.

It can further be shown, using the foregoing manipulations together with standard functional-analytic methods that, in both the above cases, L_1 dynamics and filtered L_1 dynamics, weak solutions to (138) exist in a certain Hilbert space with norm

$$\|\psi(\cdot)\| = \left(\int_{\mathcal{D}} |\nabla\psi|^2 \, \mathrm{d}\mathbf{x} \right)^{1/2} \tag{154}$$

(R. Temam, personal communication). The methods used are analogous to those described in Temam (1984), and in the case of filtered L_1 dynamics the function space is like the standard space $\mathbf{H}_0^1(\mathcal{D}) = H_0^1(\mathcal{D}) \otimes H_0^1(\mathcal{D})$ apart from the different boundary conditions involved.

B Generalized Legendre duality properties of a subclass of near-local canonical models

Consider now the class of near-local canonical models whose canonical coordinates $\mathbf{X}$ are semigeostrophy-like in a generalized sense, in that they have gradient form. One example of this is canonical coordinates $X_i(x_j(\mathbf{a}))$ that may be defined by a Legendre transformation in phase space,

$$X_i = \frac{\delta^{\mathrm{L}}\Upsilon}{\delta^{\mathrm{L}} x_i(\mathbf{a})} \ , \tag{155}$$

either generally, as in (96) and (98), or alternatively where near-local form is assumed,

$$\Upsilon[\mathbf{x}, h] = \int_{\mathcal{D}} \mathrm{d}m \ f(x_j, h, h_{,i}, h_{,ij}, \cdots) \ . \tag{156}$$

From (47)–(53) of §3, we have

$$X_i = \frac{\delta^{\mathrm{L}}\Upsilon}{\delta^{\mathrm{L}} x_i(\mathbf{a})} = \frac{\partial\Lambda}{\partial x_i} \ , \tag{157}$$

where

$$\Lambda = f + h f_{,h} - \frac{\partial}{\partial x_i}\left(h f_{,h_{,i}}\right) - \cdots \ . \tag{158}$$

From here on, we consider general near-local Λ, including, but not restricted to, the above case (158) derived from the phase-space functional Υ.

We now substitute $\mathbf{X}$ defined by (157) into the right-hand side of (96). With the use of (157) and the fact that $\partial/\partial x_j$ and δ^{E} commute, the integrand of the middle term on the right of (95) becomes

$$\tfrac{1}{2}\varepsilon_{jk}\frac{\partial\Lambda}{\partial x_k}\delta^{\mathrm{L}}\left(\frac{\partial\Lambda}{\partial x_j}\right) = \tfrac{1}{2}\varepsilon_{jk}\frac{\partial\Lambda}{\partial x_k}\left(\delta^{\mathrm{L}} x_i \frac{\partial^2\Lambda}{\partial x_i \partial x_j} + \frac{\partial\delta^{\mathrm{E}}\Lambda}{\partial x_j}\right). \tag{159}$$

Calculation of the Eulerian variations is performed in exactly the same way as was done in obtaining (105), and we obtain, after including in equation (159) the analogue of the exact one-form $\delta^{\mathrm{L}}\int_{\mathcal{D}}\mathrm{d}\mathbf{a}\ B(x_j, h, \cdots)$, the following relation between $\mathbf{u}^{\mathrm{C}}$ and Λ:

$$u_i^{\mathrm{C}} = -\tfrac{1}{2} f_0 \varepsilon_{jk}\frac{\partial\Lambda}{\partial x_k}\frac{\partial^2\Lambda}{\partial x_i \partial x_j} +$$

$$+\frac{\partial}{\partial x_i}\left[\tfrac{1}{2}f_0\varepsilon_{jk}\left\{\left(\Lambda_{,h}\frac{\partial}{\partial x_j}(h\Lambda_{,k})\right)-\frac{\partial}{\partial x_p}\left(\Lambda_{,h_{,p}}\frac{\partial}{\partial x_j}(h\Lambda_{,k})\right)+\frac{\partial^2}{\partial x_p\partial x_q}\left(\Lambda_{,h_{,pq}}\frac{\partial}{\partial x_j}(h\Lambda_{,k})\right)-\cdots\right\}\right.$$
$$\left.+\left(B+h\frac{\partial B}{\partial h}-\frac{\partial}{\partial x_j}\left(h\frac{\partial B}{\partial h_{,j}}\right)+\frac{\partial^2}{\partial x_j\partial x_k}\left(h\frac{\partial^2 B}{\partial h_{,jk}}\right)-\cdots\right)\right]. \tag{160}$$

This of course is a special case of (105), though its derivation via that route is a little longer.

The connection with Monge–Ampère equations is now evident. Note that $\zeta^{\mathrm{C}}=\nabla\times\mathbf{u}^{\mathrm{C}}$ involves only the first term on the right of (160), and it can be shown that the additional terms obtained by working through the calculation when Λ depends on the higher-order derivatives of h, do not change the structure given above, i.e.

$$u_i^{\mathrm{C}}=-\tfrac{1}{2}\varepsilon_{jk}\left[\frac{\partial\Lambda}{\partial x_k}\frac{\partial^2\Lambda}{\partial x_i\partial x_j}-\frac{\partial}{\partial x_i}F(\nabla\Lambda,\nabla\nabla\Lambda,\nabla\nabla\nabla\Lambda,\nabla\nabla\nabla\nabla\Lambda,\cdots)\right].$$

Therefore, for the class of functionals Υ considered here, we have a generalized Monge–Ampère equation for Λ given ζ^{C}

$$\det\mathrm{Hes}(\Lambda)=\det\left(\frac{\partial^2\Lambda}{\partial x_i\partial x_j}\right)=\zeta^{\mathrm{C}}(x_j,h,h_{,i},\cdots)\,. \tag{161}$$

However, these 'Υ-models' and 'Λ-models' cannot have a higher order of accuracy in Ro than semigeostrophy, for the following reason. Semigeostrophy is the case

$$\Lambda=\tfrac{1}{2}(x_1^2+x_2^2)+gh/f^2\ , \tag{162}$$

already accurate to one order in Ro . No coefficient on the right of this expression can be changed without making the accuracy worse than one order. So the only hope of reaching higher accuracy is to add terms involving the first or higher spatial derivatives of h. But any such term evidently produces third and higher derivatives of h that cannot cancel to zero on the left of (161). Consider for instance a term proportional to $\partial h/\partial x_1$ added to (162). Then there will be one and only one term in $\partial^3 h/\partial^3 x_1$, coming from $(\partial^2\Lambda/\partial^2 x_1)(\partial^2\Lambda/\partial^2 x_2)$, in (161). This term cannot cancel any other term in (161). Thus none of the new terms can take the form $\det\mathrm{Hes}(h)$, which contains only second derivatives. Thus none of the new terms can produce agreement with (114) when there is a coefficient -2 in front of its final Jacobian term, instead of the coefficient $+1$ that results from the unmodified (161).

We note for completeness that semigeostrophy is a special case of the foregoing, though with the modifications described in §11. In semigeostrophy, with Λ given by (162), (157) and (158) imply

$$\Upsilon[\mathbf{x},h]=\tfrac{1}{2}\int_{\mathcal{D}}\mathrm{d}m\left[\frac{gh}{f^2}+x_1^2+x_2^2\right],$$

so $\partial\Lambda/\partial h_{,i}=0$ etc., and from (160)

$$u_i^{\mathrm{C}}=-\tfrac{1}{2}f_0\varepsilon_{jk}\left[\frac{\partial\Lambda}{\partial x_k}\frac{\partial^2\Lambda}{\partial x_i\partial x_j}-\frac{\partial}{\partial x_i}\left(\frac{g}{f^2}\frac{\partial}{\partial x_j}(h\Lambda_{,k})\right)\right]. \tag{163}$$

From the first of (106) with $c=0$, and with $f_0=f$ a constant, we have

$$X_1\equiv\frac{\partial\Lambda}{\partial x_1}=x_1+\frac{1}{f}u^{\mathrm{G(rel)}}{}_2,\ \ X_2\equiv\frac{\partial\Lambda}{\partial x_2}=x_2-\frac{1}{f}u^{\mathrm{G(rel)}}{}_1;$$

note that $\partial X_2/\partial x_1 = \partial X_1/\partial x_2$. Using this, we can show, for example, that

$$\begin{aligned} u_1^{\mathrm{C}} &= -\tfrac{1}{2} f \left[X_2 \frac{\partial X_1}{\partial x_1} - X_1 \frac{\partial X_2}{\partial x_1} - x_2 \frac{\partial X_1}{\partial x_1} + X_2 + x_1 \frac{\partial X_2}{\partial x_1} \right], \\ &= -\tfrac{1}{2} \left[f x_2 - 2 u^{\mathrm{G(rel)}}{}_1 - \frac{1}{f} u^{\mathrm{G(rel)}}{}_1 \frac{\partial u^{\mathrm{G(rel)}}{}_2}{\partial x_1} + \frac{1}{f} u^{\mathrm{G(rel)}}{}_2 \frac{\partial u^{\mathrm{G(rel)}}{}_1}{\partial x_1} \right], \\ &= -\tfrac{1}{2} \left[f x_2 - 2 u^{\mathrm{G(rel)}}{}_1 - \frac{1}{f} u^{\mathrm{G(rel)}}{}_1 \frac{\partial u^{\mathrm{G(rel)}}{}_2}{\partial x_1} - \frac{1}{f} u^{\mathrm{G(rel)}}{}_2 \frac{\partial u^{\mathrm{G(rel)}}{}_2}{\partial x_2} \right], \end{aligned}$$

which is precisely the u_1^{C} component of Salmon's constraint (133). Roulstone & Sewell (1996*a*) give further discussion.

REFERENCES

Allen, J. S., Barth, J. A., Newberger, P. A., 1990: On intermediate models for barotropic continental shelf and slope flow fields. Part I: Formulation and comparison of exact solutions. *J. Phys. Oceanog.*, **20**, 1017–1043.

Allen, J. S., Holm, D. D., 1996: Extended-geostrophic Hamiltonian models for rotating shallow water motion. *Physica D*, to appear.

Chynoweth, S., Sewell, M. J., 1991: A concise derivation of the semi-geostrophic equations. *Q. J. Roy. Meteorol. Soc.*, **117**, 1109–1128. See equations (59), (66), (71).

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

Crighton, D. G., 1981: Acoustics as a branch of fluid mechanics. *J. Fluid Mech.*, **106**, 261–298.

Errico, R. M., 1982: Normal mode initialization and the generation of gravity waves by quasi-geostrophic forcing. *J. Atmos. Sci.*, **39**, 573–586.

Ford, R., 1994: Gravity wave generation by vortical flows in a rotating frame. *PhD Thesis* DAMTP, Cambridge University, 269pp.

Ford, R., McIntyre, M. E., Norton, W. A., 1996: Balance and the slow quasi-manifold: some explicit results. *J. Atmos. Sci.*, to be submitted.

Gotay, M. J., Nester, J. M., Hinds, G., 1978: Presymplectic manifolds and the Dirac–Bergmann theory of constraints. *J. Math. Phys.*, **19**, 2388–2399.

Holm, D. D., 1996: Hamiltonian balance equations. *Physica D*, to appear.

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

Hoskins, B. J., Draghici, I., Davies, H. C., 1978: A new look at the ω-equation. *Q. J. Roy.*

Meteorol. Soc., **104**, 31–38.

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. See also **113**, 402–404.

Kambe, T., Minota, T., Murakami, T., Takaoka, M., 1990: Oblique collision of two vortex rings and its acoustic emission. In: Topological Fluid Mechanics, ed. H. K. Moffatt, A. Tsinober; Cambridge University Press, 515–524.

Kambe, T., Minota, T., Takaoka, M., 1993: Oblique collision of two vortex rings and its acoustic emission. *Phys. Rev. E,* **48**, 1866–1881.

McIntyre, M. E., Norton, W. A., 1990*a*: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.*, **212**, 403–435. (Corrigendum 220, 693.) (G. K. Batchelor Festschrift Issue.)

McIntyre, M. E., Norton, W. A., 1990*b*: Nonlinear vorticity or potential vorticity inversion. In: *Topological Fluid Mechanics*, ed. H. K. Moffatt, A. Tsinober; Cambridge University Press, 355–358.

McIntyre, M. E., Norton, W. A., 1996: Potential-vorticity inversion on a hemisphere. *J. Atmos. Sci.*, to be resubmitted.

Misner, C. W., Thorne, K. S., Wheeler, J. A., 1973: Gravitation. San Francisco, Freeman. 1279 pp. See e.g. §2.5.

Morrison, P. J., Greene, J. M., 1980: Non-canonical Hamiltonian density formulation of hydrodynamics and ideal magnetohydrodynamics. *Phys. Rev. Lett.*, **45**, 790–794.

Norton, W. A., 1988: Balance and potential vorticity inversion in atmospheric dynamics. University of Cambridge, *PhD Thesis*, 167 pp.

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

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

Roulstone, I., Sewell, M. J., 1996*b*: The mathematical structure of theories of semi-geostrophic type. *Phil. Trans. Roy. Soc. Lond,* submitted.

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

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

SCHUBERT, W. H., MAGNUSDOTTIR, GUDRUN, 1994: Vorticity coordinates, transformed primitive equations, and a canonical form for balanced models. *J. Atmos. Sci.*, **51**, 3309–3319.

SCHUTZ, B. F., 1980: Geometrical methods of mathematical physics. Cambridge, Cambridge University Press, 250pp.

SEWELL, M. J., ROULSTONE, I., 1994: Families of lift and contact transformations. *Proc. R. Soc. Lond.*, **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.

TEMAM, R., 1984: Navier–Stokes Equations: Theory and Numerical Analysis, 3rd edn. Amsterdam, New York, North-Holland, 526pp. See pp. 24–26.

WARN, T., 1996: Nonlinear balance and quasigeostrophic sets. *Atmos.–Ocean*, in press. (This pioneering paper was written in 1983 and was the first to recognize that the 'slow quasi-manifold' of the parent dynamics is in some sense fuzzy.)

WARN, T., MÉNARD, R., 1986: Nonlinear balance and gravity-inertial wave saturation in a simple atmosperic model. *Tellus*, **38A**, 285–294.

WEINSTEIN, A., 1983: Hamiltonian structure for drift waves and geostrophic flow. *Phys. Fluids*, **26**, 388–390.

WEBSTER, R. B., 1970: Jet noise simulation on shallow water. *J. Fluid Mech.*, **40**, 423–432.

WHITAKER, J. S., 1993: A comparison of primitive and balance equation simulations of baroclinic waves. *J. Atmos. Sci.*, **50**, 1519–1530.

FIGURE CAPTION

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, for the simplest possible case of one particle in a two-dimensional configuration space. In this illustration the constraint has

the special form $p_1 = p_1^{\rm C} \equiv mu_1^{\rm C}$ = constant, $p_2 = p_2^{\rm C} \equiv mu_2^{\rm C}$ = function of x_1 alone, making (29) nonzero and giving invertible $\Omega_{ij}^{\rm C}$ provided that $\partial p_2^{\rm C}/\partial x_1 \neq 0$, i.e. $\mathcal{M}^{\rm C}$ nowhere 'horizontal'. The parent symplectic structure (corresponding to the two-form $dx_i \wedge dp_i = m dx_i \wedge du_i$ and giving rise to the left-hand side of (14) or (28)) consists of two sets of infinitesimal oriented hypertubes (e.g. Misner et al. 1973, Schutz 1980), say 'tubes' for short, one set intersecting only the x_1p_1 plane and the other intersecting only the x_2p_2 plane. 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. In the canonical-coordinate view shown, the parent tubes fill phase space homogeneously and all have the same infinitesimal cross-sectional area. 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' intersecting any specified small parallelogram A (not shown) that is arbitrarily oriented in the four-dimensional phase space, the sign being positive if the tubes' orientation or circulation agrees with that of A. (This signed total, in the continuum limit with A becoming infinitesimal, is proportional to the value of the two-form on the vector-pair 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 (10), as here, with A lying in $\mathcal{M}^{\rm C}$ (giving rise to the two-form $\frac{1}{2}\Omega_{ij}^{\rm C} dx_i \wedge dx_j$ when $\mathbf{x}$ is used, as here, to track position on $\mathcal{M}^{\rm C}$). Invertibility of $\Omega_{ij}^{\rm C}$ says, in this illustration, that when A lies in $\mathcal{M}^{\rm C}$ the signed total number of parent tubes intersecting A is generally nonzero — true here because, with $p_1^{\rm C}$ = constant, the second set of parent tubes does not intersect $\mathcal{M}^{\rm C}$ at all, but false if, for instance, $p_1^{\rm C}$ were changed to $p_1^{\rm C} = \int (\partial p_2^{\rm C}/\partial x_1) {\rm d}x_2$, in which case the two sets of parent tubes give mutually cancelling contributions for any A in $\mathcal{M}^{\rm C}$. When this picture is extended to the infinite-dimensional fluid cases of interest, $\mathcal{M}^{\rm C}$ will be, heuristically speaking, 'steeper' than the figure suggests: particles need not move far to upset near-geostrophic balance when Rossby and Froude numbers are small, ${\rm Ro} \ll 1, {\rm Fr} \ll 1$.