

Exact semi-geostrophic flows in an elliptical ocean basin*

Robert J McCann¹ and Adam M Oberman²

¹ Department of Mathematics, University of Toronto, Toronto, Ontario M5S 3G3, Canada

² Department of Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

E-mail: mccann@math.toronto.edu, aoberman@math.sfu.ca and trokhim@math.toronto.edu

Received 28 September 2003, in final form 11 June 2004

Published 27 July 2004

Online at stacks.iop.org/Non/17/1891

doi:10.1088/0951-7715/17/5/017

Recommended by F Otto

Abstract

A new family of exact solutions is analysed, which models two-dimensional circulations of an ideal fluid in a uniformly rotating elliptical tank, under the semi-geostrophic approximation from meteorology and oceanography. The fluid pressure and stream function remain quadratic functions of space at each instant in time, and their fluctuations are described by a single degree of freedom Hamiltonian system depending on two conserved parameters: domain eccentricity and the constant value of potential vorticity. These parameters determine the presence or absence of periodic orbits with arbitrarily long periods, fixed points of the dynamics, and aperiodic homoclinic orbits linking hyperbolic saddle points. The energy relative to these parameters selects the frequency and direction in which isobars nutate or precess, as well as the steady circulation direction of the fluctuating flow. The canonically conjugate variables are the moment of inertia and angle of inclination of an elliptical inverse-potential-vorticity patch evolving in dual coordinates.

Mathematics Subject Classification: 35J60, 76B60, 76U05, 86A10

(Some figures in this article are in colour only in the electronic version)

Contents

1. Introduction	1892
1.1. Conservation of potential vorticity	1894
1.2. Conservation of energy	1894
1.3. Quadratic-potential ansatz and affine symmetry	1895
1.4. An evolving dual potential-vorticity ellipse	1895

* Appendix and figures prepared by Maxim Trokhimtchouk.

1.5. Non-dimensionalization and the Rossby number	1895
1.6. Cyclonic versus anticyclonic circulation and pressure harmonicity	1896
1.7. Main results	1896
1.8. Global pressure dynamics: <i>wobbling</i> versus <i>rotating</i> pulsations	1901
1.9. Self-consistency of the SG approximation	1903
2. Dual formulation as an active scalar transport problem	1903
2.1. Wasserstein distance as Hamiltonian energy and transportation cost	1904
2.2. Physical and geometric interpretation of the Hamiltonian variables	1905
2.3. Symmetry breaking bifurcation in an optimal transport problem	1905
3. Matrix identities	1905
3.1. Canonical form for symmetric matrices	1905
4. Matrix equations	1906
4.1. Optimal map and transport cost between ellipses	1906
4.2. Dynamics of the free boundary	1909
5. Active scalar transport of an elliptical inverse-potential-vorticity patch	1909
6. Recovering the pressure dynamics (proof of theorem 1.1)	1912
7. Energy landscape and bifurcation structure (proof of theorem 1.4)	1914
Acknowledgments	1917
Appendix A. Well-posed SG circulations in an elliptical ocean (by Maxim Trokhimtchouk)	1917
References	1921

1. Introduction

This paper exploits affine invariance and conservation laws to reveal some exact motions for fluid rotating in an elliptical domain under the semi-geostrophic (SG) theory.

Euler's equation predicts the velocity evolution in a perfect fluid. However, for large-scale flows of the atmosphere and oceans, inertial effects are small, and the Coriolis force due to the Earth's rotation is roughly balanced by pressure gradients. The SG approximation and resulting fluid models (e.g. incompressible SG, shallow water SG, fully compressible SG) in two-dimensional (XY, XZ or YZ) and three-dimensional flat space, or on a sphere, have therefore been proposed to explore the dynamical meteorology and oceanography of this regime. For perspectives on this theory and some recent developments see the works of Bannon, Benamou, Blumen, Brenier, Bretherton, Chynoweth, Cullen, Douglas, Draghici, Eliassen, Gangbo, Hoskins, Magnusdottir, Maroofi, Mawson, McIntyre, Norbury, Purser, Roulstone, Rubtsov, Salmon, Schubert, Sewell, Shutts, Thorpe, Williams and others, described in [10, 14, 15, 23, 33].

This paper presents a family of exact solutions to the two-dimensional incompressible SG equation in a flat elliptical ocean basin; constant depth and strong gravitational stratification may be construed as a rationalization for modelling the flow two-dimensionally. Our solutions display smooth velocities and acceleration which depend linearly on space at each time; in this sense, they are SG analogues of flows discovered by Dirichlet [18], Dedekind [17] and Riemann [37] for self-gravitating fluid ellipsoids [3]; by Kirchhoff [26], Moore and Saffman [31] and Kida [25] for elliptical vorticity patches in a shearing field [28]; by Meacham *et al* [29] in the quasi-geostrophic (QG) cousin [35, 41] of the SG theory; and by Thacker [44], Cushman-Roisin [16] and Rogers [38] in the study of frontal, warm-core eddies using the reduced-gravity, shallow water equations. Most of the latter solutions are built around the quadratic Newtonian potential generated inside a homogeneous ellipsoid. Our new solutions

are built instead on quadratic solutions to the Monge–Ampère equation (2.4), when the domain and its gradient image (2.5) are both ellipses. As recalled below, this is because the Hessian determinant of the stream function (2.4), and not the Laplacian, gives the advected scalar quantity in the active transport (i.e. vorticity) formulation of SG theory [1, 12, 22].

The affine symmetry of the determinant is responsible for the new solutions we describe. These exhibit a rich variety of dynamics due to the nonlinear interaction of rotation with domain geometry. The non-dimensionalized dynamics form a single variable Hamiltonian system governed by two parameters, which correspond roughly to domain eccentricity and Rossby (rotation) number of the flow. As these parameters vary, the system undergoes a sequence of topological bifurcations, displaying behaviour such as stable and unstable fixed points, periodic orbits of widely varying lengths, and (aperiodic) homoclinic orbits linking hyperbolic saddle points. The analysis demonstrates much about the symmetries and the structure of the model. Besides caricaturing a polar ocean, it may serve to enhance our heuristic understanding of SG dynamics and provide benchmarks for computations.

In a coordinate system which rotates with the domain $Y \subset \mathbf{R}^2$ at frequency $\Omega/2\pi$, the Euler equations for a perfect two-dimensional fluid take the form:

$$(\partial_t + \mathbf{v} \cdot \nabla)\mathbf{v} + 2\Omega J\mathbf{v} = -\varrho^{-1}\nabla P \quad \text{in } [0, \infty[\times Y, \quad (1.1)$$

$$\operatorname{div} \mathbf{v} = 0 \quad \text{in } [0, \infty[\times Y, \quad (1.2)$$

$$\mathbf{v} \cdot \mathbf{n}_Y = 0 \quad \text{on } [0, \infty[\times \partial Y. \quad (1.3)$$

Here, $\mathbf{v}(t, \mathbf{y}) \in \mathbf{R}^2$ is the fluid velocity, $P(t, \mathbf{y})$ its pressure (or rather its pressure plus geopotential [41]), and $\varrho > 0$ its (constant) density; $\mathbf{n}_Y(\mathbf{y})$ is the normal to the boundary of the domain, $\mathbf{y} \in Y$, and J is the symplectic matrix (1.4). Although the Coriolis parameter Ω should actually depend on latitude, we hereafter assume that our ocean is small enough to treat Ω as a constant and neglect the Earth's spherical geometry. It then costs no generality to choose units of time and of mass so that $2\Omega = 1 = \varrho$. The no-flux condition (1.3) is imposed at the ocean boundary.

Incompressibility (1.2) allows the velocity $\mathbf{v}(t, \mathbf{y}) = J\nabla Q(t, \mathbf{y})$ to be expressed as the gradient of a stream function Q rotated by 90° . The Euler equations then become a system relating the stream function $Q(t, \mathbf{y})$ to the pressure $P(t, \mathbf{y})$ on the domain $Y \subset \mathbf{R}^2$:

$$\nabla \frac{\partial Q}{\partial t} + (D^2 Q + \mathbf{I})J\nabla Q - J\nabla P = 0 \quad \text{with } J := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (1.4)$$

The no-flux condition (1.3) implies that the stream function is a constant

$$Q(t, \mathbf{y}) = 0 \quad \text{if } \mathbf{y} \in \partial Y, \quad (1.5)$$

on the domain boundary.

Motivated by the limited role of inertia in determining slow, rotating, large-scale flows, one can neglect the acceleration terms entirely to obtain a *geostrophic balance* equation $Q = P$ in place of (1.4). Despite its uses, this approximation is so crude that it makes no dynamical predictions. For evolution problems, it has therefore been proposed to replace the small terms in (1.4) by their geostrophic values $Q \sim P$. This can be undertaken in different ways: compare the SG approximation developed by Eliassen [20] and Hoskins [22],

$$\nabla \frac{\partial P}{\partial t} + (D^2 P + \mathbf{I})J\nabla Q - J\nabla P = 0 \quad (1.6)$$

with its QG elder cousin [4],

$$\nabla \frac{\partial P}{\partial t} + D^2 P J\nabla P + J\nabla Q - J\nabla P = 0. \quad (1.7)$$

Both approximations are valid in regions where the Rossby number

$$R_0(t, \mathbf{y}) := \frac{|\nabla Q(t, \mathbf{y}) - \nabla P(t, \mathbf{y})|}{|\nabla Q(t, \mathbf{y})|} \ll 1 \quad (1.8)$$

is small, which means the particle accelerations are negligible compared to the Coriolis and pressure forces. The main difference between them is that SG (1.6), although harder to solve, has a wider range of asymptotic validity, and preserves the form of the advection operator and boundary conditions (1.5). It therefore remains consistent with the (observed) flow $\mathbf{v} = J\nabla Q$ of fluid parallel to *pressure fronts*, i.e. to discontinuities in ∇P ; (1.6) is known to admit non-smooth solutions, and is employed as a model for the generation and evolution of atmospheric pressure fronts by Hoskins and Bretherton [24], Cullen and Purser [7, 11], and others. Numerical evidence also supports the idea that SG is less turbulent than Euler's equation (1.4)—a feature whose desirability and appropriateness for forecasting is emphasized by Cullen *et al* [8, 14]. QG on the other hand, does not permit fronts to form, and imposes the boundary condition $P(t, \mathbf{y}) = 0$ on ∂Y . However, Constantin *et al* [5] pointed out that QG possesses certain mathematical properties which make it an interesting model for capturing some features of three-dimensional turbulence. It has attracted the attention of Cordoba and Fefferman [6] in this context, though we shall not say more about it here.

1.1. Conservation of potential vorticity

Our analysis is founded on two well-known symmetries of the SG equation (1.6), and the corresponding conservation laws laid out by Hoskins [22]. The first of these is particle relabelling symmetry, which implies that the *potential vorticity* $v^2(t, \mathbf{y})/4 := \det[D^2P + \mathbf{I}]$ remains a constant along each particle's Lagrangian trajectory. This quantity is so named because it plays a role analogous to the ordinary vorticity $\omega(t, \mathbf{y}) = \text{trace}[D^2Q + \mathbf{I}]$ in Euler's equation (1.4). The substitution of determinant for trace indicates why the Monge–Ampère equation eventually takes the place of Poisson's equation in an active scalar transport reformulation (2.2)–(2.5) of the dynamics [1, 13, 22]. Since the convexity of $P(t, \mathbf{y}) + |\mathbf{y}|^2/2$ is equivalent to the positivity of *both* eigenvalues of $D^2P + \mathbf{I}$, conservation of v also implies this convexity is preserved by the flow (1.6). Thus, the Monge–Ampère equation remains elliptic provided it is so initially. Cullen and Purser's *stability criterion* for solutions is just the convexity $D^2P(0, \mathbf{y}) > -\mathbf{I}$; outside of appendix A, we shall concentrate tacitly but exclusively on solutions to the SG equations which satisfy their criterion [11, 36, 43]. Preventing degeneration of this ellipticity is a barrier to proving that general flows remain smooth, but it will not occur in any of the solutions constructed below.

1.2. Conservation of energy

The second symmetry—time translation invariance—combines with the no-flux boundary condition (1.5) to yield global conservation of the SG *energy*:

$$H_{\text{SG}} := \frac{1}{2} \int_Y |\nabla P(t, \mathbf{y})|^2 d^2\mathbf{y}, \quad (1.9)$$

as follows readily from the same computation

$$\frac{dH_{\text{SG}}}{dt} = - \int_Y \text{div} \left[\left(\frac{|\nabla P|^2}{2} + P \right) J\nabla Q \right] d^2\mathbf{y} = 0,$$

that yields the SG Bernoulli law. Hamiltonian formulations of SG based on such energies have been proposed by Salmon [40] and Roulstone and Norbury [39].

1.3. Quadratic-potential ansatz and affine symmetry

A third symmetry, less familiar in fluid mechanics, is the affine invariance of convex functions (and ellipses) which motivates our quadratic ansatz:

$$\begin{aligned} P(t, \mathbf{y}) &= \frac{\mathbf{y}^T \mathbf{P}(t) \mathbf{y}}{2} - \mathbf{p}^T(t) \mathbf{y} - \frac{p(t)}{2}, \\ Q(t, \mathbf{y}) &= \frac{\mathbf{y}^T \mathbf{Q}(t) \mathbf{y}}{2} - \mathbf{q}^T(t) \mathbf{y} - \frac{q(t)}{2}. \end{aligned} \quad (1.10)$$

Here, $\mathbf{P}(t)$ and $\mathbf{Q}(t)$ are time-dependent symmetric 2×2 matrices, $\mathbf{p}(t), \mathbf{q}(t) \in \mathbf{R}^2$ and $p(t), q(t) \in \mathbf{R}$. We shall verify that this leads to an interesting family of special solutions by solving the ODE (1.12) when the fluid domain is an ellipse,

$$Y = \{\mathbf{F}\mathbf{z} \mid |\mathbf{z}| < 1\} \subset \mathbf{R}^2, \quad \mathbf{F} = \begin{pmatrix} \mathbf{f}e^{\varphi/2} & 0 \\ 0 & \mathbf{f}e^{-\varphi/2} \end{pmatrix} \quad (1.11)$$

with area $\pi \mathbf{f}^2$ and aspect ratio e^φ . The no-flux condition (1.5) forces $\mathbf{Q}(t) = q(t)\mathbf{F}^{-2}$ and $\mathbf{q}(t) = \mathbf{0}$. Our ansatz, therefore, constrains the fluid to slosh around on concentric ellipses in Y with a velocity given by solving the 2×2 system of ODEs

$$P'(t) + q(t)(P(t) + \mathbf{I})\mathbf{J}\mathbf{F}^{-2} - \mathbf{J}P(t) = 0 \quad (1.12)$$

for variable speed $q(t)$ and the symmetric pressure matrix $\mathbf{P}(t)$. The subsequent sections are devoted to analysing the behaviour of these four coupled equations; here $\mathbf{p}(t) = \mathbf{0}$ and $p(t) = 0$ cost no generality since $\mathbf{p}(t)$ evolves in a circular motion with period 2π independently of $\mathbf{P}(t)$ and $q(t)$. Note how affine symmetry of the determinant enters crucially here: in Euler's equation (1.4), the same ansatz yields only constant speed solutions since the vorticity $\omega(t, \mathbf{x}) = 2 + q(t)\text{trace}[\mathbf{F}^{-2}]$ is independent of time.

1.4. An evolving dual potential-vorticity ellipse

Under the quadratic ansatz (1.10) the evolving image $X(t) := (\mathbf{P}(t) + \mathbf{I})Y$ of the fluid domain also forms an ellipse with unchanging area, because of conservation of potential vorticity. In section 2 we recall how this ellipse $X(t)$ can be interpreted as an evolving (inverse potential-) vortex patch in \mathbf{R}^2 . Here, let us simply assert that its aspect ratio $a(t) = e^{\sigma(t)}$ and inclination $\theta(t)$ to the coordinate axes are convenient variables for describing the dynamics which follow. In fact $r = \cosh \sigma$ and θ turn out to be canonically conjugate when the problem is cast into a Hamiltonian form (see section 2.2). Conservation of energy implies a relationship between r and θ , which amounts to fixing the Wasserstein distance $W_2(X(t), Y)$ in (2.8) between uniform probability measures on the two ellipses.

1.5. Non-dimensionalization and the Rossby number

In addition to the deep symmetries described above, equation (1.6) enjoys simple scaling symmetries which facilitate the subsequent analysis. The SG theory is invariant under the transformation

$$\begin{aligned} \tilde{P}(t, \mathbf{y}) - \mathbf{y} \cdot \tilde{\mathbf{p}}(t) &= \frac{P(Tt, L\mathbf{y})\tilde{Q}T^2}{QL^2}, \\ \tilde{Q}(t, \mathbf{y}) &= \frac{Q(Tt, L\mathbf{y})T}{L^2}, \\ \tilde{\Omega} &= \Omega T, \end{aligned} \quad (1.13)$$

provided $\tilde{\mathbf{p}}(t)$ rotates with constant angular frequency $2\tilde{\Omega}$. As has been mentioned, it therefore costs no generality to normalize $\nabla P(0, \mathbf{0}) = \mathbf{0}$, and choose units of time so that $2\Omega = 1$, units of mass in which the fluid has density $\varrho = 1$, and presently units of length which normalize the domain area conveniently.

Note however, that there is a dimensionless version

$$\frac{1}{\lambda^2} := \left(\frac{v(t, \mathbf{y})}{4\Omega} \right)^2 = \det \left[\mathbf{I} + \frac{D^2 P(t, \mathbf{y})}{4\Omega^2 \varrho} \right] = \frac{|X(t)|}{|Y|} \quad (1.14)$$

of the potential vorticity which cannot be scaled; it compares the rotation of fluid within the domain to the rotation rate of the domain. In the context of our quadratic ansatz (1.10) and (1.11) the quantity λ is independent of space and hence—by conservation of potential vorticity—-independent of time. When pressure gradients are small, then $1 - \lambda \sim \Delta P/2$ is akin to the *Rossby number* (1.8), so we expect $\lambda \sim 1$ to be a physically relevant parameter regime. It is worth pointing out that the potential vorticity $v(t, \mathbf{y}) = 4\Omega/\lambda$ remains constant throughout $[0, \infty[\times Y$ for our solutions despite the fluctuations of pressure and velocity. To simplify formulae, it is convenient to choose units of distance so that the areas $|Y| = \pi \mathbf{f}^2$ and $|X(t)| = \pi \epsilon^2$ of the two ellipses multiply to π^2 . Then, $\epsilon \mathbf{f} = 1$ and $\lambda = \mathbf{f}/\epsilon = \mathbf{f}^2$ from (1.14); the fluid domain has area $|Y| = \pi \lambda$ and the evolving ellipse area $|X(t)| = \pi/\lambda$.

1.6. Cyclonic versus anticyclonic circulation and pressure harmonicity

Under our quadratic ansatz (1.10) and (1.11), the circulation rate $q(t) = \Delta Q(t, \mathbf{y})/2$ is determined implicitly by pressure fluctuations through (1.12). This relationship is made explicit by Maxim Trokhimtchouk (see appendix A), who deduced how conservation of energy prevents the flow from reversing directions:

$$q(t) = \frac{\lambda^2 \text{trace } P(t)}{\text{trace} [(P(t) + \mathbf{I})F^2]} = 2\lambda \frac{(4/\pi)H_{\text{SG}} - (\lambda^2 - 1) \cosh \varphi}{\text{trace}^2 [(P(t) + \mathbf{I})F^2]}, \quad (1.15)$$

both denominators being positive from Cullen and Purser's convexity criterion $D^2 P > -\mathbf{I}$. Even without this restriction, Trokhimtchouk's formula leads to an elementary direct proof of the existence of unique, global, analytic, solutions to (1.12) under the extra hypothesis $\text{trace} [(P(0) + \mathbf{I})F^2] \neq 0$ of theorem A.1. The circulation direction (1.15) is determined by the invariants (λ, φ) and H_{SG} ; when $\text{trace} [(P(0) + \mathbf{I})F^2] > 0$, it coincides with the sign of $\Delta P(0, \mathbf{0})$: cyclonic (counterclockwise) around a pressure low or subharmonic saddle and anticyclonic otherwise. The fluid remains stagnant $q(t) = 0$ if and only if it is at rest initially, which occurs precisely when the pressure forms a symmetric saddle (with orthogonal arms) at one and hence all times. Trokhimtchouk's result also shows that superharmonicity $\Delta P \leq 0$ is consistent with Cullen and Purser's criterion precisely when $H_{\text{SG}} \leq \pi(\lambda^2 - 1)(\cosh \varphi)/4$, which occurs at low energies (if and) only if $\lambda > 1$. Even though harmonicity of $P(t, \mathbf{x})$ cannot change, his numerical simulations with our quadratic ansatz show a saddle can evolve from a pressure extremum and vice versa.

1.7. Main results

We are now ready to state our main conclusions, which describe the canonical Hamiltonian dynamics of all solutions consistent with our quadratic ansatz.

Theorem 1.1 (evolution of quadratic pressure for a SG fluid in an ellipse). *Let a fluid be constrained to an ellipse (1.11) of area $\pi\lambda > 0$ and aspect ratio $e^\varphi \geq 1$. If the pressure and stream function are quadratic (1.10) at $t = 0$, with $P(0) + \mathbf{I} > 0$ and $\lambda^2 := 1/\det(\mathbf{I} + P(0))$,*

there is a SG evolution (1.5) and (1.6) remaining spatially quadratic in which the Hessian of the pressure is given by

$$P(t) = \frac{1}{\lambda z(t)} (\mathbf{S}_{-\varphi} + \mathbf{R}_{\theta(t)} \mathbf{S}_{\log a(t)} \mathbf{R}_{-\theta(t)}) - \mathbf{I}. \quad (1.16)$$

Here, \mathbf{R}_θ and \mathbf{S}_σ are rotation (3.1) and shear matrices (3.2), while $a(t), \theta(t) \in C^\infty([0, \infty[)$ and the normalization constant $z(t) = z(a, \theta; \varphi) \geq 2$ uniquely solve the ordinary differential equations

$$\frac{da}{dt} = -\frac{2\lambda \sinh \varphi}{z(a, \theta; \varphi)} a \sin(2\theta), \quad (1.17)$$

$$\frac{d\theta}{dt} = \begin{cases} 1 - \frac{\lambda}{z(a, \theta; \varphi)} \left(\cosh \varphi + \frac{a^2 + 1}{a^2 - 1} \sinh \varphi \cos(2\theta) \right) & \text{if } a \neq 1, \\ \frac{1}{2} \left(1 - \frac{\lambda \cosh \varphi}{2 \cosh(\varphi/2)} \right) & \text{if } a = 1, \end{cases} \quad (1.18)$$

$$z(a, \theta; \varphi) = \sqrt{2 + \left(a + \frac{1}{a}\right) \cosh \varphi + \left(a - \frac{1}{a}\right) \sinh \varphi \cos(2\theta)} \quad (1.19)$$

with initial conditions $a(0) \geq 1$ and $\theta(0)$ selected by $P(0)$ through (1.16). By convention, $\theta(0) := \pi/4$ if $a(0) = 1$. Naturally, the trajectories of this evolution are constrained to lie on level sets of the SG energy

$$H(a, \theta) = \frac{1}{2} \left(a + \frac{1}{a}\right) - \lambda z(a, \theta; \varphi) + \lambda^2 \cosh \varphi = \frac{4H_{\text{SG}}}{\pi}. \quad (1.20)$$

Changing variables to $r = (a + 1/a)/2$ and re-expressing this energy as a function of (r, θ) and the parameters λ and $s = \cosh \varphi \geq 1$,

$$\bar{H}(r, \theta) = \lambda^2 s + r - \lambda \left(2 + 2rs + 2 \cos(2\theta) \sqrt{(r^2 - 1)(s^2 - 1)}\right)^{1/2} \quad (1.21)$$

brings out the $r \leftrightarrow s$ symmetry while converting our evolution (1.17)–(1.19) to an autonomous Hamiltonian system governing the canonically conjugate variables $(r(t), \theta(t))$ of section 2.2:

$$\frac{dr}{dt} = -\frac{\partial \bar{H}}{\partial \theta} = -\frac{2\lambda}{\bar{z}(r, \theta; s)} \sin(2\theta) \sqrt{(r^2 - 1)(s^2 - 1)}, \quad (1.22)$$

$$\frac{d\theta}{dt} = \frac{\partial \bar{H}}{\partial r} = 1 - \frac{\lambda}{\bar{z}(r, \theta; s)} \left(s + r \cos(2\theta) \sqrt{\frac{s^2 - 1}{r^2 - 1}} \right). \quad (1.23)$$

This Hamiltonian energy is smooth on $r > 1$ and continuous up to $r = 1$, where it fails to be differentiable; $a = r = 1$ is also the location of a coordinate singularity where θ fails to be uniquely defined. To understand the global topology of the energy surface, it is helpful to logarithmically rescale the radial variable $\sigma = \log a$ and re-express the Hamiltonian again:

$$\tilde{H}(\sigma, \theta) = \lambda^2 \cosh \varphi + \cosh \sigma - \lambda \tilde{z}(\sigma, \theta; \varphi), \quad (1.24)$$

$$\frac{\tilde{z}^2}{2} = 1 + \cosh(\sigma + \varphi) \cos^2 \theta + \cosh(\sigma - \varphi) \sin^2 \theta. \quad (1.25)$$

This clearly shows the range $]2 \cosh((\sigma - \varphi)/2), 2 \cosh((\sigma + \varphi)/2)[$ of values of $z = \tilde{z}(\sigma, \theta; \varphi)$. Since the original energy

$$H(a^{-1}, \theta) = H\left(a, \theta + \frac{\pi}{2}\right) \quad (1.26)$$

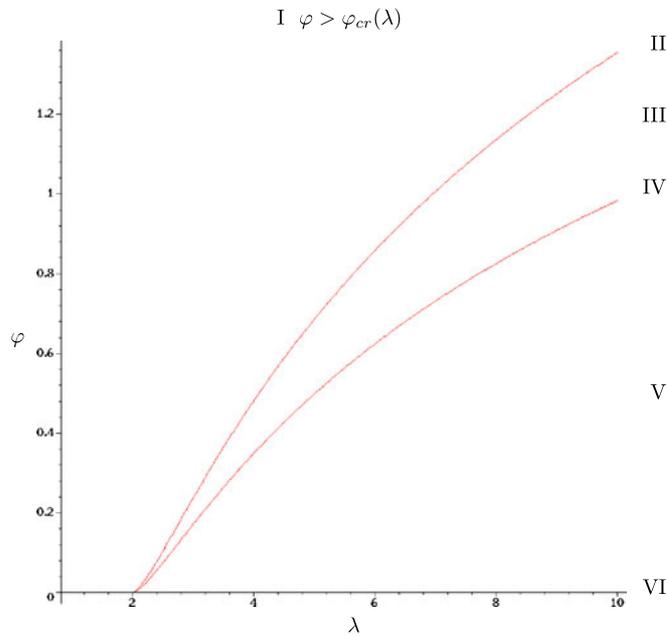


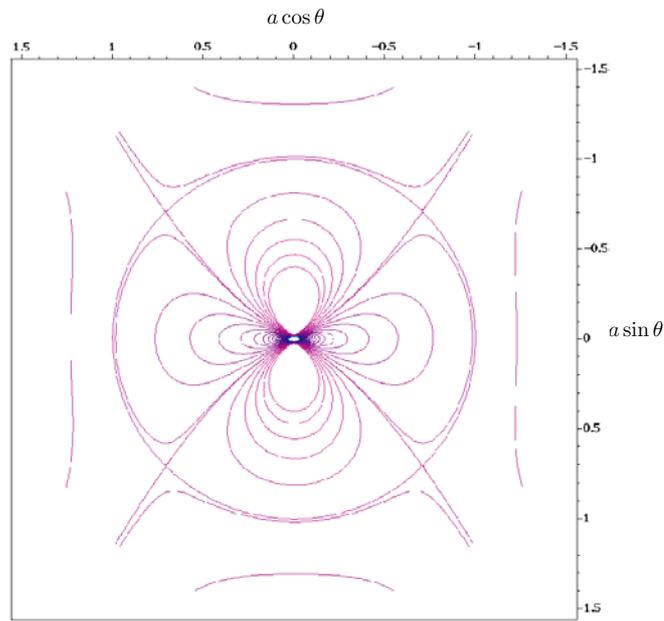
Figure 1. Bifurcation diagram illustrating the dependence of energy landscape on potential vorticity λ^{-2} and the eccentricity $\sqrt{1 - e^{-2\varphi}}$ of the fluid domain Y . Anticyclonic circulations occur precisely when the energy is small and $\lambda > 1$. The horizontal $\varphi = 0$ axis parameterizes circular domains whose area exceeds that of the evolving ellipse $X(t)$ by a factor λ^2 .

is symmetric under inversion in the unit circle when rotated by $\pi/2$, the restriction $\sigma \geq 0$ costs no generality; cf figures 2 and 3(a). Viewed as a function of the polar coordinates (σ, θ) on \mathbf{R}^2 , the topology of $\tilde{H}(\sigma, \theta) = \tilde{H}(\sigma, \theta + \pi)$ is elementary to classify, but undergoes two different bifurcations as the parameters (λ, φ) vary. These bifurcations occur along two curves II and IV described in the next theorem, which divide the parameter space into the three regions I, III, V displayed in figure 1. A topographical map displaying the level sets of $\tilde{H}(\sigma, \theta)$ for each of these five possibilities is given in figures 3(a)–(e). However, we must be careful when interpreting the orbit which passes through the pinched saddle at the origin $\sigma_1 = 0$. This saddle is not a fixed point, as the diagrams suggest, but rather lies on the unique periodic orbit in which the semi-major and minor axis of the dual ellipse $X(t)$ are perpetually exchanging roles. This orbit is more clearly displayed in figure 2, a topographical map of the original energy $H(a, \theta)$, taking (a, θ) as our polar coordinates. Here, the energy possesses four saddle points located at $a = 1$ and $\theta = \pm\pi/4, \pm 3\pi/4$, through which this special orbit threads as it weaves in and out of the symmetry circle. Squaring the radius in this final representation would facilitate visualization of the evolving dual ellipse $X(t)$, which then maintains its area π/λ as the tip of its semi-major axis traces out the level set of H .

As a prelude to characterizing the energy surface for an elliptical fluid domain, let us start by describing the analogous situation for a circular domain, figure 3(f).

Lemma 1.2 (topology of the energy landscape for a circular domain). *If $\varphi = 0$ and $\lambda > 0$ are fixed, the energy (1.24)*

$$\tilde{H}(\sigma, \theta) = 2 \left(\frac{\lambda}{2} - \cosh \frac{\sigma}{2} \right)^2 + \frac{\lambda^2}{2} - 1$$



$H(a, \theta)$ for $(\lambda, \varphi) = (3, 0.7368522964)$.

Figure 2. Contour plot of the energy $H(a, \theta)$ on Cartesian axes $(x, y) = (a \cos \theta, a \sin \theta)$. As in figure 3(a) the area $\pi\lambda$ and eccentricity of the fluid domain are given by $(\lambda, \varphi) = (3, 0.74)$. The dual ellipse has unchanging area π/λ while its aspect ratio $a(t)$ and inclination $\theta(t)$ evolve along level sets of $H(a, \theta) = H(a^{-1}, \theta + \pi/2)$.

is smooth and independent of θ . For $\lambda \leq 2$ it has a global minimum at $\sigma_1 = 0$ and no other critical points. For $\lambda > 2$ the origin $\sigma_1 = 0$ becomes a local maximum, and the only other critical points form the circle at radius $\sigma_- = 2 \cosh^{-1}(\lambda/2)$ where the global minimum is achieved.

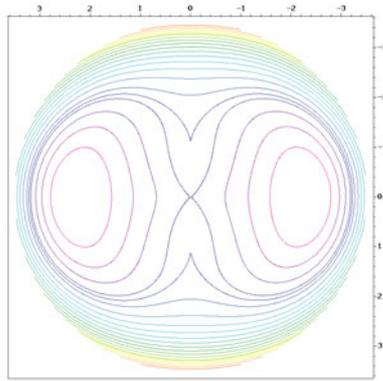
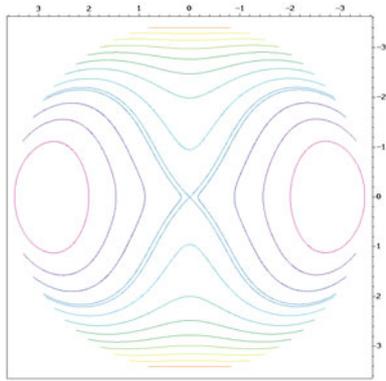
Proof. Noting that $h(\sigma) := H(\sigma, \theta)$ diverges with $\sigma \rightarrow +\infty$, we compute $h'(\sigma) = 2(\cosh(\sigma/2) - (\lambda/2)) \sinh(\sigma/2)$ and $h''(0) = 1 - \lambda/2$. The critical points and their classification then become obvious. \square

Example 1.3 (pressure and velocity dynamics in a circular domain). The resulting pressure and velocity dynamics are easy to describe. Vanishing of the stream function $2Q(\mathbf{y}) = q(t)(\lambda^{-1}|\mathbf{y}|^2 - 1)$ at the boundary of the circular domain constrains the fluid to rotate as a solid body. The local extremum $\sigma_1 = 0$ of the energy corresponds to a spherically symmetric pressure $P(t, \mathbf{y}) = (\lambda^{-1} - 1)|\mathbf{y}|^2/2$ of variable strength (but $\lambda > 0$ for Cullen and Purser's stability), which forms a steady-state solution of (1.12). The nonlinear response of the fluid to this pressure field is to rotate with angular velocity $q(t) = \lambda(1 - \lambda)$ —clockwise around a pressure maximum $\lambda > 1$ and counterclockwise around a pressure minimum $\lambda < 1$.

If, instead, the isobars are centred ellipses (or hyperbolae) corresponding to a pressure matrix $\mathbf{P}(0) + \mathbf{I} = (\mathbf{I} + \mathbf{S}_{\sigma_0})/[2\lambda \cosh(\sigma_0/2)]$, then the isobars rotate with constant angular velocity $\theta'(t) = 1 - \lambda/[2 \cosh(\sigma_0/2)]$, their shape remaining otherwise fixed. As expected, the pressure waves move quite differently from the fluid, which has angular velocity $q(t) = \lambda(1 - \lambda/\cosh(\sigma_0/2))$ from (1.12). For instance, near the isotropic limit $\sigma_0 \rightarrow 0$, the sign

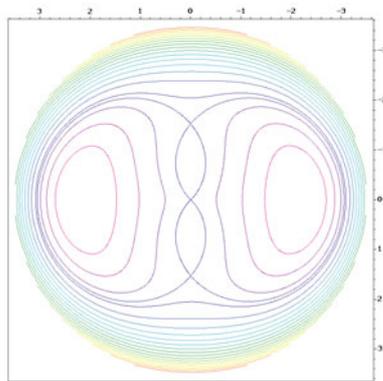
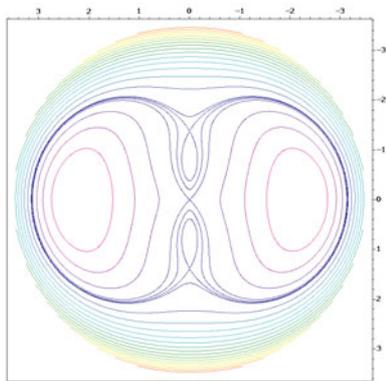
(a) $\tilde{H}(\sigma, \theta)$ for $(\lambda, \varphi) = (3, 0.7368522964)$,

(b) $(\lambda, \varphi_{cr}(\lambda)) = (3, 0.2368522964)$,



(c) $(\lambda, \varphi) = (3, 0.2068522964)$, $h_1 < h_2$,

(d) $(\lambda, \varphi_{bi}(\lambda)) = (3, 0.1709945858)$,



(e) $(\lambda, \varphi) = (3, 0.1368522964)$, $h_1 > h_2$,

(f) $(\lambda, \varphi) = (3, 0)$.

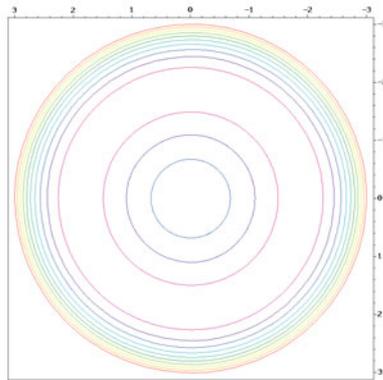
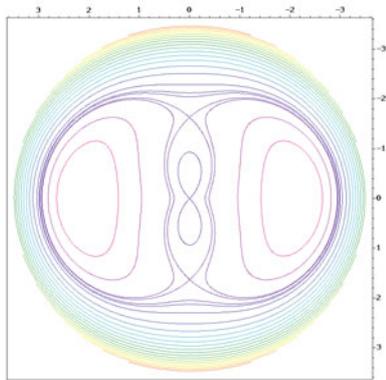


Figure 3. Contour plot of the energy $\tilde{H}(\sigma, \theta) = H(e^\sigma, \theta)$ on Cartesian axes $(x, y) = (\sigma \cos \theta, \sigma \sin \theta)$ for a sequence of domains $\lambda = 3$ times as large as the evolving ellipse, with decreasing eccentricities $\sqrt{1 - e^{-2\varphi}}$. The outer green curve encircles those orbits on which the fluid flow is anticyclonic (clockwise) through $Y \subset \mathbf{R}^2$.

of $\lambda - 2$ determines whether the isobars rotate in the direction parallel or opposite to the anticyclonic circulation of particles around a pressure maximum $\lambda > 1$. Similarly, when the energy is minimized at $\sigma_- = 2 \cosh^{-1}(\lambda/2) \in]0, \infty[$, the isobars form an unmoving set of hyperbolae while the fluid particles circulate clockwise with rate $q(t) = -\lambda < -2$.

Much of the topological variety of behaviour captured in the theorem below can be deduced from the bifurcation at $\lambda = 2$ in the circular case $\varphi = 0$, together with the pinched saddle geometry of $\tilde{H}(\sigma, 0)$ at the origin $\sigma_1 = 0$. The proof of this theorem is largely an exercise in multivariate calculus which is postponed until section 7. The location and significance of the critical thresholds which start from $\lambda = 2$ and $\lambda = 1$ are interpreted in section 2.3 and appendix A, respectively.

Theorem 1.4 (bifurcations of the energy landscape for an elliptical domain). *If $\lambda, \varphi > 0$, the function $\tilde{H} \in C(\mathbf{R}^2)$ expressed in polar coordinates by (1.24) is smooth on $\mathbf{R}^2 \setminus \{0\}$, has global minima placed symmetrically on the horizontal axis at $(\pm\sigma_-, 0)$, and a topological saddle point at $\sigma = 0$. These are the only critical points unless $\lambda > 2$ and $\varphi \leq \varphi_{cr}(\lambda)$, where*

$$\sinh \frac{\varphi_{cr}(\lambda)}{2} = \frac{1}{\lambda\sqrt{2}} \left(-1 + \sqrt{\frac{\lambda^2 - 1}{3}} \right)^{3/2}. \quad (1.27)$$

If $\varphi < \varphi_{cr}(\lambda)$ there are two additional non-degenerate critical points $(\pm\sigma_+, \pi/2)$ and $(\pm\sigma_2, \pi/2)$ on the vertical axis: a local maximum at $\sigma_+ > 0$ and a saddle point at $\sigma_2 > \sigma_+$ (plus their reflections on the negative vertical axis); these coalesce into a single degenerate critical inflection point $\sigma_+ = \sigma_2 > 0$ at the bifurcation eccentricity $\varphi = \varphi_{cr}(\lambda)$. The critical points are the solutions to

$$\sinh \sigma_{\pm} = \lambda \sinh \left(\frac{\sigma_{\pm} \mp \varphi}{2} \right) \quad (1.28)$$

and the corresponding critical values are ordered by

$$\begin{aligned} \tilde{H}(\sigma_-, 0) &=: h_- < h_1 := \tilde{H}(0, 0) < h_+ := \tilde{H}\left(\sigma_+, \frac{\pi}{2}\right), \\ h_- < h_2 &:= \tilde{H}\left(\sigma_2, \frac{\pi}{2}\right) < h_+. \end{aligned} \quad (1.29)$$

There are no further critical points. A continuous increasing curve $\varphi_{bi} :]2, \infty[\rightarrow]0, \infty[$ strictly less than $\varphi_{cr}(\lambda)$ separates the region $0 < \varphi < \varphi_{bi}(\lambda)$ of the parameter space where $h_1 > h_2$ from the region $\varphi_{bi}(\lambda) < \varphi < \varphi_{cr}(\lambda)$ where $h_1 < h_2$. The curve $\lambda \rightarrow (\varphi_{bi}(\lambda), \sigma_2(\lambda, \varphi_{bi}(\lambda)))$ simultaneously solves (1.28) and

$$\cosh \left(\frac{\sigma - \varphi}{2} \right) = \cosh \frac{\varphi}{2} + \frac{\cosh \sigma - 1}{2\lambda}. \quad (1.30)$$

If $\varphi \neq \varphi_{bi}(\lambda)$, one connected component of the level set $\{(\sigma \cos \theta, \sigma \sin \theta) \in \mathbf{R}^2 \mid \tilde{H}(\sigma, \theta) = h_1\}$ consists of a smoothly immersed figure-eight curve with orthogonal self-intersection at the origin.

1.8. Global pressure dynamics: wobbling versus rotating pulsations

From the topology of the Hamiltonian function, all SG dynamics consistent with the quadratic ansatz (1.10) can be inferred. These are summarized as follows. The orbits of the dynamical system divide into two classes according to the boundedness or unboundedness of $\theta(t)$. Apart from fixed points of the dynamics, and the aperiodic homoclinic and heteroclinic orbits which

link the saddles at $\pm\sigma_2$ on the vertical axis, the aspect ratio $a(t)$ and inclination $(\theta(t) \bmod \pi)$ of the dual potential vorticity ellipse evolve periodically. The principal axes of this ellipse (and of the isobars) either oscillate about the coordinate axes or precess eternally. We call the former motions *wobbling* and the latter modes *rotating pulsations*. For fixed parameters (λ, φ) , the range of periods represented by different orbits will be bounded below; in regime I this range is also bounded above, while it is unbounded in the other regimes. The bound from below may reflect success of the SG approximation at filtering out fast fluid motions. In what follows, we address the different dynamics in regimes I, III and V of the bifurcation diagram separately. Together with their boundaries, these regions exhaust all possible parameter values.

In each of the three regimes, the pinched saddle at $\sigma_1 = 0$ lies on an energy surface $\tilde{H}(\sigma, \theta) = h_1$ taking the form of a figure-eight curve. Due to the coordinate singularity at $\sigma_1 = 0$, this saddle is not fixed by the flow, but instead is reached by a trajectory following the figure-eight curve in finite time. The same orbit may then be viewed as continuing along either of the (equivalent) orthogonal branches of the figure-eight curve, counterclockwise in regions I and III but clockwise in region V, after the angle $\theta(t)$ jumps by $\pm\pi/2$. This jump discontinuity does not occur in the original variables $(a(t), \theta(t))$, where the semi-major and -minor axes of the evolving ellipse simply exchange roles as the aspect ratio passes instantaneously through $a(t_0) = 1$. The symmetry (1.26) shows this exchange is perpetually repeated on the distinguished figure-eight orbit—which corresponds to the unique orbit weaving in and out of the singular circle $a = 1$ in figure 2. For the other orbits of our dynamical system, assuming $a(t) > 1$ causes no loss of generality due to the symmetry (1.26). The orbits inside the figure-eight curve represent wobbling motions, in which the tip of the semi-major axis of the dual ellipse traces a closed loop around its fixed value $(e^{\sigma_{\pm}}/\lambda)^{1/2}$, counterclockwise around the energy minimum σ_- in regions I–III and clockwise around the maximum σ_+ in regime V. The orbits just outside the figure-eight curve represent rotating pulsations, in which the elliptical potential vorticity patch rotates end over end in the dual coordinates while its aspect ratio pulsates commensurately. The overall rotation is counterclockwise except in region V.

In regime I this exhausts the topological description of the dynamics. However, the remaining regimes possess a second critical curve $\tilde{H}(\sigma, \theta) = h_2$ linking two hyperbolic saddles symmetrically placed on the vertical axis. In regime III ($h_1 < h_2$) these saddles have stable and unstable manifolds which consist of one homoclinic and one heteroclinic orbit each. The homoclinic orbit encloses a local maximum σ_+ around which trajectories wobble clockwise, with arbitrarily long periods. The remaining dynamics consist of rotating pulsations whose periods diverge at the heteroclinic orbits. Note that although the net rotation is counterclockwise, trajectories close to the homoclinic orbits will temporarily experience retrograde motion as they approach the fixed saddles.

In regime V ($h_1 > h_2$) the stable manifold of each hyperbolic saddle consists of two heteroclinic orbits, which also form the unstable manifold of the second saddle point. Wobbling modes are present around both the local maxima σ_+ and global minima σ_- , although only the latter (counterclockwise) wobbles display arbitrarily long periods. Rotating pulsations of arbitrarily long periods are present in both the clockwise and counterclockwise directions; the former are sandwiched between the figure-eight curve and the nearest heteroclinic orbits, while the latter encircle all four heteroclinic orbits.

To see that the periods of all orbits remain bounded depending on (λ, φ) , we may simply observe that the Hamiltonian is smooth at the fixed points $(r_-, 0) = (\cosh \sigma_-, 0)$ and $(r_+, \pi/2) = (\cosh \sigma_+, \pi/2)$ and grows linearly as $r \rightarrow \infty$. In the small amplitude limit, the wobbling periods $T_{\pm} = 2\pi/\sqrt{\det D^2 \tilde{H}(r_{\pm}, \pi/2 \text{ or } 0)}$ around the local extrema can be deduced from (7.8) and (7.9). For example, when $\lambda = 1$, meaning the dual potential vorticity ellipse has the same area as the physical domain, the fixed energy minimum $\sigma_- = \varphi$ occurs

precisely when the two ellipses coincide; small pressure wobbles around this rest state occur with a period asymptotic to $T_- = 2\pi$. Similarly, (1.23) shows the limiting period $T_\infty = \pi/\Omega$ of rotating pulsations as $r \rightarrow \infty$ coincides with twice the Coriolis frequency $\Omega = \frac{1}{2}$. Although the period of pressure oscillations determines the period of fluctuations in the circulation speed $q(t)$ through (1.15), we caution that it is generally quite different from the period taken by physical particles to circumnavigate the domain, which is rather determined by an averaged value of $1/q(t)$.

Remark 1.5 (terminology: heteroclinic versus homoclinic). In the preceding discussion it was convenient to distinguish between orbits which are heteroclinic and homoclinic in the phase space $(r, \theta) \in [1, \infty[\times \mathbf{R}$. However, since the symmetric saddle points on the vertical axis actually represent the same Eulerian fluid configuration, one can argue that all orbits asymptotic to these saddles ought to be described as ‘homoclinic’. Similarly, the rotating pulsations and figure-eight curve are periodic in Eulerian variables despite the fact that $\theta(t)$ may be a strictly monotonic function of time. As with the rigid pendulum, it is a question of semantics whether one calls such orbits ‘periodic’ or ‘unbounded’.

1.9. Self-consistency of the SG approximation

A posteriori, it is possible to check self-consistency of the SG approximation (1.6) to the Euler equations (1.1) for the solutions which we have constructed, by computing the maximum value $R_0(t) = \|P(t)Q(t)^{-1} - \mathbf{I}\|$ of the Rossby number (1.8) over the domain Y . Since $Q(t) = q(t)F^{-2}$ is a matrix with a definite sign, we see that $R_0(t)$ cannot be small unless $P(t)$ has the same sign. A pressure saddle point at the origin violates this hypothesis. For a circular domain $\varphi = 0$ with isotropic pressure $\sigma_0 = 0$, the explicit computations of example 1.3 show the Rossby number $R_0(t) = \lambda^{-1} - 1$ is small when the evolving circle has the same area as the fluid domain.

2. Dual formulation as an active scalar transport problem

Introduce the Legendre transform

$$U(t, \mathbf{x}) = \sup_{\mathbf{y} \in Y} \mathbf{y} \cdot \mathbf{x} - V(t, \mathbf{y}) \quad (2.1)$$

of the convex function $V(t, \mathbf{y}) = P(t, \mathbf{y}) + |\mathbf{y}|^2/2$ at each instant in time. For the case when $U(t, \mathbf{x})$ and $V(t, \mathbf{y})$ are smooth and strictly convex, the evolution (1.5) and (1.6) was reformulated by Cullen and Purser [12] (see also Hoskins [22] and Schubert *et al* [42]) as a conservation law

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{u}) = 0, \quad (2.2)$$

$$\mathbf{u} = \mathbf{J}(\mathbf{x} - \nabla U(t, \mathbf{x})) = -\mathbf{J}\nabla\Phi(t, \mathbf{x}) \quad \text{in } \mathbf{R}^2 \times [0, \infty) \quad (2.3)$$

for the scalar quantity (sometimes known as *inverse potential vorticity*)

$$\rho(t, \mathbf{x}) = \det D^2 U(t, \mathbf{x}) = \det [\mathbf{I} + D^2 \Phi(t, \mathbf{x})], \quad (2.4)$$

$$\nabla U(t, \mathbf{R}^2) \subset \bar{Y}, \quad (2.5)$$

in the time-dependent dual coordinates $\mathbf{x} = \nabla V(t, \mathbf{y})$. This active scalar transport model was used by Benamou and Brenier [1], and later Lopes Filho and Nussenzweig Lopes [47] and Loeper [46], to define weak solutions and establish their global existence in the two- and three-dimensional incompressible models, and by Cullen and Gangbo [9] and Cullen

and Maroofi [10] in the shallow water and compressible theories. It is strongly analogous to the vorticity formulation of the Euler equation, except that the Monge–Ampère second boundary value problem (2.4)–(2.5) replaces Poisson’s equation relating the stream function $|\mathbf{x}|^2/2 - U(t, \mathbf{x})$ of the dual velocity (2.3) to the advected scalar $\rho(t, \mathbf{x})$. It has the advantage of decoupling the evolution of the physical pressure $P(t, \mathbf{y})$ from the physical velocity $Q(t, \mathbf{y})$. With some effort, the equivalence of (2.2)–(2.5) to (1.5) and (1.6) can be deduced by applying the material derivative $\partial_t + \mathbf{u}(t, \mathbf{x}) \cdot \nabla$ to the identity

$$1 = \det D^2V(t, \nabla U(t, \mathbf{x})) \det D^2U(t, \mathbf{x}),$$

calculating the mixed partials $(\partial \nabla U / \partial t)(t, \mathbf{x})$ and $(\partial D^2V / \partial t)(t, \mathbf{y})$ by differentiating

$$\mathbf{y} = \nabla U(t, \nabla V(t, \mathbf{y}))$$

and (1.6) with respect to time and space, respectively.

The quadratic ansatz (1.10) implies $U(t, \mathbf{x}) = \mathbf{x}^T \mathbf{U}(t) \mathbf{x} / 2$ on the evolving ellipse $X(t) = \mathbf{V}(t)Y$, where $V(t, \mathbf{y}) = \mathbf{y}^T \mathbf{V}(t) \mathbf{y} / 2$ with $\mathbf{V}(t) = \mathbf{U}(t)^{-1} = \mathbf{P}(t) + \mathbf{I}$ and $\mathbf{p}(t) = \mathbf{0}$, $p(t) = 0$ as usual. Our normalization $\det [\mathbf{V}(t)] = \det \mathbf{U}(t)^{-1} = \lambda^{-2}$ shows the inverse potential vorticity forms a vortex patch $\rho(t, \mathbf{x}) = \lambda^2 1_{X(t)}(\mathbf{x})$ given by the characteristic function of this ellipse in the dual coordinates. Our main task will be to find the positive symmetric evolution matrix $\mathbf{E}(t)$ governing the ellipse

$$X(t) = \{\mathbf{E}(t)\mathbf{z} \mid |\mathbf{z}| < 1\} = \{\mathbf{x} \in \mathbf{R}^2 \mid \mathbf{x}^T \mathbf{E}^{-2}(t) \mathbf{x} < 1\} \tag{2.6}$$

determined by (2.2)–(2.5). In this task, we are aided by global conservation of potential vorticity, which fixes the area $|X(t)| = \pi \epsilon^2 = \pi f^2 / \lambda^2$, and of SG energy (1.9), which implies that the Monge–Kantorovich–Wasserstein distance

$$\begin{aligned} |Y| \frac{W_2^2}{2} &= \frac{1}{2} \int_{X(t)} |\mathbf{x} - \nabla U(t, \mathbf{x})|^2 \rho(\mathbf{x}) \, d^2\mathbf{x} = \int_{\mathbf{R}^2} \frac{\rho |\mathbf{u}|^2}{2} \, d^2\mathbf{x} \\ &= \frac{1}{2} \int_Y |\nabla V(t, \mathbf{y}) - \mathbf{y}|^2 \, d^2\mathbf{y} = H_{SG}, \end{aligned} \tag{2.7}$$

between the physical fluid and the active scalar density $\rho(t, \mathbf{x})$ is independent of time. Thus, our solution will hinge on the explicit knowledge of the Wasserstein distance between the fixed ellipse F and the evolving ellipse $\mathbf{E}(t)$. For cultural reasons, let us therefore say a few words about the Wasserstein distance, referring the interested reader to the Villani’s book for references [45].

2.1. Wasserstein distance as Hamiltonian energy and transportation cost

The Wasserstein distance between (uniform probability measures on) two finite volume domains $Y, X \subset \mathbf{R}^n$ can be defined as the mean square transport distance required to redistribute all particles of fluid from Y uniformly throughout X : i.e. the minimum

$$W_2^2(Y, X) := \inf_{\det [Dm(\mathbf{y})] = |X|/|Y|} \int_Y |m(\mathbf{y}) - \mathbf{y}|^2 \, d^2\mathbf{y} / |Y| \tag{2.8}$$

among diffeomorphisms $m : Y \rightarrow X$ with constant Jacobian. Knott and Smith [27] understood that when such a diffeomorphism can be realized as the gradient of a convex function $V(\mathbf{y})$, then the infimum (2.8) is attained by $m = \nabla V$ (2.7). When both domains are ellipsoids, they went on to give an explicit formula for the optimal map $m(\mathbf{y}) = \mathbf{F}^{-1} \sqrt{(\mathbf{F}\mathbf{E}^2\mathbf{F})} \mathbf{F}^{-1} \mathbf{y}$ —which turns out to be linear—in terms of covariance matrices \mathbf{F}^2 and \mathbf{E}^2 ; the same formula was found by Dowson and Landau [19], Givens and Shortt [21] and Olkin and Pukelsheim [34], independently. However, the matrix square root makes this formula awkward for our purposes. In section 4 we shall derive a simpler representation of the optimal map, limited to dimension $n = 2$.

2.2. Physical and geometric interpretation of the Hamiltonian variables

Restoring correct units to the SG energy $\bar{H}_{\text{SG}} = (1/4\pi)|X(0)||Y|\bar{H}(r, \theta)$ and rescaling the time $\tau = 2t$, the canonically conjugate variables $(r_{\text{SG}}, \theta) := ((1/2\pi)|X(0)||Y|r, \theta)$ of our dynamical system (1.22)–(1.23) may be interpreted as the moment of inertia of the mass distribution $\rho(t, \mathbf{x})$, and its inclination to the coordinate axes. These appear to be physically natural variables. Geometrically, it is more appealing to take the Wasserstein distance W_2^2 separating the two ellipses as the Hamiltonian function and retain the original time scale t , in which case the conjugate variable $r_{\text{SG}}/|Y|$ to θ represents the moment of inertia of a probability distribution spread uniformly over the evolving vorticity patch $X(t)$. Either way, the same canonical variables govern the evolution of the Kirchoff and Kida elliptical vorticity patches studied by Neu [32], despite the fact that the Hamiltonian in those problems is quite different from ours. Meacham *et al* [30] have employed a reduction procedure for recovering such canonical variables in Hamiltonian subsystems; it would be interesting to know if their strategy applies in the SG context as well, where the relation between the stream function and advected scalar is nonlinear (2.4).

2.3. Symmetry breaking bifurcation in an optimal transport problem

The bifurcation commencing at $(\lambda, \varphi) = (2, 0)$ in figure 1 can now be understood as a consequence of symmetry breaking in a simple transportation problem: *among all ellipses $X \subset \mathbf{R}^2$ of fixed area π/λ^2 , find the one nearest the unit disc $Y = B_1$ in Wasserstein distance.* The answer depends on λ . If $\lambda \leq 2$, the closest ellipse is unique, according to lemma 1.2; as a circle $X = B_{1/\lambda}$, it shares the rotational symmetry of the problem. However, if $\lambda > 2$, meaning X is less than half as large as the unit disc, then instead of transporting all particles of B_1 to a ball near its centre, it becomes more efficient to stretch the ellipse out so that some of the particles near the boundary of B_1 need not be transported so far. In this case, the closest ellipses to B_1 combine aspect ratio $a_- := \exp[2 \cosh^{-1}(\lambda/2)]$ with arbitrary orientation θ ; there is a one parameter family of minimizers, but they do not share the rotational symmetry of the problem. If the same question is asked with a fixed ellipse Y instead of the unit ball, the problem is no longer symmetric, and uniqueness of the minimizer is restored: the minimizing ellipse X must be oriented parallel to Y , since $\theta_- = 0$ in theorem 1.4.

3. Matrix identities

3.1. Canonical form for symmetric matrices

In this section we establish some convenient representations for symmetric, positive definite, 2×2 matrices.

Definition 3.1 (rotation, reflection and shear matrices). *Define the rotation matrices through the angle θ ,*

$$\mathbf{R}_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \mathbf{I} \cos \theta + \mathbf{J} \sin \theta \quad \mathbf{J} = \mathbf{R}_{\pi/2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad (3.1)$$

the reflection matrices through the line with angle $\theta \in [0, \pi]$ to the x -axis,

$$\mathbf{K} := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{K}_\theta := \mathbf{R}_\theta \mathbf{K} \mathbf{R}_{-\theta}, \quad \mathbf{L} := \mathbf{K}_{\pi/4} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the shear matrix,

$$\mathbf{S}_\sigma := \begin{pmatrix} e^\sigma & 0 \\ 0 & e^{-\sigma} \end{pmatrix} = \mathbf{I} \cosh \sigma + \mathbf{K} \sinh \sigma. \quad (3.2)$$

Note that the matrices I, J, K, L form a linear basis for the 2×2 matrices, and that the following commutator identities hold

$$LJ = -JL = K, \quad JK = -KJ = L, \quad LK = -KL = J. \tag{3.3}$$

We also have

$$J^2 = -I, \quad K^2 = I, \quad L^2 = I,$$

much like the Pauli matrix or quaternion identities, except that these matrices are real.

We also have the following commutator identities for products with S_σ

$$JS_\sigma - S_\sigma J = 2L \sinh \sigma, \quad LS_\sigma + S_\sigma L = 2L \cosh \sigma, \quad KS_\sigma + S_\sigma K = 2KS_\sigma. \tag{3.4}$$

With these definitions, we have the canonical form for (2×2) symmetric matrices.

Definition 3.2 (canonical form for symmetric matrices). Write any 2×2 , symmetric positive definite matrix E with determinant ϵ^2 as

$$E = \epsilon R_\theta S_{\sigma/2} R_{-\theta}. \tag{3.5}$$

Note that

$$\begin{aligned} E^2 &= \epsilon^2 R_\theta S_\sigma R_{-\theta} \\ &= \epsilon^2 \begin{pmatrix} e^\sigma \cos^2 \theta + e^{-\sigma} \sin^2 \theta & (e^\sigma - e^{-\sigma}) \cos \theta \sin \theta \\ (e^\sigma - e^{-\sigma}) \cos \theta \sin \theta & e^\sigma \sin^2 \theta + e^{-\sigma} \cos^2 \theta \end{pmatrix} \\ &= \epsilon^2 \begin{pmatrix} \cosh \sigma + \sinh \sigma \cos 2\theta & \sinh \sigma \sin 2\theta \\ \sinh \sigma \sin 2\theta & \cosh \sigma - \sinh \sigma \cos 2\theta \end{pmatrix} \\ &= \epsilon^2 (I \cosh \sigma + K \sinh \sigma \cos 2\theta + L \sinh \sigma \sin 2\theta), \end{aligned} \tag{3.6}$$

while the inverse is given by

$$E^{-1} = \frac{1}{\epsilon} R_\theta S_{-\sigma/2} R_{-\theta}.$$

The canonical form is a convenient representation for differentiating the matrix with respect to the parameters θ and σ .

Lemma 3.3 (differentiating the canonical form). If $\theta = \theta(t)$ and $\sigma = \sigma(t)$ evolve smoothly but the determinant of $E = E(t)$ is fixed in (3.5), then

$$\frac{d}{dt} E^{-2} = \theta' (J E_t^{-2} - E_t^{-2} J) - \sigma' K_\theta E_t^{-2}. \tag{3.7}$$

Proof. This follows from the identities

$$R_\theta K = K_\theta R_\theta, \quad \frac{d}{d\theta} R_\theta = J R_\theta, \quad \frac{d}{d\sigma} S_\sigma = K S_\sigma. \quad \square$$

4. Matrix equations

4.1. Optimal map and transport cost between ellipses

In this section, we exhibit the optimal map and transportation cost between two ellipses in the plane. This amounts to solving the Monge–Ampère equation (2.4)–(2.5) when both the right-hand side $\rho(0, \mathbf{x}) = \lambda^2 1_X(\mathbf{x})$ and target $Y = F(B_1)$ are given by ellipses $X = E(B_1) \subset \mathbf{R}^2$. By specializing to two dimensions, we obtain a more convenient representation of the optimal map $m^{-1}(\mathbf{x}) = \nabla U(0, \mathbf{x})$ than the traditional one [19, 21, 27, 34]. This representation could be derived by truncating the matrix square root in $m(\mathbf{y}) = F^{-1} \sqrt{(FE^2F)} F^{-1} \mathbf{y}$ using the Cayley–Hamilton theorem, but it is simpler to address directly.

Definition 4.1 (matrix norm). Define the Hilbert–Schmidt norm on matrices M by

$$\|M\|^2 := \text{trace}(M^T M).$$

Lemma 4.2 (optimal maps between ellipses in \mathbf{R}^2). The measure-preserving diffeomorphism achieving the infimum (2.8) between the ellipses $1_Y/|Y|$ and $1_X/|X|$ from (1.11) and (2.6), is given by $\mathbf{x} = m(\mathbf{y}) = V\mathbf{y}$, where $V = U^{-1}$ is the positive symmetric matrix inverse to

$$\begin{aligned} U &:= Z^{-1} \left(E^{-2} + \frac{F^2}{\det(EF)} \right) \\ &= FR_\theta E^{-1}, \end{aligned} \quad (4.1)$$

with $\tan \theta = \text{trace}(E^{-1}F^{-1}J)/\text{trace}(E^{-1}F^{-1})$, and Z normalizes the determinant of U :

$$\begin{aligned} Z &:= \text{trace} \sqrt{(FE^2F)^{-1}} \\ &= \sqrt{\text{trace}(E^{-2}F^{-2}) + \frac{2}{\det(EF)}}, \end{aligned} \quad (4.2)$$

$$= \sqrt{\det \left(E^{-1}F^{-1} + \frac{EF}{\det(EF)} \right)}. \quad (4.3)$$

Proof. The matrix U defined by (4.1) is positive and symmetric whenever E and F are. The map $\mathbf{y} = U\mathbf{x}$ is then the gradient of a convex function $\mathbf{x}^T U\mathbf{x}/2$, which means U is an optimal map, whatever its image may be [27]. Linearity implies the image of X under U is an ellipse; we need only verify $U(X) = Y$ to conclude the lemma.

Since $E : B_1 \rightarrow X$ and $F : B_1 \rightarrow Y$, where B_1 is the unit ball, this is equivalent to showing $F^{-1}UE = R_\theta$ coincides with a rotation $R_\theta : B_1 \rightarrow B_1$ by some angle θ . Write without loss of generality

$$E^{-1} = \begin{pmatrix} a & c \\ c & b \end{pmatrix}, \quad F^{-1} = \begin{pmatrix} f_1 & 0 \\ 0 & f_2 \end{pmatrix}.$$

Multiplying (4.1) on the left by F^{-1} and on the right by E , we compute

$$\begin{aligned} F^{-1}UE &= \frac{F^{-1}E^{-1} + FE/\det(EF)}{\text{trace} \sqrt{(FE^2F)^{-1}}} \\ &= \frac{1}{\text{trace} \sqrt{(FE^2F)^{-1}}} \left(\begin{pmatrix} f_1 a & f_1 c \\ f_2 c & f_2 b \end{pmatrix} + \begin{pmatrix} f_2 b & -f_2 c \\ -f_1 c & f_1 a \end{pmatrix} \right) \end{aligned}$$

and

$$\begin{aligned} F^{-1}UE(F^{-1}UE)^T &= \frac{(FE^2F)^{-1} + FEF/\det(FE^2F) + 2I\sqrt{\det(FE^2F)^{-1}}}{(\text{trace} \sqrt{(FE^2F)^{-1}})^2} \\ &= I \end{aligned}$$

by the 2×2 matrix identity $I(\text{trace } N)^2 = N^2 + N^{-2}\det N^2 + 2I\det N$ applied to $N = \sqrt{(FE^2F)^{-1}}$.

Thus, $F^{-1}UE = R_\theta$ is indeed a rotation matrix. The unknown angle θ can be computed explicitly from

$$\tan \theta = \frac{(f_2 - f_1)c}{f_1 a + f_2 b} = \frac{\text{trace}(E^{-1}F^{-1}J)}{\text{trace}(E^{-1}F^{-1})}. \quad \square$$

The formula for the optimal map allows for the computation of the transportation cost.

Lemma 4.3 (transport cost between ellipses). *Let $Y = F(B_1)$ and $X = E(B_1)$ be the ellipses (1.11) and (2.6). The minimal quadratic cost W_2^2 of transporting one ellipse to the other (each weighted to have total mass one), is given by*

$$W_2^2(X, Y) = \frac{1}{4}(\|E\|^2 + \|F\|^2 - 2\det(EF)\text{trace}\sqrt{(EF^2E)^{-1}}), \tag{4.4}$$

$$W_2^2(X, Y) = \frac{1}{2\pi}(|X| \cosh \sigma + |Y| \cosh \varphi - \tilde{z}(\sigma, \theta; \varphi)\sqrt{|X||Y|}), \tag{4.5}$$

where $\tilde{z}(\sigma, \theta; \varphi)$, $E = R_\theta S_\sigma R_{-\theta}$ and $F = S_\varphi$ are defined by (1.25), (3.1) and (3.2).

Proof. The transport cost (2.7) and (2.8) is given by

$$W_2^2 = \frac{1}{|X|} \int_X |y - x|^2 d^2x = \frac{1}{\pi \det E} \int_X |Ux - x|^2 d^2x,$$

where $U : X \rightarrow Y$ is the optimal map (4.1). We now change the domain of integration to the unit ball $B_1 \subset \mathbf{R}^2$, by setting $x = Ez$ and $d^2x = \det E d^2z$ to get

$$W_2^2 = \frac{1}{\pi} \int_{B_1} (|UEz|^2 + |Ez|^2 - 2z^T EUEz) d^2z.$$

For any 2×2 matrix U , integration gives the identity

$$\int_{B_1} |Uz|^2 d^2z = \frac{\pi}{4} \|U\|^2. \tag{4.6}$$

Since $UE = FO$, for some orthogonal matrix O given by (4.1), the integration identity above yields

$$4W_2^2 = \|F\|^2 + \|E\|^2 - \frac{2(\|I\|^2 + (\|EF\|^2/\det(EF)))}{\text{trace}\sqrt{(EF^2E)^{-1}}}.$$

Simplifying, using the relation $(\text{trace}\sqrt{(EF^2E)^{-1}})^2 = \text{trace}(E^{-2}F^{-2}) + 2/\det(EF)$ from (4.2) and $\det(FE)^{-2}\|FE\|^2 = \|EF\|^2 = \text{trace}(E^{-2}F^{-2})$, we get

$$\|I\|^2 + \frac{\|EF\|^2}{\det(EF)} = \det(EF)(\text{trace}\sqrt{(EF^2E)^{-1}})^2,$$

which, finally, gives (4.4).

To derive (4.5), observe $\|F\|^2 = 2f^2 \cosh \varphi = (2/\pi)|Y| \cosh \varphi$ and $\|E\|^2 = (2/\pi)|X| \cosh \sigma$ follow from (1.11) and (3.5). Using the fact that L, K and $LK = J$ are traceless in (3.6) and the analogous expression for $F^{-2} = f^{-2}(I \cosh \varphi - K \sinh \varphi)$, we deduce that

$$\begin{aligned} \text{trace}(E^{-2}F^{-2}) &= \epsilon^{-2}f^{-2}\text{trace}(I^2 \cosh \sigma \cosh \varphi + K^2 \cos(2\theta) \sinh \sigma \sinh \varphi) \\ &= 2\epsilon^{-2}f^{-2}(\cosh(\sigma + \varphi) \cos^2 \theta + \cosh(\sigma - \varphi) \sin^2 \theta), \end{aligned}$$

using $K^2 = I$, whence $Z = z(\sigma, \theta; \varphi)/(\epsilon f)$ follows from (4.2) and (1.25). This completes the identification of third terms in the costs (4.4) and (4.5) thereby proving the lemma, and establishing (5.4) for later. \square

4.2. Dynamics of the free boundary

We conclude this section with the familiar observation that an ellipse convected by a spatially linear velocity field yields a family of ellipses parameterized by time, and find the differential equation relating this geometrical progression of ellipses to the vector field.

Suppose a spatially linear velocity field $\mathbf{u}(t, \mathbf{x}) = \mathbb{W}(t)\mathbf{x}$ on \mathbf{R}^2 is given by a matrix $\mathbb{W}(t)$ whose coefficients are continuously differentiable in time. The associated Lagrangian trajectories $\mathbf{X}(t, \mathbf{x}) \in \mathbf{R}^2$ are defined by integrating the differential equation

$$\begin{aligned}\mathbf{X}'(t, \mathbf{x}) &= \mathbb{W}(t)\mathbf{X}(t, \mathbf{x}), \\ \mathbf{X}(0, \mathbf{x}) &= \mathbf{x}.\end{aligned}\tag{4.7}$$

Lemma 4.4 (convecting ellipses by spatially linear velocities). *An ellipse $X \subset \mathbf{R}^2$ convected (4.7) by a linear velocity field $\mathbf{u}(\mathbf{x}, t) = \mathbb{W}(t)\mathbf{x}$ yields a family of ellipses $\mathbf{X}(t, X) = \{\mathbf{E}(t)\mathbf{z} \mid |\mathbf{z}| < 1\} \subset \mathbf{R}^2$ governed by the ordinary differential equation*

$$\frac{d}{dt}\mathbf{E}^{-2} = -\mathbb{W}^T\mathbf{E}^{-2} - \mathbf{E}^{-2}\mathbb{W}.\tag{4.8}$$

Proof. Since $\mathbf{X}(t, \mathbf{x})$ is spatially linear, it can be immediately seen that the image $X(t) := \mathbf{X}(t, X)$ of $X(0) = X$ remains an ellipse at each instant in time. To find its covariance matrix, let $\mathbf{x}(t) = \mathbf{X}(t, \mathbf{x}_0)$ be a trajectory of the ordinary differential equation (4.7), which stays on the boundary of the ellipse $\mathbf{X}(t, X)$. Differentiate the equation

$$\mathbf{x}^T\mathbf{E}^{-2}\mathbf{x} = 1$$

to get

$$(\mathbf{x}^T)'\mathbf{E}^{-2}\mathbf{x} + \mathbf{x}^T(\mathbf{E}^{-2})'\mathbf{x} + \mathbf{x}^T\mathbf{E}^{-2}\mathbf{x}' = 0.$$

Because $\mathbf{x}'(t) = \mathbb{W}(t)\mathbf{x}(t)$ we have

$$\mathbf{x}^T(\mathbb{W}^T\mathbf{E}^{-2} + (\mathbf{E}^{-2})' + \mathbf{E}^{-2}\mathbb{W})\mathbf{x} = 0,$$

for any boundary point, and hence the result. \square

5. Active scalar transport of an elliptical inverse-potential-vorticity patch

Consider the SG equations in the dual variables (2.2)–(2.5). Assume the target $Y = \mathbb{F}(B_1)$ is an ellipse (1.11) aligned with the coordinate axes. Since the advecting velocity $\mathbf{u} = -\mathbb{J}\nabla\Phi$ is divergence free, the potential vorticity $\rho(t, \mathbf{x})$ will be uniformly distributed over its evolving support if it is so initially. This corresponds to the evolution of an inverse potential vorticity patch in the plane $\mathbf{x} \in \mathbf{R}^2$. When the vorticity patch starts out in the form of an ellipse, we have just shown the velocity $\mathbf{u}(t, \mathbf{x})$ to be spatially linear, so the patch remains elliptical $X(t) = \mathbf{E}(t)(B_1)$ by lemma 4.4. Its dimensions and semi-axes are given as the eigenvalues and vectors of $\mathbf{E}(t)$. However, unlike Kirchoff's elliptic vorticity patches [26], this ellipse does not rotate uniformly with constant direction and speed. Instead, its aspect ratio $a(t)$ and inclination $\theta(t)$ fluctuate. Their nonlinear dynamics is given by the following theorem.

Theorem 5.1 (evolution of potential vorticity ellipse in SG dual coordinates). *For a fluid constrained to the ellipse $Y = \mathbb{F}(B_1)$ of (1.11), assume the potential vorticity forms an elliptical patch $X(t) = \mathbf{E}(t)(B_1)$ in SG dual coordinates at $t = 0$ and hence $t > 0$. We use the aspect*

ratio $e^{\sigma(t)}$ and inclination $\theta(t)$ of the evolving ellipse to express the matrix $E^2 = E(t)^2$ in the canonical form (3.5)

$$E^2 = \det(E)R_\theta S_\sigma R_{-\theta} = \epsilon^2 R_\theta S_\sigma R_{-\theta} \quad (5.1)$$

and the fixed ellipse in the form $F^2 = \det(F)S_\varphi = f^2 S_\varphi$ similarly. Then, the solution of the active scalar transport equations (2.2)–(2.5) with initial data $\rho(0, \mathbf{x}) = (f/\epsilon)^2 1_{X_0}(\mathbf{x})$, is given by

$$\rho(t, \mathbf{x}) = \frac{f^2}{\epsilon^2} 1_{X(t)}(\mathbf{x})$$

and

$$U(t, \mathbf{x}) = \frac{1}{2} \mathbf{x}^T U(t) \mathbf{x},$$

where $U(t)$ is a matrix given by

$$U(t) = \frac{1}{Z(t)} \left(E^{-2}(t) + \frac{F^2}{\epsilon^2 f^2} \right). \quad (5.2)$$

Whenever $\sigma(t_0) \neq 0$, the dynamical variables $(\sigma(t), \theta(t))$ which determine $E(t)$ satisfy the ordinary differential equation

$$\begin{aligned} \sigma' &= -\frac{2}{Z\epsilon^2} \sinh(\varphi) \sin(2\theta), \\ \theta' &= 1 - \frac{1}{Z\epsilon^2} \left(\cosh \varphi + \sinh(\varphi) \frac{\cosh \sigma}{\sinh \sigma} \cos(2\theta) \right) \end{aligned} \quad (5.3)$$

and the normalization constant $Z(\sigma, \theta; \varphi)$ is given by

$$\begin{aligned} Z &= \frac{\sqrt{2}}{\epsilon f} (1 + \cosh \sigma \cosh \varphi + \sinh \sigma \sinh \varphi \cos 2\theta)^{1/2} \\ &= \frac{\sqrt{2}}{\epsilon f} (1 + \cosh(\sigma + \varphi) \cos^2 \theta + \cosh(\sigma - \varphi) \sin^2 \theta)^{1/2}. \end{aligned} \quad (5.4)$$

Furthermore, the trajectories of the ODE are constrained to the level sets of the function

$$W^2 = \frac{1}{2} (\epsilon^2 \cosh \sigma + f^2 \cosh \varphi - Z\epsilon^2 f^2), \quad (5.5)$$

which is the Wasserstein distance between the two ellipses.

Proof. This proof requires a lot of manipulation of matrix identities to get the final result. The form of the normalization constant (5.4) and Wasserstein distance (5.5) were established already at the end of the proof of lemma 4.3. We defer a discussion of the course of the evolution after $\sigma(t_0) = 0$ to the next section, except to point out that this eventuality occurs only if the initial condition has one particular value h_1 of the conserved energy.

(1) Equating two different expressions (3.7) and (4.8) for the evolution of the dual ellipse under convection by the velocity field $\mathbf{u}(\mathbf{x}, t) = J(\mathbf{I} - U(t))\mathbf{x}$ in (2.3) gives the matrix differential equation

$$\theta'(JE^{-2} - F^{-2}J) - \sigma'K_\theta E^{-2} = (J(U - \mathbf{I}))^T E^{-2} + E^{-2}J(U - \mathbf{I}).$$

(2) Using $J^T = -J$ simplifies this to

$$(\theta' - 1)(JE^{-2} - E^{-2}J) - \sigma'K_\theta E^{-2} - (E^{-2}JU - UJE^{-2}) = 0.$$

(3) Now, the formula $U(t) = Z^{-1}[E^{-2} + F^2/(\epsilon^2 f^2)]$ for the optimal map (4.1) gives

$$(\theta' - 1)(JE^{-2} - E^{-2}J) - \sigma'K_\theta E^{-2} - \frac{Z^{-1}(E^{-2}JF^2 - F^2JE^{-2})}{\det(EF)} = 0. \quad (5.6)$$

We remark that at this stage of the computation we have a closed matrix equation which will give ordinary differential equations for θ and σ . The remainder of the computation requires only algebraic manipulations.

(4) First conjugate by R_θ , and multiply by the common factor $\epsilon^2 = \det(E)$, since $R_{-\theta}E^{-2}R_\theta = \epsilon^{-2}S_{-\sigma}$. Writing $F_\theta^{-2} := R_\theta F^{-2}R_{-\theta} = f^{-2}R_\theta S_{-\varphi}R_{-\theta}$ we get

$$(\theta' - 1)(JS_{-\sigma} - S_{-\sigma}J) - \sigma'KS_{-\sigma} - \frac{1}{Z\epsilon^2 f^2}(S_{-\sigma}JF_{-\theta}^2 - F_{-\theta}^2JS_{-\sigma}) = 0.$$

(5) Note that $JS_\sigma - S_\sigma J = 2L \sinh \sigma$ from (3.4).

(6) Next, we consider the term $(S_{-\sigma}JF_{-\theta}^2 - F_{-\theta}^2JS_{-\sigma})/(Z\epsilon^2 f^2)$. Factor out the determinant of F and write $f^{-2}F_{-\theta}^2$ in the canonical basis (3.6):

$$\begin{aligned} f^{-2}F_{-\theta}^2 &= \cosh(\varphi)I + \sinh \varphi \cos(2\theta)K - \sinh \varphi \sin(2\theta)L \\ &= a_1I + a_2K + a_3L. \end{aligned} \quad (5.7)$$

This last equation defines the temporary variables a_1, a_2, a_3 , which are used to save space, and whose scope is limited to this section of the proof.

Simplify the last term of (5.6) by writing

$$\begin{aligned} f^{-2}(S_{-\sigma}JF_{-\theta}^2 - F_{-\theta}^2JS_{-\sigma}) &= a_1(S_{-\sigma}J - JS_{-\sigma}) + a_2(S_{-\sigma}JK - KJS_{-\sigma}) + a_3(S_{-\sigma}JL - LJS_{-\sigma}) && \text{using (5.7)} \\ &= a_1(S_{-\sigma}J - JS_{-\sigma}) + a_2(S_{-\sigma}L + LS_{-\sigma}) - a_3(S_{-\sigma}K + KS_{-\sigma}) && \text{by (3.3)} \\ &= 2(a_1 \sinh \sigma + a_2 \cosh \sigma)L - 2a_3KS_{-\sigma} && \text{by (3.4)} \\ &= 2[\cosh \varphi \sinh \sigma + \sinh \varphi \cosh \sigma \cos(2\theta)]L + 2 \sinh \varphi \sin(2\theta)KS_{-\sigma} && \text{using (5.7)}. \end{aligned}$$

(7) Combine the results of steps 5 and 6 into (5.6) to get

$$\begin{aligned} &\left((2 \sinh(\sigma)(\theta' - 1) + \frac{2}{Z\epsilon^2}(\cosh \varphi \sinh \sigma + \sinh \varphi \cosh \sigma \cos(2\theta))) \right) L \\ &+ \left(\sigma' + \frac{2}{Z\epsilon^2} \sinh \varphi \sin(2\theta) \right) KS_{-\sigma} = 0. \end{aligned}$$

Noting that L is off-diagonal and KS_σ is diagonal gives the result

$$\begin{aligned} \sigma' &= -\frac{2}{Z\epsilon^2} \sinh(\varphi) \sin(2\theta), \\ \sinh(\sigma)(\theta' - 1) &= -\frac{1}{Z\epsilon^2}(\cosh \varphi \sinh \sigma + \sinh(\varphi) \cosh \sigma \cos(2\theta)) \end{aligned}$$

which finally gives (5.3).

(8) From lemma 4.3 we have a conserved quantity (5.5) which should hold during the evolution. As a check of our calculations we can simply differentiate

$$\frac{\partial W_2^2}{\partial \sigma} = +\frac{\epsilon^2}{2} \sinh \sigma \frac{d\theta}{dt}, \quad \frac{\partial W_2^2}{\partial \theta} = -\frac{\epsilon^2}{2} \sinh \sigma \frac{d\sigma}{dt}$$

to see that W_2^2 is unchanging:

$$\frac{dW_2^2}{dt} = \frac{\partial W_2^2}{\partial \sigma} \frac{d\sigma}{dt} + \frac{\partial W_2^2}{\partial \theta} \frac{d\theta}{dt} = \frac{\epsilon^2}{2} \sinh \sigma \left(\frac{d\theta}{dt} \frac{d\sigma}{dt} - \frac{d\sigma}{dt} \frac{d\theta}{dt} \right) = 0. \quad \square$$

6. Recovering the pressure dynamics (proof of theorem 1.1)

This section is devoted to reconstructing the pressure dynamics in physical space from our knowledge of the potential-vorticity evolution in dual space. To do this, we need mainly to establish that the Legendre conjugate functions $U(t, \mathbf{x})$ and $V(t, \mathbf{y})$ whose gradients define the change of variables relating these two spaces remain smooth and strictly convex. The only delicacy is to address the dynamics on the unique orbits which pass through the polar coordinate singularity $\sigma(t_0) = 0$ corresponding to a circular potential vorticity patch.

Proof of theorem 1.1. From our definition (2.1) of $V(t, \mathbf{y}) = P(t, \mathbf{y}) + |\mathbf{y}|^2/2$ and its Legendre transform $U(t, \mathbf{x}) = \mathbf{x}^T \mathbf{U}(t) \mathbf{x} / 2$, we obtain $\mathbf{P}(t) + \mathbf{I} = \mathbf{U}(t)^{-1}$ under the quadratic ansatz (1.10). To invert the map $U(t) : X(t) \rightarrow Y$, we simply exchange the ellipses and $\mathbf{E}(t)$ and \mathbf{F} in (5.2). Thus,

$$\begin{aligned} \mathbf{P}(t) + \mathbf{I} &= Z^{-1} \left(\mathbf{F}^{-2} + \frac{\mathbf{E}(t)^2}{(\epsilon \mathbf{f})^2} \right) \\ &= Z^{-1} \frac{\mathbf{S}_{-\varphi} + \mathbf{R}_{\theta(t)} \mathbf{S}_{\sigma(t)} \mathbf{R}_{-\theta(t)}}{\mathbf{f}^2} \end{aligned}$$

from (5.1). Since units of length were chosen to yield $\epsilon \mathbf{f} = 1$ and $\lambda = \mathbf{f}/\epsilon = \mathbf{f}^2$, comparing (1.19) with (5.4) we see $z = \epsilon \mathbf{f} Z = Z$. The identification $\sigma = \log a$ establishes (1.16). Since $H_{\text{SG}} = \pi \lambda W_2^2/2$ from (2.7), we also recognize (1.20) as (5.5) and the case $a \neq 1$ of (1.17) and (1.18) as (5.3).

Now, $\det[\mathbf{P}(t) + \mathbf{I}] = \lambda^{-2}$ is invariant, so we need only show that $\mathbf{P}(t)$ is bounded to conclude that the parabola $V(t, \mathbf{y})$ is spatially smooth and uniformly convex, independent of time. But $\log a(t)$ remains bounded since the energy is conserved, whereas the function $H(a, \theta)$ diverges at $a = 0$ and $a = \infty$. It remains only to check the smooth dependence of $(a(t), \theta(t))$ —and hence $\mathbf{P}(t)$ —on time to justify the reformulation in dual variables (2.2)–(2.5) and complete the proof of the theorem. Away from $a = 1$, this follows from smoothness of the vector field (1.17) and (1.18) generating the evolution, or equivalently, of the Hamiltonian function $\tilde{H}(r, \theta)$. The remainder of the proof is, therefore, devoted to showing that $(a(t), \theta(t))$ remain smooth even on the orbit (i.e. energy level) which includes $a = 1$. Let us first understand the structure of the energy $H(a, \theta)$ near this orbit.

Regarding $\lambda > 0$ and the aspect ratio $e^\varphi \neq 1$ of the physical domain as fixed parameters, we shall need to compute the gradient DH and the Hessian D^2H of the energy $H(a, \theta) \in C^\infty(]0, \infty[\times \mathbf{R})$ with respect to the dynamical variables (a, θ) . It is easiest to first compute the derivatives of $\tilde{H}(\sigma, \theta) \in C^\infty(\mathbf{R}^2)$ and $\tilde{z}^2(\sigma, \theta; \varphi)$ with respect to $\sigma = \log a$ and θ . From (1.24) and (1.25), we have

$$\tilde{D}\tilde{z}^2(\sigma, \theta; \varphi) := \begin{bmatrix} \frac{\partial \tilde{z}^2}{\partial \sigma} \\ \frac{\partial \tilde{z}^2}{\partial \theta} \end{bmatrix} = 2 \begin{bmatrix} \sinh \sigma \cosh \varphi + \cosh \sigma \sinh \varphi \cos(2\theta) \\ -2 \sinh \sigma \sinh \varphi \sin(2\theta) \end{bmatrix}, \tag{6.1}$$

$$\begin{aligned} \tilde{D}^2\tilde{z}^2(\sigma, \theta; \varphi) &:= \begin{bmatrix} \frac{\partial^2 \tilde{z}^2}{\partial \sigma^2} & \frac{\partial^2 \tilde{z}^2}{\partial \sigma \partial \theta} \\ \frac{\partial^2 \tilde{z}^2}{\partial \sigma \partial \theta} & \frac{\partial^2 \tilde{z}^2}{\partial \theta^2} \end{bmatrix} \\ &= 2 \begin{bmatrix} \frac{\tilde{z}^2}{2} - 1, & -2 \cosh \sigma \sinh \varphi \sin(2\theta) \\ -2 \cosh \sigma \sinh \varphi \sin(2\theta), & -4 \sinh \sigma \sinh \varphi \cos(2\theta) \end{bmatrix}, \end{aligned} \tag{6.2}$$

$$\tilde{D}\tilde{H}(\sigma, \theta)^T := \begin{bmatrix} \frac{\partial \tilde{H}}{\partial \sigma} & \frac{\partial \tilde{H}}{\partial \theta} \end{bmatrix} = \frac{\lambda}{\tilde{z}} \begin{bmatrix} \frac{\tilde{z}}{\lambda} \sinh \sigma - \frac{1}{2} \frac{\partial \tilde{z}^2}{\partial \sigma}, & -\frac{1}{2} \frac{\partial \tilde{z}^2}{\partial \theta} \end{bmatrix}, \quad (6.3)$$

$$\tilde{D}^2 \tilde{H}(\sigma, \theta) = \frac{\lambda}{4\tilde{z}} \begin{bmatrix} \frac{4\tilde{z}}{\lambda} \cosh \sigma + \left(\frac{1}{\tilde{z}} \frac{\partial \tilde{z}^2}{\partial \sigma}\right)^2 - 2 \frac{\partial^2 \tilde{z}^2}{\partial \sigma^2}, & \frac{1}{\tilde{z}^2} \frac{\partial \tilde{z}^2}{\partial \sigma} \frac{\partial \tilde{z}^2}{\partial \theta} - 2 \frac{\partial^2 \tilde{z}^2}{\partial \sigma \partial \theta} \\ \frac{1}{\tilde{z}^2} \frac{\partial \tilde{z}^2}{\partial \sigma} \frac{\partial \tilde{z}^2}{\partial \theta} - 2 \frac{\partial^2 \tilde{z}^2}{\partial \sigma \partial \theta}, & \left(\frac{1}{\tilde{z}} \frac{\partial \tilde{z}^2}{\partial \theta}\right)^2 - 2 \frac{\partial^2 \tilde{z}^2}{\partial \theta^2} \end{bmatrix}. \quad (6.4)$$

Using (6.1)–(6.4), we deduce that the smooth function $z(a, \theta) \geq 2$ and $H(a, \theta)$ on $a > 0$ both have non-degenerate saddle points at $(a, \theta) = (1, \pm\pi/4)$ and $(1, \pm 3\pi/4)$. In fact, it is easy to compute

$$Dz^2(1, \theta; \varphi)^T := \begin{bmatrix} \frac{\partial z^2}{\partial a} & \frac{\partial z^2}{\partial \theta} \end{bmatrix}_{a=1} = [2 \sinh \varphi \cos(2\theta) \quad 0],$$

$$DH(1, \theta)^T := \begin{bmatrix} \frac{\partial H}{\partial a} & \frac{\partial H}{\partial \theta} \end{bmatrix}_{a=1} = \frac{\lambda}{z(1, \theta; \varphi)} [-\sinh \varphi \cos(2\theta) \quad 0], \quad (6.5)$$

$$D^2 z^2 \left(1, \pm \frac{\pi}{4}; \varphi\right) = 2 \begin{bmatrix} \cosh \varphi & \mp 2 \sinh \varphi \\ \mp 2 \sinh \varphi & 0 \end{bmatrix},$$

$$D^2 H \left(1, \pm \frac{\pi}{4}\right) = \frac{\lambda}{\cosh(\varphi/2)} \begin{bmatrix} \frac{1}{\lambda} \cosh\left(\frac{\varphi}{2}\right) - \frac{1}{2} \cosh \varphi & \pm \sinh \varphi \\ \pm \sinh \varphi & 0 \end{bmatrix}, \quad (6.6)$$

from (6.1)–(6.4) using $\sigma = 0$, $d\sigma/da = 1/a = 1$ and $z(1, \theta; \varphi) = 2 \cosh(\varphi/2)$.

Taylor expanding $H(a, \theta)$ around $a = 1$ and applying the implicit function theorem to

$$h(a, \theta) := \frac{H(a, \theta) - H(1, \theta)}{a - 1} = \frac{H(a, \theta) - H(1, \pi/4)}{a - 1},$$

we find that, locally, the level set $H(a, \theta) = H(1, \pi/4)$ consists of two smooth curves intersecting transversally at the non-degenerate saddles. One of these curves is the circle $a = 1$; (6.6) implies that the other one has a slope

$$\left. \frac{d\theta}{da} \right|_{(a, \theta) = (1, \pm\pi/4)} = -\frac{\partial h / \partial a}{\partial h / \partial \theta} = -\frac{(1/2)\partial^2 H / \partial a^2}{\partial^2 H / \partial \theta \partial a} = \pm \frac{\lambda \cosh \varphi - 2 \cosh(\varphi/2)}{4\lambda \sinh \varphi}, \quad (6.7)$$

as it passes through the saddle $(a, \theta \bmod \pi) = (1, \pm\pi/4)$. Since (6.5) vanishes only at the four saddle points, conservation of energy and the implicit function theorem again show that no trajectory of our dynamical system can reach the unit circle $a = 1$ except by following one of these transverse segments—which from theorem 1.4 will turn out to be part of the same global curve; figures 2 and 3(a). From the radial speed (1.17) and the slope of the curve (6.7), we deduce the angular speed

$$\left. \frac{d\theta}{dt} \right|_{(a, \theta) = (1, \pm\pi/4)} = \frac{da}{dt} \frac{d\theta}{da} = \frac{1}{2} \left(1 - \frac{\lambda \cosh \varphi}{2 \cosh(\varphi/2)}\right) \quad (6.8)$$

(1.18) of this special trajectory as it crosses the singular circle $a = 1$ and continues smoothly to the other side.

Note that when $a(0) = 1$ our convention $\theta(0) = \pi/4$ forces the trajectory onto this special segment. When $a(0) \neq 1$ the unit circle will not be approached subsequently unless $(a(0), \theta(0))$ lies on the same segment, which is why our modification to (1.18) is really relevant only at the four saddles $(\theta \bmod \pi/2) = \pi/4$.

To complete the proof of the theorem, we still need to deduce that the evolution $(a(t), \theta(t))$ gives a $C^\infty([0, \infty[)$ parameterization of this smooth segment as it crosses the unit circle. This follows from the fundamental theorem of calculus if we can argue that the restriction of the vector field (1.17) and (1.18) to the segment gives a smooth tangent field. Tangency holds by construction and (6.8). From (6.4) and the smoothness of $\sigma = \log a$, we see that $\partial^2 H / \partial a \partial \theta$ is smooth and non-vanishing (6.6) near $(a, \theta) = (1, \pi/4)$, since $\varphi \neq 0$. Thus, the slope (6.7) is a smooth function of (a, θ) along the segment in question. Since the radial speed (1.17) is a smooth function of both parameters (a, θ) , we conclude that the angular speed (6.8) is also a smooth function (of say arclength) along the curve. Thus, the tangent field is as smooth as desired.

The solution to (5.3) must vary smoothly. Theorem 5.1 and the foregoing then assert that we have found smoothly evolving matrices defining quadratic solutions to the dual (2.2)–(2.5) and primal (1.5)–(1.6) dynamics except possibly at an isolated sequence of times when $\sigma(t_0) = 0$. On the other hand, theorem A.1 implies the primal dynamics has a unique quadratic solution which depends analytically on time. Since our two smooth solutions agree except at an isolated sequence of times, they must coincide. This concludes the proof of the theorem.

For use in subsequent sections, note that logarithmically rescaling the polar radius $(\sigma, \theta) = (\log a, \theta)$ of the Cartesian plane collapses the circle $a = 1$ to the origin, while the four saddles at $\pm\pi/4 \bmod \pi$ combine to form a single simple saddle at the origin whose arms cross orthogonally. \square

7. Energy landscape and bifurcation structure (proof of theorem 1.4)

Proof. Note that \tilde{H} is a smooth function in all four parameters since $\tilde{z} \geq 2$. The local saddle topology at the origin $\sigma = 0$ of our polar coordinate system was elucidated at the end of the preceding section; we seek the remaining critical points. The angular derivative from (6.1)–(6.3),

$$\frac{\partial \tilde{H}}{\partial \theta} = \frac{2\lambda}{\tilde{z}} \sinh \sigma \sinh \varphi \sin(2\theta), \tag{7.1}$$

vanishes only on the horizontal axis—where $\tilde{H}(\sigma, \cdot)$ is a minimum—and the vertical axis, where $\tilde{H}(\sigma, \cdot)$ is maximized. Let us study the radial derivatives of $\tilde{H}(\sigma, 0) =: h_-(\cosh(\sigma/2))$ and $\tilde{H}(\sigma, \pi/2) =: h_+(\cosh(\sigma/2))$ using the new variable $R = \cosh(\sigma/2) \geq 1$ to parameterize these axes. With $\cos(2\theta) = \mp 1$ we compute $\tilde{z} = 2 \cosh((\sigma \mp \varphi)/2)$ in (1.24) and (1.25), whence

$$h_{\pm}(R) = \left(R - \lambda \cosh \frac{\varphi}{2} \right)^2 + \left(\sqrt{R^2 - 1} \pm \lambda \sinh \frac{\varphi}{2} \right)^2, \tag{7.2}$$

$$\frac{dh_{\pm}}{dR} = 4R - 2\lambda \cosh \frac{\varphi}{2} \pm 2\lambda \frac{\sinh(\varphi/2)}{\sqrt{1 - R^{-2}}}, \tag{7.3}$$

$$\frac{d^2 h_{\pm}}{dR^2} = 4 \mp 2\lambda \frac{\sinh(\varphi/2)}{\sqrt{(R^2 - 1)^3}}. \tag{7.4}$$

The condition $h'_{\pm}(R) = 0$ is equivalent to (1.28). Along the horizontal axis, its second derivative shows $h_-(R)$ to be uniformly convex, so there can be at most one point $R_- > 1$ where $h'_-(R_-) = 0$; there is exactly one such point since $h'_-(1) < 0 < h'_-(\infty)$. The corresponding critical point $(\sigma_-, \theta_-) := (2 \cosh^{-1}(R_-), 0)$ must be a global minimum, since

$\tilde{z}(\sigma, \theta; \varphi) \leq \tilde{z}(\sigma, 0; \varphi)$ implies $\tilde{H}(\sigma, \theta) \geq h_-(\cosh(\sigma/2))$ enjoys unbounded radial growth, and all other critical points will be shown to be saddles or maxima.

Along the vertical axis, monotonicity of its second derivative shows $h_+(R)$ has a unique point of inflection R_* given by

$$\sinh^3 \frac{\sigma_*}{2} = (R_*^2 - 1)^{3/2} = \frac{\lambda}{2} \sinh \frac{\varphi}{2}. \quad (7.5)$$

Since $h'_+(1) = h'_+(\infty) = +\infty$, the sign of the minimum slope $h'_+(R_*)$ determines whether we have zero or two critical points $h'_+(R_+) = h'_+(R_2) = 0$.

The bifurcation from zero to two critical points can only occur along the curve in the parameter space (λ, φ) where

$$h'_+(R_*) = 4 \left(1 + \frac{(\lambda/2) \sinh(\varphi/2)}{((\lambda/2) \sinh(\varphi/2))^{1/3}} \right)^{3/2} - 2\lambda \cosh \frac{\varphi}{2}$$

vanishes. This yields an equation

$$\left(\left(\frac{\lambda}{2} \sinh \frac{\varphi}{2} \right)^{2/3} + 1 \right)^3 = \left(\frac{\lambda}{2} \right)^2 \left(1 + \sinh^2 \frac{\varphi}{2} \right),$$

which turns out to be quadratic rather than cubic in $\sinh^{2/3}(\varphi/2)$. The explicit solution φ is positive only if $\lambda \geq 2$, in which case $\varphi = \varphi_{cr}(\lambda)$ is a non-decreasing function of λ given by (1.27). To see that the two critical points $1 < R_+ < R_2$ exist if and only if $\varphi < \varphi_{cr}(\lambda)$, recall that $h_+(R)$ depends on φ and λ as well as R . It is enough to verify negativity of the mixed partial $0 > \partial^2 h_+ / \partial \lambda \partial R$ at $(\lambda, \varphi_{cr}(\lambda), R_*)$, as this states that the minimum slope of h_+ decreases through zero as λ crosses the bifurcation curve. From (7.3) we see the desired negativity is equivalent to $R_* =: \cosh(\sigma_*/2) \geq \cosh(\varphi_{cr}/2)$. This in turn can be verified directly using (7.5) and (1.27) to express the desired inequality in terms of $\lambda \geq 2$ alone.

For $\varphi < \varphi_{cr}(\lambda)$, order the critical radii $R_+ =: \cosh(\sigma_+/2) < R_* < \cosh(\sigma_2/2) := R_2$ along the positive vertical axis. Both inequalities degenerate to equalities when $\varphi = \varphi_{cr}(\lambda)$. The sign of the second derivative shows R_+ to be a local maximum and R_2 a local minimum for $h_+(R)$, so the ordering (1.29) has been established. Negativity of

$$\frac{\partial^2 \tilde{H}}{\partial \theta^2} \Big|_{(\sigma, \pi/2)} = -\frac{\lambda}{2} \left[\frac{1}{\tilde{z}} \frac{\partial^2}{\partial \theta^2} \tilde{z}^2 \right]_{(\sigma, \pi/2)} = \frac{4\lambda \sinh \sigma \sinh \varphi}{\tilde{z}} \cos(\pi) < 0 \quad (7.6)$$

from (6.2)–(6.4) shows that the bifurcation produces a local maximum of $\tilde{H}(\sigma, \theta)$ at $(\sigma_+, \pi/2)$ and a saddle point at $(\sigma_2, \pi/2)$ and that these critical points are non-degenerate except when $\varphi = \varphi_{cr}(\lambda)$. (Mixed partials of \tilde{H} vanish along both axes by reflection symmetry $\tilde{H}(\sigma, (\pi/2) + \theta) = \tilde{H}(\sigma, (\pi/2) - \theta)$.)

Let us, finally, address the relative heights h_1 and h_2 of the two saddle values in (1.29). For a circular domain $\sigma = 0$, lemma 1.2 asserts that $h_+(R)$ has no critical points save a unique global minimum at $R_- > 1$. On a near circular domain $h_+(R)$ must find its minimum nearby—at R_2 since it cannot occur at R_+ or $R_1 = 1$. Thus, $h_1 > h_2$ persists for $\sigma > 0$ small enough (depending on $\lambda > 2$), and region V := $\{(\lambda, \varphi) \mid h_1 > h_2\}$ borders the $\varphi = 0$ axis of parameter space for all $\lambda > 2$, as in figure 1. Similarly, $h_1 < h_2 = h_+$ along the critical curve $\varphi = \varphi_{cr}$, and a slight perturbation of this inequality yields a neighbourhood under the critical curve belonging to region III := $\{(\lambda, \varphi) \mid h_1 < h_2\}$. To see that a continuously increasing curve separates region III from region V, we show that the boundary between these regions can be expressed as a graph over either λ or φ . This boundary must lie strictly between the $\lambda > 2$ axis and the critical curve $\varphi = \varphi_{cr}(\lambda)$. We do this by exploiting monotonic dependences of the heights $h_1 = h_+^{\lambda, \varphi}(1)$ and $h_2 = h_+^{\lambda, \varphi}(R_2(\lambda, \varphi))$ on the parameters λ and φ .

At both saddles $i = 1, 2$, the chain rule yields

$$\begin{aligned} \frac{\partial h_i}{\partial \varphi} \Big|_{(\lambda, \varphi)} &= \frac{\partial h_+}{\partial \varphi} \Big|_{R_i} + \frac{\partial h_+}{\partial R} \Big|_{(\lambda, \varphi, R_i)} \frac{\partial R_i}{\partial \varphi} \Big|_{\lambda, \varphi} \\ &= \frac{\partial h_+}{\partial \varphi} \Big|_{R_i}. \end{aligned} \tag{7.7}$$

For one saddle, the product of derivatives vanished because $R_1 = 1$ is independent of (λ, φ) ; for the other it vanished because R_2 is a local minimum of $h_+(R)$. Thus,

$$\begin{aligned} \frac{\partial(h_2 - h_1)}{\partial \varphi} \Big|_{(\lambda, \varphi)} &= \lambda \frac{\partial \tilde{z}}{\partial \varphi} \Big|_{(\sigma_1, \pi/2; \varphi)} - \lambda \frac{\partial \tilde{z}}{\partial \varphi} \Big|_{(\sigma_2, \pi/2; \varphi)} \\ &= \lambda \sinh\left(\frac{\varphi - 0}{2}\right) - \lambda \sinh\left(\frac{\varphi - \sigma_2}{2}\right) \\ &> 0 \end{aligned}$$

from (1.24). Now, if $h_1 \leq h_2$ at some parameter values (λ, φ_{III}) , with the reverse inequality $h_1 \geq h_2$ holding at (λ, φ_V) , it follows that $\varphi_V \leq \varphi_{III}$. If $h_1 = h_2$ then $\varphi_{III} = \varphi_V$. Thus, we can find a function $\varphi_{bi} :]2, \infty[\rightarrow]0, \infty[$ strictly less than φ_{cr} , with $h_1 < h_2$ above the curve $\varphi = \varphi_{bi}(\lambda)$ and $h_1 > h_2$ below it.

The same Feynman–Hellmann trick (7.7) yields

$$\frac{\partial h_i}{\partial \lambda} \Big|_{(\lambda, \varphi)} = \frac{\partial h_+^{\lambda, \varphi}}{\partial \lambda} \Big|_{(\lambda, \varphi, R_i(\lambda, \varphi))}.$$

Combining

$$\begin{aligned} h_1 - h_2 &= \cosh(0) - \lambda \tilde{z}\left(0, \frac{\pi}{2}; \varphi\right) - \cosh \sigma_2 + \lambda \tilde{z}\left(\sigma_2, \frac{\pi}{2}; \varphi\right) \\ &< \lambda \tilde{z}\left(\sigma_2, \frac{\pi}{2}; \varphi\right) - \lambda \tilde{z}\left(0, \frac{\pi}{2}; \varphi\right) \end{aligned}$$

with

$$\begin{aligned} \frac{\partial(h_1 - h_2)}{\partial \lambda} \Big|_{(\lambda, \varphi)} &= \tilde{z}\left(0, \frac{\pi}{2}; \varphi\right) - \tilde{z}\left(\sigma_2, \frac{\pi}{2}; \varphi\right) \\ &> \frac{h_1 - h_2}{\lambda} \end{aligned}$$

gives a strict Gronwall’s inequality. Thus, moving to the right from a point (λ_{IV}, φ) where $h_1 \geq h_2$ can only yield points (λ_V, φ) where $h_1 > h_2$. This shows that there is also a function $\lambda_{bi} :]0, \infty[\rightarrow]2, \infty[$ with $h_1 > h_2$ holding to the right of the curve $\lambda = \lambda_{bi}(\varphi)$, and $h_1 < h_2$ holding to its left (wherever h_2 is defined).

Taken together, the existence of both functions $\varphi = \varphi_{bi}(\lambda)$ and $\lambda = \lambda_{bi}(\varphi)$ implies that each is monotone. The equality $h_1 = h_2$ must hold on the graph of this non-decreasing curve separating $h_1 > h_2$ from $h_1 < h_2$. Neither $\varphi_{bi}(\lambda)$ nor $\lambda_{bi}(\varphi)$ can be constant on any interval, without violating single-valuedness of the other function by the strict inequalities above. Thus, the bifurcation curve is continuous with respect to either parameter. It satisfies (1.30), which asserts the equality of $h_2 = \tilde{H}(\sigma_2, \pi/2) = \cosh(\sigma_2) + \lambda^2 \cosh \varphi - 2\lambda \cosh((\sigma_2 - \varphi)/2)$ with $h_1 = \tilde{H}(0, \pi/2)$.

Turning to the form of the saddle at the origin, we recall from the proof of theorem 1.1 that the level set $\{(\sigma \cos \theta, \sigma \sin \theta) \in \mathbf{R}^2 \mid \tilde{H}(\sigma, \theta) = h_1\}$ has a smooth orthogonal self-intersection there. The four arms of this simple saddle cannot extend to $\sigma = \infty$ since $\tilde{H}(\sigma, \theta)$ grows, nor can they end except at another saddle or degenerate critical point. We just proved

that all other such points lie at a different energy level $h_2 \neq h_1$ unless $\varphi = \varphi_{bi}(\lambda)$. Except, in this case, the four arms of the saddle are formed from two smooth curves beginning and ending at $\sigma = 0$, which do not cross each other elsewhere. The origin, therefore, lies on a connected component of the h_1 level set consisting only of a smooth figure-eight curve. It is easy to see that this figure-eight curve encloses the local maxima or minima σ_{\pm} depending on the sign of $h_1 - h_2$.

Although the theorem is now proved, for use in section 1.8, we record second derivatives with respect to the canonically conjugate variables $r = 2R^2 - 1$ along the coordinate axes $\theta_+ = \pi/2$ and $\theta_- = 0$:

$$\left. \frac{\partial^2 \bar{H}}{\partial r^2} \right|_{(r, \theta_{\pm})} = h''_{\pm}(R) \left(\frac{dR}{dr} \right)^2 + h'_{\pm}(r) \frac{d^2 R}{dr^2}.$$

Using $R'(r) = 1/4R$ and (7.4) at $r_{\pm} = \cosh(\sigma_{\pm})$ yields

$$\left. \frac{\partial^2 \bar{H}}{\partial r^2} \right|_{(r_{\pm}, \theta_{\pm})} = \frac{1}{4 \cosh^2(\sigma_{\pm}/2)} \left(1 - \frac{\lambda \sinh(\varphi/2)}{2 \sinh^3(\pm \sigma_{\pm}/2)} \right). \quad (7.8)$$

Now, (7.6) combines with non-vanishing of the mixed partials at the critical points to give the Hessian determinant:

$$\det [D^2 \bar{H}]_{(r_{\pm}, \theta_{\pm})} = - \frac{2\lambda \sinh(\pm \sigma_{\pm}) \sinh \varphi}{\cosh((\sigma_{\pm} \mp \varphi)/2)} \left. \frac{\partial^2 \bar{H}}{\partial r^2} \right|_{(r_{\pm}, \theta_{\pm})}. \quad (7.9)$$

□

Acknowledgments

It is a pleasure to acknowledge many fruitful conversations with Michael Cullen, who was also responsible for introducing us to semi-geostrophy. We are grateful to the Universities of Texas (RJM and AMO), Toronto (AMO) and Wisconsin (RJM), for their generous hospitality during the course of this work, and to United States National Science Foundation grants DMS 0074037, 0354729 and Natural Sciences and Engineering Research Council of Canada grant 217006-99 RGPIN for providing financial support. We thank Maxim Trokhimtchouk for preparing the figures, and for authoring the following appendix featuring results he discovered while working with us on an NSERC undergraduate research project.

Appendix A. Well-posed SG circulations in an elliptical ocean (by Maxim Trokhimtchouk)

This appendix collects some results obtained by Maxim Trokhimtchouk at the University of Toronto, while pursuing an NSERC summer undergraduate research project under the guidance of R J McCann.

It begins by addressing well-posedness of the nonlinear differential equation (1.12) governing a SG fluid in an ellipse under the quadratic-potential ansatz (1.10) and (1.11). A continuous function $q : \mathbf{R} \rightarrow \mathbf{R}$ and matrix $P(t)$ with continuously differentiable coefficients will be called a *classical solution* to this system if $P(t) = P(t)^T$ and (1.12) holds for all $t \in \mathbf{R}$. Reasoning directly in physical variables, the existence, uniqueness, and analyticity of classical solutions will be demonstrated whether or not Cullen and Purser's criterion $P(0) > -I$ is satisfied. Instead, the hypothesis $\det [(P(0) + I)F^2] \neq 0$ introduced below guarantees finite circulation speed $q(t)$ at one and hence all times. The use of physical instead of dual coordinates

resolves the singularity of the Hamiltonian $\bar{H}(r, \theta)$ from (1.21) at the boundary of its phase space $(r, \theta) \in [1, \infty[\times \mathbf{R}$.

Conservation of energy and potential vorticity are observed to constrain the pressure matrix coefficients to the intersection of an ellipsoid (A.6) and hyperboloid (A.7), respectively. The hyperboloid is double-sheeted, single-sheeted, or degenerates to a standard light cone, according to the sign $\kappa \in \{\pm 1, 0\}$ of the potential vorticity $\det [P(0) + I] = \kappa/\lambda^2$. Moreover, the key formula (A.1) shows that the direction of fluid circulation never reverses: it is determined by the invariants (λ, φ) and H_{SG} alone, and will be cyclonic if and only if the energy $H_{SG} > (\pi/4)(\lambda^2 - \kappa) \cosh \varphi$, or equivalently, if and only if the pressure $P(0, \mathbf{y})$ is initially sub/superharmonic—depending on the sign of $\det [(P(0) + I)F^2]$. Anticyclonic circulations account for most of the orbits depicted in figure 3, but since $\kappa = +1$ they can only occur when $\lambda > 1$ in figure 1. An auxiliary result shows that the condition $\lambda > 1$ is sufficient as well as necessary to guarantee anticyclonic circulation at low energies.

Theorem A.1 (analytic evolution of linear SG fluid circulations in an ellipse). *If $F > 0$ and $P(0)$ are symmetric matrices and $\det [(P(0) + I)F^2] \neq 0$, then the differential equation (1.12) admits a unique classical solution $q(t)$ and $P(t) = P(t)^T$, and it varies analytically with $t \in \mathbf{R}$. If $\det [P(0) + I] = \kappa/\lambda^2$ for $\kappa \in \{+1, 0, -1\}$ and $F^2 = \lambda S_\varphi$ from (3.2), then*

$$q(t) = \frac{\lambda^2 \text{trace } P(t)}{\text{trace } [(P(t) + I)F^2]} = 2\lambda \frac{(4/\pi)H_{SG} + (\kappa - \lambda^2) \cosh \varphi}{\text{trace}^2 [(P(t) + I)F^2]} \tag{A.1}$$

and the denominator is not zero unless both numerators and $q(t)$ vanish perpetually.

Proof. Rotating and rescaling space if necessary, it costs no generality to assume $F^2 = \lambda S_\varphi > 0$ is diagonal and $\det [P(0) + I] = \kappa/\lambda^2$ for some $\kappa \in \{+1, 0, -1\}$. With this normalization, we begin by deducing that the hypotheses of the theorem guarantee the four differential equations (1.12) relating $q(t)$ to the symmetric matrix $V(t) = P(t) + I$; namely

$$-q(t)JV(t)JF^{-2} = JV'(t) + P(t), \tag{A.2}$$

admits a unique solution in the classical sense, for a short time $T > 0$ whose length might depend on initial data.

Antisymmetry of J implies $JV'(t)$ is traceless, so taking traces of both sides, the identity $JF^{-2}J = -F^2/\det F^2 = -F^2/\lambda^2$ yields

$$q(t)\text{trace } [(P(t) + I)F^2] = \lambda^2 \text{trace } P(t), \tag{A.3}$$

and the first equality in (A.1) follows if its denominator is non-zero. Introducing the notation

$$P(t) = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} u - 1 & w \\ w & v - 1 \end{pmatrix} = V(t) - I, \tag{A.4}$$

(A.1) and (A.2) motivate consideration of the differential equation

$$V'(t) = J(V(t) - I) - \frac{\lambda^2 \text{trace } P(t)}{\text{trace } [(P(t) + I)F^2]} V(t) JF^{-2},$$

which takes the explicit form

$$\begin{pmatrix} u' & w' \\ w' & v' \end{pmatrix} = \begin{pmatrix} -w & 1 - v \\ u - 1 & w \end{pmatrix} - \frac{u + v - 2}{ue^\varphi + ve^{-\varphi}} \begin{pmatrix} e^{-\varphi} w & -e^\varphi u \\ e^{-\varphi} v & -e^\varphi w \end{pmatrix} \tag{A.5}$$

instead of (A.10). Consistency of the two expressions for $w'(t)$ is easily checked, so this system of four equations for three unknown functions is not overdetermined. Since the right-hand side is a rational function of (u, v, w) , the standard theory of ordinary differential equations asserts a unique, analytic solution $V(t)$ of (A.5) exists locally in time [2], as long as $u(t)e^\varphi + v(t)e^{-\varphi} \neq 0$.

Our hypothesis $u(0)e^\varphi + v(0)e^{-\varphi} \neq 0$ guarantees these smooth solutions to (A.5) exist, at least for a short time $T > 0$. Defining $q(t)$ and $P(t)$ by (A.1) and (A.4), gives the existence and uniqueness of the desired solution to (A.2) for $t \in [0, T[$.

The remainder of the argument is devoted to showing these solutions are actually global in time, meaning T can be replaced by $+\infty$. To derive a contradiction, assume $T < +\infty$ denotes the maximal time a classical solution exists for some admissible initial condition. As long as the matrices $P(t)$ and $Q(t) = q(t)F^{-2}$ satisfy the ordinary differential equation (A.2), the quadratic functions (1.10) and (1.11) satisfy the SG system (1.5) and (1.6) on the elliptical fluid domain Y . This implies that the potential vorticity $\det [P(t) + I] = \kappa/\lambda^2$ and SG energy (1.9)

$$\begin{aligned} H_{SG} &= \frac{1}{2} \int_{F(B_1)} |P(t)\mathbf{y}|^2 d^2\mathbf{y} \\ &= \frac{\det F}{2} \int_{B_1} |P(t)F\mathbf{z}|^2 d^2\mathbf{z} \\ &= \frac{\pi\lambda}{8} \text{trace} [P(t)^2 F^2] \end{aligned}$$

given by (4.6) remain invariant on the time interval $[0, T[$. In other words, the coefficients (A.4) of $P(t)$ evolve on the intersection of the ellipsoid

$$H_{SG} = \frac{\pi\lambda^2 \cosh \varphi}{4} \left(\frac{e^\varphi}{e^\varphi + e^{-\varphi}}(u - 1)^2 + \frac{e^{-\varphi}}{e^\varphi + e^{-\varphi}}(v - 1)^2 + w^2 \right) \tag{A.6}$$

with the hyperboloid

$$\frac{\kappa}{\lambda^2} = \det [I + P(t)] = uv - w^2 \tag{A.7}$$

in the vector space $(u, v, w) \in \mathbf{R}^3$ parameterizing symmetric 2×2 matrices. Since the ellipsoid is compact, the initial energy $H_{SG} < +\infty$ constrains (u, v, w) to a bounded subset of \mathbf{R}^3 , so a subsequential limit $(u(t_n), v(t_n), w(t_n)) \rightarrow (u(T), v(T), w(T))$ exists as $t_n \rightarrow T$. Summing (A.6) and (A.7) yields

$$\begin{aligned} \frac{4H_{SG}}{\pi\lambda^2 \cosh \varphi} + \frac{\kappa}{\lambda^2} - 1 &= (u + v - 2) \left(\frac{e^\varphi}{e^\varphi + e^{-\varphi}}u + \frac{e^{-\varphi}}{e^\varphi + e^{-\varphi}}v \right) \\ &= \text{trace} \left[\frac{P(t)}{\cosh \varphi} \right] \text{trace} [(P(t) + I)S_\varphi]/2. \end{aligned} \tag{A.8}$$

Compactness of the ellipsoid (A.6) guarantees both factors on the right-hand side of (A.8) remain bounded in magnitude, and therefore bounded away from zero if the constant left-hand side does not vanish. In this case, $\text{trace}[(P(t) + I)F^2] \neq 0$ for all $t \in [0, T[$, which means equation (A.2) admits a unique analytic solution on a longer time interval $[0, T + \epsilon[$, contradicting maximality of $T < +\infty$. Since the dynamics is reversible in time, the only remaining escape is for unique global solutions to exist which are analytic in $t \in \mathbf{R}$. Multiplying (A.8) by $\lambda^2 \cosh \varphi$ yields

$$\text{trace } P(t) = \frac{(4/\pi)H_{SG} + (\kappa - \lambda^2) \cosh \varphi}{\lambda \text{trace} [(P(t) + I)F^2]/2} \tag{A.9}$$

with non-zero denominator, and we recover the second identity (A.1) from the first.

This concludes the theorem, unless the left-hand side of (A.8) vanishes. In the latter case, we must first argue that $q(t) = 0$ for any classical solution. Otherwise, the continuous function $q(t)$ is non-vanishing on some interval of time, during which (A.3) implies the factors on the right-hand side of (A.8) must both vanish, since they cannot vanish separately. Vanishing of both factors determines the values $u(t) = -e^{-\varphi}/\sinh \varphi$ and $v(t) = e^\varphi/\sinh \varphi$ uniquely on

this same interval, and forces $\varphi \neq 0$. However, these unchanging values of $u(t)$ and $v(t)$ prove inconsistent with the postulated dynamics

$$\begin{pmatrix} u' & w' \\ w' & v' \end{pmatrix} = \begin{pmatrix} -w & 1-v \\ u-1 & w \end{pmatrix} - q(t) \begin{pmatrix} e^{-\varphi}w & -e^{\varphi}u \\ e^{-\varphi}v & -e^{\varphi}w \end{pmatrix}; \tag{A.10}$$

we rapidly derive $w(t) = 0$ and the contradiction $u(t) = v(t)$. The only conclusion can be that $q(t) = 0$ perpetually, in which case trace $P(t)$ vanishes according to (A.3), and (A.10) reduces to an inhomogeneous linear system which admits unique, global, explicit solutions that are analytic in time. \square

A final proposition shows that in the most relevant case $\kappa = 1$, anticyclonic circulations occur if and only if the energy is sufficiently low and $\lambda > 1$. This can be understood intuitively as follows. Since the energy H_{SG} is non-negative, $q(t)$ cannot be negative if $\lambda^2 \leq \kappa$ in (A.1). On the other hand, when $\lambda = 1$ the domain $Y \subset \mathbf{R}^2$ and dual ellipse $X(t)$ have the same area, so the minimum energy $H_{SG} = 0$ is attained when $X(t) = Y$ since the transportation cost vanishes. This is a borderline case $q(t) = 0$ for anticyclonic rotation.

Proposition A.2 (conditions for anticyclonic circulation). *Fix $\lambda \geq 1$ and $\varphi \geq 0$, and consider the function $\bar{H}(r, \theta)$ in polar coordinates (1.21). Let h_- denote its minimum value and $h_0 := (\lambda^2 - 1) \cosh \varphi$ be the energy of a fluid which circulates with the domain, so $q(t) = 0$. Then, $h_0 \geq h_-$ with equality precisely when $\lambda = 1$.*

Proof. First, recall that

$$\cosh(u + v) + \cosh(u - v) = 2 \cosh u \cosh v, \tag{A.11}$$

$$\sinh(u + v) + \sinh(u - v) = 2 \cosh v \sinh u. \tag{A.12}$$

Taking $\sigma + \varphi = 2u$ and $\sigma - \varphi = 2v$ yields

$$\sinh \sigma - \sinh \varphi = 2 \sinh \sigma - 2 \cosh \frac{\sigma - \varphi}{2} \sinh \frac{\sigma + \varphi}{2}. \tag{A.13}$$

Also, recall $h_- = \lambda^2 \cosh \varphi + \cosh \sigma_- - 2\lambda \cosh((\sigma_- + \varphi)/2)$ from (1.25), where σ_- is given by (1.28)

$$2\lambda \sinh \frac{\sigma_- + \varphi}{2} = 2 \sinh \sigma_-. \tag{A.14}$$

Hence, $h_- - h_0 = \cosh \sigma_- + \cosh \varphi - 2\lambda \cosh((\sigma_- + \varphi)/2)$, and by (A.11) it follows that

$$\begin{aligned} h_- - h_0 &= 2 \cosh \frac{\sigma_- - \varphi}{2} \cosh \frac{\sigma_- + \varphi}{2} - 2\lambda \cosh \frac{\sigma_- + \varphi}{2} \\ &= 2 \cosh \frac{\sigma_- + \varphi}{2} \left[\cosh \frac{\sigma_- - \varphi}{2} - \lambda \right]. \end{aligned} \tag{A.15}$$

Assuming $h_- - h_0 > 0$ to derive a contradiction, yields $\lambda < \cosh((\sigma_- - \varphi)/2)$. Subtracting (A.14) from (A.13) we obtain

$$\sinh \sigma_- - \sinh \varphi = 2 \left[\lambda - \cosh \frac{\sigma_- - \varphi}{2} \right] \sinh \frac{\sigma_- + \varphi}{2} < 0,$$

hence $\varphi > \sigma_-$. Therefore, $\sinh((\sigma_- + \varphi)/2) > \sinh \sigma_-$, which contradicts $\lambda \geq 1$ in (A.14). The conclusion must be that $h_- \leq h_0$. Equality implies $\lambda = \cosh((\sigma_- - \varphi)/2)$ in (A.15), and we then see from (A.13) and (A.14) that $\sigma_- = \varphi$ and $\lambda = 1$. Conversely, $\lambda = 1$ is easily seen to yield $h_- = 0 = h_0$, concluding the proposition. \square

References

- [1] Benamou J-D and Brenier Y 1998 Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem *SIAM J. Appl. Math.* **58** 1450–61
- [2] Birkhoff G and Rota G-C 1989 *Ordinary Differential Equations* 4th edn (New York: Wiley)
- [3] Chandrasekhar S 1969 *Ellipsoidal Figures of Equilibrium* (New Haven, CT: Yale University Press)
- [4] Charney J G 1947 The dynamics of long waves in a baroclinic westerly current *J. Meteorol.* **4** 135–62
- [5] Constantin P, Majda A J and Tabak E 1994 Formation of strong fronts in the 2-d quasi-geostrophic thermal active scalar *Nonlinearity* **7** 1495–533
- [6] Cordoba D and Fefferman C 2002 Growth of solutions for QG and 2D Euler equations *J. Am. Math. Soc.* **15** 665–9
- [7] Cullen M J P 1983 Solutions to a front forced by deformation *Q. J. R. Meteorol. Soc.* **109** 565–73
- [8] Cullen M J P 2002 Large-scale non-turbulent dynamics in the atmosphere *Q. J. R. Meteorol. Soc.* **128** 2623–39
- [9] Cullen M J P and Gangbo W 2001 A variational approach for the 2-D semi-geostrophic shallow water equations *Arch. Ration. Mech. Anal.* **156** 241–73
- [10] Cullen M J P and Maroofi H 2003 The fully compressible semi-geostrophic system from meteorology *Arch. Ration. Mech. Anal.* **167** 309–36
- [11] Cullen M J P and Purser R J 1984 An extended Lagrangian theory of semigeostrophic frontogenesis *J. Atmos. Sci.* **41** 1477–97
- [12] Cullen M J P and Purser R J 1989 Properties of the Lagrangian semigeostrophic equations *J. Atmos. Sci.* **46** 2684–97
- [13] Cullen M J P, Norbury J and Purser R J 1991 Generalised Lagrangian solutions for atmospheric and oceanic flows *SIAM J. Appl. Math.* **51** 20–31
- [14] Cullen M J P, Norbury J, Purser R J and Shutts G J 1987 Modelling the quasi-equilibrium dynamics of the atmosphere *Q. J. R. Meteorol. Soc.* **113** 735–57
- [15] Cullen M J P, Douglas R J, Roulstone I and Sewell M J Generalised semigeostrophic theory on a sphere *J. Fluid Mech.* submitted
- [16] Cushman-Roisin B 1987 Exact analytical solution for elliptical vortices of the shallow-water equations *Tellus A* **39** 235–44
- [17] Dedekind R 1860 Zusatz zu der vorstehenden Abhandlung *J. Reine Angew. Math.* **58** 217–28
- [18] Dirichlet G L 1860 Untersuchungen über ein Problem der Hydrodynamik *J. Reine Angew. Math.* **58** 181–216
- [19] Dowson D C and Landau B V 1982 The Fréchet distance between multivariate normal distributions *J. Multivariate Anal.* **12** 450–5
- [20] Eliassen A 1948 The quasi-static equation of motion with pressure as independent variable *Geofys. Publ.* **17** 1–44
- [21] Givens C R and Shortt R M 1984 A class of Wasserstein metrics for probability distributions *Michigan Math. J.* **31** 231–40
- [22] Hoskins B J 1975 The geostrophic momentum approximation and the semi-geostrophic equations *J. Atmos. Sci.* **32** 233–42
- [23] Hoskins B J 1982 The mathematical theory of frontogenesis *Ann. Rev. Fluid Mech.* **14** 131–51
- [24] Hoskins B J and Bretherton F P 1972 Atmospheric frontogenesis models: mathematical formulation and solutions *J. Atmos. Sci.* **29** 11–37
- [25] Kida S 1981 Motion of an elliptic vortex in a uniform shear-flow *J. Phys. Soc. Japan* **50** 3517–20
- [26] Kirchhoff G 1876 *Vorlesungen über Mathematische Physik: Mechanik* (Leipzig: Teubner)
- [27] Knott M and Smith C S 1984 On the optimal mapping of distributions *J. Optim. Theory Appl.* **43** 39–49
- [28] Majda A J and Bertozzi A L 2002 Vorticity and incompressible flow *Cambridge Texts in Applied Mathematics* vol 27 (Cambridge: Cambridge University Press)
- [29] Meacham S P, Pankratov K K, Shchepetkin A S and Zhmur V V 1994 Ellipsoidal vortices in background shear and strain flows *Dynam. Atmos. Oceans* **21** 167–212
- [30] Meacham S P, Morrison P J and Flierl G R 1997 Hamiltonian moment reduction for describing vortices in shear *Phys. Fluids* **9** 2310–28
- [31] Moore D and Saffman P 1971 Structure of a line vortex in an imposed strain *Aircraft Wake Turbulence* ed A Goldberg *et al* (New York: Plenum) pp 339–54
- [32] Neu J C 1984 The dynamics of a columnar vortex in an imposed strain *Phys. Fluids* **27** 2397–402
- [33] Norbury J and Roulstone I (ed) 2002 *Large-Scale Atmosphere-Ocean Dynamics* vol 1 *Analytical Methods and Numerical Models* vol 2 *Geometric Methods and Models* (Cambridge: Cambridge University Press)
- [34] Olkin I and Pukelsheim F 1982 The distance between two random vectors with given dispersion matrices *Linear Algebra Appl.* **48** 257–63

- [35] Pedlosky J 1987 *Geophysical Fluid Dynamics* 2nd edn (New York: Springer)
- [36] Purser R J and Cullen M J P 1987 A duality principle in semi-geostrophic theory *J. Atmos. Sci.* **44** 3449–68
- [37] Riemann B 1860 Untersuchen über die Bewegung eines flüssigen gleichartigen Ellipsoides *Abh. Königl. Gesell. Wis. zu Göttingen* **9** 3–36
- [38] Rogers C 1989 Elliptic warm-core theory: the pulsatron *Phys. Lett.* **138** 267–73
- [39] Roulstone I and Norbury J 1994 A Hamiltonian structure with contact geometry for the semi-geostrophic equations *J. Fluid Mech.* **272** 211–33
- [40] Salmon R 1988 Semigeostrophic theory as a Dirac-bracket projection *J. Fluid Mech.* **196** 345–58
- [41] Salmon R 1998 *Lectures on Geophysical Fluid Dynamics* (New York: Oxford University Press)
- [42] Schubert W H, Ciesielecki P E, Stevens D E and Kuo H-C 1991 Potential vorticity modelling of the ITCZ and the Hadley circulation *J. Atmos. Sci.* **48** 1493–509
- [43] Shutts G and Cullen M J P 1987 Parcel stability and its relation to semigeostrophic theory *J. Atmospheric Sci.* **44** 1318–30
- [44] Thacker W C 1981 Some exact solutions to the non-linear shallow-water wave equations *J. Fluid Mech.* **107** 499–508
- [45] Villani C 2003 Topics in Optimal Transportation *Graduate Studies in Mathematics* vol 58 *Am. Math. Soc.* (Providence, RI: American Mathematical Society)
- [46] Loeper G 2003 Applications de l'équation de Monge-Ampère à la modélisation des fluides et des plasmas *PhD thesis* Université de Nica-Sophia-Antipolis
- [47] Lopes Filho M C and Nussenzveig Lopes H J 2002 Existence of a weak solution for the semi-geostrophic equation with integrable initial data *Proc. R. Soc. Edinb. A* **132** 329–39