

On Hamiltonian balanced dynamics and the slowest invariant manifold

Article

Published Version

Bokhove, O. and Shepherd, T. G. ORCID:
<https://orcid.org/0000-0002-6631-9968> (1996) On Hamiltonian balanced dynamics and the slowest invariant manifold. *Journal of the Atmospheric Sciences*, 53 (2). pp. 276-297. ISSN 1520-0469 doi: [https://doi.org/10.1175/1520-0469\(1996\)053<0276:OHBDAT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1996)053<0276:OHBDAT>2.0.CO;2) Available at <https://centaur.reading.ac.uk/32869/>

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

Published version at: [http://dx.doi.org/10.1175/1520-0469\(1996\)053<0276:OHBDAT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1996)053<0276:OHBDAT>2.0.CO;2)

To link to this article DOI: [http://dx.doi.org/10.1175/1520-0469\(1996\)053<0276:OHBDAT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1996)053<0276:OHBDAT>2.0.CO;2)

Publisher: American Meteorological Society

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

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

On Hamiltonian Balanced Dynamics and the Slowest Invariant Manifold

ONNO BOKHOVE AND THEODORE G. SHEPHERD

Department of Physics, University of Toronto, Toronto, Ontario, Canada

(Manuscript received 21 October 1994, in final form 30 June 1995)

ABSTRACT

The concept of a slowest invariant manifold is investigated for the five-component model of Lorenz under conservative dynamics. It is shown that Lorenz's model is a two-degree-of-freedom canonical Hamiltonian system, consisting of a nonlinear vorticity-triad oscillator coupled to a linear gravity wave oscillator, whose solutions consist of regular and chaotic orbits. When either the Rossby number or the rotational Froude number is small, there is a formal separation of timescales, and one can speak of fast and slow motion. In the same regime, the coupling is weak, and the Kolmogorov–Arnold–Moser theorem is shown to apply. The chaotic orbits are inherently unbalanced and are confined to regions sandwiched between invariant tori consisting of quasi-periodic regular orbits. The regular orbits generally contain free fast motion, but a slowest invariant manifold may be geometrically defined as the set of all slow cores of invariant tori (defined by zero fast action) that are smoothly related to such cores in the uncoupled system. This slowest invariant manifold is not global; in fact, its structure is fractal; but it is of nearly full measure in the limit of weak coupling. It is also nonlinearly stable. As the coupling increases, the slowest invariant manifold shrinks until it disappears altogether.

The results clarify previous definitions of a slowest invariant manifold and highlight the ambiguity in the definition of “slowness.” An asymptotic procedure, analogous to standard initialization techniques, is found to yield nonzero free fast motion even when the core solutions contain none. A hierarchy of Hamiltonian balanced models preserving the symmetries in the original low-order model is formulated; these models are compared with classic balanced models, asymptotically initialized solutions of the full system, and the slowest invariant manifold defined by the core solutions. The analysis suggests that for sufficiently small Rossby or rotational Froude numbers, a stable slowest invariant manifold can be defined for this system, which has zero free gravity wave activity, but it cannot be defined everywhere. The implications of the results for more complex systems are discussed.

1. Introduction

The study of balanced dynamics is a central subject in geophysical fluid dynamics. Originally, the subject arose from the practical imperative of numerical weather prediction (e.g. Daley 1991): when raw observations are used as initial data for the primitive equations of motion, unrealistically large high-frequency oscillations develop during the first hours of integration. (The observed atmospheric flow, for example as seen in barograph traces, has comparatively little high-frequency activity.) It was realized (Charney 1949) that these excessive high-frequency oscillations are associated with free gravity waves that are triggered by errors in the observations. The question then became how to eliminate or at least control such oscillations.

One approach is to use reduced models of the primitive equations, which represent the slow-time behavior of the flow on a lower-dimensional manifold in phase space and thereby filter the high-frequency oscillations:

such models are referred to as balanced models. Balanced models consist of prognostic equations for slowly evolving variables, for example potential vorticity, and diagnostic invertibility principles relating the other variables such as winds, pressure, and density to the predicted slow variables. These invertibility principles are constraints that act to filter the high-frequency oscillations, just as the constraint of incompressibility acts to filter sound waves in the incompressible form of the Euler equations. Examples of balanced models include the barotropic vorticity equation, the quasigeostrophic equations, and the balance equations of Charney (1955).

A second approach, which very quickly replaced the first, is to use the full primitive equations of motion but to “balance” or “initialize” the initial data such that excessive high-frequency oscillations do not develop.¹ The first initialization schemes were linear; these were surpassed by nonlinear normal-mode initialization

Corresponding author address: Prof. T. G. Shepherd, Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON, M5S 1A7, Canada.

¹ This really amounts to a combination of the two approaches, because the primitive equations are hydrostatically balanced and thus already filter vertically propagating sound waves.

(Baer 1977; Baer and Tribbia 1977; Machenhauer 1977), which is still in operational use. Leith (1980) demonstrated the connection between the quasigeostrophic equations and leading-order normal-mode initialization and, together with Lorenz (1980), introduced the concept of the *slow manifold*.

As defined by Leith (1980) and Lorenz (1980), the slow manifold is a reduced, nonlinear invariant manifold in the phase space of dynamical variables whereon the motion, once initialized, is devoid of gravity wave oscillations; that is, motion on the slow manifold remains slow. Balanced dynamics is then an attempt to describe motion on the slow manifold, while initialization is an attempt to project the data onto the slow manifold. The question of the existence of the slow manifold has been debated ever since its definition (Errico 1982; Warn 1983, unpublished manuscript; Warn and Menard 1986; Lorenz 1986; Vautard and Legras 1986; Lorenz and Krishnamurthy 1987; Jacobs 1991; Lorenz 1992; Boyd 1994; Camassa 1995). This debate has been conducted mainly in the context of low-order models, and we shall continue in this vein.

The central difficulty has proven to be that of defining what is "slow" and what is "fast." Most investigators have employed perturbation expansions related to the bounded derivative method (Kreiss 1979, 1980), based on the formal separation of timescales that exists when the Rossby number or the rotational Froude number is small. It was found that such approaches are generally asymptotic, and not convergent (Warn 1983, unpublished manuscript; Kopell 1985; Warn and Menard 1986; Lorenz 1986; Vautard and Legras 1986; Lorenz and Krishnamurthy 1987), and that initialization procedures based on such approaches generally yield a nonzero amount of gravity wave activity. This is not surprising, because even a long-period closed wave-number-triad interaction in the barotropic vorticity equation, whose solution can be represented in terms of elliptic functions, has high-frequency overtones that do not satisfy the bounded derivative condition (Lorenz 1986) and that can be difficult to distinguish from free gravity wave oscillations (Errico 1982). These considerations have led to the concept of a "fuzzy manifold" (Warn 1983, unpublished manuscript; Warn and Menard 1986), or a "slow quasi manifold" (McIntyre 1994), which is regarded as a stochastic layer of varying thickness in phase space, but relatively thin in the "fast" direction, wherein the gravity wave oscillations are bounded. The fuzziness of this quasi manifold is seen to be related to the lack of a precise definition of a free gravity wave or equivalently to the lack of a precise definition of slow motion.

Another argument against the existence of a slow invariant manifold involves an appeal to notions of statistical mechanics. Errico (1984) and Warn (1986) have shown that the statistical mechanical equilibrium of a primitive-equations model involves an equiparti-

tioning of energy into all degrees of freedom, both vortical modes and gravity waves. Obviously, such a tendency toward energy equipartition works against any sort of balance. Selective dissipation of gravity waves could prevent this equilibration; however, Kopell (1985) showed that an attracting invariant slow manifold (a center manifold) existed only for damping rates that were unrealistic for the atmosphere or oceans (Warn and Menard 1986; Lorenz 1986).

Lorenz (1986) introduced a five-component model, based on a truncation of the shallow-water equations. Its conservative form represents a nonlinear vorticity triad, nonlinearly coupled to a linear gravity wave. This model has since attracted considerable attention, as it represents what is arguably the simplest possible dynamical system involving nonlinear slow motion coupled to linear fast motion. Many investigators have studied the problem of balance in the strongly forced-dissipative regime; but, as noted above, this regime is not of obvious relevance to the atmosphere or oceans. It is arguable that one should first attempt to understand the nonlinear, conservative aspects of the dynamics, whose mathematical description is at least known, before trying to parameterize forcing and dissipation for a low-order system. (We would also note that the practical advances in initialization methods described earlier were all first developed by appealing to the properties of the conservative form of the relevant governing equations.) In this paper we follow Lorenz (1986) and focus attention on the conservative form of his system.

Noting the difficulties with asymptotic approaches to balance based on a formal separation of timescales, Lorenz (1986) proposed an exact approach to balance for his model, based on periodicity. He first noted that for any balanced solution, the fast variables ought to be slaved to the slow variables via a single-valued slaving relation. He then showed that any such solution (for this system) must be periodic in time and defined such solutions as constituting the *slowest invariant manifold*. (Periodic solutions obviously lie on an invariant set.) By searching numerically for periodic solutions (which, however, were not guaranteed to have single-valued slaving relations), Lorenz mapped out his slowest invariant manifold for this system.

Yet Lorenz's analysis raises a number of questions. In particular, what is the dynamical meaning of Lorenz's slowest invariant manifold? What is its structure? Why should periodicity be related to slaving? What is the connection between Lorenz's manifold and the asymptotic approaches? Is there [as Lorenz (1992) suggests] an inherent contradiction between invariance and slowness? Does the asymptotic nature of classic initialization methods merely reflect a failure of those methods, or does it reflect something more fundamental? Can one unambiguously identify free fast motion that is independent of the slow motion? If so, can one completely eliminate it?

In this paper we attempt to answer these questions. We do so by recognizing and exploiting the Hamiltonian structure of Lorenz's (1986) five-component model. It turns out that for this model the problem of balanced dynamics falls within the province of Hamiltonian perturbation theory. A consequence is that one may use the Kolmogorov–Arnold–Moser (KAM) theorem, complemented by numerical analysis, to obtain a geometric definition of a slowest invariant manifold—a subset of Lorenz's slowest invariant manifold, it turns out—on which the free gravity wave activity is unambiguously zero. This manifold is moreover nonlinearly stable: if a small amount of gravity wave activity is introduced, it will remain forever bounded. On the other hand, this manifold is not defined smoothly throughout the slow phase space but is fractal, though it is of nearly full measure as the perturbation parameter goes to zero. Numerical analysis reveals that, as the perturbation parameter increases, the manifold gets smaller and eventually disappears. All these insights, and the understanding of Lorenz's analysis, arise from the fact that the slaving manifold is closely related to the structure of rational and irrational tori in the unperturbed system.

The plan of the paper is as follows. In section 2 we introduce Lorenz's five-component model and identify the nondimensional measure of amplitude, ϵ , that will serve as the perturbation parameter. When $\epsilon \ll 1$, there is a formal separation of timescales between the slow vorticity-triad oscillator and the fast gravity wave oscillator. The model's Hamiltonian structure and canonical formulation are provided in section 3, where it is also shown that the KAM theorem applies in the limit $\epsilon \rightarrow 0$. Numerical calculations presented in section 4 confirm the existence of invariant tori in a sea of chaotic orbits; this allows a precise geometric definition of a slowest invariant manifold. The slowest invariant manifold proposed by Lorenz is analyzed as a slaving manifold in section 5 and compared with our definition. A comparison between asymptotic initialization procedures and initialization onto the slowest invariant manifold is given in section 6. The analysis confirms that our slowest invariant manifold is completely devoid of free gravity wave oscillations. In section 7 a hierarchy of Hamiltonian balanced models is derived, and the Hamiltonian and so-called classic balanced models are compared with asymptotically initialized orbits and with orbits on the slowest invariant manifold. The implications and limitations of our analysis are discussed in section 8.

2. Lorenz's five-component model

Lorenz (1986, hereafter L86) derived the following nine-component "primitive equations" model:

$$\frac{du_i}{dT} = y_i - z_i, \quad (1a)$$

$$\frac{dy_i}{dT} = ca_i^{-1}(a_k - a_j)y_jy_k - u_i, \quad (1b)$$

$$\frac{dz_i}{dT} = g_0a_iu_i, \quad (1c)$$

where $\{i, j, k\}$ are cyclic permutations of $\{1, 2, 3\}$. The dynamical variables in (1) are the cosine coefficients of one closed wavenumber triad in a Fourier expansion of the shallow-water equations. The system describes the (simplified) interaction between a vorticity triad with coefficients y_i and a gravity wave triad associated with the divergence u_i and the height z_i . Here, a_i is the square of the i th wavenumber, $c = \frac{1}{2}[(a_1a_2 + a_2a_3 + a_1a_3) - \frac{1}{4}(a_1^2 + a_2^2 + a_3^2)]^{1/2}$, $g_0 = gH/f^2L^2$, g is the gravitational acceleration, H the mean depth, L a characteristic horizontal length scale, and f the Coriolis parameter; time T has been nondimensionalized by f^{-1} . Geostrophically balanced dynamics are obtained by elimination of the time derivative of the divergence coefficient u_i in (1a) and elimination of u_i between (1b) and (1c), giving

$$(g_0a_i + 1) \frac{dy_i}{dT} = g_0c(a_k - a_j)y_jy_k. \quad (2)$$

The system (1) has two invariants: the energy

$$E = \frac{1}{2} \sum_i (g_0a_i(u_i^2 + y_i^2) + z_i^2) \equiv \sum_i E_i$$

and enstrophy

$$Z = \sum_i a_i E_i.$$

This is in contrast to the complete triad model for the shallow-water equations based on a cosine expansion only, without the further simplifications that lead to (1), which has no energy and enstrophy invariants.

The five-component model results if (1) is replaced by (2) for wavenumbers 1 and 2 but kept for wavenumber 3. Further simplifications, and restricting attention to the case $a_1 > a_2 > a_3$, then yield

$$\begin{aligned} \frac{dx_1}{dt} &= -x_2x_3 + bx_2x_5, \\ \frac{dx_2}{dt} &= x_1x_3 - bx_1x_5, \\ \frac{dx_3}{dt} &= -x_1x_2, \\ \frac{dx_4}{dt} &= -\frac{x_5}{\epsilon}, \\ \frac{dx_5}{dt} &= \frac{x_4}{\epsilon} + bx_1x_2. \end{aligned} \quad (3)$$

L86 derived (3) without introducing the parameter ϵ (his form is recovered on taking $\epsilon = 1$); that is, his

equations were nondimensionalized but not scaled. Apart from some factors involving the wavenumber parameters a_i , which we take to be of order unity, the dependent variables x_i in L86's form of (3) all have nondimensional amplitudes, that are formally of order $\epsilon = bR/\sqrt{1 + b^2}$. In (3), $b = 1/\sqrt{g_0 a_3} = f/\kappa\sqrt{gH}$ is the rotational Froude number, with κ being the dimensional wavenumber of the gravity wave component, and $R = U/fL$ is the Rossby number, with U being a characteristic velocity scale. We have explicitly accounted for the amplitude dependence of the variables by scaling the amplitudes by ϵ relative to L86. We also scale time t by ϵ^{-1} relative to L86; in terms of the original dimensional variables, this means that time is nondimensionalized by the advective timescale L/U .

The system (3) describes the interaction between a vorticity triad with associated coefficients x_1, x_2, x_3 and a gravity wave mode associated with coefficients of divergence x_4 and geostrophic imbalance x_5 . For $b = 0$ the system decouples into the nonlinear oscillation of the vorticity triad (x_1, x_2, x_3) , which can be expressed in terms of elliptic functions and has a frequency of $O(x_3) = O(1)$ [dimensionally $O(U/L)$], and the harmonic oscillation of the variables (x_4, x_5) with frequency $1/\epsilon$ [dimensionally $f\sqrt{1 + (gH\kappa^2/f^2)}$]. The ratio of the two frequencies is $O(\epsilon)$. In the small-amplitude limit $\epsilon \rightarrow 0$, we therefore have a formal separation of timescales with "fast" gravity wave oscillations and "slow" vorticity-triad oscillations. The small-amplitude limit can be approached in two obvious ways: (i) with R finite and $b \rightarrow 0$ and (ii) with b finite and $R \rightarrow 0$. Case (i) corresponds to the low Froude number scaling of Charney (1963), and case (ii) to the low Rossby number quasigeostrophic scaling of Charney (1948). Both cases allow a meaningful definition of balanced motion in the full shallow-water equations (e.g., Spall and McWilliams 1992).

3. Hamiltonian structure and canonical formulation

The Hamiltonian structure of (3) turns out to be crucial in the identification of a slowest invariant manifold, in the construction of symmetry-preserving balanced models, and in performing symplectic numerical integrations. Thus far we have been unable to find a Hamiltonian structure for (1).

The dynamical system (3) may be represented in the Hamiltonian form (Bokhove 1993)

$$\frac{dF(\mathbf{x})}{dt} = [F, H'] \tag{4}$$

for functions of state $F(\mathbf{x})$ with variables $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, Hamiltonian

$$H' = \frac{1}{2}(x_1^2 + 2x_2^2 + x_3^2 + x_4^2 + x_5^2), \tag{5}$$

and generalized Poisson bracket $[\cdot, \cdot]$ defined by

$$\begin{aligned} [F, G] = & \frac{\partial F}{\partial x_1} x_2 \left(b \frac{\partial G}{\partial x_5} - \frac{\partial G}{\partial x_3} \right) \\ & + \frac{\partial F}{\partial x_2} x_1 \left(\frac{\partial G}{\partial x_3} - b \frac{\partial G}{\partial x_5} \right) \\ & + \frac{\partial F}{\partial x_3} \left(x_2 \frac{\partial G}{\partial x_1} - x_1 \frac{\partial G}{\partial x_2} \right) - \frac{1}{\epsilon} \frac{\partial F}{\partial x_4} \frac{\partial G}{\partial x_5} \\ & + \frac{\partial F}{\partial x_5} \left(-bx_2 \frac{\partial G}{\partial x_1} + bx_1 \frac{\partial G}{\partial x_2} + \frac{1}{\epsilon} \frac{\partial G}{\partial x_4} \right) \end{aligned} \tag{6}$$

for functions $F(\mathbf{x})$ and $G(\mathbf{x})$. The bracket (6) is skew symmetric and obeys Leibniz's rule and Jacobi's identity $[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0$ [for the mathematical background see Littlejohn (1982)]. The Casimir function $C = \frac{1}{2}(x_1^2 + x_2^2)$ arises immediately from (6) as the solution of $[C, G] = 0$ for all functions G . The invariance of C in time is guaranteed by construction and may be verified directly from (3). The vorticity-triad dynamics for $b = 0$ has two independent Hamiltonian structures, in the sense that the Casimir for each bracket is different, but in proving Jacobi's identity with $b \neq 0$ only one form of the bracket survives. In appendix A we explain the connection between the uncoupled vorticity-triad dynamics of (3) and rigid-body rotation. The Hamiltonian H' and Casimir invariant C are linearly related to analogs of the energy E and enstrophy Z of the simplified and truncated shallow-water system (1).

The existence of a Casimir invariant suggests a reduction to four variables with the constant Casimir parameter C , since $dC/dt = 0$. Making the transformation

$$x_1 = \sqrt{2C} \cos \phi, \quad x_2 = \sqrt{2C} \sin \phi \text{ yields the system}$$

$$\frac{d\phi}{dt} = x_3 - bx_5, \tag{7a}$$

$$\frac{dx_3}{dt} = -C \sin 2\phi, \tag{7b}$$

$$\frac{dx_4}{dt} = -\frac{x_5}{\epsilon}, \tag{7c}$$

$$\frac{dx_5}{dt} = \frac{x_4}{\epsilon} + bC \sin 2\phi. \tag{7d}$$

The Hamiltonian may be rewritten as

$$H' - \frac{3}{2} C \equiv H = -\frac{1}{2} C \cos 2\phi + \frac{1}{2} (x_3^2 + x_4^2 + x_5^2). \tag{8}$$

The system (7) then has the Hamiltonian H and the reduced bracket

$$[F, G] = \left(\frac{\partial F}{\partial \phi} \frac{\partial G}{\partial x_3} - \frac{\partial F}{\partial x_3} \frac{\partial G}{\partial \phi} \right) + b \left(\frac{\partial F}{\partial x_5} \frac{\partial G}{\partial \phi} - \frac{\partial F}{\partial \phi} \frac{\partial G}{\partial x_5} \right) + \frac{1}{\epsilon} \left(\frac{\partial F}{\partial x_5} \frac{\partial G}{\partial x_4} - \frac{\partial F}{\partial x_4} \frac{\partial G}{\partial x_5} \right). \quad (9)$$

A variational formulation of the dynamics (7) is provided by Hamilton's principle

$$\delta \int \left\{ -(\phi - \epsilon b x_4) \frac{dx_3}{d\tau} + \epsilon x_4 \frac{dx_5}{d\tau} - \epsilon H \right\} d\tau = 0, \quad (10)$$

subject to the endpoint conditions $\delta x_3 = 0 = \delta x_5$, where the dynamics has been expressed on the fast timescale $\tau = t/\epsilon$.

The following canonical formulation arises from Hamilton's principle (10): $q_1 = \phi - \epsilon b x_4$, $p_1 = x_3$, $q_2 = \sqrt{\epsilon} x_4$, $p_2 = -\sqrt{\epsilon} x_5$, with the Hamiltonian

$$\begin{aligned} \tilde{H} &= \epsilon H = \epsilon \frac{1}{2} p_1^2 \\ &- \frac{1}{2} \epsilon C \cos 2(q_1 + b\sqrt{\epsilon} q_2) + \frac{1}{2} (p_2^2 + q_2^2). \end{aligned} \quad (11)$$

The corresponding canonical equations are

$$\begin{aligned} \frac{dq_1}{d\tau} &= \frac{\partial \tilde{H}}{\partial p_1} = \epsilon p_1, \\ \frac{dp_1}{d\tau} &= -\frac{\partial \tilde{H}}{\partial q_1} = -\epsilon C \sin 2(q_1 + b\sqrt{\epsilon} q_2), \\ \frac{dq_2}{d\tau} &= \frac{\partial \tilde{H}}{\partial p_2} = p_2, \\ \frac{dp_2}{d\tau} &= -\frac{\partial \tilde{H}}{\partial q_2} = -q_2 - \epsilon^{3/2} b C \sin 2(q_1 + b\sqrt{\epsilon} q_2), \end{aligned} \quad (12)$$

with q_i the generalized coordinates and p_i the generalized momenta. This is a two-degree-of-freedom canonical Hamiltonian system. When $b = 0$, the two conjugate pairs decouple, and the system is integrable: (q_1, p_1) corresponds to the vorticity triad and is (for given C) equivalent to the nonlinear pendulum, while (q_2, p_2) corresponds to the gravity wave and is equivalent to the harmonic oscillator. Thus, L86's primitive-equations atmosphere corresponds to a pair of coupled oscillators: one nonlinear and the other linear. In the small amplitude limit $\epsilon \rightarrow 0$, the frequency of the nonlinear oscillator goes to zero (on the fast timescale), while the frequency of the linear oscillator remains finite. This is probably the simplest possible dynamical system for examining the interaction of fast and slow modes, which may help to explain its lasting interest.

Since balanced dynamics concerns the behavior of the system (12) when $\epsilon \ll 1$ (i.e., when there is a separation of timescales, so that the concepts of "fast" and "slow" are meaningful), and since the system (12) is integrable when $\epsilon = 0$, it is evident that for this system the problem of balanced dynamics falls within the province of Hamiltonian perturbation theory (e.g., Arnold and Avez 1968; Arnold 1978; Arnold et al. 1988). Since the two uncoupled oscillators are integrable, each may be written in action-angle variables (I_i, θ_i) [$i = 1, 2$] (e.g., Arnold 1978), where the actions I_i are constants of the motion and where the angles θ_i evolve in time according to $\theta_i(t) = \omega_i t + \beta_i$ with β_i being constants of integration and $\omega_i(I_i)$ the frequencies. Note that action-angle variables are not defined at a separatrix and that different action-angle representations are required on either side of a separatrix.

The total (i.e., perturbed) Hamiltonian (11) may be rewritten (except at the separatrix $H_1 = \frac{1}{2}C$) in terms of these action-angle variables as

$$\tilde{H} = H_0(I_2) + \epsilon H_1(I_1) + \epsilon^{3/2} b H_2(I_i, \theta_i; b, \epsilon), \quad (13)$$

where $H_0 = \frac{1}{2}(p_2^2 + q_2^2)$, $H_1 = \frac{1}{2}p_1^2 - \frac{1}{2}C \cos 2q_1$, and H_2 is periodic in the angles θ_i . The unperturbed Hamiltonian H_0 is degenerate: it depends only on I_2 . The perturbation at $O(\epsilon)$ removes the degeneracy but does not destroy integrability: the action-angle structure now involves both I_1 and I_2 . Moreover, the "intermediate" Hamiltonian $H_{\text{int}} = H_0(I_2) + \epsilon H_1(I_1)$ is isoenergetically nondegenerate (for $H_1 \neq \frac{1}{2}C$), in the sense that the ratio of the frequencies ω_1/ω_2 varies as either one of the actions is varied on a constant energy surface; this condition may be written as

$$\det \begin{vmatrix} \frac{\partial^2 H_{\text{int}}}{\partial I_1^2} & \frac{\partial^2 H_{\text{int}}}{\partial I_1 \partial I_2} & \frac{\partial H_{\text{int}}}{\partial I_1} \\ \frac{\partial^2 H_{\text{int}}}{\partial I_2 \partial I_1} & \frac{\partial^2 H_{\text{int}}}{\partial I_2^2} & \frac{\partial H_{\text{int}}}{\partial I_2} \\ \frac{\partial H_{\text{int}}}{\partial I_1} & \frac{\partial H_{\text{int}}}{\partial I_2} & 0 \end{vmatrix} \neq 0 \quad (14)$$

(e.g., Arnold et al. 1988, chap. 5). The consequence of isoenergetic nondegeneracy is that the KAM theorem applies (op. cit., chap. 5.3, theorem 14): for small enough ϵ , most of the phase space is filled by invariant tori that are close to the invariant tori (I_1, I_2) of the intermediate system. Hence, we expect the generic picture of a perturbed, nearly integrable Hamiltonian system: most of the original tori survive for a sufficiently small perturbation and on these tori the solutions remain regular, but there are regions of chaos around resonant tori, whose measure goes to zero (in this case exponentially fast, like $\exp[-c/\epsilon]$) as $\epsilon \rightarrow 0$. Indeed, Camassa (1995) has proven the existence of chaos in this system by using a Melnikov analysis in the vicinity of the separatrix.

Hereafter, we will investigate the four-component model (7) rather than the five-component model (3). For given values of the energy H and the Casimir parameter C , the accessible phase space is three-dimensional. The topology of regular orbits corresponds to a two-torus (Fig. 1a) around which the orbits wind themselves (Fig. 1b). The winding number around the two-torus can be either rational or irrational. The two-dimensional invariant tori partition the three-dimensional accessible phase space, so that the phase space on one side of the invariant surface is not accessible from the other side of the invariant surface. The invariant tori thus restrict the chaotic orbits to the phase space sandwiched between them. Energy-conserving balanced motion has only one degree of freedom, and balanced orbits are therefore regular. It follows that chaotic orbits necessarily involve both low- and high-frequency degrees of freedom and are therefore unbalanced.

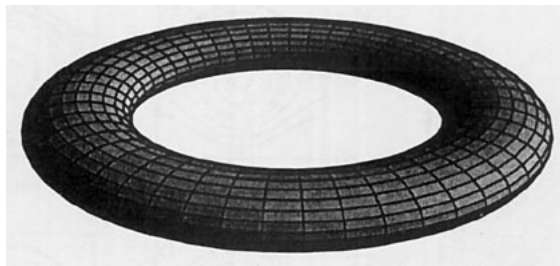
4. Regular and chaotic orbits

In this section we explore the phase space structure of regular and chaotic orbits by numerical means. This leads to a geometric characterization of a slowest invariant manifold.

Numerical integrations of (7) are performed with a fourth-order explicit symplectic integrator (Yoshida 1990) applied to (12), which preserves the phase space volume $dq_i dp_i$; control runs have been made with an implicit fourth-order Gear's algorithm (Gear 1971) and by using different time steps. Symplectic integration preserves the symplectic structure of the dynamics by defining each time step as a canonical transformation. The implicit nature of Gear's algorithm guarantees that the energy does not increase monotonically, as it would under integration by a fourth-order Runge-Kutta scheme, but oscillates around its constant analytical value. The symplectic integrator is about 10 times faster than Gear's algorithm, and the variations in the energy are smaller.

Figure 2 shows pairs of Poincaré sections for increasing values of the rotational Froude number b , for fixed $R = 1$ and $C = 1$. The upper member of each vertical pair is the section in the "slow" variables $\phi - x_3$, and the lower member the section in the "fast" variables $x_4 - x_5$. The two sections contain the same sets of orbits; since a Poincaré section is taken at a fixed energy, this means that lower-energy curves in the $\phi - x_3$ plane correspond to higher-energy curves in the $x_4 - x_5$ plane, and vice versa. The leftmost pairs correspond to the unperturbed case $b = 0$: the phase portraits of the nonlinear pendulum and the harmonic oscillator emerge. Three two by four sets of panels are shown, corresponding to different energies: (i) $H = 1$, (ii) $H = 0.5$, and (iii) $H = 0$. These values of H are chosen so that when the fast motion is of small amplitude the slow motion will be, respectively, outside, around, or within the separatrix of the nonlinear pendulum.

a



b

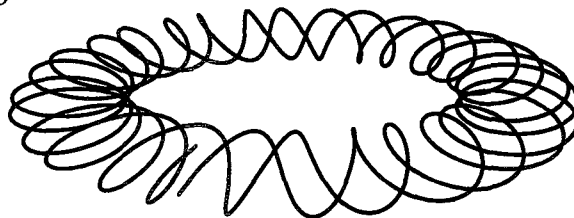


FIG. 1. a) The two-torus. b) The individual regular orbits or trajectories wind their way around the two-torus.

Moving to the right in each set of panels, the perturbation b , and therefore ϵ , increases. For $b = 0.1$, the structure of the irrational invariant tori survives in accordance with the KAM theorem. But as b increases further, the tori begin to break up over significant parts of the phase space, being replaced by chaotic regions (the sea of dots in each section). Clear areas are presumably filled with tori. For lower energies the breakup of the tori as b increases is less severe. Secondary tori (island structures) emerge in some of the sections; they arise from the breakup of rational tori. All this behavior is, of course, quite generic for perturbed Hamiltonian dynamical systems.

Figure 3 presents the same kind of picture, but for increasing values of the Rossby number R , for fixed $b = 0.5$ and $C = 1$. Again the three two by four sets of panels correspond to $H = 1$, $H = 0.5$, and $H = 0$. The same complex structure of regular and chaotic orbits is seen.

The finite-time Liapunov exponents of two orbits, one chaotic and one regular, among those of Figs. 3(i)c,d, are given in Fig. 4a. The persistent phase space stretching around the chaotic orbit is implied. In Figs. 4b,c, power spectra of the departure from the mean of x_3 are shown for the same two orbits. In the chaotic case (Fig. 4b) one sees a broadband spectrum; in the regular case (Fig. 4c) one sees a finite set of peaks corresponding to two fundamentals and their overtones.

A consideration of the phase space geometry for this system offers a compelling definition of a slowest in-

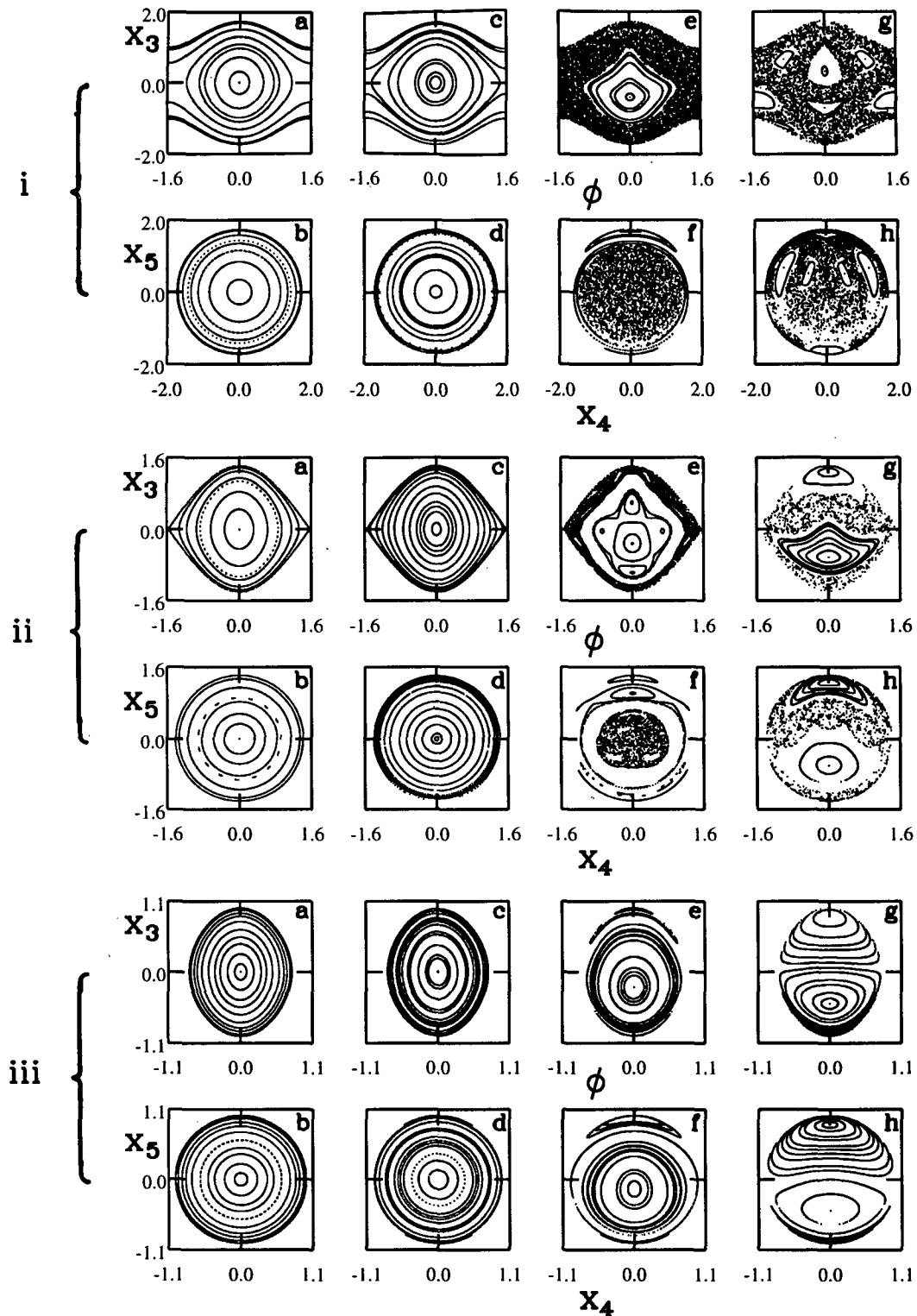


FIG. 2. Pairs of Poincaré sections; the upper panel of each pair corresponds to the section in the "slow" $\phi-x_3$ plane when $x_4 = 0$ and $\dot{x}_4 < 0$, and the lower panel to the section in the "fast" x_4-x_5 plane when $\phi = 0$ and $\dot{\phi} < 0$. Three two by four sets of panels are shown, corresponding to three decreasing energies as one moves down: (i) $H = 1$, (ii) $H = 0.5$, and (iii) $H = 0$. In each set the rotational Froude number b increases as one moves to the right: (a) and (b) $b = 0$, (c) and (d) $b = 0.1$, (e) and (f) $b = 0.5$, and (g) and (h) $b = 1$. In all cases $C = 1$ and $R = 1$. The solutions are for 3750 time units with a time step of $\Delta t = 0.0025$ (i.e., 1.5×10^6 time steps).

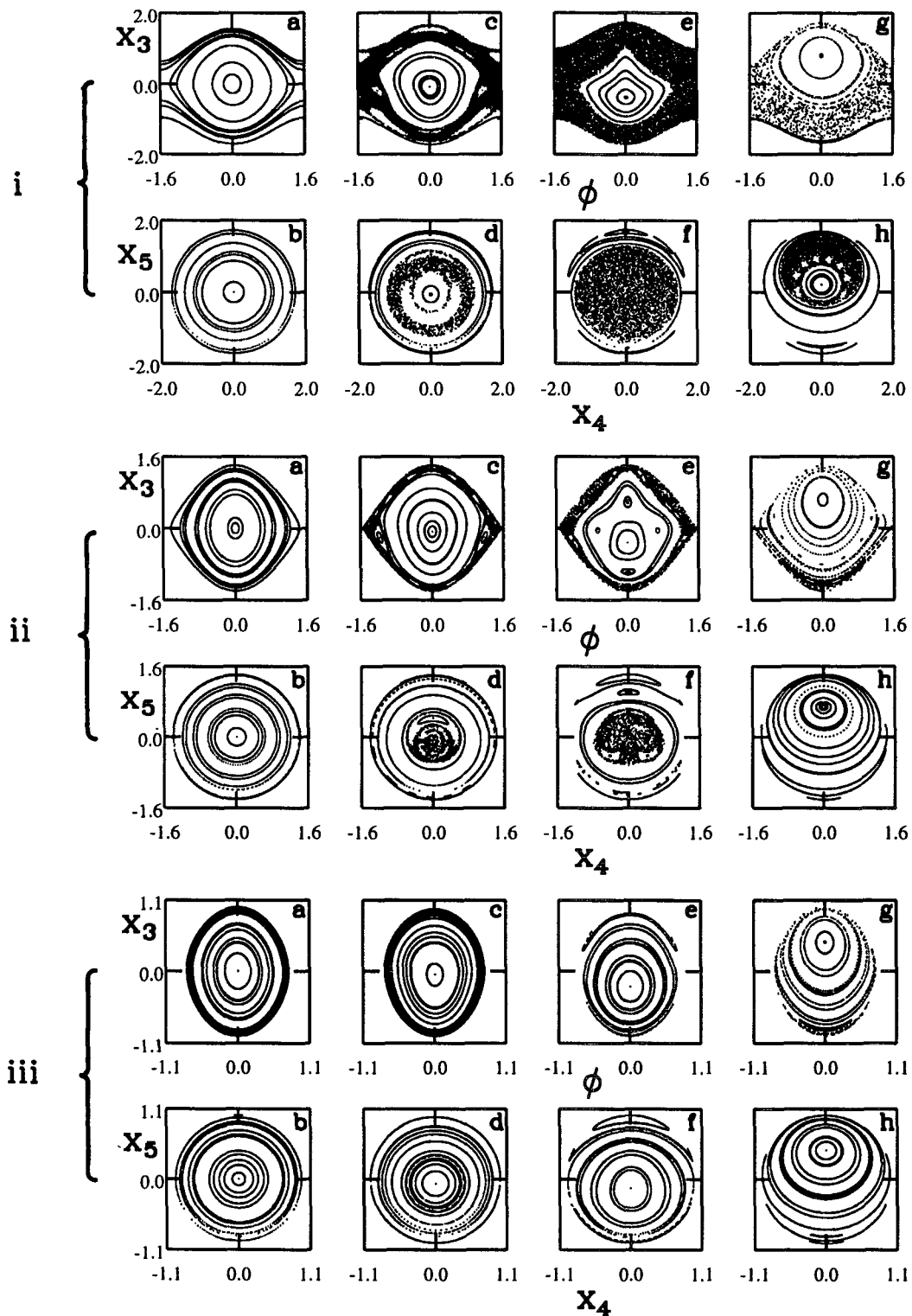


FIG. 3. As in Fig. 2 but for increasing Rossby number R : (a) and (b) $R = 0.1$, (c) and (d) $R = 0.5$, (e) and (f) $R = 1$, and (g) and (h) $R = 10$. In all cases $b = 0.5$.

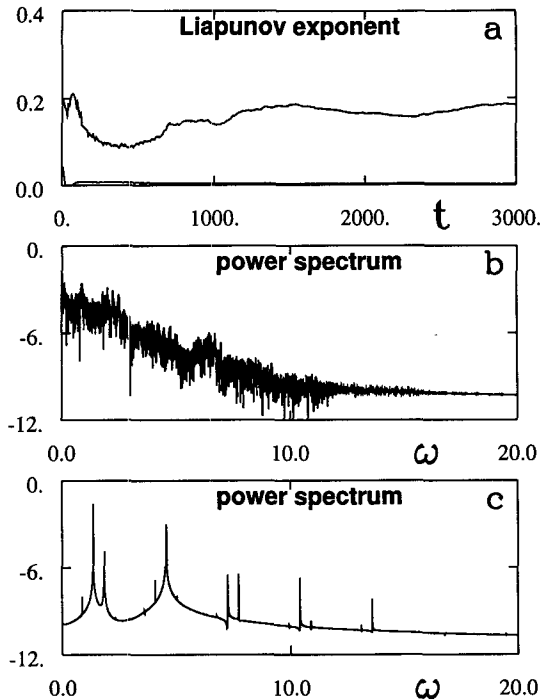


FIG. 4. (a) Finite-time Liapunov exponents, as functions of time, of a regular and chaotic orbit taken from the solutions in Fig. 3(i)c,d. (b) and (c) Power spectrum of the departure from the mean of x_3 over 3000 time units for the (b) chaotic and (c) regular orbit; the ordinates of (b) and (c) are scaled logarithmically.

variant manifold. In the uncoupled case $b = 0$, it is clear that the slowest invariant manifold is defined by setting the fast action $I_2 = 0$ for any choice of the slow variables; this corresponds to the center, or core, of the tori shown in the x_4-x_5 cross section [e.g., Fig. 2(i)b], which has zero free gravity wave activity. The question then arises: Is it possible to define corresponding core solutions $J_2 = 0$ in the *new* (i.e., perturbed) action-angle variables (J_i, φ_i) that are smoothly related to the original core solutions $I_2 = 0$ as $\epsilon \rightarrow 0$? If this could be done for any choice of the slow variables, then we would have a well-defined invariant manifold that would contain no free fast motion and would arguably be the slowest invariant manifold. Moreover, it would be nonlinearly stable: small disturbances from the core would stay small, trapped by an invariant torus of small amplitude in the J_2 direction—the amplitude of the free fast motion would be bounded. The KAM theorem guarantees that, for sufficiently small perturbations, one *can* in fact construct new action-angle variables over a significant part of the phase space, which are smooth deformations of the original variables; in other words, most invariant tori survive and are merely smoothly deformed. The preservation of the topology of the unperturbed dynamics, for most initial conditions, means that it is indeed possible to define core solutions $J_2 = 0$, for sufficiently small ϵ , that are smoothly related to the

original core solutions $I_2 = 0$.² The surviving invariant tori that can be used to define the core solutions are defined only for irrational frequency ratios [the rational tori are resonant and break up according to the Poincaré–Birkhoff fixed-point theorem (Arnold and Avez 1968)], which means that the slowest invariant manifold $J_2 = 0$ is not continuous in the slow actions. Rather, it is the complement of a Cantor-like set and has nearly full measure. (The degenerate isoenergetic version of the KAM theorem, which applies here, says that the fraction of phase space where the tori are not defined is exponentially small in the perturbation parameter.) For practical purposes then, $J_2 = 0$ effectively defines an invariant manifold wherever the tori structure remains intact.

We can now reexamine Figs. 2 and 3 to see whether these core structures can be identified. Such core structures would be seen as single dots in the center of the primary tori in the x_4-x_5 plane. In Fig. 2 the cores are clearly evident in each case for $b = 0.1$ but for larger b they are clearly identifiable only for the case $H = 0$. Of course, what appear to be dots in the figures are actually tori of exceedingly small radius in the x_4-x_5 plane, the size of which seems to be limited only by numerical precision (about 10^{-12}). Since the oscillations in the fast plane are close to being linear (i.e., with frequency independent of amplitude to leading order in ϵ), the resonance condition will be essentially independent of the amplitude of the fast motion. This means that we can identify the dots with true cores $J_2 = 0$. As the perturbation increases and the chaotic regions become more extensive, it becomes increasingly unlikely that a core can be identified for a given choice of the slow variables. In Fig. 3 the same conclusions generally apply, although chaos appears to set in more drastically with R than with b . This is not surprising in light of (13): the perturbation is $O(b^{3/2})$ relative to H_{int} in the small- b limit but only $O(R^{1/2})$ in the small- R limit.

Taken together, the KAM theorem and the numerical analysis demonstrate that for small but finite b or R one may define a slowest invariant manifold, which will cover most, but not all, of the slow phase space. It is not a simply connected manifold, because it is not continuous in the slow actions, but it is of nearly full mea-

² Since the transformation between the original variables and the action-angle variables in the unperturbed problem has a singularity at $I_2 = 0$, we leave open the question of whether the mathematical existence of such core solutions in the perturbed problem is guaranteed by the KAM theorem. What is guaranteed is the preservation of a topological structure that allows one to define the limit $J_2 \rightarrow 0$: the smoothness of the transformation to new action-angle variables guarantees that the nesting order of different tori cannot change. Since for practical purposes one is only interested in whether J_2 can be made arbitrarily small, to machine accuracy for example, we use $J_2 = 0$ throughout the paper as a shorthand for the limit $J_2 \rightarrow 0$, with the above caveat in mind.

sure. We would not wish to call it a ‘‘slow manifold,’’ for two reasons. First, even the uncoupled slow motion has high-frequency overtones. Second, it is possible that the transformation to new action-angle variables, although smooth, will introduce some fast motion in the original variables. This possibility will be explored in section 6. The set of core solutions $J_2 = 0$ is, however, devoid of free gravity waves and is arguably the slowest invariant manifold.

It should be noted that there are secondary core structures that arise in the perturbed system; a good example is provided by Fig. 2(i)h. These are associated with islands that arise from the breakup of rational tori. Such secondary core structures are described by action-angle variables that are not smoothly related to the original action-angle variables and thus would be difficult to reach by perturbative methods; moreover, they involve a nonzero winding in the x_4 - x_5 plane and thus may be regarded as containing a nonzero amount of free fast motion.

5. The slaving manifold

In this section we connect our geometric definition of a slowest invariant manifold—namely, the cores of the primary invariant tori with $J_2 = 0$, which are smoothly related to the original cores $I_2 = 0$ as $\epsilon \rightarrow 0$ —with that proposed by L86. It turns out that L86’s results can be readily understood in terms of the Hamiltonian structure of the dynamics.

We may anticipate the existence of an invariant manifold by introducing a slaving principle (Kreiss 1979, 1980; van Kampen 1985; Warn et al. 1995), which entails that the fast variables x_4, x_5 are assumed to be (possibly multivalued) functions of the slow variables ϕ and x_3 ; that is,

$$x_4 = U_4(\phi, x_3; C), \quad x_5 = U_5(\phi, x_3; C). \quad (15)$$

Conservation of energy then ensures that the slaved motion is integrable by quadratures. In principle, we would then obtain an invariant slaving manifold in phase space. There is, however, no guarantee that it would be devoid of gravity waves.

In section 6 we solve the slaving relations approximately by expanding them in a power series in ϵ ; this corresponds to well-known initialization techniques in numerical weather prediction (Lorenz 1980, L86). In the remainder of this section we look for exact slaving solutions. That such solutions exist is clear from the presence of regular orbits in the numerical solutions (section 4). Eliminating ϕ by conservation of energy, the slaving relations can be written

$$x_4 = U_4(x_3; C, H), \quad x_5 = U_5(x_3; C, H). \quad (16)$$

Substitution of (16) into (7) and further elimination of ϕ gives the so-called (Lorenz 1980) ‘‘superbalance equations’’:

$$-\frac{1}{\epsilon} U_5 = \frac{dU_4}{dt} = \mp \frac{\partial U_4}{\partial x_3} \times [C^2 - (x_3^2 + U_4^2 + U_5^2 - 2H)^2]^{1/2},$$

$$\frac{1}{\epsilon} U_4 = \frac{dU_5}{dt} \mp b[C^2 - (x_3^2 + U_4^2 + U_5^2 - 2H)^2]^{1/2}$$

$$= \mp \left(\frac{\partial U_5}{\partial x_3} + b \right) [C^2 - (x_3^2 + U_4^2 + U_5^2 - 2H)^2]^{1/2}. \quad (17)$$

Combining these two equations and making a change of variables to $U_4 = \sqrt{2U} \cos \theta$, $U_5 = \sqrt{2U} \sin \theta$, $x = x_3$, yields after some algebra

$$\frac{d\theta}{dx} = -\frac{1}{2U} \left[2Ub^2 - \left(\frac{dU}{dx} \right)^2 \right]^{1/2} \mp \frac{1}{\epsilon} \left[\frac{1}{C^2 - (x^2 + 2U - 2H)^2} \right]^{1/2}, \quad (18a)$$

$$\frac{d^2U}{dx^2} - b^2 - \frac{1}{\epsilon} \times \left[\frac{2Ub^2 - (dU/dx)^2}{C^2 - (x^2 + 2U - 2H)^2} \right]^{1/2} = 0. \quad (18b)$$

Equations (18a,b) have two solutions for $b = 0$: (i) $U = 0$ and (ii) $U = E_2$, with θ given by

$$\theta = \frac{2}{\epsilon(2E_1 + C)^{1/2}} \left[F\left(\frac{1}{2} \cos^{-1}\left(\frac{x_3^2 - 2E_1}{C}\right); m\right) - K(m) \right] + \frac{\pi}{2}, \quad (19)$$

where $F(\cdot; \cdot)$ is the elliptic integral of the first kind, $K(\cdot)$ is the complete elliptic integral of the first kind, $m = 2C/(C + 2E_1)$, and E_1 and E_2 are the energies in the uncoupled nonlinear and linear oscillators, respectively. The $\pi/2$ and $K(m)$ terms in (19) arise from imposing the boundary condition $x_4 = 0$ when $\phi = 0$, as will be justified shortly. These solutions for $b = 0$ merely reflect the action-angle structure of the uncoupled, integrable system; obviously, there can be no real dynamical slaving when $b = 0$. The case $U = 0$ corresponds to $I_2 = 0$, that is, zero gravity wave amplitude. The case $U \neq 0$ corresponds to expressing θ_2 in terms of θ_1 (and I_1) through the elimination of time.

It is clear that only slaving solutions with $I_2 = 0$ can reasonably be regarded as slow solutions. The other slaving solutions contain free fast motion, which is completely independent of the slow motion. When the ratio of the slow to the fast periods is an integer, then the fast solution is a harmonic of the slow solution and satisfies a single-valued slaving relation. When this ratio is a noninteger rational number, then the slaving relation is multivalued but with a finite number of

branches. Both of these cases correspond to rational tori. But the most likely scenario is that the ratio is irrational, in which case the slaving relation has a dense infinity of branches: the orbits densely fill a two torus in the four-dimensional phase space, and for a given value of the slow variables there is a continuous infinity of fast variables, which may be associated with it. In this case the concept of slaving would appear to lose any significance.

L86 argued that one should restrict attention to single-valued slaving relations. For $b = 0$ these include all tori (rational or irrational) with $I_2 = 0$ and all rational tori with $I_2 \neq 0$ having integer resonances $\omega_2/\omega_1 \in \mathbb{Z}$. He then proposed such solutions as the *definition* of the slowest invariant manifold in the case $b \neq 0$. By an appeal to the discrete symmetries of (7), L86 showed that all single-valued slaving solutions must be periodic in time. He then solved for periodic orbits by an iteration procedure. [Note, however, that although periodic orbits must be slaved in the sense of (15) they need not correspond to single valued slaving relations.] We have implemented L86's algorithm: Figs. 5a,b show x_5 as a function of ϵ when $\phi = x_4 = 0$ for fixed values of x_3 and with $C = 1$. The dashed lines correspond to $b = 0$, and the continuous lines or dots to $b = 0.5$. In Fig. 5a we have taken $x_3 = 1/\sqrt{0.405}$; the associated orbits lie well outside the separatrix of the nonlinear pendulum [cf. Fig. 2(i)a]. This figure corresponds to L86's Figs. 4 and 5. In Fig. 5b we have taken $x_3 = \sqrt{2.04}$; the orbits now lie close to the separatrix. For $b = 0$, the horizontal dashed line $x_5 = 0$ shows the slow slaving manifold $I_2 = 0$ of the uncoupled system, while the vertical dashed lines are the rational tori with $I_2 \neq 0$ for which the frequency ratio ω_2/ω_1 is an integer. For $b = 0.5$, the lines composed of dots show the complex character of the slaving manifold. The multi-valued nature of x_5 is especially evident in Fig. 5b.

The interpretation of Fig. 5 becomes clear by appealing to Hamiltonian perturbation theory. Since the dashed vertical lines correspond to rational tori, they break up under the perturbation. On the other hand, in the resulting band of chaos one can still find isolated stable and unstable periodic solutions. This follows from the Poincaré-Birkhoff fixed-point theorem (e.g., Arnold and Avez 1968). However, these new periodic solutions will have a high-frequency time dependence, which is associated with the original resonance. The dashed horizontal lines correspond to the core solutions $I_2 = 0$ of the uncoupled system, including the cores of both rational and irrational tori. Under the perturbation, most of the irrational cores will survive and moreover will be close to their original location. Since these solutions are just those described at the end of the previous section, we can see that our slowest invariant manifold is a subset of the slowest invariant manifold of L86. It is clear from Fig. 5 that for small ϵ the horizontal dashed line corresponding to $I_2 = 0$ deforms to

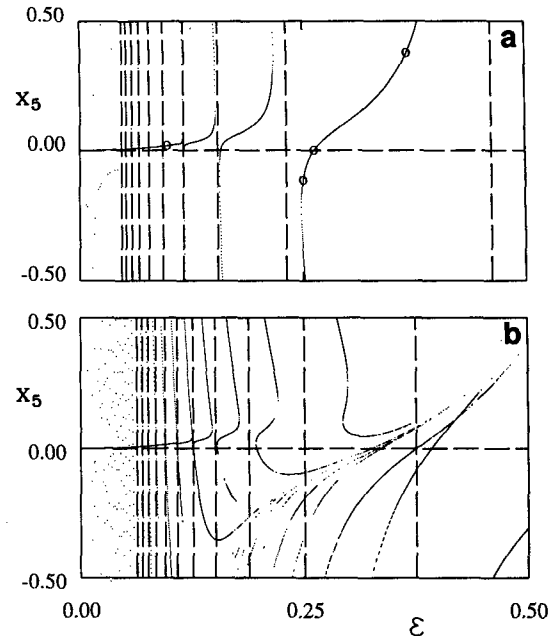


FIG. 5. The slowest invariant manifold of periodic orbits found using L86's algorithm, showing x_5 vs ϵ when $\phi = x_4 = 0$ for fixed values of x_3 and with $C = 1$ and either $b = 0$ (the dashed lines) or $b = 0.5$ (the continuous lines or dots): (a) $x_3 = 1/\sqrt{0.405}$; (b) $x_3 = \sqrt{2.04}$. Four points are marked by circles in (a) for later reference. Note that the set of dashed lines represents all integer fractions of the fundamental and is therefore incomplete; to the left of the leftmost one there is an infinity of lines, accumulating at $\epsilon = 0$.

a quasi-horizontal apparently continuous curve, which is tangent to the horizontal dashed line in the limit $\epsilon \rightarrow 0$. However, we know that the quasi-horizontal curve cannot really be continuous, because it must be punctuated by resonances—and not only the pure integer resonances corresponding to the vertical dashed lines but all resonances. In other words, the slowest invariant manifold represented by the quasi-horizontal curve cannot be defined for a dense set of ϵ 's corresponding to all the rational numbers. Of course, the rational numbers have zero measure relative to the irrational ones. As one approaches $\epsilon = 0$, the density of the resonances becomes ever greater; but their order also increases (i.e., ω_2/ω_1 gets larger), which means that more of the surrounding invariant tori survive the perturbation. And in the limit $\epsilon \rightarrow 0$, we know from the KAM theorem that the fraction of phase space that is filled with non-resonant tori approaches unity exponentially fast in ϵ . Thus, for practical purposes the slowest invariant manifold is well defined for sufficiently small (but nonzero) ϵ . As ϵ increases, however, the resonances become less frequent but much stronger; this is why the curves start to diverge in Fig. 5a. The calculation shown in Fig. 5b is deliberately performed close to the separatrix, where we expect the resonances to appear for much smaller ϵ , as is indeed observed.

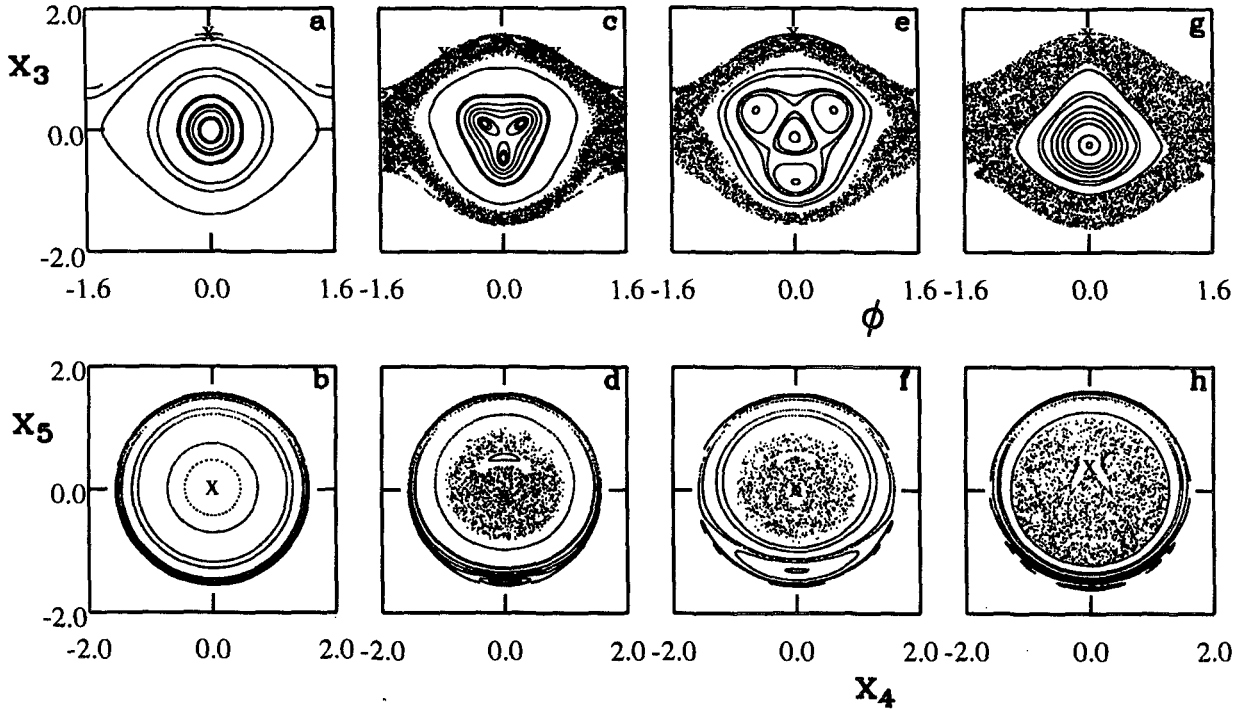


FIG. 6. Pairs of Poincaré sections, defined as in Figs. 2 and 3, of the phase space surrounding the four orbits denoted by circles in Fig. 5a. The pair (a) and (b) corresponds to the small- ϵ orbit, and the sequence of pairs (c) and (d), (e) and (f), and (g) and (h) corresponds to increasing values of x_5 along the larger- ϵ branch in Fig. 5a. In the sections these four orbits are denoted by crosses.

It is interesting to see exactly where, in the Poincaré sections, L86's periodic solutions reside. In Fig. 6 we plot Poincaré sections that include the four points marked by circles in Fig. 5a; those orbits are marked by crosses in Fig. 6. The orbit corresponding to the point with small ϵ evidently lies in the core $J_2 = 0$ of the primary tori and is therefore part of the slowest invariant manifold. The orbit corresponding to the lower point of the larger- ϵ branch in Fig. 5a is seen to lie within a sea of chaos, while the two upper points lie at the cores of visible island structures. Whether these cores are smooth deformations of the original core $I_2 = 0$ is not at all obvious. The point lying in the chaotic sea turns out to be numerically stable; in other cases, however, we found that L86's algorithm converged (for a convergence threshold of 10^{-12}) to solutions that appeared to be unstable in the time integrations.

6. Slaving and initialization

After substitution of the slaving ansatz $x_4 = U_4(\phi, x_3)$ and $x_5 = U_5(\phi, x_3)$ introduced in section 5 into (7), the superbalance equations result:

$$-\frac{1}{\epsilon} U_5 = \frac{dU_4}{dt} = \frac{\partial U_4}{\partial \phi} (x_3 - bU_5) - \frac{\partial U_4}{\partial x_3} C \sin 2\phi,$$

$$\frac{1}{\epsilon} U_4 + bC \sin 2\phi = \frac{dU_5}{dt} = \frac{\partial U_5}{\partial \phi} (x_3 - bU_5) - \frac{\partial U_5}{\partial x_3} C \sin 2\phi. \quad (20)$$

In section 5 we tried to find exact solutions of (20) by eliminating ϕ using conservation of energy. Here, we pursue a different strategy, obtaining approximate and smooth solutions of (20) by expanding the slaving relations in a series in ϵ as $U_4 = U_4^{(0)} + \epsilon U_4^{(1)} + \dots$ and $U_5 = U_5^{(0)} + \epsilon U_5^{(1)} + \dots$. To leading order, we find $U_4 = 0 = U_5$, and at the next order $U_4 = -\epsilon bC \sin 2\phi$ and $U_5 = 0$. (In appendix B we give the solutions up to the fourth order in ϵ .) Higher-order solutions are found by expressing the nonzero terms $U_4^{(2k+1)}$ and $U_5^{(2k+2)}$ as sums of products $A_{k,i_1,i_2}^{(4)} C^{i_1+i_2} \sin^{i_1}(2\phi) \cos^{i_2}(2\phi) x_3^{i_3}$ and $A_{k,i_1,i_2}^{(5)} C^{i_1+i_2} \sin^{i_1}(2\phi) \cos^{i_2}(2\phi) x_3^{i_3}$, respectively; $A^{(4)}$ and $A^{(5)}$ are coefficients depending on b ; i_1, i_2, i_3 are integers; and i_3 is given in terms of (k, i_1, i_2) . We have iteratively constructed higher-order slaving relations by calculating all unknown coefficients $A^{(4)}$ and $A^{(5)}$ in this expansion to 40th order and could proceed beyond this if desired. As a check, the algorithm reproduced the column labeled BT (Baer-Tribbia initialization) in Table 3 of L86.

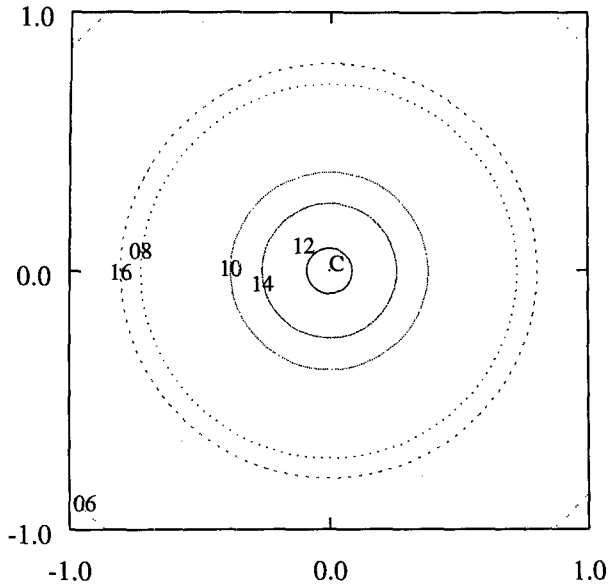


FIG. 7. Part of the composite Poincaré sections in the x_4 - x_5 plane when $\phi = 0$ and $\dot{\phi} < 0$ for asymptotic initializations of varying order in ϵ ; the numbers from 0 to 16 denote the order of the asymptotic initialization. The point marked C denotes the L86 periodic orbit initialization. The abscissa is $x_4 \times 10^{-4}$; the ordinate is $(x_5 - 0.01260708) \times 10^4$. The initial values of ϕ and x_3 are the same in all cases. The parameter values are $R = 0.2$, $b = 0.5$, and $C = 1$, with $\Delta t = 0.0025$ over ~ 1250 time units.

“Classic” balanced models follow upon substituting such approximations to $x_4 = U_4(\phi, x_3)$ and $x_5 = U_5(\phi, x_3)$ into (7a, b); namely,

$$\frac{d\phi}{dt} = x_3 - bU_5(\phi, x_3), \quad (21a)$$

$$\frac{dx_3}{dt} = -C \sin 2\phi, \quad (21b)$$

followed by a truncation to the required order in ϵ . Generalizations of this systematic procedure to construct balanced models for partial differential equations can be found in Warn et al. (1995). The procedure does not guarantee preservation of the conservation laws of the full system (7) unless the slaving relationship (20) is exact. Second-order ordinary differential equations, however, cannot exhibit chaotic behavior. Moreover, the stable fixed point of (21) remains (neutrally) stable up to at least the fourth order in ϵ . Hence, we expect that the approximate forms of (21) probably have an invariant, but phase space volume conservation and the preservation of Hamiltonian structure are not expected.

To continue the exploration of phase space, we examined composite Poincaré sections of orbits (of the full system) initialized by the series approximations of the slaving relations to various orders. Figure 7 shows a blowup of a section in the x_4 - x_5 plane (the fast vari-

ables) for initializations up to 16th order, for $b = 0.5$ and $R = 0.2$ [so $\epsilon = 1/(5\sqrt{5})$]. Here, the initial values for the slow variables are $\phi = 0$ and $x_3 = \sqrt{2.04}$, as in Fig. 5b, which lies just outside the separatrix. The initial values of the fast variables x_4 and x_5 are determined by the different initializations, from leading order up to 16th order in the asymptotic case (the numbers denoting the order of the initialization), and by L86’s algorithm (denoted by C in the figure). Evidently, the asymptotic initializations are landing on tori with nonzero J_2 . The asymptotic nature of the initialization procedure is reflected in the fact that the optimal initialization is the 12th-order one, after which the procedure begins to diverge. In contrast, L86’s initialization finds the core of the torus (the only limit to the precision of this statement being numerical). So in this case at least, the slowest invariant manifold is precisely defined even though ϵ is nonzero. (We have reproduced these results for a variety of other initial values of x_3 , both inside and outside the separatrix.) This underscores the point made by L86 that the convergence or nonconvergence of a particular algorithm has nothing to do with the existence or nonexistence of a slow manifold. (Recall that even when invariant tori exist the standard Hamiltonian perturbative approximation procedures tend to diverge; yet this does not preclude the possibility of constructing a convergent approximation. Indeed, this is precisely how the KAM theorem is proven.)

Figure 7 suggests that the asymptotic initializations contain a nonzero amount of free gravity wave activity whereas the core solution contains none. This is verified by a computation of power spectra of x_4 for the solutions corresponding to the 10th and 12th order initializations, and the core solution, shown in Fig. 8. The most noticeable feature of the power spectra is a dominant low-frequency peak with strong overtones, virtually identical in all three cases. This structure is the spectrum of the slow dynamics, essentially that of the elliptical functions. (It does not seem to be much affected by nonzero b .) Among the overtones, one can detect a small peak at $\omega = 1/\epsilon \approx 11.18$ in Fig. 8a, which decreases in Fig. 8b, and disappears altogether in Fig. 8c. This is the free gravity wave signal whose amplitude corresponds to the radius of the circles in Fig. 7. To further illustrate that this is the case, a time series of the difference in the value of x_4 between the 12th-order initialized solution and the core solution is plotted in Fig. 9. One clearly sees an oscillation with period $2\pi\epsilon \approx 0.56$. That the free gravity wave has truly zero amplitude (within numerical limitations) in the core solution is easily seen as follows: since the gravity wave frequency does not correspond to an overtone of the slow periodic motion (Fig. 8), if there were any amplitude at the gravity wave frequency, then the trajectory would have to show up either as a curve or as a set of points in Fig. 7 (depending on whether the fast/slow frequency ratio was an irrational number or a non-

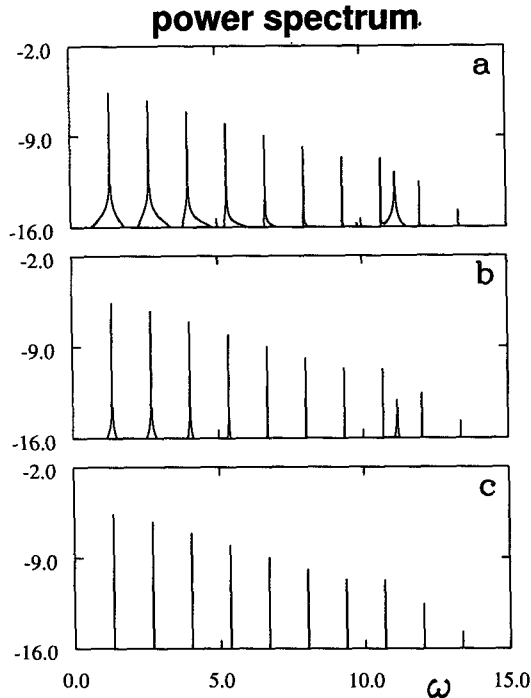


FIG. 8. Power spectra of x_4 for solutions shown in Fig. 7 initialized by (a) the 10th-order asymptotic scheme, (b) the 12th-order asymptotic scheme, and (c) the L86 scheme. Among the overtones of the slow dynamics one can see a free gravity wave peak at $\omega \approx 11.18$, which decreases between (a) and (b), and disappears altogether for (c). The power spectra are calculated from solutions that extend over ~ 950 time units, corresponding to 200 periods of the L86 solution.

integer rational number) but not as a single point (within the numerical limitations).

Figure 10 shows a logarithmic plot of the amplitude of the gravity wave activity, using similar sections as in Fig. 7, of the optimal asymptotic initialization—the order changes for different ϵ —versus $1/\epsilon$, for fixed initial values of ϕ and x_3 . For the smaller values of ϵ we were forced to reduce the time step considerably. The amplitude decreases as $\epsilon \rightarrow 0$, reaching the computational limit for the smallest value of ϵ . The results confirm the asymptotic nature of the classic initialization procedure: for fixed ϵ , there is a minimum nonzero amount of fast motion, but this amount decreases as ϵ decreases. The exponential dependence of the amplitude on $1/\epsilon$ agrees with Lorenz and Krishnamurthy (1987), who found a dependence of the form $\exp(-\pi/2\epsilon)$. The slope of -1.07 in our graph is considerably shallower than $-\pi/2$. However, the results are not strictly comparable; Lorenz and Krishnamurthy (1987) considered the fixed point at the separatrix, where chaos is strongest, and used only a first-order initialization.

There are two important points to be made here. The first is that when the structure of the primary invariant tori persists then a core solution $J_2 = 0$ exists and can

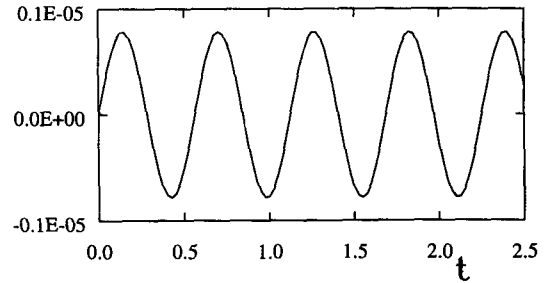


FIG. 9. Time series of the difference in the value of x_4 between the solution corresponding to the 12th-order initialization and the core solution of Fig. 7.

be found numerically. [We have used L86's algorithm to find it here, but it should be cautioned that this works only for an initial condition with $\sin 2\phi = 0$. In a general initialization problem, with given $\sin 2\phi \neq 0$, L86's algorithm could not be used, and one would have to find another convergent perturbative approach (e.g. Baranger et al. 1988; Davies et al. 1992).] This is an important conceptual point, because it seems to be a common belief that free gravity wave motion cannot be completely eliminated in this system and that the minimum gravity wave activity found by classic initialization procedures reflects an inherent "fuzziness" of a slow manifold. Instead, we find that the nonconvergence of classic initialization procedures merely reflects a failure of those procedures and that a slowest invariant manifold, having a dimension equal to that of the slow variables, can be defined for this system—though not for all values of the slow variables. But in the limit $\epsilon \rightarrow 0$, the set of values for which it can be defined approaches full measure (in accordance with the KAM theorem). Moreover, the slowest invariant manifold is nonlinearly stable, since it is surrounded by invariant

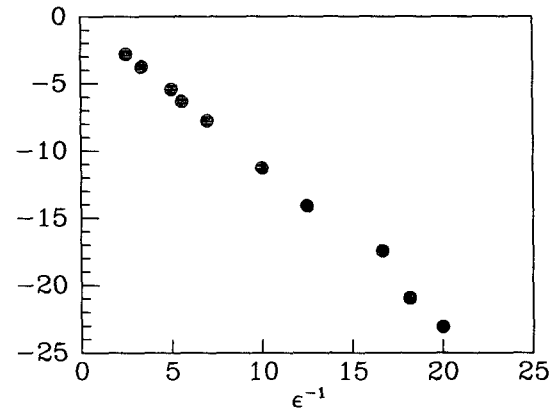


FIG. 10. Logarithmic plot of the amplitude of the gravity wave activity of the optimal asymptotic initialization vs $1/\epsilon$. The gravity wave activity is defined by the radius J_2 of the invariant torus (when it exists) in Poincaré sections similar to Fig. 7.

tori; if a small amount of gravity wave activity is introduced, it will remain bounded for all time (as is seen in the asymptotically initialized solutions).

Yet this slowest invariant manifold has a fuzziness of its own, in that for finite ϵ , it will not be defined where there are strong resonances (i.e., where the invariant tori are destroyed). This is illustrated by Fig. 11, which follows a cross section in x_3 along the $\phi = 0$ axis, with $C = 1$, $b = 0.5$, and $R = 0.5$. We compare the value of x_5 (note that $x_4 = 0$ since $\phi = 0$) corresponding to various algorithms. The dots denote L86's slowest invariant manifold, following his procedure as in Fig. 5; the asterisks denote the second-order asymptotic slaving (zeroth order is of course $x_5 = 0$); the plus signs the fourth-order asymptotic slaving; and the open circles are selected core solutions (a subset of the dots). It is clear that as one crosses the separatrix where $x_3 = \sqrt{2}$ the slowest invariant manifold ceases to be well defined: for two values of x_3 , periodic solutions could not be found, and for another value of x_3 , an unstable periodic orbit was found in addition to a stable one.

The second important point to be made, which complements the first, is that the free gravity wave activity is not the only fast motion in the system and that below a certain amplitude it is overwhelmed by the overtones of the slow dynamics. This overlap between the free gravity waves and the overtone spectrum explains why definitions of a slowest invariant manifold based on a timescale separation are imprecise: there is no timescale separation. [This point has also been made by Warn (1983, unpublished manuscript), Warn and Menard (1986), and L86.]

7. Hamiltonian balanced dynamics

Integrable Hamiltonian balanced models are constructed by imposing the two slaving relations $x_4 = U_4(\phi, x_3)$ and $x_5 = U_5(\phi, x_3)$ as constraints in Hamilton's principle (10). This is equivalent to the use of Lagrange multipliers and yields the balanced bracket

$$[F, G]^* = \left(\frac{\partial F}{\partial \phi} \frac{\partial G}{\partial x_3} - \frac{\partial F}{\partial x_3} \frac{\partial G}{\partial \phi} \right) \frac{1}{(1 + \epsilon g(\phi, x_3))}, \quad (22a)$$

with $g(\phi, x_3)$ defined by

$$g(\phi, x_3) = - \left(b + \frac{\partial U_5}{\partial x_3} \right) \frac{\partial U_4}{\partial \phi} + \frac{\partial U_4}{\partial x_3} \frac{\partial U_5}{\partial \phi}, \quad (22b)$$

together with the Hamiltonian

$$H = -\frac{1}{2} C \cos 2\phi + \frac{1}{2} (x_3^2 + U_4^2(\phi, x_3) + U_5^2(\phi, x_3)). \quad (22c)$$

The bracket (22a) obeys all mathematical requirements by construction for any $U_4(\phi, x_3)$ or $U_5(\phi, x_3)$. More generally, for example for noncanonical systems such

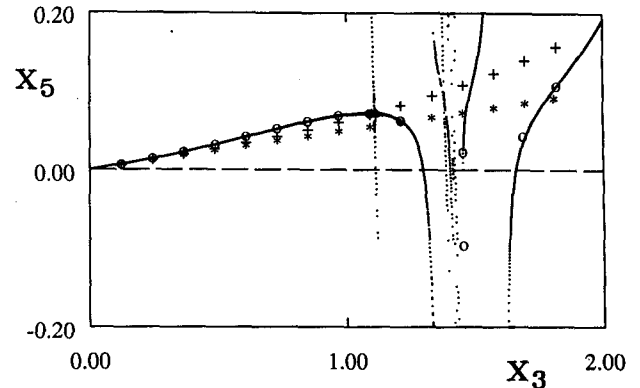


FIG. 11. A cross section of L86's slowest invariant manifold, showing x_5 vs x_3 when $\phi = x_4 = 0$ (for fixed $b = 0.5$, $R = 0.5$, $C = 1$), is given by the dots. The open circles denote selected core solutions (a subset of the dots) for later reference. The second-order slaving relation is shown by "*" and the fourth-order one by "+"; the formulas are given in appendix B. The zeroth-order solution is $x_5 = 0$. No periodic orbit was found for two of the values of x_3 , and for one value of x_3 , an unstable periodic orbit (the lower of the two circles) was found in addition to a stable one. The unperturbed separatrix corresponds to $x_3 = \sqrt{2}$.

as (3), one may appeal to Dirac's theory of Hamiltonian systems with constraints (Dirac 1950, 1958, 1964; Sudarshan and Mukunda 1974).

The Hamiltonian balanced equations corresponding to (22) are

$$\begin{aligned} \frac{d\phi}{dt} &= \frac{1}{1 + \epsilon g(\phi, x_3)} \left(x_3 + U_4 \frac{\partial U_4}{\partial x_3} + U_5 \frac{\partial U_5}{\partial x_3} \right), \\ \frac{dx_3}{dt} &= - \frac{1}{1 + \epsilon g(\phi, x_3)} \\ &\quad \times \left(C \sin 2\phi + U_4 \frac{\partial U_4}{\partial \phi} + U_5 \frac{\partial U_5}{\partial \phi} \right). \quad (23) \end{aligned}$$

If the slaving relations are exact, then (20) and (23) give (21); if, on the other hand, the slaving relations are only approximate, then a hierarchy of Hamiltonian balanced models may be obtained by approximating the slaving relations $U_4(\phi, x_3)$ and $U_5(\phi, x_3)$ in (23).

The procedure of finding asymptotic constraints to the required order, calculating $g(\phi, x_3)$, and constructing the corresponding balanced model is performed using computer algebra. The equations for the balanced dynamics at a specific order are then automatically translated into numerical code and subsequently integrated numerically (although the balanced models are in principle exactly integrable). A fourth-order, multistep, implicit Gear's algorithm is used to integrate the balanced models.

We now examine the accuracy of the various balanced models. Because all the balanced models are restricted to evolve on a phase plane, the evolution is always regular, and classic measures of predictability

are inapplicable. We choose to examine two measures of accuracy: average values of the variables over the orbits and phase shift errors along the orbits.

Figure 12 shows three pairs of panels, each pair consisting of the average values of x_3^2 and x_5^2 , shown as functions of the (positive) value of x_3 when $\phi = 0$ (thereby indicating the location of the orbit on the phase plane). Parameter values are as for Fig. 11. The three pairs of panels correspond to (a) zeroth-order, (b) second-order, and (c) fourth-order models. The solid lines, which are the same in all three cases, indicate the stable periodic core solutions marked by open circles in Fig. 11. The long-dashed lines marked by plus signs indicate the full solution initialized to the given order; there is clearly a substantial amount of gravity wave activity, as seen in $\langle x_5^2 \rangle$, excited by all initializations in the chaotic region around the separatrix. However, one

hesitates to regard this gravity wave activity as spurious, because balance is evidently not a precise concept in the vicinity of the separatrix. In general both balanced models show the same performance and improve as the order is increased.

A measure of the phase drift errors of the different balanced models and initializations is given by

$$f(t) = \frac{1}{t} \int_0^t \Delta\phi(t') dt', \quad (24)$$

where $\Delta\phi$ is the difference in ϕ relative to the stable periodic core solutions. The initialized runs of the full system are denoted by long-dashed lines, the Hamiltonian balanced models by dashed lines, and the classic balanced models by long-short dashed lines. The corresponding differences are plotted on a log-linear graph

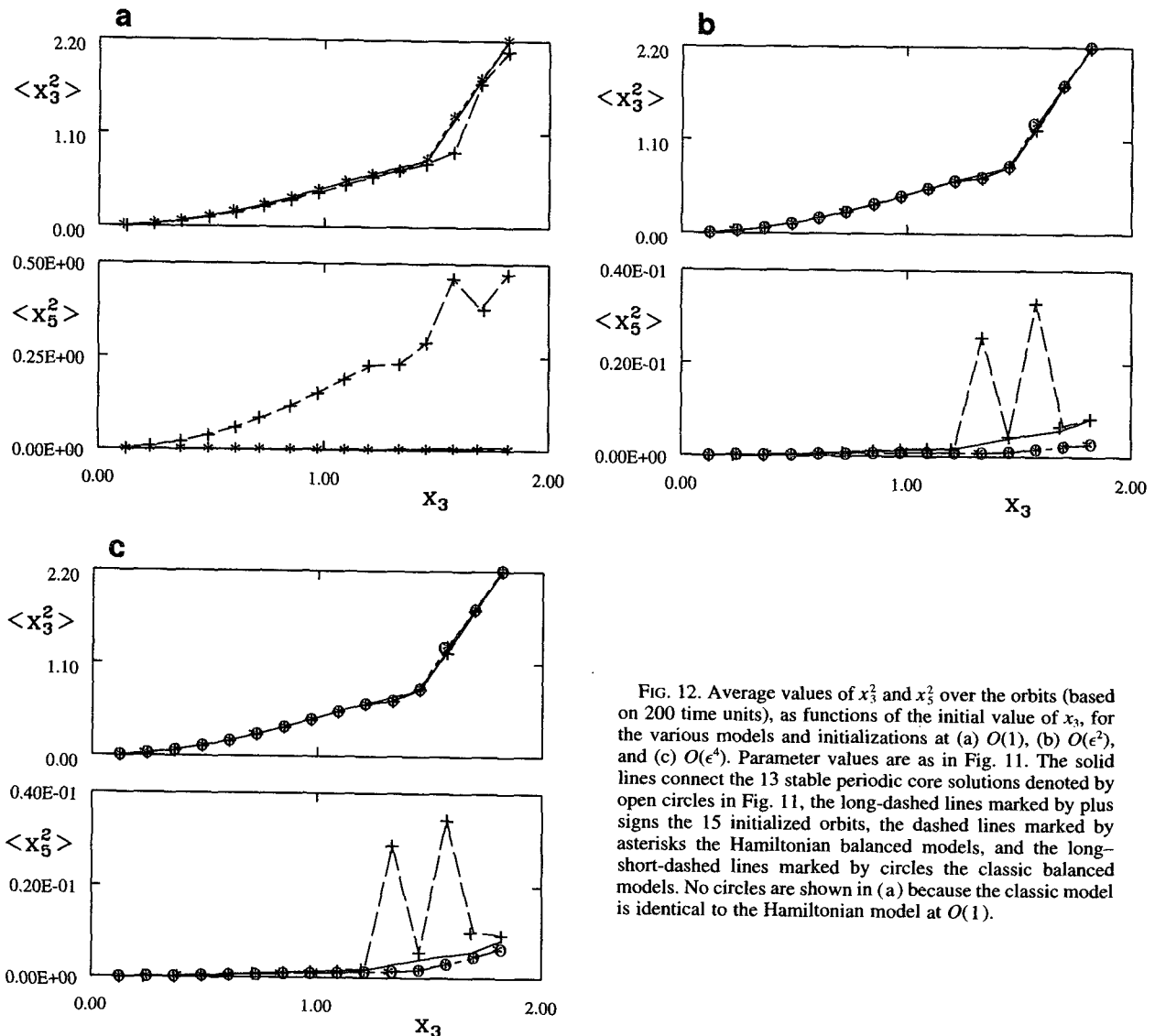


FIG. 12. Average values of x_3^2 and x_5^2 over the orbits (based on 200 time units), as functions of the initial value of x_3 , for the various models and initializations at (a) $O(1)$, (b) $O(\epsilon^2)$, and (c) $O(\epsilon^4)$. Parameter values are as in Fig. 11. The solid lines connect the 13 stable periodic core solutions denoted by open circles in Fig. 11, the long-dashed lines marked by plus signs the 15 initialized orbits, the dashed lines marked by asterisks the Hamiltonian balanced models, and the long-short dashed lines marked by circles the classic balanced models. No circles are shown in (a) because the classic model is identical to the Hamiltonian model at $O(1)$.

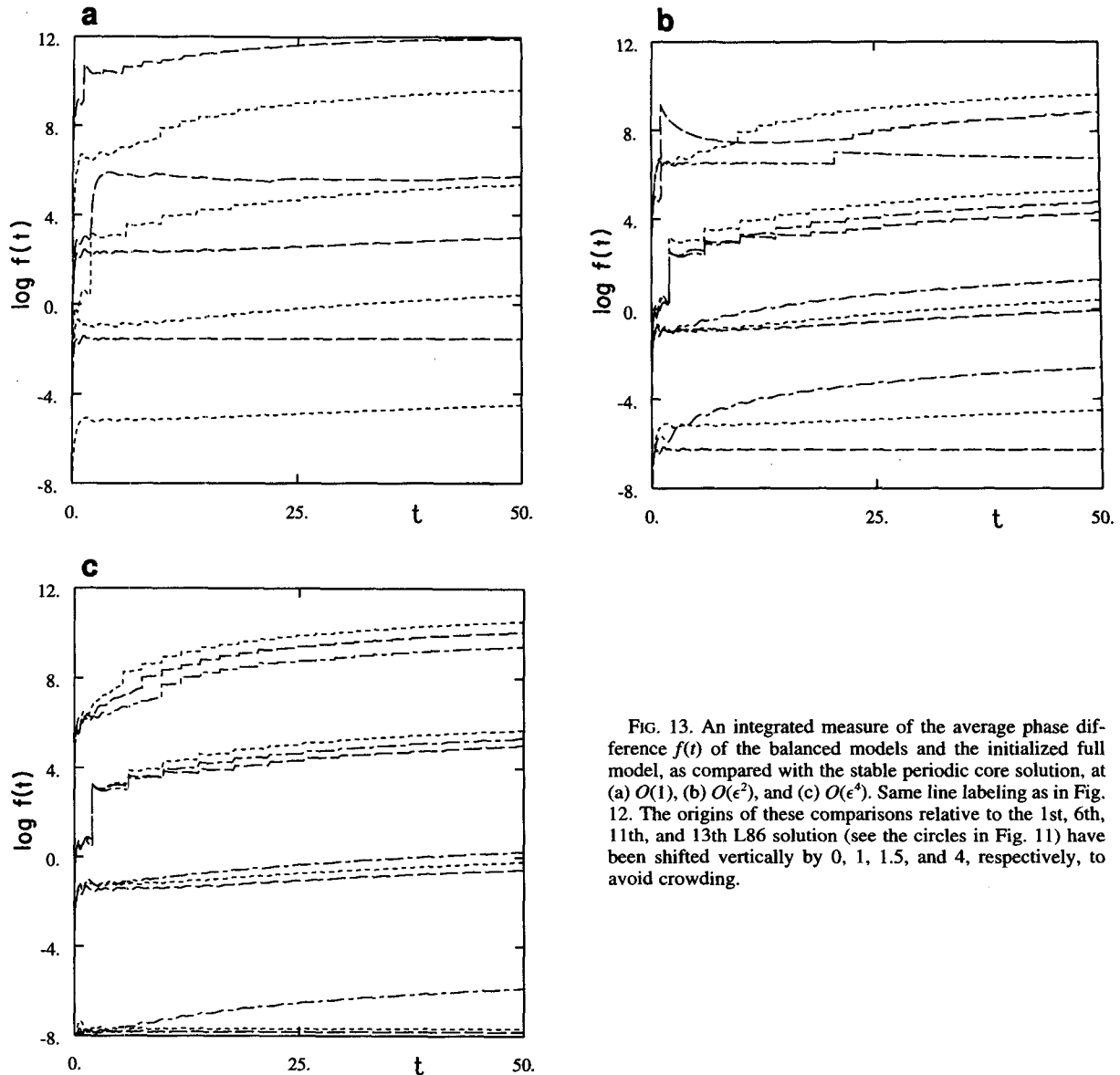


FIG. 13. An integrated measure of the average phase difference $f(t)$ of the balanced models and the initialized full model, as compared with the stable periodic core solution, at (a) $O(1)$, (b) $O(\epsilon^2)$, and (c) $O(\epsilon^4)$. Same line labeling as in Fig. 12. The origins of these comparisons relative to the 1st, 6th, 11th, and 13th L86 solution (see the circles in Fig. 11) have been shifted vertically by 0, 1, 1.5, and 4, respectively, to avoid crowding.

in Fig. 13 for (a) zeroth order, (b) second order, and (c) fourth order. We can see that the initialized models remain closest to the slowest invariant manifold, followed by the Hamiltonian balanced models, and finally by the classic balanced models—except in the chaotic region around the separatrix where both balanced models perform poorly. All differences reduce as the order is increased.

Given the asymptotic character of the small- ϵ expansion of the slaving relations, we expect that there is an optimal balanced model. From the above results we may conclude that the fourth-order balanced models are superior to the second-order models and that among the balanced models tested here the Hamiltonian fourth-order balanced model appears to be the best one.

8. Summary and discussion

We have reexamined the concept of a slowest invariant manifold in the context of the conservative low-order model of Lorenz (1986). The L86 model is shown (section 3) to be a two degree of freedom canonical Hamiltonian dynamical system, consisting of two coupled oscillators. One oscillator is the nonlinear pendulum and corresponds to a vorticity-triad interaction; the other is the harmonic oscillator and corresponds to a gravity wave. The amplitude of the motion is given by the nondimensional parameter $\epsilon = bR/\sqrt{1 + b^2}$, where b is the rotational Froude number and R is the Rossby number. (The Froude number is then bR .)

The ratio of the nonlinear frequency to the linear frequency is $O(\epsilon)$. Thus, in the small amplitude limit $\epsilon \rightarrow 0$, and *only* in the small amplitude limit, a separation of timescales exists, and we may refer to the vorticity-triad oscillation as “slow” and the gravity wave oscillation as “fast.” Balanced dynamics concerns the description of the slow dynamics under such a separation of timescales. In this model the small amplitude limit $\epsilon \rightarrow 0$ can be approached in two ways: $b \rightarrow 0$, or $R \rightarrow 0$. Both of these correspond to well-known scalings that yield balanced dynamics in the full shallow-water equations.

In the same limit $\epsilon \rightarrow 0$, the two oscillators decouple, and the system is integrable. It is shown (section 3) that the degenerate isoenergetic version of the KAM theorem applies to this system, with ϵ as the perturbation parameter. For $\epsilon = 0$, only the linear (gravity wave) oscillator enters the Hamiltonian, so the system is degenerate. The $O(\epsilon)$ correction to the Hamiltonian breaks the degeneracy but does not destroy integrability; the intermediate, $O(\epsilon)$ Hamiltonian is isoenergetically nondegenerate and consists of the two uncoupled oscillators. Solutions of the intermediate system may be expressed in terms of action-angle variables (I_i, θ_i) and lie on invariant two-tori within a three-dimensional constant energy hypersurface.

For the uncoupled system, the slow dynamics is defined by setting the action of the fast oscillator to zero: $I_2 = 0$. This corresponds to linear normal-mode initialization. Geometrically, $I_2 = 0$ defines the core of the invariant tori when viewed in a cross section in the fast-variable plane. The union of all such cores (for all possible values of the slow variables) evidently defines a “slowest invariant manifold” for the uncoupled system. We resist using the term “slow manifold” because even the “slow” motion contains high-frequency overtones, which overlap with the fast frequency.

The higher-order corrections to the Hamiltonian couple the two oscillators and render the system nonintegrable. However, the KAM theorem implies that for sufficiently small (but finite) ϵ most of the invariant tori survive. In particular, over most of phase space one can construct new action-angle variables (J_i, φ_i) that are smooth deformations of the original action-angle variables (I_i, θ_i) . In this case, “most” means a set whose complement approaches zero measure exponentially fast as $\epsilon \rightarrow 0$, like $\exp(-c/\epsilon)$. The resonant tori break up under the perturbation and around them zones of chaos emerge, containing secondary tori (island structures). The existence of chaos in this system has recently been rigorously established by Camassa (1995). For this model, chaotic motion is necessarily unbalanced, because it must involve all degrees of freedom on the constant energy hypersurface.

Wherever the primary invariant tori survive the perturbation, a natural, and precise, definition of a slowest invariant manifold is obtained by setting the new fast

action to zero: $J_2 = 0$. For sufficiently small ϵ , the KAM theorem ensures that this can be done for most values of the new slow variables (J_1, φ_1) . Such core solutions will not however be defined for J_1 corresponding to resonant tori in the intermediate system, which are distributed densely throughout phase space. This means that even for infinitesimally small ϵ the slowest invariant manifold $J_2 = 0$ will not be a simply connected manifold, because it will be punctured densely with gaps, and its structure will be fractal. However, the total measure of these gaps goes to zero exponentially fast as $\epsilon \rightarrow 0$. Moreover, the KAM theorem implies that this slowest invariant manifold is nonlinearly stable: if a small amount of free gravity wave activity is introduced by setting $J_2 \neq 0$, then its amplitude must remain forever bounded. This applies even within the chaotic regimes, which are sandwiched between invariant tori.

We have explored the phase space geometry of regular and chaotic orbits numerically (section 4). It is found that for $\epsilon = 0.1$ one can define such a slowest invariant manifold over a significant fraction of the slow phase space on either side of the separatrix of the nonlinear oscillator, but as ϵ gets much larger, this becomes increasingly difficult to do: the slowest invariant manifold evaporates.

Our results shed light on previous studies of the L86 model. The fact that a slowest invariant manifold can be defined only where the invariant tori survive is consistent with the demonstration by Lorenz and Krishnamurthy (1987) that a slow manifold cannot be global: they follow the unstable manifold emanating out of the unstable fixed point $(x_3, \phi) = (0, \pi/2)$ and find that fast oscillations develop. From the standpoint of Hamiltonian perturbation theory this makes perfect sense: Lorenz and Krishnamurthy were attempting to follow the separatrix solution, which of course breaks up under the perturbation. Indeed, the slowest invariant manifold *cannot* be defined everywhere—most noticeably in the vicinity of the separatrix (Fig. 11). This point has been further developed by Camassa (1995).

L86 observed that slaved solutions of his system are necessarily periodic and proposed that the slowest invariant manifold should be defined to be the set of all periodic solutions having single-valued slaving relations between the fast and slow variables. Such a manifold is certainly invariant; the question is whether it is in any sense slow. In a similar vein Boyd (1994) proposed that *all* periodic solutions should constitute the slowest invariant manifold.

These definitions, however, are seen to be problematic in light of our results. In the uncoupled case there are two kinds of periodic solutions: those corresponding to rational tori and those for which one of the actions is zero (including both rational and irrational tori). Boyd’s definition of the slowest invariant manifold would include all such solutions, even though most will have nonzero free fast motion. Lorenz’s (1986)

definition of the slowest invariant manifold, which requires a single-valued slaving relation, eliminates the rational tori with noninteger frequency ratios and the solutions with zero slow action. Yet it is clear that in the uncoupled case the only solutions with zero free fast motion are those with zero fast action: $I_2 = 0$.

In the coupled case, the continuous branches of periodic solutions corresponding to the rational tori breakup under the perturbation, but periodic solutions can still be found in the same region of phase space. This is not surprising: the Poincaré–Birkhoff fixed-point theorem states that rational tori break up into stable and unstable periodic solutions, which are surrounded by chaos. But these periodic solutions are not smoothly related to the rational tori, would be difficult to find by perturbative methods based on the uncoupled solutions, and contain nonzero amounts of free fast motion in that they wind around the primary tori. In contrast, the core solutions $I_2 = 0$ are smoothly transformed to new core solutions $J_2 = 0$ for most irrational tori, which means for most values of J_1 .

The problem with definitions of a slowest invariant manifold based on periodicity is that periodicity can reflect a kind of “false slaving.” In the small-amplitude regime $\epsilon \ll 1$, the full system is close to being integrable and most orbits are regular. For a regular orbit any variable can be written in terms of time and thus can be written in terms of any other variable through the elimination of time. Hence, any regular orbit is formally a slaved solution. This is demonstrated explicitly in the uncoupled case (section 5). Yet regular orbits generally contain nonzero free fast motion. When the regular orbit is quasi-periodic, then the slaving relation has a dense infinity of branches and is meaningless as a reduced manifold. When the regular orbit is periodic, then the slaving relation will generally be multivalued. When the winding number of the orbit is an integer, then the slaving relation is single valued. But such an orbit may still contain a nonzero amount of free fast motion, which just happens to be an exact harmonic of the slow motion. Thus, any slaving that comes purely from integrability, through the elimination of time, is really a false slaving.

Our results also shed light on asymptotic approaches to defining “slowness” based on the formal separation of timescales when $\epsilon \ll 1$. Even when the core solutions $J_2 = 0$ can be unambiguously identified, classic perturbative approximations are only asymptotic; solutions initialized by such procedures contain a nonzero amount of fast action J_2 , as is clearly reflected in composite Poincaré sections (Fig. 7). This minimum amount of free gravity wave activity that is excited by the asymptotic initialization procedure goes to zero as $\epsilon \rightarrow 0$ (Fig. 10), like $\exp(-c/\epsilon)$, but is finite for any finite ϵ . This gravity wave activity has been previously taken by some to indicate the inherent “fuzziness” of the slow manifold. This interpretation may be correct for more complex systems (see further discussion be-

low) but it is definitely not correct for this system: as L86 points out, the convergence or nonconvergence of a particular approximation procedure has nothing to do with the existence or nonexistence of a slow manifold. We show (Fig. 8) that the power spectrum of the optimal asymptotic initialization contains a nonzero peak at the gravity wave frequency, which disappears altogether for the core solution $J_2 = 0$ while the rest of the spectrum (consisting of the slow frequency and its overtones) remains unchanged; the difference between the two solutions consists of a clean sinusoidal wave at the gravity wave frequency (Fig. 9). On the other hand, the fact that the gravity wave frequency may be surrounded by overtones of the slow frequency of equal or even larger amplitude explains why perturbative definitions of a slow manifold based on a timescale separation are imprecise: there really is no timescale separation.

A rather different argument against the existence of a precise slowest invariant manifold is that of Errico (1984) and Warn (1986) that the tendency toward statistical mechanical equilibrium will inevitably excite gravity waves. It has been shown that this argument does not apply to the L86 system, because the dynamics is constrained by KAM tori and is decidedly nonergodic. Such behavior is, of course, generic for perturbed, nearly integrable Hamiltonian systems—the classic example being the Fermi–Pasta–Ulam system of coupled oscillators, which even for a large number of oscillators is constrained by KAM tori and does not thermalize (for sufficiently weak coupling).

The Hamiltonian structure of the full dynamics permits the construction of a hierarchy of Hamiltonian balanced models (section 7), up to arbitrarily high order in ϵ . These models have been compared with so-called classic balanced models (analogous to the Charney balance model), which are generally not Hamiltonian; with asymptotic initializations of the full system; and with the core solutions. Regarding the core solutions as the exact slowest invariant manifold whose dynamics we wish to simulate as accurately as possible, the Hamiltonian models show good accuracy in average amplitude and in phase drift, with well-bounded activity in the fast variables, when compared with the other approximate balanced models and solutions.

This study has focused on a single model dynamical system and thus has many limitations. Our reason for focusing on the L86 model is twofold: first, it is probably the simplest dynamical system containing the essential ingredients of a fast linear and a slow nonlinear oscillator, coupled nonlinearly, with physically meaningful dimensionless parameters, and second, it has become a paradigm system for recent investigations of balanced dynamics—yet despite much study, many questions previously remained unanswered.

The single most important limitation is the integrability of the uncoupled slow dynamics of the L86 model. The consequence of integrability is that the slow

dynamics has a frequency spectrum consisting of a fundamental and its higher-frequency overtones. For most slow solutions these overtones can be distinguished from the gravity wave frequency; overlap will occur only for a resonance. The KAM theorem implies that for a sufficiently small perturbation most nonresonant tori will survive the perturbation, and one can therefore continue to distinguish the spectrum of the free fast dynamics (which now will contain overtones of its own) from that of the slow dynamics. It is this distinction that allows one to unambiguously set the amplitude of the free fast motion to zero, thereby obtaining a slowest invariant manifold. But for a sufficiently complex system (and certainly for a continuous system), the uncoupled slow dynamics will be chaotic, with a full spectrum. The mathematical consequence of this is that the KAM theorem cannot be invoked; the physical consequence is that one cannot unambiguously identify free fast motion.

A second limitation is that the L86 model contains only one gravity wave. The full nine-component model (1) would seem in principle to be a more representative model of "primitive equations" dynamics because like the full shallow-water equations it has one gravity wave (i.e., two fast degrees of freedom) for every vortical mode. However, even if a Hamiltonian representation of (1) could be found, the KAM theorem would not apply because the intermediate (uncoupled) system would contain three (linear) gravity waves and would therefore be degenerate. However, this might only be a technical problem: many perturbed Hamiltonian systems, for example the Hénon–Heiles system, exhibit KAM-like behavior even though the KAM theorem does not strictly apply.

A third limitation is that the accessible phase space (namely the energy–Casimir hypersurface) of the L86 model is three dimensional, so the invariant tori partition the space. The KAM theorem may well establish the presence of invariant tori for higher-dimensional systems, but these tori would no longer partition the space and orbits could drift significantly, a phenomenon known as "Arnold diffusion." The implication is that the core solutions described here would no longer be nonlinearly stable. However, this may also only be a technical problem because the diffusion speed is found in practice to be superexponentially small in terms of the closest distance to an invariant torus (Nekhoroshev 1977; Morbidelli and Giorgilli 1995).

A fourth limitation is that a low-order model has a restricted range of wavenumbers, which suppresses gravity wave generation by the process of "Lighthill emission" (M. E. McIntyre 1995, personal communication). In Lighthill emission (so named because of the parallel with Lighthill's theory of aerodynamic sound generation) low Froude number vortical motion at a given horizontal scale can directly excite gravity waves at much longer wavelengths, by matching frequencies. In this case, of course, even a formal separation

of timescales cannot be guaranteed if one has a sufficiently wide range of length scales available in the problem.

A rather different limitation concerns our investigation of only the conservative form of the L86 model. Many previous investigations of balanced dynamics in low-order models have relied strongly on forcing and dissipation (e.g., Kopell 1985; Jacobs 1991). But the appropriate parametrization of forcing and dissipation for a low-order model is highly problematical, to say the least. The fundamental ingredient in balanced dynamics is the nonlinear interaction between fast and slow degrees of freedom, which is part of the conservative, Hamiltonian structure of geophysical fluid dynamics (e.g., Shepherd 1990). In this regard we note that the development of balanced models, and of linear and nonlinear normal-mode initialization methods, has occurred through the theoretical analysis of the conservative form of the relevant governing equations. Of course, dissipation may ultimately be important in controlling the fast degrees of freedom but this dissipation probably occurs either through radiation of the fast waves to infinity (classic geostrophic adjustment) or through a nonlinear cascade of fast energy to small scales (Sadourny 1975; Warn 1986)—neither of which can be said to be appropriately represented by dissipation in low-order models. So while we accept that the conservative framework is limiting, it is not at all clear to us how one should deal with this limitation in the context of low-order models.

Acknowledgments. The authors benefitted from discussions with P. J. Kushner, M. E. McIntyre, P. J. Morrison, K. Ngan, G. K. Vallis, and T. Warn. We are indebted to the referees for pointing out some valuable references. This work represents part of OB's Ph.D. thesis at the University of Toronto. OB is supported by a Simcoe Special Fellowship; TGS is supported by grants from the Natural Sciences and Engineering Research Council and the Atmospheric Environment Service of Canada.

APPENDIX A

Rigid-Body Rotation for $b = 0$

For $b = 0$ the vorticity-triad dynamics of (3) is given by

$$\frac{dx_1}{dt} = -x_2 x_3, \quad \frac{dx_2}{dt} = x_1 x_3, \quad \frac{dx_3}{dt} = -x_1 x_2, \quad (\text{A1})$$

with bracket

$$[F, G]_1 = -\frac{\partial F}{\partial x_1} x_2 \frac{\partial G}{\partial x_3} + \frac{\partial F}{\partial x_2} x_1 \frac{\partial G}{\partial x_3} + \frac{\partial F}{\partial x_3} \left(x_2 \frac{\partial G}{\partial x_1} - x_1 \frac{\partial G}{\partial x_2} \right) \quad (\text{A2})$$

or

$$[F, G]_2 = -\frac{1}{2} \frac{\partial F}{\partial x_1} x_3 \frac{\partial G}{\partial x_2} + \frac{1}{2} \frac{\partial F}{\partial x_2} \times \left(x_3 \frac{\partial G}{\partial x_1} + x_1 \frac{\partial G}{\partial x_3} \right) - \frac{1}{2} \frac{\partial F}{\partial x_3} x_1 \frac{\partial G}{\partial x_2} \quad (\text{A3})$$

for arbitrary functions F and G , and Hamiltonian

$$H' = \frac{1}{2} (x_1^2 + 2x_2^2 + x_3^2). \quad (\text{A4})$$

The Casimir invariants of (A2) and (A3), namely, the solutions of $[C, G]_{1,2} = 0$ for arbitrary G , are $C_1 = \frac{1}{2}(x_1^2 + x_2^2)$ and $C_2 = \frac{1}{2}(x_3^2 - x_1^2)$, respectively. The brackets $[F, G]_1$ and $[F, G]_2$ are different in the sense that the associated Casimir invariants are different but both brackets satisfy Jacobi's identity. One may also use the Hamiltonians $H_1 = H' - 2C_1 = C_2$ and $H_2 = H' - C_2 = 2C_1$ for the brackets (A2) and (A3), respectively.

The standard rigid-body dynamics (Arnold 1978) of the angular momenta $M = (M_1, M_2, M_3)^T$ has the Hamiltonian representation

$$\frac{dF}{d\tau} = [F, H_R]_R, \quad (\text{A5})$$

with the bracket $[\cdot, \cdot]_R$ given by

$$[F, G]_R = \frac{\partial F}{\partial M_1} \left(M_3 \frac{\partial G}{\partial M_2} - M_2 \frac{\partial G}{\partial M_3} \right) + \frac{\partial F}{\partial M_2} \left(M_1 \frac{\partial G}{\partial M_3} - M_3 \frac{\partial G}{\partial M_1} \right) + \frac{\partial F}{\partial M_3} \left(M_2 \frac{\partial G}{\partial M_1} - M_1 \frac{\partial G}{\partial M_2} \right) \quad (\text{A6})$$

and the Hamiltonian H_R given by

$$H_R = \frac{1}{2} \left(\frac{M_1^2}{I_1} + \frac{M_2^2}{I_2} + \frac{M_3^2}{I_3} \right), \quad (\text{A7})$$

with I_i ($i = 1, 2, 3$) being the moments of inertia. The corresponding Casimir function C_R is

$$C_R = \frac{1}{2} (M_1^2 + M_2^2 + M_3^2).$$

It may be verified that with the identification

$$M_1 = x_1, \quad M_2 = -\sqrt{2}x_2, \quad M_3 = x_3, \quad \text{and} \quad t = \sqrt{2}\tau, \quad (\text{A8})$$

as well as the choice

$$\frac{1}{I_1} = \frac{1 + 2I_3}{I_3}, \quad \frac{1}{I_2} = \frac{1 + I_3}{I_3}, \quad I_3 = I_3,$$

the dynamics (A5) yields the system (A1).

APPENDIX B

Slaving Relations to Fourth Order

In this appendix we give the asymptotic series solutions of the superbalance equations (20) up to fourth order in the Rossby number. At leading order in ϵ the solution of (20) is

$$U_4 = U_5 = 0; \quad (\text{B1})$$

at first order

$$U_4 = -\epsilon b C \sin(2\phi), \quad U_5 = 0; \quad (\text{B2})$$

at second order

$$U_4 = -\epsilon b C \sin(2\phi), \\ U_5 = 2\epsilon^2 b x_3 C \cos(2\phi); \quad (\text{B3})$$

at third order

$$U_4 = -\epsilon b C \sin(2\phi) [1 + 4\epsilon^2 x_3^2 + 2\epsilon^2 C \cos(2\phi)], \\ U_5 = 2\epsilon^2 b x_3 C \cos(2\phi); \quad (\text{B4})$$

and at fourth order

$$U_4 = -\epsilon b \sin(2\phi) C [1 + 4\epsilon^2 x_3^2 + 2\epsilon^2 C \cos(2\phi)], \\ U_5 = 2\epsilon^2 b x_3 C [\cos(2\phi) - 2\epsilon^2 b^2 C \cos^2(2\phi) \\ + 4\epsilon^2 x_3^2 \cos(2\phi) + 2\epsilon^2 C \cos^2(2\phi) \\ - 6\epsilon^2 C \sin^2(2\phi)]. \quad (\text{B5})$$

REFERENCES

- Arnold, V. I., 1978: *Mathematical Methods of Classical Mechanics*. Springer-Verlag, 508 pp.
- , and A. Avez, 1968: *Ergodic Problems of Classical Mechanics*. Addison-Wesley, 286 pp.
- , V. V. Kozlov, and A. I. Neishtadt, 1988: *Dynamical Systems III. Mathematical Aspects of Classical and Celestial Mechanics*. Springer-Verlag, 291 pp.
- Baer, F., 1977: Adjustments of initial conditions required to suppress gravity oscillations in nonlinear flows. *Beitr. Phys. Atmos.*, **50**, 350–366.
- , and J. J. Tribbia, 1977: On complete filtering of gravity modes through nonlinear initialization. *Mon. Wea. Rev.*, **105**, 1536–1539.
- Baranger, M., K. T. R. Davies, and J. H. Mahoney, 1988: The calculation of periodic trajectories. *Ann. Phys.*, **186**, 95–110.
- Bokhove, O., 1993: On Hamiltonian balanced models. Preprint, *Ninth Conf. on Atmospheric and Oceanic Waves and Stability*, San Antonio, TX, Amer. Meteor. Soc., 367–368.
- Boyd, J. P., 1994: The slow manifold of a five-mode model. *J. Atmos. Sci.*, **51**, 1057–1064.
- Camassa, R., 1995: On the geometry of an atmospheric slow manifold. *Physica D*, **84**, 357–397.
- Charney, J. G., 1948: On the scale of atmospheric motions. *Geophys. Publik.*, **17**, 251–265.
- , 1949: On a physical basis for numerical prediction of large-scale motions in the atmosphere. *J. Meteor.*, **6**, 371–385.
- , 1955: The use of the primitive equations of motion in numerical prediction. *Tellus*, **7**, 22–26.
- , 1963: A note on large-scale motion in the tropics. *J. Atmos. Sci.*, **20**, 607–609.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

- Davies, K. T. R., T. E. Huston, and M. Baranger, 1992: Calculations of periodic trajectories for the Hénon–Heiles Hamiltonian using the monodromy method. *Chaos*, **2**, 215–230.
- Dirac, P. A. M., 1950: Generalized Hamiltonian dynamics. *Can. J. Math.*, **2**, 129–148.
- , 1958: Generalized Hamiltonian dynamics. *Proc. Roy. Soc. London, Ser. A*, **246**, 326–332.
- , 1964: Lectures on quantum mechanics. Belfer Graduate School of Science, Yeshiva University, 87 pp.
- Errico, R. M., 1982: Normal mode initialization and the generation of gravity waves by quasi-geostrophic forcing. *J. Atmos. Sci.*, **39**, 573–586.
- , 1984: The statistical equilibrium solution of a primitive-equations model. *Tellus*, **36A**, 42–51.
- Gear, C. W., 1971: *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice-Hall, 253 pp.
- Jacobs, S. J., 1991: Existence of a slow manifold in a model system of equations. *J. Atmos. Sci.*, **48**, 893–901.
- Kopell, N., 1985: Invariant manifolds and the initialization problem for some atmospheric equations. *Physica D*, **14**, 203–215.
- Kreiss, H.-O., 1979: Problems with different time scales for ordinary differential equations. *SIAM J. Numer. Anal.*, **16**, 980–998.
- , 1980: Problems with different time scales for partial differential equations. *Comm. Pure Appl. Math.*, **33**, 399–439.
- Leith, C. E., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.*, **37**, 958–968.
- Littlejohn, R. G., 1982: Singular Poisson tensors. *Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems*, Vol. 88, M. Tabor and Y. M. Treve, Eds., Amer. Inst. Phys., 47–66.
- Lorenz, E. N., 1980: Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.*, **37**, 1685–1699.
- , 1986: On the existence of a slow manifold. *J. Atmos. Sci.*, **43**, 1547–1557.
- , 1992: The slow manifold—What is it? *J. Atmos. Sci.*, **49**, 2449–2451.
- , and V. Krishnamurthy, 1987: On the nonexistence of a slow manifold. *J. Atmos. Sci.*, **44**, 2940–2950.
- Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with application to normal mode initialization. *Beitr. Phys. Atmos.*, **50**, 253–271.
- McIntyre, M. E., 1994: Numerical weather prediction: An updated vision of the future. *The Life Cycles of Extratropical Cyclones*, Vol. 1, S. Grønås and M. Shapiro, Eds., University of Bergen, 275–286.
- Morbidelli, A., and A. Giorgilli, 1995: Superexponential stability of KAM tori. *J. Stat. Phys.*, **78**, 1607–1617.
- Nekhoroshev, N. N., 1977: An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. *Russ. Math. Surv.*, **32**, 1–65.
- Sadourny, R., 1975: The dynamics of finite-difference models of the shallow-water equations. *J. Atmos. Sci.*, **32**, 680–689.
- Shepherd, T. G., 1990: Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Advances Geophysics*, Vol. 32, Academic Press, 287–338.
- Spall, M. A., and J. C. McWilliams, 1992: Rotational and gravitational influences on the degree of balance in the shallow-water equations. *Geophys. Astrophys. Fluid Dyn.*, **64**, 1–29.
- Sudarshan, E. C. G., and N. Mukunda, 1974: *Classical Dynamics: A Modern Perspective*. John Wiley & Sons, 615 pp.
- van Kampen, N. G., 1985: Elimination of fast variables. *Phys. Rep.*, **124**, 69–160.
- Vautard, R., and B. Legras, 1986: Invariant manifolds, quasi-geostrophy and initialization. *J. Atmos. Sci.*, **43**, 565–584.
- Warn, T., 1986: The statistical mechanical equilibria of the shallow water equations. *Tellus*, **38A**, 1–11.
- , and R. Menard, 1986: Nonlinear balance and gravity–inertial wave saturation in a simple atmospheric model. *Tellus*, **38A**, 285–294.
- , O. Bokhove, T. G. Shepherd, and G. K. Vallis, 1995: Rossby-number expansions, slaving principles, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*, **121**, 723–739.
- Yoshida, H., 1990: Construction of higher order symplectic integrators. *Phys. Lett. A*, **150**, 262–268.