Transport, Stretching, and Mixing in Classes of Chaotic Tangles

Thesis by
Darin Beigie

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1993

Submitted June 2, 1992

Acknowledgements

I offer thanks to my advisors Professors Anthony Leonard and Stephen Wiggins for their guidance and support. This work was supported by a National Science Foundation Graduate Fellowship, the National Science Foundation Presidential Young Investigator Program, the Office of Naval Research Young Investigator Program, and the Air Force Office of Scientific Research, Grant No. AFOSR-91-0241.

Abstract

We use global stable and unstable manifolds of invariant hyperbolic sets as templates for studying the dynamics within classes of homoclinic and heteroclinic chaotic tangles, focusing on transport, stretching, and mixing within these tangles. These templates are exploited in the context of *lobes* in phase space mapping within invariant lobe structures formed out of the intersecting global stable and unstable manifolds. Our interest lies in: (a) extending the templates and their applications to fundamentally larger classes of dynamical systems, (b) expanding the description of dynamics offered by the templates, and (c) applying the templates to the study of various nonlinear physical phenomena, such as stirring and mixing under chaotically advecting fluids and molecular dissociation under external electromagnetic forcing. These and other nonlinear physical phenomena are intimately connected to the underlying chaotic dynamics, and describing these processes encourages study of finite-time, or transient, phenomena as well as asymptotics, the former being much more virgin territory from a dynamical systems perspective. Under the rubric of themes (a)-(c) we offer five studies.

(i) One of the canonical classes of dynamical systems in which these templates have been exploited is defined by 2D time-periodic vector fields, where the analysis reduces to a 2D Poincaré map. In this instance, one is well-equipped with basic elements of dynamical systems theory associated with 2D maps, such as the Smale horseshoe map paradigm, KAM-tori, hyperbolic fixed points and their global stable and unstable manifolds that define the tangle boundaries, classical Melnikov theory, and so on. Our first study performs a systematic extension of the dynamical system constructs associated with 2D time-periodic vector fields to apply to 2D vector fields with more complicated time dependences. In particular, we focus on 2D vector fields with quasiperiodic, or multiple-frequency, time dependence. Any extension past the time-periodic case requires the fundamental generalization from 2D maps to *sequences* of 2D nonautonomous maps. To large extent the constructs associated with 2D Poincaré maps are found to be robust under this generalization. For example, the Smale horseshoe map generalizes to a *traveling horseshoe map sequence*, hyperbolic fixed points generalize to points that live on *invariant normally hyperbolic tori*, and invariant 2D chaotic tangles generalize to *sequences* of 2D chaotic tangles derived from an invariant tangle embedded in a higher-dimensional phase space. It is within the setting of 2D lobes mapping within a *sequence* of 2D lobe structures that one has a template for systematic study of the dynamics

generated by multiple-frequency vector fields. Dynamical systems tools with which to study these systems include: (i) a *generalized Melnikov theory* that offers an approximate analytical measure of stable and unstable manifold separation in the tangles, the basis for a variety of analytical studies, and (ii) a *double phase slice sampling method* that allows for numerical computation of *precise 2D slices* of the higher-dimensional invariant chaotic tangles, the basis for numerical work. The Melnikov function defines *relative scaling functions* which give an analytical measure of the relative importance of each frequency on manifold separation. With the template and tools in hand, we study multiple-frequency dynamics and compare with single-frequency dynamics. We recast lobe dynamics under a bi-infinite sequence of nonautonomous maps in closed form by exploiting underlying periodicity properties of the vector field, and present numerical simulations of sequences of chaotic tangles and lobe dynamics within these tangles. In contrast to lobes of equal area mapping within a *fixed* 2D lobe structure found under single-frequency forcing, we find lobes of varying areas mapping within a sequence of lobe structures that are distorting and breathing from one time sample to the next, affording greater freedom in the nature of the dynamics. Our primary focus in this new setting is on *phase space transport* (we consider *stretching* and *mixing* in other contexts in later studies). The non-integrable motion in chaotic tangles allows for *transport* between various regions of phase space, in particular, between regions corresponding to qualitatively different types of motion, such as bounded and unbounded motion. This transport is intimately connected to basic physical processes, such as the fluid mixing and molecular dissociation processes. Transport theory refers to the enterprise where one uses a combination of invariant manifold theory, Melnikov theory, numerical simulation and/or approximate models such as Markov models, to partition phase space into regions of qualitatively different behavior (such as bounded and unbounded motion), establish complete and partial barriers between the regions, identify the *turnstile lobes* that are the gateways for transport across partial barriers, and then study in the context of lobe dynamics such phase space transport issues as *flux* and *escape rates* from a particular region. The formal construction of a transport theory for multiple-frequency vector fields is more involved than in the single-frequency case, as a consequence of more complicated manifold geometry. This geometry is uncovered and explored, however, via theorems and numerical studies based on Melnikov theory. We then partition phase space and define turnstiles in the higher-dimensional autonomous setting, and from this obtain the sequence of partitions and turnstiles in the 2D nonautonomous setting. A

main new feature of transport is its manifestation in the context of a sequence of *time-dependent* regions, and we argue this is consistent with a Lagrangian viewpoint. We then perform a detailed study of such transport properties as flux, lobe geometry, and lobe content. In contrast to the single-frequency case, where a single flux suffices, in the multiple-frequency case a variety of fluxes are allowed, such as different types of instantaneous, finite-time average, and infinite-time average flux. We find for certain classes of multiple-frequency forcing that *infinite-time* average flux is maximal in a particular *single-frequency limit*, but that the spatial variation of lobe areas found in multiple-frequency systems affords greater freedom to enhance or diminish *finite-time* transport quantities. We illustrate our study with a quasiperiodically oscillating vortical flow that gives rise to chaotic fluid trajectories and a quasiperiodically forced Duffing oscillator. We explain how the analysis generalizes to vector fields with more complicated time dependences than quasiperiodic.

(ii) Besides the destruction of phase space barriers, allowing for phase space transport, other essential features of the dynamics in chaotic tangles include greatly enhanced *stretching* and *mixing*. Our second study returns to 2D time-periodic vector fields and uses invariant manifolds as templates for a global study of stretching and mixing in chaotic tangles. The analysis here thus complements the one of *transport* via *invariant manifolds*, and can essentially be viewed as a generalization of the horseshoe map construction to apply to entire material interfaces inside the tangles. Given the dominant role of the unstable manifold in chaotic tangles, we study the stretching of a material interface originating on a segment of the unstable manifold associated with a *turnstile lobe*. We construct a symbolic dynamics formalism that describes the evolution of the *entire* material curve, which is the basis for a global understanding of the stretch processes in chaotic tangles, such as the topology of stretching, the mechanisms for good stretching, and the statistics of stretching. A central interest will be in understanding the *stretch profile* of the material interface, which is the graph of finite-time stretch experienced as a function of location on the interface. In a near-integrable setting (meaning we add a perturbation to the vector field of an originally integrable system) we argue how the *perturbed* stretch profile can be understood in terms of a corresponding *unperturbed* stretch profile approximately repeating itself on smaller and smaller scales, as described by the symbolic dynamics. The basic interest is in how the *non-uniformity* in the unperturbed stretch profile can approximately carry through to the non-uniformity in the perturbed stretch profile, and this non-uniformity can play a basic role in mixing properties and

stretch statistics. After the stretch analysis we then add to the deterministic flows a small stochastic component, corresponding for example to molecular diffusion (with small diffusion coefficient D) in a fluid flow, and study the diffusion of passive scalars across material interfaces inside the chaotic tangles. For sufficiently thin diffusion zones, the diffusion of passive scalars across interfaces can be treated as a one-dimensional process, and diffusion rates across interfaces are *directly related* to the stretch history of the interface. Our understanding of stretching thus directly translates into an understanding of mixing. However, a notable exception to the thin diffusion zone approximation occurs when an interface folds on top of itself so that neighboring diffusion zones overlap. We present an analysis which takes into account the overlap of neighboring diffusion zones, capturing a *saturation effect* in the diffusion process relevant to *efficiency of mixing*. We illustrate the stretching and mixing study in the context of two oscillating vortex pair flows, one corresponding to an open heteroclinic tangle, the other to a closed homoclinic tangle. Though we focus here on single-frequency systems, from the previous study the extensions to multiple-frequency systems should be clear.

(iii) We then study stretching from a different perspective, focusing on *rates of strain* experienced by infinitesimal line elements as they evolve under near-integrable chaotic flows associated with 2D time-periodic velocity fields. We introduce the notion of *irreversible rate of strain* responsible for net stretch, study the role of hyperbolic fixed points as engines for good irreversible straining, and observe the role of *turnstiles* as mechanisms for enhancing straining efficiency via *re-orientation* of line elements and transport of line elements to regions of superior straining.

(iv) The remaining two studies can be viewed as applications of the material developed in the previous studies, although both applications develop new theory and/or new ideas as well. The first application studies the dynamics associated with a quasiperiodically forced Morse oscillator as a classical model for molecular dissociation under external quasiperiodic electromagnetic forcing. The forcing entails destruction of phase space barriers, allowing escape from bounded to unbounded motion, and we study this transition in the context of our quasiperiodic theory, comparing with single-frequency forcing. New and interesting features of this application beyond the subject matter of the previous quasiperiodic study includes that the relevant fixed point of the unforced system is non-hyperbolic and at infinity, and the study of additional transport issues, such as *escape* (implying dissociation) from a particular level set of the unforced Hamiltonian system corresponding to a quantum state. We find for example that though

infinite-time average flux can be maximal in a single-frequency limit, escape from a level set, or equivalently lobe penetration, can be maximal in the multiple-frequency case.

(v) The second application studies statistical relaxation of distributions of finite-time Lyapunov exponents associated with interfaces evolving within the chaotic tangles of 2D time-periodic vector fields. Whereas recent studies claim or give evidence that distributions of finite-time Lyapunov exponents are essentially Gaussian, our previous analysis of stretching via the symbolic dynamics construction shows the wide variety of stretch processes and stretch scales involved in the tangle, motivating our further study of stretch statistics. In particular, we focus on the high-stretch tails of finite-time Lyapunov exponents, which have relevance in incompressible flows to the mixing properties and multifractal characteristics of passive scalars and vectors in the limit of small spatial scales. Previous studies of stretch distributions consider a *fixed* number of points, thus lacking adequate resolution to study these tails. Instead, we use a dynamic point insertion scheme to maintain adequate interfacial covering, entailing extremely good resolution at high-stretch tails. These tails show a great range in behavior, varying from essentially Gaussian to nearly exponential, and these non-Gaussian deviations can have a significant effect on interfacial stretching, one that persists *asymptotically*. These non-Gaussian deviations can be associated with very small probabilities, thus indicating the need for highly-resolved numerical studies of stretch statistics. We explain the nearly exponential tail in a particular limiting regime corresponding to highly non-uniform stretch profiles, and explore how the full statistics might be captured by elementary models for the stretch processes.

Contents

Chapter 1
Introduction

Though linear theories have had great success in describing physical reality, a vast array of nonlinear phenomena pervade our world. These phenomena are difficult to characterize and analyze, and it is only in the last few decades, with the advent of sophisticated nonlinear mathematics and powerful supercomputers, that a hope to adequately address these phenomena has dawned. The scope of nonlinear phenomena is vast, and a single unified treatment impossible. A variety of disciplines offer a mixture of approaches, and this mixture is particularly fruitful, allowing one to appreciate complicated physical processes from several vantages.

A basic and fascinating aspect of nonlinear systems is that they can evolve in an irregular or *chaotic* fashion (a precise definition of this will be offered later). Though much of the initial study of chaotic phenomena has focused on chaos *per se* - its meaning and implications, its proper characterization, the criteria for its existence, and so on - there has been growing activity concentrating on the fundamental role played by chaotic phenomena in a wide spectrum of physical processes: transport and mixing in fluid flows, ionization and dissociation in atomic and molecular systems, magnetic field amplification in kinematic dynamos, and so on.

Dynamical systems analysis is a branch of mathematics that offers a theoretical framework with which to study chaotic phenomena. Conventionally, the theory addresses *systems of ordinary differential equations (ODE's)*

$$\dot{x} = f(x), \qquad x \in \mathcal{M}^n, \tag{1.1}$$

and *maps*

$$x \longrightarrow f(x), \qquad x \in \mathcal{M}^n, \tag{1.2}$$

where $\mathcal{M}^n$ is some n-dimensional *phase space* (for example, $\mathbb{R}^n$). Given the prescribed dynamical system *vector field* $f(x)$, the goal is to understand the

evolution of x in its n-dimensional phase space, *i.e.*, to understand the *dynamics* of x. Rather than concentrating on individual solutions of (1.1) and (1.2) corresponding to particular initial conditions, dynamical systems theory studies evolution in phase space from a *global, geometrical* perspective, a basic goal being to establish *phase space structure* that offers a *template* with which to study the evolution of the system. What we mean by this should become apparent as we go along, and is best conveyed in the context of an elementary example, which we discuss here in some detail as a means for introducing in a concrete setting the language, means of description, and principal themes of our study.

Consider the two-dimensional fluid flow induced by a pair of equal and opposite point vortices.[1] In the frame comoving with the vortex pair the flow is steady, and the equations of motion for fluid particle trajectories can be written in the time-independent, or *autonomous*, form

$$\begin{aligned} \dot{x} &= \frac{\partial \psi_0(x,y)}{\partial y} \\ \dot{y} &= -\frac{\partial \psi_0(x,y)}{\partial x}, \end{aligned} \tag{1.3}$$

where (x, y) denotes the spatial coordinates of the fluid and $\psi_0(x, y)$ is the *stream function* of the steady flow (the precise form of ψ_0 is given later in the text). Equation (1.3) is an example of a *Hamiltonian system*, which defines an *area-preserving* flow (or a *volume-preserving* flow in a higher-dimensional context), a major and extremely active subset of dynamical systems. In particular, equation (1.3) defines a *one degree-of-freedom* Hamiltonian system, where the stream function plays the role of the Hamiltonian and y the conjugate momentum, and this class of Hamiltonian systems is guaranteed to be *integrable*, since an integral of the motion is provided by the Hamiltonian. The *level sets* of the Hamiltonian $\psi_0(x, y) = constant$ define *streamlines* (see Figure 1.1), which are *invariant manifolds* - any point starting on a given manifold remains on it - and

these curves thus act as a *template* for studying the fluid motion. Though one could of course perform a variety of numerical experiments to study the evolution of the fluid, the *phase portrait* offered by these invariant manifolds affords a basic understanding of the flow. Of particular note in the phase portrait are the *hyperbolic fixed points*, where the velocity field vanishes and the eigenvalues associated with the linearized flow about the fixed points have nonzero real components (these eigenvalues are of course real for our fluid example, and have opposite sign as a consequence of area preservation). These fixed points are connected by *heteroclinic manifolds* that define *separatrices*, two of which join to divide bounded from unbounded fluid motion (another important class of separatrices are formed out of *homoclinic manifolds* when the separatrix forms a closed loop originating and terminating at the same fixed point). Inside the separatrix fluid particles move along closed streamlines (1-*tori*), and outside the separatrix fluid particles translate from right to left along open streamlines. No particle can cross the separatrix, which thus acts as a *complete barrier* in phase space. Within the region of bounded motion one can transform via a *generating function* to a new set of coordinates, referred to as *action-angle coordinates*, which reduces the motion to its most elementary representation:

$$(x, y) \in \mathbb{R}^2 \longrightarrow (I, \theta) \in \mathbb{R}^+ \times T^1,$$

where T^1 denotes the 1-torus $[0, 2\pi)$, to give

$$\begin{aligned} \dot{I} &= 0 \\ \dot{\theta} &= \omega_0(I), \end{aligned} \tag{1.4}$$

with solutions

$$\begin{aligned} I(t) &= I_0 \\ \theta(t) &= \omega_0(I_0)t + \theta_0. \end{aligned} \tag{1.5}$$

On each level set, then, particles move with constant *action* I_0 and steady rotation of the *angle* $\theta(t)$. Each torus is thus specified by $(I_0, \omega_0(I_0))$, an action

and a rotation frequency, and $\omega_0(I_0) \to \infty$ as the level set defined by I_0 approaches the separatrix. The evolution of integrable systems is well-understood and pretty mundane. For example, here there is no *transport* of fluid across the separatrix, or any streamline for that matter, *stretching* is generically asymptotically linear in time, and if we choose to add molecular diffusion to the problem (see Chapter 3), the resulting *mixing* is poor.

Suppose we now add to the velocity field a *time-dependent perturbing straining field* of the form

$$\varepsilon\psi_p(x,y,t) = \varepsilon xy\sin(\omega t + \theta_0), \tag{1.6}$$

where ε is the perturbation amplitude, ω is the perturbation frequency, and θ_0 is the initial phase of the perturbation. This straining field oscillates periodically in time, and can arise, for example, if one embeds the vortex pair in a wavy-walled channel. The vortices respond to the external perturbation by oscillating periodically in time, and the net result is a time-periodic perturbation to the original velocity field, so that the equations governing particle motion now have the form

$$\begin{aligned} \dot{x} &= \frac{\partial\psi_0(x,y)}{\partial y} + \varepsilon\frac{\partial\tilde{\psi}(x,y,\omega t+\theta_0;\varepsilon)}{\partial y} \\ \dot{y} &= -\frac{\partial\psi_0(x,y)}{\partial x} - \varepsilon\frac{\partial\tilde{\psi}(x,y,\omega t+\theta_0;\varepsilon)}{\partial x}, \end{aligned} \tag{1.7}$$

where $\tilde{\psi}(x,y,\omega t+\theta_0;\varepsilon) = \tilde{\psi}(x,y,\omega t+\theta_0+2\pi;\varepsilon)$. This time-dependent or *nonautonomous* two-dimensional system can be trivially recast as a three-dimensional autonomous system (or, equivalently, a one-and-a-half degree-of-freedom Hamiltonian system)

$$\begin{aligned} \dot{x} &= \frac{\partial\psi_0(x,y)}{\partial y} + \varepsilon\frac{\partial\tilde{\psi}(x,y,\theta;\varepsilon)}{\partial y} \\ \dot{y} &= -\frac{\partial\psi_0(x,y)}{\partial x} - \varepsilon\frac{\partial\tilde{\psi}(x,y,\theta;\varepsilon)}{\partial x} \\ \dot{\theta} &= \omega, \end{aligned} \tag{1.8}$$

(where $\theta \in T^1$), thus taking on the form of equation (1.1) (the reason for such a reformulation will become apparent later if it seems opaque now). Another reformulation is to recast the ODE system (1.7) or (1.8) as a *two-dimensional map* P_ϵ (referred to as a *Poincaré map*) by strobing the fluid at the perturbation frequency, *i.e.*, by sampling the particle trajectories periodically in time:

$$\begin{aligned} P_\epsilon &: \Sigma \longrightarrow \Sigma \\ (x(t), y(t)) &\longrightarrow (x(t + \frac{2\pi}{\omega}), y(t + \frac{2\pi}{\omega})), \end{aligned} \tag{1.9}$$

where Σ denotes the two-dimensional *Poincaré section*, representing the fluid at any one of the sample times, on which the Poincaré map acts. Equation (1.9) has the form of equation (1.2), and it is thus by periodic sampling that one reduces the analysis of the ODE system to one of a 2D map, which considerably simplifies the analysis, as we shall see. This simplification is unique to time-periodic vector fields, and it is precisely this reduction scheme that has made periodically forced systems exceedingly popular, subject to vast amounts of literature, and one of the canonical classes of dynamical systems out of which many of the initial paradigms and constructs of the field have been built.

This seemingly innocuous addition to the velocity field of a small time-periodic perturbation entails drastic changes in the fluid motion, as portrayed in the Poincaré section of Figure 1.2. In particular, the resulting one-and-a-half degree-of-freedom Hamiltonian system is not integrable. In some regions of phase space regular motion gives way to *irregular* or *chaotic* motion, as indicated by the "fuzzy" regions of the Poincaré section. These chaotic zones come in two classes.

(*i*) *Resonance bands.* When the forcing frequency ω is *commensurate* with (rationally related to) a closed level set rotation frequency $\omega_0(I_0)$ of the *unperturbed* flow, the level set breaks up under the perturbation into a resonance band.

(*ii*) *Heteroclinic and homoclinic tangles.* For any forcing frequency ω the heteroclinic and homoclinic manifolds will break up into a heteroclinic and homoclinic tangle, respectively (in this example there are only heteroclinic tangles).

Figure 1.2 shows a substantial heteroclinic tangle associated with separatrix break-up, one that has in fact "eaten up" a significant amount of nearby resonance bands. At the resolution of the portrayal in Figure 1.2, resonance bands outside the heteroclinic tangle are too thin to decipher, although a more highly resolved study indeed shows their presence. The heteroclinic tangle is the overwhelmingly dominant chaotic region, and though both classes of chaotic zones are studied by similar techniques, we will be focusing our attention primarily on homoclinic/heteroclinic tangles. The "fuzzy" portrayal of these chaotic zones is generated by the orbit under successive iterates of the Poincaré map of one or more initial conditions inside the chaotic zone. In the chaotic domains of resonance bands a single orbit will, instead of being confined to a curve, map to successive locations in a rather erratic fashion, filling up an available *area.* Orbits in the heteroclinic tangle exhibit similar behavior, except for the fact they will asymptote to $x = -\infty$ since this particular heteroclinic tangle is unbounded; hence, one initializes more than one orbit to obtain a sufficiently dense portrayal of the "fuzz" (see the caption for further details on this point).

These chaotic zones correspond to radically different fluid motion from the integrable flow, accompanied by greatly enhanced transport, stretching, and mixing. For example, the destruction of the unperturbed separatrix entails that open flow fluid can be *entrained into* the bounded region, and closed flow fluid can be *detrained from* the bounded region. The entrained fluid is stretched and folded in a violent manner until it is eventually detrained, and the result is a *stirring* process or (when we later add molecular diffusion) a *mixing process* in

the heteroclinic tangle. This stirring process is inherently related to the chaotic advection of the fluid, as we shall discuss in great length through the course of the text. At the outset we make two observations about the appearance of chaotic regions under the presence of the perturbing straining field.

(*i*) The appearance of chaos is *immediate* for an *arbitrarily small* perturbing straining field. No *route to chaos* consisting of progressively complicated motions as one increases some perturbation parameter is needed here, although this latter scenario seems to have been much more widely disseminated among the general scientific community.

(*ii*) The recognition that extremely simple, non-turbulent fluid velocity fields can evolve exceedingly complicated, indeed chaotic, fluid motion, a phenomenon commonly referred to as *chaotic advection*, has stirred much excitement and activity in the fluid mechanics community since the mid-eighties. Though such a phenomenon might at first glance seem paradoxical to one accustomed to the study of complicated (*i.e.*, turbulent) velocity fields, it is important to recognize that the whole point of dynamical systems analysis since its inception has been that simple vector fields can evolve complicated dynamics.

Between the chaotic zones lie regions of regular motion, as indicated by the persistence of invariant tori in the Poincaré section of Figure 1.2. Though trajectories may wander freely throughout individual chaotic zones, they may not pass through one of these invariant curves, which thus act as *complete barriers* between distinct regions of irregular motion. The entire phase portrait is thus filled with interspersed regions of regular and irregular motion, and there is a well-developed theory describing the intertwining of these regions. Most notably, the response of tori to external forcing is described by KAM (Kolmogorov-Arnold-Moser) theory (see Berry[2] for a nice introduction).

Loosely speaking, for the ratio between the perturbation frequency ω and the unperturbed level set frequency $\omega_0(I_0)$ "sufficiently irrational" (a precise notion in the theory), the torus persists under sufficiently small forcing; otherwise the torus breaks up into a resonance band. The role of the frequency ratio arises in the attempt to express the perturbed system in integrable form by recasting in action-angle coordinates. Small divisors in the generating function cause difficulties in the transformation, and the vanishing of this divisor is directly related to the above frequency ratio. There has also been considerable numerical study portraying these intertwining regular and irregular regions, offering rich and intricate pictures (for example, there is a rich set of *islands* of regular motion nested in the resonance bands). These numerical portrayals of chaos are extremely popular in the literature even to this day. A major limitation of such "fuzzy" portrayals of the chaotic zones, however, is that they basically offer for chaotic zones just a *big mess*. One gains little understanding of the dynamics *inside* regions of chaos, *e.g.*, of the stirring or mixing process in the oscillating vortex pair flow. In a similar manner, some of the more celebrated early work in dynamical systems has concentrated on issues such as criteria for existence (or rather *non*-existence) of chaos, rather than a detailed understanding of the motion within the chaotic regions. The principal interest of this study will be in the dynamics within the chaotic regions, in particular, in the idea that underlying the chaotic regions is a *skeletal backbone* of *invariant manifolds* which act as a *template* with which to study the dynamics, as shown in Figure 1.3. These manifolds are the global stable and unstable manifolds of hyperbolic fixed points that exist in the perturbed setting, as will be explained in more detail in the rest of the text. These invariant manifolds criss-cross each other *ad infinitum* to form the boundary of a *chaotic tangle* and to form *lobes* in phase space (see Figures 1.4 and 1.5). Lobes map from one to another under

successive iterates of the Poincaré map, and it is within this context of *lobe dynamics* within the invariant chaotic tangle that one has a template for studying fundamental aspects of the dynamics within the tangle, such as *transport, stretching,* and *mixing* (see Figures 1.5 and 1.6). Exploiting this idea of lobe dynamics in a variety of contexts will be a central theme of this study. Both resonance bands and homoclinic/heteroclinic tangles can in principal be studied in this context, though the tangles in resonance bands can be geometrically more complex. Our attention will focus on homoclinic/heteroclinic tangles. We also note that the oscillating vortex pair example considers a heteroclinic tangle in a *near-integrable* setting, meaning we add a small perturbation to the vector field of an initially integrable system. The subject matter of the study will be described primarily in this setting, since it simplifies the description of the dynamics and allows for additional results, such as *Melnikov theory* (to be discussed later). We stress, however, the idea of a template formed out of invariant manifolds does not necessitate small perturbations or even a perturbative setting.

Though we have introduced the setting for this study in the context of a fluid flow, we wish to stress the variety of physical applications in the above setting. For example, in a later chapter we will perform a classical study of dissociation of diatomic molecules under external electromagnetic forcing in the context of a forced Morse oscillator, described by the dynamical system equations:

$$\begin{aligned} \dot{x} &= p/m \\ \dot{p} &= -2D_0 a(e^{-ax} - e^{-2ax}) + \varepsilon E d\cos(\omega t + \theta_0), \end{aligned} \tag{1.10}$$

$(x, p) \in \mathbb{R}^2$. One can think of $x \equiv r - r_e$ as the separation, r, of a two-atom molecule minus an equilibrium separation, r_e, with p the relative momentum. The system then corresponds to a non-rotating pair of atoms interacting under a Morse potential and forced by an external electromagnetic field with ampli-

tude εE. The parameters a and D_0 correspond to the range parameter and unperturbed dissociation energy, respectively, of the Morse potential, and d is the effective charge, or dipole gradient. Figure 1.7 shows the phase portrait associated with the unperturbed ($\varepsilon = 0$), or integrable, system. The noteworthy feature of this phase portrait is the *separatrix* which divides bounded molecular motion from unbounded molecular motion (there is a slight subtlety associated with the fact that the fixed point associated with this separatrix is non-hyperbolic and at infinity, but this is dealt with in a later chapter). The external forcing again entails destruction of integrability, in particular destruction of the separatrix, allowing for *escape* from bounded to unbounded motion, corresponding to *molecular dissociation.* The template offered by the global stable and unstable manifolds in the perturbed setting will be the key to studying this dissociation process, as shall be described in detail later.

We stress not only the variety of physical applications, but the fact that basic nonlinear physical processes, such as stirring, mixing, dissociation, magnetic field amplification, *etc.*, are *inherently and fundamentally related to the underlying chaotic dynamics.* We also emphasize that, though the dynamical systems community has given great attention to the study of *asymptotics*, *i.e.*, the behavior in the infinite-time limit, many of the above physical processes are indeed *finite-time* processes (*e.g.*, for a mixing experiment one may be interested in 20 - 40 perturbation periods). Hence, asymptotics and long-time behavior may have little relevance to the problem, and much of our study will be explicitly concerned with finite-time or transient phenomena, relatively virgin territory from a dynamical systems perspective.

The existence of these invariant stable and unstable manifolds in chaotic tangles dates all the way back to Poincaré[3] at the turn of the century; however, using them as a template for systematic study of chaotic dynamics and

to study basic physical processes is a relatively recent pursuit, active primarily since the mid-eighties. Indeed it is rather unfortunate to note the *Dark Age* in nonlinear dynamics spanning several decades following the nonperturbative vision of Poincaré, as a vast majority of physicists and mathematicians focused their attention on more pressing phenomena describable in a linear or perturbative framework. This *Dark Age*, however, is not so dark as some have the impression, for though *Chaos* burst on the scene with great hoopla in the seventies, becoming quite popular by the eighties, many of the rigorous foundations for nonlinear chaotic dynamics, such as the modern foundations of Hamiltonian dynamics, had been carefully laid prior without great fanfare by some impressive mathematical schools. It is also unfortunate to note the extent to which present-day physicists and engineers who study chaotic phenomena ignore the role of invariant manifolds. We hope to stress in this study, then, a role of a dynamical systems theorist as a *painter of phase space structure via invariant manifolds,* and the fundamental role these manifolds play in the study of dynamics.

Our interest in nonlinear dynamics and chaos will lie somewhere in the middle of the two extremes of a wide range of approaches. At the one extreme, there is a substantial dynamical systems community primarily concerned with rigorous mathematical foundation, leaving physical applications to others (fair enough). At the other extreme, since dynamical systems and chaos is a popular subject readily accessible to anyone with a personal computer, there is a community of some physicists and engineers who prefer to study chaotic phenomena unencumbered by any rigorous mathematical foundation. The study offered here will to large extent attempt to respect, and indeed sometimes extend, rigorous mathematical foundation; however, emphasis will also be placed on physical applications, and the fundamental role chaos plays in a variety of

physical processes. Indeed, there is a uniquely fruitful partnership between dynamical systems analysis and chaotic physical phenomena, as we shall see. The global, geometrical framework of dynamical systems theory offers new ways with which to conceptualize and quantify nonlinear physical processes. Conversely, the effort to understand physical processes motivates new features of dynamical systems analysis. There is also a fruitful partnership between analytical and numerical study of chaotic dynamics. A role of a dynamical systems theorist being to establish invariant structure in phase space, computers are essential for visualizing and conceptualizing these structures, and using them to uncover practical physical information. The geometrical constructs offered by dynamical systems theory typically offer a setting with which to most effectively perform numerical computations. Conventional mathematics disciplines have been hesitant about the role of the computer in basic mathematics, but given a growing awareness of the limitations of analysis in nonlinear systems, there has been growing acceptance of the computer as a basic mathematical tool in conjunction with analytical work. The present study attempts to exploit this partnership between analytical and numerical work.

We are now in position to introduce in fairly concrete manner the subject matter of this study, offering first the principal themes, followed by a chapter-by-chapter description of the subject matter. We use invariant manifolds as a *template* for studying the chaotic dynamics within classes of homoclinic and heteroclinic tangles. Our principal interest lies in:

(i) extending these templates and their applications to fundamentally larger classes of dynamical systems;

(ii) expanding the description of dynamics offered by these templates; and

(iii) strengthening the interplay between dynamical systems concepts and the description of basic nonlinear physical phenomena.

We focus primarily on three basic features of the dynamics in chaotic tangles, which we introduce in the context of prior work, followed by motivations for the present study.

(i) *Transport.* The destruction of integrability implies the destruction of *phase space barriers*, which allows *transport* between various regions of phase space, in particular *between regions corresponding to qualitatively different types of motion*, such as bounded and unbounded motion. For example, in the oscillating vortex pair flow the perturbing straining field destroys the separatrices that divide the open and closed flow, allowing for fluid to be *transported* into and out of (*i.e., entrained into* and *detrained from*) the core region, entailing a *stirring* process or (when we later add molecular diffusion) a *mixing process.* Similarly, subjecting the molecular system modeled by the Morse oscillator to external forcing destroys the separatrix that divides bounded and unbounded molecular motion, allowing for ensembles of molecules to be transported from inside to outside the bounded region, entailing a *dissociation process.* Phase space transport plays a fundamental role in these physical processes. A *phase space transport theory* has been developed since the mid-eighties for dynamical systems defined by 2D time-periodic vector fields, or equivalently 2D maps. One can divide this theory into three steps, as will be described in great detail in later chapters.

(a) Use segments of invariant manifolds to *partition* phase space into regions of qualitatively different motion (*e.g.*, bounded and unbounded motion). These partitions in the non-integrable case will not be *complete barriers*, but rather *partial barriers*, allowing for *transport* across the partitioning.

(b) Identify the mechanism for transport across the partial barriers. One finds that a special set of lobes in the tangles, referred to as *turnstile lobes*, are *solely* responsible for this transport.[4,5]

(c) Having identified the partitions and turnstiles, perform a systematic study of various features of transport across the partitions. There are all sorts of features one might be interested in investigating. An immediate example is the *flux* from one region to another, determined by the turnstile lobe area. This area is computable from a numerical determination of lobe boundaries in phase space, or, more elegantly, from analytical expressions obtained by *Melnikov theory* (for an introductory account of this theory, see Guckenheimer and Holmes[6]). Loosely speaking, this theory provides an $O(\varepsilon)$ expression of the separation of the chaotic tangle's stable and unstable manifolds (where ε is the perturbation amplitude), and is thus a basic analytical tool for studying several features of the tangles, including lobe area and flux. A more detailed study of transport considers also the *species content* of turnstile lobes. For example, suppose in the oscillating vortical flow we wish to determine the amount of fluid initially in the core that has been detrained after a given time interval. Similarly, suppose in the Morse oscillator we wish to determine the percentage of initially bounded molecules that have been dissociated after a given time interval. Turnstile lobe area will not suffice to address these questions, since at a given time a portion of the turnstile lobe associated with detrainment or dissociation may contain points that were not *initially* in the bounded region, but started outside and were entrained or captured later. Conceptually, to study such a transport problem one could initialize phase space with a distribution of particles of different

species according to the region from which they originate. The goal then is to determine the species content of the turnstile lobes at any given time. An approximate solution for lobe content can be obtained using a Markov method approach;[4,7,8] however this approximation necessitates a rather artificial assumption and can give poor results. A framework for exact solution of turnstile lobe content is offered by addressing the geometry of how lobes in the chaotic tangle intersect with one another,[1,9] or more simply how the lobes intersect with the partitioned regions.

A formal construction of transport theory for a non-integrable system is a major success, and the basic question that remains is then: how robust is this theory? In particular:

(a) Can one construct a transport theory under more complicated classes of vector fields, for example, with higher dimensions and more complicated time dependences? The theory has been extended to certain classes of higher-dimensional Hamiltonian systems,[10] and a primary aim of this study will be to extend transport theory to more complicated vector field time dependences.

(b) Given a class of vector fields, can one extend present transport theory to address other aspects of transport? For example, the above-mentioned features of transport are concerned with ensembles of points distributed uniformly *throughout* phase space, but one might be interested in a transport theory corresponding to other ensemble distributions, such as one originating on a level set of the unforced Hamiltonian, as is relevant to the problem of dissociation from a particular quantum state.

(ii) *Stretching.* Chaotic systems exhibit good (*i.e.*, exponential) stretching on a nontrivial subset of phase space, entailing sensitive dependence on initial conditions and *deterministic randomness*: given an *arbitrarily small* but finite uncertainty in the initial conditions, one will *not* be able to predict the behavior of the deterministic system after a *finite* time interval. This observation has profound impact on the understanding of the evolution of deterministic systems. A basic dynamical system measure of stretching is offered by the *Lyapunov exponent* associated with an infinitesimal line element initially at $x(0) \in \mathcal{M}^n$ with orientation $m(0) \equiv \delta x(0)/||\delta x(0)||$,

$$\lambda(x(0), m(0)) \equiv \lim_{\substack{t \longrightarrow \infty \\ ||\delta x(0)|| \longrightarrow 0}} \frac{1}{t} \ln\Big(\frac{||\delta x(t)||}{||\delta x(0)||}\Big), \tag{1.11}$$

and a positive Lyapunov exponent entails exponential stretching. Loosely speaking, then, chaotic dynamics is indicated by the presence of positive Lyapunov exponents on some *nontrivial* subset of phase space. We note that such a subset can even be of measure zero, as is the case for the Cantor set constructed from the Smale horseshoe map (discussed momentarily). One possible study of stretching under a flow is of course a numerical computation of Lyapunov exponents as a function of initial condition, but numerics can sometimes be deceptive and rather unilluminating in this instance. For sufficiently simple vector fields one has certain paradigms and constructs with which to obtain a profound and detailed understanding of stretching, sensitive dependence on initial conditions, and chaos. For example, 2D time-periodic vector fields have as a paradigm the *Smale horseshoe map*, described later in the text. The dynamics generated by this map acting on an appropriately constructed *Cantor set* is *topologically conjugate* to a *shift map acting on a bi-infinite sequence of two-symbols* (we defer the description of this to later chapters). In this context, one can recast the evolution on the Cantor set in a *symbolic dynamics formalism*,

which portrays the chaotic dynamics in its most basic and elementary form, indeed affords a precise definition of chaos, and renders proofs for various theorems involving the dynamics quite plain. The existence of the Smale horseshoe map has been established for the chaotic tangles generated by 2D time-periodic vector fields or 2D maps (see, *e.g.*, ref. 6 for a discussion). Though the Smale horseshoe map paradigm is extremely insightful, it is also extremely limited in scope since:

(a) it applies to a limited class of vector fields (2D time-periodic or 2D maps), and

(b) it describes the dynamics on only a rather artificial subset of phase space that consists of a Cantor set, *i.e.*, a set of measure zero.

Given an appreciation of the existence of chaotic dynamics in a limited class of homoclinic/heteroclinic tangles, one would like to obtain a much more robust understanding of the stretch processes inside chaotic tangles. In particular one would like to know:

(a) What, if any, paradigms apply to more complicated classes of vector fields? How is one to understand *chaos* in more general circumstances?

(b) Is there a way to obtain a *global appreciation* of stretching *throughout* the chaotic tangle, rather than on a Cantor set of measure zero? Such an appreciation might include an understanding of the global topology of stretching, the mechanisms for good stretching, the statistics of stretching, and so on, of which the horseshoe map gives either no or partial means for appreciation.

(*iii*) *Mixing.* This is a vast subject, and we have here a specific interest. Equations (1.1) and (1.2) consider deterministic evolution under a prescribed vector field. It is interesting to add to the vector fields in (1.1) and (1.2) a

stochastic term in order to study the effects of such stochastic phenomena as molecular diffusion or heat transfer. We refer to the stochastic evolution of passive scalars or vectors simply as *mixing*. A common approach is to explicitly add the stochastic term to vector field and perform a numerical study:

$$\dot{x} = f(x,t) + \varepsilon\eta(t), \tag{1.12}$$

where $\eta(t)$ is some random function of time with a Gaussian probability distribution. A principal goal is to study the interplay between stochastic effects (mixing properties) and characteristics of the underlying deterministic flows (*e.g.*, stretching properties). If possible, one would like to go beyond numerical studies of (1.12), and pursue analytical studies exploring this interplay, even if in a limited context.

We now proceed to outline the chapter-by-chapter contents of this study. As remarked earlier, 2D time-periodic vector fields, or equivalently 2D maps, define one of the canonical classes of dynamical systems, out of which many of the constructs and paradigms of the field have been built. Chapter 2 studies the chaotic dynamics evolved by 2D vector fields with more complicated time dependences than periodic. In particular, we consider the dynamics evolved by 2D *quasiperiodic* or *multiple-frequency* vector fields (with great emphasis on *two*-frequency systems). The motivation for such an extension is three-fold.

(i) From a mathematical perspective, any extension past the time-periodic case requires a fundamental generalization, since one *cannot* in general reduce the analysis to iterates of a 2D map. Hence all the dynamical system constructs and paradigms that apply to 2D maps, such as the Smale horseshoe map, KAM tori, hyperbolic fixed points and their 1D stable and unstable manifolds which define the boundaries of chaotic tangles, classical Melnikov theory, and so on, *are no longer applicable.*

(ii) From a physical perspective, one would obviously like to employ dynamical systems theory to study vector fields with more complicated time dependences. For example, though there has been considerable recent literature on chaotically advecting fluids under time-periodic velocity fields, the fluid mechanics community is primarily interested in turbulent velocity fields. Hence, there is strong motivation to take any steps one can in bridging the wide gap between these disparate classes of velocity fields, and studying multiple-frequency velocity fields is a small step in this direction. For essentially any other physical system, the motivation to relax the constraint of time-periodicity in the vector field is obvious. For example, when in a later chapter we study molecular dissociation under external electromagnetic forcing in the context of a forced Morse oscillator, being able to address multiple-frequency perturbations allows one to study forcing by several lasers and compare with forcing by one.

(iii) Underlying both the mathematical and physical interests is a basic question regarding the interplay between the dynamical system vector field and the resulting dynamics. We know that complicated vector fields can evolve complicated dynamics, and also that simple vector fields can. As one increases the complexity of the *vector field* (in this instance temporal complexity), how does this affect the complexity of the *dynamics*?

Consider for sake of introduction a dynamical system defined by the two-frequency vector field:

$$\begin{aligned} \dot{x} &= \frac{\partial H(x,y)}{\partial y} + \varepsilon g_1(x,y,\omega_1 t + \theta_{1_0}, \omega_2 t + \theta_{2_0}; \varepsilon) \\ \dot{y} &= -\frac{\partial H(x,y)}{\partial x} + \varepsilon g_2(x,y,\omega_1 t + \theta_{1_0}, \omega_2 t + \theta_{2_0}; \varepsilon), \end{aligned} \tag{1.13}$$

where g_i, $i = 1, 2$, is 2π-periodic in $\omega_j t + \theta_{j_0}$, $j = 1, 2$, and may or may not be Hamiltonian. For any sort of discrete-time sampling, *e.g.*, sampling at the second forcing frequency, the system reduces to a *sequence of 2D nonautonomous*

maps:

$$T_\varepsilon(\,\cdot\,;n) : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$
$$(x,y) \longrightarrow \phi\Big(\frac{2\pi}{\omega_2}(n+1), \frac{2\pi}{\omega_2}n, x, y\Big), \tag{1.14}$$

where $\phi(t,t_0,x_0,y_0) \in \mathbb{R}^2$ denotes the solution to (1.13) satisfying the initial condition $\phi(t_0,t_0,x_0,y_0) = (x_0,y_0)$, and $T_\varepsilon(\,\cdot\,;n)$ is the 2D map which evolves the system from time $t = 2\pi n/\omega_2$ to time $t = 2\pi(n+1)/\omega_2$, $n \in Z$.

The goal for this class of vector fields is to establish *phase space structure* that acts as a *template* for studying the evolution of the system. In direct analogy with the single-frequency case, we search for phase space structure by recasting the system in a higher-dimensional autonomous form

$$\begin{aligned}
\dot{x} &= \frac{\partial H(x,y)}{\partial y} + \varepsilon g_1(x,y,\theta_1,\theta_2;\varepsilon) \\
\dot{y} &= -\frac{\partial H(x,y)}{\partial x} + \varepsilon g_2(x,y,\theta_1,\theta_2;\varepsilon) \\
\dot{\theta}_1 &= \omega_1 \\
\dot{\theta}_2 &= \omega_2,
\end{aligned} \tag{1.15}$$

where $(\theta_1,\theta_2) \in T^2$, the two-torus. The autonomous system phase space is *four-dimensional*, and periodic sampling of phase space trajectories at the second frequency ω_2 defines a *three-dimensional* Poincaré section. We will establish phase space structure and invariant chaotic tangles in this *three-dimensional* setting, and then relate the constructs back to our original two-dimensional physical system (such as our two-dimensional fluid): the sequence of nonautonomous 2D maps in equation (1.14) is understood in terms of the 3D Poincaré map acting on a *sequence of two-dimensional slices of the 3D Poincaré section.* Hence one derives from invariant 3D tangles ($(n+1)$-dimensional tangles for an n-frequency problem) a sequence of time-dependent 2D tangles, and it is within this framework that one has a template for studying chaotic dynamics under multiple-frequency vector fields, and can systematically generalize the

canonical dynamical systems constructs associated with the single-frequency case. One finds the single-frequency constructs to be essentially robust under generalization. For example, the Smale horseshoe map generalizes to what we call a *traveling horseshoe map sequence*, hyperbolic fixed points generalize to points which lie on *invariant normally hyperbolic tori*, invariant 1D stable and unstable manifolds and the tangle boundaries they define generalize to *sequences* of 1D stable and unstable manifolds and *sequences* of tangle boundaries. The template for dynamics in the multiple-frequency case is described in terms of two-dimensional lobes mapping withing a *sequence* of 2D lobe structures. Given this new template, we then discuss the dynamical systems tools available with which to study multiple-frequency systems. A *generalized Melnikov theory*[11] offers in the multiple-frequency case an analytical $0(\varepsilon)$ measure of stable and unstable manifold separation in the tangles that is the basis for a variety of analytical studies, and a *double phase slice sampling method* allows for numerical computation of *precise 2D slices* of the invariant higher-dimensional chaotic tangles, and is thus the basis for all numerical work. We define from the Melnikov function a *relative scaling function* which gives an analytical $O(\varepsilon)$ measure of the relative importance of each frequency on manifold separation. With the template and tools in hand, we study multiple-frequency dynamics and compare with single-frequency dynamics. We show how lobe dynamics under a bi-infinite sequence of maps can be specified in closed form by exploiting underlying periodicity properties of the vector field, and offer several illustrations. We then present numerical simulations of sequences of chaotic tangles and lobe dynamics within these sequences, and make some qualitative observations and comparisons with single-frequency forcing. In contrast to lobes of equal area mapping within a *fixed* 2D lobe structure found under conventional single-frequency forcing, we find lobes of varying areas mapping within

a sequence of lobe structures that are distorting and breathing from one time sample to the next, affording greater freedom in the nature of the dynamics. Our primary focus in this new setting is on *transport* (we consider *stretching* and *mixing* in other contexts in later chapters). The formal construction of a transport theory is more involved than in the single-frequency case, as a consequence of the more complicated invariant manifold geometry in the higher-dimensional setting. This geometry is uncovered and explored, however, via theorems and numerical studies based on Melnikov theory. We then partition phase space and define turnstiles in the autonomous setting, and from this obtain the sequence of partitions and turnstiles in the non-autonomous setting. A main new feature of transport is its manifestation in the context of a sequence of *time-dependent* regions, and we argue this is consistent with a Lagrangian viewpoint. We then perform a detailed study of such transport properties as flux, lobe geometry, and lobe content. In contrast to the single-frequency case, where a single flux suffices, in the multiple-frequency case a variety of fluxes are allowed, such as different types of instantaneous, finite-time average, and infinite-time average flux. We find for certain classes of multiple-frequency forcing that *infinite-time* average flux is maximal in a *single-frequency limit*, but that the spatial variation of lobe areas found in multiple-frequency systems affords greater freedom to enhance or diminish *finite-time* transport quantities. We illustrate our study with a quasiperiodically oscillating vortical flow and a quasiperiodically forced Duffing oscillator. Though much discussion in this chapter is in terms of the two-frequency case, all results are formally developed in the context of the general n-frequency problem, and we explain how the analysis generalizes to vector fields with more complicated time dependences than quasiperiodic.

Whereas Chapter 2 extends templates and their applications to larger classes of systems, Chapter 3 expands the description of dynamics offered by

these templates. We in fact return to 2D time-periodic vector fields and use invariant manifolds as templates for a global study of *stretching* and *mixing* in chaotic tangles. The analysis in the chapter thus complements the one of *transport* via invariant manifolds, and can essentially be viewed as a generalization of the horseshoe map construction to apply to entire material interfaces inside the tangles. Given the dominant role of the unstable manifold in chaotic tangles, we study the stretching of a material interface originating on a segment of the unstable manifold associated with a turnstile lobe. Since the dynamics within chaotic tangles is describable in terms of lobes mapping from one to another within an invariant lobe structure, studying the stretching of a *single* lobe boundary provides the basis for what is, though not an exhaustive study of stretching in chaotic tangles, certainly a much more global study than afforded by the horseshoe map paradigm (for example, we are considering a subset of phase space that is a one-dimensional curve rather than a set of measure zero). We construct a symbolic dynamics formalism that describes the evolution of the *entire* material curve, which is the basis for a global understanding of the stretch processes in chaotic tangles, such as the topology of stretching, the mechanisms for good stretching, and, as we shall address in a later chapter, the statistics of stretching. Stable and unstable manifold geometry plays a fundamental role in this symbolic dynamics: for example, the intersection points between the stable and unstable manifolds are shown to have particular relevance to the stretch processes. A central interest will be in understanding the *stretch profile* of the material interface, which is the graph of finite-time stretch experienced as a function of location on the interface. We argue how the *perturbed* stretch profile can be understood in terms of a corresponding *unperturbed* stretch profile approximately repeating itself on smaller and smaller scales, as described by the symbolic dynamics. The basic interest is how the *non-uniformity* in

the unperturbed stretch profile can approximately carry through to the non-uniformity in the perturbed stretch profile, and this non-uniformity will play a basic role in mixing properties and stretch statistics, as we shall see later in the chapter, and in Chapter 6, respectively. After the stretch analysis we then add to the deterministic flows a small stochastic component, corresponding for example to molecular diffusion (with small diffusion coefficient D) in a fluid flow, and study the diffusion of passive scalars across material interfaces inside the chaotic tangles. For sufficiently thin diffusion zones, the diffusion of passive scalars across interfaces can be treated as a one-dimensional process, and diffusion rates across interfaces are *directly related* to the stretch history of the interface. Our understanding of interfacial stretching thus directly translates into an understanding of mixing. However, a notable exception to the thin diffusion zone approximation occurs when an interface folds on top of itself so that neighboring diffusion zones overlap. We present an analysis which takes into account the overlap of neighboring diffusion zones, capturing a *saturation effect* in the diffusion process relevant to *efficiency of mixing*. We illustrate the results in the context of two oscillating vortex pair flows, one corresponding to an open heteroclinic tangle, the other to a closed homoclinic tangle. Though this chapter focuses on single-frequency systems, from the previous chapter the extensions to multiple-frequency systems should be clear.

Chapter 4 studies stretching from a different perspective than Chapter 3, focusing on *rates of strain* experienced by individual infinitesimal line elements as they evolve under the chaotic flow, a pursuit more closely related to conventional fluid studies. We again consider 2D time-periodic vector fields (indeed the same two oscillating vortical flows as in Chapter 3) in a near-integrable setting. We introduce the notion of *irreversible rate of strain* responsible for net stretch, study the role of hyperbolic fixed points as engines for good irreversible

straining, and observe the role of *turnstiles* as mechanisms for enhancing straining efficiency via *re-orientation* of line elements and *transport* of line elements to regions of superior straining. Turnstiles have been shown to play a fundamental role in *phase space transport*, and we argue here they play a fundamental role in *phase space stretching.*

Chapters 5 and 6 can be viewed as applications of the material developed in the previous chapters, although they both develop new theory and/or ideas as well. Chapter 5 studies the dynamics associated with a quasiperiodically forced Morse oscillator as a classical model for molecular dissociation under external quasiperiodic electromagnetic forcing. The forcing entails destruction of phase space barriers, allowing escape from bounded to unbounded motion, and we study this transition via the theory developed in Chapter 2, comparing with single-frequency forcing. We stress here how this transport theory allows for a *rigorous and precise construction* of a transition theory for non-integrable classical systems between bounded and unbounded motion. New and interesting features of this study beyond the subject matter of Chapter 2 include:

(i) The relevant fixed point of the unforced system is *non-hyperbolic* and at infinity, rendering the conventional theorems associated with hyperbolic fixed points inadequate. The fixed point is of saddle-type stability, however, and we show how the same tangle construction goes through. For example, conventional Melnikov theory is shown to apply, and stable and unstable manifold separation determined by Melnikov theory and by explicit numerical calculation are in extremely good agreement up to sizable perturbations.

(ii) Besides the transport issues of Chapter 2, we additionally study transport in the context of *escape from a particular level set of the unforced Hamiltonian system (corresponding to a quantum state).* Such a study requires

different considerations of lobe geometry, in particular lobe *penetration* into the action or energy coordinate, and is an example of the effort to extend the scope of transport theory and the description of dynamics offered by the template of invariant manifolds. We find that even though infinite-time average *flux* can be maximal in a single-frequency limit, escape rates from a level set can be maximal in the multiple-frequency case.

(iii) We address the approximate treatment of transport via Markov models, including its extension to the multiple-frequency case, and a brief comparison with exact transport theory via lobe intersection geometry.

(iv) Unlike the vortical flows, one can solve analytically the Melnikov amplitude integrals rather than having to resort to numerical integration.

(v) We study and interpret the interesting behavior of the Morse oscillator in the limit as forcing frequency goes to zero.

Chapter 6 returns to stretching under 2D time-periodic vector fields, this time focusing on the *statistics* associated with interfacial stretching in chaotic tangles. More precisely, we study distributions of *finite-time Lyapunov exponents* $\lambda(t)$

$$\lambda(t) = \frac{1}{t}\ln\Big(\frac{ds(t)}{ds(0)}\Big), \tag{1.16}$$

where $ds(t)$ denotes the length of an infinitesimal line element along the interface. These exponents are thus finite-time versions of conventional Lyapunov exponents. Though interest in asymptotics has led to considerable study of conventional Lyapunov exponents, there has been in recent years growing interest in the finite-time analogues, since many basic physical processes depending on stretching, such as fluid mixing experiments, are very much finite-time ventures. Whereas recent studies[12,13] claim or give evidence that distributions of finite-time Lyapunov exponents are essentially Gaussian, the analysis of Chapter 3 shows the wide variety of stretch processes and stretch scales involved in

the tangles, motivating our further study of stretch statistics. In particular, we focus on *high-stretch tails* of finite-time Lyapunov exponent distributions. Previous studies of these distributions consider a *fixed* number of points, thus lacking the resolution to study these tails. Instead, we use a point insertion scheme to maintain adequate interfacial covering, entailing extremely good resolution at high-stretch tails. We thus explore the statistics with truly 1D probe, and want to stress the difference from a probe of measure zero (a collection of points lacking adequate resolution). The high-stretch tails show great range in behavior, varying from essentially Gaussian to nearly exponential, and these non-Gaussian deviations can have a significant effect on interfacial stretching, one that persists *asymptotically*. We explain the nearly exponential tail in a particular limiting regime corresponding to highly non-uniform stretch profiles (refer back to Chapter 3), and explore how the full statistics might be captured by elementary models for the stretch processes.

Chapter 7 offers a few brief concluding remarks. As a final introductory remark we should warn that the subsequent chapters are essentially a compilation of different publications and preprints. As a result, the tone and approach can vary from one chapter to the next depending on the intended journal audience, and the chapters do not follow one another as smoothly as if the study were written as one coherent work. We apologize for the somewhat disjointed presentation.

References

[1] V. Rom-Kedar, A. Leonard, and S. Wiggins, J. Fluid Mech. **214**, 347 (1990).

[2] M.V. Berry, in *AIP Conference Proceedings No. 46. Topics in nonlinear dynamics: a tribute to Sir Edward Bullard*, S. Jorna editor (AIP, New York, 1978).

[3] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Gauthier-Villars, Paris, 1899).

[4] R. S. MacKay, J.D. Meiss, and I.C. Percival, Physica D **13**, 55 (1984).

[5] D. Bensimon and L.P. Kadanoff, Physica D **13**, 82 (1984).

[6] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer-Verlag, New York, 1983).

[7] R.S. MacKay, J.D. Meiss, and I.C. Percival, Physica D **27**, 1 (1987).

[8] J.D. Meiss and E. Ott, Physica D **20**, 387 (1986).

[9] V. Rom-Kedar and S. Wiggins, Arch. Rat. Mech. Anal. **109**, 239 (1990).

[10] S. Wiggins, Physica D **44**, 471 (1990).

[11] S. Wiggins, *Global Bifurcations and Chaos - Analytical Methods* (Springer-Verlag: New York, 1988).

[12] F.J. Muzzio, P.D. Swanson, and J.M. Ottino, Phys. Fluids A **3**, 822 (1991).

[13] F. Városi, T.M. Antonsen, Jr., and E. Ott, Phys. Fluids A **3**, 1017 (1991).

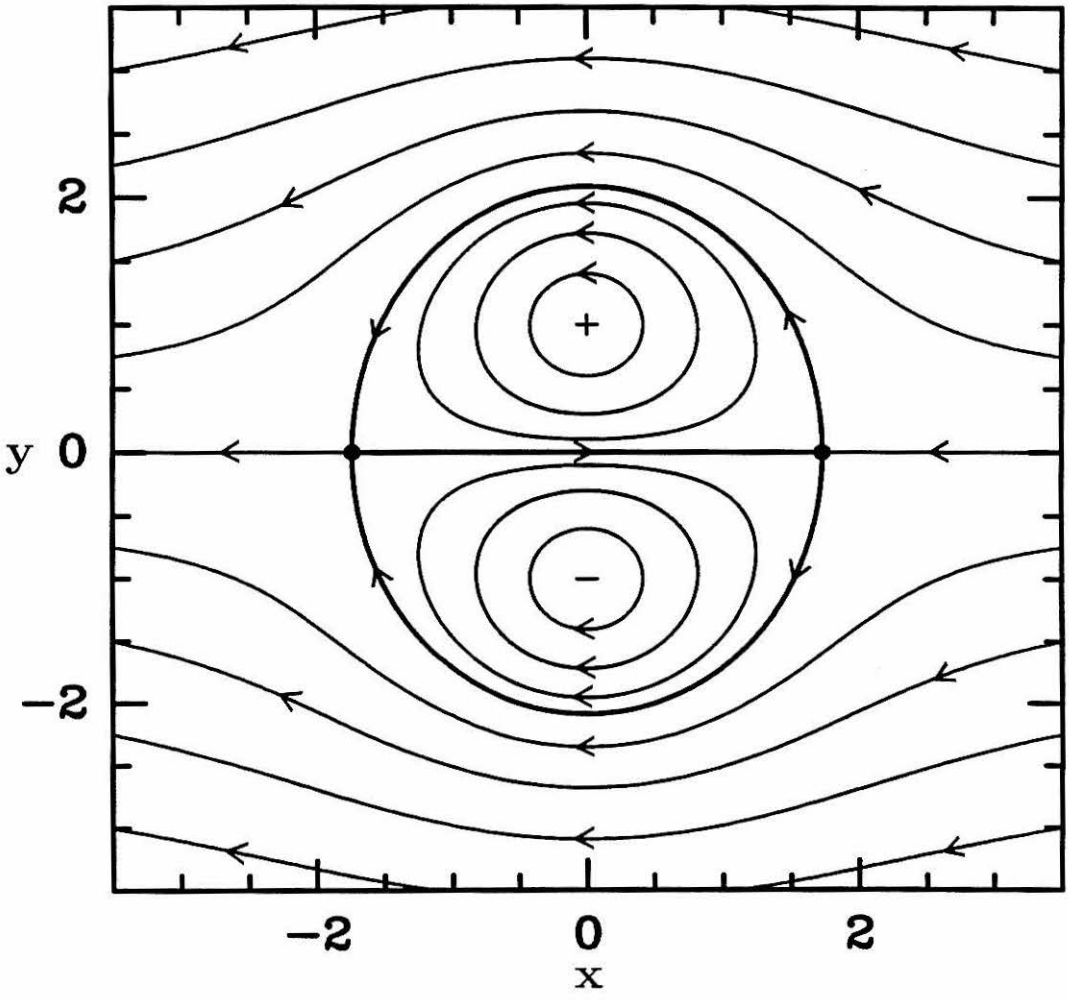


Figure 1.1 The streamlines of the unperturbed vortex pair flow in the frame comoving with the vortex pair (the vortices are situated at $(x, y) = (0, \pm 1)$).

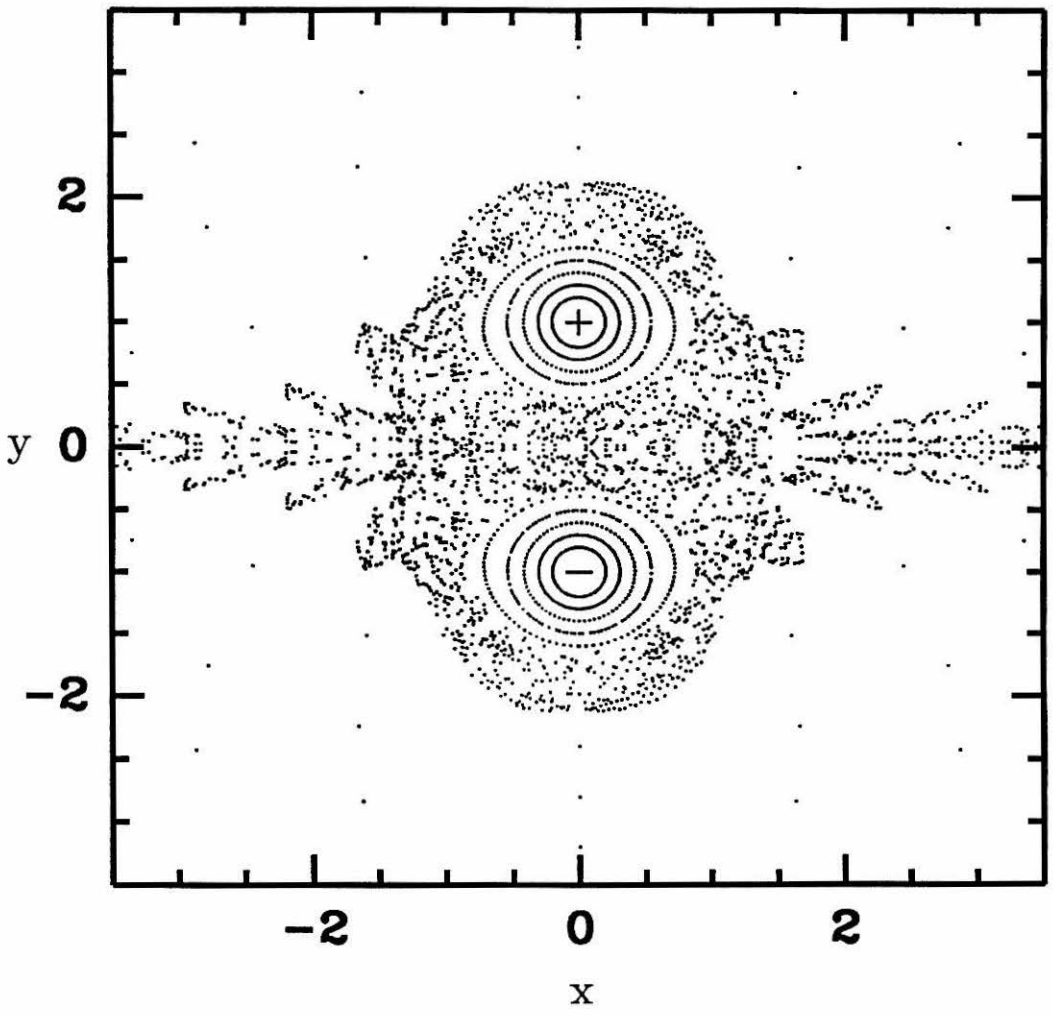


Figure 1.2 The two-dimensional Poincaré section associated with the oscillating vortex pair flow exhibits regions of regular and irregular (*i.e.*, chaotic) motion (system parameters are given later in the text). A substantial chaotic zone is associated with separatrix break-up. The "fuzzy" portrayal of this zone is the result of evolution forwards and backwards in time of a set of fifty points initialized in a particular subset of the zone. This subset is the *entraining turnstile lobe* of the chaotic tangle, whose meaning will become quite clear in later chapters.

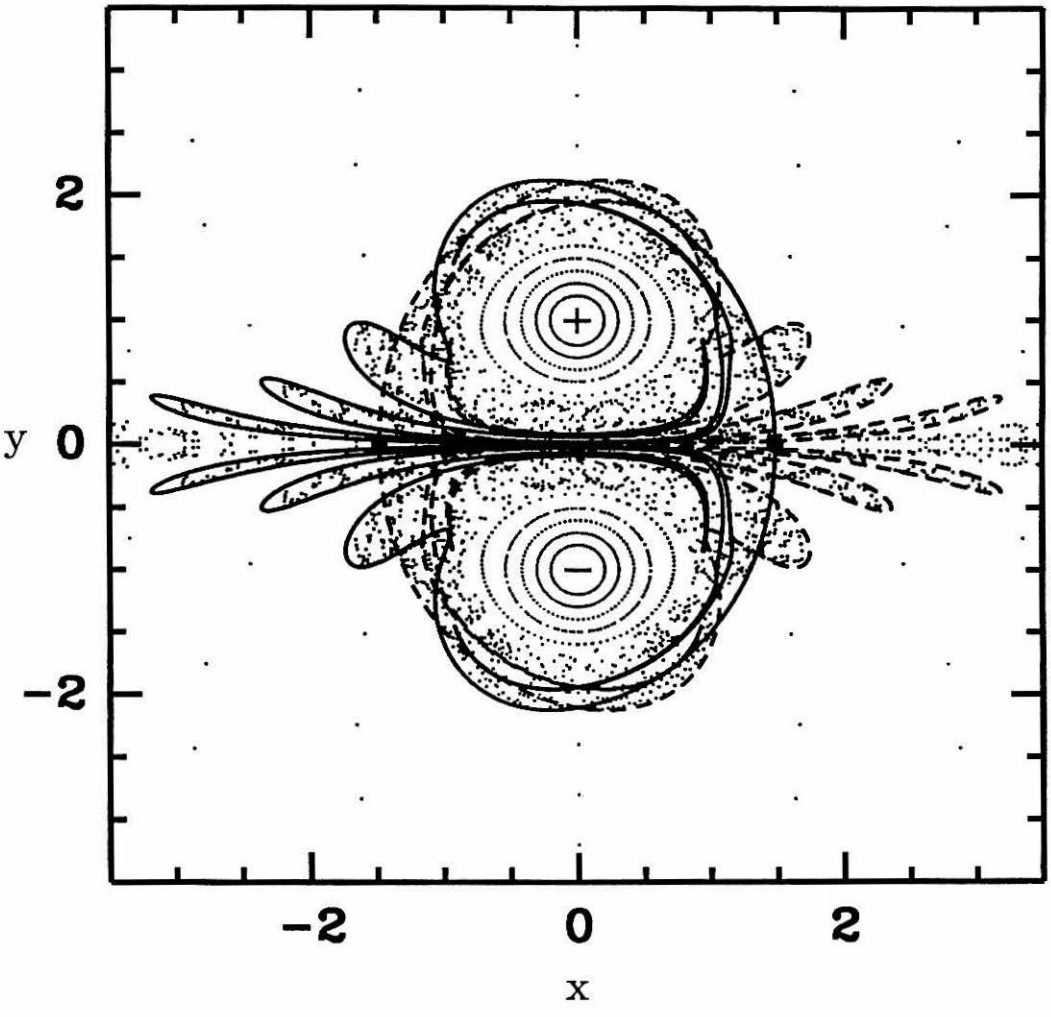


Figure 1.3 Underlying the "fuzzy" portrayal of the chaotic zones is a skeletal backbone formed out of intersecting global stable (dashed lines) and unstable (solid lines) manifolds of persisting hyperbolic fixed points. These invariant manifolds offer a *template* with which to study the chaotic motion. We are of course showing only finite segments of these manifolds.

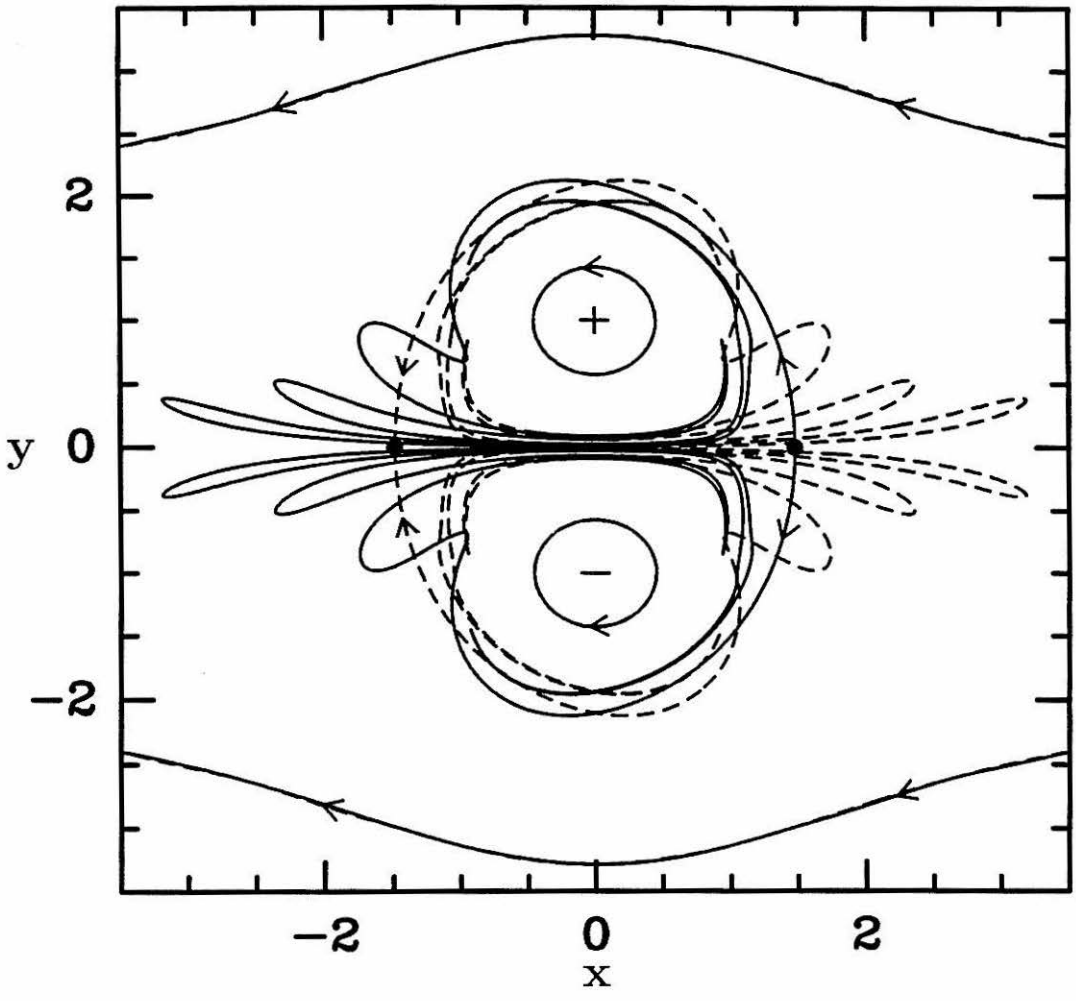


Figure 1.4 The skeletal backbone without the "fuzz". The two solid dots denote hyperbolic fixed points of the Poincaré map which persist under the perturbation, the solid lines denote *invariant global unstable manifolds* of the right fixed point, and the dashed lines denote *invariant global stable manifolds* of the left fixed point. These manifolds criss-cross each other ad-infinitum to form the boundary of an *invariant heteroclinic tangle.*

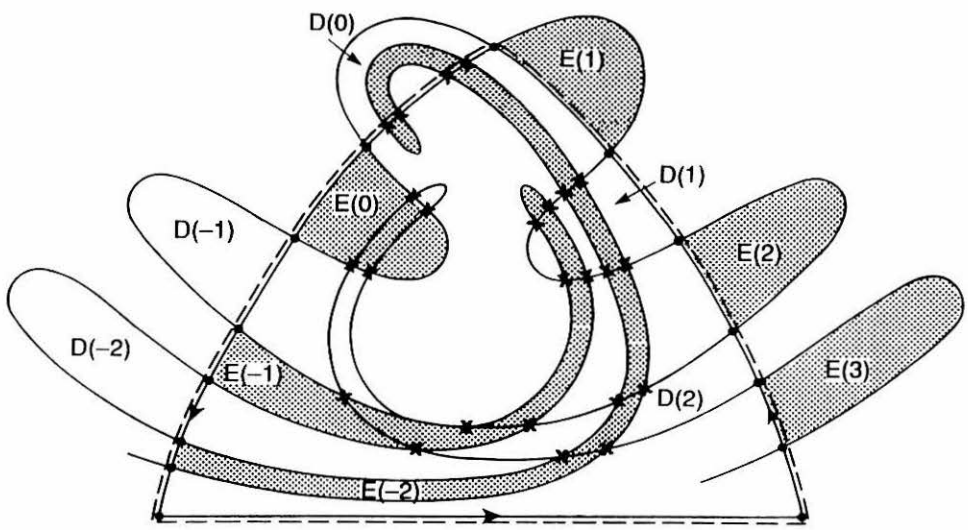


Figure 1.5 The stable and unstable manifolds criss-cross each other *ad infinitum* to form an *invariant lobe structure.* A finite number of lobes of the upper heteroclinic tangle are portrayed here. The lobes are labelled according to the dynamics they undergo, as will be explained in detail in the body of the thesis: the shaded lobes are those associated with *entrainment*, the white lobes with *detrainment*, and under the Poincaré map they evolve according to $E(i) \longrightarrow E(i-1)$ and $D(i) \longrightarrow D(i-1)$.

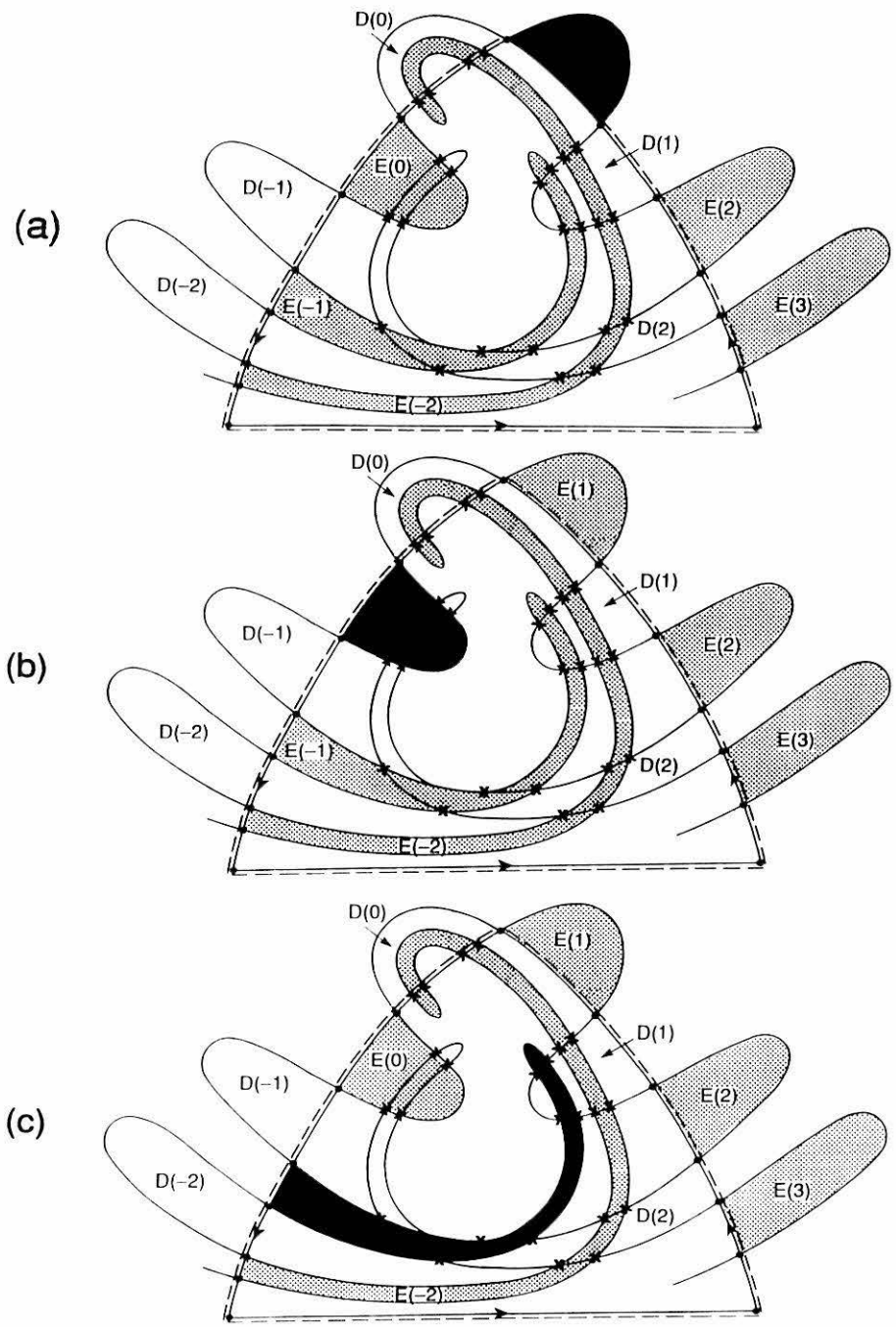


Figure 1.6 A *lobe* (shown in black) mapping within the invariant lobe structure. The lobe is portrayed at (a) $t = 0$, (b) $t = \frac{2\pi}{\omega}$, and (c) $t = 2\frac{2\pi}{\omega}$.

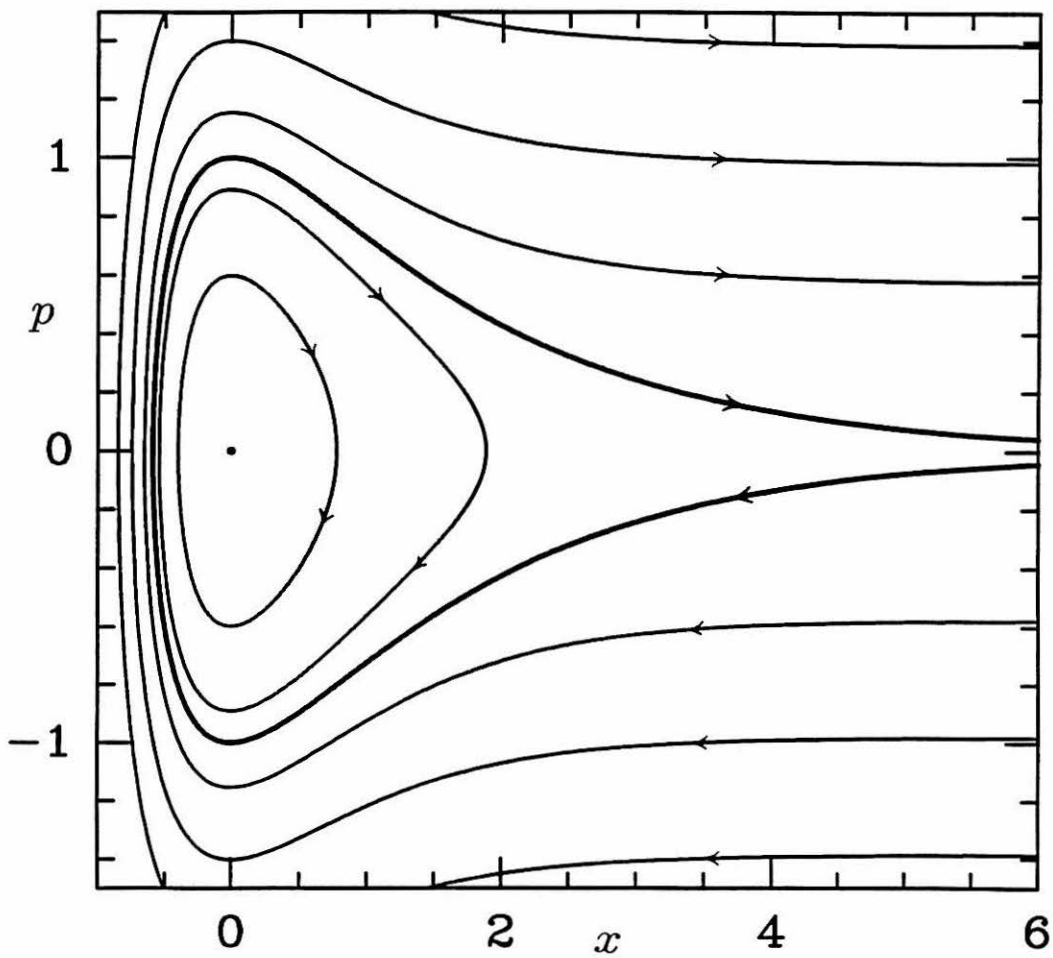


Figure 1.7 The phase portrait associated with the unforced Morse oscillator. The separatrix (boldface) divides bounded from unbounded molecular motion. The dimensionless coordinates for this portrait are explained in Chapter 5.

Chapter 2

The dynamics associated with the chaotic tangles of quasiperiodically forced two-dimensional dynamical systems: theory and applications

Based upon articles that appear in:

Nonlinearity 4, 775-819 (1991)

and

Nonlinear Phenomena in Atmospheric and Oceanic Sciences
(IMA Volumes in Mathematics and its Applications - Volume 40)
Editors G.F. Carnevale and R.T. Pierrehumbert
(Springer-Verlag: New York, Berlin, Heidelberg, 1992) 47-138.

Abstract

We generalize notions of transport in phase space associated with the classical Poincaré map reduction of a periodically forced two-dimensional system to apply to a sequence of nonautonomous maps derived from a quasiperiodically forced two-dimensional system. We obtain a global picture of the dynamics in homoclinic and heteroclinic tangle regions using a sequence of time-dependent two-dimensional lobe structures derived from the invariant global stable and unstable manifolds of one or more normally hyperbolic invariant sets in a Poincaré section of an associated autonomous system phase space. We discuss transport in the context of a few example systems, with emphasis on advection in fluids. Transport is specified in terms of two-dimensional lobes mapping from one to another within the sequence of lobe structures, and we argue that these lobe structures, and the invariant manifolds from which they are derived, are the dominant structure by which to understand chaotic transport in the tangle regions. We present computer simulations of lobes mapping from one to another within the sequence of lobe structures and discuss qualitatively the essential features of the dynamics in the tangle region: entrainment and detrainment with respect to a sequence of time-dependent cores with boundaries formed by segments of stable and unstable manifolds (for concreteness we focus mainly on the prototypical situation of a single core sequence and hence borrow the fluid mechanical terminology of entrainment and detrainment to make the discussion of transport more apparent), and repeated stretching and folding to give chaotic dynamics. These features are studied in a more exact fashion via the lobe structures. Instantaneous and average flux are quantified analytically in the near-integrable case through the use of a generalized Melnikov function, and the computational prescription for determining flux in systems that are not near-integrable is provided; the time average of the flux is proportional to an average of areas over appropriate regions of the Poincaré section of the autonomous system phase space. We explain how answering basic questions about transport in phase space (both in the near-integrable and general case) necessitates a consideration of issues such as the geometry of intersections of lobes with each other, with the core, and with level sets of

the unperturbed Hamiltonian, and the variation of lobe areas. We compare transport rates in the single-frequency and multiple-frequency cases. Though the comparisons depend on choices of normalization, for some reasonable normalizations the average flux of a class of multiple-frequency systems is found to be maximal in a certain single-frequency limit. The variation of lobe areas in multiple-frequency systems, however, gives one the freedom to enhance or diminish aspects of transport. The chaotic nature of the dynamics is understood in the framework of a *traveling horseshoe map sequence.* A search through parameter space to determine when a quasiperiodically forced system is chaotic can uncover simple, practical results: for example, a relative phase shift in the forcing can lead to suppression of chaos in a certain class of systems. We explain the extension of the analysis to more general time dependences.

2.1 Introduction

Some dynamical systems that exhibit chaotic behavior evolve according to sufficiently simple vector fields that one can obtain a fundamental and detailed global picture of the chaotic dynamics. A canonical example is a periodically forced two-dimensional system, which can be reduced by time-periodic sampling of trajectories to a two-dimensional Poincaré map (for a detailed account see Guckenheimer and Holmes [1983]). Consider the often studied case where the Poincaré map of the unforced system contains a hyperbolic fixed point connected to itself by a homoclinic orbit, or a pair of hyperbolic fixed points connected to each other by heteroclinic orbits (see Figure 2.1.1(i)). Periodic forcing drastically alters the dynamics near the unperturbed homoclinic/heteroclinic orbits (also called separatrices). For small enough perturbations the Poincaré map's fixed points survive, and, if their global stable and unstable manifolds intersect once, they will intersect infinitely many times to produce the boundary of a complicated two-dimensional lobe structure (the so-called *homoclinic/heteroclinic tangle region*) that is invariant under the Poincaré map (see Figure 2.1.1(ii)). Lobes map from one to another within the invariant lobe structure with each application of the Poincaré map, and this provides a global picture of the dynamics, *i.e.*, of *transport in phase space.* A variety of physical phenomena can be studied in this context, a most notable example in recent years being advection in fluids, where the dynamical system phase space is physical space. The idea that relatively innocuous velocity fields can produce chaotic advection has made quite a stir among the fluid mechanics community in the last few years (see for example Ottino [1990]), even though the observation that simple vector fields can evolve complicated dynamics has been at the heart of dynamical systems analysis since its inception. Another noteworthy example involves multiphoton ionization of excited atoms and dissociation of diatomic and polyatomic molecules (see Chapter 5). On the one hand, the fairly detailed understanding that one can obtain (as we shall see) of homoclinic and heteroclinic tangle regions is

quite impressive, given the complicated nature of the dynamics. This understanding prevents the need to resort to statistical theories, and provides a framework for computing exact quantities, such as transport in and out of a defined set of regions in phase space (which has relevance for quantifying a number of physical phenomena, such as roll-to-roll transport in a Rayleigh-Bénard flow (see Camassa and Wiggins [1991]) and dissociation rates of diatomic molecules in the presence of an electromagnetic field (see Chapter 5)). However, for periodically forced systems the vector field that evolves the system is quite simple, and one is studying transport under iterates of a map. One would like to extend the results to vector fields with as complicated time dependences as possible, while still retaining a fairly complete global picture of the dynamics; this will entail extending the study of maps to a study of *sequences of maps*. Our goal is to generalize some notions of transport in phase space associated with iterates of a Poincaré map derived from a time-periodic two-dimensional vector field to apply to a bi-infinite sequence of nonautonomous maps derived from a two-dimensional quasiperiodic vector field (*i.e.*, a vector field whose time dependence involves ℓ frequencies, $\ell \geq 2$). As we shall see, the concepts involved in the generalization are robust and apply to vector fields with more general time dependences. We thus wish to introduce these concepts in the context of a familiar and appealing class of systems.

Let us first introduce some of the features of the dynamics in the time-periodic case. These features are highlighted by a simulation of a typical system with a heteroclinic tangle region, an oscillating vortex pair (OVP) flow (studied by Rom-Kedar et al. [1990]) that is defined along with its quasiperiodic generalization in Appendix 2.A1. The unperturbed system is a pair of equal and opposite point vortices whose streamlines in the comoving frame are shown in Figure 2.1.2. Open flow passes above and below an inner core of rotating fluid bounded by two heteroclinic orbits. The perturbed system consists of the vortices oscillating periodically in response to a time-periodic external strain-rate field. Figure 2.1.3 shows, for seven perturbation periods, snapshots of a material curve initially on the upper

unperturbed heteroclinic orbit (the perturbed situation is symmetric about $y = 0$). With each snapshot a new lobe of fluid is entrained into the core region. The lobe stretches as it winds around the core, and then it folds across another entrained lobe, part of it escaping to be advected steadily to the left, part of it remaining in the core to repeat the stretching and folding process. This process continues ad infinitum to give chaotic advection. The material curve is being essentially "attracted" to the upper unstable manifold of the Poincaré map's right fixed point. Since the dynamical system is Hamiltonian (the fluid is incompressible), the unstable manifold is of course not an attractor; however, the strong stretching and contraction of the fluid lobes implies that, though area elements do not shrink to zero, they tend to be stretched in one direction and contracted in another such that the unstable manifold dominates the evolution of material curves in the tangle region (see Figure 2.1.4 for a heuristic portrayal). The stable and unstable manifolds of the Poincaré map's fixed points are thus understood to be the dominant structure by which the dynamics in the tangle regions should be studied. Recent comparisons of fluid mixing in experiments with numerical simulations of the associated unstable manifold show this to be the case (see Rom-Kedar et al. [1990] for a discussion of these examples).

Our simulation shows the essential features of the dynamics in the tangle region: (i) entrainment and detrainment, *i.e.*, the back and forth motion across boundaries in phase space that were barriers to transport in the unperturbed case, and (ii) repeated stretching and folding to give chaotic dynamics. These features can be examined via the dynamical system's associated Poincaré map acting on the invariant lobe structure. Computer simulation of a finite number of the lobes, coupled with a knowledge of how lobes map from one to another, provides a direct portrayal of the transport. Further, one can establish that there are regions in phase space on which a sufficiently large number of iterates of the Poincaré map acts as a horseshoe map (this is essentially a consequence of the stretching and folding of lobes as shown in the simulation). One can then use the standard horseshoe map construction

to state in a precise way what is meant by chaotic dynamics in the tangle region. Though establishing chaos, this construction provides only a limited understanding of the topology of stretching in chaotic tangles; however, a study of lobe behavior allows one to obtain a global understanding of the stretch properties in chaotic tangles (see Chapter 3). In addition, segments of stable and unstable manifolds of the Poincaré map's fixed points can be used to define the boundary of a homoclinic or heteroclinic core region (approximately equal to the unperturbed core), and the mapping of lobes from one to another with each application of the Poincaré map can be used to define and quantify notions of entrainment and detrainment with respect to this core. The issue of transport in phase space, under iterates of a map, across boundaries formed by segments of stable and unstable manifolds of invariant hyperbolic sets has been studied by various investigators (see MacKay et al. [1984] and Bensimon and Kadanoff [1984]). Rom-Kedar and Wiggins [1990] provide a more detailed study of transport under iterates of a two-dimensional map, one which considers the global geometry of the invariant lobe structures that govern the transport, and Wiggins [1990] generalizes this study to a class of k degree-of-freedom Hamiltonian systems ($2 \leq k < \infty$) with hyperbolicity in two dimensions. We should stress that much of the analysis of transport across boundaries formed by segments of stable and unstable manifolds applies to more general circumstances than to a single core of a near-integrable map. Indeed, it can apply to an arbitrary number of simply connected regions bounded by segments of stable and unstable manifolds of an arbitrary number of invariant hyperbolic sets of maps that need not be near-integrable.

Let us now address the extension of the transport features in the time-periodic case to multiple-frequency vector fields. The addition of even one more frequency to the vector field's time dependence requires a fundamental departure from the time-periodic analysis: the evolution of trajectories from one time sample to the next (for whatever choice of sampling) must now in general be described by a different map for each time sample. One can thus no longer construct an invariant

two-dimensional lobe structure upon which to study phase space transport, or invoke the standard horseshoe map construction to establish that there is chaotic motion. How then does one obtain a detailed global picture of the dynamics? To study the dynamics of a nonautonomous system that evolves according to a two-dimensional, ℓ-frequency vector field, we consider in Section 2.2 an associated $(\ell+2)$-dimensional autonomous system in order to construct an invariant $(\ell+1)$-dimensional lobe structure in a Poincaré section of the enlarged phase space (the lobe boundaries being formed out of intersecting global stable and unstable manifolds of normally hyperbolic invariant $(\ell-1)$-tori in the defined Poincaré section). We derive from this invariant lobe structure a sequence of two-dimensional time-dependent lobe structures in the nonautonomous system phase space that will be used to analyze transport under a sequence of nonautonomous maps. The invariant lobe structure will prove instrumental throughout our analysis in understanding transport relative to the sequence of two-dimensional time-dependent lobe structures, and we will see how the transport properties of the nonautonomous system are related to transport properties of the associated autonomous system. For concreteness, we will focus on quasiperiodically forced systems whose unforced solution contains in the Poincaré section of the associated autonomous system a single normally hyperbolic invariant $(\ell-1)$-torus connected to itself by a homoclinic orbit, or a pair of such tori connected by two or more heteroclinic orbits. Appendix 2.A1 discusses two such systems (the OVP flow mentioned earlier and a Duffing oscillator) that will be used throughout the chapter to illustrate our discussion. Though it will be clear as we proceed how many of the arguments extend to the case of an arbitrary number of normally hyperbolic invariant $(\ell-1)$-tori of a map that is not necessarily near-integrable, we will be able to give a more complete global picture of transport in the two prototypical situations we have introduced. For example, a generalized Melnikov function will repeatedly help us understand in the near-integrable case the nature of the invariant $(\ell+1)$-dimensional lobe structure, which has more complicated geometrical considerations than in the single-frequency

case. In particular, in Section 2.2 we study the zero sets of the Melnikov function to determine the geometry of intersections of stable and unstable manifolds, and hence the geometry of the invariant lobes. Section 2.3 describes how the lobes of one time sample of the nonautonomous system map to the lobes of the next time sample; we define a sequence of time-dependent cores and turnstile lobes, explain the procedure for determining the number of turnstile lobes for each time sample, and illustrate this procedure with some two-frequency examples. The turnstile lobes are the mechanism for entrainment and detrainment relative to the core sequence (since for simplicity we are dealing with a single core sequence, we borrow the fluid mechanical terminology of entrainment and detrainment to make the discussion of transport in and out of the core sequence, respectively, more apparent). The main extension from the previously mentioned studies of transport across boundaries is the notion of a sequence of time-dependent cores, and we discuss how such a notion is reasonable, and how the stable and unstable manifolds in the Poincaré section of the autonomous system phase space provide the natural definition of a sequence of time-dependent cores. Indeed, we discuss how for near-integrable systems whose unperturbed separatrix divides bounded and unbounded motion, it is segments of stable and unstable manifolds of the perturbed problem, and *not* the unperturbed separatrix, which form the "last frontier" between bounded and unbounded motion in the perturbed problem. We conclude Section 2.3 by presenting computer simulations of some sequences of lobe structures and of lobes mapping from one to another within these sequences, and discussing qualitatively the essential features of transport in the time-dependent tangle region. The sampling method used for the simulations, which we refer to as a *double phase slice method*, contrasts with a previous suggestion by Moon and Holmes [1985] for sampling quasiperiodically forced systems to find underlying structure, and is explained in Appendix 2.A4. The next three sections examine more closely the features discussed qualitatively in Section 2.3. Section 2.4 focuses on entrainment and detrainment relative to a sequence of two-dimensional time-dependent cores in the nonautonomous system phase space.

We calculate analytically instantaneous and average flux with respect to the core sequence in near-integrable systems, and present the computational prescription for calculating these quantities when the system is not near-integrable. We compare average flux in the single-frequency and multiple-frequency cases; though the comparisons depend on the normalization of perturbation amplitudes, we shall see how, for reasonable normalizations, the average flux for a class of multiple-frequency systems is maximal in a particular single-frequency limit. We explain how answering basic questions about transport in phase space (both in the near-integrable and general case) necessitates a consideration of more than just flux, and then study such issues as the geometry of intersections of lobes with each other, with the core, and with level sets of the unperturbed Hamiltonian. For example, we will see that even though average flux may decrease in quasiperiodic systems, other factors can outweigh this effect to make it easier for inner level sets of the unperturbed Hamiltonian to escape the core. More generally, *the variation of lobe areas gives one added freedom to enhance or diminish transport properties over finite time scales for a fixed average flux.* Section 2.5 deals with the description of chaos associated with the stretching and folding of lobes in the framework of a *traveling horseshoe map sequence.* The chaotic dynamics of the nonautonomous system is understood relative to a time-dependent Cantor set of points. Other studies, with different methods, of the chaotic nature of the dynamics under perturbations with time dependences more complicated than periodic, or, equivalently, under a sequence of nonautonomous maps, are given by Scheurle [1986], Stoffer [1988a], [1988b], and Meyer and Sell [1989]. We will see here how the notion of a traveling horseshoe map sequence and of chaotic dynamics relative to a time-dependent Cantor set of points follows quite naturally from previous work by Wiggins (see Wiggins [1987] and [1988]). The generalized Melnikov function allows one to search in parameter space of the quasiperiodic problem to determine when one can and cannot have chaos. Such a search can have simple, practical consequences: for example, we shall see how a relative phase shift in the forcing can lead to suppression of chaos in a class

of multiple-frequency vector fields. Section 2.6 returns to a transport problem discussed in Section 2.4 and treats the problem in a more general way with a formalism that focuses on lobe intersections (for simplicity we consider in Section 2.4 transport between only two regions, but the framework easily extends to an arbitrary number of regions (such as found in the previously mentioned Rayleigh-Bénard flow), and this extension is part of what is considered in Section 2.6). We conclude in Section 2.7 with an explanation of how the analysis extends to vector fields with more general time dependences than quasiperiodic. We will emphasize throughout the chapter the two-frequency case as a workable and illuminating example of transport in phase space under a bi-infinite sequence of nonrepeating maps.

2.2 Defining the chaotic tangle and studying its geometry

We consider two-dimensional nonautonomous systems of the form

$$\dot{x} = JDH(x) + \varepsilon g(x, t; \mu, \varepsilon), \tag{2.2.1}$$

where $x \in \mathbb{R}^2, H : X \mapsto \mathbb{R}^1$ is $C^{r+1} (r \geq 2)$ on some open set $X \subset \mathbb{R}^2, g : X \times \mathbb{R}^1 \times V \times \mathbb{R}^1 \mapsto \mathbb{R}^2$ is $C^r (r \geq 2)$ with parameter space $V \subset \mathbb{R}^p$ an open set, $\varepsilon \in \mathbb{R}^1$, and

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$DH(x) = \begin{pmatrix} \frac{\partial H}{\partial x_1} \\ \frac{\partial H}{\partial x_2} \end{pmatrix}.$$

Equation (2.2.1) represents a class of perturbed single degree-of-freedom Hamiltonian systems (note that the perturbation $\varepsilon g(x, t; \mu, \varepsilon)$ may or may not be Hamiltonian). We restrict our interest to perturbations periodic or quasiperiodic in time for fixed $x \in X, \mu \in V, \varepsilon \in \mathbb{R}^1$, and hence we can rewrite the governing equation as

$$\dot{x} = JDH(x) + \varepsilon g(x, \theta_1(t), \theta_2(t), \ldots, \theta_\ell(t); \mu, \varepsilon), \tag{2.2.2}$$

where $\theta_i(t) = \omega_i t + \theta_{i_0}$ for $i = 1$ to ℓ (some positive integer ≥ 1), and g is 2π periodic in each of $\theta_i(t)$. Though we focus on quasiperiodic systems ($\ell \geq 2$), it should be clear as we proceed how our $\ell \geq 2$ analysis applies to the familiar time-periodic ($\ell = 1$) case, and we will refer from time to time to the single-frequency case.

Let $x(t) = \phi(t, t_0, x_0)$ denote the solution to (2.2.2) which satisfies the initial condition $\phi(t_0, t_0, x_0) = x_0$. We shall also refer to $\phi(t, t_0, x_0)$ as the *trajectory of* (2.2.2) *through* x_0 *at time* t_0. To study the dynamics of (2.2.2), it will be advantageous to sample the trajectories at discrete time intervals, the interval being one of the perturbation periods, say $2\pi/\omega_\ell$. For time-periodic vector fields, it is well understood that periodic sampling filters out redundant dynamical information to obtain a simpler underlying structure with which to study the dynamics; for quasiperiodic vector fields, periodic sampling will in a similar way simplify the underlying structure with which the dynamics is studied. The evolution of system (2.2.2) from time $t = t_0 + \frac{2\pi}{\omega_\ell} n$ ($n \in Z$) to time $t = t_0 + \frac{2\pi}{\omega_\ell}(n+1)$ defines a map on X:

$$T_\varepsilon(\,\cdot\,; n) \,:\, X \mapsto X$$

$$x \mapsto \phi(\frac{2\pi}{\omega_\ell}(n+1), \frac{2\pi}{\omega_\ell} n, x),$$

where for simplicity we have set $t_0 = 0$, which is not restrictive since the θ_{i_0} are arbitrary. Our goal then is to study the dynamics in X under the bi-infinite sequence of nonautonomous maps $\{T_\varepsilon(\,\cdot\,; n); n \in Z\}$. To do this, we employ a standard technique for the study of nonautonomous systems (see Guckenheimer and Holmes [1983] or Wiggins [1988]), that is to examine an associated autonomous system in order to construct invariant manifolds that are fixed in the enlarged phase space. These manifolds form barriers which constrain the motion in the autonomous system phase space and hence provide a global picture of the dynamics. Having obtained these manifolds, we will use them to define for each time $t = \frac{2\pi}{\omega_\ell} n$ a set of curves in X that will be used in later sections to study the dynamics in X under the bi-infinite sequence of maps $\{T_\varepsilon(\,\cdot\,; n); n \in Z\}$.

2.2.1 Constructing an invariant tangle and studying its geometry

We rewrite equation (2.2.2) in the autonomous form

$$\begin{aligned}
\dot{x} &= JDH(x) + \varepsilon g(x, \theta_1, \theta_2, \ldots, \theta_\ell; \mu, \varepsilon) \\
\dot{\theta}_1 &= \omega_1 \\
\dot{\theta}_2 &= \omega_2 \\
&\vdots \\
\dot{\theta}_\ell &= \omega_\ell.
\end{aligned} \tag{2.2.3}$$

The autonomous system phase space is $\mathbb{R}^2 \times T^\ell$ (where T^ℓ is the ℓ-torus). The expression

$$(x(t), \theta_1(t), \theta_2(t), \ldots, \theta_\ell(t)) = (\phi(t, t_0 = 0, x_0), \omega_1 t + \theta_{1_0}, \omega_2 t + \theta_{2_0}, \ldots, \omega_\ell t + \theta_{\ell_0}) \tag{2.2.4}$$

solves equation (2.2.3). Let us denote this solution by $\Phi(t, t_0 = 0, x_0, \theta_{1_0}, \theta_{2_0}, \ldots, \theta_{\ell_0})$. We shall also refer to this as the *trajectory of* (2.2.3) *through* $(x_0, \theta_{1_0}, \theta_{2_0}, \ldots, \theta_{\ell_0})$ *at time* $t = 0$. From equation (2.2.4), the trajectory of (2.2.2) through x_0 at time $t = 0$ is the x component of the trajectory of (2.2.3) through $(x_0, \theta_{1_0}, \theta_{2_0}, \ldots, \theta_{\ell_0})$ at time $t = 0$, and hence the study of trajectories of the nonautonomous system is related to the study of the appropriate trajectories of the autonomous system.

We study cases in which the unperturbed system has in X:

(a) a hyperbolic fixed point x_h which is connected to itself by a homoclinic trajectory $x_h(t)$ (*i.e.*, $\lim_{t \to \pm\infty} x_h(t) = x_h$), or

(b) a pair of hyperbolic fixed points x_h^a, x_h^b connected to each other by heteroclinic trajectories $x_h^i(t)$ (i labels different trajectories); for each i either $\{\lim_{t \to -\infty} x_h^i(t) = x_h^a$ and $\lim_{t \to \infty} x_h^i(t) = x_h^b\}$ or $\{\lim_{t \to -\infty} x_h^i(t) = x_h^b$ and $\lim_{t \to \infty} x_h^i(t) = x_h^a\}$.

From these basic structures, the results will generalize to an arbitrary number of hyperbolic fixed points, each connected to itself and/or other fixed points by homoclinic/heteroclinic trajectories. In the unperturbed homoclinic case, the autonomous system phase space thus contains a normally hyperbolic ℓ-torus of the form

$$T_0 = \{(x, \theta_1, \theta_2, \ldots, \theta_\ell) | x = x_h\},$$

whose $(\ell + 1)$-dimensional stable and unstable manifolds, denoted $W^s(T_0)$ and $W^u(T_0)$, coincide along the $(\ell + 1)$-dimensional homoclinic manifold given by

$$W^s(T_0) \cap W^u(T_0) = \{(x, \theta_1, \theta_2, \ldots, \theta_\ell) | x = x_h(t),\ t \in \mathbf{R}\}.$$

Note that "normal hyperbolicity" refers to the fact that expansion and contraction rates normal to T_0 dominate those tangent to it, as to be expected since the θ_i motion is only linear in time (see Wiggins [1988] for further discussion of normal hyperbolicity). The unperturbed heteroclinic case has two normally hyperbolic ℓ-tori of the same form as the homoclinic case, with the $(\ell + 1)$-dimensional stable manifold of one torus and the $(\ell + 1)$-dimensional unstable manifold of the other torus coinciding along an $(\ell + 1)$-dimensional heteroclinic manifold

$$W^s\left(T_0^a\right) \cap W^u\left(T_0^b\right) = \{(x, \theta_1, \theta_2, \ldots, \theta_\ell) | x = x_h^i(t),\ t \in \mathbf{R}\}.$$

Figure 2.2.1 shows the manifolds in $X \times T^\ell$ (in the forthcoming diagrams, we will not keep presenting both the homoclinic and heteroclinic case, but rather show one or the other interchangeably). The parameter s specifies the location in X along the invariant unperturbed homoclinic or heteroclinic manifold: it represents the time it takes for the point $x(-s)$ on the unperturbed manifold in X to move according to the flow (2.2.3) to the point $x(s = 0)$ (which we define to be the spatial midpoint of the unperturbed manifold).

We define a global cross section of the autonomous system phase space

$$\Sigma^{\theta_{\ell_0}} = \{(x, \theta_1, \theta_2, \ldots, \theta_\ell) | \theta_\ell = \theta_{\ell_0}\},$$

and the associated Poincaré map generated by the flow of (2.2.3) is given by

$$P_\epsilon : \Sigma^{\theta_{\ell_0}} \mapsto \Sigma^{\theta_{\ell_0}}$$

$$(x, \theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) \mapsto (\phi(\frac{2\pi}{\omega_\ell}(n+1), \frac{2\pi}{\omega_\ell}n, x), \theta_0 + 2\pi\frac{\omega}{\omega_\ell}(n+1)), \tag{2.2.5}$$

where $\theta_0 \equiv (\theta_{1_0}, \theta_{2_0}, \ldots, \theta_{(\ell-1)_0})$ and $\omega \equiv (\omega_1, \omega_2, \ldots, \omega_{\ell-1})$. Note that the map is independent of n since it is derived from an autonomous system. Studying the flow of (2.2.3) via this Poincaré map is equivalent to sampling the trajectories of (2.2.3) at time intervals equal to $2\pi/\omega_\ell$. The Poincaré map of the unperturbed homoclinic system has a normally hyperbolic invariant $(\ell-1)$-torus of the form

$$\tau_0 = T_0 \cap \Sigma^{\theta_{\ell_0}} = \{(x, \theta_1, \theta_2, \ldots, \theta_\ell) | x = x_h, \theta_\ell = \theta_{\ell_0}\},$$

with an ℓ-dimensional homoclinic manifold

$$\begin{aligned} W^s(\tau_0) \cap W^u(\tau_0) &= \{W^s(T_0) \cap W^u(T_0)\} \cap \Sigma^{\theta_{\ell_0}} \\ &= \{(x, \theta_1, \theta_2, \ldots, \theta_\ell) | x = x_h(t), \theta_\ell = \theta_{\ell_0}, t \in \mathbb{R}\}. \end{aligned}$$

The heteroclinic case is understood similarly. See Figure 2.2.2 for a portrayal of homoclinic manifolds in the full phase space and in the Poincaré section $\Sigma^{\theta_{2_0}}$ for the two-frequency case (which will be studied in particular examples in later sections). Note how, in the second portrayal of $W^s(\tau_0) \cap W^u(\tau_0)$ in Figure 2.2.2(b), the θ_2 specifier is dropped. In the diagrams and notation of the geometrical constructs that clearly lie in the Poincaré section $\Sigma^{\theta_{\ell_0}}$, we will leave out the cumbersome θ_{ℓ_0} specifier; one should remember, however, that these constructs are in $X \times T^\ell$ and correspond to $\theta_\ell = \theta_{\ell_0}$.

To establish for the perturbed system the existence of a lobe structure in $\Sigma^{\theta_{\ell_0}}$, we invoke the following persistence result for small perturbations.

Proposition 2.2.1. If $P_{\varepsilon=0}$ contains a normally hyperbolic invariant $(\ell-1)$-torus τ_0, then for ε sufficiently small, P_ε possesses a C^r normally hyperbolic invariant $(\ell-1)$-torus, τ_ε, whose local ℓ-dimensional C^r stable and unstable manifolds, $W^s_{loc}(\tau_\varepsilon)$ and $W^u_{loc}(\tau_\varepsilon)$, are C^r ε-close to $W^s_{loc}(\tau_0)$ and $W^u_{loc}(\tau_0)$, respectively.

Proof: This is an immediate consequence of the persistence theory for normally hyperbolic invariant manifolds. See Fenichel [1971], Hirsch et al. [1977], or Wiggins [1988].

In the usual way we can define global stable and unstable manifolds, $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$, from the local ones:

$$W^s(\tau_\varepsilon) = \bigcup_{n \leq 0} P_\varepsilon^n \left(W^s_{loc}(\tau_\varepsilon)\right)$$

$$W^u(\tau_\varepsilon) = \bigcup_{n \geq 0} P_\varepsilon^n \left(W^u_{loc}(\tau_\varepsilon)\right),$$

and these manifolds will soon be seen to provide an invariant lobe structure.

At this point we introduce a generalized Melnikov function (Wiggins [1988]), which provides a criterion for existence of the lobe structure and a tool for studying the geometry of the lobe structure in the near-integrable case:

$$M(s,\theta_1,\theta_2,\ldots,\theta_\ell;\nu) = \int_{-\infty}^{\infty} DH(x_h(t)) \cdot g(x_h(t),\omega_1 t + (\omega_1 s + \theta_1), \omega_2 t + (\omega_2 s + \theta_2),\ldots,\omega_\ell t + (\omega_\ell s + \theta_\ell);\mu,\varepsilon = 0)dt, \tag{2.2.6}$$

where $(s,\theta_1,\theta_2,\ldots,\theta_\ell)$ specify a point on $W^s(T_0^a) \cap W^u(T_0^b)$ (and are thus constant in the integral expression), and $\nu \equiv (\mu,\omega_1,\omega_2,\ldots,\omega_\ell)$. Note that we leave out the i superscript on $x_h(t)$; in the heteroclinic case, it is understood that $x_h(t)$ represents one orbit out of a set of possible heteroclinic orbits. Further, whenever we write T_0^a and T_0^b, or τ_0^a and τ_0^b, it is understood that in the homoclinic case $a = b$.

Wiggins [1988] gives a full discussion of this function; further, it is explained there how redundancy due to uniqueness of solutions allows one to eliminate one of the parameters when studying global bifurcations, and the $s = 0$ case is thus examined. Here we choose to let the s parameter vary, but set $\theta_\ell = \theta_{\ell_0}$. In the Poincaré section $\Sigma^{\theta_{\ell_0}}$, the first order in ε approximation to the signed distance between the global stable and unstable manifolds is

$$d\left(s, \theta_1, \theta_2, \ldots, \theta_{(\ell-1)}; \nu, \varepsilon\right) = \frac{\varepsilon M(s, \theta_1, \theta_2, \ldots, \theta_{(\ell-1)}, \theta_{\ell_0}; \nu)}{\| DH(x_h(-s)) \|} + 0(\varepsilon^2) \qquad (2.2.7)$$

(see Wiggins [1988]). More specifically, d is the signed distance, measured along the normal to the unperturbed homoclinic or heteroclinic orbit, between the "first" intersection of $W^u(\tau_\varepsilon^b)$ with the normal and the "first" intersection of $W^s(\tau_\varepsilon^a)$ with the normal (see Figure 2.2.3). "First" is understood relative to moving along $W^u(\tau_\varepsilon^b)$ normally away from τ_ε^b and along $W^s(\tau_\varepsilon^a)$ normally away from τ_ε^a. Note that $\| DH(x_h(-s)) \| \to 0$ as $s \to \pm\infty$, so the distance function blows up with $|s|$, making the first order in ε approximation in (2.2.7) a poor one for sufficiently large $|s|$. The approximation will be valid however over a limited domain: the manifolds $W^s(\tau_\varepsilon^a)$, $W^u(\tau_\varepsilon^b)$ can be approximated uniformly on semi-infinite time intervals (see Guckenheimer and Holmes [1983] and Wiggins [1988]), and on these time intervals the manifolds move only an $O(\varepsilon)$ amount from the unperturbed manifolds. Hence for sufficiently small perturbations the manifold separations will be well approximated by the $0(\varepsilon)$ term over a finite domain. The approximation will prove quite useful in later sections, even though it is valid over a limited domain.

For illustration purposes we consider systems with perturbations of the form

$$g(x, t; \mu, \varepsilon) = F_0(x; \mu) + \mathcal{F}(x; \mu) \sum_{i=1}^{\ell} F_i(\mu, \omega_i) \sin(\omega_i t + \theta_{i_0}) + O(\varepsilon), \qquad (2.2.8)$$

where F_i, $1 \leq i \leq \ell$, is scalar. It will be important to keep in mind that the θ_{i_0} specify the initial phases of the perturbation. The generalized Melnikov function

associated with these perturbations has the form

$$\begin{aligned} M(s,\theta_1,\ldots,\theta_{\ell-1},\theta_{\ell_0};\nu) =& A_0(\mu) + \sum_{i=1}^{\ell-1} A_i(\mu,\omega_i)\sin(\omega_i s + \theta_i) \\ &+ A_\ell(\mu,\omega_\ell)\sin(\omega_\ell s + \theta_{\ell_0}) + \sum_{i=1}^{\ell-1} B_i(\mu,\omega_i)\cos(\omega_i s + \theta_i) \\ &+ B_\ell(\mu,\omega_\ell)\cos(\omega_\ell s + \theta_{\ell_0}). \end{aligned} \tag{2.2.9}$$

In particular, Appendix 2.A1 gives the equations of motion for the two-frequency OVP flow mentioned in the introduction and a two-frequency Duffing oscillator, and derives their generalized Melnikov functions. In both these examples, as is often the case, the unperturbed homoclinic/heteroclinic orbits have a symmetry, and hence only some of the amplitudes in equation (2.2.9) are non-zero. (Due to the symmetry in the unperturbed Duffing oscillator, the perturbation is written in terms of cosines, which of course satisfies (2.2.8) with the proper phase shift.) We shall refer to the ratio of the Melnikov function amplitude, $A_i(\mu,\omega_i)$ or $B_i(\mu,\omega_i)$, to the corresponding relative perturbation amplitude, $F_i(\mu,\omega_i)$, as a *relative scaling factor*:

$$\frac{A_i(\mu,\omega_i)}{F_i(\mu,\omega_i)} \text{ or } \frac{B_i(\mu,\omega_i)}{F_i(\mu,\omega_i)} \equiv \textit{relative scaling factor associated with frequency } \omega_i.$$

As seen in the examples in Appendix 2.A1, the relative scaling factors are in general *frequency dependent*, which will be relevant to the study of how transport rates depend on single-frequency and multiple-frequency forcing. As illustrated in Appendix 2.A1, all scaling factors are determined by two *relative scaling functions* (one for the A's, one for the B's) which are derived from the generalized Melnikov function (in fact all the B's vanish for the examples in Appendix 2.A1, so there is only a single scaling function for these examples). We should point out that if the perturbation amplitudes in (2.2.8) each had a different spatial dependence, the corresponding Melnikov function would still be expressed in the form (2.2.9), since the spatial dependences are integrated out. We choose for concreteness, however, to

refer to the class of perturbations (2.2.8) for illustration (whose attractive features include being able to define relative scaling factors).

To define and study an invariant lobe structure necessitates a consideration of the geometry of intersection manifolds of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$, and the following theorem allows one to use the Melnikov function for this consideration.

Theorem 2.2.2. Suppose there exists a point $(\tilde{s}, \tilde{\theta}_1, \ldots, \tilde{\theta}_{(\ell-1)})$ for some $\tilde{\nu}$ such that

1) $M(\tilde{s}, \tilde{\theta}_1, \ldots, \tilde{\theta}_{(\ell-1)}, \theta_{\ell_0}; \tilde{\nu}) = 0$,

2) $D_{(s,\theta_1,\ldots,\theta_{(\ell-1)})} M(\tilde{s}, \tilde{\theta}_1, \ldots, \tilde{\theta}_{(\ell-1)}, \theta_{\ell_0}; \tilde{\nu})$ is of rank one.

Then for ε sufficiently small $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ intersect transversely near (within $O(\varepsilon)$) this point, and this intersection can be continued to an $(\ell-1)$-dimensional manifold in $\Sigma^{\theta_{\ell_0}}$.

Proof: This is a straightforward consequence of the implicit function theorem (see Wiggins [1988]).

Hence within $O(\varepsilon)$ of the zero sets of the Melnikov function which satisfy the above derivative criterion are intersection manifolds of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$. Note that, though the $M = 0$ sets can imply nearby intersection sets, not all intersection manifolds need be near the $M = 0$ sets, as we will describe later. Further note that we are making no assumptions on the geometry of the $(\ell-1)$-dimensional intersection manifolds; in particular, unlike Wiggins [1988] and Ide and Wiggins [1989], we do not assume them to be $(\ell-1)$-tori. We approximately portray (*i.e.*, to $0(\varepsilon)$) some intersection manifolds by plotting the zero sets of a two-frequency Melnikov function of the form

$$M(s, \theta_1, \theta_{2_0} = 0; \nu) = A_0(\mu) + A_1(\mu, \omega_1)\sin(\omega_1 s + \theta_1) + A_2(\mu, \omega_2)\sin(\omega_2 s). \quad (2.2.10)$$

(The systems described in Appendix 2.A1 have Melnikov functions of this form.) To do so, we first flatten out the (s, θ_1) coordinates onto a plane, as shown in Figure 2.2.4. Figures 2.2.5 (a) to (g) show the zero sets near $s = 0$ for a range of A_1/A_2 values and $A_0 = 0$, $\omega_2/\omega_1 = g^{-1}$, where g is the golden mean $((\sqrt{5}-1)/2)$. For

$A_1/A_2 < 1$ the intersection manifolds are non-intersecting 1-tori, for $A_1/A_2 > 1$ they are non-intersecting spirals, and for $A_1/A_2 = 1$ they are intersecting spirals (or equivalently intersecting 1-tori). For other ratios of ω_2/ω_1 (both commensurate and incommensurate), the geometry of the intersection manifolds as A_1/A_2 is varied (keeping $A_0 = 0$) is qualitatively similar, as will become more apparent with some additional figures in Section 2.3. There is a slightly technical point about the non-generic case of $A_1/A_2 = 1$: that the zero sets cross does not necessarily imply that the intersection manifolds cross; however, we have shown that for each crossing of the $M = 0$ sets the intersection manifolds cross within $0(\varepsilon)$, and for the systems defined in Appendix 2.A1 each such crossing occurs for the same value $A_1/A_2 = 1 + 0(\varepsilon)$ (see Appendix 2.A2). Figure 2.2.5(h) shows a case with $A_0 = \pm 1.5$, $A_1 = A_2 = \pm 1$, and $\omega_2/\omega_1 = 2$: the intersection manifolds exist in a *subset* of $[0, 2\pi)$. As will be proven in an upcoming theorem, the geometry of intersection manifolds for $A_0 \neq 0$ may be qualitatively different for commensurate and incommensurate frequencies, since it is in the commensurate frequency case alone that for certain parameter values the intersection manifolds exist on a subset of $[0, 2\pi)$. We remark parenthetically that the study of zero sets of generalized Melnikov functions is proving to be useful to a number of problems involving transport within chaotic tangle regions. For example, in k degree-of-freedom systems ($2 \leq k < \infty$) with hyperbolicity in two dimensions (see Wiggins [1990]), the definition of the so-called turnstile lobes is a somewhat delicate issue, and necessitates a consideration of the geometry of the zero sets of the Melnikov function.

With these simple plots as illustration, we define *primary intersection manifolds*, study their geometry, and then define *lobes*. The definitions will be sufficiently general to allow arbitrary geometries of intersection manifolds found in systems that satisfy the conditions of Proposition 2.2.1 and Theorem 2.2.2. Suppose that the point $(\tilde{s}, \tilde{\theta}_1, \ldots, \tilde{\theta}_{\ell-1})$ of Theorem 2.2.2 satisfies $\frac{\partial M}{\partial s}(\tilde{s}, \tilde{\theta}_1, \ldots, \tilde{\theta}_{\ell-1}, \theta_{\ell_0}; \tilde{\nu}) \neq 0$. Then by the implicit function theorem there exists a C^r function $s(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$

on some domain

$$\mathcal{Z} \equiv (\alpha_1, \beta_1) \times \cdots \times (\alpha_{\ell-1}, \beta_{\ell-1}) \in T^{\ell-1},$$

where $\tilde{\theta}_i \in (\alpha_i, \beta_i) \subset [0, 2\pi)$, $i = 1, 2, \ldots, \ell - 1$, such that

$$M(s(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0}), \theta_1, \ldots, \theta_{\ell-1}, \theta_{\ell_0}; \tilde{\nu}) = 0$$

(for simplicity of future discussion the dependence of the above function s on the parameters $\tilde{\nu}$ is not explicitly included in the notation). We denote the closure of $\mathcal{Z}$ by

$$\bar{\mathcal{Z}} \equiv [\alpha_1, \beta_1] \times \cdots \times [\alpha_{\ell-1}, \beta_{\ell-1}] \in T^{\ell-1},$$

and if $\beta_i = 2\pi$ for some $i = 1, \ldots, \ell - 1$ then we define $[\alpha_i, \beta_i] = [\alpha_i, 2\pi)$.

Definition 2.2.1 Let $s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ with domain $\mathcal{Z}^i$, $i = 1, \ldots, p$, be functions as defined above with $\mathcal{Z}^i \cap \mathcal{Z}^j = \emptyset$ and $\bar{\mathcal{Z}}^i \cap \bar{\mathcal{Z}}^j \equiv \mathcal{Z}^{ij}$ $\forall i, j$ $i \neq j$. Then from Theorem 2.2.2 the set

$$\{(s, \theta_1, \ldots, \theta_{\ell-1}) \,|\, s = s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0}), (\theta_1, \ldots, \theta_{\ell-1}) \in \bar{\mathcal{Z}}^i, i = 1, \ldots, p\}$$

parametrizes an $(\ell-1)$-dimensional surface contained in $W^s(\tau_\varepsilon^a) \cap W^u(\tau_\varepsilon^b)$. In order for this surface to be a single-valued graph over $\bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$ we further specify

$$s = s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0}) \text{ on } \mathcal{Z}^{ij} \qquad i \neq j$$

(provided $\mathcal{Z}^{ij} \neq \emptyset$). We refer to this surface as a *primary intersection manifold (PIM)*, which we denote by τ.

This definition allows several geometrical possibilities, and we list some fundamental ones below.

(i) $p = 1$ and $\bar{\mathcal{Z}}^1 = T^{\ell-1}$. In this case τ is either an $(\ell - 1)$-torus or an $(\ell - 1)$-dimensional segment of a spiral manifold. See Figures 2.2.5(a),(b),(c) for tori, and Figure 2.2.5(g) for spirals. Note that one could define spiral PIM's in

Figure 2.2.5(f) as well, but we will use this figure to illustrate another type of definition of PIM's.

(ii) $p > 1$, $\bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p = T^{\ell-1}$, and $s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0}) = s^j(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ on $\mathcal{Z}^{ij}$, $i \neq j$. In this case τ is again either an $(\ell-1)$-torus or an $(\ell-1)$-dimensional segment of a spiral manifold. What typically happens in this situation is that $s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ and $s^j(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ undergo a bifurcation on $\mathcal{Z}^{ij}$ and τ is formed by piecing together the functions $s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$, $i = 1, \ldots, p$, at the surfaces where they bifurcate (note that the dimension of $\mathcal{Z}^{ij}$ is generically $(\ell - 2)$, and it is necessary for $D_{(s,\theta_1,\ldots,\theta_{\ell-1})}M(s, \theta_1, \ldots, \theta_{\ell-1}, \theta_{\ell_0}; \nu)$ to have rank zero on these surfaces). See Figure 2.2.5(d) for an example.

(iii) $p > 1$, $\bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p = T^{\ell-1}$, and $s^i(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0}) \neq s^j(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ on $\mathcal{Z}^{ij}$, $i \neq j$. In this case τ is discontinuous on the $\mathcal{Z}^{ij}$; however, τ is a single-valued graph over $T^{\ell-1}$ from our previous condition on the definition of PIM's. See Figures 2.2.5(e) and (f) for examples. At present, there may seem to be no motivation for such a choice of PIM's in Figure 2.2.5(f); however, what may seem geometrically unappealing in $\Sigma^{\theta_{\ell_0}}$ will turn out to be geometrically appealing in the nonautonomous system phase space.

(iv) $p > 1$ and $\mathcal{Z}^{ij} = \emptyset$. Unlike the previous cases, $\bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subset T^{\ell-1}$ and τ contains gaps. We define a PIM on a *subset* of $T^{\ell-1}$ when intersection manifolds do not cover $T^{\ell-1}$, as in Figure 2.2.5(h).

These four cases are not exhaustive: for example, one may have combinations of (ii),(iii), and (iv). Namely, τ may be nondifferentiable and/or discontinuous and/or possess gaps on $\mathcal{Z}^{ij}$.

As for time-periodic vector fields, general properties of $W^s(\tau_\epsilon^a) \cap W^u(\tau_\epsilon^b)$ can be inferred from the Melnikov function, so that we can understand the geometry associated with different types of perturbations. We can describe some of these properties after a definition and two lemmas.

Definition 2.2.2 For $\bar{\theta} \in T^{\ell-1}$, we refer to the plane $\chi(\bar{\theta}) \equiv \{(x,\theta)|\theta = \bar{\theta}\}$ as a *phase slice* of the Poincaré section $\Sigma^{\theta_{\ell_0}}$.

Lemma 2.2.3 If there exists an $(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1})$ for some $\bar{\nu}$ such that

$$M(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1}, \theta_{\ell_0}; \bar{\nu}) = 0$$

$$\frac{\partial M}{\partial s}(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1}, \theta_{\ell_0}; \bar{\nu}) \neq 0$$

then there exists in the phase slice $\chi(\bar{\theta})$ a countable infinity of transversal intersection points of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$.

Proof: See Appendix 2.A3.

Note how, as illustrated by the cross points in Figure 2.2.5(d), there can be, in addition to the countable infinity of transversal intersection points in a phase slice, one or more non-transversal intersection points in that phase slice. For the study of transport it does not matter whether the intersections are transversal or nontransversal, so we henceforth refer to intersection points without distinguishing between what type they are.

Lemma 2.2.4 (a)If the frequencies are all mutually incommensurate then a countable infinity of intersection points of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ in *one* phase slice of $\Sigma^{\theta_{\ell_0}}$ implies a countable infinity of such intersection points in *all* phase slices of $\Sigma^{\theta_{\ell_0}}$; (b) if one or more pairs of frequencies are commensurate then a countable infinity of intersection points of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ in *one* phase slice of $\Sigma^{\theta_{\ell_0}}$ implies a countable infinity of such intersection points in phase slices defined either over all of $T^{\ell-1}$ or some subset of $T^{\ell-1}$.

Proof: See Appendix 2.A3.

These lemmas imply the following theorem.

Theorem 2.2.5 Suppose there exists a point $(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1})$ for some $\bar{\nu}$ such that

1) $M(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{(\ell-1)}, \theta_{\ell_0}; \bar{\nu}) = 0$,

2) $\frac{\partial M}{\partial s}(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{(\ell-1)}, \theta_{\ell_0}; \bar{\nu}) \neq 0$.

If the frequencies are all mutually incommensurate, then $W^s(\tau_\varepsilon^a)$ intersects $W^u(\tau_\varepsilon^b)$ in a countable infinity of $(\ell-1)$-dimensional surfaces that can be represented as graphs over $T^{\ell-1}$. If one or more pairs of frequencies are commensurate, then $W^s(\tau_\varepsilon^a)$ intersects $W^u(\tau_\varepsilon^b)$ in a countable infinity of $(\ell-1)$-dimensional surfaces that can be represented as graphs over all $T^{\ell-1}$ or some subset of $T^{\ell-1}$.

Proof: See Appendix 2.A3.

With a definition of PIM's, and an understanding of their geometry, we can now define and study the invariant lobe structure. To deal with the arbitrary geometries of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$, it will prove convenient to define a *lobe* as an appropriate union over phase slices of a simply connected two-dimensional region in each phase slice. The first two of the following definitions thus pertain to phase slices.

Definition 2.2.3 For $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, the intersection of the PIM τ with the phase slice $\chi(\bar{\theta}) \equiv \{(x, \theta) | \theta = \bar{\theta}\}$ defines a unique point. We refer to this point $p(\bar{\theta}) = \tau \cap \chi(\bar{\theta})$ as a *primary intersection point (PIP).*

Definition 2.2.4 Let $p_1(\bar{\theta})$ and $p_2(\bar{\theta})$ denote two PIP's in the phase slice $\chi(\bar{\theta})$. Then let $S[p_1(\bar{\theta}), p_2(\bar{\theta})]$ and $U[p_1(\bar{\theta}), p_2(\bar{\theta})]$ denote the segments of $W^s(\tau_\varepsilon^a) \cap \chi(\bar{\theta})$ and $W^u(\tau_\varepsilon^b) \cap \chi(\bar{\theta})$, respectively, from $p_1(\bar{\theta})$ to $p_2(\bar{\theta})$. We say the two PIP's are *adjacent* if $S[p_1(\bar{\theta}), p_2(\bar{\theta})]$ and $U[p_1(\bar{\theta}), p_2(\bar{\theta})]$ contain no other PIP's than $p_1(\bar{\theta})$ and $p_2(\bar{\theta})$.

Definition 2.2.5 For all $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, let $p_1(\bar{\theta})$, $p_2(\bar{\theta})$ denote adjacent PIP's in the phase slice $\chi(\bar{\theta})$. A *lobe* $\mathcal{L}$ is an $(\ell+1)$-dimensional region in $\Sigma^{\theta_{\ell_0}}$ such that: (a) for all $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, $\mathcal{L} \cap \chi(\bar{\theta})$ is the region in $\chi(\bar{\theta})$ bounded by $S[p_1(\bar{\theta}), p_2(\bar{\theta})] \cup U[p_1(\bar{\theta}), p_2(\bar{\theta})]$, and

(b) the sign of $M(s, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1}, \theta_{\ell_0}; \nu)$ in the interval $s(\bar{\theta}) \in [s_1(\bar{\theta}), s_2(\bar{\theta})]$, where $s_i(\bar{\theta})$ is the s value associated with $p_i(\bar{\theta})$, for $i = 1, 2$, is independent of $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$.

The motivation for condition (b) should be clear from the familiar two-dimensional lobes associated with time-periodic vector fields: there are two classes of lobes, as distinguished by the sign of the Melnikov function associated with the lobe. When we form the union of regions over different phase slices, we want to choose regions from the same class.

This general definition of a lobe, as with the definition of a PIM, allows several geometrical possibilities; however, our understanding of the geometry of PIM's translates into an understanding of the geometry of lobes. Indeed, for the examples in Figure 2.2.5, let us choose, for all $\bar{\theta}_1 \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, $p_1(\bar{\theta}_1) = \tau_1 \cap \chi(\bar{\theta}_1)$ and $p_2(\bar{\theta}_1) = \tau_2 \cap \chi(\bar{\theta}_1)$. From the PIM's in Figure 2.2.5 one can visualize a lobe by remembering that its "thickness", *i.e.*, the separation of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ normal to $W^s(\tau_0^a) \cap W^u(\tau_0^b)$, comes out of the page (see Figure 2.2.6). For the toral PIM's of Figures 2.2.5(a),(b),(c), the three-dimensional lobe is a simply connected region whose boundary divides $\Sigma^{\theta_{2_0}}$ into an inside and outside; for the spiral PIM's of Figure 2.2.5(g), the lobe has a discontinuity at $\theta_1 = 0$; for the bifurcating PIM's of Figure 2.2.5(d) the lobe has no discontinuity but "pinches off" to zero volume at an isolated θ_1 value; for the PIM's of Figures 2.2.5(e) and (f) with one or more discontinuities in $[0, 2\pi)$, the lobe has discontinuities in $[0, 2\pi)$; for the PIM's of Figure 2.2.5(h) which contain gaps in T^1, the lobes contain gaps in T^1.

From the definition of lobes and Theorem 2.2.5, when the system satisfies the conditions of Theorem 2.2.5 then $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ form the boundaries of a countable infinity of $(\ell+1)$-dimensional lobes in $\Sigma^{\theta_{\ell_0}}$, and hence we refer to $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ as the boundary of an *invariant lobe structure* in $\Sigma^{\theta_{\ell_0}}$ (see Figure 2.2.7), or of a *homoclinic/heteroclinic tangle* in $\Sigma^{\theta_{\ell_0}}$.

2.2.2 Deriving a sequence of two-dimensional time-dependent tangles from the invariant $(\ell+1)$-dimensional tangle

2.2.2a The autonomous system

Our eventual goal is to study the dynamics of the nonautonomous system (2.2.2) under the sequence of maps $\{T_\varepsilon(\cdot\,;n);n\in Z\}$. From expression (2.2.4), and the comments that followed, the study of trajectories of (2.2.2) through x_0 at time $t=0$ for all x_0 is equivalent to the study of the x component of the trajectories of the autonomous system (2.2.3) through $(x_0,\theta_{1_0},\ldots,\theta_{\ell_0})$ at time $t=0$ for all x_0. The coordinates of the solution $\Phi(t,t_0=0,x_0,\theta_{1_0},\ldots,\theta_{\ell_0})$ at time $t=\frac{2\pi}{\omega_\ell}n$ $(n\in Z)$ for all x_0 lie in the plane

$$\{(x,\theta_1,\ldots,\theta_\ell)|\theta_i=\theta_{i_0}+2\pi\frac{\omega_i}{\omega_\ell}n;i=1,\ldots,\ell\}.$$

We shall refer to this plane as the n^{th} *time slice* of the phase space of the autonomous system, which is of course identical to a phase slice. To study the dynamics of (2.2.2), we are interested in the sequence of time slices of the autonomous system phase space, and the intersection of the invariant lobe structure with these time slices defines a sequence of two-dimensional time-dependent lobe structures. For the single-frequency case, each time slice is identical to the plane defined by the Poincaré section $\Sigma^{\theta_{1_0}}$. Hence a two-dimensional lobe structure in this plane is immediately provided by the invariant lobe boundaries $W^s(\tau_\varepsilon^a)\cup W^u(\tau_\varepsilon^b)$ in $\Sigma^{\theta_{1_0}}$ (here the τ's refer to points). For the two-frequency case, the n^{th} time slice is a plane in the three-dimensional Poincaré section $\Sigma^{\theta_{2_0}}$ defined by

$$\chi(\theta_{1_0}+2\pi\frac{\omega_1}{\omega_2}n)=\{(x,\theta_1)|\theta_1=\theta_{1_0}+2\pi\frac{\omega_1}{\omega_2}n\}$$

(see Figure 2.2.8). For $\theta_{1_0}+2\pi\frac{\omega_1}{\omega_2}n\in\bar{\mathcal{Z}}^1\times\cdots\times\bar{\mathcal{Z}}^p$ the intersection of this plane with the invariant lobe boundaries $W^s(\tau_\varepsilon^a)\cup W^u(\tau_\varepsilon^b)$ in $\Sigma^{\theta_{2_0}}$ defines the boundary

of a two-dimensional lobe structure:

$$\Gamma(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n) = \left\{W^s\left(\tau_\varepsilon^a\right) \cup W^u\left(\tau_\varepsilon^b\right)\right\} \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$$

(we henceforth assume that $\theta_{1_0} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, *i.e.*, that the initial time slice contains a lobe structure, and hence, by invariance under P_ε, all time slices have a lobe structure). The intersection of the set of PIM's with the plane $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ defines a countable infinity of PIP's, and a countable infinity of two-dimensional lobes is defined by regions in $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ bounded by segments of stable and unstable manifolds in that plane that connect adjacent PIP's. The geometry of the two-dimensional lobe structure in each time slice is thus the same as the familiar invariant lobe structure of the time-periodic case. The two-frequency case easily extends to the ℓ-frequency case by replacing θ_1, ω_1, and θ_{1_0} with $\theta \equiv (\theta_1, \theta_2, \ldots, \theta_{(\ell-1)})$, $\omega \equiv (\omega_1, \omega_2, \ldots, \omega_{(\ell-1)})$, and $\theta_0 \equiv (\theta_{1_0}, \theta_{2_0}, \ldots, \theta_{(\ell-1)_0})$, respectively. Hence the two-dimensional lobe structure in $\Sigma^{\theta_{\ell_0}}$ for the ℓ-frequency problem is defined for time $t = \frac{2\pi}{\omega_\ell}n$ by the boundary

$$\Gamma(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) = \left\{W^s\left(\tau_\varepsilon^a\right) \cup W^u\left(\tau_\varepsilon^b\right)\right\} \cap \chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n),$$

where

$$\chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) = \{(x,\theta) | \theta = \theta_0 + 2\pi\frac{\omega}{\omega_\ell}n\}.$$

We will frequently refer to the expression $\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n$, so it will be useful to remember that it specifies, along with θ_{ℓ_0}, the phases of the perturbation at the n^{th} time slice.

2.2.2b The nonautonomous system

The sequence of two-dimensional lobe structures defined in the autonomous system phase space will suffice to study the trajectories of (2.2.2). Indeed some might prefer to remain in the autonomous system phase space since, after all, the invariant manifolds that constrain the motion are in this space. However, when we study the motion of a two-dimensional nonautonomous physical system of the form

(2.2.2), there is conceptual appeal in returning to the physical space, X, rather than dealing with time slices of an $(\ell + 2)$-dimensional phase space. This is a simple matter of projecting the coordinates of the autonomous solution $\Phi(t, t_0 = 0, x_0, \theta_{1_0}, \theta_{2_0}, \ldots, \theta_{\ell_0})$ onto X, since the x component of this solution solves the nonautonomous system (2.2.2). Let $\hat{X}$ denote the projection operator onto X:

$$\hat{X} : X \times T^{\ell} \mapsto X$$

$$(x, \theta_1, \theta_2, \ldots, \theta_{\ell}) \mapsto x.$$

Then the boundaries of the two-dimensional lobe structure in X at time $t = \frac{2\pi}{\omega_\ell} n$ are given by

$$\gamma(n) = \hat{X}(\Gamma(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n)),$$

as shown in Figure 2.2.9. *We stress at the outset that for the sake of simple notation, when describing the projected quantities here and elsewhere, the dependence on initial phases* $(\theta_0, \theta_{\ell_0})$ *and frequencies* (ω, ω_ℓ) *is to be understood implicitly.* Some PIP's and two-dimensional lobes are shown explicitly in Figure 2.2.10. Earlier (following Theorem 2.2.2) we stated that not all intersection sets need be continued from the $M = 0$ sets, and in Figure 2.2.10 we explicitly see the two types of intersection sets: PIP's, marked by crosses, which are continued from, and hence within $O(\varepsilon)$ of, the zero sets of the Melnikov function, and SIP's (secondary intersection points), marked by solid dots, which need not be within $0(\varepsilon)$ of the zero sets. Similarly, in $\Sigma^{\theta_{\ell_0}}$ one has, in addition to PIM's, secondary intersection manifolds (SIM's). We refer to the time-dependent lobe structure in X as a *homoclinic or heteroclinic tangle region.* We proceed to study the dynamics in this region by using the invariance under P_ε of the higher-dimensional lobe structure in $\Sigma^{\theta_{\ell_0}}$. Note that, since in the single-frequency case the time slices and hence the lobe boundaries in each slice are the same for all values of n, the projected lobe structure on X recovers the familiar invariant lobe structure of the time-periodic case.

2.3 Describing phase space transport in the tangle regions

We proceed to describe the global dynamics within the homoclinic and heteroclinic tangle regions of (2.2.2). The global stable and unstable manifolds in $\Sigma^{\theta_{\ell_0}}$ are by definition invariant under P_ε. Hence, by (2.2.5):

$$P_\varepsilon\big(\{W^s(\tau_\varepsilon^a) \cup W^u(\tau_\varepsilon^b)\} \cap \chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big) = \{W^s(\tau_\varepsilon^a) \cup W^u(\tau_\varepsilon^b)\} \cap \chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}(n+1)), \tag{2.3.1}$$

where $\theta_0 \equiv (\theta_{1_0}, \theta_{2_0}, \ldots, \theta_{(\ell-1)_0})$ and $\omega \equiv (\omega_1, \omega_2, \ldots, \omega_{\ell-1})$ (henceforth, when we use θ, ω, or θ_0 they are understood to represent $(\ell-1)$-dimensional vectors as defined here). The n^{th} time slice lobe boundaries are thus mapped under P_ε to the $(n+1)^{th}$ time slice lobe boundaries:

$$P_\varepsilon\big(\Gamma(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big) = \Gamma(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}(n+1)). \tag{2.3.2}$$

Projecting (2.3.2) onto X and using that

$$\hat{X} \circ P_\varepsilon\big(\chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big) = T_\varepsilon\big(\hat{X}(\chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)); n\big) \tag{2.3.3}$$

gives

$$T_\varepsilon\big(\gamma(n); n\big) = \gamma(n+1). \tag{2.3.4}$$

So the lobe boundaries in X at time $t = \frac{2\pi}{\omega_t}n$ map under $T_\varepsilon(\,\cdot\,; n)$ to the lobe boundaries in X at time $t = \frac{2\pi}{\omega_t}(n+1)$. By the invariance of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ under P_ε, if an intersection point in the n^{th} time slice satisfies the criterion for a PIP, its image will satisfy the criterion. Hence the n^{th} time slice PIP's map to the $(n+1)^{th}$ time slice PIP's. Further, the n^{th} time slice segments of the stable and unstable manifolds that connect adjacent PIP's map to the $(n+1)^{th}$ time slice segments of stable and unstable manifolds connecting adjacent PIP's. Thus, each lobe in the n^{th} time slice maps under P_ε to a lobe in the $(n+1)^{th}$ time slice.

Equivalently, projecting onto X and using (2.3.3), each lobe in X at time $t = \frac{2\pi}{\omega_t} n$ maps under $T_\varepsilon(\,\cdot\,; n)$ to a lobe in X at time $t = \frac{2\pi}{\omega_t}(n+1)$ (see Figure 2.3.1). The mapping of one lobe to another under successive time samples provides a global picture of the dynamics, *i.e.*, of *transport in phase space*, in the tangle region (see Figure 2.3.2). We concentrate in this section on describing transport in terms of two-dimensional lobes mapping from one to another. We facilitate this discussion with the following definition.

Definition 2.3.1 We refer to the adjacent PIP's that are used to define a two-dimensional lobe in a phase slice as the *bounding PIP's* of that lobe. Further, we say a lobe is *between* two arbitrary PIP's in a phase slice if its bounding PIP's lie on or between these two PIP's (relative to the curve defined by either the stable or unstable manifold in that phase slice).

2.3.1 Transport in the time-periodic case

Before discussing transport in the $\ell \geq 2$ case, let us summarize the results of the single-frequency case, illustrated by the OVP flow mentioned in the introduction and described more fully in Appendix 2.A1. Figure 2.3.3 shows the invariant two-dimensional lobe structure in the $x_2 \geq 0$ region of $\Sigma^{\theta_{10}=0}$. An upper core boundary, shown by the dashed-dotted line, is defined by

$$C = U[q_\varepsilon^b, p_c] \cup S[p_c, q_\varepsilon^a] \cup x_1[q_\varepsilon^a, q_\varepsilon^b],$$

where p_c denotes the PIP at $x_1 = 0$ and $x_1[q_\varepsilon^a, q_\varepsilon^b]$ denotes the segment along the x_1 axis from the left fixed point q_ε^a to the right fixed point q_ε^b (the meaning of $U[q_\varepsilon^b, p_c]$ and $S[p_c, q_\varepsilon^a]$ should be clear from Section 2.2). One can formally choose any PIP to play the role of p_c in defining the core; however, p_c is chosen here to give a core that best approximates the unperturbed core. Regions in $\Sigma^{\theta_{10}=0}$ are *entrained into/ detrained from* the core if they are mapped from outside/inside the core to inside/outside the core. The lobes between $P_\varepsilon^{-m}(p_c)$ and $P_\varepsilon^{-(m+1)}(p_c)$ are

mapped under P_ε to the lobes between $P_\varepsilon^{-(m-1)}(p_c)$ and $P_\varepsilon^{-m}(p_c)$ such that they preserve orientation and relative ordering (this preservation is a consequence of the fact that the Poincaré map is obtained by discrete time sampling of trajectories of ODE's, as discussed in Wiggins [1988]). By the definition of the core boundary, the lobes between p_c and $P_\varepsilon^{-1}(p_c)$ reverse their orientation relative to the core boundary under application of P_ε; all other lobes preserve under P_ε their orientation relative to the core boundary. Borrowing the terminology from MacKay et al. [1984], we refer to the lobes between p_c and $P_\varepsilon^{-1}(p_c)$ as *turnstile lobes.* We assume that each turnstile lobe and its image is entirely inside or entirely outside the core (which is usually the case), with the understanding that if any one of the lobes has pieces both inside and outside the core, one can redefine the lobes to be the pieces of the old lobes contained in the appropriate region (inside or outside the core). Each turnstile lobe is either entrained into or detrained from the core under P_ε; these lobes are the *only* mechanism for transport inside and outside the core under P_ε. This motivates the following definition of *sets* of lobes: for $m \in Z$, $m > 0$, $E(m)$ and $D(m)$, respectively, denote the set of lobes that are entrained into/detrained from the core upon the m^{th} iterate of P_ε; for $m \leq 0$ they denote the set of lobes that are detrained from/entrained into the core upon the $(|m|+1)^{th}$ iterate of P_ε^{-1}. We refer to $E(m)$ and $D(m)$ as sets of *entraining* and *detraining* lobes, respectively. From previous remarks, $E(m) \cup D(m)$ is the set of lobes between $P_\varepsilon^{-(m-1)}(p_c)$ and $P_\varepsilon^{-m}(p_c)$. For the single-frequency OVP flow studied by Rom-Kedar et al. [1990], $E(m)$ and $D(m)$ each contain one lobe for all $m \in Z$. For single-frequency forcing the number of lobes is (generically) independent of m and of whether the lobes are entraining or detraining. In near-integrable systems, the Melnikov function can be used to determine the number of lobes in each set $E(m)$ and $D(m)$; otherwise one can numerically compute the lobe structure. We will discuss these procedures in the multiple-frequency case, which will also cover the single-frequency case. By their

definition the sets of lobes map under P_ε according to

$$\begin{aligned} P_\varepsilon\big(E(m)\big) &= E(m-1) \\ P_\varepsilon\big(D(m)\big) &= D(m-1) \end{aligned} \qquad \forall m \in Z. \tag{2.3.5}$$

Since the sets of lobes in this case are identical to individual lobes, equation (2.3.5) specifies the dynamics of individual lobes. In our discussion of the multiple-frequency case we will explain the more complicated scenario when the sets contain more than one lobe. Since the Poincaré section is two-dimensional, it is a trivial matter to project onto the "physical" space X, so we will not bother to do it here.

2.3.2 Transport in the quasiperiodic case

We describe transport in a manner that is step by step analogous to the approach in the time-periodic case. We illustrate our discussion with two-frequency near-integrable systems with a Melnikov function of the form

$$M(s, \theta_1, \theta_{2_0} = 0; \nu) = A_1(\mu, \omega_1)\sin(\omega_1 s + \theta_1) + A_2(\mu, \omega_2)\sin(\omega_2 s) \tag{2.3.6}$$

(refer to Section 2.2 and Appendix 2.A1). We stress, though, that the description of transport does not rely on near-integrability. To illustrate transport using a three-dimensional portrayal of the lobe structure in $\Sigma^{\theta_{2_0}=0}$ would certainly be cumbersome; we thus suppress the dimension normal to the unperturbed homoclinic/heteroclinic manifold, as done when we plotted the zero sets of the Melnikov function in the previous section. As described in that section, the set of $M = 0$ curves portray to $O(\varepsilon)$ the set of PIM's, and hence portray the dividing lines between invariant three-dimensional lobes, whose thickness comes out of the page (refer back to Figure 2.2.6). Of course, the $M = 0$ curves give no information on the thickness of the lobes normal to the unperturbed manifolds, and they only approximate the PIM's location, but they are ideally suited for our purposes in this section, that is, to illustrate globally (*i.e.*, for all θ_1) how lobes map from one time slice to the next by allowing one to visualize how the PIM's map from one to

another with each iterate of P_ε. Consider some simple choices of A_i and ω_i shown in Figures 2.3.4(a)-(g). The shading and labeling of these plots will be explained as our notation is developed. We include commensurate frequencies along with incommensurate ones: even though the commensurate cases are really periodic, they are still illuminating for our two-frequency prescription. Indeed, for many commensurate frequency cases, the common period of the two frequencies is much longer than the individual periods, so there is motivation to sample at one of the two periods and hence use the two-frequency formalism.

We first define an invariant core in $\Sigma^{\theta_{\ell_0}}$, which necessitates the definition of a manifold that plays the role that p_c does in the time-periodic case. Let τ_c be an $(\ell-1)$-dimensional PIM.

Definition 2.3.2 A *homoclinic core boundary* $\mathcal{C}$ in $\Sigma^{\theta_{\ell_0}}$ is the ℓ-dimensional surface $\{U[p_\varepsilon(\bar{\theta}), p_c(\bar{\theta})] \cup S[p_c(\bar{\theta}), p_\varepsilon(\bar{\theta})] \,|\, \bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}\}$, where $p_\varepsilon(\bar{\theta}) = \tau_\varepsilon \cap \chi(\bar{\theta})$ and $p_c(\bar{\theta}) = \tau_c \cap \chi(\bar{\theta})$.

The heteroclinic core boundary involves two heteroclinic orbits, and hence has more than one possible form. We will worry about only one of the heteroclinic orbits and hence make the following definition.

Definition 2.3.3 A *heteroclinic core boundary* $\mathcal{C}$ in $\Sigma^{\theta_{\ell_0}}$ is the ℓ-dimensional surface $\{U[p^b_\varepsilon(\bar{\theta}), p_c(\bar{\theta})] \cup S[p_c(\bar{\theta}), p^a_\varepsilon(\bar{\theta})] \cup O[p^a_\varepsilon(\bar{\theta}), p^b_\varepsilon(\bar{\theta})] \,|\, \bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}\}$, where $O[p^a_\varepsilon(\bar{\theta}), p^b_\varepsilon(\bar{\theta})]$ denotes the segment of the core boundary in the phase slice $\chi(\bar{\theta})$ due to the other heteroclinic orbit, which may be the same as the unperturbed orbit (as in the OVP flow), or may involve another heteroclinic tangle.

As done in the time-periodic case with p_c, we specialize our concern to a specific choice of τ_c, and hence of the core:

$$\tau_c = \{(x, \theta) | x = \zeta(\theta), \theta \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}\},$$

where $\zeta(\theta = \bar{\theta})$ is the x value of the PIP in the phase slice $\chi(\bar{\theta})$ such that $\mathcal{C} \cap \chi(\bar{\theta})$ *best approximates the unperturbed core in the phase slice* $\chi(\bar{\theta})$. For example, the

unperturbed heteroclinic and homoclinic orbits in X of the OVP flow and the Duffing equation are symmetric about their spatial midpoint ($s = 0$), and a natural choice for τ_c is for $\zeta(\theta = \bar{\theta})$ to be the x value of the PIP in the $\chi(\bar{\theta})$ plane whose s value is closest to zero. Figure 2.3.4 shows this choice of τ_c for the two-frequency examples: for (c),(d), and (g), τ_c is a 1-torus, while in (a),(b),(e) and (f) it is a piecewise continuous set of segments of spirals. Note that, for visual clarity, the piecewise continuous sets are joined by vertical dashed lines at the points of discontinuity; it should be understood that τ_c is really single-valued everywhere. The invariant core for this choice of τ_c immediately defines a two-dimensional core in any phase slice, $\chi(\bar{\theta})$, $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}$, and equivalently in X for any time sample:

$$C(\bar{\theta}) = \mathcal{C} \cap \chi(\bar{\theta})$$

$$c(n) = \hat{X}\big(C(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n)\big). \tag{2.3.7}$$

We stress two fundamental points regarding the invariant core and the resulting sequence of cores. First, since the invariant core boundary in general may be piecewise continuous in θ, it does not necessarily divide the Poincaré section $\Sigma^{\theta_{\ell 0}}$ into an inside and outside. This would present a problem for more general systems, such as the k degree-of-freedom systems discussed in Wiggins [1990]. *However, due to the trivial nature of the dynamics in θ (everything maps from phase slice to phase slice), all that is needed of the invariant core boundary is that it divides each phase slice into an inside and an outside, which our piecewise continuous core boundary indeed does.* By their definitions, for each n, $C(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)$ divides its two-dimensional time slice into an inside and an outside, and $c(n)$ divides X into an inside and outside. Second, though τ_c and hence the invariant core boundary may in certain cases seem complicated and geometrically unappealing, by its very definition it ensures that for each n the core $c(n)$ most closely resembles the unperturbed core boundary. The aim then in the definition of τ_c is geometric appeal in each phase slice.

We now define entrainment and detrainment with respect to the core sequence.

Definition 2.3.4 The regions in the n^{th} time slice $\chi(\theta_0 + 2\pi\frac{\omega}{\omega_\iota}n)$ of $\Sigma^{\theta_{\iota_0}}$ that are *entrained into/detrained from the core under* P_ε are the regions in that slice that are outside/inside $C(\theta_0 + 2\pi\frac{\omega}{\omega_\iota}n)$ and are mapped to inside/outside $C(\theta_0 + 2\pi\frac{\omega}{\omega_\iota}(n+1))$ in the $(n+1)^{th}$ time slice. Equivalently, the regions in X that are *entrained into/detrained from the core under* $T_\varepsilon(\,\cdot\,;n)$ are the regions in X that are outside/inside $c(n)$ and are mapped to inside/outside $c(n+1)$.

We have defined entrainment and detrainment with respect to a core that changes shape and area from one time sample to the next, which is markedly different from the time-periodic case, where the core corresponds to a fixed spatial region. Thus, before continuing we should motivate such a definition. Imagine for example a two-dimensional material blob that pulsates in some quasiperiodic fashion, as shown in the sequence of snapshots in Figure 2.3.5(a). We do not want to say there is any entrainment or detrainment here. Though we could define a fixed spatial core and monitor the back and forth motion across this core due to the pulsation, we prefer to define a time-dependent core boundary as the material blob boundary, and say that there is no entrainment/detrainment with respect to this pulsating core. Now consider a material blob which, on top of a net pulsation, shows lobe formations that one would like to associate with entrainment/detrainment, as shown in Figure 2.3.5(b). Again, here our preferred way to monitor entrainment/detrainment is with respect to a pulsating core boundary, as shown by the dashed line in Figure 2.3.5(b), since this prevents us from counting net pulsation as entrainment/detrainment. The obvious complication is: how do we distinguish between "overall pulsation" and "lobe formation"? When a material blob deforms from one time sample to the next, to distinguish between the "reversible" pulsation deformations and the "irreversible" lobe-like deformations, we need to know the long-time behavior of the material (see Figure 2.3.6(a)). An essential feature of our lobe structure comes into play here: the invariance under P_ε of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ implies that these manifolds contain

the long term information needed to provide a natural way of distinguishing between net pulsation and lobe formation for system (2.2.2). For example, returning to the OVP flow, in the region in X at time $t = \frac{2\pi}{\omega_2}(n+1)$ where $T_\varepsilon\big(c(n);n\big)$ shows lobe-like deformations, the curve $\hat{X}\big(W^s(\tau_\varepsilon^a) \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+1))\big)$ provides the natural dividing line between "inside" and "outside" since by definition the points on $W^s(\tau_\varepsilon^a)$ asymptote to τ_ε^a (see Figure 2.3.6(b)); fluid particles immediately to the left of the highlighted $\hat{X}\big(W^s(\tau_\varepsilon^a) \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+1))\big)$ segment are destined to move further and further in the $-x_1$ direction (their x_1 component eventually going to $-\infty$), while fluid particles immediately to the right are destined to move further and further in the $+x_1$ direction until they are rotated around the core to be either detrained or rotated around the core again, and so on. The invariant manifolds $W^s(\tau_\varepsilon^a)$, $W^u(\tau_\varepsilon^b)$ thus provide a natural way to define a time-dependent core in X with which to monitor entrainment/detrainment for the sequence of nonautonomous maps $\{T_\varepsilon(\,\cdot\,;n); n \in Z\}$. Though for concreteness we have phrased our motivation in the context of a fluid, we do not wish the reader to lose sight of the significance of a time-dependent core in relation to the fundamental notion of bounded and unbounded motion in non-integrable systems. Indeed points within the core may escape, and hence their motion may not be bounded; however, we shall soon see that they can escape only via the detraining turnstile lobes, and hence, when detrainment leads to unbounded motion (as is the case in the OVP flow, but not in the Duffing oscillator), it is meaningful to regard the time-dependent core boundary as the *last frontier* between bounded and unbounded motion and to view the points within the core as bounded until they escape via the detraining turnstile lobes to unbounded motion.

With a motivated definition of a core sequence, we address the global dynamics associated with lobes mapping from one time slice to the next. The two-dimensional lobes in $\chi(\bar{\theta})$, $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, between $P_\varepsilon^{-m}(\tau_c) \cap \chi(\bar{\theta})$ and $P_\varepsilon^{-(m+1)}(\tau_c) \cap \chi(\bar{\theta})$ are mapped under P_ε to the lobes in $\chi(\bar{\theta} + 2\pi\frac{\omega}{\omega_t})$ between $P_\varepsilon^{-(m-1)}(\tau_c) \cap \chi(\bar{\theta} + 2\pi\frac{\omega}{\omega_t})$ and $P_\varepsilon^{-m}(\tau_c) \cap \chi(\bar{\theta} + 2\pi\frac{\omega}{\omega_t})$ such that they preserve their orientation and relative

ordering (this preservation is again a consequence of the fact that P_ε is obtained by discrete time sampling of the trajectories of ODE's). By the definition of the invariant core boundary, the lobes in $\chi(\bar{\theta})$ between $\tau_c \cap \chi(\bar{\theta})$ and $P_\varepsilon^{-1}(\tau_c) \cap \chi(\bar{\theta})$ reverse their orientation relative to the core boundary under the application of P_ε; all other lobes in that phase slice preserve under P_ε their orientation relative to the core boundary. Hence we refer to the lobes in $\chi(\bar{\theta})$ between $\tau_c \cap \chi(\bar{\theta})$ and $P_\varepsilon^{-1}(\tau_c) \cap \chi(\bar{\theta})$ as the *turnstile lobes* of that phase slice. Again we assume that each two-dimensional turnstile lobe and its image is entirely inside or outside the core, with the understanding that we can redefine the lobes as in the time-periodic case if the need arises. Hence the turnstile lobes of each phase slice are either entrained into or detrained from the core under P_ε; again we stress these lobes are the *only* mechanism for entrainment and detrainment relative to the sequence of cores. Previous investigators have studied escape from bounded to unbounded motion in low-dimensional chaotic systems by focusing on such notions as overlap of neighboring resonance zones (see for example Goggin and Milonni [1988a] and [1988b]) and cantori as partial barriers to transport (see MacKay and Meiss [1988]). These methods provide only an approximate framework for escape to unbounded motion, in contrast to a theory dealing with the lobes of the homoclinic/heteroclinic tangles, which provides a framework for exact computation of escape from the core, and which, as we shall see, can offer simple explanations of transport phenomena that heretofore have had insufficient explanation. The behavior of the turnstile lobes motivates the following definition of *sets* of two-dimensional lobes in a given phase slice, and hence invariant $(\ell+1)$-dimensional lobes in $\Sigma^{\theta_{t_0}}$.

Definition 2.3.5 For $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}$, $m \in Z$, $m > 0$, $E(m,\bar{\theta})$ and $D(m,\bar{\theta})$ denote, respectively, the set of lobes in $\chi(\bar{\theta})$ that are entrained into/detrained from the core upon the m^{th} iterate of P_ε; for $m < 0$ they denote the set of lobes in $\chi(\bar{\theta})$ that are detrained from/ entrained into the core upon the $(|m|+1)^{th}$ iterate of P_ε^{-1}. We refer to $E(m,\bar{\theta})$ and $D(m,\bar{\theta})$ as sets of *entraining* and *detraining* lobes,

respectively. Invariant $(\ell+1)$-dimensional lobes in $\Sigma^{\theta_{\ell 0}}$ are defined by:

$$E(m) = \{E(m,\bar{\theta}) \,|\, \bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}\}$$

$$D(m) = \{D(m,\bar{\theta}) \,|\, \bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}\}.$$

From previous remarks, $E(m,\bar{\theta}) \cup D(m,\bar{\theta})$ denotes the set of lobes in $\chi(\bar{\theta})$ between $P_\varepsilon^{-(m-1)}(\tau_c) \cap \chi(\bar{\theta})$ and $P_\varepsilon^{-m}(\tau_c) \cap \chi(\bar{\theta})$. In the multiple-frequency case the number of lobes in $E(m,\bar{\theta})$ and $D(m,\bar{\theta})$, denoted $\mathcal{N}_E(m,\bar{\theta})$ and $\mathcal{N}_D(m,\bar{\theta})$, respectively, in general depend on m, $\bar{\theta}$, and whether the set is entraining or detraining. At times we will want to refer to individual lobes, which we denote by $E(m,\bar{\theta};i)$, $D(m,\bar{\theta};i)$. Hence for all $m \in Z$, $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}$:

$$E(m,\bar{\theta}) = E(m,\bar{\theta};1) \cup \cdots \cup E(m,\bar{\theta};\mathcal{N}_E(m,\bar{\theta}))$$

$$D(m,\bar{\theta}) = D(m,\bar{\theta};1) \cup \cdots \cup D(m,\bar{\theta};\mathcal{N}_D(m,\bar{\theta})).$$

From previous comments on the dynamics:

$$\begin{aligned} \mathcal{N}_E(m,\bar{\theta}) &= \mathcal{N}_E(1,\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell}(m-1)) \\ \mathcal{N}_D(m,\bar{\theta}) &= \mathcal{N}_D(1,\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell}(m-1)). \end{aligned} \tag{2.3.8}$$

Hence, determining the number of turnstile lobes in each phase slice determines the number of lobes in each set of lobes for each phase slice; further it tells us how many lobes in each phase slice are entrained into/detrained from the core under P_ε. This number can be determined visually in each phase slice for the examples in Figure 2.3.4 by the portrayals of τ_c and $P_\varepsilon^{-1}(\tau_c)$; however, one need not determine these quantities from plots of PIM's. In the near-integrable case, the Melnikov function provides a computational tool for determining these quantities. If $\mathcal{N}(\bar{\theta})$ is the number of zeros of the Melnikov function in $\chi(\bar{\theta})$ between (and not counting) τ_c and $P_\varepsilon^{-1}(\tau_c)$, then

$$\mathcal{N}_E(\bar{\theta}) + \mathcal{N}_D(\bar{\theta}) = \mathcal{N}(\bar{\theta}) + 1,$$

where for simplicity of notation we have written $\mathcal{N}_E(1,\bar{\theta}) \equiv \mathcal{N}_E(\bar{\theta})$ and $\mathcal{N}_D(1,\bar{\theta}) \equiv \mathcal{N}_D(\bar{\theta})$. If $\mathcal{N}(\bar{\theta})+1$ is even then $\mathcal{N}_E(\bar{\theta}) = \mathcal{N}_D(\bar{\theta}) = (\mathcal{N}(\bar{\theta})+1)/2$. If $\mathcal{N}(\bar{\theta})+1$ is odd then one of $\mathcal{N}_E(\bar{\theta})$, $\mathcal{N}_D(\bar{\theta})$ equals $\mathcal{N}(\bar{\theta})/2$, the other $\mathcal{N}(\bar{\theta})/2+1$; which one equals which can be determined by the sign of the Melnikov function between the zeros. If $s_c(\bar{\theta})$ denotes the s value of $\tau_c \cap \chi(\bar{\theta})$, then $(s_c(\bar{\theta}+2\pi\frac{\omega}{\omega_\ell})+\frac{2\pi}{\omega_\ell})$ is the s value of $P_\varepsilon^{-1}(\tau_c) \cap \chi(\bar{\theta})$, and hence one determines $\mathcal{N}(\bar{\theta})$ by computing the number of zeros of $M(s,\bar{\theta}_1,\ldots,\bar{\theta}_{\ell-1},\theta_{\ell_0};\nu)$ between $s_c(\bar{\theta})$ and $(s_c(\bar{\theta}+2\pi\frac{\omega}{\omega_\ell})+\frac{2\pi}{\omega_\ell})$. When systems are not near-integrable, one can resort to explicit simulation of two-dimensional lobe structures in a sequence of time slices. We discuss and present computer simulations of lobe structures in Section (2.3.3) and explain the sampling methods for simulation in Appendix 2.A4. Once one chooses a PIP $\in \tau_c$ for the n^{th} time slice, $\tau_c \cap \chi(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)$, one easily determines the PIP $\in P_\varepsilon^{-1}(\tau_c)$ by applying P_ε^{-1} to the PIP $\in \tau_c$ in the $(n+1)^{th}$ time slice, $\tau_c \cap \chi(\theta_0+2\pi\frac{\omega}{\omega_\ell}(n+1))$.

The sets of lobes map according to

$$\begin{aligned} P_\varepsilon(E(m,\bar{\theta})) &= E(m-1,\bar{\theta}+2\pi\frac{\omega}{\omega_\ell}) \\ P_\varepsilon(D(m,\bar{\theta})) &= D(m-1,\bar{\theta}+2\pi\frac{\omega}{\omega_\ell}). \end{aligned} \tag{2.3.9}$$

The action of P_ε on m follows by the definition of the lobes, and is hence not very informative; this is because we have chosen a lobe labeling that incorporates the dynamics, which will facilitate later discussions of transport. Note though that a determination of the number of turnstile lobes in each phase slice in principle specifies the number of lobes in each set of lobes, and thus along with equation (2.3.9) specifies how individual lobes map from one to another; in particular, it specifies how many lobes are entrained and detrained with each iterate of P_ε. For all useful transport expressions, however, we will be interested in how *sets* of lobes map from one to another.

Projecting onto X immediately labels individual lobes and sets of lobes in X for any time sample:

$$e(m,n;i) = \hat{X}(E(m,\theta_0+2\pi\frac{\omega}{\omega_\ell}n;i))$$

$$d(m,n;i) = \hat{X}(D(m,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n;i)) \qquad (2.3.10),$$

and

$$e(m,n) = e(m,n;1) \cup \cdots \cup e(m,n;\mathcal{N}_e(m,n))$$

$$d(m,n) = d(m,n;1) \cup \cdots \cup d(m,n;\mathcal{N}_d(m,n)),$$

where $\mathcal{N}_e(m,n) = \mathcal{N}_E(m,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)$, $\mathcal{N}_d(m,n) = \mathcal{N}_D(m,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)$ (see Figure 2.3.7). The projection further specifies how sets of lobes map from one to another under the sequence of maps $\{T_\varepsilon(\,\cdot\,;n)\,|\,n \in Z\}$:

$$\begin{aligned} T_\varepsilon(e(m,n);n) &= e(m-1,n+1) \\ T_\varepsilon(d(m,n);n) &= d(m-1,n+1). \end{aligned} \qquad (2.3.11)$$

Note that, for a general bi-infinite sequence of nonrepeating maps, to specify the number of turnstile lobes at each time sample necessitates determining a countable infinity of numbers. However, multiple-frequency perturbations offer a fundamental simplification due to the compactness of $T^{\ell-1}$: if one determines the number of turnstile lobes in each phase slice $\chi(\bar{\theta})$ for all $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}$, then by (2.3.10) this determines the number of turnstile lobes for each map in the nonrepeating sequence.

Figure 2.3.4 illustrates our description of transport. The defined τ_c and a few iterates of $P_\varepsilon^{\pm 1}$ on τ_c are portrayed; the number of turnstile lobes for each phase slice is immediately deduced by looking at the number of PIP's between τ_c and $P_\varepsilon^{-1}(\tau_c)$ in that phase slice. The $\omega_1 = \omega_2 \equiv \omega$ ($A_1 = A_2 = 1$ until otherwise stated) case in Figure 2.3.4(a) shows a single pair of entraining/detraining turnstile lobes for all $\theta_1 \in [0,2\pi)$ (except for the isolated value $\theta_1 = \pi$), and the sets of lobes map according to

$$\begin{aligned} P_\varepsilon\big(E(m,\theta_1)\big) &= E(m-1,\theta_1+2\pi) = E(m-1,\theta_1) \\ P_\varepsilon\big(D(m,\theta_1)\big) &= D(m-1,\theta_1+2\pi) = D(m-1,\theta_1) \end{aligned} \qquad \forall\,\theta_1 \in [0,2\pi).$$

One can easily visualize the lobes mapping under P_ε in Figure 2.3.4(a) by recognizing that $P_\varepsilon^{-1}(\tau_c)$ maps to τ_c; the mapping is portrayed in all of Figures 2.3.4 (a)

to (g) by showing successive images of a single two-dimensional lobe, represented in the (s, θ_1) plane by a dot between the lobe's two bounding PIP's (everything in the (s, θ_1) plane moves rigidly with this dot). The two-dimensional lobes of Figure 2.3.4(a) remain in the same phase slice and thus map within an invariant two-dimensional lobe structure. Hence for $\theta_{1_0} = 0$ we recover the time-periodic result given earlier for the OVP flow. Exact cancellation of two sine functions out of phase by π gives the vertical zero set at $\theta_1 = \pi$, and, as we shall see in Section 2.4, lobe areas vary smoothly and monotonically from zero at $\theta_1 = \pi$ to maxima at $\theta_1 = 0$ and 2π.

The $\omega_1 = 2\omega_2 \equiv 2\omega$ case of Figure 2.3.4(b) has two pairs of entraining/detraining turnstile lobes for all θ_1 (since we are sampling at the slower frequency), except for the isolated values where the PIM's intersect at $\theta_1 = \pi/2, 3\pi/2$ (we henceforth do not keep pointing out the isolated values). The sets of lobes map according to:

$$\begin{aligned} P_\varepsilon\big(E(m,\theta_1)\big) &= E(m-1, \theta_1 + 2\pi \cdot 2) = E(m-1,\theta_1) \\ P_\varepsilon\big(D(m,\theta_1)\big) &= D(m-1, \theta_1 + 2\pi \cdot 2) = D(m-1,\theta_1) \end{aligned} \qquad \forall\, \theta_1 \in [0, 2\pi).$$

As in the previous figure, P_ε acts on the indice m, but leaves (θ_1, θ_2) unchanged, so that the two-dimensional lobes remain in the same phase slice. This is an example of a general property of the commensurate frequency case: if $\omega_1/\omega_2 = i/j$, $i \in Z$, $j \in Z$, then there exists an $N \in Z$ (for example $N = i \cdot j$) such that P_ε^N leaves (θ_1, θ_2) unchanged. This operator P_ε^N thus acts on the two-dimensional lobes the same way as does the familiar Poincaré map derived from time-periodic vector fields. The price of this correspondence is that the time between iterates of P_ε^N, $\Delta t = N 2\pi/\omega_2$, can be extremely long. The $2\omega_1 = \omega_2 \equiv 2\omega$ case of Figure 2.3.4(c) is the same problem as Figure 2.3.4(b) sampled at the faster frequency (we purposely consider both samplings for comparison). Now there is only one pair of entraining/detraining turnstile lobes for all $\theta_1 \in [0, 2\pi)$, and the transport equations are given by

$$\begin{aligned} P_\varepsilon\big(E(m,\theta_1)\big) &= E(m-1, \theta_1 + \pi) \\ P_\varepsilon\big(D(m,\theta_1)\big) &= D(m-1, \theta_1 + \pi) \end{aligned} \qquad \forall\, \theta_1 \in [0, 2\pi).$$

The lobes now map back and forth between two phase slices separated by $\Delta\theta_1 = \pi$. In general for an arbitrary sampling of a commensurate frequency case, the lobes will visit a finite number of θ_1 values in $[0, 2\pi)$.

Figure 2.3.4(d) shows an incommensurate frequency case $\omega_1 = g\omega_2 \equiv \omega$, where g is the golden mean, so the sequence of nonautonomous maps is now truly nonrepeating. The PIM's appear qualitatively quite similar to the previous commensurate example. In fact, for all θ_1 values there is one pair of entraining and detraining turnstile lobes; hence a nonrepeating map sequence can have underlying PIM's with such simple geometry that the lobes map from one to another in a very simple manner. The lobe sets map according to

$$\begin{aligned} P_\varepsilon\big(E(m,\theta_1)\big) &= E\left(m-1, \theta_1 + 2\pi g\right) \\ P_\varepsilon\big(D(m,\theta_1)\big) &= D\left(m-1, \theta_1 + 2\pi g\right) \end{aligned} \qquad \forall\, \theta_1 \in [0, 2\pi).$$

The essential difference in transport compared with the previous example is that the time slices will now visit $\theta_1 \in [0, 2\pi)$ densely, as will always be the case with incommensurate frequencies. Figure 2.3.4(e) shows the case $g\omega_1 = \omega_2 \equiv \omega$, where the number of turnstile lobes depends on θ_1:

$$\mathcal{N}_E(\theta_1) = \begin{cases} 2 & \theta_1 \in \left[0,\ \frac{\pi(1+g)}{2}\right) \\ 1 & \theta_1 \in \left[\frac{\pi(1+g)}{2},\ \frac{\pi(5-3g)}{2}\right) \\ 2 & \theta_1 \in \left[\pi\left(\frac{5-3g}{2}\right),\ 2\pi\right) \end{cases}$$

$$\mathcal{N}_D(\theta_1) = \begin{cases} 2 & \theta_1 \in \left[0,\ \frac{\pi(3-g)}{2}\right) \\ 1 & \theta_1 \in \left[\frac{\pi(3-g)}{2},\ \frac{\pi(7-5g)}{2}\right) \\ 2 & \theta_1 \in \left[\pi\left(\frac{7-5g}{2}\right),\ 2\pi\right). \end{cases}$$

The lobe sets map according to

$$\begin{aligned} P_\varepsilon\big(E(m,\theta_1)\big) &= E\big(m-1, \theta_1 + 2\pi/g\big) \\ P_\varepsilon\big(D(m,\theta_1)\big) &= D\big(m-1, \theta_1 + 2\pi/g\big) \end{aligned} \qquad \forall \theta_1 \in [0, 2\pi).$$

The last two examples deal with simpler and yet generic examples of unequal Melnikov function amplitudes, and hence PIM's that are non-intersecting tori or spirals.

Figure 2.3.4(f) shows a large amplitude, slow frequency term modulated by a small amplitude, fast frequency term $\left(A_1 = (2\pi A_2) = 1, \omega_1 = \frac{\omega_2}{2\pi} \equiv \omega\right)$. The lobe structure is essentially determined by the large amplitude component, and we sample here at the faster frequency, so only a small set of θ_1 values actually have turnstile lobes. Here we have

$$\mathcal{N}_E(\theta_1) = \begin{cases} 1 & \theta_1 \in [0.5\,,\, 1.5) \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{N}_D(\theta_1) = \begin{cases} 1 & \theta_1 \in [2\pi - 2.5\,,\, 2\pi - 1.5) \\ 0 & \text{otherwise}\,. \end{cases}$$

The lobe sets map according to

$$P_\varepsilon\big(E(m,\theta_1)\big) = E\big(m-1, \theta_1 + 1\big) \qquad \forall \theta_1 \in [0, 2\pi).$$

$$P_\varepsilon\big(D(m,\theta_1)\big) = D\big(m-1, \theta_1 + 1\big)$$

Figure 2.3.4(g) is the previous system sampled at the slower frequency (*i.e.*, $(2\pi A_1) = A_2 = 1, \frac{\omega_1}{2\pi} = \omega_2 \equiv \omega$). Now the PIM's are 1-tori, rather than segments of spirals, and there is again a single pair of entraining/ detraining turnstile lobes for all $\theta_1 \in [0, 2\pi)$. The lobe sets map according to:

$$\begin{aligned} P_\varepsilon\big(E(m,\theta_1)\big) &= E(m-1, \theta_1 + (2\pi)^2) \\ P_\varepsilon\big(D(m,\theta_1)\big) &= D(m-1, \theta_1 + (2\pi)^2) \end{aligned} \qquad \forall\, \theta_1 \in [0, 2\pi).$$

For all the examples $\mathcal{N}_E(\bar{\theta}_1)$, $\mathcal{N}_D(\bar{\theta}_1)$ are determined for all $\bar{\theta}_1 \in [0, 2\pi)$, and this directly determines $\mathcal{N}_E\left(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n\right)$, $\mathcal{N}_D\left(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n\right)$, and hence $\mathcal{N}_e(n)$, $\mathcal{N}_d(n)$, for all $n \in Z$. Hence the number of entraining and detraining turnstile lobes is determined for *all* time samples, due to the compactness of $\bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p \subseteq T^{\ell-1}$.

In summary then, the transport in the tangle region of the nonautonomous system (2.2.2) is understood in terms of the two-dimensional lobes of the n^{th} time

slice of $\Sigma^{\theta_{t_0}}$ mapping under P_ε to the lobes of the $(n+1)^{th}$ time slice, and we now know how to describe such a map.

2.3.3 Computer simulation of lobe boundaries and qualitative discussion of lobe transport

As in the time-periodic case, there is for quasiperiodic vector fields a generic picture of transport associated with lobes in phase space mapping from one to another in the manner just described. Some essential features of the picture can be described in a precise way, and in the next three sections we will address some of these features. A good prelude to such a study is the computer simulation of the lobe structure. This provides a direct portrayal of transport in the tangle region, and it will make obvious some features of transport that will be studied in later sections. Our sampling method used for computer simulation, which we refer to as a *double phase slice method*, contrasts with a previous suggestion by Moon and Holmes [1985] for analyzing the dynamics under quasiperiodic vector fields, a double Poincaré map method which essentially attempts to treat the system as periodic. We defer a discussion of this procedure to Appendix 2.A4. We present simulations of two cases of the two-frequency OVP flow, one with a 1:2 frequency ratio in the forcing term (the oscillating strain-rate field),

$$\psi_{forcing} = 0.12 x_1 x_2 \{2 \cdot 0.4 \sin(2t) + 4 \cdot 1.05 \sin(4t)\}, \tag{2.3.12}$$

and one with a 1:g^{-1} frequency ratio,

$$\psi_{forcing} = 0.12 x_1 x_2 \{2 \cdot 0.4 \sin(2t + \theta_{1_0}) + 2g^{-1} \cdot 0.95 \sin(2g^{-1} t)\}, \tag{2.3.13}$$

where $\theta_{1_0} = 2\pi[8g - 4] \approx 5.933$ ($\psi_{forcing}$ is the streamfunction of the forcing term for the *nondimensional* equations 2.A1.1 in Appendix 2.A1). For each forcing term, the arguments of the first and second sine functions will be referred to as θ_1 and θ_2, with frequencies ω_1 and ω_2, respectively. We focus on the heteroclinic example here

because it has a larger core region with which to portray lobes. Further, we include a commensurate frequency example with a simple ratio because it contains many of the essential features found in the case of incommensurate frequencies, and its simplicity highlights these features. For comparison with the two-frequency case, Figure 2.3.8 shows some lobes of two single-frequency cases: (a) shows the lobes in X at $t = n\pi$ for the case $\psi_{forcing} = 0.12x_1x_2\{2 \cdot 0.4\sin(2t)\}$, and (b) shows the lobes in X at $t = n\frac{\pi}{2}$ for the case $\psi_{forcing} = 0.12x_1x_2\{4 \cdot 1.05\sin(4t)\}$. Figure 2.3.9 shows for the perturbation (2.3.12) lobes in two phase slices of $\Sigma^{\theta_{20}=0}$, $\chi(\bar{\theta}_1 = 0)$ and $\chi(\bar{\theta}_1 = \pi)$. The stable and unstable manifold separations in the phase slice $\chi(\bar{\theta}_1 = 0)$ are seen to be essentially a superposition of the manifold separations of the two lobe structures in Figure 2.3.8, as predicted by the linear Melnikov approximation of manifold separation in equation (2.2.7), valid for sufficiently small perturbations. Figures 2.3.10 and 2.3.11 show for the perturbations in equations (2.3.12) and (2.3.13), respectively, sequences of four time samples and how the lobes of fluid map within these sequences.

The time-dependent lobe structure in X for any time $t = \frac{2\pi}{\omega_2}n$ has many similarities with the invariant two-dimensional lobe structure of the time-periodic case. There is a reasonable looking core region and a bi-infinite sequence of lobes that extend from the inside and outside of the core boundary, each of which gets thinner and longer as its base (*i.e.*, where it extends from the core) gets closer to the points that lie on the normally hyperbolic invariant 1-tori. The thinner and longer the entraining/detraining lobes that extend from inside the core boundary are, the more they wind about the core and wrap around other entraining/detraining lobes and intersect other detraining/entraining lobes (this happens since $W^s(\tau_\varepsilon^a)$, $W^u(\tau_\varepsilon^b)$ cannot intersect themselves, but can intersect each other). That the time-dependent lobe structure in X has the same geometrical constraints as the invariant lobe structure of the time-periodic case will prove instrumental in easily extending some lobe intersection analysis to be discussed in Sections 2.4 and 2.6. A notable difference from the single-frequency OVP flow, however, is the variation of lobe areas, and, as

we shall see, this variation can have a significant effect on transport, for it gives one the freedom to enhance or diminish features of transport over finite time scales.

A sequence of snapshots of the lobe structure in X shows notable differences from the time-periodic case. Of course, the lobe structure now varies with each sample time $t = \frac{2\pi}{\omega_2}n$. The core region varies in shape and area with each sample, as do lobe areas relative to their ordering with respect to the PIP $\in \tau_c$ (which does not contradict that lobes of fluid conserve area as they map from one lobe structure to the next). In the 1:2 frequency ratio case (sampled at the faster frequency), the lobe structure oscillates with successive time samples between two forms: a "tall, skinny" one ($\theta_1 = 0$) and a "short, fat" one ($\theta_1 = \pi$). In the 1:g^{-1} frequency ratio case (sampled at the faster frequency) the lobe structure varies in a nonrepeating fashion with successive time samples. As should be clear from the previous sections, the key to understanding the time-dependent structure is to recognize that it is the intersection of a time slice with an invariant structure in a higher-dimensional Poincaré section; for example, we stress how there are no fixed points in the two-dimensional structure, but rather points on a normally hyperbolic invariant 1-torus in $\Sigma^{\theta_{20}}$.

As lobes of fluid map in X from one lobe structure to the next, their behavior is qualitatively similar to that found in the time-periodic case. They stretch in one direction and contract in another to produce the two essential features of lobe dynamics found in the time-periodic case: entrainment and detrainment (back and forth motion across boundaries that were barriers to transport in the unperturbed case) and repeated stretching and folding to produce chaotic advection. In Figure 2.3.12 we explicitly show the entrainment and detrainment process for the time-dependent core associated with the system described by equation (2.3.12). These features need to be understood in the framework of the sequence of nonautonomous maps on X, which will be the focus of the next three sections. An essential aspect of the dynamics to recognize right away, however, is that, because the repeated stretching and folding of lobes carries over to the quasiperiodic case (see Section 2.5), material

curves in the tangle region tend to get "attracted" to $W^u(\tau_\varepsilon^b)$, just like in the time-periodic case (refer back to Section 2.1). Since $\hat{X}\left(W^u(\tau_\varepsilon^b) \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)\right)$ varies with n, the "attracting" structure in X is time-dependent. The time-dependent lobe structure is thus the dominant structure by which to understand motion in the tangle region, and it will allow us to embrace rather than avoid the time-dependent nature of the more complicated transport issues under quasiperiodic perturbations. We wish to stress the notion of a time-dependent attractor. Often a physical system is a point in an autonomous system phase space, and so there is little motivation to conceptualize anything but the invariant attractor in the autonomous system phase space. In chaotic advection, however, where physical space at sample times is a time slice of the autonomous system phase space, the time-dependent "attractor" associated with each time slice does have physical meaning.

2.4 Entrainment and detrainment

2.4.1 Instantaneous and average flux

Several features of entrainment and detrainment can be studied via the lobe structures. We start with lobe areas, which provide a measure of *flux* across the core boundaries. For simplicity of discussion, let us first consider Hamiltonian perturbations, *i.e.*, sequences of area-preserving maps on X.

Definition 2.4.1(a) The *instantaneous flux* associated with P_ε acting on the phase slice $\chi(\bar{\theta})$, denoted by $\phi_E(\bar{\theta})$ and $\phi_D(\bar{\theta})$ for entrainment and detrainment, respectively, is the area of the regions in the phase slice that are entrained into/detrained from the core under P_ε, divided by the sampling period.

Hence expressions for instantaneous flux associated with P_ε acting on the n^{th} time

slice are given by

$$\phi_E(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) = \frac{\omega_\ell}{2\pi}\mu\big(E(1,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big)$$
$$\phi_D(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) = \frac{\omega_\ell}{2\pi}\mu\big(D(1,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big), \tag{2.4.1}$$

where $\mu(\,\cdot\,)$ denotes the area of the set of lobes within the parentheses. Recall that Section 2.3 explained how to identify the turnstile lobes in each time slice.

Definition 2.4.1(b) The *instantaneous flux* associated with $T_\varepsilon(\,\cdot\,;n)$ acting on X, denoted by $\phi_e(n)$ and $\phi_d(n)$ for entrainment and detrainment, respectively, is the area of the regions in X that are entrained into/detrained from the core under $T_\varepsilon(\,\cdot\,;n)$, divided by the sampling period.

Hence the expressions for instantaneous flux associated with $T_\varepsilon(\,\cdot\,;n)$ acting on X are given by

$$\phi_e(n) = \frac{\omega_\ell}{2\pi}\mu\big(e(1,n)\big)$$
$$\phi_d(n) = \frac{\omega_\ell}{2\pi}\mu\big(d(1,n)\big). \tag{2.4.2}$$

Since $e(i,n) = \hat{X}\big(E(i,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big)$ and $d(i,n) = \hat{X}\big(D(i,\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big)$, $\phi_e(n) = \phi_E(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)$ and $\phi_d(n) = \phi_D(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)$, so for the remainder of this section we deal just with slices of the autonomous system phase space, with the understanding of how the flux results apply directly to the nonautonomous system. Note that, in contrast to the single-frequency case, instantaneous flux will in general vary from one time slice to the next, and will be different for entrainment and detrainment.

In addition, we will also be concerned with *average flux*.

Definition 2.4.2 The *average flux* associated with P_ε acting on the bi-infinite sequence of time slices, or equivalently with the bi-infinite sequence of maps $\{T_\varepsilon(\,\cdot\,;n)\,|\,n \in Z\}$ acting on X, is the average of the instantaneous flux over all visited time slices.

Since P_ε preserves area as it acts on a time slice, the average flux is identical for entrainment and detrainment, and is thus denoted by a single symbol Φ:

$$\Phi(\theta_0) = \{< \phi_E(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) >_n + < \phi_D(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n) >_n\}/2, \tag{2.4.3}$$

where $<>_n$ denotes the average over $n \in Z$ (when one or more frequencies are mutually commensurate this average will depend on θ_0, and when the frequencies are all mutually incommensurate it will be independent of θ_0).

For sufficiently small perturbations, the Melnikov approximation of manifold separation in equation (2.2.7) provides, for any phase slice $\chi(\bar{\theta})$ of $\Sigma^{\theta_{\ell_0}}$, $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, a first order in ε approximation of areas of the individual turnstile lobes. From the comments in Section 2.2 on uniform approximation of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ on semi-infinite time intervals, for sufficiently small perturbations the boundaries of the turnstile lobes are $O(\varepsilon)$ close to the unperturbed manifold. The area element of one of these lobes, as shown in Figure 2.4.1, has an area

$$dA = \left[\frac{\varepsilon|M(s,\bar{\theta}_1,\ldots,\bar{\theta}_{\ell-1},\theta_{\ell_0};\nu)|}{\|DH(x_h(-s))\|} + O(\varepsilon^2)\right]\left[d\lambda(1+O(\varepsilon))\right]$$

$$= \frac{\varepsilon|M(s,\bar{\theta}_1,\ldots,\bar{\theta}_{\ell-1},\theta_{\ell_0};\nu)|}{\|DH(x_h(-s))\|}\,d\lambda + O(\varepsilon^2),$$

where $d\lambda$ is an element of arc length, centered at $x_h(-s)$, along the unperturbed homoclinic or heteroclinic manifold in $\chi(\bar{\theta})$. Using that

$$d\lambda = \frac{d\lambda}{ds}\,ds = \|DH(x_h(-s))\|\,ds$$

gives

$$\mu(\text{ individual turnstile lobe } (\bar{\theta})) = \varepsilon\int_{s_a(\bar{\theta})}^{s_b(\bar{\theta})} |M(s,\bar{\theta}_1,\ldots,\bar{\theta}_{\ell-1},\theta_{\ell_0};\nu)|ds + O(\varepsilon^2), \tag{2.4.4}$$

where $s_a(\bar{\theta})$ and $s_b(\bar{\theta})$ are the s values of the bounding PIP's of the individual turnstile lobe in $\chi(\bar{\theta})$. Equations (2.4.1) and (2.4.4) provide an analytical prescription

for determining instantaneous flux, and we do so shortly for the examples given in Section 2.3.

If all the frequencies are mutually commensurate, then $\theta(t = \frac{2\pi}{\omega_\ell} n) = \theta_0 + 2\pi \frac{\omega}{\omega_\ell} n$ will visit a finite number of values, N, in $T^{\ell-1}$, and the average flux is given by

$$\Phi(\theta_0) = \frac{\omega_\ell}{2\pi} \frac{\varepsilon}{2N} \sum_{n=0}^{N-1} \Big\{ \int_{s_0(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n)}^{s_{-1}(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n)} |M(s, \theta_{1_0} + 2\pi \frac{\omega_1}{\omega_\ell} n, \ldots, \theta_{(\ell-1)_0} + 2\pi \frac{\omega_{\ell-1}}{\omega_\ell} n, \theta_{\ell_0}; \nu)| ds + O(\varepsilon) \Big\}, \tag{2.4.5}$$

where

$$s_0(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n) = s \text{ value of } \tau_c \cap \chi(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n)$$
$$s_{-1}(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n) = s \text{ value of } P_\varepsilon^{-1}(\tau_c) \cap \chi(\theta_0 + 2\pi \frac{\omega}{\omega_\ell} n).$$

Given a choice of sampling phase θ_{ℓ_0}, the average flux is in general a function of the initial phases θ_0. If all the frequencies are mutually incommensurate, then $\theta(t = \frac{2\pi}{\omega_\ell} n) = \theta_0 + 2\pi \frac{\omega}{\omega_\ell} n$ will visit $T^{\ell-1}$ densely and uniformly, and hence the average flux is given by

$$\Phi = \frac{\omega_\ell}{(2\pi)^\ell} \frac{\varepsilon}{2} \int_{T^{\ell-1}} \int_{s_0(\bar{\theta})}^{s_{-1}(\bar{\theta})} |M(s, \bar{\theta}_1, \ldots, \bar{\theta}_{(\ell-1)}, \theta_{\ell_0}; \nu)| \, ds \, d\bar{\theta} + O(\varepsilon^2), \tag{2.4.6}$$

where we have left out the θ_0 argument in Φ since the average is now independent of the initial phases. The invariant lobe structure in $\Sigma^{\theta_{\ell_0}}$ is thus seen to provide direct long term information about the nonautonomous system (2.2.2), since average flux is proportional to volumes of the invariant turnstile lobes in $\Sigma^{\theta_{\ell_0}}$.

For the two-frequency case there is only one frequency ratio, either commensurate or incommensurate, and so expressions (2.4.5) and (2.4.6) exhaust all possibilities. Figures 2.4.2(a) to (i) show, for the two-frequency examples of Section 2.3, plots of individual turnstile lobe areas and of total areas of entrainment and detrainment for all phase slices (often the total areas are just individual lobe areas), as well as the averages of these areas over the visited time slices. These plots specify (by dividing the area in the appropriate time slices by the sampling period) the

instantaneous flux for all time slices and the average flux. We should remark that in addition to the average of flux over infinite time, another relevant characteristic is how long a system takes for its average over finite time to converge to the average over infinite time. This depends on how quickly $\theta_{n+1} = \theta_n + 2\pi \frac{\omega}{\omega_\ell}$ "spreads out" across $T^{\ell-1}$, as well as how the turnstile lobe areas vary along $T^{\ell-1}$. For example, in the two-frequency case, one can choose an incommensurate frequency ratio arbitrarily close to 1:1, so that, though the average over infinite time would be the average over θ_1 of the plot in Figure 2.4.2(a), it takes an arbitrarily long time for the finite-time average to converge to this mean. In contrast, the case where the frequency ratio equals $1 : g^{-1}$ spreads out quite quickly along the unit circle, and hence the finite-time average converges quickly to the infinite-time average (see Figure 2.4.3).

For more than two frequencies some of the frequency ratios may be commensurate, some incommensurate, so in general the expressions for flux will be "in between" (2.4.5) and (2.4.6), involving a combination of sums and integrals. Since the Melnikov approximation of manifold separation is valid only over a limited domain of s, one might think that we will have to content ourselves with summing and/or integrating over θ to find average flux, which would be cumbersome. However, using that by its definition the Melnikov function satisfies for all $m \in Z$

$$M(s, \theta_1, \ldots, \theta_{\ell-1}, \theta_{\ell_0}; \nu) = M(s - m\frac{2\pi}{\omega_\ell}, \theta_1 + m\frac{2\pi\omega_1}{\omega_\ell}, \ldots, \theta_{\ell-1} + m\frac{2\pi\omega_{\ell-1}}{\omega_\ell}, \theta_{\ell_0}; \nu), \tag{2.4.7}$$

and employing a simple change of variables, one can rewrite the sum and/or integral over θ as an integral over s in a constant phase slice. When all frequencies are mutually commensurate one obtains (see Appendix 2.A5):

$$\Phi(\theta_0) = \frac{\omega_\ell}{2\pi}\frac{\varepsilon}{2N}\int_{s_0(\theta_0)}^{s_{-N}(\theta_0)} |M(s, \theta_{1_0}, \ldots, \theta_{(\ell-1)_0}, \theta_{\ell_0}; \nu)| ds + O(\varepsilon^2), \tag{2.4.8}$$

where $s_{-N}(\theta_0)$ is the s value of $P_\varepsilon^{-N}(\tau_c) \cap \chi(\theta_0)$ and N is a positive integer such that $\theta_0 + 2\pi\frac{\omega}{\omega_\ell}N = \theta_0$. Recall from Section 2.3.2 that P_ε^N maps the two-dimensional lobes

within the same phase slice, and hence acts on these lobes the same way as does the familiar Poincaré map derived from time-periodic vector fields. Hence the integral runs from a chosen PIP $p_0(\theta_0)$ (with s value $s_0(\theta_0)$) to $(P_\varepsilon^N)^{-1}(p_0(\theta_0))$; it should thus be apparent that the commensurate frequency expression (2.4.8) recovers the standard result for time-periodic vector fields. A more elegant expression is obtained by letting $N \to \infty$:

$$\Phi(\theta_0) = \lim_{T \to \infty} \frac{\varepsilon}{2T} \int_0^T |M(s, \theta_{1_0}, \ldots, \theta_{(\ell-1)_0}, \theta_{\ell_0}; \nu)| ds + O(\varepsilon^2). \qquad (2.4.9)$$

(That the lower limit is 0 and not $s_0(\theta_0)$ introduces an error that goes to zero as $T \to \infty$). The case of one or more incommensurate frequencies transforms to (2.4.9) using the same method as with commensurate frequencies (equation (2.4.7) and a simple change of variables as described in Appendix 2.A5). *We should stress that the transformations do not depend on the choice of τ_c, so (2.4.9) gives the unique value of average flux for any choice of τ_c.* We also point out that for the heteroclinic core there can of course be multiple sets of turnstiles (potentially one set associated with each unperturbed heteroclinic orbit) and the above flux expressions apply to only one set of turnstiles.

With this more convenient expression, one can use the generalized Melnikov function to perform a detailed study of average flux as a function of system parameters, especially to compare single- and multiple-frequency forcing. There are three ingredients to understanding such a comparison:

1) the relative scaling factors are frequency dependent;
2) a sum of sinusoidal functions interfere;
3) one must choose a normalization of perturbation amplitudes in order to compare single- and multiple-frequency systems.

Let us compare single-frequency and two-frequency forcing in the following way (from our two-frequency comparisons, the extension to ℓ-frequency comparisons should be clear). Consider the perturbation in equation (2.2.8) with $F_0(x; \mu) = 0$

(which the systems defined in Appendix 2.A1 satisfy with $F_i(\mu,\omega_i) = \omega_i f_i$ for the OVP flow and $F_i(\mu) = f_i$ for the nondissipative Duffing oscillator). For any frequency pair (ω_1,ω_2), study the dependence of Φ on F_1, with F_2 chosen by some normalization condition. For example, suppose $F_1 + F_2 = 1$. For equal relative scaling factors, both single-frequency cases ($F_1 = 0$ and $F_1 = 1$) will correspond to absolute maxima, due to interference effects (see the solid line in Figure 2.4.4(a)). For unequal relative scaling factors (say $A_1/F_1 > A_2/F_2$ without loss of generality), the single-frequency case associated with the larger relative scaling factor (the case $F_1 = 1$) will correspond to an absolute maximum of average flux (see the dashed line in Figure 2.4.4(a)). This holds for arbitrary relative scaling factors (see Appendix 2.A6). *The interpretation for this normalization is then this: for a two-frequency perturbation, in general one of the frequencies has a greater relative scaling factor and hence a larger contribution to manifold separation and lobe area for a given forcing amplitude; to include the other frequency in the perturbation only detracts from average flux – the best one can do is the single-frequency case with the larger relative scaling factor.* Of course, the comparisons depend on the choice of normalization, and it is possible to choose a normalization such that the perturbation amplitudes are large enough at mid-range values of $F_1 \in [0,1]$ that the effects due to interference and different relative scaling factors are outweighed, and then the multiple-frequency case has maximal rates. Preferably the dynamical system equations describe a physical situation which motivates a normalization. For example, as shown in Rom-Kedar et al. [1990], the OVP flow can be thought of as a vortex pair moving in a wavy-wall channel. Suppose we choose to keep the root-mean-square wall amplitude constant for all $f_1 \in [0,1]$ for any given (ω_1,ω_2). For the long wavelength limit of the wall amplitude oscillations, the wall amplitudes are simply related to the perturbation amplitudes, and the normalization becomes $(F_1/\omega_1)^2 + (F_2/\omega_2)^2 = f_1^2 + f_2^2 = 1$. Figure 2.4.4(b) shows plots of average flux for two choices of (ω_1,ω_2). The profiles are similar to those in Figure 2.4.4(a), except that the interference dips are not quite so pronounced as with the previous normal-

ization ($F_1 + F_2 = 1$). Here too then average flux is maximal in a single-frequency limit. Note that for perturbations (2.2.8) the single-frequency (ω_i) flux is proportional to A_i, and hence to $F_i\mathcal{RSF}(\omega_i)$, where $\mathcal{RSF}(\omega_i)$ denotes the relative scaling factor. Thus, whereas for the first normalization the maximal single-frequency limit corresponds to the frequency ω_i with greater value of $\mathcal{RSF}(\omega_i)$, for the second normalization it corresponds to the frequency ω_i with greater value of $\omega_i\mathcal{RSF}(\omega_i)$.

Perhaps more important than making a comparison between the single-frequency and multiple-frequency case, however, is the ability to search through the parameter space of the multiple-frequency problem. *Hence, just as one can use the generalized Melnikov function to study where in parameter space chaos can occur (see Ide and Wiggins [1989], and Section 2.5), one can in addition search these chaotic regions to study the magnitude of the average flux.* As should be clear after Section 2.5, the amount of material in phase space that is repeatedly stretched and folded is related to the average flux (*i.e.*, to average lobe areas), and so this magnitude can be viewed as a basic measure of how "chaotic" a system is. One could of course produce endless plots like those in Figure 2.4.4 to sweep through parameter space, but it is not our goal here to do a detailed numerical study. Before one performs such a study, however, we cannot overemphasize the usefulness of the relative scaling function. *A plot of this function, along with a chosen normalization, immediately indicates the effectiveness of each frequency at producing manifold separations, which allows one to deduce such practical results as which frequencies have the greatest contribution to average flux, and when one frequency component will overwhelm the other, so that the two-frequency forcing is essentially single-frequency forcing.* For example, in the Duffing oscillator (refer to Appendix 2.A1, specifically to Figure 2.A1.4), the relative scaling function indicates that if one of the frequencies is of order one, then an additional forcing term with the same amplitude and $\omega > 5$ is negligible.

We emphasize that the analytical expressions we have provided are $O(\varepsilon)$ approximations and hence are good approximations only in the near-integrable case. An exact determination of lobe areas, and hence flux, necessitates straightforward,

yet laborious, computation. One needs to determine the turnstile lobe boundaries in each time slice, and this can be done by the procedure described in Appendix 2.A4. Then it is a simple matter of computing the areas between boundaries in a plane. Other investigators (for example MacKay and Meiss [1988]) point out in the context of time-periodic perturbations that the Melnikov function provides an *exact* expression for flux through the *unperturbed separatrix*. However, as ε grows, the core boundary formed by segments of stable and unstable manifolds can move further away from the unperturbed separatrix, and hence the difference between the flux associated with the unperturbed separatrix and the flux associated with the core sequence can grow (the former becoming smaller and smaller compared to the latter). The larger flux associated with the core sequence would be a truer measure of the magnitude of the back and forth motion in the tangle region; further, from previous comments in Section 2.3 it is the perturbed manifolds which should provide the criterion for entrainment and detrainment and act as a "last frontier" between bounded and unbounded motion.

For non-Hamiltonian perturbations there is the additional complication that the two-dimensional turnstile lobes do not conserve area as they are entrained into or detrained from the core under P_ε. The definitions and analysis for flux extend to non-Hamiltonian perturbations if we replace everywhere the area of the turnstile lobes by the area of the *images* of the turnstile lobes. This replacement introduces *no* changes to the near-integrable $O(\varepsilon)$ approximations for flux (see Remark 2 of Appendix 2.A5). When the system is not near-integrable, one needs to resort to numerical simulation of the boundaries of the *images* of the turnstile lobes.

2.4.2 Lobe geometry

The average flux provides a measure of the back and forth motion across a reasonable boundary in the tangle region. However, to answer basic questions about transport in phase space necessitates a consideration of lobe geometry and the variation of lobe areas. For example, in the case of a forced system whose unperturbed

separatrix divides phase space into bounded and unbounded motion, a fundamental concern (with a variety of physical applications) is the probability of escape from the core after n time samples. This necessitates a consideration of *lobe content*, rather than just lobe areas, which can be determined by addressing the intersection of lobes with each other and with the core. An additional concern (motivated for example by the classical treatment of molecular dissociation) is the probability of escape for points that lie on a particular level set of the unperturbed Hamiltonian. This necessitates a consideration of the intersection of lobes with the level sets of the unperturbed Hamiltonian. Both these transport problems are affected not only by the average of lobe areas, but by the *variation* of lobe areas.

Let us consider these transport problems. As mentioned in Section 2.3.3, entraining and detraining lobes can intersect each other; from these intersections we obtain a more detailed knowledge of transport in the tangle region. For example, the $t = 0$ sample of Figure 2.3.11, which we present again here in Figure 2.4.5 with labels on some lobes, shows lobe $d(3,0)$ intersecting with lobe $e(1,0)$; the fluid in $d(3,0) \cap e(1,0)$ will not only be entrained into the core in the next time sample, it will then be detrained from the core between the $n = 2$ and $n = 3$ time sample. Similarly $e(-2,0)$ intersects $d(1,0)$: the fluid in $e(-2,0) \cap d(1,0)$ will be detrained from the core in the next time sample and was entrained into the core between the $n = -3$ and $n = -2$ time sample. These concerns relate to the *content* of those lobes about to be entrained/detrained in the next time sample: examining the intersections of these lobes with other lobes provides an understanding of the history of the fluid particles, *i.e.*, when they might have been previously entrained/detrained or in the future when they may be entrained/detrained. An example of such a concern is our first question about escape from the core. We begin by examining this question in the context of the multiple-frequency OVP flow, *i.e.*, an area-preserving system whose unperturbed solution contains a pair of heteroclinic orbits that divide bounded and unbounded motion (since the system is symmetric about $x_2 = 0$, again we consider just the $x_2 > 0$ half). We can phrase the transport problem in

the following way: suppose at $t = 0$ fluid of one type (A) is in the core, and fluid of another type (B) is outside the core, then at any time $t = \frac{2\pi}{\omega_t} n$ how much of fluid A and B is inside and outside the core, *i.e.*, how much of fluid A and B have been *transported* out of and into the core, respectively? For the OVP system, the $e(m,n)$ lobes for $m \geq 1$ and the $d(m,n)$ lobes for $m \leq 0$ are contained entirely outside the core. The remaining lobes will in general lie both inside and outside the core, except for the lobes $e(0,n)$ and $d(1,n)$ by our previous turnstile assumption (see Figure 2.4.6). The particle content of a lobe of fluid is of course determined by its intersection with the core at time $t = 0$. If $\mu_A(\cdot)$, $\mu_B(\cdot)$ denote the areas of the pieces of the lobe in the parentheses containing fluid A and fluid B, respectively, then

$$\begin{aligned}
\mu_A\big(e(m,0)\big) &= \mu\big(e(m,0) \cap c(0)\big) \\
\mu_B\big(e(m,0)\big) &= \mu\big(e(m,0)\big) - \mu\big(e(m,0) \cap c(0)\big) \\
\mu_A\big(d(m,0)\big) &= \mu\big(d(m,0) \cap c(0)\big) \\
\mu_B\big(d(m,0)\big) &= \mu\big(d(m,0)\big) - \mu\big(d(m,0) \cap c(0)\big).
\end{aligned} \tag{2.4.10}$$

These expressions can be rewritten in terms of lobe intersections alone. For $m \geq 1$:

$$\begin{aligned}
\mu_A\big(e(m,0)\big) &= 0 \\
\mu_B\big(e(m,0)\big) &= \mu\big(e(m,0)\big) \\
\mu_A\big(d(m,0)\big) &= \mu\big(d(m,0)\big) - \sum_{i=1}^{m-1} \mu\big(d(m,0) \cap e(i,0)\big) \\
\mu_B\big(d(m,0)\big) &= \sum_{i=1}^{m-1} \mu\big(d(m,0) \cap e(i,0)\big),
\end{aligned} \tag{2.4.11}$$

where for $m = 1$ we replace the two sums by zero. The last two equations follow from the fact that the stable manifold cannot intersect itself, so the only way for $d(m,0)$ ($m \geq 1$) to be outside the core is for it to be in the $e(i,0)$ lobes for $i \geq 1$ (see Figure 2.4.6). The $i \geq m$ intersections are disallowed because of our assumption that the detraining turnstile lobes are completely inside the core (which is valid for the OVP flow up to quite large perturbations).

Now we wish to specify the change of core content between any two time samples. If $\Delta A_{\text{core}}(n)$ and $\Delta B_{\text{core}}(n)$ are the area changes under $T_\varepsilon(\,\cdot\,;n)$ of fluid type A and B, respectively, in the core, then

$$\begin{aligned}\Delta A_{\text{core}}(n) &= \mu_A\big(e(1,n)\big) - \mu_A\big(d(1,n)\big)\\ \Delta B_{\text{core}}(n) &= \mu_B\big(e(1,n)\big) - \mu_B\big(d(1,n)\big).\end{aligned} \tag{2.4.12}$$

From previous comments, for all positive $i \in Z$ and all $n \in Z$

$$\begin{aligned}\mu_A\big(e(i,n)\big) &= 0\\ \mu_B\big(e(i,n)\big) &= \mu\big(e(i,n)\big).\end{aligned} \tag{2.4.13}$$

To deal with $d(i,n)$, recall that $T_\varepsilon(\,\cdot\,;n)$ is area-preserving for all $n \in Z$, so that

$$\begin{aligned}\mu_A(\,\cdot\,) &= \mu_A\big(T_\varepsilon(\,\cdot\,;n)\big)\\ \mu_B(\,\cdot\,) &= \mu_B\big(T_\varepsilon(\,\cdot\,;n)\big).\end{aligned} \tag{2.4.14}$$

Using (2.4.14) and (2.4.10) gives expressions for the $d(i,n)$ lobes which, when plugged into (2.4.12) along with (2.4.13), give

$$\begin{aligned}\Delta A_{\text{core}}(n) &= -\,\mu\big(d(1+n,0)\cap c(0)\big)\\ \Delta B_{\text{core}}(n) &= \mu\big(e(1,n)\big) - \mu\big(d(1,n)\big) + \mu\big(d(1+n,0)\cap c(0)\big).\end{aligned} \tag{2.4.15}$$

This specifies the area changes in terms of areas of lobes and of intersections of lobes with the core. Alternatively, one could use (2.4.11) instead of (2.4.10) to obtain

$$\begin{aligned}\Delta A_{\text{core}}(n) &= -\mu\big(d(1,n)\big) + \sum_{j=0}^{n-1} \mu\big(d(1,n)\cap e(-j,n)\big)\\ \Delta B_{\text{core}}(n) &= \mu\big(e(1,n)\big) - \sum_{j=0}^{n-1} \mu\big(d(1,n)\cap e(-j,n)\big)\end{aligned} \tag{2.4.16}$$

(for $n = 0$ replace the sums over j by zero). This specifies the area changes in terms of areas of turnstile lobes and of intersections of turnstile lobes with other lobes. *The right hand side of (2.4.16) has the same form as in the time-periodic case with the second lobe argument 'n' simply going along for the ride. This is due to the fact*

that, as mentioned in Section 2.3, the geometrical constraints of the time-dependent lobe structures in X for the quasiperiodic case are the same as those of the invariant lobe structure of the time-periodic case. Summing the first expression of (2.4.15) or (2.4.16) over the first $\mathcal{N}$ time samples and dividing by the initial core area gives the percentage of material initially in the core that escapes after $\mathcal{N}$ time samples.

It is particularly easy to discuss lobe content in the OVP example since there is no re-entrainment (once fluid escapes the core, it never re-enters) and the sequence of maps is area-preserving. For Hamiltonian systems which involve re-entrainment, such as the non-dissipative Duffing oscillator, to specify area changes in terms of intersections of lobes with other lobes is more difficult. However, to specify area changes in terms of intersections of lobes with the core is no more difficult: one still maps lobes back to the $t = 0$ sample and sees how much of the lobe is inside and outside the $t = 0$ core. Since the lobes are no more fundamental an entity than the core, to specify area changes in terms of intersections of lobes with the core is no less fundamental than specification in terms of intersections of lobes with other lobes. Further, if one wishes to compute a lobe's content, it is easier to deal with a single intersection of the lobe mapped back to $t = 0$ with the core at $t = 0$ than with expressions involving a sum of intersections of the lobe with other lobes. Nevertheless, there is still interest in understanding the geometry of intersections of lobes, and we will discuss this point further in Section 2.6. The additional complexity of non-Hamiltonian perturbations is of course that the two-dimensional lobes do not preserve area as they map from one time sample to the next, and we also defer this point to Section 2.6.

The main difference between single-frequency and multiple-frequency systems in regard to this transport problem is the *variation* of lobe areas: *given a fixed average flux, the variation of lobe areas gives one the freedom to enhance or diminish transport in a given direction over a finite time.* For example, if one wishes to enhance immediate detrainment over the first few time samples, one can choose a perturbation time dependence such that the first few detraining lobes are large

(relative to the mean) and the first few entraining lobes are small (which forces the first few detraining lobes to start off mainly in the core and hence take more of the initial core material with it). Of course the fact that average flux is typically worse in multiple-frequency systems than in a corresponding single-frequency system (as explained earlier) tends to diminish the transport in both directions, but over a finite time interval this can be outweighed by the above effect due to variation of lobe areas.

The second transport problem we introduced in this section deals with the probability of escape for points that lie on a *particular level set of the unperturbed Hamiltonian.* For example, Noid and Stine [1979] and Goggin and Milonni [1988a] and [1988b] study dissociation of diatomic molecules due to forcing with one and two lasers, as modelled by the Morse oscillator; a classical approach to this problem is concerned with the probability that an ensemble of points at a particular level set of the unperturbed Hamiltonian (for example the one corresponding to the quantum mechanical ground state) escape from bounded to unbounded motion. They find in their numerical investigations that the two laser system tends to *enhance* dissociation (*i.e.*, escape from a particular level set to unbounded motion). This contrasts with our earlier result that two-frequency systems tend to diminish the average flux, which one might think would *hinder* dissociation. However, average flux is only one of many factors in this transport problem; one must also consider the geometry of lobes and the variation of lobe areas. For example, though the absolute value of the two-frequency Melnikov function may have a lower average over s than in a corresponding single-frequency case, for some normalizations it can have a larger maximum, and hence the lobes may better penetrate inner level sets of the unperturbed Hamiltonian. Further, as mentioned briefly in Noid and Stine [1979] and elaborated upon by Goggin and Milonni [1988b], additional forcing frequencies entail additional resonance bands within the core. Though no mention is made of the homoclinic tangle, it is clear that more efficient destruction of KAM tori within the core can allow the lobes of the homoclinic tangle to better penetrate

inner level sets. Additionally, for penetration of a fixed unperturbed level set, the previously discussed pulsation and "breathing" of the lobe structure can be significant. We see then that though average flux may decrease, there are other factors that can affect the ability of the lobes to penetrate a particular inner level set in a multiple-frequency problem. As a simple example, consider Figure 2.4.7, which shows again in (a) some lobes for the $n = 2$ time slice of the system portrayed in Figure 2.3.11, and in (b) some lobes of a corresponding single-frequency system ($f_1^2 + f_2^2$ is the same for both systems). Though the two-frequency system has a smaller average lobe area, its first detraining lobe is large relative to the mean and intersects with more inner level sets than does the first detraining lobe of the single-frequency case. Of course, this does not imply that the first $\mathcal{N}$ detraining lobes of this two-frequency system intersect with more of the level set shown than do the lobes of the single-frequency system; we are focusing for visual simplicity on the first detraining lobe. We should point out that the studies of Noid and Stine [1979] and Goggin and Milonni [1988b] consider a fairly specific situation: they focus on the *threshold intensity* needed for the ground state level set to achieve any dissociation in the long time limit. Escape over *all* time scales (including small and medium time scales) for a forcing amplitude that does not happen to be right at threshold for the concerned level set is a more complicated problem for which conclusions are less easily drawn and a study in terms of lobe dynamics is motivated. In particular, one would expect the variation of lobe areas to affect any studies on finite time scales.

The framework in terms of lobe dynamics is easily written down. One can write for the OVP flow (or equivalently the Morse oscillator, or any system with similar geometry) an expression for probability of escape for this second transport problem that is quite similar to the first expression of equation (2.4.15). If $\mathcal{L}_0$ denotes the level set of interest of the unperturbed Hamiltonian, and $\Delta\mathcal{L}_0(n)$ denotes the number of points initially on the level set that escape under $T_\varepsilon(\,\cdot\,;n)$, where the

points are initialized on the level set according to some density, then

$$\Delta\mathcal{L}_0(n) = \mu\big(d(1+n,0) \cap \mathcal{L}_0\big), \tag{2.4.17}$$

where $\mu(\,\cdot\,)$ denotes the number of points on the curve within the parentheses. Summing $\Delta\mathcal{L}_0(n)$ over the first $\mathcal{N}$ time samples and dividing by $\mu(\mathcal{L}_0)$ gives the percentage of points on the level set that have escaped after $\mathcal{N}$ time samples. Equations (2.4.15) to (2.4.17) thus provide a framework for computing answers to the two transport problems. *Once one recognizes the mechanism for, and topology of, transport, there are rich possibilities for determining answers to transport questions by explicit simulation of a finite number of lobes in the homoclinic/heteroclinic tangle.*

We conclude this section by remarking that a more detailed study of this second transport problem is given in Chapter 5.

2.5 Chaos

2.5.1 The periodic case

The chaotic nature of the dynamics in homoclinic and heteroclinic tangle regions under iterates of a two-dimensional Poincaré map derived from a time-periodic vector field is well established (see Guckenheimer and Holmes [1983] or Wiggins [1988]). Recall the snapshot sequence, shown in Figure 2.1.3, of the time-periodic OVP flow. As mentioned in the introduction, with each time sample a new lobe of fluid is entrained into the core; the lobe stretches as it winds around the core, and at the fifth time sample the first entrained lobe wraps around the fifth entrained lobe. The wrapping of lobes around one another in the invariant lobe structure in the two-dimensional Poincaré section for time-periodic vector fields entails horseshoe maps, as shown heuristically in Figure 2.5.1, where $P_\varepsilon^3(D)$ intersects D in the shape of a horseshoe. Using techniques such as those found in Wiggins [1988], one can rigorously establish the existence of horseshoe maps in the tangle region of time-periodic vector fields (one does this by considering a region sufficiently close

to, or containing, the hyberbolic fixed point). One can associate with this map the usual symbolic dynamics formalism and establish the existence of a Cantor set $\Lambda \in \Sigma^{\theta_1 0}$ on which P_ϵ^k (for large enough k) is topologically conjugate to a full shift on the bi-infinite sequence of two symbols, as shown in the commuting diagram

$$\begin{array}{ccc} \Lambda & \xrightarrow{P_\epsilon^k} & \Lambda \\ \downarrow \phi & & \downarrow \phi \\ \Sigma & \xrightarrow{\sigma} & \Sigma, \end{array}$$

where σ is the shift map acting on the set of bi-infinite sequences of two symbols Σ, and ϕ is a homeomorphism that takes each point in Λ to a sequence in Σ. This directly entails some standard features of chaotic dynamics:

1) a countable infinity of periodic points of all possible periods;
2) an uncountable infinity of nonperiodic points;
3) a dense orbit.

2.5.2 The quasiperiodic case

For quasiperiodic systems (2.2.2), the map of X from time $t = \frac{2\pi}{\omega_t} n$ to time $t = \frac{2\pi}{\omega_t}(n+1)$, $T_\epsilon(\,\cdot\,; n)$, depends on n. So we of course cannot develop the usual two-dimensional horseshoe map construction. *So in what sense is the dynamics in X chaotic?* Again we use the autonomous system (2.2.3) to construct an invariant structure with which to understand the dynamics, and then we project the sequence of time slices onto X to obtain a sequence of time-dependent structures from the invariant one. As a preview of what is to come, one can imagine, rather than a single horseshoe map, a bi-infinite sequence S_H of different "horseshoe maps" $H(\,\cdot\,; j) : X \to X$,

$$S_H = \{\ldots, H(\,\cdot\,; -j), \ldots, H(\,\cdot\,; -1), H(\,\cdot\,; 0), H(\,\cdot\,; 1), \ldots, H(\,\cdot\,; j), \ldots\},$$

and a bi-infinite sequence S_D of different domains $D(j) \in X$,

$$S_D = \{\ldots, D(-j), \ldots, D(-1), D(0), D(1), \ldots, D(j), \ldots\},$$

such that $H(D(j); j)$ intersects $D(j+1)$ in the shape of a horseshoe (see Figure 2.5.2). There is thus a sequence of formed horseshoes landing on different regions of X; each time the horseshoe happens to land on the region that will next form a horseshoe, and land in such a way that the stretched direction "aligns" with the direction about to be stretched. We refer to this as a *traveling horseshoe map sequence.* It is clear that this sequence of maps retains the essential ingredient of chaos – repeated stretching and folding, and hence sensitive dependence on initial conditions. Though our discussion here is heuristic, Figure 2.5.2 should make apparent what we mean.

A rigorous construction of a traveling horseshoe map sequence can be made for systems of the form (2.2.2) which possess an invariant lobe structure. For simplicity of discussion, let us consider the two-frequency homoclinic case. In Section 2.3.3 we discussed how the two-dimensional lobes in any phase slice of $\Sigma^{\theta_{20}}$ fold and wrap violently around one another just as in the invariant lobe structure of the time-periodic case. These lobes exist in all phase slices $\chi(\bar{\theta}_1)$, $\bar{\theta}_1 \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, to give in $\Sigma^{\theta_{20}}$ a three-dimensional lobe structure that folds and wraps violently around itself in the direction "normal" to τ_ε. The Poincaré section $\Sigma^{\theta_{20}}$ thus contains a three-dimensional region $\mathcal{R}$ whose image under $P_\varepsilon^k(\mathcal{R})$ (for k sufficiently large) intersects $\mathcal{R}$ such that, for any $\bar{\theta}_1 \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, $P_\varepsilon^k(\mathcal{R}) \cap \chi(\bar{\theta}_1)$ intersects $\mathcal{R} \cap \chi(\bar{\theta}_1)$ in the shape of a horseshoe (see Figure 2.5.3). Using techniques such as those found in Wiggins [1988], one can rigorously establish the existence of such a region $\mathcal{R}$ (in a manner similar to the periodic case, one needs to consider a region sufficiently close to or containing the normally hyperbolic invariant 1-torus). The geometry of $\mathcal{R} \cap P_\varepsilon^k(\mathcal{R})$, and of the resulting Cantor set, is determined by the geometry of the PIM's (just as was the geometry of the lobes in Section 2.2). When the PIM's are non-intersecting 1-tori (this case is dealt with rigorously in Wiggins [1988]), $\mathcal{R} \cap P_\varepsilon^k(\mathcal{R})$ is a simply connected region whose boundaries divide $\Sigma^{\theta_{20}}$ into an inside

and outside, and there is a Cantor set of 1-tori, $\tilde{\Lambda}$, on which P_ε^k is topologically conjugate to a full shift on the bi-infinite sequence of two symbols:

$$\begin{array}{ccc} \tilde{\Lambda} & \xrightarrow{P_\varepsilon^k} & \tilde{\Lambda} \\ \downarrow \tilde{\Phi} & & \downarrow \tilde{\Phi} \\ \Sigma & \xrightarrow{\sigma} & \Sigma, \end{array} \tag{2.5.1}$$

where $\tilde{\Phi}$ is a homeomorphism that takes each torus in $\tilde{\Lambda}$ to a sequence in Σ. Note how by a "Cantor set of 1-tori" we mean a set of 1-tori whose intersection with $\chi(\bar{\theta}_1)$ for any $\bar{\theta}_1 \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$ defines a Cantor set of points. The role of points in the time-periodic case thus applies to 1-tori in the two-frequency case (see Figure 2.5.4), and the dynamics on the Cantor set of 1-tori is thus understood to be chaotic, with:

1) a countable infinity of periodic 1-tori of all possible periods;
2) an uncountable infinity of nonperiodic 1-tori;
3) a 1-torus whose orbit under P_ε^k is dense in $\tilde{\Lambda}$.

Heuristically, then, points which lie on this set of 1-tori move chaotically normal to τ_ε as they move in a regular manner "along" τ_ε (*i.e.*, in the θ_1 direction). Just as one can establish chaos for time-periodic vector fields in which $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ intersect non-transversally (see Guckenheimer and Holmes [1983]), one can construct an invariant Cantor set in $\Sigma^{\theta_{20}}$ when the toral PIM's meet at isolated points to give non-transversal intersections at that point. When the PIM's are segments of spirals (either intersecting or non-intersecting), then $P_\varepsilon^k(\mathcal{R}) \cap \mathcal{R}$ will in general consist of piecewise continuous segments of "spiral" volumes. Figure 2.5.3 shows the case where $\mathcal{R}$ is a segment of a "spiral" volume from 0 to 2π; hence $\mathcal{R}$ is discontinuous at $\theta_1 = 0$. Applying P_ε^k to $\mathcal{R}$ sends this discontinuity to a new θ_1 value, and the intersection of $P_\varepsilon^k(\mathcal{R})$ with $\mathcal{R}$ creates another discontinuity at $\theta_1 = 0$. Explicit construction of a Cantor set by repeated application of P_ε^k and P_ε^{-k} introduces a new discontinuity in θ_1 with each application, so that the resulting Cantor set $\tilde{\Lambda}$ in $\Sigma^{\theta_{20}}$ will consist of a countable infinity of piecewise continuous segments of spirals that intersect each phase slice, $\chi(\bar{\theta}_1)$, $\bar{\theta}_1 \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$,

in a Cantor set of points. Note that for those perturbations for which the lobe structure exists on only a subset of T^1, the Cantor set will exist on a subset of T^1 (further note that this occurs only in the commensurate frequency case, which from previous discussion can be described by the time-periodic formalism, and hence by the standard horseshoe map construction). For example, in Figure 2.3.4(a), where the lobe structure vanishes in the plane $\chi(\bar{\theta}_1 = \pi)$, we have to exclude from our consideration a finite but arbitrarily small window in θ_1 around $\theta_1 = \pi$ to obtain a finite k in P_ε^k. More significantly, in Figure 2.2.5(h) there are gaps in θ_1 for which there are no intersection manifolds and hence no lobe structures, and there are thus gaps in the resulting Cantor set.

Regardless of the geometry of $\mathcal{R} \cap P_\varepsilon^k(\mathcal{R})$ in the Poincaré section $\Sigma^{\theta_{2_0}}$, the derivation of a traveling horseshoe map sequence is straightforward. From the three-dimensional region $\mathcal{R}$ we can define a two-dimensional region in any phase slice $\chi(\bar{\theta}_1)$, $\bar{\theta}_1 \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, by $R(\bar{\theta}_1) = \mathcal{R} \cap \chi(\bar{\theta}_1)$. That $P_\varepsilon^k(\mathcal{R})$ intersects $\mathcal{R}$ in any of the above phase slices in the shape of a horseshoe directly implies that $P_\varepsilon^k\big(R(\bar{\theta}_1)\big)$ intersects $R\left(\bar{\theta}_1 + 2\pi \frac{\omega_1}{\omega_2} k\right)$ in the shape of a horseshoe. Projecting onto X,

$$r(n) = \hat{X}\big(R(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n)\big),$$

and using (2.3.3) gives that $T_\varepsilon(\,\cdot\,; n+k-1) \circ \ldots \circ T_\varepsilon(\,\cdot\,; n+1) \circ T_\varepsilon(r(n); n)$ intersects $r(n+k)$ in the shape of a horseshoe (see Figure 2.5.5). We thus have our defined sequence of traveling horseshoe maps, with

$$D(j) \equiv r(j \cdot k)$$

$$H(\,\cdot\,; j) \equiv T_\varepsilon\big(\,\cdot\,; (j+1)k - 1\big) \circ \cdots \circ T_\varepsilon(\,\cdot\,; j \cdot k + 1) \circ T_\varepsilon(\,\cdot\,; j \cdot k).$$

Of course nothing magical is happening here: it is just a matter of images of two-dimensional lobes still folding and wrapping around one another ad infinitum even though the entire lobe structure is varying from one time sample to the next. In fact, if one thinks of, say, a mixing fluid, the case where fluid lobes always wrap around each other with each time sample in the exact same way seems more of an anomaly than the case we have here.

We can also use the invariant Cantor set in $\Sigma^{\theta_{2_0}}$ to define for each time slice a Cantor set of points (when the lobe structure and hence the Cantor set has gaps, recall that we assume we start in a time slice with a lobe structure). We define the Cantor set of points in the phase slice $\chi(\bar{\theta}_1)$ to be $\Lambda(\bar{\theta}_1) = \tilde{\Lambda} \cap \chi(\bar{\theta}_1)$. From the commutating diagram (2.5.1), we directly obtain

$$\begin{array}{ccc} \Lambda(\bar{\theta}_1) & \xrightarrow{P_\epsilon^k} & \Lambda(\bar{\theta}_1 + 2\pi \frac{\omega_1}{\omega_2} k) \\ \downarrow \Phi & & \downarrow \Phi \\ (\Sigma, \bar{\theta}_1) & \xrightarrow{\hat{\sigma}} & \left(\Sigma, \bar{\theta}_1 + 2\pi \frac{\omega_1}{\omega_2} k\right), \end{array} \tag{2.5.2}$$

where

$$\hat{\sigma}\left(\{\cdots s_{-n} \cdots s_{-1} \cdot s_0 s_1 \cdots s_n \cdots\} \equiv s, \bar{\theta}_1\right) = \left(\{\cdots s_{-n} \cdots s_{-1} s_0 \cdot s_1 \cdots s_n \cdots\}, \bar{\theta}_1 + 2\pi \frac{\omega_1}{\omega_2} k\right)$$

and

$$\Phi^{-1}(s, \bar{\theta}_1) = \tilde{\Phi}^{-1}(s) \cap \chi(\bar{\theta}_1)$$

(see Guckenheimer and Holmes [1983] or Wiggins [1988] for a discussion of the standard notation for the symbolic dynamics; further note how the s variable in this context is of course different from the parameter for the unperturbed homoclinic/heteroclinic manifold seen elsewhere). The operator $\hat{\sigma}$ is similar to the "extended shift map" of Stoffer [1988a,b]. Projecting (2.5.2) onto X, we have:

$$\begin{array}{ccc} \lambda(j) & \xrightarrow{T_\epsilon(\,\cdot\,;(j+1)k-1) \circ \cdots \circ T_\epsilon(\,\cdot\,;j\cdot k)} & \lambda(j+1) \\ \downarrow \phi(\,\cdot\,;j) & & \downarrow \phi(\,\cdot\,;j+1) \\ \Sigma & \xrightarrow{\sigma} & \Sigma, \end{array} \tag{2.5.3}$$

where

$$\lambda(j) = \hat{X}\left(\Lambda(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2}(j \cdot k))\right)$$

$$\phi^{-1}(s; j) = \hat{X}\left(\tilde{\Phi}^{-1}(s) \cap \chi\left(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2}(j \cdot k)\right)\right).$$

Some may feel more comfortable with topological conjugacy to an "extended shift map", shown in (2.5.2), than with the diagram (2.5.3), but both relations say the same thing: the dynamics in X from one time sample to the next can be described by a shift map relative to a time-dependent Cantor set of points. Properties such as periodic points of all periods or dense orbits are thus to be understood relative to this time-dependent set, rather than fixed spatial coordinates. Note how, though for simplicity we have discussed the two-frequency case throughout this section, the results hold as well for the general ℓ-frequency problem, where the Cantor set in $\Sigma^{\theta_{\ell_0}}$ consists of $(\ell-1)$-dimensional objects (for example $(\ell-1)$-tori or $(\ell-1)$-dimensional segments of spirals).

Having discussed chaos, we reiterate our previous comments on searching parameter space to determine when one can and cannot have chaos. From previous remarks, such a search is equivalent to a search in parameter space to determine whether a lobe structure exists, which in the near-integrable case is equivalent to determining whether zero sets of the Melnikov function exist. We stress that such a search can have simple, practical consequences. For example, in commensurate two-frequency systems the lobe structures, and hence Cantor sets, can have gaps in θ_1 (refer back to Figure 2.2.5(h)): hence the choice of θ_{1_0} affects whether the system has chaotic dynamics (if θ_{1_0} lies in a gap there will be no chaos). *Thus a shift in the relative initial phase $\theta_{1_0} - \theta_{2_0}$ can suppress chaos.*

2.6 Specifying transport via lobe intersections

In the last part of Section 2.4, we addressed the question for the two-frequency OVP flow: if at time $t = 0$ there is fluid of type A in the core and type B outside the core, then how much of A and B is inside and outside the core at any later time $t = \frac{2\pi}{\omega_2}n$? We answered this question two ways: in terms of (i) intersections of lobes with the core, and (ii) intersections of lobes with each other. As was mentioned, the first approach is easily extended to more complicated systems such

as the Duffing oscillator, where re-entrainment can occur. There is still interest, however, in understanding the geometry of lobe intersections with other lobes, and how this applies to transport questions such as the one above. Rom-Kedar and Wiggins [1990] give a detailed study of transport under two-dimensional maps by focusing on intersections of lobes with other lobes. The classical Poincaré map reduction of periodically forced systems can thus be described by this analysis. The analysis applies to an arbitrary number of regions of phase space bounded by pieces of stable and unstable manifolds of hyperbolic fixed points, and does not require the map to be area preserving or near-integrable. The generalization to a sequence of maps derived from a quasiperiodic vector field is fairly easy, as we explain in this section. Our goal here is certainly not to go through the same detailed and laborious analysis given for the time-periodic case; we wish to merely summarize some main results of Rom-Kedar and Wiggins [1990] and explain how the formulas extend to the quasiperiodic case with little change.

2.6.1 The periodic case

Consider a $C^r, r \geq 1$, orientation preserving map $T : X \mapsto X$. Let p_i, $i = 1, 2, \ldots, N$, denote N hyperbolic fixed points of T, with stable and unstable manifolds denoted by $W^s(p_i)$ and $W^u(p_i)$, respectively. As done when we defined in X the core region of the single-frequency OVP flow, we can divide X into *regions*, *i.e.*, simply connected domains of phase space with boundaries consisting of boundaries of phase space and segments of stable and unstable manifolds starting at hyperbolic fixed points and ending at either PIP's or at the boundary of phase space (which can be infinity). Suppose one divides X into N_R disjoint regions, denoted R_i, $i = 1, 2, \ldots, N_R$, and that initially particles of species S_i are uniformly distributed throughout region R_i, $i = 1, 2, \ldots, N_R$. Then the goal is to compute the area occupied by species S_i in region R_j (for any i and j) after an arbitrary number of iterations of the map. As has been explained, the turnstile lobes are the only mechanism for transport across the boundaries; one can identify these lobes

and from lobe intersection analysis specify the particle content of these lobes. For $m \in Z$ and $m \geq 1$, let $L_{k,j}(m)$ denote the set of lobes that will leave R_k and enter R_j upon the m^{th} iterate of T; for $m \leq 0$ let $L_{k,j}(m)$ denote the set of lobes that leave R_j and enter R_k upon the $(|m|+1)^{th}$ iterate of T^{-1}. Further, let $L^i_{k,j}(m)$ denote the portion of these lobes that contain species S_i (*i.e.*, at $t = 0$, $L_{k,j}(m) \cap R_i$). (Note how, as described briefly in Section 2.3 and more fully in Rom-Kedar and Wiggins [1990], typically the turnstile lobes $L_{k,j}(1)$ are contained entirely in R_k; we assume this to be the case, being aware that for cases where $L_{k,j}(1)$ is contained in regions other than R_k we can redefine the lobes to be the pieces of $L_{k,j}(1)$ that are in R_k.) Then for area-preserving maps (we mention later the case where area is not preserved), the area occupied by species S_i in region R_j immediately after the n^{th} iterate of T $(n \geq 1)$ is given by

$$T_{i,j}(n) = \delta_{ij}\mu(R_j) + \sum_{k=1}^{N_R} \sum_{m=1}^{n} \left[\mu\left(L^i_{k,j}(m)\right) - \mu\left(L^i_{j,k}(m)\right) \right], \tag{2.6.1}$$

where δ_{ij} is the Kronecker delta and

$$\mu\left(L^i_{k,j}(m)\right) = \sum_{s=1}^{N_R} \sum_{p=1}^{m} \mu\left(L_{k,j}(m) \cap L_{i,s}(p)\right) - \sum_{s=1}^{N_R} \sum_{p=1}^{m-1} \mu\left(L_{k,j}(m) \cap L_{s,i}(p)\right). \tag{2.6.2}$$

Note how (2.6.2) is purely a geometrical result. The area of the portion of a given lobe that contains particle species S_i is simply the area of the intersection of region R_i with the lobe mapped back to time $t = 0$. This is rewritten in terms of intersections of lobes with other lobes based on the geometrical constraints of the lobe structure: for example, that the stable and unstable manifold can never intersect themselves, but can intersect each other, and that $L_{k,j}(m)$ (by our assumption) will after $(m-1)$ iterations of T be contained entirely in region R_k (this determines the truncation of the sum in (2.6.2)). Using that $T^q(L_{k,j}(m)) = L_{k,j}(m-q)$ and that T is area preserving, (2.6.2) can be rewritten as

$$\mu\left(L^{i}_{k,j}(m)\right) = \sum_{s=1}^{N_R}\sum_{p=0}^{m-1}\mu\left(L_{k,j}(1)\cap L_{i,s}(1-p)\right) - \sum_{s=1}^{N_R}\sum_{p=1}^{m-1}\mu\left(L_{k,j}(1)\cap L_{s,i}(1-p)\right). \tag{2.6.3}$$

This rewrites areas in terms of intersections of the turnstile lobes with other lobes.

We conclude the discussion of the periodic case by noting that, though not explicitly stated in Rom-Kedar and Wiggins [1990], these transport equations in general apply to a division of phase space such that there is at most a single set of turnstiles between each pair of regions; one can of course relax this restriction by considering at the end of the transport calculation some particle species types to be identical, which is equivalent to grouping together previously distinct regions to form more complicated regions that allow multiple sets of turnstiles between pairs of regions.

2.6.2 The quasiperiodic case

Let P_ε now have N normally hyperbolic invariant $(\ell-1)$-tori in $\Sigma^{\theta_{\iota_0}}$, denoted by τ^i_ε, $i = 1,2,\cdots,N$, with stable and unstable manifolds denoted by $W^s(\tau^i_\varepsilon)$ and $W^u(\tau^i_\varepsilon)$, respectively. Further we relax the condition that P_ε be derived from a near-integrable system (though we retain the same notation). As done when we defined the core in Section 2.3, we can divide each phase slice $\chi(\bar{\theta})$, $\bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p$, of $\Sigma^{\theta_{\iota_0}}$ into N_R two-dimensional disjoint regions, denoted $R_i(\bar{\theta})$, $i = 1,2,\ldots,N_R$, that are simply connected domains of the phase slice with boundaries consisting of boundaries of phase space and segments of stable and unstable manifolds in that phase slice starting at normally hyberbolic invariant $(\ell-1)$-tori and ending at either PIM's or at the boundaries of phase space. Invariant $(\ell+1)$-dimensional regions in $\Sigma^{\theta_{\iota_0}}$ are then defined by

$$\mathcal{R}_i = \{R_i(\bar{\theta}) \,|\, \bar{\theta} \in \bar{\mathcal{Z}}^1 \times \cdots \times \bar{\mathcal{Z}}^p\}.$$

Projecting the n^{th} time slice onto X, we have regions in X for the nonautonomous system (2.2.2) at time $t = \frac{2\pi}{\omega_\ell} n$ defined by

$$r_i(n) = \hat{X}\big(R_i(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big)$$

(assuming that the initial time slice is a phase slice with a set of lobe structures). Note how in the single core case we defined a core region in $\Sigma^{\theta_{\ell 0}}$ (by choosing a specific τ_c) that was fairly easy to deal with and, we argued, that was a meaningful region with which to monitor transport. By our above general definition of a region one can of course construct invariant regions in $\Sigma^{\theta_{\ell 0}}$ whose piecewise continuous variation in θ behaves wildly or whose shape in the phase slice $\chi(\bar{\theta})$ is quite contorted, so that, though one could still monitor what goes in and out of these regions, the analysis would be cumbersome and the results not very useful. The goal of course is to use the definition to construct reasonable regions in the spirit of Section 2.3. For $m \in Z$, $m \geq 1$, we denote by $L_{k,j}(m, \bar{\theta})$ the set of two-dimensional lobes in $\chi(\bar{\theta})$ that leave $R_k(\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell}(m-1))$ and enter $R_j(\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell}m)$ (or, equivalently, leave $\mathcal{R}_k$ and enter $\mathcal{R}_j$) upon the m^{th} iterate of P_ε; for $m \leq 0$, $L_{k,j}(m, \bar{\theta})$ denotes the set of lobes in $\chi(\bar{\theta})$ that leave $R_j(\bar{\theta} - 2\pi\frac{\omega}{\omega_\ell}|m|)$ and enter $R_k(\bar{\theta} - 2\pi\frac{\omega}{\omega_\ell}(|m|+1))$ upon the $(|m|+1)^{th}$ iterate of P_ε^{-1}. Projecting onto X, we obtain for the n^{th} time sample of the nonautonomous system (2.2.2) a labeling of lobes $\ell_{k,j}(m, n)$:

$$\ell_{k,j}(m, n) = \hat{X}\big(L_{k,j}(m, \theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)\big).$$

As in the periodic case, the area of the portion of a given lobe that contains particles of species S_i is the area of the intersection of the region $r_i(0)$ with the lobe mapped back to $t = 0$. The geometrical constraints of the lobes in X for any time sample are identical to those of the time-periodic case: the two-dimensional lobes are bounded by stable and unstable manifolds which cannot intersect themselves, but which intersect each other, with the added assumption of the turnstile lobes $\ell_{k,j}(1, n)$

being contained totally in $r_k(n)$. So the same arguments that give (2.6.1) and (2.6.2) for the time-periodic case give for the quasiperiodic case

$$T_{i,j}(n) = \delta_{ij}\mu\left(r_j(0)\right) + \sum_{k=1}^{N_R}\sum_{m=1}^{n}\left[\mu\left(\ell^i_{k,j}(m,0)\right) - \mu\left(\ell^i_{j,k}(m,0)\right)\right], \tag{2.6.4}$$

where $T_{i,j}(n)$ is the area occupied by species S_i in region $r_j(n)$ at the n^{th} time sample, and

$$\mu\left(\ell^i_{k,j}(m,0)\right) = \sum_{s=1}^{N_R}\sum_{p=1}^{m}\mu\left(\ell_{k,j}(m,0)\cap\ell_{i,s}(p,0)\right) - \sum_{s=1}^{N_R}\sum_{p=1}^{m-1}\mu\left(\ell_{k,j}(m,0)\cap\ell_{s,i}(p,0)\right). \tag{2.6.5}$$

Applying $T_\varepsilon(\,\cdot\,;m-2)\circ\cdots\circ T_\varepsilon(\,\cdot\,;0)$ to the right hand side of (2.6.5) and using that $T_\varepsilon(\,\cdot\,;n)$ is area-preserving for all n (we are still considering the area-preserving case), we obtain

$$\begin{aligned}\mu\left(\ell^i_{k,j}(m,0)\right) = &\sum_{s=1}^{N_R}\sum_{p=0}^{m-1}\mu\left(\ell_{k,j}(1,m-1)\cap\ell_{i,s}(1-p,m-1)\right)\\ &-\sum_{s=1}^{N_R}\sum_{p=1}^{m-1}\mu\left(\ell_{k,j}(1,m-1)\cap\ell_{s,i}(1-p,m-1)\right).\end{aligned} \tag{2.6.6}$$

This rewrites areas in terms of intersections of the turnstile lobes in the $(m-1)^{th}$ time sample with the other lobes in the $(m-1)^{th}$ time sample. *Equation (2.6.6) is identical in form to equation (2.6.3), with the* $(m-1)$ *indice going along for the ride.*

The non-Hamiltonian case is extended from time-periodic to quasiperiodic vector fields in a similar manner, but we will not go through the formulas here: we refer the reader to Rom-Kedar and Wiggins [1990] for the time-periodic case (the additional complexity is that one must now of course keep track of the variation of a lobe's area as it maps from one time slice to the next).

2.7 Conclusions and explanation of extending the analysis to more general time dependences

On the one hand, as is the case with turbulent fluid flows, the behavior of nonlinear physical systems is often so complex that one can obtain only a very qualitative or partial understanding of the motion. On the other hand, there do exist nonlinear systems which display complicated dynamics, and yet with which one can obtain a fundamental and detailed global picture of the dynamics. At present, there is a wide gap between these two types of systems, and one would of course like to bridge this gap by extending some of the initial paradigms of low-dimensional dynamical systems analysis to more complicated scenarios. It is in this spirit that we have extended some notions of transport in phase space associated with the classical Poincaré map reduction of two-dimensional periodically forced systems to apply to a bi-infinite sequence of nonautonomous maps derived from a quasiperiodic two-dimensional vector field. In a certain sense, this is a mild extension, since it still deals with a highly limited class of perturbation time dependences. We have deliberately focused on a simple extension (especially the two-frequency case) since any nontrivial advance past the ubiquitous time-periodic case immediately entails a fundamental departure in the analysis (going from maps to sequences of maps) and the new concepts in the simple two-frequency extension are for the most part robust (as we shall describe momentarily), applicable to more general time dependences. Our goal then has been to discuss these new concepts in the context of a simple class of systems.

Suppose we wish to consider more complicated perturbation time dependences and remove the restriction that $g(x,t;\mu,\varepsilon)$ takes on the form in (2.2.2). We can still always introduce the associated autonomous system:

$$\begin{aligned} \dot{x} &= JDH(x) + \varepsilon g(x,\theta;\mu,\varepsilon) \\ \dot{\theta} &= 1. \end{aligned} \tag{2.7.1}$$

The phase space is now $X \times \mathbb{R}$ (the θ phase space is no longer compact), and we

write $\theta = t + \theta_0$. An unperturbed homoclinic orbit in the autonomous system phase space is now as shown in Figure 2.7.1(a). If one can establish for one's perturbation, as done in Section 2.2 for quasiperiodic perturbations, persistence of one or more invariant normally hyperbolic sets, and the existence of infinitely many intersections of global stable and unstable manifolds in a given time slice $\chi(\bar{\theta}) = \{(x, \theta) | \theta = \bar{\theta}\}$ (and hence trivially in all time slices), then one has an invariant lobe structure upon which to define transport (see Figure 2.7.1(b)). One can define an invariant core boundary and label the lobes in $X \times \mathbb{R}$ just as done in the quasiperiodic case (things will thus be piecewise continuous in θ). A sequence of monotonically increasing real numbers defines a sequence of sample times, and the evolution under (2.7.1) from one time slice to the next defines a sequence of maps. For any given time slice a two-dimensional lobe structure, a core boundary, and lobe labeling is defined by the intersection of the time slice with the invariant objects in $X \times \mathbb{R}$. Using the invariance of the lobe structure in the autonomous system phase space, one can monitor how the lobes map from one time slice to the next, identify turnstile lobes, and from this define and quantify entrainment and detrainment with respect to the time-dependent core, and so forth. One can simulate the lobe structure in any time slice using a method just like the double phase slice method described in Appendix 2.A4 (more appropriately called a *double time slice method* here).

Meyer and Sell [1989] establish persistence of invariant normally hyperbolic manifolds for all perturbations that are bounded and uniformly continuous in time, and then construct a Melnikov functional. Stoffer [1988a], [1988b] defines a Melnikov function for nonautonomous maps and shows that the sequence of maps admits a hyperbolic set (which is equivalent to the presence of an invariant lobe structure) if the Melnikov function has infinitely many simple zeros with derivatives bounded away from zero. Their Melnikov analysis applying to (2.7.1) can be thought of as providing the following function:

$$M(s, \theta; \mu) = \int_{-\infty}^{\infty} DH(x_h(t)) \cdot g(x_h(t), t + s + \theta; \mu, \varepsilon = 0) dt \qquad (2.7.2)$$

(see Figure 2.7.1(a) for s and θ). The same arguments that lead to the expression for average flux (2.4.9) with the quasiperiodic Melnikov function (2.2.6) gives expression (2.4.9) with the Melnikov function (2.7.2) substituted:

$$\begin{aligned} \Phi &= \lim_{T \to \infty} \frac{\varepsilon}{2T} \int_0^T |M(s,\theta;\mu)| ds + O(\varepsilon^2) \\ &= \lim_{T \to \infty} \frac{\varepsilon}{2T} \int_0^T |M(s,\theta;\mu)| d\theta + O(\varepsilon^2), \end{aligned} \tag{2.7.3}$$

(instead of using equation (2.4.7), one uses $M(s,\theta;\mu) = M(s-\tau,\theta+\tau;\mu)$, $\tau \in \mathbb{R}$, and then again employs a simple change of variables). Thus the expression for average flux extends trivially. Note that, from equation (2.7.2), the first and second expressions of equation (2.7.3) will be independent of θ and s, respectively. In addition, the geometrical constraints of the two-dimensional lobes in any time slice are identical to those of the time-periodic case (refer back to Section 2.6), so the transport equations of Sections 2.4 and 2.6 are identical under more general time dependences. As we see, then, the ideas for quasiperiodic perturbations are robust and extend easily to more general time dependences. For the more general perturbations, however, the global picture of the dynamics may be less apparent than in the quasiperiodic case, especially the two-frequency case, since the θ variable is no longer compact. For instance, to find the turnstile lobes in each time slice one will truly have to deal with an infinite domain $\theta \in \mathbb{R}$, and one will not be able to give a global portrayal of the choice of τ_c and the invariant core boundary. Thus, because of their simplicity and geometric appeal, we wish to stress two-frequency systems as workable and illuminating examples of chaotic transport under a bi-infinite sequence of nonrepeating maps.

Appendices

2.A1 A quasiperiodically oscillating vortex pair flow and a quasiperiodiclly forced Duffing oscillator

We define two dynamical systems that are referred to throughout the chapter, and derive their generalized Melnikov functions and relative scaling functions.

2.A1.1 An oscillating vortex pair (OVP) flow

Consider the quasiperiodic generalization of the oscillating vortex pair flow studied by Rom-Kedar et al. [1990]. As mentioned in the introduction, this two-dimensional fluid flow consists of a pair of point vortices of equal and opposite strength $\pm\Gamma$ in the presence of an oscillating strain-rate field which perturbs the vortex motion and which shall be referred to as the forcing term (even though it is understood that the net perturbation of the fluid flow is a sum of this forcing term and the effects of the vortex response). The streamfunction of the quasiperiodic forcing case is, in the comoving frame,

$$\psi(x_1,x_2,t) = \frac{-\Gamma}{4\pi}\ln\left[\frac{(x_1-x_1^v(t))^2+(x_2-x_2^v(t))^2}{(x_1-x_1^v(t))^2+(x_2+x_2^v(t))^2}\right] - V_v x_2 + \varepsilon x_1 x_2\left(\sum_{i=1}^{\ell}\omega_i f_i \sin(\omega_i t+\theta_{i_0})\right),$$

where $(x_1^v(t), \pm x_2^v(t))$ are the vortex positions, V_v is the average velocity of the vortex pair in the lab frame, and $\varepsilon\omega_i f_i$ is the strain rate amplitude associated with the i^{th} frequency (εf_i is non-dimensional). For $\varepsilon = 0$, $(x_1^v, x_2^v) = (0, d)$ and $V_v = \Gamma/4\pi d$. Let us specialize to the two-frequency case and use the dimensionless variables $x_1/d \to x_1$, $x_2/d \to x_2$, $\Gamma t/2\pi d^2 \to t$, $V_v 2\pi d/\Gamma \to v_v$, $2\pi\omega_i d^2/\Gamma \to \omega_i$ ($i =$

1, 2). The flow is then specified by

$$\dot{x}_1 = -\left[\frac{x_2 - x_2^v(t)}{(x_1 - x_1^v(t))^2 + (x_2 - x_2^v(t))^2} - \frac{x_2 + x_2^v(t)}{(x_1 - x_1^v(t))^2 + (x_2 + x_2^v(t))^2}\right]$$

$$- v_v + \varepsilon x_1 \{\omega_1 f_1 \sin(\omega_1 t + \theta_{1_0}) + \omega_2 f_2 \sin(\omega_2 t + \theta_{2_0})\}$$

$$\dot{x}_2 = (x_1 - x_1^v(t))\left[\frac{1}{(x_1 - x_1^v(t))^2 + (x_2 - x_2^v(t))^2} - \frac{1}{(x_1 - x_1^v(t))^2 + (x_2 + x_2^v(t))^2}\right]$$

$$- \varepsilon x_2 \{\omega_1 f_1 \sin(\omega_1 t + \theta_{1_0}) + \omega_2 f_2 \sin(\omega_2 t + \theta_{2_0})\}, \tag{2.A1.1}$$

where

$$x_1^v(t) = e^{-\varepsilon\left(f_1 \cos(\omega_1 t + \theta_{1_0}) + f_2 \cos(\omega_2 t + \theta_{2_0}) - f\right)} \cdot \int \left(\frac{1}{2} - v_v e^{\varepsilon\left(f_1 \cos(\omega_1 t + \theta_{1_0}) + f_2 \cos(\omega_2 t + \theta_{2_0}) - f\right)}\right) dt$$

$$x_2^v(t) = e^{\varepsilon\left(f_1 \cos(\omega_1 t + \theta_{1_0}) + f_2 \cos(\omega_2 t + \theta_{2_0}) - f\right)},$$

and v_v is chosen so that

$$\lim_{T \to \infty} \int_0^T \left(\frac{1}{2} - v_v e^{\varepsilon\left(f_1 \cos(\omega_1 t + \theta_{1_0}) + f_2 \cos(\omega_2 t + \theta_{2_0}) - f\right)}\right) dt = 0.$$

The vortex solutions $(x_1^v(t), \pm x_2^v(t))$ are found using that each vortex is advected by the flow due to the other vortex and the oscillating strain-rate field, so the vortices satisfy

$$\frac{dx_1^v}{dt} = \frac{1}{2x_2^v} - v_v + \varepsilon x_1^v \{\omega_1 f_1 \sin(\omega_1 t + \theta_{1_0}) + \omega_2 f_2 \sin(\omega_2 t + \theta_{2_0})\}$$
$$\frac{dx_2^v}{dt} = -\varepsilon x_2^v \{\omega_1 f_1 \sin(\omega_1 t + \theta_{1_0}) + \omega_2 f_2 \sin(\omega_2 t + \theta_{2_0})\}.$$

Rom-Kedar et al. [1990] choose, for the single-frequency analysis with $\theta_{1_0} = 0$, the initial conditions $(x_1^v(0), x_2^v(0)) = (0, 1)$, which guarantees a vortex response symmetric about $x_1 = 0$ to first order in ε. A simple quasiperiodic generalization will

need to keep this symmetry, which is done by choosing the constant of integration in the x_1^v expression to be zero. The x_2^v behavior is determined uniquely by the choice of f: a choice of $f = f_1 \cos\theta_{1_0} + f_2 \cos\theta_{2_0}$ or $f = f_1 + f_2$ follows the spirit of Rom-Kedar et al. [1990]; alternatively, one could choose $f = 0$ to obtain a perturbation of mean zero in the x_2 direction (not necessary, as is the case for x_1, but nevertheless appealing). We choose $f = f_1 + f_2$ for our simulations. The periodic forcing simulation shown in the introduction corresponds to equations (2.A1.1) with $f_2 = 0$, $\varepsilon f_1 = 0.1$, $\omega_1 = 2.5$, $\theta_{1_0} = 0$, $\varepsilon f = 0.1$.

The equations for x_1 and x_2 in (2.A1.1) define a two-dimensional nonautonomous dynamical system, and the motion of (x_1, x_2) for a given $\varepsilon, f_1, f_2, \omega_1, \omega_2$, $\theta_{1_0}, \theta_{2_0}$ and choice of initial vortex conditions is the fluid flow whose transport we address. The net perturbation is a sum of the forcing term, linear in ε, and the vortex response, nonlinear in ε; by Taylor expanding the vortex term about $\varepsilon = 0$, the governing equations in the single-frequency case can be put in the autonomous form

$$\begin{aligned} \dot{x} &= JDH(x) + \varepsilon g^{per}(x, \theta_1; \omega_1) + 0(\varepsilon^2) \\ \dot{\theta}_1 &= \omega_1, \end{aligned} \tag{2.A1.2}$$

(see Rom-Kedar et al. [1990]), and in the two-frequency case they will have the form

$$\begin{aligned} \dot{x} &= JDH(x) + \varepsilon g^{qp}(x, \theta_1, \theta_2; f_1, f_2, \omega_1, \omega_2) + 0(\varepsilon^2) \\ \dot{\theta}_1 &= \omega_1 \\ \dot{\theta}_2 &= \omega_2, \end{aligned} \tag{2.A1.3}$$

where $x = (x_1, x_2)$. The quasiperiodic forcing term, and thus the first order expansion term in the vortex response, is a superposition of two periodic forcing terms. Hence

$$g^{qp}(x, \theta_1, \theta_2; f_1, f_2, \omega_1, \omega_2) = f_1 g^{per}(x, \theta_1; \omega_1) + f_2 g^{per}(x, \theta_2; \omega_2). \tag{2.A1.4}$$

Figure 2.A1.1(a) shows some level sets in $\Sigma^{\theta_{2_0}}$ of the unperturbed ($\varepsilon = 0$) autonomous system (2.A1.3).

The Melnikov function for the single-frequency case (2.A1.2),

$$M^{per}(s,\theta_1;\omega_1) \equiv \int_{-\infty}^{\infty} DH(x_h(t))\cdot g^{per}(x_h(t),\, \omega_1 t + (\omega_1 s + \theta_1);\, \omega_1)\, dt, \quad (2.A1.5)$$

is, for $x_h(t)$ equal to the upper unperturbed heteroclinic orbit in X,

$$M^{per}(s,\theta_1\,;\,\omega_1) = \omega_1 F(\omega_1^{-1})\sin(\omega_1 s + \theta_1)\,, \quad (2.A1.6)$$

where $F(\omega_1^{-1})$ is shown in Figure 2.A1.2 (see Rom-Kedar et al. [1990]). From Section 2.2, the two-frequency Melnikov function is

$$M^{qp}(s,\theta_1,\theta_2;f_1,f_2,\omega_1,\omega_2) \equiv \int_{-\infty}^{\infty} DH(x_h(t))\cdot g^{qp}(x_h(t),\, \omega_1 t+\omega_1 s+\theta_1, \omega_2 t+\omega_2 s+\theta_2; f_1,f_2,\omega_1,\omega_2)\, dt,$$

which by (2.A1.4) and (2.A1.5) satisfies for the same $x_h(t)$ as above

$$M^{qp}(s,\theta_1,\theta_2;f_1,f_2,\omega_1,\omega_2) = f_1 M^{per}(s,\theta_1\,;\omega_1) + f_2 M^{per}(s,\theta_2\,;\omega_2)\,.$$

Hence by (2.A1.6)

$$\begin{aligned} &M^{qp}(s,\theta_1,\theta_2;f_1,f_2,\omega_1,\omega_2) \\ &\quad = f_1\omega_1 F(\omega_1^{-1})\sin(\omega_1 s+\theta_1) + f_2\omega_2 F(\omega_2^{-1})\sin(\omega_2 s+\theta_2)\,. \end{aligned} \quad (2.A1.7)$$

From Section 2.2 we refer to the ratio of each Melnikov function amplitude $f_i\omega_i F(\omega_i^{-1})$ to the corresponding relative perturbation amplitude $\omega_i f_i$ as the *relative scaling factor* associated with frequency ω_i. The fact that the relative scaling factors $F(\omega_i^{-1})$ are frequency dependent is pertinent to the study of transport rates, in particular to a comparison of average flux between single- and multiple-frequency forcing. Note how all scaling factors are determined by the single *relative scaling function* $F(\omega^{-1})$. M^{qp} is easily seen to have zero sets in all phase slices of $\Sigma^{\theta_{20}}$, and hence the upper and lower stable and unstable manifolds in $\Sigma^{\theta_{20}}$ of the perturbed flow form the boundary of the defined three-dimensional heteroclinic lobe structure (see Figure 2.A1.1(b)) and intersect each phase slice to give the boundary of a two-dimensional lobe structure. The physical situation can be pictured as the

two-dimensional fluid mapping from one time slice to the next in $\Sigma^{\theta_{20}}$ with each successive time sample. The lobe structure in $\Sigma^{\theta_{20}}$ thus constrains the motion of the fluid in the tangle region, as described in Section 2.3.

2.A1.2 The Duffing oscillator

In contrast to the OVP flow, consider the quasiperiodically forced Duffing oscillator:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= \beta x_1 - x_1^3 + \varepsilon \left[\sum_{i=1}^{\ell} f_i \cos(\omega_i t + \theta_{i_0}) - \gamma x_2 \right].\end{aligned} \tag{2.A1.8}$$

Consider again the two-frequency case $\ell = 2$. The physical system is now a point in X, rather than all of X (as in the fluid example). However, the chaotic nature of the oscillator's homoclinic tangle region (see Section 2.5) encourages that we still develop a global picture of the dynamics in X. Though the previous example involved an incompressible fluid and hence a Hamiltonian (*i.e.*, area-preserving) perturbation, there is now a dissipative term ($-\varepsilon\gamma x_2$) in the perturbing vector field. The lobe structure of Section 2.2 did not assume Hamiltonian perturbations, so we can have a lobe structure upon which to discuss transport even in non-Hamiltonian systems.

The two-frequency autonomous system (with $\beta = 1$) is

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_1 \left(1 - x_1^2\right) + \varepsilon[f_1 \cos\theta_1 + f_2 \cos\theta_2 - \gamma x_2] \\ \dot{\theta}_1 &= \omega_1 \\ \dot{\theta}_2 &= \omega_2.\end{aligned}$$

The generalized Melnikov function is

$$M\left(s, \theta_1, \theta_2; f_1, f_2, \omega_1, \omega_2, \gamma\right) =$$
$$\int_{-\infty}^{\infty} DH(x_h(t)) \cdot g^{qp}\left(x_h(t), \omega_1 t + (\omega_1 s + \theta_1), \omega_2 t + (\omega_2 s + \theta_2); f_1, f_2, \gamma\right) dt,$$

where

$$H(x_1, x_2) = \frac{x_2^2}{2} - \frac{x_1^2}{2} + \frac{x_1^4}{4}$$

and

$$g^{qp}(x_1, x_2, \theta_1, \theta_2; f_1, f_2, \gamma) \equiv (g_1, g_2) = (0, f_1 \cos\theta_1 + f_2 \cos\theta_2 - \gamma x_2)\,.$$

Figure 2.A1.3(a) shows in $\Sigma^{\theta_{2_0}}$ some level sets of the unperturbed system. Using that the x component of the unperturbed homoclinic orbits is

$$x_h(t) = (x_{1_h}(t), x_{2_h}(t)) = \pm\sqrt{2}\,\mathrm{sech}(t)(1, -\tanh(t)),$$

the Melnikov function is

$$M\,(s, \theta_1, \theta_2; f_1, f_2, \omega_1, \omega_2, \gamma) = -\frac{4\gamma}{3} \pm \sqrt{2}\,\pi f_1 \omega_1\,\mathrm{sech}\left(\frac{\pi\omega_1}{2}\right)\sin(\omega_1 s + \theta_1)$$
$$\pm \sqrt{2}\,\pi f_2 \omega_2\,\mathrm{sech}\left(\frac{\pi\omega_2}{2}\right)\sin(\omega_2 s + \theta_2). \quad (2.A1.9)$$

Figure 2.A1.4 shows a plot of the relative scaling function $\sqrt{2}\pi\omega\mathrm{sech}(\pi\omega/2)$ versus ω. Zero sets of the Melnikov function are not guaranteed here for $\varepsilon > 0$ because of the presence of the dissipative term; Ide and Wiggins [1989] study where in parameter space the zero sets do occur. For $\gamma = 0$ (the Hamiltonian case), zero sets are guaranteed for all phase slices in $\Sigma^{\theta_{2_0}}$ and hence the homoclinic lobe structure exists in this case in each phase slice (see Figure 2.A1.3(b)).

The OVP fluid and the nondissipative Duffing oscillator both have Melnikov functions in $\Sigma^{\theta_{2_0}=0}$ of the form

$$M(s, \theta_1, \theta_{2_0} = 0; \nu) = A_1(\mu, \omega_1)\,\sin(\omega_1 s + \theta_1) + A_2(\mu, \omega_2)\,\sin(\omega_2 s), \quad (2.A1.10)$$

where $A_1(\mu, \omega_1), A_2(\mu, \omega_2)$ are independent amplitudes. Since the Poincaré maps defined for any two cross sections $\Sigma^{\theta_{2_0}}$, $\Sigma^{\theta'_{2_0}}$ are topologically conjugate to one another, we choose for simplicity to set $\theta_{2_0} = 0$ in our transport examples.

2.A2 The persistence of zero set crossings in parametrized families

Let

$$\tilde{d}(s,\theta_1;A_1,\varepsilon) = \frac{M(s,\theta_1,\theta_{2_0}=0;A_1,\omega_1,\omega_2)}{||DH(x_h(-s))||} + O(\varepsilon),$$

where

$$M(s,\theta_1,\theta_{2_0}=0;A_1,\omega_1,\omega_2) = A_1\sin(\omega_1 s+\theta_1) + (2-A_1)\sin(\omega_2 s)$$

and the denominator is defined in Section 2.2. The conditions for the zero sets of the Melnikov function crossing at $(\bar{s},\bar{\theta}_1;\bar{A}_1)$ are

$$\begin{aligned} \tilde{d}(\bar{s},\bar{\theta}_1;\bar{A}_1,0) &= 0 \\ \frac{\partial\tilde{d}}{\partial s}(\bar{s},\bar{\theta}_1;\bar{A}_1,0) &= 0 \\ \frac{\partial\tilde{d}}{\partial\theta_1}(\bar{s},\bar{\theta}_1;\bar{A}_1,0) &= 0. \end{aligned} \qquad (2.A2.1)$$

Since $det\{D_{(s,\theta_1,A_1)}f(\bar{s},\bar{\theta}_1;\bar{A}_1,0)\} \neq 0$, where f denotes the three component vector that is the left hand side of equation (2.A2.1), the implicit function theorem implies that equations (2.A2.1) can be continued to non-zero ε. Hence the intersection manifolds cross within $0(\varepsilon)$ of $(\bar{s},\bar{\theta}_1)$ for some $A_1 = \bar{A}_1 + 0(\varepsilon)$.

Further, with the systems defined in Appendix 2.A1, for each crossing of the $M=0$ sets, the intersection manifolds will cross for the same value of A_1. Using the symmetries of the full equations, one can easily establish that if the intersection manifolds cross within $0(\varepsilon)$ of $(\bar{s},\bar{\theta}_1)$ at some value of A_1, they will also cross within $0(\varepsilon)$ of $(-\bar{s},-\bar{\theta}_1)$ at the same value of A_1. Applying $P_\varepsilon^{\pm n}$ to these cross points implies that for each crossing of the $M=0$ sets, the intersection manifolds will cross for the same value of A_1.

2.A3 Proofs for Section 2.2

Lemma 2.2.3

Suppose there exists an $(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1})$ for some $\bar{\nu}$ such that

$$\begin{aligned} M(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1}, \theta_{\ell_0}; \bar{\nu}) &= 0 \\ \frac{\partial M}{\partial s}(\bar{s}, \bar{\theta}_1, \ldots, \bar{\theta}_{\ell-1}, \theta_{\ell_0}; \bar{\nu}) &\neq 0. \end{aligned} \tag{2.A3.1}$$

Theorem 2.2.2 implies that there exists a $\Delta\theta \in T^{\ell-1}$ such that defined on the interval $(\bar{\theta} - \Delta\theta, \bar{\theta} + \Delta\theta)$ there is a C^r function $s(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ that parametrizes a segment of an $(\ell - 1)$-dimensional PIM. From the definition of the generalized Melnikov function (2.2.6) it immediately follows that

$$\begin{aligned} M(\bar{s} + \frac{2\pi}{\omega_\ell} n, \bar{\theta}_1 - 2\pi\frac{\omega_1}{\omega_\ell} n, \ldots, \bar{\theta}_{\ell-1} - 2\pi\frac{\omega_{\ell-1}}{\omega_\ell} n, \theta_{\ell_0}; \bar{\nu}) &= 0 \\ \frac{\partial M}{\partial s}(\bar{s} + \frac{2\pi}{\omega_\ell} n, \bar{\theta}_1 - 2\pi\frac{\omega_1}{\omega_\ell} n, \ldots, \bar{\theta}_{\ell-1} - 2\pi\frac{\omega_{\ell-1}}{\omega_\ell} n, \theta_{\ell_0}; \bar{\nu}) &\neq 0 \end{aligned} \qquad \forall n \in Z. \tag{2.A3.2}$$

When all the frequencies are mutually commensurate, $\bar{\theta} - 2\pi\frac{\omega}{\omega_\ell} n$, $n \in Z$, visits $\bar{\theta}$ identically a countable infinity of times, and hence by (2.A3.2) there are an infinity of s values among $\{\bar{s} + \frac{2\pi}{\omega_\ell} n\,;\, n \in Z\}$ that satisfy (2.A3.1). Hence, by Theorem 2.2.2 there exists a countable infinity of transversal intersection points of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ in the phase slice $\chi(\bar{\theta})$. When one or more pairs of frequencies are incommensurate, it is still true that for $\Delta\theta' \in T^{\ell-1}$ *arbitrarily small* there are a countable infinity of $n \in Z$ such that $\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell} n \in [\bar{\theta} - \Delta\theta', \bar{\theta} + \Delta\theta']$. By (2.A3.2) and the comments that precede it, for each such n there exists a C^r function $s_n(\theta_1, \ldots, \theta_{\ell-1}; \theta_{\ell_0})$ defined on the interval $[\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell} n - \Delta\theta, \bar{\theta} + 2\pi\frac{\omega}{\omega_\ell} n + \Delta\theta]$ that parametrizes a segment of an $(\ell - 1)$-dimensional PIM. As long as we choose $\Delta\theta' < \Delta\theta$ (by this we mean each component of $\Delta\theta'$ is smaller than the corresponding component of $\Delta\theta$), the interval $[\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell} n - \Delta\theta, \bar{\theta} + 2\pi\frac{\omega}{\omega_\ell} n + \Delta\theta]$ contains $\bar{\theta}$, and hence $\chi(\bar{\theta})$ contains a countable infinity of transversal intersection points of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ in the phase slice $\chi(\bar{\theta})$.

Lemma 2.2.4

This follows immediately from the invariance of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ under P_ε. An intersection of $W^s(\tau_\varepsilon^a)$ and $W^u(\tau_\varepsilon^b)$ remains an intersection under $P_\varepsilon^{\pm 1}$. Hence, if there exists a countable infinity of intersection points in the phase slice $\chi(\bar{\theta})$, there will be a countable infinity of intersection points in the phase slice $\chi(\bar{\theta} + 2\pi\frac{\omega}{\omega_\ell}n)$ for all $n \in Z$. If all frequencies are mutually incommensurate, then $\{\bar{\theta}+2\pi\frac{\omega}{\omega_\ell}n; n \in Z\}$ is dense in $T^{\ell-1}$, which implies that there is a countable infinity of intersection points in all phase slices of $\Sigma^{\theta_{\ell_0}}$. If one or more pairs of frequencies are commensurate then $\bar{\theta} - 2\pi\frac{\omega}{\omega_\ell}n$ does *not* in general visit $T^{\ell-1}$ densely, and hence we can only conclude that there exists a countable infinity of intersection points in all phase slices defined either over all of $T^{\ell-1}$ or some subset of $T^{\ell-1}$.

Theorem 2.2.5 follows immediately from the lemmas.

2.A4 Computing lobe structures using a double phase slice sampling method

Suppose we wish to portray lobe boundaries in X at time $t = \frac{2\pi}{\omega_\ell}n$, or, equivalently, the invariant lobe boundaries in the time slice $\chi(\theta_0 + 2\pi\frac{\omega}{\omega_\ell}n)$. For simplicity of discussion, let us consider the two-frequency homoclinic case. Evolving the curve

$$W^u_{loc}(\tau_\varepsilon) \cap \chi\big(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n-i)\big)$$

forward in time according to (2.2.3) for i sample periods (where i is some positive integer) gives a curve which extends from

$$\tau_\varepsilon \cap \chi\big(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n\big)$$

along a finite length of

$$W^u(\tau_\varepsilon) \cap \chi\big(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n\big)$$

(the greater i is, of course, the greater the length will be). Similarly, evolving the curve

$$W^s_{loc}(\tau_\varepsilon) \cap \chi\big(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2}(n+i)\big)$$

backwards in time according to (2.2.3) for i sample periods gives a curve which extends from

$$\tau_\varepsilon \cap \chi\big(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n\big)$$

along a finite length of

$$W^s(\tau_\varepsilon) \cap \chi\big(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n\big)$$

(see Figure 2.A4.1). The resulting two curves form the boundary of a finite number of lobes of the two-dimensional lobe structure in the time slice $\chi\big(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n\big)$.

We still need to be able to find $W^u_{loc}(\tau_\varepsilon)$ and $W^s_{loc}(\tau_\varepsilon)$ in the appropriate slices. The procedure for doing this is a straightforward generalization of the standard trial and error procedure used in the time-periodic case, so our discussion here will only be heuristic. Suppose for some arbitrary phase slice $\chi(\bar{\theta}_1)$ of $\Sigma^{\theta_{2_0}}$ we wish to find $\tau_\varepsilon \cap \chi(\bar{\theta}_1)$ and one or both of $W^u_{loc}(\tau_\varepsilon) \cap \chi(\bar{\theta}_1)$ and $W^s_{loc}(\tau_\varepsilon) \cap \chi(\bar{\theta}_1)$. Let $C_\pm$ be a closed curve in the phase slice $\chi\left(\bar{\theta}_1 \pm 2\pi \frac{\omega_1}{\omega_2} j\right)$ that contains the point $\tau_\varepsilon \cap \chi\left(\bar{\theta}_1 \pm 2\pi \frac{\omega_1}{\omega_2} j\right)$ and is pierced by $W^u_{loc}(\tau_\varepsilon)$ and $W^s_{loc}(\tau_\varepsilon)$, where j is some positive integer (see Figure 2.A4.2). Due to the normal hyperbolicity of τ_ε, $P^j_\varepsilon(C_-)$ will be stretched along $W^u(\tau_\varepsilon) \cap \chi(\bar{\theta}_1)$ and $P^{-j}_\varepsilon(C_+)$ will be stretched along $W^s(\tau_\varepsilon) \cap \chi(\bar{\theta}_1)$. The region in $\chi(\bar{\theta}_1)$ bounded by $P^j_\varepsilon(C_-)$ will intersect with the region in $\chi(\bar{\theta}_1)$ bounded by $P^{-j}_\varepsilon(C_+)$ in one or more disjoint regions, one of which, $\mathcal{Q}(\bar{\theta}_1)$, will contain $\tau_\varepsilon \cap \chi(\bar{\theta}_1)$ and shrink to zero area as $j \to \infty$ (again see Figure 2.A4.2). Thus, the trial and error procedure, similar in spirit to the time-periodic case, is to make reasonable tries at $C_\pm$ and evolve them to $\chi(\bar{\theta}_1)$ as described above to pinpoint $\tau_\varepsilon \cap \chi(\bar{\theta}_1)$ and its local stable and unstable manifolds by watching how the two curves stretch and intersect. By choosing $C_\pm$ sufficiently small, we can obtain an arbitrarily good approximation of $W^u_{loc}(\tau_\varepsilon) \cap \chi(\bar{\theta}_1)$ and $W^s_{loc}(\tau_\varepsilon) \cap \chi(\bar{\theta}_1)$. When the system is near-integrable, one can make a good initial estimate of $C_\pm$ due to the

knowledge of the location of the fixed point of the unperturbed system; otherwise one must employ a more arduous trial-and-error effort, which would be the case under time-periodic vector fields as well.

The procedure for simulating global stable and unstable manifolds in the n^{th} time slice of the autonomous system phase space is thus quite similar to the procedure for simulating the invariant lobe structure in the time-periodic case, except that one has to take into account that curves are, with each application of P_ϵ, mapped around T^1 in an enlarged phase space. Hence one finds the intersection of $W^u_{loc}(\tau_\epsilon)$ and $W^s_{loc}(\tau_\epsilon)$ with an appropriate pair of phase slices, which in turn are used to simulate a finite length of $W^u(\tau_\epsilon)$ and $W^s(\tau_\epsilon)$ in the desired phase slice. From our discussion of the two-frequency homoclinic case, the procedure for more frequencies and for the heteroclinic case should be clear. Note how the described procedure contrasts with a previous suggestion by Moon and Holmes [1985] for analyzing the dynamics under quasiperiodic vector fields, a double Poincaré map method which essentially attempts to treat the system as periodic. Samples of an equation like (2.2.2) are taken only when both $\theta_2 = \bar{\theta}_2$ and $\theta_1 \in [\bar{\theta}_1 - \beta, \bar{\theta}_1 + \beta]$ (for some choice of $\bar{\theta}_1, \bar{\theta}_2$ and $\beta \ll 2\pi$), and the results are summed. The time between samples can be much longer with this approach, and there is a "fuzziness" of the resulting structure due to the finite width of the sampling window, 2β. In contrast to this method, we shall refer to our approach as a *double phase slice method*, since to simulate the lobe structure in any phase slice one evolves two curves, each originating in a different phase slice.

2.A5 Derivation of equation (2.4.8) from equations (2.4.5) and (2.4.7)

From equation (2.4.5):

$$\Phi(\theta_0) = \frac{\omega_\ell}{2\pi}\frac{\varepsilon}{2N}\sum_{n=0}^{N-1} I_n + O(\varepsilon^2) \tag{2.A5.1}$$

where

$$I_n = \int_{s_0(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)}^{s_{-1}(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)} |M(s, \theta_{1_0} + 2\pi\frac{\omega_1}{\omega_\ell}n, \ldots, \theta_{(\ell-1)_0} + 2\pi\frac{\omega_{\ell-1}}{\omega_\ell}n, \theta_{\ell_0}; \nu)|ds.$$

From equation (2.4.7)

$$I_n = \int_{s_0(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)}^{s_{-1}(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)} |M(s + \frac{2\pi}{\omega_\ell}n, \theta_{1_0}, \ldots, \theta_{(\ell-1)_0}, \theta_{\ell_0}; \nu)|ds.$$

Letting $s' = s + n\frac{2\pi}{\omega_\ell}$,

$$I_n = \int_{s_0(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)+\frac{2\pi}{\omega_\ell}n}^{s_{-1}(\theta_0+2\pi\frac{\omega}{\omega_\ell}n)+\frac{2\pi}{\omega_\ell}n} |M(s', \theta_{1_0}, \ldots, \theta_{(\ell-1)_0}, \theta_{\ell_0}; \nu)|ds',$$

which gives

$$I_n = \int_{s_{-n}(\theta_0)}^{s_{-(n+1)}(\theta_0)} |M(s', \theta_{1_0}, \ldots, \theta_{(\ell-1)_0}, \theta_{\ell_0}; \nu)|ds', \tag{2.A5.2}$$

where

$$s_{-j}(\theta_0) = s \text{ value of } P_\varepsilon^{-j}(\tau_c) \cap \chi(\theta_0) \quad \forall\, j \in Z.$$

Substituting (2.A5.2) into (2.A5.1) gives (2.4.8).

Remark 1: The proof of equation (2.4.9) for one or more incommensurate frequency ratios is the same as the above proof with $N \to \infty$.

Remark 2: Note that (2.A5.2) with $n = -1$ implies that for non-Hamiltonian perturbations the turnstile lobes and their images have the same area through $O(\varepsilon)$.

2.A6 Maximal average flux in a single-frequency limit

Consider the perturbation in (2.2.8) with $F_0(x;\mu) = 0$, $F_1 + F_2 = 1$, and the larger relative scaling factor at $F_1 = 1$. The average flux is given by

$$\Phi(\theta_{1_0}) = \frac{\varepsilon}{2} < |a_1 F_1 \sin(\omega_1 s + \theta_{1_0}) + a_2 F_2 \sin(\omega_2 s)| >_s, \qquad (2.A6.1)$$

where a_1 and a_2 denote the relative scaling factors associated with ω_1 and ω_2, respectively, and $< >_s$ denotes the average over s. We rewrite this as

$$\Phi(\theta_{1_0}) = \frac{\alpha}{2} < |\beta_1 \sin(\omega_1 s + \theta_{1_0}) + \beta_2 \sin(\omega_2 s)| >_s, \qquad (2.A6.2)$$

where

$$\alpha = \varepsilon(a_1 F_1 + a_2 F_2)$$
$$\beta_i = \frac{a_i F_i}{a_1 F_1 + a_2 F_2} \text{ for } i = 1, 2.$$

The new amplitudes β_i satisfy $\beta_1 + \beta_2 = 1$ and $(\beta_1(F_1 = 0), \beta_2(F_1 = 0)) = (0, 1)$, $(\beta_1(F_1 = 1), \beta_2(F_1 = 1)) = (1, 0)$; hence the expression $< | \cdot | >_s$ in (2.A6.2) has the profile of the equal relative scaling factor case, but skewed in F_1; it will thus still have absolute maxima at $F_1 = 0$ and 1. The term α is monotonic increasing in F_1, hence $\Phi(\theta_{1_0})$ will have an absolute maximum at $F_1 = 1$. Thus the single-frequency case with the larger relative scaling factor corresponds to an absolute maximum of average flux for the above class of systems.

References

D. Beigie, A. Leonard, and S. Wiggins [1991a]. A global study of enhanced stretching and diffusion in chaotic tangles. *Phys. Fluids A* **3**, 1039-1050.

D. Beigie, A. Leonard, and S. Wiggins [1991b]. Chaotic transport in the homoclinic and heteroclinic tangle regions of quasiperiodically forced two-dimensional dynamical systems. *Nonlinearity* **4**, 775-819.

D. Beigie, A. Leonard, and S. Wiggins [1992a]. The dynamics associated with the chaotic tangles of two-dimensional quasiperiodic vector fields: theory and applications, in **Nonlinear Phenomena in Atmospheric and Oceanic Sciences (IMA Volumes in Mathematics and its Applications - Volume 40)**. Editors G.F. Carnevale and R. Pierrehumbert (Springer-Verlag: New York, Berlin, Heidelberg) 47-138.

D. Beigie and S. Wiggins [1992b]. The dynamics associated with a quasiperiodically forced Morse oscillator: Application to molecular dissociation. *Phys. Rev. A* **45**, 4803-4827.

D. Bensimon and L.P. Kadanoff [1984]. Extended Chaos and Disappearance of KAM Trajectories. *Physica* **13D**, 82-89.

R. Camassa and S. Wiggins [1991]. Chaotic advection in a Rayleigh-Bénard flow. *Phys. Rev. A* **43**, 774 -797.

M.J. Davis [1985]. Bottlenecks to intramolecular energy transfer and the calculation of relaxation rates. *J. Chem. Phys.* , **83**, 1016-1031.

R. W. Easton [1986]. Trellises Formed by Stable and Unstable Manifolds in the Plane. *Trans. Am. Math. Soc.,* **294**, 719-732.

N. Fenichel [1971]. Persistence and Smoothness of Invariant Manifolds for Flows. *Ind. Univ. Math J.,* **21**, 193-226.

M.E. Goggin and P.W. Milonni [1988a]. Driven Morse oscillator: Classical chaos,

quantum theory, and photodissociation. *Phys. Rev. A* **37**, 796-806.

M.E. Goggin and P.W. Milonni [1988b]. Driven Morse oscillator: Classical chaos and quantum theory for two-frequency excitation. *Phys. Rev. A* **38**, 5174-5181.

J. Guckenheimer and P. Holmes [1983]. **Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.** Springer-Verlag: New York, Heidelberg, Berlin.

M.W. Hirsch, C.C. Pugh, and M. Shub [1977]. **Invariant Manifolds.** Springer Lecture Notes in Mathematics, Vol. 583. Springer-Verlag: New York, Heidelberg, Berlin.

K. Ide and S. Wiggins [1989]. The bifurcation to homoclinic tori in the quasiperiodically forced Duffing oscillator. *Physica* **34D**, 169-182.

A. Leonard, V. Rom-Kedar, and S. Wiggins [1987]. Fluid Mixing and Dynamical Systems. *Nuclear Physics B (Proc. Suppl.)*, **2**, 179-190.

R.S. MacKay, J.D. Meiss, and I.C. Percival [1984]. Transport in Hamiltonian Systems. *Physica* **13D**, 55-81.

R.S. MacKay and J.D. Meiss [1988]. Relation between quantum and classical thresholds for multiphoton ionization of excited atoms. *Phys. Rev. A* **37**, 4702-4706.

K.R. Meyer and G.R. Sell [1989]. Melnikov transforms, Bernoulli bundles, and almost periodic perturbations. *Trans. Am. Math. Soc.* **314**, 63-105.

F.C. Moon and W.T. Holmes [1985]. Double Poincaré sections of a quasi-periodically forced, chaotic attractor. *Phys. Lett.* **111A**, 157-160.

D.W. Noid and J.R. Stine [1979]. Infrared multiphoton dissociation with one and two lasers. *Chem. Phys. Lett.* **65**, 153-157.

J.M. Ottino [1990]. Mixing, Chaotic Advection, and Turbulence. *Annu. Rev. Fluid Mech.* **22**, 207-254.

V. Rom-Kedar, A. Leonard, and S. Wiggins [1990]. An analytical study of transport,

mixing and chaos in an unsteady vortical flow. *J. Fluid Mech.* **214**, 347-394.

V. Rom-Kedar and S. Wiggins [1990]. Transport in Two Dimensional Maps. *Arch. Rat. Mech. Anal.* **109** 3, 239-298.

J. Scheurle [1986]. Chaotic solutions of systems with almost periodic forcing. *J. Appl. Math. and Phys. (ZAMP)*, **37**, 12-26.

T.H. Solomon and J. P. Gollub [1988]. Chaotic particle transport in time-dependent Rayleigh-Bénard convection. *Phys. Rev. A* **38** 6280-6286.

D. Stoffer [1988a]. Transversal homoclinic points and hyperbolic sets for non-autonomous maps I. *J. Appl. Math. and Phys. (ZAMP)*, **39**, 518-549.

D. Stoffer [1988b]. Transversal homoclinic points and hyperbolic sets for non-autonomous maps II. *J. Appl. Math. and Phys. (ZAMP)*, **39**, 783-812.

S. Wiggins [1987]. Chaos in the quasiperiodically forced Duffing oscillator. *Phys. Lett.* **124A**, 138-142.

S. Wiggins [1988]. **Global Bifurcations and Chaos – Analytical Methods.** Springer-Verlag: New York, Heidelberg, Berlin.

S. Wiggins [1990]. On the Geometry of Transport in Phase Space, I. Transport in k-Degree-of-Freedom Hamiltonian Systems, $2 \leq k < \infty$. *Physica D* **44**, 471-501.

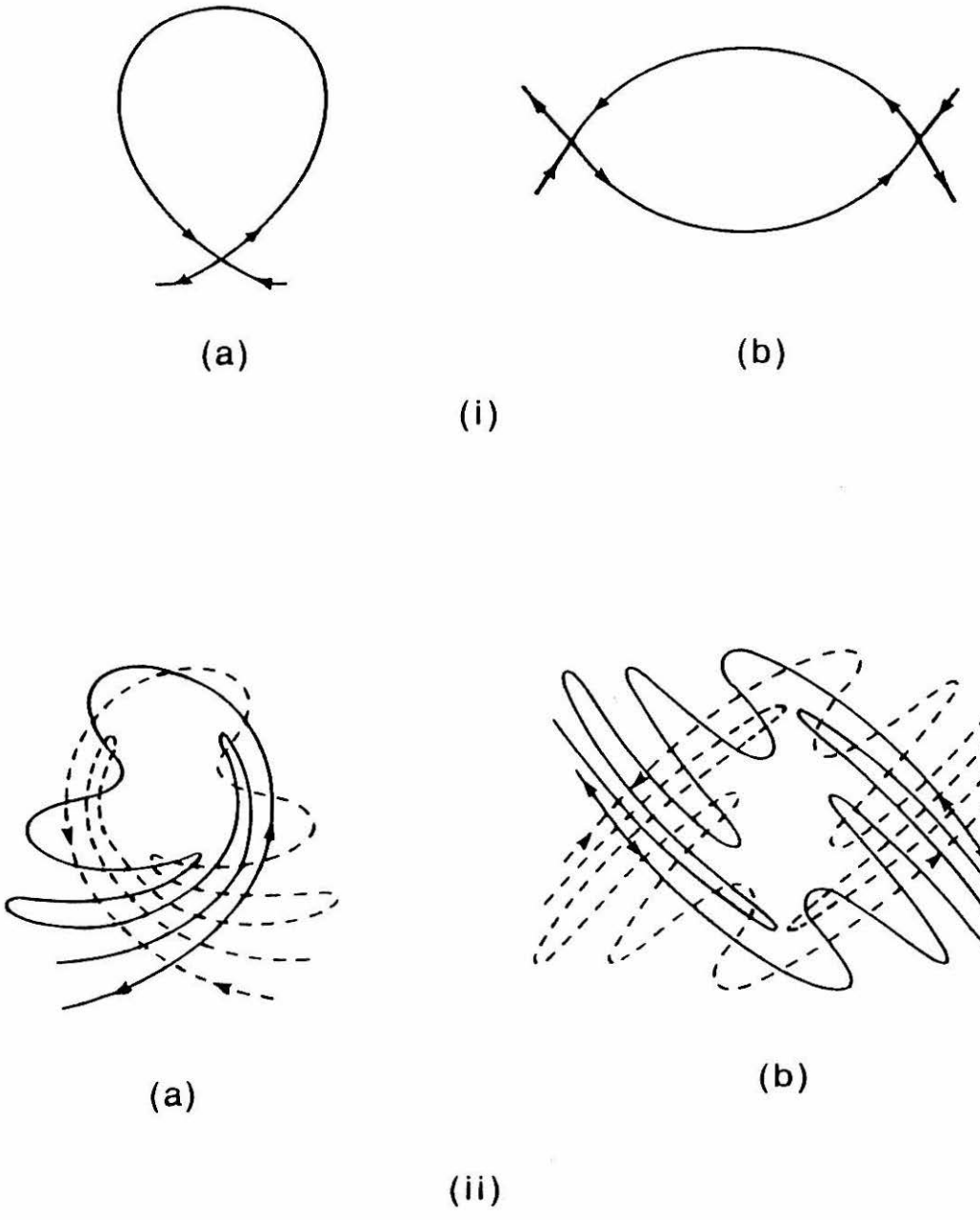


Figure 2.1.1 (i) (a) An unperturbed homoclinic orbit, and (b) a pair of unperturbed heteroclinic orbits. (ii) The tangle region of the perturbed system for (a) the homoclinic case, and (b) the heteroclinic case. The dashed and solid lines are stable and unstable manifolds, respectively, of the surviving hyperbolic fixed points.

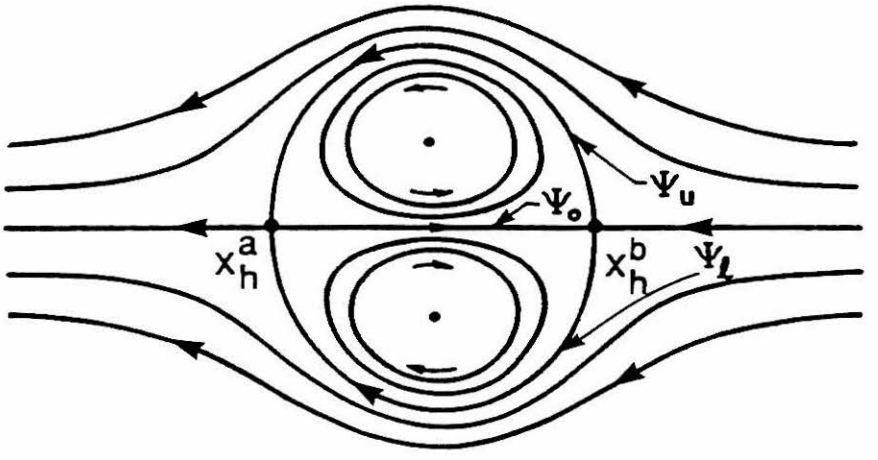


Figure 2.1.2 Streamlines in the comoving frame of the unperturbed OVP flow. The hyperbolic fixed points are marked by x_h^a, x_h^b; the heteroclinic orbits are marked by $\Psi_u, \Psi_o, \Psi_\ell$.

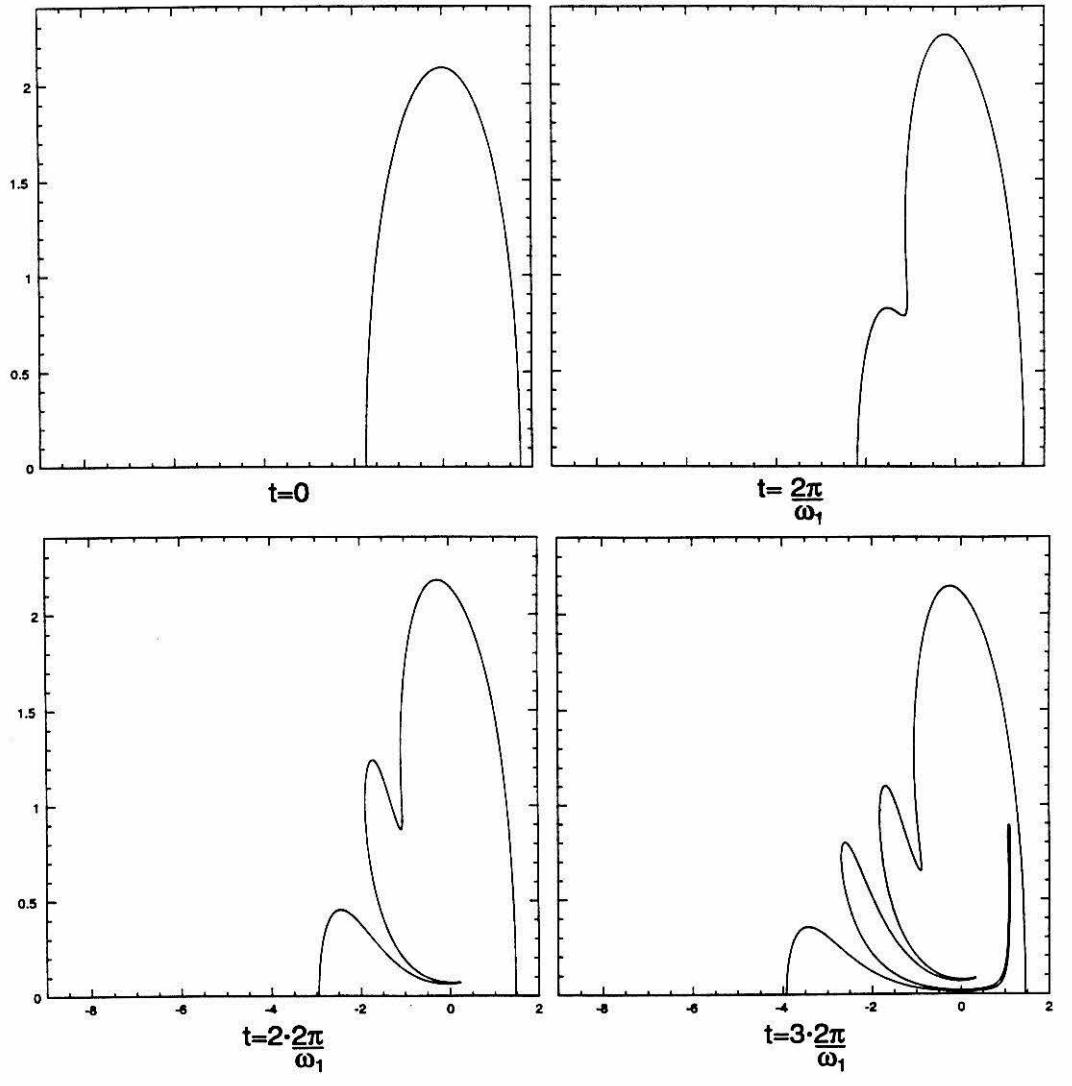


Figure 2.1.3 Evolution, according to the single-frequency OVP flow, of a material curve that initially lies on Ψ_u. The system parameters are given in Appendix 2.A1.

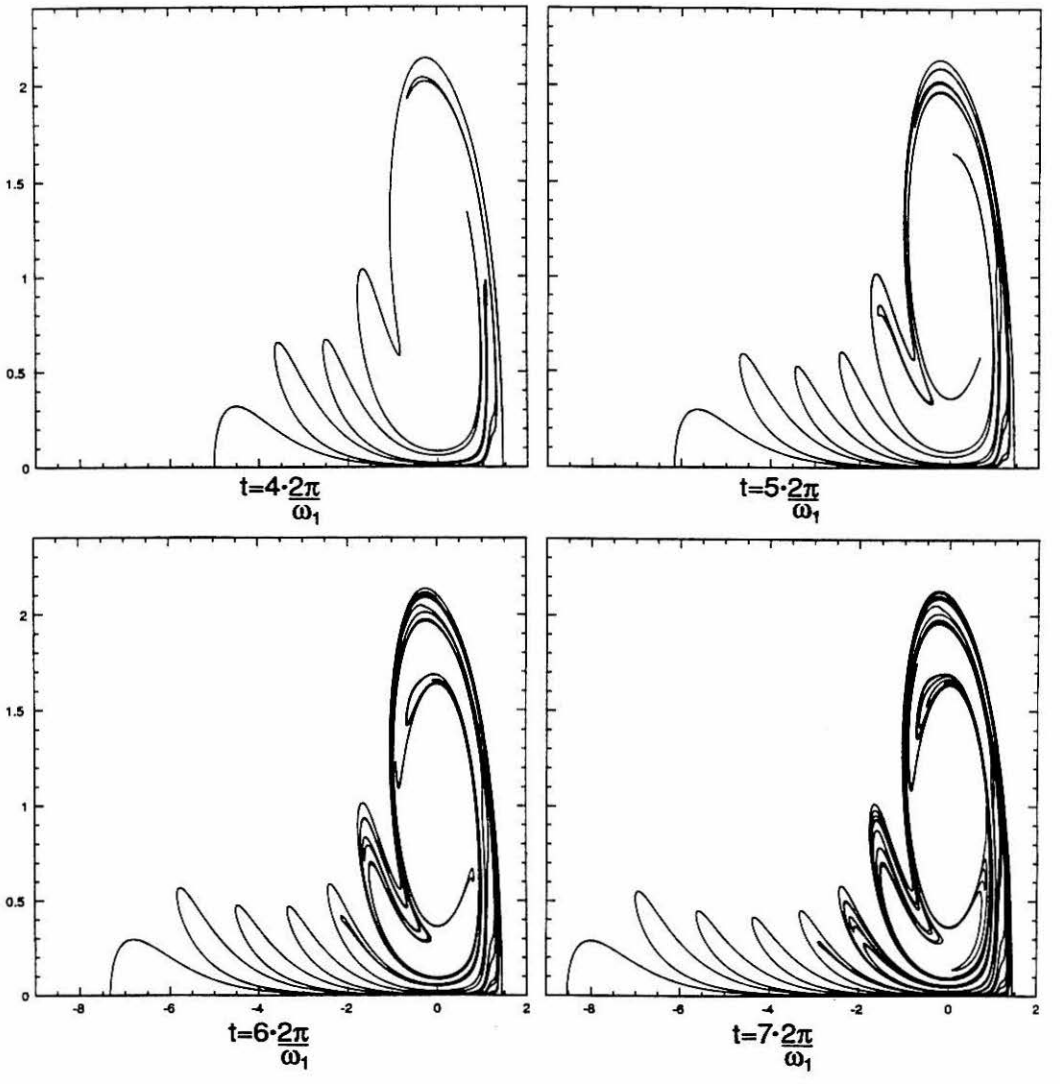


Figure 2.1.3 *Continued.*

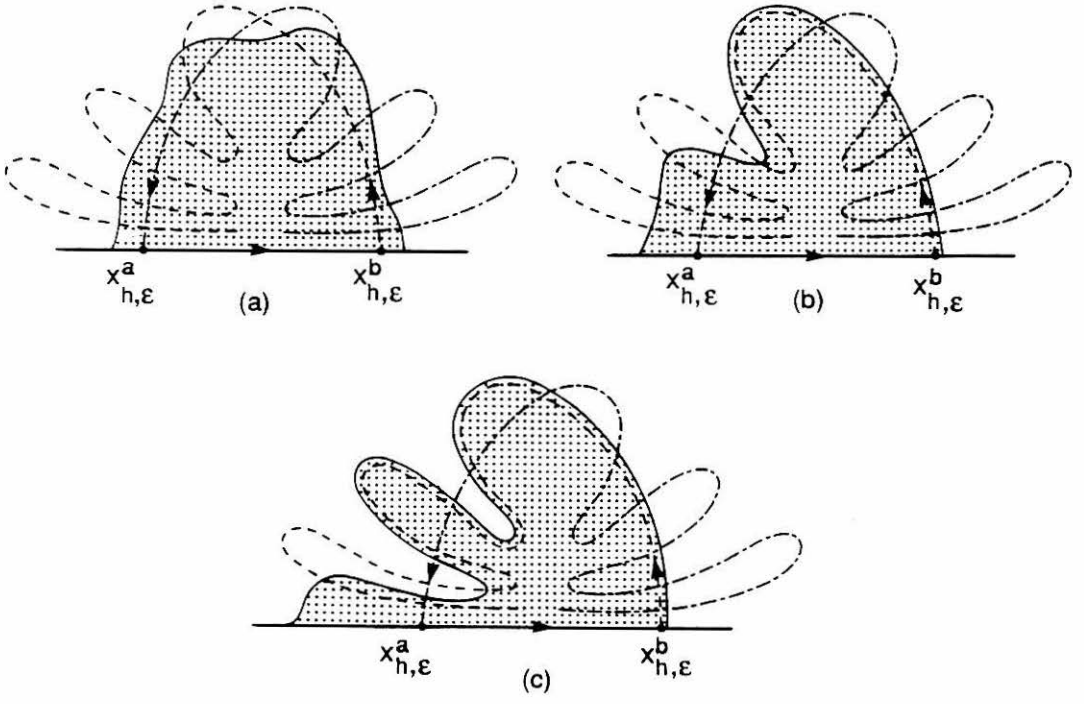
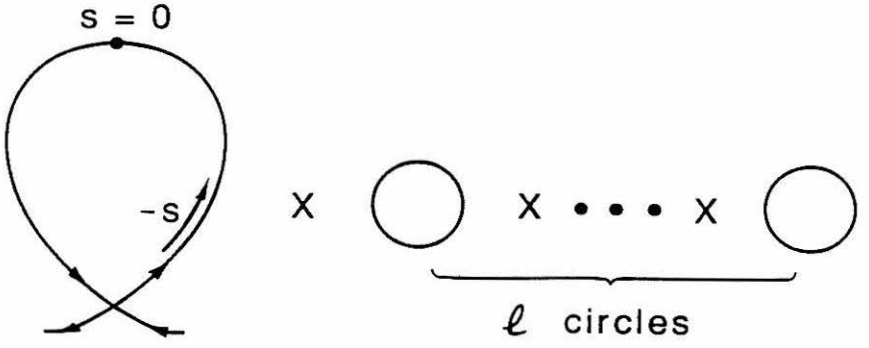
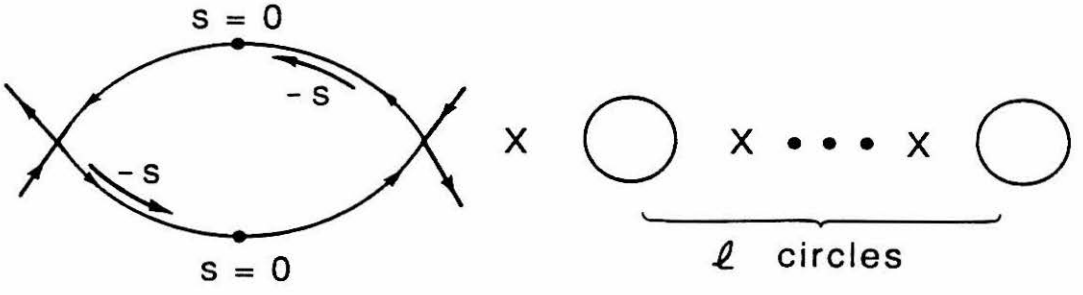


Figure 2.1.4 Three successive time samples showing a material curve being "attracted" to the unstable manifold. The dashed-dotted line is the stable manifold, the dashed line is the unstable manifold, and the solid line is the material curve. Section 2.2 will make clear the notation involving ε. A similar portrayal appears in Rom-Kedar et al. [1990].

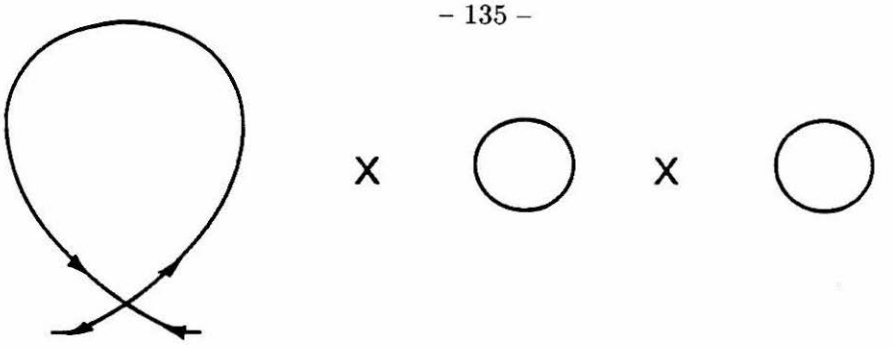


(a)

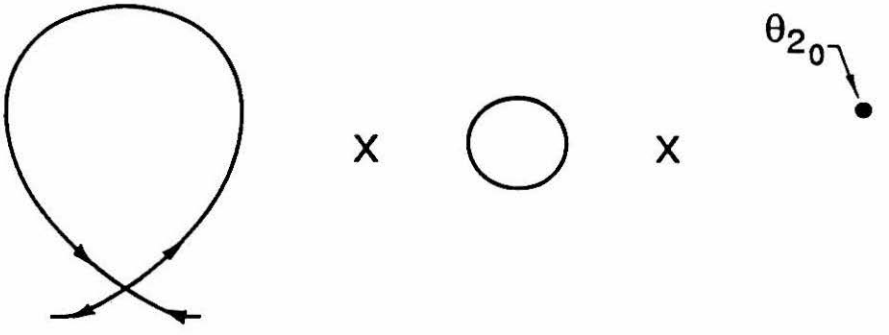


(b)

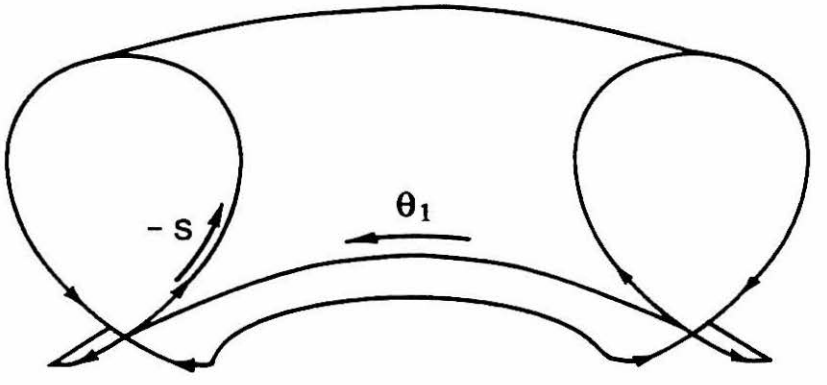
Figure 2.2.1 (a) A homoclinic manifold and (b) two heteroclinic manifolds in the autonomous system phase space of the ℓ-frequency case.



(a)



or



(b)

Figure 2.2.2 The $\ell = 2$ homoclinic manifold in (a) the full phase space and (b) the Poincaré section $\Sigma^{\theta_{2_0}}$ (the second portrayal shows a cut-away view).

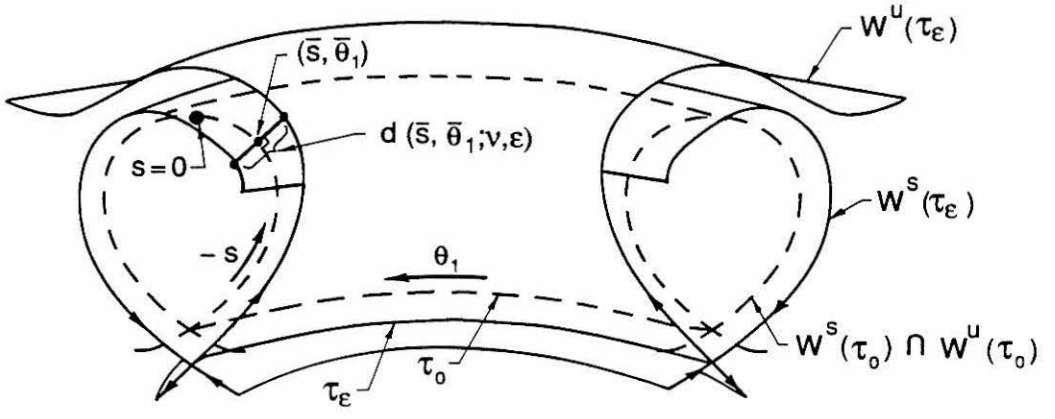


Figure 2.2.3 The signed distance function between the stable and unstable manifolds in $\Sigma^{\theta_{20}}$ for the perturbed two-frequency homoclinic system.

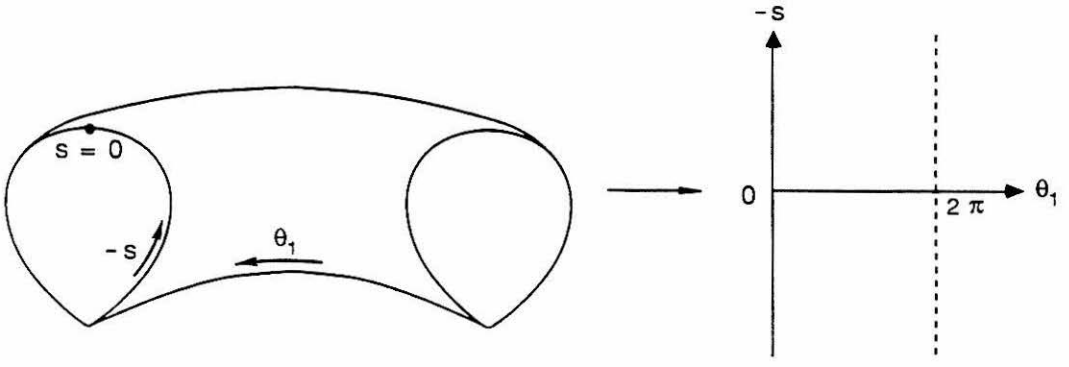


Figure 2.2.4 Flattening the coordinates of an unperturbed homoclinic orbit onto a plane.

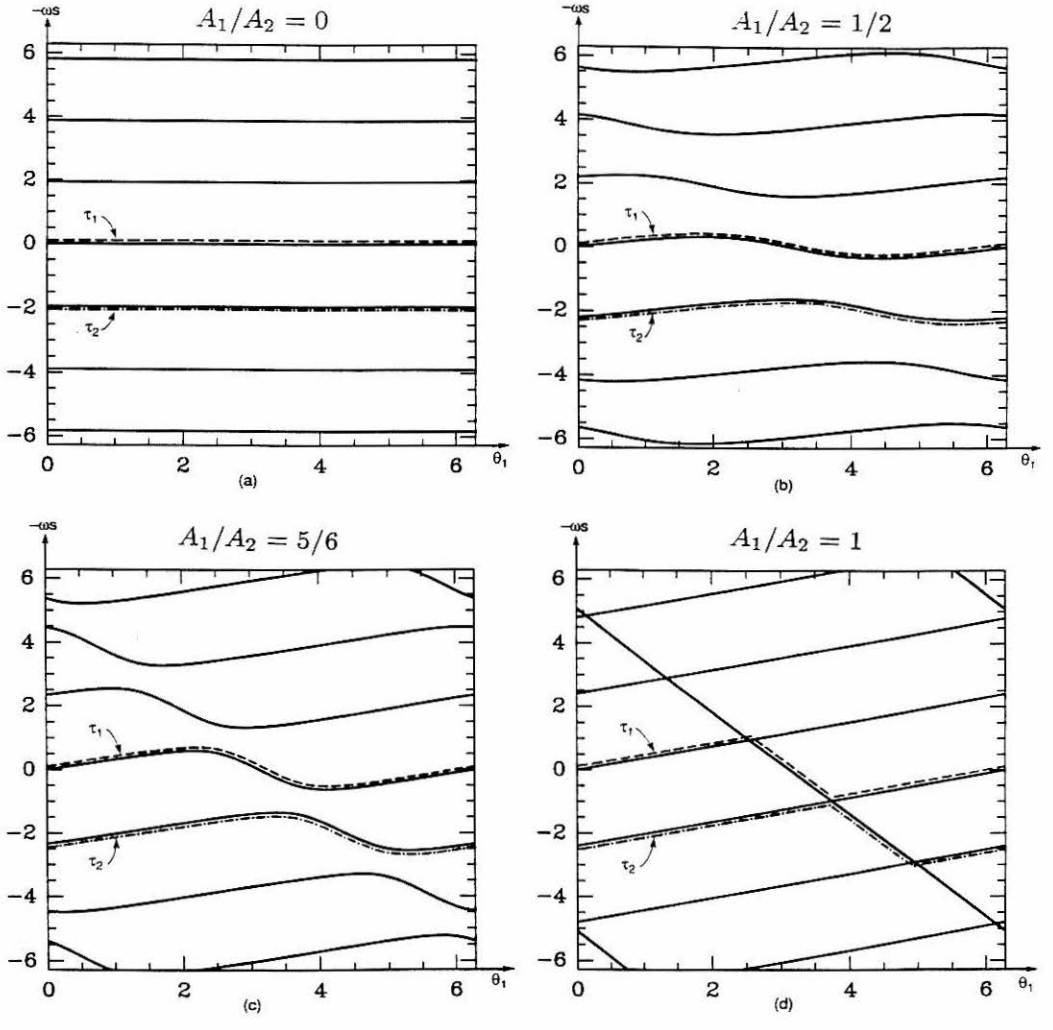


Figure 2.2.5 Some zero sets of the Melnikov function in equation (2.2.10), for a range of parameter values ($\omega_1 = g\omega_2 \equiv \omega$, except for (h) where $\omega_1 = \omega_2/2 \equiv \omega$). The dashed and dashed-dotted lines represent two choices of PIM's according to Definition 2.2.1.

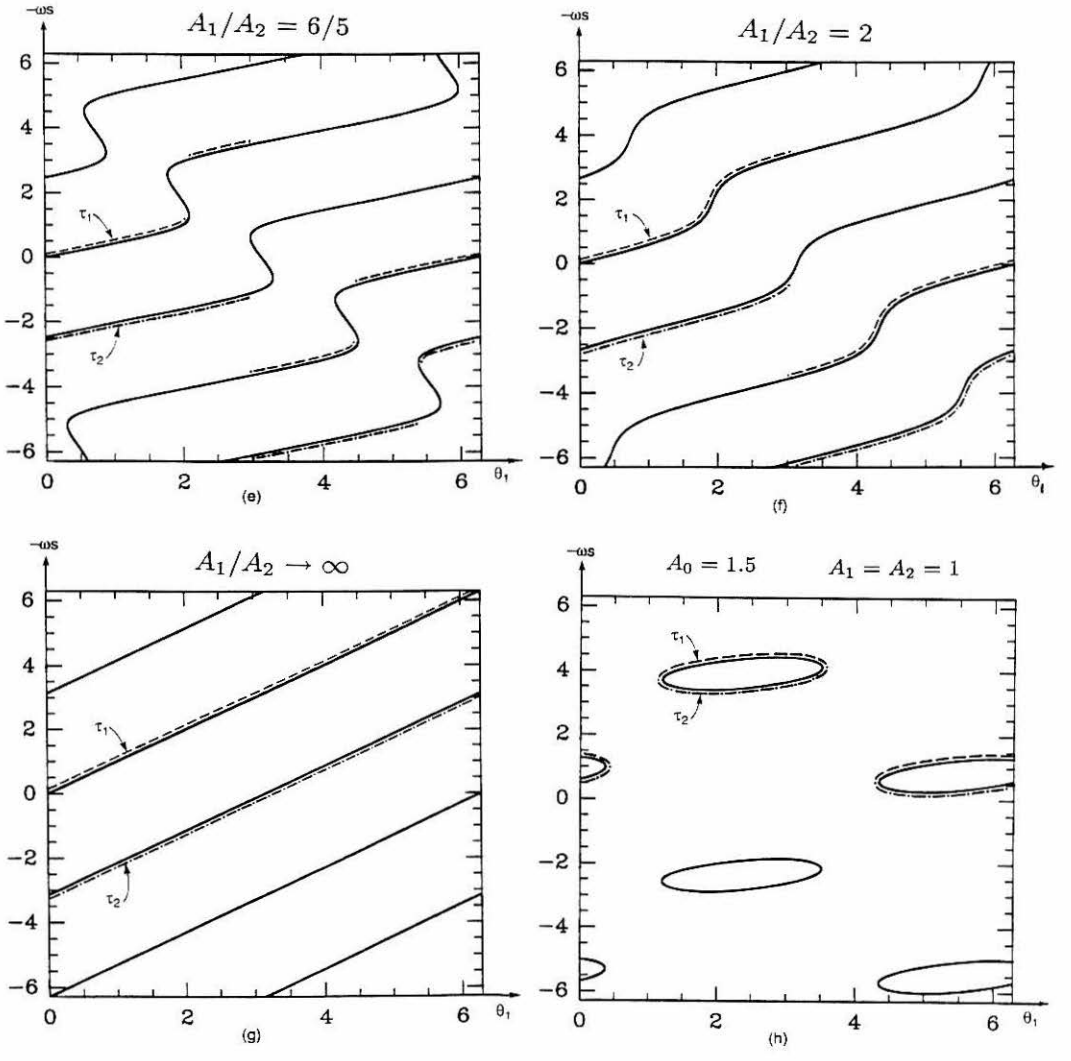


Figure 2.2.5 *Continued.*

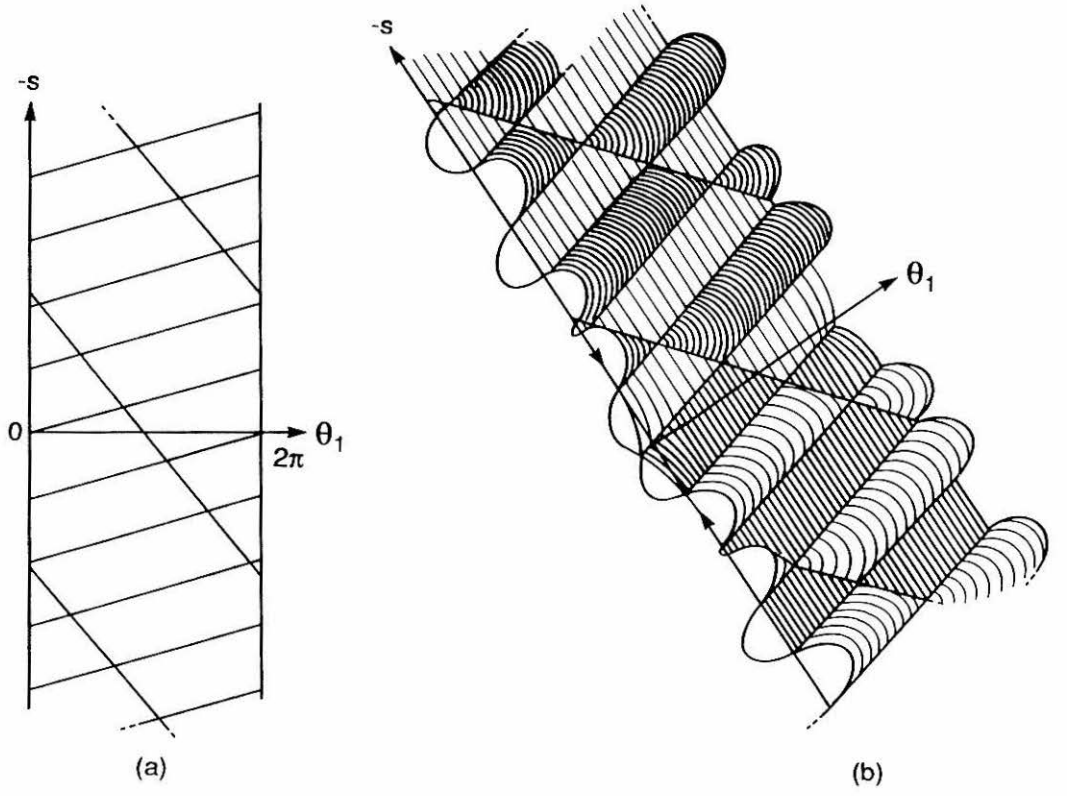


Figure 2.2.6 Visualizing lobes by showing the suppressed dimension.

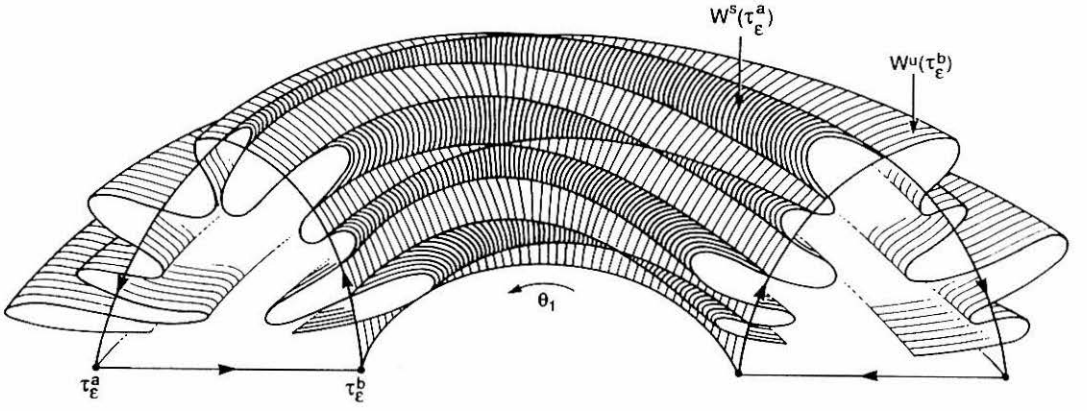


Figure 2.2.7 The boundaries of a three-dimensional invariant lobe structure in $\Sigma^{\theta_{20}}$ for a two-frequency heteroclinic case.

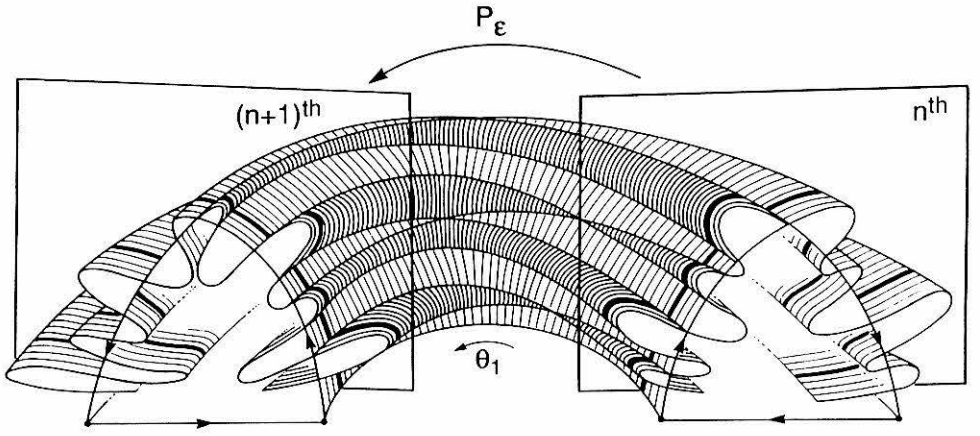


Figure 2.2.8 The n^{th} and $(n+1)^{th}$ time slice of $\Sigma^{\theta_{20}}$.

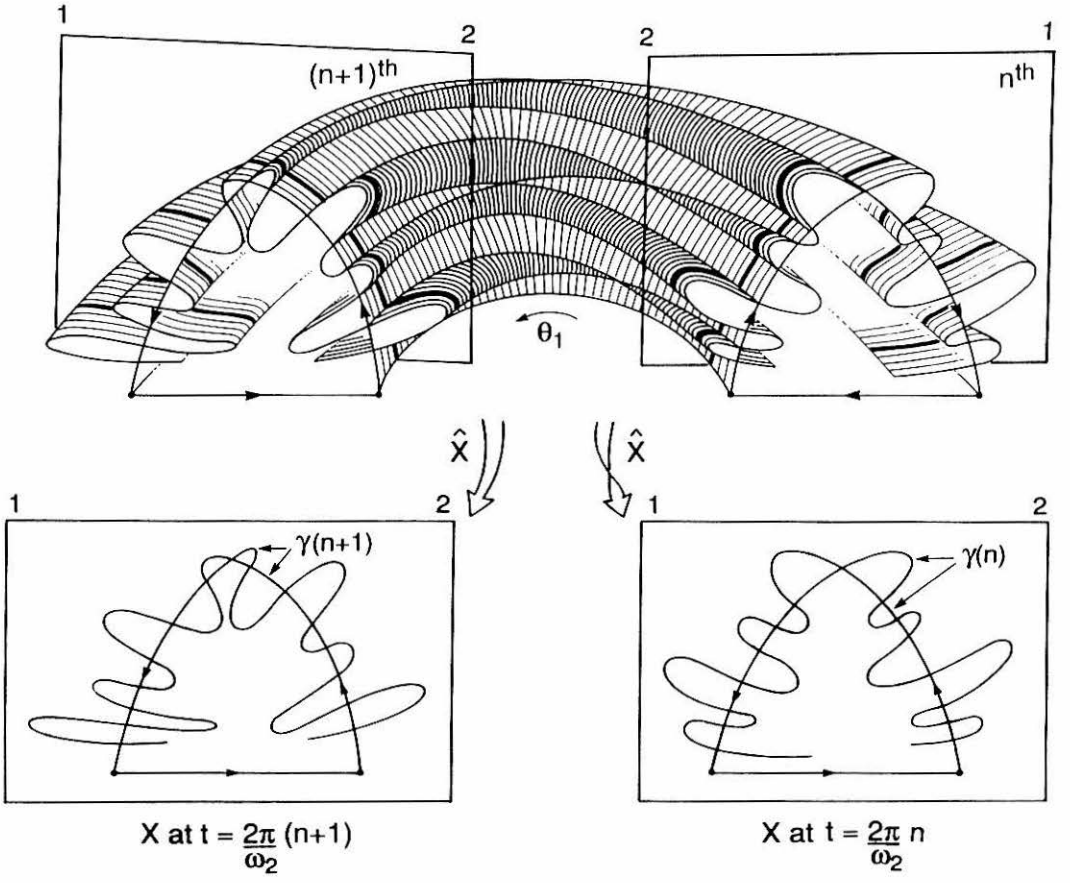


Figure 2.2.9 Projecting onto X to obtain the time-dependent two-dimensional lobe structures in the nonautonomous system phase space.

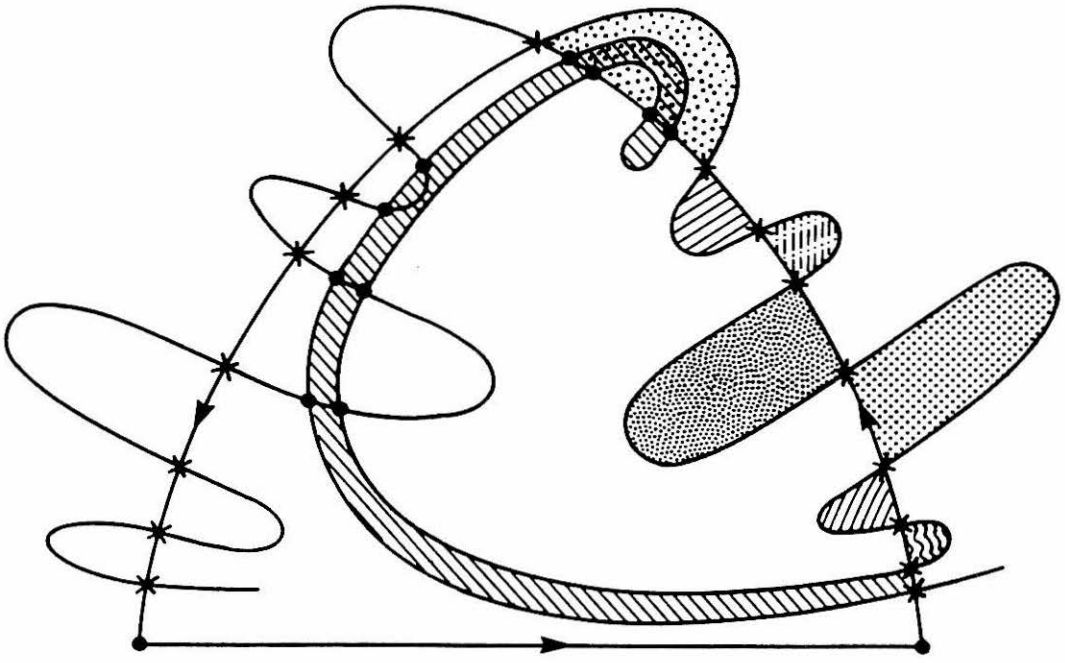


Figure 2.2.10 Showing explicitly some intersection points and two-dimensional lobes in X for a given time sample. The PIP's are marked by crosses, and the SIP's are marked by dots. Each lobe is filled in differently.

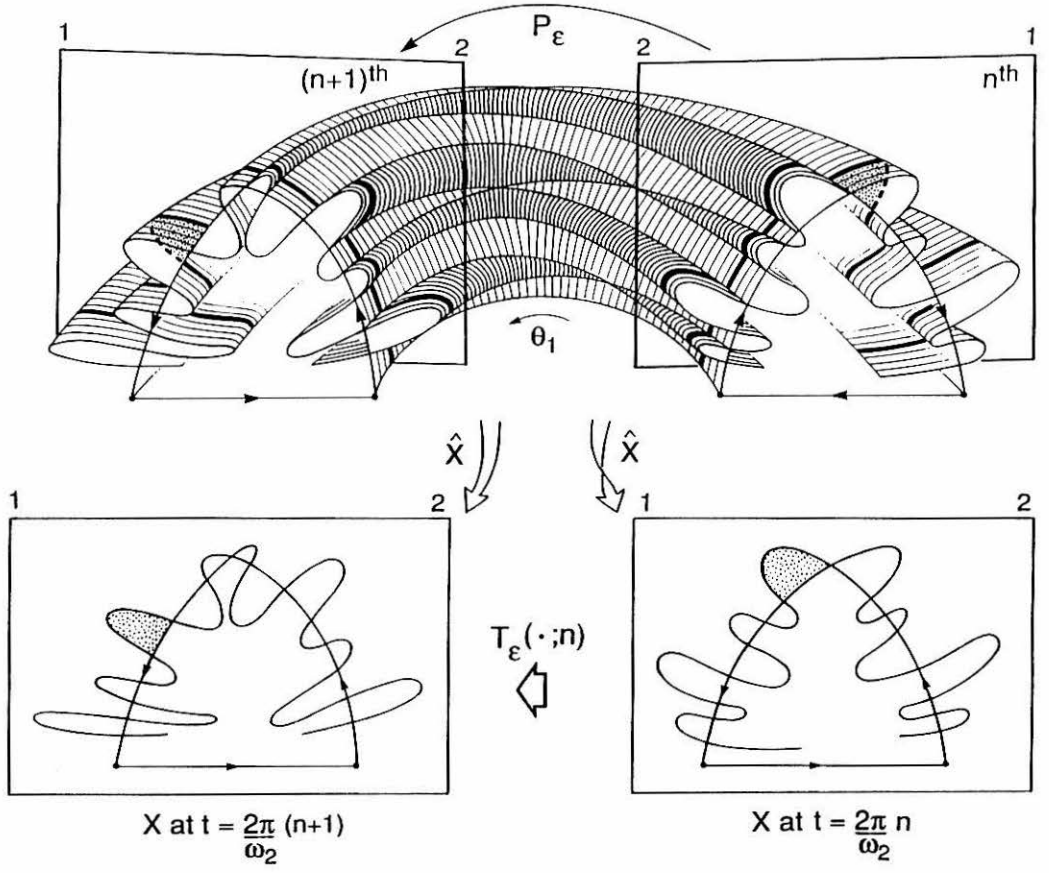


Figure 2.3.1 A lobe mapping between successive time slices of $\Sigma^{\theta_{20}}$, or, equivalently, between successive time samples of X.

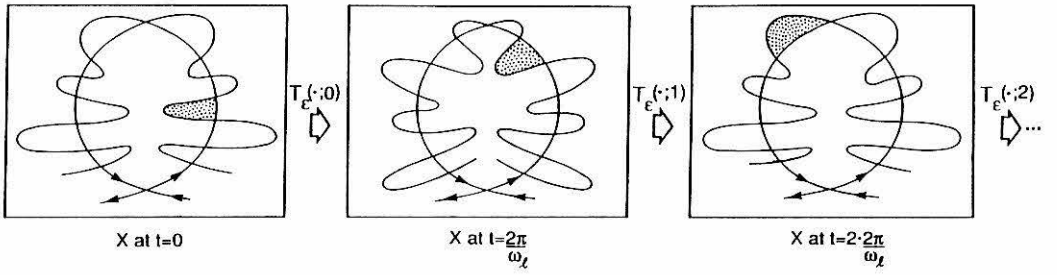


Figure 2.3.2 A lobe mapping within the sequence of lobe structures in X for an ℓ-frequency homoclinic case.

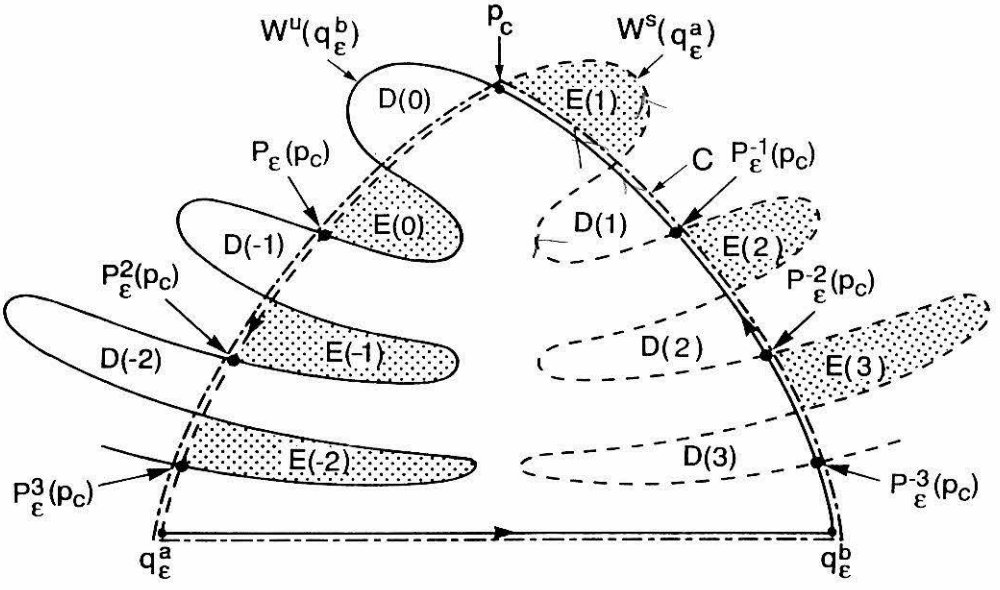


Figure 2.3.3 The invariant lobe structure in $\Sigma^{\theta_{10}=0}$ for the time-periodic OVP flow.

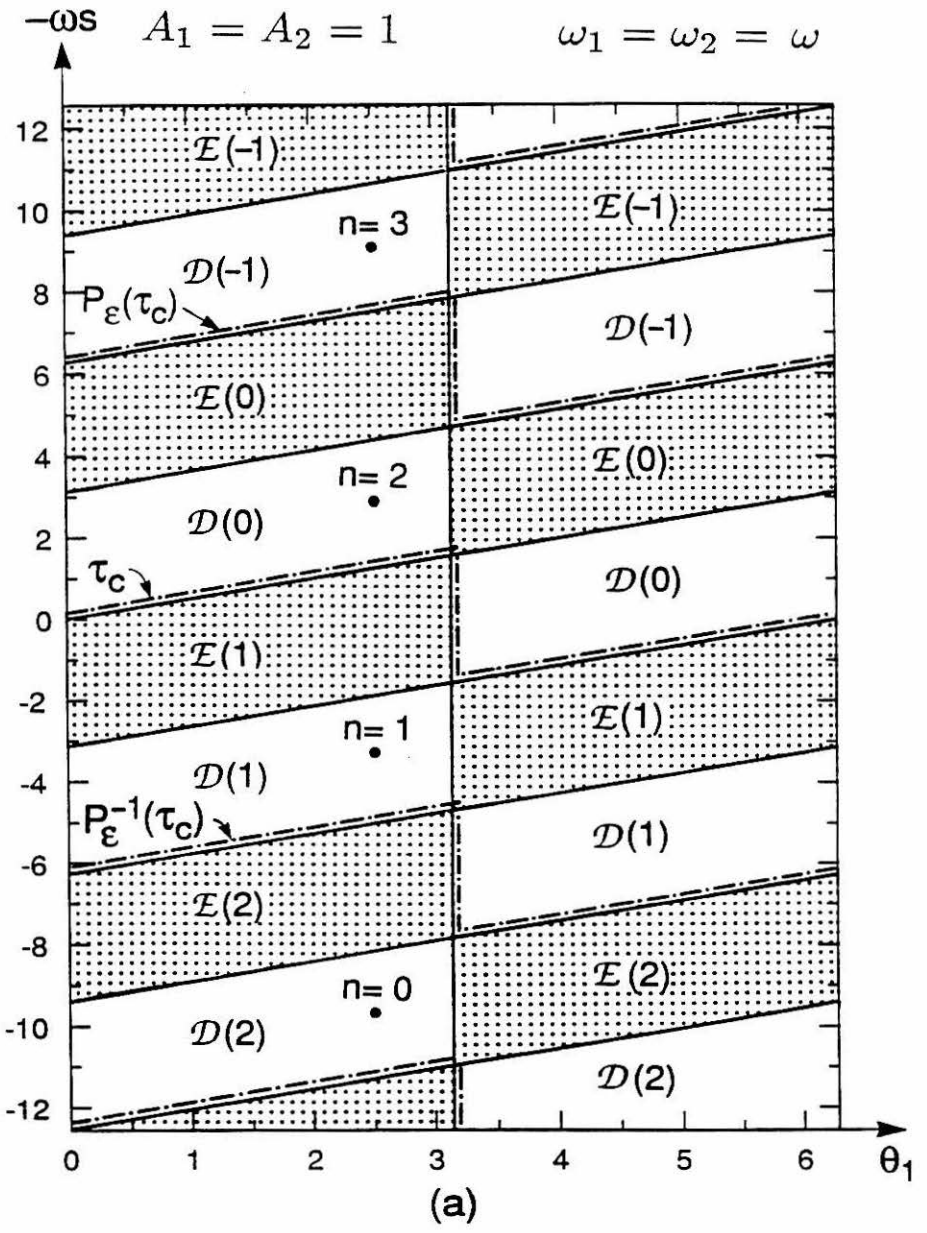


Figure 2.3.4 Portrayal and labeling, in (s, θ_1) space, of the invariant three-dimensional lobes, PIM's, τ_c and $P_\varepsilon^{\pm 1}(\tau_c)$ for systems with Melnikov function of the form (2.3.6).

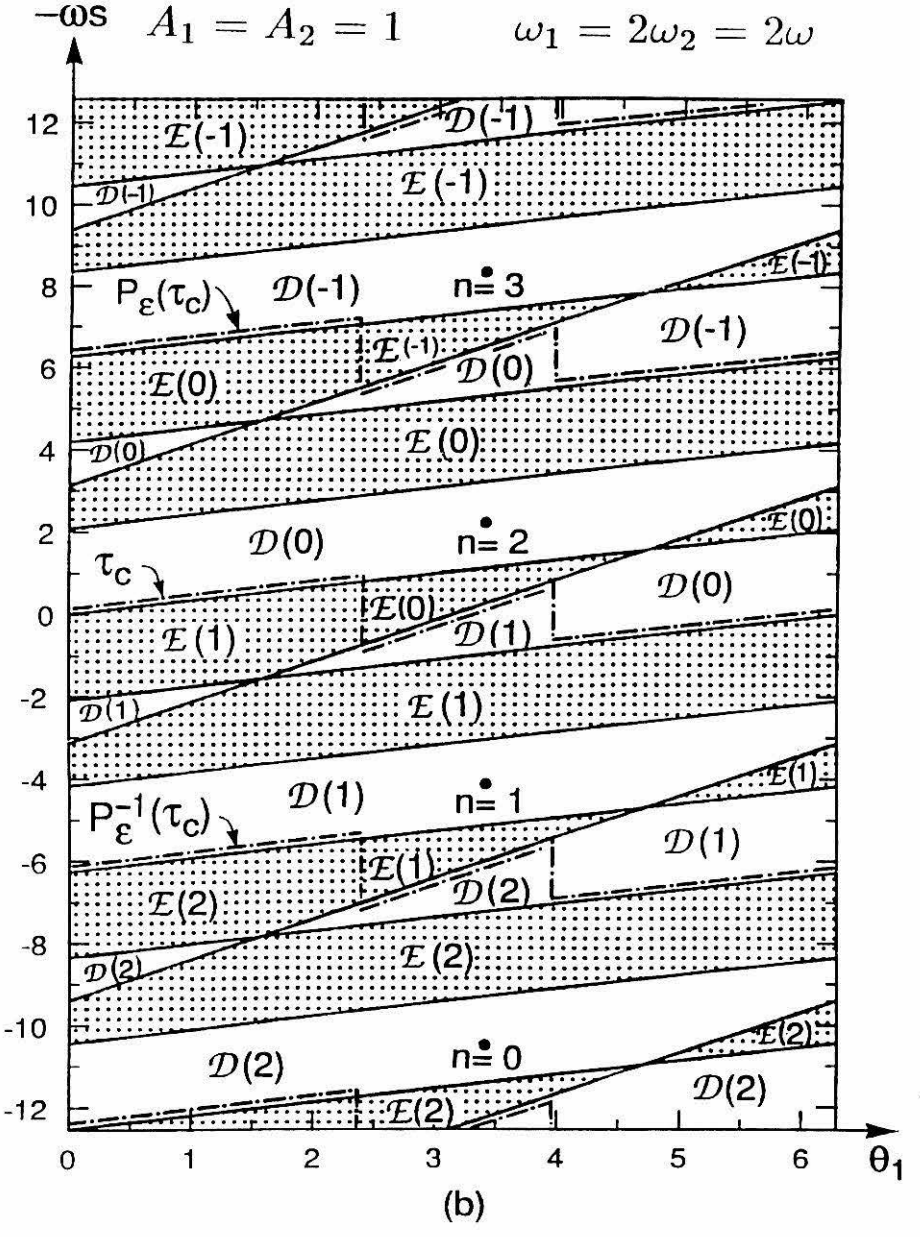


Figure 2.3.4 *Continued.*

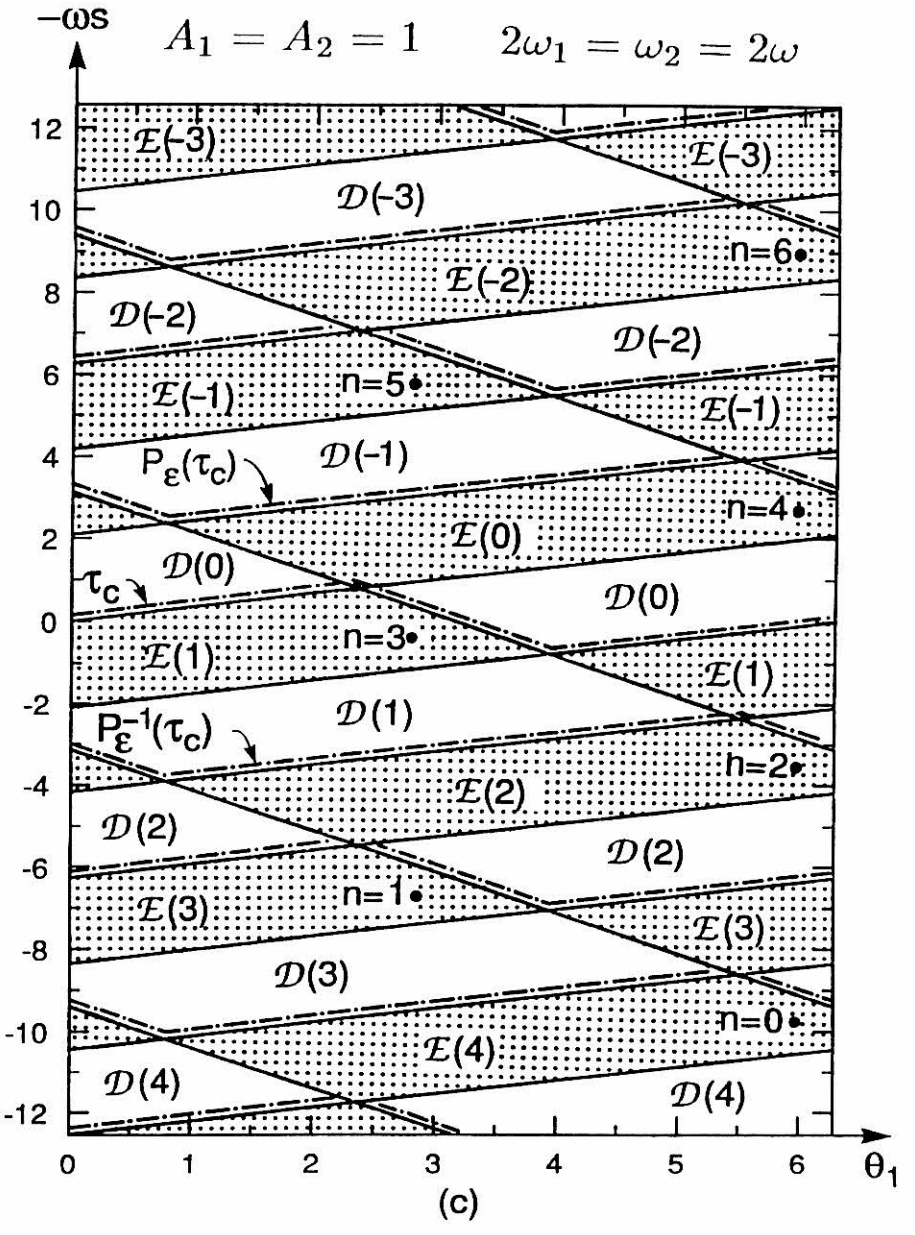


Figure 2.3.4 *Continued.*

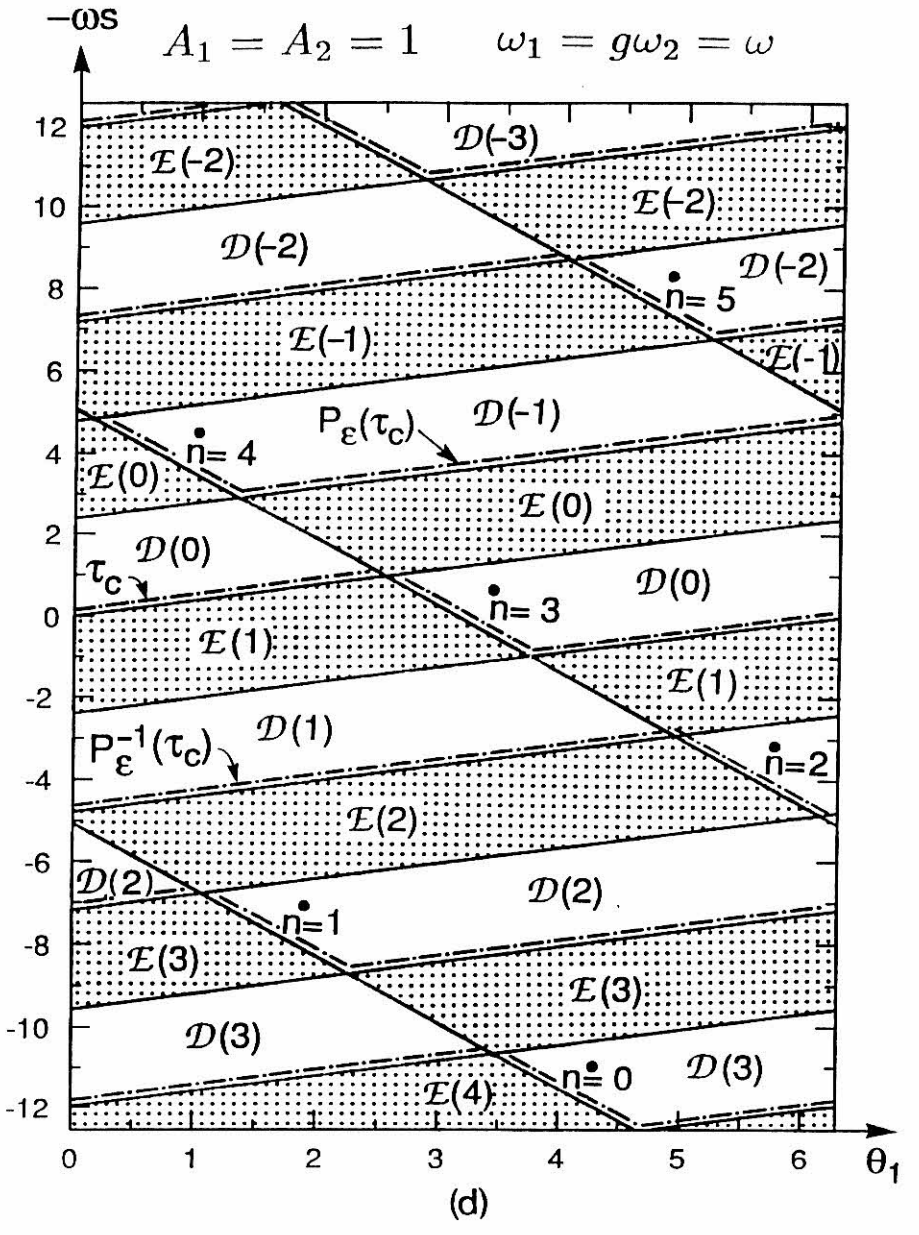


Figure 2.3.4 *Continued.*

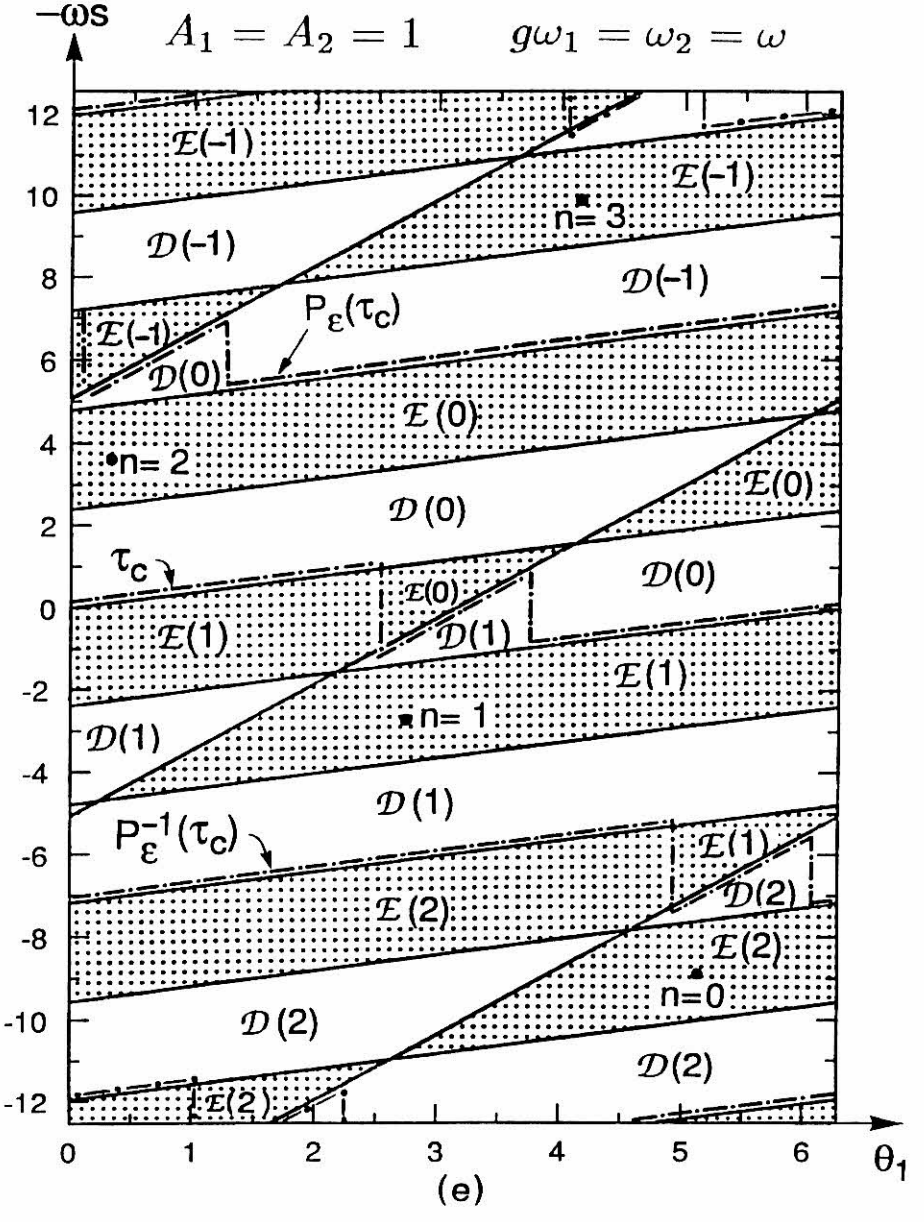


Figure 2.3.4 *Continued.*

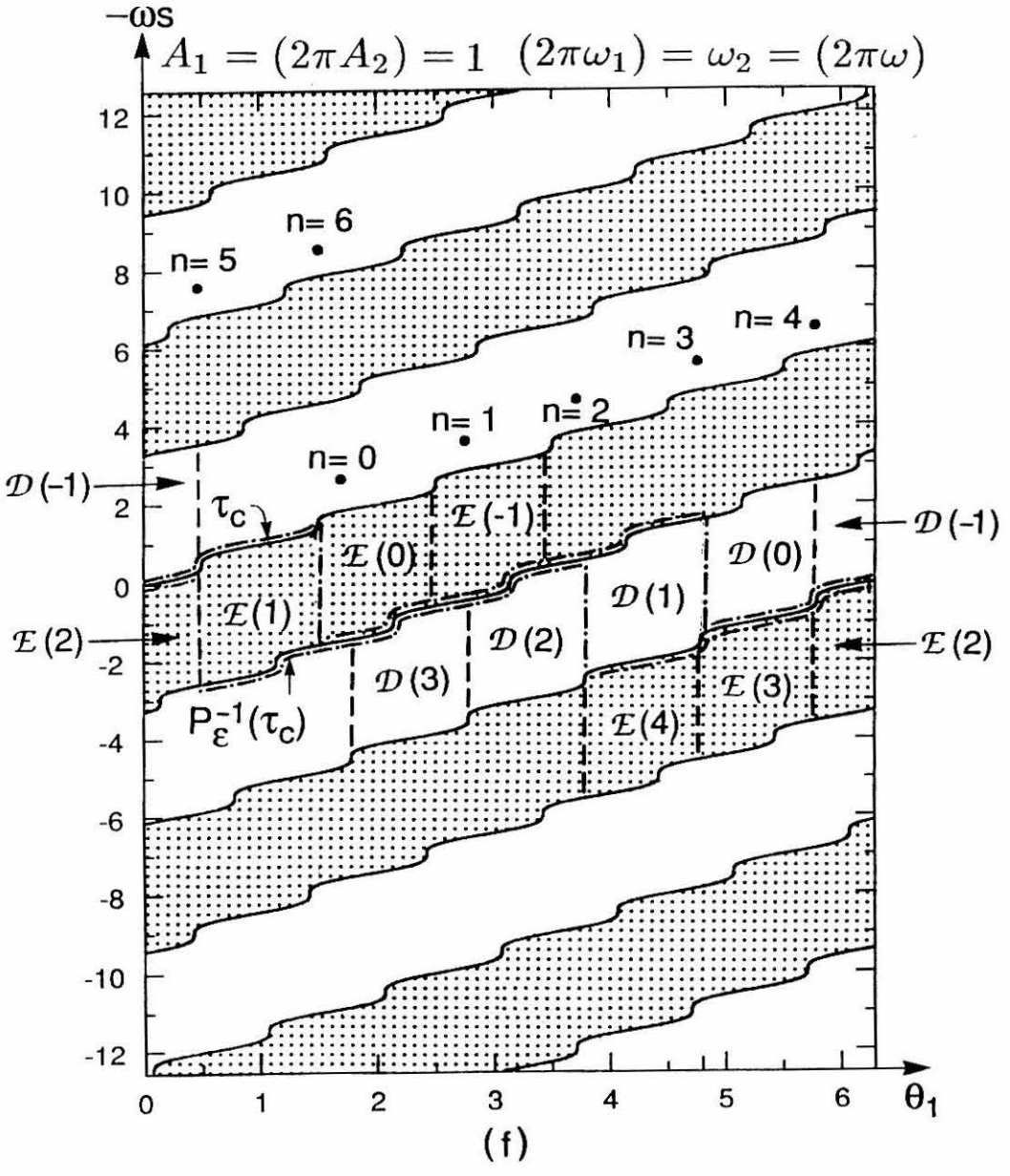
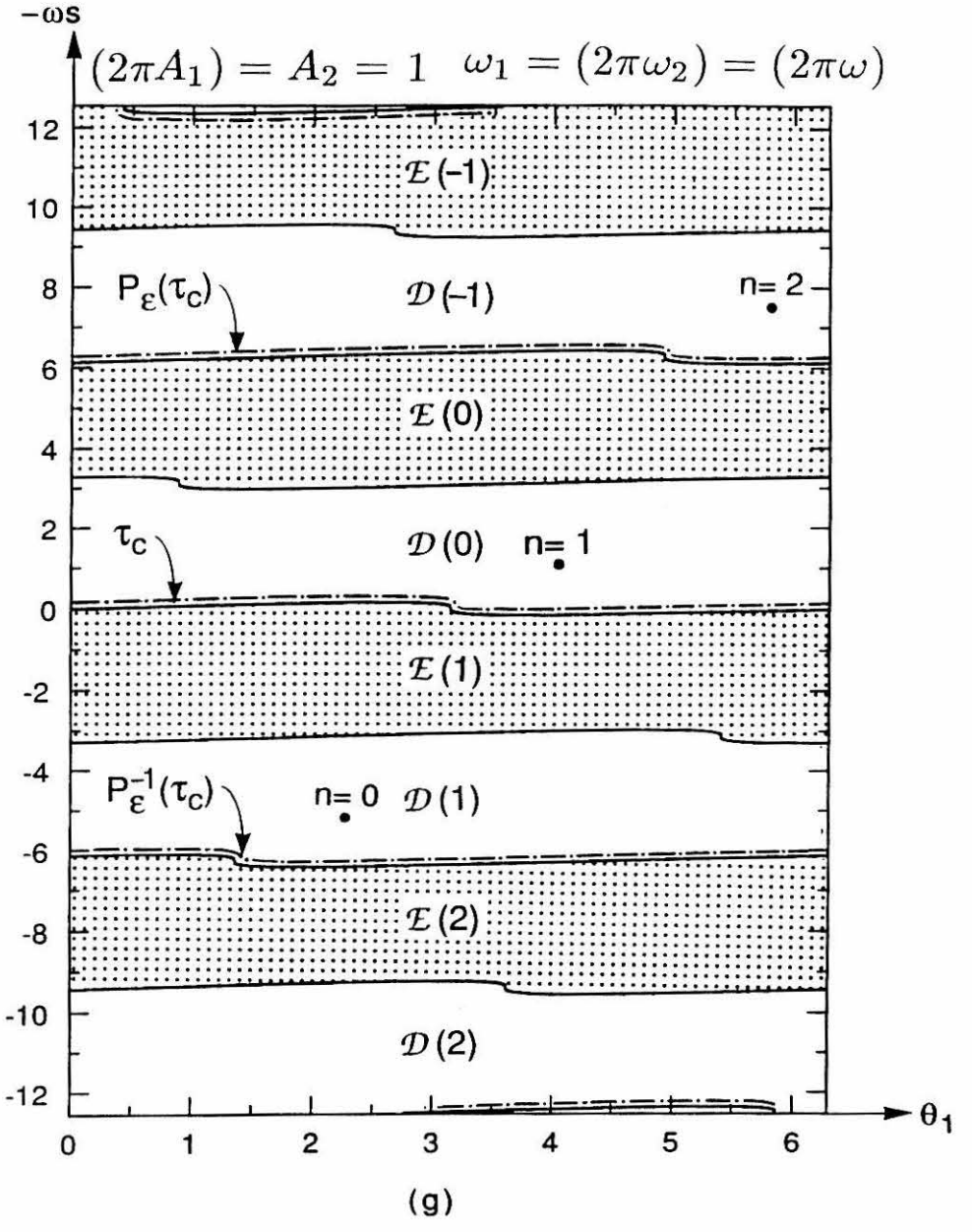


Figure 2.3.4 *Continued.*



(g)

Figure 2.3.4 *Continued.*

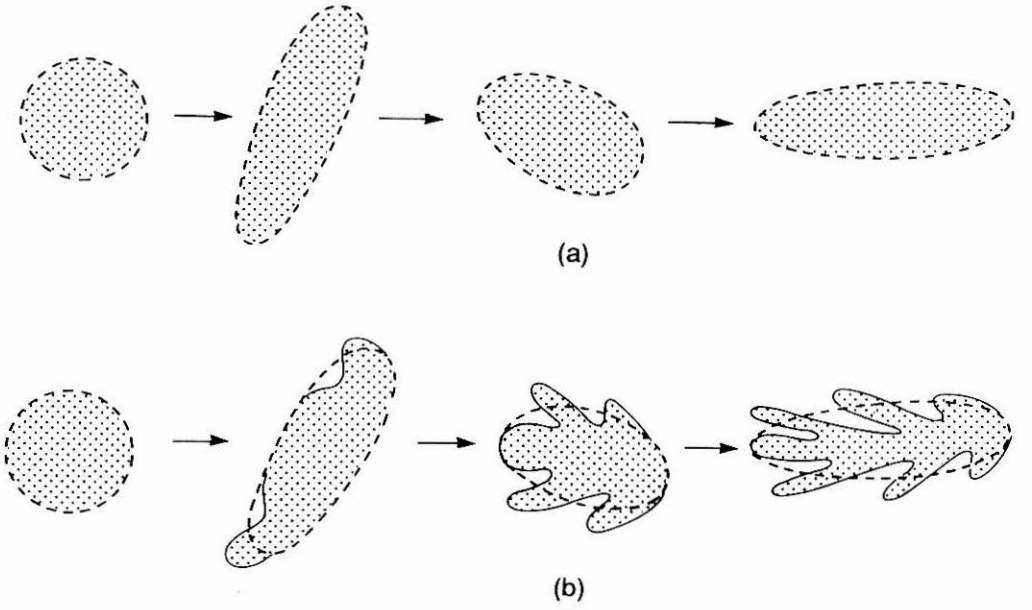


Figure 2.3.5 (a) A pulsating material blob and (b) a pulsating and lobe-forming material blob. A defined core boundary is shown by the dashed line.

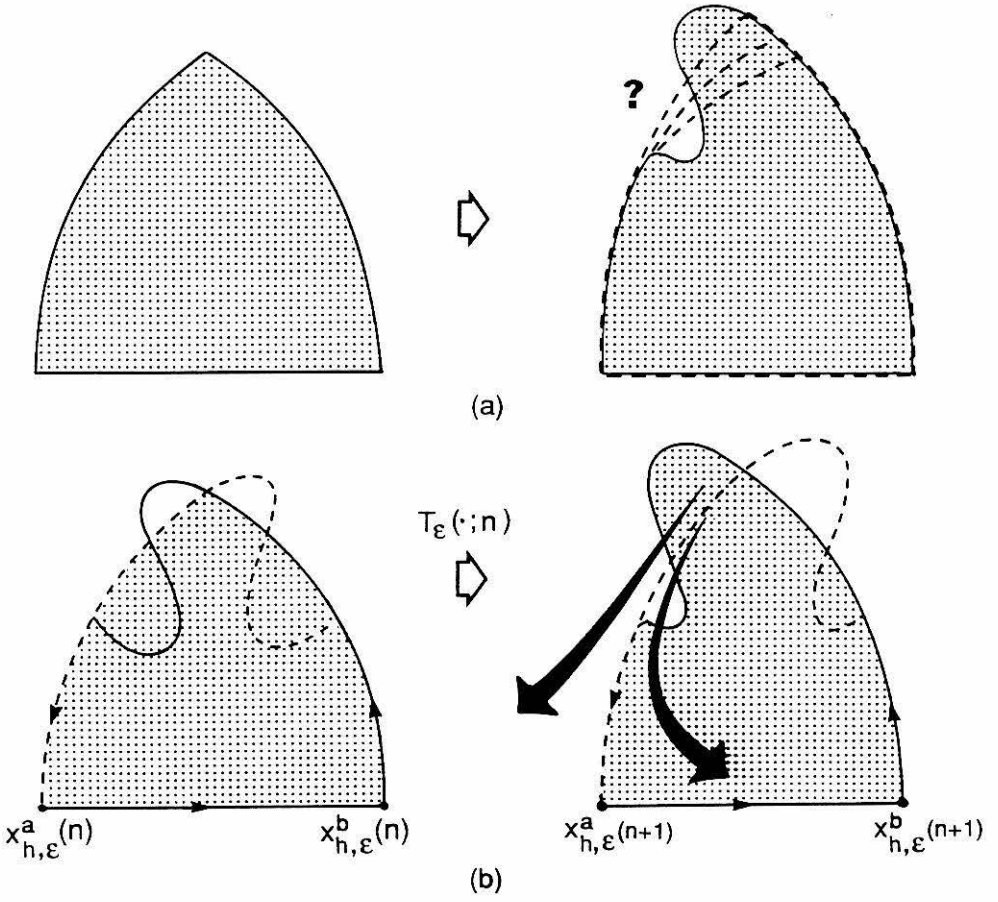


Figure 2.3.6 (a) A material blob deforms; the dashed lines show possible dividing lines between "inside" and "outside". (b) The stable manifold (the dashed line) provides a natural dividing line where the lobes form between "inside" and "outside".

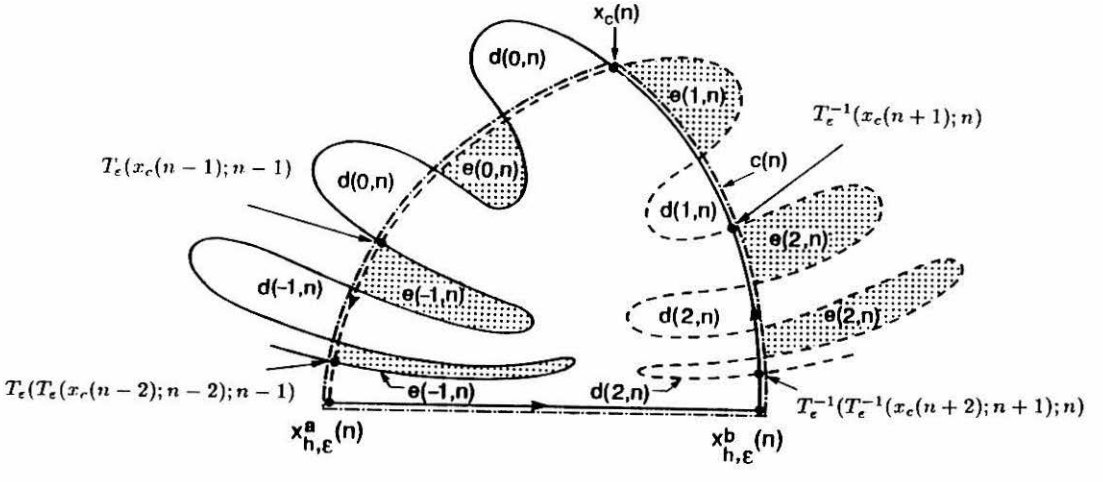


Figure 2.3.7 Portraying the core and labeling the sets of lobes in X for the n^{th} time sample of the two-frequency OVP flow ($x_c(n) = \hat{X}(\tau_c \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n))$, $x^i_{h,\varepsilon}(n) = \hat{X}(\tau^i_\varepsilon \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n))$ for $i = a, b$, and $T_\varepsilon^{-1}(\,\cdot\,;n)$ maps from $t = (n+1)\frac{2\pi}{\omega_2}$ to $t = n\frac{2\pi}{\omega_2}$).

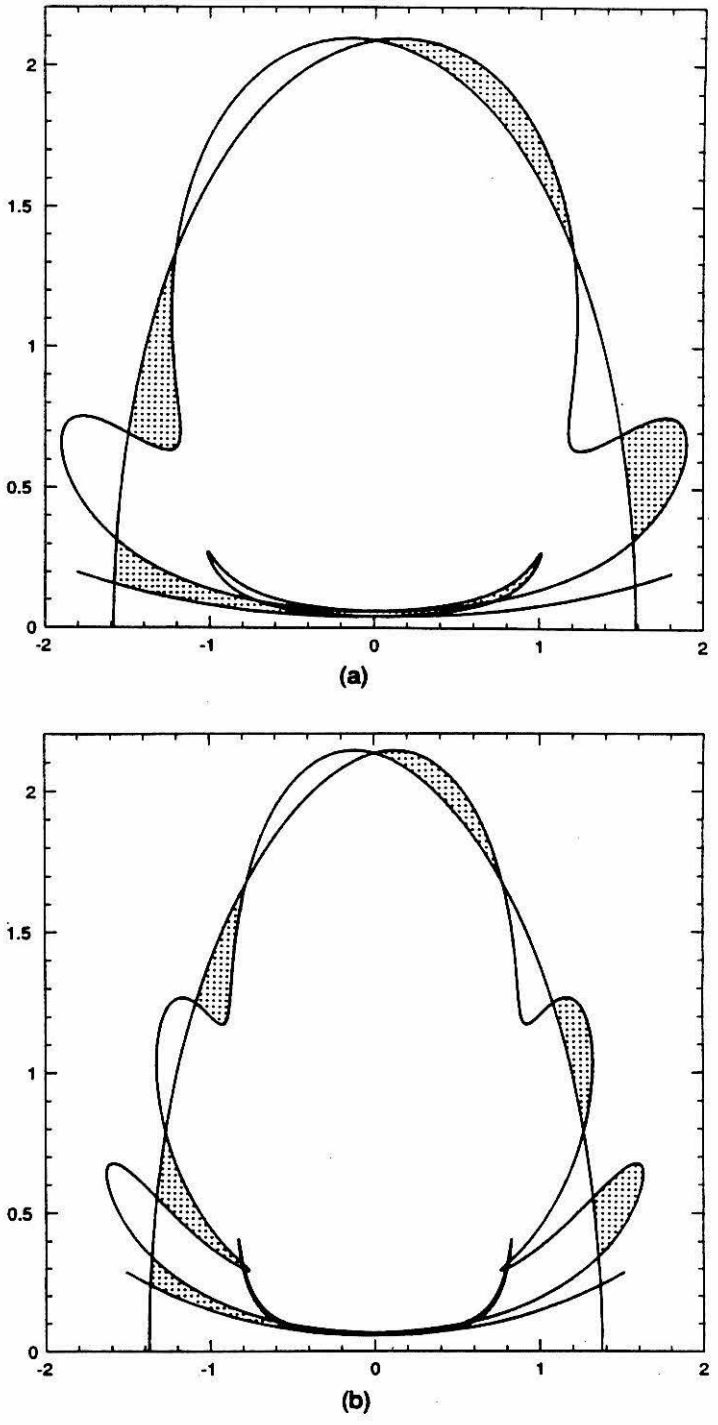


Figure 2.3.8 Some lobes from the OVP lobe structure in X at times (a) $t = n\pi$ for $\psi_{forcing} = 0.12x_1x_2\{2 \cdot 0.4\sin(2t)\}$ and (b) $t = n\frac{\pi}{2}$ for $\psi_{forcing} = 0.12x_1x_2\{4 \cdot 1.05\sin(4t)\}$. The entraining lobes are shaded.

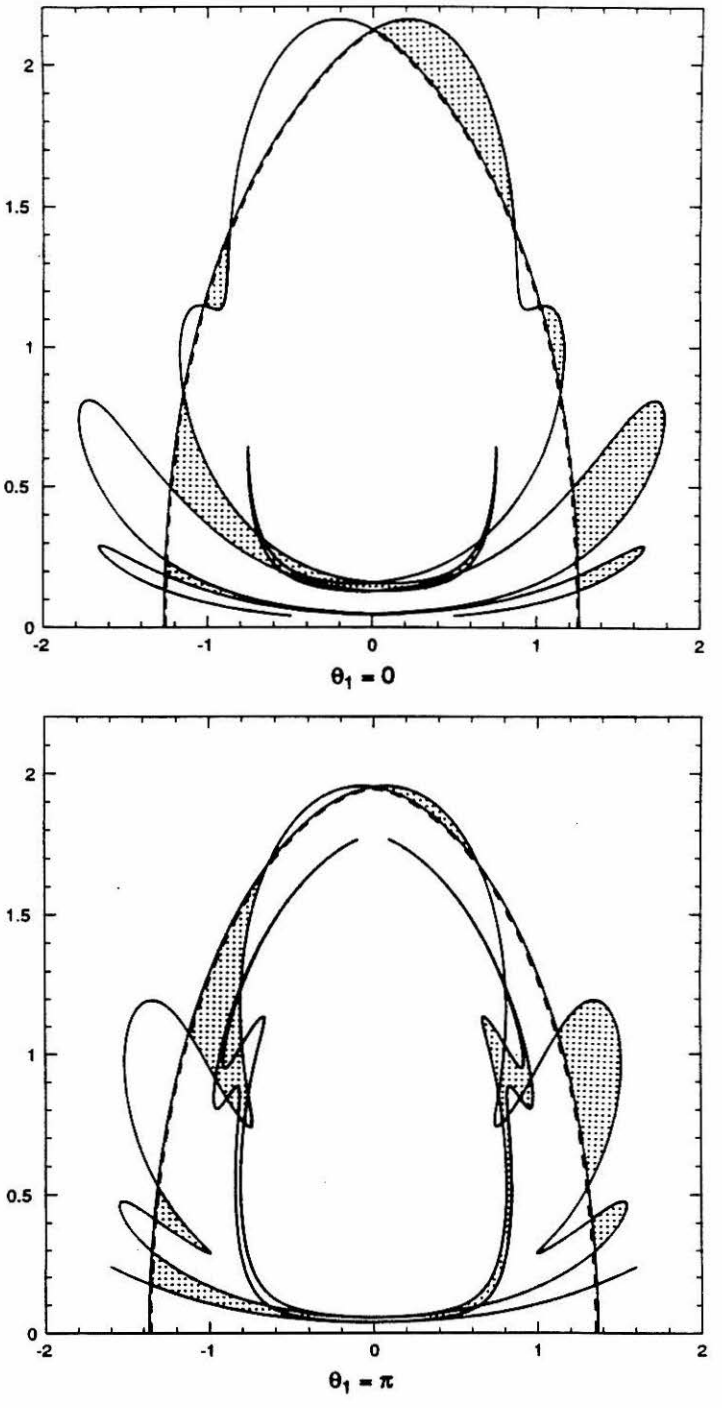


Figure 2.3.9 Some lobes from the OVP lobe structure in the phase slices $\chi(\bar{\theta}_1 = 0)$ and $\chi(\bar{\theta}_1 = \pi)$ of $\Sigma^{\theta_{2_0}=0}$ for $\psi_{forcing}$ given by equation (2.3.12). The entraining lobes are shaded.

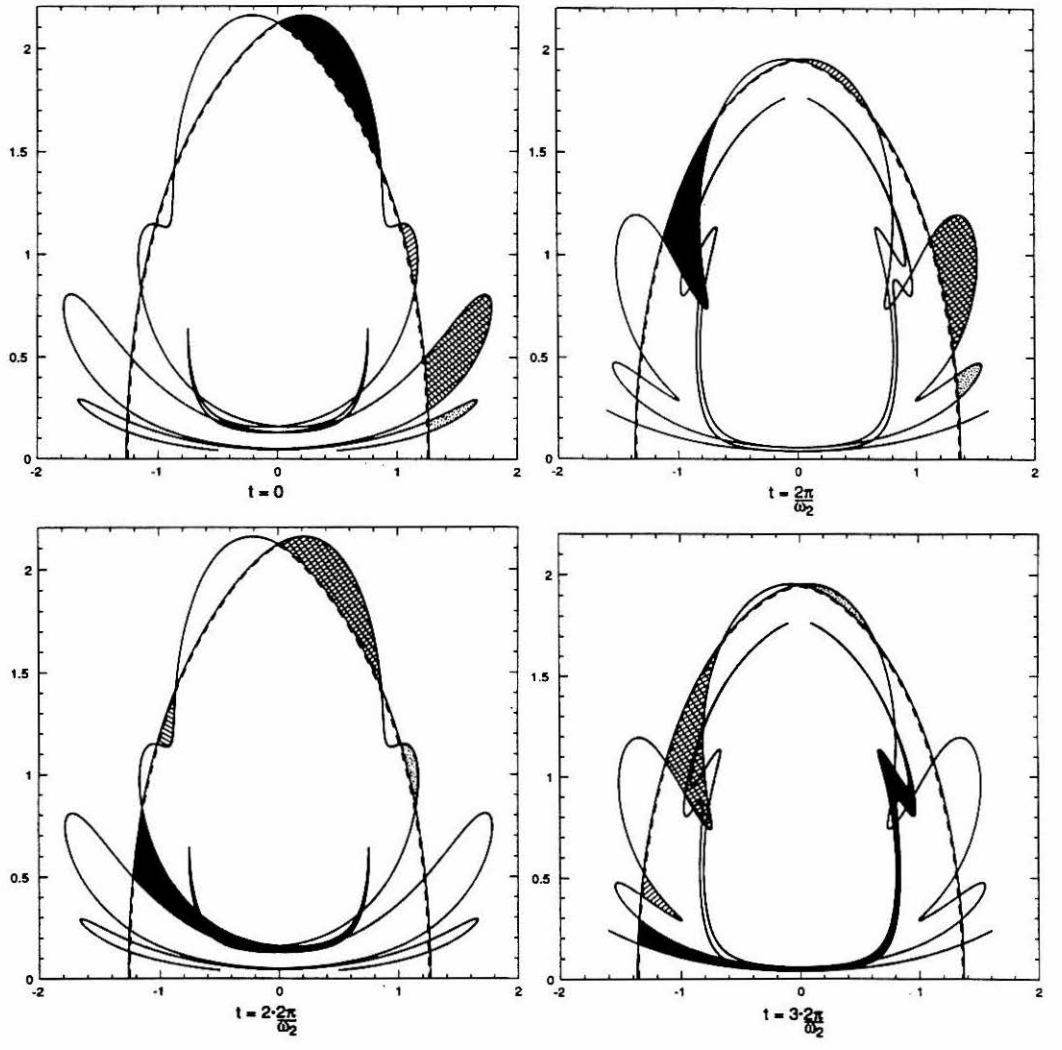


Figure 2.3.10 A sequence of four time samples of the OVP flow forced according to equation (2.3.12). Four material lobes (entraining lobes) are shaded so that we can monitor their transport. The core boundaries are marked by dashed lines.

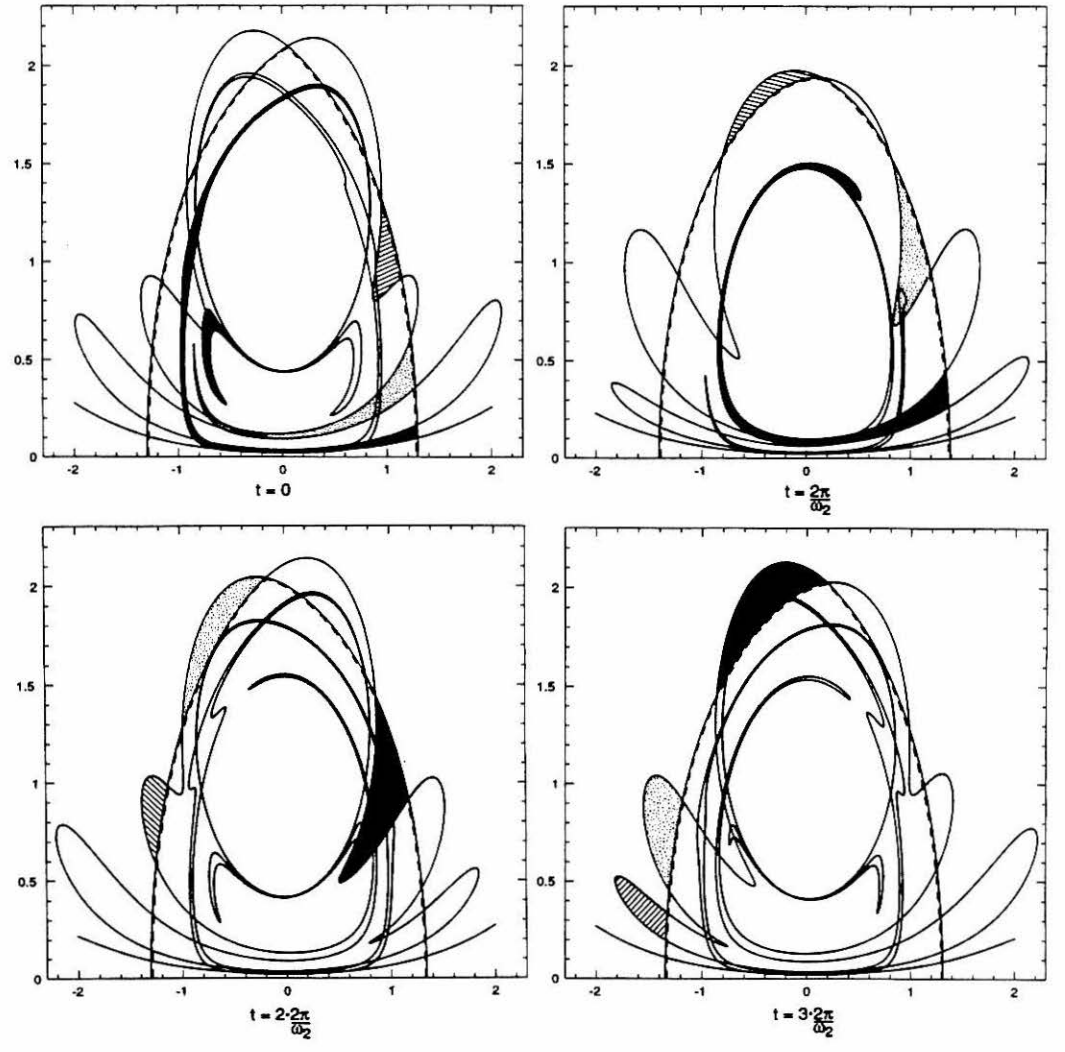


Figure 2.3.11 A sequence of four time samples of the OVP flow forced according to equation (2.3.13). Three material lobes (detraining lobes) are shaded so that we can monitor their transport. The core boundaries are marked by dashed lines.

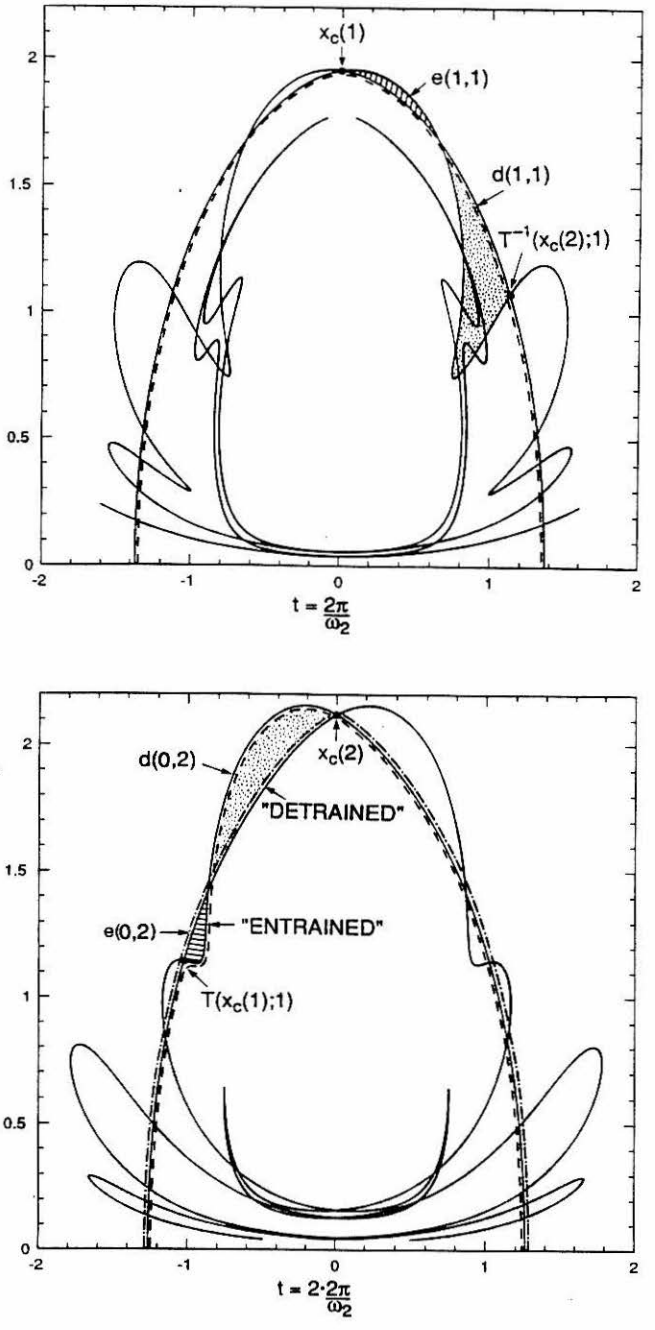


Figure 2.3.12 Showing explicitly entrainment and detrainment between the $t = 2\pi/\omega_2$ sample and the $t = 2 \cdot 2\pi/\omega_2$ sample of the OVP flow with forcing given by equation (2.3.12). The dashed line is $c(1)$ in the $t = 2\pi/\omega_2$ sample and $T(c(1);1)$ in the $t = 2 \cdot 2\pi/\omega_2$ sample. The dashed-dotted line is $c(2)$.

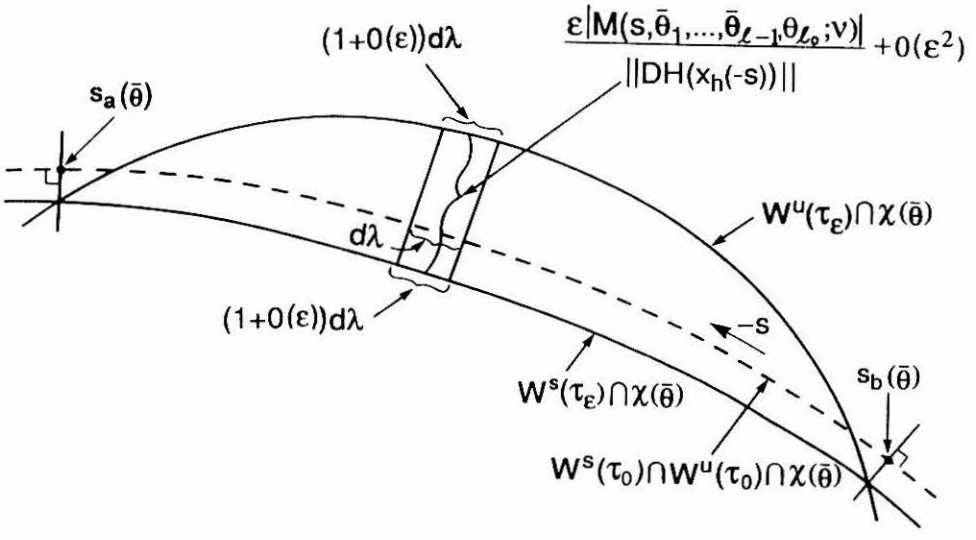


Figure 2.4.1 An area element of a two-dimensional turnstile lobe in a given phase slice for an ℓ-frequency homoclinic tangle.

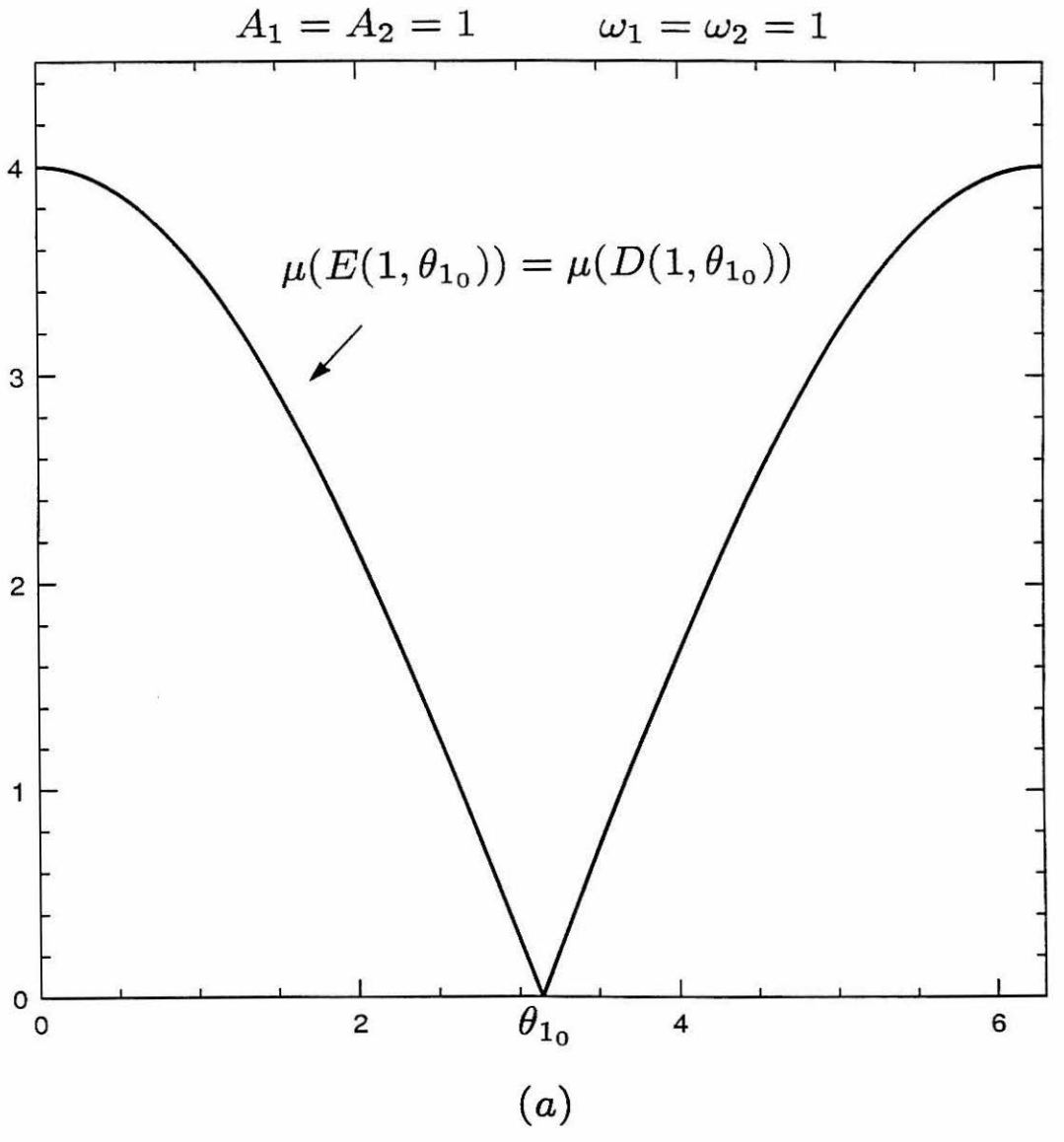


Figure 2.4.2 Turnstile lobe areas for the examples of Section 2.3 (the vertical scale is per unit ε and $< >_n$ denotes the average over all n). For the labeling of individual lobes, $E(m, \bar{\theta}_1; i)$ and $D(m, \bar{\theta}_1; i)$, we choose i to increase in the same direction as m. Note that the labeling applies to *types* of lines; for example, in (b) $\mu(E(0, \theta_1; 1))$ applies to the solid line, which is discontinuous.

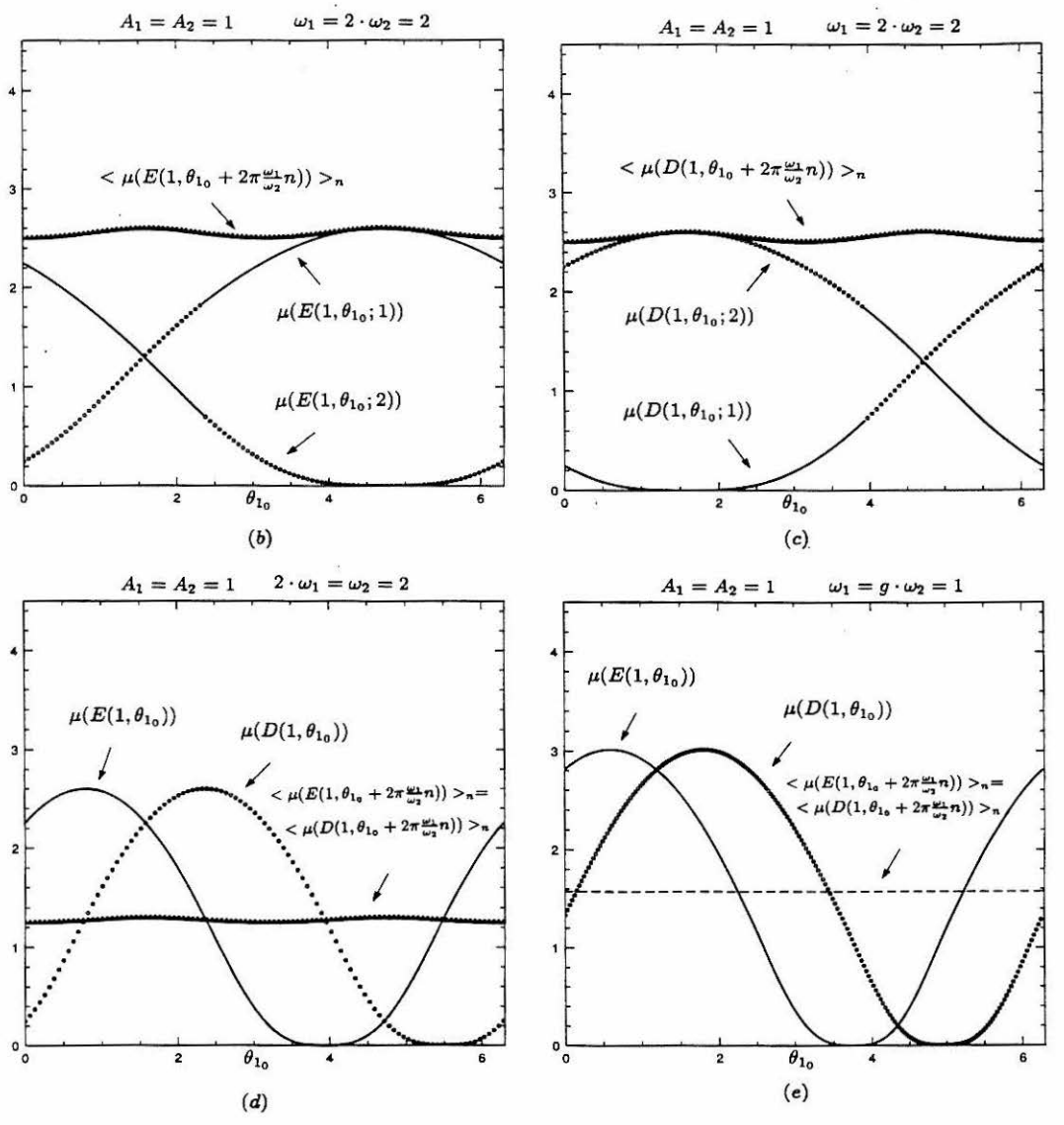


Figure 2.4.2 *Continued.*

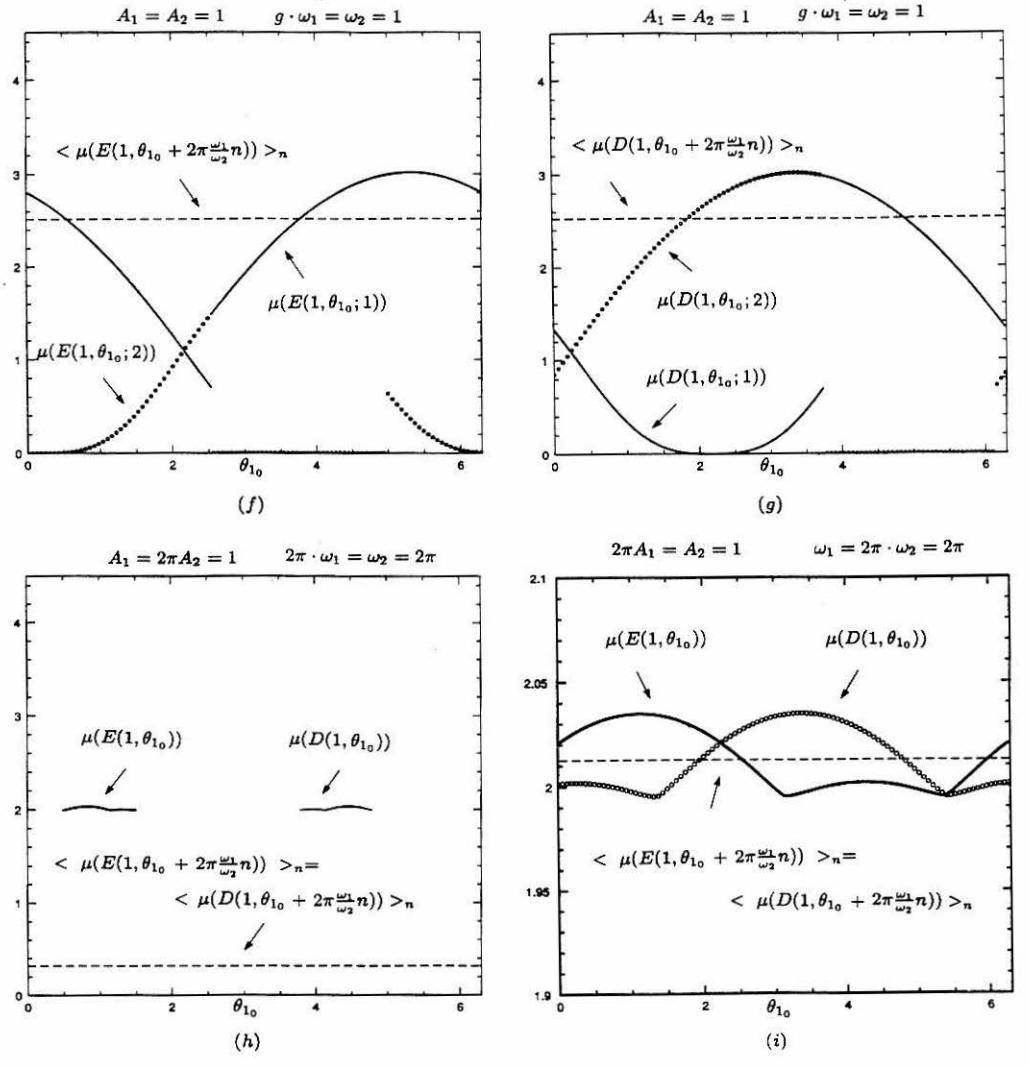


Figure 2.4.2 *Continued.*

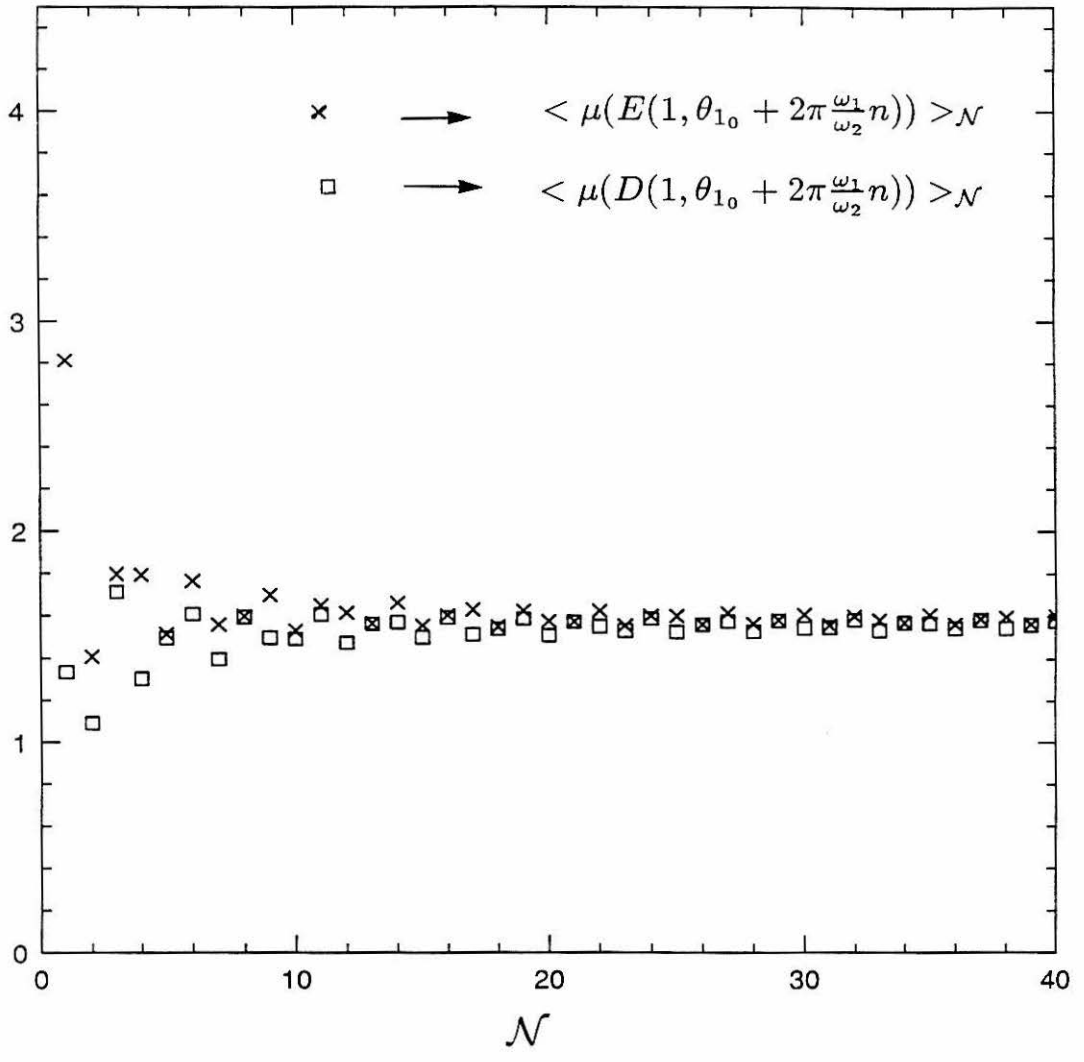
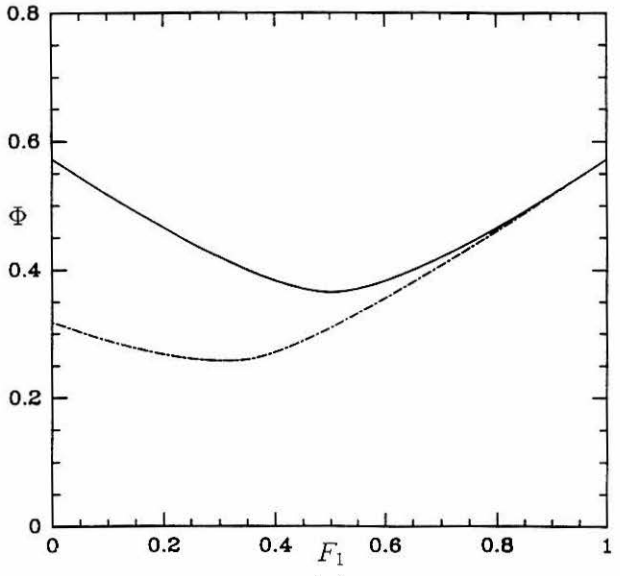
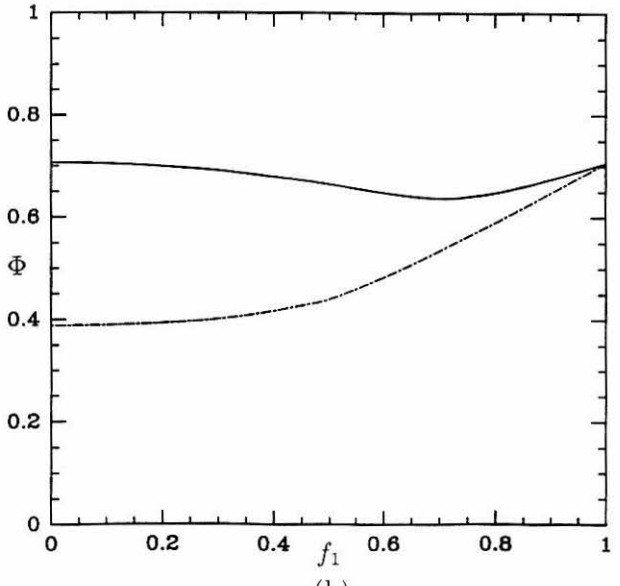


Figure 2.4.3 Finite-time average areas for the fourth example of Section 2.3 ($\theta_{1_0} = 0$). The average over the first $\mathcal{N}$ time samples is denoted by $< >_{\mathcal{N}}$. Again the vertical scale is per unit ε.



(a)



(b)

Figure 2.4.4 (a) Average flux as a function of F_1 for $F_1 + F_2 = 1$, with $(\omega_1, \omega_2) = (1.48, 0.78)$, $(A_1/F_1, A_2/F_2) = (1.80, 1.80)$ for the solid line and with $(\omega_1, \omega_2) = (1.48, 0.66)$, $(A_1/F_1, A_2/F_2) = (1.80, 1.00)$ for the dashed-dotted line (A_i/F_i is taken from the OVP flow: see Appendix 2.A1). (b) Average flux for the OVP flow as a function of f_1 for $f_1^2 + f_2^2 = 1$, with $(\omega_1, \omega_2) = (1.00, 1.94)$, $(A_1/f_1, A_2/f_2) = (2.22, 2.22)$ for the solid line and $(\omega_1, \omega_2) = (1.00, 0.75)$, $(A_1/f_1, A_2/f_2) = (2.22, 1.22)$ for the dashed-dotted line. In both plots the vertical scale is per unit ε and $\theta_{1_0} = \theta_{2_0} = 0$.

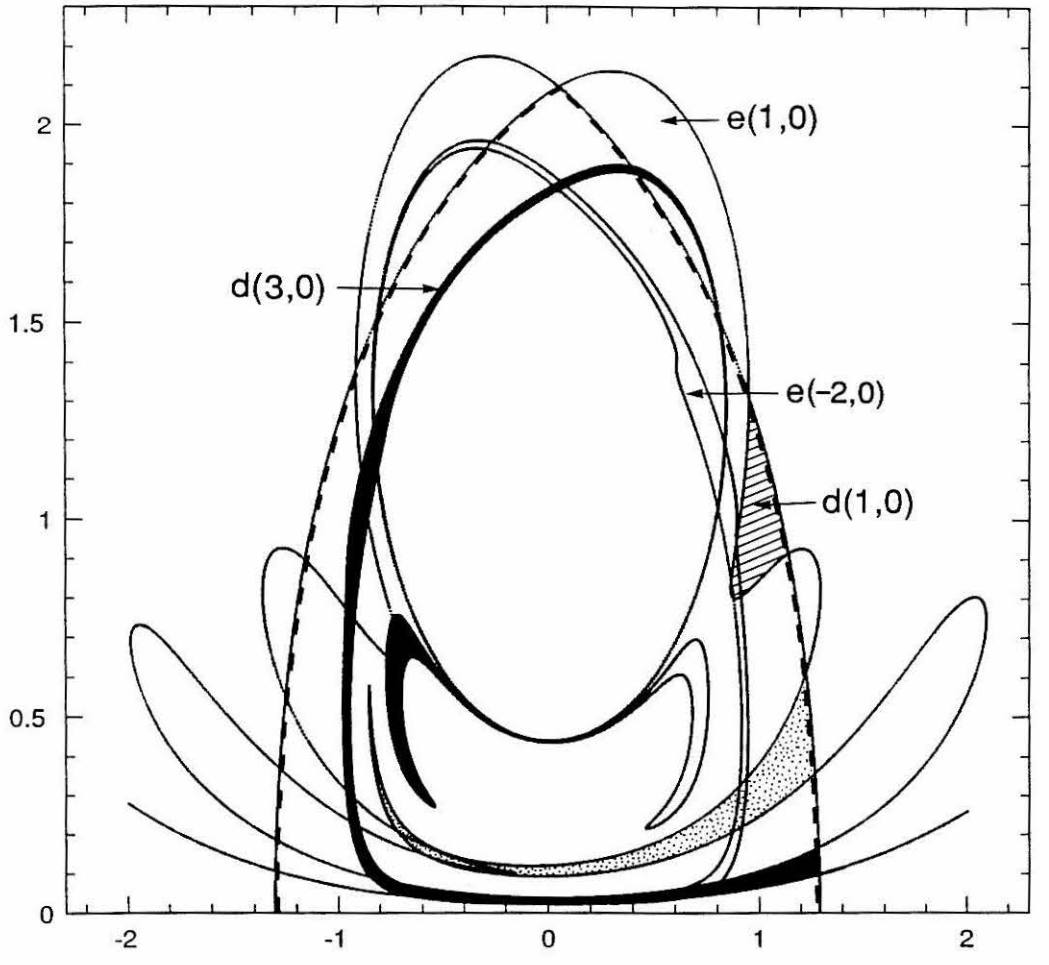


Figure 2.4.5 The lobe structure of the $t = 0$ sample of Figure 2.3.11, shown again with the necessary labeling.

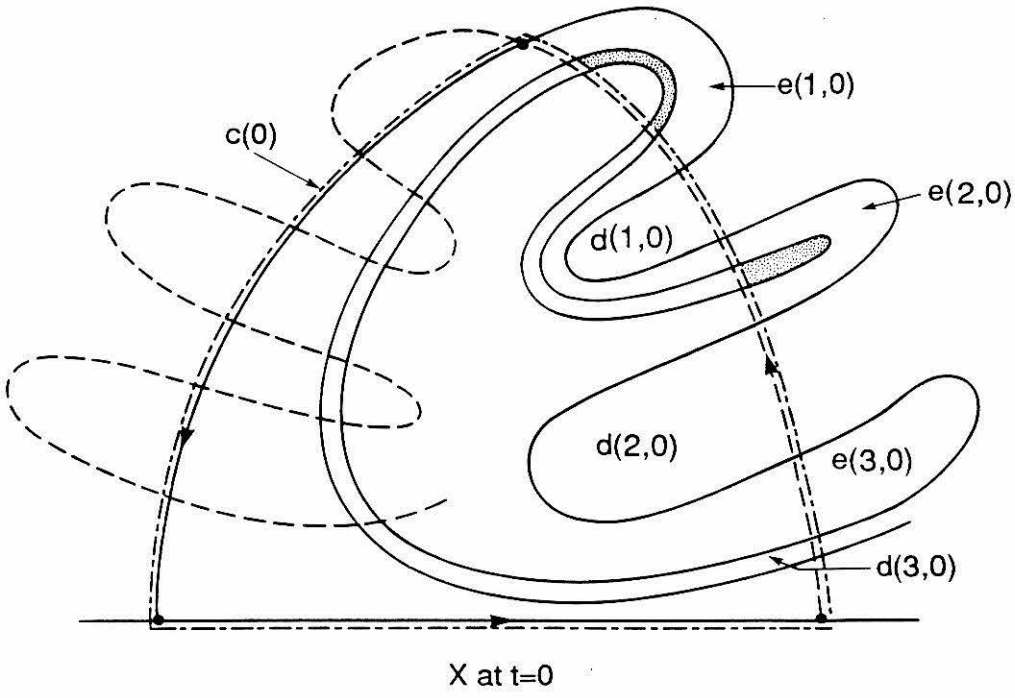
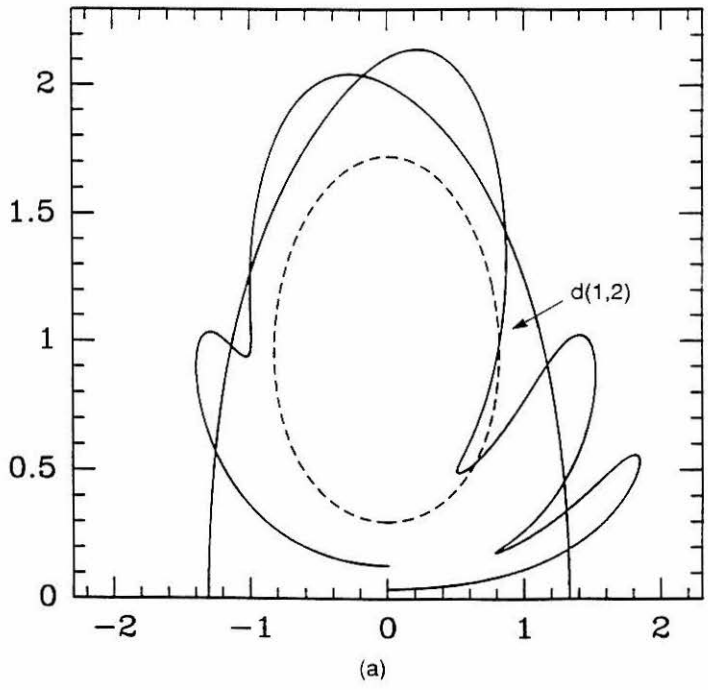


Figure 2.4.6 Schematic diagram of some lobes of the OVP flow in the $t = 0$ sample. The $d(3,0)$ lobe is contained partially inside and partially outside the $t = 0$ core. The portion of the lobe with type A particles is white, and the portion with type B particles is shaded.



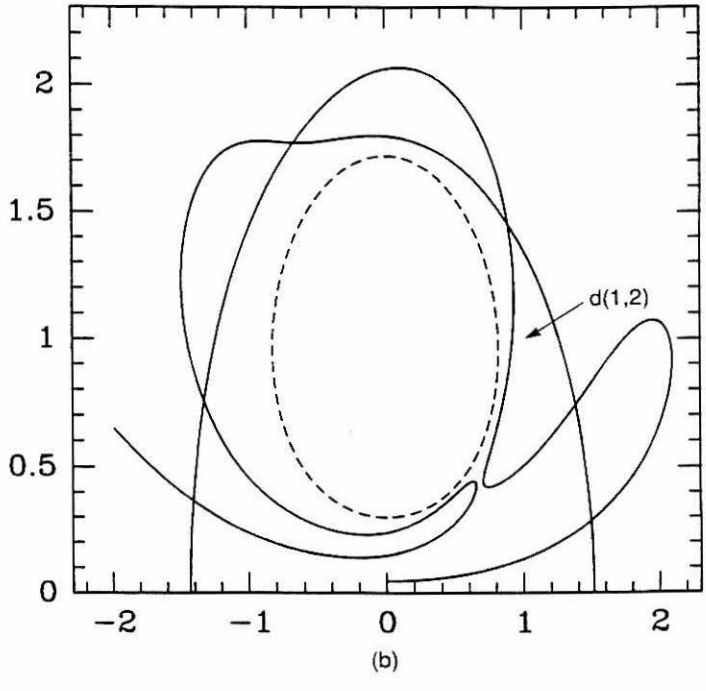


Figure 2.4.7 Some lobes in the $n = 2$ time sample of (a) the system portrayed in Figure 2.3.11 and (b) a corresponding single-frequency system ($\omega = 2$).

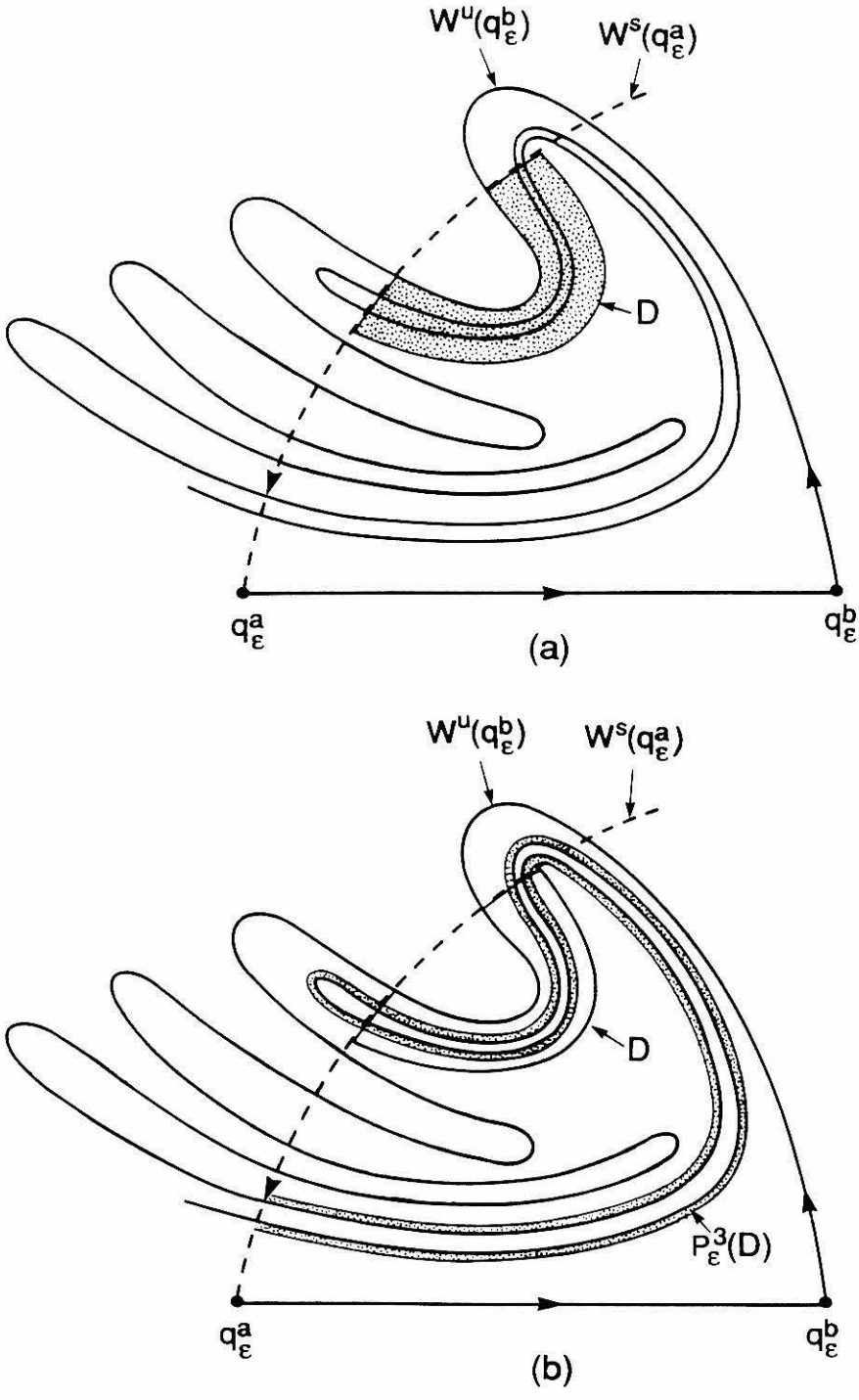


Figure 2.5.1 A horseshoe map in the heteroclinic tangle region of $\Sigma^{\theta_{10}=0}$ in the time-periodic case.

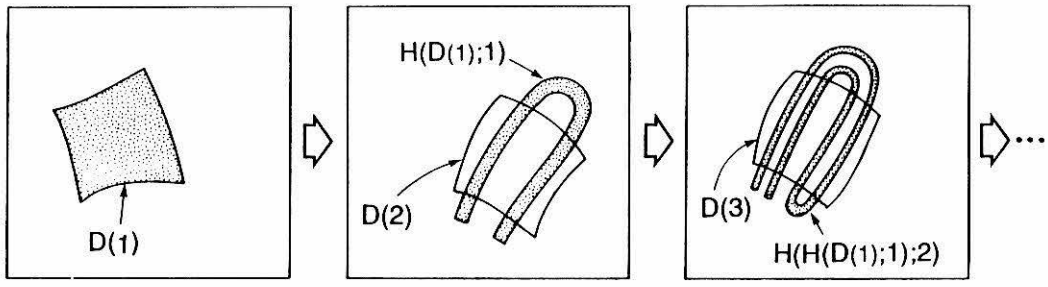


Figure 2.5.2 A traveling horseshoe map sequence.

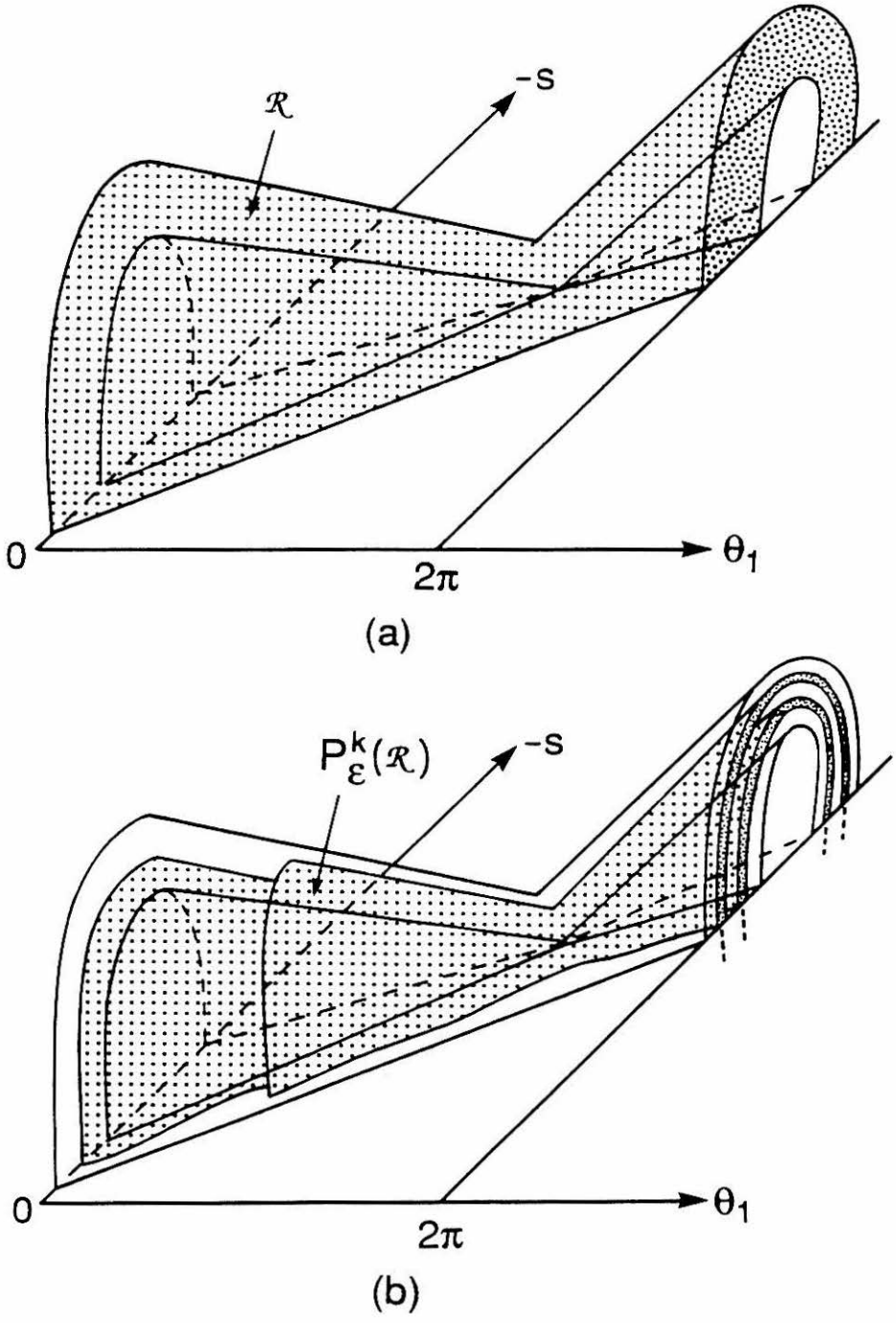


Figure 2.5.3 $P_\varepsilon^k(\mathcal{R})$ intersects $\mathcal{R}$ in a horseshoe fashion. Note how the initial discontinuity of $\mathcal{R}$ at $\theta_1 = 0$ is mapped to another θ_1 value.

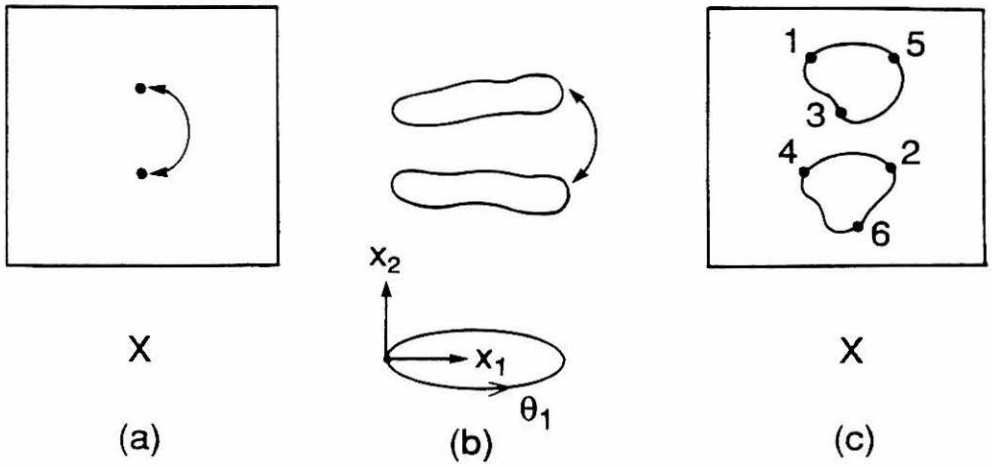


Figure 2.5.4 (a) A period two point in X for the time-periodic case; (b) a period two 1-torus in $\Sigma^{\theta_{20}}$ for the two-frequency case; (c) motion in X under $\{T_\varepsilon(\,\cdot\,;n);n\in Z\}$ of a point that initially lies on the period two 1-torus of (b).

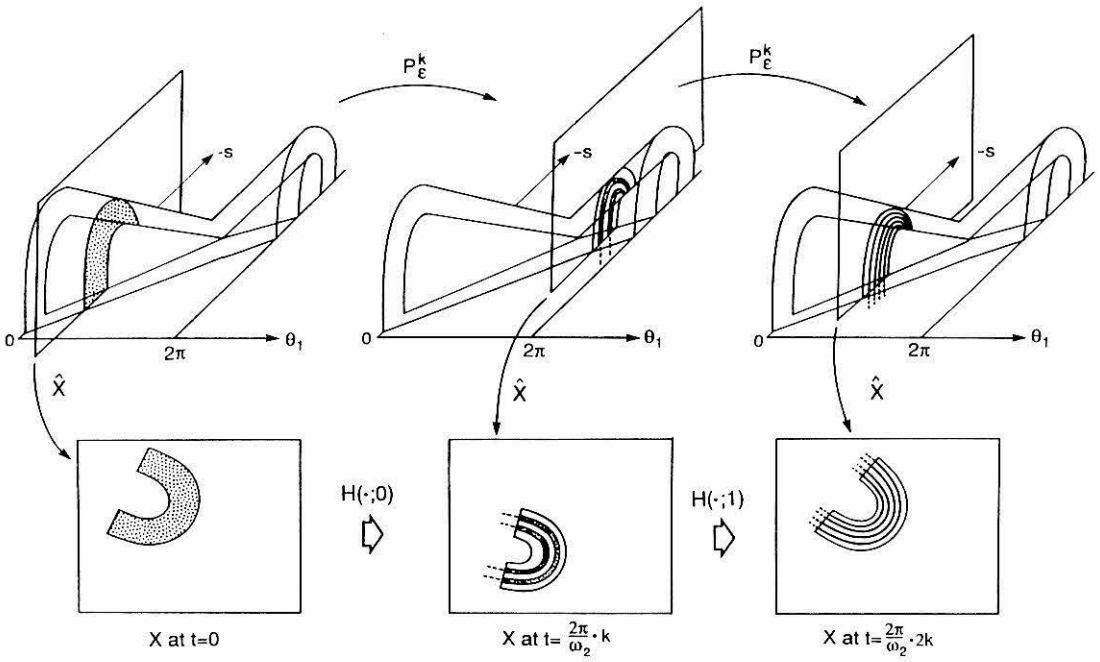


Figure 2.5.5 Obtaining a traveling horseshoe map sequence from $\mathcal{R} \cap P_\varepsilon^k(\mathcal{R})$.

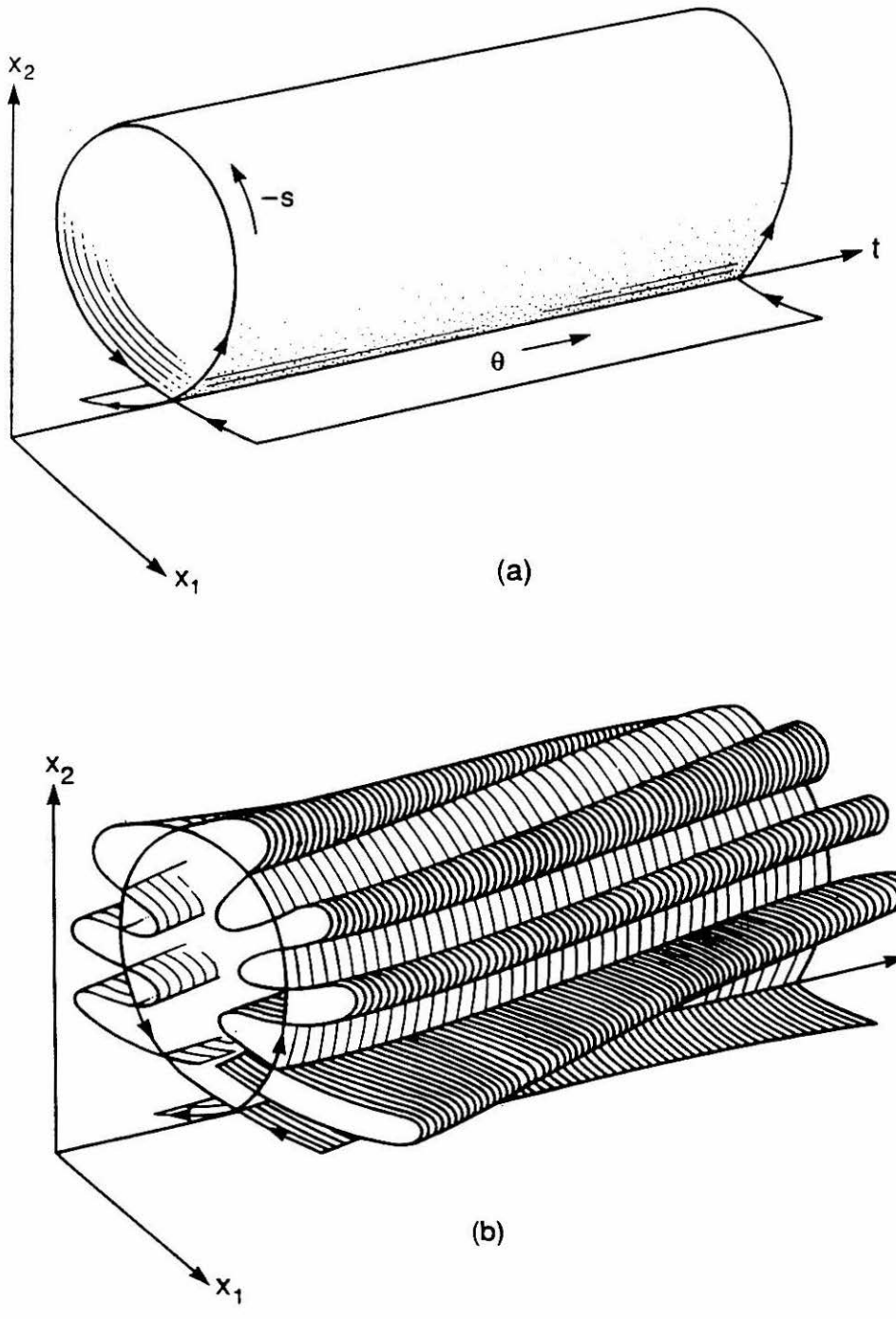


Figure 2.7.1 (a) An unperturbed homoclinic manifold in $X \times \mathbb{R}$; (b) an invariant lobe structure in $X \times \mathbb{R}$ for the perturbed case.

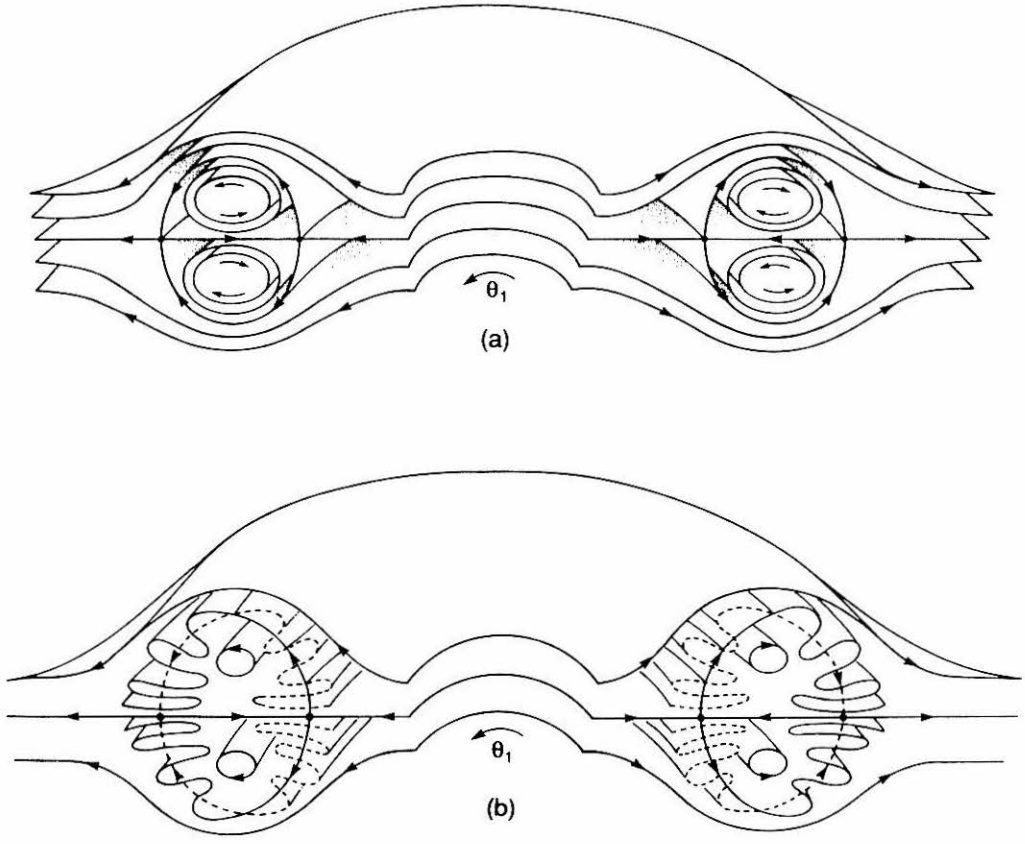


Figure 2.A1.1 (a) Some level sets in $\Sigma^{\theta_{2_0}}$ of the unperturbed OVP flow. (b) The lobe structure in $\Sigma^{\theta_{2_0}}$ of the perturbed OVP flow.

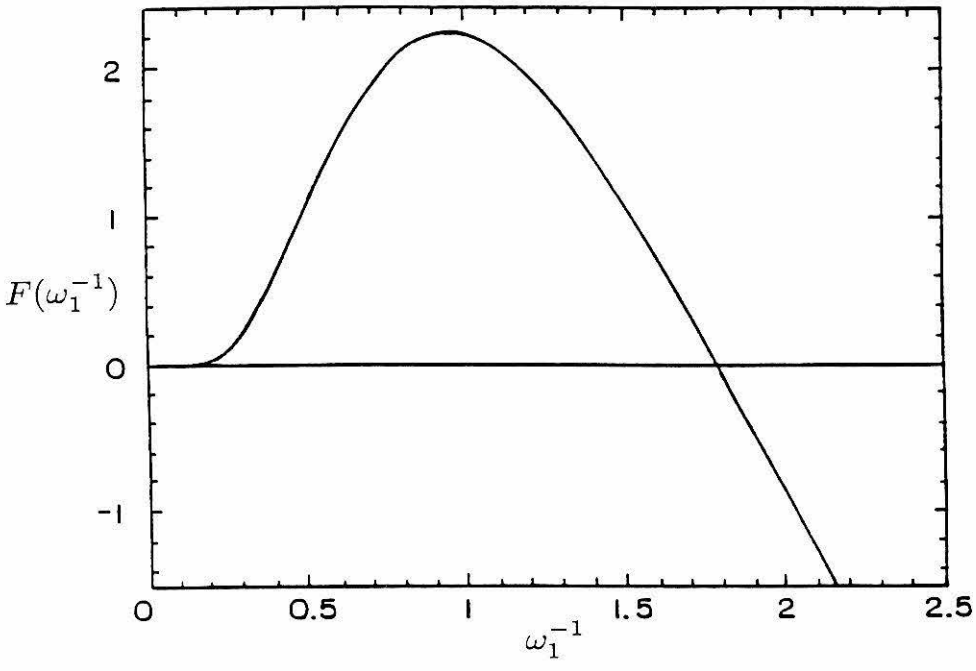


Figure 2.A1.2 The plot of $F(\omega_1^{-1})$ in equation (2.A1.6). Taken from Rom-Kedar et al. [1990].

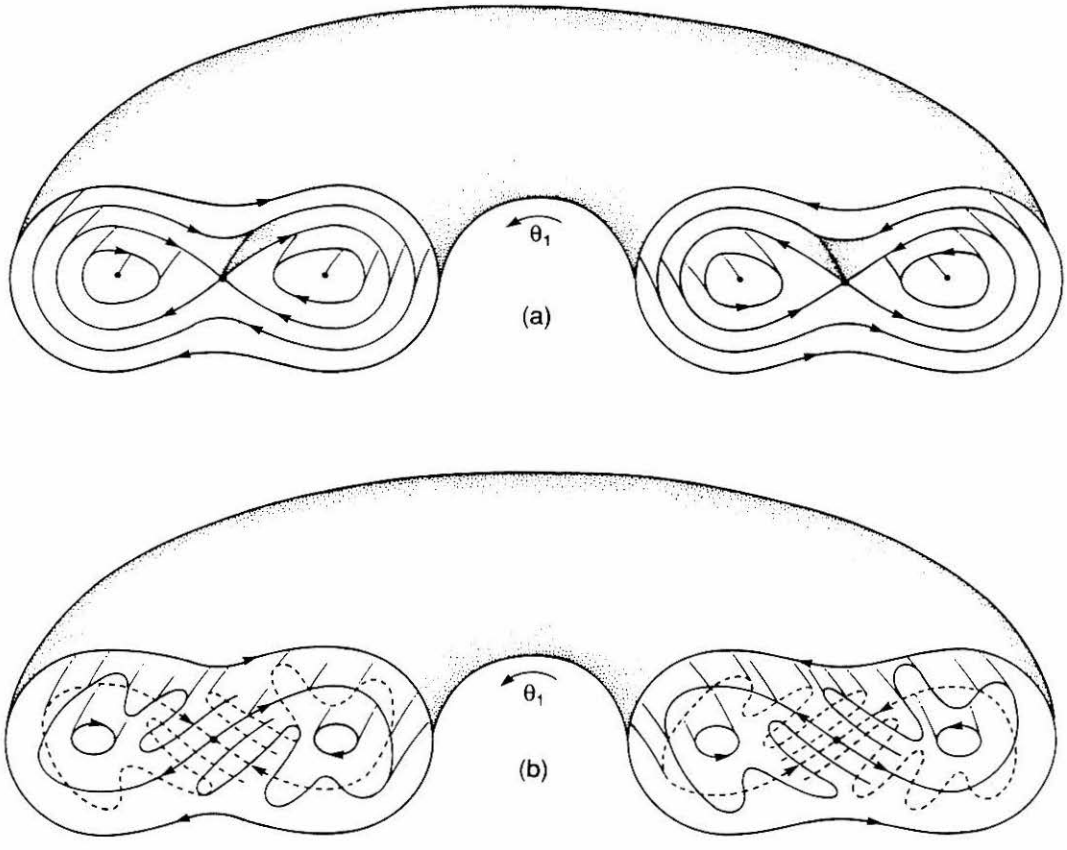


Figure 2.A1.3 (a) Some level sets in $\Sigma^{\theta_{20}}$ of the unperturbed Duffing oscillator. (b) The lobe structure in $\Sigma^{\theta_{20}}$ of the perturbed nondissipative Duffing oscillator.

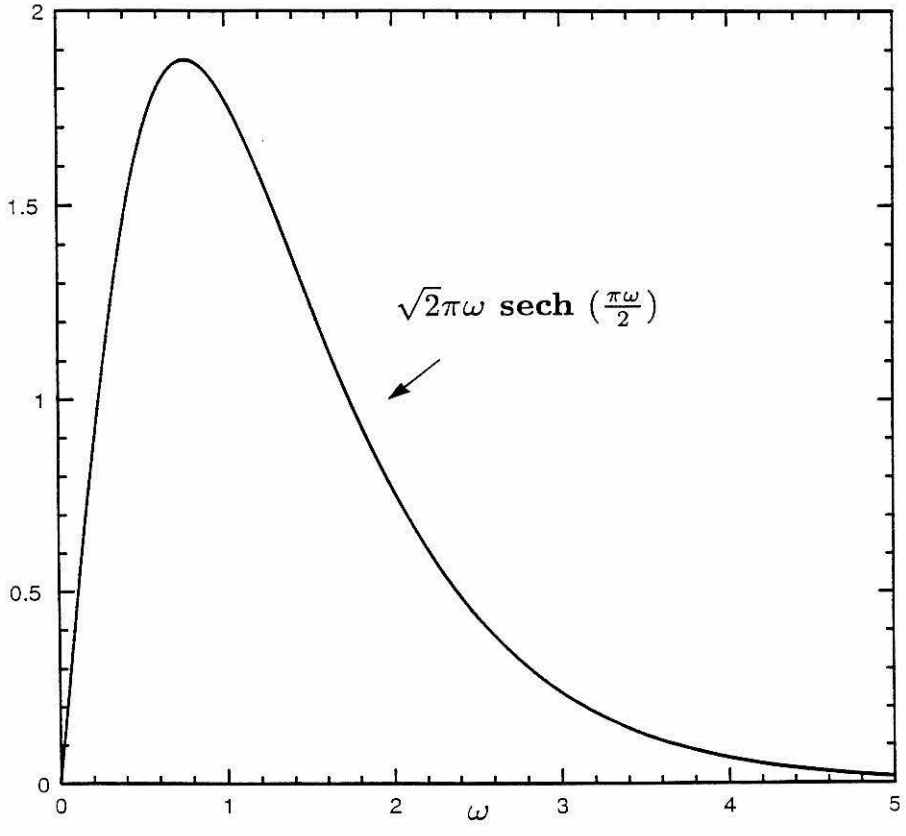


Figure 2.A1.4 The relative scaling function for the Duffing oscillator.

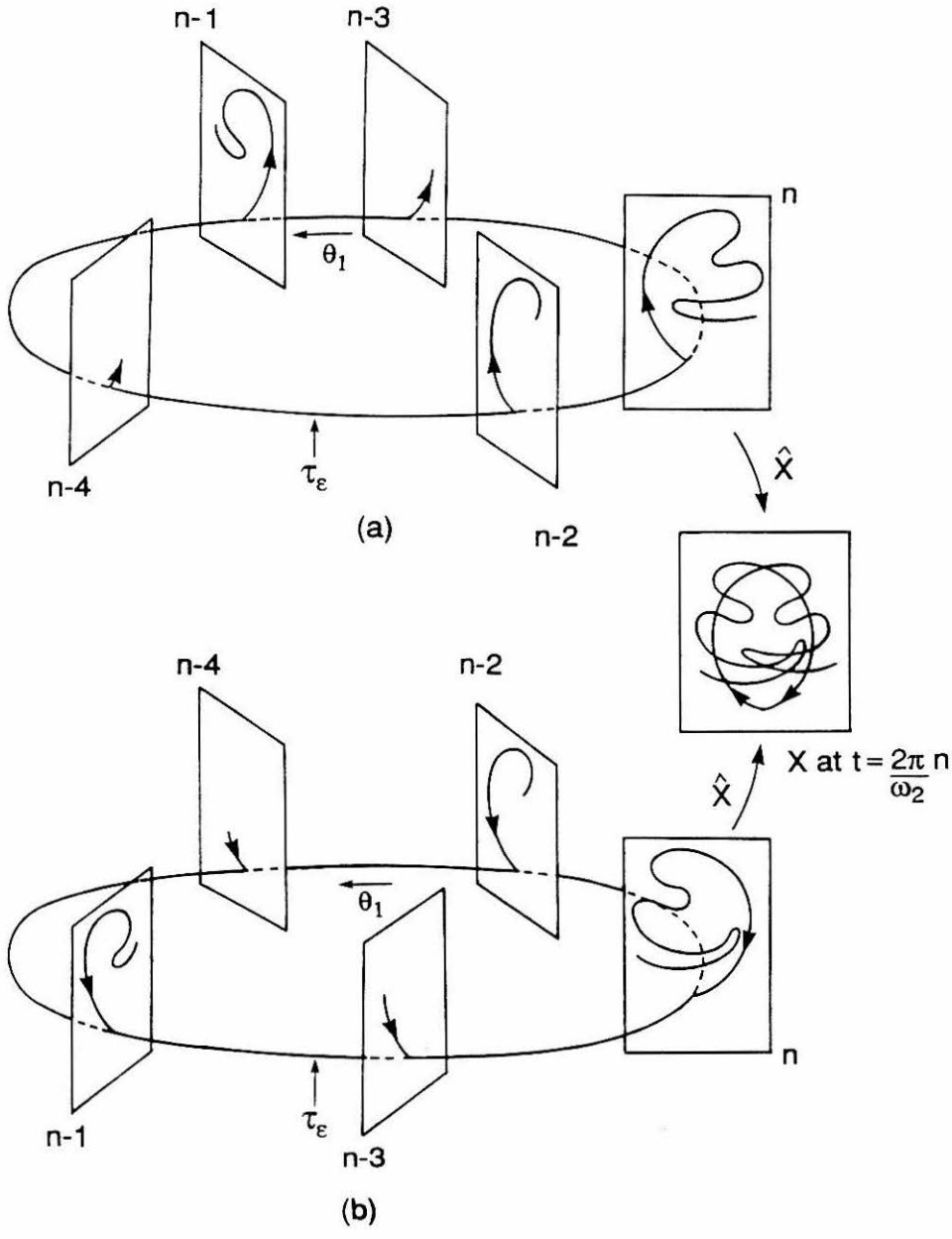


Figure 2.A4.1 Simulating a finite length of (a) $W^u(\tau_\varepsilon)$ and (b) $W^s(\tau_\varepsilon)$ in the n^{th} time slice for a two-frequency homoclinic case.

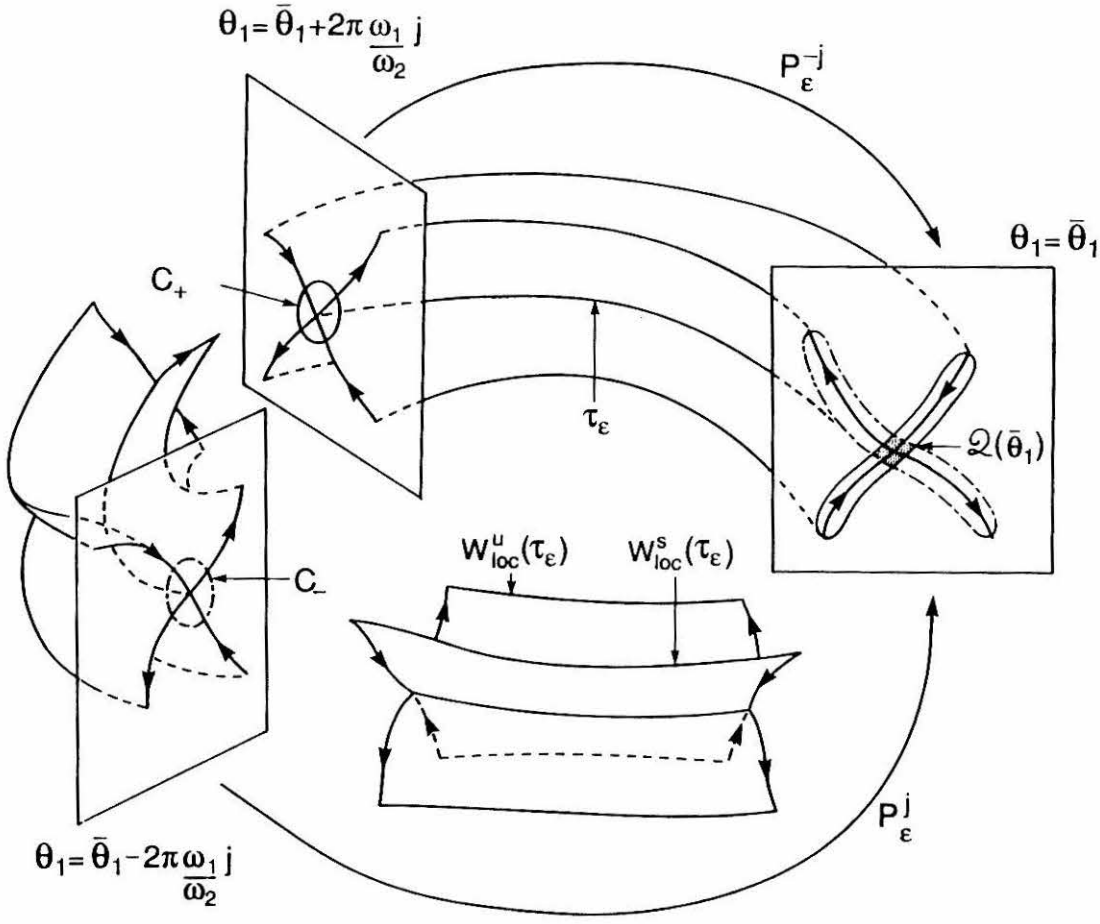


Figure 2.A4.2 Evolving $C_\pm$ to the phase slice $\chi(\bar{\theta}_1)$ in $\Sigma^{\theta_{2_0}}$ to obtain τ_ε and its local stable and unstable manifolds in that plane.

Chapter 3

A global study of enhanced stretching and mixing in chaotic tangles

Based upon an article that appears in:

Physics of Fluids A **3**, 1039-1050 (1991).

Abstract

We present results of a global, finite-time study of stretching and diffusion in a class of chaotic tangles associated with fluids described by periodically forced two-dimensional dynamical systems. Invariant lobe structures formed by intersecting global stable and unstable manifolds of persisting invariant hyperbolic sets provide the geometrical framework for studying stretching of interfaces and diffusion of passive scalars across these interfaces. In particular, we focus on the material curve that initially lies on the unstable manifold segment of the boundary of the *entraining turnstile* lobe. A knowledge of the stretch profile of a corresponding curve that evolves according to the unperturbed flow, coupled with an appreciation of a symbolic dynamics that applies to the *entire* original material curve in the perturbed flow, provides the framework for understanding the mechanism for, and topology of, enhanced stretching in chaotic tangles. Secondary intersection points (SIP's) of the stable and unstable manifolds are particularly relevant to the topology, and the perturbed stretch profile is understood in terms of the unperturbed stretch profile approximately repeating itself on smaller and smaller scales. For sufficiently thin diffusion zones, diffusion of passive scalars across interfaces can be treated as a one-dimensional process, and diffusion rates across interfaces are directly related to the stretch history of the interface. An understanding of interface stretching thus directly translates to an understanding of diffusion across interfaces. However, a notable exception to the thin diffusion zone approximation occurs when an interface folds on top of itself so that neighboring diffusion zones overlap. We present an analysis which takes into account the overlap of nearest neighbor diffusion zones, which is sufficient to capture new phenomena relevant to *efficiency of mixing*. The analysis adds to the concentration profile a saturation term which depends on the distance between neighboring segments of the interface. Efficiency of diffusion thus depends not only on efficiency of stretching along the interface, but on how this stretching is distributed relative to the distance between neighboring segments of the interface.

3.1 Introduction

The study of fluid mixing under chaotic advection encourages a global, finite-time analysis of dynamics. Such analysis requires recasting some standard methods of investigation. For example:

(i) Much of dynamical systems analysis, in fact the definition of chaos and many of the tools for measurement, such as Lyapunov exponents, involve infinite-time concepts; however, infinite-time measures may have little relevance to fluid mixing on finite time scales.

(ii) Typically one establishes analytically the presence of chaos in two-dimensional flows via the horseshoe map construction. The chaotic motion, and the symbolic dynamics used to describe the motion, thus applies to an invariant Cantor set in phase space. To study fluid mixing, however, one is less interested in the dynamics of a Cantor set of points than in the motion of entire interfaces or material blobs.

(iii) Some common measures of mixing, such as efficiency of stretching (see Ottino[1,2]), are local in nature. The study of enhanced stretching is greatly facilitated by a consideration of the global geometry underlying the chaotic zones.

We present results of a global, finite-time study of stretching and diffusion in a class of chaotic flows: two-dimensional fluids that evolve according to time-periodic velocity fields. We specialize to near-integrable systems whose unperturbed solution contains a hyperbolic fixed point connected to itself by one or more homoclinic orbits, or a pair of hyperbolic fixed points connected to each other by two or more heteroclinic orbits. In particular, we consider two fluid flows induced by oscillating vortex pairs as prototypes of fluid mixing in open and closed flows. Though we consider a highly specialized class of velocity fields, we wish to emphasize the robustness of the analysis. For example, the

framework for studying global dynamics within chaotic tangles via invariant manifolds extends to velocity fields with more complicated time dependences[3] (as discussed in Chapter 2), to certain classes of higher-dimensional flows,[4] and to cases with an arbitrary number of invariant hyperbolic sets of flows that need not be near-integrable[5] (simple examples of these extensions include two-frequency quasiperiodic velocity fields, and three-dimensional fluids with normally hyperbolic invariant 1-tori rather than hyperbolic fixed points). For concreteness, however, we focus on a simple class of velocity fields.

Consider the two-dimensional fluid flows induced by (i) a pair of equal and opposite point vortices and (ii) a pair of identical point vortices oscillating periodically in response to a time-periodic strain-rate field. We refer to these two flows as the open and closed flow, respectively, whose stream functions, in the comoving and corotating frames, respectively, can be written as

$$\psi_o = -\frac{\Gamma}{4\pi}\Big(\ln\big[(x-x_v)^2+(y-y_v)^2\big] - \ln\big[(x-x_v)^2+(y+y_v)^2\big]\Big) + \varepsilon x y \sin(2\pi t) - \tilde{v} y \tag{3.1a}$$

for the open flow and

$$\psi_c = -\frac{\Gamma}{4\pi}\Big(\ln\big[(x-x_v)^2+(y-y_v)^2\big] + \ln\big[(x+x_v)^2+(y+y_v)^2\big]\Big) + \varepsilon x y \sin(2\pi t) + \frac{\tilde{\omega}}{2}(x^2+y^2) \tag{3.1b}$$

for the closed flow, where (x_v, y_v) represents the spatial coordinates of one of the vortices in the pair (whose evolution is easily determined from the above streamfunctions with the self-contribution term ignored - refer back to Appendix 2.A1 for the open flow, which should make the closed flow apparent), Γ is the magnitude of circulation associated with each vortex, ε is the strain rate amplitude, $\tilde{v}$ is the mean vortex pair velocity in the open flow lab frame, and $\tilde{\omega}$ is the mean vortex pair rotation frequency in the closed flow lab frame. The streamlines of the unperturbed systems in these time-independent frames are shown in Figure 3.1(a). Periodic forcing drastically alters the dynamics

near the unperturbed separatrices, and it is convenient to study the dynamics by sampling trajectories periodically in time (with period equal to that of the strain-rate field, $\Delta t = 1$), which defines a Poincaré map, denoted by P_ϵ for the perturbed system. For small enough perturbations the hyperbolic fixed points of the map persist, and if their global stable and unstable manifolds intersect once they intersect in a countable infinity of points to produce the boundary of a complicated two-dimensional lobe structure (the so-called heteroclinic/homoclinic tangle) that is invariant under the Poincaré map (see Figure 3.1(b)). With each application of P_ϵ (*i.e.*, from one time sample to the next), lobes map from one to another within the invariant lobe structure in an orientation preserving manner, and this provides a global picture of the dynamics, *i.e.*, of *transport in phase space* (which is identical to transport in *physical* space in the case of chaotic advection). There is a wealth of features of chaotic dynamics that can be studied within this framework; we concentrate here on the stretching of interfaces and the diffusion of passive scalars across these interfaces.

3.2 Interface dynamics

3.2.1 Lobe dynamics

We first provide a more detailed description of the global dynamics in chaotic tangles with the help of Figure 3.2. An invariant core boundary, approximately equal to the unperturbed core boundary formed by the separatrices, is defined by segments of the stable and unstable manifolds, as shown by the dashed lines. The stable and unstable manifolds intersect in two classes of points: *primary intersection points (PIP's)* (marked by dots) and *secondary intersection points (SIP's)* (marked by crosses). The nature of these two types of intersection points should be clear from the figure (see Wiggins[7] for a discussion of intersection points); both the PIP's and SIP's will be seen to be relevant

to understanding the topology of enhanced stretching. There are two types of lobes, which we distinguish as *entraining* and *detraining*: the m^{th} ($m \geq 1$) entraining lobe, $E(m)$, is *entrained into* the core (mapped from outside to inside the core) upon the m^{th} iterate of P_ε; the m^{th} ($m \geq 1$) detraining lobe, $D(m)$, is *detrained from* the core (mapped from inside to outside) upon the m^{th} iterate of P_ε (in a similar manner one defines the lobes for $m \leq 0$, as should be clear from Figure 3.2). One pictures transport in phase space in Figure 3.2 by recognizing that the lobes map according to:

$$P_\varepsilon(E(m)) = E(m-1)$$

$$P_\varepsilon(D(m)) = D(m-1)$$

(these transport equations follow from the invariance of the manifolds, the orientation-preserving nature of the Poincaré map, and the choice of sampling, as explained in Rom-Kedar et al.[6]). We refer to $E(1)$ and $D(1)$ as *turnstile lobes*, and these lobes are the *only* mechanism for transport in and out of the core under P_ε. The segments of the lobe boundaries formed by the stable manifold behave asymptotically in a simple manner: they asymptote to the left hyperbolic fixed point and contract asymptotically exponentially in time. The segments of the lobe boundaries formed by the unstable manifold behave in a complicated manner, and it is these interfaces whose dynamics we wish to study. Indeed, an understanding of the stretch properties of these interfaces provides a good understanding of stretching throughout the tangle region, since the unstable manifold acts somewhat like an attractor.[3,6] Of course, since the dynamical system describing the fluid motion is Hamiltonian (the fluid is incompressible), the unstable manifold is not truly an attractor; however, the strong stretching and contraction of the fluid lobes implies that, though area elements do not shrink, they tend to be stretched in one direction and contracted in another such that the unstable manifold dominates the evolution of material curves in the tangle region.

3.2.2 Stretch profiles and a symbolic dynamics for the entire interface

From the way the lobes map from one to another, it suffices to consider the segment of the unstable manifold along the boundary of the *turnstile* lobes; for simplicity we consider the material curve that starts off ($t = 0$) on the segment of the unstable manifold that bounds the entraining turnstile lobe, which we denote by $r_0(t)$ (the endpoints of this line segment at $t = 0$ are the PIP's $p1$ and $p2$, as marked in Figure 3.2). The behavior of r_0 is best understood by first making a comparison with an unperturbed case; for example, consider the evolution according to the *unperturbed* flow of a material curve $\tilde{r}_0(t)$ which at $t = 1$ lies on the segment of the *perturbed* unstable manifold that is essentially $P_\epsilon(r_0(t = 0))$, except that its endpoints $\tilde{p}1$, $\tilde{p}2$ are the points which asymptote to the unperturbed fixed point rather than the perturbed fixed point (see Figure 3.3). This material curve winds around the core ad infinitum, never intersecting the unperturbed core boundary, except at $\tilde{p}1$ and $\tilde{p}2$. Since the endpoints asymptote to a hyperbolic fixed point, the infinitesimal line elements of $\tilde{r}_0$ associated with the endpoints stretch asymptotically exponentially in time; however, overall $\tilde{r}_0$ stretches only asymptotically linearly in time. Hence as time progresses, the stretch profiles in Figure 3.3 will be peaked more and more towards the endpoints, with an overall dip in the middle corresponding to poor stretching. In the perturbed case, r_0 is entrained into the core at $t = 1$, and then for near-integrable flows starts to wind around the core not much unlike $\tilde{r}_0$ in the unperturbed case, except that an arbitrarily small time-periodic modulation of the velocity field suffices to cause r_0 to intersect the stable manifold segment of the perturbed core boundary (these intersection points are SIP's), as shown heuristically in Figure 3.2 for the open flow (numerical simulations are shown presently). The SIP's have a two-fold significance: first, like the PIP's they asymptote to a hyperbolic fixed point and thus the infinitesimal line

elements of r_0 associated with the SIP's stretch asymptotically exponentially in time; second, they are the dividing points between segments of r_0 that behave in a qualitatively different fashion, as shown in Figure 3.2:

(i) A "returned" part r_0r_3, at $t = 3$ between SIP's a, b and e, f, and which at $t = 3$ is close to $r_0(t = 0)$, and hence *will evolve from time* $t = 3$ *approximately the way* r_0 *does from time* $t = 0$ *(the SIP's playing the role that the PIP's did originally for* r_0*).* The subscript on the second r specifies the number of time samples it takes for $r_0r_3(t = 0)$ to "return" to $r_0(t = 0)$ (the subscript on future symbols should be clear from this comment).

(ii) A detrained part r_0d_3, at $t = 3$ between SIP's b, c and d, e, which will remain outside the core and stretch asymptotically linearly in time.

(iii) A tail, r_0t_3, at $t = 3$ between SIP's c, d, which will wind around the core as its SIP's asymptote to the left fixed point, but in a way that is qualitatively different from r_0. (Though a general scheme needs to include a tail, for some system parameters there will be no tail, as can occur for example if in Figure 3.2 the tip of $E(-2)$ lies in $D(0)$.)

(iv) The remainder of r_0, to which we add no further labels for the moment.

In this manner we build up a symbolic dynamics, so to speak, that applies to the *entire material curve*, as shown heuristically in Figure 3.4. To understand how this is built up, one needs to recognize:

(i) With each additional application of P_ε, a portion of the remainder of r_0 will be "returned" and another portion detrained. Hence, there will be an r_0r_4 and an r_0d_4 after the fourth iterate of P_ε, then an r_0r_5 and an r_0d_5 after the fifth iterate of P_ε, and so on. Referring to Figure 3.2, it is clear from the lobe dynamics that at $t = 3$ r_0r_n, $n \geq 3$, is a portion of the boundary of $E(-2)$ between $D(n-3)$ and $D(n-2)$, and r_0d_n is a portion of the boundary of $E(-2)$ inside $D(n-3)$. Hence the r_0r_n's and r_0d_n's join

contiguously to cover r_0, as shown by the upper line in Figure 3.4. These pairs of symbols on the upper line pertain to the first revolution around the core of successive pieces of r_0.

(ii) The tail and returned parts will further subdivide qualitatively the same way as does r_0 starting at $t = 0$; the detrained parts do not keep subdividing. When we say that further subdivisions are qualitatively similar, we should note that successive revolutions around the core may take a different number of iterates of P_ϵ for the first return to occur, and that these numbers may in fact differ (by at most one) on either side of the tail (when this occurs we split the tail in two and have the t subscript on each side agree with that of the adjacent detrained part). Also, for large enough perturbations, though the tail subdivisions will for large enough n settle down to the alternating sequence of $r_0 \ldots t_m r_n$, $r_0 \ldots t_m d_n$, $r_0 \ldots t_m r_{n+1}$, $r_0 \ldots t_m d_{n+1}$, *etc.*, for small n the ordering may be a bit erratic, due to the possibly complicated behavior of the tail. These minor complications do not detract from the overall usefulness of the symbolic dynamics in understanding the topology of enhanced stretching.

The result is a dense labeling of the interface, consisting of a string of symbols that terminate if a 'd' is reached. For example, the string $r_0 r_7 r_{20} d_5$ denotes the segment of r_0 which takes 7 time samples to wind around the core and "return" to r_0, then another 20 time samples wind around the core once more and "return", and then 5 time samples to wind around the core one last time to be detrained. The dynamics is thus described by a string of symbols, with three types of entries: r, d, and t. The differences from conventional symbolic dynamics are that the strings may be finite (if they contain a d), and that there is a subscript on each entry which specifies the time interval between successive entries. These differences in time scales, as we shall later discuss, are highly relevant to understanding the dynamics within the tangle regions. One can

construct a similar symbolic dynamics for the closed flow, with two differences:

(i) Detrained segments can be re-entrained, so detrained segments will further subdivide, and the string of symbols will not terminate at a d.

(ii) The segment r_0 can pass through both the upper and lower core, which can be kept track of by using, say, small letters for the upper core and capital letters for the lower core.

The symbolic dynamics, coupled with a knowledge of the unperturbed stretch profile, provides a framework for studying the mechanism for, and topology of, enhanced stretching in chaotic tangles. As described earlier, the "returned" segments evolve approximately like $r_0(t = 0)$, so one understands the enhancement of the stretch profile in terms of the unperturbed profile approximately repeating itself on smaller and smaller scales, the topology of this repetition understood via the symbolic dynamics (refer back to Figure 3.4, and see Figure 3.5 for a heuristic portrayal). For the open flow, the detrained segments stretch asymptotically linearly in time, so the stretch profile will contain gaps of poor stretching, the ordering of which is again understood by the symbolic dynamics.

We present in Figure 3.6 plots of r_0 and its associated stretch profile with symbolic dynamics labeling for selected time samples of a short-time simulation of the open and closed flow. Figures 3.7 and 3.8 show *each* time sample of the simulation, without detailed labeling, to help convey the nature of the interfacial stretching. The closed flow clearly shows the highly non-uniform unperturbed stretch profile approximately repeating itself on smaller and smaller scales. The SIP's entering the region local to the hyperbolic fixed point play the role of the endpoints in the unperturbed case, and with increasing time there appear sharp peaks of extremely good stretching around these SIP's. One even sees the small dip in the unperturbed profile near the endpoints appear near the SIP's in the perturbed profile. Since the open flow unperturbed profile is more

uniform, it is more difficult to decipher the perturbed stretch profile without a somewhat belabored discussion, but with the help of SIP portrayal and symbolic dynamics labeling in Figure 3.6 one can make some basic observations. The detrained segments clearly begin to form pockets of poor stretching, and the returned segments do show evidence of approximate self-similar behavior. An obvious example of this latter behavior is the sharp dip of poor stretching associated with the "tip" of the lobe formed by $r_0(t)$ appearing in the returned segments when they perform another revolution around the heteroclinic core. More specifically, note in Figure 3.6 the open flow stretch profile of $r_0(t = 3)$ approximately repeating itself on smaller scale in the stretch profile of $r_0r_3(t = 6)$: both show four significant peaks separated by significant dips, the middle dip being quite pronounced. Of course the repetition is indeed only approximate. For example, the middle dip of $r_0r_3(t = 6)$ associated with the "tip" is much wider than the corresponding dip of $r_0(t = 3)$, since the tail of the former is much larger as a consequence of $r_0r_3(t = 3)$ lying further inside the heteroclinic core than $r_0(t = 0)$. Additionally, since the stretch profile of $r_0r_3(t = 3)$ has a mild non-uniformity, this non-uniformity carries through to give an additional mild non-uniformity in $r_0r_3(t = 6)$ not found in $r_0(t = 3)$.

It is with these qualifying comments that we hope the spirit of our notion of approximate self-similar repetition is understood. Within this setting, we make a number of observations about the stretching.

(i) Notice the relationship between the non-uniformity of the unperturbed stretch profile and the non-uniformity of the perturbed profile. For example, the unperturbed closed flow profile, for the parameters chosen, has good stretching highly localized around the endpoints, with a large dip of poor stretching between the endpoints; hence the perturbed profile has good stretching highly localized around the SIP's, with significant dips of poor stretching between SIP's. In contrast, the unperturbed open flow

profile is more even (due in part to the presence of the other hyperbolic fixed point); hence the perturbed profile has a more even distribution of good stretching. We see then that *the unperturbed stretch profile provides qualitative and yet basic information about the non-uniformity of the perturbed profile*, which is highly relevant to efficiency of mixing in the context of diffusion across interfaces, as we will describe in the next section, and to the statistics of stretching, as we will study in Chapter 6. Primarily in this context is our interest in approximate repetition on smaller and smaller scales.

(ii) The mechanism for enhanced stretching can be understood partially in a variety of ways, and the different flows highlight these different ways. For example, consider *residence time near the hyperbolic fixed points.* The creation of SIP's by external forcing entails a creation of a countable infinity of points along the interface that eventually reside near the fixed point for all time, and hence entails an overall increase in the interface's residence time near the hyperbolic fixed points. This mechanism is dominant in the closed flow, where the stretching is highly localized around SIP's. Consider also *the repeated stretch, fold and return mechanism*, which need not be tied to residence near the hyperbolic fixed point. For example, referring back to Figure 3.2, after three perturbation periods r_0 has been stretched and folded. The portion $r_0 r_3$ will repeat the stretch and fold process, as will then $r_0 r_3 r_3$, and so on. As long as the concerned segment stretches by a factor greater than one with each fold and return, it will be undergoing exponential stretching. Here, then, exponential stretching is not necessarily due to good local stretching near a hyperbolic fixed point, but to the repeated stretch and fold mechanism. This mechanism is highlighted by the open flow, where stretching is more evenly distributed between the SIP's. Note that the stretch profile of $r_0 r_3 (t = 3)$ need not be uniformly

greater than one, so that one will have to compute the profile (and then further profiles) to determine which portions of r_0 are candidates for exponential stretching. In such a manner one determines "where the horseshoe maps are", so to speak. A third explanation is that turnstile lobes act in a certain sense as *reorientation* lobes. As segments of r_0 wind around the core, their stretch efficiency decays not much unlike the unperturbed case; however, when they reach the turnstile lobes they are then mapped by P_ϵ in such a way as to reorient many of the line elements to increase stretch efficiency (as described in Chapter 4). All of the above mentioned mechanisms for stretch enhancement are really part of the same overall mechanism we previously mentioned, that is the *approximate self-similar behavior of* r_0 *as it evolves.*

(iii) Notice the great variation of stretch histories (the vertical scale of the profiles in Figures 3.6-3.8 is logarithmic) and the spatial complexity of the interface produced in a very short time by a very simple velocity field. As we shall discuss, mixing in the context of diffusion of passive scalars across an interface depends on the stretch *histories* of the interface, and, as the simulations in Figures 3.6-3.8 suggest, there may be little obvious connection between a velocity field's properties and the corresponding advection properties. In particular, simple velocity fields can of course produce complicated advection, and more generally the scaling properties of turbulent velocity fields may have little to say about the mixing properties. In addition, the presence of hyperbolicity in the chaotic tangles implies substantial temporal variation of the strain experienced by an infinitesimal line element, as well as great spatial variation of strain history along a material curve. In studies of turbulent mixing, sometimes an *assumption* of constant strain is made at some point (see for example Batchelor[8] for one of the early well-known works using this assumption); such an assump-

tion certainly would not apply in chaotic tangles, and we see no particular reason for it to be valid for turbulent flows.

(iv) Though previous investigators have commonly referred to chaotic zones as regions of "enhanced" stretching, notice how in both the open and *closed* flows there are gaps of *poor* stretching on finite time scales. As previously mentioned, these gaps in the open flow are partially explained by detrainment. For both flows, however, the additional explanation is offered by the previously described dip in the unperturbed stretch profile, which carries through to the perturbed profile by the previously described approximate self-similar enhancement of the stretch profile. Note how the regions of poor stretching typically correspond to "bulbous" parts of the lobe, and turning points such as the "tip" of the interface. These bulbous parts can rotate throughout the chaotic zones with ineffectual stretching on finite time scales, not coming close to the fixed point, and not undergoing the repeated stretch and fold mechanism. We should stress here a basic difference between the open and closed flow: in the closed flow the chaotic zone associated with the homoclinic tangle is finite, re-entrainment occurs, and poor stretching eventually gives way to enhanced stretching; in the open flow the chaotic zone associated with the heteroclinic tangle is infinite, and detrainment leads to advection out to $x = -\infty$ in an asymptotically linear manner, so that poor stretching can remain so for all time.

(v) Note the variety of stretch scales, spatial scales, and temporal scales involved in the interface stretching. Basically r_0 is undergoing a repeated stretch and fold process, with the magnitude of stretching depending on the location relative to the fixed point, and this stretch, fold, and return process takes place on all time scales (above some minimum time), as monitored by the symbolic dynamics: for example, $r_0 r_n(t = 0)$, $n > n_{min}$, takes n time samples for the first stretch, fold, and return. Hence horseshoe maps

occur on all time scales above some minimum time. The variety of stretch scales and spatial scales is indicated by the non-uniformity of the stretch profiles. For each revolution around the homoclinic/heteroclinic core, one can have anything from *contraction* on sizable spatial scales to extremely good stretching on very small spatial scales. This range of scales will be relevant to the statistical analysis of stretching given in Chapter 6.

(vi) Consider the total length of the interfaces, as shown in Figure 3.9. Up until the first "returned" portion of r_0 wraps around $E(0)$ (*i.e.*, a portion of r_0 begins its second wind around the core), the total length of r_0 does not differ significantly from that of $\tilde{r}_0$ in the unperturbed problem; however, thereafter (after $t \approx 4$ for the open flow and $t \approx 3$ for the closed flow) the length of r_0 quickly converges to mild oscillation about exponential stretching. In this context, great regularity underlies the approximate self-similar behavior of r_0.

The above observations highlight some of the relevant features of stretching as understood in the geometrical setting of invariant manifolds. With these observations as a backdrop, we wish to convey the utility of the geometrical framework offered by the invariant manifolds. Heretofore, the horseshoe map construction and its associated symbolic dynamics have been the central paradigm for enhanced stretching in the chaotic tangles of two-dimensional maps. Though establishing exponential stretching on a non-trivial subset of phase space, this construction offers little global understanding of stretching, and the study of global stable and unstable manifolds provides an extension of this analysis. The conventional horseshoe map analysis applied to chaotic tangles considers the repetition of an essentially uniform stretch and fold process in a domain that is a small subset of phase space near a hyperbolic fixed point, and ignores all pieces of the domain which map outside the domain; the analysis based on invariant manifolds considers the *approximate* repetition of highly

non-uniform stretching and folding of turnstile lobe boundaries *throughout* the chaotic zone, and monitors the *entire* material curve. For near-integrable systems, it is the unperturbed stretch profile that is approximately repeated on smaller and smaller scales, and the topology of this repetition is understood by the symbolic dynamics. The utility of both the conventional horseshoe map analysis and the invariant manifold analysis is primarily conceptual, but the latter offers a framework for more detailed quantitative analysis. Applications to realistic flows can in practice be difficult, which motivates first a study of model systems. Past models for the stretch processes under chaotic flows have typically been fairly elementary. For example, in their study of the kinematic dynamo problem under chaotic advection, Ott and Antonsen[9] assume simple enough models for the stretch processes, such as a baker's map, that stretch distributions are described by standard multifractals from a binomial multiplicative process. The invariant manifold analysis of this chapter provides a framework for much more realistic models, in which the topology of, and multiple scales involved in, fluid flows are respected. For example, one could choose a reasonable analytical model for the unperturbed stretch profile, one whose non-uniformity can be adjusted by a parameter, and have the profile repeat on smaller and smaller scales according to the symbolic dynamics (with the SIP locations chosen by some reasonable model). The result is a multifractal with a much richer construction, one that is truer to what one would expect in realistic flows. In Chapter 6 we will consider some elementary models in the context of a study of distributions of finite-time Lyapunov exponents in chaotic tangles. To study realistic flows rather than models can involve intensive computation; however, the invariant manifolds provide a framework for such computation. The goal of course, as done in Figures 3.6-3.8, is to identify the hyperbolic fixed points and local stable and unstable manifolds (for near-integrable systems there is a standard and easily implemented procedure for doing this); one then computes the global stable and unstable manifolds by evolving the local sta-

ble and unstable manifolds backwards and forwards in time, respectively, from which one can determine the turnstile lobes and hence $r_0(t=0)$ and $\tilde{r}_0(t=0)$. Simulations of $\tilde{r}_0$ illustrate the stretch profile that is approximately repeated on smaller and smaller scales in the perturbed case, and one can study and alter this profile (by changing system parameters) before going on to the perturbed simulations. Of course, to determine the topology for any given perturbed flow, such as finding the SIP's and calculating stretch profiles, one must resort to explicit numerical simulation. Such computations can become arduous for long time scales, but *it is nevertheless crucial to appreciate that an underlying geometrical framework does exist in the chaotic tangles.* At present, the studies of chaotic advection rarely, if ever, acknowledge the presence of an underlying topology in chaotic zones: typically a blob or set of points are initialized in a fairly random fashion, and the fact that finite-time stretching can vary greatly within the chaotic zones is not addressed. *A main goal of this chapter then is to encourage future studies to address the underlying geometries within chaotic tangles and the great variation of the dynamics on finite time scales.* There are a number of examples where such a consideration would be helpful. For example, in their study of separation under flow reversal, Aref and Jones[10] present a computational example illustrating that separation of a material blob is better in chaotic zones than in regular zones, and argue that this enhancement is due to the enhanced stretching in chaotic zones. A consideration of the topology of stretching on finite time scales would be useful in maximizing this effect.

We should point out that, though long-time computations can be arduous, they are not so difficult as one might first think. For example, suppose one wished to determine $\mu(r_0 r_i \ldots r_j d_k(t=n))$ for all possible combinations of $i+\ldots+j+k=n$, where $\mu(\,)$ denotes the total stretch from $t=0$. As $n \to \infty$, both the value each indice can take on and the number of possible r entries in the string go to infinity. However, it is reasonable to expect that $\mu(r_0 r_{i+1}(t=i+1))/\mu(r_0 r_i(t=i))$ asymptotes to a simple relation as $i \to \infty$ and

that for sufficiently small perturbations one can find upper and lower bounds on how the first revolution around the core (*i.e.*, $\mu(r_0 r_i(t = i))$) describes later revolutions (due to the approximate self-similar behavior of r_0). With these two properties, a finite set of quantities ($\mu(r_0 r_i(t = i))$ for a finite number of i's) provides a reasonable characterization of the stretching. We should stress however that, as with the conventional horseshoe map construction and other paradigms of dynamical systems theory, the goal of our stretch analysis is not to provide a large amount of data, but rather to provide a conceptual framework with which to understand the order that underlies the chaotic dynamics.

We end this section by stressing that for classes of three-dimensional flows and flows with more general velocity field time dependences one can define from invariant manifolds sets of lobes which map from one to another and obtain from this a global picture of enhanced stretching. To do so requires recent and fairly technical developments in dynamical systems theory, involves much more complicated geometries, and is beyond the scope of this chapter. We will briefly discuss, however, the two simple cases mentioned in the introduction. For a two-frequency velocity field,[3] one needs to generalize from *maps* to *sequences of maps* (refer back to Chapter 2). Transport in phase space is described in terms of lobes mapping within a *sequence* of lobe structures that are derived from invariant manifolds embedded in a higher-dimensional Poincaré section. For a three-dimensional system with normally hyperbolic invariant 1-tori[4] (*i.e.*, the hyperbolicity is in a plane "normal" to the 1-torus, and there is no exponential stretching "along" the torus), the extension from two dimensions is fairly robust, with three-dimensional lobes stretching and folding in two dimensions, and negligible stretching in the third dimension.

3.3 Diffusion across interfaces

Let us now consider mixing in the context of diffusion of passive scalars across an interface (the analysis extends to passive vectors, which will be the subject of future work). For example, a natural problem starts off with fluid A outside the perturbed core and fluid B inside the perturbed core; we consider a simpler scenario that starts off with fluid A in the entraining turnstile lobe and fluid B everywhere else (this allows us to concentrate on a single lobe, and from the lobe dynamics it should be clear how the latter problem relates to the former). As these fluids advect according to the open or closed flow, let them diffuse, according to the standard diffusion equation, across the interface defined by the material curve that initially lies on the entraining turnstile lobe boundary. This curve thus consists of r_0 and a segment of the stable manifold; the stable manifold part of the curve, as seen in Figures 3.6-3.8, quickly shrinks to negligible length relative to r_0, and so with negligible error we ignore the diffusion across the stable manifold segment and focus only on r_0. For thin enough diffusion zones and sufficiently steep concentration gradients across the interface, which is valid for a given diffusion coefficient for small enough times, the diffusion process is essentially normal to the interface, and can be treated as one-dimensional. The amount diffused across the interface can then be related to the stretch history of the interface.[1,2,11–13] The concentration of fluid A (and similarly fluid B) spreads with increasing time like an error function (see Figure 3.10(a)):

$$\Phi_A(\mathbf{p}, y, t) = \frac{1}{2}\Big\{1 + \mathrm{erf}\Big[\frac{S(\mathbf{p},t)y}{(4D\tau(\mathbf{p},t))^{\frac{1}{2}}}\Big]\Big\}, \tag{3.2}$$

where $\mathbf{p}$ specifies a point on the interface, y is a relative coordinate normal to the interface at $\mathbf{p}$ and pointing into the lobe, D is the diffusion coefficient, $S(\mathbf{p},t) = \delta s(\mathbf{p},t)/\delta s(\mathbf{p},t=0)$ is the interface stretch at $\mathbf{p}$, and

$$\tau(\mathbf{p},t) = \int_0^t S^2(\mathbf{p},\tilde{t})d\tilde{t}. \tag{3.3}$$

The diffusion rate per unit initial arclength across the interface at $\mathbf{p}$, $C(\mathbf{p},t)$, is given by

$$C(\mathbf{p},t) = D\frac{\partial \Phi_A}{\partial y}|_{y=0} S = \frac{1}{2}\left[\frac{D}{\pi\tau}\right]^{\frac{1}{2}} S^2. \tag{3.4}$$

The amount of fluid A diffused across the interface at $\mathbf{p}$ per unit initial arc length is then

$$\mathcal{A}(\mathbf{p},t) = \int_0^t C(\mathbf{p},\tilde{t})d\tilde{t} = \left[\frac{D\tau}{\pi}\right]^{\frac{1}{2}}. \tag{3.5}$$

Hence, the local time-dependent diffusion rate and the amount diffused across the interface is completely determined by the local stretch history of the interface. The total amount of fluid diffused out of the lobe is found by integrating equation (3.5) over all $\mathbf{p}$. Enhanced interface stretching corresponds to enhanced diffusion across the interface, and a global understanding of stretching directly translates into an understanding of diffusion. However, this analysis is valid only for sufficiently thin diffusion zones, and a notable exception to this case occurs when the interface folds back onto itself so that the diffusion zones overlap.[13] As should be clear from the plots in Figures 3.6-3.8, in chaotic zones interfaces quickly fold and wrap around themselves in a violent manner, and one would thus like to employ an analysis that addresses overlap of neighboring diffusion zones. A full analysis that takes into account the overlap of all neighboring diffusion zones would be quite laborious. Instead we employ an analysis that takes into account the overlap of a *nearest neighbor* diffusion zone, which is sufficient to capture additional phenomena relevant to the notion of *efficiency of mixing*. The key to such an analysis is to notice, as seen in Figures 3.6-3.8, that much of the neighboring sections of the interface are essentially parallel to one another and have small curvature (typically these portions of the interface have good stretching); in addition, the parts of high curvature tend to be bulbous and hence have less of a tendency for overlap of diffusion zones (typically these portions have poor stretching). Under the assumption that overlap of neighboring diffusion zones is not significant until the neighboring segments of

the interface are essentially parallel, which is valid for a small enough diffusion coefficient, the diffusion process can still be treated as one-dimensional. We address overlap of diffusion zones when it is due to the lobe being sufficiently thin; we do not worry about overlap due to the lobe folding on top of itself. When there is overlap at $\mathbf{p}$, equation (3.2) is replaced by a superposition of two error functions

$$\Phi_A(\mathbf{p},y,t) = \frac{1}{2}\Big\{\mathrm{erf}\Big[\frac{Sy}{(4D\tau)^{1/2}}\Big] - \mathrm{erf}\Big[\frac{S_d(y-d)}{(4D\tau_d)^{1/2}}\Big]\Big\}, \tag{3.6}$$

where $d = d(\mathbf{p},t)$ is the normal distance at $\mathbf{p}$ between the two neighboring segments of the interface whose diffusion zones overlap, and S is understood to represent $S(\mathbf{p},t)$, while $S_d \equiv S(\mathbf{p}+d\hat{y},t)$ (and similarly for τ and τ_d). Note that in the limit $d \to \infty$ (with y fixed) the second error function goes to -1, and hence equation (3.5) recovers equation (3.2). The concentration spreads with increasing time like a sum of two error functions, one associated with each interface (see Figure 3.10(b)). By a calculation similar to that in equation (3.4), except that one has to now take into account the presence of the other interface, the diffusion rate out of the lobe per unit initial length of neighboring segments of the lobe boundary is

$$C(\mathbf{p},t) = \frac{1}{2}\Big[\frac{D}{\pi}\Big]^{\frac{1}{2}}\Big\{\frac{S^2}{\tau^{\frac{1}{2}}}\Big(1-\exp\Big[\frac{-(Sd)^2}{4D\tau}\Big]\Big) + \frac{{S_d}^2}{{\tau_d}^{\frac{1}{2}}}\Big(1-\exp\Big[\frac{-(S_d d)^2}{4D\tau_d}\Big]\Big)\Big\}. \tag{3.7}$$

The presence of a neighboring interface thus adds to the diffusion process a saturation term of the form $(1-\exp[\frac{-(Sd)^2}{4D\tau}])$, so that the separation distance d is a factor in the diffusion process. To find the total amount of fluid diffused out of the lobe, a convenient trick, with physical meaning, is to separate the two terms in equation (3.7) and then integrate over *all* $\mathbf{p}$ in the interface, rather than integrate (3.7) over half the $\mathbf{p}$ of neighboring segments of the interface whose diffusion zones overlap. Hence for all $\mathbf{p}$ we write

$$C(\mathbf{p},t) = \frac{1}{2}\Big[\frac{D}{\pi\tau}\Big]^{\frac{1}{2}} S^2\Big(1-\exp\Big[\frac{-(Sd)^2}{4D\tau}\Big]\Big), \tag{3.8}$$

where it is understood that when there is negligible overlap at $\mathbf{p}$ we send $d \to \infty$. The amount of fluid A diffused across the interface at $\mathbf{p}$ per unit initial arc length, $\mathcal{A}(\mathbf{p}, t)$, is then given by integrating $C(\mathbf{p}, t)$ in time as done in equation (3.5), and then the total amount of fluid A that leaks out of the lobe is given by integrating $\mathcal{A}(\mathbf{p}, t)$ over *all* $\mathbf{p}$. Note that, to determine $\mathcal{A}(\mathbf{p}, t)$ at a particular time, one can of course also deal with a spatial integral of the concentration $\Phi_A(\mathbf{p}, y, t)$ rather than a time integral of $C(\mathbf{p}, t)$; the advantage of the former approach is that one need calculate $d(\mathbf{p}, t)$ only at that given time.

We present in Figure 3.11 some plots of $\mathcal{A}(\mathbf{p}, t)$ when r_0 advects under the same velocity fields as considered in the examples of Figures 3.6-3.8. Figure 3.12 shows the total amounts of fluid A that diffuse out of the lobe. The saturation term is clearly relevant to the efficiency of the diffusion process. For example, in the closed flow simulation, though the interface stretches more than in the open flow, this stretching is highly localized in a few regions where the lobe is extremely thin, and the diffusion process quickly saturates since there is little fluid available to diffuse out of these parts of the lobe. Hence, we see from $t = 4$ to $t = 5$ there is little enhancement of diffusion since, though there is a great amount of stretching going on, it is localized in regions where the diffusion process has saturated (note that as more SIP's asymptote to the hyperbolic fixed point, the diffusion process will again be enhanced, so that the overall enhancement is staggered in time). This is in contrast to the open flow example, where the stretching is more evenly distributed, and saturation is less of a factor. *One sees then that, in the context of diffusion, the efficiency of mixing depends not only on the efficiency of stretching along the interface, but on how this stretching is distributed relative to the separation of neighboring segments of the interface.* This is highlighted in Figure 3.12, where we see that, even though r_0 stretches more in the closed flow, the open flow has a greater amount of diffusion on the time scale shown (the greater saturation in the closed flow is due in part to the smaller turnstile lobe area, but more importantly to

the greater non-uniformity of the stretch distribution). On longer time scales the differences between a finite and infinite mixing region (the tangle) of the closed and open flow, respectively, will become more apparent. As mentioned earlier, regions of poor stretching in the closed flow, due to re-entrainment, eventually give way to enhanced stretching, which is not necessarily the case in the open flow. However, the finiteness of the closed flow mixing region implies that diffusion zone overlap, and hence saturation, will become more of a factor than in the open flow (especially when one considers more than just nearest neighbor overlap). Also, in the original problem of fluid B in the core and fluid A outside the core, with successive time samples the open flow entraining turnstile lobe will always contain only fluid A, except for the amount diffused out, while the closed flow entraining turnstile lobe will contain less and less of fluid A due to more and more re-entrainment of fluid B via this lobe. In this context, open flow mixing involves more fluid and lasts longer.

With Figures 3.11 and 3.12 in hand, we emphasize the spirit of our investigation. Certainly the approach to studying diffusion is an approximate one, with several sources of error: we consider only nearest neighbor overlap, we ignore diffusion across the stable manifold boundary, and the assumption of parallel neighboring segments of the interface and a one-dimensional diffusion process break down at sharp turning points of the interface, such as the tip of the lobe. However, the errors regarding the stable manifold and sharp turning points involve a truly negligible amount of fluid, and nearest neighbor overlap is sufficient to approximately capture the effect of saturation. Our main goal is not to focus on a computational prescription for exact numerical results, but rather to take a first step past the non-overlapping thin diffusion zone theory, and to show how the efficiency of diffusion is affected by this extension. Until now our discussion of "efficiency" has been informal; though there is more than one candidate for a formal definition of efficiency of diffusion in this context, none seem adequate for a *global* definition. For example, it seems natural to

address the non-uniformity of stretching and interface separation, and hence to compare the amount of fluid diffused out of the lobe with an idealized case that has the same lobe area and lobe boundary length, but spatially uniform stretching and manifold separation. Dividing the non-uniform result by the uniform result could define an "efficiency" of diffusion, except that the result is not guaranteed to be less than one: though $\Delta A_o(t)/\Delta A_{no}(t)$ can be *decreased* by nonuniformity, where ΔA_o, ΔA_{no} denote the total amount of A diffused out of the lobe in the overlapping and nonoverlapping theory, respectively, $\Delta A_{no}(t)$ can be ***increased*** by nonuniformity, so that the above "efficiency" can be greater than one or less than one (indeed for Figure 3.12 that above "efficiency" is found to be greater than one for the open flow and less than one for the closed flow for $t \geq 3$). For a global consideration of diffusion, it is not clear that there is a "best case" on which to base a definition of efficiency. Rather than focusing on a formal definition, and exact quantification, of efficiency, however, we would rather highlight the qualitative, practical aspects of the problem. For example, it is clear that the efficiency of diffusion across an interface is affected by the non-uniformity of stretching in the perturbed case, which, as discussed before, is related to the unperturbed stretch profile, which in turn is affected by the geometry of the unperturbed phase portrait (for example, the presence of an additional hyperbolic fixed point in the open flow tended to make the stretch distribution more even), and by the unperturbed system parameters (such as the strength of the vorticity). Thus one can alter the efficiency of the diffusion process by changing the geometry of the unperturbed phase portrait or the system parameters. In addition, we see from Figures 3.6 and 3.11 that the local stretch and diffusion rates vary greatly with initial conditions, and by straightforward calculations one can identify the regions of good stretching and mixing; for example, in the closed flow simulation good stretching and mixing is isolated near the SIP's, so identifying these points locates in phase space the "seeds" of good stretching and mixing in chaotic tangles. Though exact results can only

come from explicit numerical simulation, we feel that a knowledge of the unperturbed stretch profile, coupled with an appreciation of the symbolic dynamics that applies to material interfaces, provides the framework for studying many aspects of stretching and mixing in chaotic tangles. Indeed Chapter 6 offers an example in the context of distributions of finite-time Lyapunov exponents.

References

[1] J.M. Ottino, J. Fluid Mech. **114**, 83 (1982).

[2] J.M. Ottino, *The kinematics of mixing: stretching, chaos, and transport* (Cambridge University Press, Cambridge, 1989).

[3] D. Beigie, A. Leonard, and S. Wiggins, Nonlinearity, **4**, 775 (1991).

[4] S. Wiggins, Physica D **44**, 471 (1990).

[5] V. Rom-Kedar and S. Wiggins, Arch. Rat. Mech. and Anal. **109**, 239 (1990).

[6] V. Rom-Kedar, A. Leonard, and S. Wiggins, J. Fluid Mech. **214**, 347 (1990).

[7] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Springer-Verlag, New York, Berlin, Heidelberg, 1990).

[8] G.K. Batchelor, J. Fluid Mech. **5**, 113 (1958).

[9] E. Ott and T.M. Antonsen, Jr., Phys. Rev. A **39**, 3660 (1989).

[10] H. Aref and S.W. Jones, Phys. Fluids A **1**, 470, (1989).

[11] G.F. Carrier, F.E. Fendell, and F.E. Marble, SIAM J. Appl. Math. **28**, 463 (1975).

[12] F.E. Marble, in *Recent Advances in the Aerospace Sciences*, edited by C. Casci (Plenum, 1985), pp. 395-413.

[13] A. Leonard, V. Rom-Kedar, and S. Wiggins, Nucl. Phys. B (Proc. Suppl.) **2**, 179 (1987).

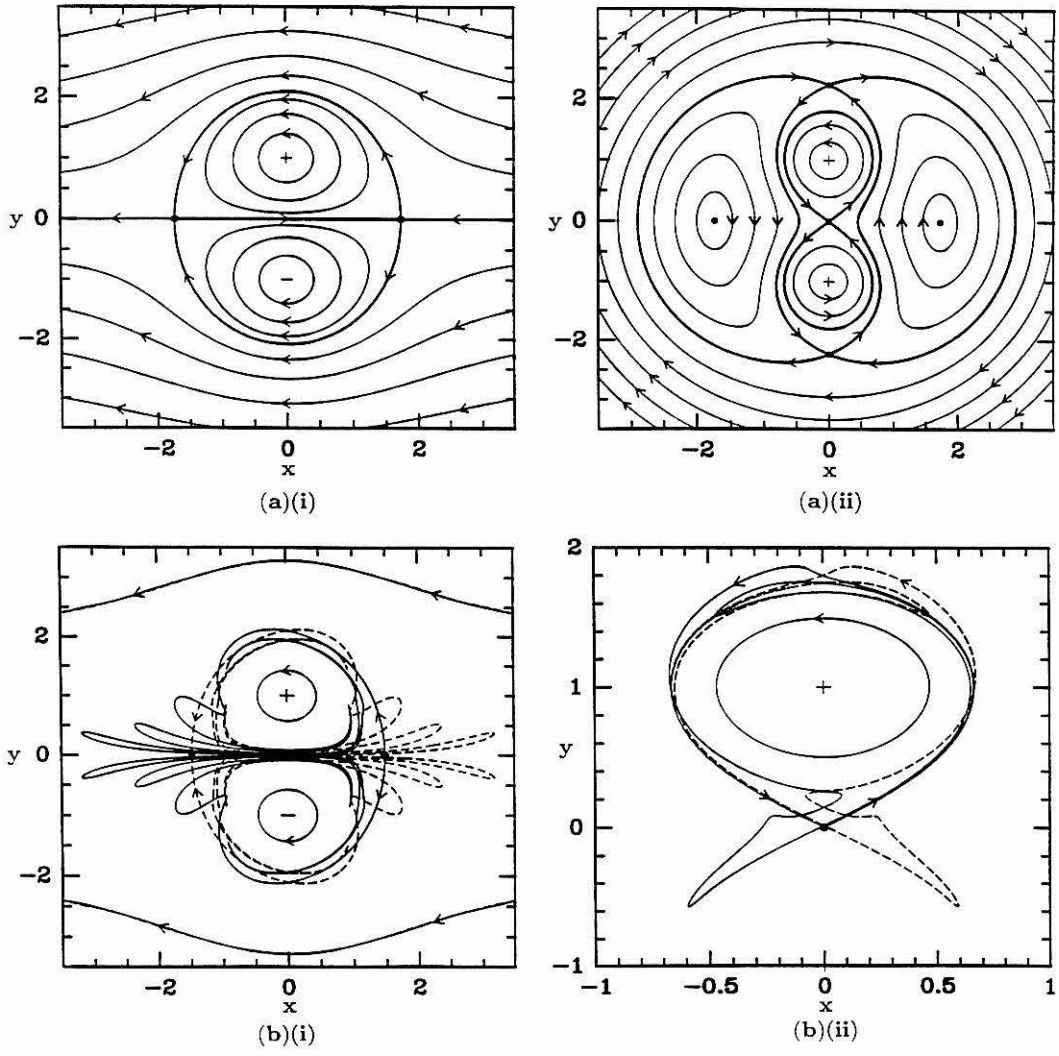


Figure 3.1 (a) The streamlines of the unperturbed (i) open and (ii) closed flows with the vortices at $(0, \pm 1)$. The fixed points are marked by dots and the separatrices are shown in boldface. (b) The (i) heteroclinic and (ii) upper homoclinic tangles of the perturbed flow. The persisting hyperbolic fixed points are marked by dots, and their stable and unstable manifolds are marked by dashed and solid lines, respectively; other invariant curves are marked by solid lines. The closed flow has several tangles, and we focus on the homoclinic tangle associated with the innermost separatrix. See Figure 3.6 for the perturbed flow parameters.

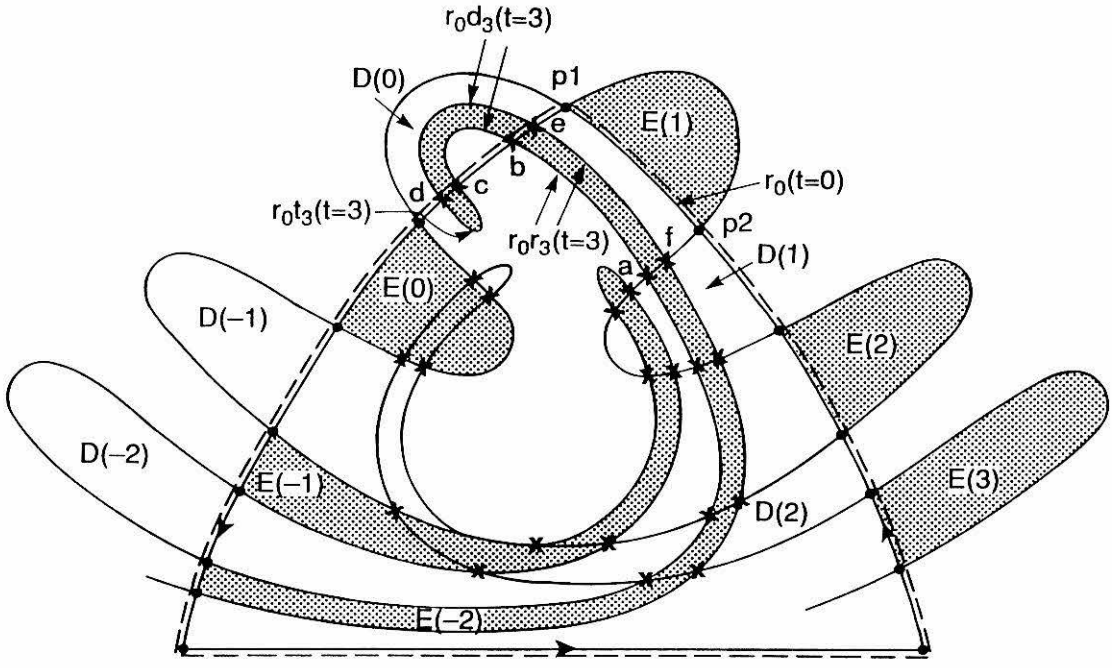


Figure 3.2 The upper ($y \geq 0$) invariant lobe structure of the perturbed open flow (the system is symmetric about $y = 0$).

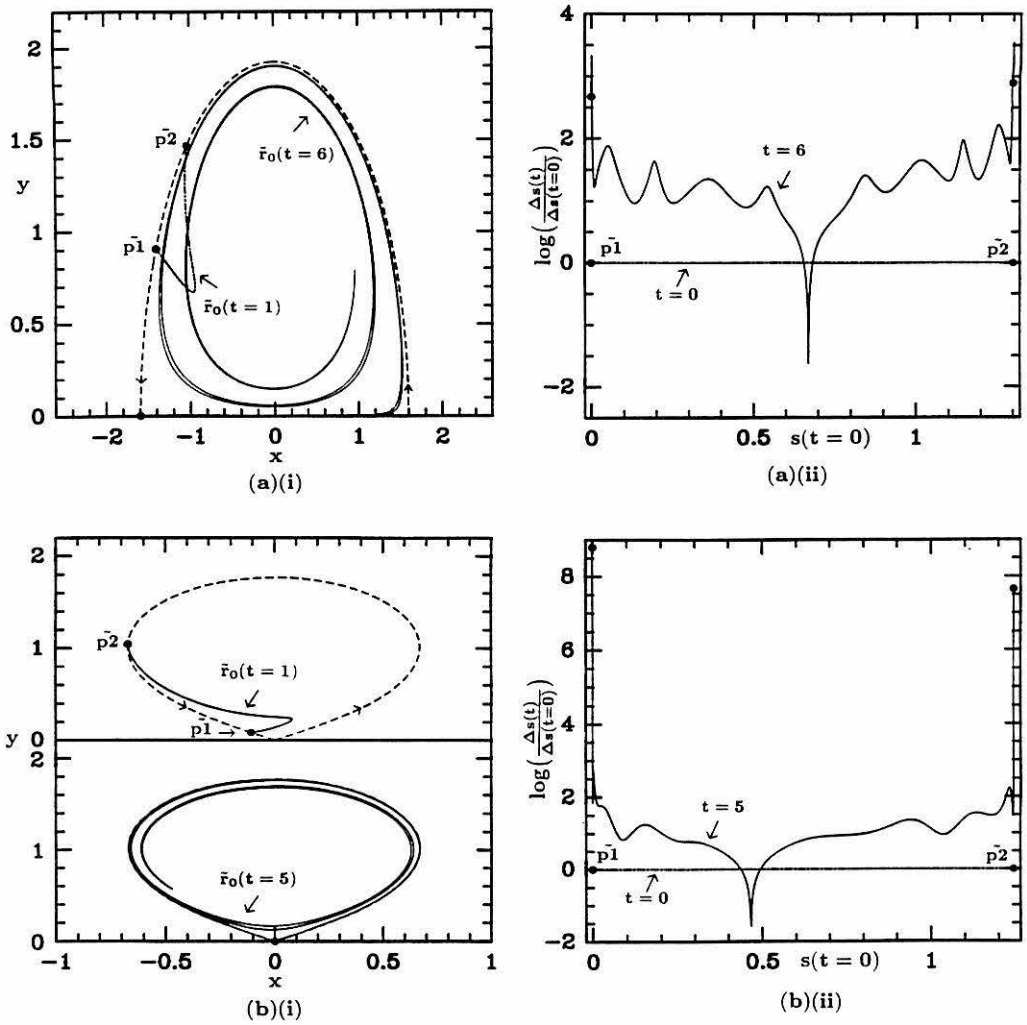


Figure 3.3 Evolution, according to the unperturbed (a) open and (b) closed flows, of the material curve $\tilde{r}_0(t)$. The material curves are shown in (i) (the dashed line represents the unperturbed separatrix), and the associated stretch profiles are shown in (ii) (s denotes arclength along the material curve and *log* denotes the common logarithm). The magnitude of the circulation associated with each vortex is $\Gamma = 0.4(2\pi)^2$. For the stretch profiles, note that $\tilde{r}_0(t=0) = P_\varepsilon^{-1}(\tilde{r}_0(t=1))$, i.e. we evolve $\tilde{r}_0(t)$ from $t=0$ to $t=1$ according to the *perturbed* flow (see Figure 3.6 for the parameters) and from then on according to the *unperturbed* flow. For the perturbed flows, the vortices are at $(0, \pm 1)$ at $t = 0$ (following the convention of Rom-Kedar et al.[9]); for the unperturbed flows the vortices are set at the time average of the perturbed vortex coordinates, i.e. at $(0, \pm 0.92183)$ for the open flow and $(0, \pm 0.98039)$ for the closed flow.

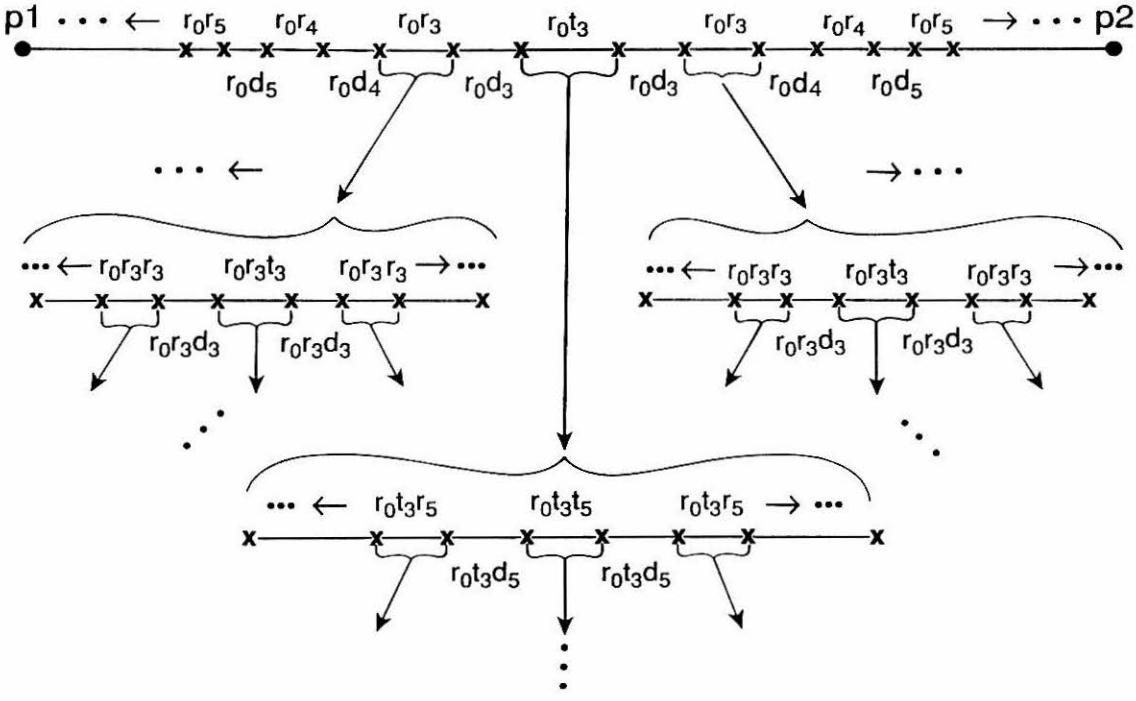


Figure 3.4 The symbolic dynamics that applies to r_0 in the open flow. The upper line corresponds to r_0, and the lower lines to enlargements of pieces of r_0.

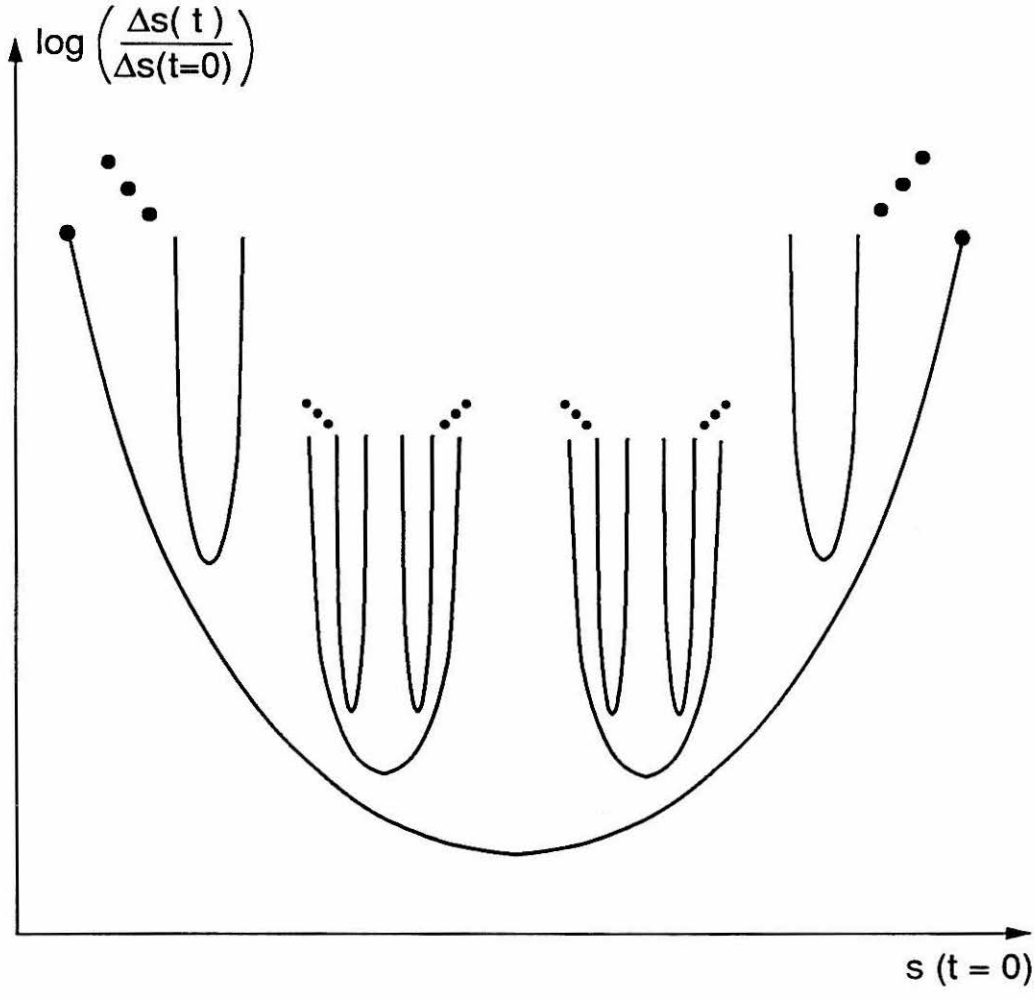


Figure 3.5 The unperturbed stretch profile approximately repeating itself on smaller and smaller scales. This diagram gives only an extremely cartoon-like portrayal of the phenomenon: the unperturbed profile is represented simply by a well-shaped curve (spatial oscillations and sharp dips are ignored) and we are indicating a sequence of events by a single diagram.

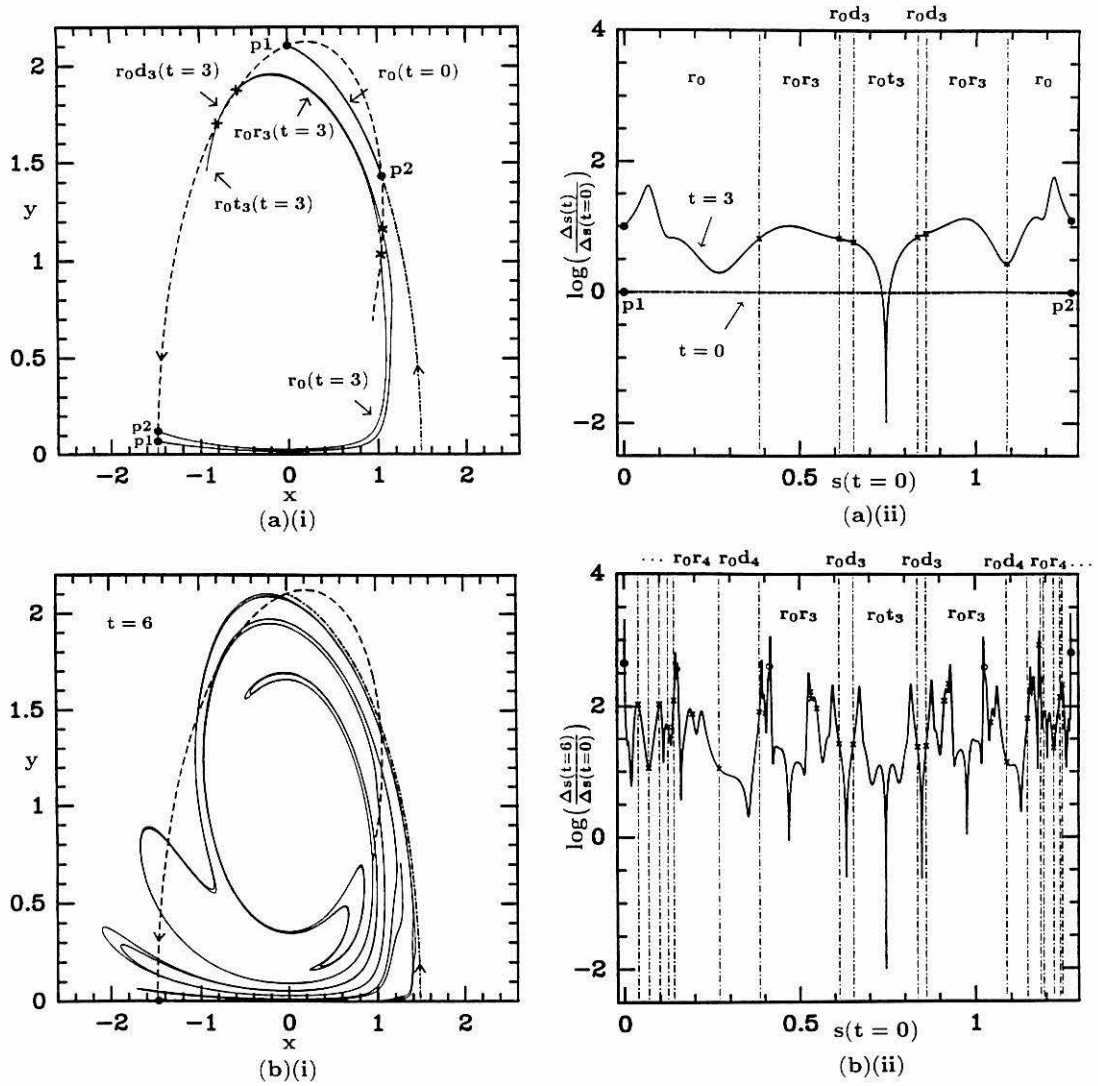


Figure 3.6 Plots of (i) r_0 and (ii) associated stretch profiles for (a) the open flow at $t = 0, 3$, (b) the open flow at $t = 6$, (c) the closed flow at $t = 0, 2$, and (d) the closed flow at $t = 5$, with symbolic dynamics labeling. In (i) the stable manifolds are shown by dashed lines, and the unstable manifolds by dashed-dotted lines; in (ii) the dashed lines indicate SIP's. In plots (b) and (d) we do not explicitly label all the subdivisions of r_0, since the labeling would become much too crowded on the scale shown (plot (b) does indicate additional SIP's by crosses, and points which are "almost SIP's" (*i.e.*, almost intersect the stable manifold segment of the core) by open dots). (*Caption continued on next page.*)

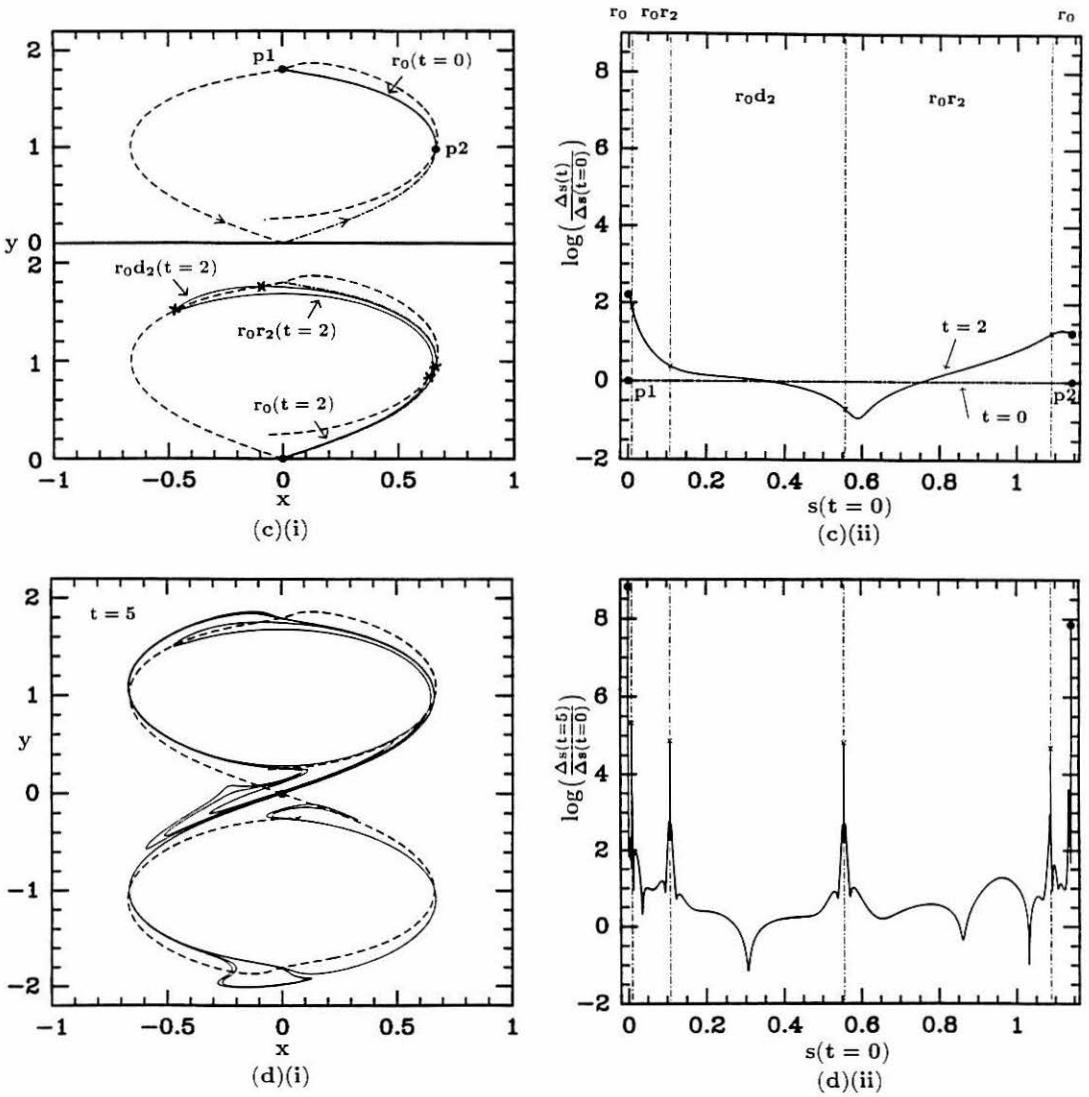


Figure 3.6 (*Caption continued from previous page.*) The flows are specified by the parameters $\Gamma = 0.4(2\pi)^2$ and (a,b) $\varepsilon = 0.085 \cdot 2\pi$, (c,d) $\varepsilon = 0.02 \cdot 2\pi$, with the vortices at $(0, \pm 1)$ at $t = n \in Z$. For these and all other simulations, interfaces are represented by a weighted distribution of passive marker particles. The weighting is determined by applying dynamic point insertion to an initially uniform distribution (starting the simulation over with the weighted distribution then eliminates any point insertion error). Particle trajectories are found by integrating the governing ODE system with a fourth order Runge-Kutta scheme.

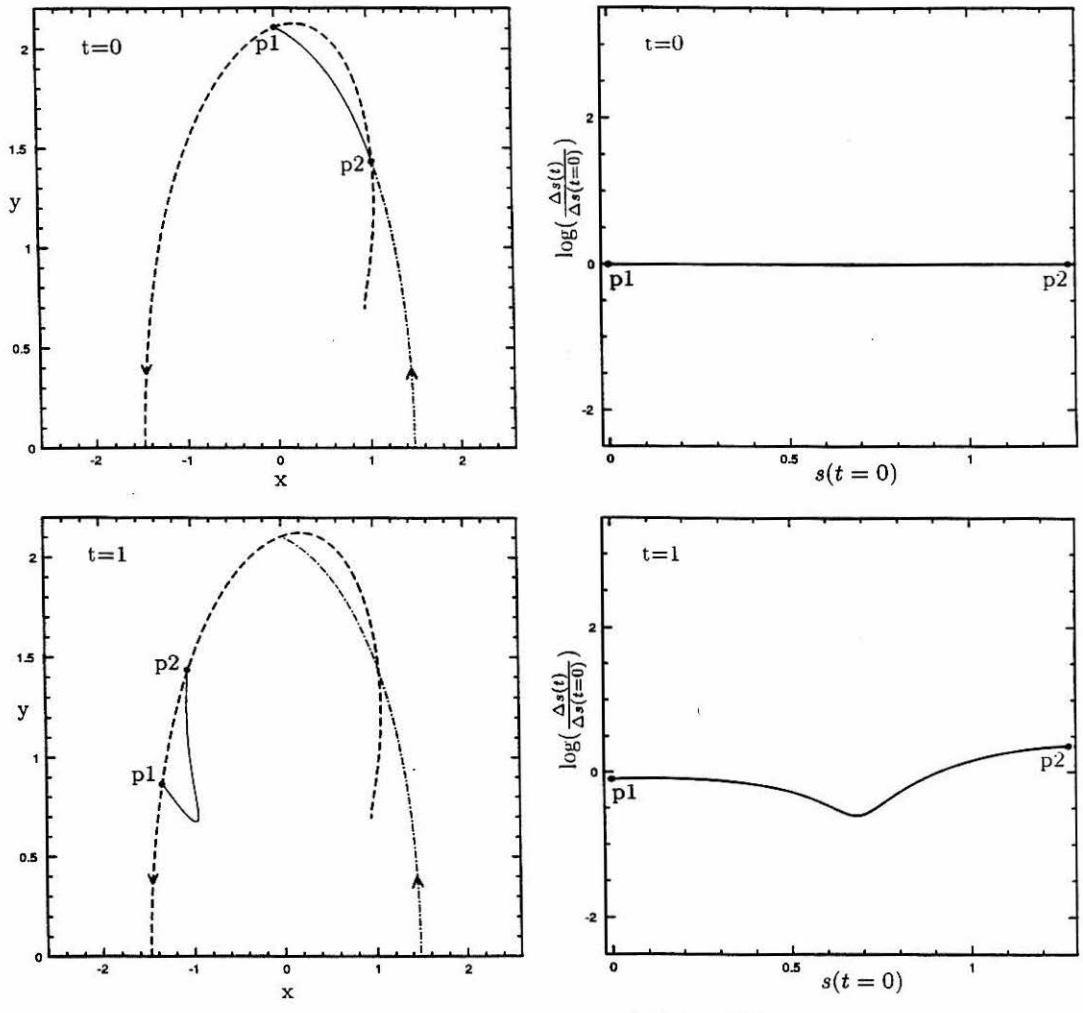


Figure 3.7 Plots of r_0 and the associated stretch profile for *each* time sample of the perturbed open flow simulation, without detailed labeling.

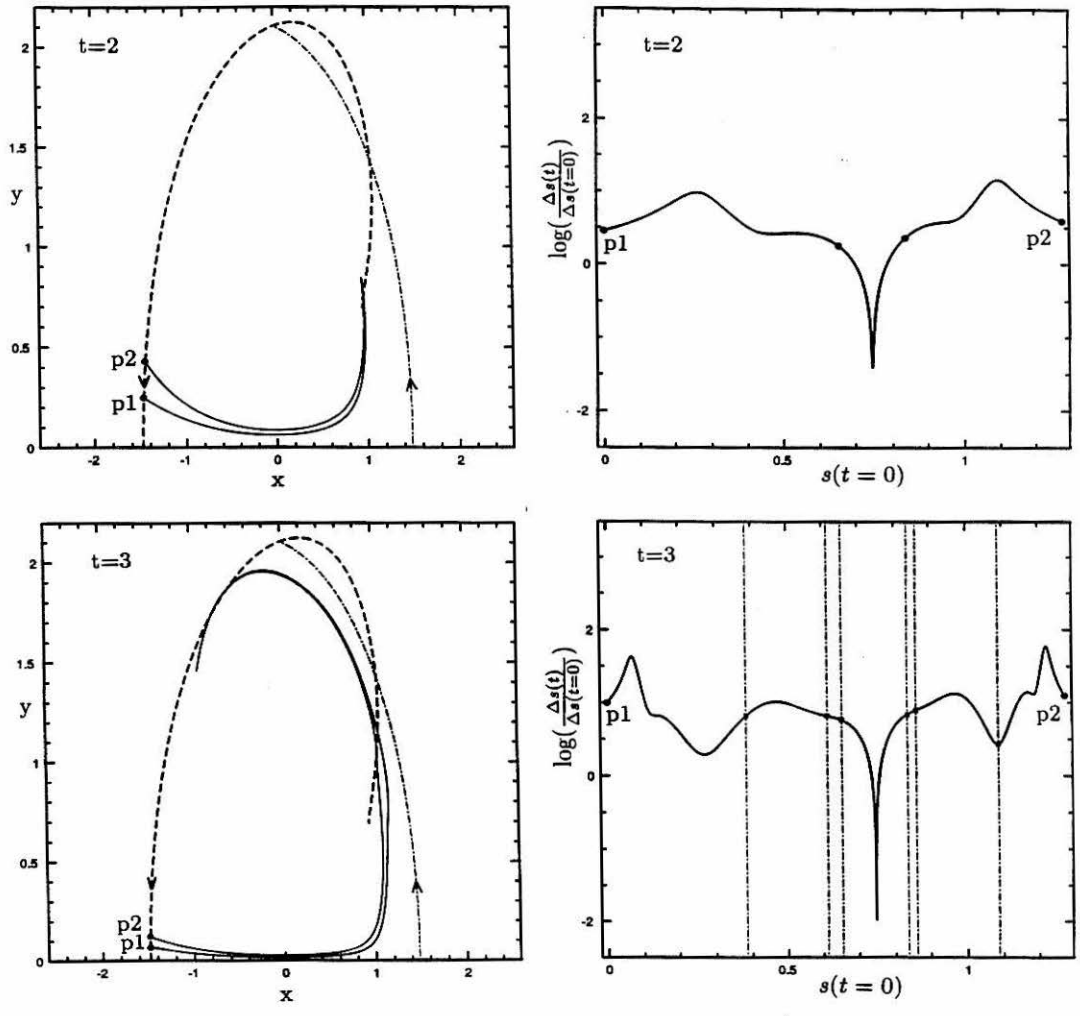


Figure 3.7 *Continued.*

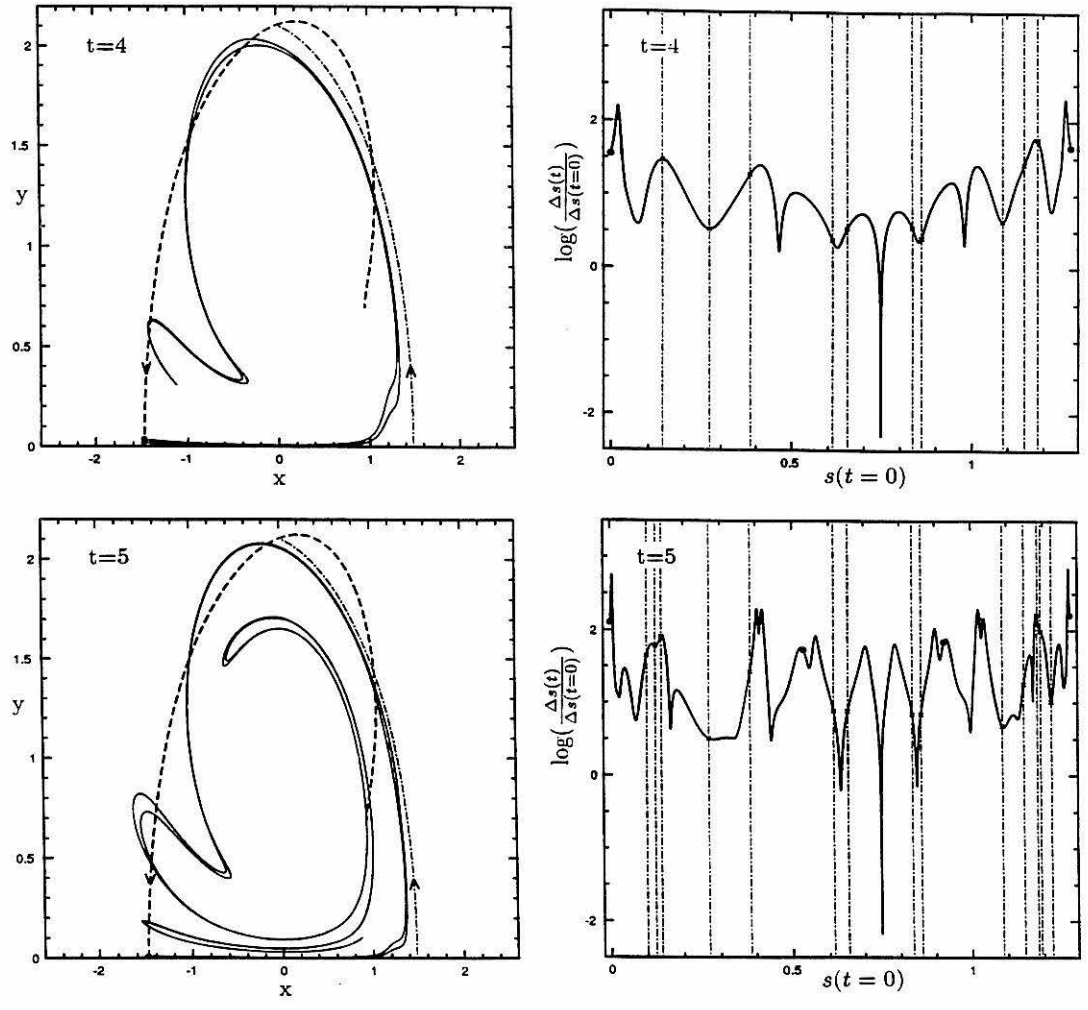


Figure 3.7 *Continued.*

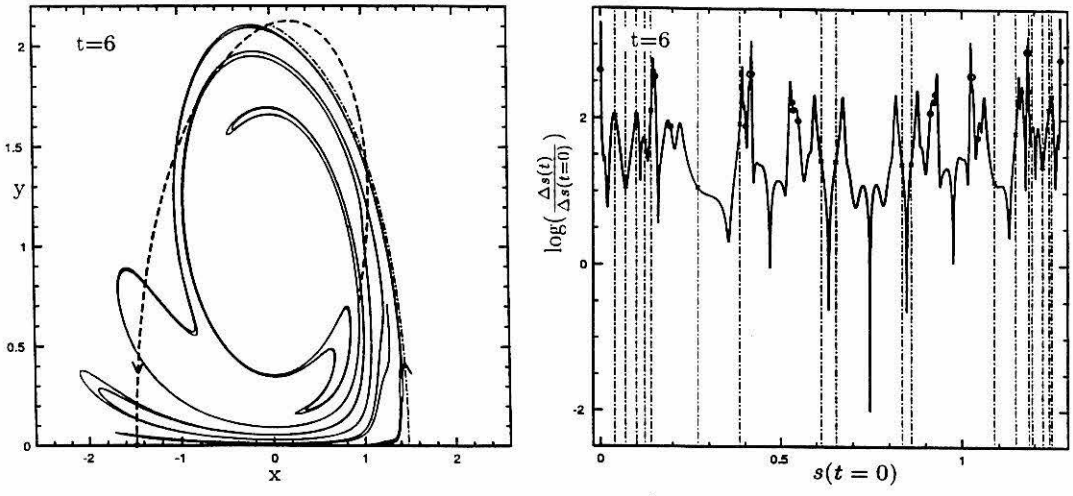


Figure 3.7 *Continued.*

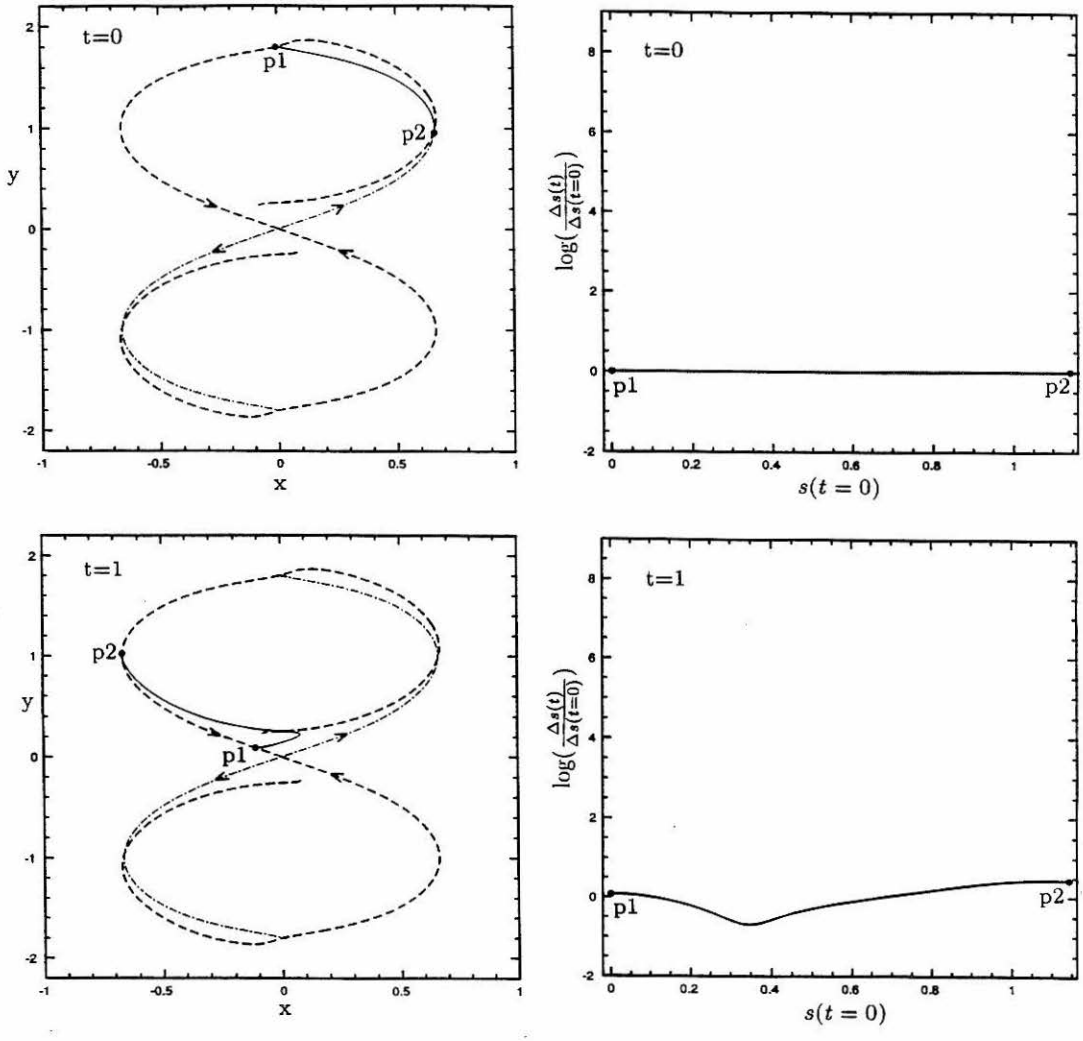


Figure 3.8 Plots of r_0 and the associated stretch profile for *each* time sample of the perturbed closed flow simulation, without detailed labeling.

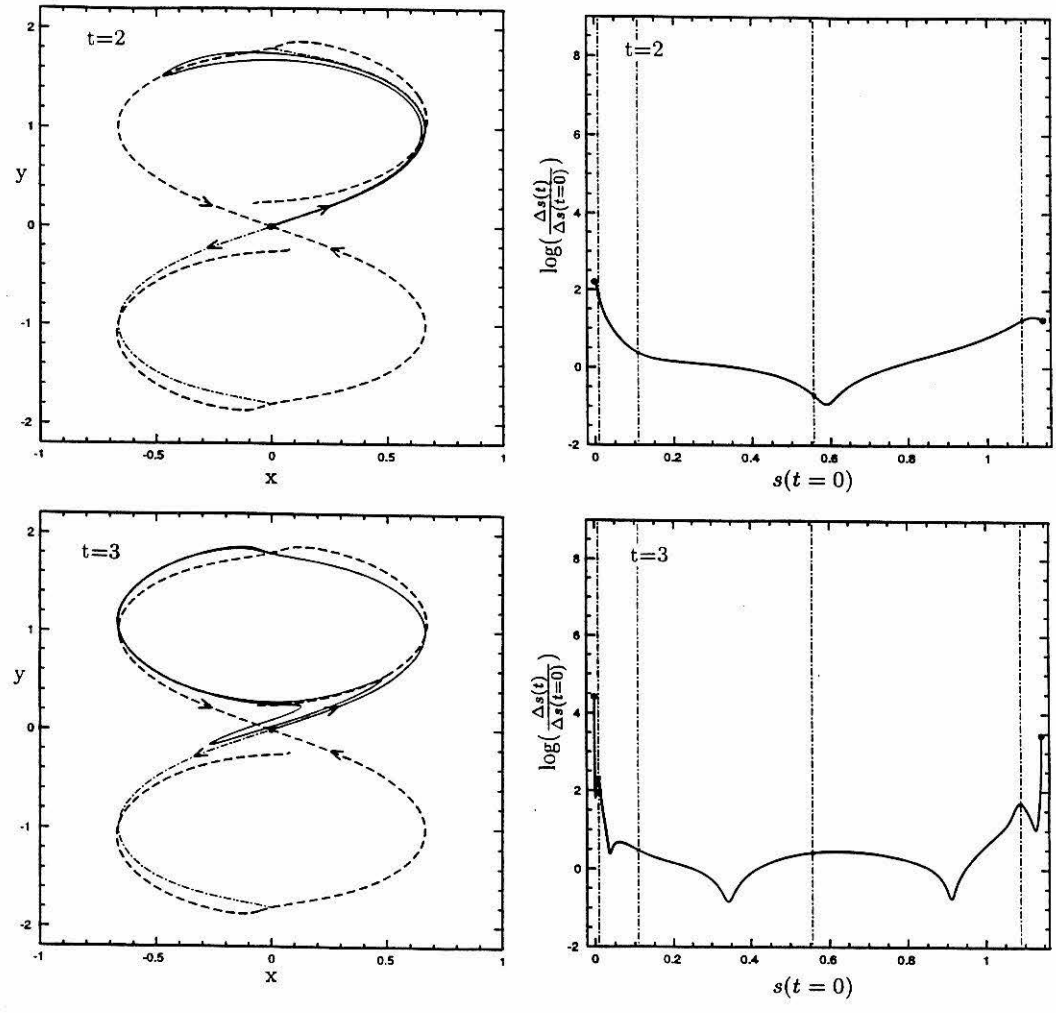


Figure 3.8 *Continued.*

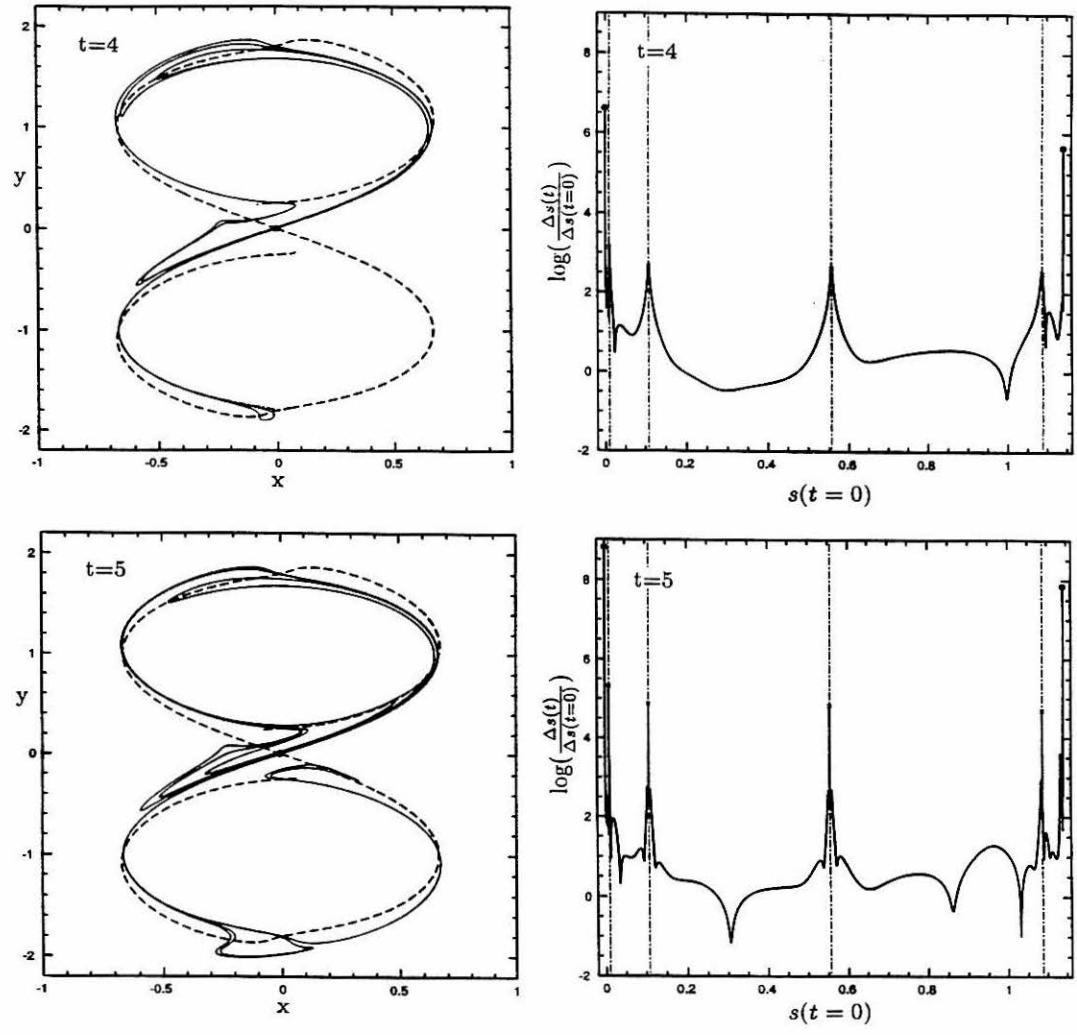


Figure 3.8 *Continued.*

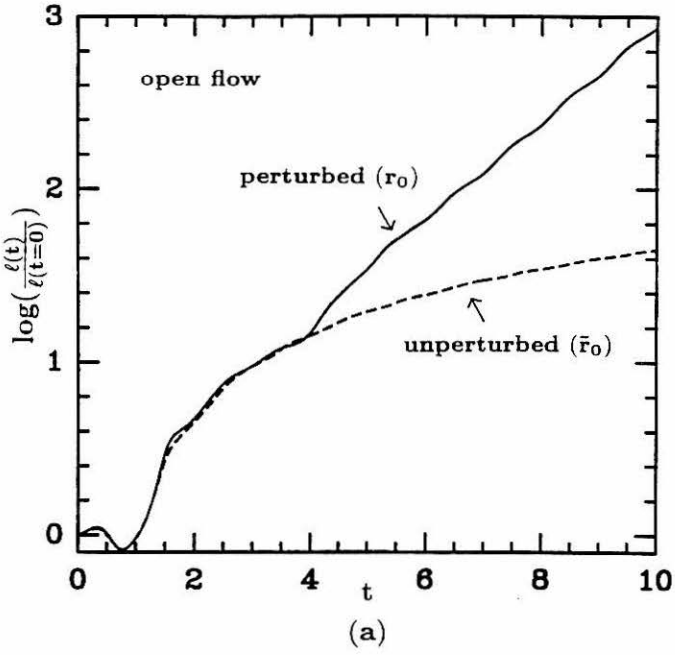
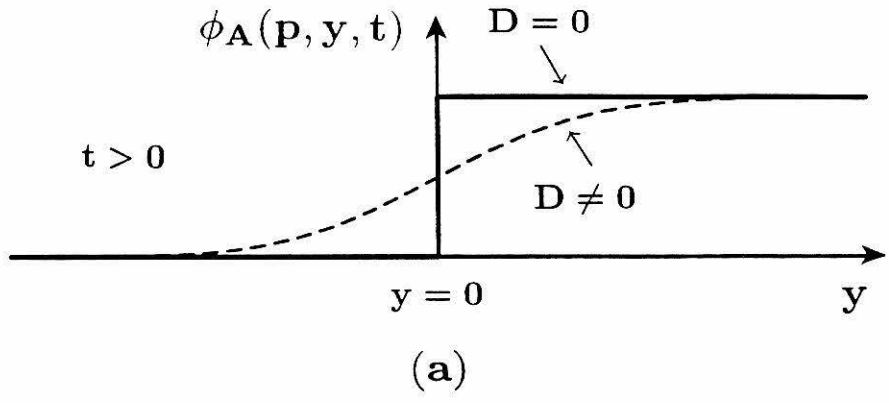
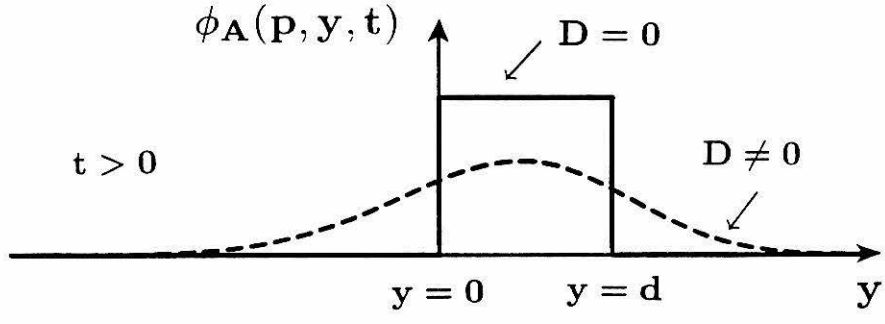


Figure 3.9 Total length $\ell(t)$ of r_0 in the perturbed case and of $\tilde{r}_0$ in the unperturbed case for (a) the open flow and (b) the closed flow.



(a)



(b)

Figure 3.10 The concentration profile of fluid A near the interface at **p** for (a) the non-overlapping thin diffusion zone approximation and (b) the overlapping thin diffusion zone approximation.

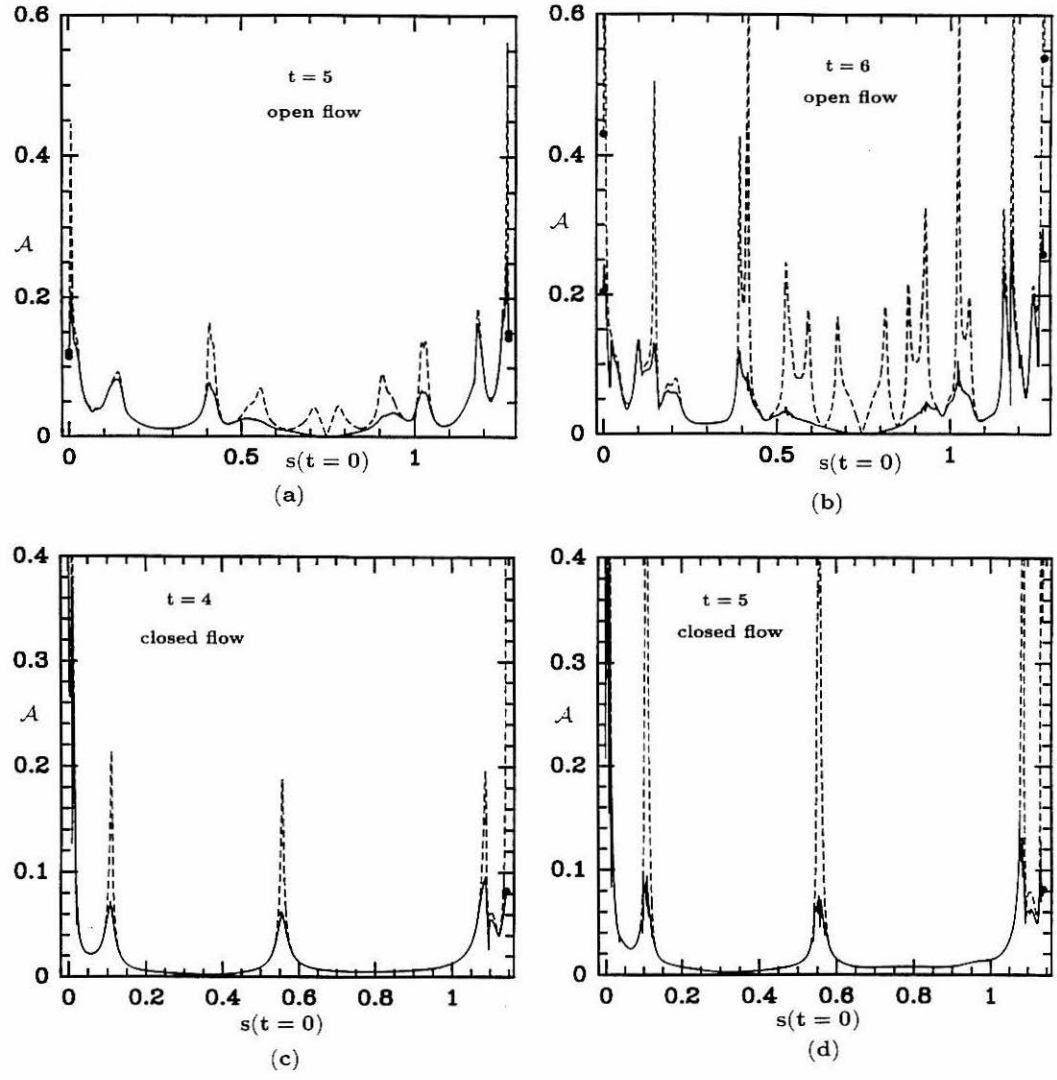


Figure 3.11 $\mathcal{A}(\mathbf{p}, t)$ when r_0 advects under the same velocity fields as considered in the examples of Figures 3.6-3.8 ($D = 2\pi \cdot 10^{-6}$ for both flows). The dashed line corresponds to the non-overlapping thin diffusion zone theory and the solid line to the overlapping thin diffusion zone theory.

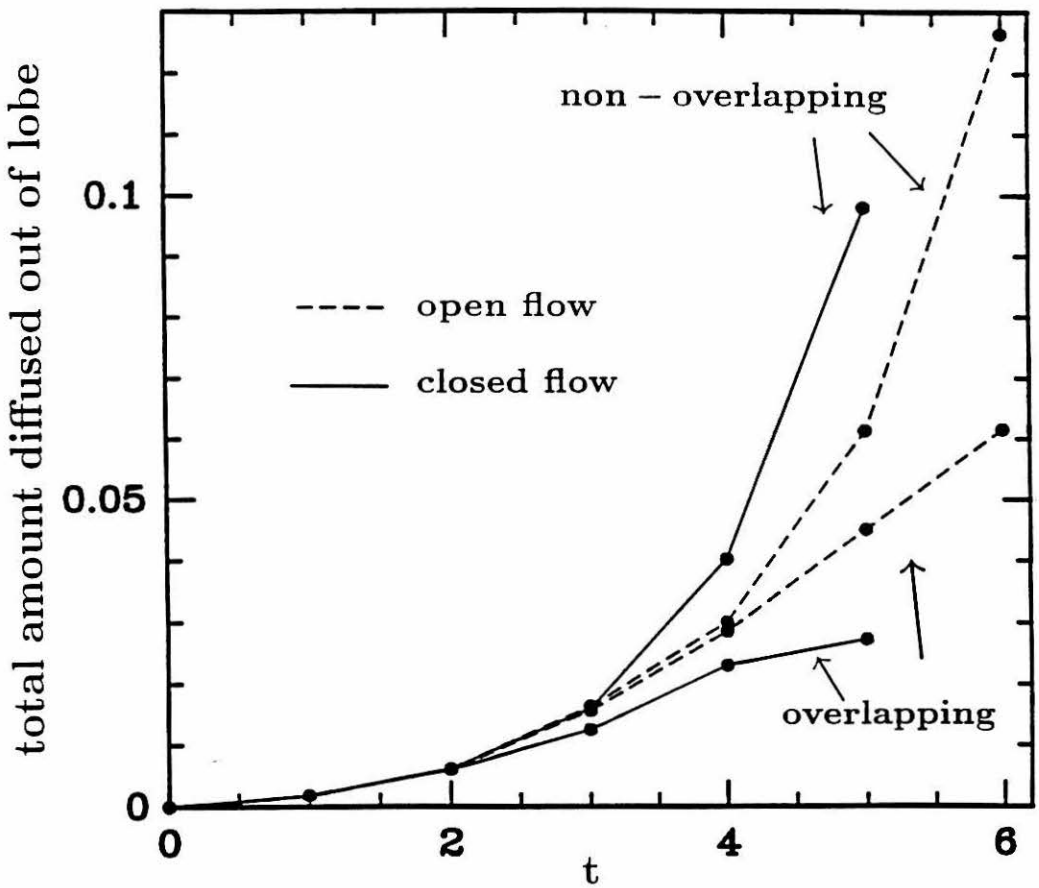


Figure 3.12 The total amount of fluid A that diffuses out of the lobe in the non-overlapping and overlapping theory. The entraining turnstile lobe area is 0.1388 and 0.0931 for the open and closed flow, respectively.

Chapter 4
Strain rates under near-integrable chaotic flows

Abstract

We study rates of strain experienced by infinitesimal line elements evolving under near-integrable chaotic flows, revisiting for sake of illustration the two oscillating vortical flows considered in Chapter 3. We introduce the notion of *irreversible rate of strain* responsible for net stretch, study the role of hyperbolic fixed points as engines for good irreversible straining, and observe the role of *turnstiles* as mechanisms for enhancing straining efficiency via *re-orientation* of line elements and *transport* of line elements to regions of superior straining. Continous-time irreversible strain rate plots associated with infinitesimal line elements exhibit for *integrable* flows isolated peaks with each pass by a hyperbolic fixed, whose amplitudes decay in time as a consequence of progressive orientation of line elements along the streamlines. For *near-integrable* flows the irreversible strain rate plots show isolated peaks which *do not* decay in time as a consequence of re-orientation of line elements by the turnstile lobes. Turnstiles have been shown to play a fundamental role in *phase space transport*, and we argue here they play a fundamental role in *phase space stretching.*

4.1 Introduction

Whereas studies of turbulent fluids sometimes *assume* a constant rate of strain (*e.g.*, see Batchelor[1] and more recently Dimotakis[2]), chaotic flows have been claimed to have rates of strain which "oscillate wildly" (*e.g.*, Dresselhaus and Tabor[3]). Given this backdrop, we study strain rates experienced by infinitesimal line elements evolving under 2D near-integrable chaotic flows. Our principal interest is in the notion of *irreversible rate of strain*, the role of hyperbolic fixed points as engines for good irreversible straining, and the role of *turnstiles* (see Chapters 2, 3, and 5) as mechanisms for increasing straining efficiency via *re-orientation* of line elements and *transport* of line elements to regions of superior straining. We illustrate our study using the same two oscillating vortical flows considered in Chapter 3. In Section 4.2 we consider the integrable (*i.e.*, unperturbed) flows, and study numerically the rate of strain associated with sets of infinitesimal line elements distributed along the material interfaces $\tilde{r}_0(t = 1)$ of Chapter 3 (refer back to Figure 3.3). Continuous-time strain rate plots contain a large-amplitude oscillatory component associated with the component of the line element along the streamlines, which has no net contribution to line element stretch or contraction. We thus subtract out this contribution to define the *irreversible rate of strain* directly responsible for net stretch. Continuous-time plots of this irreversible rate of strain make apparent the contributions to net stretching associated with the fact that the line element has a component normal to the streamlines. Distinct peaks are observed with each passage by a hyperbolic fixed point, and the peak amplitudes decay in time due to increasing line element orientation along the streamlines as time progresses. In Section 4.3 we address the perturbed flows, and discuss the enhancement of straining efficiency via re-orientation of line elements by the turnstile lobes. The role of turnstiles in the study of *transport* in phase space has had fundamental and far-reaching consequences (see Chapter 2 and espe-

cially Chapter 5); we offer here a complementary observation about turnstiles relevant to *stretching* in phase space. We study numerically the conventional rate of strain in the non-integrable flows, observe the enhancement of straining, and note the particular relevance of SIP's (secondary intersection points) discussed in Chapter 3. As in the integrable case, there is generically a large-amplitude oscillatory straining component which again complicates the study. For non-integrable flows, lacking streamlines, there is no *exact* expression afforded that one can subtract out, but one can subtract out an *approximate* contribution which can give reasonable results for near-integrable flows. In contrast to Chapter 3, the present analysis of stretching in the context of strain rates is more directly related to conventional studies of stretching under fluid flows, and, as we shall see, the results here compliment the stretch analysis of Chapter 3 (and later Chapter 6), since there we are concerned with *stretch* experienced by *ensembles* of line elements (representing an interface), and here we study stretch *rates* experienced by *individual* infinitesimal line elements.

4.2 Integrable flows

4.2.1 Conventional rate of strain

We consider the two oscillating vortical flows identical to those of Chapter 3 (same parameters, same interface). Again we begin with the integrable (unperturbed) flow, which provides the setting with which to understand the near-integrable flows. Figure 4.1 shows the initial location of the infinitesimal line elements whose straining we study, distributed along $\tilde{r}_0(t = 1)$ from Figure 3.3, with initial orientation defined by the local tangent to the material curve on which they lie. The rate of strain experienced by an infinitesimal line element $(\delta\ell_1, \delta\ell_2)$ located at (x_1, x_2) at time t in the presence of a two-dimensional

velocity field

$$\begin{aligned} u_1(x_1, x_2, t; \varepsilon) &= f_1(x_1, x_2) + \varepsilon g_1(x_1, x_2, \omega t + \theta_0; \varepsilon) \\ u_2(x_1, x_2, t; \varepsilon) &= f_2(x_1, x_2) + \varepsilon g_2(x_1, x_2, \omega t + \theta_0; \varepsilon), \end{aligned} \tag{4.2.1}$$

where g_i, $i = 1, 2$, is 2π-periodic in $\omega t + \theta_0$, is given by

$$\sigma \equiv \frac{1}{\delta\ell}\frac{d(\delta\ell)}{dt} = m_i m_j \frac{\partial u_i}{\partial x_j}(x_1, x_2, t; \varepsilon), \tag{4.2.2}$$

where

$$\delta\ell \equiv \sqrt{(\delta\ell_1)^2 + (\delta\ell_2)^2}$$

is the line element length,

$$m_i \equiv \frac{\delta\ell_i}{\delta\ell}, \qquad i = 1, 2,$$

defines the components of the unit vector oriented along the line element, and the Einstein summation convention is used (with $x_1 \equiv x$, $x_2 \equiv y$). The spatial location of the line element is evolved by

$$\dot{x}_i = u_i(x_1, x_2, t; \varepsilon), \quad i = 1, 2, \tag{4.2.3}$$

and its orientation by

$$\dot{m}_i = m_k \frac{\partial u_i}{\partial x_k} - m_k m_\ell \frac{\partial u_k}{\partial x_\ell} m_i, \quad i = 1, 2. \tag{4.2.4}$$

Evolving the line elements of Figure 4.1 according to equations (4.2.3) and (4.2.4), the continuous-time evolution of the rate of strain experienced by each line element via equation (4.2.2) is shown in Figures 4.2 and 4.3 for the open and closed flow, respectively. We make the following observations.

(i) There is a large-amplitude oscillatory component associated with the translation of the line elements around the closed streamlines. If the line elements were oriented along the streamlines, these oscillations would have zero mean and hence no net contribution to stretch or contraction. *Note*

that there is significant instantaneous straining away from the hyperbolic fixed points. Indeed, for the open flow, the most significant straining occurs away from the region local to the hyperbolic fixed point.

(ii) The oscillations are asymmetric about $\sigma = 0$; the area above $\sigma = 0$ bounded by the positive strain rates is greater than the area bounded below $\sigma = 0$ by the negative strain rate, entailing net positive stretch. The imbalance is manifested by larger positive strain amplitudes and longer time intervals over which the positive strain is defined. These asymmetries are due to the initial orientation of the line elements having a component normal to the streamlines.

(iii) The asymmetries decay in time due to the increasing orientation of line elements along the streamlines as time progresses (refer back to Figure 3.3, and see Figure 4.4), entailing that the net positive stretch is poor, only asymptotically linear in time (refer back to Chapter 3). The decay is fairly rapid with successive revolutions, entailing that the obvious asymmetries of the first revolution quickly become difficult to decipher with the backdrop of large-amplitude oscillations.

(iv) The further away the line element is from the endpoints $\tilde{p}1$, $\tilde{p}2$, (*i.e.*, the deeper the line element is into the integrable core), the higher the frequency is of the straining oscillations (since the line elements translate faster around the core).

(v) Besides the difficulty associated with the large-amplitude oscillations that have no net contribution to stretch, note the limitations of the conventional rate of strain due to the lack of a normalization - for an individual orbit one cannot identify the overall efficiency of the straining process.

The difficulties in (i), (iii), and (v) motivate an improved definition of rate of strain.

4.2.2 Irreversible rate of strain

We address the difficulties in (i), (iii), and (v) of the previous section by defining a new rate of strain. If a line element were oriented along a streamline, then due to the invariance of the streamline, the line element would be destined to remain oriented along the streamline, oscillating around the closed curve *ad infinitum* with no *net* stretching and contraction. The strain rate could of course oscillate from positive to negative values, but with zero mean. Hence, though the strain rate at any given moment might have what may seem an impressive contribution, the time integral of this contribution would offer no net stretch. For any line element, we thus subtract out this *reversible rate of strain* σ_{rev} to define the *irreversible rate of strain* σ_{irr} responsible for the net stretch

$$\sigma_{irr} \equiv \sigma - \sigma_{rev}, \tag{4.2.5}$$

where

$$\sigma_{rev} \equiv \tilde{m}_i \tilde{m}_j \frac{\partial f_i}{\partial x_j}, \tag{4.2.6}$$

and

$$\begin{aligned} \tilde{m}_1 &= \frac{f_1}{\sqrt{f_1^2 + f_2^2}} \\ \tilde{m}_2 &= \frac{f_2}{\sqrt{f_1^2 + f_2^2}} \end{aligned} \tag{4.2.7}$$

defines the orientation of an infinitesimal line element tangent to the streamline. This will allow us to ignore the large-amplitude oscillations of (i) and focus on the subtle asymmetries of (iii).

In addition to addressing (i) and (iii), one might like to address (v) by normalizing the strain rate. One choice of normalization that pervades the literature by Ottino and collaborators (see for example Ref. 4 and references therein) is to define a *stretch efficiency* by dividing the conventional rate of strain by $\sqrt{D : D}$, where D is the symmetric 2-tensor associated with spatial

derivatives of the velocity field

$$D_{ij} \equiv \frac{1}{2}\Big(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\Big). \tag{4.2.8}$$

Such a definition of stretch efficiency suffers from the property that it does not equal one in such instances where it would seem to be natural. This point is discussed in the Appendix, and is a consequence of the fact that the normalization seems motivated by the Cauchy-Schwarz inequality, and though one can obtain an equality in this relation for vectors, one cannot in general obtain such an equality for matrices. A simple example is offered in Figure 4.5, which shows an infinitesimal line element approaching a hyperbolic fixed point parallel to the local unstable manifold and normal to the local stable manifold. The strain rate experienced by the line element is the unstable eigenvalue λ_u, and this corresponds to maximum rate of strain. We would thus like to have a stretch efficiency equal to one here, and yet

$$D : D = \lambda_u^2 + \lambda_s^2,$$

so that the stretch efficiency employed by Ottino and collaborators is

$$\frac{\lambda_u}{\sqrt{\lambda_u^2 + \lambda_s^2}} < 1,$$

due to the inclusion of the *contraction* term in the normalization. Instead we choose a normalization for positive (negative) strain rate with division by the absolute value of the maximum (minimum) rate of strain a line element can experience at its given spatial location (in practice one can rotate a hypothetical line element about its spatial location to determine the orientation corresponding to maximal (minimal) strain rate). In the above example, then, one would divide by λ_u to give an efficiency of one. More formally, we define a *straining efficiency* by

$$\sigma^e \equiv \frac{\sigma}{\sigma_m} \tag{4.2.9}$$

and the *irreversible straining efficiency* by

$$\sigma^e_{irr} \equiv \frac{\sigma}{\sigma_m} - \frac{\sigma_{rev}}{\sigma_m}, \tag{4.2.10}$$

where σ_m denotes the maximum or minimum strain rate as discussed above. By definition $\sigma^e \in [-1, 1]$ and $\sigma^e_{irr} \in [-2, 2]$. The value $\sigma^e = -1$ corresponds to maximally efficient contraction, $\sigma^e = 1$ to maximally efficient stretching. The value $\sigma^e_{irr} = 0$ corresponds to rate of strain equal to the reversible rate of strain; $\sigma^e_{irr} = 2$ corresponds to maximally efficient stretching for the strain rate and maximally efficient contraction for the reversible strain rate (and conversely for $\sigma^e_{irr} = -2$).

Though a study of *efficiencies* has advantages, we should stress its limitations. Straining efficiency indicates the *percentage* of the maximum available strain rate being exploited, while ignoring what this maximum strain rate is. Hence continuous-time straining efficiency plots for an individual trajectory offer no accounting of the relative importance of a given efficiency at two different times. For example, a very efficient straining when the maximal strain rate is close to zero will appear significant in efficiency plots, but it may be negligible compared with a comparatively inefficient straining when the maximum available strain rate is quite large. A study of both strain rates and strain efficiencies will thus be useful, and we will consider the two interchangeably.

Figures 4.6 and 4.7 show continuous-time plots of the irreversible straining efficiency associated with Figures 4.2 and 4.3, respectively. Subtracting out the mean-zero oscillations and normalizing makes quite apparent the nature of the straining. We observe the following.

(i) The irreversible strain rate is always positive in these plots, and there are isolated peaks associated with passage by a hyperbolic fixed point (there are thus two peaks per core revolution in the heteroclinic tangle and one peak per revolution in the homoclinic tangle). This effect is explained momentarily in the context of the linear flow local to the hyperbolic fixed

point.

(ii) Peak amplitudes decay in time due to the increasing orientation of line elements along the streamlines as time progresses.

(iii) The further away the line element is from the endpoints $\tilde{p}1$, $\tilde{p}2$, (*i.e.*, the deeper the line element is into the integrable core), the higher is the frequency of peak occurrence (since the line elements revolve faster around the core), and yet *the smaller the peaks* (since the line elements are further away from the hyperbolic fixed points).

The reversible straining can be interpreted in the context of streamline geometry, and the irreversible straining in terms of the linear flow local to hyperbolic fixed points. First, the reversible strain rate large-amplitude oscillations are easily interpreted from Figure 4.8 in light of the incompressible nature of the flow. When neighboring streamlines squeeze together (expand apart), area elements are squeezed (expanded) normal to the adjacent streamlines, and hence stretched (contracted) parallel to the streamlines. The temporal reversible straining oscillations are thus understood from the spatial oscillations in neighboring streamline separation. We reiterate that substantial reversible strainings occur outside the region local to the hyperbolic fixed point. Second, the irreversible strain rate peaks associated with line elements having a component normal to the streamlines can be understood with the help of Figure 4.9. Consider the linear flow local to a hyperbolic fixed point in the form

$$\begin{aligned} \dot{x} &= \lambda_u x \\ \dot{y} &= \lambda_s y, \end{aligned} \tag{4.2.11}$$

where $\lambda_u > 0$ and $\lambda_s < 0$ are the unstable and stable eigenvalues, respectively, of a hyperbolic fixed point located at the origin. Let α_d denote the angle between the line element and the tangent to the streamline, referred to as the ***angle of deviation*** (see Figure 4.9(a)). We write α_d as the difference between the angles the line element and streamline tangent make with the vertical axis, α_ℓ and α_t,

respectively:

$$\alpha_d = \alpha_\ell - \alpha_t, \tag{4.2.12a}$$

where

$$\begin{aligned} \alpha_\ell &= \tan^{-1}\left(\frac{\Delta x_\ell}{\Delta y_\ell}\right) \\ \alpha_t &= \tan^{-1}\left(\frac{\Delta x_t}{\Delta y_t}\right) \end{aligned} \tag{4.2.12b}$$

and $\alpha_\ell > \alpha_t$ (see Figure 4.9(b)). For the moment, assume α_ℓ and α_t (hence α_d) are small as they enter the region local to the hyperbolic fixed point. Equation (4.2.11) implies that as the line element passes by the hyperbolic fixed point Δx_ℓ, Δx_t increase exponentially in time, and Δy_ℓ, Δy_t decrease exponentially in time; hence, using that $\tan^{-1}(\Delta x/\Delta y) \approx \Delta x/\Delta y$ for small $\Delta x/\Delta y$ one easily obtains that $\alpha_d = \alpha_\ell - \alpha_t$ grows exponentially in time from some small initial value. This growth in α_d is saturated and then reversed, due to two effects (as should be made clear in Figure 4.9(c)):

(i) The growth in each of α_ℓ and α_t is saturated at $\pi/2$ by the higher order terms in $\tan^{-1}(\Delta x/\Delta y)$.

(ii) The angle α_ℓ "saturates" at (comes close to) $\pi/2$ before α_t does, the latter continuing to grow until it then "saturates" at $\pi/2$.

The net result of the pass by the fixed point is then an isolated peak in α_d, as portrayed in Figure 4.9(d), entailing an isolated peak in the irreversible rate of strain. *Note then that the influence of hyperbolic fixed points on good stretching is not merely "good straining", but rather good irreversible straining due to the lag of α_t behind α_ℓ.* We have assumed in our explanation of the peaks that α_d and α_t are small as the line element enters the region local to the hyperbolic fixed point. For any pass by a hyperbolic fixed point but the first, this assumption on α_d is good. During the very first pass α_d may be initially large, however, which means only that one gets a "fraction" of a peak in the region local to the hyperbolic fixed point, as seen in Figures 4.6 and 4.7 and portrayed in Figure 4.9(e). The assumption on α_t is good as long as the line

element enters the region local to the hyperbolic fixed point sufficiently close to the local stable manifold; the further a line element is from the stable manifold (*i.e.*, the further into the core it is), the larger α_t will initially be, hence the smaller the lag of α_t behind α_ℓ, and the smaller the straining peak, consistent with numerical observation. With successive passes the approaching value of α_d will be increasingly smaller, making the lag of α_t behind α_ℓ progressively smaller, and hence the peaks less pronounced. Additionally, note that one can further see why the irreversible straining peaks near the fixed points must for both fixed points have *positive* amplitude in the heteroclinic case, as explained in Figure 4.10.

One needs to take care in interpreting stretching in the context of these observations before proceeding to the non-integrable case. At first glance, one might conclude that the hyperbolic fixed points are the engines for good irreversible straining, and hence a goal to maximize strain rates would be to be near these engines. Note, however, that the peaks can correspond to very sizable regions around these fixed points, as should be clear from the y component of the line elements in Figures 4.6 and 4.7. More significantly, as one moves deeper into the core, there is a balance between decreasing peak size (hinders stretching) and increasing peak frequency (helps stretching). It remains to be determined whether having a few large peaks is better or worse than having many small peaks, which necessitates integrating the strain rates, i.e., performing a study of stretch *history*. The stretch analysis given in Chapter 3 and forthcoming in Chapter 6 is thus essential for a complete understanding of the problem.

4.3 Near-integrable flows

In Chapters 2 and 5 we study the role of *turnstile lobes* as mechanisms for transport across partial barriers in phase space (as reminded us by Figure 4.11). In a similar manner, turnstiles play a fundamental role in *increasing*

straining efficiency via re-orientation of line elements and transport of line elements to regions in phase space of superior straining (as encapsulated in Figures 4.12 and 4.13). Figure 4.12(a) shows again for the perturbed open flow the material interface after one revolution around the core, a result not too dissimilar from the integrable case. Figure 4.12(b) shows the dramatic change in the interface under the next iterate of the Poincaré map due to the turnstile lobes. In contrast to the integrable case, where line elements continue to wind around the core with increasing orientation along the streamlines, significant portions of the interface have been dramatically *re-oriented* to recover a large component normal to the unperturbed streamlines (for near-integrable flows these streamlines are relevant, as we shall see). Indeed a significant portion of the interface (and successive portions with increasing time) wraps around $r_0(t = 1)$ (which recall is the portion of the unstable manifold bounding $P_\varepsilon(E(1))$), and *will thus approximately recover the orientation it had at* $t = 1$. Additionally, portions of the interface have been displaced by the turnstiles leftwards relative to the stable manifold so that they will soon enter the region local to the hyperbolic fixed point with significant re-orientation. The SIP's seen in Figure 4.12(b) that are a result of this leftward displacement are an extreme example of this effect - they will in short order enter the region local to the hyperbolic fixed point, and they indeed asymptote to the fixed point while oriented parallel to the unstable manifold, corresponding to optimal straining (or near-optimal straining when the local stable and unstable manifolds are not orthogonal). Two periods later, the consequence of this transport is apparent (see Figure 4.13) - substantial portions of the interface find themselves near the hyperbolic fixed point, oriented essentially parallel to the local unstable manifold, ideal for irreversible straining efficiency. The net result of re-orientation and transport is portrayed heuristically in Figure 4.14. In identifying these mechanisms for enhanced straining efficiency, we should stress that they of course will not necessarily enhance the straining of the *entire interface* over finite or even infinite

time intervals, but rather *portions* of the interface. For example, in the open flow, portions of the interface are *detrained* with each new time sample, and this effect of *transport* via the turnstiles is ultimately detrimental to straining (since the detrained portion asymptotes infinitely far away from the hyperbolic fixed points). In a similar vein, some of the interface that wraps around $r_0(t = 1)$ is not dramatically re-oriented (*e.g.*, near the "tip" of the lobe). These realities do not detract from our message, however, for the goal in explaining good stretching in chaotic tangles is not to argue its occurrence everywhere in the tangle on finite or even infinite time scales, but on a significant subset of the ensemble of interest. We should also stress how the two mechanisms of re-orientation and transport to regions of superior straining of course need not go hand-in-hand; significant portions of the interface can be re-oriented without any obvious gainful transport.

The results for straining are illustrated in Figures 4.15 to 4.18. Figure 4.15 shows open flow strain rate plots for the extreme case of PIP's and SIP's, corresponding to asymptotically optimal or near-optimal straining, where the non-decay of strain rate asymmetries is obvious. Though these are isolated points, they and their surrounding line elements can have a significant effect on interfacial stretching (see Chapter 6). Figure 4.16 shows more generic strain rate plots for the open and closed flow. The basic observation to make is that, as in the integrable case, there are asymmetries (more positive strain rates than negative), but in contrast to the integrable case the asymmetries do not appear to be decaying. Though this non-decay is plain to see, the large-amplitude oscillations clearly render the conventional strain rate difficult to decipher, motivating some sort of irreversible strain rate. As mentioned earlier, lacking streamlines, there is no *exact* expression to subtract out. There are several possible expressions, however, that work reasonably well in the near-integrable case. We choose here to define an irreversible strain rate simply by subtracting out the *same term as in the integrable case*, to give the expressions

identical to equations (4.2.5) to (4.2.7). We stress that, though we are in the non-integrable regime, we are subtracting out the straining term associated with the integrable flow. To indicate the validity of such an approach, Figure 4.17 shows the term we subtract out

$$\sigma_{rev} = \tilde{m}_i \tilde{m}_j \frac{\partial f_i}{\partial x_j} \tag{4.2.13}$$

and compares with another possible choice

$$\sigma_{rev} = \tilde{m}_i \tilde{m}_j \frac{\partial (f_i + \varepsilon g_i)}{\partial x_j}. \tag{4.2.14}$$

One possible concern of subtracting out expression (4.2.13) or (4.2.14) is that, since line elements in the perturbed flow are not confined to a streamline, what we are subtracting out does not have mean zero. For small perturbations, however, the expressions (4.2.13) and (4.2.14) exhibit oscillations with mean *close* to zero, as is clear from Figure 4.17, so in this context the scheme works approximately well. Indeed, the difference in the area bounded by the positive strain rate and the $\sigma = 0$ axis, $\mathcal{A}_+$, and the area bounded by the negative strain rate and the $\sigma = 0$ axis, $\mathcal{A}_-$, is for each case less than one percent of $(\mathcal{A}_+ + \mathcal{A}_-)/2$. It is also encouraging to note from Figure 4.17 that expressions (4.2.13) and (4.2.14) exhibit quite similar continuous-time behavior, indicating that in the near-integrable setting our approximate approach is fairly robust. The resulting irreversible strain rates are shown in Figure 4.18, rendering the essence of the straining process relatively apparent. We observe:

(i) For each pass by the left hyperbolic fixed point in the open flow and each sufficiently close pass by the hyperbolic fixed point in the closed flow, there is a sizable peak in the irreversible strain rate, and *these peaks are not decaying with increasing number of revolutions.*

(ii) Besides these isolated peaks, there are some additional secondary peaks, dips, or small oscillatory behavior. Though the absence of such behavior would have given a much more visually appealing plot (isolated non-decaying peaks separated by dead intervals), the reality of non-integrable

flows is the existence of these relatively minor variations. Their existence can be attributed to the following.

(a) The turnstile can re-orient line elements well before they reach the region local to the hyperbolic fixed point, and this re-orientation can entail peaks and dips before the pass by the fixed point. The major peaks, however, are associated with the pass by the hyperbolic fixed point.

(b) The overall breathing or pulsating of the lobe structure, and line elements within the structure, and the oscillating perturbing straining field can entail additional variations in the irreversible strain rates.

As an example of (ii)(a), in the closed flow the unstable manifold bounding the image of the detraining turnstile, $P_\varepsilon(D(1))$, shows significant deviation angles at very large values of y (refer back to Figure 3.6). As an example of (ii)(b), note that the peak associated with the pass by the *right* hyperbolic fixed point in the open flow (which recall has a significantly smaller peak than the pass by the left fixed point in the integrable case as a consequence of the decay process), is either reinforced or hindered depending on the phase of the perturbing straining field. The basic message of non-diminishing peaks associated with passes by hyperbolic fixed points via re-orientation by turnstiles is nevertheless clear. The effect of transport of line elements is perhaps slightly less clear, demanding a qualification in conjunction with our previous remarks about whether passing closer to the hyperbolic fixed points is necessarily more beneficial to stretching. Referring back to the stretch history analysis of Chapter 3, the closed flow's highly non-uniform stretch profile shows extremely good stretching highly localized near the PIP's and SIP's. Clearly in this case, passing near the hyperbolic fixed point is most beneficial. The more uniform open flow profile, however, indicates that some of the best stretching can occur well between the PIP's and SIP's. In this case, remaining further away from the hyperbolic fixed point is

not necessarily unhelpful to straining. The effect of transport thus depends on the spatial disparity of stretching in phase space, namely, whether or not the linearized flow about the hyperbolic fixed point corresponds to significantly better straining than elsewhere. Nonetheless, being transported near hyperbolic fixed points with proper re-orientation can of course never *hurt*, and under fairly common conditions it is unquestionably helpful. In closing we remark that our illustrations (especially the open flow) correspond to rather sizable perturbations. Even in these cases the above non-integrable complications are not overwhelming, and for smaller perturbations they can be truly minor.

Appendix - Consequence of a $\sqrt{D:D}$ normalization

If D and m each represent two-component vectors, then by the Cauchy-Schwarz inequality

$$D_i m_i \leq \sqrt{D_i D_i}\sqrt{m_j m_j}. \tag{4.A.1}$$

If m is a unit vector, then we have

$$D_1 m_1 + D_2 m_2 \leq \sqrt{D_1^2 + D_2^2}, \tag{4.A.2}$$

and it is of course possible to obtain an equality by choosing

$$m_i = \frac{D_i}{\sqrt{D_1^2 + D_2^2}} \qquad \text{for } i = 1, 2. \tag{4.A.3}$$

For D a 2×2 matrix we have

$$D_{ij} m_i m_j \leq \sqrt{D_{ij} D_{ij}}\sqrt{m_k m_\ell m_k m_\ell}. \tag{4.A.4}$$

If we again choose m to be a unit vector, and also choose D_{ij} to be diagonal, we then have

$$D_{11} m_1^2 + D_{22} m_2^2 \leq \sqrt{D_{11}^2 + D_{22}^2}. \tag{4.A.5}$$

Since m_1^2 and m_2^2 cannot be negative, for D_{11} and D_{22} oppositely signed one will not in general be able to choose an m to obtain an equality in equation (4.A.5). Hence, a $\sqrt{D:D}$ can fail to give an efficiency of one in instances where such a value would be natural (*e.g.*, see Figure 4.5).

References

[1] G.K. Batchelor, J. Fluid Mech. **5**, 113 (1958).

[2] P.E. Dimotakis, Caltech preprint (1989).

[3] Dresselhaus and Tabor, J. Phys. A: Math. Gen **22**, 971 (1989).

[4] J.M. Ottino, *The kinematics of mixing: stretching, chaos, and transport* (Cambridge University Press: Cambridge, 1989).

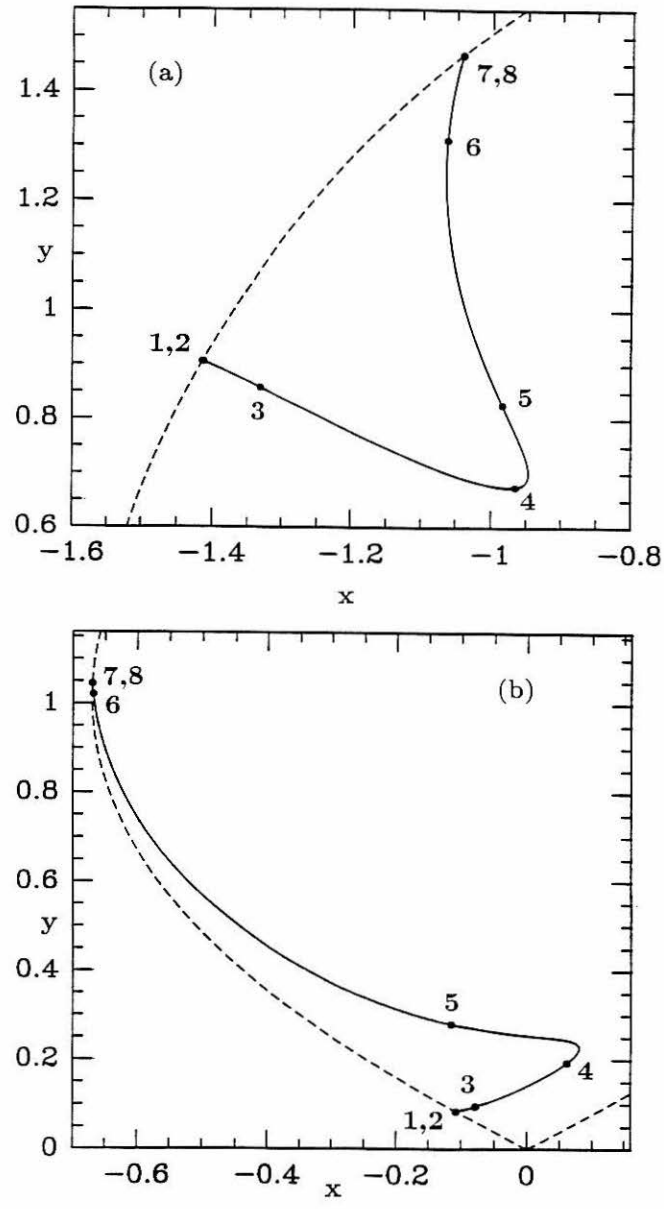


Figure 4.1 The set of infinitesimal line elements distributed along the material interface $\tilde{r}_0(t=1)$ of Figure 3.3 for (a) the unperturbed open flow and (b) the unperturbed closed flow ($x \equiv x_1, y \equiv x_2$). Points 1 and 8 are situated identically at the endpoints (*i.e.*, they are points $\tilde{p}1$ and $\tilde{p}2$, respectively), while points 2 and 7 are very close by: $s_2(0) - s_1(0) = 8.39 \times 10^{-4} (2.47 \times 10^{-8})$ and $s_8(0) - s_7(0) = 1.29 \times 10^{-3} (5.86 \times 10^{-7})$ for the open (closed) flow, where $s_i(0)$ denotes the initial arclength along the curve for the i^{th} point.

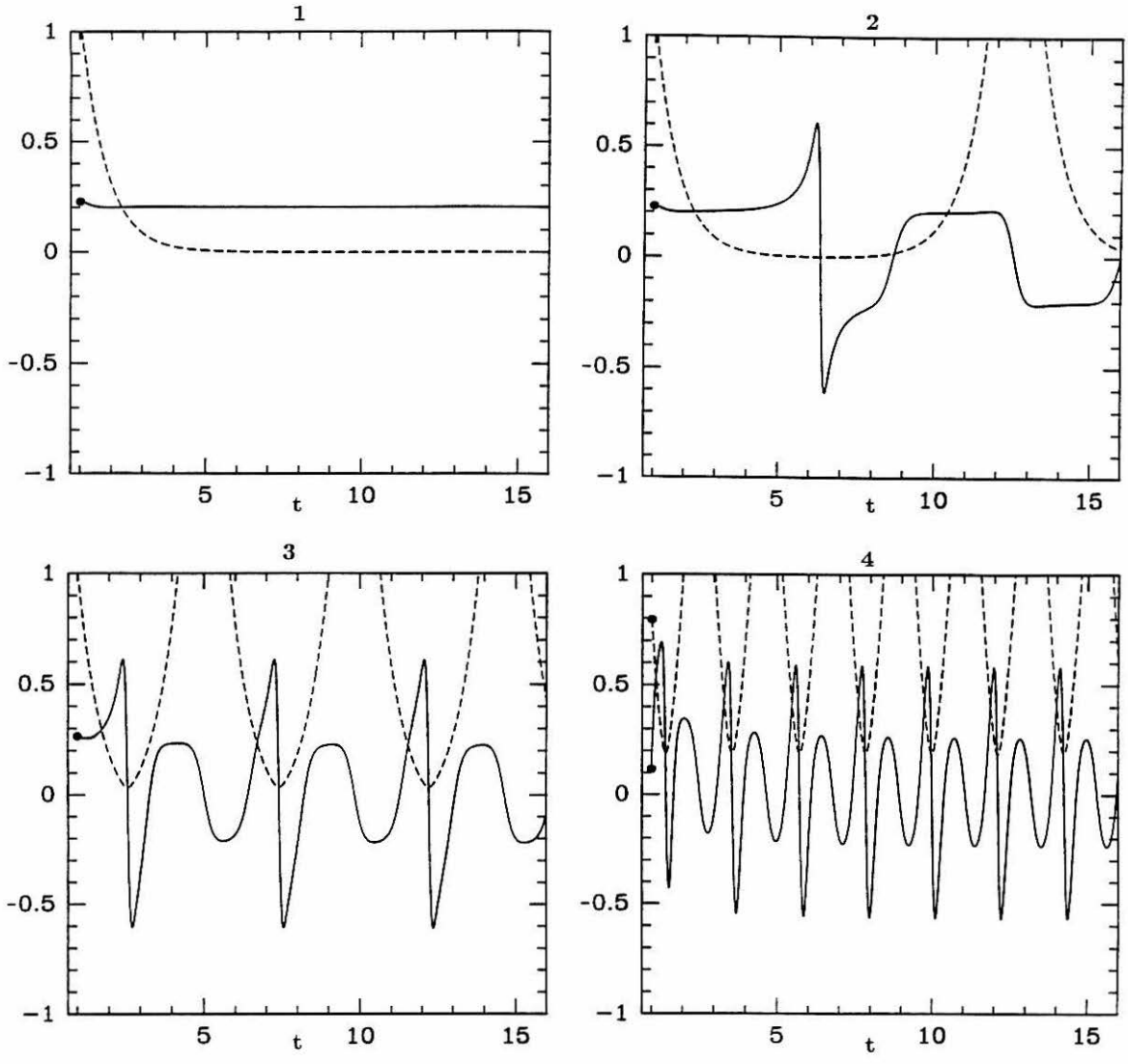


Figure 4.2 Continuous-time open flow strain rate plots for the points shown in Figure 4.1(a). The solid line is the strain rate, and the dashed line is the vertical ($x_2 \equiv y$) location of the infinitesimal line element. *Note that the strain rates are given in units of* 2π. This convention is a consequence of the fact that the units originally used in this study differed from those of Chapters 3 and 6, and the above convention renders the present units consistent with the other two chapters. Further note that the unstable eigenvalue at the hyperbolic fixed points is found analytically to be $\lambda_u = 1.281 = 0.204 \times 2\pi$.

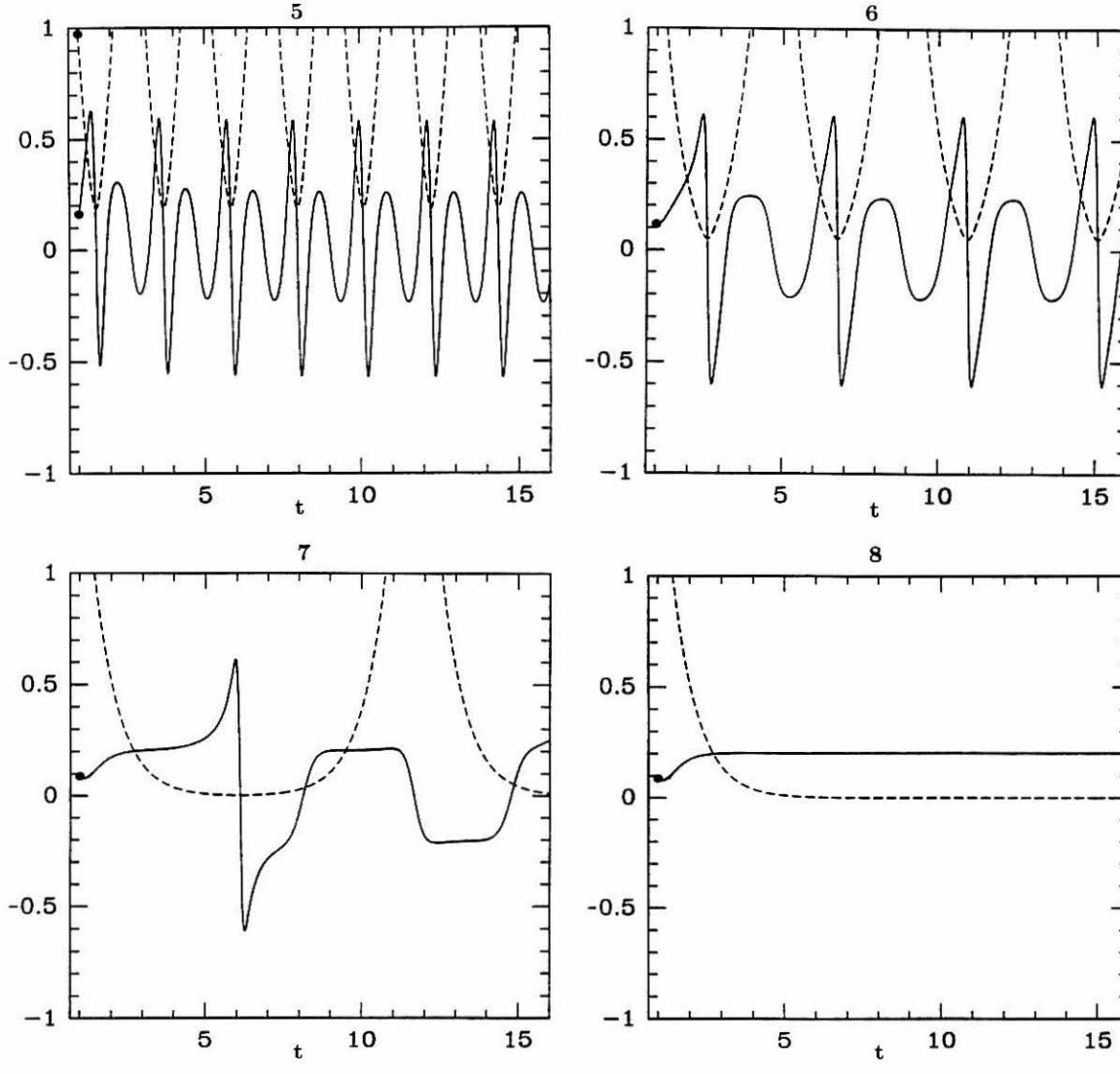


Figure 4.2 *Continued.*

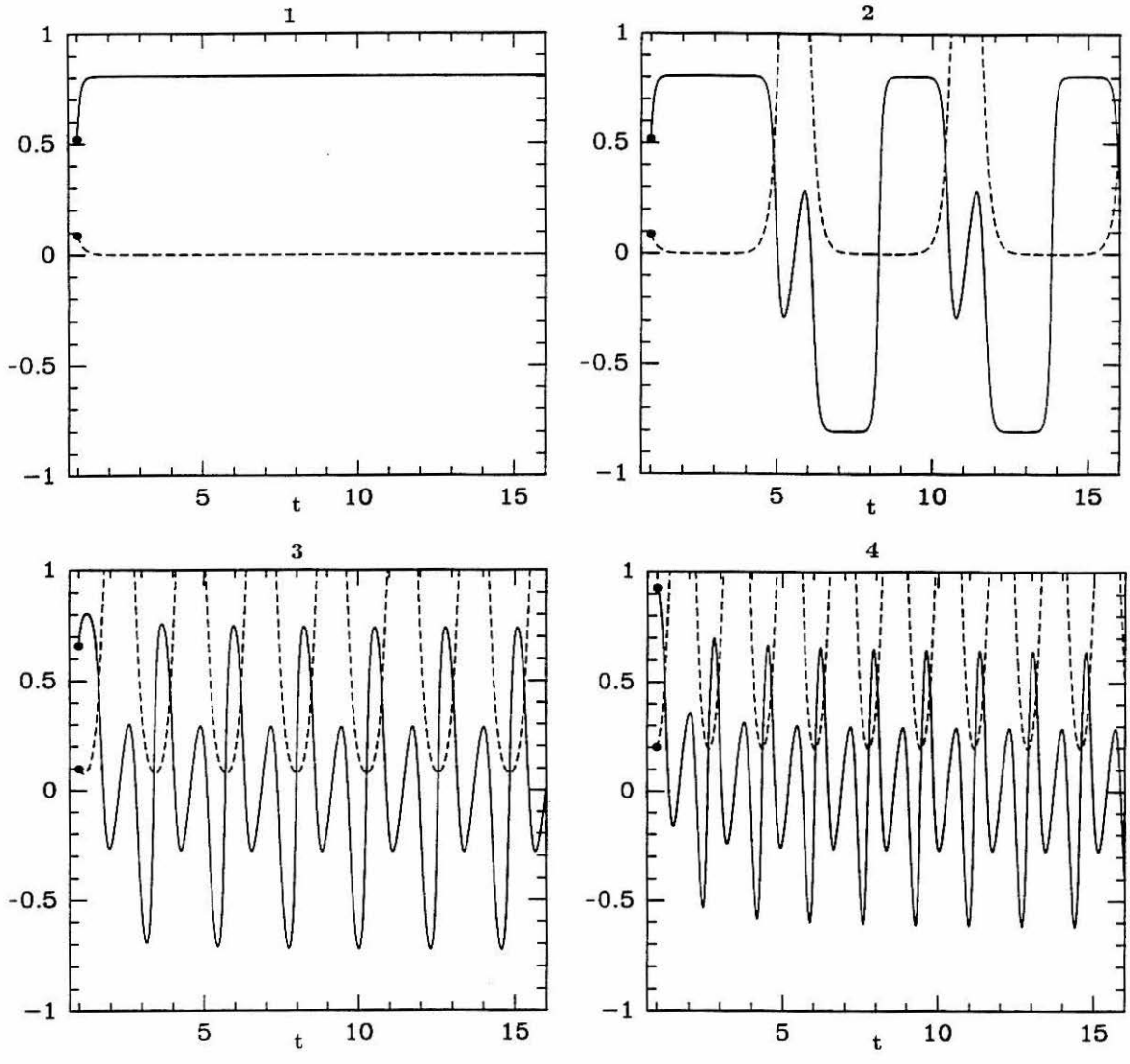


Figure 4.3 Continuous-time closed flow strain rate plots for the points shown in Figure 4.1(b). The solid line is the strain rate, and the dashed line is the vertical location of the infinitesimal line element. Note that the unstable eigenvalue of the hyperbolic fixed point located at the origin is found analytically to be $\lambda_u = 5.064 = 0.806 \times 2\pi$.

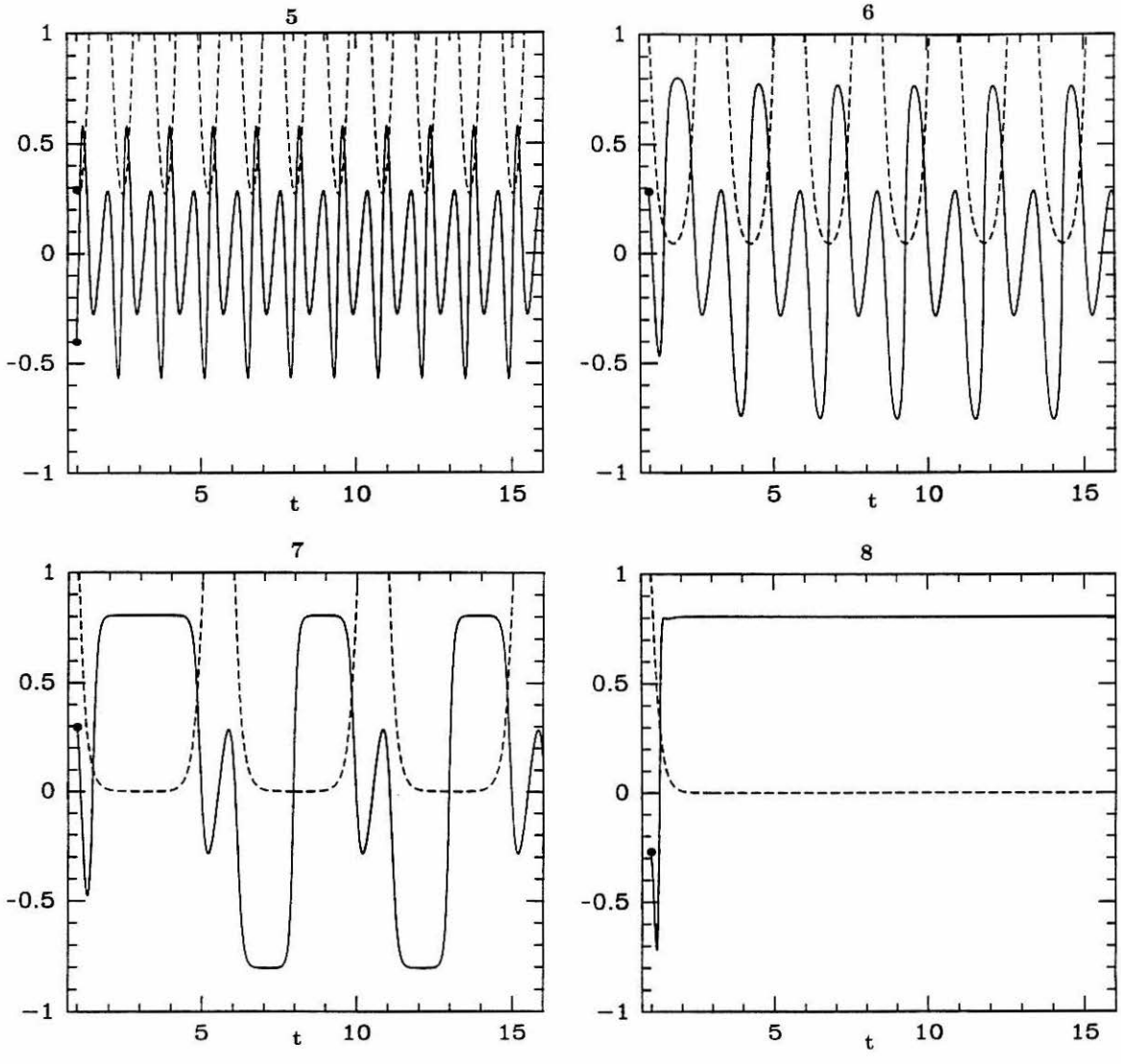


Figure 4.3 *Continued.*

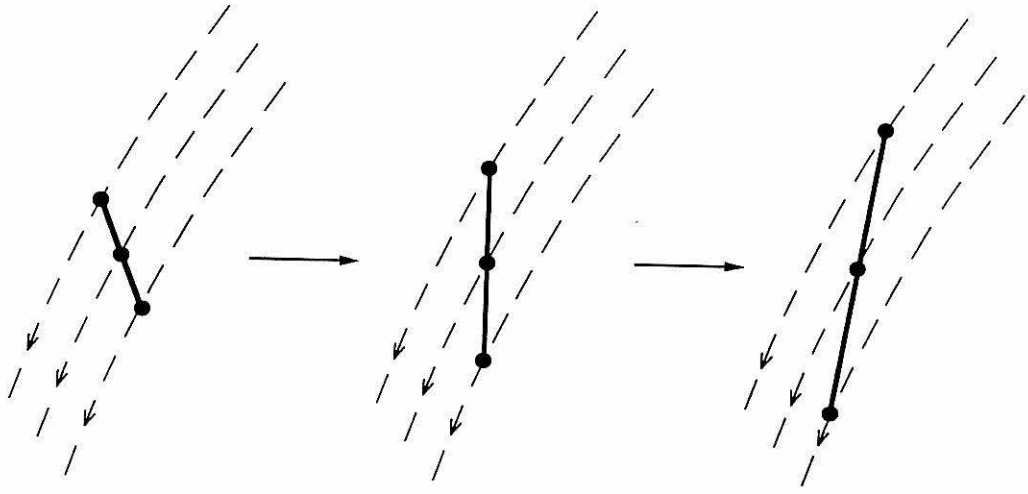


Figure 4.4 As time progresses, the line element (solid line) under the integral flow becomes progressively oriented along the streamlines (dashed lines), as a consequence of shear. The diagram shows the initial line element and the results after two successive revolutions around the closed streamline.

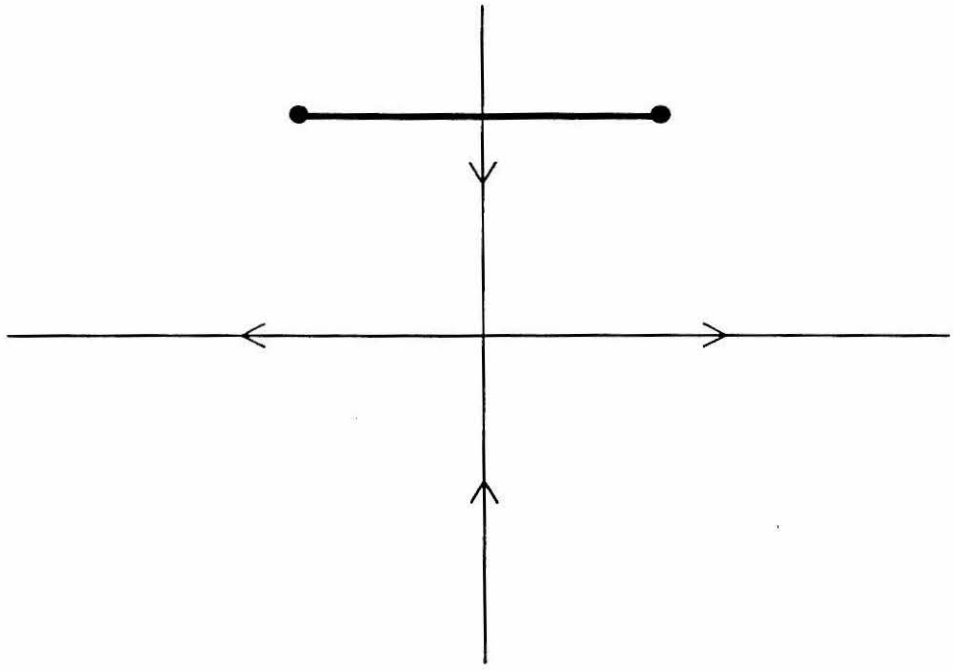


Figure 4.5 A line element approaches parallel to the local unstable manifold a hyperbolic fixed point with orthogonal local stable and unstable manifolds.

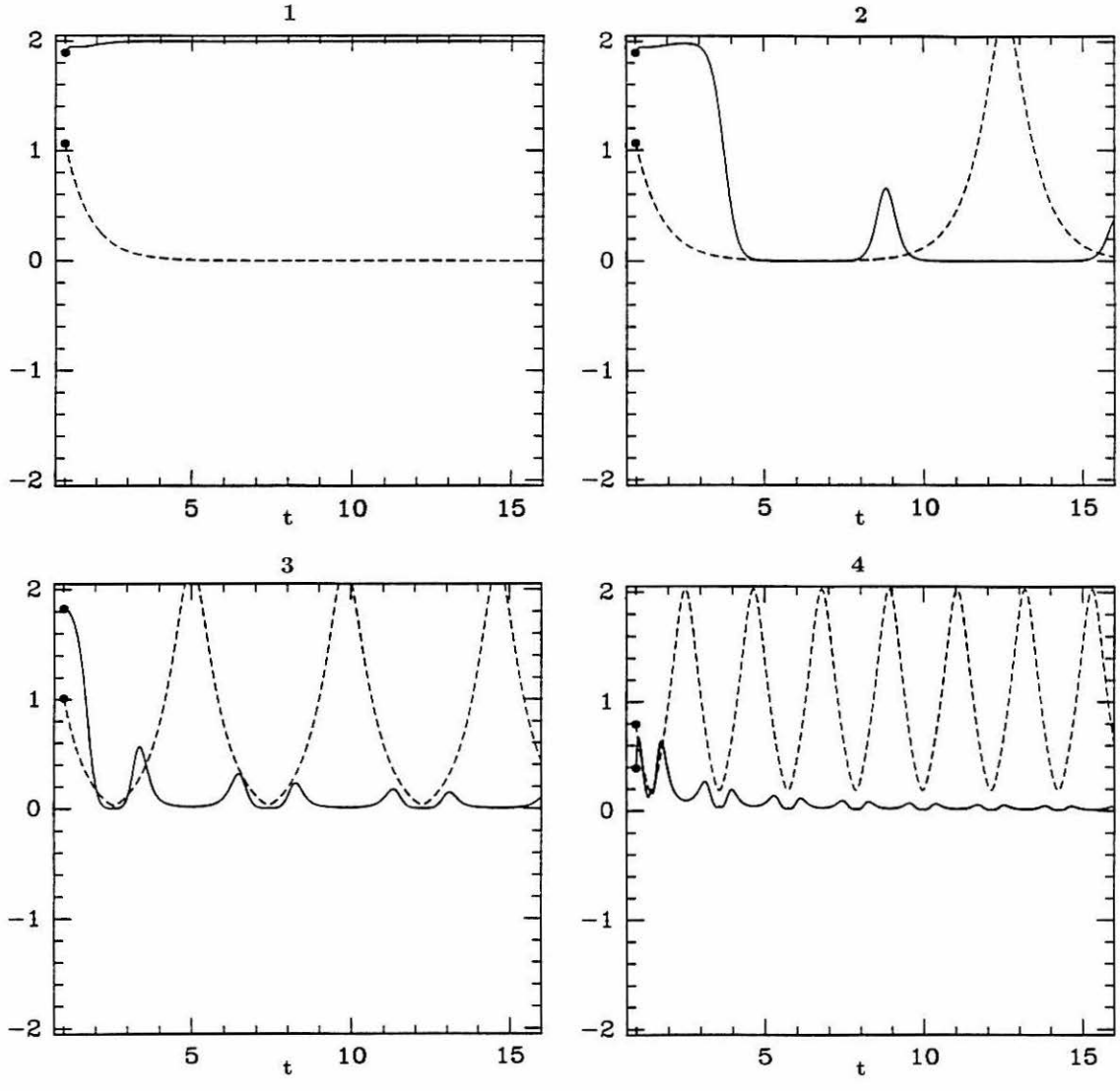


Figure 4.6 Open flow irreversible straining efficiency (solid line) for the points shown in Figure 4.1. Again the dashed line is the vertical location of the infinitesimal line element.

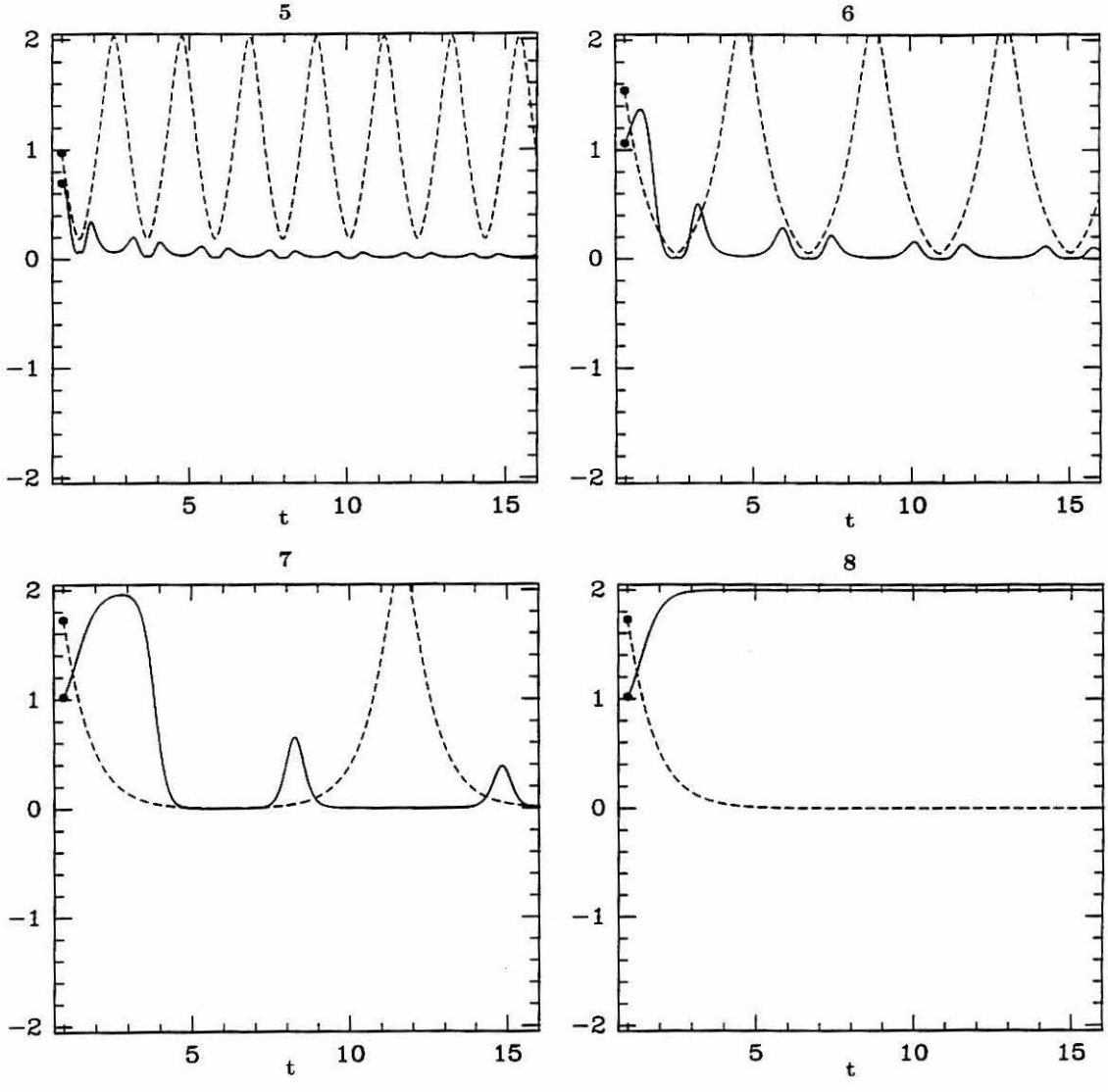


Figure 4.6 *Continued.*

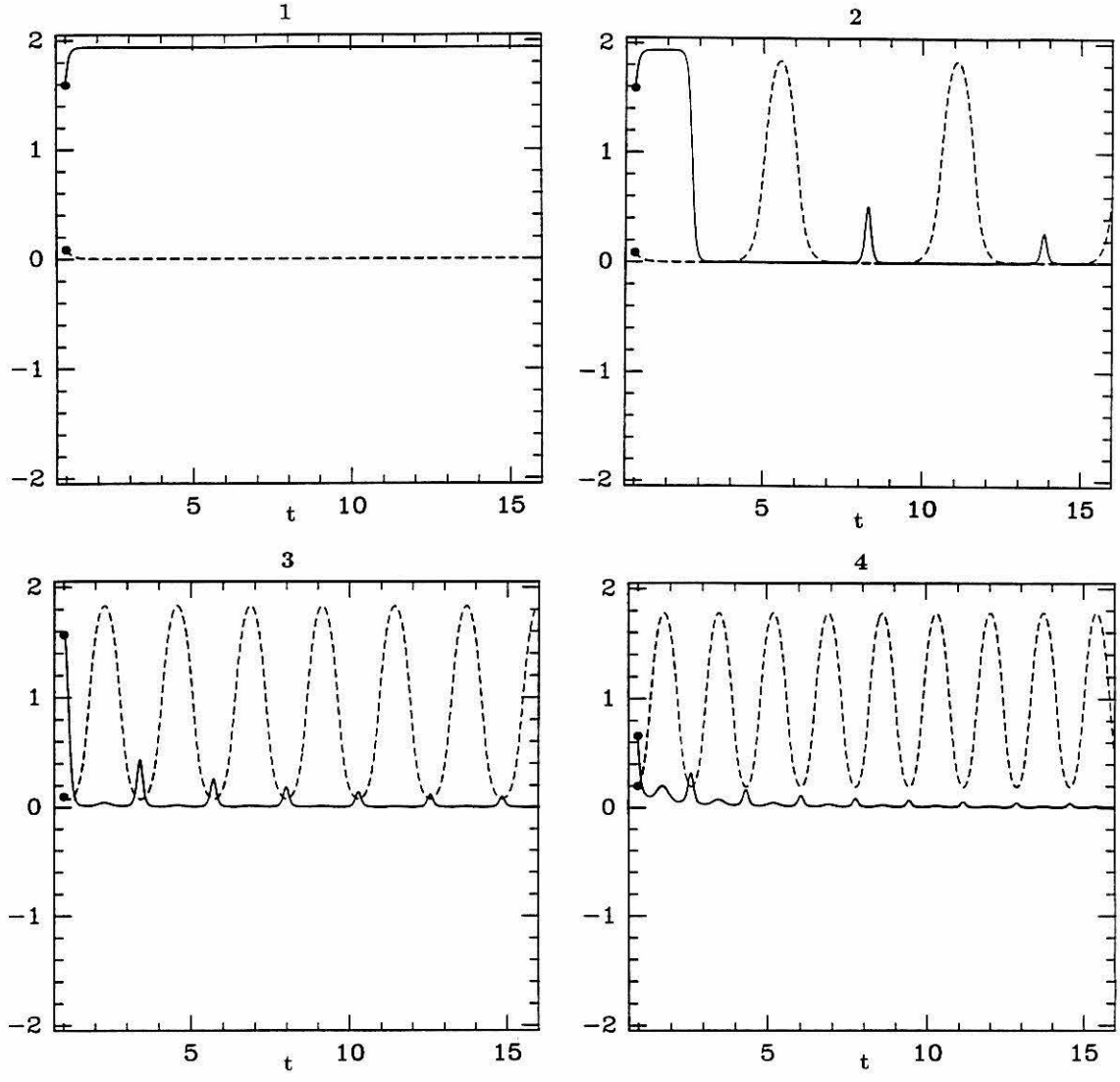


Figure 4.7 Closed flow irreversible straining efficiency (solid line) for the points shown in Figure 4.1. The dashed line is the vertical location of the infinitesimal line element. Notice that line elements asymptoting to the hyperbolic fixed point oriented along the local unstable manifold correspond to σ^e_{irr} close to two, but not identically two, a consequence of the non-orthogonality of the local stable and unstable manifolds.

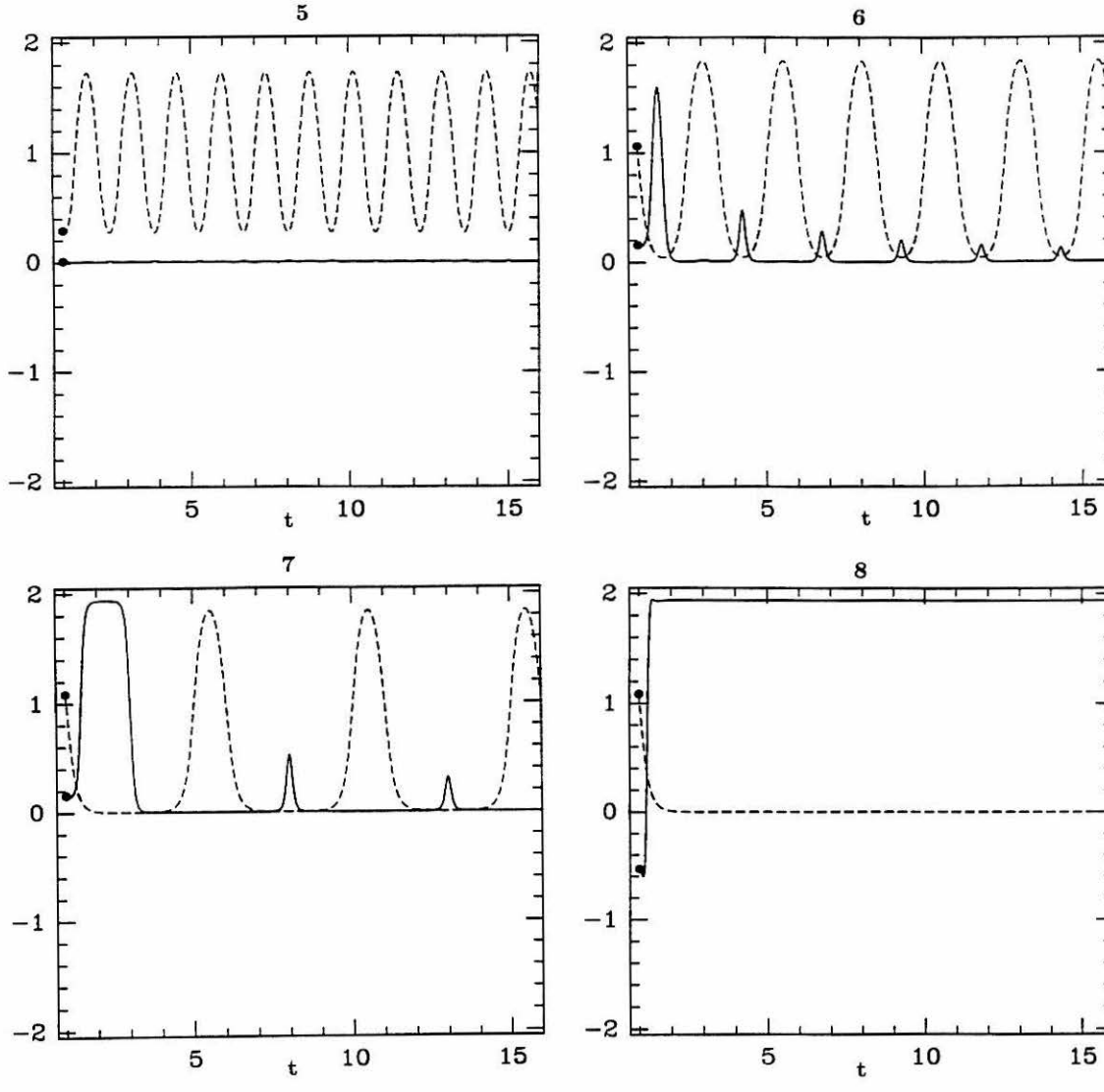
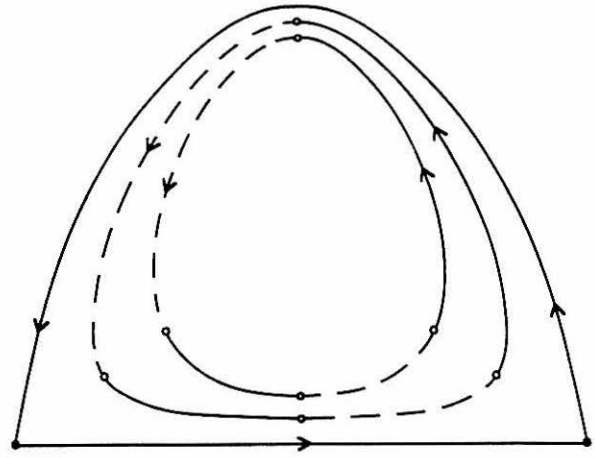
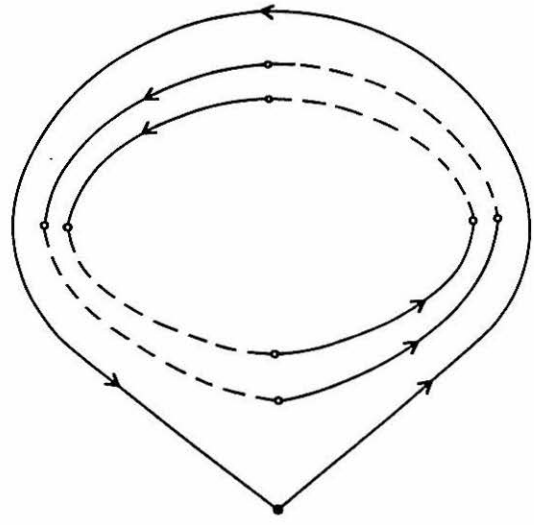


Figure 4.7 *Continued.*



(a)



(b)

Figure 4.8 Portrayal of neighboring streamlines in the (a) open and (b) closed flows, with the spatial oscillations in the streamline separation exaggerated for sake of clarity. The open dots indicate points of extremal streamline separation, corresponding to zero rate of strain along the streamlines, and thus delineating between regions of positive (solid) and negative (dashed) strain rate along the streamlines.

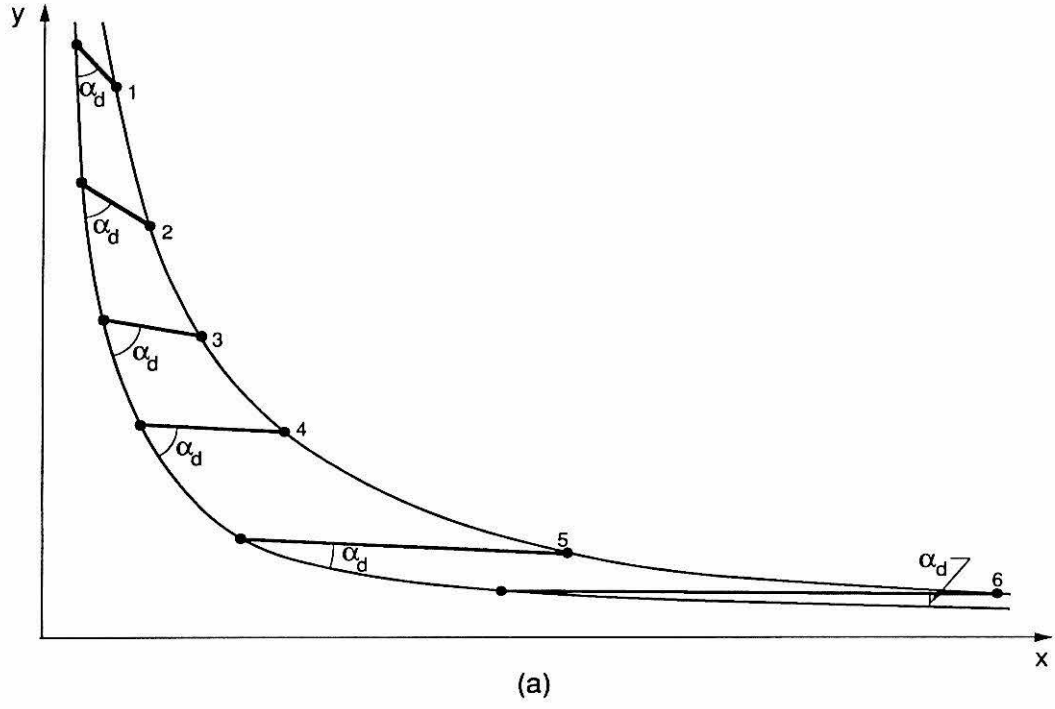


Figure 4.9 Understanding the irreversible strain rate peaks associated with passage by a hyperbolic fixed point. (a) Portrayal of the peak in the deviation angle α_d as the line element passes through the region local to the hyperbolic fixed point.

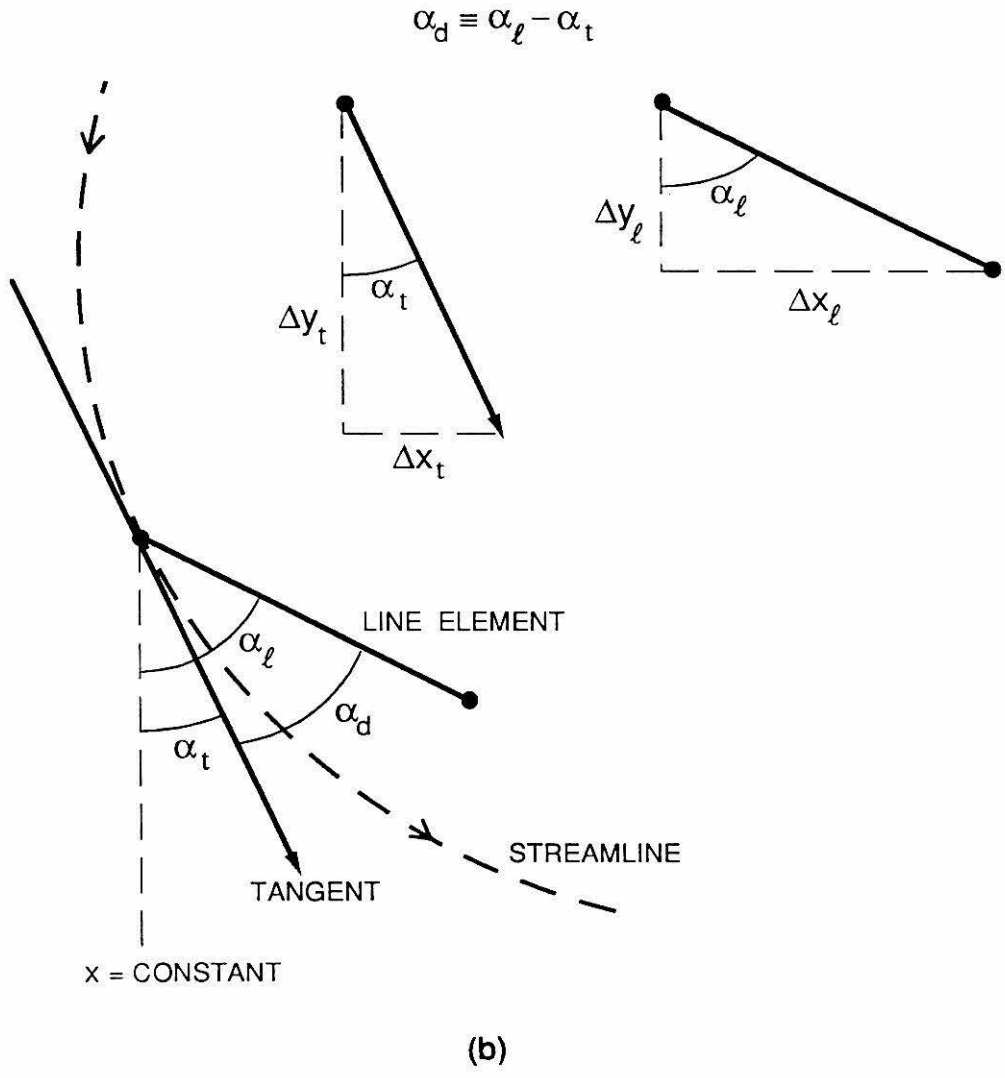
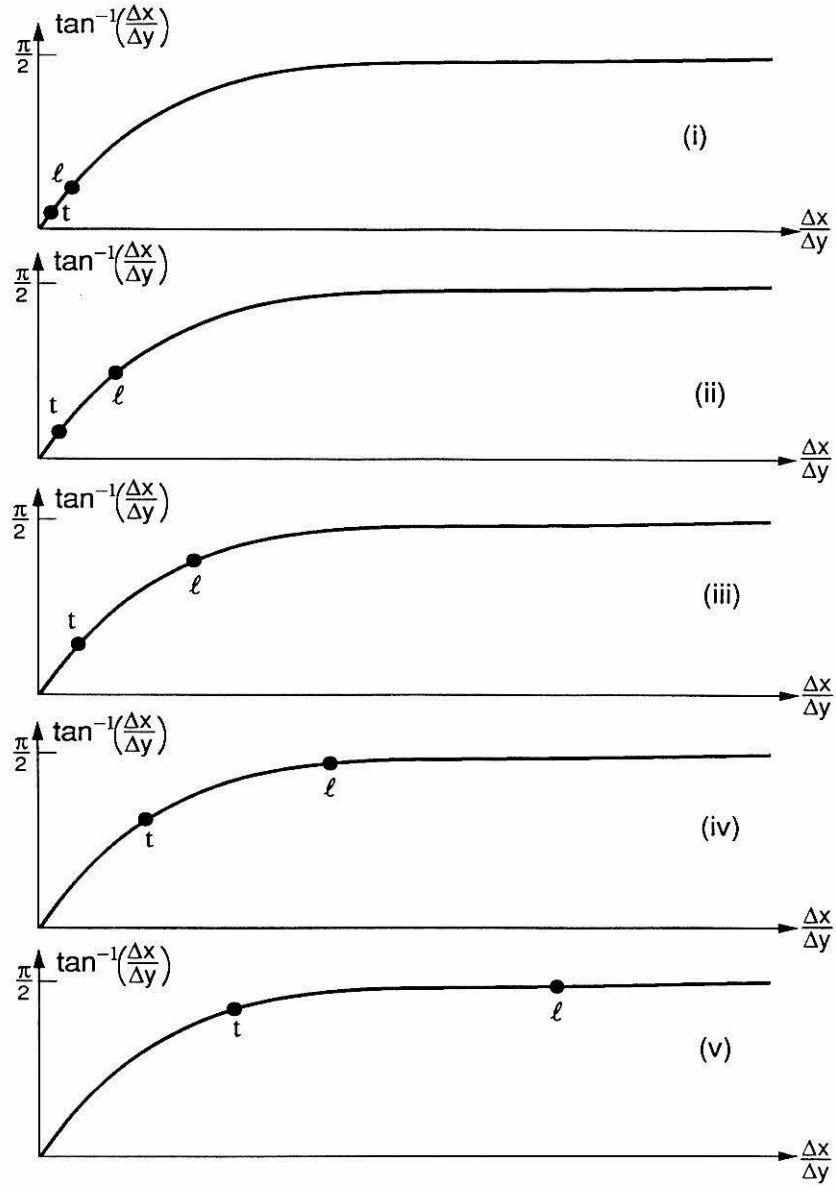
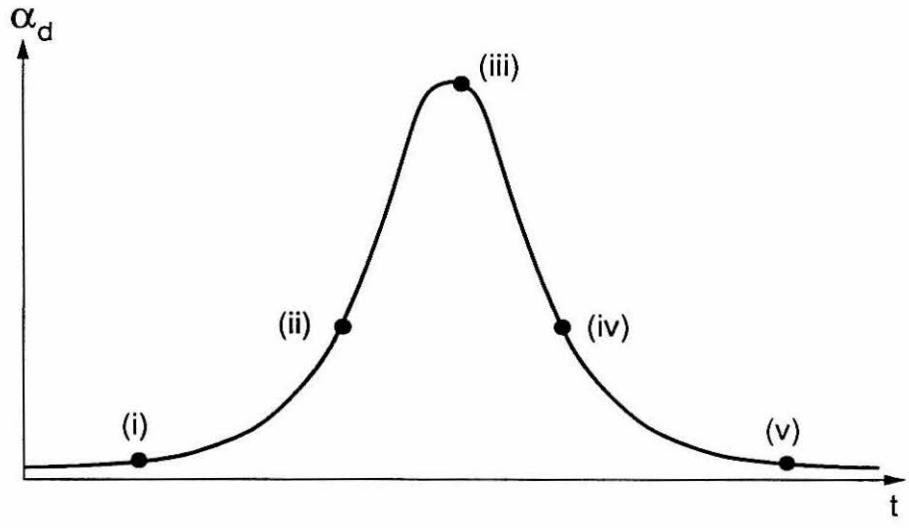


Figure 4.9 (b) Expressing the deviation angle in terms of α_t and α_ℓ.

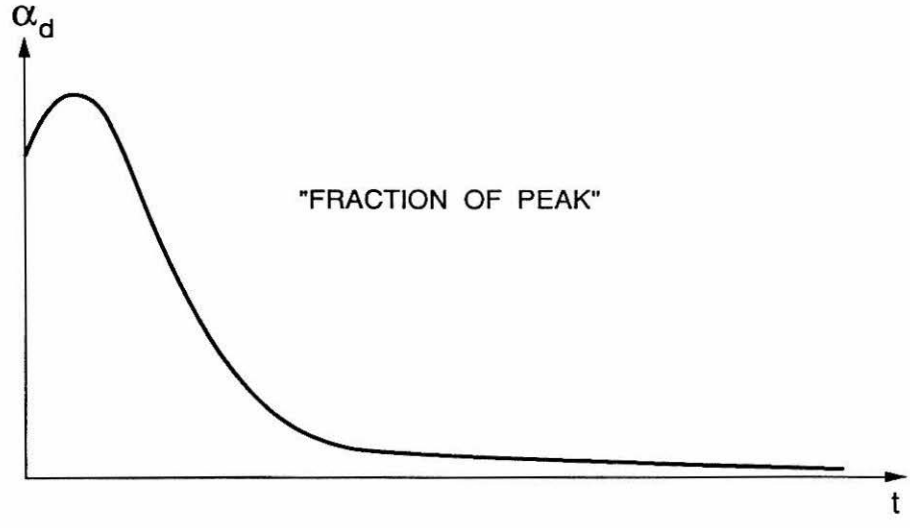


(c)

Figure 4.9 (c) Converting slope to angle via $\tan^{-1}(\Delta x/\Delta y)$, and visualizing, in terms of the lag of α_t behind α_ℓ, the growth in α_d followed by saturation and reversal.



(d)



(e)

Figure 4.9 (d) The resulting isolated peak in α_d, and (e) a "fraction" of a peak experienced in the region local to the hyperbolic fixed point when α_d is initially not small.

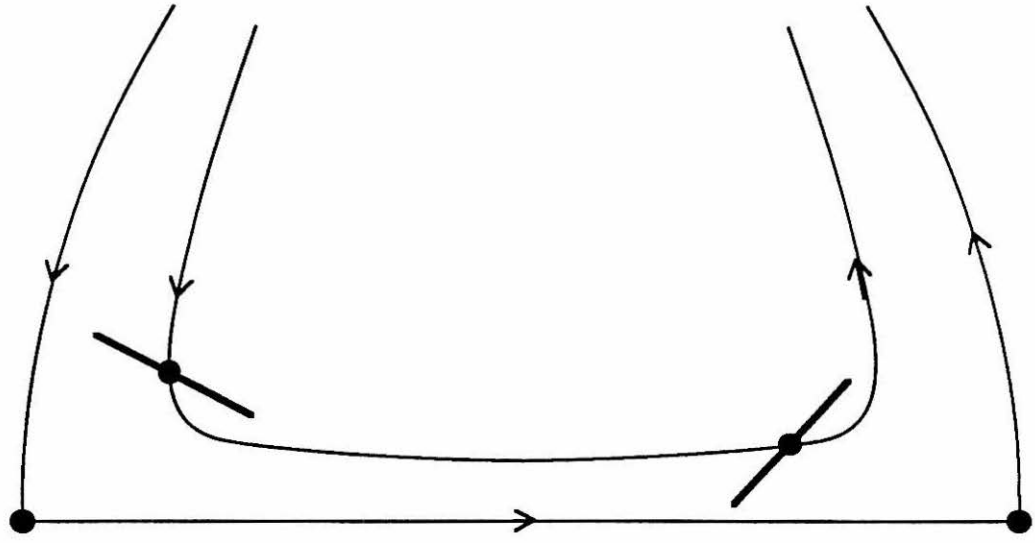


Figure 4.10 In each pass by the open flow's hyperbolic fixed points, the line element deviation from the streamline tangent *enhances* irreversible straining, rather than enhance in one case and hinder in the other. In each case shown the deviation makes the line element more parallel to the unstable manifold.

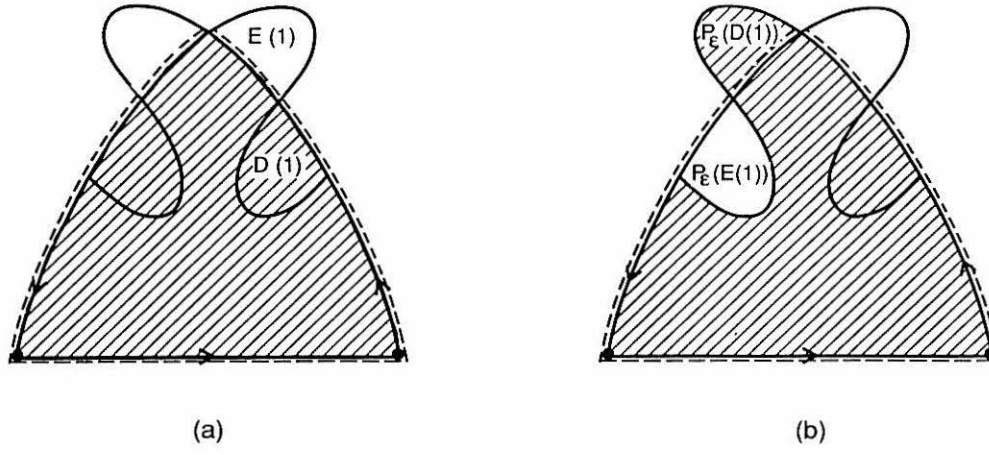


Figure 4.11 The turnstile lobes $E(1)$ and $D(1)$ are the mechanism for transport in and out of the core region defined by the dashed line. P_ε denotes the Poincaré map defined by time-periodic sampling (see Chapter 3).

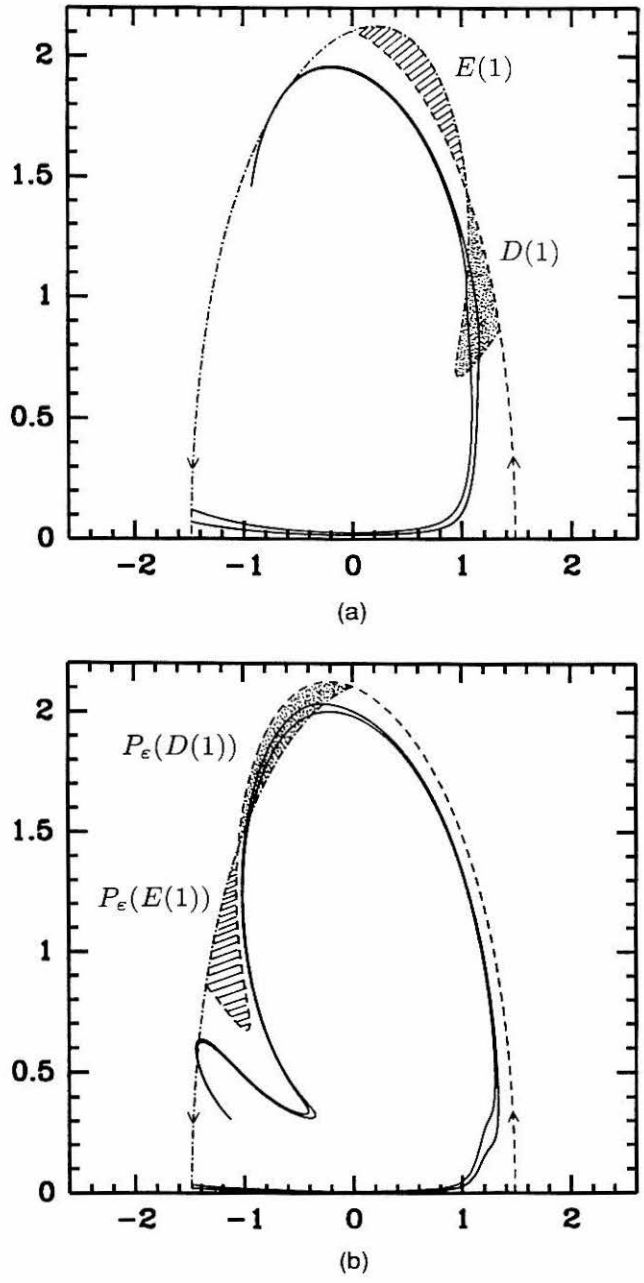
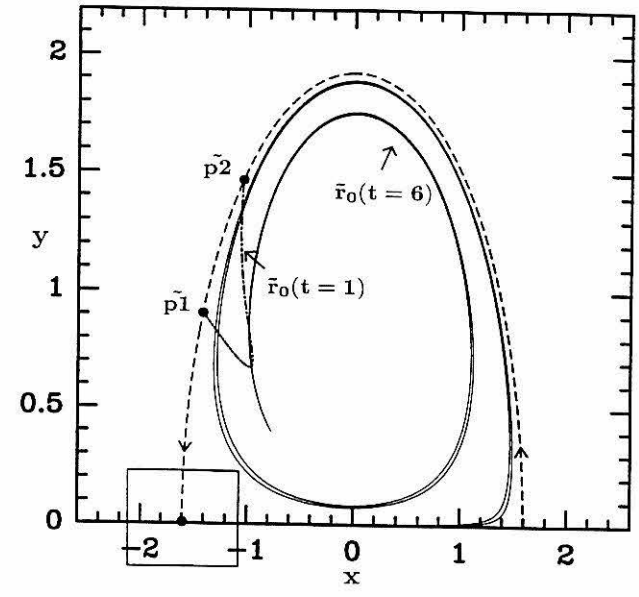
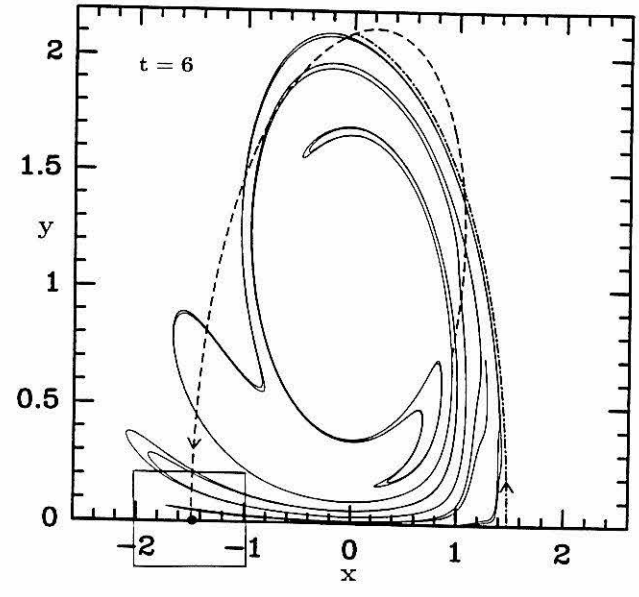


Figure 4.12 The perturbed open flow material interface at successive times (a) $t = 3$ and (b) $t = 4$. Substantial portions of the interface are re-oriented and some portions are transported to regions of superior straining. In particular a portion of the lobe wraps around $r_0(t = 1)$ (*i.e.*, around $P_\epsilon(E(1))$), thus approximately regaining its orientation at $t = 1$.



(a)



(b)

Figure 4.13 Revisiting the (a) integrable and (b) near-integrable open flow of Chapter 3 at $t = 6$. The box highlights the results of re-orientation and transport via the turnstiles: a significant portion of the interface has been re-oriented and transported to the region local to the hyperbolic fixed point.

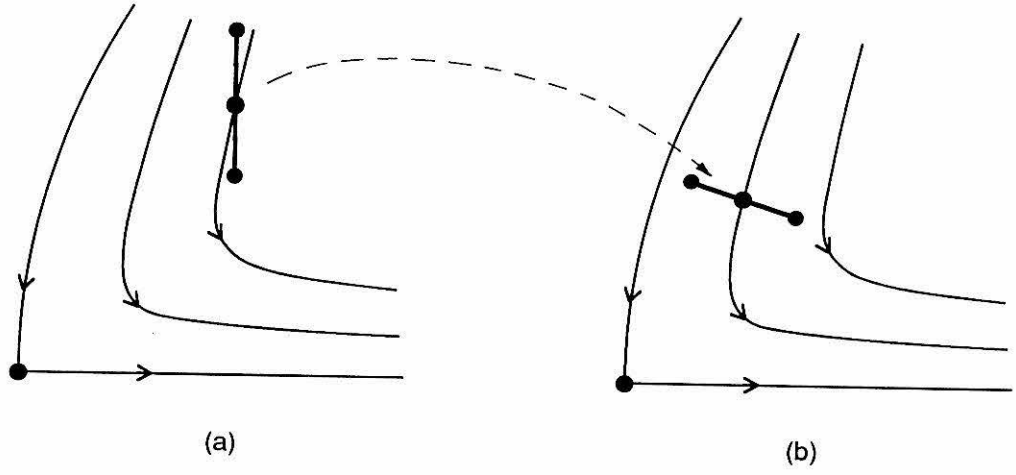


Figure 4.14 Portrayal of the effects of re-orientation and transport: (a) a line element after several revolutions around the core under the integrable flow, and (b) where that line element might end up after evolution under the near-integrable flow over the same time interval.

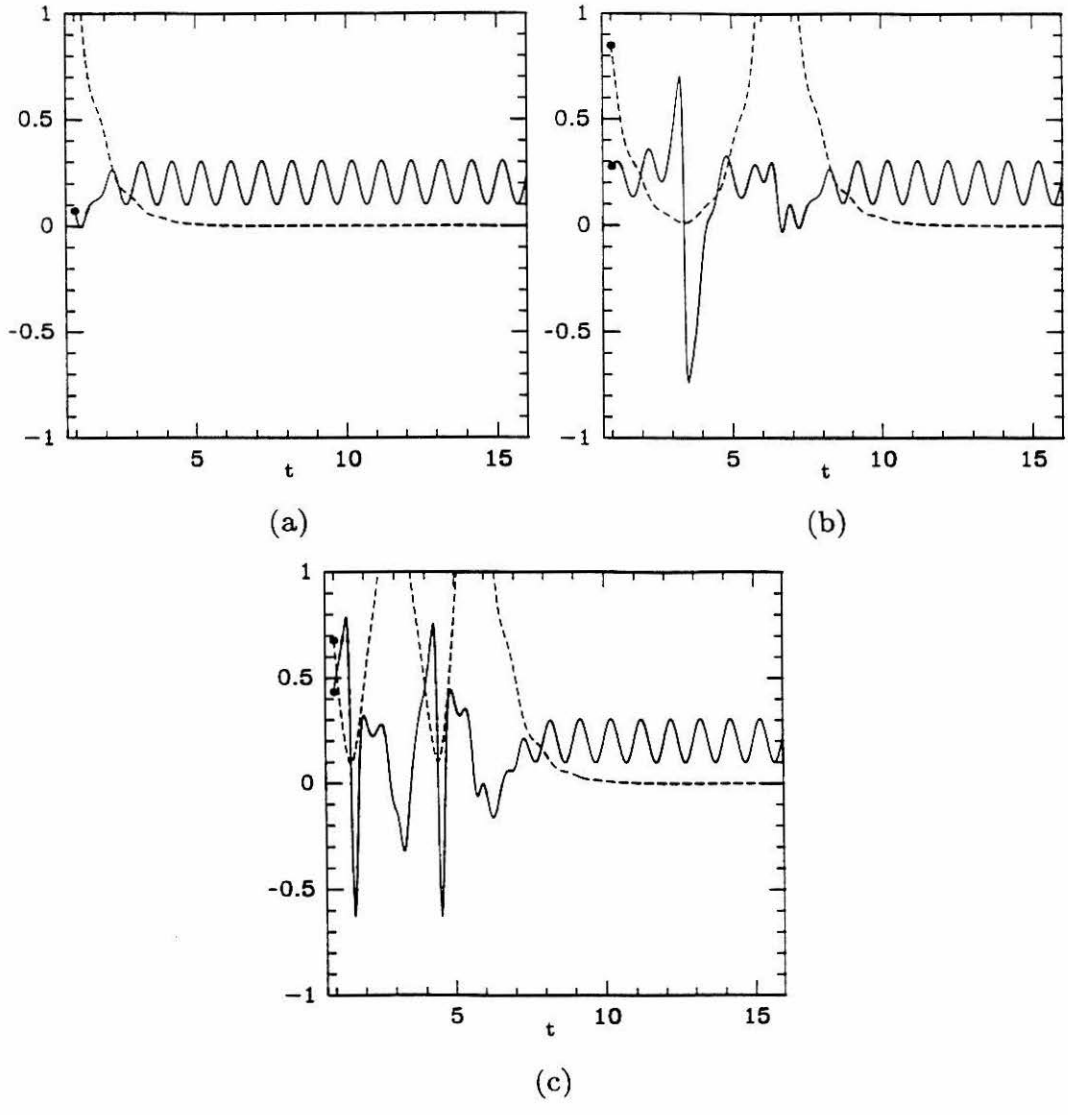
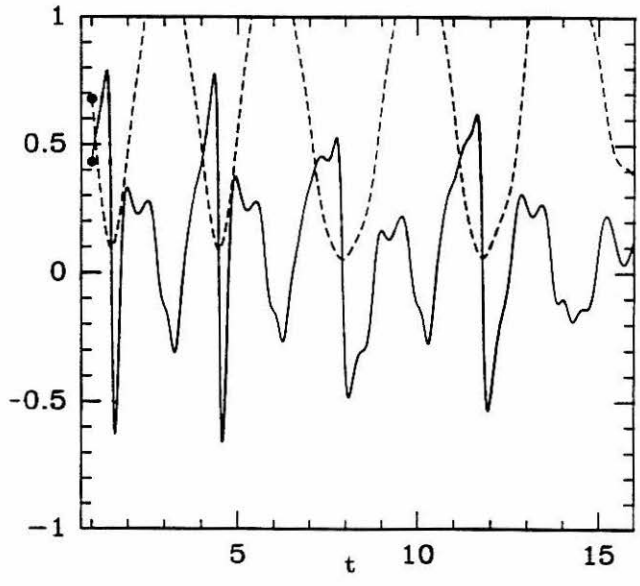
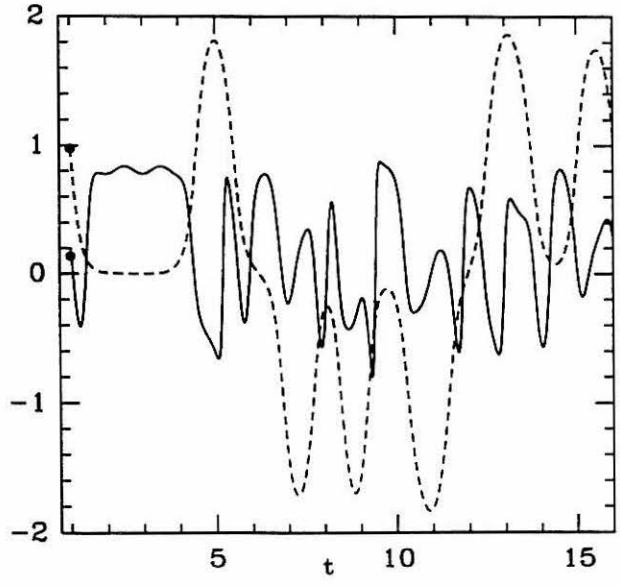


Figure 4.15 Perturbed open flow strain rate plots (solid lines) for infinitesimal line elements on r_0 associated with (a) a PIP, (b) a SIP that revolves around the core once, and (c) another SIP that revolves around the core twice. Recall from Chapter 3 that a point can revolve around the core anywhere from once to arbitrarily many times, with anywhere from some minimum to an arbitrarily long revolution time, before it is either detrained or asymptotes to a hyperbolic fixed point. Again (as for the remaining plots) the dashed line represents the vertical location of the line element.



(a)



(b)

Figure 4.16 More generic strain rate plots associated with infinitesimal line elements on r_0 for the perturbed (a) open flow and (b) closed flow. The negative values of y in the closed flow are due to the freedom of line elements to revolve around both the upper and lower cores.

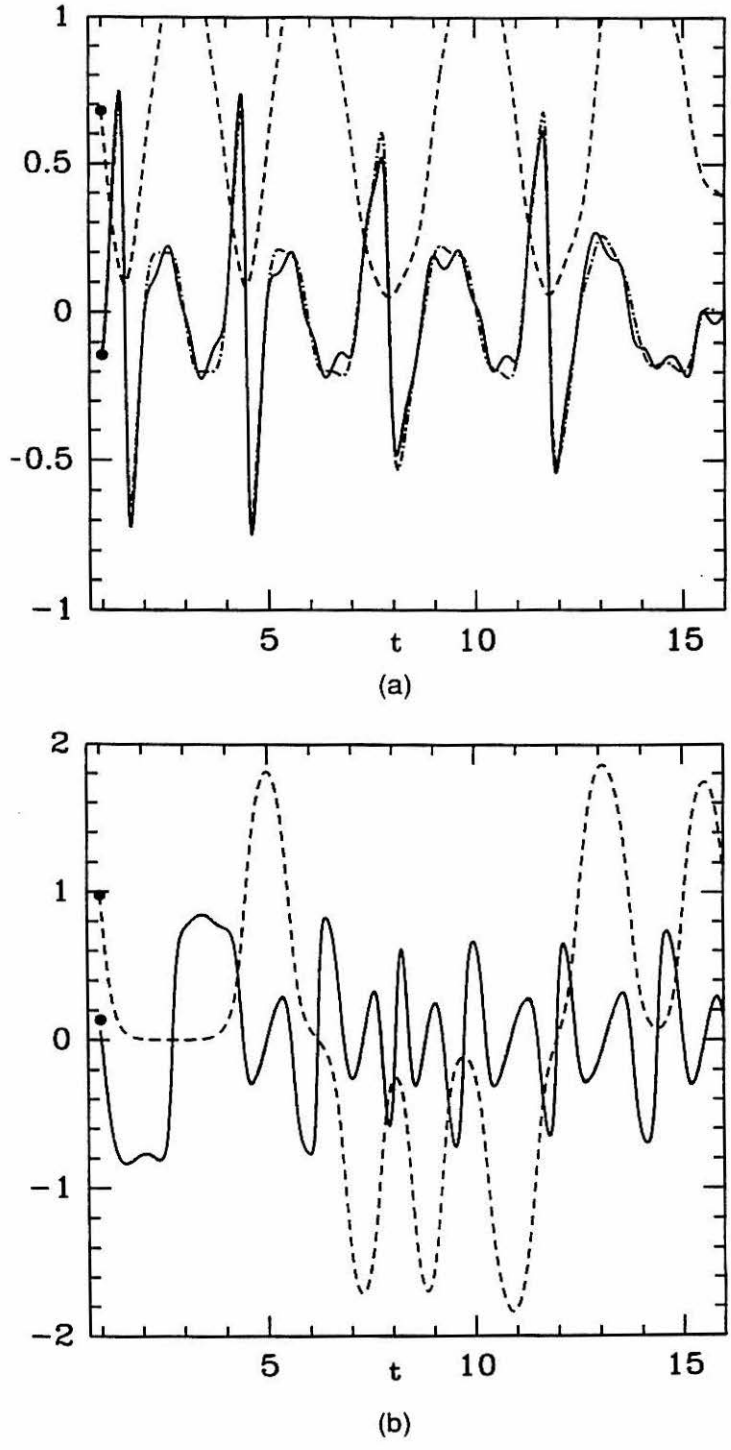


Figure 4.17 The expressions for the *reversible* strain rates defined by expressions (4.2.13) (dashed-dotted line) and (4.2.14) (solid line). In (b) the two lines are essentially on top of one another.

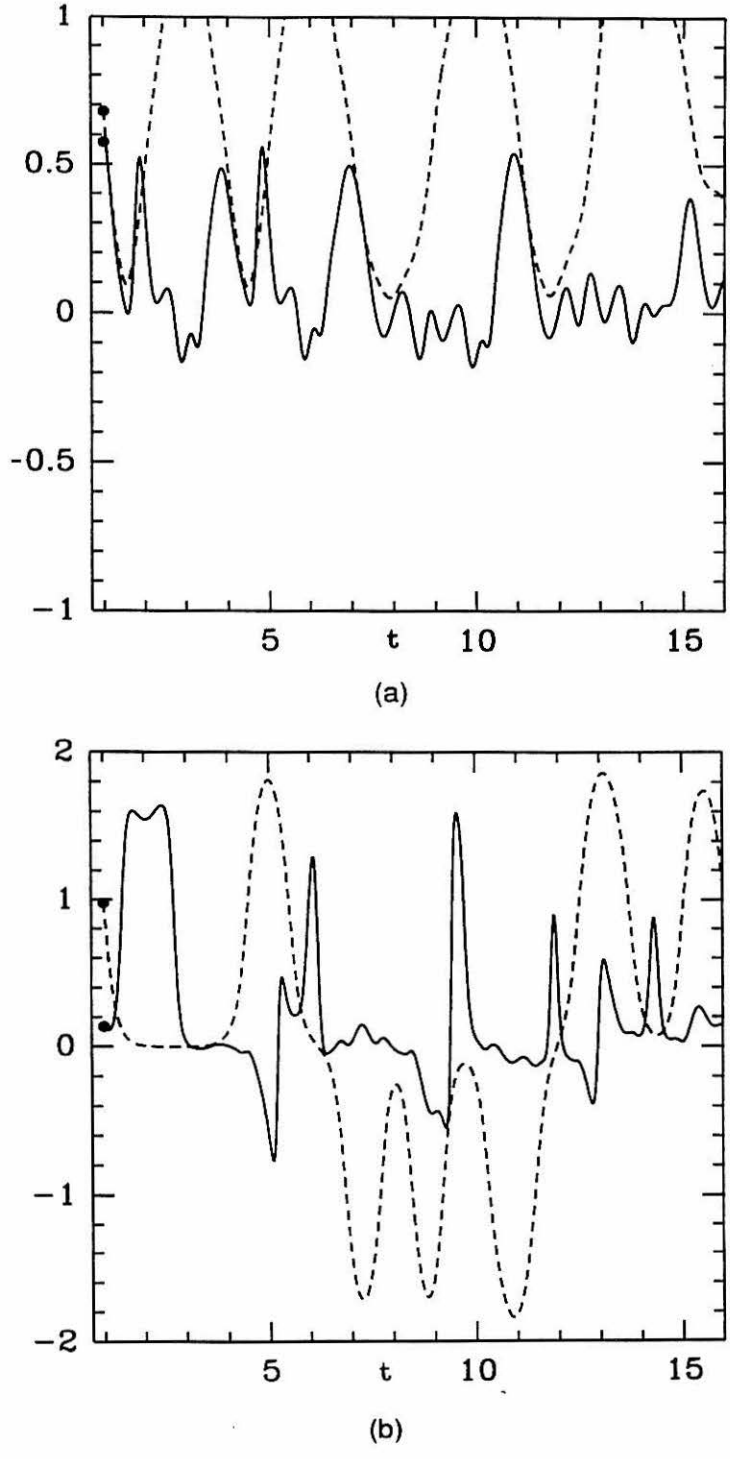


Figure 4.18 Irreversible strain rate plots corresponding to Figure 4.16 for the (a) open flow and (b) closed flow.

Chapter 5

The dynamics associated with a quasiperiodically forced Morse oscillator: application to molecular dissociation

Based upon an article that appears in:

Physical Review A **45**, 4803-4827 (1992).

Abstract

The dynamics associated with a quasiperiodically forced Morse oscillator is studied as a classical model for molecular dissociation under external quasiperiodic electromagnetic forcing. The forcing entails destruction of phase space barriers, allowing escape from bounded to unbounded motion. In contrast to the ubiquitous Poincaré map reduction of a periodically forced system, we derive a sequence of nonautonomous maps from the quasiperiodically forced system. We obtain a global picture of the dynamics, *i.e.*, of *transport in phase space*, using a sequence of time-dependent two-dimensional lobe structures derived from the invariant homoclinic tangle of a persisting invariant saddle-type torus in a Poincaré section of an associated autonomous system phase space. Transport is specified in terms of two-dimensional lobes mapping from one to another within the sequence of lobe structures, and this provides the framework for studying basic features of molecular dissociation in the context of classical phase space trajectories. We obtain a precise criterion for discerning between bounded and unbounded motion in the context of the *forced* problem. We identify and measure analytically *flux* associated with the transition between bounded and unbounded motion, and study dissociation rates for a variety of initial phase space ensembles, such as an even or weighted distribution of points in phase space, or a distribution on a particular level set of the unperturbed Hamiltonian (corresponding to a quantum state). A *double phase slice* sampling method allows exact numerical computation of dissociation rates. We compare single-frequency and two-frequency forcing. Average flux is maximal in a particular single-frequency limit; however, lobe penetration of the level sets of the unperturbed Hamiltonian can be maximal in the two-frequency case. The variation of lobe areas in the two-frequency problem gives one added freedom to enhance or diminish aspects of phase space transport on finite time scales for a fixed average flux, and for both types of forcing the geometry of lobes is relevant. The chaotic nature of the dynamics is understood in terms of a *traveling horseshoe map sequence.*

5.1 Introduction

Periodically forced nonlinear oscillators and coupled nonlinear oscillators have been studied as models for a variety of atomic and molecular phenomena, such as multiphoton ionization of excited atoms,[1–7] multiphoton dissociation of diatomic and polyatomic molecules,[8–14] and unimolecular and bimolecular reactions.[15–21] In this chapter we consider a quasiperiodically forced Morse oscillator as a classical model of molecular dissociation under external quasiperiodic electromagnetic forcing. The unforced Morse oscillator is a one degree-of-freedom Hamiltonian system whose unperturbed phase portrait contains a separatrix that divides bounded and unbounded motion, and the use of dynamical systems theory in the ubiquituous case of *periodically* forcing such a system is by now well established, indeed commonplace, being employed in a variety of physical problems.[7,22–27] The immense popularity of, and indeed almost exclusive focus on, this type of forcing is related to the fact that, by time-periodic sampling of phase space trajectories, the study reduces to that of a two-dimensional Poincaré map, so that one is fully armed with all the tools and paradigms from dynamical systems theory associated with *two-dimensional maps* (to be described momentarily). It is in this context that, without resorting to statistics, one can answer basic questions about molecular dissociation in the context of phase space trajectories, such as:

(i) Is there a precise criterion for discerning between bounded and unbounded motion in the context of the *forced, non-integrable* problem and, if so, is there a practical way to determine this criterion and study its properties?

(ii) Can one identify in a nonstatistical manner which points in phase space undergo the transition between bounded and unbounded motion, corresponding to molecular dissociation, and at what time these points make the transition?

(iii) Can one identify and measure a *flux* associated with the transition between bounded and unbounded motion? More ambitiously, is there a way to quantify dissociation rates for a variety of possible initial phase space ensembles, such as an even or weighted distribution of points in phase space, or a distribution on a particular level set of the unperturbed Hamiltonian (corresponding to a quantum state)? Can one control these rates in any way?

(iv) How does one establish the existence of a chaotic response to the forcing and appreciate its significance?

Our goal is to address these questions in the context of forcing with more complicated time dependences than periodic, focusing for concreteness on quasiperiodic (two-frequency) forcing. There is obvious motivation to relax the restriction to the special case of periodic forcing, for it allows one to study more complicated scenarios: for example, one can study forcing by two lasers at different frequencies and compare with forcing by one. The basic difficulty in the extension past periodic forcing is that, for any choice of discrete time sampling of phase space trajectories, the study does not in general reduce to a two-dimensional map, but to a *sequence* of two-dimensional maps. Hence, all the conceptual machinery and mathematical tools associated with iterates of a single two-dimensional map, such as horseshoe maps, KAM (Kolmogorov-Arnold-Moser) tori, fixed partial barriers, turnstiles and so forth, do not apply here. In extending the analysis to sequences of maps, we will in addition to questions (i)-(iv) want to ask:

(v) Is there a way to compare the relative effects of each frequency on questions (i)-(iv)?

(vi) Is there a way to compare flux and dissociation rates between single-frequency and multiple-frequency systems, and say which is "better" or "worse"?

(vii) Do more complicated forcing time dependences lead to new features in molecular dissociation, and can one exploit these features?

(viii) What is "chaos" in the context of a *sequence* of maps, where the horseshoe map paradigm does not apply, and how does multiple-frequency forcing affect chaos?

Before outlining how we will proceed to answer these questions for quasiperiodic forcing, let us first briefly summarize the approach associated with periodic forcing. Periodic forcing destroys the separatrix, which acted as a *complete barrier* between bounded and unbounded motion in the unforced case, and this destruction entails a mechanism for *escape* from bounded to unbounded motion, corresponding to molecular dissociation. Dynamical systems theory offers a conceptual framework for studying dissociation and answering questions (i) to (iv) by establishing *phase space structure* and using this structure to study *phase space transport* (what we mean by this and the brief description that follows should be made quite apparent in the systematic study of quasiperiodic forcing to follow). This structure is provided by a set of invariant manifolds of the associated Poincaré map (these manifolds are global stable and unstable manifolds of a persisting saddle-type fixed point of the Poincaré map). One can use these manifolds to precisely define *partial barriers* in phase space between regions of bounded and unbounded motion, and to identify the *turnstile lobes* that are the gateways for transport from one region to another. These barriers and turnstiles are *fixed* in the Poincaré section. Having identified the turnstiles, one can then study flux from bounded to unbounded motion, and dissociation rates. Flux is determined by measuring turnstile lobe area, which can be done by an approximate analytical method using Melnikov theory, or by exact numerical computation of the turnstile lobe boundaries. For dissociation rates associated with an initial even distribution of points in phase space, a popular approximation in chemical kinetics

studies involves a Markov model approach,[15,16,21,28,29] but the deficiencies of such an approximation will be discussed later. An exact method for calculating dissociation rates, for a variety of initial distributions, involves studying the topology of intersections of preimages of the turnstile lobes with the appropriate geometrical objects (depending on the problem at hand), such as the turnstiles themselves, or the level sets of the unperturbed Hamiltonian. Lastly, Melnikov theory allows one to study when the oscillator will respond chaotically to the forcing.

For extension to quasiperiodic forcing, the basic question we need to address, before addressing the other eight questions is: *since a sequence of maps is needed to deal with more complicated forcing time dependences, what is the phase space structure associated with a sequence of maps, and how is this structure used to study phase space transport in such a way as to quantify molecular dissociation?* Our approach to answering this question will be to first rewrite the forced system in autonomous form in order to obtain a set of invariant manifolds in a Poincaré section of the enlarged phase space (these manifolds are the global stable and unstable manifolds of a persisting invariant saddle-type 1-*torus* of the Poincaré map). The geometric structure provided in this autonomous setting is then related back to the original nonautonomous sequence of maps: one understands the sequence of maps in terms of the Poincaré map acting on a sequence of two-dimensional *slices* of the higher-dimensional Poincaré section. The intersection of the time-dependent slices with the invariant structures in the autonomous setting defines a *sequence of time-dependent structures* for the sequence of maps. These structures then generate a sequence of time-dependent partial barriers between bounded and unbounded motion, and a sequence of time-dependent turnstile lobes. The basic new feature of the analysis is thus the variation with sample time of the relevant geometrical structure: to understand bounded and unbounded motion under quasiperiodic forcing, one *must* deal with time-dependent con-

structs, and the fact that these contructs are derived from an invariant structure embedded in an enlarged phase space allows one to embrace the more complicated transport issues associated with sequences of maps.

The chapter is organized as follows. In Section 5.2 we define the sequence of maps associated with the quasiperiodically forced Morse oscillator, and then study the geometrical structure associated with the oscillator recast in the autonomous form. The geometrical possibilites are far richer than for periodic forcing, and in Section 5.3 we introduce and calculate the quasiperiodic Melnikov function and use the function to study the geometry. Intimately related to the Melnikov function are the frequency dependent *relative scaling factors*, which provide a measure of the relative effect of each forcing frequency on the geometry. In Section 5.4 we derive the *sequence* of two-dimensional time-dependent phase space structures from the invariant one in the enlarged phase space, and use this to define the sequence of partial barriers between bounded and unbounded motion and the sequence of turnstiles as the gateway for transport between these regions. Section 5.5, the bulk of the chapter, focuses on quantifying transport in phase space in order to study molecular dissociation. We first describe and use a *double phase slice* sampling method to numerically compute the partial barriers and the turnstile boundaries for an arbitrary time sample, which provides the framework for *exact* computation of transport quantities under the sequence of maps. We then describe and quantify *flux* associated with *escape* to unbounded motion (*i.e.*, molecular dissociation) and *capture* into bounded motion: relevant measures include instantaneous flux, finite-time average flux and infinite-time average flux associated with both escape and capture, and the quasiperiodic Melnikov function affords an approximate analytical expression for all these measures, good for small forcing. Note that in contrast to the case of periodic forcing, these measures of flux are in general different, with the sole exception that infinite-time average flux associated with escape and capture are identical. It is found that

the infinite-time average of the flux *is maximal in the single-frequency limit associated with the larger relative scaling factor* (*i.e.*, in this context single-frequency forcing is the best one can do). Though flux is a popular concept in molecular dissociation problems, we assert its limitations and consider several other transport features that help describe dissociation. In particular, we address the question of dissociation probability for an even or weighted initial distribution of points in phase space, which has relevance to the dissociation of an ensemble of excited diatomic molecules. As mentioned before, a popular approach to this dissociation problem in the context of maps employs a Markov model; we extend such an approach to sequences of maps, and then stress the deficiencies of such an approach, and the need to consider the geometry associated with the intersections of preimages of the turnstile lobes with the turnstiles. This lobe intersection analysis, which provides the basis for *exact* computation of the dissociation rates, extends from maps to sequences of maps in a straightforward manner. The main new feature associated with sequences of maps is the *variation of lobe areas*: this variation gives one added freedom to alter finite-time transport quantities for a fixed infinite-time average flux. As an additional dissociation problem, we discuss penetration of the unperturbed energy levels by the turnstiles and their preimages (which has relevance to forcing a system prepared in a particular quantum state). In contrast to the flux result, it is possible for lobe penetration to be maximal in the two-frequency case. Lastly we close with a brief discussion of the nature of chaos in these systems: the horseshoe map paradigm generalizes to a *traveling horseshoe map sequence.* We point out that Noid and Stine,[9] and more recently Goggin and Milonni,[13] have also studied the quasiperiodically forced Morse oscillator; their approach differs from ours in that it is primarily a numerical study of absorption and the dissociation *threshold* for points initialized on the level set of the unperturbed Hamiltonian that corresponds to the *ground state*, and does not address the global picture of phase space

structure and transport afforded by the global stable and unstable manifolds. There have also been experimental studies of quasiperiodic electromagnetic forcing, such as Moorman et al.[30] on ionization of highly excited hydrogen atoms under a bichromatic microwave field.

We should stress at the outset the intent of this chapter, for the scope of recent investigations of dissociation problems is quite vast and not without controversy, covering classical and/or quantum mechanical studies (with or without comparisons) of regimes ranging from low intensity microwave ionization of highly excited states to high intensity infrared dissociation of ground states, along with chemical kinetics problems that cover a range of situations. Our interest here is in a classical study alone (*i.e.*, the study of ensembles of classical phase space trajectories), the reasoning being that before performing the difficult and somewhat controversial task of comparing classical and quantum mechanical treatments of a classically chaotic system, it would be helpful to have a firm grounding in the classical extension from maps to sequences of maps.

5.2 The quasiperiodically forced Morse oscillator and its phase space structure

We consider a quasiperiodically forced Morse oscillator:

$$\begin{aligned}\dot{x} &= \frac{p}{m} \\ \dot{p} &= -2D_0 a(e^{-ax} - e^{-2ax}) + \varepsilon d(E_1 \cos(\omega_1 t + \theta_{1_0}) + E_2 \cos(\omega_2 t + \theta_{2_0})),\end{aligned} \tag{5.2.1}$$

$(x, p) \in \mathbb{R}^2$. One can think of $x \equiv r - r_e$ as the separation, r, of a two-atom molecule minus an equilibrium separation, r_e, with p the relative momentum. The system then corresponds to a non-rotating pair of atoms interacting under a Morse potential and forced by an external two-frequency electromagnetic field with amplitudes εE_1 and εE_2. The parameters a and D_0 correspond

to the range parameter and *unperturbed* dissociation energy, respectively, of the Morse potential, and d is the effective charge, or dipole gradient. The initial phases associated with the forcing are given by θ_{1_0}, θ_{2_0}. For concreteness, one can think of, say, an HF molecule, and hence set: $m = 0.9571\text{amu}$, $D_0 = 6.125eV$, $a = 1.1741 r_B^{-1}$ ($r_B \equiv$ Bohr radius), and $d = 0.7876 D_0 r_B^{-1}$. Note that, though for concreteness we consider dissociation of a diatomic molecule, a forced Morse oscillator can model other molecular phenomena, such as pumping a local mode of a polyatomic molecule by an infrared laser.[10] Further note that the rotating diatomic problem can be studied in the context of the k degree-of-freedom transport theory of Wiggins.[31]

To study the dynamics of (5.2.1) it is advantageous to sample phase space trajectories at discrete time intervals, the interval being one of the forcing periods, say $\frac{2\pi}{\omega_2}$. For periodic forcing it is well understood that periodic sampling of trajectories leads to the study of a Poincaré map, which simplifies the underlying geometrical structure with which to study the motion; for quasiperiodic forcing, periodic sampling will in a similar way simplify the underlying structure to be studied. However, because of the lack of periodicity in the vector field time dependence, in this case we will be lead to the study of a bi-infinite sequence of maps. The evolution of system (5.2.1) from time $t = \frac{2\pi}{\omega_2} n$ to time $t = \frac{2\pi}{\omega_2}(n+1)$ defines a map on the plane:

$$T_\varepsilon(\,\cdot\,;n) : \mathbb{R}^2 \to \mathbb{R}^2$$

$$\Big(x\Big(\frac{2\pi}{\omega_2}n\Big), p\Big(\frac{2\pi}{\omega_2}n\Big)\Big) \to \Big(x\Big(\frac{2\pi}{\omega_2}(n+1)\Big), p\Big(\frac{2\pi}{\omega_2}(n+1)\Big)\Big). \tag{5.2.2}$$

The goal then is to study the dynamics under the bi-infinite sequence of maps $\{T_\varepsilon(\,\cdot\,;n); n \in Z\}$. To do this, we rewrite the Morse oscillator in autonomous form in order to construct invariant manifolds that are fixed in the resulting enlarged phase space. These manifolds form barriers which constrain the motion in the autonomous system phase space and hence form *structure* that governs motion in phase space. Having obtained these manifolds, we will use

them to define for each time $t = \frac{2\pi}{\omega_2} n$ a set of curves in (x, p) space that will be used later to study the dynamics under the sequence of maps; in particular, to define the sequence of partial barriers between bounded and unbounded motion and the sequence of turnstiles.

The Morse oscillator in autonomous form is given by:

$$\begin{aligned} \dot{x} &= \frac{p}{m} \\ \dot{p} &= -2D_0 a(e^{-ax} - e^{-2ax}) + \varepsilon d(E_1 \cos\theta_1 + E_2 \cos\theta_2) \\ \dot{\theta}_1 &= \omega_1 \\ \dot{\theta}_2 &= \omega_2, \end{aligned} \tag{5.2.3}$$

where $(\theta_1, \theta_2) \in T^2$, the 2-torus. We then define a Poincaré section:

$$\Sigma^{\theta_{2_0}} = \{(x, p, \theta_1, \theta_2) \,|\, \theta_2 = \theta_{2_0}\},$$

and the associated Poincaré map generated by the flow of (5.2.3) is given by:

$$P_\varepsilon : \Sigma^{\theta_{2_0}} \rightarrow \Sigma^{\theta_{2_0}} \tag{5.2.4}$$

$$(x(0), p(0), \theta_{1_0}) \rightarrow \Big(x\Big(\frac{2\pi}{\omega_2}\Big), p\Big(\frac{2\pi}{\omega_2}\Big), \theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}\Big).$$

Studying the flow of (5.2.3) via this three-dimensional Poincaré map is equivalent to sampling the trajectories of (5.2.3) at time intervals equal to $2\pi/\omega_2$.

The phase portrait of the unperturbed Poincaré map $P_{\varepsilon=0}$ is portrayed in Figure 5.2.1. There is a neutrally stable 1-torus at $\{(x, p, \theta_1) | x = p = 0\}$ and a non-hyperbolic invariant 1-torus (of saddle-type stability) of the form

$$\tau_{\varepsilon=0} = \{(x, p, \theta_1) \,|\, \lim x \rightarrow \infty,\ p = 0\}. \tag{5.2.5}$$

The global stable and unstable manifolds of $\tau_{\varepsilon=0}$, denoted $W^s(\tau_{\varepsilon=0})$ and $W^u(\tau_{\varepsilon=0})$ respectively, coincide to form a two-dimensional separatrix which *separates bounded and unbounded motion*, and which asymptotes with increasing x to $p = 0$ (see the Appendix for the analytical expression of the separatrix). Any point initialized inside the separatrix evolves on a 2-torus and

corresponds to a molecule which does not dissociate. Any point initialized outside the separatrix evolves on an unbounded two-dimensional surface, and corresponds to a molecule that is dissociated, asymptoting to infinite separation. As shown in Figure 5.2.1, the separatrix is parametrized by (s, θ_1), where s is a time-like coordinate, described in the caption ($s = 0$ is chosen to correspond to the point of symmetry $(x = -\frac{\ln(2)}{a}, p = 0)$). In Sections 5.4 and 5.5 we will be particularly interested in *phase slices* of the Poincaré section $\Sigma^{\theta_{2_0}}$, defined as $\chi(\bar{\theta}_1) \equiv \{(x, p, \theta_1) \mid \theta_1 = \bar{\theta}_1\}$, and Figure 5.2.2 shows the phase portrait in an arbitrary phase slice (for the unforced problem, the phase plane decouples from θ_1). The area enclosed by the separatrix in each phase slice is equal to $4\pi D_0/\omega_0$, where $\omega_0 \equiv a(\frac{2D_0}{m})^{\frac{1}{2}}$ is the frequency associated with simple harmonic motion near the bottom of the Morse well (this is in contrast to the study of MacKay and Meiss,[7] who have an unbounded area enclosed by the separatrix for a Coulomb potential).

Quasiperiodic forcing causes the separatrix to undergo a *global bifurcation*, leading to non-integrable motion near the unperturbed separatrix, and providing a mechanism for *escape* from bounded to unbounded motion, and *capture* from unbounded to bounded motion. The study of this bifurcation is complicated by the fact that the invariant saddle-type 1-torus is non-hyperbolic with infinite radius, and the relevant theorems, such as persistence of the invariant 1-torus, have been proven for normally hyperbolic tori with finite radii. For the Morse oscillator (5.2.3) we prove that the invariant 1-torus persists under forcing and obtain an expression for the 1-torus (see the Appendix). Given for the perturbed problem the existence of global differentiable stable and unstable manifolds of the surviving invariant 1-torus, the remaining theorems, such as those necessary for the generalized Melnikov construction, easily extend (see the Appendix). To avoid a lengthy digression to a rigorous mathematical discussion, we assume the existence of global differentiable stable and unstable manifolds, denoted $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ respectively, and present later

some simple numerical computations that confirm the validity of the assumption. For $\varepsilon > 0$, $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ are no longer identical, and Figure 5.2.3 shows a case where the manifolds intesect in an infinity of 1-tori to produce the boundary of a complicated geometrical construct referred to as a *homoclinic tangle*. The manifolds criss-cross each other ad infinitum to form the boundary of three-dimensional regions referred to as *lobes*, which shall be discussed more carefully as we go along. It is the homoclinic tangle and its lobes that form the phase space structure needed to study molecular dissociation, and so as a first step we need to understand the geometry of this structure. For two-dimensional time-periodic systems (whose associated Poincaré map is two-dimensional), the geometrical possibilities are simple: the manifolds are one-dimensional and either they never intersect, or they intersect in an infinity of points to form two-dimensional lobes. For more complicated time dependences and higher dimensions, the geometrical possibilities are far richer, but a generalized Melnikov theory provides an analytical tool for studying these geometries.

5.3 Generalized Melnikov theory and the geometry of the homoclinic tangle

The generalized Melnikov function[32] is defined to be

$$M(s,\theta_1,\theta_2;\nu) = \int_{-\infty}^{\infty} DH(x_h(t),p_h(t)) \cdot g(x_h(t),p_h(t), \omega_1 t + (\omega_1 s + \theta_1), \omega_2 t + (\omega_2 s + \theta_2); \mu) dt, \tag{5.3.1}$$

where

$$H(x,p) = \frac{p^2}{2m} + D_0(1 - e^{-ax})^2$$

is the Hamiltonian of the unforced Morse oscillator, D denotes the derivative operator $(\frac{\partial}{\partial x}, \frac{\partial}{\partial p})$ acting on $H(x,p)$,

$$\varepsilon g(x,p,\theta_1,\theta_2;\mu) = (0, \varepsilon d(E_1 \cos\theta_1 + E_2 \cos\theta_2))$$

is the perturbation, (x_h, p_h) are the (x,p) coordinates of the unperturbed homoclinic separatrix, μ represents the perturbation parameters d, E_1, and E_2, and ν represents both the perturbation parameters and the other system parameters, m, D_0, and a. Expression (5.3.1) thus involves a time-like integral, along the *unperturbed* homoclinic separatrix, of a dot product between the perturbing vector field and the gradient of the *unperturbed* Hamiltonian (see Guckenheimer and Holmes[22] for an introduction to Melnikov theory and Wiggins[32] for an extensive study of generalized Melnikov theory). The main use of the Melnikov function is its provision of an $0(\varepsilon)$ estimate of the separation $d(s, \theta_1, \theta_{2_0}; \nu, \varepsilon)$, normal to the unperturbed separatrix, of the perturbed stable and unstable manifolds in $\Sigma^{\theta_{2_0}}$[32]:

$$d(s, \theta_1, \theta_{2_0}; \nu, \varepsilon) = \frac{\varepsilon M(s, \theta_1, \theta_{2_0}; \nu)}{||DH(x_h(-s), p_h(-s))||} + O(\varepsilon^2) \qquad (5.3.2)$$

(see the Appendix and Figure 5.3.1). This analytical estimate of manifold separation is a powerful tool for studying many features of the homoclinic tangle, such as its geometry and, later on, the flux between bounded and unbounded motion. It is a straightforward consequence of the implicit function theorem[32] that within $O(\varepsilon)$ of a zero of the Melnikov function (with $D_{(s,\theta_1)}M$ of rank one at the zero) is a transversal intersection of the stable and unstable manifolds, and we refer to this class of intersections as *primary intersection manifolds (PIM's)*, which play a basic role in understanding tangle geometry. Hence the zero sets of the Melnikov function allow one to study the approximate location and the *exact* geometry of the PIM's.

For the quasiperiodically forced Morse oscillator (5.2.3) the Melnikov function is

$$M(s, \theta_1, \theta_2; \nu) = -\frac{2\pi}{a}\{E_1 d e^{-\omega_1/\omega_0}\sin(\omega_1 s + \theta_1) + E_2 d e^{-\omega_2/\omega_0}\sin(\omega_2 s + \theta_2)\} \qquad (5.3.3)$$

(see the Appendix). Note that (5.3.3) is valid for $\omega_i \neq 0$, $i = 1, 2$; as explained in the Appendix, for either $\omega_1 = 0$ or $\omega_2 = 0$ (giving a constant

term in the perturbation), the corresponding term in the Melnikov function vanishes. The Melnikov function is thus discontinuous at $\omega_i = 0$, since the Melnikov amplitudes limit to a non-zero value as $\omega_i \to 0$, and we will discuss the physical implications of this discontinuity momentarily. Figure 5.3.2 shows comparisons between manifold separation computed by the Melnikov approximation (5.3.2) and by numerical computation of the manifolds using a *double phase slice* sampling method, which we describe later in the chapter. The trigonometric dependence of the perturbation on $(t, \theta_{1_0}, \theta_{2_0})$ in (5.2.1) carries through to a similar trigonometric dependence of M on (s, θ_1, θ_2) in (5.3.3) (the cosines go to sines). We refer to the absolute value of the ratio of the Melnikov amplitude in (5.3.3), $A_i \equiv -\frac{2\pi}{a} E_i d e^{-\omega_i/\omega_0}$, $i = 1, 2$, to the corresponding relative forcing amplitude $E_i d$ in (5.2.1) or (5.2.3) as a *relative scaling factor*:

$$\begin{aligned} \textit{relative scaling factor} &\equiv \left| \frac{\textit{Melnikov amplitude for } \omega_i}{\textit{relative forcing amplitude for } \omega_i} \right| \\ &= \frac{2\pi}{a} e^{-\omega_i/\omega_0}. \end{aligned} \tag{5.3.4}$$

The dependence of these factors on forcing frequency is described by a single *relative scaling function*

$$\mathcal{RSF}(\omega; a, \omega_0) = \frac{2\pi}{a} e^{-\omega/\omega_0}. \tag{5.3.5}$$

The relative scaling function provides an approximate ($O(\varepsilon)$) measure of the effectiveness of a forcing frequency in producing manifold separation, and, as such, is a basic tool in understanding how each frequency affects dissociation. Since the function's exponential decay depends only on ω_0, the period associated with simple harmonic motion at the bottom of the Morse well provides the relevant time scale for the forcing's effectiveness at producing manifold separation. Given any two forcing frequencies, one immediately knows the relative importance of each one; for example, if one of the frequencies is at ω_0 and the other at $4\omega_0$, and the amplitudes of the two forcing terms are

identical, then due to the exponential decay of the relative scaling function, the second term has a relatively negligible effect on manifold separation, and hence, as we shall later see, on certain transport properties, so that the problem is essentially one of single-frequency forcing. As another example, in the microwave limit $\omega_i << \omega_0$ the relative scaling factor associated with each frequency will be approximately $\frac{2\pi}{a}$, essentially frequency independent. We remark that, though the nonvanishing of the Melnikov amplitudes in the limit $\omega_i \to 0$ is not common (often they vanish in the limit of infinitely slow and infinitely fast forcing), this nonvanishing does occur in other systems, such as the forced Josephson junction.[24,25]

In Figure 5.3.3 we approximately portray (*i.e.*, to $O(\varepsilon)$) some PIM's in the Poincaré section $\Sigma^{\theta_{2_0}=0}$ by plotting the zero sets of $M(s, \theta_1, \theta_{2_0} = 0; \nu)$ for a particular frequency ratio $\omega_2/\omega_1 = g^{-1}$, where g is the golden mean $(\sqrt{5}-1)/2$, and a range of forcing amplitudes. Note that the Melnikov function, and hence the zero sets, are specified relative to the two-dimensional unperturbed separatrix, so that one interprets the plots in Figure 5.3.3 in terms of the separatrix, shown by the dashed line in Figure 5.3.1, being flattened out onto a plane. Hence, though the PIM's live in the three-dimensional Poincaré section, they are close to the unperturbed separatrix and for ease of portrayal we suppress the dimension normal to the unperturbed separatrix. Figures 5.3.3(a) to (g) show PIM's for a range of ratios of Melnikov amplitudes. For $|A_1/A_2| < 1$ the PIM's are non-intersecting 1-tori, for $|A_1/A_2| > 1$ they are non-intersecting spirals, and for $|A_1/A_2| = 1$ they are intersecting spirals (or equivalently intersecting 1-tori). For other ratios of ω_2/ω_1 (both commensurate and incommensurate), the PIM geometry as A_1/A_2 is varied is qualitatively similar. In Chapter 2 (equivalently refs. 33, 34) we provide a more detailed study of PIM geometry via theorems involving the Melnikov function. The basic result relevant to the Morse oscillator is the following. The manifolds $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ intersect if the Melnikov function (5.3.3)

has a zero with $D_{(s,\theta_1)}M$ of rank one. One can easily see that the Melnikov function (5.3.3) passes through zero a countable infinity of times for each $\theta_1 \in [0, 2\pi)$. Thus, regardless of PIM geometry, the set of PIM's intersects each phase slice $\chi(\bar{\theta}_1)$ of the Poincaré section $\Sigma^{\theta_{20}}$ in a countable infinity of points for $\bar{\theta}_1 \in [0, 2\pi)$.

With an appreciation of PIM geometry, it should not be difficult to visualize the nature and geometry of the previously mentioned three-dimensional lobes in $\Sigma^{\theta_{20}}$ by remembering that the separation of $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$, and hence the thickness of the lobes, is normal to the page for Figure 5.3.3 (see Figure 5.3.4). As one might expect from the figures, the geometries of the invariant lobes can take on a rich variety in quasiperiodic problems, and thus a precise definition of an invariant three-dimensional lobe in $\Sigma^{\theta_{20}}$ necessitates fairly careful development and a rather abstract description. We merely say here that for the case where PIM's are 1-tori, a *lobe* in $\Sigma^{\theta_{20}}$ is the region bounded by segments of $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ between adjacent PIM's; for other PIM geometries a definition similar in spirit applies and we refer the reader to Chapter 2.

5.4 The homoclinic tangle as a template for studying molecular dissociation

5.4.1 Using the invariant homoclinic tangle to study the sequence of maps

We have a set of two-dimensional invariant manifolds $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ fixed in the three-dimensional Poincaré section $\Sigma^{\theta_{20}}$ for the Morse oscillator in autonomous form (5.2.3), and the geometry of the manifolds is studied via the generalized Melnikov function. The trick is then to relate things back to the original physical problem, that of a sequence of two-dimensional maps on (x, p) space. This is done in a simple manner, as described heuristically here and

rigorously in Chapter 2. At each sample time $t = \frac{2\pi}{\omega_2}n$, the phase associated with the first forcing frequency is $\theta_1(n) = \theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n$; hence one can think of the original two-dimensional Morse oscillator (5.2.1) at this sample time as a *two-dimensional slice* of the three-dimensional Poincaré section $\Sigma^{\theta_{2_0}}$, defined by the *time slice*, or equivalently the *phase slice*

$$\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n) = \{(x, p, \theta_1)|\theta_1 = \theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n\}. \tag{5.4.1}$$

Hence the sequence of maps (5.2.2) on (x, p) space can be understood in terms of P_ε acting on a sequence of phase slices of $\Sigma^{\theta_{2_0}}$, as shown in Figure 5.4.1 (for incommensurate frequencies the phase slices will visit $\theta_1 \in [0, 2\pi)$ uniformly and densely, and for commensurate frequencies they will visit a finite number of θ_1 values). The intersection of each time slice with $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ defines *time-dependent* one-dimensional manifolds in (x, p) space at the appropriate sample time (the manifolds vary with the time slice since $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ vary with θ_1). Figure 5.4.1 illustrates this in the case where the invariant manifolds intersect in 1-tori, and the result for each time slice is a time-dependent two-dimensional homoclinic tangle with the same topological constraints as in the familiar case of periodic forcing: the two manifolds criss-cross ad infinitum, intersecting each other but never intersecting themselves, to define a countable infinity of two-dimensional *lobes*. More precisely, the set of PIM's intersect the time slice $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ in a countable infinity of *primary intersection points (PIP's)*, and the segments of $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ in that time slice between adjacent PIP's define the boundary of a two-dimensional lobe in that slice. The template for transport in (x, p) space under the sequence of maps, and hence the key to studying molecular dissociation, is thus the time-dependent two-dimensional lobe structure associated with each sample time. It is a straightforward consequence of the invariance of $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ that each lobe in the n^{th} time slice maps under $T_\varepsilon(\,\cdot\,; n)$ to a lobe in the $(n+1)^{th}$ time slice. *Hence the global picture for transport*

in (x,p) space under the sequence of maps $\{T_\varepsilon(\cdot;n)\,|\,n \in Z\}$ is expressed in terms of lobes mapping within a sequence of two-dimensional lobe structures derived from invariant manifolds embedded in a three-dimensional Poincaré section (as shown in Figure 5.4.2). We should stress that, as a consequence of invariance, for any invariant manifold geometry there is a two-dimensional lobe structure for *each* time slice in any particular sequence of time slices.

To describe transport in phase space necessitates specifying how the two-dimensional lobes map from one to another under the sequence of maps. At first this may seem a daunting task, since one needs to address a bi-infinite sequence of maps. However this task is greatly simplified by the fact that the $(\dot{x},\dot{p})$ equations in (5.2.3) depend periodically on θ_1 for fixed (x,p,θ_2), so we only need consider the finite interval $\theta_1 \in [0,2\pi)$; hence one can describe any map in the sequence of maps in terms of P_ε acting on $\chi(\bar{\theta}_1)$, for some $\bar{\theta}_1 \in [0,2\pi)$. The invariant manifolds are central to such a description, for once one knows how the three-dimensional lobes in $\Sigma^{\theta_{20}}$ evolve under the *single* map P_ε, one can easily deduce how the two-dimensional lobes evolve under any of the maps in the sequence of maps (see Chapter 2). Recognizing the importance of the invariant manifolds, we proceed to discuss the aspects of transport that are relevant to the study of molecular dissociation. The key issue is to use the sequence of lobe structures to *precisely* define boundaries between bounded and unbounded motion and then *precisely* identify the regions in phase space which cross these boundaries under each map.

5.4.2 Deriving from the homoclinic tangle the sequence of partial barriers between regions of bounded and unbounded motion and the sequence of turnstiles

By definition, points on $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ asymptote with positive and negative time, respectively, to the invariant 1-torus at infinity, τ_ε, so that in a

manner similar to the separatrix of the unforced system, $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ play the role of dividing surfaces between bounded and unbounded motion. For the unforced problem, where $W^s(\tau_{\varepsilon=0})$ and $W^u(\tau_{\varepsilon=0})$ are identical, the *entire* manifolds play the role of the dividing surface and they act as a *complete barrier*. For the forced problem, where $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$ are no longer identical, *segments* of the manifolds, joined together at an intersection set, play the role of the dividing surface between bounded and unbounded motion; it is possible for points to move across this surface, and hence the dividing surface acts as a *partial barrier*. We give a careful discussion of the construction of these dividing surfaces in Chapter 2, and the essentials of this construction applied to the Morse oscillator can be conveyed here in a simple manner with the help of some figures. One chooses the above-mentioned intersection set to be a one-dimensional subset of the set of PIM's that is piecewise continuous and intersects each phase slice $\chi(\bar{\theta}_1)$, $\bar{\theta}_1 \in [0, 2\pi)$, in a single point. Examples are illustrated in Figure 5.3.3, denoted by τ_c (there are of course other possible choices, and we will discuss our particular choice momentarily). If we let $S[\tau_\varepsilon, \tau_c]$ and $U[\tau_\varepsilon, \tau_c]$ denote the segments of $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$, respectively, from τ_ε to τ_c, then $\mathcal{C} \equiv U[\tau_\varepsilon, \tau_c] \cup S[\tau_\varepsilon, \tau_c]$ denotes a two-dimensional surface in $\Sigma^{\theta_{2_0}}$ that divides *each phase slice* $\chi(\bar{\theta}_1)$, $\bar{\theta}_1 \in [0, 2\pi)$, into two regions, as illustrated in Figure 5.4.3. For the case of toral PIM's, illustrated in Figure 5.4.3(a), one can choose τ_c to be a 1-torus, and $\mathcal{C}$ in fact divides $\Sigma^{\theta_{2_0}}$ into two regions; for the case of spiral PIM's, illustrated in Figure 5.4.3(b), one must choose τ_c to have a discontinuity, and hence $\mathcal{C}$ is discontinuous and does *not* divide $\Sigma^{\theta_{2_0}}$ into two regions, since there are gaps at the region of discontinuity. However, for all cases, $\mathcal{C}$ divides each *phase slice* $\chi(\bar{\theta}_1)$, $\bar{\theta}_1 \in [0, 2\pi)$, into two regions, which is all one is interested in for the sequence of maps. At each sample time $t = \frac{2\pi}{\omega_2} n$, we thus have in (x, p) space a time-dependent boundary $\mathcal{C}(n) \equiv \mathcal{C} \cap \chi(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n)$ that divides the space into two regions, denoted $R_b(n)$ and $R_u(n)$, as illustrated in Figure 5.4.3 (note that invariant

three-dimensional regions in $\Sigma^{\theta_{2_0}}$, R_u and R_b, are then defined as the union of the corresponding two-dimensional regions over the phase slices defined by $\bar{\theta}_1 \in [0, 2\pi)$). In the context of this time-dependent boundary we can now explain the choice of τ_c in Figure 5.3.3, and hence the choice of the sequence of boundaries. We chose τ_c such that for each $\bar{\theta}_1 \in [0, 2\pi)$ the s parameter associated with $\tau_c \cap \chi(\bar{\theta}_1)$ is the closest to zero out of all the PIM intersections with that phase slice. The aim of this choice is for *the boundary $\mathcal{C}(n)$ to most closely resemble the unperturbed separatrix at each time slice.*

As we shall explain, points can cross $\mathcal{C}$ *only* via the *turnstile lobes*, which we now define and identify. It is a straightforward consequence of the orientation preserving nature of P_ε that the lobes between τ_c and $P_\varepsilon^{-1}(\tau_c)$, which we refer to as the *turnstiles*, change their orientation *relative to* $\mathcal{C}$ under P_ε. What we mean by this can be easily visualized in Figure 5.4.4, where we show an example with toral PIM's and one pair of turnstile lobes. It should be clear from the figure that points in the turnstiles, and *only* these points, cross the invariant boundary $\mathcal{C}$ under P_ε, mapping from outside to inside or vice-versa. At each sample time $t = \frac{2\pi}{\omega_2} n$, we thus have in (x, p) space time-dependent turnstiles defined by

$$\begin{aligned} L_{bu}(1; n) &\equiv L_{bu}(1) \cap \chi(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n) \\ L_{ub}(1; n) &\equiv L_{ub}(1) \cap \chi(\theta_{1_0} + 2\pi \frac{\omega_1}{\omega_2} n), \end{aligned} \tag{5.4.2}$$

where $L_{bu}(1)$ and $L_{ub}(1)$ are the three-dimensional turnstile lobes in $\Sigma^{\theta_{2_0}}$ that map under one iterate of P_ε from R_b to R_u and from R_u to R_b, respectively. Hence the turnstile $L_{bu}(1; n)$ ($L_{ub}(1; n)$) is the set of lobes which map under $T_\varepsilon(\,\cdot\,; n)$ from *inside* (*outside*) $R_b(n)$ to *outside* (*inside*) $R_b(n+1)$, and we shall refer to these two processes as *escape* and *capture*, repectively. The points in $R_b(n)$ at time $t = \frac{2\pi}{\omega_2} n$ are destined to oscillate in a bounded fashion until at some future sample time $t = \frac{2\pi}{\omega_2} \tilde{n}$, $\tilde{n} > n$ (which may or may not ever occur), they *escape* under $T_\varepsilon(\,\cdot\,; \tilde{n})$ *via the turnstile lobes* $L_{bu}(1; \tilde{n})$ to $R_u(\tilde{n} + 1)$ and

henceforth asymptote to infinite separation (in a similar manner one describes capture). Hence, it is correct to interpret $R_b(n)$, $R_u(n)$ as the regions of bounded and unbounded motion, respectively, and the turnstile lobes $L_{bu}(1;n)$ and $L_{ub}(1;n)$ as the *sole* mechanism for transport between the bounded and unbounded regions.

We conclude this section with three remarks:

(i) Previous studies which do not address invariant manifolds are forced to use approximate and somewhat ad hoc criteria for dissociation. For example, in their study of ionization and dissociation in terms of a forced nonlinear oscillator, Leopold and Percival[1] (Coulomb potential, periodic forcing), and later Goggin and Milonni[12,13] (Morse potential, periodic and quasiperiodic forcing) use for their dissociation criterion a time-dependent "compensated energy": any point with energy greater than this compensated energy is considered dissociated. As we shall see with some simulations in Section 5.5.1, it is possible for their criterion to significantly differ from the true criterion offered by the invariant manifolds. Even if one is in a regime where such an approximation works well, it is still important to appreciate the true dissociation criterion.

(ii) We stress the new feature in the generalization from transport under *maps* to transport under *sequences of maps*: one is not concerned with transport with respect to a fixed region in the plane, as is the case with two-dimensional maps, but with transport with respect to a sequence of regions, which vary in shape and size from one time sample to the next. This is a consequence of the fact that the relevant dividing lines between bounded and unbounded motion, made up of points which asymptote to the invariant torus τ_ε as $x \to \infty$, vary from one time sample to the next, due to the more complicated time dependence in the forcing. That the time-dependent template with which to study transport, the sequence of

lobe structures, is derived from a set of invariant manifolds embedded in an enlarged phase space, allows one to embrace rather than avoid the more complicated transport issues associated with quasiperiodic forcing.

(iii) Though not explicitly stated, our discussion of turnstile lobes assumed that each lobe was entirely contained in one region or the other. This is the case for small enough forcing (which in practice can be quite large), and when it is not the case one can redefine the turnstiles in an appropriate manner to be the portions in one of the regions.[33,34,35]

5.4.3 A comment on boundedness and unboundedness in the quantum mechanical treatment

For a classical approach, it is the segments of the invariant manifolds of the *perturbed* system, and *not* the unperturbed separatrix, that are relevant to the criterion for dissociation. For example, in the forced system a point may oscillate in a bounded manner over some finite time interval such that it repeatedly crosses back and forth over the unperturbed separatrix. Bearing this in mind, consider the conventional approach used in the quantum mechanical treatment (as described, for example, in Goggin and Milonni[12,13] and Lu et al.,[14] or elementary quantum mechanics texts for that matter). Let $\{|b_i>\}$ denote the set of bound states of the *unforced* Morse oscillator. If the system is initially in some bound state $|\psi(t=0)>$ then the dissociation probability at time t for the forced problem is

$$P_d(t) = 1 - \sum_j |<b_j|\psi(t)>|^2. \tag{5.4.3}$$

Hence one finds the dissociation probability by subtracting from one the probability of being in one of the $\{|b_i>\}$ states. The obvious question that arises for this approach is: does one want to interpret $\{|b_i>\}$ as the set of bound states for the *forced* problem? For time-independent perturbations (Zeeman

effect, Stark effect, and so forth) one routinely solves for the perturbed eigenvalues and eigenstates; however, for the more difficult case of time-dependent perturbations it has been common to study transitions relative to the *unperturbed* basis states according to (5.4.3) and the question of the interpretation of bounded and unbounded motion in the time-dependent problem remains largely ignored.

5.5 Quantifying flux and dissociation rates

5.5.1 Computing the partial barriers between bounded and unbounded motion, the number of turnstile lobes, and the turnstile boundaries

A first step towards quantifying molecular dissociation via phase space transport is the numerical computation of the partial barriers and turnstiles, which provides a direct portrayal of phase space structure and allows exact computation of flux and various dissociation rates. The procedure for numerical computation of the invariant stable and unstable manifolds of saddle-type fixed points of maps at finite (x, p) values is well-known and straightforward; the added complications here are of course that we are dealing with sequences of maps and an invariant 1-torus with infinite radius. One can study sequences of maps by employing a *double phase slice sampling method* (refer back to Chapter 2), which is similar to the sampling method for maps, except that one takes into account that the evolution from one time sample to the next is understood in terms of phase slices of the Poincaré section $\Sigma^{\theta_{2_0}}$ mapping from one to another under P_ϵ. Hence to compute segments of the global stable and unstable manifolds in some finite window $x \in [x_a, x_b]$ at a given time sample $t = \frac{2\pi}{\omega_2} n$, *i.e.*, at a given phase slice $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2} n)$, one evolves under the dynamical system vector field backward and forward in time, respectively, over some time interval $|\Delta t| = \frac{2\pi}{\omega_2} j$, small segments of the

stable and unstable manifolds that originate at large x values in the *different* phase slices $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+j))$ and $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n-j))$, respectively. If τ_ε had a finite radius, one would take these initial segments to be the local stable and unstable manifolds of τ_ε in the appropriate phase slices. The p coordinate of τ_ε is known exactly in any phase slice (see equation (5.*A*.4) in the Appendix), so in practice one can take the initial segments to be horizontal segments at large x values (say $x > 20$) slightly above and below the p value of $\tau_\varepsilon \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+j))$ and $\tau_\varepsilon \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n-j))$ for $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$, respectively. The error incurred by this approximation can be made negligible by choosing for the line segment x sufficiently large and p sufficiently close to the p coordinate of τ_ε in the appropriate phase slice (for example, in the Melnikov calculations in Figure 5.3.2 we insure that the error due to the initial segment approximation is orders of magnitude smaller than the true manifold separation). In making this approximation, we stress the need to respect the variation of the p coordinate of τ_ε with the choice of Poincaré section and phase slice: in the context of periodic forcing, Bruhn[36] takes the initial segment to lie on the corresponding *unperturbed* manifolds, but for reasonably sized perturbations a line element that lies on $W^s(\tau_{\varepsilon=0})$ can lie *below* $W^u(\tau_\varepsilon)$, and hence move *opposite* to the intended direction.

With the above procedure one can calculate with arbitrary precision segments of the invariant manifolds for the n^{th} time sample, and from this immediately calculate the partial barriers and turnstile boundaries by an appropriate choice of $\tau_c \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ and $\tau_c \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+1))$, and then the calculation of $P_\varepsilon^{-1}(\tau_c \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+1)))$. Figure 5.5.1 shows a simple illustration with a sequence of four time slices. Note how the number, size, and shape of the turnstile lobes can change from one time sample to the next, which contrasts sharply with the time-periodic case, where the turnstiles are independent of sample time. By explicit computation of the manifolds as just described here, one can determine the number of turnstile

lobes for each time slice. However, the Melnikov function provides an analytical tool for determining turnstile lobe number, which will be relevant to some of the flux calculations to come. If $s = s_c(\theta_1)$ denotes the s values associated with the choice of τ_c, then the turnstile lobes at the n^{th} time slice lie between $s_c(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ and $\{s_c(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}(n+1)) + 2\pi/\omega_2\}$. Every time the Melnikov function crosses through zero it corresponds to a nearby crossing of the stable and unstable manifolds; hence, if $\mathcal{N}(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ denotes the number of zero crossings of $M(s, \theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n, \theta_{2_0}; \nu)$ between (and not counting) the above two endpoints, then the number of turnstile lobes in that time sample is (generically) given by $\mathcal{N}(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n) + 1$ (and by the sign of the Melnikov function one can determine how many of these lobes correspond to escape and how many to capture, as described in Chapter 2). Bearing this in mind, the plots of the Melnikov zero sets in Figure 5.3.3, along with the portrayal of τ_c and $P_\varepsilon^{-1}(\tau_c)$, allow one to immediately deduce the number of turnstile lobes in each phase slice. For example, in Figure 5.3.3(d) there is seen to be two turnstile lobes (one capture and one escape) for all time slices (except the isolated cases where the PIM's cross). In contrast, in Figure 5.5.1, which is the same example, except sampled now at the slower rather than faster frequency, there is a different number of turnstiles lobes (ranging from two to four) for each time slice. Chapter 2 gives a careful comparison of the effect of different sampling choices; for either choice, however, the same transport formalism goes through and we will on purpose consider both types of sampling interchangeably.

In Figure 5.5.2 we compare for a particular example the dissociation criterion based on numerical computation of segments of the invariant manifolds versus Leopold and Percival's "compensated energy". The compensated energy includes a time-dependent term that simply shifts the unperturbed separatrix up and down in p with time, and there is no accounting for the overall distortion and "breathing" of the true partial barriers.

5.5.2 Turnstile lobe area and flux

A central transport quantity in previous studies of molecular dissociation in the context of maps is the *flux* between regions of bounded and unbounded motion.[7,15–21,37–39] Though flux describes an important feature of transport to and from regions of bounded and unbounded motion, we shall later see its limitations. In contrast to the case of area-preserving maps, where a single measure of flux suffices, for area-preserving sequences of maps there are five relevant measures of flux. The first two are the *instantaneous flux*, denoted $\phi_e(n)$ and $\phi_c(n)$ for escape and capture, respectively:

$$\begin{aligned} \phi_e(n) &= \frac{\omega_2}{2\pi}\mu(L_{bu}(1;n)) \\ \phi_c(n) &= \frac{\omega_2}{2\pi}\mu(L_{ub}(1;n)), \end{aligned} \tag{5.5.1}$$

where $\mu(\,\cdot\,)$ denotes the area of the lobes within the parentheses. Instantaneous flux thus refers to the area in phase space, per sample time, that escapes or is captured between the n^{th} and $(n+1)^{th}$ time samples, and is in general different for each time sample and for escape versus capture, since the partial barriers vary with each time slice and the area of the region they enclose can change. The next two are the *finite-time average flux*, denoted $\Phi_e(n)$ and $\Phi_c(n)$ for escape and capture, respectively, which are simply the average of the instantaneous flux over the first $n+1$ time samples. This as well is different for each n and for escape versus capture, and though both quantities will converge to the same value in the limit $n \to \infty$, the convergence time can vary from being quite short to quite long. The final measure is the *infinite-time average flux*

$$\Phi = \lim_{N \to \infty} \frac{1}{2(N+1)} \sum_{n=0}^{N} \{\phi_e(n) + \phi_c(n)\}, \tag{5.5.2}$$

which is the same for escape and capture, since the area of $R_b(n)$ remains bounded, as should be plain from the definition of $R_b(n)$. One can exactly

compute the above fluxes by identifying the turnstiles and computing their boundaries for each time slice, as described in the previous section.

One can approximately (to $O(\varepsilon)$) compute the fluxes using the Melnikov function. Integrating the distance expression over the arc length of the unperturbed separatrix $d\lambda = ||DH(x_h(-s), p_h(-s))||ds$ gives the area of a turnstile lobe in the n^{th} time slice $\chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$:

$$\mu(\text{turnstile lobe}(n)) = \varepsilon \int_{s_a(n)}^{s_b(n)} |M(s, \theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n, \theta_{2_0}; \nu)|ds + O(\varepsilon^2), \quad (5.5.3)$$

where $s_a(n) < s_b(n)$ are the s values of the bounding PIP's of the lobe. Using (5.5.3), one can calculate instantaneous flux and finite-time average flux by determining the areas of the appropriate turnstile lobes (see Figures 5.5.3 and 5.5.4). For the infinite-time average, one obtains a particularly simple expression:

$$\Phi = \lim_{T \to \infty} \frac{\varepsilon}{2T} \int_0^T |M(s, \theta_{1_0}, \theta_{2_0}; \nu)|ds + O(\varepsilon^2). \quad (5.5.4)$$

The expression for Φ really involves a sum over phase slices of the turnstile lobe areas, for which the $0(\varepsilon)$ term of (5.5.3) is a good approximation for ε not too large and ω_1, ω_2 not too small, but this is converted to an integral over a *single* phase slice by elementary periodicity properties of the Melnikov function arguments (refer back to Chapter 2). Note that, for a given sampling phase θ_{2_0}, for commensurate frequencies the average flux depends on the initial phase θ_{1_0}, *i.e.*, $\Phi = \Phi(\theta_{1_0})$, but for incommensurate frequencies it is independent of θ_{1_0}.

One would like to compare the flux associated with periodic and quasiperiodic forcing. A normalization criterion for the forcing amplitudes is needed to decide upon "equivalent" periodic and quasiperiodic perturbations that can be compared. For example, let us choose the criterion $E_1^2 + E_2^2 = \textit{constant} \equiv E^2$ (*i.e.*, as E_1 and E_2 are varied, we keep constant the sum of the intensities associated with each amplitude in the electromagnetic field). For fixed ω_1, ω_2,

we then vary, say, E_1 from 0 to E, with E_2 determined by the normalization criterion. The question then is: how does the infinite-time average flux vary as a function of E_1? The variation depends on two properties: interference effects and relative scaling factors. Referring to Figure 5.5.5, at the single-frequency limits, associated with the endpoints $E_1 = 0$ and $E_1 = E$, the average flux Φ is proportional to the relative scaling factor associated with the corresponding frequency. Interference effects cause the flux profile to dip below a linear increase from the lower to higher Φ. The single-frequency limit associated with the larger relative scaling factor thus corresponds to an absolute maximum of the average flux (the absolute minimum may or may not be at the other single-frequency limit, depending on the difference between the two relative scaling factors and the size of the dip due to interference effects). In this context, the best one can do is single-frequency forcing. In other contexts, however, including other aspects of flux, multiple-frequency forcing can enhance certain transport quantities. In particular, the variation of lobe areas gives one the freedom to vary instantaneous and finite-time average flux for a fixed infinite-time average. Relevant to the variation of finite-time average flux is the time scale for convergence to the infinite-time average. This time scale can be anywhere from relatively short to arbitrarily long (refer back to Figure 5.5.4). We wish to stress the importance of the finite-time quantities, for, as we shall see, dissociation rates of course decay in time, so the dissociation process occurs over a finite time interval, and the infinite-time average flux may or may not have strong relevance to the problem at hand.

Just as the relative scaling function helps one understand the multiple-frequency case, so too does the flux associated with single-frequency forcing (see Figure 5.5.6). The single-frequency infinite-time average flux given by (5.5.4) increases monotonically with decreasing forcing frequency, ω, corresponding to infinite turnstile lobe area and maximal flux in the limit $\omega \to 0$. Such a result would seem to deserve an explanation. In comparison, MacKay

and Meiss[7] obtain for a Coulomb potential a turnstile lobe area and flux which *both* diverge as $\omega \to 0$ for fixed ε. In their calculations they (along with Casati et al.[6]) consider dissociation of a prepared highly excited state ($n_0 = 66$), and they choose both ε and ω to depend on n_0; if one maintained this dependence as $n_0 \to \infty$ then *both* ε and ω would limit to zero, in such a way as to lead to divergent lobe area but zero flux. For the Morse potential we briefly discuss two cases of the limit $\omega \to 0$: (i) ε fixed and (ii) lim $\varepsilon \to 0$. For the case of ε fixed note from the Appendix (equation (5.A.3)) that as $\omega \to 0$ the amplitude of oscillation of the p coordinate of τ_ε diverges, so that for sufficiently small ω the problem becomes no longer near-integrable. Figure 5.5.7 shows the turnstile lobe $L_{bu}(1)$ for four successively smaller values of ω. The lobe grows in area and spirals around R_b. There is no obvious reason why the lobe should not remain in a spiral as ω further decreases, and given that the stable manifold cannot intersect itself, $L_{bu}(1)$ would then be trapped entirely in R_b, allowing for two possibilities as $\omega \to 0$: either the area of the lobe and hence of R_b diverge, or the area of R_b will remain finite, forcing the turnstile lobe area to remain finite (this would not contradict Melnikov theory since we are not in the near-integrable regime). The plots suggest that the areas will keep growing, though a much more intensive numerical study would be needed to assure this. Even if this were the case, however, one would of course not conclude from a maximal flux result that the aim for good dissociation rates is to make the forcing frequency arbitrarily small:

(i) That R_b grows with decreasing frequency implies that more and more of the phase space is eaten up by the bounded region.

(ii) One sees from Figure 5.5.7 that as the frequency decreases a larger percentage of $L_{bu}(1)$ intersects with $P_\varepsilon(L_{ub}(1))$, so that a larger percentage of phase space trajectories escaping under the next iterate of P_ε were just captured at that time sample.

(iii) $L_{bu}(1)$ does not necessarily penetrate any better the inner level sets of the unperturbed Hamiltonian as ω decreases.

It is precisely issues such as these, such as the intersection of the turnstile lobes with other lobes and with level sets of the unperturbed Hamiltonian, that need to be dealt with for a complete study of dissociation rates, and we address these issues in the next two sections. We lastly remark here that for the case $\lim \varepsilon \to 0$ the flux will fall to zero, and as long as ε/ω is kept sufficiently small the problem remains near-integrable (one would be interested in this situation for low intensity microwave ionization or dissociation of highly excited states).

5.5.3 Dissociation rates

The main limitation of flux is that it involves phase space *areas* with no regard for the histories and initial conditions of phase space trajectories within these areas. In the next two sections we study two basic phase space transport questions relevant to molecular dissociation which address the history and location of phase space trajectories. The question addressed in this section is: at time $t = \frac{2\pi}{\omega_2} n$, how many of the initially *bounded* points (*i.e.*, those in $R_b(0)$ at $t = 0$) have *escaped* (*i.e.*, are in $R_u(n)$)? For $n > 0$ the turnstile $L_{bu}(1;n)$ may in general contain points which were in $R_b(0)$ and $R_u(0)$ at $t = 0$, so to answer this question one must know the history of the points in the turnstile lobe. Such a question is studied in the following language. At time $t = 0$ let the regions $R_b(0)$ and $R_u(0)$ be covered *uniformly* with particles of species S_b and S_u, respectively, and let $\mathcal{T}_{bu}(n)$ ($\mathcal{T}_{ub}(n)$) denote the total amount of species S_b (S_u) contained in region $R_u(n)$ ($R_b(n)$) at $t = \frac{2\pi}{\omega_2} n$. Then for area-preserving sequences of maps:

$$\mathcal{T}_{bu}(n+1) - \mathcal{T}_{bu}(n) = \mu_b(L_{bu}(1;n)) - \mu_b(L_{ub}(1;n)), \tag{5.5.5}$$

where $\mu_b(\cdot)$ denotes the area of the region within the parentheses that contains S_b. Equation (5.5.5) then quantifies a dissociation rate, and summing (5.5.5) over n (and using that $T_{bu}(0) = 0$)), one obtains an expression for $T_{bu}(n)$ for any n (of course $T_{ub}(n) = T_{bu}(n)$, so we concentrate on $T_{bu}(n)$). Solving equation (5.5.5) boils down to computing lobe *content* $\mu_b(L_{bu}(1;n))$ and $\mu_b(L_{ub}(1;n))$. We will discuss two approaches at solving for lobe content, one based on a Markov model approach, applied to periodic forcing by MacKay, Meiss, Ott, and Percival[28,29] and extended here to quasiperiodic forcing, and one based on the study of the topology of lobe intersections with the bounded and unbounded regions and with other lobes, applied to periodic forcing by Rom-Kedar et al.[26] and Rom-Kedar and Wiggins[35] and extended to quasiperiodic forcing in Chapter 2.

The phase space transport question studied in this section, that of an initial *uniform* distribution of points in phase space, is popular in chemical kinetics studies,[15–21] but we point out that a study of an initial distribution that depends on energy levels would also seem worthwhile, and we will discuss this latter problem as well.

5.5.3a Markov models

The Markov model approach provides an approximate answer to the first transport question in the context of periodic forcing, and has gained some popularity in the application to molecular dissociation. Its advantages are its conceptual simplicity and possible ease of implementation; its disadvantages are that it can lead to a poor approximation, one that can be even qualitatively incorrect and can miss basic features of transport. For a two-dimensional Poincaré map derived from time-periodic forcing, one has for the Poincaré section the same constructs as in the time slices of the quasiperiodic case, except that everything is independent of the sample time; hence, we use the

same notation as in the quasiperiodic case, with the parameter n dropped. The Poincaré section is divided into two regions R_b and R_u, and for an elementary version of the Markov method often employed in chemical kinetics studies one further specifies for each of these regions the stochastic part that can participate in transport between R_u and R_b, denoted r_u and r_b, respectively. Thus r_b is the portion of R_b outside the outermost surviving KAM torus, minus the islands of stability, and calculating the area of this region *approximately* is not too difficult. One then assumes that once a lobe maps from one region to another, its contents instantaneously diffuse throughout the region it maps into. This rather unrealistic assumption allows one to determine lobe content in a simple manner. If we let $p \equiv \mu(L_{ub}(1))/\mu(r_b)$, *i.e.*, p is the percentage of r_b occupied by the captured turnstile lobe, then *at the* n^{th} *time sample*

$$\mu_b(L_{bu}(1)) = (1-p)^n \mu(L_{bu}(1)), \quad \mu_b(L_{ub}(1)) = 0, \tag{5.5.6}$$

and one finds

$$\mathcal{T}_{bu}(n) = \{1-(1-p)^n\}\mu(r_b). \tag{5.5.7}$$

For a sequence of maps the Markov method generalizes in a simple manner: if we let $p_i \equiv \mu(L_{ub}(1;i))/\mu(r_b(i+1))$, then the expression for $\mu_b(L_{bu}(1;n))$ becomes for $n \geq 1$

$$\mu_b(L_{bu}(1;n)) = (1-p_0)\cdot(1-p_1)\cdot\ldots\cdot(1-p_{(n-1)})\mu(L_{bu}(1;n)) \tag{5.5.8}$$

and one has

$$\mathcal{T}_{bu}(n) = \{1-(1-p_0)\cdot\ldots\cdot(1-p_{(n-1)})\}\mu(r_b(n)). \tag{5.5.9}$$

Since in general the p_i are different, one has for the sequence of maps added freedom to enhance or diminish aspects of transport over finite time scales. The previously mentioned influence of interference effects and relative scaling factors on average lobe area are relevant to expressions (5.5.8) and (5.5.9),

since the p_i are proportional to turnstile lobe area. One must also consider the variation of $\mu(r_b(i))$ under the Markov model, but as should soon become clear, the dependence on $\mu(r_b(i))$ is a rather artificial consequence of the unrealistic approximation of infinitely fast diffusion. We lastly remark that for $p_0+p_1\ldots+p_{(n-1)} = constant$, the product $(1-p_0)\cdot\ldots\cdot(1-p_{(n-1)})$ is maximal in the limit $p_0 = p_1 = \ldots = p_{(n-1)}$ (this is just the simple result that the maximum volume of an n-sided box is associated with a cube), *i.e.*, this product is maximal, given the constraint, in the single-frequency limit.

5.5.3b The topology of lobe intersections

An exact determination of lobe content, and hence an exact answer to the first transport question, necessitates a consideration of the topology of lobe intersections. Until now, we have discussed only the turnstile lobes $L_{bu}(1;n)$ and $L_{ub}(1;n)$, *i.e.*, the lobes of the n^{th} time slice that escape/are captured between the n^{th} and $(n+1)^{th}$ time slice; we now must consider other lobes $L_{bu}(m;n)$ and $L_{ub}(m;n)$, defined to be for $m \geq 1$ the lobes of the n^{th} time slice which escape/are captured between the $(n+m-1)^{th}$ and $(n+m)^{th}$ time slice (see Figure 5.5.8). For quasiperiodic forcing, to find the true content of the turnstile lobes at $t = \frac{2\pi}{\omega_2}n$, one maps the lobes back to $t = 0$ and sees how much of them are inside and outside $R_b(0)$:

$$\begin{aligned}\mu_b(L_{bu}(1;n)) &= \{T_\varepsilon^{-1}(\cdot\,;0)\circ\ldots\circ T_\varepsilon^{-1}(L_{bu}(1;n);n-1)\}\cap R_b(0)\\ &= L_{bu}(1+n;0)\cap R_b(0) \qquad (5.5.10)\\ \mu_b(L_{ub}(1;n)) &= 0,\end{aligned}$$

where $T_\varepsilon^{-1}(\cdot\,;n)$ denotes the map from the $(n+1)^{th}$ to the n^{th} time slice. Since the stable manifold can never intersect itself, the only way for $L_{bu}(1+n;0)$, $n \geq 1$, to be outside $R_b(0)$ is for it to be inside the lobes $L_{ub}(m;0)$ for $1 \leq m \leq n$, as shown in Figure 5.5.9 (note that it is due to the fact

that the one-dimensional manifolds of each time slice are subject to the same geometrical constraints as those manifolds in the familiar case of periodic forcing that allows a straighforward extension of lobe intersection analysis to the quasiperiodic case). Hence one can rewrite the first expression of (5.5.10) as

$$\mu_b(L_{bu}(1;n)) = \mu(L_{bu}(1;n)) - \sum_{m=1}^{n} \mu(L_{bu}(1+n;0) \cap L_{ub}(m;0)) \qquad (5.5.11)$$

(for $n = 0$ replace the sum by zero). As n increases, the lobe intersections become exceedingly complicated, and the topology of the intersections is better appreciated by recognizing the approximate self-similar behavior of the preimages of the lobes as one evolves them backwards in time. These issues are discussed in Chapter 3 (equivalently ref. 40) in the context of two-dimensional maps, and we convey the essence of the situation for seqences of maps in Figure 5.5.10. Once a lobe has completed one revolution around the bounded region, it begins to wrap around other turnstile lobes, as shown in Figure 5.5.9 and highlighted by the shaded region in Figure 5.5.10(a). This shaded region then itself repeats the process of winding around the core under further evolution backwards in time, as shown in Figure 5.5.10(b). One then understands the lobe intersections in terms of pieces of $L_{bu}(1;n)$ repeatedly revolving around the core under evolution backward in time until they intersect the turnstile $L_{ub}(1;\tilde{n})$, $\tilde{n} < n$: for each such piece the number of revolutions before the turnstile intersection can vary from 1 to ∞, and the period of each revolution can vary from some minimum number to ∞. One can construct a symbolic dynamics to describe the topology of these pieces,[40] but we will not dwell on this here, our goal being merely to point out the complicated nature of the intersections, and yet the existence of a relatively simple framework with which to understand the topology of the intersections.

The intersection of lobes with $R_u(n)$ and $R_b(n)$, or equivalently with other lobes, provides the framework for exact computation of transport under

quasiperiodic forcing, and we compare this framework with the approximate Markov model, and with periodic forcing. Our main interest is in the latter comparison, and we do not wish to digress too much with the former. A careful discussion of the deficiencies of the approximate Markov model has been given by Rom-Kedar and Wiggins[41] and Wiggins[23] (and is also pointed out by Zembekov[42]), and we only highlight here some of the results. Previous studies employing the Markov approximation show agreement with exact results only at short time scales.[21,43] Indeed any practical implementation of the Markov method entails exponential decay in transport rates, whereas the exact theory allows both exponential and power law decay and the latter can indeed occur.[26] Even the short time agreements can be viewed as rather accidental, for there is no reason to expect that the way lobes intersect one another bears any relation to the rather unrealistic process that assumes infinitely fast diffusion and disregards lobe intersections. In fact, we will soon mention in the context of quasiperiodic forcing a simple example where short time agreement is not expected.

In comparing the exact transport theory for maps and sequences of maps, a central difference to recognize is the additional freedom in the latter case to enhance or diminish aspects of transport over finite time scales due to the variation of lobe areas. One can study this variation and exploit it to one's advantage by numerical simulation of the lobes. In Figure 5.5.11 we show two simple examples that differ only in reversal of the sign of the amplitude for the first forcing frequency. In Figure 5.5.11(a) the areas of the first six escape lobes are almost twice that of the average lobe areas, while the areas of the first six capture lobes are close to zero. Hence at short time scales the dissociation rates are almost twice as large as the infinite-time average flux. In Figure 5.5.11(b) the situation is reversed: hence, after five time samples the content of the bounded region has changed significantly. A Markov model approach would imply that the turnstile lobe $L_{bu}(1;5)$ contains a significant

fraction of points that were not initially in the bounded region, but this is in fact not true since $L_{bu}(1;5)$ intersects with none of the captured lobes (on the scale shown it looks possible that $L_{bu}(1;5)$ just touches one of the capture lobes, but at greater resolution this is seen to not happen, which is consistent with the $n = 0$ plot). Despite the additional freedom of variation of lobe areas, one must also bear in mind that average turnstile lobe area is maximal in a single-frequency limit, for lobe area does play a central role in phase space transport.

To numerically determine lobe content at any sample time $t = \frac{2\pi}{\omega_2}n$ for an initial uniform distribution in phase space, one could initialize the turnstile with a uniform distribution of points and then evolve the system backwards in time to $t = 0$ and determine how many points are inside and outside $R_b(0)$. As mentioned earlier, one might also be interested in an initial distribution that depends on energy levels: for this one would employ the same procedure as above, except that at $t = 0$ one would then weight the points in $R_b(0)$ according to the chosen distribution.

5.5.4 Lobe penetration of energy levels and dissociation rates of a microcanonical ensemble

In addition to lobe *area* and lobe *content*, a third feature of dissociation to study in terms of phase space transport is lobe *penetration*, meaning the ability of the lobes to penetrate along the action, or energy, coordinate. Indeed, there has been considerable interest in the ionization and dissociation of a system prepared in a particular quantum state,[1–7,9,12,13] which in a classical context corresponds to the escape of an ensemble of points initialized on a particular level set of the unperturbed Hamiltonian. Previous studies have made no real connection to lobe dynamics within the homoclinic tangle; often there is discussion of such features as the presence of cantori and overlap of

neighboring resonances,[7,13] but these discussions are sometimes rather heuristic. As we have stressed, the sole mechanism for dissociation is the turnstiles of the homoclinic tangle, so any discussion of dissociation ultimately needs to be related to preimages of these turnstiles.

The transport equations, similar in spirit to Section 5.5.3, are easily written down. Let $\mathcal{L}_0$ denote the level set of interest of the unperturbed Hamiltonian, and $\Delta\mathcal{L}_0(n)$ denote the percentage of points initially on the level set that escape between the n^{th} and $(n+1)^{th}$ time sample, then

$$\Delta\mathcal{L}_0(n) = \frac{\mu(L_{bu}(1+n;0) \cap \mathcal{L}_0)}{\mu(\mathcal{L}_0)}, \tag{5.5.12}$$

where $\mu(\,)$ denotes the number of points along the specified curve. One typically chooses the initial distribution on the level set to be uniform along the angle coordinate in action-angle space (which of course leads to a weighted distribution in (x,p) space). Given this framework, one can divide the study of lobe intersections with the unperturbed level sets into three regimes, depending on the time scale $t = \frac{2\pi}{\omega_2}n$.

5.5.4a Short time scale - small n

For small n the shapes of the lobes $L_{bu}(1+n;0)$ are not very convoluted and their thickness normal to the unperturbed separatrix is well approximated by the Melnikov expression (5.3.2). In the small n regime we can identify two aspects of the lobes that affect penetration.

(a) The Melnikov function provides a measure of *relative penetration*, as described below. Let

$$q_\varepsilon^{s,u}(-s,\theta_1,\theta_{2_0};\nu) = q_{\varepsilon=0}(-s) + \varepsilon\frac{\partial q_\varepsilon^{s,u}}{\partial\varepsilon}(-s,\theta_1,\theta_{2_0};\nu)|_{\varepsilon=0} + O(\varepsilon^2)$$

denote the (x,p) component of a point on $W^s(\tau_\varepsilon)$ and $W^u(\tau_\varepsilon)$, respectively, where $q_{\varepsilon=0}(-s)$ is the (x,p) component of a point on the unperturbed sepa-

ratrix. Then

$$\begin{aligned} H(q_\varepsilon^u) - H(q_\varepsilon^s) =& H(q_{\varepsilon=0}) + \varepsilon DH(q_{\varepsilon=0}) \cdot \frac{\partial q_\varepsilon^u}{\partial \varepsilon}|_{\varepsilon=0} + O(\varepsilon^2) \\ & - H(q_{\varepsilon=0}) - \varepsilon DH(q_{\varepsilon=0}) \cdot \frac{\partial q_\varepsilon^s}{\partial \varepsilon}|_{\varepsilon=0} - O(\varepsilon^2) \qquad (5.5.13) \\ =& \varepsilon DH(q_{\varepsilon=0}) \cdot (\frac{\partial q_\varepsilon^u}{\partial \varepsilon} - \frac{\partial q_\varepsilon^s}{\partial \varepsilon})|_{\varepsilon=0} + O(\varepsilon^2). \end{aligned}$$

By definition[22,32] the first order term in the last line of (5.5.13) is $\varepsilon M(s, \theta_1, \theta_{2_0}; \nu)$. Hence:

$$H(q_\varepsilon^u(-s, \theta_1, \theta_{2_0}; \nu)) - H(q_\varepsilon^s(-s, \theta_1, \theta_{2_0}; \nu)) = \varepsilon M(s, \theta_1, \theta_{2_0}; \nu) + O(\varepsilon^2). \qquad (5.5.14)$$

The interpretation is thus that εM gives the first order measure of how the lobe boundaries span the energy levels of the *unperturbed* Hamiltonian. In the small n regime, then, the previously discussed influence of relative scaling factors and interference effects on the Melnikov function have relevance on the lobes' ability to span the unperturbed energy levels. Recall from Section 5.5.2 that, if for any pair of frequencies (ω_1, ω_2) we vary the amplitudes (E_1, E_2) with $E_1^2 + E_2^2$ fixed, then $< |\varepsilon M| >_s$ is maximal in the single-frequency limit corresponding to the larger relative scaling factor (where $< >_s$ denotes the average over $s \in (-\infty, \infty)$). For the small n regime we are interested in a finite s interval, so that, though $< |\varepsilon M| >_s$ can have an influence in this regime, it is also balanced by the added freedom of two-frequency forcing to enhance or diminish $|\varepsilon M|$ over a finite s interval for a fixed $< |\varepsilon M| >_s$. Additionally relevant to lobe penetration is the *maximum* value of $|\varepsilon M|$: though interference effects cause $< |\varepsilon M| >_s$ to dip *below* a linear increase as the (E_1, E_2) weighting is varied from the lower to higher relative scaling factor (refer back to Figure 5.5.5), $|\varepsilon M|_{max}$ will be *raised* above a linear increase, so that the span of the unperturbed energy levels can be maximal in the two-frequency case (see Figure 5.5.12).

(b) In addition to a measure of the range of unperturbed energy levels spanned by the lobes, *i.e.*, the *relative* penetration, one is also interested in the *absolute*

penetration of energy levels, *i.e.*, in how the lobes penetrate any particular level set, which is affected not only by lobe width along the energy coordinate, but also the previously mentioned net oscillation and "breathing" of the lobe structure. A measure of the net oscillation in the p coordinate is given by the expression for τ_ε (see equation (5.A.4) in the Appendix): the amplitudes of the p_ε oscillation are $\varepsilon dE_i/\omega_i$ for the ω_i component. The "breathing" of the lobe structure is more difficult to study and, as previously explained, one will typically need to resort to numerical simulation of the lobe structure. We wish to stress the importance of the breathing and oscillations of the lobe structure, for it can have a significant effect on dissociation: for example, it can allow portions of a high quantum number bounded level set to initially be already *outside* $R_b(0)$. Though not explicitly mentioned in the next two regimes, it should be understood that the net breathing and oscillations are relevant there too.

5.5.4b Intermediate time scale - medium n

In this regime n is, on the one hand, large enough that the lobes $L_{bu}(1+n;0)$ wind around $R_b(0)$, wrapping around other escape lobes and intersecting other capture lobes (recall the earlier discussion of the approximate self-similar behavior of the lobes to help make sense of this intermediate regime); however, n is still small enough that the penetration process has not yet approached saturation. It is in this intermediate regime that explicit numerical simulation of the lobes is most worthwhile, for the analytical tools and constructs applicable to the limiting regimes of small n and large n (to be discussed next) have only partial relevance here. The Melnikov function, useful in the small n regime, offers no accounting for how the lobes further penetrate the energy levels by repeatedly winding around the core and wrapping around other lobes. Additionally, until the penetration process approaches saturation, the large n constructs we will mention (such as the outermost surviving KAM-type torus)

may have little relevance. The constructs of both limiting regimes will have *some* relevance, though: the intermediate n lobes wrap around the small n lobes, and hence the penetration of these small n lobes does affect the penetration of the intermediate n lobes. Conversely, the large n constructs, such as the outermost surviving KAM-type torus, will serve as an inner barrier that limits penetration. The single-frequency and two-frequency comparisons will thus be influenced by the comparisons in the two limiting regimes, but a true comparison will necessitate numerical simulations on a case by case basis.

5.5.4c Long time scale - large n

For large enough n the ability of the lobes to penetrate the unperturbed energy levels saturates. In principle one could study dissociation in terms of lobe dynamics until saturation, but to do this would be quite computer intensive (a large n calculation might be interesting nevertheless). Other constructs are useful in this regime and have been employed in the single-frequency case: in particular, the outermost surviving KAM-type torus provides a phase space boundary that encloses a region in phase space whose points remain bounded for all time. Hence the escape lobes can never penetrate past this torus, which thus acts as a barrier to lobe penetration. In this respect there is interest in the overlap of neighboring resonances,[13] for such an overlap disallows a surviving KAM-type torus between the two resonances, and thus can be used as a tool for estimating the location of the outermost surviving KAM-type torus. We remark parenthetically that Arnold [44] indicates how original KAM theory extends to quasiperiodically forced one degree-of-freedom Hamiltonian systems.

Goggin and Milonni[12,13] study one-frequency and two-frequency dissociation *thresholds* for an ensemble of points on a level set corresponding to the ground state of the Morse oscillator: they numerically compute the forcing

intensity needed to dissociate five percent of the level set once the dissociation process has been saturated. They provide numerical evidence that two-frequency forcing has a *lower* threshold intensity, and argue that this lowering is a consequence of the additional resonances created by the additional forcing frequency, which facilitates resonance overlap and hence dissociation. We add two qualifications to their results. First, the sole numerical comparison for dissociation they offer keeps one of the frequencies in the two-frequency problem fixed at what is almost the "best" available frequency (*i.e.*, lowest single-frequency threshold intensity), so that much of the comparison amounts to one between a relatively poor single-frequency forcing (high threshold intensity) and a combination of one good and one poor frequency in the two-frequency problem, a somewhat unfair comparison. Nevertheless, their comparison offers *some* evidence of a lower two-frequency threshold, though not so striking as their visual comparison suggests at first glance. Second, the presence of additional resonances alone does not ensure enhanced resonance overlap, for if one has, for example, twice as many resonances at half the width, it is not obvious that one has gained anything. Since forcing intensity is proportional to the sum of the *squares* of the electric fields, however, it is possible to have in the two-frequency problem at least twice as many resonances at *more* than half the width for the same intensity as the single-frequency problem. It is thus plausible that with the ideal frequency combinations one will have enhanced resonance overlap, and a lower threshold intensity.

After distinguishing between the above three regimes, we point out that much of the previous studies of ionization and dissociation focus on a fairly specific situation: the *threshold regime*, where the forcing intensity is at the threshold where ionization and dissociation just begin. This corresponds to the situation where the lobes $L_{bu}(1+n;0)$ intersect the level set only in the saturation regime (large n). Another interesting situation that deserves study

is what we will call the *post-threshold* regime, where the forcing intensity is larger than the threshold intensity for the level set concerned, and the escape lobes $L_{bu}(1+n;0)$ thus play a role in dissociation for small and medium values of n. As should be clear from our discussion, the threshold and post-threshold regimes can have qualitatively different behavior and necessitate different methods of investigation. We end this section with three remarks to highlight these differences and some of the issues in the post-threshold regime.

(i) Constructs relevant to the threshold regime, such as cantori, resonance overlap, and the outermost surviving KAM-type torus, may have little relevance to the post-threshold regime. For example, in the case where the concerned level set intersects the turnstiles, it should be clear that issues such as resonance overlap can have little relevance.

(ii) The penetration dependence on forcing frequency in the threshold and post-threshold regime may differ. In the threshold regime, the frequency of the concerned level set can have particular relevance; for example Goggin and Milonni[13] offer evidence that for one-frequency forcing it is best to force near this frequency (*i.e.*, one obtains lowest threshold intensity near the level set frequency). A reasonable argument for this phenomenon is that, as one increases the forcing intensity to find the threshold regime, and hence moves the outermost surviving KAM torus to smaller and smaller action values, having a nice fat 1:1 resonance at the concerned level set awaiting the incoming KAM torus will facilitate resonance overlap and hence facilitate moving the KAM torus inside the level set. For the post-threshold regime, however, the relevant neighboring resonances will have already overlapped and the level set frequency may have less relevance.

(iii) Consider a particular scenario as an example of the more varied questions one can address in the post-threshold regime (also of interest for

comparing with quantum mechanical calculations). Suppose we force at some fixed intensity and some fixed frequency tuned, say, to the ground state dissociation energy, and ask how dissociation rates change for an initial distribution of points on a high quantum number level set as we increase the quantum number, *i.e.*, move closer to the separatrix. On the one hand, for a fixed forcing intensity and frequency the turnstile lobes are fixed, so that as the levet set of interest approaches the separatrix its intersection with the turnstile can approach a constant length, so that the numerator of (5.5.12) approaches a constant. However the arclength and period of the level set increases to infinity, so that the transition rate falls to zero.

5.5.5 Chaos

We close with a comment on the chaotic nature of the dynamics associated with quasiperiodic forcing, or equivalently, with sequences of maps, and the ability to detect a chaotic response to quasiperiodic forcing. It has of course been widely appreciated for a long time that the chaotic nature of the dynamics in homoclinic and heteroclinic tangles of two-dimensional *maps* is understood in terms of the presence of a Smale horseshoe map. How does one understand chaos for *sequences of maps*? As described rigorously in Chapter 2, and as discussed heuristically here, chaos is understood in terms of the presence of a *traveling horseshoe map sequence* $\{H(\cdot\,;n)|n \in Z\}$ (see Figure 5.5.13). The sequence consists of a bi-infinite sequence of domains $\{\ldots, D(-n), \ldots, D(-1), D(0), D(1), \ldots, D(n), \ldots\}$ such that $T_\varepsilon(D(n);n)$ intersects $D(n+1)$ in a horseshoe-like fashion. The result is chaotic dynamics relative to a $time-dependent$ Cantor set of points.

5.6 Concluding remarks

Invariant manifolds offer a basic framework for studying classes of non-integrable systems. For a periodically or quasiperiodically forced nonlinear oscillator such as the Morse oscillator, the stable and unstable manifolds of the corresponding Poincaré map's homoclinic tangle provide a precise criterion for the partial barriers between bounded and unbounded motion and for the sole mechanism of transport across these partial barriers, the turnstiles. Having identified these phase space structures one can study a variety of transition rates between bounded and unbounded motion, and hence perform a classical study of molecular dissociation.

Addressing the quasiperiodically forced Morse oscillator necessitates a basic extension past the ubiquitous single-frequency case, for one must generalize from two-dimensional *maps* to *sequences of two-dimensional maps*, or equivalently to a three-dimensional Poincaré map acting on a sequence of two-dimensional phase slices. As we have seen, however, the constructs associated with iterates of a two-dimensional map are fairly robust: a fixed set of partial barriers and turnstiles generalizes to a *sequence* of partial barriers and turnstiles, and the Smale horseshoe map generalizes to a *traveling horseshoe map sequence*. A generalized Melnikov function and a *double phase slice sampling method* provide the analytical and computational tools for studying two-frequency dissociation in the context of sequences of maps, and comparing with the single-frequency case. On the one hand, infinite-time average flux is seen to be maximal in the single-frequency limit associated with the larger relative scaling factor; however, lobe penetration of the level sets of the unperturbed Hamiltonian can be maximal in the two-frequency case. Additionally, the variation of lobe areas in the two-frequency problem gives one added freedom to enhance or diminish aspects of transport over finite time scales for a fixed infinite-time average flux.

Melnikov theory and the results based on this theory are valid in the near-integrable regime; however, in practice fairly large perturbations are allowed (refer, for example, to Figure 5.3.2). More significantly, we stress that the basic framework used here - the use of invariant manifolds to partition phase space, identify partial barriers and turnstiles, and study molecular dissociation via phase space transport and lobe dynamics - does not require near-integrability.

For simplicity of discussion, we have focused on two-frequency forcing; however, a similar analysis applies to multiple-frequency systems with more than two forcing frequencies, and indeed to more complicated forcing time dependences than quasiperiodic, as explained in Chapter 2.

Appendix

We include here a few mathematical details about the quasiperiodically forced Morse oscillator.

(i) The expression for an orbit on the unperturbed separatrix $(x_h(t-s), p_h(t-s))$ is solved by setting $H = D_0$, where H is the unperturbed Hamiltonian, and using $\dot{x} = p/m$ to obtain

$$\dot{x} = \pm\sqrt{\frac{2D_0}{m}}(2e^{-ax} - e^{-2ax})^{1/2}.$$

One can integrate this to obtain x and hence $p = m\dot{x}$:

$$\begin{aligned} x_h(t-s) &= \frac{1}{a}\ln\Big(\frac{1+(\omega_0(t-s))^2}{2}\Big) \\ p_h(t-s) &= 2\sqrt{2mD_0}\,\frac{\omega_0(t-s)}{1+(\omega_0(t-s))^2}, \end{aligned} \tag{5.A.1}$$

where $\omega_0 \equiv a\sqrt{\frac{2D_0}{m}}$.

(ii) The expression for the invariant 1-torus τ_ε in $\Sigma^{\theta_{2_0}}$ is solved by setting $z = 1/x$ and studying the Morse system at $z = 0$:

$$\begin{aligned} \dot{z} &= 0 \\ \dot{p} &= \varepsilon d(E_1\cos(\omega_1 t + \theta_{1_0}) + E_2\cos(\omega_2 t + \theta_{2_0})), \end{aligned} \tag{5.A.2}$$

which is solved by

$$\begin{aligned} z(t) &= 0 \\ p(t) &= \varepsilon d(\frac{E_1}{\omega_1}\sin(\omega_1 t + \theta_{1_0}) + \frac{E_2}{\omega_2}\sin(\omega_2 t + \theta_{2_0})). \end{aligned} \tag{5.A.3}$$

The expression for τ_ε is thus given by

$$(x_\varepsilon(\theta_1), p_\varepsilon(\theta_1)) = (\lim x \to \infty, \varepsilon d(\frac{E_1}{\omega_1}\sin(\theta_1) + \frac{E_2}{\omega_2}\sin(\theta_{2_0}))). \tag{5.A.4}$$

(iii) Calculation of the Melnikov function.

From (5.3.1) $M(s,\theta_1,\theta_2;\nu) = I_1 + I_2$ where

$$I_i = \frac{E_i d}{m}\int_{-\infty}^{\infty} p_h(t)\cos(\omega_i t + (\omega_i s + \theta_i))dt \quad i = 1,2. \tag{5.A.5}$$

Using (5.A.1) and that $p_h(t)$ is odd in t:

$$I_i = -\frac{4E_i d}{a}\sin(\omega_i s + \theta_i)\int_0^{\infty} \frac{\omega_0 t}{1+(\omega_0 t)^2}\sin(\omega_i t)d(\omega_0 t) \quad i = 1,2. \tag{5.A.6}$$

Solving the integral gives

$$I_i = -\frac{2\pi}{a}E_i d e^{-\frac{\omega_i}{\omega_0}}\sin(\omega_i s + \theta_i) \quad i = 1,2. \tag{5.A.7}$$

Note that (5.A.7) is valid for $\omega_i > 0$; for $\omega_i = 0$ the integrand in (5.A.6) is identically zero and hence I_i vanishes. The Melnikov function is thus discontinuous at $\omega_i = 0$ due to the nonvanishing of the Melnikov amplitudes in the limit $\omega_i \to 0$. This nonvanishing is due to the fact that $p_h(t) \sim 1/(\omega_0 t)$ for large $|t|$.

(iv) The Melnikov approximation of manifold separation.

If we assume differentiable stable and unstable manifolds we obtain by Taylor expanding in ε

$$d(s,\theta_1,\theta_{2_0};\nu,\varepsilon) = \varepsilon\frac{\Delta^u(0,s,\theta_1,\theta_{2_0};\nu) - \Delta^s(0,s,\theta_1,\theta_{2_0};\nu)}{||DH(q_h(-s))||} + O(\varepsilon^2), \tag{5.A.8}$$

where $q_h(-s) \equiv (x_h(-s), p_h(-s))$ and

$$\Delta^{s,u}(t,s,\theta_1,\theta_2;\nu) \equiv DH(q_h(t-s))\cdot\frac{\partial q_\varepsilon^{s,u}}{\partial\varepsilon}(t,s,\theta_1,\theta_2;\nu)|_{\varepsilon=0}$$

with $q_\varepsilon^{s,u}$ the (x,p) component of an orbit on the perturbed stable and unstable manifold, respectively, in the full autonomous system phase space. Employing the standard procedure of deriving and solving a linear differential equation for $\Delta^{s,u}$ in t[22,32], one obtains

$$\begin{aligned}\Delta^u(0,s,&\theta_1,\theta_{2_0};\nu) - \Delta^s(0,s,\theta_1,\theta_{2_0};\nu) = \\ &M(s,\theta_1,\theta_{2_0};\nu) + \lim_{t\to-\infty}\Delta^u(t,s,\theta_1,\theta_{2_0};\nu) \\ &- \lim_{t\to\infty}\Delta^s(t,s,\theta_1,\theta_{2_0};\nu).\end{aligned} \tag{5.A.9}$$

The non-hyperbolicity and infinite radius of τ_ϵ means that the convergence of the Melnikov integral and the vanishing of the two limits is not guaranteed without additional analysis. The Melnikov integral was explicitly shown to converge in (iii). The two limits in (5.A.9) vanish as long as

$$\lim_{|t| \to \infty} DH(q_h(t-s)) = 0$$

$$\lim_{t \to \infty, -\infty} \frac{\partial q_\epsilon^{s,u}}{\partial \varepsilon}(t, s, \theta_1, \theta_{2_0}; \nu) \text{ is bounded.}$$

The first condition is easily seen to hold, and it should be noted that it does not require hyperbolicity, only that τ_ϵ be of saddle-type stability. The second condition is seen to be satisfied for the component $p_\epsilon^{s,u}$ due to (5.A.3). Both $x_{\varepsilon>0}^{s,u}$ and $x_{\varepsilon=0}^{s,u}$ limit to infinity as $t \to \infty, -\infty$, but their difference is

$$x_{\varepsilon>0}^{s,u} - x_{\varepsilon=0}^{s,u} = \frac{1}{m} \int (p_{\varepsilon>0}^{s,u} - p_{\varepsilon=0}^{s,u}) dt, \qquad (5.A.10)$$

and hence the bounded oscillations of p in (5.A.3) imply bounded oscillations in x at infinity. The constant of integration in (5.A.10) may become unbounded in the concerned limit, but since $\frac{\partial H}{\partial x}(q_h(t-s)) \sim 1/(t-s)^2$ for large $|t-s|$, all we really need is for $\frac{\partial x_\epsilon^{s,u}(t,s,\theta_1,\theta_{2_0};\nu)}{\partial \varepsilon}$, or equivalently $x_{\varepsilon>0}^{s,u}(t, s, \theta_1, \theta_{2_0}; \nu) - x_{\varepsilon=0}^{s,u}(t-s)$, not to diverge faster than t^2 with increasing $|t|$, which is guaranteed by the fact that, for large $|t|$, $x_{\varepsilon=0}^{s,u}(t-s) \sim 2a^{-1}\ln|t-s|$ and $|\dot{x}_{\varepsilon>0}^{s,u}(t, s, \theta_1, \theta_{2_0}; \nu)|$ decays to bounded oscillations with zero mean as t asymptotes to ∞,$-\infty$ (indeed, numerical studies (see Figure 5.A.1) show that the x component of points on the perturbed stable and unstable manifolds also grows as $2a^{-1} \ln|t-s|$ for large $|t-s|$ with increasing and decreasing time, respectively, which would imply that the constant of integration in (5.A.10) remains bounded anyway). We remark that the excellent agreement in Figure 5.3.2 between the exact manifold separation and the Melnikov approximation (5.3.2) can be taken as confirmation that both limits in (5.A.9) indeed vanish, and that the assumption of differentiable stable and unstable manifolds is valid.

References

[1] J.G. Leopold and I.C. Percival, Phys. Rev. Lett. **41**, 944 (1978).

[2] J.G. Leopold and I.C. Percival, J. Phys. B **12**, 709 (1979).

[3] R.V. Jensen, Phys. Rev. A **30**, 386 (1984); in *Proceedings of the Oji International Seminar on Highly Excited States of Atoms and Molecules*, edited by S. Kano and M. Matsuzawa (Fuji-Yoshida, Japan, 1986), p. 149.

[4] J.G. Leopold and D. Richards, J. Phys. B **18**, 3369 (1985).

[5] J.N. Bardsley and M.J. Comella, J. Phys. B **19**, L565 (1986).

[6] G. Casati, B.V. Chirikov, D.L. Shepelyansky, and I. Guarneri, Phys. Rev. Lett. **57**, 823 (1986).

[7] R.S. MacKay and J.D. Meiss, Phys. Rev. A. **37**, 4702 (1988).

[8] R.B. Walker and R.K. Preston, J. Chem. Phys. **67**, 2017 (1977).

[9] D.W. Noid and J.R. Stine, Chem. Phys. Lett. **65**, 153 (1979).

[10] G.C. Lie and J.M. Yuan, J. Chem. Phys. **84**, 5486 (1986).

[11] M. Tung and J.M. Yuan, Phys. Rev. A **36**, 4463 (1987).

[12] M.E. Goggin and P.W. Milonni, Phys. Rev. A **37**, 796 (1988).

[13] M.E. Goggin and P.W. Milonni, Phys. Rev. A **38**, 5174 (1988).

[14] Z.M. Lu, J.F. Heagy, M. Vallières and J.M. Yuan, Phys. Rev. A **43**, 1118 (1991).

[15] M.J. Davis, J. Chem. Phys. **83**, 1016 (1985).

[16] M.J. Davis and S.K. Gray, J. Chem. Phys. **84**, 5389 (1986).

[17] S.K. Gray, S.A. Rice, and M.J. Davis, J. Phys. Chem. **90**, 3470 (1986).

[18] M.J. Davis, J. Chem. Phys. **86**, 3978 (1986).

[19] R.T. Skodjie and M.J. Davis, J. Chem. Phys. **88**, 2429 (1988).

[20] P. Gaspard and S.A. Rice, J. Phys. Chem. **93**, 6947 (1989).

[21] R.E. Gillilan and G.S. Ezra, J. Chem. Phys. **94**, 2648 (1991).

[22] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer-Verlag, New York, 1983).

[23] S. Wiggins, *Chaotic Transport in Dynamical Systems* (Springer-Verlag, New York, 1991).

[24] F.M.A. Salam and S.S. Sastry, IEEE Trans. Circ. Sys. **32**, 784 (1985).

[25] K. Hockett and P. Holmes, Ergod. Th. & Dynam. Sys. **6**, 205 (1986).

[26] V. Rom-Kedar, A. Leonard, and S. Wiggins, J. Fluid Mech. **214**, 347 (1990).

[27] I.S. Kang and L.G. Leal, J. Fluid Mech. **218**, 41 (1990).

[28] R.S. MacKay, J.D. Meiss, and I.C. Percival, Physica D **13**, 55 (1984).

[29] J.D. Meiss and E. Ott, Physica D **20**, 387 (1986).

[30] L. Moorman, E.J. Galvez, B.E. Sauer, A. Mortazawi-M., K.A.H. van Leeuwen, G.v. Oppen, and P.M. Koch, Phys. Rev. Lett. **61**, 771 (1988).

[31] S. Wiggins, Physica D **44**, 471 (1990).

[32] S. Wiggins, *Global Bifurcations and Chaos – Analytical Methods* (Springer-Verlag: New York, 1988).

[33] D. Beigie, A. Leonard, and S. Wiggins, Nonlinearity **4**, 775 (1991).

[34] D. Beigie, A. Leonard, and S. Wiggins, in *Nonlinear Phenomena in Atmospheric and Oceanic Sciences (IMA Volumes in Mathematics and its Applications - Volume 40)*, edited by G.F. Carnevale and R. Pierrehumbert (Springer-Verlag: New York, 1992) 47-138.

[35] V. Rom-Kedar and S. Wiggins, Arch. Rat. Mech. Anal. **109**, 239 (1990).

[36] B. Bruhn, Ann. der Phys. **46**, 367 (1989).

[37] D. Bensimon and L.P. Kadanoff, Physica D **13**, 82 (1984).

[38] R.S. MacKay and J.D. Meiss, J. Phys. A: Math. Gen. **19**, L225 (1986).

[39] R.S. MacKay, J.D. Meiss, and I.C. Percival, Physica D **27**, 1 (1987).

[40] D. Beigie, A. Leonard, and S. Wiggins, Phys. Fluids A **3**, 1039 (1991).

[41] V. Rom-Kedar and S. Wiggins, Physica D **51**, 248 (1991).

[42] A.A. Zembekov, Phys. Rev. A **42**, 7163 (1990).

[43] R. Camassa and S. Wiggins, Phys. Rev. A **43**, 774 (1991).

[44] V.I. Arnold, Russian Mat. Surveys, **18**, 85 (1963).

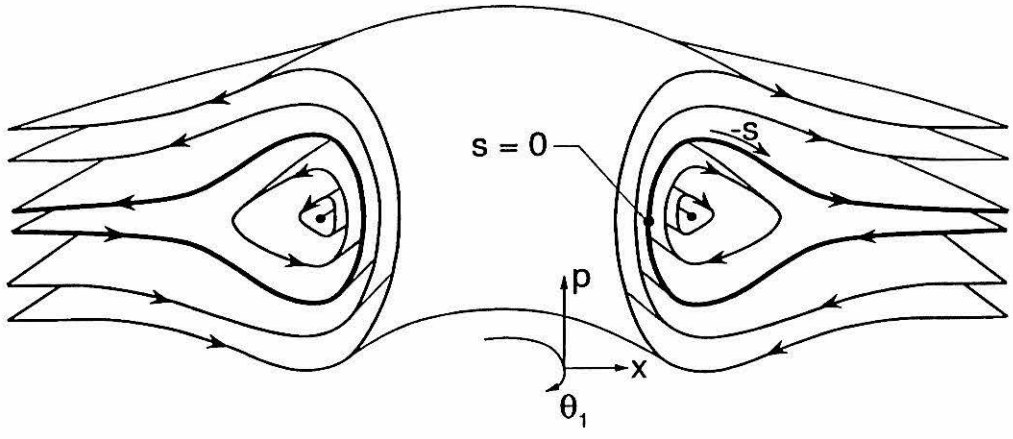


Figure 5.2.1 Invariant manifolds of the unperturbed Poincaré map $P_{\varepsilon=0}$, shown with a cut-away half-view. The separatrix is shown in boldface and is parametrized by (s, θ_1), where s is the time it takes for the point on the separatrix with (x, p) component $(x(-s), p(-s))$ to move to the point with (x, p) component $(x(0), p(0))$ (negative time intervals are included in this definition).

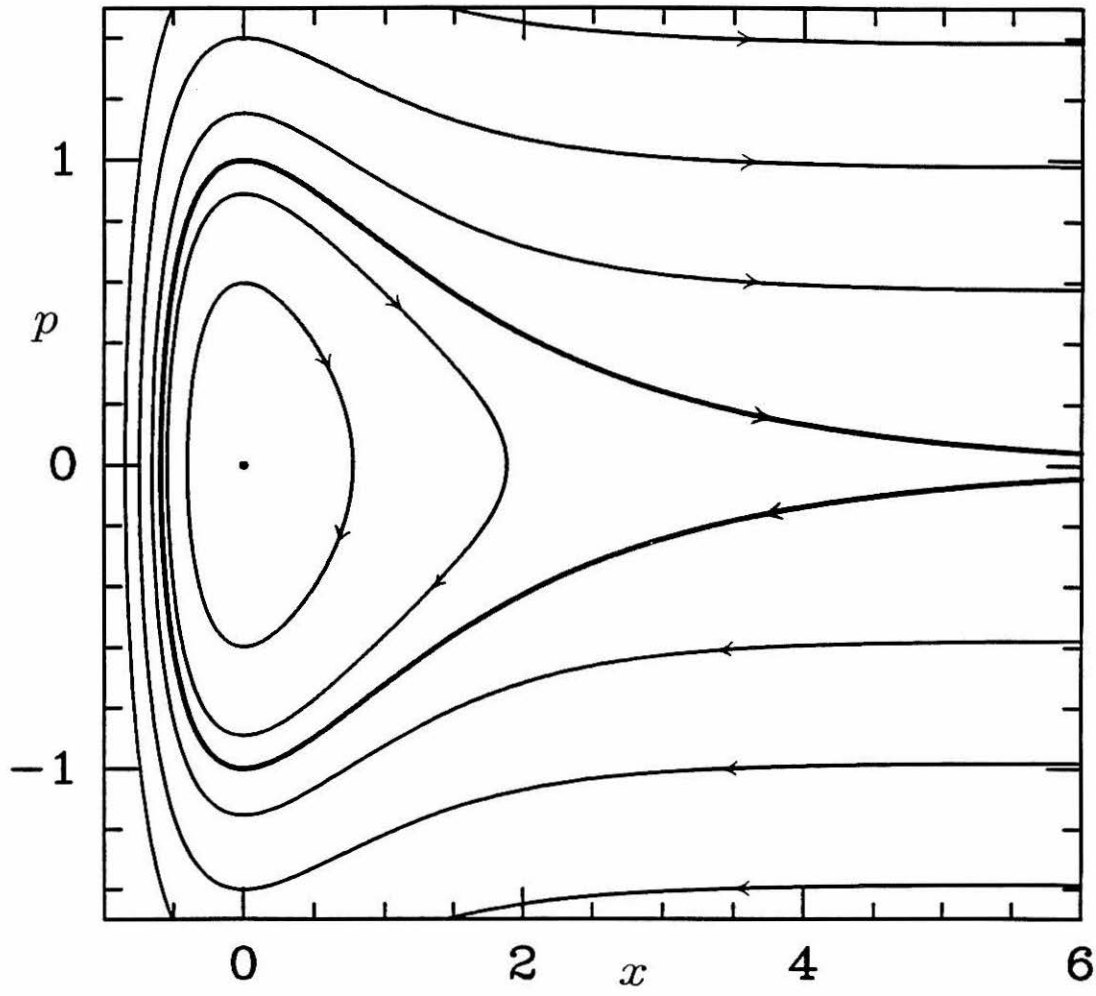


Figure 5.2.2 The phase portrait in the two-dimensional phase slice $\chi(\bar{\theta}_1)$ of $\Sigma^{\theta_{20}}$. As shall be the case with all figures, x and p are plotted in units of Bohr radii and $\sqrt{2mD_0}$, respectively.

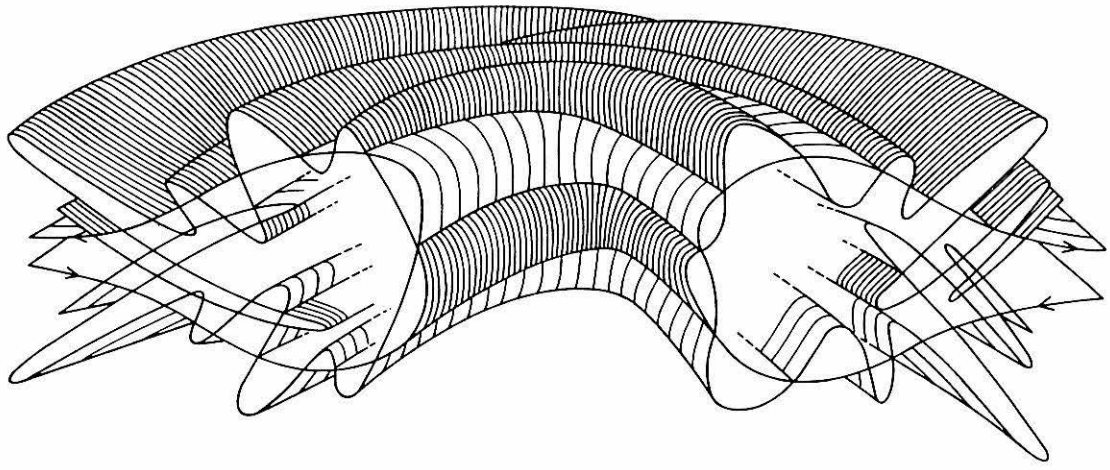


Figure 5.2.3 A possible stable and unstable manifold geometry in $\Sigma^{\theta_{20}}$: the manifolds intersect in an infinity of 1-tori. There are other possible geometries, as discussed in Section 5.3.

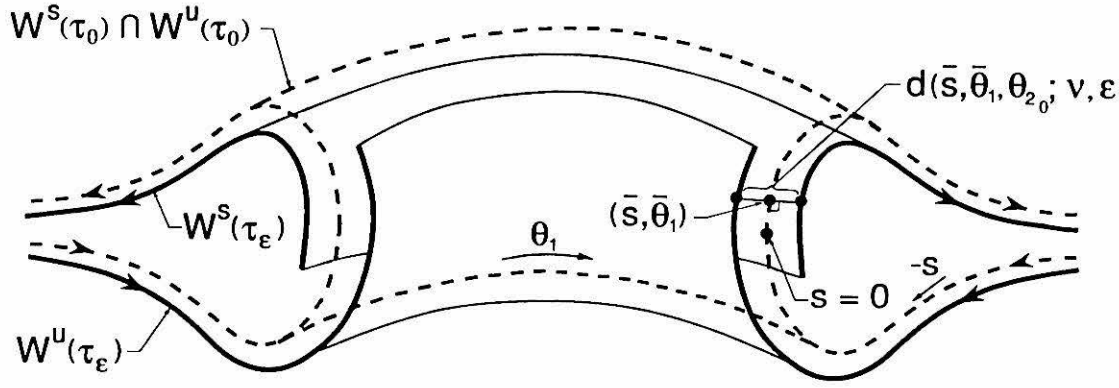


Figure 5.3.1 The manifold separation $d(s, \theta_1, \theta_{2_0}; \nu, \varepsilon)$ normal to the unperturbed separatrix.

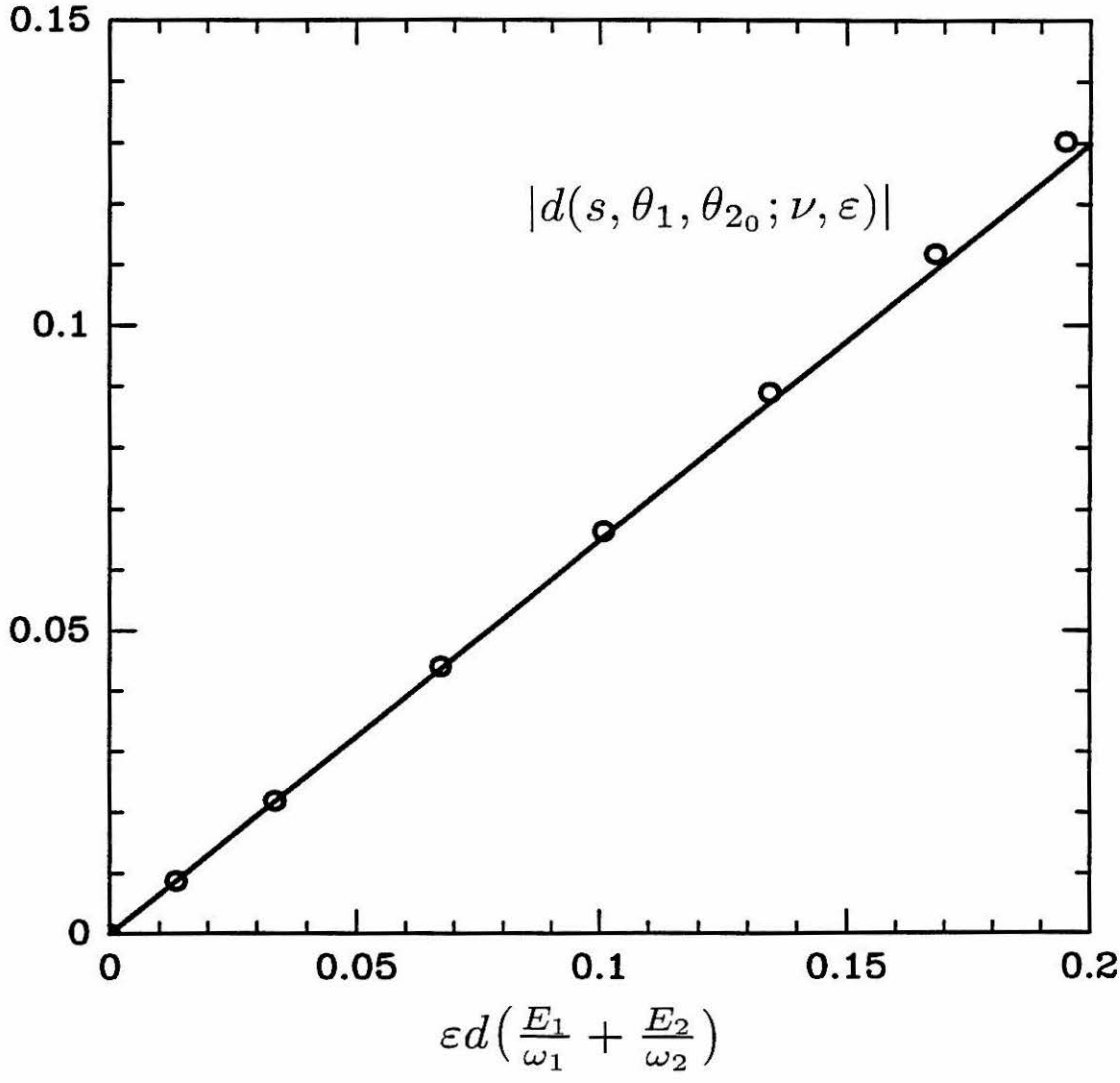


Figure 5.3.2 A comparison between the manifold separation $|d(s, \theta_1, \theta_{2_0}; \nu, \varepsilon)|$ at $(s, \theta_1, \theta_{2_0}) = (0, \frac{13}{8}\pi, \frac{3}{2}\pi)$, as computed by the Melnikov approximation (5.3.2) (the solid line) and by numerical simulation of the manifolds (open dots). For this example $(\omega_1, \omega_2) = 2.6\omega_0(g, 1)$, where g is the golden mean, $E_1 d = 2.047\omega_0\sqrt{2mD_0}$, $E_2 d = 3.685\omega_0\sqrt{2mD_0}$. The horizontal axis is the maximum value of the p coordinate of τ_ε (see equation (5.A.4) in the Appendix), in units of $\sqrt{2mD_0}$.

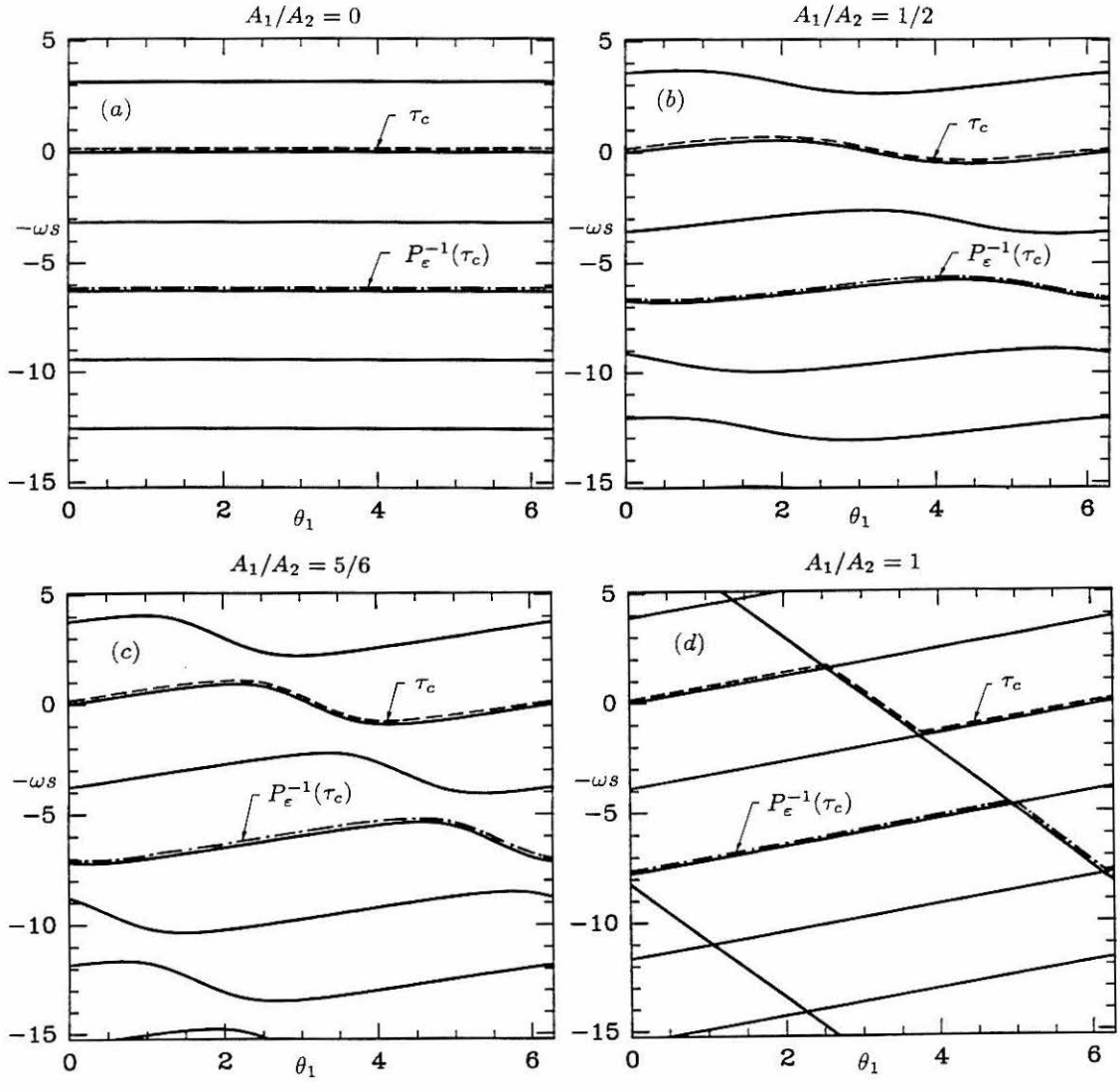


Figure 5.3.3 Zero sets of the Melnikov function (5.3.3) for various ratios of Melnikov amplitudes A_1/A_2 and $\omega_1/g = \omega_2 \equiv \omega$. The dashed and dashed-dotted lines, τ_c and $P_\varepsilon^{-1}(\tau_c)$, are defined and discussed in Section 5.4.2. Note that these lines are really single-valued everywhere, and the vertical lines are merely for visual clarity.

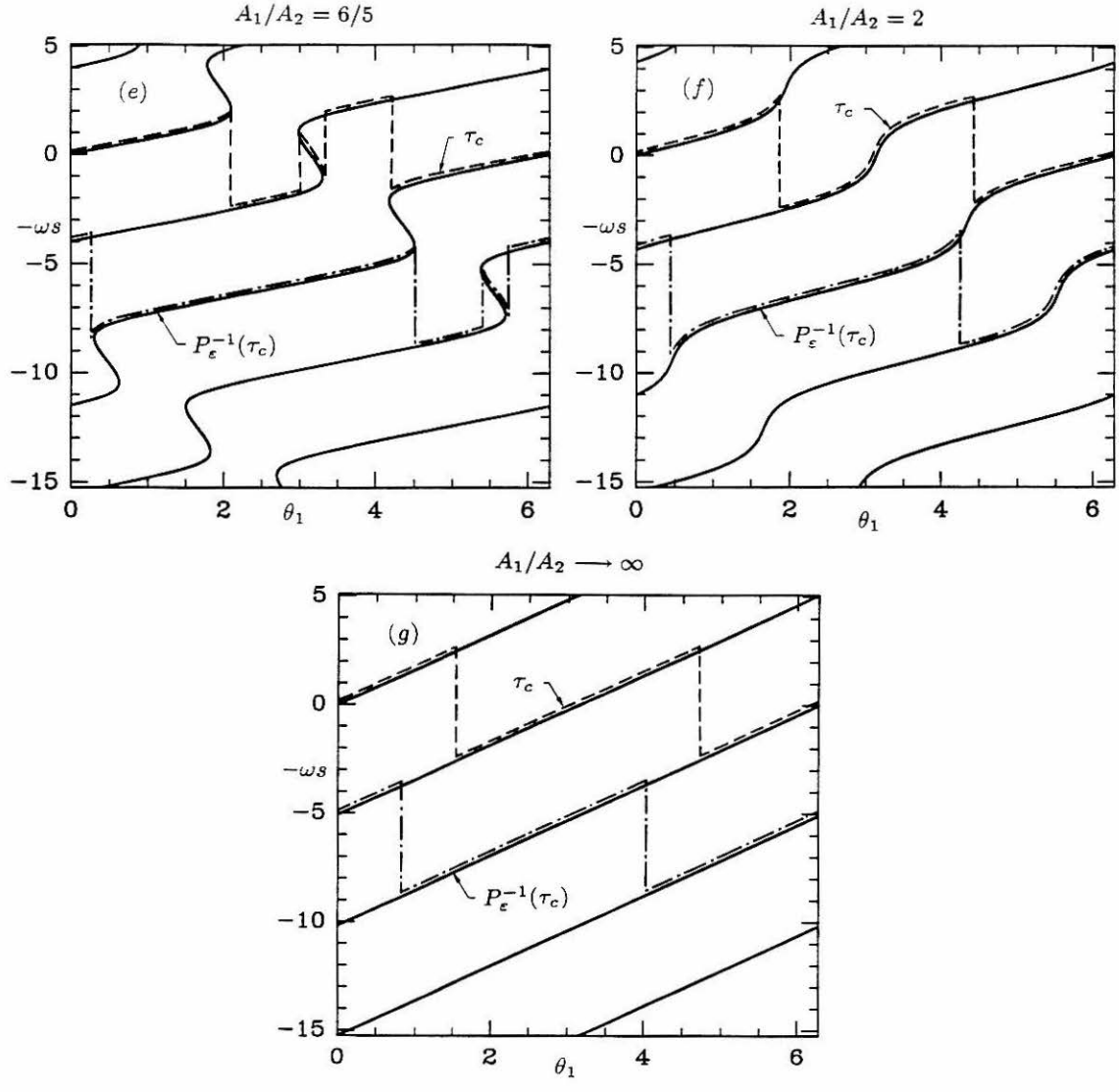


Figure 5.3.3 *Continued.*

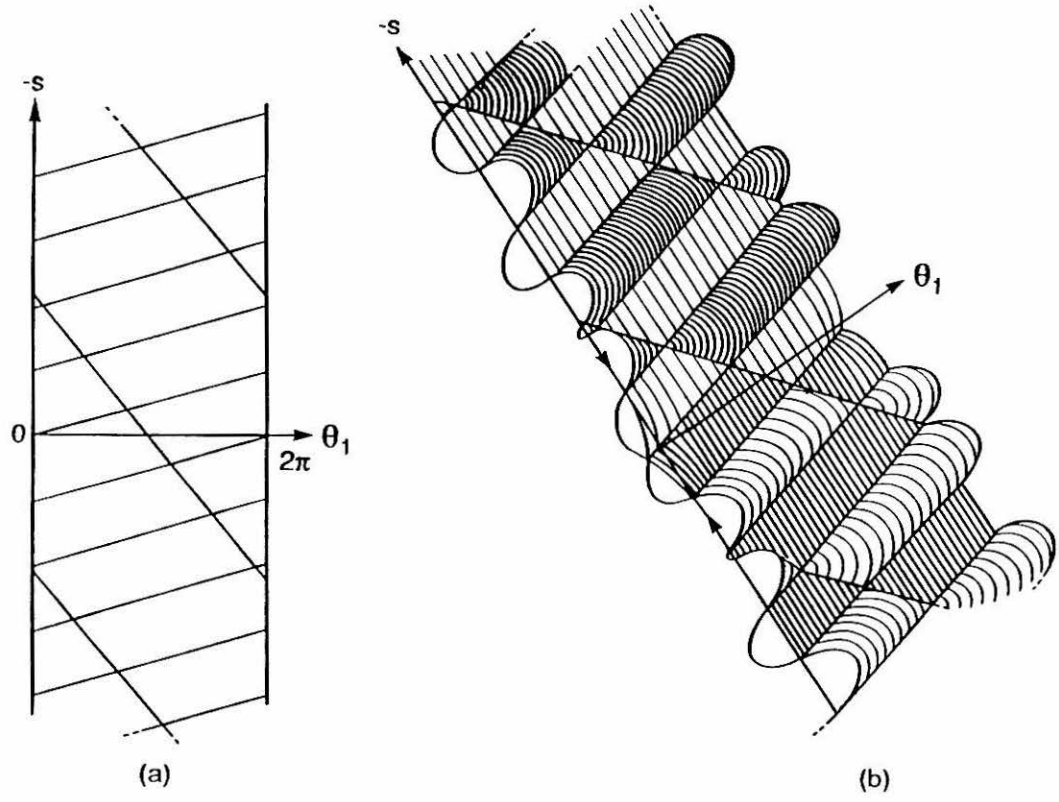


Figure 5.3.4 Visualizing three-dimensional lobes by showing the suppressed dimension.

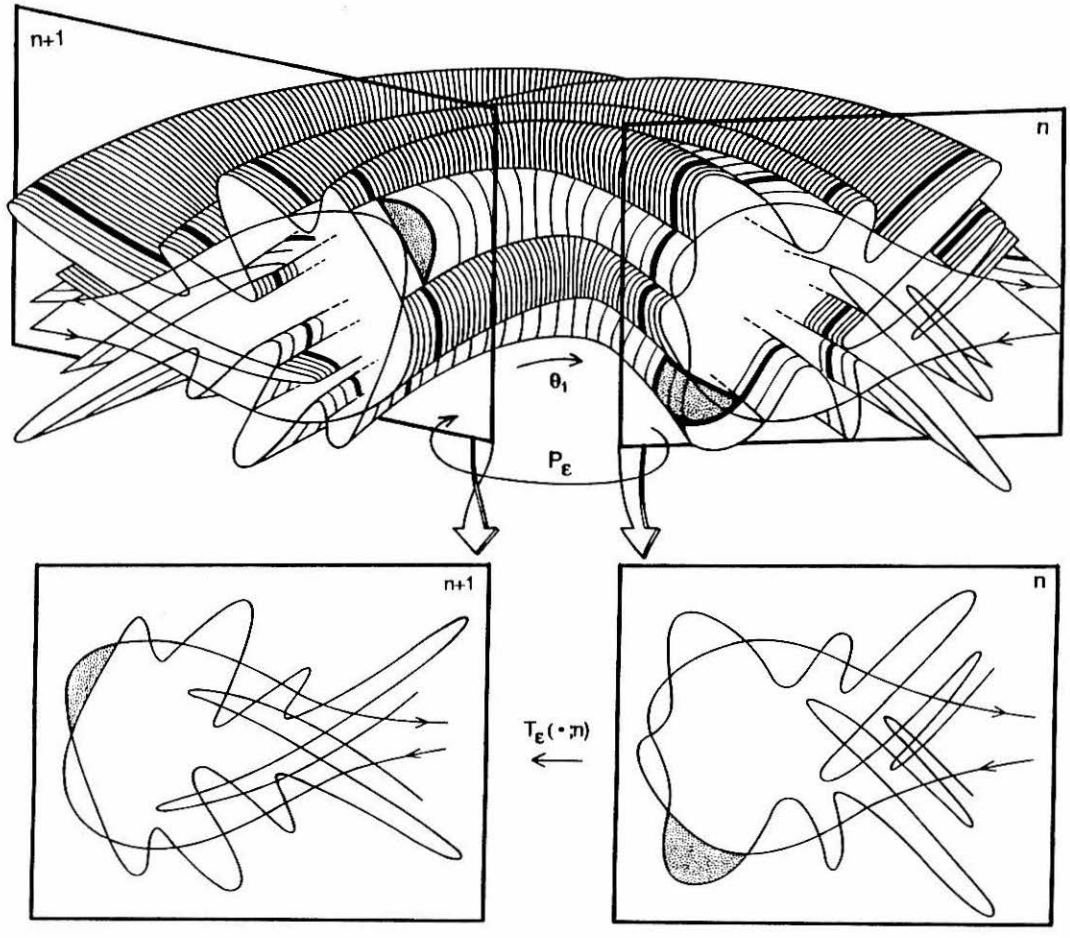


Figure 5.4.1 Two successive time slices of the invariant three-dimensional lobe structure in $\Sigma^{\theta_{20}}$ defines two successive time-dependent two-dimensional lobe structures in (x, p) space.

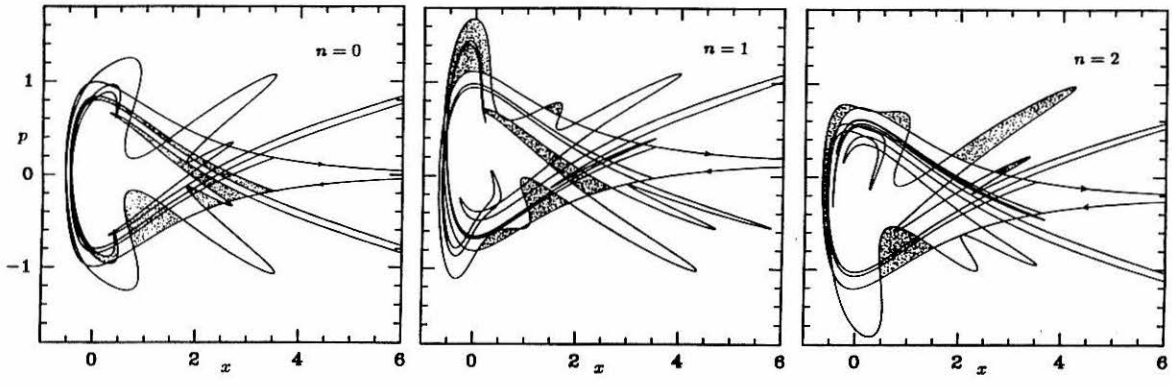


Figure 5.4.2 Two-dimensional lobes mapping within the sequence of lobe structures (the shaded lobes map in an orientation and order preserving manner). Numerical computation of the lobe structures is discussed in Section 5.5.1. For this example $(\omega_1, \omega_2) = (1, g)2.665\omega_0$, $\varepsilon d(E_1, E_2) = (0.590, 0.213)\omega_0\sqrt{2mD_0}$, and $(\theta_{1_0}, \theta_{2_0}) = (\pi, 0)$. Since we will return to this case, note that the perturbation amplitudes are chosen to give equal Melnikov amplitudes, so that the problem is similar to the one in Figure 5.3.3(d), except that here we sample at the slower frequency.

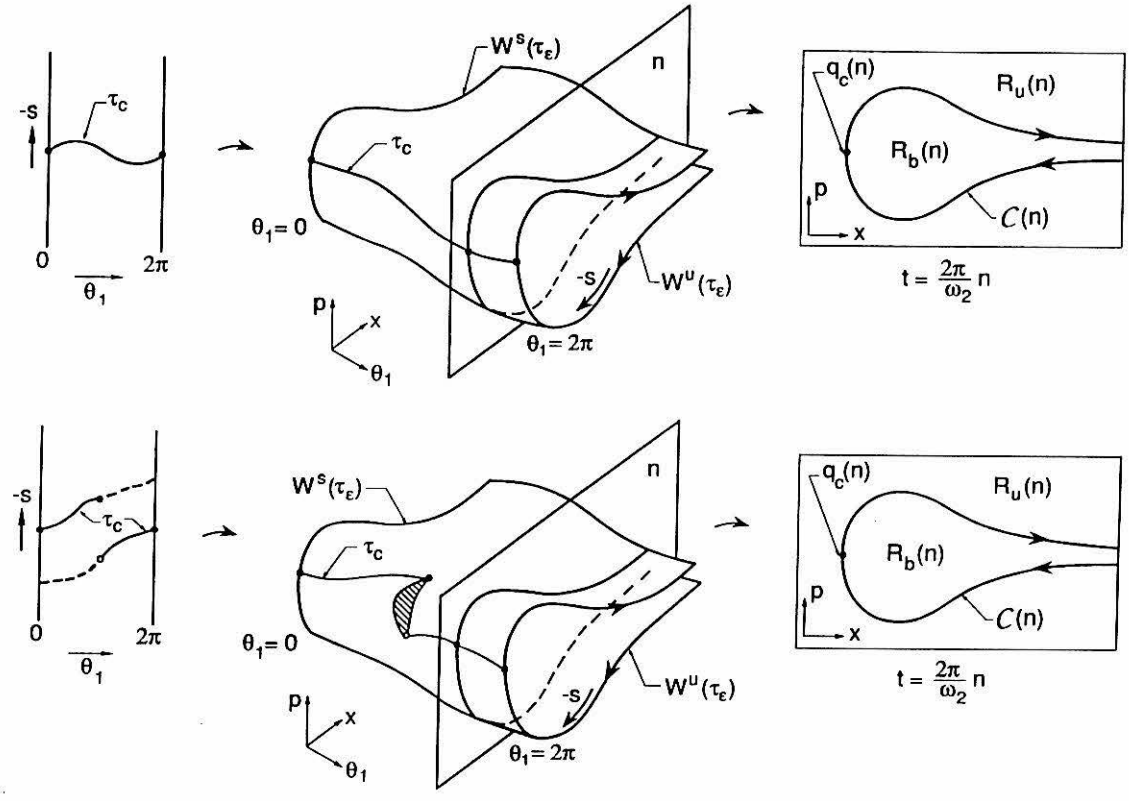


Figure 5.4.3 The invariant boundary $\mathcal{C}$ divides each phase slice into two regions: an illustration for (a) toral PIM geometry and (b) spiral PIM geometry ($q_c(n) \equiv \tau_c \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$). The shaded region shows a gap.

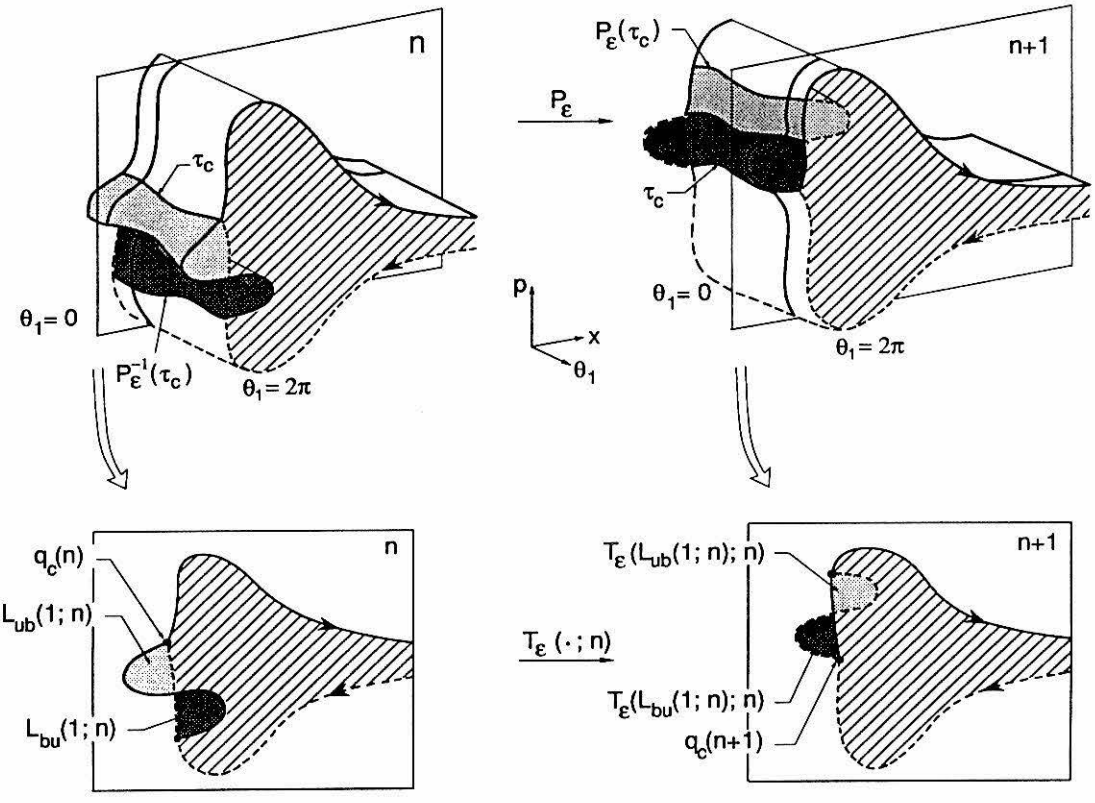


Figure 5.4.4 The turnstiles as the sole mechanism for transport between the bounded and unbounded regions.

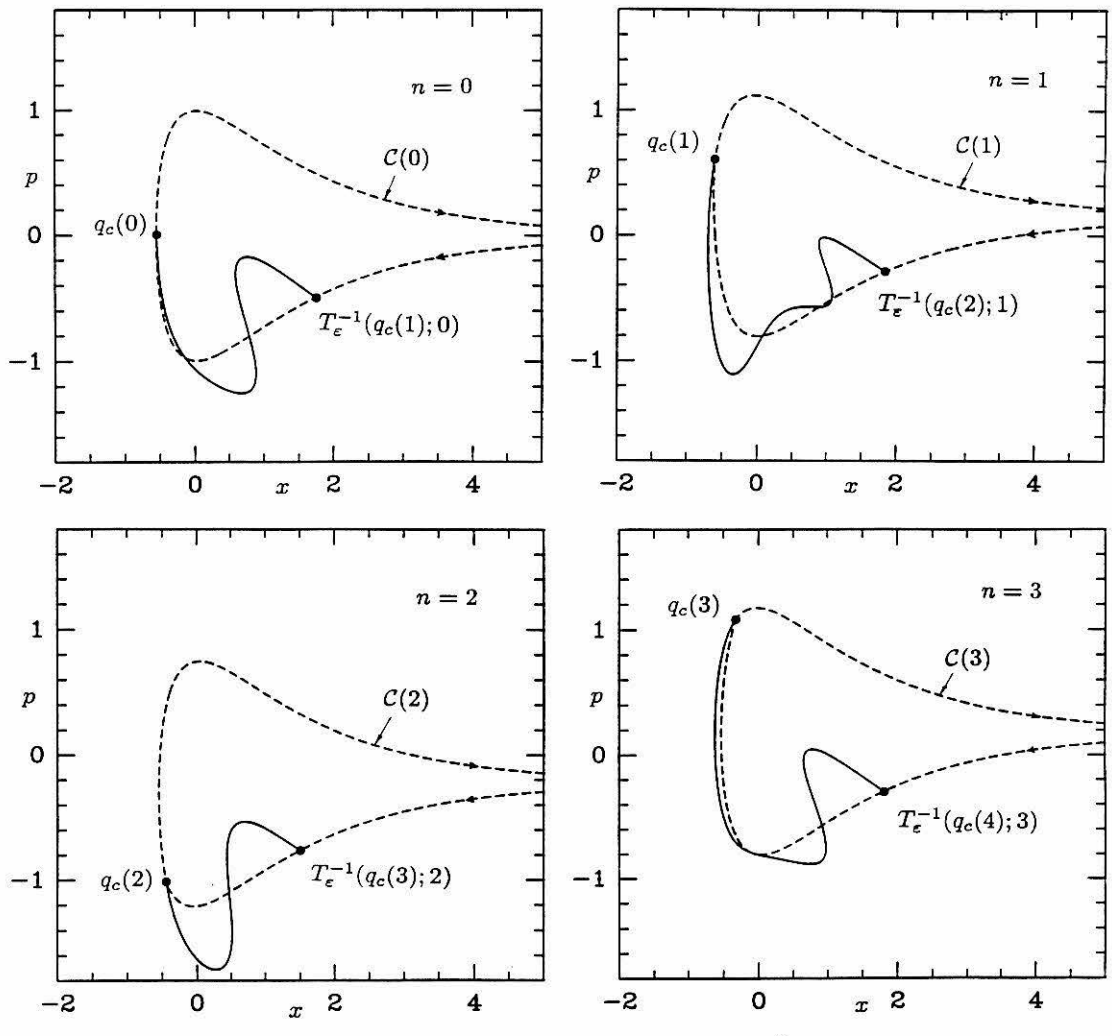


Figure 5.5.1 The turnstiles and core boundaries for a sequence of four time slices ($q_c(n) \equiv \tau_c \cap \chi(\theta_{1_0} + 2\pi\frac{\omega_1}{\omega_2}n)$ and the parameters are the same as in Figure 5.4.2).

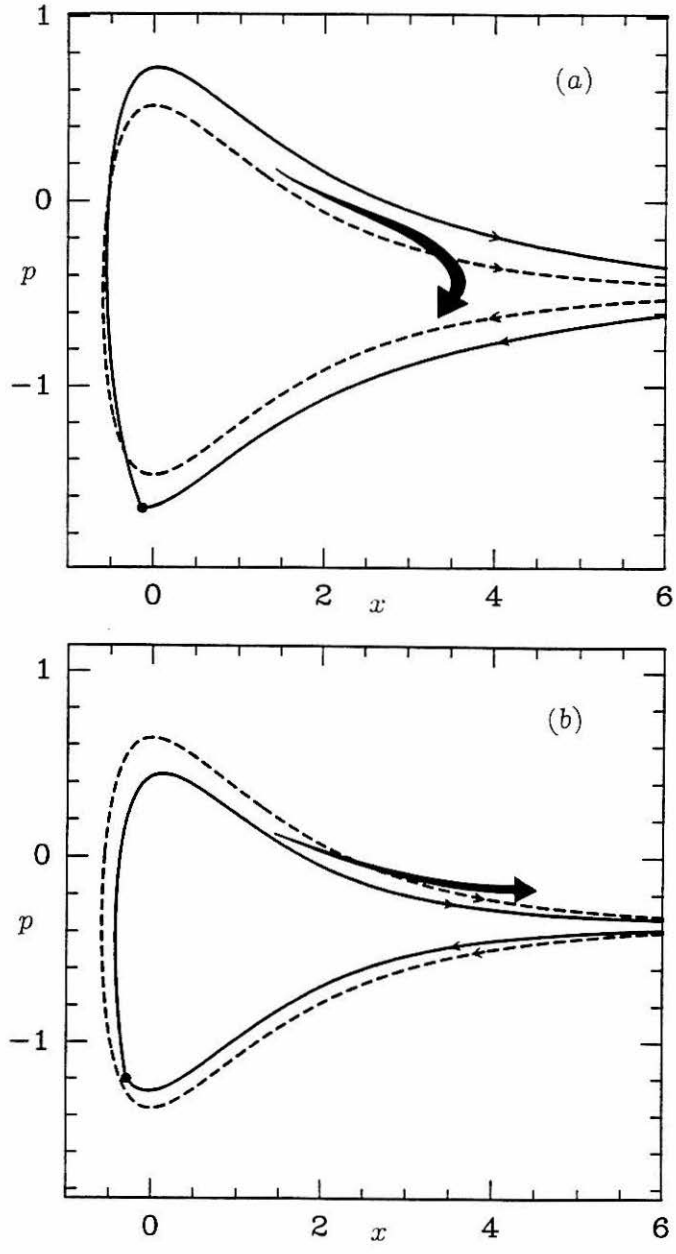


Figure 5.5.2 The dissociation criterion based on segments of the stable and unstable manifolds (the solid line) and Leopold and Percival's compensated energy (the dashed line). For this example we are in the phase slice $\chi(\theta_1 = \frac{23}{22}\pi)$ of the Poincaré section $\Sigma^{\theta_{20}=3\pi/2}$, and the parameters are: (a) $(\omega_1, \omega_2) = \omega_0(0.231, 2.618)$, $\varepsilon d(E_1, E_2) = (0.102, 1.105)\omega_0\sqrt{2mD_0}$; (b) the same as (a) except $E_1 \rightarrow -E_1$. Note that for this particular example we choose $q_c(n)$ whose s value is *second* closest to zero, which leads to a smoother core boundary in this case.

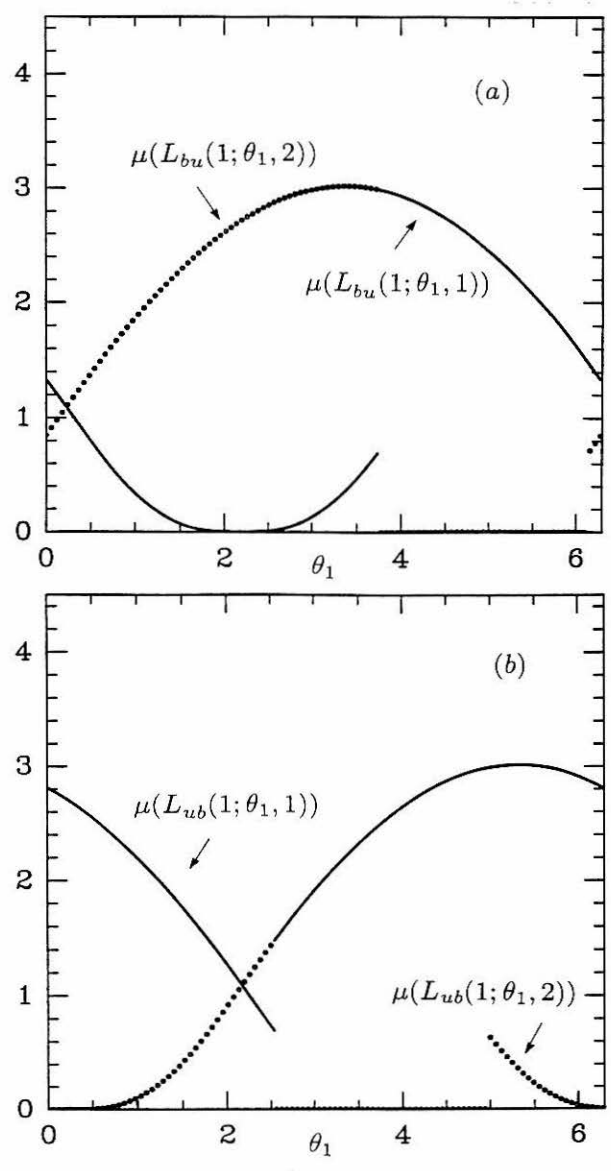


Figure 5.5.3 The variation of turnstile lobe area with the phase θ_1. In this example there are one or two of each of the escape and capture turnstile lobes, depending on the phase slice: $L_{bu}(1;\theta_1,j)$ and $L_{ub}(1;\theta_1,j)$ denote the j^{th} escape and capture turnstile lobes, respectively, in the phase slice defined by θ_1, where j increases with s. Note that different lobes correspond to different line types, which are discontinuous in θ_1. For this example we are in the Poincaré section $\Sigma^{\theta_{2_0}=0}$ with Melnikov amplitudes $(A_1, A_2) \equiv -(A, A)$, $A > 0$, and $(\omega_1, \omega_2) = (g^{-1}, 1)\omega_0$. The areas for the escape and capture lobes are given in (a) and (b), respectively, and are per unit $\varepsilon A/\omega_0$.

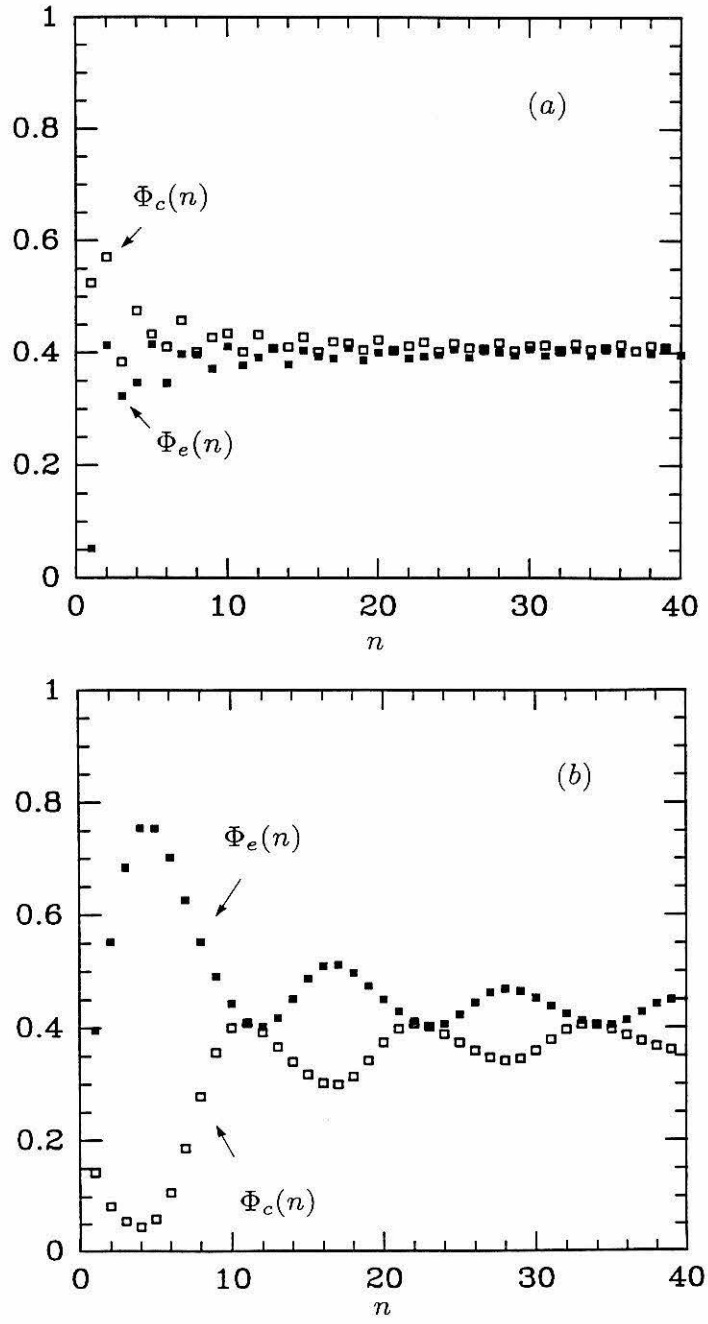


Figure 5.5.4 The variation of convergence rate for finite-time average flux. The black dots correspond to escape, the white dots to capture, and the flux is per unit εA. For this example $(A_1, A_2) \equiv -(A, A)$, $A > 0$, and (a) $(\omega_1, \omega_2) = (g, 1)\omega_0$, $(\theta_{1_0}, \theta_{2_0}) = (\pi, 0)$, (b) $(\omega_1, \omega_2) = (0.231, 2.618)\omega_0$, and $(\theta_{1_0}, \theta_{2_0}) = (\frac{23}{22}\pi, \frac{3}{2}\pi)$.

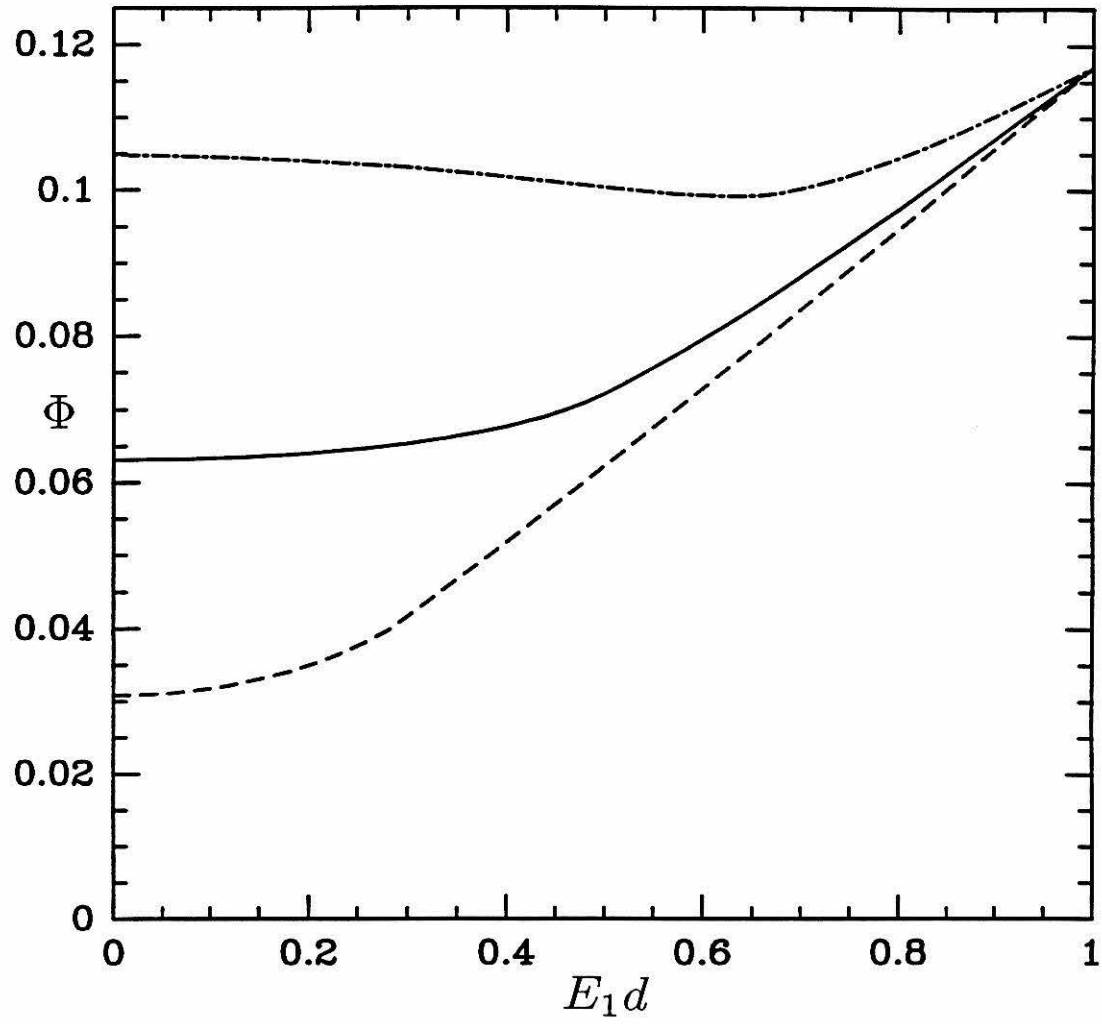


Figure 5.5.5 Infinite-time average flux as a function of $E_1 d$, with $(E_1^2 + E_2^2)d^2 \equiv E^2 d^2$. The frequency ω_1 is fixed at ω_0; the frequency ω_2 is $\frac{7}{3}\omega_0$ (dashed), ω_0/g (solid), and $\frac{10}{9}\omega_0$ (dashed-dotted). The flux is per unit $\varepsilon \frac{2\pi}{a} Ed$, the horizontal axis is per unit Ed, and $\theta_{1_0} = \theta_{2_0} = 0$.

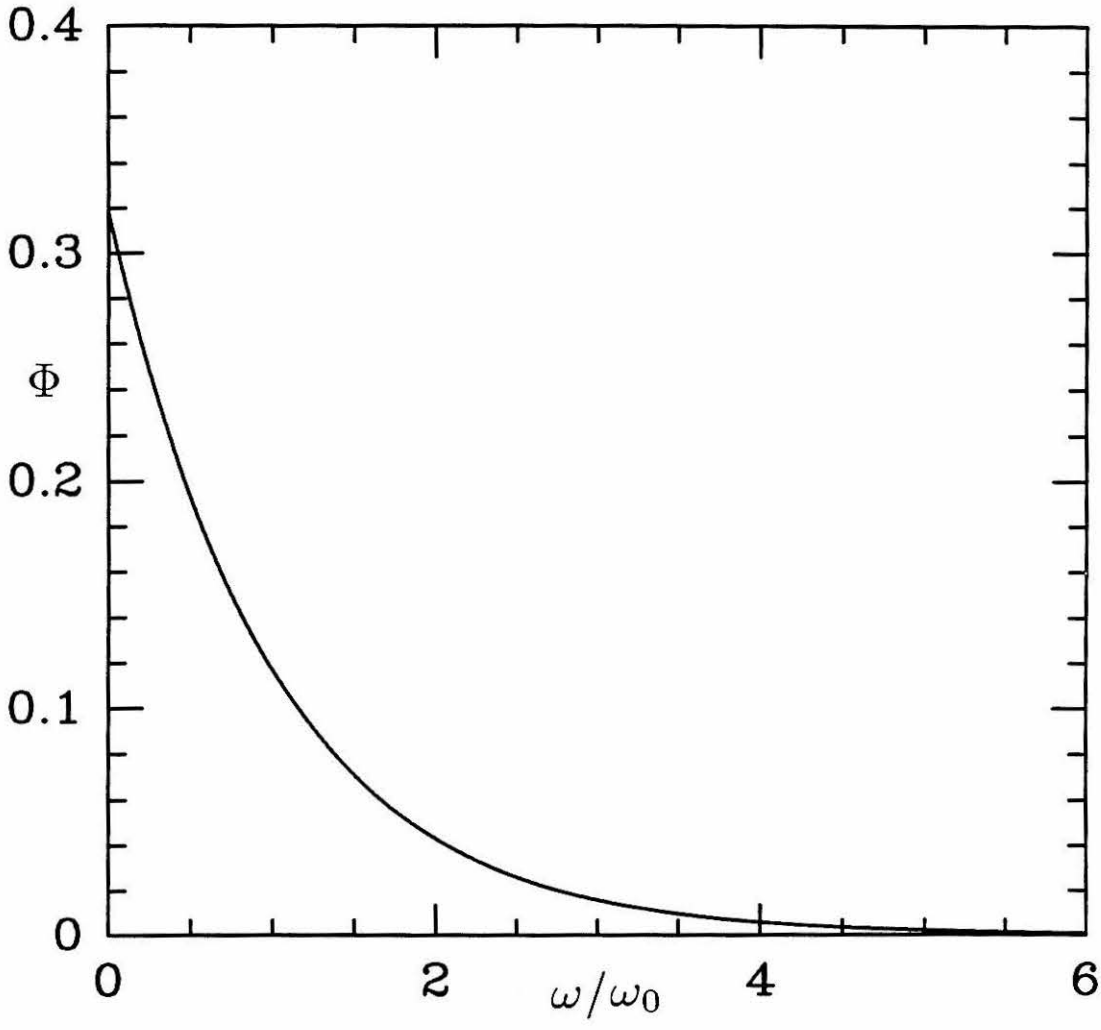


Figure 5.5.6 The single-frequency infinite-time average flux as calculated by the Melnikov approximation (5.5.4). The flux is per unit $\varepsilon \frac{2\pi}{a} Ed$.

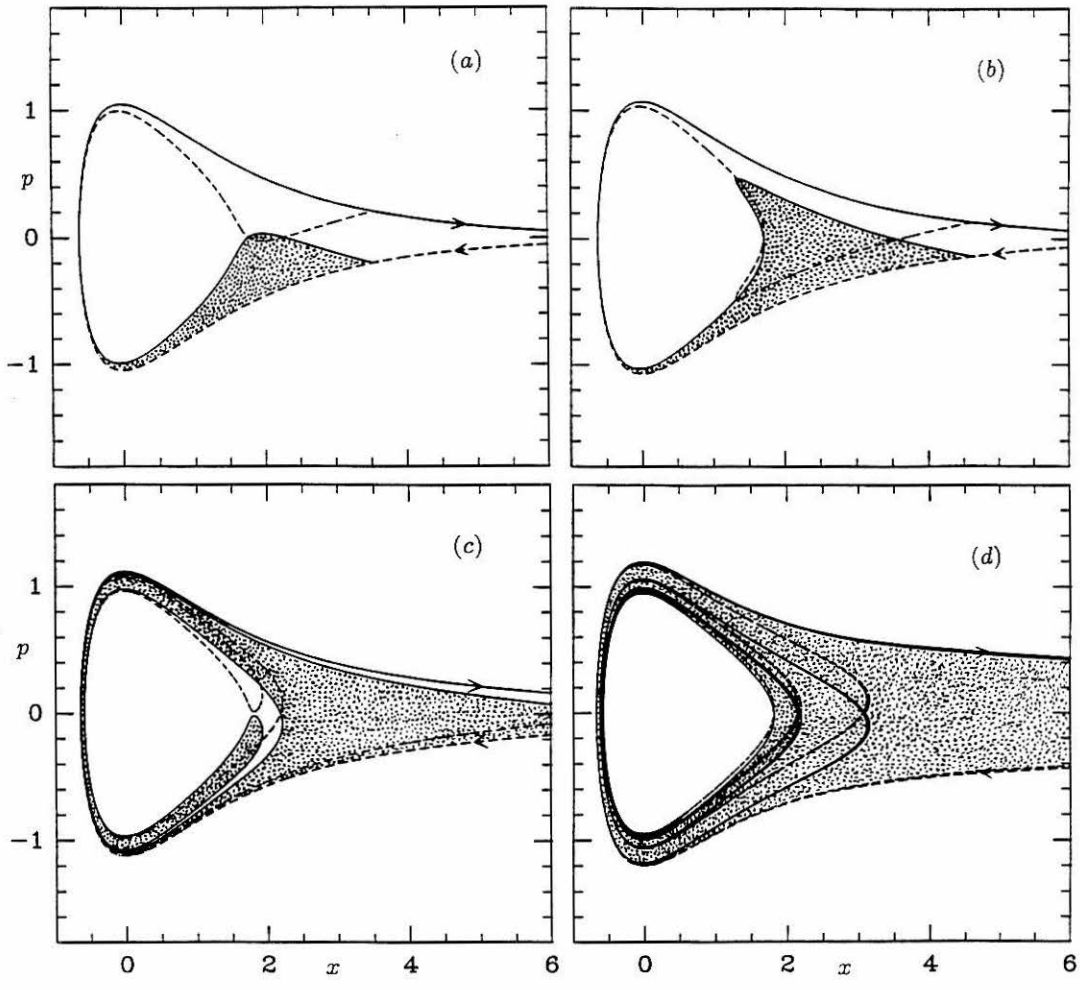


Figure 5.5.7 The single-frequency lobe structure in $\Sigma^{\theta_{10}=\pi}$ for four successively smaller forcing frequencies, with εEd fixed. The shaded lobe is the escape turnstile lobe $L_{bu}(1)$, and the white lobe is the image of the capture turnstile lobe $P_\varepsilon(L_{ub}(1))$. For these examples $\varepsilon Ed = 0.037\omega_0\sqrt{2mD_0}$, and (i) $\omega = 0.333\omega_0$, (ii) $\omega = 0.222\omega_0$, (iii) $\omega = 0.111\omega_0$, and (iv) $\omega = 0.0555\omega_0$.

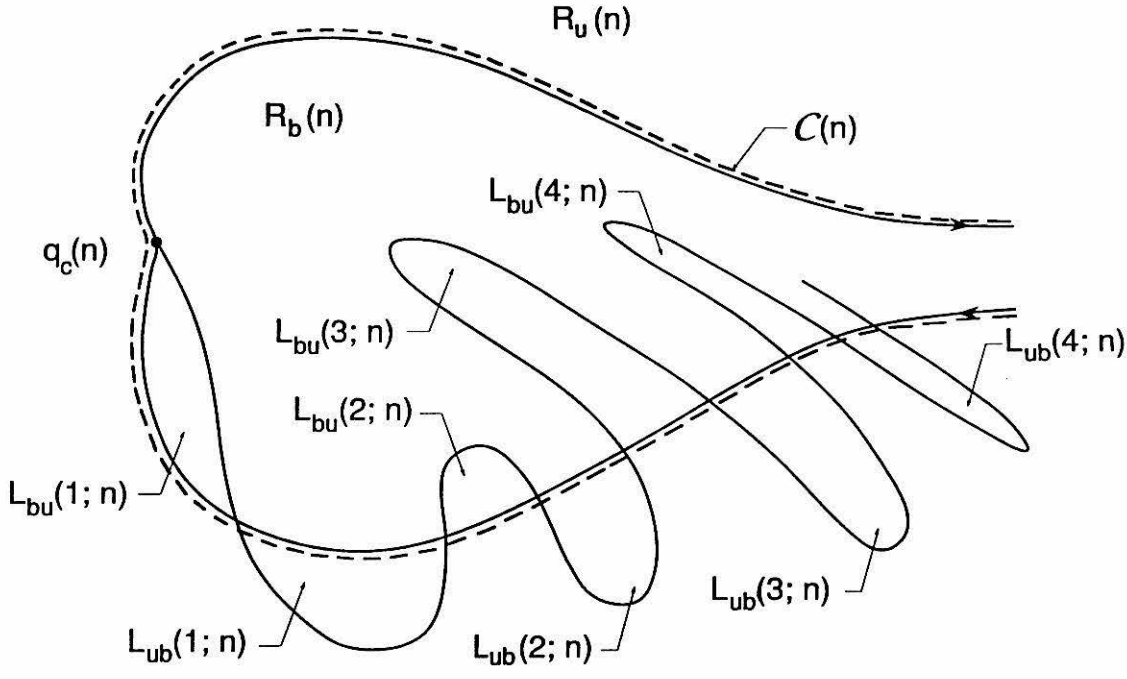


Figure 5.5.8 Labeling the other lobes in the lobe structure. For simplicity of illustration we consider a case where there is only one lobe for each $L_{bu}(m;n)$, $L_{ub}(m;n)$, and we only portray the lobes for $m \geq 1$.

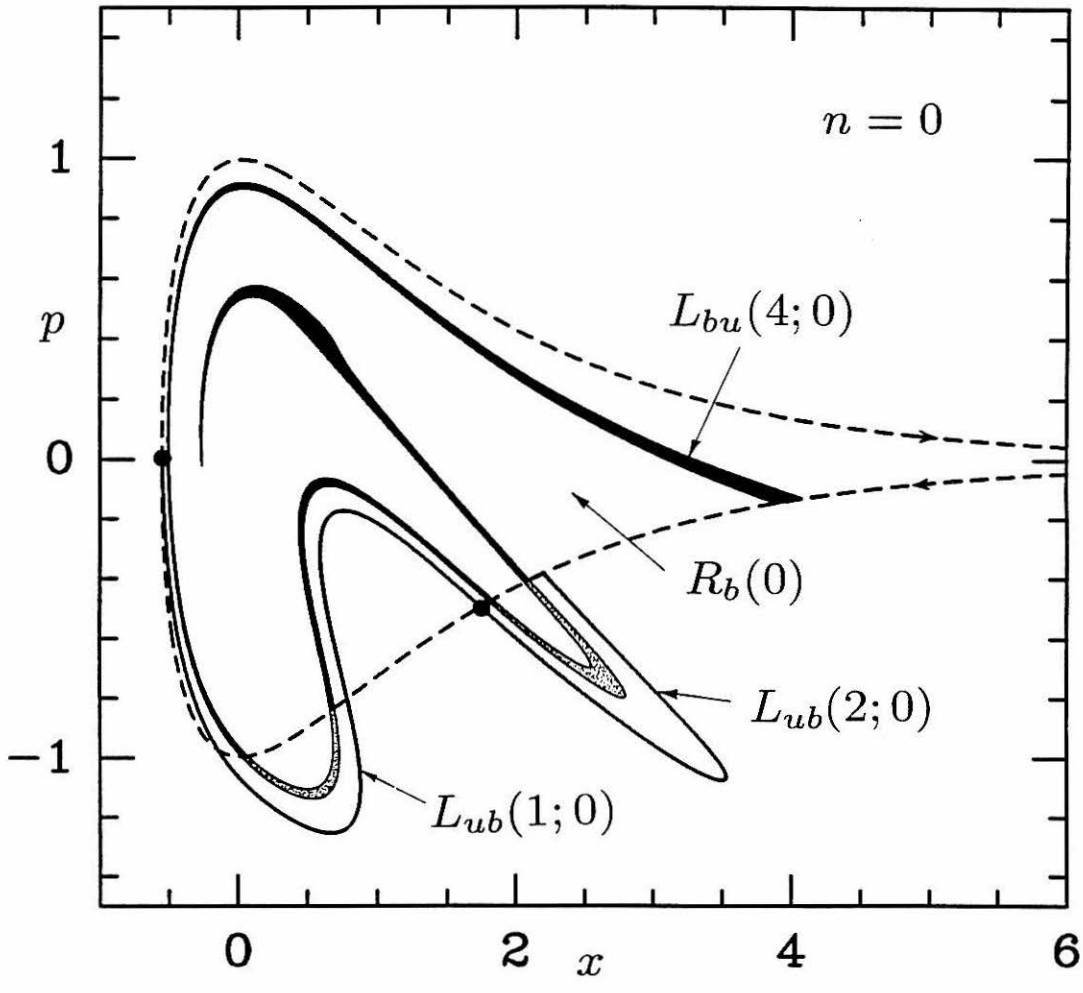


Figure 5.5.9 Finding the content of a turnstile lobe. The parameters are the same here as in Figures 5.4.2 and 5.5.1. In this particular example, we wish to determine the content of the escape turnstile lobe in the $n = 3$ time slice, $L_{bu}(1;3)$. Hence we map the lobe back to the $n = 0$ time slice (which gives the lobe $L_{bu}(4;0)$): the black part of the lobe is in $R_b(0)$, and the speckled part is outside $R_b(0)$ and inside the lobes $L_{ub}(1;0)$ and $L_{ub}(2;0)$.

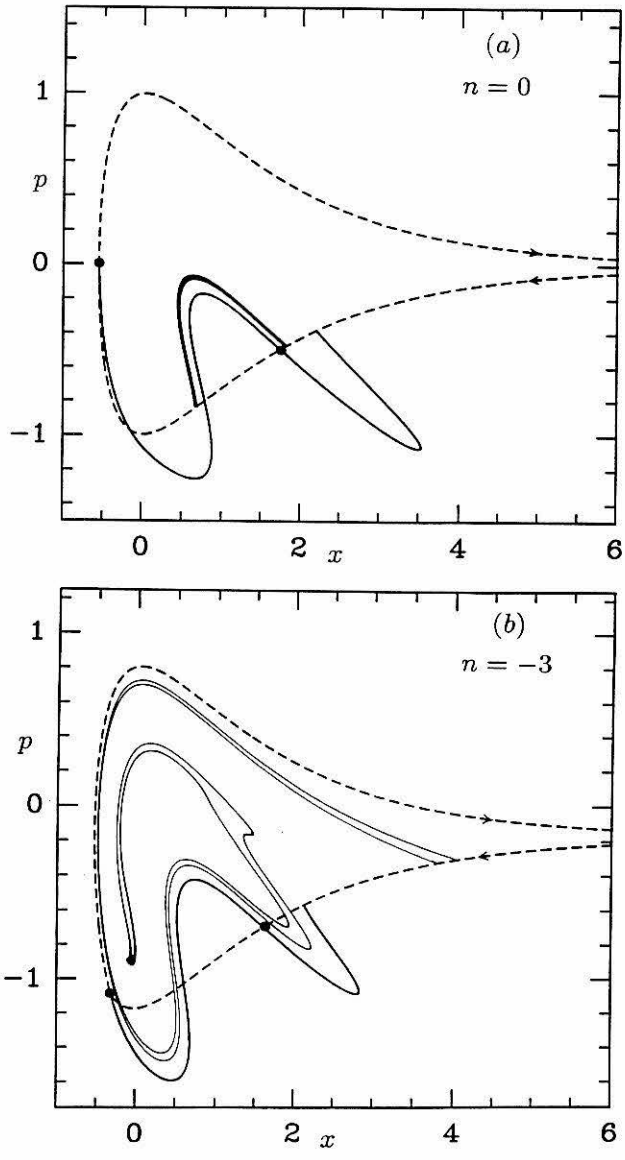


Figure 5.5.10 Illustrating the approximate self-similar behavior as one maps a lobe backwards in time. In (a) we show a piece of the lobe $L_{bu}(4;0)$ in Figure 5.5.9 that wraps around one of the $L_{bu}(1;0)$ lobes: we map this piece of lobe back another three periods to obtain (b). The resulting piece of lobe looks approximately like the boundary of the lobe $L_{bu}(1;3)$ mapped back three periods (compare with Figure 5.5.9). This is an example where the approximate self-similarity is quite good: in general, since one has lobes of varying shape wrapping around one another, the approximate self-similarity is to be understood only very loosely.

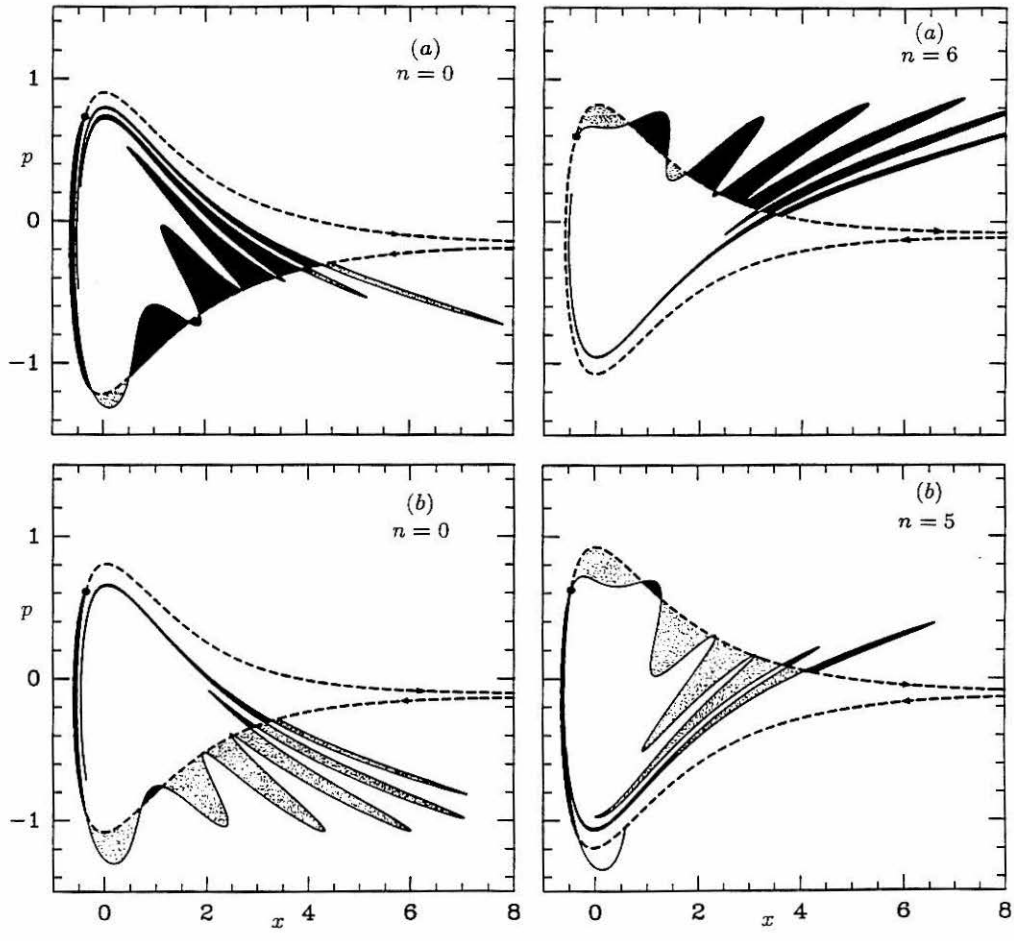


Figure 5.5.11 Using variation of lobe areas to enhance or diminish finite-time escape or capture for a fixed infinite-time average flux. The parameters for (a) and (b) are identical to those in Figure 5.5.2, except that ε is three times smaller here. The solid dot represents our choice of $q_c(n)$ for these examples, and in the $n = 5$ sample for (b), $L_{bu}(1;5)$ is the thin black lobe that touches the dot.

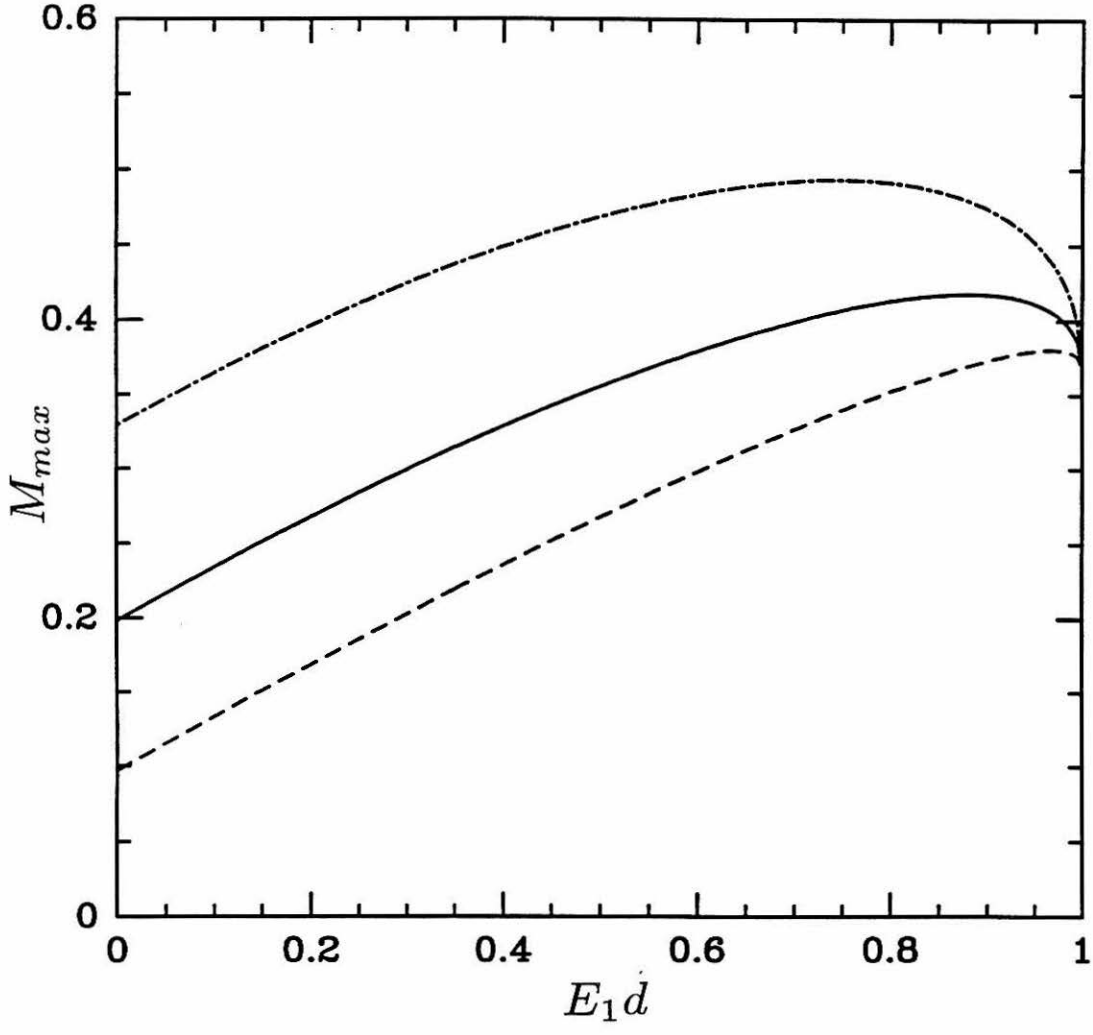


Figure 5.5.12 Plot of $M_{\max}$ as a function of $E_1 d$, with $(E_1^2 + E_2^2)d^2 \equiv E^2 d^2$. The parameters are identical to those of Figure 5.5.5. M_{max} is per unit $\frac{2\pi}{a} Ed$, and the horizontal axis is per unit Ed.

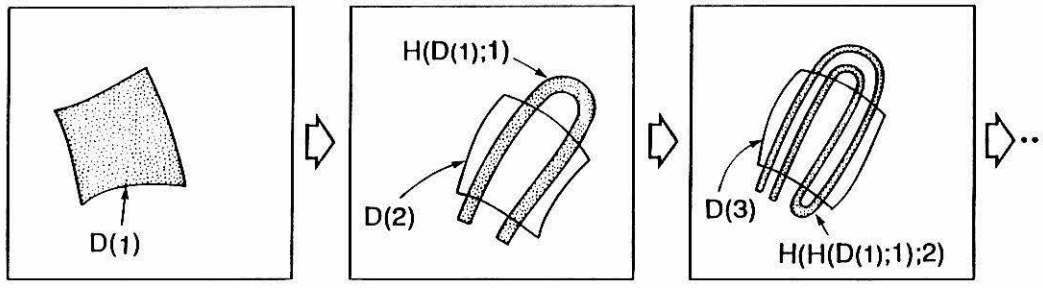


Figure 5.5.13 A traveling horseshoe map sequence.

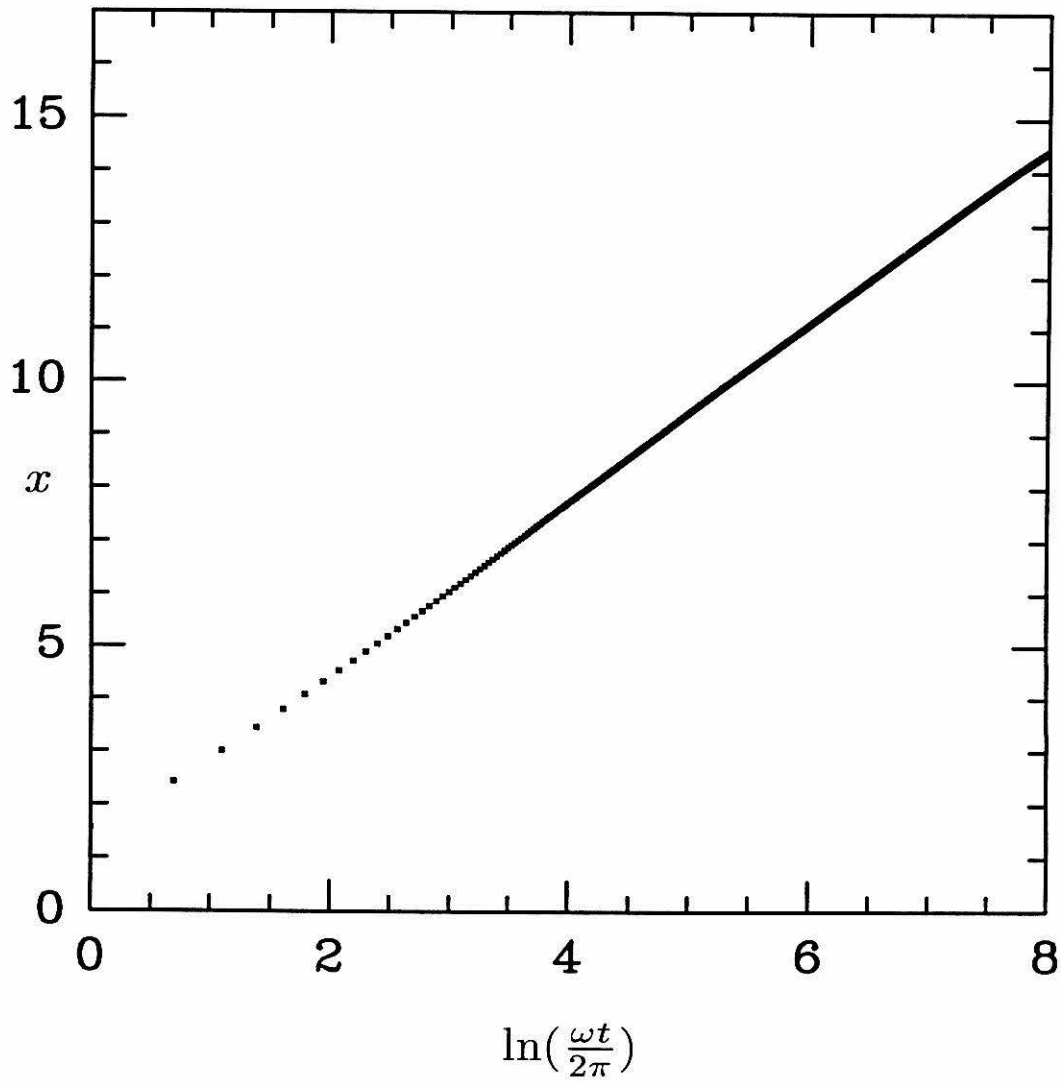


Figure 5.A.1 The x component of a point on the *perturbed* stable manifold grows logarithmically in time with successive iterates of the Poincaré map, and the slope in the above graph is $2a^{-1}$, as is the case for the unperturbed system (recall that $a = 1.1741 r_B^{-1}$ and x is given in units of r_B). Hence $x^{s,u}_{\varepsilon>0} - x^{s,u}_{\varepsilon=0}$ remains bounded in the limit $t \to \infty, -\infty$. Here for simplicity we consider a periodically forced example with $\omega = \omega_0$, $\varepsilon E d = 0.0577\omega_0\sqrt{2mD_0}$, and $\theta_{1_0} = 0$.

Chapter 6

Statistical relaxation under non-turbulent chaotic flows: non-Gaussian high-stretch tails of finite-time Lyapunov exponent distributions

Abstract

We observe that high-stretch tails of finite-time Lyapunov exponent distributions associated with interfaces evolving under a class of non-turbulent chaotic flows can range from essentially Gaussian tails to nearly exponential tails, and that the non-Gaussian deviations can have a significant effect on interfacial evolution that persists asymptotically. This range of behavior is understandable in the context of our Chapter 3 analysis of interfacial stretching in chaotic tangles, which shows the rich variety of stretch scales, spatial scales, and temporal scales associated with the stretch processes in chaotic tangles. Different scale combinations can lead to different high-stretch tails; in particular, highly non-uniform interfacial *stretch profiles* can entail significant non-Gaussian high-stretch tails. We interpret the results in the context of some elementary stretch models. That the significant non-Gaussian deviations can occur at only very small probability values of the stretch distributions encourages the use of highly resolved finite-time numerics over the usual more weakly resolved asymptotic studies, and we employ a dynamic point insertion scheme to maintain good interfacial covering and highly resolved high-stretch tails (right up to the maximum stretch value). These high-stretch statistics have relevance for incompressible flows to the mixing properties and multifractal characteristics of passive scalars and vectors in the limit of small spatial scales.

6.1 Introduction

We investigate probability distributions of finite-time Lyapunov exponents (or simply *stretch distributions*) associated with interfaces evolving within the chaotic tangles of 2D time-periodic vector fields. Such statistics are relevant, for example, to the mixing process in fluid flows, where the stretching of an interface between two species affects the rate of mixing, and to kinematic dynamos, where magnetic field amplification is intimately connected to the underlying stretch processes. Typically the interest in finite-time Lyapunov exponents is to search for and study asymptotically invariant distributions under proper scaling, as well as studying short- and medium-time behavior.[1–9] The stretch distributions of Városi and collaborators[5] are well-described by an effectively Gaussian fit, and Muzzio and collaborators[4] claim that for chaotic flows without islands the stretch distributions should be "essentially Gaussian". Finn and Ott[8] and Ott and Antonsen[9] motivate Gaussian stretch distributions with a simple model of the chaotic stretch processes, a binomial multiplicative process, which gives a binomial ("approximately Gaussian") stretch distribution. Sepúlveda and collaborators[2] and Horita and collaborators[3] report non-Gaussian behavior at the *low*-stretch tails due to the presence of KAM-tori. The prevailing opinion is generally that, except for effects due to the "stickiness" of KAM-tori, one can view stretching under chaotic flows as the product of weakly correlated stretches, leading to essentially Gaussian statistics. Though the *possibility* of non-Gaussian behavior at either tail has not been disallowed, the high-stretch tails have remained virtually unstudied, investigators typically seeming content to conclude "essentially Gaussian" statistics. Indeed, previous studies exclusively consider stretch statistics associated with a *fixed* number of points initialized upon a domain of interest, such as an area or line segment,[1–7] and such an approach is typically doomed to have poor resolution at the high-stretch tail.

Our interest is in the high-stretch statistics, which are relevant for several reasons. First, though the high-stretch tails correspond to small probability values, they correspond to *large* stretch values, and we find they can play a significant role in interfacial stretching. Second, for incompressible flows these tails correspond to the limit of small spatial scales (small striation width), and thus have direct impact on the multifractal characteristics of passive scalars in the small striation width regime. Third, there has been significant recent interest in non-Gaussian, indeed exponential, tails in the turbulence literature,[11–16] and it would be interesting to understand the tails associated with *non*-turbulent chaotic flows. We thus perform a high-resolution numerical study of these high-stretch tails by implementing a dynamic point insertion scheme to maintain good interfacial covering, and find that these tails can take on a great range of behavior, varying from essentially Gaussian to nearly exponential. This range of behavior is understandable in the context of our Chapter 3 analysis of interfacial stretching in chaotic tangles, which shows via a symbolic dynamics construction how, though one can indeed view stretching in chaotic tangles in terms of products of weakly correlated events, there is a *rich variety* of stretch scales, spatial scales, and temporal scales associated with these products. Different scale combinations can lead to different high-stretch tails; in particular, highly non-uniform interfacial *stretch profiles* can entail significantly non-Gaussian high-stretch tails. We give evidence in a *scaling-independent* context of the significance of these non-Gaussian deviations on interfacial stretching by exploiting rapid convergence of a scalar quantity associated with the interface (the exponent associated with the rate of increase of interfacial length).

This chapter is organized as follows. We re-introduce in Section 6.2 the two chaotic fluid flows studied in Chapter 3, and then perform in Section 6.3 a numerical study of the high-stretch statistics associated with interfacial stretching, implementing a dynamic point insertion scheme which maintains

good interfacial covering, allowing us to explore the chaotic tangle with a truly one-dimensional probe, rather than a collection of points. This difference in probe dimension can entail a difference in the statistics uncovered, and we wish to emphasize the relevance of probe dimension. The obvious consequence of such a one-dimensional probe is an exponentially growing number of points, which disallows a long-time experiment. Whether this is a factor depends on convergence times associated with statistical relaxation. Exploiting rapid convergence of the exponent associated with the rate of increase of interface length, we study in Section 6.4 the significance of the high-stretch behavior. In Section 6.5 we observe that the non-uniformity of the *stretch profiles* studied in Chapter 3 can entail non-Gaussian high-stretch tails, and then study elementary models of stretch processes to determine if they can capture the numerically observed statistical behavior.

6.2 Two chaotic flows

We consider interfacial stretching within 2D chaotic tangles, focusing for illustration on two well-studied oscillating vortex pair flows induced by (i) a pair of equal and opposite point vortices and (ii) a pair of identical point vortices oscillating periodically in response to a time-periodic straining field[10,17] (refer back to Chapter 3). We henceforth refer to these two flows as the open and closed flow, respectively, whose stream functions, in the comoving and corotating frames, respectively, can be written as

$$\psi_{o,c} = -\frac{\Gamma}{4\pi}\Big(\ln\big[(x-x_v)^2+(y-y_v)^2\big] \mp \ln\big[(x\mp x_v)^2+(y+y_v)^2\big]\Big) + \varepsilon xy\sin(2\pi t) + \psi_{tr}, \tag{6.1}$$

where (x_v, y_v) represents the spatial coordinates of one of the vortices in the pair, Γ is the magnitude of circulation associated with each vortex, and ψ_{tr} is the stream function associated with transforming to the comoving and corotating frame (refer back to Chapter 3). It is well understood that in these periodically oscillating vortex flows one obtains *chaotic tangles* (regions in physical

space where there is chaotic motion), more specifically *heteroclinic* and *homoclinic tangles* for the open and closed flow, respectively.[10,17] The boundaries of these tangles are defined by the global stable and unstable manifolds of hyperbolic fixed points in a 2D Poincaré section Σ, which represents the fluid at discrete sample times $t = n \in Z$ (see Figure 6.1). A particularly relevant interface is the one defined by the segment of the unstable manifold associated with one of the so-called *turnstile lobes*, as shown in Figure 6.1. The symbolic dynamics construction of Chapter 3 for the *entire* interface explicitly portrays the rich variety of *stretch* scales, *spatial* scales, and *temporal* scales associated with stretch processes in the tangles. Faced with a wide variety of stretch processes, it is by no means obvious that previous simple multifractal models, such as one in which the underlying stretch process is modeled by a binomial multiplicative process,[8,9] capture the true stretch statistics. For example, consider in Figure 6.2 some *stretch profiles* for the open and closed flow, *i.e.*, plots of stretch versus initial arclength along the interface (refer back to Chapter 3). These two profiles are markedly different as a simple consequence of different flow geometry and flow parameters (*e.g.*, the hyperbolic fixed point's unstable eigenvalue is significantly higher in the closed flow, and there are more fixed points in the open flow tangle), which determine the non-uniformity of the profiles, as described in Chapter 3. The open flow profile is, loosely speaking, evolved from a fairly even mix of scales and a more uniform stretch profile, and seems a candidate for Gaussian stretch statistics. In stark contrast, the closed flow shows a highly non-uniform stretch profile with *very good* stretching highly localized on *very small* spatial scales. Additionally, the vertical dashed lines in the figures denote points on the interface which intersect the stable manifold, and hence asymptote to a hyperbolic fixed point, so that as one gets closer to these points along the interface, the time interval associated with revolution around the homoclinic/heteroclinic core goes to infinity. It seems plausible that the stretch statistics are qualitatively different for the

two flows, motivating a numerical study.

6.3 Distributions of finite-time Lyapunov exponents

Let $P(\lambda(n);n)$ denote the probability distribution at the n^{th} cycle ($t = n \in Z$) of the finite-time Lyapunov exponent $\lambda(n) \equiv \ln(\delta s(n)/\delta s(0))/n$, where $\delta s(n)$ denotes the length at the n^{th} cycle of an infinitesimal line segment that originates along our interface. Studies of finite-time Lyapunov exponent distributions that monitor a fixed number of particles would approximate $P(\lambda(n);n)$ as the percentage of particles per bin width $\delta\lambda$ at a given $\lambda(n)$ bin. In contrast, we employ a point insertion scheme that maintains the covering of our interface to within a certain width (chosen to be about one percent of the mean vortex separation). When we insert a new point along the interface, its associated stretch history is interpolated from the two neighboring points (accurate for a dense enough interfacial grid), and the initial arclengths of the line segments associated with the inserted and neighboring points are adjusted. The result at the n^{th} cycle is thus a partitioning of the interface, of total length $s_{tot}(n)$, into a set of *small* line elements $\{\delta s_i(n) \mid \sum_i \delta s_i(n) = s_{tot}(n)\}$, each of which has a specified *stretch* $\delta s_i(n)/\delta s_i(0)$ and initial arclength $\delta s_i(0)$. The stretch and initial arclength can vary greatly along the interface; hence, an improved approximation of $P(\lambda(n);n)$ is given by the *percentage of the interface's initial arclength per bin width* $\delta\lambda$ *in the* $\lambda(n)$ *bin*:

$$P(\lambda(n);n) = \frac{1}{\delta\lambda} \sum_{\substack{i \\ \lambda_i(n)\in\lambda(n)\text{ bin}}} \frac{\delta s_i(0)}{s_{tot}(0)}, \tag{6.2}$$

where $\lambda_i(n)$ is the finite-time Lyapunov exponent associated with the i^{th} line element.

We find for our class of chaotic tangles that high-stretch tails of the stretch distributions can range from essentially Gaussian to nearly exponential. This

range of behavior is illustrated by plotting the distributions on a logarithmic scale and then scaling the vertical axis with division by n, which exploits our high resolution calculation at the high-stretch tail (see Figure 6.3). The open flow tail, which corresponds to a more uniform stretch profile, is by $n = 10$ essentially Gaussian. The closed flow tail, which corresponds to a more non-uniform stretch profile with extremely good stretching highly localized in small spatial scales, appears by $n = 10$ to be almost exponential. With increasing n the closed flow tail, in addition to smoothing out, appears to be flattening and extending downwards and to the right. This transient behavior is easily understood by the fact that it takes a few cycles for the first few portions of the interface to approach the region local to the hyperbolic fixed point; hence with increasing n the line segments in this local region will have spent a larger percentage of time there and their associated $\lambda(n)$ increases. The rightmost point in the distribution is thus asymptoting to $\lambda \simeq 5$ associated with the linearized flow local to the hyperbolic fixed point. By $n = 10$ the transient behavior of the scaled distributions has decayed considerably, so that, for example, the $n = 9$ and $n = 10$ distributions lie essentially on top of one another. One can thus confidently claim in the closed flow example significant non-Gaussian behavior on short and medium time scales, and it seems plausible this behavior persists asymptotically under the present scaling. Though one might be tempted to draw a definitive conclusion about the asymptotic distribution, we avoid this temptation for two basic reasons. First, an appearance of convergence can be deceptive in the context of stretch distributions, for the distributions can vary slowly over long time scales; for example, because all particles on the open flow interface (except those which intersect the stable manifold) asymptote infinitely far away from the hyperbolic fixed points, the maximum of the *open* flow distribution asymptotes *very slowly* to $\lambda = 0$, and asymmetries in the distribution can slowly set in. Hence, comparing a few successive iterates for convergence can be meaningless in the context of dis-

tributions. Second, asymptotic results can depend on choice of scaling. Our scaling via contraction of the vertical scale (dividing by n) is consistent with that of previous investigators[1,5] (we neglect a small transient term that some choose to keep in the scaling). There are of course other possible scalings, however, such as the one employed in the central limit theorem, where one expands the horizontal scale $\lambda(n) \to (\lambda(n) - \mu(n))\sqrt{n}/\sigma(n)$, with $\mu(n)$ the mean and $\sigma(n)$ the standard deviation of the distribution. Different scalings can give different asymptotic results; for example, a binomial distribution asymptotes to a Gaussian over any finite interval of its domain under the above horizontal scaling, but *not* under the previous vertical scaling. This horizontal scaling result (in keeping with the central limit theorem) is caused by pushing any tail deviations out to infinity, a wise scaling if one wishes to ignore vanishingly small probability values. However, in the context of stretching, it is not obvious that we wish to scale these deviations away, since small probability values are associated with high stretch values. Hence the vertical scaling is useful, and we are additionally motivated to study the effects of tail deviations in a *scaling-independent* framework, which is the focus of the next section.

6.4 Asymptotic result

If a small portion of a physical ensemble has a profound effect on the evolution of the ensemble, one should presumably be interested in this small portion. The closed flow non-Gaussian deviations associated with the high-stretch tail indeed correspond to a small percentage of the initial interface arclength, as is clear in Figure 6.4. Hence we ask: can these non-Gaussian deviations have a significant effect on interfacial evolution? This question can be studied in a simple setting that exploits rapid convergence of a single scalar quantity, the exponent associated with the growth rate of the total length of the interface. Let $\ell_a(n)$ denote the total length of the *actual interface*, and $\ell_g(n)$ denote the total length of the *interface whose stretch statistics are*

described by the corresponding Gaussian approximation. A measure of the relevance of the non-Gaussian deviations is thus the length ratio $\ell_g(n)/\ell_a(n)$. If this ratio is significantly less than one, then the non-Gaussian deviations have a significant effect on the interface. For the open flow interface the values of $\ln(\ell_a(n))/n$ and $\ln(\ell_g(n))/n$ are essentially equal over the short-time calculation (which went to $n = 11$), as one would expect from Figure 6.3(a); hence, one cannot easily conclude that any non-Gaussian deviations are significant. For the closed flow, however, Figures 6.5(a) and (b) indicate for $\ell_a(n)$ and $\ell_g(n)$ fairly rapid convergence with increasing time to the relations

$$\begin{aligned} \frac{\ln(\ell_a(n))}{n} &\longrightarrow 0.87 \\ \frac{\ln(\ell_g(n))}{n} &\longrightarrow 0.65. \end{aligned} \tag{6.3}$$

This would indicate rapid convergence of the length ratio to a time dependence

$$\frac{\ell_g(n)}{\ell_a(n)} \sim e^{-0.22n}. \tag{6.4}$$

Hence it appears that the length ratio asymptotes to zero, giving strong indication of the relevance of the non-Gaussian tail in the evolution of the closed flow interface. *This observation indicates the need for highly-resolved finite-time numerical studies of statistical distributions; poorly resolved asymptotic studies may miss relevant features of the statistics.*

6.5 Models

Given a numerical observation of non-Gaussian high-stretch tails, and an indication of their significance, one would like to have a better appreciation of the range of high-stretch statistics. A key feature in determining the statistics is seen to be the non-uniformity of the perturbed stretch profiles, which in turn is related to the non-uniformity of the corresponding *unperturbed* stretch profiles, as described in Chapter 3, and portrayed again in Figure 6.6. As the

unperturbed profiles become more non-uniform, good stretching is weighted more towards the *endpoints* of the profiles (see Figure 6.7). As a result, the perturbed profiles become more non-uniform, with good stretching weighted more towards the *secondary intersection points (SIP's)* of the profiles (refer back to the examples in Chapter 3), entailing progressively better stretching associated with progressively smaller spatial scales (smaller initial arclength of the interface), and progressively longer time intervals for revolution around the homoclinic/heteroclinic core.

One approach to understanding the stretch statistics exhibited by interfaces evolving in chaotic tangles is to search for elementary models for the stretch processes that capture the observed statistical behavior, the goal being to obtain the simplest model possible which respects the essential features of the actual stretch processes. We consider here some elementary models for stretch processes under chaotic flows consisting of iterative processes acting on the unit interval [0,1]. The domain represents an interface, parametrized by initial arclength, and the value defined over this domain represents the stretch experienced by the interface, taken initially to be one throughout the domain for all the models. The simplest model, motivated by the horseshoe map, would have a *single* stretch p acting over the entire unit interval. The result would be a finite-time Lyapunov exponent equal to $\ln(p)$ defined over the entire domain, giving for a stretch distribution a delta function centered at $\lambda(n) = \ln(p)$. Such a model is obviously inadequate as a consequence of containing only a single stretch, so we proceed to a more complicated model that has *two* stretches, entailing a *binomial multiplicative process* (see Figure 6.8). Partition the unit interval into two equal halves and multiply by p on one half, $q < p$ on the other half. Repeat this process to each half, then to each resulting quarter, and so on. This procedure is a canonical construction of a *multifractal*, which is heuristically described as a set of fractals, each with its own weighting and dimension (see, for example, Feder[18] for an introductory

account). The result at the n^{th} step is a set of stretches

$$\{p^i q^{n-i}; i \in [0,n]\} \tag{6.5}$$

with corresponding domain widths given by the binomial distribution

$$\left\{\frac{n!}{i!(n-i)!}\frac{1}{2^n}; i \in [0,n]\right\}. \tag{6.6}$$

These expressions give a set of finite-time Lyapunov exponents

$$\left\{\lambda_i(n) = \frac{i}{n}\ln(p) + (1-\frac{i}{n})\ln(q); i \in [0,n]\right\} \tag{6.7}$$

with corresponding asymptotically invariant scaled probability distribution in the large n limit

$$\frac{\ln(P(\lambda(n);n))}{n} = -\{\xi\ln(\xi) + (1-\xi)\ln(1-\xi) + \ln(2)\}, \tag{6.8}$$

where

$$\xi \equiv \frac{\lambda(n) - \lambda(n)_{min}}{\lambda(n)_{max} - \lambda(n)_{min}}, \tag{6.9}$$

as follows from Stirling's formula and sending the transients to zero. The Gaussian fit to this distribution, determined by requiring the same curvature at the maximum, is

$$\frac{\ln(P_{gauss}(\lambda(n);n))}{n} = -2(\xi - \frac{1}{2})^2. \tag{6.10}$$

The resulting stretch distribution for any $p \neq q$ is shown in Figure 6.9. The model obviously lacks the freedom to capture the range of high-stretch tails found earlier. For example, there is no way to get a nearly exponential tail out of this model in a limiting regime. This model appears inadequate since, although it contains two *stretch* scales, it contains only one *spatial* scale, thus not respecting the possible occurrence of very good stretching distributed over very small spatial scales.

An improved model might be then to make the division of the domain and successive subdomains weighted by some ratio $\beta : (1-\beta)$, $0 < \beta < 1$, as shown in Figure 6.10. The result at the n^{th} step is then a set of stretches

$$\{p^i q^{n-i}; i \in [0,n]\} \tag{6.11}$$

with corresponding domain widths

$$\Big\{\frac{n!}{i!(n-i)!}\beta^i(1-\beta)^{n-i}; i \in [0,n]\Big\}. \tag{6.12}$$

For example, one might take $\beta/(1-\beta) \propto q/p$, so that as the disparity in the two stretches increases, the disparity in spatial scales increases, and in particular better stretching is associated with smaller spatial scales. Equation (6.11) again gives a set of finite-time Lyapunov exponents

$$\Big\{\lambda_i(n) = \frac{i}{n}\ln(p) + (1-\frac{i}{n})\ln(q); i \in [0,n]\Big\},$$

with corresponding asymptotically invariant scaled probability distributions in the large n limit given by

$$\frac{\ln(P(\lambda(n);n))}{n} = -\{\xi\ln(\xi)+(1-\xi)\ln(1-\xi)-\xi\ln(\beta)-(1-\xi)\ln(1-\beta)\}, \tag{6.13}$$

again from using Stirling's formula and sending transients to zero, and where ξ is again given by equation (6.9). The Gaussian approximation, using the same prescription as in the previous binomial model, is

$$\frac{\ln(P_{gauss}(\lambda(n);n))}{n} = -\frac{1}{2}\Big\{\frac{1}{\beta}+\frac{1}{1-\beta}\Big\}(\xi-\beta)^2. \tag{6.14}$$

The resulting stretch distributions are shown in Figure 6.11 for decreasing values of β. On the one hand, the results look more promising, for in the limit $\beta \to 0$ one recovers an exponential tail. Hence by varying β from 1/2 to 0, one can vary the high-stretch tail from nearly Gaussian to nearly exponential. In the limit $\beta << 1$, equation (6.13) gives

$$\frac{\ln(P(\lambda(n);n))}{n} \approx \xi\ln(\beta), \tag{6.15}$$

valid for $\xi > \xi_{min}$, where ξ_{min} is some minimum value that vanishes in the limit $\beta \to 0$. If we take $\beta \sim p^{-1}$ in the limit $\beta << 1$, a reasonable choice for a closed incompressible flow and consistent with the numerical observation of good stretching localized on small spatial scales, then $\lambda(n) \approx -\xi \ln(\beta)$ (again for $\xi > \xi_{min}$, where $\xi_{min} \to 0$ as $\beta \to 0$) and equation (6.15) gives

$$\frac{\ln(P(\lambda(n);n))}{n} \approx -\lambda(n), \tag{6.16}$$

an exponential tail with slope minus one. Clearly, however, this weighted binomial multiplicative process does not capture the numerically observed statistics, since it approaches an exponential high-stretch tail at the expense of the domain of the Gaussian hump, and for no parameter values do we find a nearly exponential tail with a Gaussian hump defined over a sizable domain, *i.e.*, we can find nothing resembling the closed flow distribution.

We thus again search for an improved model, and a *trinomial multiplicative process* (with weighted divisions of the domain) seems promising, the rationale being that as we let one of the three stretches become very large, we will have *two* other stretches to help us control the Gaussian hump. An example of this process is shown in Figure 6.12. In an effort to capture the closed flow statistics, we will take p and q to be reasonably small ($O(1)$) with r much greater ($r >> 1$). We take the width of the r strip to be r^{-1}, and the widths of the p, q strips to each be $(1 - r^{-1})/2$. This process then attempts to model a highly non-uniform stretch process, with good stretching highly localized on very small spatial scales. The result at the n^{th} step is then a set of stretches

$$\{p^i q^j r^{n-i-j}; i, j \in [0, n], i + j \leq n\} \tag{6.17}$$

with corresponding domain widths

$$\left\{\frac{n!}{i!j!(n-i-j)!}\frac{(1-r^{-1})^{i+j}}{2^{i+j}}\frac{1}{r^{n-i-j}}; i, j \in [0, n], i + j \leq n\right\}. \tag{6.18}$$

The resulting set of finite-time Lyapunov exponents is

$$\left\{\lambda_{i,j}(n) = \frac{i}{n}\ln(p)+\frac{j}{n}\ln(q)+\left(1-\frac{i}{n}-\frac{j}{n}\right)\ln(r); i,j \in [0,n], i+j \leq n\right\}. \quad (6.19)$$

Since this set depends on two independent variables i and j, we need an intermediate step to get the asymptotically invariant probability distribution in the large n limit. We first write

$$\begin{aligned}\frac{\ln(P(\lambda_{i,j}(n);n))}{n} = -\Big\{&\frac{i}{n}\ln\Big(\frac{i}{n}\Big) + \frac{j}{n}\ln\Big(\frac{j}{n}\Big) + \Big(1-\frac{i}{n}-\frac{j}{n}\Big)\ln\Big(1-\frac{i}{n}-\frac{j}{n}\Big)\\ &+\frac{i}{n}\ln\Big(\frac{2}{(1-r^{-1})}\Big) + \frac{j}{n}\ln\Big(\frac{2}{(1-r^{-1})}\Big)\\ &+\Big(1-\frac{i}{n}-\frac{j}{n}\Big)\ln(r)\Big\},\end{aligned} \quad (6.20)$$

followed by

$$P(\lambda(n);n)\Delta\lambda(n) = \sum_{\substack{i,j \\ \lambda_{i,j}(n)\in\lambda(n)\ \text{bin}}} P(\lambda_{i,j}(n);n)\Delta\lambda_{i,j}(n). \quad (6.21)$$

Figure 6.13 shows the scaled probability distribution determined by the above prescription. Figure 6.13(a) is the result of an effort to obtain ballpark agreement with the closed flow distribution. For example, we take $r \approx \exp(\lambda_u)$, where λ_u is the unstable eigenvalue of the closed flow hyperbolic fixed point at the origin, and p, q are chosen with the aim of obtaining a similar Gaussian hump. For this still rather crude model we do not try to account for contraction, however, and simply deal with three stretches. At first glance the distribution compares well with that of the closed flow. The curvature of the high-stretch tails are nevertheless still significantly different, as highlighted in Figure 6.14. As in the weighted binomial process, we can make the high stretch tail of the trinomial model arbitrarily close to exponential, but at the expense of either

(i) making the width of the Gaussian hump smaller, by decreasing $(p-q)/r$,

or

(ii) having the peak of the Gaussian hump occur at a smaller $\lambda(n)$, by decreasing $(p+q)/r$.

An example of (i) is shown in Figure 6.13(b) - by making $(p-q)/r$ smaller, the high-stretch tail is slightly more exponential, but at the expense of a much skinnier Gaussian hump; indeed, non-Gaussian deviations exist for significant probability values, which is why we do not try to approximate the distribution with a Gaussian. For the weighted trinomial stretch process, then, one can choose parameter values to obtain approximate agreement with the closed flow distribution, but it appears as though small but fundamental differences remain. There are several avenues one may pursue with more complicated models having multiple stretch, spatial, and even temporal scales. Such avenues are best pursued in the context of a more systematic numerical study of the stretch models (to obtain a precise idea of what needs to be modeled) and more serious effort at constructing models that adequately respect the fairly complicated underlying symbolic dynamics. We thus end our study of stretch models here, the primary purpose of the chapter being to point out the range of high-stretch statistics afforded by the fairly complicated stretch processes found in chaotic tangles, the significance of these high-stretch tails, and the need for high-resolution numerical studies of stretch statistics.

6.6 Concluding remarks

There is a rich variety of stretch, spatial, and temporal scales associated with stretch processes in chaotic tangles, and this entails finite-time Lyapunov exponent distributions that show great variation in the nature of the high-stretch tails, ranging from essentially Gaussian to nearly exponential. In particular, highly non-uniform interfacial stretch profiles (that contain for example very good stretching defined over very small spatial scales) can entail significantly non-Gaussian high-stretch tails, that have a significant effect on

the evolution of the system, such as the rate of interfacial elongation, and have direct impact on the multifractal characteristics of striation widths in the limit of small spatial scales. Since these deviations may be apparent only at very small probability values, high-resolution studies of stretch statistics are motivated.

References

[1] P. Grassberger, R. Baddi, and A. Politi, J. Stat. Phys. **51**, 135 (1988).

[2] M.A. Sepúlveda, R. Badii, and E. Pollak, Phys. Rev. Lett. **63**, 1226 (1989).

[3] T. Horita, H. Hata, R. Ishizaki, and H. Mori, Prog. Theor. Phys. **83**, 1065 (1990).

[4] F.J. Muzzio, P.D. Swanson, and J.M. Ottino, Phys. Fluids A **3**, 822 (1991).

[5] F. Városi, T.M. Antonsen, Jr., and E. Ott, Phys. Fluids A **3**, 1017 (1991).

[6] R.T. Pierrehumbert, Phys. Fluids A **3**, 1250 (1991).

[7] H.D.I. Abarbanel, R. Brown, and M.B. Kennel, J. Nonlinear Sci. **1**, 175 (1991).

[8] J.M. Finn and E. Ott, Phys. Fluids **31**, 2992 (1988).

[9] E. Ott and T.M. Antonsen, Jr., Phys. Rev. A **39**, 3660 (1989).

[10] D. Beigie, A. Leonard, and S. Wiggins, Phys. Fluids A **3**, 1039 (1991).

[11] S. Balachandar and L. Sirovich, Phys. Fluids A **3**, 919 (1991).

[12] A. Vincent and M. Meneguzzi, J. Fluid Mech. **225**, 1 (1991).

[13] Z. She and S.A. Orszag, Phys. Rev. Lett. **66**, 1701 (1991).

[14] R. Benzi, L. Biferale, G. Paladin, A. Vulpiani, and M. Vergassola, Phys. Rev. Lett. **67**, 2299 (1991).

[15] J.P. Gollub, J. Clarke, M. Gharib, B. Lane, and O.N. Mesquita, Phys. Rev. Lett. **67**, 3507 (1991).

[16] E.S.C. Ching, University of Chicago preprint.

[17] V. Rom-Kedar, A. Leonard, and S. Wiggins, J. Fluid Mech. **214**, 347 (1990).

[18] J. Feder, *Fractals* (Plenum Press, New York, 1988).

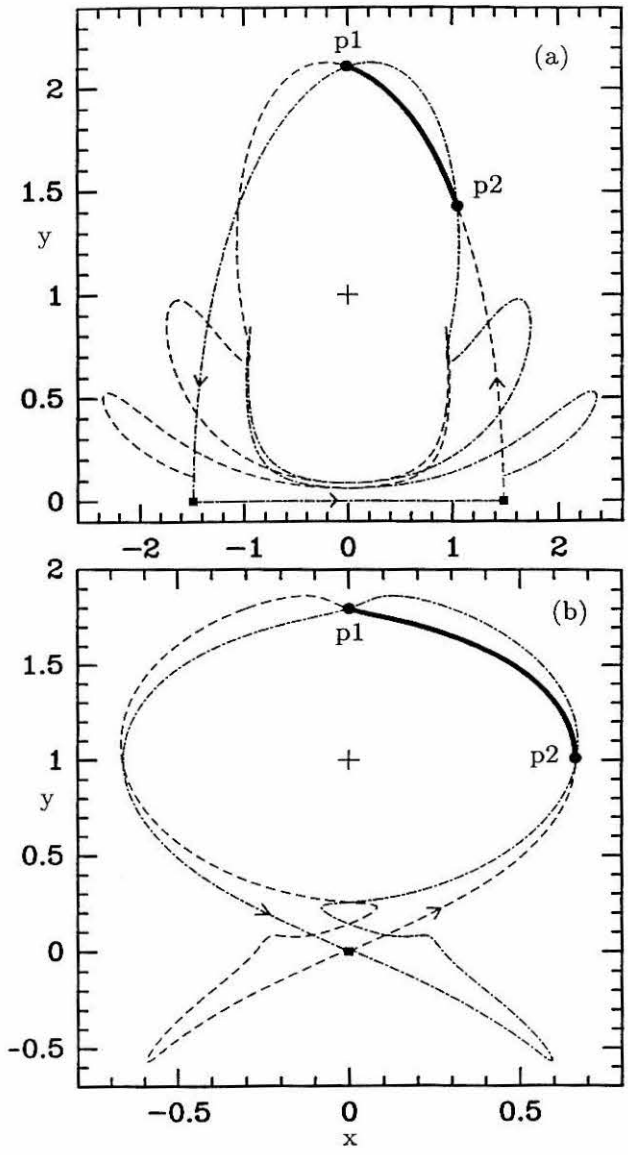


Figure 6.1 Segments of stable (dashed-dotted) and unstable (dashed) manifolds of hyperbolic fixed points (solid squares) in Σ, and our chosen interface (solid line) for the (a) open and (b) closed flow. The vortices are at $(x, y) = (0, \pm 1)$ at $t = n \in Z$, with magnitude of circulation $\Gamma = 0.4(2\pi)^2$ and (a) $\varepsilon = 0.085 \cdot 2\pi$, (b) $\varepsilon = 0.02 \cdot 2\pi$.

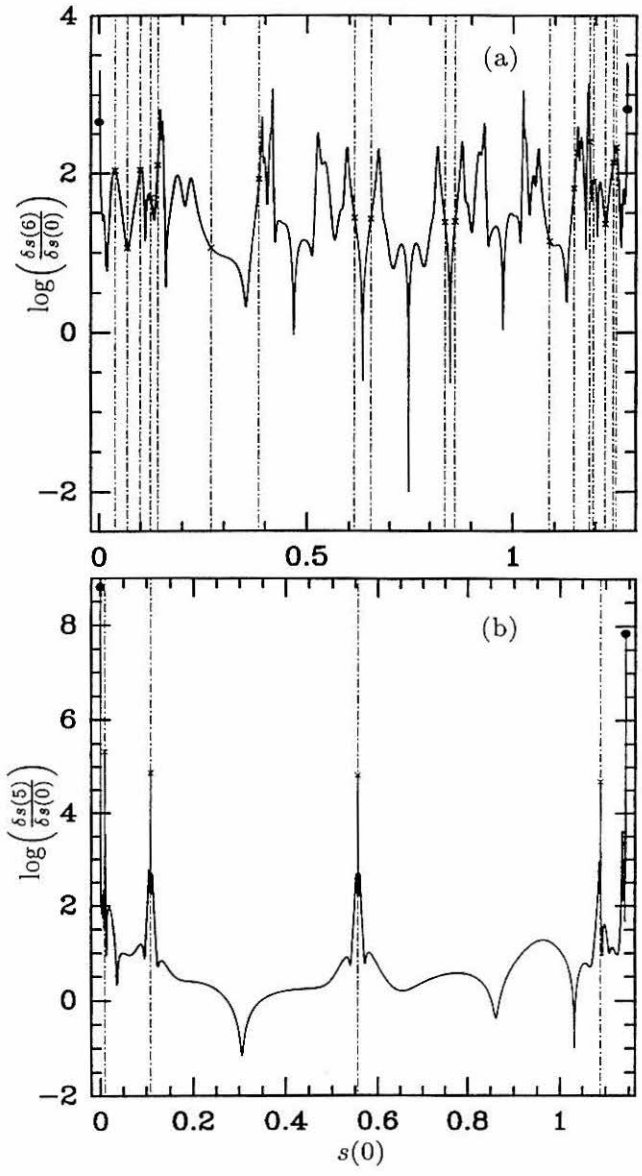


Figure 6.2 Plots of *stretch profiles* for (a) the open flow at $n = 6$ and (b) the closed flow at $n = 5$, where $s(n)$ denotes arclength along the interface (increasing from $p1$ to $p2$). The *stretch profiles* are plots of *stretch* $\delta s(n)/\delta s(0)$ on a common logarithmic scale versus initial arclength of the interface $s(0)$.

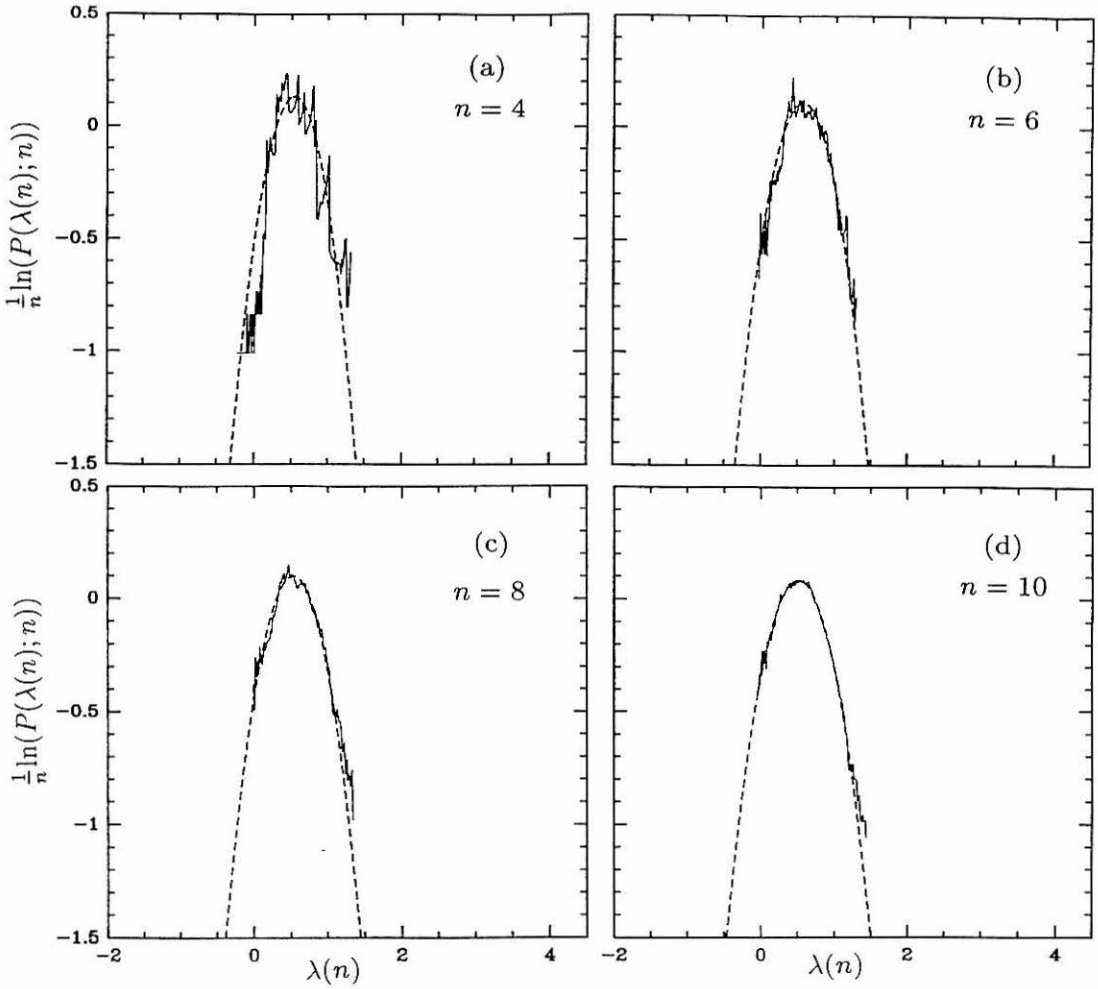


Figure 6.3 (a) Scaled open flow stretch distributions (solid lines) at sample times $n = 4, 6, 8, 10$. The dashed lines are Gaussian approximations, defined by having the same mean and standard deviation of the actual distributions.

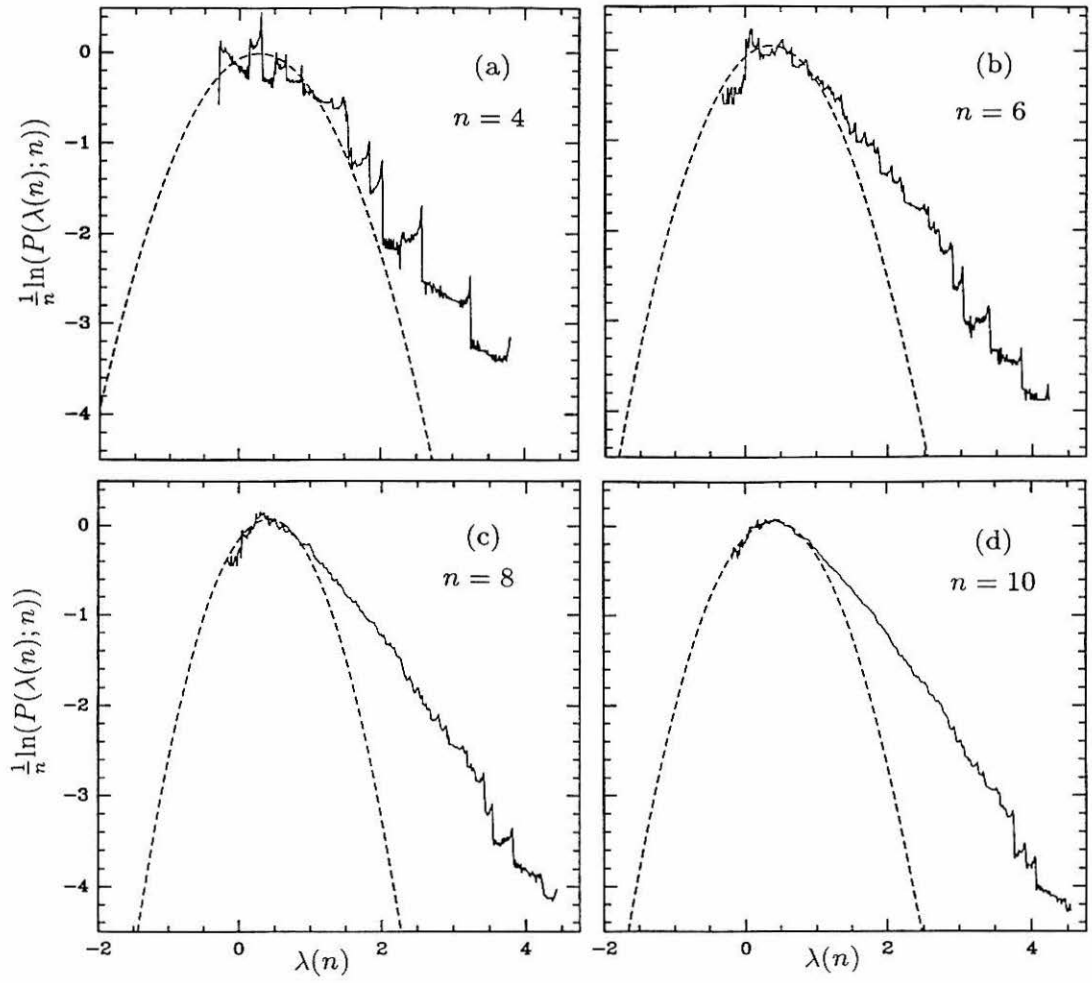


Figure 6.3 (b) Scaled closed flow stretch distributions (solid lines) at sample times $n = 4, 6, 8, 10$. The dashed lines are Gaussian approximations, defined by having the same mean and standard deviation of the actual distributions. The Gaussian approximation *fits the actual distributions well except for very small probability values.*

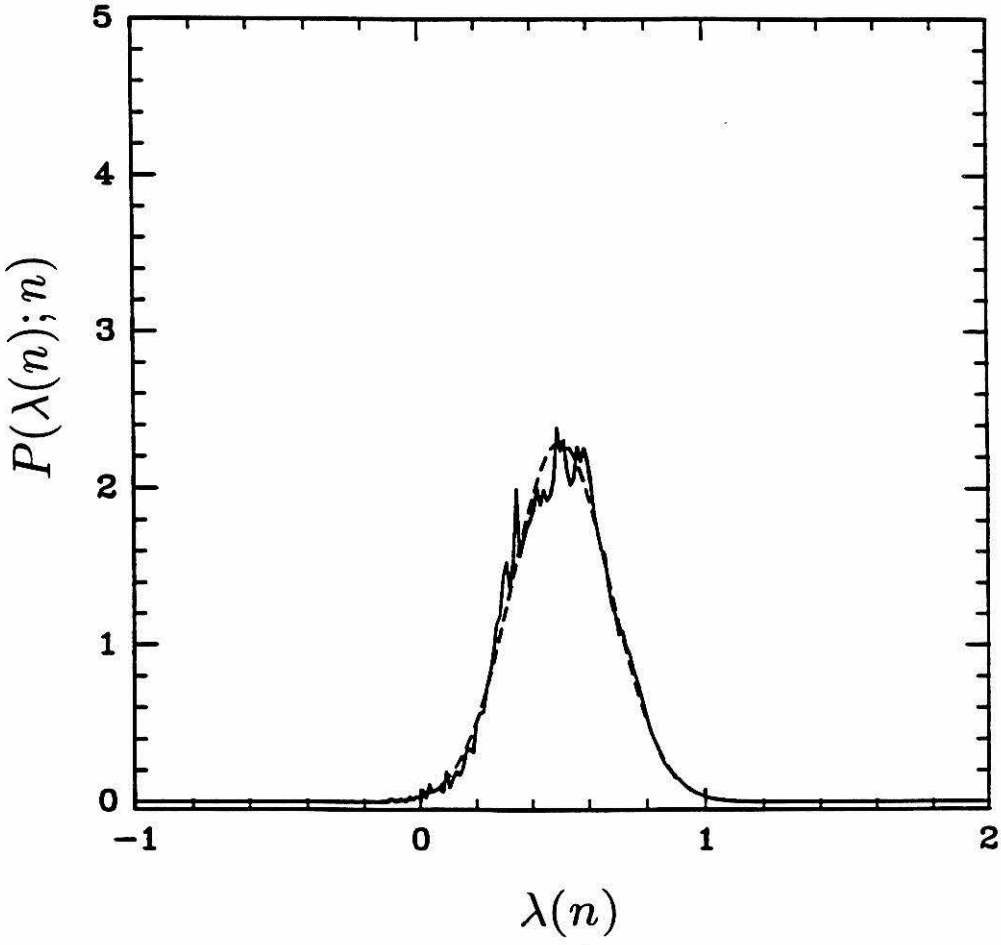


Figure 6.4 The closed flow stretch distribution (solid line) at $n = 10$ agrees well with the Gaussian approximation (dashed line) for the range of probabilities shown.

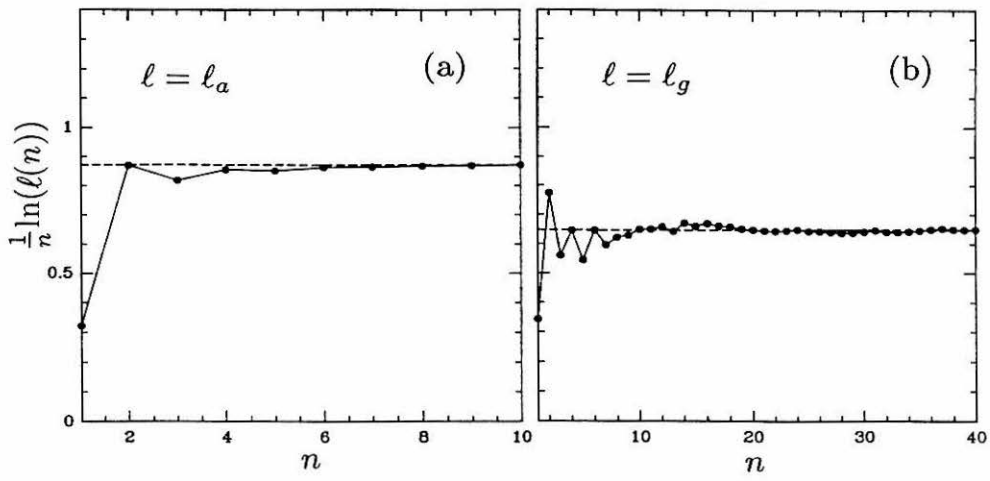


Figure 6.5 Exponents associated with the growth rate of interfacial length. Note that the closed flow computation of $\ell_g(n)$ is performed over 40 cycles to ensure convergence. This computation is performed with a *fixed* number of points; we verified that it does not affect the statistics associated with the Gaussian approximation $\ell_g(n)$ (since ignoring the high-stretch tail does not significantly alter the Gaussian approximation).

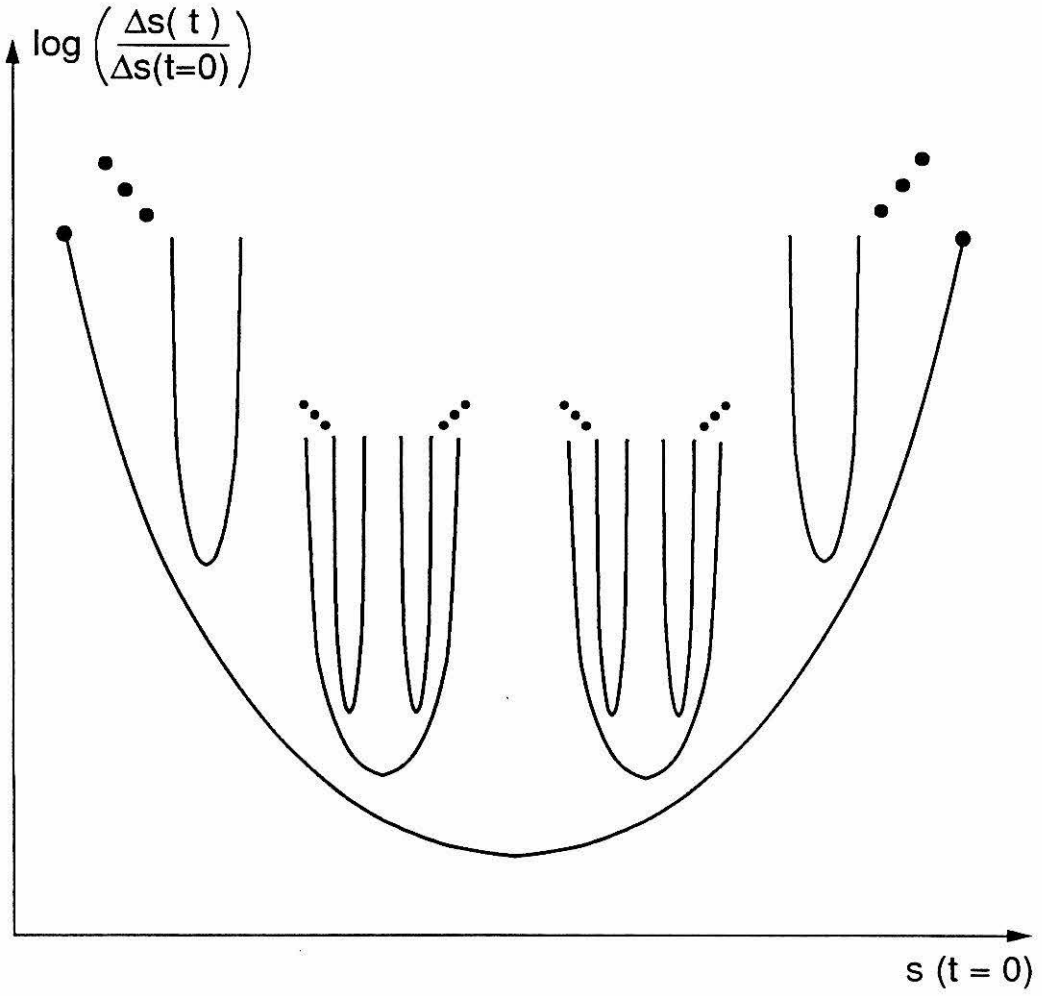
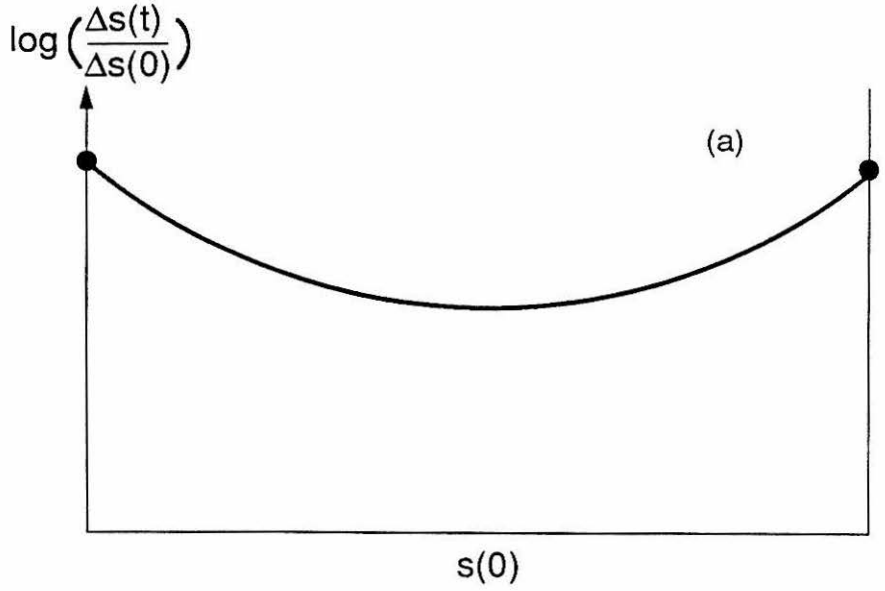


Figure 6.6 Heuristic portrayal of the unperturbed stretch profile approximately repeating itself on smaller and smaller scales (refer back to Chapter 3, and recall the rather cartoon-like nature of this portrayal).



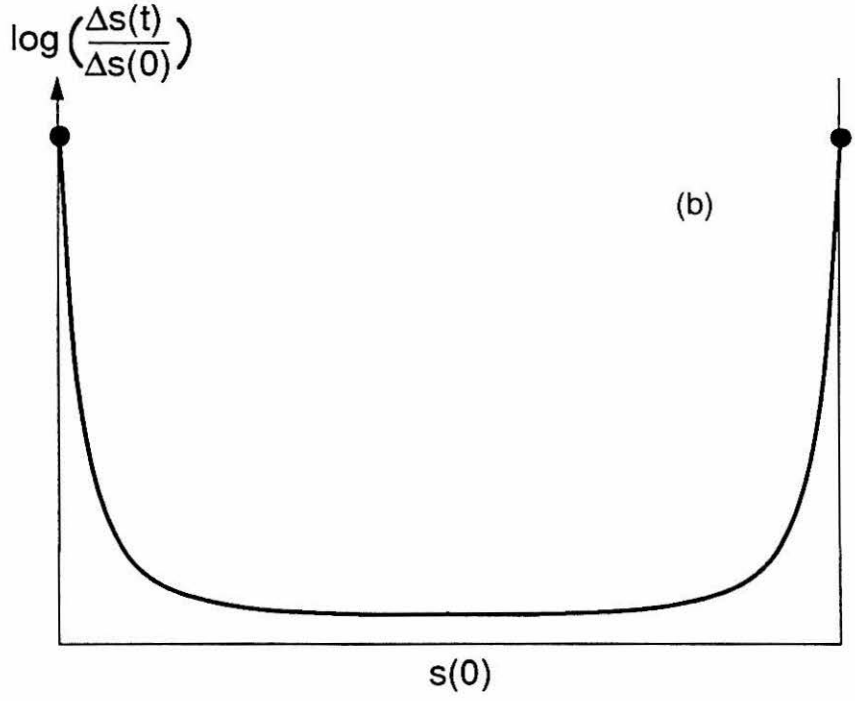


Figure 6.7 As the unperturbed stretch profile becomes more non-uniform, good stretching is weighted more towards the endpoints, entailing that good stretching will be weighted more towards the SIP's in the perturbed case.

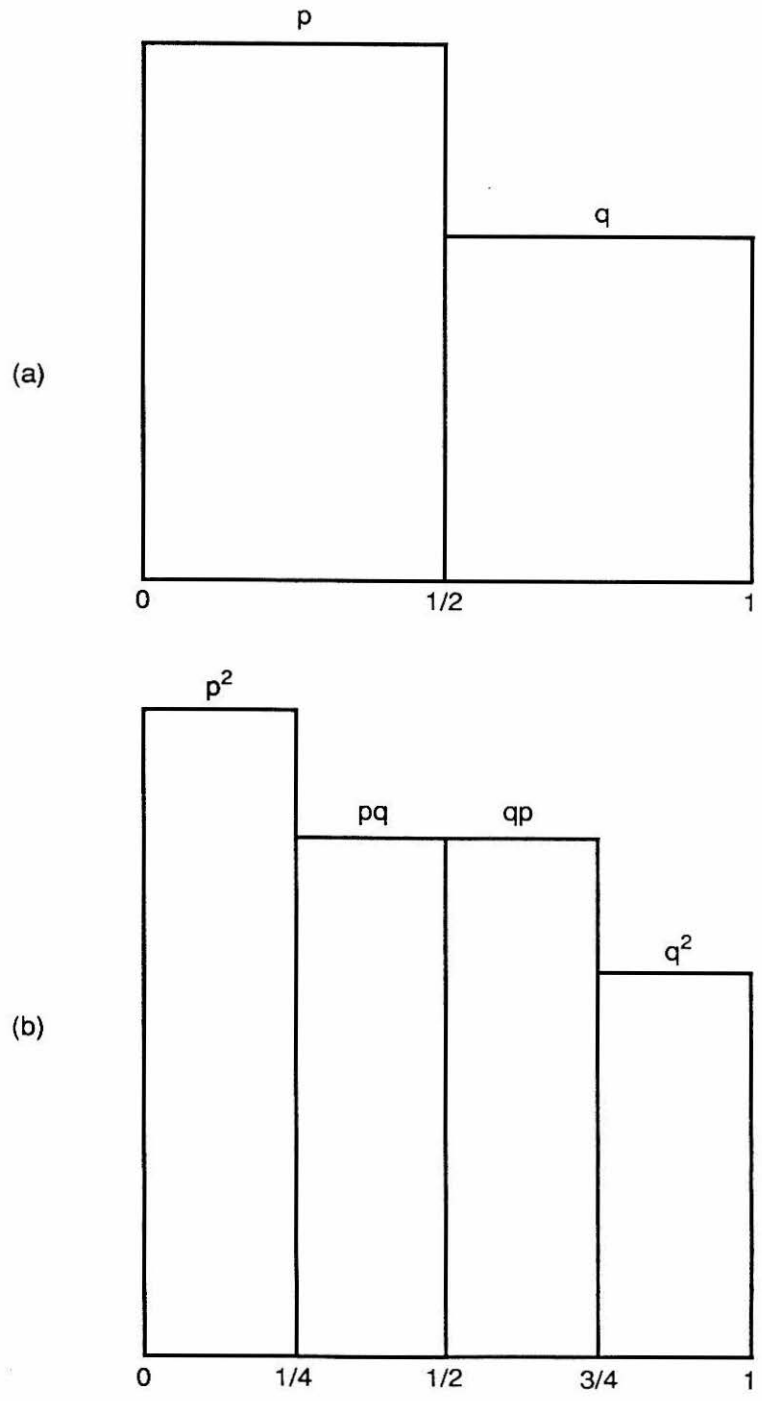


Figure 6.8 The (a) first and (b) second iterate of the binomial multiplicative process.

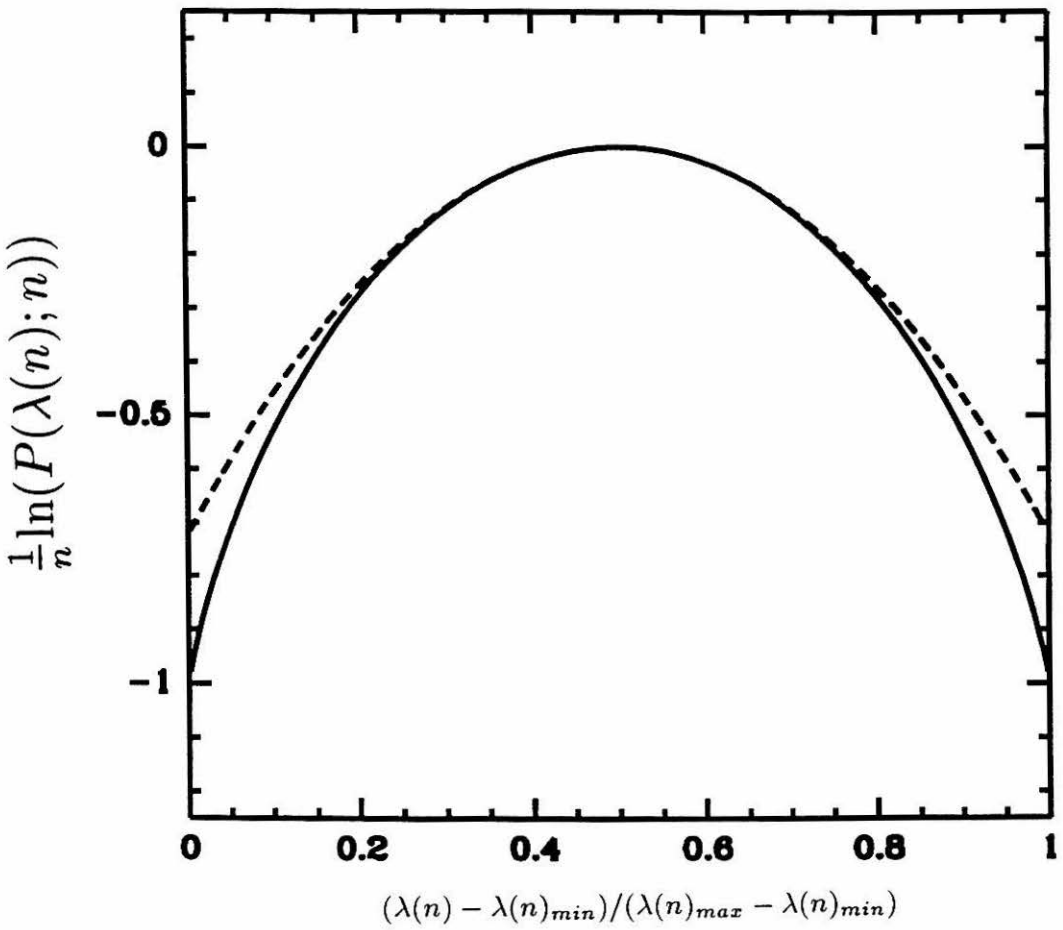


Figure 6.9 The asymptotically invariant form of the scaled stretch distribution associated with the binomial multiplicative process for arbitary $p \neq q$. The vertical scale is in units of $|\ln(P(\lambda(n);n)_{min})|$, and the dashed line shows the Gaussian approximation.

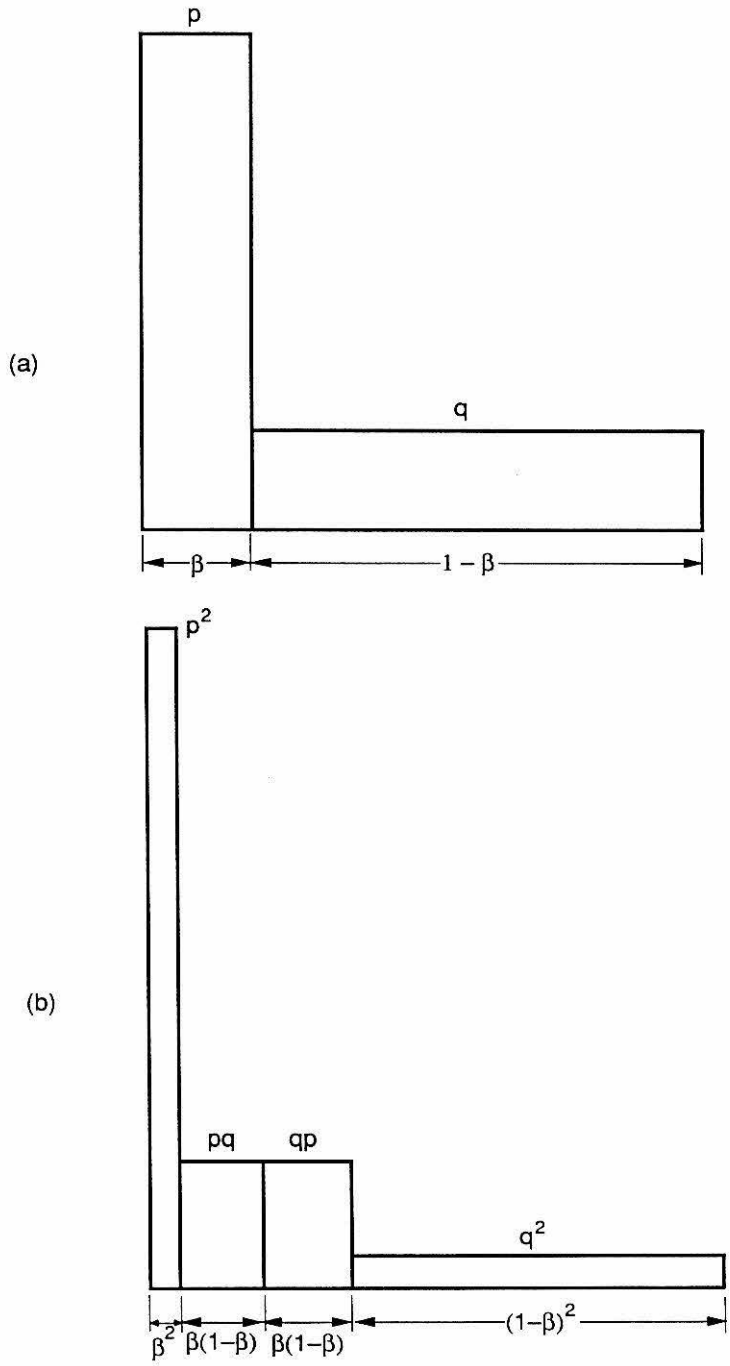


Figure 6.10 The (a) first and (b) second iterate of the binomial multiplicative process with weighted subdomains.

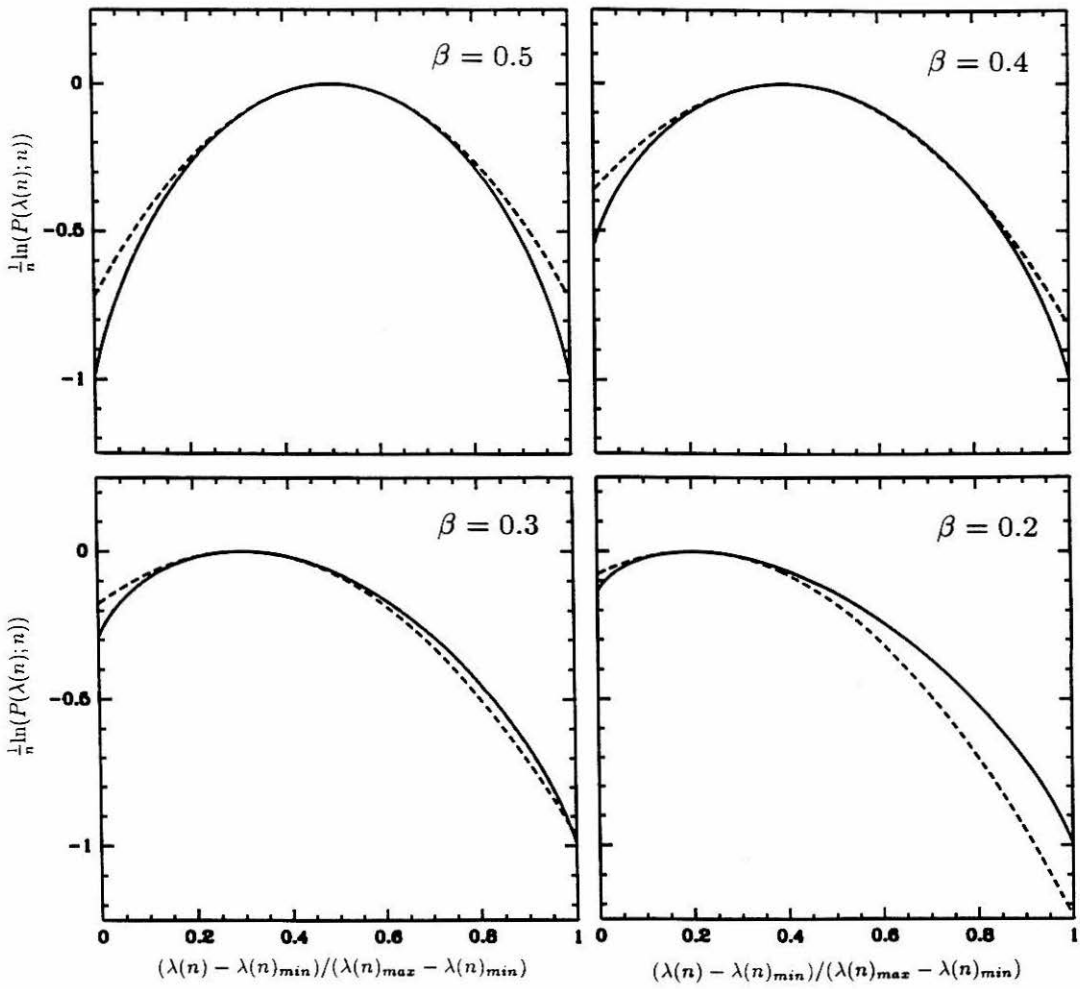


Figure 6.11 Scaled probability distributions (solid lines) associated with the weighted binomial multiplicative process for decreasing values of β. The vertical scale is in units of $|\ln(P(\lambda(n); n)_{min})|$, and the dashed lines show the Gaussian approximation.

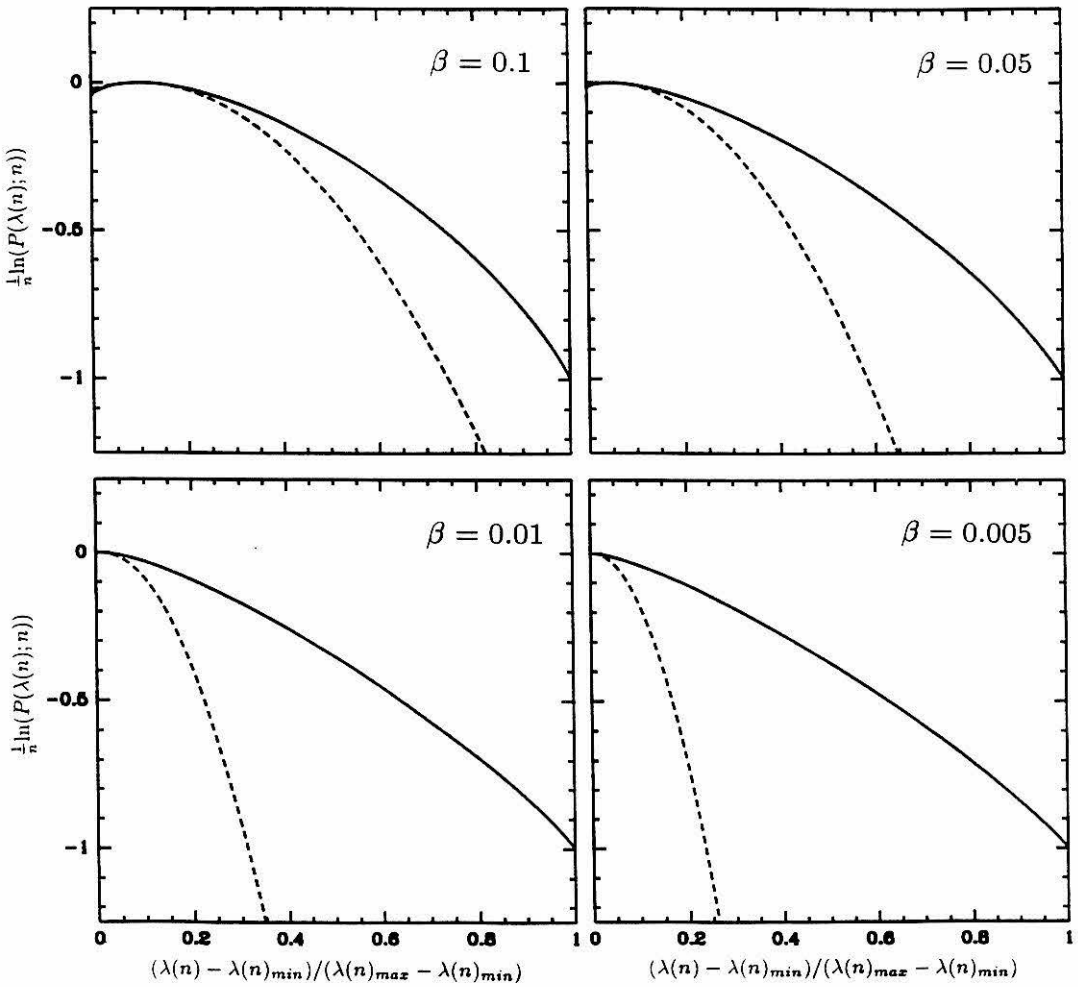


Figure 6.11 *Continued.*

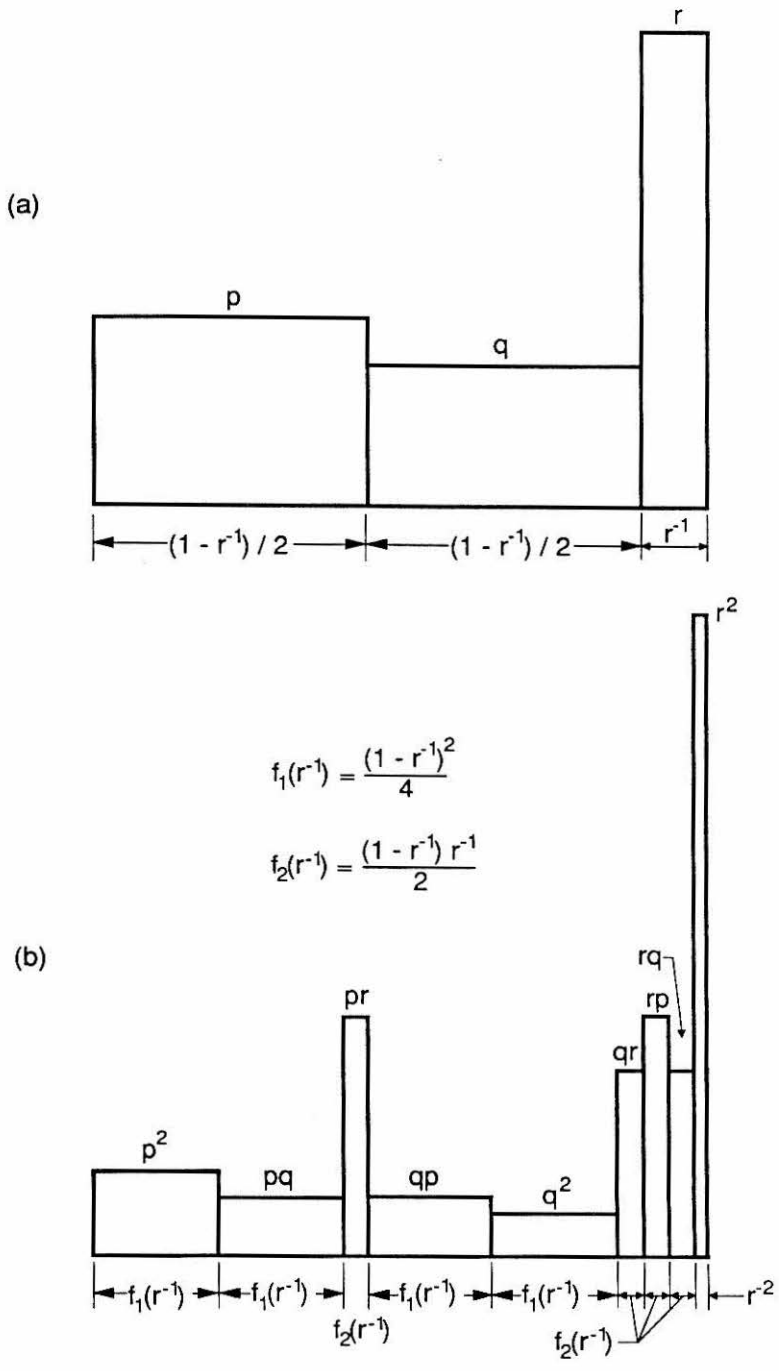


Figure 6.12 The (a) first and (b) second iterate of the trinomial stretch process, with one stretch r much bigger than the other two stretches p, q.

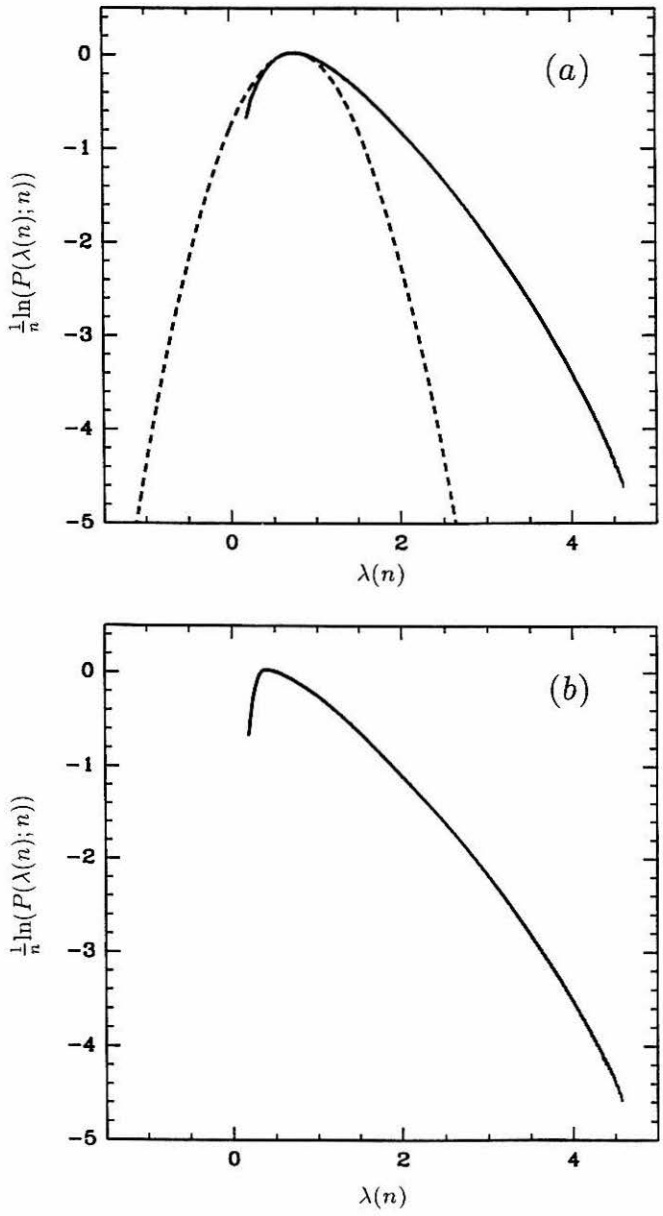


Figure 6.13 Scaled probability distributions associated with the trinomial stretch process for (a) $p = 1.2$, $q = 3.4$, and $r = 100$, and (b) the same parameters as (a) except now $q = 1.7$. In (a) the Gaussian approximation is given by the dashed line; in (b) the approximation is omitted since the distribution is obviouly non-Gaussian even at substantial probability values.

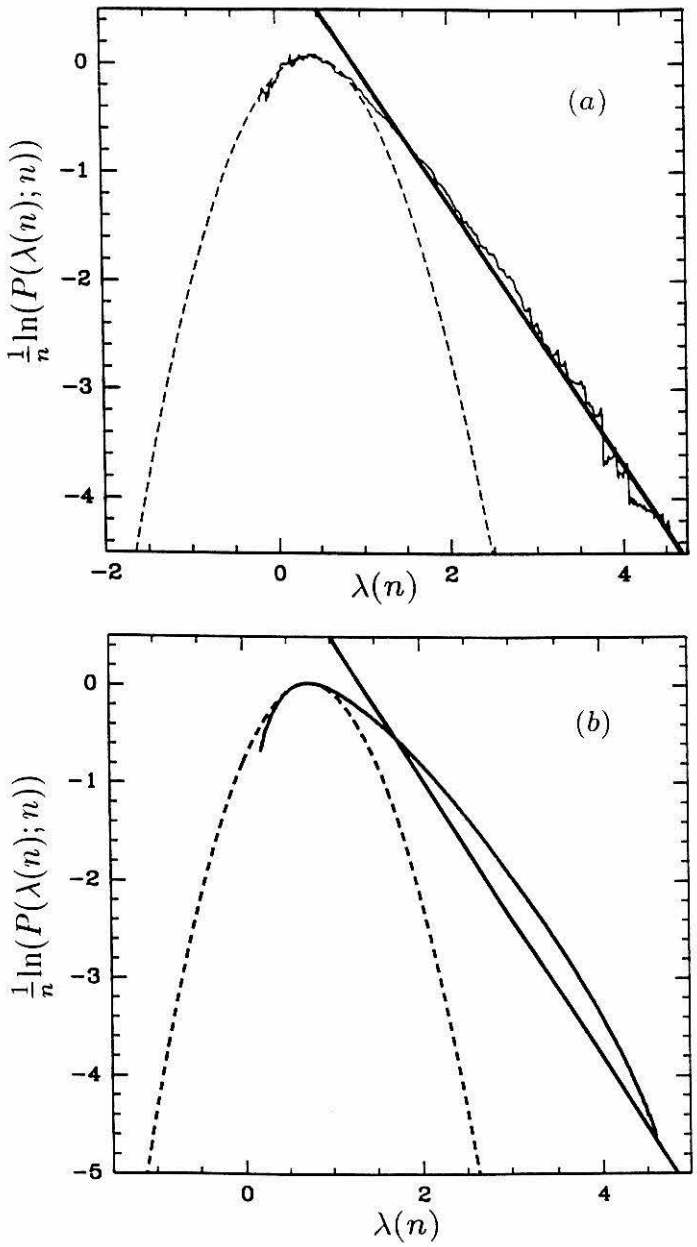


Figure 6.14 (a) Attempting to draw a straight line through the nearly exponential tail of the closed flow at the $n = 10$ sample, and (b) a similar line for the trinomial distribution of Figure 6.13(a).

Chapter 7
Concluding remarks

The principal theme of this study has been the use of global stable and unstable manifolds of invariant hyperbolic sets as templates for studying the dynamics within classes of homoclinic and heteroclinic chaotic tangles, focusing on transport, stretching, and mixing within these tangles. These templates are exploited in the context of *lobe dynamics* within invariant lobe structures formed out of intersecting global stable and unstable manifolds. The principal goals have been three-fold.

(i) *Extending the templates and their applications to fundamentally larger classes of dynamical systems.* In particular, we have generalized many of the canonical dynamical systems constructs associated with 2D time-periodic vector fields to apply to 2D vector fields with more complicated time dependences, focusing on multiple-frequency time dependence. The encouraging result is that most of the single-frequency constructs are robust, entailing some appealing generalizations. The study offered here is of course a preliminary step in what should be fairly promising subject matter, with much remaining to be pursued. A variety of other extensions have been and remain to be performed. For example, the one degree-of-freedom transport theory has been generalized to classes of multiple degree-of-freedom Hamiltonian systems with hyperbolicity in one degree-of-freedom and toroidal intersection manifold geometry,[1] and a significant achievement would be to further extend to more general classes of multiple degree-of-freedom Hamiltonian systems.

(ii) *Expanding the description of dynamics offered by the templates.* Prior work in the context of lobe dynamics has been principally concerned with phase space transport associated with ensembles of points uniformly distributed throughout phase space. Hence our study of stretching and mixing in the context of lobe dynamics can be viewed as an example of such an expansion. Indeed examples are found even with phase space transport, such

as the study of lobe penetration into the action or energy coordinate, or equivalently of transport of an ensemble of points that originate on a level set of the unforced Hamiltonian. Additionally, though a selling point of lobe dynamics studies is the nonstatistical framework, a pursuit of statistical studies within the backdrop of lobe dynamics, such as distributions of finite-time Lyapunov exponents, affords another promising example.

(iii) *Strengthening the interplay between dynamical systems analysis and chaotic nonlinear physical phenomena.* We have been primarily interested in chaotic fluid flows and molecular dynamics problems, and there are of course many other physical examples, such as kinematic dynamo problems, celestial mechanics problems, and so on. We stress how basic nonlinear physical processes - mixing, stirring, dissociation, magnetic field amplification, *etc.* - are inherently and fundamentally related to the underlying chaotic dynamics. The interplay between dynamical systems theory and nonlinear physical processes is indeed quite fruitful. The global, geometrical approach of dynamical systems analysis motivates new ways of thinking about nonlinear physical phenomena, an example being the use of invariant manifolds in *non*-integrable systems to distinguish between qualitatively different types of motion, such as bounded and unbounded motion, and to identify and study the transitions between the two. Conversely, physical phenomena motivate new dynamical systems analysis. For example, the desire to study chaotic advection under more realistic fluid velocity fields motivated the multiple-frequency extension, and the desire to study interfacial stretching and the mixing of passive scalars across interfaces in fluid flows motivated the horseshoe map extension. Though the dynamical systems community has conventionally been interested in asymptotic behavior and infinite-time dynamics, the finite-time nature of basic nonlinear physical processes, such as mixing and dissociation, implies that much of

our study of dynamics is explicitly interested in finite-time dynamics and transient phenomena, relatively virgin territory from a dynamical systems perspective.

Though dynamical systems theory can be elegant and insightful, it nevertheless applies to fairly limited contexts, and the general pursuit along the lines of (i) - (iii) offers a promising endeavour in the effort to make dynamical systems theory apply more robustly to descriptions of physical reality. There will of course always be fundamental restrictions on a dynamical systems description of nonlinear behavior, but this will be the case for any nonlinear theory or method. Though conventional disciplines have a tendency to judge dynamical systems theory on what it *cannot* do for any particular discipline, there is merit in judging the theory on what it *can* do, and seeing how far dynamical systems theory can go as a framework for studying a wide spectrum of nonlinear physical phenomena.

References

[1] S. Wiggins, Physica D **44**, 471 (1990).