Geometric Singular Perturbation Theory

Christopher K.R.T. Jones
Division of Applied Mathematics
Brown University
Providence, RI 02912

Contents

The research described here was supported by the National Science Foundation under grant DMS-9403774 and the Office of Naval Research under grant N00014-92-J-1401. This manuscript was written, and some of the research described herein was completed, partly while the author was supported by a Humboldt Award at the Universität Stuttgart and partly while a Visiting Professor of the Consiglio Nazionale Delle Richerche at Università di Roma II. The author wishes to thank Jonathan Rubin for his careful reading of the original manuscript.

Chapter 1

Introduction

The goal of these lectures is an exposition of the geometric approach to singular perturbation problems. Singularly perturbed equations gain their special structure from the presence of differing time scales. The fundamental tool in their analysis, from the perspective taken here, is the set of theorems due to Fenichel. The first step is then to explain these theorems and their significance. At the same time, new proofs of Fenichel's three main results will be outlined.

1.1 Background and motivation

The basic equations we consider are of the form

$$\begin{aligned} x' &= f(x,y,\epsilon) \\ y' &= \epsilon g(x,y,\epsilon), \end{aligned} \tag{1.1}$$

where $' = \frac{d}{dt}$, $x \in \mathbf{R}^n$, $y \in \mathbf{R}^l$ and ϵ is a real parameter. We shall compile various hypotheses about the system (1.1), which are denoted with the letter H.

(H1) The functions f and g are both assumed to be C^∞ on a set $U \times I$ where $U \subset \mathbf{R}^N$ is open, with $N = n + l$, and I is an open interval, containing 0.

Note that we are assuming full smoothness on the nonlinear terms which is unnecessary but greatly simplifies the discussion. If less smoothness is present in a given problem the precise smoothness required can be easily retraced through the proofs.

System (1.1) can be reformulated with a change of time-scale as

$$\begin{aligned} \epsilon \dot{x} &= f(x,y,\epsilon) \\ \dot{y} &= g(x,y,\epsilon), \end{aligned} \tag{1.2}$$

where $\dot{} = \frac{d}{d\tau}$ and $\tau = \epsilon t$. The time scale given by τ is said to be slow whereas that for t is fast, as long as $\epsilon \neq 0$ the two systems are equivalent. Thus we call

(1.1) *the fast system* and (1.2) *the slow system.* We have two distinguished limits for these equations, one naturally associated with each scaling as $\epsilon \to 0$. In (1.1) letting $\epsilon \to 0$ we obtain the system

$$\begin{aligned} x' &= f(x,y,0) \\ y' &= 0. \end{aligned} \tag{1.3}$$

According to (1.3) the variable x will vary while y will remain constant. Thus x is called the fast variable. If we let $\epsilon \to 0$ in (1.2), the limit only makes sense if $f(x,y,0) = 0$ and is thus given by

$$\begin{aligned} &f(x,y,0) = 0 \\ &\dot{y} = g(x,y,0). \end{aligned} \tag{1.4}$$

One thinks of the condition $f(x,y,0) = 0$ as determining a set on which the flow is given by $\dot{y} = g(x,y,0)$. It is natural to attempt to solve x in terms of y from the equation $f(x,y,0) = 0$ and plug it into the second equation of (1.4) (the reader should check that the dimensions are right to expect such a solution if non-degeneracy conditions hold). Notice that this set is exactly the set of critical points for (1.3). We thus have the "formal" picture that (1.3) has large sets of critical points and that (1.4) blows the flow on this set up to produce non-trivial behavior.

In either limiting formulation, one pays a price. On the (large) set $f(x,y,0) = 0$ *the flow is trivial* for (1.3). Whereas under (1.4) the flow is non-trivial on this set, but *the flow is not defined off this set.* The primary mathematical goal of *geometric singular perturbation theory,* henceforth denoted by the acronym GSP, is to realize both these aspects (i.e., fast and slow) simultaneously. This apparently contradictory aim will be accomplished within the phase space of (1.1) (or, equivalently, (1.2)) for ϵ *non-zero but small.*

There are two basic reasons why GSP is a powerful tool for analyzing high-dimensional systems:

1. In many applications, quantities will vary on *widely differing time scales,* and thus are naturally formulated in the form (1.1).

2. It affords a reduction of a possibly high-dimensional system, such as (1.1), into the lower-dimensional systems (1.3) and (1.4).

The first reason given above justifies the theory from an applied point of view, but also offers us the opportunity of invoking many different applications as examples to guide the theory. The second rationale means that we can hold the hope of analyzing singularly perturbed systems of twice the size of those analyzed without this theory. For example if $n = l = 2$, we would study 2-dimensional systems, the analysis of which is well understood, and, through GSP, make conclusions about 4-dimensional systems, which are ostensibly far less tractable. Moreover, the resulting behavior is not restricted to being merely a shadow of that present in (1.3) and (1.4), for new structures can result from the patching together of solutions of (1.3) and (1.4). The early examples will not reflect such effects, but we will progressively develop richer dynamical behavior. The later applications will exhibit *intrinsically higher-dimensional phenomena*, despite their being rendered susceptible to analysis by reduction to low-dimensional systems. It should be noted here that the theory of singular perturbations commands a large literature that does not fit into the category of GSP as discussed here. Indeed, this classical theory predates GSP and the interested reader is referred to [11] and [44] for good expositions of this theory.

The phenomena that we will isolate for (1.1) will generally involve the construction of specific orbits, such as *homoclinic, heteroclinic, or periodic orbits.* These are constructed by *following certain invariant manifolds* (for instance, stable and unstable manifolds of critical points) through their ambient phase space and using the reductions i.e., (1.3) and (1.4), to keep track of their position and configuration at different points of their travel. These "special" orbits may be of significance due to their rôle in the overall dynamics of the equation, for instance homoclinic orbits are often a signature of chaotic motion, or as special solutions, such as travelling waves, of a related partial differential equation.

Summary of goals

- Determination of flow near sets $f(x,y,0)=0$ for (1.1):
  - Fenichel's theorems,
  - Fenichel coordinates and normal form,
  - slow manifold flow.
- Effective tracking of invariant manifolds through the phase space of (1.1):
  - use of differential forms,
  - transversality,
  - exchange lemmas.
- Applications to the existence and properties of special orbits of (1.1):
  - perturbed slow structures,
  - travelling waves and stability,
  - homoclinic abundance.

1.2 Fenichel's first theorem

The set of critical points $f(x,y,0)=0$ for (1.3) is formed by solving n equations in $\mathbf{R}^N$, where $N = n + l$, and thus is expected to be, at least locally, an l-dimensional manifold. Indeed, it is natural to expect it to have a parametrization by the variable y. We shall thus assume that we are given an l-dimensional manifold, possibly with boundary, M_0 which is contained in the set $\{f(x,y,0)=0\}$. The fundamental hypothesis on M_0 will be that, as a set of critical points, the directions normal to the manifold will correspond to eigenvalues that are not neutral. In the following, the notation $\mathcal{R}(\lambda)$ denotes the real part of λ.

Definition 1 *The manifold M_0 is said to be normally hyperbolic if the linearization of (1.1) at each point in M_0 has exactly l eigenvalues on the imaginary axis $\mathcal{R}(\lambda) = 0$.*

Fenichel's first theorem asserts the existence of a manifold that is a perturbation of M_0. It will be connected with the flow of (1.1) when $\epsilon \neq 0$. We need a definition to clarify this connection. The notation $x \cdot t$ is used to denote the application of a flow after time t to the intial condition x. The existence of a flow for (1.1) follows from the basic theorems of ODE.

Definition 2 *A set M is locally invariant under the flow from (1.1) if it has neighborhood V so that no trajectory can leave M without also leaving V. In other words, it is locally invariant if for all $x \in M$, $x \cdot [0,t] \subset V$ implies that $x \cdot [0,t] \subset M$, similarly with $[0,t]$ replaced by $[t,0]$ when $t < 0$.*

Fenichel's theorems will actually address the perturbation of a subset of $\hat{M}_0$, because of technical difficulties near the boundary.

(H2) The set M_0 is a compact manifold, possibly with boundary, and is normally hyperbolic relative to (1.3).

The set M_0 will be called the critical manifold. We are now in a position to state the first theorem that Fenichel proved, under the hypotheses (H1) and (H2).

Theorem 1 (Fenichel's Invariant Manifold Theorem 1) *If $\epsilon > 0$, but sufficiently small, there exists a manifold M_ϵ that lies within $O(\epsilon)$ of M_0 and is diffeomorphic to M_0. Moreover it is locally invariant under the flow of (1.1), and C^r, including in ϵ, for any $r < +\infty$.*

This theorem follows from Fenichel's early work, [15], as singular perturbations are a special case of the more general decomposition by exponential rates that he considered in that context. However, his later paper, see [18], specifically addresses singular perturbations. There are a number of alternative formulations and proofs of this basic theorem, see, for instance, the work of Sakamoto [51].

The manifold M_ϵ will be called the slow manifold. It should be noted that the only connection to the flow is through the statement that the perturbed manifold M_ϵ is locally invariant. This seems weak but, in fact, is not as it entails that we can restrict the flow to this manifold, which is lower-dimensional, in order to find interesting structures. The fact that the manifold is *locally invariant*, and not *invariant*, is due to the (possible) presence of the boundary and the resulting possibility that trajectories may fall out of M_ϵ by escaping through the boundary. This cannot be avoided as most applications do indeed supply us with manifolds that have boundaries. A comment is also in order about normal hyperbolicity; there are many applications in which interesting phenomena occur because normal hyperbolicity of manifolds breaks down, such as in relaxation-oscillations, but we shall not consider such cases in these lectures. As is the case for center manifolds, it should be noted here that the exponent r cannot be set to $+\infty$.

In order to significantly simplify the notation, as well as the structure of the proofs, we shall restrict attention throughout these lectures to the case that M_0 is given as the graph of a function of x in terms of y. That is we assume there is a function $h^0(y)$, defined for $y \in K$, with K being a compact domain in $\mathbf{R}^l$, and so that

$$M_0 = \left\{(x, y) : x = h^0(y)\right\}.$$

This is a natural assumption as it can always be satisfied for M_0 *locally*. Indeed, on account of normal hyperbolicity (H2) the matrix

$$D_x f(\hat{x}, \hat{y}, 0)$$

is invertible for any $(\hat{x}, \hat{y}) \in M_0$ and hence x can locally be solved for y by the Implicit Function Theorem. We are thus just assuming that such a solution can be made *globally* over M_0.

Thus, consider $x = h^0(y)$ wherein $y \in K$ and make the following assumption.

(H3) The set M_0 is given as the graph of the C^∞ function $h^0(y)$ for $y \in K$. The set K is a compact, simply connected domain whose boundary is an $(l-1)$-dimensional C^∞ submanifold.

Under the hypotheses (H1)-(H3), we can restate Fenichel's first theorem in terms of the graph of a function.

Theorem 2 *If $\epsilon > 0$ is sufficiently small, there is a function $x = h^\epsilon(y)$, defined on K, so that the graph*

$$M_\epsilon = \{(x, y) : x = h^\epsilon(y)\},$$

is locally invariant under (1.1). Moreover h^ϵ is C^r, for any $r < +\infty$, jointly in y and ϵ.

Remark The diffeomorphism between M_ϵ and M_0 follows easily in this formulation through the diffeomorphism of the graph to K.

An equation on M_ϵ can easily be calculated using Theorem 1. We substitute the function $h^\epsilon(y)$ into (1.1) and see that the y equation will decouple from that of the x equation. We thus obtain an equation for the variation of the variable y. Since y parametrizes the manifold M_ϵ, this equation will suffice to describe the flow on M_ϵ. It is given by

$$y' = \epsilon g(h^\epsilon(y), y, \epsilon). \tag{1.5}$$

In the alternative slow scaling we can recast (1.6) as

$$\dot{y} = g(h^{\epsilon}(y), y, \epsilon), \tag{1.6}$$

which has the distinct advantage that a limit exists as $\epsilon \to 0$, given by

$$\dot{y} = g(h^{0}(y), y, 0), \tag{1.7}$$

which naturally describes a flow on the critical manifold M_0, and is exactly the second equation in (1.2). Using this theorem and this resulting equation (1.6), the problem of studying (1.1), at least on M_ϵ is reduced to a *regular perturbation problem.* In the next three sections we shall give examples in which this is applied.

1.3 An equation from phase-field theory

An equation with spatial derivatives of even powers in formulated by Caginalp and Fife [7] to describe the behavior of phase transitions. In a model case in which a scalar equation can be used as a reasonable model, Gardner and Jones [19] studied the stability of travelling waves. As an example of the above theory, it will be shown here how to construct the basic travelling wave. Consider the equation

$$\frac{\partial \phi}{\partial t} = \epsilon^4 \frac{\partial^6 \phi}{\partial x^6} + \epsilon^2 A \frac{\partial^4 \phi}{\partial x^4} + \frac{\partial^2 \phi}{\partial x^2} + f(\phi), \tag{1.8}$$

where $f(\phi) = \phi(\phi - a)(1 - \phi)$ is the bistable nonlinearity with $a < \frac{1}{2}$, and $A > 0$. The parameter ϵ is intended to be small, see [19], and thus (1.8) is easily seen to be a perturbation of the well-known, scalar bistable reaction-diffusion equation. We shall seek travelling wave solutions of (1.8), namely $\phi(\xi) = \phi(x - ct)$, satisfying

$$\phi(\xi) \to \begin{cases} 0 & \text{as} \quad \xi \to -\infty \\ 1 & \text{as} \quad \xi \to +\infty. \end{cases} . \tag{1.9}$$

The wave is then seen to satisfy the ODE

$$- c\phi^{(1)} = \epsilon^4 \phi^{(6)} + \epsilon^2 A \phi^{(4)} + \phi^{(2)} + f(\phi), \tag{1.10}$$

where the derivatives are taken with respect to the travelling wave variable ξ. The equation (1.10) can be rewritten as a system of six equations

$$\begin{aligned} \dot{u}_1 &= u_2 \\ \dot{u}_2 &= u_3 \\ \epsilon\dot{u}_3 &= u_4 \\ \epsilon\dot{u}_4 &= u_5 \\ \epsilon\dot{u}_5 &= u_6 \\ \epsilon\dot{u}_6 &= -Au_5 - u_3 - cu_2 - f(u_1), \end{aligned} \tag{1.11}$$

where $\dot{} = \frac{d}{d\tau}$. This system is already formulated in slow variables and we have replaced ξ by τ as the independent variable to conform to our established notation. The correspondence with our notation is:

$$x = \begin{pmatrix} u_3 \\ u_4 \\ u_5 \\ u_6 \end{pmatrix}$$

is the fast variable, and

$$y = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

is the slow variable. The critical manifold M_0 can be taken as any compact subset of

$$\{u_4 = u_5 = u_6 = 0,\ u_3 = -cu_2 - f(u_1)\},$$

which shall be chosen to be large enough to contain any of the dynamics of interest. The eigenvalues of the linearization at any point of M_0, other than the double eigenvalue at 0 are seen to be solutions of the quartic

$$\mu^4 + A\mu^2 + 1 = 0,$$

which are not pure imaginary if $0 < A < 2$.

The equations for the slow flow on the critical manifold M_0 are given by

$$\begin{aligned} \dot{u}_1 &= u_2 \\ \dot{u}_2 &= -cu_2 - f(u_1). \end{aligned} \tag{1.12}$$

The slow manifold M_ϵ, which exists by virtue of Theorem 1, is given by the equations

$$(u_3, u_4, u_5, u_6) = h^\epsilon(u_1, u_2) = (-cu_2 - f(u_1), 0, 0, 0) + O(\epsilon).$$

and the equations on M_ϵ are

$$\begin{aligned} \dot{u}_1 &= u_2 \\ \dot{u}_2 &= -cu_2 - f(u_1) + O(\epsilon). \end{aligned} \tag{1.13}$$

It is a well-known fact that (1.12) has a heteroclinic orbit connecting the critical point $(0,0)$ at $-\infty$ with $(1,0)$ at $+\infty$, for a particular value of c, say $c = c^*$. One checks easily that $(0,0)$ and $(1,0)$ are still critical points of (1.13), for ϵ sufficiently small (why?). The strategy is then to show that there is a $c = c(\epsilon)$ defined for ϵ small, with $c(0) = c^*$, at which there is such a heteroclinic orbit for (1.13). The idea is to show that the heteroclinic orbit for (1.12) exists by virtue of a transverse intersection of stable and unstable manifolds and thus perturbs.

Appending an equation for c to (1.12), the heteroclinic, on M_0, can be viewed as the intersection of the unstable manifold of the curve of critical points $\{(0,0,c) : |c - c^*| \text{ small }\}$, say W^- with the stable manifold of the curve

$\{(1,0,c) : |c-c^*| \text{ small }\}$, say W^+, see Figure 1. this intersection will be viewed in the plane $u = a$. In $u = a$, W^- is given by the graph of a function, say $u_2 = h^-(c)$, and W^+ is given by the graph of another function, say $u_2 = h^+(c)$. These curves are each monotone, which is the usual proof of the uniqueness of the wave and its speed, see for instance Aronson and Weinberger [2]. The transverse intersection is related to this fact and it will be shown specifically in Chapter 4 that the following quantity

$$\left(\frac{\partial h^-}{\partial c} - \frac{\partial h^+}{\partial c}\right)|_{c=c^*} \neq 0. \tag{1.14}$$

This is a Melnikov type calculation.

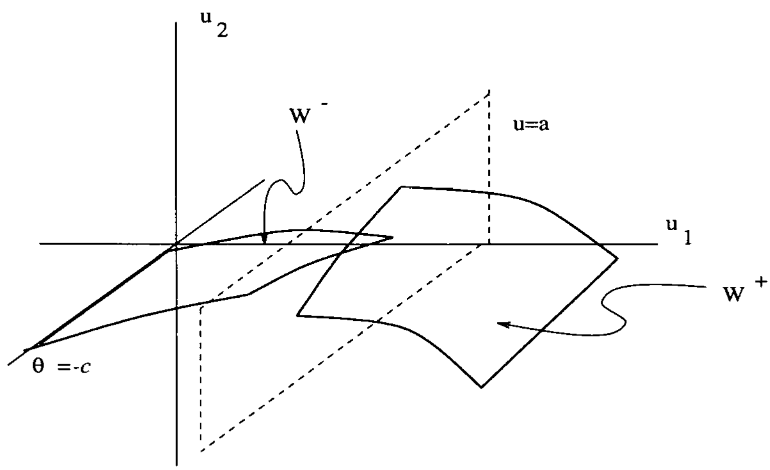

Figure 1
The intersection of the unstable and stable manifolds.

The next step is to check whether this intersection perturbs to M_ϵ. Indeed, on M_ϵ the relevant unstable and stable manifolds will again be given as graphs. The manifold $W^- \cap \{u_1 = a\}$ will be given by $u_2 = h^-(c,\epsilon)$ and $W^+ \cap \{u_1 = a\}$ will be given by $u_2 = h^+(c,\epsilon)$. An intersection point is found by solving these equations simultaneously for $u_2 = u_2^*(\epsilon)$ and $c = c^*(\epsilon)$. This will follow from the Implicit Function Theorem if the determinant of the matrix

$$\begin{pmatrix} 1 & \frac{\partial h^-}{\partial c} \\ 1 & \frac{\partial h^+}{\partial c} \end{pmatrix},$$

at $c = c^*$, and $\epsilon = 0$, is non-zero. But this is exactly the statement (1.14).

1.4 A travelling wave in semiconductor theory

An example due to Szmolyan, see [53], concerning the problem of finding a travelling wave in a system of equations governing the behavior of a semiconductor material will be presented. In the following, E is the electric field, n is the total concentration of *all* the electrons and u is the concentration of one individual species of electrons. There are two species of electrons present, and so the concentration of the other species is $n - u$. The concentration variable u satisfies a second order equation. We bypass the PDE's and formulate immediately the travelling wave equations, which are thus a system of four equations, given by

$$\begin{aligned} u' &= w \\ w' &= \epsilon(\nu_1(E) - c)w + \epsilon^2\nu_1'(E)u\tfrac{n-1}{\lambda^2} + (1+\alpha(E))u - n \\ E' &= \epsilon\tfrac{n-1}{\lambda^2} \\ n' &= \epsilon\left(\nu_1(E)u + \nu_2(E)(n-u) - cn + \gamma\right), \end{aligned} \tag{1.15}$$

where $\nu_i(E)$ and $\alpha(E)$ are phenomenologically determined, positive, smooth functions, the exact structure of which is not directly important. The nature of a certain combination of these functions that appears in the slow equations will be of most relevance. Of interest will be orbits of (1.15) that are homoclinic to critical points. These correspond to travelling waves of the original PDE that decay to a fixed constant state at $\pm\infty$. Clearly the variables E and n are slow, while u and w are fast. We thus have the correspondence $x = (u, w)$ and $y = (E, n)$ with the notation above. The critical manifold M_0 will be given by the equations

$$w = 0, \ u = \frac{n}{1+\alpha(E)} \tag{1.16}$$

and is thus easily seen to be defined, for our purposes, on any compact domain of $\mathbf{R}^2$. The eigenvalues of the linearization at any critical point of M_0, apart from the double eigenvalue at 0, are $\pm(1+\alpha(E))^{\frac{1}{2}}$. Since α is positive, these are then non-zero and the uniform normal hyperbolicity assumption is satisfied.

The limiting slow equations i.e., equations (1.7) for this example, are given by

$$\begin{aligned} \dot{E} &= \tfrac{n-1}{\lambda^2} \\ \dot{n} &= (G(E) - c)\,n + \gamma, \end{aligned} \tag{1.17}$$

where

$$G(E) = \frac{\nu_1(E) + \alpha(E)\nu_2(E)}{1+\alpha(E)}. \tag{1.18}$$

The graph of $G(E)$ will be important and is given in Figure 2.

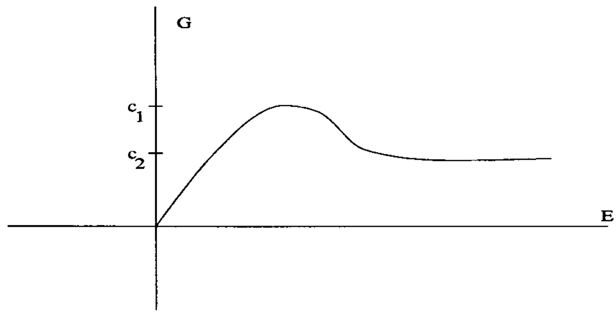

Figure 2
The graph of the function G.

The equations on the perturbed manifold M_ϵ, which exists by Theorem 1 for ϵ positive, but sufficiently small, are given by

$$\begin{aligned} \dot{E} &= \tfrac{n-1}{\lambda^2} \\ \dot{n} &= (G(E) - c)\, n + \gamma + O(\epsilon). \end{aligned} \tag{1.19}$$

The goal is to find an orbit for (1.19) that is homoclinic to a rest state. This orbit will be the desired travelling wave, as it is a homoclinic orbit for the system that happens to live on M_ϵ. The strategy is to find a homoclinic orbit for (1.17) and prove that it perturbs to such for (1.19) when ϵ is sufficiently small.

We first analyze (1.17) when γ is also set equal to 0. In this case a simple transformation converts it into a Hamiltonian system, namely we let $m = \log n$, and from (1.17) we obtain

$$\begin{aligned} \dot{E} &= \tfrac{e^m - 1}{\lambda^2} \\ \dot{m} &= G(E) - c. \end{aligned} \tag{1.20}$$

Since the variable n is a concentration, we are only interested in solutions with $n > 0$, and thus the transformation is valid for the solutions that will ultimately be of interest to us. It is easily checked that the function

$$H(E, m) = \frac{e^m - m}{\lambda^2} - \Gamma(E),$$

where $\Gamma'(E) = G(E) - c$, is a Hamiltonian for (1.20). If c is in the interval (c_1, c_2), see Figure 2, then (1.20) has 2 critical points, the left one being a saddle, which we denote $\hat{y}$, and the right one a center. It is an exercise (left to the reader-see [53]) to check that, from the Hamiltonian above, one can conclude the existence of an orbit homoclinic to $\hat{y}$, see Figure 3. In principle, this could be concluded from sketching the level curves of H, but this can be avoided by invoking some qualitative arguments about the nature of the level curve containing the saddle.

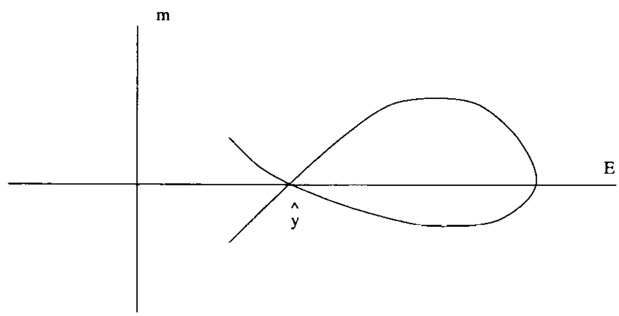

Figure 3
The homoclinic orbit in M_0.

One would not expect this homoclinic orbit to survive a perturbation, such as to equation (1.19), as it is not caused by a transversal intersection. However, we have not used the parameter γ. Undoing the transformation we see that (1.17) will also have a homoclinic orbit to a saddle, which, with an abuse of notation, we continue to denote by $\hat{y}$, but now $\hat{y}$ has its second coordinate given by $n = 1$. Since this point is a saddle an application of the Implicit Function Theorem shows that there is a nearby saddle critical point for γ sufficiently small (note that n stays equal to 1). We denote this curve of critical points by $\mathcal{C}$. The next step is to show that the unstable manifold of the curve $\mathcal{C}$ intersects its stable manifold transversely in (E, n, γ)-space at $\gamma = 0$ (we append the equation $\gamma' = 0$). There is then a hope of its perturbing to M_ϵ.

As is common in transversality arguments, we consider the intersection of these manifolds with the set $n = 1$. Again by the Implicit Function Theorem, it is easily checked that, for γ sufficiently small, these intersections are, indeed, curves. We denote them by

$$E = h^-(\gamma) \text{ and } E = h^+(\gamma),$$

for $W^u(\mathcal{C}) \cap \{n = 1\}$ and $W^s(\mathcal{C}) \cap \{n = 1\}$ respectively, see Figure 4. The intersection is transversal if

$$M = \left(\frac{\partial h^+}{\partial \gamma} - \frac{\partial h^-}{\partial \gamma}\right) |_{\gamma=0} \neq 0. \tag{1.21}$$

This is a type of Melnikov calculation and can be checked to hold. The reader is asked to have faith that such a result holds, or can consult [53].

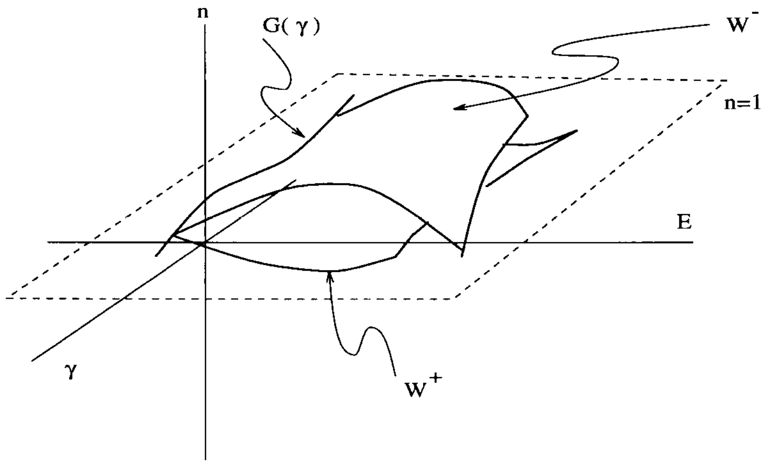

Figure 4
The The stable and unstable manifolds in $n = 1$.

It remains to show that, for some γ, there is a homoclinic orbit for (1.19). By the same argument as above the saddle perturbs for all γ and ϵ sufficiently small. We now think of the system (1.19) with equations for γ and ϵ appended. The unstable manifold of this surface of critical points, call it again W^u, will intersect its stable manifold W^s when $\gamma = 0$. We wish to find a curve of intersections given by γ as a function of ϵ. If W^u is given by $E = h^-(\gamma, \epsilon)$, and W^s by $E = h^+(\gamma, \epsilon)$, inside the set $n = 1$. We need to simultaneously solve the equations

$$\begin{aligned} E - h^-(\gamma,\epsilon) &= 0 \\ E - h^+(\gamma,\epsilon) &= 0, \end{aligned} \tag{1.22}$$

by E and γ as functions of ϵ. This can be achieved, by the Implicit Function Theorem, exactly when the determinant of the matrix

$$\begin{pmatrix} 1 & \frac{\partial h_-}{\partial \gamma} \\ 1 & \frac{\partial h_+}{\partial \gamma} \end{pmatrix}$$

is non-zero. But this is, again, exactly the condition $M \neq 0$. The transversality condition thus indeed supplies us with a homoclinic orbit.

A note on notation used in the examples is in order here. Many of the examples have the goal of finding travelling waves of a certain PDE. We usually use the variable ξ for the travelling wave variable i.e., $\xi = x - ct$, as the variable t has a meaning in the original PDE. However, we shall abuse the notation here and always revert to independent variables that conform to our general framework, once the ODE's in a given example are derived. As a result the variable t may have nothing to do with "time" in the original PDE.

The above examples are pleasing applications of GSP but do not use the perturbation that is supplied by the original equations to the equation on the slow manifold. Indeed, all the information is present in the limiting slow equation and it is merely checked that the object of interest, namely the homoclinic or heteroclinic orbit, is not destroyed by the perturbation. In the next subsection, we consider an application that actually uses the perturbing terms i.e., the order ϵ terms for the equations on M_ϵ to create the homoclinic orbit.

1.5 Solitary waves of the KdV-KS equation

We shall base this section on a paper by Ogawa, see [47]. The results, and approach, are very similar in spirit to the work of Ercolani et al., see [13], except that in the latter work periodic orbits are considered. The basic equations are a perturbed form of the Korteweg-deVries equations. The higher order terms perturbing the KdV part are characteristic of the Kuramoto-Sivashinsky equations, and the full model has arisen in a number of places, including in models of shallow water on tilted planes, see [56]. The partial differential equations are then

$$U_t + UU_x + U_{xxx} + \epsilon(U_{xx} + U_{xxxx}) = 0, \tag{1.23}$$

where $x \in \mathbf{R}$ and $t \geq 0$. We seek travelling wave solutions of (1.23). These will be solutions of (1.23) that are functions of the single variable $\xi = x - ct$. We are specifically interested in those that are asymptotic to the rest state $u = 0$ as $\xi \to \pm\infty$, these will then be solitary waves. The wave $U = U(\xi)$ must satisfy the ODE

$$-cU^{(1)} + UU^{(1)} + U^{(3)} + \epsilon\left(U^{(2)} + U^{(4)}\right) = 0, \tag{1.24}$$

where ${}^{(1)} = \frac{d}{d\xi}$. Using the boundary condition at $-\infty$, (1.24) can be integrated once to yield the equation

$$-cU + \frac{U^2}{2} + U^{(2)} + \epsilon\left(U^{(1)} + U^{(3)}\right) = 0. \tag{1.25}$$

This, in turn, we rewrite as a system of ODE's, wherein $u = U/c$

$$\begin{aligned} \dot{u} &= v \\ \dot{v} &= w \\ \epsilon\dot{w} &= \tfrac{1}{\sqrt{c}}\left(u - \tfrac{u^2}{2} - w - \tfrac{\epsilon}{\sqrt{c}}v\right), \end{aligned} \tag{1.26}$$

where $\dot{} = \frac{d}{d\tau}$ and $\tau = \sqrt{c}\xi$. Note the location of the small parameter ϵ means that (1.26) is already formulated on a slow time scale. The corresponding fast equations are

$$\begin{aligned} u' &= \epsilon v \\ v' &= \epsilon w \\ w' &= \tfrac{1}{\sqrt{c}}\left(u - \tfrac{u^2}{2} - w - \tfrac{\epsilon}{\sqrt{c}}v\right). \end{aligned} \tag{1.27}$$

The critical manifold M_0 is given by the conditions $w = u - \frac{u^2}{2}$ suitably restricted to any compact domain K of (u, v) space. Since (1.27) has only three equations, there is only one normal direction to this 2-dimensional manifold. The derivative of the fast part in the w-direction is $-1/\sqrt{c}$ and hence M_0 is normally hyperbolic, in fact attracting. Fenichel's Invariant Manifold Theorem then guarantees the existence of M_ϵ. The flow on M_ϵ is found by writing out (1.6) for this case

$$\begin{aligned} \dot{u} &= v \\ \dot{v} &= u - \tfrac{u^2}{2} + O(\epsilon). \end{aligned} \tag{1.28}$$

This equation has the limiting form, on M_0, of

$$\begin{aligned} \dot{u} &= v \\ \dot{v} &= u - \tfrac{u^2}{2}, \end{aligned} \tag{1.29}$$

which can be easily analyzed as it is a simple one-degree of freedom Hamiltonian system. Indeed, (1.29) has a homoclinic orbit to the critical point $(0, 0)$, see Figure 5.

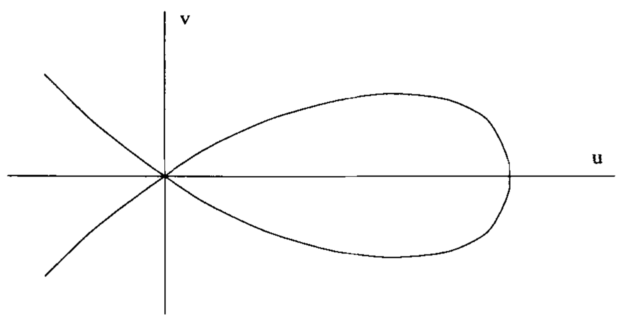

Figure 5
The homoclinic orbit in M_0.

However, it is inevitably not transversal. Moreover, the constant c offers no relief to this dilemma as it does not enter into (1.29). This is the point at which this example departs from its similarity to those of the preceding sections. We must then consider the $O(\epsilon)$ terms in (1.28). We know that M_ϵ is given by a function $w = h(u, v, \epsilon)$ and, by smoothness, can be expanded in ϵ, so that

$$w = u - \frac{u^2}{2} + \epsilon h_1(u, v) + O(\epsilon^2). \tag{1.30}$$

We need to calculate the term $h_1(u, v)$, which is also likely to depend on the parameter c. The only remaining information about M_ϵ is the local invariance relative to the equation and this must then be used to evaluate h_1.

To this end, we differentiate (1.30)

$$w' = u' - uu' + \epsilon\left(\frac{\partial h_1}{\partial u}u' + \frac{\partial h_1}{\partial v}v'\right) + O(\epsilon^2). \tag{1.31}$$

We substitute the expressions for u', v' and w', from (1.28), and also the expression for w, given by (1.30) into (1.31), and, after cancelling the $O(1)$ terms, we have

$$\frac{1}{\sqrt{c}}\left(-(\epsilon h_1 + O(\epsilon^2)) - \frac{\epsilon}{\sqrt{c}}v\right) = \epsilon v - \epsilon uv + O(\epsilon^2). \tag{1.32}$$

Equating the terms of $O(\epsilon)$ in (1.32) we obtain

$$h_1 = \sqrt{c}\left(uv - v\left(1 + \frac{1}{\sqrt{c}}\right)\right). \tag{1.33}$$

We cannot expect (1.28) to have a homoclinic orbit forced merely by adding the $O(\epsilon)$ term. The parameter c will also need to be used, in the jargon: it is a codimension two problem. We shall augment the system (1.28) with both equations for ϵ and c

$$\begin{aligned} \dot{u} &= v \\ \dot{v} &= u - \tfrac{u^2}{2} + \epsilon\sqrt{c}\left(u - \left(1 + \tfrac{1}{\sqrt{c}}\right)\right)v + O(\epsilon^2) \\ \dot{\epsilon} &= 0 \\ \dot{c} &= 0. \end{aligned} \tag{1.34}$$

We seek homoclinic orbits for (1.34) with small ϵ. These will be found at values of c that depend on ϵ. From the original equations one can see that 0 remains a critical point and must lie on M_ϵ (why?). We thus look for orbits homoclinic to 0. The critical point 0 can, in reference to (1.34), be construed as a surface of critical points, say $\mathcal{S}$, parametrized by c, ϵ. This in turn spawns an unstable manifold $W^u(\mathcal{S})$ and stable manifold $W^s(\mathcal{S})$ which meet in a curve at $\epsilon = 0$, namely the homoclinic orbits found already, see Figure 5. In the set $v = 0$ we parametrize W^u and W^s respectively, near the intersection away from the critical point, as $u = h^-(c, \epsilon)$ and $u = h^+(c, \epsilon)$.

We next define

$$d(c, \epsilon) = h^-(c, \epsilon) - h^+(c, \epsilon),$$

and observe that zeroes of d render homoclinic orbits. Since there are homoclinic orbits independently of c when $\epsilon = 0$, we have that $d(c, 0) = 0$, and thus that $d(c, \epsilon) = \epsilon\tilde{d}(c, \epsilon)$. The Melnikov function is here given by

$$\tilde{d}(c, 0) = M(c) = \left(\frac{\partial h^+}{\partial \epsilon} - \frac{\partial h^-}{\partial \epsilon}\right)|_{\epsilon=0}. \tag{1.35}$$

It is a simple application of the Implicit Function Theorem to see that there is a curve of homoclinic orbits given by $c = c(\epsilon)$ for ϵ small, if there exists a c $(= c(0))$, at which

$$M(c) = 0 \text{ and } M'(c) \neq 0. \tag{1.36}$$

The function $M(c)$ can be calculated explicitly, see Lecture 3 below, as

$$M(c) = \frac{1}{\alpha}\left\{\frac{1}{\sqrt{c}}\left\{c\int_{-\infty}^{+\infty} \ddot{u}^2 - \int_{-\infty}^{+\infty} \dot{u}^2\right\}\right\}, \tag{1.37}$$

where u comes from the underlying, already known, homoclinic orbit and $\alpha \neq 0$. It is clear then that (1.36) at a unique value of c.

It is interesting to note in this application that the perturbing terms supply a speed selection that is not evident without them. The reduced equations i.e., those on M_0, are the travelling wave equations for the KdV equation. No particular speed for this equation is determined by the travelling waves as they exist at every speed. However, when the perturbing terms, supplied by the Kuramoto-Sivashinsky formulation, are added a specific wave speed is selected.

Chapter 2

Invariant Manifold Theorems

Fenichel's Theorem as stated in the first lecture gives us only part of the picture in a neighborhood of a slow manifold. The existence of the slow manifold is guaranteed by the Theorem and the equation on the manifold can be computed as shown in the examples. However, at this point, we know nothing of the flow off the slow manifold and this must now be addressed. Our goal here is the derivation of a normal form, that we shall call "Fenichel Normal Form" for the equations near a slow manifold. This goal will be reached in the third lecture. In this lecture, Fenichel's Second Theorem that describes the stable and unstable manifolds of a slow manifold will be presented. These are perturbations of the stable and unstable manifolds, respectively, of the critical manifold. They are related to those invariant manifolds in the same way that the slow manifold is related to the critical manifold.

2.1 Stable and unstable manifolds

The slow manifold discussed in the first lecture possesses attendant stable and unstable manifolds that are perturbations of the corresponding manifolds when $\epsilon = 0$. The following theorem holds under (H1)-(H3) and its conclusion is depicted in Figure 6.

Theorem 3 (Fenichel Invariant Manifold Theorem 2) *If $\epsilon > 0$ but sufficiently small, there exist manifolds $W^s(M_\epsilon)$ and $W^u(M_\epsilon)$ that lie within $O(\epsilon)$ of, and are diffeomorphic to, $W^s(M_0)$ and $W^u(M_0)$ respectively. Moreover, they are each locally invariant under (1.1), and C^r, including in ϵ, for any $r < +\infty$.*

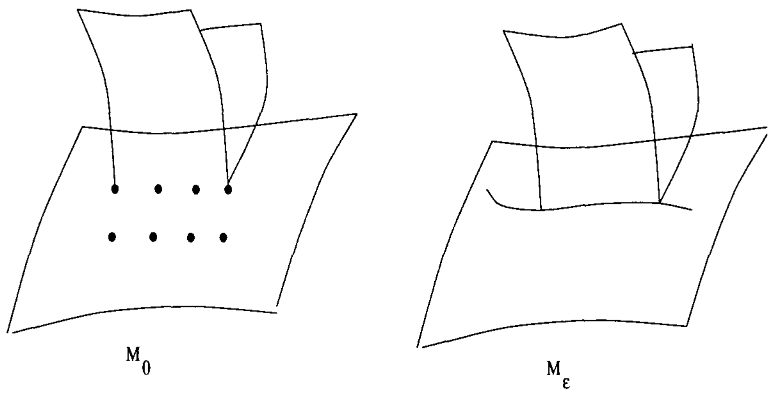

Figure 6
The slow manifold and its stable and unstable manifolds.

The Fenichel Normal Form will be produced through a series of coordinate changes. Initially, these will be made with the purpose of facilitating the proofs of the theorems. The final set of coordinate changes will follow from the theorems themselves. We shall restrict, as stated earlier, to the case of critical manifolds that are given as graphs.

Without loss of generality, we can assume that $h^0(y) = 0$ for all $y \in K$. Indeed, we can replace x by $\tilde{x} = x - h^0(y)$ and recompute the equations. For each point $y \in K$, there are subspaces $S(y)$ and $U(y)$, corresponding, respectively, to stable and unstable eigenvalues. Since the eigenvalues are bounded uniformly away from the imaginary axis over K, the dimensions of $S(y)$ and $U(y)$ are independent of y. Let $\dim S(y) = m$ and $\dim U(y) = k$. Since K is simply-connected by (H3), we can smoothly choose bases for $S(y)$ and $U(y)$. Changing the coordinates to be in terms of these new bases, we can set $x = (a, b)$, where $a \in \mathbf{R}^k$ and $b \in \mathbf{R}^m$, so that our equations have the form

$$\begin{aligned} a' &= A(y)a + F_1(x, y, \epsilon) \\ b' &= B(y)b + F_2(x, y, \epsilon) \\ y' &= \epsilon g(x, y, \epsilon), \end{aligned} \tag{2.1}$$

where the spectrum of the matrix $A(y)$ lies in the set $\{\lambda : \mathcal{R}(\lambda) > 0\}$ and the spectrum of the matrix $B(y)$ lies in $\{\lambda : \mathcal{R}(\lambda) < 0\}$. Both F_1 and F_2 are higher order in x and ϵ; to be precise, we have the estimates

$$|F_i| \leq \gamma(|x| + \epsilon), \tag{2.2}$$

$i = 1, 2$ and γ can be taken to be as small as desired by restricting to a set with $|a|$ and $|b|$ small.

With this notation established, we can determine $W^s(M_\epsilon)$ and $W^u(M_\epsilon)$ as graphs and give the following restatement of Theorem 3

Theorem 4 *If $\epsilon > 0$ is sufficiently small, then, for some $\Delta > 0$,*

(a) *there is a function* $a = h_s(b, y, \epsilon)$ *defined for* $y \in K$ *and* $|b| \leq \Delta$, *so that the graph*

$$W^s(M_\epsilon) = \{(a, b, y) : a = h_s(b, y, \epsilon)\}$$

is locally invariant under (2.1). Moreover, $h_s(b, y, \epsilon)$ *is* C^r *in* (b, y, ϵ) *for any* $r < +\infty$.

(b) *there is a function* $b = h_u(a, y, \epsilon)$ *defined for* $y \in K$ *and* $|a| \leq \Delta$, *so that the graph*

$$W^u(M_\epsilon) = \{(a, b, y) : b = h_u(a, y, \epsilon)\}$$

is locally invariant under (2.1). Moreover, $h_u(a, y, \epsilon)$ *is* C^r *in* (a, y, ϵ) *for any* $r < +\infty$.

These theorems also apply when $\epsilon = 0$ and render the stable and unstable manifolds of the known critical manifold, the existence of which is also guaranteed by the usual stable and unstable manifold theorems at critical points (their smooth variation in y requires a little work to show but follows from Theorem 4). These latter two theorems then assert that these manifolds perturb. At this point, there is little justification for naming these manifolds as stable and unstable, other than their status gained as perturbations of the $\epsilon = 0$ case. It will be seen below, see Theorem 5, that they enjoy certain decay and growth estimates respectively.

Theorem 1 can be concluded from Theorem 3 by taking the intersection of $W^s(M_\epsilon)$ with $W^u(M_\epsilon)$. Locally, the Implicit Function Theorem gives the intersection as a graph, and these functions can be patched together since K is a compact set. Moreovwe, we need only give the construction of the stable manifold, as that of the unstable manifold follows immediately by a reversal of time. The proof to be given is very geometric in flavor and is based on the use of cones. The immediate result will be of a Lipschitz manifold and the smoothness proof for these manifolds will only be sketched here.

It is appropriate at this point to say some words about the history of these invariant manifold theorems, although one can only address such a task in an incomplete manner. There are two approaches taken to proving invariant manifold theorems and both have an extensive history. The first is that due to Hadamard, see [21], and relies on the geometry present in the splitting due to the decay rates. The second approach is due to Perron, see [48], and is based on proving the existence of the invariant manifold as a fixed point of a certain integral equation. Fenichel adopted the Hadamard approach in his seminal papers, see [15, 16, 17, 18]. These lectures are, to a great extent, based on his adaptation of the method to the case of singularly perturbed ODE's [18]. Simultaneous to Fenichel's work, Hirsch, Pugh and Shub [25] used the more analytic approach to achieve related results. Sakamoto [51] used the Lyapounov-Perron approach to derive Fenichel's results. An extensive exposition of Fenichel's Theorems, as well as their proofs, are given by Wiggins [58]. Other results in this direction have been obtained by many different authors including, but not limited to, Knobloch, Lin and Szmolyan.

The proofs given here are an extreme geometric version of Fenichel's. Many of the ideas lying behind the proofs were learnt by the author from Conley in his lectures on dynamical systems at Wisconsin. They have also been used by McGehee [42] and Bates and Jones [3] in the case of single fixed points.

2.2 Preparation of equations

The set K will need to be somewhat enlarged. Since M_0 is given by h^0, which is assumed to be C^∞ on K, which is compact, a set $\hat{K}$ can be found so that $K \subset \text{int}\hat{K}$ and h^0 is defined, and C^∞ for all $y \in \hat{K}$. Moreover, $\hat{M}_0 = \left\{(x,y) : x = h^0(y), ; y \in \hat{K}\right\}$ is a set of critical points and we can choose $\hat{K}$ so that $\hat{M}_0$ is normally hyperbolic.

The equations will be further prepared before the proofs can be given. The coordinates given above in terms of a and b separate the stable and unstable parts, but do not necessarily give good estimates for decay and growth. We set the quantities $\lambda_+ > 0$ and $\lambda_- < 0$ so that

$$\lambda_+ < \mathcal{R}(\lambda) \text{ for any } \lambda \in \sigma\left(A(y)\right) \text{ and } y \in \hat{K}, \tag{2.3}$$

$$\lambda_- > \mathcal{R}(\lambda) \text{ for any } \lambda \in \sigma\left(B(y)\right) \text{ and } y \in \hat{K}, \tag{2.4}$$

We shall refine the coordinates for a and b so that appropriate decay estimates on the linear parts are exposed.

Lemma 1 *Coordinates can be chosen so that, in the new inner product, the following estimates hold*

$$< a, A(y)a > \geq \lambda_+ < a, a >, \tag{2.5}$$

$$< b, B(y)b > \leq \lambda_- < b, b > . \tag{2.6}$$

Proof The coordinates can be found locally using ϵ-Jordan form. These can be patched together over all of K using a partition of unity.

In all proofs of the Center Manifold Theorem a modification has to be made to the equation in the center directions. This serves the purpose of mitigating its neutral character. We must perform the same modification here to deal with the slow directions, which are, effectively, center directions. The set $\hat{K}$ can further be chosen so that its boundary is given by the condition $\hat{\nu}(y) = 0$ for some C^∞ function $\hat{\nu}(y)$ and $\hat{\nu}(y)$ satisfies $\nabla\hat{\nu}(y) \neq 0$ for all $y \in \partial\hat{K}$. The function $\hat{\nu}(y)$ is assumed to have been normalized so that $\nabla\hat{\nu}(y) = n_y$ is a unit outward normal for $\partial\hat{K}$. We let $\rho(y)$ be a C^∞ function that has the following values

$$\rho(y) = \begin{cases} 1 & \text{if} \quad y \in \hat{K}^c, \\ 0 & \text{if} \quad y \in K, \end{cases} . \tag{2.7}$$

The existence of such a function can be achieved locally with a C^∞ bump function and then the full function is created with a partition of unity. We now modify the third equation of (2.1) by adding the term $\delta\rho(y)n_y$, where δ is some number that remains to be chosen.

We shall need to append an equation for the small parameter ϵ. However, for the purpose of making estimates later, we shall actually use a multiple of ϵ as the new auxiliary variable. Thus, we set $\epsilon = \eta\sigma$ and append the equation $\eta' = 0$ to the system (2.1). We then arrive at the system

$$\begin{aligned} a' &= A(y)a + F_1(x,y,\epsilon) \\ b' &= B(y)b + F_2(x,y,\epsilon) \\ y' &= \eta\sigma g(x,y,\epsilon) + \delta\rho(y)n_y \\ \eta' &= 0, \end{aligned} \tag{2.8}$$

in which it is understood that x is a function of a and b, and ϵ is a function of η. Clearly if Theorem 4 is restated with ϵ replaced by η and proved in that formulation, the original version of Theorem 4 can easily be recaptured by substituting ϵ back in.

2.3 Proof of Theorem 4

We are now ready for the proof of Theorem 4. The strategy will be to find a function $a = h_s(b,y,\eta)$ defined for $y \in \hat{K}$ and then restrict it to K, where the new equation agrees with the old as $\rho = 0$ in K.

Proof The first step is to set the neighborhood of $\hat{M}_0$ in which we shall work. The neighborhood will be called $\hat{D}$ and is determined by the conditions:

$$y \in \hat{K},\ |a| \le \Delta,\ |b| \le \Delta,\ \eta \in [0,\eta_0].$$

Define the set

$$\Gamma_s = \left\{(a,b,y,\eta) : (a,b,y,\eta)\cdot t \in \hat{D},\ \text{for all } t \ge 0\right\}. \tag{2.9}$$

We shall prove that Γ_s is the graph of a function given by a in terms of the remaining variables and this will be the function $h_s(b,y,\eta)$. The next step is to show that Γ_s contains the graph of a function. Set $\zeta = (b,y,\eta)$ and let $\hat{D}_{\hat{\zeta}}$ denote the cross-section of $\hat{D}$ at fixed $\zeta = \hat{\zeta}$.

$$\hat{D}_{\hat{\zeta}} = \left\{(a,\hat{\zeta}) : |a| \le \Delta\right\},$$

as depicted in Figure 7.

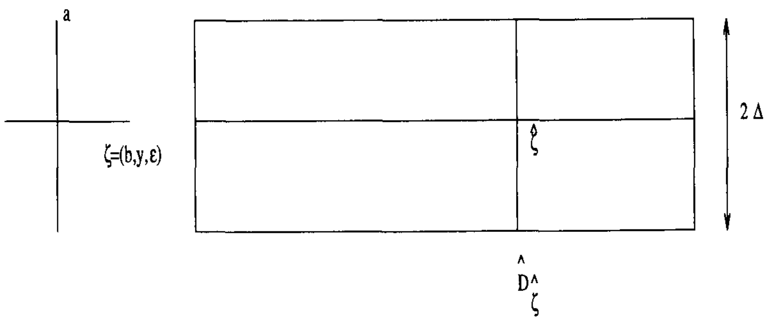

Figure 7
The neighborhood $\hat{D}$ and the cross-section.

We need to show that there is at least one point $(a, \hat{\zeta}) \in \hat{D}$ for which $(a, \hat{\zeta})\cdot t \in \hat{D}$ for all $t \geq 0$. To achieve this, the Wazewski Principle is used. Let $\hat{D}^+$ be the immediate exit set of $\hat{D}$ and $\hat{D}^0$ be the eventual exit set. If $\hat{D}^+$ is closed relative to $\hat{D}^0$ then $\hat{D}$ is called Wazewski set and the map $W : \hat{D}^0 \to \hat{D}^+$, that takes each point to the first from which it exits $\hat{D}$, is continuous. We need to check the boundary of $\hat{D}$ and find the immediate exit set.

$|a| = \Delta$:

$$\begin{aligned} < a, a >' &= 2\left\{< a, A(y)a > + < a, F_1 >\right\}, \\ &\geq 2\left\{\lambda_+ \Delta^2 - \Delta\gamma(\Delta + \epsilon_0)\right\}, \end{aligned} \tag{2.10}$$

using Lemma 1 and (2.2). If ϵ_0 is chosen less than Δ, we see that

$$< a, a >' \geq 2(\lambda_+ - 2\gamma)\Delta^2 > 0, \tag{2.11}$$

if γ is chosen small enough, which can be achieved by choosing Δ and ϵ_0 sufficiently small. The set $|a| = \Delta$ is then a part of the immediate exit set, as $|a|$ increases there.

$|b| = \Delta$:
It can been similarly that, with small enough Δ and ϵ_0,

$$< b, b >' < 0.$$

$y \in \partial\hat{K}$:

$$< y', n_y > = \epsilon < g(x, y, \epsilon), n_y > + \delta < n_y, n_y >,$$

since on $\partial\hat{K}$ $\rho = 1$. Setting $M = \sup_{\hat{D}} \{|g|, |Dg|\}$, the above can be estimated as

$$< y', n_y > \geq \delta - \epsilon_0 M > 0,$$

if $\delta > \epsilon_0 M$, which can be assumed as δ was arbitrary.

$\eta = 0$ or $\eta = \eta_0$:
both of these sets are invariant and thus render neither entrance nor exit sets.

The immediate exit set is thus seen to be the set $|a| = \Delta$, and $\hat{D}$ is easily checked to be a Wazewski set. The set $\hat{D}_{\hat{\zeta}}$ is a ball of dimension k. Suppose that $\Gamma_s \cap \hat{D}_{\hat{\zeta}} = \emptyset$, then $\hat{D}_{\hat{\zeta}} \subset \hat{D}^0$ and $\hat{D}_{\hat{\zeta}}$ lies in the domain of the Wazewski map W, so, restricting W, we have

$$W : \hat{D}_{\hat{\zeta}} \to \hat{D}^+ = \{|a| = \Delta\}.$$

If we follow this by a projection $\pi(a, \zeta) = a$, we see that $\pi \circ W$ maps a k- ball onto its boundary, while keeping that boundary fixed. This contradicts the No-Retract Theorem, which is equivalent to the Brouwer Fixed-Point Theorem, see, for instance, [41]. Thus there is a point in $\hat{D}_{\hat{\zeta}} \cap \Gamma_s$. Since $\hat{\zeta}$ was arbitrary, this gives, at least, one value for a as a function of (b, y, η), and we name it $h_s(b, y, \eta)$.

The next step will be to show that the graph of the above derived function is all of Γ_s. At the same time, it will be shown that the function is, in fact, Lipschitz with Lipschitz constant equal to 1. A comparison between the growth rates in different directions will be derived in the next lemma. Let $(a_i(t), \zeta_i(t))$, $i = 1, 2$ be two solutions of (2.8), set $\Delta a = a_2(t) - a_1(t)$ and $\Delta\zeta = \zeta_2(t) - \zeta_1(t)$. Further, we define

$$M(t) = |\Delta a|^2 - |\Delta\zeta|^2 .$$

Lemma 2 *If $M(t) = 0$ then $M'(t) > 0$, as long as the two solutions stay in $\hat{D}$, unless $\Delta a = 0$.*

Proof The lemma follows from estimates that we will make on each of the quantities $< \Delta a, \Delta a >$ etc. The equation for Δa is

$$\Delta a' = A(y_2)a_2 - A(y_1)a_1 + F_1(x_2, y_2, \eta_2\sigma) - F_1(x_1, y_1, \eta_1\sigma), \tag{2.12}$$

which we rewrite as

$$\Delta a' = A(y_2)\Delta a + [A(y_2) - A(y_1)]\, a_1 + \Delta_x F_1 + \Delta_y F_1 + \Delta_\epsilon F_1, \tag{2.13}$$

where

$$\Delta_x F_1 = F(x_2, y_2, \epsilon_2) - F_1(x_1, y_2, \epsilon_2),$$
$$\Delta_y F_1 = F_1(x_1, y_2, \epsilon_2) - F_1(x_1, y_1, \epsilon_2)$$

and

$$\Delta_\epsilon F_1 = F_1(x_1, y_1, \epsilon_2) - F_1(x_1, y_1, \epsilon_1).$$

Using the fact that F_1 involves only higher order terms, one can derive the estimates

$$|\Delta_x F_1| \leq \gamma\, |\Delta x|\,, \tag{2.14}$$

and

$$|\Delta_\epsilon F_1| \leq \sigma\gamma\,|\Delta\eta|\,, \tag{2.15}$$

wherein γ can be made as small as desired by reducing the defining parameters of $\hat{D}$. Since $F_1(0, y, 0) = 0$, we can write

$$F_1(x, y, \epsilon) = x\tilde{F}_1(x, y, \epsilon) + \epsilon\hat{F}_1(x, y, \epsilon),$$

from which we obtain the estimate

$$|\Delta_y F_1| \leq C\,\{\Delta\,|\Delta y| + \epsilon_0\,|\Delta y|\}\,. \tag{2.16}$$

We can estimate $< \Delta a, \Delta a >' = 2 < \Delta a', \Delta a >$ by taking the inner product of (2.13) with Δa. Each term can then be estimated using Lemma 1, (2.14), (2.16), (2.15) and the continuity of A in y. We then obtain

$$\begin{aligned} < \Delta a', \Delta a > \quad & \geq \lambda_+\,|\Delta a|^2 - \{c_1\Delta\,|\Delta y|\,|\Delta a| + \gamma\,|\Delta x|\,|\Delta a| + \\ & c_2(\Delta + \epsilon_0)\,|\Delta y|\,|\Delta a| + \sigma\gamma\,|\Delta\eta|\,|\Delta a|\}\,, \end{aligned} \tag{2.17}$$

for some constants c_1 and c_2. We can bound $|\Delta x|$ by $c_3\,(|\Delta a| + |\Delta b|)$ and each term with $|\Delta b|$, $|\Delta y|$ or $|\Delta\eta|$ by $|\Delta\zeta|$. The estimate (2.17) can then be written as

$$< \Delta a, \Delta a >' \geq 2\,(\lambda_+ - \beta_1)\,|\Delta a|^2 - \beta_2\,|\Delta a|\,|\Delta\zeta|\,, \tag{2.18}$$

where β_1 and β_2 can be made small. If $M(t) = 0$, we can then replace $|\Delta\zeta|$ throughout by $|\Delta a|$. The net result is that

$$< \Delta a, \Delta a >' \geq 2\,\{\lambda_+ - (\beta_1 + \beta_2)\}\,|\Delta a|^2\,. \tag{2.19}$$

Note that if Δ and ϵ_0 are chosen sufficiently small, then the coefficient of $|\Delta a|^2$ in (2.19) can be made positive, say greater than $\lambda_+ - \beta$.

The estimate on $|\Delta\zeta|$ must be broken down into pieces. In a similar fashion to the above we can estimate

$$< \Delta b, \Delta b >' \leq 2\,\{\lambda_- + \alpha\}\,|\Delta a|^2\,, \tag{2.20}$$

where α can be made as small as desired by choosing the parameters of the neighborhood small. We can write, as above,

$$\begin{aligned} \Delta y' = \quad & \sigma\Delta\eta g(x_2, y_2, \eta_2\sigma) + \eta\sigma\,\{\Delta_x g + \Delta_y g + \Delta_\epsilon g\} \\ & +\delta\,(\rho(y_2)n_{y_2} - \rho(y_1)n_{y_1})\,. \end{aligned} \tag{2.21}$$

Moreover, the terms in parentheses can be estimated as follows:

$$|\Delta_x g| \leq \hat{\gamma}\,|\Delta x|\,, \tag{2.22}$$

$$|\Delta_y g| \leq M\,|\Delta y|\,, \tag{2.23}$$

$$|\Delta_\epsilon g| \leq \sigma\hat{\gamma}\,|\Delta\eta|\,, \tag{2.24}$$

where $\hat{\gamma}$ can be made as small as desired by reducing the size of the neighborhood. The following estimate can then be deduced

$$\begin{aligned} < \Delta y', \Delta y > \quad & \leq \sigma M \left|\Delta \eta\right| \left|\Delta y\right| + \eta_1 \sigma \left\{\hat{\gamma} c_3(\left|\Delta a\right| + \left|\Delta b\right|) + \right. \\ & \left. M \left|\Delta y\right| + \sigma \hat{\gamma} \left|\Delta \eta\right|\right\} \left|\Delta y\right| + c_4 \left|\Delta y\right|^2 , \end{aligned} \tag{2.25}$$

and, again in a similar fashion to the above estimates, we can conclude that

$$< \Delta y, \Delta y >' \leq 2\sigma \hat{C} \left|\Delta a\right|^2 . \tag{2.26}$$

Combining the estimates (2.19), (2.20), (2.26) and the fact that $\Delta \eta' = 0$, we conclude that

$$M'(t) \geq \left\{\lambda_+ - \beta - \left(\lambda_- + \alpha + \sigma \hat{C}\right)\right\} \left|\Delta a\right|^2 , \tag{2.27}$$

when $M(t) = 0$. The coefficient of the right hand side is

$$\lambda_+ - \lambda_- - \left(\beta + \alpha + \sigma \hat{C}\right) .$$

Since α and β can be made as small as desired by adjusting the parameters of the neighborhood and σ can be made small, this quantity is positive and the Lemma follows.

The decomposition of ϵ into η and σ is used at this last step. It seems somewhat artificial but, in fact, is not, as it corresponds to imposing an "ϵ-Jordan form" on the neutral directions coming from y and ϵ.

The Lemma can be interpreted in terms of "moving cones". This notion will be important in the next lecture and so it is worth exposing further at this point. Define the cone

$$\mathcal{C} = \{(a, \zeta) : |a| \geq |\zeta|\} , \tag{2.28}$$

and then Lemma 2 can be restated in terms of $\mathcal{C}$ as follows: If $z_2 \in z_1 + \mathcal{C}$ then $z_2 \cdot t \in z_1 \cdot t + \mathcal{C}$ so long as $z_2 \cdot t$ and $z_1 \cdot t$ stay in $\hat{D}$, see Figure 8.

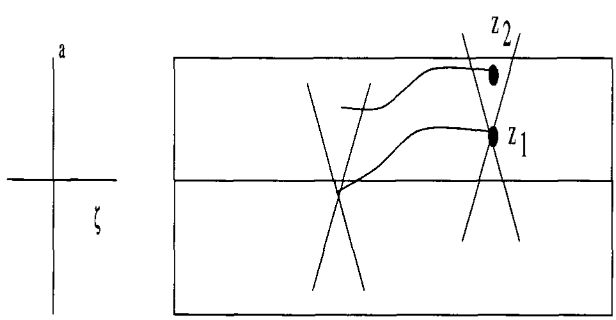

Figure 8
The moving cones.

The proof of Theorem 4 can now be completed, at least for the Lipschitz case. The set Γ_s will be our stable manifold. We have shown that it contains the graph of a function, which we denote by $a = h_s(b, y, \eta)$. Suppose that Γ_s contains more than one point with the same values of b, y and η. There would then be a_1 and a_2 so that both (a_1, b, y, η) and (a_2, b, y, η) lie in Γ_s. At $t = 0$, we would then have $|\Delta a| \geq |\Delta\zeta|$. By Lemma 2,

$$|\Delta a(t)| \geq |\Delta\zeta(t)|,$$

for all $t \geq 0$. In the estimate (2.18) we can then replace $|\Delta\zeta|$ by $|\Delta a|$ to obtain

$$\left\{|\Delta a|^2\right\}' \geq (\lambda_+ - \beta)\,|\Delta a|^2 .$$

From which it can be easily concluded that Δa grows exponentially, which contradicts the hypothesis that both points stay in $\hat{D}$ for all $t \geq 0$.

The same argument can be used to show that h_s is also Lipschitz. If (a_1, ζ_1) and (a_2, ζ_2) are both in Γ_s and $|a_2 - a_1| \geq |\zeta_2 - \zeta_1|$, then $|a_2 - a_1|$ can be seen to grow exponentially, contradicting the hypothesis again that both points lie in Γ_s. We have now shown that the set Γ_s is the graph of a Lipschitz function. This manifold is $W^s(M_\epsilon)$, when y is restricted to the set K, in which the modified equation agrees with the original.

2.4 Decay estimates

Some justification should be given to the terminology "stable manifold". Since the base manifold M_ϵ no longer consists of equilibria, we cannot characterize $W^s(M_\epsilon)$ as the stable manifold of a set of critical points, however we can say that the solutions in $W^s(M_\epsilon)$ will decay to M_ϵ at an exponential rate, with the caveat that the decay will only last as long as the solution under consideration stays in the neighborhood D. In the following $d(.,.)$ is the Euclidean distance.

Theorem 5 *There are $\kappa_s > 0$ and $\alpha_s < 0$ so that if $v \in W^s(M_\epsilon)$ and $v \cdot [0, t] \subset D$, with $t > 0$, then*

$$d(v \cdot t, M_\epsilon) \leq \kappa_s \exp\{\alpha_s t\}. \tag{2.29}$$

Furthermore, there are $\kappa_u > 0$ and $\alpha_u > 0$ so that if $v \in W^u(M_\epsilon)$ and $v \cdot [t, 0] \subset D$, with $t < 0$, then

$$d(v \cdot t, M_\epsilon) \leq \kappa_u \exp\{\alpha_u t\}. \tag{2.30}$$

The proof follows from the results of the next lecture.

Chapter 3

Fenichel Normal Form

In this lecture, the third Theorem of Fenichel will be presented. This result gives a more detailed picture of the structure of the flow on the stable and unstable manifolds. An application will be given to settling in a cellular flow field in which this extra structure is used to show that the full flow in question is slaved in a precise manner to a two-dimensional flow. This two-dimensional flow can be analysed to reveal the salient features of the flow.

Of central interest will be, however, the derivation of a normal form for singularly perturbed equations in the neighborhood of a slow manifold. The derivation of this normal form, which we call Fenichel Normal Form, will rest on Fenichel's Third Theorem and will be the main goal of this chapter.

We shall thus focus on the proof of the existence of the stable manifold in Theorem 3, the proof for the unstable manifold follows immediately by a reversal of time.

3.1 Smoothness of invariant manifolds

The change of variables will need to be smooth and thus we need that the manifolds constructed in the Fenichel Theorems are smooth. In order to prove the smoothness of the invariant manifolds, the variational equation is used. This procedure will be sketched without details. The equation of variations of (2.1) (with an equation for ϵ and the modification added), which was the last version of the equation before Fenichel's Theorems were used, is given by

$$\begin{aligned} \delta a' &= A(y)\delta a + D\left(A(y)a\right)\delta y + DF_1\delta z \\ \delta b' &= B(y)\delta b + D\left(B(y)b\right)\delta y + DF_2\delta z \\ \delta y' &= \epsilon Dg\delta z + g\delta\epsilon + \delta\rho'(y)\delta y \\ \delta\epsilon' &= 0, \end{aligned} \tag{3.1}$$

where $\delta z = (\delta a, \delta b, \delta y, \delta\epsilon)$. We imagine coupling this with the underlying equation to achieve a system in $\mathbf{R}^{2N+2}$. The linearization of this big system at a point in M_0 is easily seen to be in block form, each block having dimension $N+1$. This matrix has the form

$$\begin{pmatrix} A(y) & & & & & & & \\ & B(y) & & & & & & \\ & & 0 & \sigma g(\hat{x},\hat{y},0) & & & & \\ & & & 0 & & & & \\ & & & & A(y) & & & \\ & & & & & B(y) & & \\ & & & & & & 0 & \sigma g(\hat{x},\hat{y},0) \\ & & & & & & & 0 \end{pmatrix}, \tag{3.2}$$

with 0's in all the vacant places.

A perusal of (3.2) gives that the block associated with $(\delta a, \delta b, \delta y, \delta\epsilon)$ is, in fact, exactly the same as the linearization of (2.1) at a point in M_0. The strategy for smoothness is then to reconstruct the invariant manifolds for this larger system, taking care to balance the unboundedness due to the variational equation by its linearity. In particular, for the stable manifold, this renders a function

$$\delta a = H_s(a,b,y,\epsilon,\delta b,\delta y,\delta\epsilon),$$

and the proof follows by showing that H_s is, in fact, the derivative of h_s i.e.,

$$H_s(a,b,\epsilon,\delta b,\delta y,\delta\epsilon) = Dh_s(a,b,y,\epsilon)(\delta a,\delta b,\delta y,\delta\epsilon).$$

This is achieved by using the characterization of the manifolds in terms of the set Γ_s and its uniqueness properties.

3.2 Straightening the invariant manifolds

The first part of Fenichel Normal Form can be implemented from the theorems already proved. Indeed, we shall straighten out the stable and unstable manifolds of the slow manifold M_ϵ. Using the functions that give these manifolds, we shall transform them to coordinate planes. First, set

$$\begin{aligned} a_1 &= a - h_s(b,y,\epsilon), \quad & b_1 &= b, \\ y_1 &= y, & \epsilon_1 &= \epsilon, \end{aligned} \tag{3.3}$$

which has the effect of transforming $W^s(M_\epsilon)$ to the subspace $a_1 = 0$. This transformation is invertible by inspection and is as smooth as h_s. Next, set

$$\begin{aligned} a_2 &= a_1, \quad & b_2 &= b_1 - h_u(a_1 + h_s(b_1,y_1,\epsilon_1), b_1, \epsilon_1), \\ y_2 &= y_1, & \epsilon_2 &= \epsilon_1, \end{aligned} \tag{3.4}$$

which has the effect of moving $W^u(M_\epsilon)$ to the subspace $b_2 = 0$. This latter transformation can also be checked to be invertible, using the fact that $W^s(M_\epsilon)$ is tangent to $b = 0$ along M_0, and obviously as smooth as h_u and h_s. We shall drop the subscripts and revert to the notation (a,b,y,ϵ) for a point in the new coordinate system. We thus have that the sets $a = 0$ and $b = 0$ are invariant in D, it thus must follow that $a = 0$ implies $a' = 0$ and $b = 0$ implies $b' = 0$. This imposes a certain character on the equations. Indeed, the variable a can be factored out of the equation for a', and analogously for b'. We thus arrive at

$$\begin{aligned} a' &= \Lambda(a, b, y, \epsilon)a \\ b' &= \Gamma(a, b, y, \epsilon)b \\ y' &= \epsilon g(a, b, y, \epsilon), \end{aligned} \tag{3.5}$$

where Λ and Γ are matrices with $\Lambda(0, 0, y, 0) = A(y)$ and $\Gamma(0, 0, y, 0) = B(y)$. Note that the function g has been transformed appropriately and, with an abuse of notation, is also denoted by g. The matrix Λ therefore inherits the spectral properties of A and Γ those of B, if a, b and ϵ are all sufficiently small. As a side benefit of this coordinate change, we have obtained that M_ϵ is given by $a = b = 0$.

3.3 Fenichel fibering

The above change of coordinates has refined the stable and unstable directions to the point that estimates can be easily invoked from the linearized system. We still need to refine the equations for the slow directions. This involves using what has become known as "Fenichel Fibering", see, for instance, Wiggins [57]. It can be motivated by asking a question: we have seen that the stable and unstable manifolds of M_0 perturb to analogous objects when ϵ is sufficiently small, do the individual stable and unstable manifolds of points in M_0 also perturb? The answer would appear to be negative as the base points themselves do not perturb as critical points. However, this judgement is premature.

A minor technical difficulty arises here on account of the modification that we have performed to the equations. The equations (2.8) agree in D with the original equation (1.1) and we can restrict our attention to that set. However, points may leave D but re-enter it at a later time. Once trajectories of (2.8) have left D, their evolution is no longer governed by the original equation and they are no longer of interest. We must, therefore, restrict attention to the solutions while they are only in D. To facilitate this discussion, we need a definition.

Definition 3 *The forward evolution of a set $A \subset D$ restricted to D is given by the set*

$$A \cdot_D t = \{x \cdot t : x \in A \text{ and } x \cdot [0, t] \subset D\}.$$

With this definition in hand, we can state Fenichel's Third Invariant Manifold Theorem. In the following $v_\epsilon \in M_\epsilon$ is smooth in ϵ, including $\epsilon = 0$, and we are assuming (H1)-(H3).

Theorem 6 (Fenichel Invariant Manifold Theorem 3) *For every $v_\epsilon \in M_\epsilon$, there is an m-dimensional manifold*

$$W^s(v_\epsilon) \subset W^s(M_\epsilon),$$

and an l-dimensional manifold

$$W^u(v_\epsilon) \subset W^u(M_\epsilon),$$

lying within $O(\epsilon)$ of, and diffeomorphic to, $W^s(v_0)$ and $W^u(v_0)$ respectively. Moreover, they are C^r for any r, including in v and ϵ. The family $\{W^s(v_\epsilon) : v_\epsilon \in M_\epsilon\}$ is invariant in the sense that

$$W^s(v_\epsilon) \cdot_D t \subset W^s(v_\epsilon \cdot t), \tag{3.6}$$

if $v_\epsilon \cdot s \in D$ for all $s \in [0,t]$, and the family $\{W^u(v_\epsilon) : v_\epsilon \in M_\epsilon\}$ is invariant in the sense that

$$W^u(v_\epsilon) \cdot_D t \subset W^u(v_\epsilon \cdot t), \tag{3.7}$$

if $v_\epsilon \cdot s \in D$ for all $s \in [t,0]$.

Naturally the Theorem will again be proved in the case that M_0 is given by a function over K, and we shall produce a function to describe the fiber $W^s(v_\epsilon)$. It will also be assumed that the above coordinate changes have been made so that M_ϵ is given by $a = b = 0$, and $W^s(M_\epsilon)$ is given by $a = 0$. Some further technical difficulties are caused by the fact that the flow has been modified near the boundary of $\hat{D}$. To alleviate this difficulty we shall assume that there is a compact set $\tilde{K}$ so that $K \subset \operatorname{int} \tilde{K} \subset \tilde{K} \subset \operatorname{int} \hat{K}$ and $\rho = 0$ on $\tilde{K}$. The original equations are then seen to hold on the (larger) set $\tilde{K}$.

Just as for Theorem 1 and Theorem 3, we shall actually prove, and later use, this theorem in the case that the invariant manifolds can be given by graphs. We are assuming the form of the equations given by (3.5).

Theorem 7 *If $\epsilon > 0$ but sufficiently small, then*

(a) *in $a = 0$ (which is $W^s(M_\epsilon)$) there is, for each $v = v_\epsilon = (\hat{y}, \epsilon) \in M_\epsilon$, a function $y = h_s^v(b)$ defined for $|b| \leq \Delta$, so that the graphs*

$$W^s(v) = \{(0, b, y, \epsilon) : y = h_s^v(b)\}$$

form a locally invariant family as in (3.6). Moreover, $h_s^v(b)$ is C^r in v and ϵ jointly for any $r < +\infty$.

(b) *in $b = 0$ (which is $W^u(M_\epsilon)$) there is, for each $v = v_\epsilon = (\hat{y}, \epsilon) \in M_\epsilon$, a function $y = h_u^v(b)$ defined for $|a| \leq \Delta$, so that the graphs*

$$W^u(v) = \{(a, 0, y, \epsilon) : y = h_u^v(a)\}$$

form a locally invariant family as in (3.7). Moreover, $h_u^v(b)$ is C^r in v and ϵ jointly for any $r < +\infty$.

Proof of Theorem 7 We work entirely inside $W^s(M_\epsilon)$, which has become $a = 0$. The arguments for $W^u(M_\epsilon)$ are analogous. In $a = 0$, the variables (b, y, η) will suffice. The splitting of decay rates between b and $\zeta = (y, \eta)$ will be crucial. Define the cone

$$C = \left\{(b, \zeta) : |b| \geq \left|\zeta - \hat{\zeta}\right|\right\},$$

where $\hat{\zeta} = (\hat{y}, \hat{\eta})$ is fixed. Set $v = (0, \hat{\zeta}) = (0, \hat{y}, \hat{\eta})$ and $\hat{\eta} = \epsilon\sigma$. In the, by now familiar, strategy we shall characterize $W^s(v)$ as

$$\Gamma_v = \{u = (b, y, \eta) : u \cdot t \in v \cdot t + C, \text{ for all } t \geq 0\}. \tag{3.8}$$

We use the modified equations exactly as in the proof of Theorem 4 so that $W^s(M_\epsilon)$ is positively invariant and (3.8) is well-defined. We need to show that the set Γ_v is the graph of a Lipschitz function. As in the proof of Theorem 3, we take cross-sections, but in this case of the cone $v + C$. Fix $\hat{b}$ and set

$$S_{\hat{b}} = \left\{(b, y, \eta) \in v + C : b = \hat{b}\right\}.$$

The first task is to show that there is a point $u \in S_{\hat{b}}$ for which $u \cdot t \in v \cdot t + C$ for all $t \geq 0$. A lemma analogous to Lemma 2 can be proved here and will be stated without proof. If $(b_i(t), \zeta_i(t))$ are solutions of (2.8) with $i = 1, 2$, set $M(t) = |b_2(t) - b_1(t)|^2 - |\zeta_2(t) - \zeta_1(t)|^2$.

Lemma 3 *If $M(t) = 0$ then $M'(t) > 0$, unless $b_2 = b_1$.*

One can now apply the flow to the set $\Sigma = S_{\hat{b}}$. By a similar topological argument to that used in the proof of Theorem 4, it can be seen that, for each $t \geq 0$ there is at least one point in the set $\{v \cdot t + C\} \cap \Sigma \cdot t$. Call this point u_t and consider the set $\{u_t \cdot (-t) : 0 < t < +\infty\} \subset \Sigma$. Since Σ is compact, we can find a sequence $t_n \to +\infty$ so that $u_{t_n} \cdot (-t_n)$ converges to, say, $\hat{u}$. One can then see that $\hat{u}$ has the desired property, namely that $\hat{u} \cdot t \in v \cdot t + C$ for all $t \geq 0$. This argument constructs a point $\zeta = (y, \eta)$ for each $b = \hat{b}$ so that $|b| \leq \Delta$. The y component of ζ is $h_s^v(b)$. The uniqueness argument folows in the same way as in the proof of Theorem 4. Converting back to the variable ϵ gives the functions of Theorem 7

The invariance of the family follows from the cone characterization of the fiber, and the proof of the existence of Lipschitz fibers is complete.

The fibers give a very useful matching between the points in $W^s(M_\epsilon)$ and partners they have in M_ϵ. One can then see that the decay of points in $W^s(M_\epsilon)$ to M_ϵ is actually to the base point of the fiber, this gives a decay result with "asymptotic phase"; similarly for points in $W^u(M_\epsilon)$. The proof of Theorem 5 actually follows from Corollary 1

Corollary 1 *$\kappa_s > 0$ and $\alpha_s < 0$ so that if $u \in W^s(v)$ then*

$$|u \cdot t - v \cdot t| \leq \kappa_s \exp\{\alpha_s t\},$$

for all $t \geq 0$ for which $v \cdot [0,t] \subset D$ and $u \cdot [0,t] \subset D$. Furthermore, there are $\kappa_u > 0$ and $\alpha_u < 0$ so that if $u \in W^u(v)$ then

$$|u \cdot t - v \cdot t| \leq \kappa_u \exp\{\alpha_u t\},$$

for all $t \geq 0$ for which $v \cdot [t,0] \subset D$ and $u \cdot [t,0] \subset D$.

Proof A differential inequality on $|b_2(t) - b_1(t)|$, where $v \cdot t = (b_1(t), \zeta_1(t))$ and $u \cdot t = (b_2(t), \zeta_2(t))$, can be derived using the fact that $u \cdot t \in v \cdot t + K$ for all $t \geq 0$. This then leads to the decay estimate.

The cone characterization of the fibers is not the usual approach taken. Fenichel constructs a graph transform map and most other authors follow this lead. However, the cone approach has a very appealing intuition and gives a real characterization of the fibers in terms of the flow. I believe that Bates (private communication) was the first to observe the relevance of cones in this context.
Remark We shall also use the fibers to construct stable and unstable manifolds of subsets of M_ϵ, so that if $A \subset M_\epsilon$,

$$W^u(A) = \cup_{v \in A} W^u(v), \tag{3.9}$$

$$W^s(A) = \cup_{v \in A} W^s(v). \tag{3.10}$$

3.4 Settling in a cellular flow field

The problem of the settling under the influence of gravity of particles, with small inertia, through a fluid flow field can be formulated as a singular perturbation problem. The underlying flow will be assumed to be two-dimensional and of a particular form, namely a cellular fluid flow. Stommel [52] studied this situation in the case of zero inertia and concluded that both suspension in the cells as well as settling could take place. In the following model, $y(t)$ is the position of the center of the particle in space ($y = (y_1, y_2)$).

$$\begin{aligned} \dot{y_1} &= v_1 \\ \dot{y_2} &= v_2 \\ \epsilon \dot{v_1} &= -v_1 + \sin y_1 \cos y_2 \\ \epsilon \dot{v_2} &= -v_2 - W - \cos y_1 \sin y_2, \end{aligned} \tag{3.11}$$

where $v(t)$ is therefore the velocity of the particle and W is the settling velocity scaled so that $0 < W < 1$. The small parameter $\epsilon > 0$ is the Stokes' number and measures the inertial response time of the medium to the particle.

The question of interest here is whether the suspension of particles can still occur if inertial effects are included. In the limit that Stommel studied inertial effects were neglected and the equations were those governing the motion of a fluid particle. In other words, the particle in the fluid was considered to be behaving as a fluid particle would under the combined influence of the fluid flow and the gravitational field. The introduction of inertial effects, when small, supplies a singular perturbation of the case considered by Stommel.

The critical manifold is any appropriate subset M_0 as follows

$$M_0 \subset \{v_1 = \sin y_1 \cos y_2,\ v_2 = -W - \cos y_1 \sin y_2\}. \tag{3.12}$$

The normal eigenvalues are both -1 and hence M_0 is not only normally hyperbolic, but, in fact, attracting. The $W^s(M_0)$ thus fills an entire neighborhood of M_0. Note that the equations (3.11) are periodic in both y_1 and y_2 and thus M_0 can be restricted to any domain that contains a fundamental domain and conclusions that are global in y_1, y_2 can be drawn. Setting $\epsilon = 0$ in (3.11), it is then not hard to see that any initial condition has its ω-limit set in the set of critical points. Thus the stable manifolds of these critical points fills the entire space. The slow manifold M_ϵ is then given by the equations

$$\begin{aligned} v_1 &= \sin y_1 \cos y_2 + O(\epsilon) \\ v_2 &= -W - \cos y_1 \sin y_2 + O(\epsilon), \end{aligned} \tag{3.13}$$

over an appropriate domain. It is also attracting and, if the flow is considered on the space with both y_1 and y_2 identified modulo 2π, is globally attracting. To see this last point, one follows a trajectory with given initial condition by using the approximation of the $\epsilon = 0$ flow to get close to M_ϵ i.e., until the trajectory lies in $W^s(M_\epsilon)$.

The equations on M_ϵ are given by

$$\begin{aligned} \dot{y}_1 &= \sin y_1 \cos y_2 + O(\epsilon) \\ \dot{y}_2 &= -W - \cos y_1 \sin y_2 + O(\epsilon), \end{aligned} \tag{3.14}$$

which reduce to the following on M_0

$$\begin{aligned} \dot{y}_1 &= \sin y_1 \cos y_2 \\ \dot{y}_2 &= -W - \cos y_1 \sin y_2. \end{aligned} \tag{3.15}$$

This latter system is exactly the Stommel model, confirming the expectation that the zero inertia case should appear as the singular limit of the small inertia case. An interesting point to note here is that the variables of physical space, namely y_1, y_2 parametrize the slow manifold and thus the analysis of the trajectories on this manifold has a pleasing interpretation in terms of the flow of particles in physical space.

The phase portrait of (3.15) was given by Stommel and is shown in Figure 9. The space is divided into two distinct types of behavior. Inside each cell, there is a region in which the particles are trapped, this is bounded by the heteroclinic orbits, see Figure 9. The other particle trajectories will settle through the cells and never be trapped. The size of the trapped region increases as W decreases to 0, giving in the limit the usual cellular flow, as expected as gravity is then no longer present, in which all trajectories are trapped. The issue is, to what extent, these proportions of trapped versus settling particles change when ϵ is introduced.

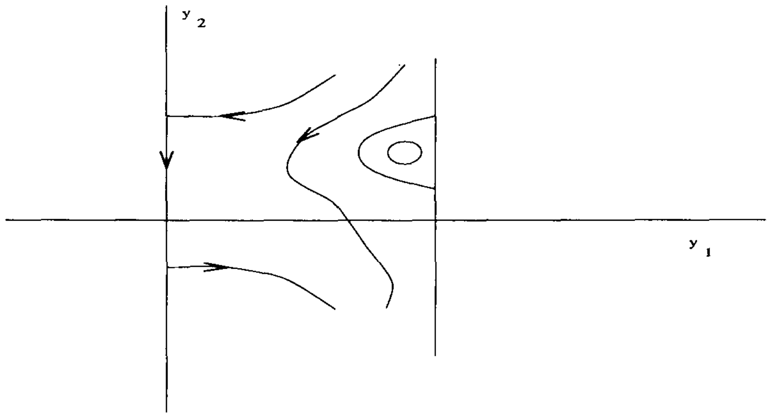

Figure 9
The Stommel flow.

The fate of the heteroclinic orbits surrounding the trapped region when ϵ is turned on must be determined. The following analysis is taken from the paper by Jones, Maxey and Rubin [31]. Indeed, there are two possible scenarios: if it opens so that the unstable manifold is "inside" the stable manifold then more particles can be trapped in a cell by staying inside the heteroclinic orbit. Moreover, particles that were settling can become trapped in the cell by being "caught" by a heteroclinic in some lower cell. However, if the heteroclinic breaks as shown in Figure 10 then particles that are settling will not be trapped in a cell but will continue to pass through. There is a possibility that some particles are still trapped. Indeed the critical point interior to the cell, which necessarily survives the perturbation (why?), will obviously be trapped, but if there is a surviving periodic orbit that surrounds this critical point then it will bound a region of trapped particles. This appears not to occur, as shown by numerical investigations.

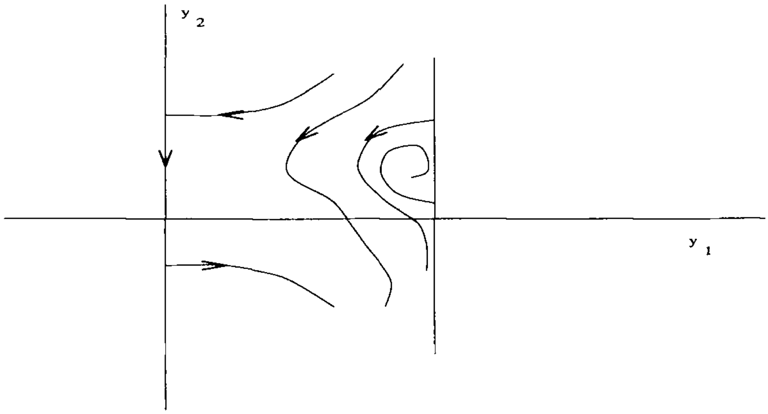

Figure 10
The case in which settling predominates.

As for the Ogawa waves, see Chapter 1, one must calculate the $O(\epsilon)$ terms in order to see whether the heteroclinic orbit persists. the equations up to $O(\epsilon)$ are given by

$$\begin{aligned} \dot{y}_1 &= \sin y_1 \cos y_2 + \epsilon \sin y_1 (W \sin y_2 + \cos y_1) + O(\epsilon^2) \\ \dot{y}_2 &= -W - \cos y_1 \sin y_2 + \epsilon \cos y_2 (W \cos y_1 + \sin y_2) + O(\epsilon^2). \end{aligned} \tag{3.16}$$

We fix attention on the cell $0 \leq y_1 \leq \pi$, $0 \leq y_2 \leq \pi$. Let $y^-(\epsilon)$ be the critical point so that $y^-(0) = (0, \hat{y}_2)$ and $\hat{y}_2 \in [\frac{\pi}{2}, \pi]$ and $y^+(\epsilon)$ be the analogous point with $\hat{y}_2 \in [0, \frac{\pi}{2}]$. Further, let W^- be the unstable manifold of $y^-(\epsilon)$ and W^+ the stable manifold of $y^+(\epsilon)$. Consider the intersection of these manifolds with the set $y_1 = \frac{\pi}{2}$. Each will be given as the graphs of functions, so that $W^- \cap \{y_1 = \frac{\pi}{2}\}$ is given by $y_1 = h^-(\epsilon)$ and $W^+ \cap \{y_1 = \frac{\pi}{2}\}$ is given by $y_1 = h^+(\epsilon)$. To see which way the heteroclinic opens it suffices to calculate the quantity

$$K = \left(\frac{\partial h^-}{\partial \epsilon} - \frac{\partial h^+}{\partial \epsilon} \right) |_{\epsilon=0}.$$

Indeed, If $K > 0$ then the heteroclinic opens in such a way as to facilitate settling and inhibit trapping. It is shown in [31] that

$$\begin{aligned} \frac{\partial h^-}{\partial \epsilon}|_{\epsilon=0} = W \int_{\hat{y}_1}^{\pi} \Big[& -\sin 2y_1 + y_1(\cos^2 y_1 + 1) \\ & + W^2 y_1 \frac{\sin^2 y_1 - y_1^2}{\sin^2 y_1 \sqrt{(\sin^2 y_1 - W^2 y_1^2)}} \Big] dy_1. \end{aligned} \tag{3.17}$$

It is shown in [31] that $\frac{\partial h^-}{\partial \epsilon} > 0$, and, by symmetry, $\frac{\partial h^+}{\partial \epsilon} < 0$. It follows that $M > 0$ and the flow on M_ϵ is as shown in Figure 10.

The above analysis shows that on the slow manifold M_ϵ the settling of particles is facilitated. Indeed, it is more likely that particles with small inertia will settle than those with no inertia i.e., fluid particles. Thus a striking change to Stommel's conclusions occurs when inertia is taken into account.

Somewhat more about the flow on the slow manifold is shown in [31]. The slow manifold M_ϵ can be viewed as a torus by identifying both y_1 and y_2 modulo 2π. It is shown injonrubmax that on this torus there are only finitely many periodic orbits that wind around the torus. These are settling trajectories and are expected to be the asymptotic motion of the particles. If W is sufficiently small, it appears, from numerical computations, that there is a unique such periodic trajectory. In fact, there is one periodic orbit that has a special status as the fixed point of a certain map and all the others are obtained as period doubling bifurcations from this base orbit, see [31].

Since the slow manifold is attracting as discussed above, this predominance of settling will hold for the entire system. It is important to consider here exactly what more one obtains from knowing that the stable manifold, now filling a neighborhood of M_ϵ, is fibered by the individual stable manifolds. To each point $z = (y_1, y_2, v_1, v_2) \in \mathbf{R}^4$ the Fenichel fiber map assigns a point $\pi^-(z) \in M_\epsilon$ so that

$$\left| z \cdot t - \pi^-(z) \cdot t \right|$$

decays exponentially. This means that the point z will inherit all the asymptotic characteristics of the point $\pi^-(z)$. For instance, asymptotic settling rates can be concluded from the period of the attracting periodic orbits on M_ϵ. These settling rates will then also be valid for all initial conditions, including those off M_ϵ. If any non-trivial region of initial conditions is trapped in the cells, other than the critical point, on M_ϵ, then the corresponding region in $\mathbf{R}^4$ is determined to be the stable fibers to the points in this set. If, as expected, almost everything, except for the trapped critical points, settles in M_ϵ, then so does everything off M_ϵ except for the stable fiber of that critical point. When there is a unique periodic orbit on the torus (M_ϵ), almost all trajectories on M_ϵ will tend to this orbit, and thus so will almost all off M_ϵ. The only remaining possibility is that a periodic orbit surrounding the critical point persists then it, and its interior, will consist of trapped points. Thus the union of the stable fibers to this set would render the full set of initial conditions that lead to trapping.

We have used, in this example, the Fenichel Theorems to great effect in showing that even the smallest inertial effects of particles settling under a gravitational cellular flow field will encourage their settling and inhibit any potential trapping. Moreover, the Fenichel fibers have allowed us to see that the motion on M_ϵ determines all the asymptotic features of the system.

3.5 Normal form

Using the fibers constructed in Theorem 7, we can complete the transformation to Fenichel Normal Form. Indeed, Theorem 7 gives a map from (b, v) to a point $(b, h_s^v(b))$. The inverse of this map will send each point in $W^s(M_\epsilon)$ to the base point of its fiber. The fact that the inverse exists can be seen by observing that the construction of the fiber did not use the fact that $v_\epsilon \in M_\epsilon$, and could easily have been based also at the point $u = (b, \zeta)$, where $\zeta = h_s^v(b)$. It then follows that v would lie on the fiber for the point u and would give the inverse of the map. We denote this inverse map $(b, y, \epsilon) \mapsto \hat{y}$ by π^-.

The final step in deriving the Fenichel Normal Form is to straighten out the fibers inside each of $W^s(M_\epsilon)$ and $W^u(M_\epsilon)$. To this end, set

$$\begin{array}{ll} a_3 = a, & b_3 = b, \\ y_3 = \pi^-(b, y, \epsilon) & \epsilon_3 = \epsilon, \end{array} \tag{3.18}$$

so that the y-coordinate of each point is changed into that of its fiber base-point. The fibers on the unstable manifold are straightened out analogously by

$$\begin{array}{ll} a_4 = a_3, & b_4 = b_3, \\ y_4 = \pi^+(a_3, y_3, \epsilon_3) & \epsilon_4 = \epsilon_3, \end{array} \tag{3.19}$$

which takes each point in $W^u(M_\epsilon) = \{b = 0\}$ to the base point of its unstable fiber.

The transformations to arrive at the coordinates $(a_4, b_4, y_4, \epsilon_4)$ have now modified the equations on both $W^s(M_\epsilon)$ $(a_4 = 0)$ and $W^u(M_\epsilon)$ $(b_4 = 0)$. We shall drop the subscripts and, with an abuse of notation, revert to the use of the original letters, with the understanding that the new coordinates are being used. On these sets, the slow flow has become independent of both a and b. It follows that if, either $a = 0$ or $b = 0$ we have $g(a, b, y, \epsilon)$ is a function only of y and ϵ, Thus we can write

$$g(a, b, y, \epsilon) = h(y, \epsilon) + H(a, b, y, \epsilon)(a, b), \tag{3.20}$$

where $H(a, b, y, \epsilon)$ is a bilinear function of a and b.

Putting the pieces together, we can give the final Fenichel Normal Form for singularly perturbed equations in the neighborhood of a slow manifold

$$\begin{aligned} a' &= \Lambda(a, b, y, \epsilon)a \\ b' &= \Gamma(a, b, y, \epsilon)b \\ y' &= \epsilon \left\{h(y, \epsilon) + H(a, b, y, \epsilon)(a, b)\right\}, \end{aligned} \tag{3.21}$$

which holds in the set $D = \{(a, b, y, \epsilon) : |a| \leq \Delta, |b| \leq \Delta, y \in K, \epsilon \in [0, \epsilon_0]\}$. A simplified version of this normal form was derived by Jones and Kopell, see [28]. The current version was derived by Jones, Kaper and Kopell, see [27], and simultaneously discovered by Sandstede (private communication). Tin [54] has used it extensively and derived it in the more general context of perturbed invariant manifolds without a singular structure in which Fenichel first derived his Theorems.

Chapter 4

Tracking with Differential Forms

The Fenichel Normal Form will be a key in proving the Exchange Lemma. The Exchange Lemma concerns the passage of certain invariant sets (which will, in fact, be manifolds in their own right) near a slow manifold. In order to track the appropriate information about these invariant sets during the passage, we shall use differential forms to quantify this information.

4.1 Motivation

With the equations in Fenichel Normal Form (3.21) in a neighborhood D of M_ϵ, when ϵ is sufficiently small, a picture can be drawn in which the stable and unstable manifolds are coordinate planes and the spine, along which they intersect, is the manifold M_ϵ. Indeed, as shown in Figure 11, the set $a = 0$ is $W^s(M_\epsilon)$, $b = 0$ is $W^u(M_\epsilon)$ and $a = b = 0$ is M_ϵ. Of interest will be the situation in which a locally invariant manifold will be followed from some remote part of the phase space and studied as it passes near M_ϵ. We shall call this manifold the "shooting manifold" and usually denote it by Σ_ϵ. This manifold should not be confused with any of the manifolds that have been constructed above under the guise of the Fenichel Invariant Manifold Theorems. The purpose of following such a manifold will be to construct a homoclinic, or heteroclinic, orbit. The shooting manifold will then be the unstable manifold of a certain invariant set, such as a curve of critical points or periodic orbit. This invariant set will, in general, be unrelated to the slow manifold it is passing.

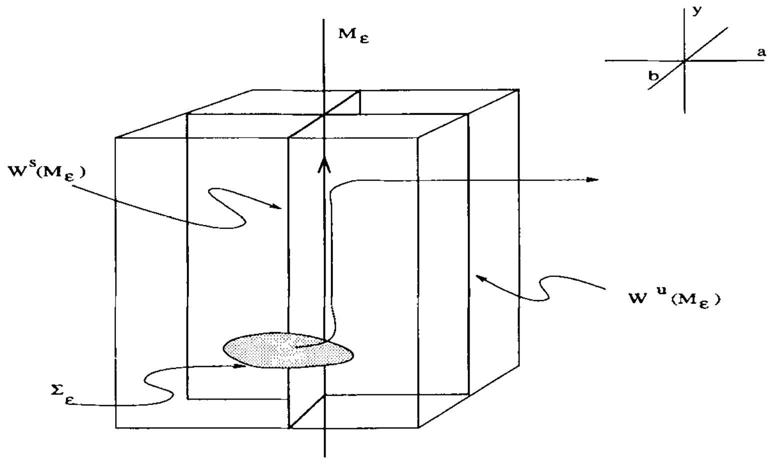

Figure 11
The passage near a slow manifold.

The shooting manifold will enter the neighborhood D and pass through D being modified by virtue of its flight near M_ϵ. It is viewed as entering with certain information and exiting with certain other information. This "exchange" of information will be the subject of the Exchange Lemma. It is instructive to consider what information is significant. A typical method for constructing homoclinic, or heteroclinic, orbits is by locating the (transverse) intersections of relevant stable and unstable manifolds. To determine a transverse intersection requires precise knowledge of the tangent spaces to the manifolds. It is the information encoded in the tangent spaces to the shooting manifold that will thus be of interest. The shooting manifold arrives at the boundary of D carrying certain tangent vectors and, during its passage through D, they will be exchanged for other tangent vectors. The goal of the ensuing analysis will thus be to see how to figure out the tangent vectors upon exit from D in terms of those at the entrance to D.

The general problem of following tangent spaces through phase space can be attacked in numerous ways. The approach adopted here will be to use differential forms that give "coordinates" to subspaces of a given space. These are particularly well suited to studies involving singular perturbations. Indeed, individual tangent vectors are hard to follow as vectors can switch from being predominantly in fast directions to the slow directions. The use of differential forms gives a satisfactory resolution to this problem as they afford a way of tracking the entire tangent space, without reference to individual vectors.

The consideration of an example will put this challenge of tracking invariant manifolds during their passage near a slow manifold in perspective. The canonical example that is the easiest to visualize is that of finding the travelling pulse solution in the FitzHugh-Nagumo equations and this will be described in the next section.

4.2 FitzHugh-Nagumo equations

The paradigmatic example for the construction of homoclinic orbits in singularly perturbed systems is the travelling pulse problem of the FitzHugh-Nagumo equations. These equations arose originally as a simplification to the Hodgkin-Huxley equations, formulated independently by FitzHugh and Nagumo. This is the first example in which both fast and slow structure will appear in the orbits of interest. The mathematical proof of the existence of these travelling pulses was originally given independently by Carpenter [10] and Hastings [24]. A geometric proof was later given by Langer [38], which paper spawned the Exchange Lemma through attempts to simplify difficult parts of Langer's proof and generalize the construction. It should be noted here that the geometric approach is needed in order to assess the stability of the travelling wave. The topological construction of the wave is inadequate for proving stability, see [26]. The full partial differential equations are

$$\begin{aligned} u_t &= u_{xx} + f(u) - w \\ w_t &= \epsilon(u - \gamma w), \end{aligned} \tag{4.1}$$

where $f(u) = u(u-a)(1-u)$ is the usual bistable nonlinearity and ϵ is a small parameter. The travelling pulse is a solution of (4.1) which is a function only of $\xi = x - ct$, and thus satisfies the system

$$\begin{aligned} u' &= v \\ v' &= -cv - f(u) + w \\ w' &= \tfrac{\epsilon}{\theta}(u - \gamma w) \\ c' &= 0, \end{aligned} \tag{4.2}$$

so that (u, v) are fast variables and (w, c) are slow variables. The problem here is to construct an orbit homoclinic to the rest state $u = v = w = 0$. The first step is to construct a singular orbit consisting of fast transitions (heteroclinic orbits) between critical manifolds with intervening trajectories of the slow system. The critical manifolds must lie in the set $\{v = 0,\ w = f(u)\}$, which is the graph of a cubic inside the plane $v = 0$. Pieces of this critical set will form critical manifolds, namely in regions where $f'(u) \neq 0$. Of particular interest will be critical manifolds in the left branch of the cubic, say M_0^L, which is defined as the graph of the cubic restricted to an interval $[u_1, u_2]$ where $u_1 < 0$ and $f'(u) < 0$ on this interval, and also in the right branch, say M_0^R, defined similarly on $[u_3, u_4]$, see Figure 12. Both M_0^L and M_0^R are easily seen to be normally hyperbolic with 1-dimensional stable manifolds and 1-dimensional unstable manifolds.

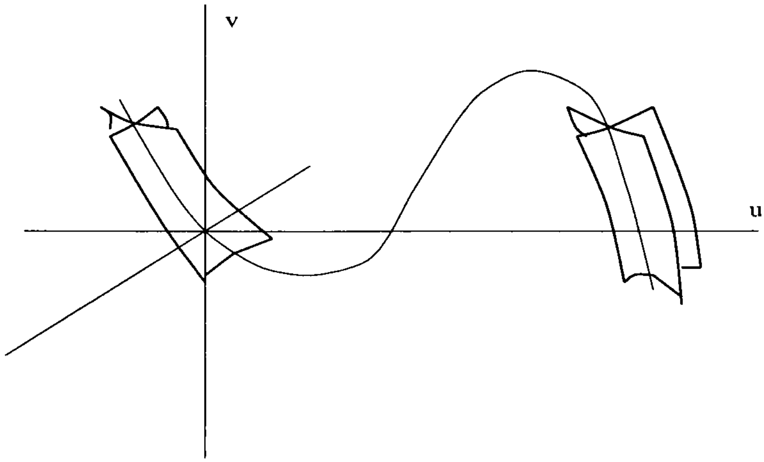

Figure 12
The slow manifolds and their stable
and unstable manifolds
for the FitzHugh-Nagumo system.

The singular homoclinic orbit, which exists at $\epsilon = 0$, will form a template on which the full orbit is built when $\epsilon > 0$. It consists of two fast pieces, $\mathcal{F}_1$, $\mathcal{F}_2$, and two slow pieces $\mathcal{S}_1$ and $\mathcal{S}_2$. These are determined as follows, see Figure 13.

$\mathcal{F}_1$ heteroclinic orbit when $\epsilon = 0$, with $w = 0$ and $c = c^*$, exactly as constructed for the scalar bistable reaction-diffusion equation,

$\mathcal{F}_2$ heteroclinic orbit when $\epsilon = 0$ from M_0^R to M_0^L, exists at given $c = c^*$ and fixed $w = w^*$ which is determined by the construction of this heteroclinic orbit,

$\mathcal{S}_1$ solution of limiting slow system $\dot{w} = -\frac{1}{c}(u - \gamma w)$ on M_0^R, connecting end of $\mathcal{F}_1$ to beginning of $\mathcal{F}_2$,

$\mathcal{S}_2$ solution of limiting slow system $\dot{w} = -\frac{1}{c}(u - \gamma w)$ on M_0^R, connecting end of $\mathcal{F}_1$ to rest state.

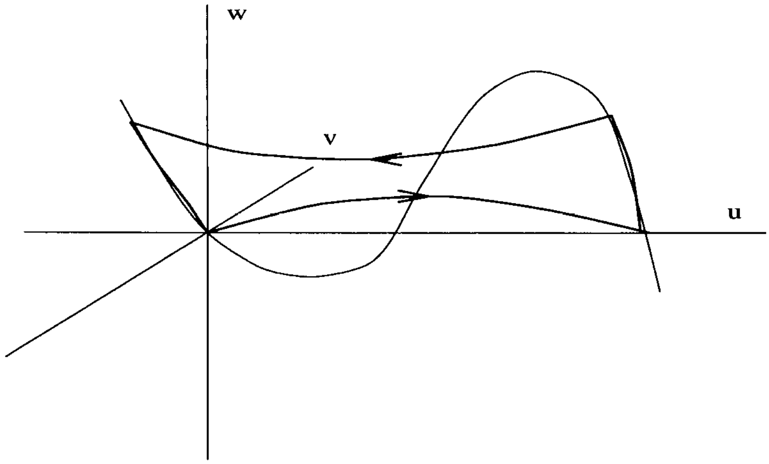

Figure 13
The singular homoclinic orbit.

The idea for constructing the true travelling pulse (homoclinic orbit) at a value of c near to c^* is to carry W_-^u, the unstable manifold of the curve of critical points $u = v = w = 0$ and $|c - c^*| \leq \delta$ for some fixed $\delta > 0$, around the phase space when $0 < \epsilon \ll 1$, see Figure 14.

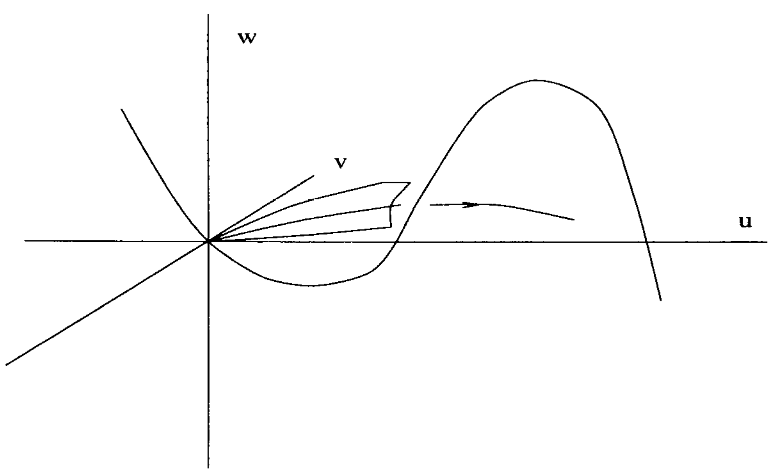

Figure 14
The shooting manifold.

It will be controlled by the information available from the $\epsilon = 0$ case. Indeed, the singular solution gives a template along which the manifold W_-^u is carried.

The ultimate goal will be to force it to intersect the stable manifold of the same curve of critical points.

The transversality shown above for the bistable equation shows that, when $\epsilon = 0$, W^u_- will transversely intersect $W^s(M_0^R)$. It will then also follow, by, for instance, the Implicit Function Theorem, that, when $\epsilon > 0$ but small, W^u_- transversely intersects $W^s(M_\epsilon^R)$. We need to see how W^u_- is affected during its time near the slow manifold. This is necessarily an $\epsilon \neq 0$ consideration, as when $\epsilon = 0$, W^u_- will stay in $w = 0$. The Exchange Lemma, see [29], [28], will precisely answer this question. It is expected that W^u_- will emerge from a neighborhood of M_0^R near to the point where the singular orbit exits the neighborhood, at least if ϵ is small enough, which is given as the point at which $\mathcal{F}_2$ emerges. What is needed is the tangent space to W^u_- at this exit point as then the next stage of carrying the shooting manifold over the next fast jump can be realized and a transversality set up to see how the manifold intersects the stable manifold of the slow manifold M_0^L.

To see that the problem of determining the tangent space to W^u_- upon exit from D is non-trivial, one should consider the dimensions of the manifolds involved. The shooting manifold W^u_- is 2-dimensional, while the manifold M_0^R is also 2- dimensional. But then the unstable manifold to the slow manifold, namely $W^u(M_0^R)$ is 3-dimensional. It is natural to expect that W^u_- will be crushed against the unstable manifold of M_0^R. This intuition comes from the λ-Lemma, see [49], which applies when M_0^R is a single critical point. The λ-Lemma has been generalized to the case of critical manifolds, for the purpose of applying it to singular perturbation problems, by Deng, see [12], and this has the consequence that, indeed, W^u_- does become close to $W^u(M_0^R)$. However, this information is insufficient as $W^u(M_0^R)$ is 3-dimensional and W^u_- is only 2-dimensional. Clearly the fast unstable direction must be present in the tangent space to W^u_-. The question is: which slow direction is picked out by the tangent space to W^u_- at the exit point? The Exchange Lemma will give a precise answer to this question and we will then be able to finish the construction of the FitzHugh-Nagumo homoclinic orbit.

4.3 Variational equation and differential forms

The statement of the Exchange Lemma and the sketch of its proof will be postponed until after the technique of using differential forms has been introduced. Consider a general ODE

$$z' = F(z), \tag{4.3}$$

where F is a smooth (C^∞) function on an open subset U of $\mathbf{R}^N$. The variational equation of (4.3) can be written as

$$p' = DF(z)p, \tag{4.4}$$

where z satisfies (4.3) and $p \in \mathbf{R}^N$. The coordinates of p can be conveniently expressed using differential forms, $p_i = dz_i(p)$, recalling that dz_i is a linear form on tangent vectors. Since (4.4) can be rewritten, using the summation convention, as

$$p_i' = \partial_j F_i(z) p_j, \tag{4.5}$$

we can also write

$$dz_i(p)' = \partial_j F_i(z) dz_j(p). \tag{4.6}$$

Moreover, we will usually suppress the tangent vector p itself and write

$$dz_i' = \partial_j F_i(z) dz_j, \tag{4.7}$$

which ostensibly is nonsensical as the forms dz_i are constant, but whenever a derivative of dz_i appears, it should be understood as applied to whatever particular tangent vector is currently under consideration. We can also write (4.7) in shorthand as

$$dz' = DF(z)dz. \tag{4.8}$$

4.4 Tracking tangent spaces

We have seen that the variational equation can conveniently be expressed using differential 1-forms. We shall now put this to good use by using it as a vehicle to calculate equations on higher order forms. It is natural to compute such equations in order to ascertain how information about tangent spaces to invariant manifolds is carried under the flow. If invariant sets, or more specifically manifolds, are to be tracked under the influence of the flow, their configuration at a certain time is locally encoded in the tangent space to the invariant manifold at the relevant underlying point of the flow. The "coordinates " of a tangent space are given by projecting a (unit) cube in that space onto each of the coordinate subspaces of the same dimension as the tangent spaces themselves. The "volume" of the resulting object is the value of that coordinate. These quantities can be algebraically calculated using differential forms as they are indeed the values of the various k-forms for manifolds of dimension k. In other words, if Π is a k-dimensional subspace of $\mathbf{R}^N$, using coordinates $z = (z_1, \ldots, z_N)$, the coordinates of Π are given by

$$dz_{i_1} \wedge dz_{i_2} \wedge \ldots \wedge dz_{i_k} (\Pi),$$

for all choices of $(i_1, \ldots, i_k)$ (without repetition or permutation). If Π is spanned by $\{v_1, \ldots, v_k\}$ then

$$dz_{i_1} \wedge \ldots \wedge dz_{i_k} (\Pi) = \sum_\pi (-1)^{\mathrm{sgn}\,\pi} dz_{i_1}(v_{\pi(1)}) dz_{i_2}(v_{\pi(2)}) \ldots dz_{i_k}(v_{\pi(k)}), \tag{4.9}$$

where π is a permutation of $(1, \ldots, k)$. This is exactly the volume of the cube in Π, that is determined by the spanning vectors $v_1, \ldots, v_k$, projected onto the $(z_{i_1}, \ldots z_{i_k})$ subspace.

If $p \cdot t$ is a solution of (4.3) and $p \cdot t$ belongs to some invariant manifold, to which $\Pi(t)$ is the tangent space at $p \cdot t$, we should be able to calculate an equation for the coordinates of $\Pi(t)$. This can indeed be done and is related to the variational equation as a k-form version of it. The calculation of this equation in full generality is tedious and thus we shall carry out such a computation only on examples. Hopefully, these examples will make the general structure clear, in which there are, associated with any given equation, flows on the entire exterior algebra; in other words, flows on the space of k-forms, for any $1 \leq k \leq N$.

4.5 Example

We shall refresh this abstract discussion by considering a specific example and calculating the flow on tangent spaces using differential forms for it. We shall apply it to the derivation of an important transversality condition. The example will be that of the travelling wave for the bistable reaction-diffusion equation. This has appeared twice already in these lectures. In the FitzHugh-Nagumo travelling pulse problem, these equations arose as the $\epsilon = 0$, $w = 0$ subsystem that gave the fast jump $\mathcal{F}_1$. This fast jump is a heteroclinic orbit and a key point about the FitzHugh- Nagumo pulse is that it is constructed as the transverse intersection of stable and unstable manifolds. It has also appeared as the limiting slow equation for the phase field model considered in the first lecture, namely (1.12). Indeed, a transversality condition was given in that lecture for the heteroclinic orbit to persist to such an orbit on the slow manifold M_ϵ. This transversality condition will be derived using differential forms. This example has thus arisen as both the slow equation, for the phase field problem, and the fast equation, for the FitzHugh-Nagumo pulse!

Consider then the travelling wave problem for the bistable reaction-diffusion equation with an equation for the speed parameter appended

$$\begin{aligned} u' &= v \\ v' &= -cv - f(u) \\ c' &= 0, \end{aligned} \tag{4.10}$$

where $f(u) = u(u-a)(1-u)$ and $a < \frac{1}{2}$. We shall show that W^-, the unstable manifold of the curve $\{(0,0,c) : c \text{ near } c*\}$, and W^+, the stable manifold of $\{(1,0,c) : c \text{ near } c*\}$, intersect transversely at $q^* \in \{u = a\}$ at the value of c, say $c*$, at which the heteroclinic orbit exists.

The variational equation for (4.10), in the differential form notation of the above section, can be calculated as

$$\begin{aligned} du' &= dv \\ dv' &= -cdv - Df(u)du - vdc \\ dc' &= 0. \end{aligned} \tag{4.11}$$

Since W^- and W^+ are both 2-dimensional manifolds, we will need to track 2-dimensional subspaces and thus should calculate the equations for the various different 2-forms. There are three different 2-forms, namely $du \wedge dv$, $dv \wedge dc$ and

$du \wedge dc$. The equations for the variation of these 2-forms can be calculated as shown in the preceding section. The product rule is used to see that

$$\begin{aligned}(du \wedge dv)' &= du' \wedge dv + du \wedge dv', \\ &= dv \wedge dv + du \wedge (-cdv - Df(u)du - vdc) \\ &= -cdu \wedge dv - vdu \wedge dc,\end{aligned} \tag{4.12}$$

where it has been used that $du \wedge du = dv \wedge dv = 0$. The calculation of the equations for the remaining two forms is left as an exercise.

Recall the reduction of the transversality to calculating the sign of the quantity (rewritten in the current notation)

$$\left(\frac{\partial h^-}{\partial c} - \frac{\partial h^+}{\partial c}\right)|_{c=c*}, \tag{4.13}$$

where the intersections of $W^\pm$ with $\{u = a\}$ are given, respectively by $v = h^\pm(c)$. As a procedure for verifying transversality, the following has general applications, and a very definite structure to the argument, thus we divide it into steps that can be repeated in other contexts, such as for Melnikov calculations, see below.
Step 1 Observe first that the vectors

$$\eta_1^\pm = \left(0, \frac{\partial h^\pm}{\partial c}, 1\right)$$

are tangent respectively to $W^\pm$. We seek a 2-form that renders a multiple of $\frac{\partial h^\pm}{\partial c}$ when applied to the tangent space of $W^\pm$. We know another vector that is tangent to both $W^\pm$, namely the vector field itself

$$\eta_2 = (v, -cv - f(u), 0).$$

We then see that

$$= du \wedge dv(\eta_1^\pm, \eta_2) = v\frac{\partial h^\pm}{\partial c}, \tag{4.14}$$

at $q = q^*$.
Step 2 Equation (4.12) would be very useful for evaluating the left hand side of (4.14) were the last term known. The difficulty is that the value of $\eta_1^\pm \cdot t$, where the flow referred to is that of the variational equation over the underlying heteroclinic orbit, is not known for $t \neq 0$. However we do know that

$$dc(\eta_1^\pm \cdot t) = 1$$

as the c-component is invariant from (4.11). Moreover $\eta_2 \cdot t$ is known for all t as it is exactly the vector field, that is $\eta_2 \cdot t = (v, -cv - f(u), 0)$. This is sufficient to compute $du \wedge dc$, indeed

$$du \wedge dc(\eta_1^\pm \cdot t, \eta_2 \cdot t) = -v,$$

and so, setting $\omega = du \wedge dv(\Pi^\pm(t))$, we obtain

$$\omega' = -c\omega - v^2. \tag{4.15}$$

Step 3 It is an exercise to check that $\exp(ct)\,\omega \to 0$ as $t \to -\infty$. Equation (4.15) can then be explicitly solved to render

$$\omega = e^{-ct} \int_{-\infty}^{t} \left\{ -e^{ct} v^2 \right\} dt, \tag{4.16}$$

which shows that $\omega(0) = -v(0)\frac{\partial h^-}{\partial c} < 0$. Since $v(t) > 0$ for all t, we conclude that

$$\frac{\partial h^-}{\partial c}|_{c=c^*,q=q^*} > 0.$$

A similar argument shows that

$$\frac{\partial h^+}{\partial c}|_{c=c^*,q=q^*} < 0.$$

Putting these two inequalities together, we have the desired result that

$$\left(\frac{\partial h^-}{\partial c} - \frac{\partial h^+}{\partial c} \right) |_{c=c^*,q=q^*} > 0, \tag{4.17}$$

as desired.

4.6 Transversality for the KdV-KS waves

As a second application, the transversality condition involved in the existence of solitary waves in the KdV-KS equations as discussed in the first lecture, will be derived. This will give an application of the above prescription that is very close to the Melnikov method and the reader is invited to check that the usual Melnikov conditions can be derived using this differential form approach, see [20].

Consider the equations (1.34), for which an intersection of $W^u(\mathcal{S})$ and $W^s(\mathcal{S})$ is sought. We need to calculate $M(c)$ as defined in (1.36).
Step 1 A 3-form must be found that renders the quantities $\frac{\partial h^\pm}{\partial \epsilon}$ when applied to the tangent spaces $\Pi^\pm(0)$ to the invariant manifolds $W^u(\mathcal{S})$ and $W^s(\mathcal{S})$. There are three tangent vectors to $W^u(\mathcal{S})$ and $W^s(\mathcal{S})$ at $t = 0$ that are easily found. They are given by

$$\begin{aligned} \eta_1 &= \left(\tfrac{\partial h^\pm}{\partial \epsilon}, 0, 1, 0 \right) \\ \eta_2 &= \left(v, u - \tfrac{u^2}{2}, 0, 0 \right) = (0, \alpha, 0, 0) \\ \eta_3 &= (0, 0, 0, 1), \end{aligned} \tag{4.18}$$

where $\alpha < 0$. It can then be checked that

$$du \wedge dv \wedge dc\,(\eta_1, \eta_2, \eta_3) = \alpha \frac{\partial h^\pm}{\partial \epsilon},$$

Step 2 The equation for the form $du \wedge dv \wedge dc$ can be calculated as

$$(du \wedge dv \wedge dc)^{\cdot} = \sqrt{c}\left(u - (1 + \frac{1}{c})\right) v du \wedge d\epsilon \wedge dc.$$

Step 3 As in the previous section, the form $du \wedge dv \wedge dc$, when applied to the subspace $\Pi^{\pm}(t)$, can actually be calculated. Since

$$\eta_1 \cdot t = (*, *, 1, 0),$$

$$\eta_2 \cdot t = (v, u - \frac{u^2}{2}, 0, 0),$$

and

$$\eta_3 \cdot t = (*, *, 0, 1),$$

it can be seen that $du \wedge d\epsilon \wedge dc(\eta_1 \cdot t, \eta_2 \cdot t, \eta_3 \cdot t) = v$. It follows that

$$(du \wedge dv \wedge dc)^{\cdot} = -\sqrt{c}\left(u - (1 + \frac{1}{c})\right) v^2.$$

From which, one obtains that

$$\alpha M(c) = -\sqrt{c}\left[\int_{-\infty}^{+\infty} (u-1)v^2 - \frac{1}{c}\int_{-\infty}^{+\infty} v^2\right].$$

The expression for $M(c)$ given in (1.37) is now easily derived using the equation.

Chapter 5

Exchange Lemma

With the above motivation, we are now in a position to give the simplest version of the Exchange Lemma. This will be the one originally formulated by Jones and Kopell, see [28]. In the next lecture, more sophisticated versions will be presented, but today's version is already powerful enough to prove a general theorem on the existence of heteroclinic orbits.

5.1 $k+1$ Exchange Lemma

In the following, we are assuming the standard singular perturbation set-up and that the slow manifold under consideration has a k-dimensional unstable manifold, as usual. We shall track another (locally) invariant manifold, say Σ_ϵ during its passage near the slow manifold M_ϵ; this should not be confused with the manifold M_ϵ itself, or its attendant stable, unstable manifolds or fibers. Recall that the manifold Σ_ϵ will be generated, most probably, at some other part of the phase space; the example of the FitzHugh-Nagumo equations should be kept in mind here, wherein the invariant manifold is the unstable manifold of a curve of critical points lying on a *different* slow manifold. We make the following hypothesis, in which the notation $\cap_T$ means that the intersection is transversal. The set D is the standard neighborhood of M_ϵ in when the coordinates of Fenichel Normal Form are used i.e.,

$$D = \{(a, b, y) : |a| \leq \Delta, |b| \leq \Delta, y \in K\}.$$

(H4) There is a $(k+1)$-dimensional, locally invariant manifold Σ_ϵ, defined for $0 < \epsilon \ll 1$, and smooth in ϵ, so that

$$\Sigma_0 \cap_T W^s(M_0) \neq \emptyset,$$

at a point $q \in \partial D$.

As a technical point, we shall always assume that Σ_0 is cut off so as to intersect $W^s(M_0)$ only in a neighborhood of q.

Since Σ_0 is $(k+1)$-dimensional and $W^s(M_0)$ is $(l+m)$-dimensional, in the intersection of these two manifolds we have available $k+l+m+1$ directions. Since the dimension of the phase space is $N = k+l+m$ the intersection will be 1-dimensional under the transversality hypothesis (H3). But this is optimal as both manifolds are locally invariant and hence their intersection must contain trajectories, which means that it must be, at least, 1-dimensional. From (H4) it then follows that $\Sigma_0 \cap W^s(M_0)$ must live in $W^s(J_0)$ for some point $J_0 \in M_0$. Alternatively, we can write

$$J_0 = \omega\left(\Sigma_0 \cap W^s(M_0)\right). \tag{5.1}$$

The geometry is shown in Figure 15. It is also useful to note here that $J_0 = \pi^-\left(\Sigma_0 \cap W^s(M_0)\right)$, where π^- is the map sending each point in $W^s(M_0)$ to the base point of its fiber.

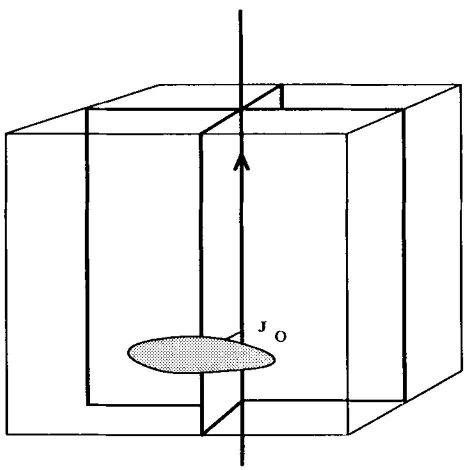

Figure 15
The point J_0.

Of central importance will be the trajectory of the limiting slow flow through the point J_0. Let $\hat{\tau}$ be some fixed (slow) time, choose $0 < \eta_- < \hat{\tau}$ and $\eta_+ > 0$. Set $I = [\hat{\tau} - \eta_-, \hat{\tau} + \eta +]$. Assume that these quantities are chosen so that $J_0 \circ \tau$ is defined for all $\tau \in I$, where "$\circ$" refers to the action of the limiting slow flow i.e., that associated with (1.7) on M_0. In the following, we shall consider the set

$$J_0 \circ I \subset M_0. \tag{5.2}$$

When $\epsilon > 0$ but sufficiently small, the manifold Σ_ϵ will intersect $W^s(M_\epsilon)$ transversely. However the point $p_\epsilon \in \Sigma_\epsilon \cap W^s(M_\epsilon)$ will not ultimately be the point of interest as the trajectory emanating from this point will not leave D near

$W^u(M_\epsilon)$. Indeed, it will either never leave D or else leave it near the boundary of M_0. With reference to the FitzHugh-Nagumo example, if we wish to construct an orbit close to the singular orbit, the true orbit will have to leave D near the fast jump away from the critical manifold, and thus be close to $W^u(M_0)$. The notation q_ϵ will refer to a point in $\Sigma_\epsilon \cap \partial D$. This will be the point along the trajectory of which we shall track the manifold $\Sigma_\epsilon \cdot t$. We shall assume that the trajectory through q_ϵ does indeed leave D in forward time after (fast) time $T_\epsilon > 0$, so that $q_\epsilon \cdot T_\epsilon \in \partial D$. We need to able to control where the trajectory leaves D. Recall that the unstable manifold for a subset of M_0 is just the union of the unstable manifolds of its elements, see (3.10), which is here only used in the $\epsilon = 0$ case but also makes sense when $\epsilon \neq 0$.

Proposition 1 *Given $r_0 \in W^u(J_0 \circ I) \cap \partial D$, there is a $q_\epsilon \in \Sigma_\epsilon \cap \partial D$ and a $T_\epsilon > 0$ so that $q_\epsilon \cdot T_\epsilon \in \partial D$ and*

$$|q_\epsilon \cdot T_\epsilon - r_0| = O(\epsilon). \tag{5.3}$$

The proof of this Proposition uses the Wazewski Principle, in other words it is a topological shooting argument. The proof will be omitted. From this proposition we can see that trajectories can be found that exit D near prescribed points, for instance the point where a fast jump (when $\epsilon = 0$) will leave the neighborhood of the critical manifold. It does not, however, tell us how the manifold Σ_ϵ is configured in a neighborhood of this point. Thus, we can think of the above Proposition as a C^0-Exchange Lemma. The C^1-Exchange Lemma determines the configuration of the tangent space to Σ_ϵ upon exit from D. We shall call it the $(k+1)$-Exchange Lemma as it refers to the case in which Σ_ϵ is $(k+1)$-dimensional. The hypotheses (H1)-(H4) are assumed to hold.

Lemma 4 (**$(k+1)$-Exchange Lemma**) *The manifold $\Sigma_\epsilon \cdot T_\epsilon$ is C^1 $O(\epsilon)$ close to $W^u(J_0 \circ I)$ in a neighborhood of the point $\hat{q}_\epsilon = q_\epsilon \cdot T_\epsilon$.*

Lemma 4 can be restated as the claim that

$$d\left(T_{\hat{q}_\epsilon}\left(\Sigma_\epsilon \cdot T_\epsilon\right), T_{r_0}\left(W^u\left(J_0 \circ I\right)\right)\right) = O(\epsilon). \tag{5.4}$$

The Exchange Lemma tells us how the manifold is configured upon exit from the neighborhood of the slow manifold. Indeed, it gives precise information about the directions present in the tangent space at that point. At the risk of being repetitious, to understand the lemma, some thought should again be given to the dimensions of the manifolds involved. The shooting manifold is $k+1$-dimensional and the unstable manifold of M_0 is $k+l$-dimensional. As commented earlier, the Exchange Lemma addresses which of the (slow) l directions is picked out by the manifold at its exit point from D. The set $J_0 \circ I$ is the (slow) trajectory through the point J_0 and is thus 1-dimensional. Its unstable manifold, $W^u(J_0 \circ I)$ is $k+1$-dimensional. The dimensions thus agree and the Exchange Lemma tells us that the extra direction picked out is exactly that determined by the slow flow itself. A picture is useful here. The basic Figure (11) that we have been using is deceptive as we can only represent there a 1- dimensional slow manifold. In Figure 16, the stable manifold is omitted and the case of a 2-dimensional M_0 is depicted with a 3-dimensional unstable manifold. The manifold $W^u(J_0 \circ I)$ is shown and the Exchange Lemma says that the shooting manifold Σ_ϵ will exit D with a nearby tangent space.

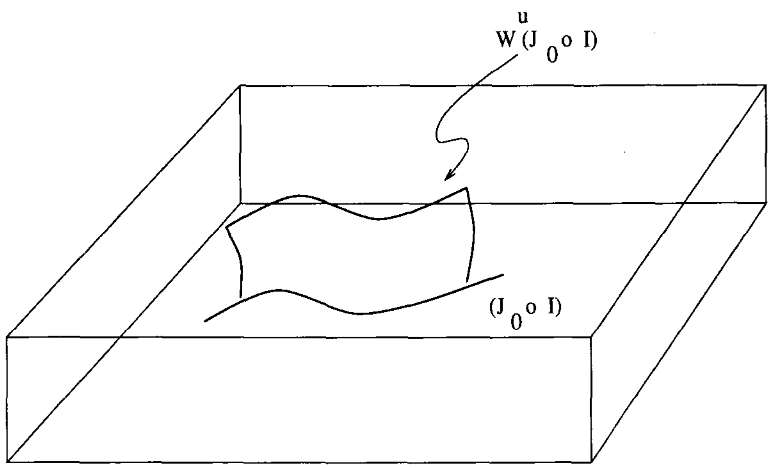

Figure 16
The unstable manifold to $J_0 \circ I$.

The application to the FitzHugh-Nagumo pulse will be given below, as well as the formulation of a general theorem concerning the construction of homoclinic orbits. However, we will first show why differential forms are relevant to the Exchange Lemma.

5.2 Differential forms and the Exchange Lemma

In order to show the C^1-closeness that is given in the Exchange Lemma, the linear spaces $T_{r_0}(W^u(J_0 \circ I))$ and $T_{\hat{q}_\epsilon}(\Sigma_\epsilon \cdot T_\epsilon)$ must be shown to be $O(\epsilon)$ close. This can be achieved by showing that the Plücker coordinates given by applying all the basis of $(k+1)$-forms to these tangent spaces are (projectively) $O(\epsilon)$ close. In other words it is not necessary that the length of the Plücker vectors are close, but only that the associated directions of these vectors are close. Amongst the Plücker coordinates there are two distinguished sets, we call block-1 forms and block-2 forms, as the resulting linear equations on these coordinates are a perturbation of a system in block form. The block-1 coordinates are those resulting from an application of the forms

$$da_1 \wedge da_2 \wedge \ldots \wedge da_k \wedge dy_i, \quad i = 1, \ldots, l, \tag{5.5}$$

and the block-2 forms are all the others. When the block-2 forms are applied to the tangent space $T_{r_0}(W^u(J_0 \circ I))$ the result is always 0. One part of the proof of the Exchange Lemma is thus to show that the block-2 forms applied to $T_{\hat{q}_\epsilon}(\Sigma_\epsilon \cdot T_\epsilon)$ are $O(\epsilon)$.

The equation for the block-1 forms must then be calculated. An approximation can be easily derived using the variational equation of the Fenichel Normal Form. Indeed, this calculation is instructive as it shows clearly why this normal form is so crucial. First, the variational equation in Fenichel coordinates can be obtained by differentiating (3.21)

$$\begin{aligned} da' &= \Lambda da + D_z(\Lambda dz)a \\ db' &= \Gamma db + D_z(\Gamma dz)b \\ dy' &= \epsilon\left\{D_y h(y,\epsilon) + H(da,b) + H(a,db) + D_z H(a,b)(a,b,dz)\right\}, \end{aligned} \tag{5.6}$$

where $z = (a, b, y)$ and the arguments of H, Λ and Γ have been suppressed. A simple approximation to (5.6) can be obtained if the underlying orbit actually lies on M_ϵ (which is not true for the orbit of interest but it does give the lowest order approximation). On M_ϵ we have that $a = b = 0$ and (5.6) simplifies to

$$\begin{aligned} da' &= A^\epsilon(y)da \\ db' &= B^\epsilon(y)db \\ dy' &= \epsilon\left\{D_y h(y,\epsilon)dy\right\}, \end{aligned} \tag{5.7}$$

where $A^\epsilon(y) = \Lambda(0,0,y,\epsilon)$ and $B^\epsilon(y) = \Gamma(0,0,y,\epsilon)$. Notice that the simple structure of (5.7) follows precisely from the elements of the Fenichel Normal Form. For instance, that the slow equation is not influenced by any fast (infinitesmal) variables comes directly from the bilinear form H. For this approximation, the equation for the block-1 forms can be easily calculated using the product rule.

$$\begin{aligned} (da_1 \wedge \ldots \wedge da_k \wedge dy_i)' &= \textstyle\sum_{j=1}^k da_1 \wedge \ldots \wedge da_j' \wedge \ldots \wedge da_k \wedge dy_i \\ &\quad + da_1 \wedge \ldots \wedge da_k \wedge dy_i', \\ &= da_1 \wedge \ldots \wedge A^\epsilon_{jl}(y)da_l \wedge \ldots \wedge da_k \wedge dy_i \\ &\quad + da_1 \wedge \ldots \wedge da_k \wedge \epsilon[\tfrac{\partial}{\partial y_l}h_i(y,\epsilon)dy_l], \end{aligned} \tag{5.8}$$

where the summation convention is being used, A^ϵ_{jl} is the jlth entry in A and h_i denotes the ith component of $h(y, \epsilon)$. In the first part, the only terms that will remain are those including $A^\epsilon_{jl}(y)da_l$, all the others will cancel. It follows that

$$\begin{aligned} (da_1 \wedge \ldots \wedge da_k \wedge dy_i)' &= \operatorname{Tr} A^\epsilon(y)[da_1 \wedge \ldots \wedge da_k \wedge dy_i] \\ &\quad + da_1 \wedge \ldots \wedge da_k \wedge \epsilon\tfrac{\partial}{\partial y_l}h_i(y,\epsilon)[da_1 \wedge \ldots \wedge da_k \wedge dy_l]. \end{aligned} \tag{5.9}$$

Now setting

$$\Omega = \begin{pmatrix} da_1 \wedge \ldots \wedge da_k \wedge dy_1 \\ \vdots \\ da_1 \wedge \ldots \wedge da_k \wedge dy_l \end{pmatrix},$$

(5.9) can be abbreviated as

$$\Omega' \;=\; \text{Tr } A^\epsilon(y)\Omega + \epsilon D_y h(y,\epsilon)\Omega. \tag{5.10}$$

Since Tr $A^\epsilon(y)$ is a scalar, an integrating factor can be introduced, which does not change the Plücker coordinates projectively. An equation for

$$\hat{\Omega} = \exp\left\{-\int_0^t \text{ Tr } A^\epsilon(y)\right\}\Omega$$

is easily computed as

$$\hat{\Omega}' = \epsilon D_y h(y,\epsilon)\hat{\Omega}, \tag{5.11}$$

which is exactly the variational equation for the slow flow! This fact is the key to the Exchange Lemma is used in the proof, sketched below, to show that the slow direction is that picked out for the remaining tangent vector for the shooting manifold after its passage near the slow manifold.

5.3 The FitzHugh-Nagumo pulse

In this section, we shall show how the Exchange Lemma can be used to complete the construction of the FitzHugh-Nagumo pulse. The shooting manifold Σ^ϵ in this case is W^u_-, the unstable manifold of the curve of rest states, which is 2-dimensional ($2 = k+1$ since $k = 1$ here). In order to emphasize the dependence on ϵ we shall write $W^{u,\epsilon}_-$. In order to apply the Exchange Lemma, we first need to see that

$$W^{u,0}_- \cap_T W^s(M^R_0),$$

but this is a statement about the $\epsilon = 0$ flow, wherein $W^{u,0}_-$ lies in the plane $w = 0$. The flow in $w = 0$ is exactly given by the equation for the travelling waves of the scalar reaction-diffusion equation, namely (4.10). Furthermore, $W^s(M^0_R) \cap \{w = 0\}$ is exactly W^s_+ which is the stable manifold of the curve $\{(1,0,c) : c \text{ near } c^*\}$. It therefore follows that $W^{u,0}_-$ transversely intersects $W^s(M - 0^R)$ inside $w = 0$. Since the full space only adds the w-direction, which is in $W^s(M^R_0)$, it follows that

$$W^{u,0}_- \cap_T W^s(M^R_0),$$

and the hypothesis of the Exchange Lemma holds.

Next, set $r_0 \in \mathcal{F}_2 \cap \partial D$, which is the intersection of the singular orbit with the boundary of D along the fast jump as it veers away from M^R_0. From Proposition 1, it follows that there is a point $\hat{q}_\epsilon = q_\epsilon \cdot T_\epsilon$ in $W^{u,0}_- \cap \partial D$ which lies at most $O(\epsilon)$ from r_0.

In order to carry the manifold $W^{u,\epsilon}_-$ over the fast jump away from M^R_0, we need to assess the configuration of $W^{u,\epsilon}_-$ at this exit point $\hat{q}_\epsilon$. This is precisely the information offered by the Exchange Lemma. Indeed, from Lemma 4 it follows that $W^{u,\epsilon}_-$ at $\hat{q}_\epsilon$ is $O(\epsilon)$ close to $W^u(J_0 \circ I)$.

The next step is then to show that

$$W^u(J_0 \circ I) \cap_T W^s(M_0^L),$$

which is a statement about the $\epsilon = 0$ flow with, moreover, c fixed at $c = c^*$, see Figure17.

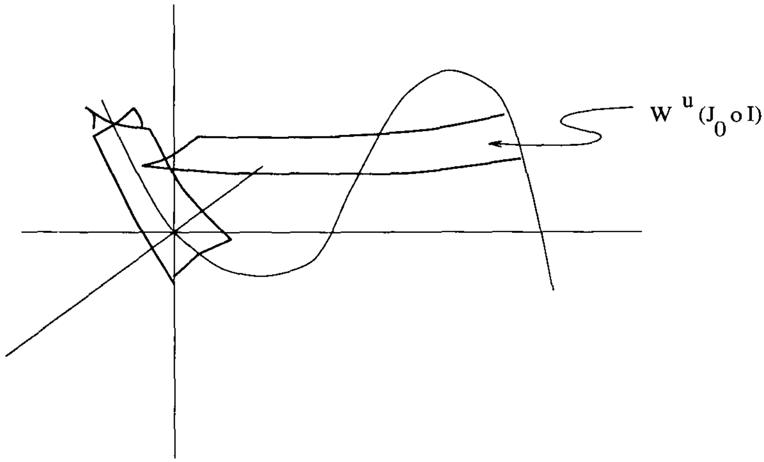

Figure 17
The transversality over the return jump.

The equation to be considered to verify this transversality is then

$$\begin{aligned} u' &= v \\ v' &= -c^* v - f(u) + w \\ w' &= 0. \end{aligned} \tag{5.12}$$

The transversality argument is then very similar to that given, except that w is now the parameter and c is fixed, over the first fast piece and is omitted here.

Since $W_-^{u,\epsilon}$ is $O(\epsilon)$ from $W^u(J_0 \circ I)$ and $W^s(M_0^L)$ is $O(\epsilon)$ from $W^s(M_\epsilon^L)$, there will also be a transversal intersection between $W_-^{u,\epsilon}$ and $W^s(M_\epsilon^L)$. It can easily be checked that $W^s(M_\epsilon^L)$ is actually the stable manifold of the curve of critical points $\{(0,0,c) : c \text{ near } c^*\}$. Thus we have constructed a homoclinic orbit from this curve of critical points to itself. Since c is actually a parameter, the orbit must lie in a fixed c slice and hence we have a homoclinic orbit to the rest state for some c near c^*.

Remark It is instructive to see how the "exchange" of information occurs in this passage near the slow manifold. All information in the speed parameter c is lost as the shooting manifold veers near to M_ϵ. However new information, namely in the w-direction is acquired. This is exactly the direction of the slow flow in the manifold M_0^R. If multiple passages near slow manifolds occur then each passage is characterized by a loss of information acquired at the previous such passage and new information from the current slow flow is substituted.

5.4 General application

The Exchange Lemma will be used in this section to derive a very general theorem on the existence of homoclinic orbits to a 1-dimensional invariant set, due to Jones and Kopell, see [28]. We consider again the general singularly perturbed system (1.1) and suppose that there is a particular slow manifold M_ϵ^1 which contains an invariant curve P_ϵ for ϵ sufficiently small, and moreover that P_ϵ is attracting relative to the (slow) flow on M_ϵ. One should think of P_ϵ as being either a curve of critical points or a periodic orbit.

The singular orbit will be constructed with (arbitrarily) many fast jumps and intervening slow pieces. We shall assume that the following ingredients are given from which this will be put together. Let ρ be some fixed integer, which will be the number of critical manifolds visited by the singular orbit.

(A1) The following sets are assumed to exist for the equation (1.1):

M_0^j, $0 \leq j \leq \rho$: These are each normally hyperbolic critical manifolds, given, as usual, by the graphs of functions i.e., each satisfying (H1)-(H3). These are not necessarily distinct and the number of normal stable and unstable directions is independent of j. Moreover $M_0^\rho = M_0^0$.

$\mathcal{F}_0^j$, $1 \leq j \leq \rho$ Each being a heteroclinic orbit from M_0^{j-1} to M_0^j, including its α and ω-limit sets.

$\mathcal{S}_0^j$, $0 \leq j \leq \rho$ If $j = 0$, $S_0^j = P_0$. If $0 < j < \rho$, S_0^j is a trajectory of the slow flow (1.7) which connects the end-point of $\mathcal{F}_0^j$ to the beginning point of $\mathcal{F}_0^{j+1}$. If $j = \rho$, it is a trajectory of (1.7) on M_0^0 connecting the end-point of $\mathcal{F}_0^\rho$ to P_0. The notation $\hat{\mathcal{S}}_0^j$ will refer to the trajectories of (1.7) extended beyond the jump points.

The singular orbit can now be constructed in the obvious way by piecing together the fast jumps with in the intervening slow pieces. We call this singular homoclinic orbit $\mathcal{H}$, see Figure 18, so that

$$\mathcal{H} = \cup_{j=1}^{j=\rho} \mathcal{F}_0^j \cup \mathcal{S}_0^j. \tag{5.13}$$

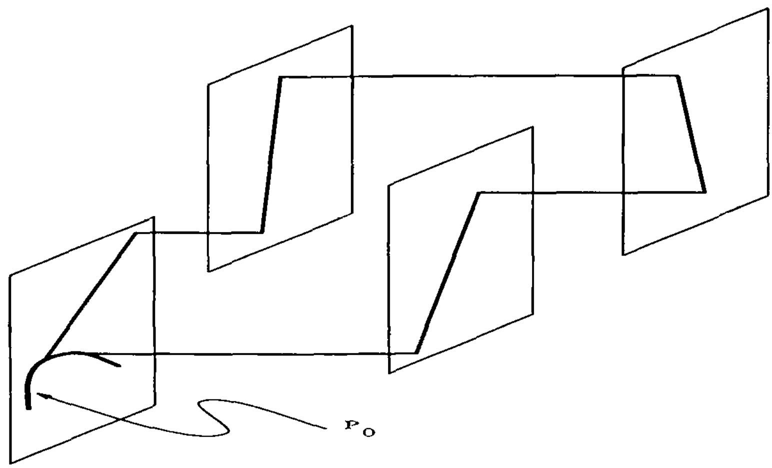

Figure 18
The singular homoclinic orbit.

A transversality assumption must also be made.

(A2) For every j so that $1 \leq j \leq \rho$,

$$W^u(\hat{\mathcal{S}}_0^{j-1}) \cap_T W^s(M_0^j). \tag{5.14}$$

Then we have the following theorem, see [28]

Theorem 8 *Under the assumptions (A1)-(A2), for $\epsilon > 0$, but sufficiently small, there is an orbit homoclinic to the set P_ϵ and $O(\epsilon)$ close to $\mathcal{H}$. Moreover, there is a neighborhood of $\mathcal{H}$ in which it is the unique such homoclinic orbit.*

Proof The proof is very simple, given the Exchange Lemma, and follows exactly the lines of the argument for the FitzHugh-Nagumo case. One takes $W^u(P_\epsilon)$, the existence of which is guaranteed by Theorem 6, and follows it around the phase space. The first transversality hypothesis $j = 1$ allows us to invoke the Exchange Lemma (Lemma 4) and conclude that we can make P_ϵ exit $O(\epsilon)$ close to the next fast jump where it exits the neighborhood of M_0^1. At that point, it is a consequence of Lemma 4 that it is $O(\epsilon)$ close to $W^u(\mathcal{S}_0^1)$ and the next transversality assumption can be used to get it by the next manifold. One continues in this fashion until $W^u(P_\epsilon)$ is seen to transversely intersect $W^s(M_0^\sigma)$. Since P_ϵ is attracting in M_0^σ, it is easily checked that the resulting intersection gives an orbit homoclinic to P_ϵ.

Theorem 8 can be applied to many problems. For instance, the reader can construct examples in which Silnikov orbits are easily seen to exist, see [57] for a discussion of the consequences of such orbits. Homoclinic orbits for the

Hodgkin-Huxley equations can be constructed using this Theorem. Bose [5] has constructed pulses for the Keener model [34] of two coupled nerve fibers each governed by the FitzHugh-Nagumo system.

5.5 Sketch of proof

The proof of the Exchange Lemma involves a fair amount of estimation and can only be sketched here. We will show, however, how the idea introduced above concerning the use of differential forms can be made into a proof. The proof given here is based on the approach developed by Tin, see [54]

Recall from the above that the forms applied to $T_{q_\epsilon} \cdot t$ are divided into two blocks. Let $Z(t)$ denote the vector of block-1 forms and $X(t)$ denote the vector of block-2 forms, the equations can then be calculated, with all the higher order terms appearing as

$$\begin{aligned} Z' &= (\text{ Tr } \Lambda + \phi + \epsilon[D_y h(y,\epsilon) + \Theta_1])\, Z + \Theta_2 X \\ X' &= (G + \Psi_1)\, X + \Psi_2 Z, \end{aligned} \tag{5.15}$$

where ϕ, Θ_i and Ψ_i are higher order terms. To be precise,

$$\begin{aligned} \|\Theta_1\| \le c_1\,|a|\,|b|\,,\ \ \|\Theta_2\| &\le c_2\,|a|\,,\ \ \|\Psi_1\| \le c_3(|a|+|b|), \\ \|\Psi_2\| \le c_4\,|b|\,, \qquad & |\phi| \le c_5\,|a|\,, \end{aligned} \tag{5.16}$$

where each of these estimates follows from the structure of the Fenichel Normal Form. Using an integrating factor as before, and keeping in mind that we only care about vectors being close in a projective sense, we can scale Z and X to obtain

$$\begin{aligned} \hat{Z}' &= \epsilon\,[D_y h(y,\epsilon) + \Theta_1]\,\hat{Z} + \Theta_2 \hat{X}, \\ \hat{X}' &= (G - \text{ Tr } \Lambda - \phi + \Psi_1)\,\hat{X} + \Psi_2 \hat{Z}. \end{aligned} \tag{5.17}$$

The dominant term in the second equation is $G-$ Tr Λ, and the idea is that its spectrum has negative real part, on account of the fact that Tr Λ is the sum of the real parts of the eigenvalues with positive real part (for the case that the underlying point lies in M_0). This can be used to show that $\hat{X}$ is forced to decay exponentially, although the actual estimates are quite difficult due to the coupling.

We need to determine $\hat{Z}(t)$ and this is achieved by a series of approximation. Consider the equation

$$\hat{\Omega}' = \epsilon D_y h(\hat{y},\epsilon)\hat{\Omega}, \tag{5.18}$$

where $\hat{y}(t)$ is a solution of (3.21) chosen as follows: let p_ϵ be chosen in $\Sigma_\epsilon \cap W^s(M_\epsilon)$ exponentially close to q_ϵ, which is possible (why?). Next set $p_\epsilon \cdot t = (\hat{x}(t), \hat{y}(t), \epsilon)$, and check that there are $\kappa_1 > 0$ and $\beta_1 < 0$ so that

$$|y(t) - \hat{y}(t)| \le \kappa_1 e^{-\beta_1 t}. \tag{5.19}$$

Concerning $\hat{Z}(t)$ and $\hat{\Omega}(t)$ we then have the

Lemma 5 $\left|\hat{Z}(t) - \hat{\Omega}(t)\right| \leq \kappa_1 e^{-\beta_1 t}$.

The idea behind the proof of Lemma 5 is to estimate the terms in the equation satisfied by $\hat{Z}(t) - \hat{\Omega}(t)$, which is

$$\begin{aligned} \left(\hat{Z} - \hat{\Omega}\right)' = \quad & \epsilon \Big[D_y h(y,\epsilon)(\hat{Z} - \hat{\Omega}) \\ +(D_y h(y,\epsilon) \quad & -D_y h(\hat{y},\epsilon))\hat{\Omega} + \Theta_1 \hat{Z}\Big] + \Theta_2 \hat{X}. \end{aligned} \tag{5.20}$$

One then uses the known estimates on the last three terms followed by the Gronwall inequality. Each of these terms is exponentially small, provided $\hat{Z}$ and $\hat{\Omega}$ are bounded. The first on account of (5.19), the second because $\|\Theta_1\| \leq c\,|a|\,|b|$ and ab is exponentially small as both are bounded and one of them is always exponentially small. Similarly for the last term as $\hat{X}$ is exponentially decaying.

It should be noted here that this is the key point at which the structure of the bilinear term, which encodes the Fenichel fibering, is used. Indeed, the bound on $\|\Theta_1\|$ in terms of ab is essential.

The next step in the proof of the Exchange Lemma is to introduce an approximation to $\hat{\Omega}(t)$. Let $\tilde{\Omega}(t)$ satisfy the same equation as $\hat{\Omega}(t)$, namely (5.18), but with $\tilde{\Omega}(0)$ being the block-1 forms of the space $T_{\pi^+(\hat{q}_\epsilon)} W^u(S_{\pi^+(\hat{q}_\epsilon)})$, where S_y is the slow trajectory through the point y on M_ϵ. Note that this space has block-2 forms all being 0. Note that $\pi^+(\hat{q}_\epsilon) = \hat{y}(0)$ and hence $\tilde{\Omega}(t)$ will be the block-1 forms of

$$T_{(0,\hat{y}(t),\epsilon)} W^u(S_{\hat{y}(t)}),$$

which are $O(\epsilon)$ close to those of

$$T_{(0,\hat{y}(t),0)} W^u(J_0 \circ I),$$

for appropriate choice of the interval I and normalization. It thus remains to estimate $\hat{\Omega}(t) - \tilde{\Omega}(t)$. But it can be shown that $\left|\hat{\Omega}(0) - \tilde{\Omega}(0)\right|$ is exponentially small and, since they satisfy the same equation, we can conclude that, for appropriate values of t $\left|\hat{\Omega}(t) - \tilde{\Omega}(t)\right|$ is also exponentially small. The steps are now completed by invoking Lemma 5, which implies the Exchange Lemma provided that $\hat{Z}(0)$ is bounded away from 0, otherwise the estimates may be vacuous. But this follows from the transversality hypothesis.

Chapter 6

Generalizations and Future Directions

The Exchange Lemma given in the previous lecture is inadequate in two respects.

1. *The* $(k+1)$ *Exchange Lemma will be inadequate for many applications.* For instance suppose that the periodic orbit or curve of critical points in the example above were *not attracting* in the *slow manifold*, we would then have to follow an invariant manifold that was larger than $k+1$ as it would have more slow directions. The same would be true if the invariant sets, to which we sought homoclinic orbits, were higher-dimensional.

2. $O(\epsilon)$ *is not good enough.* If, for instance, we are considering a perturbed Hamiltonian system the transversality of the shooting manifold Σ_ϵ with the stable manifold to the slow manifold, namely $W^s(M_\epsilon)$ may only be of $O(\epsilon)$ and not $O(1)$, as we have used above. This turns out not to be a problem in reaching the conclusion of the Exchange Lemma, but that conclusion will itself be useless in this case as we would be following an invariant manifold by an $O(\epsilon)$ approximate version and the approximate version would itself intersect the next stable manifold only at $O(\epsilon)$, since these might cancel a more accurate estimate must be found in the Exchange Lemma.

In the first few sections of this last lecture, I will give the generalizations of the Exchange Lemma that address these points. In the last sections, I will discuss various applications and the directions for further study that they suggest.

6.1 The $(k+\sigma)$ Exchange Lemma

We suppose now that the shooting manifold Σ_ϵ is a $(k+\sigma)$-dimensional locally invariant manifold, again smooth in ϵ. Recall that k is the number of unstable directions for the slow manifold. The cases of interest are when $1 < \sigma < l$, for then the Exchange Lemma of the last lecture does not apply and yet the dimension of the shooting manifold is still not equal to the full dimension of the unstable manifold of the slow manifold. However the Exchange Lemma I shall state in fact applies for $0 \leq \sigma \leq l$.

An assumption must again be made as to how the shooting manifold enters a neighborhood of the slow manifold. Indeed, we assume

(H5) There is a $(k + \sigma)$-dimensional, locally invariant manifold Σ_ϵ, defined for $0 < \epsilon \ll 1$, and smooth in ϵ, so that

$$\Sigma_0 \cap_T W^s(M_0) \neq \emptyset,$$

at a point $q \in \partial D$.

Note that in this case the intersection will be more than a trajectory indeed, by a dimension count, we expect it to be σ-dimensional. Let V be some suitably chosen neighborhood of the point $q \in \Sigma_0 \cap W^s(M_0) \cap \partial D$ and consider the set

$$\begin{aligned} J_0 &= \omega\left(\Sigma_0 \cap W^s(M_0) \cap V\right) \\ &= \pi^-\left(\Sigma_0 \cap W^s(M_0) \cap V\right), \end{aligned} \tag{6.1}$$

in M_0. By the transversality hypothesis (H5) the set J_0 will be a $\sigma - 1$ dimensional submanifold of M_0. We need a transversality hypothesis on the slow flow.

(H6) The set J_0 is a $(\sigma - 1)$-dimensional manifold and the slow flow i.e., that associated with (1.7), is *not tangent* to J_0, in other words

$$g(\pi^-(q), 0) \notin T_{\pi^-(q)} J_0. \tag{6.2}$$

With the same notation as in the previous section, I can state the $(k + \sigma)$-Exchange Lemma, which is due to Tin and the author, see [30].

Lemma 6 **$(k + \sigma)$-Exchange Lemma** *The manifold $\Sigma_\epsilon \cdot T_\epsilon$ is C^1 $O(\epsilon)$ close to $W^u(J_0 \circ I)$ in a neighborhood of the point $\hat{q}_\epsilon = q_\epsilon \cdot T_\epsilon$.*

The proof follows the same lines as the sketch of the $k + 1$ case given in the previous lecture, see [30] or [54]. In his thesis, Tin [54], formulated a general theorem concerning the existence of homoclinic orbits to invariant subsets of a slow manifold. Let P_0 be a γ-dimensional invariant, compact submanifold of M_0^1 that is normally hyperbolic under the slow flow on M_0^1. Further, let U_0 be the $\gamma + \eta$-dimensional unstable manifold of P_0 in M_0^1, in the slow flow on M_0^1, given as usual by (1.7). In a similar fashion to the application given in the previous lecture, we assume the presence of the objects given in the next hypothesis.
(A3) The following sets exist as stated:

M_0^j, $0 \leq j \leq \rho$: These are each normally hyperbolic critical manifolds, given, as usual, by the graphs of functions i.e., satisfying (H1)-(H3). These are not necessarily distinct and the number of normal stable and unstable directions is independent of j. Moreover $M_0^\rho = M_0^0$.

$\mathcal{F}_0^j$, $1 \leq j \leq \rho$: Each being a heteroclinic orbit from M_0^{j-1} to M_0^j.

$\mathcal{S}_0^j$, $0 \leq j \leq \rho$: If $j = 0$, S_0^j is P_0 together with a curve in U_0^j that connects P_0 to the beginning point of $\mathcal{F}_0^1$. If $0 < j < \rho$, S_0^j is a trajectory of the slow flow (1.7) which connects the end-point of $\mathcal{F}_0^j$ to the beginning point of $\mathcal{F}_0^{j+1}$. If $j = \rho$, it is a trajectory of (1.7) on M_0^0 connecting the end-point of $\mathcal{F}_0^\rho$ to P_0.

The singular orbit is given similarly by the expression (5.13).

Since P_0 is assumed to be normally hyperbolic in M_0^0, see Fenichel [15], it will perturb to a (locally) invariant manifold in M_ϵ^0, say P_ϵ. The next theorem will give conditions under which the singular orbit $\mathcal{H}$ perturbs to an actual homoclinic orbit to P_ϵ when $\epsilon > 0$ but sufficiently small.

Needless to say, there are transversality hypotheses to be satisfied along each of the fast jumps. However, another collection of sets need to be determined in order to express these transversality conditions.

(A4) The following sets exist as stated:

U_0^j, $0 \leq j \leq \rho - 1$: Subsets of M_0^j. If $j = 0$, U_0^j is the unstable manifold in M_0^1 of P_0 under the flow of (1.7). They are then defined inductively along with the sets below by

$$U_0^j = J_0^j \circ I, \tag{6.3}$$

where $\circ$ refers to the slow flow on M_0^j, and $I = (\hat{\tau}_j - \eta_j, \hat{\tau}_j + \eta_j)$, with $\hat{\tau}_j$ chosen so that the beginning point of $\mathcal{F}_0^{j+1}$ is contained in U_0^j.

J_0^j, $1 \leq j \leq \rho - 1$: Subsets of M_0^j, that are defined by

$$\omega \left(W^u(U_0^{j-1}) \cap W^s(M_0^j) \right), \tag{6.4}$$

inductively along with the above.

For these sets to be well-defined and to be able to apply the Exchange Lemma, the following transversality hypotheses must hold.

(A5) If $1 \leq j \leq \rho - 2$ then

$$W^u(U_0^j) \cap_T W^s(M_0^{j+1}) \neq \emptyset, \tag{6.5}$$

along $\mathcal{F}_0^j$.

(A6) If $1 \leq j \leq \rho - 2$ then the slow vector field is not tangent to the set J_0^{j+1}, which is assumed to be a $(\sigma - 1)$-dimensional manifold.

A final condition is needed for the last jump. The set V_0 is the stable manifold of P_0 in M_0^0, relative to the slow flow.

(A7) Along $\mathcal{F}_0^\rho$

$$W^u(U^{\rho-1})_0 \cap_T W^s(V_0) \neq \emptyset.$$

The following theorem has been proved by Tin, see [54].

Theorem 9 *Under the assumptions (A3)-(A7), if $\epsilon > 0$, but sufficiently small, then there is an orbit homoclinic to P_ϵ that is within $O(\epsilon)$ of the singular homoclinic orbit $\mathcal{H}$.*

6.2 Exponentially small Exchange Lemma

The estimate in the Exchange Lemma can be significantly tightened if we are willing to drop the comparison with the $\epsilon = 0$ case. Indeed, a perusal of the proof shows that the $O(\epsilon)$ estimate comes in when going from the tangent space to the unstable manifold of the appropriate subset of M_ϵ to that of the slow trajectory in M_0. Since the Fenichel theory does supply us with a full structure when $\epsilon \neq 0$, we can easily make the comparison with the $\epsilon \neq 0$ object.

Furthermore, the transversality at entry to D can be significantly weakened, as we only need to assume that the transversality occurs when $\epsilon \neq 0$. The transversality can be measured by taking bases for the tangent spaces and wedging the entire set of vectors (taking care not to repeat the vector field and keeping the basis bounded away from 0 and ∞ as $\epsilon \to 0$) to make a volume form. For the Exchange Lemma with Exponentially Small Error we make the following hypothesis.

(H6) There is a $(k + \sigma)$-dimensional, locally invariant manifold Σ_ϵ, defined for $0 < \epsilon \ll 1$, and smooth in ϵ, so that

$$\Sigma_\epsilon \cap_T W^s(M_\epsilon) \neq \emptyset,$$

at a point $q \in \partial D$ and the transversality is of $O(\epsilon^\rho)$ for some ρ.

Set

$$J_\epsilon = \pi^- \left(\Sigma_\epsilon \cap W^s(M_\epsilon)\right), \tag{6.6}$$

where again π^- is the Fenichel map sending points to the base points of their fiber. Notice that, in this case, the Fenichel map cannot be replaced by an ω-limit set as there is a genuine flow on the manifold M_ϵ. By the same token, the set J_ϵ contains directions from the slow flow and thus J_ϵ has one higher dimension than J_0. We can then replace $J_0 \circ I$ in the statement of the Exchange Lemma by $J_\epsilon \cdot T$ for some appropriately chosen T.

Lemma 7 Exchange Lemma with Exponentially Small Error *The manifold $\Sigma_\epsilon \cdot T_\epsilon$ is C^1 $O(\exp\{\frac{-c}{\epsilon}\})$ close to $W^u(J_\epsilon \cdot T)$ at $q_\epsilon \cdot T_\epsilon$, for some $c > 0$.*

This Lemma was first proved by Jones, Kaper and Kopell [27] in the $\sigma = 1$ case and later by Jones and Tin [30] for the general case. It is crucial that the full structure of the Fenichel Normal Form be used in order to obtain this result. In particular, the decomposition of the slow vector field as a bilinear form, which comes from the fibering, is crucial. The exponential closeness, and its relation to the bilinear form, was noticed independently by Sandstede (private communication).

6.3 Pendulum forced by two frequencies

In this section, I will outline an application that uses the Exchange Lemma with Exponentially Small Error. Consider the pendulum forced by two separate frequencies

$$\begin{aligned} q'' + \sin q &= \epsilon\,(\delta q' + \gamma(\tau_1, \tau_2)) \\ \tau_1' &= \epsilon\omega_1(\tau_1, \tau_2) \\ \tau_2' &= \epsilon\omega_2(\tau_1, \tau_2), \end{aligned} \tag{6.7}$$

where γ and ω_i, $i = 1, 2$ are all periodic, with period 2π in each of their arguments and also assumed to be smooth. Equation (6.7) is rewritten as a system

$$\begin{aligned} q' &= p \\ p' &= -\sin q + \epsilon\,(\delta p + \gamma(\tau_1, \tau_2)) \\ \tau_1' &= \epsilon\omega_1(\tau_1, \tau_2) \\ \tau_2' &= \epsilon\omega_2(\tau_1, \tau_2). \end{aligned} \tag{6.8}$$

The fast flow is obtained, as usual, by setting $\epsilon = 0$. There are critical manifolds given by any of the critical points of the pendulum. We set then

$$M_0^\pm \subset \{p = 0,\ q = \pm\pi\}. \tag{6.9}$$

Since both of these points are saddles of the pendulum, the manifolds $M_0^\pm$ are normally hyperbolic and hence the Fenichel Theorems apply. Various identifications can be made in the phase space. Since the vector field is periodic in τ_1 and τ_2, both τ_1 and τ_2 can be identified modulo 2π. The critical manifolds would then become tori. We have not set the theory up to cover such manifolds and hence we shall take this point of view. I shall assume that $M_0^\pm$ is defined on a sufficiently large region that it will contain a fundamental domain of the torus. Thus $M_\epsilon^\pm$ can be viewed as a torus after the fact by identifying τ_1 and τ_2 after the results have been obtained.

The slow flow on $M_0^\pm$ is given by the equations:

$$\begin{aligned} \dot\tau_1 &= \omega_1(\tau_1, \tau_2) \\ \dot\tau_2 &= \omega_2(\tau_1, \tau_2), \end{aligned} \tag{6.10}$$

which are also the exact equations on $M_\epsilon^\pm$ as the fast variables do not appear in the slow equations at all. I shall assume that, on $M_0^\pm$ there are two periodic orbits, one attracting, say P_a, and the other repelling P_r, see Figure 19. Of interest will be the construction of orbits homoclinic to P_a, which will require use of the $(k+1)$- Exchange Lemma with the exponentially small error. From the point of view of applying the Exchange Lemma, the construction of orbits homoclinic to P_r, although possible, is not as interesting (why?).

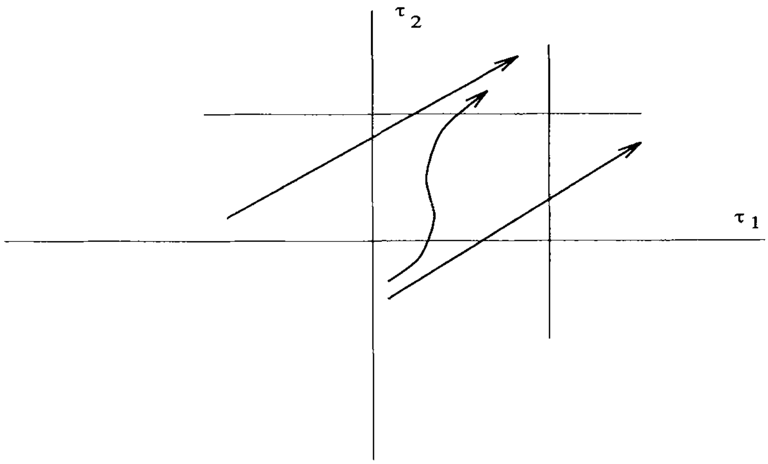

Figure19
The slow flow.

Note that the fast flow decouples completely from the slow variables, when $\epsilon = 0$. Since there is a heteroclinic orbit from $(0, -\pi)$ to $(0, \pi)$ every point on M_0^- is connected to its partner on M_0^+. There is therefore a problem in even constructing the singular orbit as it has to be decided which heteroclinic orbit is picked out. One can formulate a Melnikov function $\Gamma^+(\gamma, \tau_1, \tau_2)$ whose zeroes indicate the location of heteroclinics that exist for small values of ϵ from M_ϵ^- to M_ϵ^+. Another Melnikov function $\Gamma^-(\gamma, \tau_1, \tau_2)$ indicates the potential heteroclinic orbits from M_ϵ^+ back to M_ϵ^-. I assume that the function γ and the quantity δ are chosen so that the zero sets $\Gamma^- = 0$ and $\Gamma^+ = 0$ cross P_a transversely, for an example see [27]. The quantity δ is then fixed.

A singular orbit is constructed with ρ jumps as follows. If j is odd then $M_0^j = M_0^-$ and if j is even $M_0^j = M_0^+$. The singular orbit is then composed of the following pieces

$\mathcal{F}_0^j$ $1 \leq j \leq \rho$, being each the heteroclinic jump from M_0^j to M_0^{j+1} from a point (τ_1, τ_2) where $\Gamma^+ = 0$, if j is odd and from a point where $\Gamma^- = 0$ if j is even.

$\mathcal{S}_0^j$ $1 \leq j \leq \rho$ are pieces of P_a between end-points of $\mathcal{F}_0^j$ and the beginning point of $\mathcal{F}_0^{j+1}$.

The singular orbit is constructed as usual

$$\mathcal{H} = \cup_{i=1}^{\rho} \mathcal{F}_0^j \cup \mathcal{S}_0^j.$$

It is shown in [27] that there is a (real) homoclinic orbit nearby this singular orbit. The strategy is, by now, standard. One follows the unstable manifold of P_a around the phase space using the singular orbit as a template. The Melnikov calculation gives the required transversality upon entering each slow manifold and the exponentially small error in the Exchange Lemma Lemma 7 allows the passage of the unstable manifold near each slow manifold and sets the shooting manifold up for its next jump. The exponentially small error is needed as the transversality over the jumps is only $O(\epsilon)$. If the unstable manifold were only tracked up to $O(\epsilon)$ during its passage near the slow manifold, a transversal intersection with the next manifold could not be guaranteed. The details are given in [27]

6.4 Recent results and new directions

In this final section, I shall highlight some recent pieces of work that indicate directions in which the theory described above needs to be extended.

6.4.1 Orbits homoclinic to resonance

The problem of orbits homoclinic to resonance has been considered recently by many authors. The equations are as follows

$$\begin{aligned} \dot{x} &= JD_xH(x,I) + \delta f_1(x,I,\theta,\lambda) \\ \dot{I} &= \delta g_1(x,I,\theta,\lambda) \\ \dot{\theta} &= D_IH(x,I) + \delta g_2(x,I,\theta,\lambda), \end{aligned} \tag{6.11}$$

where all the functions are smooth and periodic of period 2π in θ The resonance occurs at a point $(\hat{x},\hat{I})$ which is a critical point of the $\epsilon = 0$ system, that is assumed to be a saddle point for the x equation. This is only a circle of critical points, but, as is typical in such resonance problems, can be blown up to expose some interesting structure. Setting $I = \hat{I} + \sqrt{\delta}h$, one arrives at the system, with $\epsilon = \sqrt{\delta}$

$$\begin{aligned} \dot{x} &= JD_xH(x,h,\epsilon) + \epsilon^2 f_1(x,h,\theta,\lambda,\epsilon) \\ \dot{h} &= \epsilon g_1(x,h,\theta,\lambda,\epsilon) \\ \dot{\theta} &= D_IH(x,\hat{I}) + \epsilon g_3(x,h,\theta,\lambda,\epsilon), \end{aligned} \tag{6.12}$$

which has a manifold of critical points M_0 given by $x = \hat{x}$, this is now an annulus. The system (6.12) is not in the usual singular perturbation form due to the term $D_IH(x,\hat{I})$, but it does possess a manifold of critical points parametrized by θ and I. In fact, this is all that is used by the results given in these lectures and everything applies to (6.12) as if both h and θ were slow variables. The normal hyperbolicity of $M_0 \subset \{x = \hat{x}\}$ follows from the fact that $\hat{x}$ is a saddle of the x-equation.

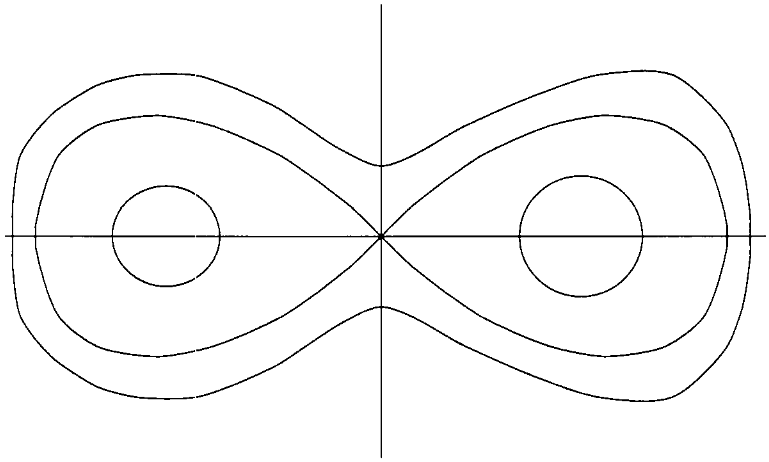

Figure 20
The scaled slow flow.

The original motivation for the study of (6.11) is a 2-mode truncation model of the nonlinear Schrödinger equation. It is desired to find orbits that are observed in certain spatio-temporal chaotic patterns, see McLaughlin et al. [4]. The slow flow will have the form of a "bow-tie" as shown in Figure 20 and the fast flow will be a "fish", see Figure 21. Of interest are orbits that are homoclinic to various different invariant sets in the slow flow, in particular: the saddle, the center(s) and the periodic orbits surrounding the centers. This has been treated by many authors using a combination of Hamiltonian techniques and those of singular perturbation theory. In the earliest work, Kovacic and Wiggins [35] found orbits connecting the center to itself involving one fast jump, however they needed some negative damping to get the orbit. Mclaughlin et al. [43] found saddle to saddle connections, with one fast jump in the case of dissipative perturbations. Simultaneously, Kovačič also solved this problem [36]. For Hamiltonian perturbations, Kovačič [37] and Haller and Wiggins [22] found periodic to periodic orbits, amongst others, with one fast jump. Haller and Wiggins, in further work [23], found multi-jump orbits for the problem with Hamiltonian perturbations. These are different, however, from those that one would construct using an Exchange Lemma argument as the time spent near the slow manifold is not great. Tin and Camassa [9] have found multi-jump orbits in a problem that is very close to the above but in the non-resonant case. In some ways, this case is harder as it is not a singular perturbation problem. This, and related work [8], [55], have made an important contribution to the problem of whether slow manifolds in atmospheric flows exist, see [40]. Two aspects of this problem are worthy of mention in the current context. First, Tin [55] developed a new strategy for proving transversality as the traditional Melnikov method was not applicable. Second, an Exchange Lemma that works for perturbed invariant manifolds in more general situations than found in singular perturbation theory was needed. Tin [54] developed such an Exchange Lemma for passages near "non-slow" manifolds.

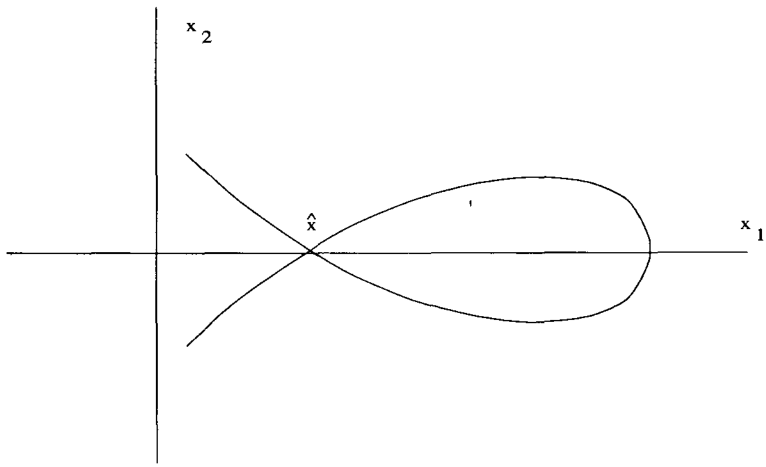

Figure 21
The fast flow.

In the most recent contribution to the orbits homoclinic to resonance problem, Kaper and Kovacic [32] have applied the Exchange Lemma to build on the single jump pulses and found multi-jump orbits homoclinic to all the available different invariant sets, for the details see [32].

The problems considered here only apply to an approximation to the real problem, namely the 2-mode truncation. The real problem is in infinite dimensions and this needs to be resolved. In particular, the Fenichel structure are not clearly understood in infinite dimensions. The Exchange Lemma also appears to be hard to generalize.There are many problems and issues in this area that, I believe, will receive much attention over the coming years.

6.4.2 Stability of travelling waves

Much of the motivation for the Exchange Lemma came from the need to compute directions of transversality for assessing the stability of travelling waves. As seen in many of the examples above, travelling waves are constructed as homoclinic orbits for particular systems of ODE's. These will be constructed then as intersections of stable and unstable manifolds of the relevant critical point. A parameter is obviously needed to make this situation robust. In most cases of travelling waves the parameter is supplied by the speed. It was shown by Evans [14] that the direction in which the unstable manifold crosses the stable manifold as the speed parameter varies renders some crucial information about the stability of the wave. Indeed, it determines the parity of the number of eigenvalues of the linearization of the PDE at the wave that lie in the right half-plane. Determining the nature of the transversal intersection in many of these travelling wave problems then becomes an important issue.

This direction is implicit in the Exchange Lemma analysis but has not, up to this point, been explicitly incorporated into the theory. In recent work, Bose and Jones [6] have studied the example of travelling pulses in coupled nerve fibers. Keener [34] formulated a model for this situation in terms of a pair of coupled FitzHugh-Nagumo equations. The fibers are coupled by reciprocal diffusive coupling. The PDE's are

$$\begin{aligned} u_{1t} &= u_{1xx} + f(u_1) - w_1 + d(u_2 - u_1) \\ w_{1t} &= \epsilon(u_1 - \gamma w_1) \\ u_{2t} &= u_{2xx} + f(u_2) - w_2 + d(u_1 - u_2) \\ w_{2t} &= \epsilon(u_2 - \gamma w_2). \end{aligned} \tag{6.13}$$

It is clear that there will be an "in-phase" wave, when ϵ is sufficiently small, for which $u_2 = u_1$, which is just the individual FitzHugh-Nagumo pulse constructed above on each fiber. However, its stability is not obvious. Bose and Jones [6] have proved the stability using the appropriate generalization of Evans' idea due to Alexander, Gardner and Jones [1]. The Exchange Lemma is crucial in tracking the relevant unstable manifold and studying its intersection with the stable manifold. This contains much more information than is in the individual travelling pulses. However, the transversality given by the Exchange Lemma does not readily give the "direction" information required in applying the Evans stability idea. Bose [5] developed a way of keeping track of this sign information in the proof of the Exchange Lemma and this is a key ingredient in the stability analysis of these "in-phase" waves.

A general theory is needed for tracking the invariant manifolds in the Exchange Lemma including the "sign" of a relevant basis. I anticipate that the Exchange Lemma will have many further applications in the stability analysis of travelling waves. Indeed, the eigenvalue equations that occur in the stability analysis are ODE's living naturally in the tangent bundle to the travelling wave phase space. If the underlying problem is singularly perturbed then so are the eigenvalue equations. The eigenvalues occur at values of the eigenvalue parameter at which certain heteroclinic orbits exist. The structure of these orbits, and consequently the distribution of eigenvalues, promises to be significantly illuminated by direct application of the Exchange Lemma.

Bibliography

[1] J.Alexander, R.Gardner & C.Jones, A topological invariant arising in the stability analysis of travelling waves, *J. reine angew. Math.* 410 (1990) 167-212.

[2] D.Aronson & H.Weinberger, Nonlinear diffusion in population genetics, combustion and nerve propagation, in *Partial Differential Equations and Related Topics*, J.Goldstein ed., Lecture Notes in Mathematics, 446 (1975), Springer-Verlag, New York.

[3] P.Bates & C.Jones, Invariant manifolds for semilinear partial differential equations, *Dynamics Reported vol.2*, Wiley (1989), 1-38.

[4] A.R.Bishop, M.G.Forest, D.W.McLaughlin and E.A.Overman II, Correlations between chaos in a perturbed Sine-Gordon equation and a truncated model system, *SIAM J. Math. Anal.* (1990) 1511-1536.

[5] A.Bose, Existence and stability of travelling waves for coupled nerve axon equations, Ph.D. thesis, Brown U., 1993.

[6] A.Bose & C.Jones, Stability of the in-phase travelling wave solution in a pair of coupled nerve fibers, preprint (1994).

[7] G.Caginalp & P.Fife, Higher order phase field models and detailed anisotropy, *Phys. Rev. B*, 34 (1986) 4940-4943.

[8] R.Camassa, On the geometry of a slow manifold, preprint (1994).

[9] R.Camassa & S.-K.Tin, The global geometry of the slow manifold in Lorenz-Krishnamurthy model, preprint (1994).

[10] G.Carpenter, A geometric approach to singular perturbation problems with applications to nerve impulse equations, *JDE* 23 (1977) 335-367.

[11] K.W.Chang & F.A.Howes, *Nonlinear Singular Perturbation Phenomena: Theory and Applications*, Springer-Verlag, New York, 1984.

[12] B.Deng, The existence of infinitely many traveling front and back waves in the FitzHugh-Nagumo equations, *SIAM J. Math. Anal.* 6 (1991).

[13] N.Ercolani, D.McLaughlin & H.Roitner, Attractors and transients for a perturbed perioidc KdV equation: a nonlinear spectral analysis, *J. Nonlin. Sci.* in press.

[14] J.W.Evans, Nerve impulse equations I-IV, *Indiana Univ. Math. Journal*, 21,22,24 (1972-5).

[15] N.Fenichel, Persistence and smoothness of invariant manifolds for flows, *Indiana Univ. Math. Journal*, 21 (1971) 193-226.

[16] N.Fenichel, Asymptotic stability with rate conditions, *Indiana Univ. Math. Journal*, 23 (1974) 1109-1137.

[17] N.Fenichel, Asymptotic stability with rate conditions II, *Indiana Univ. Math. Journal*, 26 (1977) 81-93.

[18] N.Fenichel, Geometric singular perturbation theory for ordinary differential equations, *J. Diff. Eq.* 31 (1979) 53-98.

[19] R.Gardner & C.Jones, Traveling waves of a perturbed diffusion equation arising in a pahse field model, *Indiana U. Math. J.* 38 (1989) 1197-1222.

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

[21] J.Hadamard, Sur l'iteration et les solutions asymptotiques des equations differentielles, *Bull. Soc. Math. France* 29 (1901) 224-228.

[22] G.Haller & S.Wiggins, Orbits homoclinic to resonances: the Hamiltonian case, *Physica D* 66 (1993) 298-346.

[23] G.Haller & S.Wiggins, N-pulse homoclinic orbits in perturbations of resonant Hamiltonian systems, to appear in *Arch. Rat. Mech. Anal.*(1995).

[24] S.Hastings, On travelling wave solutions of the Hodgkin-Huxley equations, *Arch. Rat. Mech. Anal.* 60 (1976) 229-257.

[25] M.Hirsch, C.Pugh & M.Shub, *Invariant Manifolds*, Lecture Notes in Mathematics, 583 (1977), Springer-Verlag, New York.

[26] C.Jones, Stability of the travelling pulse of the FitzHugh-Nagumo system, Trans. AMS 286 (1984) 431-469.

[27] C.Jones, T.Kaper & N.Kopell, Tracking invaraint manifolds up to exponentially small errors, *SIAM J. Math. Anal.*, to appear.

[28] C.Jones & N.Kopell, Tracking invariant manifolds with differential forms in singularly perturbed systems, *J. Diff. Eq.*, in press.

[29] C.Jones, N.Kopell & R.Langer, Construction of the FitzHugh- Nagumo pulse using differential forms, in *Patterns and Dynamics in Reactive Media*, H.Swinney, G.Aris, and D.Aronson eds., IMA Volumes in Mathematics and its Applications, 37 (1991), Springer-Verlag, New York.

[30] C.Jones & S.-K.Tin, On the dynamics of tangent spaces near a normally hyperbolic manifold, in preparation.

[31] C.Jones, J.Rubin & M.Maxey, Settling and asymptotic motion of aerosol particles in a cellular flow field, preprint (1994).

[32] T.J.Kaper & G.Kovačič, Multi-bump orbits homoclinic to resonance bands, preprint (1993).

[33] T.J.Kaper & S.Wiggins, On the structure of separatrix swept regions in singularly perturbed Hamiltonian systems, *J. Diff. and Int. Eq.* 5 (1992) 1363- 1381.

[34] J.Keener, Frequency decoupling of parallel excitable fibers, *SIAM J. Appl. Math.* 49 (1989) 211-230.

[35] G.Kovačič & S.Wiggins, Orbits homoclinic to resonance with an application to chaos in a model of the forced damped sine-Gordon equation, *Physica D* 57 (1992) 185-225.

[36] G.Kovačič, Singular perturbation theory for homoclinic orbits in a class of near-integrable dissipative systems, to appear in *SIAM J. Math. Anal.*

[37] G.Kovačič, Singular perturbation theory for homoclinic orbits in a class of near-integrable Hamiltonian systems, *J. Dyn. Diff. Eq.* 5 (1993) 559-597.

[38] R.Langer, Existence and Uniqueness of Pulse Solutions to the FitzHugh-Nagumo Equations, Ph.D. Thesis, Northeastern University, 1980.

[39] M.Levi, *Qualitative Analysis of the periodically forced relaxation oscillations*, Vol.244 in Memoirs AMS, AMS, Providence, RI, 1981.

[40] E.N.Lorenz, The slow manifold-What is it?, *J. Atmos. Sci.* 49 (1992) 2449-2451.

[41] W.S.Massey, *Singular Homology Theory*, Graduate Texts in Mathematics 70, Springer-Verlag, New York, 1980.

[42] R.McGehee, The stable manifold theorem via an isolating block, Symposium on ODE, W.Harris & Y.Sibuya eds., Springer-Verlag,New York (1973) 135-44.

[43] D.W.McLaughlin, E.A.Overman II, S.Wiggins & C.Xiong, Homoclinic orbits in a four dimensional model of a perturbed NLS equation: a geometric singular perturbation study, to appear in *Dynamics Reported.*

[44] E.F.Mischenko & N.Rozov, *Differential Equations with small parameters and realaxtion oscillations*, Plenum Press, New York, 1980.

[45] Y.Nishiura & H.Fujii, Stability of singularly perturbed solutions to a system of reaction-diffusion equations, *SIAM J. Math. Anal.*, 18 (1987) 1726- 1770.

[46] Y.Nishiura & M.Mimura, Layer oscillations in reaction-diffusion systems, *SIAM J. Appl. Math.* 49 (1989) 481-514.

[47] T.Ogawa, Travelling wave solutions to perturbed Korteweg-de Vries equations, preprint (1993), to appear in Hiroshima J. Math..

[48] O.Perron, DieStabilitätsfrage bei Diffferentialgleichungsysteme, *Math. Zeit.* 32 (1930) 703-728.

[49] D.Ruelle, *Elements of Diffrentiable Dynamics and Bifurcation Theory*, Academic Press, San Diego, 1989.

[50] Yu.Rzhanov, H.Richardson, A.Hagberg & J.Moloney, Spatio- temporal oscillations in a semiconductor etalon, preprint (1992).

[51] K.Sakamoto, Invariant manifolds in singular perturbation problems for ordinary differential equations, *Proc. Roy. Soc. Ed.*, 116A (1990) 45-78.

[52] H.Stommel, Trajectories of small bodies sinking slowly through convection cells, *J. Mar. Res.*, 8 (1949) 24-29.

[53] P.Szmolyan, Analysis of a singularly perturbed travelling wave problem, *SIAM J. Appl. Math.* (1992)

[54] S.-K.Tin, On the dynamics of tangent spaces near a normally hyperbolic manifold, Ph.D. Thesis, Brown University, 1994.

[55] S.-K.Tin, Transversality of double-pulse homoclinic orbits in some atmospheric equations, preprint (1994).

[56] J.Topper & T.Kawahara, Approximate equations for long nonlinear waves on a viscous fluid, *J. Phys. Soc. Japan* 44 (1978) 663-666.

[57] S.Wiggins, *Global Bifurcations and Chaos*, Springer-Verlag, New York, 1988.

[58] S.Wiggins, *Normally Hyperbolic Invariant Manifolds in Dynamical Systems*, Springer-Verlag, New York, 1994.