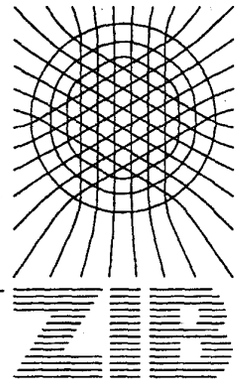
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10, D-1000 Berlin 31



Bernold Fiedler* Jürgen Scheurle **

Discretization of Homoclinic Orbits, Rapid Forcing and "Invisible" Chaos

* Mathematisches Institut A
Universität Stuttgart
Pfaffenwaldring 57
D-7000 Stuttgart 80
Germany

Fellow of the Konrad-Zuse-Zentrum Berlin

** Institut für Angewandte Mathematik
Universität Hamburg
Bundesstraße 55
D-2000 Hamburg 13
Germany

Preprint SC 91-5 (June 1991)

Contents

Abstract

One–step discretizations of order p and step size ϵ of ordinary differential equations can be viewed as time–ϵ maps of

$$\dot{x}(t) = f(\lambda, x(t)) + \epsilon^p g(\epsilon, \lambda, t/\epsilon, x(t)), \quad x \in \mathbf{R}^N, \lambda \in \mathbf{R},$$

where g has period ϵ in t. This is a rapidly forced nonautonomous system.
We study the behavior of a homoclinit orbit Γ for $\epsilon = 0$, $\lambda = 0$, under discretization. Under generic assumptions we show that Γ becomes transverse for positive ϵ. The transversality effects are estimated from above to be exponentially small in ϵ. For example, the length $\ell(\epsilon)$ of the parameter interval of λ for which Γ persists can be estimated by

$$\ell(\epsilon) \leq C \exp(-2\pi\eta/\epsilon),$$

where C, η are positive constants. The coefficient η is related to the minimal distance from the real axis of the poles of $\Gamma(t)$ in the complex time domain.
Likewise, the region where complicated, "chaotic" dynamics prevail is estimated to be exponentially small, provided $x \in \mathbf{R}^2$ and the saddle quantity of the associated equilibrium is nonzero.
Our results are visualized by high precision numerical experiments. The experiments show that, due to exponential smallness, homoclinic transversality becomes practically invisible under normal circumstances, already for only moderately small step sizes.

Key words: Homoclinic orbit, ordinary diffferential equations, discretization, transversality, averaging, exponential smallness, chaos

Subject Classifications (AMS): 34C15, 34C35, 58F14, 65L60

1. Introduction and main results

Numerically speaking, continuous time dynamical systems do not exist. Rather, a discretized version is studied and interpreted in analogy to the continuous time dynamical system. Over fixed finite time intervals, this analogy is quite close and well understood in terms of discretization errors and sophisticated discretization schemes. Over large or infinite time intervals, this analogy is not so clear, because discretization errors tend to accumulate exponentially with time. In this paper, we specifically investigate the correspondence between continuous and discrete time dynamical systems for *homoclinic orbits*. By definition, these are orbits which tend to the same fixed stationary point for both large positive and large negative times.

For illustration, we consider the following example from HOLMES ET AL., §7 (1988).

$$\ddot{\xi} + \lambda\dot{\xi} - \xi/2 + \xi^2 + \xi^3 = 0\,. \tag{1.1}$$

Picking the value $\lambda = 0$ for the real damping parameter λ, this second order ordinary differential equation becomes Hamiltonian; for a phase portrait in $(\xi,\dot{\xi})$–coordinates see Fig. 1.1.

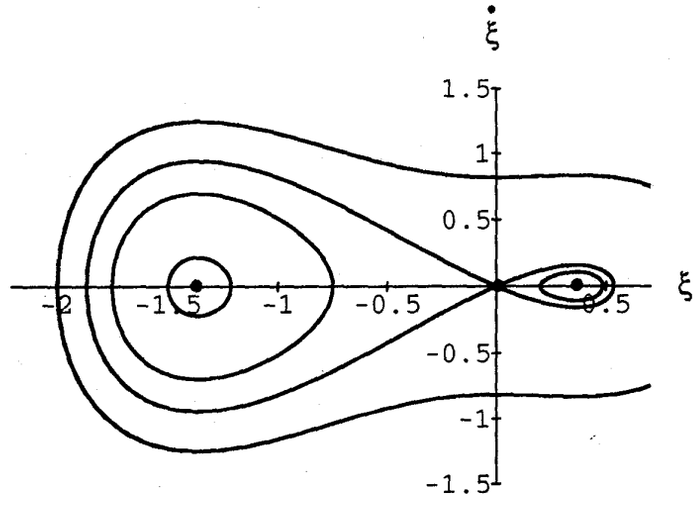


Figure 1.1: Phase portrait of (1.1) for $\lambda = 0$.

The half–plane $\xi > 0$ contains a unique homoclinic orbit $\Gamma(t) = (\xi(t),\dot{\xi}(t))$ given explicitly by

$$\xi(t) = 2e^{t/\sqrt{2}} \Big/ \left(\left(e^{t/\sqrt{2}} + 2/3\right)^2 + 1\right). \tag{1.2}$$

More abstractly, we consider the unstable and stable manifolds (separatrices) W^u_λ and W^s_λ of the equilibrium $\xi = \dot{\xi} = 0$. See CHOW & HALE (1982), HIRSCH ET AL. (1977), PALIS & DE MELO (1982) for a general reference. At $\lambda = 0$ and for $\xi > 0$, these manifolds intersect along the homoclinic orbit Γ. For $\lambda \neq 0$, $\xi > 0$ they miss each other as follows: W^u_λ spirals inward while W^s_λ spirals outward, for positive friction λ. For $\lambda < 0$ the roles of W^u_λ and W^s_λ interchange. In fact, the phase portraits for λ and $-\lambda$ are the same after a reflection through the ξ–axis and reversing the flow direction. See Fig. 1.2. It is now evident how W^u_λ and W^s_λ cross each other as λ increases through $\lambda = 0$, giving rise to a homoclinic orbit precisely at $\lambda = 0$.

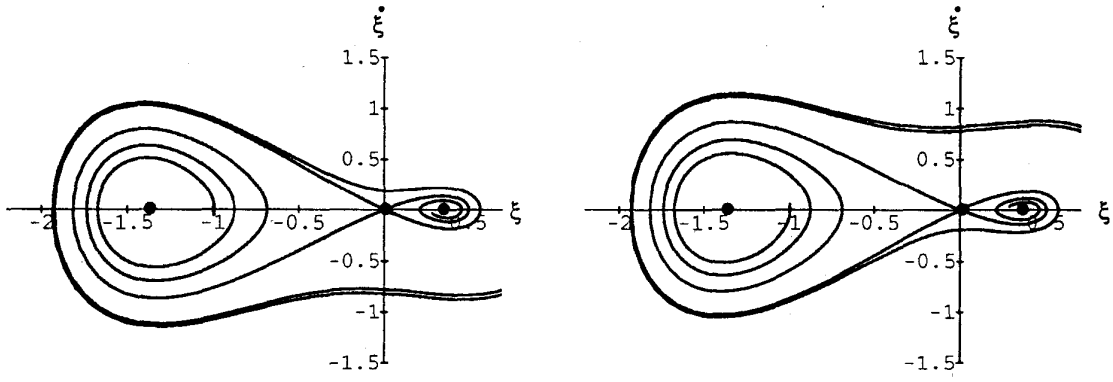


Figure 1.2: Phase portraits of (1.1) for $\lambda \neq 0$.

What happens under discretization? Let us rewrite (1.1) as a first order system

$$\dot{x} = f(\lambda, x) \tag{1.3}$$

for $x = (\xi, \dot{\xi})$. We may discretize, simple-mindedly, by explicit Euler steps of size $\varepsilon > 0$, to obtain the discrete time dynamical system

$$x_{n+1} = \Phi(\varepsilon, \lambda, x_n) := x_n + \varepsilon f(\lambda, x_n), n \in \mathbb{Z} . \tag{1.4}$$

(Note that $\Phi(\varepsilon, \lambda, \cdot)$ is a diffeomorphism, locally, if f is continuously differentiable and $\varepsilon > 0$ is chosen small enough. Therefore, negative n actually make sense.)

Naively, we might expect (1.4) to possess a perturbed homoclinic orbit

$$\Gamma^\varepsilon : \ x_n^\varepsilon = \Gamma_n^\varepsilon \tag{1.5}$$

near Γ for values $\lambda = \lambda(\varepsilon)$ near $\lambda_0 = 0$ such that for $\varepsilon \searrow 0$

$$\begin{aligned} \operatorname{dist}(\Gamma_n^\varepsilon, \Gamma) &= O(\varepsilon) , \\ |\lambda(\varepsilon) - \lambda_0| &= O(\varepsilon) . \end{aligned} \tag{1.6}$$

In fact, let $x(t)$ and x_n denote solutions of (1.3) and (1.4), respectively, with the same initial condition $x(0) = x_0$. Then

$$|x(t) - x_{[t/\varepsilon]}| = O(\varepsilon)$$

holds, uniformly for bounded sets of t and x_0. However, the constant in the $O(\varepsilon)$ term typically grows with $|t|$ like $e^{M|t|}$, for some $M > 0$. Still, (1.6) is more than just a risky conjecture; see for example the results on discretization of homoclinic orbits in BEYN (1990), DOEDEL & FRIEDMAN (1990a), DOEDEL & FRIEDMAN (1990b). Indeed, let $W^u_{\varepsilon,\lambda}$ and $W^s_{\varepsilon,\lambda}$ denote the unstable and stable manifolds, respectively, of the fixed point $x = 0$ of (1.4). It turns out that these manifolds are locally $O(\varepsilon)$ close to their continuous counterparts W^u_λ and W^s_λ. In particular, homoclinic orbits Γ^ε do exist for some λ-value $\lambda = \lambda(\varepsilon)$ near $\lambda_0 = 0$, whenever W^u_λ and W^s_λ cross each other at $\lambda = 0$ as in our example.

But there is a marked difference between system (1.3) and its discretization (1.4): homoclinic points Γ_n^ε of (1.4) can be transverse, which they never are for (1.3). Here we call the homoclinic point $\Gamma_n^\varepsilon \in W^u_{\varepsilon,\lambda} \cap W^s_{\varepsilon,\lambda}$ *transverse*, if the tangent spaces of $W^u_{\varepsilon,\lambda}$ and of $W^s_{\varepsilon,\lambda}$ at Γ_n^ε, together, span the

whole x–space. We then also call the homoclinic orbit Γ^ϵ transverse, since all Γ^ϵ_n are transverse homoclinic if one of them is. Transverse homoclinic orbits are accompanied by shift–type chaotic dynamics in Smale horseshoes. For a modern account of this observation, which originally goes back to Poincaré, Birkhoff, and Smale, see MOSER (1973). Then, for example, for fixed λ, ε and any large positive integer m there also exist homoclinic orbits besides Γ^ϵ, which follow the loop Γ closely for m complete revolutions before settling down into equilibrium. Also, there are other nonhomoclinic orbits which are getting arbitrarily close to *all* of the above homoclinic orbits, as $|n|$ increases.

Now suppose that one of the homoclinic orbits Γ^ϵ in our example actually is transverse, for some ε. Fix such an ε. By the implicit function theorem, or by roughness of transverse intersection under perturbations, it then follows that (transverse) homoclinic points of (1.4) persist for at least some open interval of λ–values. This behavior is in marked contrast to the continuous time limit (1.3), for which the homoclinic orbit Γ disappears as soon as $\lambda \neq 0$. As a main objective of the present paper, we try to estimate the magnitude of the above transversality effects, for ε tending to zero.

We now introduce several geometric splitting quantities in order to measure the magnitude of transversality effects. For later use we describe the necessary construction in complete generality, digressing from the specific example (1.3), (1.4). We consider $\lambda \in [-\lambda^0, \lambda^0]$, $x \in \mathbb{R}^N$ and a sufficiently smooth vector field $f = f(\lambda, x)$ with a nondegenerate homoclinic orbit Γ at $\lambda = 0$, as explained below. Also we discretize the corresponding flow $x(t) = F(t, \lambda, x(0))$ by a sufficiently smooth diffeomorphism $x \mapsto \Phi(\varepsilon, \lambda, x)$, associated to step–size ε, such that for some $p \geq 1$

$$\Phi(\varepsilon, \cdot, \cdot) - F(\varepsilon, \cdot, \cdot) = O(\varepsilon^{p+1})$$

in C^2_{loc} with respect to λ, x. For completely detailed and, in fact, slightly stronger assumptions, we refer to (1.10)–(1.13) below.

Fix a point $\Gamma_0 = \Gamma(0)$ on the homoclinic orbit of f at $\lambda = 0$. Let B denote a closed, local flow box with coordinates $(\vartheta, y) \in \mathbb{R} \times \mathbb{R}^{N-1}$ such that

$$\dot{x} = f(\lambda = 0, x)$$

corresponds to

$$\begin{aligned} \dot{\vartheta} &= 1\,, \\ \dot{y} &= 0 \end{aligned}$$

in B. Further, let B be chosen small enough such that at $\lambda = 0$ the local components of W^u, W^s in B which contain Γ_0, denoted by $W^{u,\text{loc}}$ and $W^{s,\text{loc}}$ respectively, are parametrized over their respective tangent spaces T_u and T_s at Γ_0. Accordingly, we may coordinatize the cross section coordinate $y \in \mathbb{R}^{N-1}$ of the box B by

$$y = (y_u, y_s, y_\perp)\,.$$

For example , $W^{u,\text{loc}}$ is given by those (ϑ, y) for which $y_s = 0$, $y_\perp = 0$. By nondegeneracy of the homoclinic orbit, we may assume $y_\perp \in \mathbb{R}$ is one–dimensional.

For small enough $|\lambda| \leq \lambda^0$, $0 \leq \varepsilon \leq \varepsilon^0$, this description persists, qualitatively, for the continued manifolds $W^{u,\text{loc}}_{\varepsilon,\lambda}$, $W^{s,\text{loc}}_{\varepsilon,\lambda}$. For example, $W^{u,\text{loc}}_{\varepsilon,\lambda} \subseteq B$ is described by those $(\vartheta, y) \in B$ for which y_s, $y_\perp$ are certain continuous, and with respect to $(\lambda, \vartheta, y_u)$ differentiable, functions

$$\begin{aligned} y_s &= y^u_s(\varepsilon, \lambda, \vartheta, y_u)\,, \\ y_\perp &= y^u_\perp(\varepsilon, \lambda, \vartheta, y_u)\,; \end{aligned}$$

analogously for $W^{s,\text{loc}}_{\varepsilon,\lambda}$ with s and u interchanged. This claim will be proved below.

We are now ready to define the *splitting quantities* ℓ, d, ω. Let $I(\varepsilon)$ denote the set of those $\lambda \in [-\lambda^0, \lambda^0]$ for which

$$W^{u,\text{loc}}_{\varepsilon,\lambda} \cap W^{s,\text{loc}}_{\varepsilon,\lambda} \neq \emptyset$$

yields some homoclinic points $(\vartheta, y) \in B$. That is $y = (y_s, y_u, y_\perp)$ with

$$y_s = y_s^u(\varepsilon, \lambda, \vartheta, y_u) , \tag{1.7.a}$$

$$y_u = y_u^s(\varepsilon, \lambda, \vartheta, y_s) , \tag{1.7.b}$$

and

$$y_\perp^u(\varepsilon, \lambda, \vartheta, y_u) - y_\perp^s(\varepsilon, \lambda, \vartheta, y_s) = 0 . \tag{1.7.c}$$

We define the *splitting length* $\ell(\varepsilon)$ by

$$\ell(\varepsilon) := \text{ length of } I(\varepsilon) . \tag{1.8.a}$$

We call $\ell(\varepsilon)$ the length, rather than the diameter of $I(\varepsilon)$, because $I(\varepsilon)$ turns out to be an interval; see Theorem 1.1 below. Note that $I(0) = \{0\}$, $\ell(0) = 0$, by nondegeneracy of the homoclinic orbit. To define a splitting distance $d(\varepsilon)$ we solve (1.7.a,b) for y_s, y_u, that is

$$\begin{aligned} y_s &= y_s(\varepsilon, \lambda, \vartheta) , \\ y_u &= y_u(\varepsilon, \lambda, \vartheta) , \end{aligned}$$

by the implicit function theorem. This is feasible for $0 \le \varepsilon \le \varepsilon^0$, $|\lambda| \le \lambda^0$, $|\vartheta| \le \vartheta^0$ small enough, because at $\varepsilon \equiv 0$

$$y_s^u \equiv 0 \quad \text{and} \quad y_u^s \equiv 0 .$$

We define

$$d(\varepsilon, \lambda, \vartheta) := y_\perp^u(\varepsilon, \lambda, \vartheta, y_u) - y_\perp^s(\varepsilon, \lambda, \vartheta, y_s) \in \mathbb{R} ,$$

plugging in the above expressions for y_s and y_u. Geometrically, $|d(\varepsilon, \lambda, \vartheta)|$ measures the distance between $W_{\varepsilon,\lambda}^{u,\text{loc}}$ and $W_{\varepsilon,\lambda}^{s,\text{loc}}$ in the ϑ–section of the flow box B. Thus d is related to the classical Melnikov function, see CHOW & HALE (1982). For $\lambda \in I(\varepsilon)$ we define the *splitting distance* $d(\varepsilon, \lambda)$ by

$$d(\varepsilon, \lambda) := \max_{|\vartheta| \le \vartheta^0} |d(\varepsilon, \lambda, \vartheta)| . \tag{1.8.b}$$

Finally, we define the *splitting angle* $\omega(\varepsilon, \lambda)$ for $\lambda \in I(\varepsilon)$ at homoclinic points by

$$\omega(\varepsilon, \lambda) := \max_{\vartheta : d(\varepsilon, \lambda, \vartheta) = 0} |d_\vartheta(\varepsilon, \lambda, \vartheta)| . \tag{1.8.c}$$

Here the partial derivative d_ϑ measures the minimal angle between the intersecting manifolds $W_{\varepsilon,\lambda}^{u,\text{loc}}$ and $W_{\varepsilon,\lambda}^{s,\text{loc}}$, up to an almost constant factor. Again, $d = \omega = 0$ at $\varepsilon = \lambda = 0$. For later use we also introduce the *splitting slope*

$$\overline{\omega}(\varepsilon, \lambda) := \max_{\vartheta} |d_\vartheta(\varepsilon, \lambda, \vartheta)| , \tag{1.8.d}$$

where the maximum is taken over all ϑ, not necessarily just homoclinic ones.

The above definitions of the splitting quantities depend on the choice of $\Gamma(0)$ on the homoclinic orbit Γ, on the size of the flow box B, and on the particular coordinates used. Fortunately, it will turn out below that the asymptotic behavior of the splitting quantities $\ell(\varepsilon)$, $d(\varepsilon, \cdot)$, $\omega(\varepsilon, \cdot)$ $\overline{\omega}(\varepsilon, \cdot)$ for $\varepsilon \searrow 0$ is essentially independent of these somewhat arbitrary choices.

To justify definitions (1.8.a–c) of the splitting quantities, we will now prove that the invariant manifolds $W_{\varepsilon,\lambda}^{s,\text{loc}}$, $W_{\varepsilon,\lambda}^{u,\text{loc}}$ are C^2–close to their counterparts $W^{s,\text{loc}}$, $W^{u,\text{loc}}$ at $\varepsilon = \lambda = 0$. For more refined statements see FONTICH & SIMÓ (1990b). The invariant manifolds are usually constructed by a contraction mapping or an implicit function argument, see e.g. CHOW & HALE (1982), HIRSCH (1976), PALIS JR. & DE MELO (1982). The parameters $\varepsilon > 0$ and λ enter differentiably, via $\Phi(\varepsilon, \lambda, x)$. Hence differentiable dependence on $\varepsilon > 0$ and λ does not present any difficulty. Only

the case $\varepsilon = 0$ requires a little extra attention. Consider the local flow $F(t, \lambda, x)$ of $\dot{x} = f(\lambda, x)$, inducing local diffeomorphisms

$$x \mapsto F(t, \lambda, x) \ , \ 0 < a \leq t \leq b \ ,$$

with fixed parameters t, λ and some $0 < a < b$. The associated manifolds $W^{u,\mathrm{loc}}_{\varepsilon=0,\lambda}$ of these diffeomorphisms are independent of the choice of t, of course. Also consider manifolds $W^{u,\mathrm{loc}}_{\Psi}$ associated to diffeomorphisms of Ψ in a C^2–neighborhood of the set of diffeomorphisms $F(t, \lambda, \cdot)$. The manifolds $W^{u,\mathrm{loc}}_{\Psi}$ then depend continuously on Ψ in the C^2–topology. But $\Phi(\varepsilon, \lambda, \cdot)$ is assumed $O(\varepsilon^{p+1})$–close to $F(\varepsilon, \lambda, \cdot)$ in C^2. For $0 < \varepsilon < b - a$, we now choose integers n such that

$$a \leq n\varepsilon \leq b \ .$$

Then the n–th iterate $\Psi := \Phi(\varepsilon, \lambda, \cdot)^n$ possesses the same unstable manifold $W^{u,\mathrm{loc}}_{\varepsilon,\lambda}$ as $\Phi(\varepsilon, \lambda, \cdot)$ itself, and is C^2–close of order $O(\varepsilon^p)$ to $F(t, \lambda, \cdot)$ with $t = n\varepsilon$, since $p \geq 1$. For the latter statement, see any book on the numerical analysis of ordinary differential equations, e.g. HAIRER, ET AL. (1987). Consequently the associated unstable manifolds are also C^2–close. To the stable manifolds, the same argument applies. This completes our definition of the splitting quantities ℓ, d, and ω.

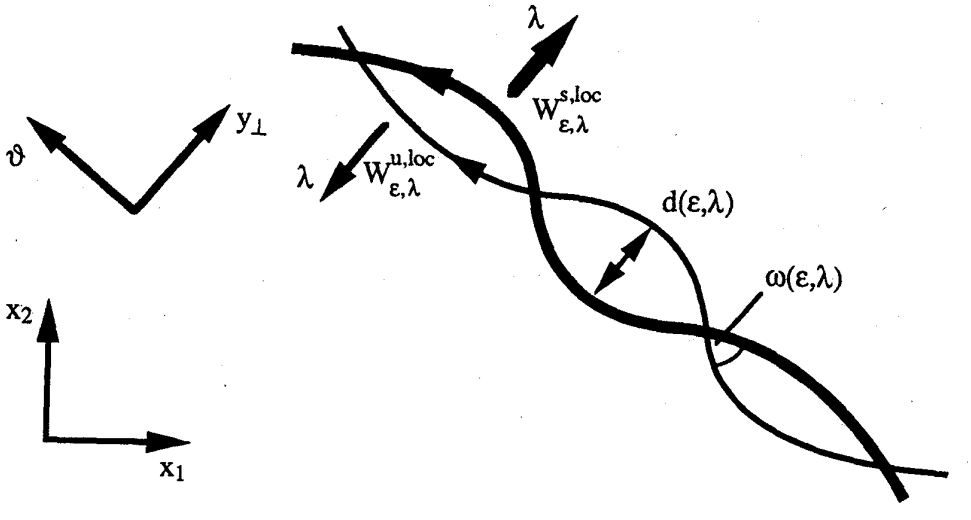


Figure 1.3: Transverse homoclinic points, splitting distance $d(\varepsilon, \lambda)$, and splitting angle $\omega(\varepsilon, \lambda)$.

In Fig. 1.3 we visualize the splitting quantities d and ω for the planar case. Here $y = y_\perp$ is one–dimensional and $y_s = y_u = 0$ can be omitted. The two arrows indicate the direction of displacement by increasing λ.

Note that the splitting length $\ell(\varepsilon)$ yields an upper estimate of $d(\varepsilon, \lambda)$, $\overline{\omega}(\varepsilon, \lambda)$, up to a constant C,

$$d(\varepsilon, \lambda) \leq C\ell(\varepsilon) \ , \tag{1.9.a}$$

$$\omega(\varepsilon, \lambda)^2 \leq \overline{\omega}(\varepsilon, \lambda)^2 \leq C\ell(\varepsilon) \ . \tag{1.9.b}$$

Indeed, (1.9.a) holds, since the speed of displacement with respect to λ is finite. (1.9.a) implies (1.9.b), since the C^2–norm of $\vartheta \mapsto d(\varepsilon, \lambda, \vartheta)$ is bounded, uniformly with respect to ε, λ (and in fact

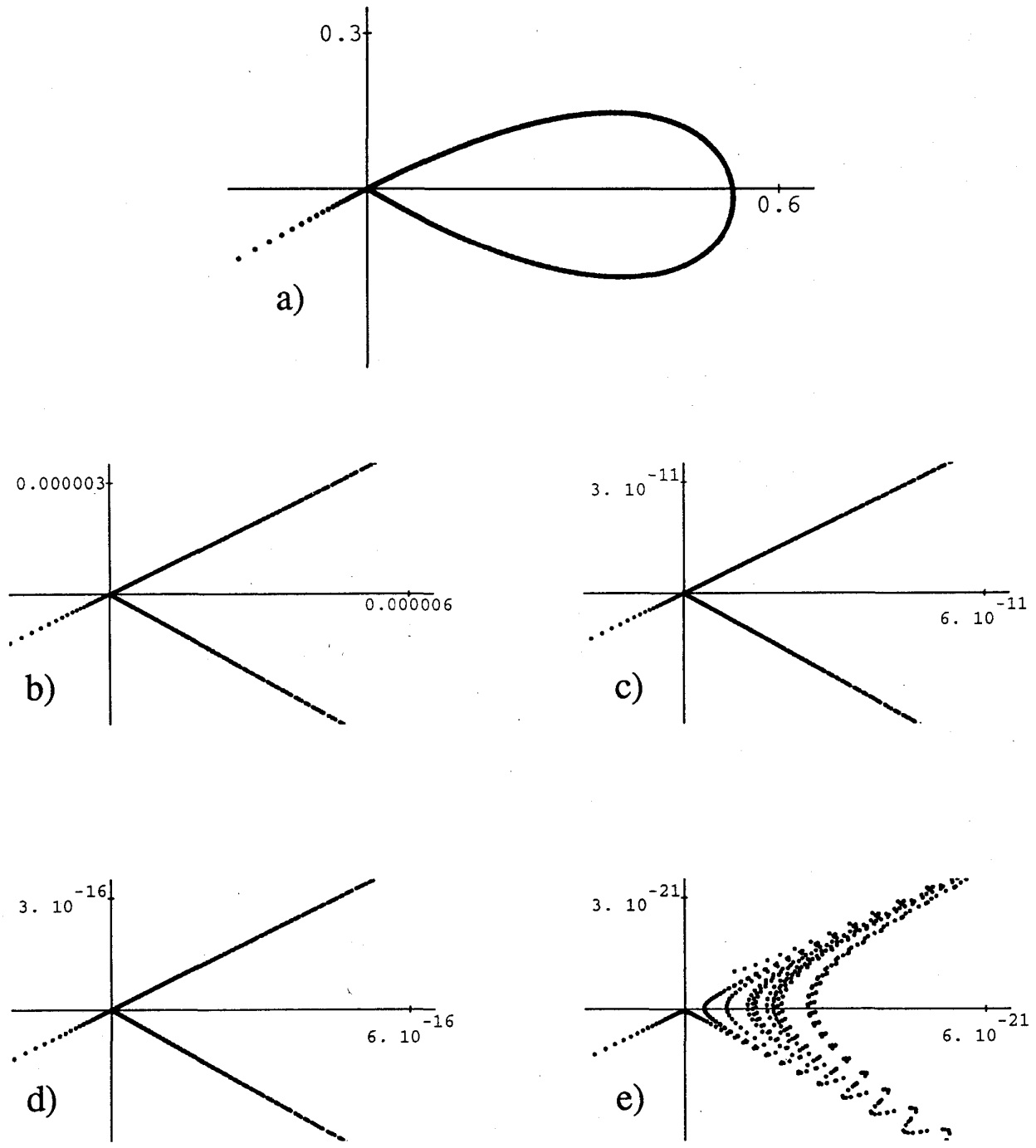


Figure 1.4: **a)** Euler's method, applied to (1.1); see text. **b–e)** Successive enlargements by a factor of 10^5.

of order $O(\varepsilon^p)$ for $\varepsilon \searrow 0$). Moreover (1.9.a,b) remain true, even for the general case $x \in \mathbb{R}^N$. Our estimates on $\overline{\omega}$, d, ℓ given in Theorem 1.1 below will of course be compatible with (1.9.a) but will in fact improve much on (1.9.b).

But what *is* the asymptotic behavior of the splitting quantities ℓ, d, ω for $\varepsilon \searrow 0$? Before stating precise results (see Theorems 1.1–1.3 below), we present some numerical evidence. In Fig. 1.4.a, we show a run of Euler's method (1.4) for equation (1.3), (1.1) with

$$\begin{aligned} \text{step size } \varepsilon &= 0.2\,, \\ \text{parameter } \lambda &= 0.090\ 164\ 641\ 779\ 388 \\ &\quad\ 258\ 080\ 324\ 281\ 175 \\ &\quad\ 394\ 773\ 960\ 000\ \ldots \end{aligned}$$

and initial condition

$$\begin{aligned} \xi = x^1 &= 10^{-21} \\ \xi = x^2 &= 10^{-21} \end{aligned}$$

near the unstable manifold $W^u_{\varepsilon,\lambda}$ near $x = 0$.

Fig. 1.4.b–e show successive enlargements of a window around $x = 0$ by a factor of 10^5. Figs. 1.4.a–d seem to indicate a perfect homoclinic orbit without any visible transversality effects. Only Fig. 1.4.e, with a window diameter of the order

$$10^{-20}$$

reveals oscillations in the trajectory which can be interpreted as tracing oscillations in $W^u_{\varepsilon,\lambda}$ due to transverse intersections with $W^s_{\varepsilon,\lambda}$. In fact, the trajectory escapes from a neighborhood of the homoclinic orbit Γ after seventeen cycles and more than

$$n = 13000$$

iterates. For more detailed numerical experiments see §6. All simulations were performed with *Mathematica* on a Mac II, in Fig. 1.4 with a precision of 120 floating decimal points to control round–off error. Fig. 1.4 clearly demonstrates that the minute chaotic effects due to discretization easily escape attention, even at moderately small step sizes. Therefore, we call such an effect "invisible chaos". Our main results, Theorems 1.1–1.3 below, indicate that the splitting quantities ℓ, d, $\overline{\omega}$ as well as the region of chaotic effects, while generically existent, are exponentially small of order

$$O(e^{-2\pi\eta/\varepsilon})$$

for some $\eta > 0$. In particular, chaoticity is small of infinite order in ε, even for a discretization of only first order. The apparent "absence" of chaoticity was noticed for example in BEYN (1987).

Before we state our main results, we fix the precise setting and assumptions for the remainder of this paper. We consider complex vector fields

$$\dot{x} = f(\lambda, x)\,, \quad |\lambda| \le \lambda^0\,, \quad \lambda \in \mathbb{C}\,, \quad x \in \mathbb{C}^N \tag{1.10}$$

and assume

$$f \text{ is complex analytic with respect to } \lambda \text{ and } x\,. \tag{1.11.a}$$

Moreover, f can be viewed as the complex extension of a parametrized real vector field, that is

$$f(\lambda, x) \in \mathbb{R}^N \text{ for } \lambda \in [-\lambda^0, \lambda^0]\,, \quad x \in \mathbb{R}^N\,.$$

We also assume that f has a real nondegenerate homoclinic orbit Γ to the equilibrium $x = 0$ at $\lambda = \lambda_0 = 0$. In detail:

$$f(0,0) = 0, \text{ and } D_x f(0,0) \text{ does not possess eigenvalues on the imaginary axis ;} \tag{1.11.b}$$

$$\text{for } x = 0, \text{ there exists a real solution } x(t) = \Gamma(t) \text{ of (1.10) which is homoclinic to } x = 0, \text{ that is } \Gamma(t) \not\equiv 0 \text{ and } \lim_{t \to \pm\infty} \Gamma(t) = 0 . \tag{1.11.c}$$

To formulate the nondegeneracy condition on Γ, we consider the linearization of (1.10) along Γ as an operator

$$\begin{aligned} \mathcal{L} : C^1(\mathbb{R}, \mathbb{R}^N) &\to C^0(\mathbb{R}, \mathbb{R}^N) , \\ x(\cdot) &\mapsto \dot{x}(\cdot) - Ax(\cdot) . \end{aligned}$$

Here C^k denotes the space of functions with uniformly bounded derivatives up to order k, endowed with the usual sup-norm. The matrix function $A = A(t)$ is defined as the linearization of f along the homoclinic orbit:

$$A(t) := D_x f(\lambda, x) \text{ at } \lambda = 0 , \quad x = \Gamma(t)$$

It is well known that $\mathcal{L}$ is a Fredholm operator of Fredholm index zero by (1.11.b.c), see the proof of PALMER (1984), Lemma 4.2. We assume

$$C^0(\mathbb{R}, \mathbb{R}^N) = \text{range}(\mathcal{L}) \oplus \text{span}(f_\lambda(0, \Gamma(\cdot)) . \tag{1.11.d}$$

Geometrically, this condition means that at some (and hence, at any) point x of Γ the tangent spaces of the stable and unstable manifold intersect only along $\dot{\Gamma}$,

$$T_x W^s_{\lambda=0} \cap T_x W^u_{\lambda=0} = \text{span}(\dot{\Gamma}) .$$

Moreover, as in Fig. 1.2, these manifolds cross each other with nonvanishing speed as λ increases through zero. See Section 3 for a more detailed analysis.

We now fix our assumptions on the p-th order discretized system

$$x_{n+1} = \Phi(\varepsilon, \lambda, x_n) , \tag{1.12}$$

where $0 \le \varepsilon \le \varepsilon^0$ is real, $|\lambda| \le \lambda^0$ is complex, $x \in \mathbb{C}^N$, and n is integer. We assume

$$\Phi \text{ is, jointly, real analytic in } 0 \le \varepsilon \le \varepsilon^0 \text{ and complex analytic with respect to } (\lambda, x) . \tag{1.13.a}$$

Denoting the (local) flow $x(t)$ of (1.10) by

$$x(t) = F(t, \lambda, x(0)) ,$$

we also assume the p-th order estimate

$$\text{there exists a continuous, increasing function } C : [0, \infty) \to [0, \infty) \text{ such that}$$

$$|\Phi(\varepsilon, \lambda, x) - F(\varepsilon, \lambda, x)| \le C(|x|)\varepsilon^{p+1} \tag{1.13.b}$$

$$\text{holds, for some } p \ge 1, \text{ and for all } \varepsilon, \ \lambda, \ x \text{ for which the left-hand side is defined.}$$

Without loss of generality, we may assume (1.13.b) to hold for all $|x| < \rho_\varepsilon$ and some radii ρ_ε, increasing with $\varepsilon \searrow 0$ up to $\rho_0 = \infty$. Note that (1.13.b) implies

$$\Phi(\varepsilon = 0,\ \lambda,\ x) = x$$

and

$$\Phi_\varepsilon(\varepsilon = 0,\ \lambda,\ x) = f(\lambda, x)\ .$$

Indeed, Φ describes a general numerical one–step discretization scheme of order p and fixed step–size ε. For example, $p = 1$ for Euler's method (1.4), and $p = 4$ for the classical Runge–Kutta scheme. Also note that (1.13.b) implies locally uniform C^k–convergence with respect to (λ, x), for any finite k by analyticity.

As a final preparation to Theorem 1.1, we consider the analytic extension of the homoclinic orbit $\Gamma(t)$ to complex time t. By ordinary (real) local existence theory, such an extension is possible along complex C^1–arcs, locally, as long as the complex solutions remain bounded. In simply connected domains this extension is defined consistently, by Cauchy's theorem. The positive radius of convergence of the associated power series expansion centered at some real $t = t_0$ may be finite, but tends to infinity as $t_0 \to \pm\infty$. Indeed, $\Gamma(t_0) \to 0$ for $t_0 \to \pm\infty$ and $f(\lambda = 0,\ x = 0) = 0$, so that a complex extension becomes possible over ever larger complex distances. We now pick and fix any positive real η such that

$$t \mapsto \Gamma(t)\ ,\ \ t \in \mathbb{R}\ , \tag{1.14}$$

possesses a complex analytic extension to the closed strip $\Sigma_\eta := \{t \in \mathbb{C} \mid |\mathrm{Im}\ t| \le \eta\}$.

Theorem 1.1 *Let assumptions* (1.11.a–d), (1.13.a,b), (1.14) *hold. In a fixed box B consider the splitting length $\ell(\varepsilon)$ of the homoclinicity set $I(\varepsilon)$ and, for $\lambda \in I(\varepsilon)$, also the splitting distance $d(\varepsilon, \lambda)$, the splitting angle $\omega(\varepsilon, \lambda)$ and the splitting slope $\overline{\omega}(\varepsilon, \lambda)$ as defined in* (1.8.a–d).

Then there exists an $\varepsilon^0 > 0$ and a constant $C > 0$ such that for all real $0 \le \varepsilon \le \varepsilon^0$, $\lambda \in I(\varepsilon)$ the following holds. The set $I(\varepsilon)$ is an interval;

$$0 \le \ell(\varepsilon) \le C \exp(-2\pi\eta/\varepsilon)\ ; \tag{1.15.a}$$

$$0 \le d(\varepsilon, \lambda) \le C \exp(-2\pi\eta/\varepsilon)\ ; \tag{1.15.b}$$

$$0 \le \omega(\varepsilon, \lambda) \le \overline{\omega}(\varepsilon, \lambda) \le C \exp(-2\pi\eta/\varepsilon)\ . \tag{1.15.c}$$

Moreover, there exists a continuous curve $\lambda = \lambda(\varepsilon)$, $0 \le \varepsilon \le \varepsilon^0$, real analytic for positive ε, such that for all ε

$$\lambda(\varepsilon) \in I(\varepsilon) \tag{1.16.a}$$

indicates a homoclinic orbit of $\Phi(\varepsilon, \lambda, \cdot)$, and

$$|\lambda(\varepsilon) - \lambda_0| \le C\varepsilon^p\ . \tag{1.16.b}$$

Naively, the p–th order curve $\lambda = \lambda(\varepsilon)$ with homoclinic orbits could have been expected for a p–th order discretization Φ of f. In addition, we see that certain homoclinic orbits, namely those represented by our construction of $W^{u,\mathrm{loc}}_{\varepsilon,\lambda}$, $W^{s,\mathrm{loc}}_{\varepsilon,\lambda}$ within the flow box B, are confined to occur only in an exponentially thin wedge around that curve. Their splitting quantities are therefore practically "invisible".

We have not claimed, so far, that any splitting does occur at all. In other words, is $\ell(\varepsilon)$ at all positive? Our answer, given in Theorem 1.2 below, states that indeed $\ell(\varepsilon) > 0$ for all $\varepsilon > 0$ provided f is "generic".

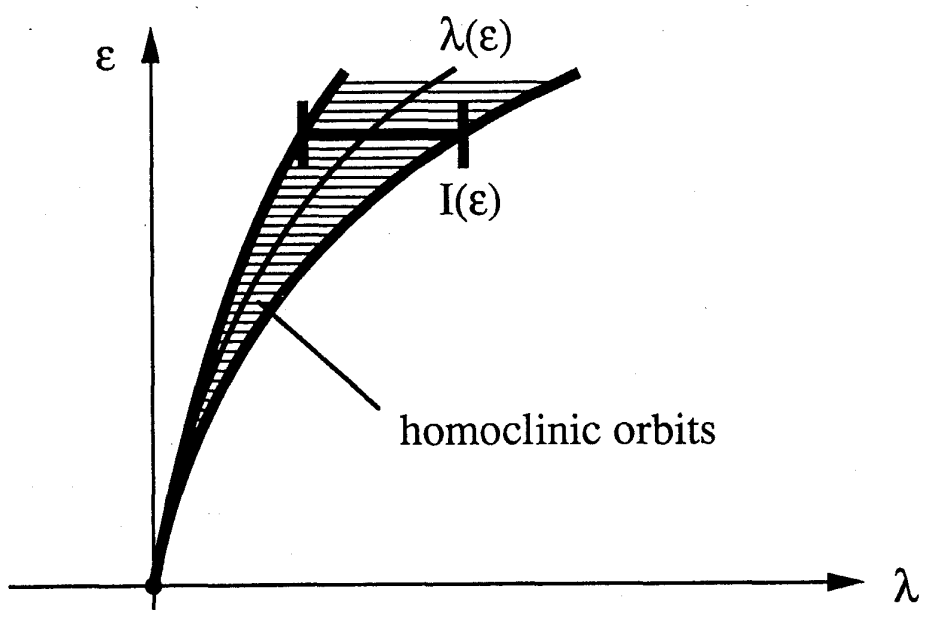


Figure 1.5: Exponentially thin tongue of homoclinic orbits.

Let us fix our notion of genericity, first. We define:

$$\begin{aligned} &C^\omega \text{ is the space of vector fields } f(\lambda, x) \text{ which are complex analytic} \\ &\text{with respect to } |\lambda| \leq \lambda^0 \text{ and } x \in \mathbb{C}^N . \end{aligned} \tag{1.17}$$

In particular, we may write $f \in C^\omega$ as a power series

$$f(\lambda, x) = \sum_k a_k \, (\lambda, x)^k$$

with multi-index $k = (k_0, \ldots, k_n)$ and

$$(\lambda, x)^k = \lambda^{k_0} \cdot x_1^{k_1} \cdot \ldots \cdot x_N^{k_N} .$$

We define a (strong Whitney type) topology on C^ω, which is generated by the following open sets $U_\delta(f)$, $\delta = (\delta_k)_k$ with $\delta_k > 0$ for all k. An element $\widetilde{f} = \sum_k \widetilde{a}_k (\lambda, x)^k$ of C^ω lies in $U_\delta(f)$, if and only if

$$|\widetilde{a}_k - a_k| < \delta_k \tag{1.18}$$

holds for all multi-indices k. With this topology, C^ω becomes a *Baire space*, that is, countable intersections of open and dense sets are still dense, see HIRSCH (1976). We call such countable intersections *residual* sets. We call a property *generic*, if it holds for all elements f of a residual set. We then also call the elements f "generic". For example, it can be proved that for generic one-parameter vector fields f, our nondegeneracy assumptions (1.11.b,d) hold for any orbit, homoclinic to an equilibrium which is not a saddle-node. This follows along the fairly standard lines of the proof of the well-known Kupka-Smale Theorem, see e.g. ABRAHAM & ROBBIN (1967) although the technical setting differs slightly. We do not intend to pursue this direction, here.

As we have mentioned above, Theorem 1.2 below claims splitting quantities to be positive, for generic vector fields. One counterexample is obvious: let

$$\Phi(\varepsilon, \lambda, x) := F(\varepsilon, \lambda, x) ,$$

where $F(t, \lambda, x)$ is the flow f itself. Then Φ does not possess transverse homoclinic orbits, being exactly the time–ε map of a continuous time flow and therefore its best possible "discretization". Since we intend to make a genericity assumption only on f, but not on Φ, we have to be able to infer some genericity properties of Φ from those of f. Indeed Φ is determined by f, when we think of discretization schemes. Writing $\Phi = \Phi(f; \varepsilon, \lambda, x)$ to emphasize this dependence, we assume

$$\begin{array}{rcl} \Phi^\varepsilon : C^\omega & \to & C^\omega \\ f & \mapsto & \Phi(f; \varepsilon, \cdot, \cdot) \end{array} \tag{1.19}$$

maps arbitrarily small neighborhoods of f onto neighborhoods of

$\Phi(f; \varepsilon, \cdot, \cdot)$, for any f and any fixed positive real $\varepsilon \leq \varepsilon^0(f)$.

Clearly, this assumption holds for the explicit Euler scheme

$$\Phi(f; \varepsilon, \lambda, x) = x + \varepsilon f(\lambda, x) .$$

Checking assumption (1.19) for implicit schemes or higher order schemes, however, is quite involved. A working alternative, in these cases, will be sketched in our discussion in Section 7.

Theorem 1.2 *Let f_0, Φ_0 satisfy the assumptions of Theorem* 1.1 *and relation* (1.19). *Fix a small enough neighborhood $\mathcal{V}$ of f_0 in C^ω and some $\varepsilon^0 > 0$ small enough. Then, for f in some residual subset of $\mathcal{V}$* (cf. (1.17), (1.18)), *that is, for generic f in $\mathcal{V}$, we have*

$$\ell(\varepsilon) > 0 , \quad \textit{for all } \varepsilon \in (0, \varepsilon^0) . \tag{1.20.a}$$

Furthermore, for any $\varepsilon \in (0, \varepsilon^0)$

$$d(\varepsilon, \lambda(\varepsilon)) > 0 , \tag{1.20.b}$$

where $\lambda(\varepsilon)$ is constructed as in Theorem 1.1. *In addition, for some $\lambda \in I(\varepsilon)$,*

$$\omega(\varepsilon, \lambda) > 0 . \tag{1.20.c}$$

Note that our lower estimate is not quantitative. But the splitting effects due to discretization actually are present, at least generically. Concerning the difficulties with sharp upper and lower estimates in general, cf. SCHEURLE ET AL. (1991). We now return to upper estimates of the splitting effects. In Theorem 1.1, we have traced homoclinic points by continuation of the local unstable and stable manifolds of $x \equiv 0$. Transversality of these homoclinic points leads to Smale horseshoes, nearby, and to very complicated recurrent dynamics, as we have argued above. It is not at all clear, how far this more complicated behavior extends in phase space $x \in R^N$ or how far it persists in parameter space $\lambda \in \mathbb{R}$, depending on $\varepsilon > 0$. In particular, we mention a subtle phenomenon discovered by GREBOGI ET AL. (1987) in a study of *metamorphoses* of basin boundaries of attractors: as λ increases, a first homoclinic tangency occurs, say at $\lambda = \lambda^*$. For small $\lambda - \lambda^* > 0$, homoclinic points are also found at a distance $\delta > 0$ from the original homoclinic tangency. But this distance δ does not shrink to zero, for $\lambda \searrow \lambda^*$. Rather, δ remains uniformly positive. Theorem 1.3 below implies that all these effects, although generically present are "invisible" (that is, confined to an exponentially thin wedge in parameter space (ε, λ)), under a non–resonance condition.

For simplicity of presentation and concepts, we only consider planar systems $x \in \mathbb{R}^2$. Under the hyperbolicity assumption (1.11.b) of Theorem 1.1, the equilibrium $x = 0$ of f at $\lambda = 0$ then has two simple eigenvalues $-\mu < 0 < \nu$. We impose the *non–resonance condition*

$$\mu \neq \nu . \tag{1.21}$$

Notably, this condition excludes the planar Hamiltonian case.

Theorem 1.3 *Let the assumptions of Theorem* 1.1 *and the non-resonance condition* (1.21) *hold with* $\mu > \nu > 0$. *Then there exist a neighborhood* U *of the homoclinic orbit* $\Gamma \cup \{0\}$, *some* $\varepsilon_0, \varepsilon_1$ *positive, and a constant* $C' > C > 0$, *such that, in the notation of Theorem* 1.1, *the following holds for all real* $0 < \varepsilon \leq \varepsilon_0$, $\lambda \in [-\varepsilon_1, \varepsilon_1]$.

The only orbits of iteration (1.12) *which remain in* U *for all (positive and negative) integers* n *are*

(i) *a hyperbolic equilibrium, and*
(ii) *possibly a unique attracting, invariant, closed Lipschitz curve* γ,

provided

$$|\lambda - \lambda(\varepsilon)| \geq C' \exp(-2\pi\eta/\varepsilon) \,. \tag{1.22}$$

In the opposite case, $\nu > \mu > 0$, *the same statements remain true, replacing "attracting" by "repelling" in (ii).*

We postpone a detailed discussion of our assumptions, of variants and generalizations of our Theorems, of related earlier work, and of open problems to Section 7. In Section 6 we make the exponential smallness results visible by numerical simulations of high machine precision. Theorem 1.k, $k = 1, 2, 3$ is proved in Section $k + 2$. In Section 2, as a prelude, we construct a non–autonomous vector field

$$\dot{x}(t) = f(\lambda, x) + \varepsilon^p g(\varepsilon, \lambda, t/\varepsilon, x) \,, \tag{1.23}$$

periodic of period 1 in the forcing variable $\tau = t/\varepsilon$, such that the time–ε map of (1.23) coincides with

$$\Phi(\varepsilon, \lambda, x) \,.$$

The exponential smallness claims of HOLMES ET AL., §5, Remark 2 (1988) then essentially apply to (1.23), yielding Theorem 1.1. Still we give an independent proof in Section 3.

2. Discretization and rapid forcing

In this section we only assume that the vector field f of

$$\dot{x}(t) = f(\lambda, x(t)), \ |\lambda| \leq \lambda^0, \ \lambda \in \mathbb{C}, \ x \in \mathbb{C}^N \tag{2.1}$$

and its p–th order discretization of step–size ε

$$x_{n+1} = \Phi(\varepsilon, \lambda, x_n), \ |\varepsilon| \leq \varepsilon^0, \ \varepsilon \in \mathbb{C}, \ |\lambda| \leq \lambda^0, \ \lambda \in \mathbb{C}, \ x \in \mathbb{C}^N \tag{2.2}$$

satisfy assumptions (1.11.a) and (1.13.a,b). We prove in Proposition 2.1 that any p–th order discretization Φ can equivalently be viewed as the time–ε period map of a suitable ε–periodic non–autonomous perturbation

$$\dot{x}(t) = f(\lambda, x(t)) + \varepsilon^p g(\varepsilon, \lambda, t/\varepsilon, x(t)) \, . \tag{2.3}$$

This fact will be basic to our proof of Theorem 1.1 given in Section 3 below. As in Section 1, let $x(t) = F(t, \lambda, x(0))$ denote the (local) flow of (2.1). Let

$$x(t) = G(t, s, \varepsilon, \lambda, x(s))$$

denote the (local) evolution of (2.3).

Proposition 2.1 *Let* (1.11.a) *and* (1.13.a,b) *hold for f and Φ. Then there exists $\varepsilon_0 > 0$, a non–autonomous vector field*

$$g = g(\varepsilon, \lambda, \tau, x) \, ,$$

and a continuous nonincreasing function

$$\rho : [0, \varepsilon_0] \to [0, \infty]$$

with $\rho(0) = \infty$, such that the following holds:

- *(i)* *$g(\varepsilon, \lambda, \tau, x) \in \mathbb{C}^N$ is defined for all real τ, $0 \leq \varepsilon \leq \varepsilon_0$, and all complex λ, x with $|\lambda| \leq \lambda^0$, $|x| < \rho(\varepsilon)$;*
- *(ii)* *g is C^∞–smooth in all variables, and g and all its τ–derivatives are analytic in $(\varepsilon, \lambda, x)$;*
- *(iii)* *g has period 1 in τ;*
- *(iv)* *$G(\varepsilon, 0, \varepsilon, \lambda, x) = \Phi(\varepsilon, \lambda, x)$.*

Statements (ii)–(iv) hold for all ε, λ, τ, x satisfying (i).

Proof. The idea is simple. Omitting argumets ε, λ, x, for a moment, we interpolate between *id* and Φ by a curve $G(t, 0)$, $0 \leq t \leq \varepsilon$, in the space of diffeomorphisms. To define an evolution, we put

$$G(t, s) := G(t, 0) \circ G(s, 0)^{-1} \, ,$$

and extend to $t \geq \varepsilon$ by iteration.

The details are a little more technical. In Step 1, we interpolate between the points y and $\Phi(\varepsilon, \lambda, y)$ by a C^∞–curve $t \mapsto G(t, 0, \varepsilon, \lambda, y)$, $t \geq 0$, such that

$$y = G(0, 0, \varepsilon, \lambda, y) \, , \tag{2.4.a}$$

$$\Phi(\varepsilon, \lambda, y) = G(\varepsilon, 0, \varepsilon, \lambda, y) \, . \tag{2.4.b}$$

In Step 2, we define $f + \varepsilon^p g$ as the time–t derivative of the curve G. More precisely, we let

$$\varepsilon^p g(\varepsilon, \lambda, t/\varepsilon, x) := -f(\lambda, x) + D_t G(t, 0, \varepsilon, \lambda, y) \,, \tag{2.5.a}$$

where D_t denotes the partial derivative with respect to the first entry and $y = y(t/\varepsilon, \varepsilon, \lambda, x)$ is determined implicitly by

$$x = G(t, 0, \varepsilon, \lambda, y) \,. \tag{2.5.b}$$

By construction (2.4.a,b), property (iv) then holds. The remaining properties (i–iii) of g are established in step 3. For simplicity of notations we will suppress the variable λ in f, F, G below, and also the initial time $s = 0$ in $G(t, s, \varepsilon, \lambda, y)$. For example,

$$G = G(t, \varepsilon, y) \,.$$

Step 1: Construction of the interpolant G.

We define the curve $G(\cdot, \varepsilon, y)$ by an interpolation of Hermite type. Let

$$\chi_0 : \mathbb{R} \to [0, 1]$$

be a C^∞ cut-off function such that

$$\begin{aligned} &\chi_0(\tau) \equiv 1 \text{ for } \tau \le 0 \,, \\ &\chi_0(\tau) \equiv 0 \text{ for } \tau \ge 1 \,, \end{aligned} \tag{2.6.a}$$

$$\chi_0 \text{ is real analytic for } \tau \neq 0, 1 \,, \tag{2.6.b}$$

and denote $\chi_1(\tau) := 1 - \chi_0(\tau)$. For example, we could take

$$\chi_0(\tau) = (1 + \tanh(\cot(\pi\tau)))/2 \,, \text{ for } 0 < \tau < 1 \,.$$

For $0 \le t \le \varepsilon$ and $y \in \mathbb{C}^N$ with $|y| < \rho(\varepsilon)$ we define

$$G(t, \varepsilon, y) := \chi_0(t/\varepsilon) F(t, y) + \chi_1(t/\varepsilon) F(t - \varepsilon, \Phi(\varepsilon, y)) \,. \tag{2.7.a}$$

The restriction $|y| < \rho(\varepsilon)$, as in the statement of the proposition becomes necessary because F is only a local flow. Note that (2.4.a,b) hold with this definition, that is

$$\begin{aligned} &G(0, \varepsilon, y) = F(0, y) = y \,, \text{ and} \\ &G(\varepsilon, \varepsilon, y) = F(0, \Phi(\varepsilon, y)) = \Phi(\varepsilon, y) \,. \end{aligned}$$

Let $[\tau]$ denote the largest integer not exceeding $\tau \in \mathbb{R}$, and let $\Phi^n(\varepsilon, \cdot)$ denote the n–th iterate of $\Phi(\varepsilon, \cdot)$, $n \ge 1$. Then we may extend our definition of G to all $t \ge 0$, putting

$$G(t, \varepsilon, y) := G(t - [t/\varepsilon]\varepsilon, \varepsilon, \Phi^{[t/\varepsilon]}(\varepsilon, y)) \,. \tag{2.7.b}$$

Obviously this definition implies

$$G(t, \varepsilon, G(n\varepsilon, \varepsilon, y)) = G(t + n\varepsilon, \varepsilon, y) \,, \text{ for all } t \ge 0 \,,\ n \in \mathbb{N} \,,\ \varepsilon, y \,, \tag{2.8}$$

as is appropriate for the evolution of an ε–periodic, non–autonomous system.

We claim that the curve

$$t \mapsto G(t, \varepsilon, y) \,, \ \ t \ge 0 \,,$$

is C^∞, and thus represents a C^∞–interpolation of the orbit

$$\Phi^n(\varepsilon, y) = G(n\varepsilon, \varepsilon, y) \,, \ \ n \in \mathbb{N} \,.$$

Indeed, even analyticity is obvious for noninteger t/ε. By recursion (2.8), it is sufficient to prove C^∞–regularity at $t = \varepsilon$. Let $D_\pm^m$ denote the m–th t–derivative from above (+), respectively from below (−). Then (2.6.a,b), (2.7.a,b) imply

$$\begin{aligned} D_+^m \mid_{t=\varepsilon} G(t,\varepsilon,y) &= D_+^m \mid_{t=0} G(t,\varepsilon,\Phi(\varepsilon,y)) \\ &= D_+^m \mid_{t=0} (\chi_0(t/\varepsilon)F(t,\Phi(\varepsilon,y)) + \chi_1(t/\varepsilon)F(t-\varepsilon,\Phi^2(\varepsilon,y))) \\ &= D_+^m \mid_{t=0} F(t,\Phi(\varepsilon,y)) = D_-^m \mid_{t=0} F(t,\Phi(\varepsilon,y)) \\ &= D_-^m \mid_{t=\varepsilon} (\chi_0(t/\varepsilon)F(t,y) + \chi_1(t/\varepsilon)F(t-\varepsilon,\Phi(\varepsilon,y))) \\ &= D_-^m \mid_{t=\varepsilon} G(t,\varepsilon,y)\,, \end{aligned}$$

proving C^∞–regularity.

Step 2: Construction of the perturbation g.

To construct g, we switch to the scaled time variable $\tau = t/\varepsilon$. Let

$$\tilde{G}(\tau,\varepsilon,y) := G(\varepsilon\tau,\varepsilon,y) = \chi_0(\tau)F(\varepsilon\tau,y) + \chi_1(\tau)F(\varepsilon\tau-\varepsilon,\Phi(\varepsilon,y))\,, \tag{2.9}$$

the second equality holding only for $0 \le \tau \le 1$. Note that $\tilde{G}$ is analytic for $\varepsilon \ge 0$, $y \in \mathbb{C}^N$ with $|y| < \rho(\varepsilon)$, and noninteger $\tau > 0$. We define

$$g(\varepsilon,\tau,x) := \varepsilon^{-p}(-f(x) + 1/\varepsilon D_\tau \tilde{G}(\tau,\varepsilon,y)) \tag{2.10.a}$$

where $y = y(\varepsilon,\tau,x)$ is given implicitly by

$$x = \tilde{G}(\tau,\varepsilon,y)\,. \tag{2.10.b}$$

Of course, this definition is equivalent to the original one given in (2.5.a,b).

We claim that (2.10.b) can be solved for y, by the implicit function theorem. Indeed, $\tilde{G}$ is C^∞ in τ, ε, y, and analytic in ε, y, where defined. The trivial solution of (2.10.b) for $\varepsilon = 0$ is

$$y = y(0,\tau,x) = x$$

because $\tilde{G}(\tau,0,y) = y$. Likewise, at $\varepsilon = 0$, we compute the partial derivative

$$\tilde{G}_y(\tau,0,y) = id\,,$$

which is invertible. Therefore (2.10.b) can be solved for $y = y(\varepsilon,\tau,x)$, uniformly for $0 \le \varepsilon < \varepsilon_*(\rho)$, as long as $|x| \le \rho$ and provided $\tilde{G}$ is defined. In particular, y is defined in a possibly reduced domain of the form $0 \le \tau \le 2$, $0 \le \varepsilon \le \varepsilon_0$, $|y| < \rho(\varepsilon)$.

Step 3: Properties of the perturbation g.

So far, $g = g(\varepsilon,\tau,x)$ is defined only for $0 \le \tau \le 2$, $\varepsilon > 0$. We first claim periodicity in τ, as in (iii), so that a 1–periodic extension of g to all real τ is consistent with our previous definition. Indeed, let $0 \le \tau \le 1$, $\varepsilon > 0$. Then

$$\Phi(\varepsilon,y(\tau+1,\varepsilon,x)) = y(\tau,\varepsilon,x) \tag{2.11}$$

because (2.7.b) implies

$$\tilde{G}(\tau,\varepsilon,\Phi(\varepsilon,y(\tau+1,\varepsilon,x))) = \tilde{G}(\tau+1,\varepsilon,y(\tau+1,\varepsilon,x)) = x\,.$$

Therefore we conclude from (2.7.b), (2.10.a,b), (2.11)

$$\begin{aligned}\varepsilon^{p+1}(g(\varepsilon,\tau+1,x)-g(\varepsilon,\tau,x)) &= D_\tau\widetilde{G}(\tau+1,\varepsilon,y(\tau+1,\varepsilon,x))-D_\tau\widetilde{G}(\tau,\varepsilon,y(\tau,\varepsilon,x))\\ &= D_\tau\widetilde{G}(\tau,\varepsilon,\Phi(\varepsilon,y(\tau+1,\varepsilon,x)))-D_\tau\widetilde{G}(\tau,\varepsilon,y(\tau,\varepsilon,x))\\ &= D_\tau\widetilde{G}(\tau,\varepsilon,y(\tau,\varepsilon,x))-D_\tau\widetilde{G}(\tau,\varepsilon,y(\tau,\varepsilon,x))=0\ .\end{aligned}$$

This proves 1–periodicity for $\varepsilon > 0$.

Next we extend g down to $\varepsilon = 0$. Here and below, let r_j denote various remainder terms which are real analytic with respect to $0 \le \varepsilon \le \varepsilon_0$, complex analytic for $|x| < \rho(\varepsilon)$ and C^∞ in τ. Because Φ is a p–th order discretization of the flow F, by assumption (1.13.b), we can write

$$\Phi(\varepsilon,x) = F(\varepsilon,x)+\varepsilon^{p+1}r_1\ . \tag{2.12}$$

Together with definition (2.9) of $\widetilde{G}$, this yields

$$\begin{aligned}\widetilde{G}(\tau,\varepsilon,x) &= \chi_0(\tau)F(\varepsilon\tau,x)+\chi_1(\tau)F(\varepsilon\tau-\varepsilon,\Phi(\varepsilon,x))\\ &= \chi_0(\tau)F(\varepsilon\tau,x)+\chi_1(\tau)F(\varepsilon\tau-\varepsilon,F(\varepsilon,x))+\varepsilon^{p+1}r_2\\ &= F(\varepsilon\tau,x)+\varepsilon^{p+1}r_2\ .\end{aligned} \tag{2.13}$$

Replacing x by y and using that $\widetilde{G}(\tau,\varepsilon,y(\varepsilon,\tau,x)) = x$, by definition, this implies

$$y(\varepsilon,\tau,x) = F(-\varepsilon\tau,x)+\varepsilon^{p+1}r_3\ . \tag{2.14}$$

Combining (2.13) and (2.14), we get

$$\begin{aligned}\varepsilon^{-1}D_\tau\widetilde{G}(\tau,\varepsilon,y(\varepsilon,\tau,x)) &= f(F(\varepsilon,\tau,y(\varepsilon,\tau,x)))+\varepsilon^pD_\tau r_2\\ &= f(F(\varepsilon\tau,F(-\varepsilon\tau,x)))+\varepsilon^pr_3\\ &= f(x)+\varepsilon^pr_3\ .\end{aligned}$$

Putting $g = r_3$, we see that all the differentiability requirements on g hold. This proves (ii). Also, by continuous extension of g down to $\varepsilon = 0$, periodicity property (iii) is proved. Finally, g is now defined on the domain specified in (i). Property (iv) holds by construction. Since all arguments are uniform in the suppressed variable λ, the proof is now complete. ■

Clearly, Φ does not determine g uniquely. Conversely, however, g with the above properties (i)–(iii) defines a corresponding time–ε map Φ via (iv). Then Φ can be viewed as some p–th order discretization of the flow F in the sense of assumption (1.13.b). Also, a similar construction applies to perturbations of normal forms by higher order terms which are not in normal form. Finally, we note that multi–parameter versions like $\lambda \in \mathbb{C}^m$ or λ in a Banach space can be proved analogously, as well as versions which do not require analyticity. The reason for our analyticity assumptions will become transparent in the next section — hopefully.

3. Exponential smallness

We prove Theorem 1.1 by Ljapunov–Schmidt reduction. Let the assumptions of Theorem 1.1 hold throughout this section, see (1.11.a–d), (1.13.a,b), (1.14). Our strategy of proof is an adaptation of previous work by Chow, Hale, and Mallet–Paret; see CHOW & HALE, Theorem 11.3.3 (1982). We will proceed as follows. In Section 2, we have seen how any p–th order discretization of

$$\dot{x} = f(\lambda, x) \tag{3.1}$$

by the iteration

$$x_{n+1} = \Phi(\varepsilon, \lambda, x_n) \tag{3.2}$$

is equivalent to the time–ε map of a suitable non–autonomous perturbation

$$\dot{x} = f(\lambda, x) + \varepsilon^p g(\varepsilon, \lambda, t/\varepsilon, x) \,. \tag{3.3}$$

Here g satisfies properties (i)–(iv) of Proposition 2.1; in particular g has period 1 in $\tau = t/\varepsilon$. The forcing in (3.3) is thus ε–periodic in τ. We call g a *rapid forcing* term. In contrast, CHOW & HALE (1982) have considered a forcing term g which has constant period 1 in t. It is precisely the structure of a rapid forcing which generates the exponentially small splittings in Theorem 1.1. Still, we will proceed pretty much as in CHOW & HALE (1982). We introduce a real time–shift β as a parameter and solve (3.3), locally and in suitable function spaces, except for a one–dimensional component. See Lemma 3.2 for this Ljapunov–Schmidt reduction. The remaining scalar reduced equation

$$\mathcal{B} = 0$$

is analyzed in the crucial Lemma 3.4. In particular, exponential smallness of $\mathcal{B}$ emerges. Relating back the reduced equation $\mathcal{B} = 0$ to the geometrically defined splitting quantities of homoclinic orbits, in Lemma 3.5, this will imply exponential smallness of the splitting quantities. The proof of the crucial Lemma 3.4 requires an extension to complex time, which is new to the method. In fact, we will introduce two time–shift variables, α and β, one being complex and the other one real, instead of just a single time–shift. For motivation, recall that g is only C^∞ in $\tau = t/\varepsilon$, whereas it is complex analytic in $t \mapsto x = \Gamma(t)$ along the original homoclinic orbit Γ. Below, α will keep track of analyticity whereas β traces periodicity of the forcing. Only in the reduced equation $\mathcal{B} = 0$, both aspects recombine to provide exponential smallness:

$$\text{analyticity} \;+\; \text{periodicity} \;=\; \text{exponential smallness} \,.$$

At the end of the section we complete the proof of Theorem 1.1, summarizing the arguments in a rigorous manner.

To set up our Ljapunov–Schmidt reduction of (3.3), we rewrite the equation in terms of a (small) perturbation z around the original homoclinic orbit $x = \Gamma(t)$ at $\varepsilon = \lambda_0 = 0$. Writing $z = x - \Gamma(t)$ and $A(t) = D_x f(\lambda_0, \Gamma(t))$, equation (3.3) reads

$$\dot{z} = A(t)z + (f(\lambda, \Gamma(t) + z) - f(\lambda_0, \Gamma(t)) - A(t)z + \varepsilon^p g(\varepsilon, \lambda, t/\varepsilon, \Gamma(t) + z)) \,.$$

Below, we will solve the system

$$(\mathcal{L}(\alpha)z)(t) - \mathcal{F}(\varepsilon, \lambda, \alpha, \beta, z(\cdot))(t) = 0 \,, \tag{3.4.a}$$

$$\int_{-\infty}^{+\infty} (z(t), \dot{\Gamma}(t + \alpha))dt = 0 \,, \tag{3.4.b}$$

where

$$(\mathcal{L}(\alpha)z)(t) := \dot{z}(t) - A(t + \alpha)z(t) \,, \tag{3.5.a}$$

$$\begin{aligned}\mathcal{F}(\varepsilon,\lambda,\alpha,\beta,z(\cdot))(t) \;:=\; & f(\lambda,\Gamma(t+\alpha)+z(t))-f(\lambda_0,\Gamma(t+\alpha)) \\ & -A(t+\alpha)z(t)+\varepsilon^p g(\varepsilon,\lambda,t/\varepsilon+\beta,\Gamma(t+\alpha)+z(t))\;.\end{aligned} \tag{3.5.b}$$

This equation makes sense for α in the closed complex strip Σ_η and for real β, t. Of course, z and Γ are considered with values in $\mathbb{C}^N$ and $(\cdot,\cdot)$ denotes the standard hermitian product. The treatment in CHOW & HALE (1982) corresponds to the special case $\alpha = 0$. Such an approach would not utilize our assumption (1.14), by which $\Gamma(\cdot)$ is complex analytic in the complex strip Σ_η.

We will solve (3.4.a) with values in the Banach space $C^0(\mathbb{R},\mathbb{C}^N)$, considering $z(\cdot)\in C^1(\mathbb{R},\mathbb{C}^N)$; as before C^k denotes k–times continuously differentiable functions with uniformly bounded derivatives and the usual sup–norm.

We collect some properties of $\mathcal{L}$ next.

Lemma 3.1 *The operator $\mathcal{L}(\alpha)$ defined above is bounded linear, Fredholm of index zero, and analytic in $\alpha\in\Sigma_\eta$ as a map*

$$\mathcal{L}(\alpha): C^1(\mathbb{R},\mathbb{C}^N)\to C^0(\mathbb{R},\mathbb{C}^N)\;. \tag{3.6}$$

The kernel of $\mathcal{L}(\alpha)$ is spanned by

$$t\mapsto \dot\Gamma(t+\alpha)\;\; in\;\; C^1(\mathbb{R},\mathbb{C}^N)\;. \tag{3.7}$$

The L^2–orthogonal complement to the range of $\mathcal{L}(\alpha)$ is spanned by

$$t\mapsto \psi(t+\alpha)\;,\;\; in\;\; C^0(\mathbb{R},\mathbb{C}^N)\;, \tag{3.8}$$

where $\psi(\cdot)$ is the analytic extension to Σ_η of the unique (up to scalar multiples) nontrivial bounded solution of the adjoint equation

$$\dot\psi(t) = -A(t)^*\psi(t)\;,\;\; t\in\mathbb{R}\;. \tag{3.9}$$

(As usual, $A^* := \overline{A}^T$).

Proof. To see that $\mathcal{L}(\alpha)$ is bounded linear, it suffices to prove that $\Gamma(\cdot)$, and therefore $A(\cdot) = D_x f(\lambda_0,\Gamma(\cdot))$, is continuous and uniformly bounded on Σ_η. Continuity is obvious. Boundedness is obvious on the real line and on compact subsets of Σ_η. But the differential equation for $\Gamma(\cdot)$ implies that

$$\lim\Gamma(t)=0\;,\;\; \lim\dot\Gamma(t)=0 \tag{3.10}$$

also holds in Σ_η if only the real part of t tends to $\pm\infty$. Indeed, this follows from the usual flow properties of a differential equation considered for complex times in the closed strip Σ_η. Therefore $\mathcal{L}(\alpha)$ is bounded.

Analyticity of $\alpha\mapsto\mathcal{L}(\alpha)$ in the operator norm follows from analyticity of $\alpha\mapsto A(\cdot+\alpha)\in C^0$.

It was observed in PALMER, Section 4 (1984) that $\mathcal{L}(0)$ is Fredholm of index zero. By continuation, all $\mathcal{L}(\alpha)$ are Fredholm of index zero, see KATO (1966).

By nondegeneracy assumption (1.11.d), the zero index Fredholm operators $\mathcal{L}(\alpha)$ have (also complex) one–dimensional kernel: the trivial one, given by $\dot\Gamma(\cdot+\alpha)$.

Finally, our claims on ψ are well–known for $\alpha=0$, see e.g. PALMER, Section 4 (1984). They extend to $\alpha\in\Sigma_\eta$ by analytic continuation of (3.8), (3.9) from real to complex time–shifts α. This proves the Lemma. ∎

We now follow the usual procedure of Ljapunov–Schmidt reduction. Let P_α denote L^2–orthogonal projection onto the co–kernel of $\mathcal{L}(\alpha)$,

$$P_\alpha z(\cdot) := \langle z(\cdot),\psi(\cdot+\alpha)\rangle\psi(\cdot+\alpha)\;. \tag{3.11}$$

Here $\langle \cdot, \cdot \rangle$ denotes the standard complex L^2–product over $\mathbb{R}$ and ψ has been L^2–normalized. We decompose system (3.4.a–b) into

$$(Id - P_\alpha)(\mathcal{L}(\alpha)z - \mathcal{F}(\varepsilon, \lambda, \alpha, \beta, z)) = 0\ ,\ \ \langle z(\cdot), \dot{\Gamma}(\cdot + \alpha)\rangle = 0 \tag{3.12.a}$$

and

$$P_\alpha(\mathcal{L}(\alpha)z - \mathcal{F}(\varepsilon, \lambda, \alpha, \beta, z)) = -P_\alpha \mathcal{F}(\varepsilon, \lambda, \alpha, \beta, z) = 0\ . \tag{3.12.b}$$

Lemma 3.2 *System* (3.12.a) *can be solved uniquely for* z *near* $\varepsilon = 0$, $\lambda = \lambda_0 = 0$, $z = 0$, *by the implicit function theorem. More precisely, there exist* $\varepsilon_0 > 0$, $\delta_0 > 0$, $C > 0$, *such that* (3.12.a) *has a unique solution*

$$z(\cdot) = z(\varepsilon, \lambda, \alpha, \beta; \cdot) \in C^1(\mathbb{R}, \mathbb{C}^N)$$

with $\|z\|_{C^1} < \delta_0$ *and the following properties:*

(i) z *is defined for all real* $0 \le \varepsilon \le \varepsilon_0$, $0 \le \beta \le 2$ *and all complex* $|\lambda| \le \varepsilon_0$, $\alpha \in \Sigma_\eta$ *with* $0 \le \Re\alpha \le 2$;

(ii) *down to* $\varepsilon = 0$, z *and all its derivatives with respect to* λ, α, β *(but not* ε*) are continuous and with respect to* λ, α *analytic.*

(iii) $\|z(\varepsilon, \lambda, \alpha, \beta; \cdot) - z(0, \lambda, \alpha, \beta; \cdot)\|_{C^1} \le C\varepsilon^p$ *for all* $\varepsilon, \lambda, \alpha, \beta$ *as above;*

(iv) z *has period* 1 *in* β;

(v) $z(\varepsilon, \lambda, \alpha, \beta; t + t') = z(\varepsilon, \lambda, \alpha + t', \beta + t'/\varepsilon; t)$ *for all* $\varepsilon, \lambda, \alpha, \beta$ *as above and real* t, t' *such that both sides are well-defined.*

Proof. To solve (3.12.a) for $z \in C^1$, we check the assumptions of the implicit function theorem. Obviously, $z = 0$ is a solution of (3.12.a) for $\varepsilon = 0$, $\lambda = \lambda_0 = 0$ and all shift parameters α, β; see the relevant definitions in (3.5.a,b), (3.8), (3.9), (3.11). The linearization of system (3.12.a) with respect to z at $z = 0$, $\varepsilon = 0$, $\lambda = \lambda_0$ is given by the bounded operator

$$\begin{array}{rccl} \widehat{\mathcal{L}}\colon & C^1 & \to & (Id - P_\alpha C^0) \times \mathbb{R} \\ & \zeta & \mapsto & ((Id - P_\alpha)\mathcal{L}(\alpha)\zeta, \langle \zeta, \dot{\Gamma}(\cdot + \alpha)\rangle)\ . \end{array} \tag{3.13}$$

We claim that $\widehat{\mathcal{L}}$ is invertible. Indeed, $\widehat{\mathcal{L}}$ is Fredholm of index zero, because $\mathcal{L}(\alpha)$ has index zero and $(Id - P_\alpha)$, $(Id - P_\alpha)\mathcal{L}(\alpha)$ are Fredholm of index one. Therefore, it suffices to show that $\widehat{\mathcal{L}}$ has trivial kernel. By definition, P_α annihilates the range of $\mathcal{L}(\alpha)$. Therefore the kernels of $(Id - P_\alpha)\mathcal{L}(\alpha)$ and of $\mathcal{L}(\alpha)$ coincide, they are spanned by $\dot{\Gamma}(\cdot + \alpha)$ see Lemma 3.1. But

$$\langle \zeta, \dot{\Gamma}(\cdot + \alpha)\rangle \neq 0$$

for $\zeta = \dot{\Gamma}(\cdot + \alpha)$. This proves invertibility of $\widehat{\mathcal{L}}$. Consequently, (3.12.a) can be solved for $z \in C^1$ as indicated in (i).

The differentiability properties (ii) of z follow from those of $\mathcal{L}$, $\dot{\Gamma}$ and $\mathcal{F}$. Differentiability with respect to ε is governed by the term

$$\varepsilon^p g(\varepsilon, \lambda, t/\varepsilon + \beta, \Gamma(t + \alpha) + z(t)) \tag{3.14}$$

contained in $\mathcal{F}$, cf. (3.5.b).

The p–th order estimate (iii) follows trivially from the uniform boundedness of $\mathcal{F}$ in the relevant region, its $\varepsilon^p g$ term, and invertibility of $\widehat{\mathcal{L}}$.

Periodicity (iv) of z in β follows from the same periodicity of $\mathcal{F}$ and g.

To check (v) we just shift time t in (3.12.a), (3.4.a,b) by t'. This is equivalent to solving (3.12.a), (3.4.a,b) with no time shift but, instead, α and β replaced by $\alpha + t'$ and $\beta + t'/\varepsilon$, respectively. Note that the shift invariant form of the normalization

$$\langle z(\cdot + t'), \dot{\Gamma}(\cdot + \alpha + t') \rangle = \langle z(\cdot), \dot{\Gamma}(\cdot + \alpha) \rangle$$

is used here. This proves the lemma. ∎

Solving system (3.12.a) for $z = z(\varepsilon, \lambda, \alpha, \beta; t)$, we have lost all differentiability with respect to ε; see (3.14). To recover even analyticity for strictly positive ε, we rescale time t as $\tau := t/\varepsilon$. For all $\varepsilon > 0$ and $\tau \in \mathbb{R}$ let

$$\tilde{z}(\varepsilon, \lambda, \alpha, \beta; \tau) := z(\varepsilon, \lambda, \alpha, \beta; \varepsilon\tau) \,. \tag{3.15}$$

Clearly, $\tilde{z}(\cdot) = \tilde{z}(\varepsilon, \lambda, \alpha, \beta; \cdot)$ satisfies the rescaled version of (3.4.a,b).

$$\begin{aligned} \tilde{z}' \;=\; & \varepsilon \big[A(\varepsilon\tau + \alpha)\tilde{z} + f(\lambda, \Gamma(\varepsilon\tau + \alpha) + \tilde{z}) - f(\lambda_0, \Gamma(\varepsilon\tau + \alpha)) \\ & - A(\varepsilon\tau + \alpha)\tilde{z} + \varepsilon^p g(\varepsilon, \lambda, \tau + \beta, \Gamma(\varepsilon\tau + \alpha) + \tilde{z}) \big] \,, \end{aligned} \tag{3.16.a}$$

$$\int_{-\infty}^{+\infty} (\tilde{z}(\tau), \dot{\Gamma}(\varepsilon\tau + \alpha)) d\tau = 0 \,, \tag{3.16.b}$$

where $'$ denotes differentiation with respect to τ. Viewing (3.16.a,b) as an operator equation, in analogy to (3.4.a,b), (3.12.a,b), we see that $\tilde{z}$ is analytic in ε for $\varepsilon > 0$, since all operators now depend analytically on ε. Indeed, note that the terms

$$t\dot{\Gamma}(t) \,, \quad t\dot{A}(t)$$

tend to zero exponentially for $|\operatorname{Re} t| \to \infty$, and t in a sufficiently narrow sector

$$\Sigma_{\eta,\delta} := \Sigma_\eta \cup \{ t \in \mathbb{C} \mid |\operatorname{Im} t| \le \delta |\operatorname{Re} t| \} \,,$$

where δ is chosen small enough. This observation holds true because these quantities decay exponentially for real $t \to \pm\infty$, and grow at most exponentially (at a possibly much larger but uniform rate) in the imaginary time direction, as long as $\Gamma(t)$ stays in some small neighborhood of $x = 0$. The observation implies complex differentiability with respect to ε in a sector

$$|\operatorname{Im} \varepsilon| < \delta \, |\operatorname{Re} \varepsilon| \,,$$

and thereby analyticity of $\tilde{z}$. However, even continuity with respect to ε is lost at $\varepsilon = 0$ in this new scaling, in general, since $\tilde{z} \equiv z(0)$ at $\varepsilon = 0$. For later use, we rephrase Lemma 3.2 in terms of $\tilde{z}$.

Corollary 3.3 *Let $\tilde{z} = \tilde{z}(\varepsilon, \lambda, \alpha, \beta; \tau)$ be defined by* (3.15). *Then*

- *(i) the domains of definition of z and $\tilde{z}$ coincide;*
- *(ii) for ε positive, $\tilde{z}$ is C^∞ in all variables, and all derivatives are analytic with respect to λ, α, and ε;*
- *(iii)* $\|\tilde{z}(\varepsilon, \lambda, \alpha, \beta; \cdot)\|_{C^1} \le C\varepsilon^p$,
- *(iv) $\tilde{z}$ has period 1 in β, and*
- *(v)* $\tilde{z}(\varepsilon, \lambda, \alpha, \beta; \tau + \tau') = \tilde{z}(\varepsilon, \lambda, \alpha + \varepsilon\tau', \beta + \tau'; \tau)$, *as in Lemma* 3.2.

Proof. Obvious from the above remarks. ∎

To complete our solution of (3.4.a,b) it remains to solve (3.12.b)

$$P_\alpha(\mathcal{L}(\alpha)z - \mathcal{F}(\varepsilon,\lambda,\alpha,\beta,z)) = 0\ ,$$

putting $z = z(\varepsilon,\lambda,\alpha,\beta;\cdot)$. Writing out the projection P_α as in (3.11), this is equivalent to the scalar complex *reduced equation*

$$\mathcal{B}(\varepsilon,\lambda,\alpha,\beta) = 0\ , \tag{3.17}$$

where

$$\mathcal{B}(\varepsilon,\lambda,\alpha,\beta) := \int_{\mathbb{R}} \overline{\psi}(t+\alpha)^T \mathcal{F}(\varepsilon,\lambda,\alpha,\beta,z(\varepsilon,\lambda,\alpha,\beta;\cdot))(t)\,\mathrm{d}t\ . \tag{3.18}$$

We collect some properties of the *bifurcation function* $\mathcal{B}$.

Lemma 3.4:

(i) *$\mathcal{B}$ has real period ε in α and 1 in β. Therefore $\mathcal{B}$ extends canonically to all real $\beta, 0 \le \varepsilon \le \varepsilon_0$, and all complex $|\lambda| \le \varepsilon_0$, $\alpha \in \Sigma_\eta$.*

(ii) *Down to $\varepsilon = 0$, $\mathcal{B}$ and all its derivatives with respect to λ, α, β (but not ε) are continuous and with respect to λ, α, β analytic. For ε positive, $\mathcal{B}$ is analytic in all variables.*

(iii) *$\mathcal{B}(\varepsilon,\lambda,\alpha,\beta) = \mathcal{B}(\varepsilon,\lambda,\alpha-\varepsilon\beta,0)$, for all arguments.*

(iv) *$\mathcal{B}$ possesses a Fourier expansion*

$$\mathcal{B}(\varepsilon,\lambda,\alpha,\beta) = \sum_{k\in\mathbb{Z}} \mathcal{B}^k(\varepsilon,\lambda) e^{2\pi i k((\alpha/\varepsilon)-\beta)} \tag{3.19}$$

which converges absolutely and uniformly. The coefficients $\mathcal{B}^k$ are as regular with respect to (ε,λ) as $\mathcal{B}$ itself. For sufficiently small $\delta > 0$ there exists a uniform constant $C_\delta > 0$ such that

$$|\mathcal{B}^k(\varepsilon,\lambda)| \le C_\delta e^{-2\pi(\eta+\delta)|k|/\varepsilon} \tag{3.20}$$

holds for all ε, λ, k.

(v) *$\mathcal{B}(0,\lambda,\alpha,\beta) = \mathcal{B}^0(0,\lambda)$, $\mathcal{B}^0(0,\lambda) = 0$, at $\lambda = \lambda_0 = 0$, $\mathcal{B}^0_\lambda(0,\lambda) \neq 0$, at $\lambda = \lambda_0 = 0$. There exists a constant C such that for all ε*

$$|\mathcal{B}^0(\varepsilon,\lambda) - \mathcal{B}(0,\lambda)| \le C\varepsilon^p\ .$$

Proof. We will show (iii) first. Because $\mathcal{F}$ and z are 1–periodic in β, this implies (i). Regularity (ii) of $\mathcal{B}$ will follow from Lemma 3.2 and Corollary 3.3, because $\mathcal{F}$ is analytic in $z(\cdot)$. We then establish the exponential smallness estimate (3.20) on the Fourier coefficients $\mathcal{B}^k$. Finally, we prove properties (v) of the averaged term $\mathcal{B}^0$ in the Fourier expansion (3.19) of $\mathcal{B}$. Excepting (v), we will suppress λ notationally.

To show (iii),

$$\mathcal{B}(\varepsilon,\alpha,\beta) = \mathcal{B}(\varepsilon,\alpha-\varepsilon\beta,0) \tag{3.21}$$

for all arguments, let

$$S_{t'} z(\varepsilon,\alpha,\beta;\cdot) := z(\varepsilon,\alpha,\beta;\cdot + t') \tag{3.22}$$

denote the time shift. For example

$$S_{t'}\mathcal{F}(\varepsilon,\alpha,\beta,z) = \mathcal{F}(\varepsilon,\alpha+t',\beta+t'/\varepsilon,S_{t'}z)$$

for real t' by definition (3.5.b) of $\mathcal{F}$. Also, by Lemma 3.2 (v)

$$S_{t'} z(\varepsilon, \alpha, \beta; \cdot) = z(\varepsilon, \alpha + t', \beta + t'/\varepsilon; \cdot) .$$

Together this implies

$$\begin{aligned} \mathcal{B}(\varepsilon, \alpha, \beta) &= \langle \mathcal{F}(\varepsilon, \alpha, \beta, z(\varepsilon, \alpha, \beta; \cdot)), \psi(\alpha + \cdot) \rangle \\ &= \langle S_{t'} \mathcal{F}(\varepsilon, \alpha, \beta, z(\varepsilon, \alpha, \beta; \cdot)), S_{t'} \psi(\alpha + \cdot) \rangle \\ &= \langle \mathcal{F}(\varepsilon, \alpha + t', \beta + t'/\varepsilon, z(\varepsilon, \alpha + t', \beta + t'/\varepsilon; \cdot)), \psi(\alpha + t' + \cdot) \rangle \\ &= \mathcal{B}(\varepsilon, \alpha + t', \beta + t'/\varepsilon) , \end{aligned} \tag{3.23}$$

for real t' and all positive ε. Putting $t' := -\varepsilon\beta$ this proves our claim (3.21), provided $\varepsilon > 0$. For $\varepsilon = 0$, (3.21) then holds by continuity of $\mathcal{B}$. This proves (iii).

Concerning regularity of $\mathcal{B}$, Lemma 3.2 and analyticity of $\mathcal{F}$ with respect to z imply that $\mathcal{B}$ and all its derivatives with respect to λ, α, β are continuous and in λ, α, β analytic, down to $\varepsilon = 0$. To prove analyticity including ε, for $\varepsilon > 0$, we invoke Corollary 3.3. Indeed, rescaling the integration variable t to become $\tau = t/\varepsilon$ and plugging this into expression (3.5.b) for $\mathcal{F}$, we have

$$\begin{aligned} \mathcal{B}(\varepsilon, \alpha, \beta) = \varepsilon \int_{-\infty}^{+\infty} & \overline{\psi}(\varepsilon\tau + \alpha)^T (f(\Gamma(\varepsilon\tau + \alpha) + \tilde{z}(t)) - f(\Gamma(\varepsilon\tau + \alpha)) \\ & - A(\varepsilon\tau + \alpha)\tilde{z}(\tau) + \varepsilon^p g(\varepsilon, \tau + \beta, \Gamma(\varepsilon\tau + \alpha) + \tilde{z}(\tau))) \, d\tau . \end{aligned}$$

Still, λ, λ_0 are suppressed and $\tilde{z}(\tau)$ stands for $\tilde{z}(\varepsilon, \alpha, \beta; \tau)$. Now analyticity of $\mathcal{B}$ in $\varepsilon > 0$ follows as in Corollary 3.3.

To prove (iv), we first note that $\mathcal{B}$ has real period ε in α, by (3.21), since $\mathcal{B}$ is 1–periodic in β. Because Γ is, in fact, analytic in a slightly enlarged closed complex strip $\sum_{\eta+\delta}$, for some $\delta > 0$, $\mathcal{B}$ is also complex analytic for $\alpha \in \sum_{\eta+\delta}$. Therefore, we may expand $\mathcal{B}(\varepsilon, \alpha, 0)$ into a Laurent series with respect to $q = e^{2\pi i \alpha/\varepsilon}$, for $e^{-2\pi(\eta+\delta)/\varepsilon} < |q| < e^{2\pi(\eta+\delta)/\varepsilon}$. This establishes the Fourier series (3.19) for $\mathcal{B}(\varepsilon, \alpha, \beta)$ with the desired regularity properties of the coefficients $\mathcal{B}^k$, $k \in \mathbb{Z}$. It remains to estimate these coefficients, uniformly. Indeed,

$$\mathcal{B}^k(\varepsilon) = \frac{1}{\varepsilon} \int_0^\varepsilon e^{2\pi i k \alpha/\varepsilon} \mathcal{B}(\varepsilon, \alpha, 0) d\alpha . \tag{3.24}$$

Since $\alpha \mapsto \mathcal{B}(\varepsilon, \alpha, 0)$ is complex analytic in $\sum_{\eta+\delta}$, and has real period ε in α, we may evaluate (3.24), as well, by integrating along the line from $0 \pm i(\eta + \delta)$ to $\varepsilon \pm i(\eta + \delta)$ rather than along $[0, \varepsilon]$. By continuity of $\mathcal{B}$, this yields the uniform exponential estimate

$$|\mathcal{B}^k(\varepsilon)| \leq C_\delta e^{-2\pi|k|(\eta+\delta)/\varepsilon}$$

for all k, ε, and (suppressed) λ. This proves (3.20) and uniform convergence of the Fourier series.

It remains to prove (v). The Fourier expansion (3.19), together with continuity of $\mathcal{B}$, $\mathcal{B}^k$ and the exponential estimate (3.20) imply that

$$\mathcal{B}(0, \lambda, \alpha, \beta) = \mathcal{B}^0(0, \lambda)$$

is independent of α, β. Denoting

$$z(0, \lambda, 0, 0; \cdot) = z^0(\lambda; \cdot) = z^0(\cdot)$$

and re–introducing λ, we find

$$\begin{aligned} \mathcal{B}^0(0, \lambda) &= \mathcal{B}(0, \lambda, \alpha = 0, \beta = 0) = \int_{-\infty}^{+\infty} \overline{\psi}(t)^T \mathcal{F}(0, \lambda, \alpha = 0, \beta = 0, z^0)(t) dt \\ &= \int_{-\infty}^{+\infty} \overline{\psi}(t)^T (f(\lambda, \Gamma(t) + z^0(t)) - f(\lambda_0, \Gamma(t)) - A(t) z^0(t)) dt . \end{aligned} \tag{3.25}$$

In particular, $\mathcal{B}^0(0,\lambda_0) = 0$ since $z^0 = 0$ at $\lambda = \lambda_0$ (cf. the proof of Lemma 3.2). Differentiating with respect to λ at $\varepsilon = 0$, $\lambda = \lambda_0 = 0$, and using $A(t) := D_x f(\lambda_0, \Gamma(t))$ we find

$$\begin{aligned} \mathcal{B}^0_\lambda(0,\lambda_0) &= \int_{-\infty}^{+\infty} \overline{\psi}(t)^T (f_\lambda + D_x f z^0_\lambda(t) - A(t) z^0_\lambda(t)) dt \\ &= \int_{-\infty}^{+\infty} \overline{\psi}(t)^T f_\lambda(0, \Gamma(t)) dt \neq 0 \,, \end{aligned} \tag{3.26}$$

because $f_\lambda(0,\Gamma(\cdot)) \notin \text{range}(\mathcal{L}) = \text{span}(\psi)^\perp$, by non-degeneracy assumption (1.11.d). Finally, we obtain $\mathcal{B}^0(\varepsilon,\lambda)$ from (3.25), if we replace t by $t+\alpha$, z^0 by $z(\varepsilon,\lambda,\alpha,\beta=0;\cdot)$, re-introduce the omitted term $\varepsilon^p g$, and average over $\alpha \in [0,\varepsilon]$. All these modifications are bounded of order $O(\varepsilon^p)$, in case of z by Lemma 3.2 (iii). Therefore,

$$|\mathcal{B}^0(\varepsilon,\lambda) - \mathcal{B}^0(0,\lambda)| \leq C\varepsilon^p \,,$$

and the proof is complete. ∎

We note that the above simple complex Fourier integration trick is the essential reason for all the exponential smallness results in this paper. We repeat that our separate treatment of (complex) time shifts α versus (real) time shifts β is motivated by this step: *α carries analyticity while β carries periodicity. Only in the reduced equation $\mathcal{B} = 0$, both properties recombine via Lemma* 3.4 (iii).

Lemmas 3.2 and 3.4 have reduced our original problem of finding homoclinic orbits of

$$\dot{x} = f(\lambda, x) + \varepsilon^p g(\varepsilon, \lambda, t/\varepsilon, x) \tag{3.27}$$

near $\Gamma(\cdot)$. In fact, for real λ, α,

$$x(t) := \Gamma(t+\alpha) + z(t)$$

with the L^2-normalization

$$\langle x(\cdot) - \Gamma(\cdot+\alpha)\,,\ \dot{\Gamma}(\cdot+\alpha)\rangle_{L^2} = 0$$

is homoclinic, C^1-near $\Gamma(\cdot+\alpha)$, if and only if

$$z(t) = z(\varepsilon,\lambda,\alpha,\beta=0;t) \tag{3.28}$$

and ε, λ, α satisfy the reduced equation

$$\mathcal{B}(\varepsilon,\lambda,\alpha,\beta=0) = 0 \tag{3.29}$$

Here and below, α will only be real.

As a final preparation to the proof of Theorem 1.1, we relate the reduced equation

$$\mathcal{B}(\varepsilon,\lambda,\alpha,\beta=0) = 0 \tag{3.30}$$

back to the original, geometric setting (1.8.a–d) which aims at measuring the distance $d(\varepsilon,\lambda,\vartheta)$ between the (local) stable and unstable manifolds in the flow box B. As an intermediate, we define a modified reduced equation

$$\mathcal{B}^*(\varepsilon,\lambda,\alpha,\beta=0) = 0 \tag{3.31}$$

as follows. We use the same decomposition (3.12.a,b) as for $\mathcal{B}$, except for the normalization

$$\langle z(\cdot), \dot{\Gamma}(\cdot+\alpha)\rangle_{L^2} = 0 \,,$$

which we replace by orthogonality at $t = 0$ (in the flow box B):

$$z(0)^T \dot{\Gamma}(\alpha) = 0 \,.$$

This modification corresponds to a change of bases in (α, z)–space, since either condition guarantees

$$z(\cdot) \notin \operatorname{span}\left(\frac{d}{d\alpha}\Gamma(\cdot + \alpha)\right) = \ker \mathcal{L}(\alpha) .$$

Consequently, (3.12.a) can still be solved for z, locally:

$$z = z^*(\varepsilon, \lambda, \alpha, \beta; \cdot) .$$

Following (3.18), we thus define

$$\mathcal{B}^*(\varepsilon, \lambda, \alpha, \beta) := \int_{\mathbb{R}} \overline{\psi}(t+\alpha)^T \mathcal{F}(\varepsilon, \lambda, \alpha, \beta, z^*(\varepsilon, \lambda, \alpha, \beta; \cdot))(t) dt .$$

The advantage of $\mathcal{B}^*$ over $\mathcal{B}$ is its close relation to the geometric distance function $d(\varepsilon, \lambda, \vartheta)$. Picking the coordinates (ϑ, y) in the flow box B such that $\vartheta = 0$, $y = 0$ at $\Gamma(0)$, it follows from the definitions that

$$d(\varepsilon, \lambda, \vartheta) = 0 \tag{3.32}$$

holds if and only if

$$\mathcal{B}^*(\varepsilon, \lambda, \vartheta, \beta = 0) = 0 .$$

Indeed, both equations describe the same fact: a homoclinic orbit $x(t)$ of (3.27) passes through the $\vartheta = \text{const.}$ plane in the box B at time $t = 0$.

However, a disadvantage of $\mathcal{B}^*$ also arises from the normalization

$$z(0)^T \cdot \dot{\Gamma}(\alpha) = 0 .$$

Unlike the L^2–normalization, this pointwise condition is not invariant under a time–shift $S_{t'}$ in both z and $\dot{\Gamma}$. As a consequence, unlike $\mathcal{B}$, the modified $\mathcal{B}^*$ cannot be expected to be ε–periodic in α. This impairs a direct proof of exponential smallness of $\mathcal{B}^*$ (which, however, does hold true).

Lemma 3.5 *Each of the equations* (3.30)–(3.32) *above can be solved for* λ, *locally. That is*

$$\mathcal{B}(\varepsilon, \lambda, \alpha, \beta = 0) = 0 \Longleftrightarrow \lambda = \lambda(\varepsilon, \alpha) ; \tag{3.33.a}$$

$$\mathcal{B}^*(\varepsilon, \lambda, \alpha, \beta = 0) = 0 \Longleftrightarrow \lambda = \lambda^*(\varepsilon, \alpha) ; \tag{3.33.b}$$

$$d(\varepsilon, \lambda, \vartheta) = 0 \Longleftrightarrow \lambda = \widehat{\lambda}(\varepsilon, \vartheta) ; \tag{3.33.c}$$

for $0 \leq \varepsilon, \alpha, \vartheta \leq \varepsilon_0$ *small enough. The functions* λ, λ^*, $\widehat{\lambda}$ *and their derivatives with respect to* α *rep.* ϑ *are continuous, and analytic in* α *down to* $\varepsilon = 0$. *For positive* ε, *these functions are analytic in both variables.*

The functions λ, λ^*, $\widehat{\lambda}$ *are mutually related as follows:*

$$\lambda(\varepsilon, \alpha) = \lambda^*(\varepsilon, \vartheta(\varepsilon, \alpha)) , \tag{3.34.a}$$

$$\lambda^*(\varepsilon, \vartheta) = \widehat{\lambda}(\varepsilon, \vartheta) , \tag{3.34.b}$$

where, with respect to the coordinates (ϑ, y) *in the flow box* B, *the transformation* $\vartheta = \vartheta(\varepsilon, \alpha)$ *is given explicitly by*

$$\vartheta = \alpha + z(\varepsilon, \lambda(\varepsilon, \alpha), \alpha, \beta = 0; t = 0)^T \cdot \dot{\Gamma}(\alpha) . \tag{3.35}$$

The transformation $\vartheta(\cdot, \cdot)$ *is as regular as* $\lambda(\cdot, \cdot)$.

Proof. Local solvability for λ follows because $\mathcal{B}_\lambda$, $\mathcal{B}^*_\lambda$, d_λ are all nonzero, locally. Indeed,

$$\mathcal{B}(0,\lambda,\alpha,\beta) = \mathcal{B}^0(0,\lambda)$$

and

$$\mathcal{B}^0_\lambda(0,\lambda_0) \neq 0$$

by Lemma 3.4 (v). Similarly,

$$\mathcal{B}^*_\lambda(0,\lambda_0,\alpha,\beta) \neq 0$$

follows from the reasoning in (3.25), (3.26), which did not depend on periodicity in α. Also $d_\lambda \neq 0$, because nondegeneracy assumption (1.11.d) implies that the stable and unstable manifolds for $\varepsilon = 0$ cross each other with nonvanishing speed as λ increases through zero.

The differentiability properties of λ, λ^* follow from those of $\mathcal{B}$, $\mathcal{B}^*$ by the arguments in the proof of Lemma 3.4. Because $d(\varepsilon,\lambda,\vartheta) = 0$ if and only if $\mathcal{B}^*(\varepsilon,\lambda,\vartheta,\beta = 0) = 0$, we obtain (3.34.b), that is, λ^* and $\widehat{\lambda}$ are in fact the same function.

It remains to prove (3.34.a). We claim that

$$\lambda(\varepsilon,\alpha) = \lambda^*(\varepsilon,\vartheta) \tag{3.36}$$

if ϑ and α are related by (3.35). Indeed, $\mathcal{B}(\varepsilon,\lambda(\varepsilon,\alpha),\alpha,\beta = 0)$ implies that (3.27) possesses a homoclinic orbit

$$x(t) = \Gamma(t+\alpha) + z(\varepsilon,\lambda(\varepsilon,\alpha),\alpha,\beta = 0;t)\ ,$$

with the L^2–normalization for z. Now suppose (3.35), in short: $\vartheta = \alpha + z(0)^T \cdot \dot{\Gamma}(\alpha)$. Define

$$z^*(0) := z(0) - (z(0)^T \cdot \dot{\Gamma}(\alpha)) \cdot \dot{\Gamma}(\alpha)\ .$$

In the flow box B we then have

$$x(0) = \Gamma(\alpha) + z(0) = \alpha\dot{\Gamma} + z(0) = (\vartheta - z(0)^T \cdot \dot{\Gamma})\dot{\Gamma} + z(0) = \Gamma(\vartheta) + z^*(0)$$

since $\dot{\Gamma}$ is a constant unit vector in the ϑ–direction within B. Moreover

$$z^*(0)^T \cdot \dot{\Gamma} = 0\ .$$

Therefore, the solution $x(t)$ of (3.27), can also be written as

$$x(t) = \Gamma(t+\vartheta) + z^*(\varepsilon,\lambda(\varepsilon,\alpha),\vartheta,\beta = 0;t)\ ,$$

and hence $\mathcal{B}^*(\varepsilon,\lambda(\varepsilon,\alpha),\vartheta,\beta = 0) = 0$. This implies $\lambda(\varepsilon,\alpha) = \lambda(\varepsilon,\vartheta)$.

The regularity properties of $\vartheta = \vartheta(\varepsilon,\alpha)$ defined by (3.35) follow from those of z, $\tilde{z}$, and $\lambda(\cdot,\cdot)$ as specified above and in Lemma 3.2, and Corollary 3.3. We only note that

$$\vartheta(0,\alpha) = \alpha\ , \tag{3.37}$$

since $\lambda(0,\alpha) = \lambda_0 = 0$ and $z = z^* = 0$ there. Hence (3.35) in fact defines a local diffeomorphism

$$\alpha \mapsto \vartheta$$

for small positive ε. This completes the proof. ∎

Proof of Theorem 1.1. We first prove the exponential estimates (1.15.a–c) of the splitting length $\ell(\varepsilon)$ of the splitting interval $\lambda \in I(\varepsilon)$, of the splitting distance $d(\varepsilon, \lambda)$, and of the splitting slope $\overline{\omega}(\varepsilon, \lambda)$, as defined in (1.8.a–d). We then define the curve $\lambda = \lambda(\varepsilon) \in I(\varepsilon)$ and prove the p–th order estimate (1.16.b) for $\lambda(\varepsilon)$, $\varepsilon \searrow 0$.

Consider the length $\ell(\varepsilon)$ of the interval $I(\varepsilon)$. Note that

$$I(\varepsilon) = \{\lambda(\varepsilon, \alpha) | 0 \leq \alpha \leq \varepsilon\} \ .$$

Here $\lambda(\varepsilon, \alpha)$ solves

$$\mathcal{B}(\varepsilon, \lambda(\varepsilon, \alpha), \alpha, \beta = 0) = 0 \tag{3.38}$$

as in Lemma 3.5 and inherits real period ε in α from $\mathcal{B}$; see Lemma 3.4. Therefore,

$$\ell(\varepsilon) \leq C \ \exp(-2\pi\eta/\varepsilon)$$

follows once we show

$$|\lambda_\alpha(\varepsilon, \alpha)| \leq C \ \exp(-2\pi\eta/\varepsilon) \ , \tag{3.39}$$

for some constants C. To prove (3.39), we differentiate (3.38) with respect to α, implicitly:

$$\lambda_\alpha = -\mathcal{B}_\alpha / \mathcal{B}_\lambda \ .$$

By Lemma 3.4, $\mathcal{B}_\lambda$ is bounded away from zero. Moreover, $\mathcal{B}_\alpha$ (replacing λ_α) satisfies an estimate (3.39), due to the uniformly convergent Fourier expansion (3.19) and estimate (3.20) on the Fourier coefficients. Therefore, an estimate (3.39) also holds for λ_α. This proves exponential smallness of $\ell(\varepsilon)$.

Let us estimate $|d(\varepsilon, \lambda, \vartheta)|$ and $d_\vartheta(\varepsilon, \lambda, \vartheta)|$ for $\lambda \in I(\varepsilon)$ next. (We recall that the maximum over ϑ of these quantities defines $d(\varepsilon, \lambda)$, $\overline{\omega}(\varepsilon, \lambda)$, respectively.) Since $\ell(\varepsilon)$ is exponentially small and d, d_ϑ have uniformly bounded partial derivatives with respect to λ, it suffices to show exponential smallness of

$$d(\varepsilon, \lambda, \vartheta) \ , \quad d_\vartheta(\varepsilon, \lambda, \vartheta)$$

at the particular values $\lambda = \widehat{\lambda}(\varepsilon, \vartheta) \in I(\varepsilon)$ introduced in Lemma 3.5. By definition,

$$d(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta) = 0 \ . \tag{3.40}$$

This proves exponential smallness of $|d(\varepsilon, \lambda, \vartheta)|$, $|d(\varepsilon, \lambda)|$.

Exponential smallness of $d_\vartheta(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta)$ is less trivial. Differentiating implicitly in (3.40) and in Lemma 3.5, we find

$$\begin{aligned} d_\vartheta(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta) &= -d_\lambda \cdot \widehat{\lambda}_\vartheta = -d_\lambda \cdot \lambda_\vartheta^* = -(d_\lambda / \vartheta_\alpha) \cdot \lambda_\alpha \\ &= -\big(d_\lambda(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta) \ / \ \vartheta_\alpha(\varepsilon, \alpha)\big) \cdot \lambda_\alpha(\varepsilon, \alpha) \ . \end{aligned}$$

Here ϑ and α are related by (3.35) and, for readability, we have transiently suppressed some function arguments. By (3.37), ϑ_α is uniformly bounded away from zero. Hence exponential smallness (3.39) of λ_α implies exponential smallness of d_ϑ, $\overline{\omega}$ and ω. This proves the exponential estimates (1.15.a–c) of Theorem 1.1.

It remains to construct the curve $\varepsilon \mapsto \lambda(\varepsilon) \in I(\varepsilon)$ of order p in ε. We define $\lambda(\varepsilon)$ by solving the equation $\mathcal{B}^0(\varepsilon, \lambda) = 0$ for the constant coefficient $\mathcal{B}^0$ in the Fourier expansion of $\mathcal{B}$ (Lemma 3.4):

$$\mathcal{B}^0(\varepsilon, \lambda(\varepsilon)) := 0 \ . \tag{3.41}$$

Recall that indeed $\mathcal{B}^0_\lambda \neq 0$. Thus $\lambda(\varepsilon)$ exists, is continuous down to $\varepsilon = 0$, and analytic for positive ε.

We prove $\lambda(\varepsilon) \in I(\varepsilon)$ next. By definition,

$$0 = \mathcal{B}^0(\varepsilon, \lambda(\varepsilon)) = \frac{1}{\varepsilon} \int_0^\varepsilon \mathcal{B}(\varepsilon, \lambda(\varepsilon), \alpha, \beta = 0) d\alpha \ ,$$

and the integrand is continuous in α. In particular, there exists an α for which

$$\mathcal{B}(\varepsilon, \lambda(\varepsilon), \alpha, \beta = 0) = 0 \ ,$$

because the integrand cannot be strictly of one sign. Therefore,

$$\lambda(\varepsilon) = \lambda(\varepsilon, \alpha) \in I(\varepsilon) \ .$$

It remains to prove that $\lambda(\varepsilon)$ is of order p in ε. Invoking

$$|\mathcal{B}^0(\varepsilon, \lambda) - \mathcal{B}^0(0, \lambda)| \leq C' \varepsilon^p$$

from Lemma 3.4 (v) yields

$$|\mathcal{B}^0(0, \lambda(0)) - \mathcal{B}^0(0, \lambda(\varepsilon))| = |\mathcal{B}^0(\varepsilon, \lambda(\varepsilon)) - \mathcal{B}^0(0, \lambda(\varepsilon))| \leq C' \varepsilon^p \ .$$

Since $\mathcal{B}^0_\lambda$ is uniformly bounded away from zero, this implies a p–th order estimate

$$|\lambda(\varepsilon) - \lambda(0)| \leq C \varepsilon^p \ .$$

The proof of Theorem 1.1 is therefore complete. ■

4. Genericity of positive splitting

Our proof of Theorem 1.2 is somewhat monolithic. Before unraveling the details, we provide some background and then give an outline. Throughout this section, let the assumptions of Theorem 1.2 hold.

Generic transversality of homoclinic points of analytic maps was established by ZEHNDER (1973) already, even in a symplectic context which greatly reduces the class of admissible perturbations. In our non–symplectic context, the basic geometric reason for genericity of transversality is simple. Fix $\varepsilon > 0$, λ, for a moment, and consider the iteration by $\Phi = \Phi(\varepsilon, \lambda, \cdot)$ in the flow box B; see Section 1, (1.7.a–c),(1.8.a–c) and Fig. 1.3 for notation. Transverse intersection of W^u_{loc} and W^s_{loc} means that the splitting angle is positive at the intersection point $\vartheta = \vartheta_0$, that is

$$d(\vartheta, \lambda, \vartheta_0) = 0 \neq d_\vartheta(\varepsilon, \lambda, \vartheta_0) \,. \tag{4.1}$$

By an analytic small perturbation of $\Phi(\varepsilon, \lambda, \cdot)$, in the (strong Whitney type) topology of C^ω, see (1.17),(1.18), we can achieve this as illustrated in Fig. 4.1. Construct a C^∞–perturbation of Φ,

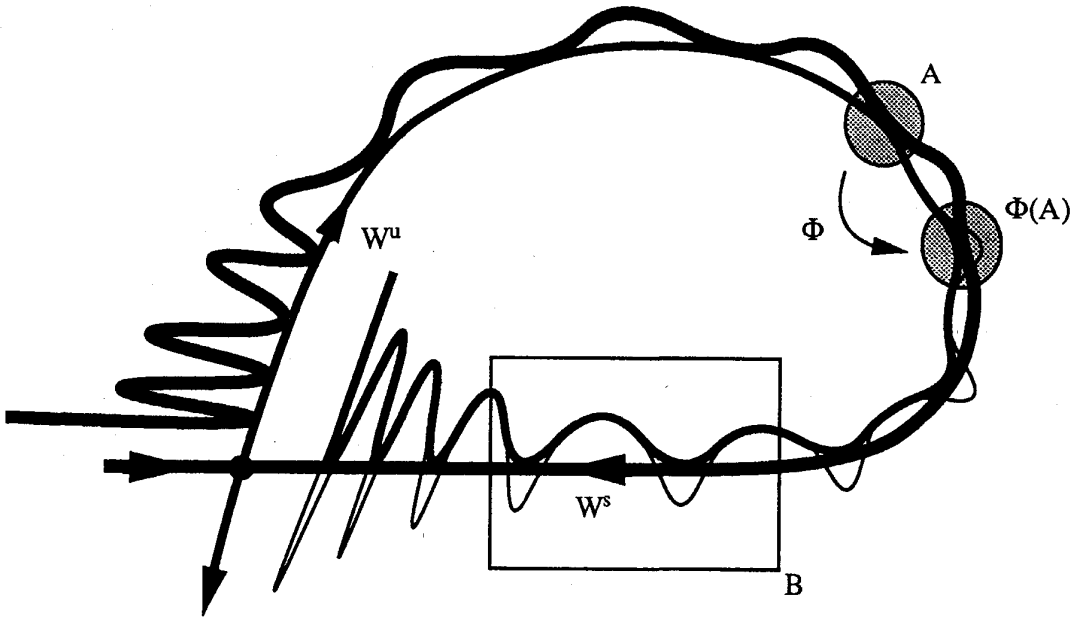


Figure 4.1: Perturbation of the unstable manifold W^u_{loc}.

first. Cutting off the perturbation outside a small region A, we may perturb W^u_{loc} in the disjoint region $\Phi(A)$ arbitrarily. But W^s_{loc} is unchanged in $\Phi(A)$. In particular W^u_{loc} and W^s_{loc} can be forced to intersect transversely, also in B, by an arbitrarily small perturbation $\Phi + \delta\Psi$. Approximating the C^∞ cut–off by an analytic function in C^1, locally in a neighborhood of the homoclinic orbit Γ, and then choosing δ small enough, the same holds true for arbitrarily small analytic perturbations $\Phi + \delta\Psi$. For a numerical experiment, implementing this idea, see BEYN, Fig. 3 (1987).

It would be a fairly standard application of transversality theory, ABRAHAM & ROBBIN (1967), to conclude $\ell(\varepsilon) > 0$, provided we could arbitrarily perturb the family of maps $\Phi(\varepsilon, \lambda, \cdot)$, depending

on both λ and ε. However, perturbations are admissible for fixed ε, only, in our setting. Therefore, we have to set up our arguments more carefully, exploiting analyticity.

We now explain the bulk part of the proof. Fix an open neighborhood $\mathcal{V}$ of f_0 in C^ω such that, for f in $\mathcal{V}$, the box B, the splitting distance function $d(\varepsilon, \lambda, \vartheta)$ and the curve $\lambda = \lambda(\varepsilon)$ from Theorem 1.1 exist for $0 \leq \varepsilon \leq \varepsilon_0$, $-\varepsilon_0 \leq \lambda \leq \varepsilon_0$, with a uniform small $\varepsilon_0 > 0$ Let ε_0 also be chosen small enough that

$$d_\lambda(\varepsilon, \lambda, \vartheta) \neq 0 \tag{4.2}$$

holds for all ε, λ, ϑ. Recall that (4.2) holds for f_0 at $\varepsilon = 0$, by nondegeneracy assumption (1.11.d). Hence

$$\ell(\varepsilon) = 0 \text{ iff } d(\varepsilon, \lambda(\varepsilon), \cdot) \equiv 0 \tag{4.3}$$

by the implicit function theorem.

Below we write $\ell(f, \varepsilon)$, $d(f, \varepsilon, \lambda, \vartheta)$, $\lambda(f, \varepsilon)$, to emphasize dependence on the vector field $f \in \mathcal{V}$. Let

$$\mathcal{G} := \{f \in \mathcal{V} \,|\, \ell(f, \varepsilon) > 0 \text{ for all } 0 < \varepsilon \leq \varepsilon_0\} \,. \tag{4.4}$$

We have to show that $\mathcal{G}$ contains a residual subset, in order to prove claim (1.20.a) of Theorem 1.2. For notational convenience, let

$$D(f, \varepsilon, \vartheta) := d(f, \varepsilon, \lambda(f, \varepsilon), \vartheta) \,. \tag{4.5}$$

For integers $m > 1/\varepsilon_0$ we define

$$\mathcal{G}_m := \{f \in \mathcal{V} \,|\, D(f, \varepsilon, 0) = 0 \Rightarrow D_\vartheta(f, \varepsilon, 0) \neq 0 \,, \quad \text{for all } \varepsilon \in [1/m, \varepsilon_0]\} \,. \tag{4.6}$$

By (4.3), (4.4), this implies

$$\mathcal{G} \supseteq \bigcap_m \mathcal{G}_m \,. \tag{4.7}$$

Obviously each set $\mathcal{G}_m$ is open. To prove genericity of $\mathcal{G}$, alias (1.20.a), it only remains to show that each set $\mathcal{G}_m$ is also dense, because the neighborhood $\mathcal{V} \subseteq C^\omega$ is a Baire space: then $\bigcap_m \mathcal{G}_m$ is residual.

To show $\mathcal{G}_m$ is dense, it is sufficient to show that any (small) open neighborhood V of the, so far arbitrary, vector field f in $\mathcal{V}$ contains an element $\tilde{h}$ of $\mathcal{G}_m$. Before defining $\tilde{h}$, first pick some $h \in V$ for which the number $q(h)$ of zeros of

$$\varepsilon \mapsto D(h, \varepsilon, \vartheta = 0) \,, \quad 1/m \leq \varepsilon \leq \varepsilon_0 \,, \tag{4.8}$$

is finite. Here zeros are counted without their algebraic multiplicity.

We claim that such an h exists. Indeed, the map (4.8) is analytic in ε since d and, by Theorem 1.1, $\lambda(\cdot)$ are. Fixing $\varepsilon = \varepsilon_0$, there exists $h \in V$ for which

$$D(h, \varepsilon_0, \vartheta = 0) \neq 0 \,, \tag{4.9}$$

by a perturbation of $\Phi(\varepsilon_0, \lambda(\varepsilon_0), \cdot)$ as indicated earlier in this section. The perturbation of Φ at level $\varepsilon = \varepsilon_0$ can be achieved by a perturbation of f, due to assumption (1.19). By analyticity of the map (4.8), this proves existence of h, viz. finiteness of $q(h)$. Note that finiteness persists on an open neighborhood V' of h in V, because the zeros of (4.8) also have finite algebraic multiplicity. Without loss of generality, let $q = q(h)$ be maximal on V'. Denote the zeros of (4.8) by $\varepsilon_1, \ldots, \varepsilon_q$.

To complete the proof of density of $\mathcal{G}_m$, we perturb $\Phi(\varepsilon_1, \lambda(\varepsilon_1), \cdot)$ such that

$$D = 0 \neq D_\vartheta \tag{4.10}$$

at $\vartheta = 0$, $\varepsilon = \varepsilon_1$. This can be achieved via a perturbation of h in V', again by assumption (1.19). Perturbing further and even less, within V', we can achieve that (4.10) holds at both $\varepsilon = \varepsilon_1$

and $\varepsilon = \varepsilon_2$. Proceeding iteratively, we arrive at an element $\widetilde{h}$ of V' for which (4.10) holds at $\varepsilon = \varepsilon_1, \ldots, \varepsilon_q$. By maximality of $q = q(h) = q(\widetilde{h})$, these enumerate all zeros of

$$\varepsilon \mapsto D(\widetilde{h}, \varepsilon, \vartheta = 0) \,, \quad 1/m \leq \varepsilon \leq \varepsilon_0 \,.$$

By (4.3), this implies $\widetilde{h} \in \mathcal{G}_m$, proving density of $\mathcal{G}_m$. This completes the proof of genericity of $\mathcal{G}$, alias (1.20.a).

To complete the proof of Theorem 1.2 we claim that for $0 < \varepsilon \leq \varepsilon_0$ and some $\lambda(\varepsilon) \in I(\varepsilon)$

$$d(\varepsilon, \lambda(\varepsilon)) = \max_{\vartheta} |d(\varepsilon, \lambda(\varepsilon), \vartheta)| > 0 \tag{1.20.b}$$

$$\omega(\varepsilon, \lambda(\varepsilon)) = \max_{\vartheta : d = 0} |d_\vartheta(\varepsilon, \lambda(\varepsilon), \vartheta)| > 0 \tag{1.20.c}$$

also hold, if $f \in \bigcap_m \mathcal{G}_m$. It is sufficient to consider $f \in \mathcal{G}_m$ and show the claim for $1/m \leq \varepsilon \leq \varepsilon_0$.

As before, let $\varepsilon_1, \ldots, \varepsilon_q$ denote the zeros of

$$\varepsilon \mapsto D(f, \varepsilon, \vartheta = 0) \,, \quad 1/m \leq \varepsilon \leq \varepsilon_0 \,.$$

Then $f \in \mathcal{G}_m$ implies for all j

$$\omega(\varepsilon_j, \lambda(\varepsilon_j)) \geq |D_\vartheta(f, \varepsilon_j, \vartheta = 0)| > 0 \,, \tag{4.11}$$

$$d(\varepsilon_j, \lambda(\varepsilon_j)) = \max |D(f, \varepsilon_j, \cdot)| > 0 \,. \tag{4.12}$$

If, on the other hand, $\varepsilon \neq \varepsilon_j$, $j = 1, \ldots, q$, then

$$d(\varepsilon, \lambda(\varepsilon)) \geq |D(f, \varepsilon, \vartheta = 0)| > 0 \,. \tag{4.13}$$

Together, (4.12) and (4.13) prove claim (1.20.b).

To prove (1.20.c) we solve $d(\varepsilon, \lambda, \vartheta) = 0$ for $\lambda = \widehat{\lambda}(\varepsilon, \vartheta)$, as in Lemma 3.5:

$$d(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta) = 0 \tag{4.14}$$

for all ε, ϑ. Recall that $d_\lambda \neq 0$ and $\widehat{\lambda}(\varepsilon = 0, \vartheta) \equiv \lambda_0 = 0$; see (4.2). As before, we find that $I(\varepsilon)$ is the image of $\widehat{\lambda}(\varepsilon, \cdot)$. Differentiating (4.14) with respect to ϑ, implicitly, we find

$$\omega(\varepsilon, \lambda(\varepsilon, \vartheta)) \geq |d_\vartheta(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta)| = |d_\lambda(\varepsilon, \widehat{\lambda}(\varepsilon, \vartheta), \vartheta)| \cdot |\widehat{\lambda}_\vartheta(\varepsilon, \vartheta)| \,. \tag{4.15}$$

For any $\varepsilon \in [0, \varepsilon_0]$ there exists some ϑ such that $\widehat{\lambda}_\vartheta(\varepsilon, \vartheta) \neq 0$, because $\ell(\varepsilon)$, the length of $I(\varepsilon)$ is positive. Because $d_\lambda \neq 0$, throughout, this proves (1.20.c) picking $\lambda(\varepsilon) = \widehat{\lambda}(\varepsilon, \vartheta)$. Theorem 1.2 is therefore proved.

■

5. Estimating the chaotic wedge

To prove Theorem 1.3, we investigate the recurrent behavior in small tubular neighborhoods $U' \supseteq U$ of the original, planar homoclinic orbit Γ. Our geometric approach is illustrated in Fig. 5.1 and outlined as follows. Without loss of generality, we only consider the fast contraction case $0 < \nu < \mu$. The opposite, fast expansion, case $0 < \mu < \nu$ reduces to the former by reversing time. We construct a Poincaré type section B^{in} to the stable manifold W^s_{loc} such that any orbit of the discretization Φ which keeps cycling in U keeps hitting B^{in}. The coordinates in B^{in} will be suitably normalized. In Lemma 5.1 we prove that any iterate of $\Phi(\varepsilon, \lambda, \cdot)$, which maps a point of B^{in} back into B^{in}, contracts area, locally at that point. If, in fact, W^u_{loc} misses W^s_{loc} sufficiently above, intersecting the box B^{in}, then we will be able to construct a continuous return map Θ. This graph transform Θ acts on a suitable compact set $\Lambda \subseteq C^0$ of uniformly Lipschitz continuous curves in B^{in}. Fixed points of Θ in Λ correspond to closed Lipschitz curves γ in U which are invariant under the discretization Φ. In Lemma 5.2 we show that Θ can have at most one such fixed point. In Lemma 5.3 we invoke Schauder's fixed point theorem to prove that a fixed point of Θ actually exists, provided

$$\Theta : \Lambda \to \Lambda \tag{5.1}$$

maps Λ into itself. In Lemma 5.6 we finally establish (5.1), after some quantitative estimates concerning the return slope of W^u_{loc} and the inclination or λ-Lemma. See Lemmata 5.4 and 5.5. In Lemma 5.7 we show that the invariant Lipschitz curve γ indeed attracts all points which stay in U, excepting W^s of course. In contrast, Lemma 5.8 shows that all orbits escape if W^u_{loc} misses W^s_{loc} sufficiently below. Finally, we summarize all our choices of small parameters. Showing consistency of these choices will complete the proof of Theorem 1.3.

We fix the following normalizations and coordinates throughout this section. We first reparametrize λ in $\Phi = \Phi(\varepsilon, \lambda, \cdot)$ such that $\lambda(\varepsilon) \equiv 0$ in Theorem 1.1, replacing λ by $\lambda - \lambda(\varepsilon)$. At $\varepsilon = 0$, the flow of $\dot{x} = f(\lambda, x)$ will also be locally C^1-linearized with respect to x at $x = 0$ by a global C^1-diffeomorphism which depends C^1 on λ in the C^1-topology. Denoting coordinates $x = (x_1, x_2)$ after linearization by (ϑ, y) we then have

$$\begin{aligned} \dot{\vartheta} &= -\mu(\lambda)\vartheta\,, \\ \dot{y} &= \nu(\lambda) y\,. \end{aligned} \tag{5.2}$$

We also stretch the (ϑ, y)-coordinates such that (5.2) holds in a unit box $|\vartheta|, |y| \leq 1$. The above linearization is assumed for notational simplification only. For a reference see BELITSKII (1973) and DENG (1989). Note differentiable dependence on λ; indeed the normalizing diffeomorphisms are obtained by a contraction mapping argument where λ enters differentiably. For $\varepsilon > 0$, we still normalize the stable and unstable manifold of the (likewise normalized) hyperbolic fixed point $\vartheta = y = 0$ to be given locally by the ϑ-axis and the y-axis up to $\vartheta = \pm 1$, $y = \pm 1$, respectively.

Note that the first reentry of W^u_{loc} into the unit box is parametrized over $1/2 \leq \vartheta \leq 1$ as the graph of a function

$$y = d(\varepsilon, \lambda, \vartheta) \tag{5.3}$$

where d is related to the distance in Section 1, by a diffeomorphism. Indeed, differently from Section 1, the (ϑ, y)-coordinates employed here do not exactly define a local flow box B but, instead, a linear flow. Anyway, the relation is so close that we deliberately put up with a possibly arising notational confusion. For the flow, that is for $\varepsilon = 0$, we further normalize λ by requiring

$$d(\varepsilon = 0, \lambda, \vartheta = 0) = \lambda\,. \tag{5.4}$$

With the above normalizations we define the box $B^{\text{in}} = B^{\text{in}}_\delta$, for fixed $\varepsilon > 0$ and λ, by its boundary as follows. The right boundary is given by the vertical line $\vartheta = 1$, $0 \leq y \leq \delta$. The left boundary is obtained by applying $\Phi(\varepsilon, \lambda, \cdot)$ to that line. And the top/bottom boundaries are straight lines connecting the top/bottom end points of the vertical line and its image. We note that B^{in} is Lipschitz

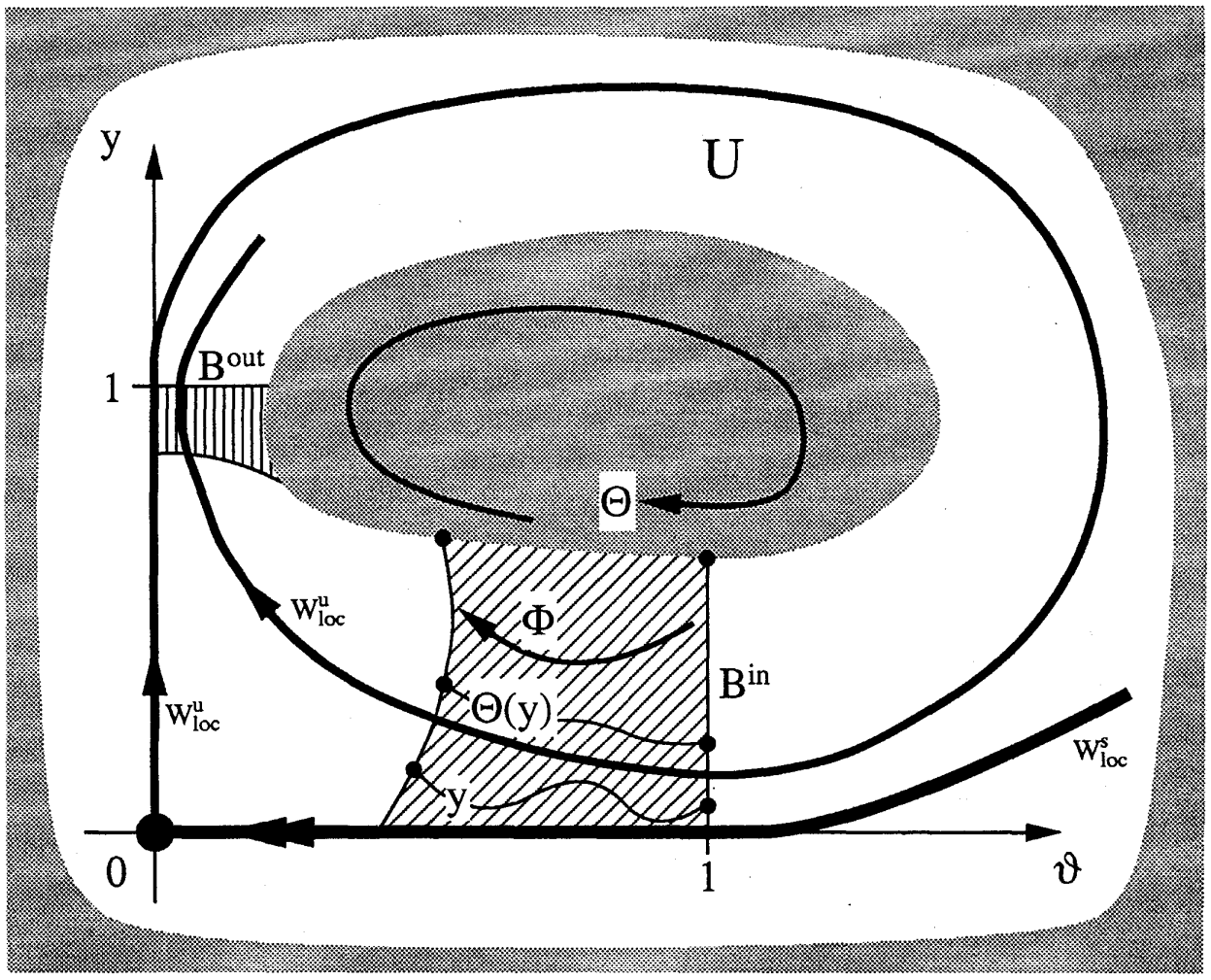


Figure 5.1: Recurrence in U.

homeomorphic to a convex set: the quadrangle determined by its four corners or, equivalently, a rectangle over the ϑ–axis.

The compact set of Λ of Lipschitz segments is defined next. It consists of graphs of Lipschitz functions

$$\vartheta \mapsto y(\vartheta)\,, \quad \tfrac{1}{2} \le \vartheta \le 1\,,$$

which are constant outside of B^{in}_δ. For the slopes we require

$$-\frac{1}{k} \le \frac{y(\vartheta_1) - y(\vartheta_2)}{\vartheta_1 - \vartheta_2} \le -k\,, \tag{5.5}$$

for all $\vartheta_1 \neq \vartheta_2$ such that $(\vartheta_\iota, y(\vartheta_\iota)) \in B^{\mathrm{in}}$, $\iota = 1, 2$. Like the height δ of B^{in}_δ, the parameter $k > 0$ will be chosen appropriately below. We finally require that the intersections of each graph in Λ with the vertical boundaries of B^{in} are related by Φ, just as these boundaries themselves are.

Note that Λ, equipped with the sup-norm topology, is a compact subset of the Banach space C^0 by Arzela–Ascoli's theorem. Moreover, Λ is homeomorphic to a convex set, becaus B^{in} is. Thus Λ

lends itself to an application of Schauder's fixed point theorem. For a map

$$\Theta : \Lambda \to C^0 \tag{5.6}$$

we take the graph transform on Λ induced by the return map of $\Phi(\varepsilon, \lambda, \cdot)$ on B^{in}, see Fig. 5.2. Of course, this map Θ will not be well-defined for all λ, but only for $\lambda \geq C' e^{-2\pi\eta/\varepsilon}$, according to the Lemmata below.

Lemma 5.1 *Fix a small enough neighborhood U' of Γ in $\mathbb{R}^2$ and some contraction constant $0 < \kappa < 1$. Then there exist small positive constants ε_0, ε_1 such that for all $0 < \varepsilon < \varepsilon_0$, $|\lambda| < \varepsilon_1$ and for all choices $\delta > 0$ the following holds.*
Let $x(\vartheta, y) < B^{\text{in}}_\delta$ be a point which returns to B^{in}_δ within U' after $n \geq 2$ iterates, that is

$$\Phi^n(\varepsilon, \lambda, x) \in B^{\text{in}} .$$

Then $\Phi^n(\varepsilon, \lambda, \cdot)$ is a local area contraction, that is

$$|\det D_x(\Phi^n(\varepsilon, \lambda, x))| \leq \kappa < 1 . \tag{5.7}$$

Proof. We have assumed that the "saddle quantity" $\nu - \mu$ is negative. Our claim on area contraction is well known, under this assumption. We still give a proof, setting up some more notation. The basic idea is to make U' so small that a large number of area contracting iterations within the unit box $B_1 = \{|\vartheta|, |y| \leq 1\}$ compensates for all possibly expanding iterations outside of the unit box B_1.

Consider the flow for $\varepsilon = 0$ first. Fixing U' small, the time an orbit can spend in $U' \setminus B_1$ is uniformly bounded above. Likewise, an area can expand at most by a factor less than some constant c under the linearized flow $\dot{y} = D_x f(\lambda, x) y$ during that time. By approximation, the same holds true for $0 < \varepsilon < \varepsilon_0$ and the linearization of iterations in $U' \setminus B_1$ of the discretization $\Phi(\varepsilon, \lambda, \cdot)$.

It remains to establish area contraction within the unit box B_1 by a factor at least κ/c, by choosing U' and $\varepsilon_0 = \varepsilon_1$ small enough. Fix constants $\overline{\mu}, \overline{\nu}$ such that

$$0 < \nu < \overline{\nu} < \overline{\mu} < \mu . \tag{5.8}$$

By approximation, we may choose ε_0 such that

$$0 < \det D_x \Phi(\varepsilon, \lambda, x) < 1 - \varepsilon(\overline{\mu} - \overline{\nu}) \tag{5.9}$$

for all $0 < \varepsilon < \varepsilon_0$, $|\lambda| < \varepsilon_0$, $x \in B_1$. Given any "time" $s > 0$ we may also choose U' so small that any $x \in B^{\text{in}} \cap U'$ spends at least

$$n_s := [s/\varepsilon] \tag{5.10}$$

iterates of $\Phi(\varepsilon, \lambda, \cdot)$ in the unit box B_1 before leaving B_1. In view of (5.9), this implies an area contraction by a factor at least

$$(1 - \varepsilon(\overline{\mu} - \overline{\nu}))^{n_s} < e^{-(\overline{\mu} - \overline{\nu})(s - \varepsilon)} . \tag{5.11}$$

Choosing s large, a uniform contraction factor κ/c can easily be achieved. This proves the lemma. ∎

Lemma 5.2 *Let U', ε_0, ε_1 be fixed as in Lemma 5.1 and consider $0 < \varepsilon < \varepsilon_0$, $|\lambda| < \varepsilon_1$. Then $\Phi(\varepsilon, \lambda, \cdot)$ possesses at most one closed invariant Jordan curve in U'.*

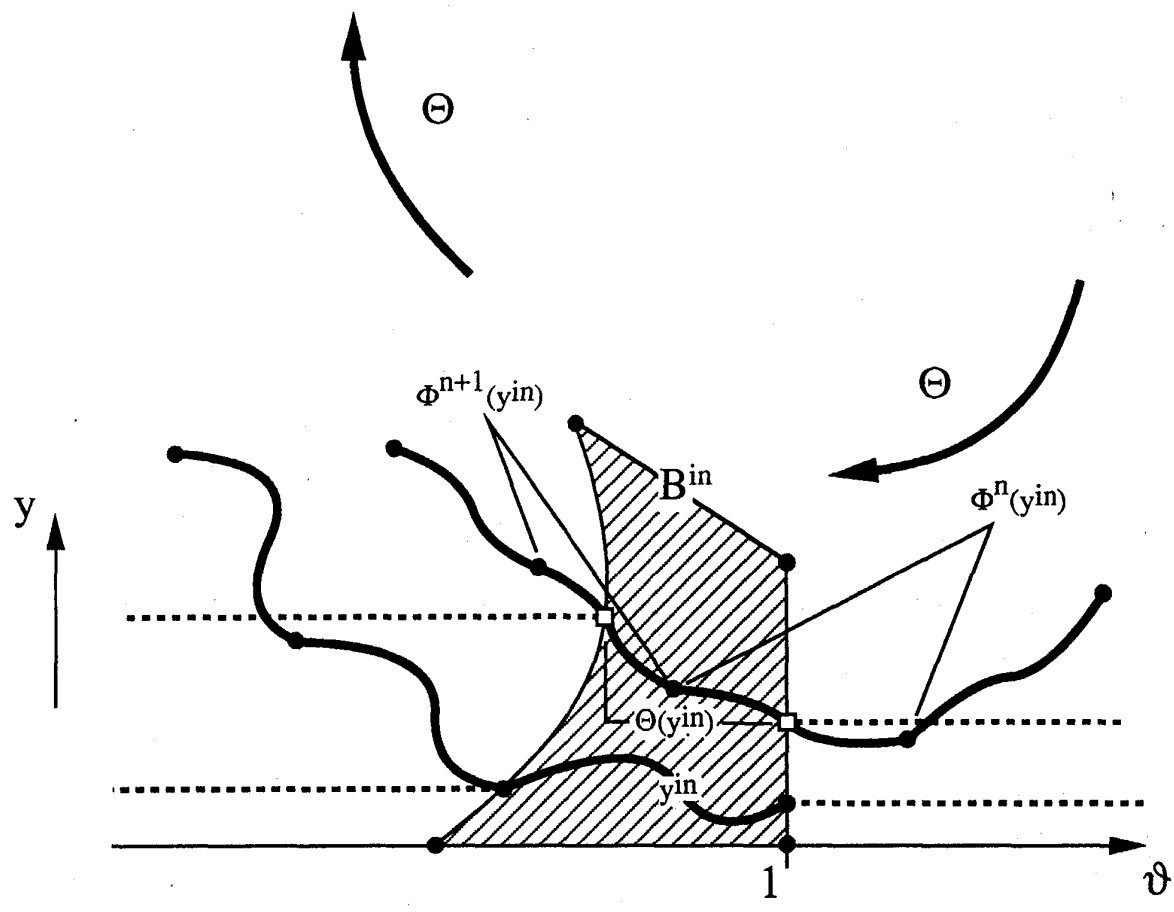


Figure 5.2: The graph transform Θ.

Proof. Suppose there are two such curves. By invariance, both curves intersect B_1. Again by invariance, we can choose δ such that both curves intersect the section B^{in}. Let J denote the part of the region between the two curves which intersects B^{in}. Any point of J eventually returns to J under both forward and backward iteration. By Lemma 5.1, area decreases locally, under this return map, by a factor at least $\kappa < 1$. This implies

$$\text{area } (J) \leq \kappa \cdot \text{area}(J) = 0 \,, \tag{5.12}$$

even though the number of iterates required for a return may differ from point to point. Therefore, the two curves must coincide in B^{in} and, a forteriori, everywhere in U'. This proves the lemma. ∎

Lemma 5.3 *Suppose ε, λ, δ and k are chosen such that the graph transform Θ, introduced in* (5.6), *is well-defined and maps the compact set Λ of graphs in B^{in} into itself (see Lemmata* 5.4–5.6 *below). Then Θ possesses a fixed point in Λ. If $B^{\text{in}} \subseteq U'$ then this fixed point is also unique.*

Proof. As we have already noticed above, Λ is homeomorphic to a convex subset of C^0. Since Λ is also compact, Schauder's fixed point theorem (see e.g. DEIMLING (1985)) implies that Θ possesses a fixed point, provided that Θ is known to be continuous on Λ. Uniqueness then follows from Lemma 5.2.

To prove continuity of Θ, we consider some graph $(\vartheta, y(\vartheta))$ in Λ. To determine its image under Θ, it suffices to trace the iterates under Φ of its piece y^{in} in B^{in}. Since the left intersection with the box B^{in} is the image under Φ of the right end point $(1, y(1))$, the forward iterates of that piece define an extended Lipschitz continuous curve which may reenter the B^{in}–box, say, after a finite number n of iterations of Φ. After $n+1$ iterations, the extended curve leaves B^{in} again, at least partially. Now perturb y^{in} slightly to become $\tilde{y}^{\text{in}}$, and extend forward correspondingly. Since the extended $\tilde{y}$–curve stays uniformly close to the extended y–curve, over the first $n+1$ iterates of Θ, their Θ–images in B^{in} are also close. This proves continuity of Θ and, thereby, the lemma. ■

As we have outlined above, we complete the construction of the unique closed invariant Lipschitz curve in U for $0 < \varepsilon < \varepsilon_0$, $C' e^{-2\pi\eta/\varepsilon} < \lambda < \varepsilon_1$ in Lemmata 5.4–5.6 below. In particular we show that indeed the graph transform Θ is well–defined and maps Λ into itself, for some uniform ε_0, ε_1 and suitable choices of δ, k. In Lemma 5.4 we study return slopes of W^u_{loc} in B^{in}.

Here and below let $\preceq$ denote an inequality which holds up to some suitable constant positive factor; $\approx$ indicates that both $\preceq$ and $\succeq$ hold with possibly different constants.

Lemma 5.4 *As in* (5.3), (5.4) *let* $y = d(\varepsilon, \lambda, \vartheta)$, $\frac{1}{2} \le \vartheta \le 1$, *describe* W^u_{loc} *in* B^{in}. *Then*

$$d(\varepsilon = 0, \lambda, \vartheta) = \lambda \cdot \vartheta^{-\nu(\lambda)/\mu(\lambda)} \tag{5.13.a}$$

$$|d(\varepsilon, \lambda(\varepsilon) = 0, \vartheta)| \preceq e^{-2\pi\eta/\varepsilon} \tag{5.13.b}$$

$$|d_\vartheta(\varepsilon, \lambda(\varepsilon) = 0, \vartheta)| \preceq e^{-2\pi\eta/\varepsilon} \tag{5.13.c}$$

Both statements hold for $\varepsilon_0, \varepsilon_1$ *small enough and for all* $0 \le \varepsilon < \varepsilon_0$ $|\lambda| < \varepsilon_1$, $\frac{1}{2} \le \vartheta \le 1$. *In particular*

$$|d(\varepsilon, \lambda, \vartheta) - \lambda| < |\lambda|/2\,, \tag{5.14.a}$$

if

$$|\lambda| \succeq e^{-2\pi\eta/\varepsilon}\,. \tag{5.14.b}$$

Likewise, for all $k > 0$ *small enough*

$$\operatorname{sign}\lambda \cdot d_\vartheta(\varepsilon, \lambda, \vartheta) < -2k \tag{5.14.c}$$

provided that

$$|\lambda| \succeq k + e^{-2\pi\eta/\varepsilon}\,. \tag{5.14.d}$$

Proof. Claim (5.13.a) follows directly from our normalizations (5.2) and (5.4). Claims (5.13.b,c) follow from Theorem 1.1, see (1.15.b,c).

To prove (5.14.c,d) we compute

$$d_{\lambda\vartheta}(\varepsilon = 0, \lambda, \vartheta) = -\tfrac{\nu}{\mu}\vartheta^{-\frac{\nu}{\mu}-1} + O(\lambda)$$

where ν, μ are evaluated at $\lambda = 0$ and the error estimate is uniform. In particular, in our notation

$$d_{\lambda\vartheta}(\varepsilon, \lambda, \vartheta) \approx -1\,,$$

for all ε, λ, ϑ used here. Together with the exponential estimate (5.13.c) this proves that (5.14.d) implies (5.14.c). Similarly, (5.13.a,b) entail that (5.14.b) implies (5.14.a) . This proves the lemma. ■

For later use we note here that the slope of the top boundary of the box B^{in} is negative, for a similar reason.

Loosely speaking, the λ–Lemma states that transverse sections to W^s_{loc} approach W^u locally uniformly in C^1, under forward iteration. We derive a more quantitative version next.

Lemma 5.5 *Let δ_0 be chosen small enough and consider $x = (\vartheta, y) \in B^{\text{in}}_\delta$ for some $0 < \delta < \delta_0$. Suppose x re-enters B^{in}_δ, after n iterates of $\Phi(\varepsilon, \lambda, \cdot)$, at $x' = (\vartheta', y')$. Let σ denote the slope of a curve $y = y(\vartheta)$ through x. Denote by σ' the slope at x' of the corresponding image under Φ^n.*

Then

$$|y' - d(\varepsilon, \lambda, \vartheta')| \preceq y^{\bar\mu/\nu} \tag{5.15.a}$$

$$|\sigma' - d_\vartheta(\varepsilon, \lambda, \vartheta')| \preceq y^{\bar\mu/\nu}\left(1 + \frac{y}{|\sigma|}\right) \tag{5.15.b}$$

where $\bar\mu$, $\bar\nu$ are fixed suitably such that $0 < \nu < \bar\nu < \bar\mu < \mu$.

Proof. Let m denote the maximal number of iterates that $x = (\vartheta, y)$ spends in the unit box B_1, initially. Let $x^{\text{out}} := (\vartheta^{\text{out}}, y^{\text{out}})$ denote the m–th iterate of x, and

$$\sigma^{\text{out}} := d\vartheta^{\text{out}}/dy^{\text{out}}$$

the corresponding slope at x^{out} induced by Φ^m, applied to the curve $y(\vartheta)$. By the same reasoning as in the proof of Lemma 5.1 it is sufficient to show

$$|\vartheta^{\text{out}}| \le y^{\bar\mu/\nu} \tag{5.16.a}$$

$$\sigma^{\text{out}} \le y^{1+\bar\mu/\nu}\sigma^{-1}\,, \tag{5.16.b}$$

and ignore the return map from B^{out} to B^{in} altogether.

To prove (5.16.a) note that

$$y \ge e^{-\nu\varepsilon(m+1)}\,,$$

similarly as for the limiting ODE at $\varepsilon = 0$. Analogously, we obtain

$$0 \le \vartheta^{\text{out}} \le e^{-\bar\mu\varepsilon m} \le y^{\bar\mu'/\nu} \tag{5.17.a}$$

with $\bar\mu' := \bar\mu m/(m+1)$. This proves (5.16.a) with $\bar\mu'$ instead of $\bar\mu$.

To prove (5.16.b) we also fix $0 \le \underline{\nu} < \nu$ and argue analogously for the slopes

$$|\sigma^{\text{out}}| = |d\vartheta^{\text{out}}/dy^{\text{out}}| \le \frac{e^{-\bar\mu\varepsilon m}}{e^{\underline{\nu}\varepsilon m}} \cdot \sigma^{-1}$$

$$|\sigma^{\text{out}}| \le y^{(\bar\mu+\underline{\nu})/\nu} \le y^{1+\bar\mu'/\nu}\sigma^{-1} \tag{5.17.b}$$

again for some new constants $\bar\nu < \bar\mu' < \bar\mu$. Since $\bar\mu$, $\bar\nu$, $\underline{\nu}$ were fixed arbitrarily in the appropriate ranges, we may omit primes in (5.17.a,b). This proves (5.16.a,b) and the lemma. ■

Lemma 5.6 *Fix $\bar\mu$, $\bar\nu$ such that Lemma 5.5 holds. There exist constants ε_0, ε_1, $C' > 0$ such that for $0 < \varepsilon < \varepsilon_0$, $C'e^{-2\pi\eta/\varepsilon} < \lambda < \varepsilon_1$ the following holds.*

There exist δ, $k > 0$, depending on $\bar\mu$, $\bar\nu$, ε, λ, such that the set Λ is nonempty and the continuous graph transform Θ, defined in (5.6), maps Λ into itself. In particular, $\Phi(\varepsilon, \lambda, \cdot)$ admits a closed, invariant Lipschitz Jordan curve γ.

Proof. We first show that Λ is not empty; specifically, $W^u_{\rm loc}$ defines an element of Λ. Let $\alpha := \frac{1}{2}(1+\overline{\mu}/\overline{\nu}) > 1$ and define

$$\begin{aligned} \delta &:= 2\lambda \\ k &:= \delta^\alpha . \end{aligned} \tag{5.18}$$

Consider only

$$\lambda \succeq e^{-2\pi\eta/\epsilon} \tag{5.19}$$

and choose ε_1 so small that in addition

$$(2\lambda)^\alpha = \delta^\alpha = k \preceq \lambda . \tag{5.20}$$

for all $0 < \lambda < \varepsilon_1$. The constants of $\preceq$ in (5.19), (5.20) will be chosen such that (5.14.b,d) hold; in particular

$$\lambda \succeq k + e^{-2\pi\eta/\epsilon} . \tag{5.21}$$

Then (5.14.c) implies

$$-\frac{1}{2k} < d_\vartheta(\varepsilon,\lambda,\vartheta) < -2k . \tag{5.22}$$

(Note that the left inequality holds trivially for small enough $\varepsilon_1 > \lambda = \frac{1}{2}k^{1/\alpha}$.) Also, (5.14.a), (5.18) imply

$$0 < \tfrac{1}{2}\lambda < d(\varepsilon,\lambda,\vartheta) < \tfrac{3}{2}\lambda = \tfrac{3}{4}\delta . \tag{5.23}$$

By (5.22), (5.23) the returning segment of $W^u_{\rm loc}$ described by d defines an element of Λ. In particular, Λ is not empty.

We now invoke the λ-Lemma 5.5 to prove that Λ is also invariant under Θ. Consider $y = y(\vartheta)$ in Λ. Then the slopes σ of y in $B^{\rm in}$ are between $-\frac{1}{k}$ and $-k$, and $0 \le y(\vartheta) \le 2\delta$ can be assumed to hold for all ϑ. By (5.15.a), this implies (for small δ)

$$|y' - d(\varepsilon,\lambda,\vartheta')| \preceq \delta^{\mu/\nu} < \delta/4 \tag{5.24}$$

for the return points (ϑ', y'). Thus the extended Lipschitz curve does reenter $B^{\rm in}$, in view of (5.23). Moreover, the slopes σ' of $y'(\vartheta')$ in $B^{\rm in}$ can be estimated by (5.15.b).

$$\begin{aligned} |\sigma' - d_\vartheta(\varepsilon,\lambda,\vartheta')| &\preceq \delta^{\mu/\nu} + \tfrac{\delta^{2\alpha}}{|\sigma|} \\ &\le \delta^{\mu/\nu} + \tfrac{\delta^{2\alpha}}{k} \preceq \delta^\alpha = k . \end{aligned} \tag{5.25}$$

In particular, choosing ε_1 small enough, this implies

$$-\frac{1}{k} \le \sigma' \le -k .$$

Therefore, Θ indeed maps Λ into itself, and the lemma is proved. ∎

So far, we have constructed a unique fixed point of Θ in Λ, corresponding to a unique closed, invariant, Lipschitz Jordan curve γ, for $\lambda > C'e^{-2\pi\eta/\epsilon}$. We now show that orbits in some neighborhood U of Γ either converge to zero, or to γ, or else leave U, in forward time.

Lemma 5.7 *There exists a neighborhood U of Γ in $\mathbb{R}^2$ and some constants ε_0, ε_1, $C' > 0$ as in Lemma 5.6 such that for $0 < \varepsilon < \varepsilon_0$, $C'e^{-2\pi\eta/\epsilon} < \lambda < \varepsilon_1$ the following alternative holds for the forward iterates x_n, $n \ge 0$ of $x = x_0 \in U$ under $\Phi(\varepsilon,\lambda,\cdot)$.*

Either

$$\lim_{n\to\infty} x_n = 0 , \quad \textit{or} \tag{5.26.a}$$

$$\lim_{n\to\infty} \operatorname{dist}(x_n,\gamma) = 0 , \quad \textit{or} \tag{5.26.b}$$

$$x_n \textit{ leaves } U, \textit{ for some } n . \tag{5.26}$$