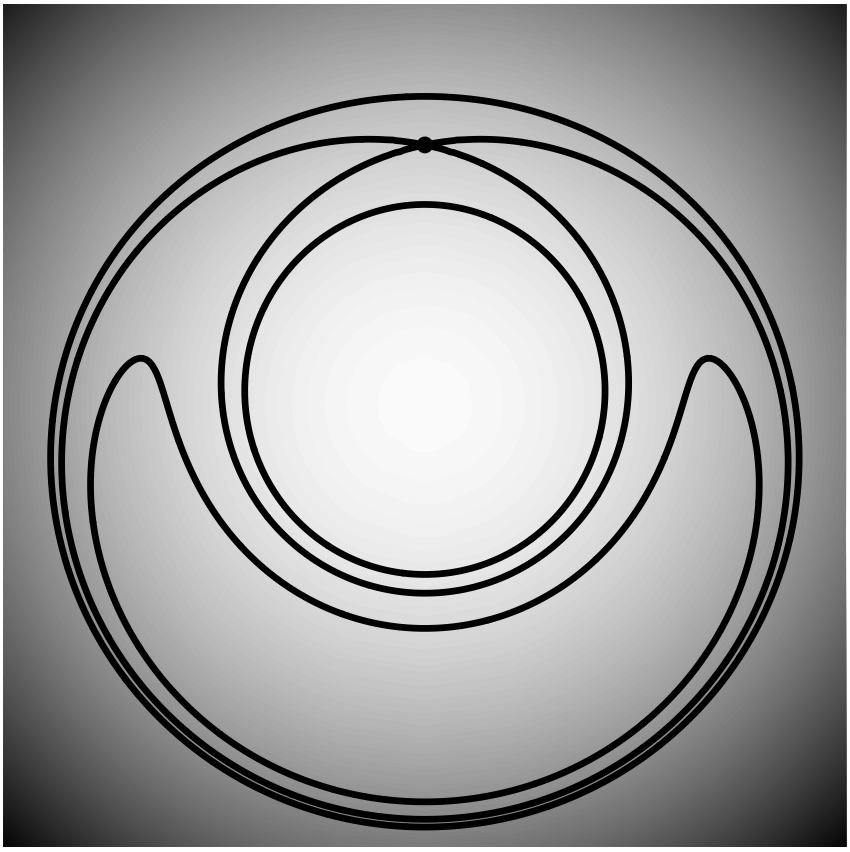
Chaotic behaviour of quasi-periodically driven Helmholtz oscillators

Femius Koenderink

May 1999



Supervisors: Prof. Dr. A. Doelman (UvA)
Dr. L. R. M. Maas (NIOZ)

Universiteit Utrecht

Chaotic behaviour of quasi-periodically driven Helmholtz oscillators

Femius Koenderink

April 1999

Supervisors: Prof. Dr. A. Doelman (UvA)
Dr. L. R. M. Maas (NIOZ)

Universiteit Utrecht

This report was typeset using LaTeX-2ε and Adobe-postscript fonts. Illustrations were produced using Mathematica 3.0 and CorelXara 1.1.

Abstract

A basin with sloping walls which is connected to the sea by a narrow strait responds nonlinearly to tidal forcing. An ordinary differential equation is used to describe the response. A small amplitude approximation to this weakly damped nonlinear oscillator equation is calculated using second order averaging in the case of weak periodic forcing close to the primary resonance. The averaged system, which is (up to scaling) the same for all surface area to depth relations, possesses multiple equilibria and two orbits homoclinic to one saddle point in the hamiltonian case without friction. The Melnikov method is employed and extended to study different types of homoclinic chaos which arise from homoclinic intersections if an extra perturbative and nearly resonant forcing is introduced. A complete classification of the different types of behaviour is accompanied by a geometrical discussion.

Contents

Introduction

Nonlinear ordinary differential equations are essential to the modeling of all sorts of phenomena in exact sciences. In contrast to the theory of linear differential equations there are few constructive 'receipes' to obtain solutions of nonlinear equations. Most of the theory of nonlinear differential equations focuses on the topological properties of the phase space, e.g the existence and stability of stationary solutions, the existence and stability of periodic solutions, et cetera. A vast range of qualitatively different phenomena may occur, which can only be studied on a case by case basis by restricting the attention to one differential equation only, or to a small family of similar differential equations. In some weakly nonlinear cases asymptotic methods may be used to approximate the solutions of the system. A very special category of nonlinear systems is the class of hamiltonian systems, which possess a first integral. In the case of a system with two degrees of freedom the orbits of the system are simply the level sets of the hamiltonian. Such hamiltonian systems often arise in the context of classical mechanics in systems without friction.

A fundamental part of the field of nonlinear differential equations is the theory of chaotic dynamics. I will not try to provide a precise definition of the term chaos, because it is one of those concepts in mathematics which is excessively hard to define. In a sense it seems to be one of those phenomena in mathematics which are 'discovered' and not 'invented'. This explains the enormous public interest in very much related subjects like fractals and the beauty of the Mandelbrot set or the Julia set. Many nonlinear hamiltonian systems which are perturbed by a time dependent forcing exhibit chaotic behaviour. This is a fact which has far reaching consequences in physical applications. At the same time it poses a fundamental problem to the mind of a physicist. A physical system in classical mechanics is deterministic, since it is described by evolution equations and initial conditions which guarantee a unique solution. Apparently it is a fundamental property of such deterministic systems to exhibit chaotic behaviour. Many physicists argue that if the initial conditions of a system can be measured with sufficient accuracy the state of the system can be predicted for as long a time span as one wants. This may be true from a mathematical point of view but is at the same time a physically irrelevant remark. It will for instance never be possible to measure the global atmospheric condition with sufficient accuracy to predict the weather for an arbitrarily long period.

Chaotic dynamics is relevant in a wide area of applications, for instance the theory of chemical reactions, population dynamics in biology as well as electric circuit theory and classical mechanics in physics. The last category again contains a variety of widely disparate subjects, such as celestial mechanics, the theory of nonlinear oscillators, fluid dynamics (for instance turbulence), atmospheric dynamics and tidal dynamics. The work which is presented in this report touches on the last field, since it analyzes a nonlinear system which originates from the description of tidal motions in nearly enclosed basins which are connected to the sea by a narrow strait. Observations of highly irregular tides in basins have been reported already from the start of this century. An important aspect of these irregular tides is that the sea elevation in such basins may be very small but are accompanied by large currents through the strait. Apparently even the dynamics of small amplitude elevations, which are studied in this report are physically relevant because sediment and nutrient transport is affected by the resulting currents.

On the one hand an attempt has been made to write an essay readable to fellow students who have followed a basic course on nonlinear dynamical systems. Therefore, a short summary of the averaging techniques in chapter 2 is provided, as well as some background concerning the Melnikov method employed in chapter 4 and the dynamics studied in chapter 5. On the other hand I do hope to have made a valuable contribution to the work on this subject initiated by my supervisors, Arjen Doelman and Leo Maas. Thanks are especially due to Arjen for his helpful remarks and the stimulation needed to obtain the presented results,

Femius Koenderink

CHAPTER 1

PHYSICAL MODEL

The differential equation which is discussed in the subsequent chapters originates from an application in oceanography, as reported in [1]. It is well known that tidal elevations can be amplified due to coastal geometry. A famous example is the Bay of Fundy where tidal elevations up to eleven meters are reported [2]. This enormous amplitude should be compared with roughly half a meter which would be the surface elevation if the whole earth would be covered by water [2].

The extreme elevation amplifications are caused by resonant behaviour, which occurs if typical length scales of the coastal geometry match with the length scale of tidal waves. Such resonances exist in narrow sea straits, bays and fjords, closed basins and nearly closed basins, which are typically regions confined between respectively two, three, and four coasts. The oscillatory behaviour of the water level and currents in straits, bays and fjords is driven by the tidal elevations of the sea to which these basins are connected. These co-oscillatory tides should be contrasted to those of a totally enclosed basin in which case tidal movements can only be caused by the direct gravitational interaction with the moon and sun [2],[3].

In bays and fjords the resonant modes are similar to those of half-open organ pipes, e.g. modes where the basin length is a quarter of the tidal wavelength (determined by $\lambda \approx T\sqrt{gH}$, where T is the tidal period (12 h 25 min. for the tide generated by the moon), g is the gravitational acceleration and H is the basin depth). The aforementioned Bay of Fundy for instance, has dimensions which (nearly) result in such a quarter wavelength resonance. Water is transferred from the sea into the basin and back again every tidal period.

In contrast a closed basin is characterized by eigen modes which have the property that the total amount of water in the basin is conserved. This is obvious from the fact that no water can be exchanged with external reservoirs. Consequently, only modes with at least one line of zero elevation exist; water is exchanged between parts of the basin on different sides of these zero elevation lines. In many real cases the direct tidal forcing is not resonant with any of the eigen modes of a basin (for instance the Black Sea, [2]) because the basin lengths are much smaller than a tidal wavelength.

If a basin is nearly enclosed, but connected to the sea by a narrow strait these modes remain, and furthermore a new mode is added. This so called Helmholtz mode [4] is characterized by a periodic exchange of water between basin and sea and is driven through the pressure difference over the strait due to differences in tidal elevations of the sea and the basin. In the case of a basin which is short and narrow the surface elevation of this mode is spatially uniform. Roughly the dynamics is ruled by mass conservation, which states that the rate of change of the amount of water in the basin equals the influx of water, and the momentum equation for the current through the strait. In terms of the excess volume v of water in the basin (e.g. the amount of water in excess of the water present in equilibrium (zero elevation)) mass conservation reads

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -u_* bH$$

where H is the depth of the channel, b the width and u_* the depth-averaged flow through the strait (directed out of the basin for positive sign). The geometry is illustrated by figure 1.1 This equation is valid if the strait is narrow enough to assume a uniform flow in the strait and if the surface elevation in the strait is small compared to the depth H.

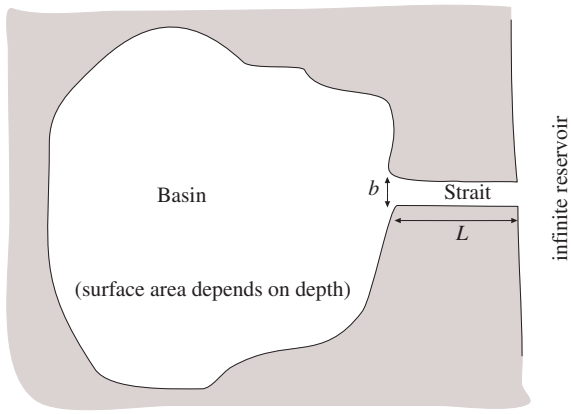


FIGURE 1.1: *A basin is considered with a surface area that varies as a function of the vertical distance to the deepest point of the basin. The basin is connected to a reservoir containing an infinite amount of water (e.g. the sea) through a narrow strait with horizontal dimensions L and b. The surface elevation of the sea is assumed to be a known function of time, not influenced by the dynamics of the basin.*

The average flow u_* is obtained by integrating the one-dimensional linearized momentum equation for the current velocity $u(x)$ through the strait in terms of the free surface elevation ζ and the gravitational acceleration g

$$\frac{\partial u}{\partial t} + g\frac{\mathrm{d}\zeta}{\mathrm{d}x} = 0$$

over the length L of the channel, resulting in

$$L\frac{\mathrm{d}u_*}{\mathrm{d}t} + g(\zeta_e(t) - \zeta(v)) = 0.$$

The elevation $\zeta_e(t)$ is the elevation of the sea, which is supposed to be a prescribed function of time while $\zeta(v)$ is the surface elevation in the basin, dependent on the excess volume v. This expression shows that (under these assumptions) the flow is directly driven by the pressure gradient that results from the elevation difference.

By differentiating the first equation with respect to t once and substituting the second equation the ordinary differential equation

$$\frac{\mathrm{d}^2 v}{\mathrm{d}t^2} + \gamma\zeta(v) = \gamma\zeta_e(t)$$

is obtained where $\gamma = bHg/L$. If the basin has vertical walls and a surface area A_0 the surface elevation $\zeta(v)$ as a function of excess volume v is simply $\zeta(v) = v/A_0$. In this case the dynamics is that of a driven harmonic oscillator

$$\frac{\mathrm{d}^2 v}{\mathrm{d}t^2} + \sigma_H^2 v = \gamma\zeta_e(t). \tag{1.1}$$

with eigen frequency (in this case called Helmholtz frequency) $\sigma_H = \sqrt{bHg/(LA_0)}$. This frequency is often smaller than the eigen frequencies of the mass-preserving modes (sloshing modes) and closer to the frequency of the tidal forcing. If the Helmholtz frequency is close to the tidal frequency the Helmholtz mode may be amplified, while the sloshing modes are choked because they are driven far away from resonance. The relevance of the Helmholtz mode to the Wadden Sea is addressed by [1] and the mechanism may be used to explain the tidal elevations in several other basins. Since the Wadden Sea possesses many tidal flats and drying banks (see [5]) it is important to incorporate sloping basin sidewalls in the model.

In a more general case the free surface elevation $\zeta(v)$ as a function of the excess volume v is defined implicitly through

$$v = \int_0^\zeta A(z)\,\mathrm{d}z$$

where $A(z)$ is the basins surface area at depth z measured with respect to the equilibrium elevation at $z = 0, A(0) = A_0$. For general $A(z)$ the restoring term $\zeta(v)$ is nonlinear. For small amplitudes $\zeta(V)$ can be approximated as $\zeta(0) + v\frac{\mathrm{d}\zeta}{\mathrm{d}v}(0)$ which simply equals v/A_0. To first order in v the Helmholtz frequency of a general basin is therefore still specified by $\sqrt{bgH/(LA_0)}$.

Apart from the assumptions mentioned earlier, effects of earth rotation were neglected, as well as those of frictional effects, such as bottom friction and radiation damping (loss of energy through waves radiated from the basin into the sea). To incorporate some of the aspects of damping a linear (viscous) damping term $-c\frac{\mathrm{d}v}{\mathrm{d}t}$ is added to the right hand side of (1.1). Though bottom friction, which is quadratic in the current velocity, may be important in applications the discussion is limited to linear velocity friction to simplify the calculations and to restrict the interest to the effects of nonlinearity of the restoring term.

This system has been studied in the special case of a surface area A which increases linearly as a function of the water level by for instance [1], [6], [7].

A more rigorous approach, including quadratic damping, to the derivation of (1.1) is provided by [1]. By employing a rescaling $t \to t/\sigma_H$, $\zeta \to \zeta H$ and $A(z) \to A_0 A(zH)$ the equation is cast into the dimension-less form

$$\frac{\mathrm{d}^2 v}{\mathrm{d}t^2} + \zeta(v) = \zeta_e(t) - C\frac{\mathrm{d}v}{\mathrm{d}t}$$

which is a standard form for a nonlinear damped and forced oscillator[1]. It should be noted that the model imposes physical limits on v and ζ; because the basin can not contain a negative amount of water the excess volume is limited to $v > \int_0^{-1} A(z)\,\mathrm{d}z$, which corresponds to the restriction $\zeta(v) > -1$.

Because of the general form of this equation the model is applicable in many other contexts, including for instance the forced Duffing equation.

[1]Note that the sign of the surface elevation and the excess volume are always the same, which means that the $\zeta(v)$ term is always restoring.

CHAPTER 2

AVERAGING

2.1 Introduction

2.1.1 Full system

As a simple model for waves in a short basin the equation

$$\frac{\mathrm{d}^2 v}{\mathrm{d}t^2} + \zeta(v) = f(t) - c\frac{\mathrm{d}v}{\mathrm{d}t} \tag{2.1}$$

was derived, after nondimensionalizing, in chapter 1. The function $\zeta(v)$ represents the elevation of the free surface with respect to the equilibrium situation as a function of the excess volume $v(t)$, which is the amount of water in the basin at time t minus the amount of water which the basin contains when the free surface is located at $\zeta = 0$. In the nondimensionalized system the bottom of the basin is located at a depth $z = -1$, while the surface area of the basin was scaled to equal unity at $z = 0$. In the case of a sloping bottom or sloping sides the restoring term $\zeta(v)$ is nonlinear. It can be calculated from the hypsometry (area-to-depth relation) specific to the basin considered through

$$v(\zeta) = \int_{z=0}^{\zeta} A(z)\,\mathrm{d}z \tag{2.2}$$

where the scaled surface area $A(z)$ as a function of depth z satisfies $A(-1) = 0, A(0) = 1$. By inverting this relation the restoring term $\zeta(v)$ is found. The forcing $f(t)$ which will be considered is assumed to be periodic (or quasi-periodic) with period ω. Finally the constant $c \geq 0$ is a linear friction coefficient, possibly obtained after linearization of damping in the physical model.

2.1.2 Small amplitude oscillations

In the remaining sections the special case of small amplitude oscillations will be considered. These are expected in the case of weak forcing and weak damping. The forcing term is therefore taken as

$$f(t) = \epsilon^3 F \cos(\omega t + \theta) \tag{2.3}$$

where θ is some phase angle and $0 < \epsilon \ll 1$. In imitation of [7] the coefficient of friction c is scaled as $c = \epsilon^2 C$. The assumption of small amplitude oscillations of the excess volume v is incorporated into the model by writing $v = \epsilon V$. This scaling was chosen because the effects of nonlinearity, friction and forcing will appear at the same order in the final evolution equations.

If $A(z)$ is sufficiently smooth the elevation $\zeta(v)$ of the free surface can be expanded in a power series in V centered at $V = 0$:

$$\zeta(V) = \sum_{n=0}^{N} \frac{\mathrm{d}^n \zeta}{\mathrm{d}v^n}(0)\frac{\epsilon^n V^n}{n!} + O(\epsilon^{N+1})$$

From the definition of $v(\zeta)$ in Eq. (2.2) which implies $v(\zeta=0)=0$ it is concluded that $\zeta(v=0)=0$ (zero excess volume when the basin is in equilibrium). Furthermore the choice $A(z=0)=1$ (after nondimensionalizing) implies

$$\left.\frac{\mathrm{d}\zeta}{\mathrm{d}v}\right|_{v=0}=\left(\left.\frac{\mathrm{d}v}{\mathrm{d}\zeta}\right|_{\zeta=0}\right)^{-1}=\left(\left.\frac{\mathrm{d}}{\mathrm{d}\zeta}\int_0^{\zeta}A(z)\,\mathrm{d}z\right|_{\zeta=0}\right)^{-1}=\left(\left.A(\zeta)\right|_{\zeta=0}\right)^{-1}=1.$$

This result reflects the fact that the rescaling procedure employed to obtain $A(z=-1)=0$ and $A(z=0)=1$ implied a rescaling of the first order linear eigenfrequency, i.e. the Helmholtz frequency σ_H, to 1. Retaining only the first four terms in the power expansion of $\zeta(V)$ results in the approximation

$$\epsilon\frac{\mathrm{d}^2V}{\mathrm{d}t^2}+\epsilon V=-\epsilon^2\frac{1}{2}\left.\frac{\mathrm{d}^2\zeta}{\mathrm{d}v^2}\right|_{v=0}V^2-\epsilon^3\frac{1}{6}\left.\frac{\mathrm{d}^3\zeta}{\mathrm{d}v^3}\right|_{v=0}V^3-\epsilon^3C\frac{\mathrm{d}V}{\mathrm{d}t}+\epsilon^3F\cos(\omega t+\theta)+O(\epsilon^4)$$

to equation (2.1). For later convenience the abbreviations

$$\alpha=-\frac{1}{2}\left.\frac{\mathrm{d}^2\zeta}{\mathrm{d}v^2}\right|_{v=0}\quad\text{and}\quad\beta=-\frac{1}{6}\left.\frac{\mathrm{d}^3\zeta}{\mathrm{d}v^3}\right|_{v=0}\tag{2.4}$$

are slipped in to obtain

$$\frac{\mathrm{d}^2V}{\mathrm{d}t^2}+V=\epsilon\alpha V^2+\epsilon^2\beta V^3-\epsilon^2C\frac{\mathrm{d}V}{\mathrm{d}t}+\epsilon^2F\cos(\omega t+\theta)+O(\epsilon^3)\tag{2.5}$$

which is a small amplitude approximation to Eq. (2.1).

2.1.3 Transforming to standard form

Without loss of generality polar coordinates $R(t)$ and $\Phi(t)$ can be introduced by defining them through the implicit definition

$$\begin{aligned}V(t)&=R(t)\cos(\omega t-\Phi(t))\\ \frac{\mathrm{d}V}{\mathrm{d}t}&=-\omega R(t)\sin(\omega t-\Phi(t))\end{aligned}\tag{2.6}$$

where $0\leq R(t)<\infty$ and $\Phi(t)\in\mathbb{R}/2\pi$. This is a well known *phase-amplitude transformation* which is used to bring Eq. (2.5) in the standard form discussed in section 2.2. This transformation, which is proposed in for instance [8] and [9] exploits the fact that the solution of the linear zeroth order equation obtained from Eq. (2.5) by setting $\epsilon=0$ is of the form $r\cos(t-\phi)$ where r,ϕ are constant. One might wonder why the frequency ω was slipped into (2.6). Indeed, this is not very useful unless the additional assumption is made that $\omega=1+\epsilon^2\sigma$, where σ is a detuning parameter. The physical interpretation of this choice is that only forcing near the primary resonance frequency is considered.

Two independent equations were introduced to define the two independent variables R and Φ. However, for this definition to make sense it is obviously required that the consistency relation

$$\frac{\mathrm{d}R}{\mathrm{d}t}\cos(\omega t-\Phi(t))+R(t)\frac{\mathrm{d}\Phi}{\mathrm{d}t}\sin(\omega t-\Phi(t))=0\tag{2.7}$$

is satisfied, which expresses the constraint that the time derivative of $V(t)$ as defined in Eq. (2.6) has to match with $-\omega R(t)\sin(\omega t-\Phi(t))$ as required by the second formula in Eq. (2.6).

Upon substitution of Eq. (2.6) the left-hand side of equation (2.5) reduces to

$$\begin{aligned}
\frac{\mathrm{d}^2V}{\mathrm{d}t^2}+V &= \frac{\mathrm{d}}{\mathrm{d}t}[-\omega R(t)\sin(\omega t-\Phi(t))]+R(t)\cos(\omega t-\Phi(t)) \\
&= -\omega^2 R(t)\cos(\omega t-\Phi(t))-\omega\frac{\mathrm{d}R}{\mathrm{d}t}\sin(\omega t-\Phi(t)) \\
&\qquad +\omega R(t)\frac{\mathrm{d}\Phi}{\mathrm{d}t}\cos(\omega t-\Phi(t))+R(t)\cos(\omega t-\Phi(t)) \\
&= \omega\left[-\frac{\mathrm{d}R}{\mathrm{d}t}\sin(\omega t-\Phi(t))+R(t)\frac{\mathrm{d}\Phi}{\mathrm{d}t}\cos(\omega t-\Phi(t))\right] \\
&\qquad +(1-\omega^2)R(t)\cos(\omega t-\Phi(t)) \\
&= \omega\left[-\frac{\mathrm{d}R}{\mathrm{d}t}\sin(\omega t-\Phi(t))+R(t)\frac{\mathrm{d}\Phi}{\mathrm{d}t}\cos(\omega t-\Phi(t))\right] \\
&\qquad -2\epsilon^2\sigma R(t)\cos(\omega t-\Phi(t))+O(\epsilon^4)
\end{aligned} \tag{2.8}$$

It is important to note that the mismatch $\epsilon^2\sigma$ between the linear resonance frequency and the forcing frequency ω now appears in the ODE at the same order as the forcing term $F(t)$ and the damping term, by the judicious choice $1-\omega=O(\epsilon^2)$.

Similarly the right hand side of the small amplitude-equation (2.5) should be transformed to polar coordinates by substitution of (2.6). It transforms into

$$\begin{aligned}
&\epsilon\alpha V^2+\epsilon^2\beta V^3+\epsilon^2 F\cos(\omega t+\theta)-\epsilon^2 C\frac{\mathrm{d}V}{\mathrm{d}t}+O(\epsilon^3) \\
&\quad =\epsilon\alpha R(t)^2\cos^2(\omega t-\Phi(t))+\epsilon^2\Big[\beta R(t)^3\cos^3(\omega t-\Phi(t))+CR(t)\omega\sin(\omega t-\Phi(t)) \\
&\qquad\qquad +F\cos(\omega t+\theta)\Big]+O(\epsilon^3).
\end{aligned} \tag{2.9}$$

At this point the transformation to polar coordinates does not seem to have resulted in a useful form of the differential equation (2.5). However the consistency relation (2.7) can be used to generate a set of two first order differential equations for $R(t)$ and $\Phi(t)$ which are in the standard form which will be required in section 2.2. This is the result of the fact that a pair of equations of the form (where $A,B,\psi,N\in\mathbb{R}$)

$$\left\{\begin{array}{lcl} A\cos\psi+B\sin\psi &=& 0 \\ -A\sin\psi+B\cos\psi &=& N \end{array}\right. \tag{2.10}$$

is equivalent to the set of equations

$$\left\{\begin{array}{lcl} A &=& -N\sin\psi \\ B &=& N\cos\psi \end{array}\right. \tag{2.11}$$

the first of which is the result of adding $\cos\psi$ times the first and $(-\sin\psi)$ times the second equation in (2.10), and the second of which is obtained by adding $\sin\psi$ times the first and $\cos\psi$ times the second equation in (2.10).

Applying this elementary result to the problem at hand, using (2.7,2.8,2.9), puts equation (2.5) into the standard form

$$\left\{\begin{array}{lcl} \dfrac{\mathrm{d}R}{\mathrm{d}t} &=& \epsilon f_1(R(t),\Phi(t),t)+\epsilon^2 g_1(R(t),\Phi(t),t)+O(\epsilon^3) \\[2ex] \dfrac{\mathrm{d}\Phi}{\mathrm{d}t} &=& \epsilon f_2(R(t),\Phi(t),t)+\epsilon^2 g_2(R(t),\Phi(t),t)+O(\epsilon^3) \end{array}\right. \tag{2.12}$$

where the functions

$$\begin{aligned} f_1(R,\Phi,t) &= -\alpha R^2\cos^2(\omega t-\Phi)\sin(\omega t-\Phi) \\ f_2(R,\Phi,t) &= \alpha R\cos^3(\omega t-\Phi) \end{aligned} \tag{2.13}$$

were introduced, as well as

$$\begin{aligned} g_1(R,\Phi,t) &= -\beta R^3\cos^3(\omega t-\Phi)\sin(\omega t-\Phi) - CR\sin^2(\omega t-\Phi) \\ &\quad -F\cos(\omega t+\theta)\sin(\omega t-\Phi) - 2\sigma R\cos(\omega t-\Phi)\sin(\omega t-\Phi) \\ g_2(R,\Phi,t) &= \beta R^2\cos^4(\omega t-\Phi) + C\sin(\omega t-\Phi)\cos(\omega t-\Phi) \\ &\quad +\frac{F}{R}\cos(\omega t+\theta)\cos(\omega t-\Phi) + 2\sigma\cos^2(\omega t-\Phi). \end{aligned} \tag{2.14}$$

For convenience an overall multiplicative factor ω^{-1} was tacitly replaced by unity, since the resulting deviation is of order ϵ^3 only. By virtue of the choice of equation (2.6) all terms in (2.12) are asymptotically of order $O(\epsilon)$ as $\epsilon \to 0$ and are furthermore periodic in t with period $T = 2\pi/\omega$. By assuming the scaling $\omega = 1+\epsilon^2\sigma$ the effect of detuning of the forcing with respect to the linear resonance frequency is of the same order in ϵ as damping and forcing.

2.2 Second order averaging

2.2.1 Survey

The fact that the leading order terms defining $\frac{\mathrm{d}R}{\mathrm{d}t}$ and $R\frac{\mathrm{d}\Phi}{\mathrm{d}t}$ are of order ϵ suggest that R and Φ vary only on a slow time-scale ($O(1)$ variations if $\epsilon t = O_s(1)$, i.e. on a time scale $t \propto 1/\epsilon$, for a definition of these concepts see [8]). There are several techniques to analyze the system (2.12) to exploit this a priori suggestion concerning the behaviour of R and Φ.

Two standard techniques are multiple-scale perturbation expansions [10] and averaging procedures [8]. Essential to the first technique is that a perturbation expansion $\sum_{n=0}^{N-1}\epsilon^n u_n$ to the exact solution u is sought where all the u_n depend on *multiple timelike variables*, corresponding to different time scales. Thus time variables $T_1 = t, T_2 = \epsilon t, \ldots T_N = \epsilon^N t$ are introduced, which are treated as independent variables. By treating the different time variables independently an assumption of decoupling of behaviour on fast and slow time scales is introduced. An extensive introduction to this approach is given in [10], together with a sample calculation for a system very similar to (2.5). This method was also employed in the specific case of a basin where $a(z) = z+1$ in [7]. Asymptotic validity of multiple time scale perturbations for more than two time scales however has not been proved rigorously according to [8].

Averaging techniques are also based on the fact that the variables to solve for (in our case R and Φ) vary slowly compared to the explicitly time-dependent terms in the differential equation. As a consequence $\frac{\mathrm{d}R}{\mathrm{d}t}$ and $R\frac{\mathrm{d}\Phi}{\mathrm{d}t}$ can be approximated to first order by averaging the functions $f_i(R,\Phi,t)$ over one fast period of time of length $T = 2\pi/\omega$. During the averaging procedure the time-dependence of R and Φ is suppressed. More precisely $f_i(R,\Phi,t)$ is averaged, not $f_i(R(t),\Phi(t),t)$. Though more implicitly than in the case of multiple scale perturbation expansion this approximation also depends on weakness of the coupling of behaviour on the fast time scale T and slow change of R and Φ on a time scale ϵt. Second order averaging is a somewhat more complicated extension of the first order technique along exactly the same ideas. A survey of averaging techniques is provided by [8], who also discuss the technique applied presently, which is second order averaging in the periodic case.

2.2.2 Specifics

Consider an initial value problem on a subset D of $\mathbb{R}^n$ of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \epsilon f(x,t) + \epsilon^2 g(x,t) + \epsilon^3 \rho(x,t,\epsilon) \quad x(0) = x_0 \tag{2.15}$$

where $f, g : D \times [0,\infty) \to \mathbb{R}^n, \rho : D \times [0,\infty) \times [0,\epsilon_0) \to \mathbb{R}^n$ where f is supposed to have a Lipschitz-continuous partial derivative with respect to x, where g and ρ are Lipschitz-continuous and where all functions involved are continuous. Assume furthermore that f and g are T-periodic in t with averages

$$f^0(x) = \frac{1}{T}\int_0^T f(x,\tau)\,\mathrm{d}\tau \quad \text{and} \quad g^0(x) = \frac{1}{T}\int_0^T g(x,\tau)\,\mathrm{d}\tau$$

and uniform boundedness of the residual term ρ on $D \times [0, \frac{L}{\epsilon}) \times [0,\epsilon_0)$ (where L is some constant related to the time-scale approximation (cf. [8])). Define u as the solution of

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \epsilon f^0(u) + \epsilon^2 \left(f^{1,0}(u) + g^0(u)\right), \quad u(0) = x_0 \tag{2.16}$$

together with

$$f^{1,0}(x,t) = \frac{1}{T}\int_0^T \nabla f(x,\tau)u^1(x,\tau) - \nabla u^1(x,\tau) f^0(x)\,\mathrm{d}\tau$$

and

$$u^1(x,t) = \int_0^t [f(x,\tau) - f^0(x)]\,\mathrm{d}\tau + a(x)$$

where $a(x)$ is some C^1 vector field such that the average of u^1 is zero. If $u(t)$ remains in the domain of definition D of x on the time scale $1/\epsilon$ then

$$u(t) + \epsilon u^1(u(t),t)$$

is an approximation to order ϵ^2 of $x(t)$ on a time scale $1/\epsilon$ (for a proof consult [8]). This result is considerably improved in the special case $f^0(x) = 0$, in which case the approximation (2.16) is valid on a time scale $1/\epsilon^2$ (cf. [8], pp. 62), at least as long as $u(t) \in D$ and as long as ρ stays uniformly bounded for $0 \le t < L/\epsilon^2$.

2.2.3 Motivational calculation

The procedure described in section 2.2.1 has a rather complex structure, which may be confusing to the reader. A rigorous proof as provided by [8] may be illustrative of general techniques used to prove approximation theorems, but does not provide much insight. Instead of a proof a motivational calculation is provided to show why the procedure seems reasonable; note that no justification of any of the approximations is presented.

Suppose an approximation to $x(t)$ of the form $u(t) + \epsilon u_1(u(t),t)$ to order ϵ^2 is to be calculated. Trivially x satisfies

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}u}{\mathrm{d}t} + \epsilon(\nabla u_1)\frac{\mathrm{d}u}{\mathrm{d}t} + \epsilon\frac{\partial u_1}{\partial t}(u,t)$$

by the chain rule. Combining this result with the differential equation (2.15) results in

$$\begin{aligned}\frac{\mathrm{d}u}{\mathrm{d}t} &= \epsilon f(u+\epsilon u_1,t) - \epsilon\frac{\partial u_1}{\partial t} - \epsilon(\nabla u_1)\frac{\mathrm{d}u}{\mathrm{d}t} + \epsilon^2 g(u,t) \\ &= \epsilon\left[f(u,t) - \frac{\partial u_1}{\partial t}\right] - \epsilon(\nabla u_1)\frac{\mathrm{d}u}{\mathrm{d}t} + \epsilon^2\left[(\nabla f)u_1 + g(u,t)\right] + O(\epsilon^3)\end{aligned} \tag{2.17}$$

Because $\frac{\mathrm{d}u}{\mathrm{d}t}$ itself is of order ϵ the only terms in the right hand side which really are of order ϵ are the ones in square brackets. By choosing

$$u_1(u,t) = \int_0^t \left[f(u,\tau) - f^0(u)\right] \mathrm{d}\tau + a(u)$$

the explicit time dependence is removed from the terms which are first order in ϵ. Substitution of this choice in (2.17) results in

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \epsilon f^0(u) + \epsilon^2 \left[(\nabla f)u_1 - (\nabla u_1)\frac{\mathrm{d}u}{\mathrm{d}t} + g(u,t)\right] + O(\epsilon^3)$$

which results in (2.16) after substituting $\frac{\mathrm{d}u}{\mathrm{d}t} = \epsilon f^0(u) + O(\epsilon^2)$ and averaging.

2.3 Application

In order to obtain an approximation to $(\frac{\mathrm{d}R}{\mathrm{d}t}, \frac{\mathrm{d}\Phi}{\mathrm{d}t})$ by second order averaging the previous section will be applied. To this end the connection between the problem (2.12) and the form required in (2.15) will be established, the conditions which are required in order to validate the result are checked and all the auxiliary functions mentioned in the previous section will be calculated for this specific system. Finally these results are combined to obtain the 'modulation equations', the averaged approximations to the evolution equations for R and Φ.

2.3.1 Conditions

Closer examination of Eq. (2.12) shows that second order averaging can be applied to $x(t) = (R(t), \Phi(t))^T$ if the domain of definition D does not contain $R = 0$. Thus $D := [\eta, C) \times \mathbb{R} \subset \mathbb{R}^2$ will do for any $\eta > 0$, and $C > \eta$. Excluding $R = 0$ is necessary to ensure Lipschitz-continuity of $g_2(R,\Phi)$ (clearly f as in section 2.2 is to be identified with $(f_1, f_2)^T$ and similarly g with $(g_1, g_2)^T$). Evidently f and g satisfy the required continuity conditions, as they are C^∞ on D. Furthermore they are T-periodic with $T = 2\pi/\omega$. This is essentially the result of the choice Eq. (2.6). Because the resulting system for $u(t) = (\bar{R}, \bar{\Phi})^T$ will turn out to be either hamiltonian or dissipative the solutions will not grow indefinitely but stay bounded. Finally the explicit time dependence of the residual term ρ only occurs in cosines and sines in which ρ is polynomial. This property ensures that ρ can be estimated by a time-independent polynomial in R. The restriction of the domain of definition from above by the bound C, which is some arbitrarily large fixed positive constant, guarantees the existence of a uniform bound on ρ.

The only condition which should be worried about is the condition that solutions of the averaged equation have to stay away from $R = 0$. This problem will be discussed later.

2.3.2 Auxiliary calculations

First the average of $f(R,\Phi,t) = (f_1(R,\Phi,t), f_2(R,\Phi,t))^T$ is calculated. To this end I integrate over one period of fast time $2\pi/\omega$ to obtain

$$\begin{aligned} f_1^0(R,\Phi) &= -\alpha R^2 \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \cos^2(\omega t - \Phi)\sin(\omega t - \Phi)\,\mathrm{d}t = 0 \\ f_2^0(R,\Phi) &= \alpha R \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \cos^3(\omega t - \Phi)\,\mathrm{d}t = 0 \end{aligned} \tag{2.18}$$

which produces $f^0(R,\Phi) = (0,0)^T$. The fact that this result simplifies the remaining calculation significantly is of minor importance compared to the conclusion that *the result which will be obtained will be an asymptotic approximation to the solution of Eq. (2.5) on a time scale* $1/\epsilon^2$, as long as all other conditions are met. The phenomenon that $f^0 = 0$ hinges on the choice $\omega = 1 + \epsilon^2\sigma$; if the scaling $\omega = 1 + \epsilon\sigma$ would

have been introduced the terms containing σ in (g_1, g_2) would have been part of (f_1, f_2). These terms do not (both) average to zero.

The average of g is obtained likewise by

$$\begin{aligned} g_1^0(R,\Phi,t) &= -\beta R^3 \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \cos^3(\omega t-\Phi)\sin(\omega t-\Phi)\,\mathrm{d}t - \\ &\qquad CR\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\sin^2(\omega t-\Phi)\,\mathrm{d}t - F\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\cos(\omega t+\theta)\sin(\omega t-\Phi)\,\mathrm{d}t - \\ &\qquad 2\sigma R\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\cos(\omega t-\Phi)\sin(\omega t-\Phi)\,\mathrm{d}t \\ &= -\frac{\beta R^3}{2\pi}\cdot 0 - \frac{CR}{2\pi}\cdot\pi + \frac{F}{2\pi}\cdot\pi\sin(\theta+\Phi) - 2\sigma^2\cdot 0 \\ &= -\frac{CR}{2} + \frac{F}{2}\sin(\theta+\Phi), \end{aligned} \tag{2.19}$$

and similarly

$$\begin{aligned} g_2^0(R,\Phi,t) &= \beta R^2\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\cos^4(\omega t-\Phi)\,\mathrm{d}t + C\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\sin(\omega t-\Phi)\cos(\omega t-\Phi)\,\mathrm{d}t + \\ &\qquad F\frac{\omega}{2\pi R}\int_0^{2\pi/\omega}\cos(\omega t+\theta)\cos(\omega t-\Phi)\,\mathrm{d}t + 2\sigma\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\cos^2(\omega t-\Phi)\,\mathrm{d}t \\ &= \frac{3\beta R^2}{8} + \frac{F}{2R}\cos(\theta+\Phi) + \sigma. \end{aligned} \tag{2.20}$$

In section 2.2 two other auxiliary functions were mentioned. In order to calculate $u^1 = (R^1, \Phi^1)^T$ it is necessary to evaluate

$$\int_0^t [f_i(R,\Phi,t) - f_i^0(R,\Phi)]\,\mathrm{d}t \qquad \text{for} \qquad i = 1,2.$$

Since f_i^0 is zero (both for $i = 1$ and $i = 2$) this reduces to

$$\int_0^t f_1(R,\Phi,\tau)\,\mathrm{d}\tau = \int_0^t -\alpha R^2\cos^2(\omega\tau-\Phi)\sin(\omega\tau-\Phi)\,\mathrm{d}\tau = \frac{\alpha R^2}{3\omega}\left[\cos^3(\omega t-\Phi)-1\right] \tag{2.21}$$

and

$$\int_0^t f_2(R,\Phi,\tau)\,\mathrm{d}\tau = \int_0^t \alpha R\cos^3(\omega\tau-\Phi)\,\mathrm{d}\tau = \frac{\alpha R}{12\omega}\left[\sin(3(\omega t-\Phi)) + 9\sin(\omega t-\Phi)\right]. \tag{2.22}$$

Since $\cos^3(\omega t-\Phi)$, $\sin(3(\omega t-\Phi))$ and $\sin(\omega t-\Phi)$ all average to zero the vector field $(R^1, \Phi^1)^T$ is obtained from Eqs. (2.21, 2.22) by deleting the constant term (absorbing it in the vector field a), resulting in the expressions

$$\begin{aligned} R^1(R,\Phi,t) &= \frac{\alpha R^2}{3\omega}\cos^3(\omega t-\Phi) \\ \Phi^1(R,\Phi,t) &= \frac{\alpha R}{12\omega}[9\sin(\omega t-\Phi) + \sin(3(\omega t-\Phi))]. \end{aligned} \tag{2.23}$$

After these tedious manipulations the function f^1 can finally be calculated. As mentioned earlier f^1 is defined as

$$f^1(R,\Phi,t) = ((\nabla f)(R,\Phi,t))\cdot(R^1,\Phi^1)^T - ((\nabla u^1)(R,\Phi,t))\cdot f^0(R,\Phi)$$

which simplifies considerably due to the fact that $f^0(R,\Phi) = (0,0)^T$ for all (R,Φ). An elementary calculation reveals that $\nabla f(R,\Phi,t)$, which is the matrix with components $\partial f_i/\partial x_j$ $(i,j \in \{1,2\})$ where $x_1 = R, x_2 = \Phi$, equals

$$\nabla f(R,\Phi,t) = \begin{pmatrix} -2\alpha R\cos^2(\omega t-\Phi)\sin(\omega t-\Phi) & \alpha R^2\frac{1}{4}[\cos(\omega t-\Phi)+3\cos(3(\omega t-\Phi))] \\ \alpha\cos^3(\omega t-\Phi) & 3\alpha R\cos^2(\omega t-\Phi)\sin(\omega t-\Phi) \end{pmatrix}.$$

By left-multiplication of $(R^1,\Phi^1)^T$ with the matrix ∇f the result

$$\begin{aligned} f_1^1(R,\Phi,t) &= \frac{\alpha^2R^3}{96\omega}[\sin(6(\omega t-\Phi))+20\sin(4(\omega t-\Phi))-27\sin(2(\omega t-\Phi))] \\ f_2^1(R,\Phi,t) &= -\frac{\alpha^2R^2}{12\omega}\cos^2(\omega t-\Phi)[\cos(4(\omega t-\Phi))+10\cos(2(\omega t-\Phi))-15] \end{aligned} \tag{2.24}$$

is obtained, after applying several trigonometric identities to reduce powers of cosines and sines. From the form of f_1^1 it is immediately clear that $f_1^{1,0} = 0$. Furthermore, using the fact that the average of $\cos^2 x = 1/2$ and the fact that the average of $\cos^2 x\cos(2x) = 1/4$ it is concluded that

$$\begin{aligned} f_1^{1,0}(R,\Phi) &= 0 \\ f_2^{1,0}(R,\Phi) &= \frac{5}{12\omega}\alpha^2R^2. \end{aligned} \tag{2.25}$$

2.3.3 Modulation equations

By collecting the various intermediary results scattered throughout the previous subsection an asymptotic approximation to order ϵ for V in terms of $\bar{R}$ and $\bar{\Phi}$ will be derived, for which two first order differential equations are constructed. According to [8] the amplitude and phase R and Φ, i.e. the solution of Eq. (2.12), can be approximated by

$$\begin{pmatrix} R \\ \Phi \end{pmatrix} = \begin{pmatrix} \bar{R} \\ \bar{\Phi} \end{pmatrix} + \epsilon\begin{pmatrix} R^1 \\ \Phi^1 \end{pmatrix} + O(\epsilon^2)$$

which results in the approximation

$$V(t) = [\bar{R}(t)+\epsilon(R^1(\bar{R},\bar{\Phi},t))]\cos(\omega t-\bar{\Phi}-\epsilon\Phi_1(\bar{R},\bar{\Phi},t))+O(\epsilon^2)$$

of the scaled excess volume V. In order to simplify this expression I first separate $\cos(\omega t-\bar{\Phi}-\epsilon\Phi^1)$ into

$$\cos(\omega t-\bar{\Phi})\cos(\epsilon\Phi^1)+\sin(\omega t-\bar{\Phi})\sin(\epsilon\Phi^1)$$

in order to isolate the ϵ-dependence. As long as $\Phi^1(\bar{R},\bar{\Phi},t)$ is bounded, independently of ϵ, on a time scale $1/\epsilon^2$ (the extent of time for which this approximation should be valid) the estimates

$$\cos(\epsilon\Phi^1) = 1+O(\epsilon^2) \quad \text{and} \quad \sin(\epsilon\Phi^1) = \epsilon\Phi^1+o(\epsilon^2) \quad \text{as } \epsilon\to 0$$

are correct. Employing these estimates results in the asymptotic approximation for V

$$V(t) = \bar{R}\cos(\omega t-\bar{\Phi})+\epsilon\left[R^1(\bar{R},\bar{\Phi},t)\cos(\omega t-\bar{\Phi})+\bar{R}\Phi^1(\bar{R},\bar{\Phi},t)\sin(\omega t-\bar{\Phi})\right]+O(\epsilon^2)$$

which can be expressed in terms of $\bar{R}$ and $\bar{\Phi}$ by recalling (2.23). Indeed substitution produces, noting that $1/\omega = 1+O(\epsilon^2)$,

$$\begin{aligned} V(t) &= \bar{R}\cos(\omega t-\bar{\Phi})+\epsilon\alpha\bar{R}^2\Big\{\frac{1}{3}\cos^4(\omega t-\bar{\Phi})+\frac{1}{12}\big[9\sin^2(\omega t-\bar{\Phi}) \\ &\qquad +\sin(\omega t-\bar{\Phi})\sin(3(\omega t-\bar{\Phi}))\big]\Big\}+O(\epsilon^2) \end{aligned} \tag{2.26}$$

$$= \bar{R}\cos(\omega t-\bar{\Phi})+\epsilon\frac{\alpha\bar{R}^2}{2}\left[1-\frac{1}{3}\cos(2(\omega t-\bar{\Phi}))\right]+O(\epsilon^2). \tag{2.27}$$

The last simplification was not approximate, but entailed the use of the trigonometric identity

$$4\cos(x)^4 + 9\sin(x)^2 + \sin(x)\,\sin(3x) = 6 - 2\cos(2x).$$

Having derived an asymptotically valid expression for the small amplitude excess volume V in terms of $\bar{R}$ and $\bar{\Phi}$ it is necessary to investigate their time dependence more closely. From section 2.2 it is clear that $\bar{R}$ and $\bar{\Phi}$ satisfy

$$\begin{aligned} \frac{\mathrm{d}\bar{R}}{\mathrm{d}t} &= \epsilon f_1^0(\bar{R},\bar{\Phi}) + \epsilon^2\left[f_1^{1,0}(\bar{R},\bar{\Phi}) + g_1^0(\bar{R},\bar{\Phi})\right] \\ \frac{\mathrm{d}\bar{\Phi}}{\mathrm{d}t} &= \epsilon f_2^0(\bar{R},\bar{\Phi}) + \epsilon^2\left[f_2^{1,0}(\bar{R},\bar{\Phi}) + g_2^0(\bar{R},\bar{\Phi})\right], \end{aligned} \tag{2.28}$$

which reveals that the terms of order ϵ disappear due to the fact that f_1 and f_2 average to zero. Substitution of the expressions for $f_i^{1,0}$ and g_i in Eqs. (2.25, 2.19, 2.20) results in

$$\begin{aligned} \frac{1}{\epsilon^2}\frac{\mathrm{d}\bar{R}}{\mathrm{d}t} &= -\frac{C}{2}\bar{R} + \frac{F}{2}\sin(\theta+\bar{\Phi}) \\ \frac{1}{\epsilon^2}\bar{R}\frac{\mathrm{d}\bar{\Phi}}{\mathrm{d}t} &= \sigma\bar{R} - \varpi\bar{R}^3 + \frac{F}{2}\cos(\theta+\bar{\Phi}) \end{aligned} \tag{2.29}$$

The effect of the nonlinear restoring term is condensed into the parameter

$$\varpi = -\left[\frac{5\alpha^2}{12} + \frac{3\beta}{8}\right].$$

For the special case of a linear slope $a(z) = z + 1$ the parameter ϖ equals $1/12$ (compare [7]). It may be rather surprising that the nonlinearity in the modulation equations doesn't only arise from the slope ($\propto \alpha$) of the walls of the basin but equally from the curvature ($\propto \beta$). This would not have been if f would not have averaged to zero.

By rescaling the time variable $T = \epsilon^2 t$ it is apparent that

$$\frac{1}{\epsilon^2}\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}T}$$

which shows that we have obtained a differential equation for $\bar{R}$ and $\bar{\Phi}$ in terms of *(very) slow* time T (the use of the symbol T will not lead to confusion with the period $T = 2\pi/\omega$ introduced earlier). The derivative of a function h with respect to slow time T will henceforth be abbreviated by $\dot{h}$.

2.3.4 Conditions again

In summary equations (2.26) and (2.29) represent an asymptotic approximation to $V = \epsilon v$ to order ϵ, valid on a time scale $1/\epsilon^2$, as long as the conditions in section 2.3.1 are fulfilled. A somewhat problematic aspect of these conditions was that the radius R was dictated to stay out of some neighbourhood of zero.

This condition was a consequence of the fact that g is not a smooth function in $R = 0$ due to the factor $1/R$ resulting from the transformation of V to polar coordinates. This factor $1/R$ only affects the evolution equation for the phase angle $\bar{\Phi}$. Since singular behaviour of $\bar{\Phi}$ does not result in unbounded behaviour of R one may guess that this problem is not essential. Furthermore the equations (2.29) will be written in a Cartesian form in the next chapter, in which form the $1/R$ singularity is absent again. Instead of performing a detailed analysis of the behaviour near $R = 0$ another method is used to show that the domain of definition can be extended to include $R = 0$ without loss of validity of the approximation.

It is possible to perform an averaging which circumvents the singularity and immediately results in the Cartesian equations which will be derived from (2.29) in the next chapter. To this end an alternative to

the phase-amplitude transformation, which is called a *van der Pol* transformation (see [11]) from V and $W = \frac{\mathrm{d}V}{\mathrm{d}t}$ to new variables p, q is introduced:

$$\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \cos(\omega t) & -\omega^{-1}\sin(\omega t) \\ -\sin(\omega t) & -\omega^{-1}\cos(\omega t) \end{pmatrix} \begin{pmatrix} V \\ W \end{pmatrix}.$$

The new coordinates form a frame of reference in which solutions of the zeroth order equation ((2.5) with ϵ is zero) are (nearly) stationary. Through the differential equation for V a first order two dimensional system of the form

$$\begin{aligned} \frac{\mathrm{d}p}{\mathrm{d}t} &= -\omega^{-1}\sin(\omega t)\left[\epsilon f(p,q) + \epsilon^2 g(p,q,t)\right] + O(\epsilon^3) \\ \frac{\mathrm{d}q}{\mathrm{d}t} &= -\omega^{-1}\cos(\omega t)\left[\epsilon f(p,q) + \epsilon^2 g(p,q,t)\right] + O(\epsilon^3) \end{aligned} \tag{2.30}$$

is obtained with

$$\begin{aligned} f(p,q) &= \alpha V^2 \\ g(p,q) &= \beta V^3 - CW + F\cos(\omega t + \theta) + 2\sigma V \end{aligned}$$

and

$$\begin{aligned} V &= p\cos(\omega t) - q\sin(\omega t) \\ W &= -\omega(p\sin(\omega t) + q\cos(\omega t)). \end{aligned}$$

The van der Pol transformation results in a system of the form (2.15), which again allows the application of second order averaging in the periodic case. In contrast to the phase-amplitude transformation, the van der Pol transformation does not introduce a singular factor that results in the necessity to exclude the origin from the domain of definition. The averaging procedure is completely analogous to the procedure outlined previously, and results in the evolution equations

$$\begin{aligned} \frac{1}{\epsilon^2}\frac{\mathrm{d}\bar{p}}{\mathrm{d}t} &= -\frac{C}{2}\bar{p} - \bar{q}(\varpi r^2 - \sigma) + \frac{F}{2}\sin(\theta) \\ \frac{1}{\epsilon^2}\frac{\mathrm{d}\bar{q}}{\mathrm{d}t} &= -\frac{C}{2}\bar{q} + \bar{p}(\varpi r^2 - \sigma) - \frac{F}{2}\cos(\theta) \end{aligned} \tag{2.31}$$

for approximations to p and q, in terms of $r = \sqrt{\bar{p}^2 + \bar{q}^2}$. An expression for V (i.e. (2.26)) can also be obtained. The validity of this approximation extends to a $1/\epsilon^2$ time scale again, because the averaged evolution equations do not contain first order terms. The new evolution equations are equivalent up to scaling to the Cartesian form of (2.29), with $\bar{p} = \bar{R}\cos(\bar{\Phi}), \bar{q} = -\bar{R}\sin(\bar{\Phi})$ (note the minus sign) which is discussed in the next chapter.

This alternative approach shows that (2.26, 2.29) remain valid if the condition that $\bar{R}$ stays away from $\bar{R} = 0$ is relaxed. The phase angle transformation does have the advantage that the polar coordinates $\bar{R}, \bar{\Phi}$ have a more obvious physical interpretation, and allow a more direct comparison with the multiple scale perturbation calculation found in the literature [10].

Multiple Scale perturbation expansion

The technique of multiple scale perturbation expansion results in the same approximation to Eq. (2.5) (see [10]), however without rigorous proof of the asymptotic validity. The calculation by second order averaging provides an a posteriori justification of the treatment in [7]. It has been demonstrated [12] that second order averaging and multiple scale perturbation expansion are formally equivalent if two time scales are employed. Though the behaviour described by the results (2.26,2.29) contains only two time scales (fast time t and slow time $T = \epsilon^2 t$) as a result of the fact that $f^0 = (0,0)^T$ the multiple scale perturbation expansion method does require the use of *three* time scales (also ϵt).

2.3.5 Nearly periodic forcing

The previous discussion was limited to forcings of the form

$$f = \epsilon^3 F \cos(\omega t + \theta)$$

where the amplitude F and the phase angle θ were assumed constant. The calculation can however be extended to include variations of F and θ on a slow time scale $\epsilon^2\tau$. This is most easily seen by writing $\tau = \epsilon^2 t$; by increasing the dimension of the system by adding the variable τ to solve for as well as the equation $\frac{\mathrm{d}\tau}{\mathrm{d}t} = \epsilon^2$ a three-dimensional, nonlinear system is obtained of the form

$$\begin{cases} \dfrac{\mathrm{d}R}{\mathrm{d}t} &= \epsilon f_1(R,\Phi,t) + \epsilon^2 g_1(t,R,\Phi,\tau) + O(\epsilon^3) \\ \dfrac{\mathrm{d}\Phi}{\mathrm{d}t} &= \epsilon f_2(R,\Phi,t) + \epsilon^2 g_2(t,R,\Phi,\tau) + O(\epsilon^3) \\ \dfrac{\mathrm{d}\tau}{\mathrm{d}t} &= \epsilon^2 \end{cases} \tag{2.32}$$

The vector field $(g_1, g_2)^T$ is the same as in (2.14), with F replaced by $F(\tau)$ and θ replaced by $\theta(\tau)$. At the cost of increasing the dimension[1] of the system the problem has been recast in the form discussed in section 2.2; the vector fields $f = (f_1, f_2, 0)^T$ and $g = (g_1, g_2, 1)^T$ are periodic in t with period $2\pi/\omega$, they still satisfy the regularity conditions if the dependence of F and θ on slow time τ is Lipschitz-continuous. Averaging with respect to fast time t produces the averaged equation $\frac{\mathrm{d}\tau}{\mathrm{d}t} = \epsilon^2$, together with the equations (2.29) and (2.26), where F and θ should be replaced by their τ-dependent counterparts.

Because the extra equation does not contain a term of order ϵ the approximation is still valid on a time scale $1/\epsilon^2$.

[1] This is a standard trick; often nonautonomous differential equations are cast into autonomous form by increasing the dimension of the system by introducing a new variable ϕ with $\dot{\phi} = 1$, replacing all explicit time-dependences by ϕ-dependences. For applications of the trick in situations similar to the present see for instance [13],[14].

CHAPTER 3

MODULATION EQUATIONS

Introduction

This chapter contains a brief description of the properties of the system of differential equations derived in chapter 2. Because the results are very similar to those presented in [7] for the special case of a basin with a linear slope lengthy derivations are avoided. The goal is to present a concise qualitative overview to those readers not familiar with [7] and to provide some of the straightforward quantitative generalizations of [7].

3.1 Frequency response

By assumption the driving frequency of the forcing was chosen close to the linear resonance frequency of the system defined by equation (2.5). As a result resonant behaviour, e.g. amplified response for some detunings σ is expected in the averaged equations. In order to investigate the relation between amplitude $\bar{R}$ and detuning σ for steady oscillatory motion, which satisfies that $\bar{R}$ and $\bar{\Phi}$ are constant, the right-hand sides of (2.29) are set equal to zero. Solving for σ determines the frequency response curve

$$\sigma(\bar{R}, F, C) = \varpi \bar{R}^2 \pm \sqrt{\frac{F^2}{4\bar{R}^2} - \frac{C^2}{4}}.$$

In effect their are (for fixed F, C) two branches of stationary solutions meeting each other at the maximum response located on the curve $\sigma = \varpi \bar{R}^2$ at a value $\sigma = \varpi \frac{F^2}{C^2}$ where $\bar{R} = |F/C|$. The two branches together form a bended resonance horn as shown in figures 3.1; for $\varpi > 0$ this resonance horn is bent towards positive detuning; in the case of negative ϖ the horn is bent towards negative detuning. A special case is $\varpi = 0$; the system is linear and the resonance horn is straight. As mentioned before the top of the resonance horn moves along a parabola ($\sigma = \varpi \bar{R}^2$) as the ratio F/C is varied.

The fact that the resonance horn is bent (for $\varpi \neq 0$) implies that there is a range of detunings for which multiple equilibria, which are either choked or amplified exist. This behaviour is similar to the properties of the Duffing equation, where the case of a left respectively right bent resonance horn is called *soft spring* respectively *hard spring* behaviour, which is terminology related to the physical model. The amplified and choked equilibria, which are respectively the largest and the smallest intersections of a line $\sigma =$ constant are stable, while the middle intersection is not. This may give rise to 'jump phenomena' if the detuning is slowly varied. If the detuning σ is slowly increased from $-\infty$ (in the case $\varpi > 0$) the system remains in the amplified regime (upper branch) until the maximum of the resonance horn is reached, at which point the system suddenly jumps into the choked regime (lower branch). On the other hand decreasing the detuning from $+\infty$, the system remains in the choked regime until the choked equilibrium and the unstable equilibrium coalesce and the system suddenly jumps to the amplified regime. This behaviour is described in [10]. It is especially relevant in for instance electric circuits which exhibit nonlinear behaviour.

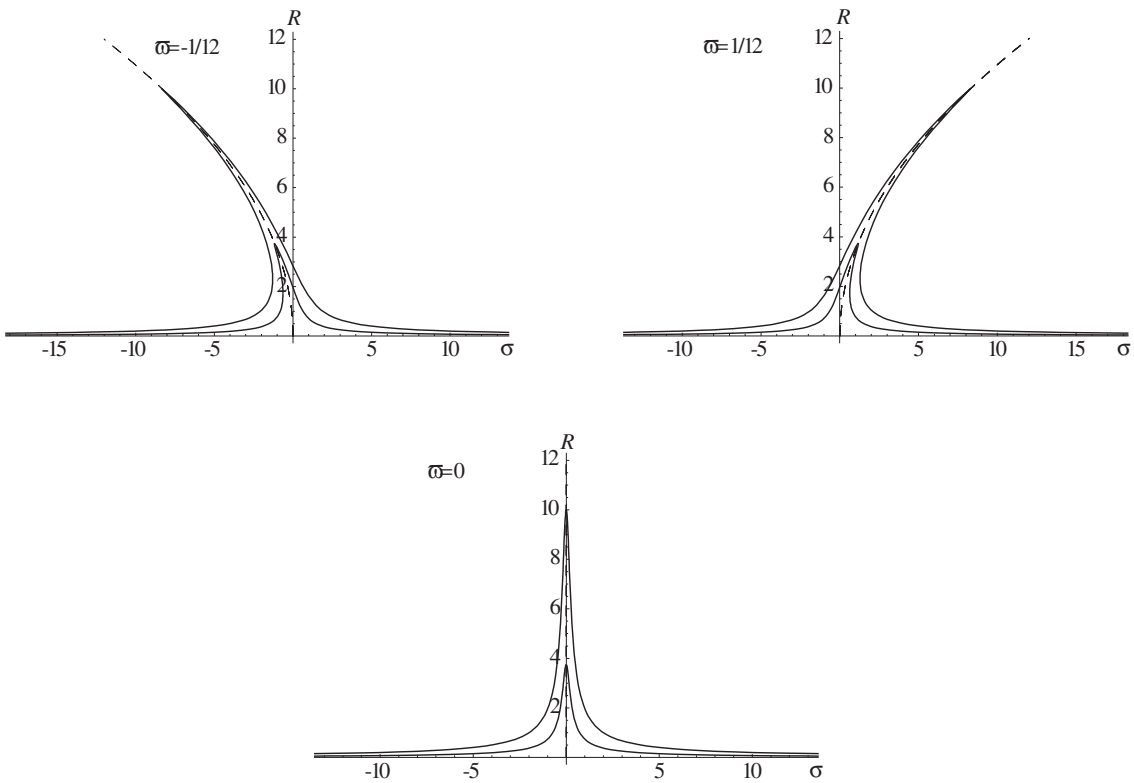


FIGURE 3.1: *Qualitatively different frequency response curves arise depending on the sign of ϖ. For $\varpi < 0$ there is an amplified response for $\sigma < 0$, for $\varpi = 0$ the maximal response occurs for $\sigma = 0$ and for $\varpi > 0$ the maximal response corresponds to some positive detuning (resp. first, third and second graph). The exact maximum of the frequency response curve depends on the ratio F/C; in each graph the upper continuous curve corresponds to $F/C = 10$ and the lower continuous curve corresponds to $F/C = 3\frac{3}{4}$. As F/C is varied the top of the resonance horn moves along the dashed curve which corresponds to $\sigma = \varpi \bar{R}^2$.*

3.2 Different signs of ϖ.

From figures 3.1 it is apparent that the two cases $\varpi < 0$ and $\varpi > 0$ result in qualitatively different phenomena. However it is possible to transform equations (2.29) for ϖ to a set of equations of the form (2.29) with positive ϖ. Define $\tilde{\sigma} = -\sigma, \tilde{\Phi} = -\bar{\Phi}, \tilde{\theta} = -\theta$ and $\tilde{F} = -F$; leaving the radial variable $\bar{R}$ unchanged, substitution produces

$$\begin{aligned} \dot{\bar{R}} &= -\frac{C}{2}\bar{R} + \frac{-\tilde{F}}{2}\sin(-\tilde{\Phi} - \tilde{\theta}) = -\frac{C}{2}\bar{R} + \frac{\tilde{F}}{2}\sin(\tilde{\Phi} + \tilde{\theta}) \\ -\bar{R}\dot{\tilde{\Phi}} &= -\tilde{\sigma}\bar{R} - (-|\varpi|)\bar{R}^3 - \frac{\tilde{F}}{2}\cos(-\tilde{\Phi} - \tilde{\theta}) \end{aligned}$$

which is a system of exactly the same form as (2.29) for the variables $\bar{R}$ and $\tilde{\Phi}$ but with a positive parameter of nonlinearity.

This property provides the possibility to restrict attention to the cases $\varpi = 0$ and $\varpi > 0$. The case $\varpi = 0$ is a rather degenerate case; furthermore, the resulting system is linear and therefore trivial to solve. From now on the discussion will be limited to systems of the form

$$\begin{aligned} \dot{R} &= -\frac{C}{2}R + \frac{F}{2}\sin(\theta + \Phi) \\ R\dot{\Phi} &= \sigma R - \varpi R^3 + \frac{F}{2}\cos(\theta + \Phi) \end{aligned} \tag{3.1}$$

where $\varpi > 0$. From now on the $^-$ on variables will be omitted.

3.3 Cartesian form

Following the derivation in [7] a rescaling is performed, such that $t \to t/\sigma, R \to \sqrt{\sigma}R, C \to \sigma C, F \to \sqrt{\sigma^3}F$. This rescaling, which actually corresponds to setting $\epsilon = \sqrt{|\omega - 1|}$ (note that $\omega - 1$ is the actual (nonscaled) detuning of the original forcing $f(t)$) , transforms equation (3.1) to

$$\left\{ \begin{aligned} \dot{R} &= -\frac{C}{2}R + \frac{F}{2}\sin(\theta + \Phi) \\ R\dot{\Phi} &= R - \varpi R^3 + \frac{F}{2}\cos(\theta + \Phi) \end{aligned} \right. \tag{3.2}$$

which is simply (3.1) with σ set equal to 1. To study the behaviour of the modulation equations in further detail it is advantageous to rewrite (3.2) in terms of Cartesian[1] coordinates $X = R\cos(\Phi)$ and $Y = R\sin(\Phi)$. Substituting (3.2) into

$$\begin{aligned} \dot{X} &= \dot{R}\cos(\Phi) - R\dot{\Phi}\sin(\Phi) \\ \dot{Y} &= \dot{R}\sin(\Phi) + R\dot{\Phi}\cos(\Phi) \end{aligned}$$

results in

$$\begin{aligned} \dot{X} &= -\frac{C}{2}R\cos(\Phi) - (1 - \varpi R^2)R\sin(\Phi) \\ &\qquad + \frac{F}{2}\big[\cos(\Phi)\sin(\Phi + \theta) - \cos(\Phi + \theta)\sin(\Phi)\big] \\ &= -\frac{C}{2}X - (1 - \varpi R^2)Y + \frac{F}{2}\sin\theta \\ \dot{Y} &= -\frac{C}{2}Y + (1 - \varpi R^2)X + \frac{F}{2}\cos\theta. \end{aligned} \tag{3.3}$$

[1] This transformation is not related to (2.6).

These equations were obtained earlier in section 2.3.4 by the alternative averaging method which employed the van der Pol transformation. These equations are equivalent to (2.31) if $\bar{p}$ and X are identified, as well as Y and $-\bar{q}$. Furthermore the detuning should be set equal to unity in equations (2.31) due to the rescaling in (3.2).

The coefficient ϖ (positive by choice of $a(z)$ or after the transformation (3.2)) can be scaled out and set equal to unity, by employing the rescaling $X \to X/\sqrt{\varpi}, Y \to Y/\sqrt{\varpi}$ and $F \to F/\sqrt{\varpi}$. Apparently the behaviour of (2.29) is, after appropriate rescalings, qualitatively the same as described in [7] for the special case of a basin with linear slope.

3.4 Dynamics

In the previous section it was concluded that the dynamics of (2.29) is, after rescaling, equivalent to the behaviour described in [7]. In order to familiarize the reader with the characteristics of the dynamical system (3.3) a concise outline, without derivations, of the conclusions in [7] is provided.

The fixed points of (3.3) are found immediately by setting the right hand sides to zero; a third order polynomial equation

$$\left[\frac{C^2}{4} + (\varpi\rho - 1)^2\right]\rho = \frac{F^2}{4} \tag{3.4}$$

in $\rho = R^2$ results, which has either one or three solutions. This result should be compared with the resonance hornes in figure 3.1. These graphs show that any line $\sigma =$ constant (vertical line) will in general intersect a resonance horn either one or three times, depending on ϖ, the ratio F/C and σ.

In terms of the rescaled system (3.2) multiple equilibria can only exist if $0 < C < \frac{2}{3}\sqrt{3}$ (independent of ϖ) as long as $F_- < F < F_+$, where

$$F_\pm^2 = \frac{8}{27\varpi}\left[1 + \frac{9C^2}{4} \pm \left(1 - \frac{3C^2}{4}\right)^{3/2}\right] \geq 0. \tag{3.5}$$

Checking the stability of the fixed points, which can be found explicitly by some algebra, is standard procedure (cf.[9]) and involves calculating the total derivative matrix (in the critical point) and examining the eigenvalues. The eigenvalue equation for critical point $n \in \{1,2,3\}$ (with R_n^2 one of the roots of (3.4)) reads

$$\left(\frac{C}{2} + \lambda\right)^2 = (3\varpi R_n^2 - 1)(1 - \varpi R_n^2).$$

A detailed analysis [7] shows that whenever (3.4) has three real roots, i.e. whenever there are three equilibria, one of them is a hyperbolic saddle point (for one of the critical points one of the eigenvalues has positive real part, and the other has negative real part).

3.5 Inviscid case

As a special case the system (3.3) is studied with $C = 0$. Because C was a coefficient of viscous damping (a damping term proportional to $\frac{\mathrm{d}V}{\mathrm{d}t}$ in the original equation) this situation will be referred to as 'the inviscid case'. If $C = 0$ the system (3.3) has the special property that it has a first integral H such that

$$\dot{X} = \frac{\partial H}{\partial Y} \qquad \dot{Y} = -\frac{\partial H}{\partial X}.$$

For simplicity the as yet arbitrary phase-angle θ is set to zero. This choice is not essential; indeed the polar form (2.29) of the differential equation shows that setting $\theta = 0$ corresponds to a shift in Φ of magnitude θ. The hamiltionian H is easily seen to be of the form

$$H(X,Y) = \frac{1}{4\varpi}\left[\varpi R^2 - 1\right]^2 - \frac{FX}{2}. \tag{3.6}$$

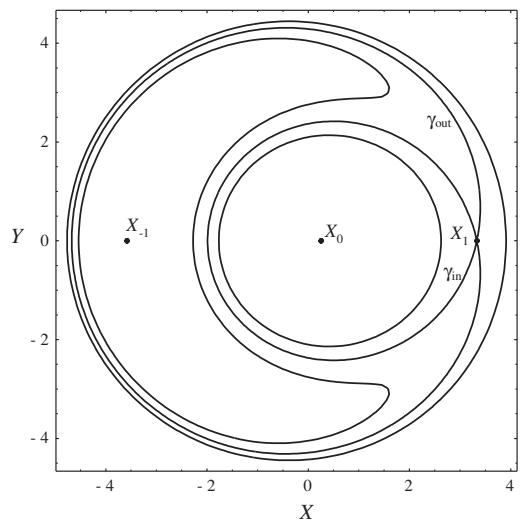


FIGURE 3.2: *Phase plane of the modulation equations (3.2) for $\varpi = 1/12$ and $F = -1/2$. Their are two homoclinic orbits emanating from the saddle-point X_1. The other stationary points are X_0 and X_1.*

As a result of the choice $\theta = 0$ the hamiltonian is symmetric in the interchange $Y \to -Y$. The critical points are located on the X-axis. From (3.5) it is apparent that multiple equilibria only exist if $0 \le |F| \le \frac{4}{3}\sqrt{1/(3\varpi)}$, which implies $F = -\frac{4}{3}\sqrt{1/(3\varpi)}\sin\alpha$ for some $\alpha \in [-\pi/2, \pi/2]$. The three steady states, which satisfy $Y = 0$ and $X(\varpi X^2 - 1) = \frac{F}{2}$, can be expressed in terms of an auxiliary angle ν, by setting

$$X_k = \sqrt{\frac{4}{3\varpi}}\sin\nu_k \qquad \nu_k = (\alpha + 2\pi k)/3$$

for $k = \{-1, 0, 1\}$. These roots are ordered in the sense that $X_{-1} < X_0 < X_1$. The saddle-point is either located at $(X_1, 0)$ if $0 < -F < \frac{4}{3}\sqrt{1/(3\varpi)}$ or at $(X_{-1}, 0)$ if F is positive. The cases $F > 0$ and $F < 0$ however are equivalent, because the equations (3.2) are invariant in the interchange $X \to -X, F \to -F$.

The phase plane is easily visualized by examining the level curves of H. Though all orbits can in fact be expressed exactly in terms of elliptic functions, only the *homoclinic* orbits will be considered explicitly. As shown in figure 3.2 there are two closed orbits emanating from the saddle point at X_1. One of these, $\gamma_{\rm out}$ encloses the other, $\gamma_{\rm in}$. The critical point X_0 is enclosed within $\gamma_{\rm in}$ while the critical point X_{-1} is contained in the region bounded by $\gamma_{\rm out}$ and not contained by $\gamma_{\rm in}$. Both X_{-1} and X_0 are centers in the inviscid case. The homoclinic orbits separate regions with $H(X,Y) > K_0$ and $H(X,Y) < K_0$ where $K_0 = H(X_1, 0)$. The regions bounded by the homoclinic orbits which contain the centers, contain periodic orbits of which the period grows to infinity as their energy level approaches K_0. A derivation of explicit expressions for the homoclinic orbit is presented by [7]. The easiest representation is in terms of the auxiliary variable

$$S = \varpi R^2 - 1 \tag{3.7}$$

which satisfies the differential equation

$$\frac{\mathrm{d}S}{\mathrm{d}T} = \varpi F Y$$

and is related to X both by (3.7) and by

$$2\varpi F X = S^2 - K$$

where $K = 4\varpi H$ is directly related to the energy level of the orbit that passes through the point specified by X and S. Combining the differential equation for S, the relation between X and S and the fact that $Y = \pm\sqrt{R^2 - X^2}$ results in

$$\frac{\mathrm{d}S}{\mathrm{d}T} = \pm\left[\varpi F^2(S+1) - \frac{1}{4}(S^2 - K)^2\right]^{1/2}.$$

For $K_0 = 4\varpi H(X_1, 0)$ the solutions with $K = K_0$ are

$$\begin{aligned} S_{\mathrm{out}}(T) &= S_1 + \left(\frac{\cosh^2(\nu T)}{S_2'} - \frac{\sinh^2(\nu T)}{S_3'} \right)^{-1}, \\ S_{\mathrm{in}}(T) &= S_1 + \left(\frac{\cosh^2(\nu T)}{S_3'} - \frac{\sinh^2(\nu T)}{S_2'} \right)^{-1}, \end{aligned} \tag{3.8}$$

specifying the outer and inner homoclinic orbits respectively. The auxiliary variables

$$S_1 = \varpi X_1^2 - 1, \quad S_{2,3} = -S_1 \pm \left(\frac{-2\varpi F}{S_1} \right)^{1/2}, \quad S_{2,3}' = S_{2,3} - S_1, \quad \text{and} \quad \nu = (-S_2' S_3')^{1/2}/4$$

were introduced to simplify the expressions. Clearly $S_{\mathrm{out,in}}(T) \to S_1$ as $|T| \to \infty$ and $S_{\mathrm{out,in}}(0) = S_{2,3}$. The points $X_{2,3}$ such that $\varpi(X_{2,3})^2 - 1 = S_{2,3}$ are the intersections of the homoclinic orbits (outer, resp. inner) with the line $Y = 0$. The explicit expressions for the homoclinic orbits will be used in chapter 4.

CHAPTER 4

MELNIKOV'S METHOD

4.1 Extra forcing

The nonlinear system (3.2) was seen to be autonomous and hamiltonian in the inviscid case; the solutions are simply level curves of the hamiltonian (3.6). Thus the single-frequency forcing assumed in (2.3) does not give rise to chaotic dynamics of the averaged small amplitude equations (also called modulation equations).

If the forcing takes a more complicated form chaotic dynamics may be present, however. A realistic situation may be a combination of forcing due to the lunar M_2 tide and the solar S_2 tide[1]; the relative frequency difference is small (of the order of 3%) and the amplitude of the lunar tide exceeds the amplitude of the solar tide significantly. Mathematically this situation can be modeled by introducing a *double-frequency forcing*:

$$f(t) = f_1 \cos(\omega_1 t + \theta_1) + f_2 \cos(\omega_2 t + \theta_2) \tag{4.1}$$

where the second component is a small perturbation to the first term, i.e. where the ratio $f_2/f_1 =: \delta \ll 1$. The dominant lunar tide is again assumed to be close to resonance with the linear Helmholtz frequency of the basin (i.e. $\omega_1 - 1 = \epsilon^2 \ll 1$). The frequency-difference between lunar and solar tide is assumed to be of the same order by setting

$$\omega_2 - \omega_1 = \epsilon^2 \Delta\Omega.$$

The forcing (4.1) can be rewritten in a form similar to (2.3) with varying F and θ by writing

$$f(t) = \tilde{f}(T) \cos(\omega_1 t + \theta(T))$$

with the help of elementary trigonometric identities. Employing the scaling $f_1 = \epsilon^3 F$ (see chapter 2) the amplitude

$$\tilde{f}(T) = f_1 \left(1 + \delta^2 + 2\delta \cos(\Delta\Omega T + \Delta\theta)\right)^{1/2} = \epsilon^3 \{F[1 + \delta \cos(\Delta\Omega T + \Delta\theta)] + O(\delta^2)\} =: \epsilon^3 F(T) + O(\delta^2)$$

and phase

$$\theta(T) = \theta_1 + \delta \sin(\Delta\Omega T + \Delta\theta) + O(\delta^2)$$

are obtained to first order in δ in terms of $T = \epsilon^2 t$ and $\Delta\theta = \theta_2 - \theta_1$.

If the modified forcing is introduced in (2.14) the vector field g is not $2\pi/\omega_1$ periodic in t anymore, seemingly prohibiting the use of the method of second order averaging in the periodic case (as described in section 2.2). However the result (2.26, 2.29) of the averaging procedure carried out in chapter 2 retains its validity if F and θ in the final result are replaced by $F(T)$ and $\theta(T)$. This is not a trivial fact; however, since the conditions that F and θ vary only on a slow time scale $T = \epsilon^2 t$ and depend smoothly on T are

[1] The notation 'M_n-tide' signifies the tide caused by the moon (due to gravitational interaction) with (roughly) n periods every 24 hours. The M_2 tide for instance is the diurnal lunar tide, with a period of 12h25 min. In contrast the S_n-tides are the tidal movements due to the sun, with the same meaning assigned to the integer n. The S_2 tide has a period of 12 hours.

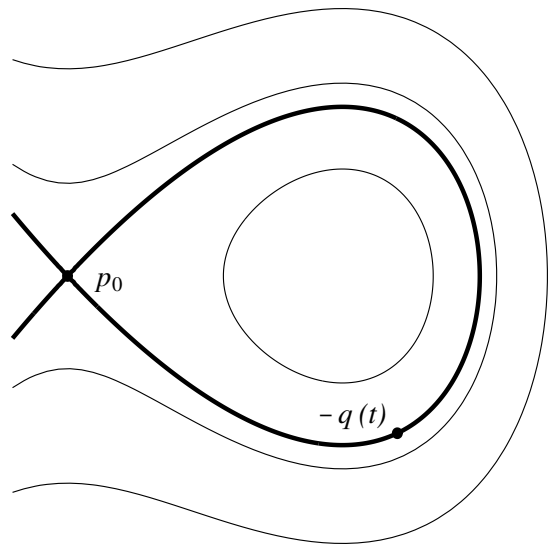


FIGURE 4.1: *The two dimensional phase space of the unperturbed system ($\delta = 0$) with one hyperbolic saddle point p_0 and one homoclinic orbit $q(t)$ such that $\lim_{t\to\infty} q(t) = \lim_{t\to-\infty} q(t) = p_0$.*

met, the procedure is justified by invoking section 2.3.5. Because F and θ vary only on the time scale $\epsilon^2 t$ (not on the time scale ϵt) the approximation remains valid on a time scale $1/\epsilon^2$.

The resulting modulation equations in cartesian form can be simplified considerably by the judicious choice $\theta_1 = 0$ and $\Delta\theta = -\pi/2$. These choices do not impose any restriction, because the first can be accomplished by a rotation in the XY-plane (equivalently a shift in fast time t) and the second is the result of a shift in slow time[2] T. The resulting system (choosing $C = \delta\tilde{C}$) has the form

$$\begin{cases} \dot{X} &= (\varpi R^2 - 1)Y + \delta\left[-\dfrac{\tilde{C}}{2}X - \dfrac{F}{2}\cos\Delta\Omega T\right] \\ \dot{Y} &= -(\varpi R^2 - 1)X + \dfrac{F}{2} + \delta\left[-\dfrac{\tilde{C}}{2}Y + \dfrac{F}{2}\sin\Delta\Omega T\right] \end{cases} \tag{4.2}$$

which is the hamiltonian system considered in section 3.5 if $\delta = 0$ (by virtue of the choice $C = \delta\tilde{C}$ (weak damping)). The assumption $0 < \epsilon \ll \delta \ll 1$ is necessary if the analysis has to be relevant to the full model (see [15]).

Such a perturbed hamiltonian system may exhibit chaotic behaviour through the mechanism described by the Smale-Birkhoff theorem (see chapter 5.4). The next section discusses the geometry of the perturbed system and the Melnikov method used to probe the occurrence of chaos analytically.

4.2 Geometry and theoretical survey

The perturbed system (4.2) is of the form

$$\dot{\mathbf{X}} = JDH(\mathbf{X}) + \delta\,\mathbf{g}(X,t)$$

where $\mathbf{X} \in \mathbb{R}^2, J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $DH(\mathbf{X})$ is the gradient of the hamiltonian of the unperturbed system. The perturbation $\mathbf{g} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ is periodic with period $2\pi/\Delta\Omega$. In this section the discussion will be

[2] As a result of the averaging procedure slow and fast time are completely uncoupled, which provides the possibility of two independent choices, which is a priori absent in the full system.

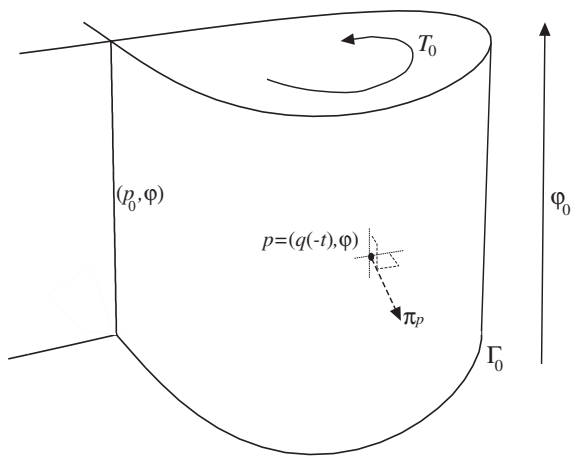


FIGURE 4.2: *The homoclinic manifold Γ_0 of the three-dimensional system (4.3) for $\delta = 0$ in $\mathbb{R}^2 \times \mathbb{R}/2\pi$. Each point on Γ_0 can be parameterized by a pair (t,φ). At a point $p = (q(-t),\varphi)$ the normal π_p is defined through the gradient of the hamiltonian.*

restricted to systems of this form, with the assumption that the unperturbed system has one hyperbolic saddle point p_0 and a homoclinic orbit q emanating from and tending to p_0. The two-dimensional phase space of such a system is shown in (figure 4.1). The perturbed system is not autonomous anymore due to the time dependence of the perturbation; however at the cost of augmenting the dimension of the system it can be cast in the autonomous form

$$\left\{ \begin{array}{rcl} \dot{\mathbf{X}} & = & JDH(\mathbf{X}) + \delta \mathbf{g}(X, \varphi/\Delta\Omega) \\ \dot{\varphi} & = & \Delta\Omega \end{array} \right. \tag{4.3}$$

where $\varphi \in \mathbb{R}/2\pi$. It may seem rather superfluous to examine the three-dimensional phase space of the new system for $\delta = 0$ because the system is not at all perturbed in this case. The three dimensional phase space of the system with $\delta = 0$ however is directly comparable to the dynamics of the perturbed system.

The saddle point p_0 of the unperturbed two dimensional system trivially corresponds to a periodic orbit $(p_0, \Delta\Omega t + \varphi_0)$, from which a two dimensional stable, as well as a two dimensional unstable invariant manifold emanate. Informally an (un)-stable invariant manifold of a saddle point p_0 or periodic orbit γ is defined as the collection of all orbits tending to the saddle point p_0 or orbit γ in the limit $t \to \infty$ resp. $-\infty$. A formal definition is presented in [9].

In the case $\delta = 0$ the stable and unstable invariant manifolds coincide. Any point p on this homoclinic invariant manifold Γ_0 can be parameterized by a pair $(t,\varphi) \in \mathbb{R} \times \mathbb{R}/2\pi$ through the map $(t,\varphi) \to (q(-t),\varphi)$ (see figures 4.1, 4.2). Of special interest is the unit normal to the homoclinic manifold at a point p. This vector will be denoted by π_p. Because Γ_0 is simply a level surface of $H(X,Y)$ the unit normal π_p at $p = (q(-t),\varphi)$ can be defined through the gradient of the hamiltonian as

$$\pi_p = \frac{(DH(q(-t)),0)}{||DH(q(-t))||}.$$

This definition makes sense for all $q(-t)$, $t \in \mathbb{R}$ since $DH(q(-t)) = (0,0)$ only occurs for the stationary point $p = p_0$ which is never reached except in the limit $|t| \to \pm\infty$.

The perturbed system also possesses a periodic orbit (structural stability [15], [16], [17],[19]), γ_δ which stays within an order δ-neighbourhood of $\{(p_0,\varphi) | \varphi \in \mathbb{R}/2\pi\}$. This fact may be formulated alternatively using the concept of the 'stroboscopic map'. In the three dimensional phase space a two dimensional section Σ^φ may be studied, which is the collection of all points with third coordinate equal to φ. The stroboscopic map, also called Poincaré map P_δ^φ is a map from Σ^φ to Σ^φ which assigns to a point $p \in \Sigma^\varphi$ its image under the flow of the perturbed differential equation. Differently said P_δ^φ assigns the point $q(t + 2\pi/\Delta\Omega)$ to the point p, where q is the solution that passes through p at time t. The Poincaré map is a discrete dynamical system on Σ^φ.

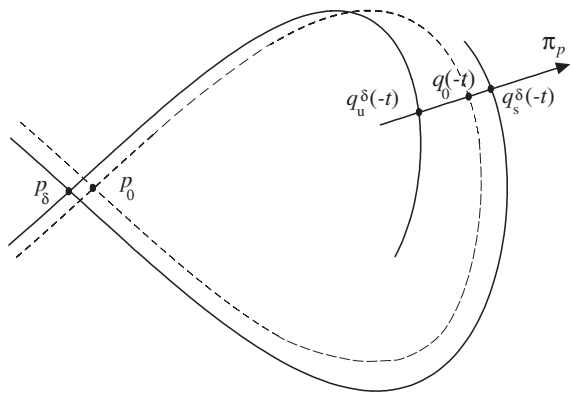


FIGURE 4.3: *A cross-section of phase space perpendicular to the φ-axis through φ_0 shows the point $p = (q(-t_0), \varphi_0)$ in the Σ^{φ_0} plane on the unperturbed homoclinic manifold. The distance in p between perturbed stable and unstable, and unperturbed manifold is measured along π_p.*

In terms of the Poincaré map P_δ^φ the intersection of γ_δ with Σ^φ is a hyperbolic saddle point (for each φ) of the Poincaré map.

The stable and unstable invariant manifolds[3] do not coincide anymore in general because the homoclinic degeneracy is not structurally stable [15], [16], [17], [19], [20]. Instead they may not intersect at all, or intersect each other transversely or tangentially. Transverse intersections are of special interest, because they guarantee the occurrence of chaotic dynamics through the Smale-Birkhoff theorem (see chapter 5).

The Melnikov method ([15], [16], [17], [18]) is an analytic tool to detect (transverse) homoclinic intersections of the perturbed stable and unstable manifolds. To this end it is relevant to note that the stable and unstable manifolds (W_s^δ and W_u^δ) of the perturbed system remain δ-close to the unperturbed homoclinic manifold Γ_0. The distance between perturbed and unperturbed invariant manifold varies from point to point. Therefore it makes sense to measure the distance at a point $p = (q(-t_0), \varphi_0)$ on Γ_0 between a perturbed manifold and the unperturbed manifold along the unit normal π_p (see figure 4.3). To this end orbits q_δ^u, q_δ^s on the unstable respectively stable perturbed manifold which pass through π_p at for instance $t = 0$ are considered. Because these orbits stay δ-close to the homoclinic invariant manifolds they may be parameterized as (see [15])

$$\begin{aligned} q_\delta^s(t, t_0, \varphi_0) &= q(t - t_0) + \delta q_1^s(t, t_0, \varphi_0) + O(\delta^2) \qquad t \in [t_0, \infty) \\ q_\delta^u(t, t_0, \varphi_0) &= q(t - t_0) + \delta q_1^u(t, t_0, \varphi_0) + O(\delta^2) \qquad t \in (-\infty, t_0] \end{aligned}$$

where the third component $\varphi_0 + \Delta\Omega t$ was not written down.

Explicitly the signed distances $d_u(t_0, \varphi, \delta)$, resp. $d_s(t_0, \varphi, \delta)$ between the unperturbed manifold and respectively the unstable and stable perturbed manifold at the point $p = (q(-t_0, \varphi_0))$ on Γ_0 are defined as

$$d_u(t_0, \varphi_0, \delta) = \frac{DH(q(-t_0)) \cdot (q_\delta^u(0, t_0, \varphi_0) - q(-t_0))}{||DH(q(-t_0))||} \tag{4.4}$$

$$d_s(t_0, \varphi, \delta) = -\frac{DH(q(-t_0)) \cdot (q(-t_0) - q_\delta^s(0, t_0, \varphi_0))}{||DH(q(-t_0))||}. \tag{4.5}$$

The distance $d(t_0, \varphi_0, \delta)$ between stable and unstable perturbed manifold along π_p is defined as the difference $d(t_0, \varphi_0, \delta) = d_u(t_0, \varphi_0, \delta) - d_s(t_0, \varphi_0, \delta)$. Melnikov's method provides a means to calculate the distances d, d_u and d_s with an order δ accuracy. Constructing a Taylor's expansion of $d_{u,s}$ with respect to δ results in

$$d_{u,s}(t_0, \varphi_0, \delta) = 0 + \delta \frac{\Delta_{u,s}(t = 0, t_0, \varphi_0)}{||DH(q(-t_0))||} + O(\delta^2)$$

[3]Note that the autonomous character of the system is essential for this notion; in a non-autonomous system orbits may intersect themselves, which makes it impossible to define a manifold as a union of orbits.

where the zeroth order term disappears because the unperturbed stable and unstable manifolds coincide. The auxiliary functions $\Delta_{u,s,t}(t=0,t_0,\varphi_0)$ can be determined by integrating the differential equation which they satisfy

$$\dot{\Delta}_{u,s}(t,t_0,\varphi_0) = DH(q(t-t_0))\cdot\mathbf{g}(q(t-t_0),\phi_0+\Delta\Omega t). \tag{4.6}$$

This nontrivial fact is derived for instance in [16], [15], [20], [19] by clever use of the Hamiltonian character of the unperturbed system in estimating the orbital derivative of the hamiltonian along an orbit of the perturbed system. Limiting the discussion to $\Delta_u(t,t_0,\varphi)$, integration from some point behind $q(-t_0)$ (i.e. some point $q(-\tau-t_0)$ with positive τ) to $q(-t_0)$ results in

$$\Delta_u(t=0,t_0,\varphi_0)-\Delta_u(-\tau,t_0,\varphi_0) = \int_{-\tau}^{0} DH(q(t-t_0))\cdot\mathbf{g}(q(t-t_0),\phi_0+\Delta\Omega t)\,\mathrm{d}t. \tag{4.7}$$

Because $||DH(q(-t))||$ goes to zero exponentially fast when $t\to\infty$ the limit $\tau\to\infty$ of the right hand side exists. In fact it can be shown that $\lim_{\tau\to\infty}\Delta_u(-\tau,t_0,\varphi_0)=0$, because $\Delta_u(-\tau,t_0,\varphi_0)=DH(q(-\tau))\cdot q_1^u(-\tau,t_0)$ which decreases exponentially as $\tau\to\infty$. A rigorous proof of the limiting behaviour of $\Delta_u(-\tau,t_0,\varphi_0)$ is provided by [15] and [16].

Similarly the deviation in energy measure of the stable perturbed orbit with respect to the homoclinic manifold reads

$$\Delta_s(t=0,t_0,\varphi_0) = -\lim_{\tau\to\infty}\int_0^{\tau} DH(q(t-t_0))\cdot\mathbf{g}(q(t-t_0),\Delta\Omega t+\varphi_0)\,\mathrm{d}t.$$

Finally the distance $d(t_0,\varphi_0)$ between the two perturbed manifolds at a point $q(-t_0,\varphi_0)$ on the homoclinic manifold reduces to

$$d(t_0,\varphi_0,\delta) = \delta\left(DH(q(-t_0))\right)^{-1}\int_{-\infty}^{\infty} DH(q(t-t_0))\cdot\mathbf{g}(q(t-t_0),\Delta\Omega t+\varphi_0)\,\mathrm{d}t. \tag{4.8}$$

The indefinite integral, often denoted by $M(t_0,\varphi_0)$, is called the 'Melnikov-function' in (t_0,φ_0).

The Melnikov-functions for the inner and outer homoclinic orbit for the perturbed system (4.2) were calculated by [7] for the case of $\varpi=1/12$. Apart from a recapitulation of the results obtained by [7] a generalization will be obtained to deal with the possibility of mixed intersections of inner invariant manifolds with outer invariant manifolds.

4.3 Application to single orbit intersections

Imitating [7] the outer homoclinic orbit is dealt with explicitly. From (4.8) one finds, after shifting the variable of integration t to $t-t_0$ (where $\tilde{C}$ is replaced by the symbol C for simplicity)

$$\begin{aligned} M_{\mathrm{out}}(t_0,\varphi_0) = &-\frac{C}{2}\int_{-\infty}^{\infty} R_0^2(\varpi R_0^2-1)\,\mathrm{d}t+\frac{FC}{4}\int_{-\infty}^{\infty} X_0\,\mathrm{d}t+ \\ &\frac{F}{2}\int_{-\infty}^{\infty} Y_0(\varpi R_0^2-1)\sin(\Delta\Omega(t+t_0)+\varphi_0)\,\mathrm{d}t \\ &+\int_{-\infty}^{\infty}\left[\frac{F^2}{4}-\frac{FX_0}{2}\left(\varpi R^2-1\right)\right]\cos(\Delta\Omega(t+t_0)+\varphi_0)\,\mathrm{d}t \end{aligned} \tag{4.9}$$

where $\gamma_{\mathrm{out}}(t)=(X_0(t),Y_0(t))$ is the outer homoclinic orbit calculated earlier and $K=K_0$ is the energy level of the homoclinic orbit (apart from a factor 4ϖ). Applying $S_0=\varpi R_0^2-1, 2\varpi FX_0=S_0^2-K$ and $\varpi FY_0=\frac{\mathrm{d}S_0}{\mathrm{d}T}$, the Melnikov function reduces to

$$\begin{aligned} M_{\mathrm{out}}(t_0,\varphi_0) \quad = \quad &-\frac{C}{8\varpi}\int_{-\infty}^{\infty}\left[3S_0^2+4S_0+K\right]\mathrm{d}t+\frac{1}{4\varpi}\int_{-\infty}^{\infty}\frac{\mathrm{d}S_0^2}{\mathrm{d}t}\sin(\Delta\Omega t+(\Delta\Omega t_0+\varphi_0))\,\mathrm{d}t+ \\ &\frac{1}{2\varpi}\int_{-\infty}^{\infty}\frac{\mathrm{d}^2S_0}{\mathrm{d}t^2}\cos(\Delta\Omega t+(\Delta\Omega t_0+\varphi_0))\,\mathrm{d}t. \end{aligned} \tag{4.10}$$

By partial integration all integrals can be written in terms of $J_n(\Delta\Omega) = \int \cos(\Delta\Omega t)\,[S_0(t) - S_1]^n\,\mathrm{d}t$ with integer n. Using the explicit formula (3.8) for S_0 these integrals reduce to an integral which can, after complexification, be calculated using the residue theorem of complex integration. The result is cited in [21] and a calculation is provided in the appendix of [7]. Collecting all the results, the expression

$$M_{\mathrm{out}}(\Delta\Omega t_0 + \varphi_0) = -\frac{2}{\varpi}\left[C\left(\psi - \frac{3\psi}{3+\tan^2\psi}\right) + \frac{\pi(\Delta\Omega)^2 e^{\psi k}}{\sinh(\pi k)}\cos(\Delta\Omega t_0 + \varphi_0)\right] \tag{4.11}$$

for the Melnikov functions concerning the splitting of the outer homoclinic manifold in $(q(-t_0), \varphi_0)$ is obtained. The auxiliary variables ψ and k are defined through $k = \Delta\Omega/(2\nu)$ and

$$F = \frac{-\frac{4}{3}\sqrt{1/(3\varpi)}\cos^2\psi}{(1+2\cos^2\psi)^{3/2}} \qquad \nu = \frac{-\tan\psi}{3+\tan^2\psi}$$

which is uniquely defined if the constraint[4] $\psi \in (\pi/2, \pi)$ is added. Using a similar procedure the analogon for the inner case is obtained

$$M_{\mathrm{in}}(\Delta\Omega t_0 + \varphi_0) = -\frac{2}{\varpi}\left[C\left(-\phi + \frac{3\tan\phi}{3+\tan^2\phi}\right) + \frac{\pi(\Delta\Omega)^2 e^{-\phi k}}{\sinh(\pi k)}\cos(\Delta\Omega t_0 + \varphi_0)\right] \tag{4.12}$$

where $\phi \in (0, \pi/2)$ equals $\pi - \psi$. The Melnikov functions M_{in} and M_{out} depend on t_0 and φ_0 only through the combination $\Delta\Omega t_0 + \varphi_0$.

Instead of analyzing the properties, specifically the zeros, of these expressions as a function of the parameters to check for the possibility of transverse intersections of the perturbed manifolds, the preceding analysis is extended to include the possibility of mixed intersections. Chapter 5 will contain the classification of parameter space in terms of the occurrence of chaotic behaviour which results from the expressions derived above and in the next section.

4.4 Intersections of inner with outer invariant manifolds

Due to the presence of *two* homoclinic orbits emanating from one saddle point in the unperturbed case the possibility of intersections of the perturbed outer stable or unstable manifolds with inner unstable or stable manifolds can not be excluded. The Melnikov method can be extended to analyze the possibilities; however a more thorough understanding of the Melnikov approach is needed.

4.4.1 Essentials

In order to generalize the Melnikov method it is essential to return to the (derivation of the) differential equation (4.6). The essential point of the Melnikov-method is, that if an orbit of the perturbed system is traced from its point of intersection with the normal π_A at some point $A = (A_X, A_Y)$ on an orbit q_0 of the unperturbed system, until it reaches the point where it intersects the unit normal π_B in $B = (B_X, B_Y)$ also[5] on q_0 the distance between perturbed and unperturbed orbit along π_B reads

$$d(t_B, \varphi_B) = \delta||DH(q_0(B_X, B_Y))||^{-1}\int_{t_B - t_A}^{0} DH(q_0(t - t_B)) \cdot \mathbf{g}(q_0(t - t_B), \Delta\Omega t + \varphi_B)\,\mathrm{d}t + O(\delta^2). \tag{4.13}$$

where t_A, t_B are such that $q_0(-t_A) = A$ and $q_0(-t_B) = B$, if arrival at π_B occurs for $\varphi = \varphi_B$. This fact is immediately apparent from the discussion in [16], because the derivation of (4.6) and (4.7) does not depend on any special assumptions that restrict the calculation to homoclinic orbits. Indeed, the property (4.13) is not only applied in Melnikov analysis of homoclinic intersections, but also in so called subharmonic Melnikov analysis, which is concerned with the detection of periodic orbits in perturbed systems.

[4] This constraint is not arbitrary but a result of details of the integration procedure.

[5] This makes sense because the orbits stay δ-close.

The discussion in chapter 5 will reveal that if, for instance, $M_{\rm out}$ has a zero as a function of t_0 and φ_0, the outer perturbed unstable manifold will intersect the outer perturbed stable manifold countably infinitely many times, and the sequence of zeroes t_0 is unbounded. In different words, there is no orbit on the unstable perturbed manifold which does not return to a neighbourhood after some finite time T. In striking contrast the perturbed unstable outer manifold will never intersect the perturbed stable outer manifold if $M_{\rm out}$ has no zeroes; instead it will wrap around the perturbed unstable inner manifold and may, or may not intersect the perturbed stable inner manifold. To analyze this situation it is necessary to

1. trace an orbit on the unstable outer manifold until it reaches some neighbourhood of the perturbed saddle-point

2. examine the behaviour in that neighbourhood

3. trace the orbit as it leaves that neighbourhood and flows along the perturbed unstable inner manifold,

under the additional assumption $M_{\rm out} \neq 0$. First of all it is necessary to define which neighbourhood of the saddle point will be considered. To this end I choose some $\eta > 0$, and consider a parallellogram in the XY-plane with vertices along the eigenvectors of $(D(JDH))(p_0)$ of length 2η centered on the saddle point p_0. The Cartesian product of this parallellogram with $\mathbb{R}/2\pi$ will be called $N(\eta)$.

A similar procedure may be used in the case that $M_{\rm in}$ has no zeroes. In that case an orbit on the inner stable manifold will (in negative time) trace the outer stable manifold after passing the saddle point. It should be noted that the outer stable manifold will not intersect the inner unstable manifold except if both $M_{\rm out}$ and $M_{\rm in}$ have zeroes. An argument to this effect is provided in chapter 5.

4.4.2 Tracking an orbit on the outer unstable manifold

We may consider an orbit starting out on the unstable outer manifold from the point of intersection with the normal π_A to the outer homoclinic manifold in A and trace it until it reaches $N(\eta)$. The extra energy separation $\Delta_{\rm out,A}$ between this orbit and an orbit q_0 of the unperturbed system starting in A reads

$$\Delta_{\rm out,A}(T_\eta, \varphi_\eta) = \int_{T_\eta - t_A}^{0} DH(q_0(t - T_\eta)) \cdot \mathbf{g}(q_0(t - T_\eta), \Delta\Omega\, t + \varphi_\eta)\, \mathrm{d}t$$

according to (4.13), where T_η is the time such that the homoclinic orbit returns to the η-neighbourhood of p_0, and φ_η is the phase at the time of arrival of the orbit at $N(\eta)$. Note that $t_A - T_\eta$ is the time of flight from A to the boundary of $N(\eta)$. By shifting the integration to match the parameterization of the homoclinic orbit q_0 defined through (3.8) with $t = 0$ in X_2 one finds

$$\Delta_{\rm out,A}(\tau_\eta, \varphi_\eta) = \int_{-\tau_A}^{\tau_\eta} DH(\gamma_{\rm out}(t)) \cdot \mathbf{g}(\gamma_{\rm out}(t), \Delta\Omega\, t + (\varphi_\eta - \Delta\Omega\, \tau_\eta))\, \mathrm{d}t$$

where τ_A is the time-of-flight from A to X_2, and τ_η is the time of flight from X_2 to the boundary of $N(\eta)$. The quantity $\Delta_{\rm out,A}$ measures the extra energy splitting between perturbed and unperturbed orbit acquired on the way from the point on π_A, the normal in A to the homoclinic orbit, to $N(\eta)$, and is of no use if the energy splitting at the point of departure A is unknown. This problem may be remedied by taking the limit $\tau_A \to \infty$, because the energy splitting goes to zero in this limit (see [16]).

In terms of the Melnikov function $M_{\rm out}$ this limiting process results in the energy separation at $(\gamma_{\rm out}(\tau_\eta), \varphi_\eta)$ of magnitude

$$\Delta_{\rm out}(\tau_\eta, \varphi_\eta) = M_{\rm out}(\varphi_\eta - \Delta\Omega\, \tau_\eta) - \int_{\tau_\eta}^{\infty} DH(\gamma_{\rm out}(t)) \cdot \mathbf{g}(\gamma_{\rm out}(t), \Delta\Omega\, t + (\varphi_\eta - \Delta\Omega\, \tau_\eta))\, \mathrm{d}t.$$

There may be some confusion about the fact that a minus sign has suddenly entered the argument of the Melnikov-function; this is a result of the fact that the energy splitting is measured in $\gamma_{\rm out}(\tau_\eta)$, while $M_{\rm out}(\Delta\Omega\, t_0 + \varphi_0)$ relates to the point $\gamma_{\rm out}(-t_0)$.

For the moment the quantity of interest is the second term, which is the deviation of the energy splitting at the boundary of $N(\eta)$ with $M_{\rm out}$ at the boundary of $N(\eta)$. This deviation is of order η when τ_η is of order $|\log(\eta)|$, since

$$\begin{aligned} 0 &\leq \left|\int_{\tau_\eta}^{\infty} DH(\gamma_{\rm out}(t))\cdot\mathbf{g}(\gamma_{\rm out}(t),\Delta\Omega t+\varphi_\eta-\Delta\Omega\tau_\eta)\,{\rm d}t\right| \\ &\leq \int_{\tau_\eta}^{\infty} |DH(\gamma_{\rm out}(t))\cdot\mathbf{g}(\gamma_{\rm out}(t),\Delta\Omega t+\varphi_\eta-\Delta\Omega\tau_\eta)|\,{\rm d}t \\ &\leq \int_{\tau_\eta}^{\infty} ||DH(\gamma_{\rm out}(t))||\;||\mathbf{g}(\gamma_{\rm out}(t),\Delta\Omega t+\varphi_\eta-\Delta\Omega\tau_\eta)||\,{\rm d}t \\ &\leq \int_{\tau_\eta}^{\infty} ||DH(\gamma_{\rm out}(-t))||\left[\frac{C^2}{4}R^2+\frac{F^2}{4}+\frac{FC}{2}(X_0+Y_0)\right]{\rm d}t \end{aligned} \tag{4.14}$$

Clearly the term in square brackets is uniformly bounded on the homoclinic orbit by some constant. Furthermore the asymptotic behaviour of $||DH(\gamma_{\rm out}(t))||$ is $||DH(\gamma_{\rm out}(t))|| \propto e^{-2\nu t}(1+O(e^{-2\nu t}))$ as $t\to\infty$, which is apparent after substitution of the parameterization (3.8) of the outer homoclinic orbit into $||DH(\gamma_{\rm out}(t))||$

$$||DH(\gamma_{\rm out}(t))|| = \sqrt{(SY)^2+\left(\frac{F}{2}-SX\right)^2} = \frac{1}{\varpi F}\sqrt{S^2\left(\frac{{\rm d}S}{{\rm d}t}\right)^2+\frac{1}{4}\left[(S_1^3-S^3)-K(S_1-S)\right]^2}$$

and noting that $|S-S_1|\propto e^{-2\nu t}$ as $t\to\infty$.

Tracing an orbit on the outer homoclinic manifold until it reaches the boundary of $N(\eta)$ at $(\gamma_{\rm out}(\tau_\eta),\varphi_\eta)$ thus results in a deviation in energy measure with respect to the homoclinic manifold of

$$\Delta_{\rm out}(\tau_\eta,\varphi_\eta) = M_{\rm out}(\varphi_\eta-\Delta\Omega\tau_\eta)+O(e^{-2\nu\tau_\eta}) \quad \text{as} \quad \tau_\eta\to\infty \tag{4.15}$$

which is in itself a rather useless expression because τ_η depends on η.

Because $|S-S_1|\propto e^{-2\nu t}$ as $t\to\infty$ the distance $||\gamma_{\rm out}(t)-p_0||$ decreases as $e^{-2\nu t}$ as $t\to\infty$. If the choice $\eta\ll 1$ and $\eta\gg\delta$ is made the orbit on the unstable outer perturbed manifold enters $N(\eta)$ after some finite time $\tau_\eta = O(|\log\eta|)\ll 1/\delta$ and deviates from the homoclinic manifold by a distance

$$\delta\left(M_{\rm out}(\varphi_\eta-\Delta\Omega\tau_\eta)+O(\eta)\right) \tag{4.16}$$

in energy.

4.4.3 Passing the saddle point

After a time τ_η the orbit on the perturbed unstable outer manifold reaches the $N(\eta)$-neighbourhood ($0\ll\eta\ll 1$) of p_0. In this small neighbourhood of p_0 the behaviour is dominated by the linear flow (see [17], [18]) generated by $D(JDH)(p_0)$. In terms of coordinates x and y along the eigenvectors of $D(JDH)(p_0)$, with x the component along the stable and y along the unstable manifolds, the flow is of the form

$$\begin{aligned} \dot{x} &= -\left(\lambda+\frac{\delta C}{2}\right)x+\delta\frac{F}{2}\cos(\Delta\Omega T) \\ \dot{y} &= \left(\lambda-\frac{\delta C}{2}\right)y+\delta\frac{F}{2}\sin(\Delta\Omega T) \end{aligned}$$

where $\lambda=\sqrt{|S_1(3S_1+2)|}$. The initial conditions are $x(0)=\beta\eta$ and $y(0)=\frac{\delta}{\eta}\alpha$ with β and α some constants. The first initial condition simply reflects the fact that the orbit enters the $N(\eta)$ neighbourhood at a (euclidean) distance of order η measured along the stable eigenvector of the stationary point by definition of $N(\eta)$. The second condition is a consequence of Eq. (4.16), which states that the distance of the

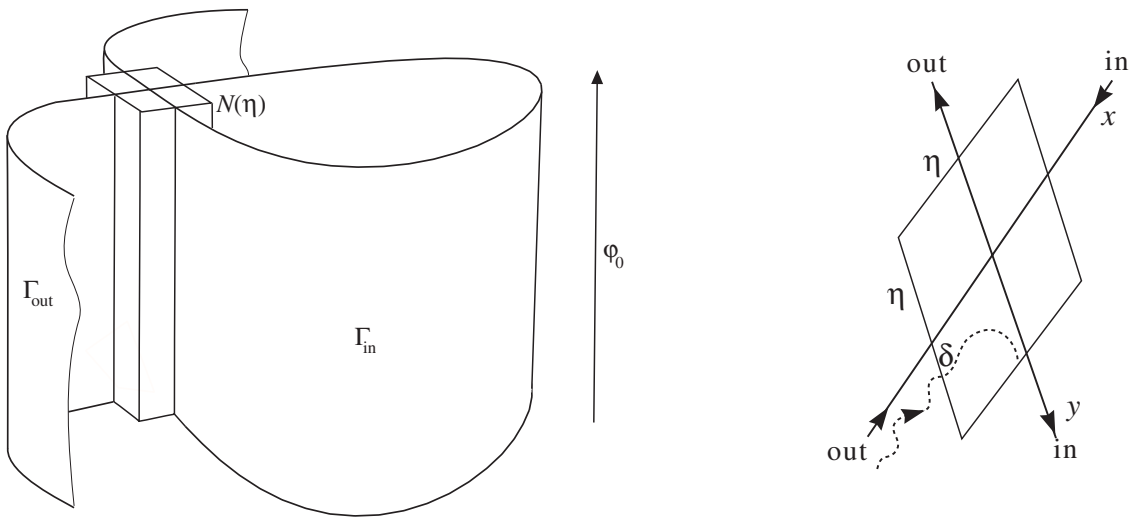


FIGURE 4.4: *The behaviour of an orbit on the outer unstable manifold is first considered by applying the Melnikov method to an orbit close to the outer homoclinic manifold* $\Gamma_{\rm out}$ *(left), until the orbit reaches the* $N(\eta)$ *neighbourhood, the Cartesian product of a parallellogram with sides of order* η*, parallel to the tangents of the homoclinic manifolds, centered on* p_0*, and* $\mathbb{R}/2\pi$*. It enters at a distance of order* η *measured along the stable manifold (*x*-coordinate) and at a distance* $\delta/\eta \ll \eta$ *along the* y*-coordinate (the eigenvector of* $D(JDH)(p_0)$ *with positive real part). After passing through* $N(\eta)$ *the Melnikov method is used to trace the orbit along the inner homoclinic manifold i.e.* $\Gamma_{\rm in}$*.*

orbit to the outer homoclinic manifold equals $\delta(M_{\rm out}(\varphi_\eta - \Delta\Omega\tau_\eta) + O(\eta))/||DH(\gamma_{\rm out}(\tau_\eta))||$ measured perpendicularly to the stable homoclinic manifold (see figure 4.4). The factor $||DH(\gamma_{\rm out}(\tau_\eta))||^{-1}$ is necessary because the usual euclidean concept of distance instead of the energy separation should be used, and results in a factor $1/\eta$. Of course this approach only makes sense if $\delta/\eta \ll \eta$, which requires $\delta \ll \eta^2$; otherwise the orbit does not enter $N(\eta)$ close to the x-axis. Due to the assumption $M_{\rm out} \neq 0$ we have $|\alpha| \neq 0$.

Problems of interest are whether it only takes a finite time to leave the neighbourhood $N(\eta)$ again, by which amount the energy H changes, and if so whether this depends crucially on η or not.

Solving these uncoupled linear inhomogeneous differential equations is a trivial exercise (see [22]) and results in expressions of the form

$$x(T) = \left(\beta\eta - \frac{\delta F}{2}\frac{\lambda+\delta C/2}{\Delta\Omega^2+(\lambda+\delta C/2)^2}\right)e^{-(\lambda+\delta C/2)T} + \frac{\delta F}{2}\{\ldots\} \tag{4.17}$$

$$y(T) = \left(\alpha\frac{\delta}{\eta} + \frac{\delta F}{2}\frac{\Delta\Omega}{\Delta\Omega^2+(\lambda-\delta C/2)^2}\right)e^{(\lambda-\delta C/2)T} + \frac{\delta F}{2}\{\ldots\} \tag{4.18}$$

where the expressions in braces are sums of $\cos(\Delta\Omega T)$ and $\sin(\Delta\Omega T)$ with constant order 1 amplitudes. The *departure condition* $|y(T_{\rm d})| \approx \eta$ for the time of departure $T_{\rm d}$ reveals that the orbit on the perturbed unstable outer manifold will leave the $N(\eta)$-neighbourhood after some finite time $T_{\rm d}$ of order $|\log\eta^2/\delta|$ (see (4.18), the exponential term dominates the terms in braces and $\lambda \gg \delta C/2$).

Furthermore the departure occurs *at the same side of the unperturbed homoclinic manifold where the orbit entered the* $N(\eta)$ *neighbourhood* because the y-coordinate can not change sign. If it would, $M_{\rm out}$ would have a zero at the point of intersection which is in contradiction with the assumption $M_{\rm out} \neq 0$. Because the orbit on the outer unstable manifold stays on the inside of the outer stable manifold (which is apparent from the sign of $M_{\rm out}$), this observation implies that the orbit will wrap around the inner homoclinic manifold instead of trailing outwards, out of the outer homoclinic manifold.

Finally the energy difference $H(q(T_{\rm d})) - H(q(0))$ should be considered. To this end I investigate the evolution of H on the orbit $q = (q_1, q_2)$ in $N(\eta)$ to which (x, y), the solution of the linearized system, is

an approximation. The fact that the orbit q satisfies (with g_1, g_2 containing the damping and forcing)

$$\begin{aligned} \dot{q_1} &= \frac{\partial H(q(t))}{\partial Y} + \delta g_1(q,t) \\ \dot{q_2} &= -\frac{\partial H(q(t))}{\partial X} + \delta g_2(q,t) \end{aligned}$$

implies that

$$\begin{aligned} \frac{\mathrm{d}H(q(t))}{\mathrm{d}t} &= \frac{\partial H(q(t))}{\partial X}\dot{q_1} + \frac{\partial H(q(t))}{\partial Y}\dot{q_2} \\ &= \delta\left[\frac{\partial H(q(t))}{\partial X} g_1(q(t),t) + \frac{\partial H(q(t))}{\partial Y} g_2(q(t),t)\right]. \end{aligned}$$

The fact that q is at a distance of at most $\sqrt{2}\eta$ of the saddle point p_0 suggests a Taylor expansion of the terms involving the Hamiltonian, resulting in

$$\frac{\mathrm{d}H(q(t))}{\mathrm{d}t} = \delta\left\{\left[D\left(\frac{\partial H}{\partial X}\right)(p_0)\cdot q(t)\right] g_1(q,t) + \left[D\left(\frac{\partial H}{\partial Y}\right)(p_0)\cdot q(t)\right] g_2(q,t) + O(\eta^2)\right\}.$$

This estimate should not be surprising to the reader, since it is simply a linearization of (4.6). Since $\frac{\mathrm{d}H}{\mathrm{d}t}$ is evidently of order $\delta\eta$ as long as q is within the $N(\eta)$-box, the energy deviation can not amount to more than

$$\left|\int_0^{T_\mathrm{d}} \frac{\mathrm{d}H(q(T))}{\mathrm{d}t}\,\mathrm{d}t\right| \leq \delta\eta T_\mathrm{d}$$

in absolute value. If the choice $\eta = \delta^k, k > 0$ is made this error is $O(\delta^{k+1}|\log\delta|)$. Of course the restriction $\delta \ll \eta^2$ should be taken into account, which dictates that k should be strictly smaller than $1/2$.

Note that the extra change in energy is $o(\delta)$, which is much smaller than the energy shift $\delta M_\mathrm{out}(\varphi_\eta - \Delta\Omega\tau_\eta)$ acquired before. This means that the orbit on the unstable perturbed outer manifold cannot cross the unstable inner manifold. Of course this is not a surprising observation because manifolds of like stability can not intersect, except if they coincide. As a proof of this statement, consider a point of intersection of two unstable manifolds and note that the backward iterate of this point under the stroboscopic map is not uniquely defined, except if it is again a point of intersection. Hence their must be an infinite number of intersections between the specifically chosen point of intersection and the saddle points from which the two manifolds emanate. Except if the two manifolds (and the stationary points) actually coincide this situation is not allowed because two curves starting from two different points (which is what the intersection of the manifolds with the Poincaré section is) must have a 'first point of intersection', i.e. a point of intersection closest along the first curve to the first starting point and closest to the second starting point along the second curve. The same reasoning can be applied to stable manifolds using forward iterations.

In summary an orbit on the perturbed unstable manifold reaches the η-neighbourhood (where $0 \leq \delta \ll \eta \ll 1$) within a time $\tau_\eta = O(|\log\delta|)$, and passes through $N(\eta)$ within a time $T_\mathrm{d} = O(|\log\delta|)$, leaving the $N(\eta)$ neighbourhood at a distance in energy measure

$$\delta\left[M_\mathrm{out}(\varphi_\eta - \Delta\Omega\tau_\eta) + O(\delta^k|\log(\delta)|)\right] \tag{4.19}$$

from the inner homoclinic manifold. Note that the error is dominated by the term due to the $N(\eta)$ neighbourhood, because of the $|\log(\delta)|$.

4.4.4 Along the inner manifold

After departure from the $N(\eta)$ neighbourhood of p_0 Melnikov's method can be used again to track the orbit as it wraps around the inner homoclinic manifold. For simplicity the orbit is tracked until it reaches the X-axis; this is not essential but serves to reduce the number of free parameters. The point of departure from $N(\eta)$ is a point δ-close to γ_in (separation dictated by (4.19)). Furthermore the phase coordinate at

the point of exit is $\tilde{\varphi} = \varphi_\eta + \Delta\Omega T_{\rm d}$. The distance $d_{\rm u,out\to in}$ between homoclinic inner manifold and the perturbed orbit on the folded unstable outer manifold thus reads (at $\gamma_{\rm in}(t=0) = X_3$)

$$d_{\rm u,out\to in} = \delta \frac{\Delta_{\rm u,out\to in}(\varphi_\eta)}{||DH(X_3)||} + O(\delta^2)$$

with

$$\begin{aligned}\Delta_{\rm u,out\to in}(\varphi_\eta) &= M_{\rm out}(\varphi_\eta - \Delta\Omega\,\tau_\eta) + O(\delta^k|\log(\delta)|) \\ &\qquad + \int_{-\tilde{\tau}_\eta}^{0} DH(\gamma_{\rm in}(t)) \cdot \mathbf{g}(\gamma_{\rm in}(t), \Delta\Omega(t+\tilde{\tau}_\eta) + \tilde{\varphi})\,{\rm d}t \\ &= M_{\rm out}(\varphi_\eta - \Delta\Omega\,\tau_\eta) + \frac{1}{2} M_{\rm in}(\Delta\Omega\,\tilde{\tau}_\eta + \tilde{\varphi}) + O(\delta^k|\log(\delta)|). \end{aligned} \tag{4.20}$$

Here the time of flight $\tilde{\tau}_\eta$ from $N(\eta)$ along $\gamma_{\rm in}$ to X_3 was introduced, and the integral was approximated by $\frac{1}{2}M_{\rm in}$ using the same estimate used to derive (4.15). The phase angle in $\mathbf{g}$ may be puzzling, but can easily be checked by noting that it should equal $\varphi_\eta + \Delta\Omega T_{\rm d}$ at $t = -\tilde{\tau}_\eta$, which is simply the phase angle at the time of entrance of the $N(\eta)$-box plus the phase increment acquired in the time span $T_{\rm d}$.

Finally it might not be clear why the two energy splittings add up; however the extra splitting $\Delta_{\rm u,out\to in}(t)$ at a point $\gamma_{\rm in}(t)$ must satisfy the ODE (4.6) together with the initial condition $\Delta_{\rm u,out\to in} = M_{\rm out}(\varphi_\eta - \Delta\Omega\,\tau_\eta)$ at $\gamma_{\rm in}(-\tilde{\tau}_\eta)$.

Summarizing the approach presented in the previous subsections, an orbit was traced on the outer unstable manifold, which wrapped itself close to the inner homoclinic manifold after reaching a small neighbourhood of the saddle point, finally reaching the Y-axis at $\varphi = \tilde{\varphi} + \Delta\Omega\,\tilde{\tau}_\eta$. To compare this orbit to an orbit on the stable inner manifold it is more useful to watch for intersections taking place at a specified $\varphi = \varphi_0$. This condition is used to eliminate φ_η and to write

$$\Delta_{\rm u,out\to in}(\varphi_0) = M_{\rm out}(\varphi_0 - \Delta\varphi) + \frac{1}{2} M_{\rm in}(\varphi_0) + O(\delta^k|\log(\delta)|)$$

for the energy splitting at (X_3, φ_0) between folded orbit on the unstable outer manifold and inner homoclinic manifold where $\Delta\varphi = \Delta\Omega(\tau_\eta + \tilde{\tau}_\eta + T_d)$ is some phase angle which is finite, though not explicitly known.

The angle $\Delta\varphi$ is a result of the fact that the orbit on the outer unstable manifold lingers near p_0 a finite, though long time with respect to the period of $\mathbf{g}$ due to the proximity of the stationary point. Theoretically the value of $\Delta\varphi$ may be calculated, but it must be noted that the time $T_{\rm d}$ is of order $|\log(\delta)|$, which is large compared to $2\pi/\Delta\Omega$.

4.4.5 Intersections with the inner stable manifold

To detect an intersection at X_3 of the folded unstable outer manifold with the inner stable manifold it is necessary to compare $\Delta_{\rm u,out\to in}(\varphi_0)$ with the distance between the stable inner manifold and the homoclinic manifold. This energy separation reads (to order δ)

$$-\Delta_{\rm in,s}(\varphi_0) = \int_0^\infty DH(\gamma_{\rm in}(t)) \cdot \mathbf{g}(\gamma_{\rm in}(t), \Delta\Omega\,t + \varphi_0)\,{\rm d}t = \frac{1}{2} M_{\rm in}(\varphi_0)$$

by standard Melnikov theory. The minus sign on the left can be explained by noting that the right hand side actually equals

$$\lim_{\tau\to\infty} \Delta_{\rm in,s}(\tau, 0, \varphi_0) - \Delta_{\rm in,s}(t=0, 0, \varphi_0)$$

using the notation of section 4.2 and the fundamental theorem of calculus. The splittings $\Delta_{\rm in,s}(\varphi_0)$ and $\Delta_{\rm u,out\to in}$ should be subtracted to result in the Melnikov function for mixed intersection in the case of $M_{\rm out} \neq 0$:

$$M^{\rm mixed}_{\rm u,out\to in}(\varphi_0) = M_{\rm out}(\varphi_0 - \Delta\varphi) + M_{\rm in}(\varphi_0). \tag{4.21}$$

This Melnikov function measures the distance in normal measure between folded outer unstable manifold and stable inner manifold at (X_3, φ_0) through

$$d_{\mathrm{u,out}\to\mathrm{s,in}} = \delta \frac{M^{\mathrm{mixed}}_{\mathrm{u,out}\to\mathrm{in}}(\varphi_0)}{||DH(X_3)||} + \mathcal{O}(\delta^{k+1}|\log(\delta)|).$$

The possibility of zeroes of this Melnikov function will be investigated in the next chapter.

4.4.6 The other way around

Previously only a Melnikov function describing the distance between outer unstable manifold and the inner stable manifold was considered. The remaining possibility that the inner unstable manifold folds along the outer stable manifold can be investigated (if M_{in} has no zeroes) along the same lines; indeed transcribing the previous calculation with all subscripts "in" interchanged with "out", and X_2 interchanged with X_3 suffices. Explicitly this Melnikov function (at (X_2, φ_0) is specified by

$$M^{\mathrm{mixed}}_{\mathrm{u,in}\to\mathrm{out}}(\varphi_0) = M_{\mathrm{in}}(\varphi_0 - \Delta\hat{\varphi}) + M_{\mathrm{out}}(\varphi_0)$$

where $\Delta\hat{\varphi}$ is again some unknown phase shift. By a shift of the coordinate φ_0 this Melnikov function can be transformed into (4.21). Again the Melnikov function measures the energy splitting with an error $\mathcal{O}(\delta^k|\log(\delta)|)$.

CHAPTER 5

CHAOS

5.1 Relevance of homoclinic intersections

In the previous chapter great pains were taken to calculate the distance between the stable and unstable manifolds of the perturbed system (4.2). The properties of the Melnikov functions derived before will be investigated to detect transverse homoclinic intersections. These are intersections of stable and unstable manifolds of the perturbed saddle point such that the manifolds are not tangential at the points of intersection. Before embarking on a quantitative analysis some motivation is presented on the relevance of homoclinic intersections.

First of all it is useful to recall the geometry of the system. In terms of the three-dimensional autonomous formulation of the previous chapter, the perturbed system has a periodic orbit which is δ-close to (p_0, φ). The intersection of this orbit with any section Σ^{φ_0} is a saddle point of the Poincaré map $P_\delta^{\varphi_0}$. Emanating from the periodic orbit are the inner stable, inner unstable, outer stable and outer unstable manifold, which are δ-close to the inner and outer homoclinic manifolds. The stable and unstable manifolds may intersect; if they do so transversely, they will not intersect in just one point, but along a one-dimensional curve through that point. This is evident from the fact that an intersection of the manifolds corresponds to a zero of a Melnikov function. If a Melnikov function $M(\Delta\Omega T + \varphi_0)$ has a zero in x, the manifolds will intersect along a one-dimensional curve defined through $\Delta\Omega T + \varphi_0 = x$.

If the stable and unstable manifold (the outer ones, for instance) intersect each other at a point (q_1, φ_0), i.e. in $q_1 \in \Sigma^{\varphi_0}$ they must intersect in an infinite sequence of other points in Σ^{φ_0}. To understand this it should be noted that q_1 is on the invariant stable manifold, which means that images of q_1 of all forward and backward iterates of the Poincaré map are also located on the stable manifold. The same reasoning holds for the unstable manifold, which leads to the conclusion that the intersection of the stable and unstable manifolds not only contains q_1 but also $(P_\delta^{\varphi_0})^n(q_1)$ for all $n \in \mathbb{Z}$. All iterates differ, because there are no stationary points in Σ^{φ_0} except the saddle point $(\gamma_\delta(t)) \cap \Sigma^{\varphi_0}$. Because each point of intersection in the plane Σ^{φ_0} corresponds to a one-dimensional curve of intersection in the three dimensional phase space, the manifolds will intersect along a countably infinite number of intersection curves if they intersect at all. This situation is illustrated in figure 5.1.

By now it may be clear that the dynamics is more easily studied in a section Σ^{φ_0} using the Poincaré map. In such a section the two dimensional manifolds correspond to one dimensional curves. Because the forward iterates of q_1 pile up on the stable manifold, near the saddle point (note that the nth iterate is past the $n-1$th iterate, following the direction of the homoclinic orbit) and the backward iterates of q_1 pile up on the unstable manifold, also coming arbitrarily close to the saddle point, the unstable manifold must form a large amount of folds closer and closer together near the saddle point, each time intersecting the stable manifold transversely. The same holds, *mutatis mutandis* for the stable manifold. Thus complicated so called 'homoclinic' tangles of intersecting crinkles arises. The Smale-Birkhoff theorem [15], [16], [17] establishes the topological equivalence of the dynamics due to homoclinic intersections and the so called Smale horseshoe map which has a chaotic invariant set (reviewed in section 5.4). Thus homoclinic intersections give rise to a chaotic invariant set (which need not be attracting however). A detailed analysis

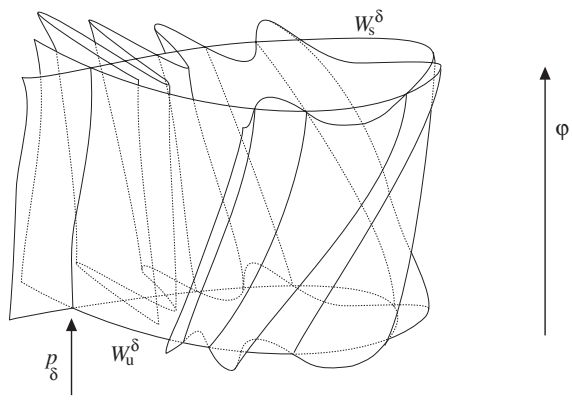


FIGURE 5.1: *The perturbed system (4.2) can be interpreted as a three dimensional autonomous system on $\mathbb{R}^2 \times \mathbb{R}/2\pi$. Limiting the attention to a perturbation of just one homoclinic manifold, the stable and unstable invariant manifold may intersect transversely. If they do so in some point they do so along a one dimensional curve through that point. Furthermore a countably infinite number of such intersection curves must exist. Because the third variable φ is taken modulo 2π the top plane and bottom plane should be identified.*

of the dynamics in the intersections of tangles is presented in [16]. Some qualitative features of the tangles will be reviewed in section 5.3.

5.2 Parameter ranges

5.2.1 Critical damping

The Melnikov functions $M_{\rm in}$, $M_{\rm out}$ and $M^{\rm mixed}$ provide an analytic tool to study the occurrence of homoclinic transverse intersections. The conditions

$$M_{\rm any}(x) = 0 \qquad \frac{{\rm d}M_{\rm any}}{{\rm d}x} \neq 0$$

where the subscript 'any' may denote 'out', 'in' or 'mixed', are sufficient to guarantee a transverse intersection at x (see [16]) of the manifolds to which $M_{\rm any}$ is relevant. The occurrence of chaotic behaviour as a function of the parameters $F, C, \Delta\Omega$ can be checked by examining the parameter range for which the Melnikov functions have zeroes. Regions of interest in parameter space are

1. the range where none of the Melnikov functions have zeroes, corresponding to the absence of a chaotic invariant set.

2. the range where both $M_{\rm in}$ and $M_{\rm out}$ have zeroes. In this case none of the two mixed Melnikov functions has any meaning. Intersection of outer with inner manifolds does occur, however (see section 5.3).

3. the region where only $M_{\rm in}$ has zeroes.

4. the region where only $M_{\rm out}$ has zeroes.

5. the region where $M_{\rm out}$ and $M^{\rm mixed}_{\rm u,in\to out}$ have zeroes and $M_{\rm in}$ does not.

6. the region where only $M_{\rm in}$ and $M^{\rm mixed}_{\rm u,out\to in}$ have zeroes and $M_{\rm out}$ does not.

7. the region where $M_{\rm in}$ and $M_{\rm out}$ do not have zeroes but $M^{\rm mixed}_{\rm u,out\to in}$ and $M^{\rm mixed}_{\rm u,in\to out}$ do.

The dynamics corresponding to these seven regions will be called respectively, *no chaos (1), double chaos (2), inner chaos (3), outer chaos (4), outer and mixed chaos ((5)* in short *no inner chaos), inner and mixed chaos ((6)* in short *no outer chaos)* and finally *only mixed chaos (7)*, in spite of the fact that these abbreviations are rather inadequate.

In fact the parameter ranges where $M^{\text{mixed}}_{\text{u,in}\to\text{out}}$ and $M^{\text{mixed}}_{\text{u,out}\to\text{in}}$ have zeroes are equal, because these are the same functions apart from a shift of variable. However the first of these Melnikov functions has no meaning if M_{in} has zeroes and vice versa the second one is useless if M_{out} has zeroes, as noted in the analysis in the previous chapter. To abbreviate the notation somewhat the term M^{mixed} will be used for both functions in the forthcoming analysis.

The parameter range in which zeroes of M_{in} and M_{out} exist are extremely easy to find because both functions are of the form (with $x = \Delta\Omega t_0 + \varphi_0$)

$$M(x) = a(b + c\cos(x))$$

where a, b and c are constants. Zeroes as a function of x occur if and only if $|c/b| \geq 1$; in the case of equality the zeroes do not satisfy the transversality condition $\frac{\mathrm{d}M}{\mathrm{d}x} \neq 0$. Transverse homoclinic intersections therefore require $|c/b| > 1$. Solving for the coefficient of friction C (and noting that negative friction coefficients are physically forbidden) the range

$$0 \leq C < C_{\text{out}} = \pi(\Delta\Omega)^2 \left| \frac{e^{\psi k}}{\sinh \pi k} \left(\psi - \frac{3\tan\psi}{3+\tan^2\psi} \right)^{-1} \right|.$$

for non-degenerate zeroes of M_{out} is found as a function of $\Delta\Omega$ and ψ from formula (4.11). Similarly M_{in} has non degenerate zeroes if

$$0 \leq C < C_{\text{in}} = \pi(\Delta\Omega)^2 \left| \frac{e^{-\phi k}}{\sinh \pi k} \left(-\phi + \frac{3\tan\phi}{3+\tan^2\phi} \right)^{-1} \right|,$$

which is obtained from equation (4.12). Slightly more work is needed to obtain a similar expression for C_{mixed}, the critical damping relevant to zeroes of M^{mixed} (whichever of the two is appropriate). The Melnikov function M^{mixed} is, possibly after a shift in x, of the form $M^{\text{mixed}}(x) = a(b + c\cos(x) + d\cos(x+\Delta x))$ ($\Delta x \in \mathbb{R}$ not known) with

$$\begin{aligned} a &= 2/\varpi, \\ b &= \left(\pi - 2\phi + \frac{6\tan\phi}{3+\tan^2\phi} \right) \\ c &= \pi(\Delta\Omega)^2 \frac{e^{\psi k}}{\sinh(\pi k)} \\ d &= \pi(\Delta\Omega)^2 \frac{e^{-\phi k}}{\sinh(\pi k)}. \end{aligned}$$

Use was made, and will be made again of the fact that $\psi + \phi = \pi$. The expressions may be simplified by using the trigonometric identity $\cos(x+\Delta x) = \cos(x)\cos(\Delta x) - \sin(x)\sin(\Delta x)$ and separating the Δx dependence through the formula $p\cos(x) + q\sin(x) = \sqrt{p^2+q^2}\cos(x+u)$ with $u = \arctan(q/p)$ a phase angle. Performing this trick results in $M^{\text{mixed}} = a[b + \sqrt{(c+d\cos(\Delta x))^2 + (d\sin(\Delta x))^2}\cos(x+u)]$ which equals $a[b + d\sqrt{1+2e^{\pi k}\cos(\Delta x) + e^{2\pi k}}\cos(x+u)]$. As a function of x this expression can only have a zero irrespective of Δx if it has a zero for the Δx at which the amplitude of the cosine reaches a minimum, i.e. $\Delta x = \pi$. The condition which results is

$$|d(1-e^{\pi k})| > |b|.$$

In terms of $\Delta\Omega$ and F (through ϕ) the parameter range for zeroes of M^{mixed} is bounded by

$$0 \leq C < C_{\text{mixed}} = \pi(\Delta\Omega)^2 \frac{e^{-\phi k}|1-e^{\pi k}|}{|\sinh(\pi k)|} \left(\left| \pi - 2\phi + \frac{6\tan\phi}{3+\tan^2\phi} \right| \right)^{-1}. \tag{5.1}$$

The three dimensional parameter space $(\Delta\Omega, F, C)$ with $C > 0$ and $-\frac{4}{3}\sqrt{1/(3\varpi)} < F < 0$ can now be divided into regions of the seven different types defined earlier. For reasons of presentation it is easiest to consider two dimensional sections of constant forcing. Figure 5.2, top left shows the surface of critical damping separating the region in parameter space which allows homoclinic intersections of any type (below the surface) from the region which does not allow homoclinic intersections. This surface is determined by the maximum of $C_{\rm in}(\Delta\Omega, F), C_{\rm out}(\Delta\Omega, F), C_{\rm mixed}(\Delta\Omega, F)$. A general characteristic is that the critical damping is identically zero for $\Delta\Omega = 0$. This case corresponds to simple single-frequency forcing which was studied in chapter 3 and indeed does not allow chaotic solutions. For $F \uparrow 0$ the critical damping also tends to 0, also consistent with chapter 3 (taking $F = 0$). Geometrical insights concerning the limiting behaviour is contained in [7]. The volume below this surface of critical damping contains the six other regions determined by

2. double chaos: $C < \min(C_{\rm in}, C_{\rm out})$
3. only inner chaos: $\max(C_{\rm out}, C_{\rm mixed}) < C < C_{\rm in}$
4. only outer chaos: $\max(C_{\rm in}, C_{\rm mixed}) < C < C_{\rm out}$
5. outer and mixed chaos: $C_{\rm in} < C < \min(C_{\rm out}, C_{\rm mixed})$
6. inner and mixed chaos: $C_{\rm out} < C < \min(C_{\rm in}, C_{\rm mixed})$
7. only mixed chaos: $\max(C_{\rm in}, C_{\rm out}) < C < C_{\rm mixed}$.

Some representative sections for constant forcing are also shown in figure 5.2. Two features which are especially striking are the relatively large size of the region of only inner chaos and the apparent absence of only outer chaos. Intersections of the outer manifolds indeed only occur together with mixed intersections.

General characteristics are the fact that the region corresponding to inner chaos consists of two lobes, one on either side of $\Delta\Omega = 0$. This is also true for double chaos. Furthermore inner and mixed chaos only occurs for $\Delta\Omega < 0$, and outer and mixed chaos only occurs for $\Delta\Omega > 0$. Finally only mixed chaos only occurs for $\Delta\Omega > 0$. In the left lobe the behaviour is ordered in the sense that there is a region of mixed and inner chaos above double chaos, with a large region of only inner chaos on top. The lobe on the right is similarly ordered, with double chaos, outer and mixed chaos and only mixed chaos in ascending order, and only inner chaos at the side close to $\Delta\Omega = 0$.

5.2.2 A problematic region

One subtlety has been overlooked in the construction of the parameter ranges presented in the previous subsection and figure 5.2. From the Melnikov function concerning mixed intersection the critical damping $C_{\rm mixed}$ proposed in (5.1) was derived, which was subsequently used in dividing parameter space into regions corresponding to different behaviour. However this critical damping bounds the region of parameters for which the Melnikov function for mixed intersections will always have zeroes *irrespective of the unknown phase parameter* $\Delta\varphi$ from above. One may similarly investigate the parameter range for which the Melnikov function for mixed intersections will never have a zero, irrespective of $\Delta\varphi$. It is easy to show along the lines of the previous paragraph that the damping range for this region is bounded from below; there will never be zeroes, irrespective of $\Delta\varphi$ if

$$C > C_{\rm no\ mixed} = \pi(\Delta\Omega)^2 \frac{e^{-\phi k}|1+e^{\pi k}|}{|\sinh(\pi k)|} \left(\left| \pi - 2\phi + \frac{6\tan\phi}{3+\tan^2\phi} \right| \right)^{-1}.$$

Obviously the behaviour of $C_{\rm no\ mixed}$ at $\Delta\Omega = 0$ is less degenerate than that of $C_{\rm mixed}$. If the parameters are such that $\max(C_{\rm mixed}, \min(C_{\rm out}, C_{\rm in})) < C \leq C_{\rm no\ mixed}$ knowledge of $\Delta\varphi$ is necessary to obtain certainty about whether or not mixed intersections will occur. In practice this means that the classification in figure 5.2 is correct for all regions showing mixed intersections. Above the surface of critical damping for inner and mixed chaos (left lobe), and mixed chaos only (lobe on the right) there is a small region of parameter space which necessitates a more detailed analysis of $\Delta\varphi$.

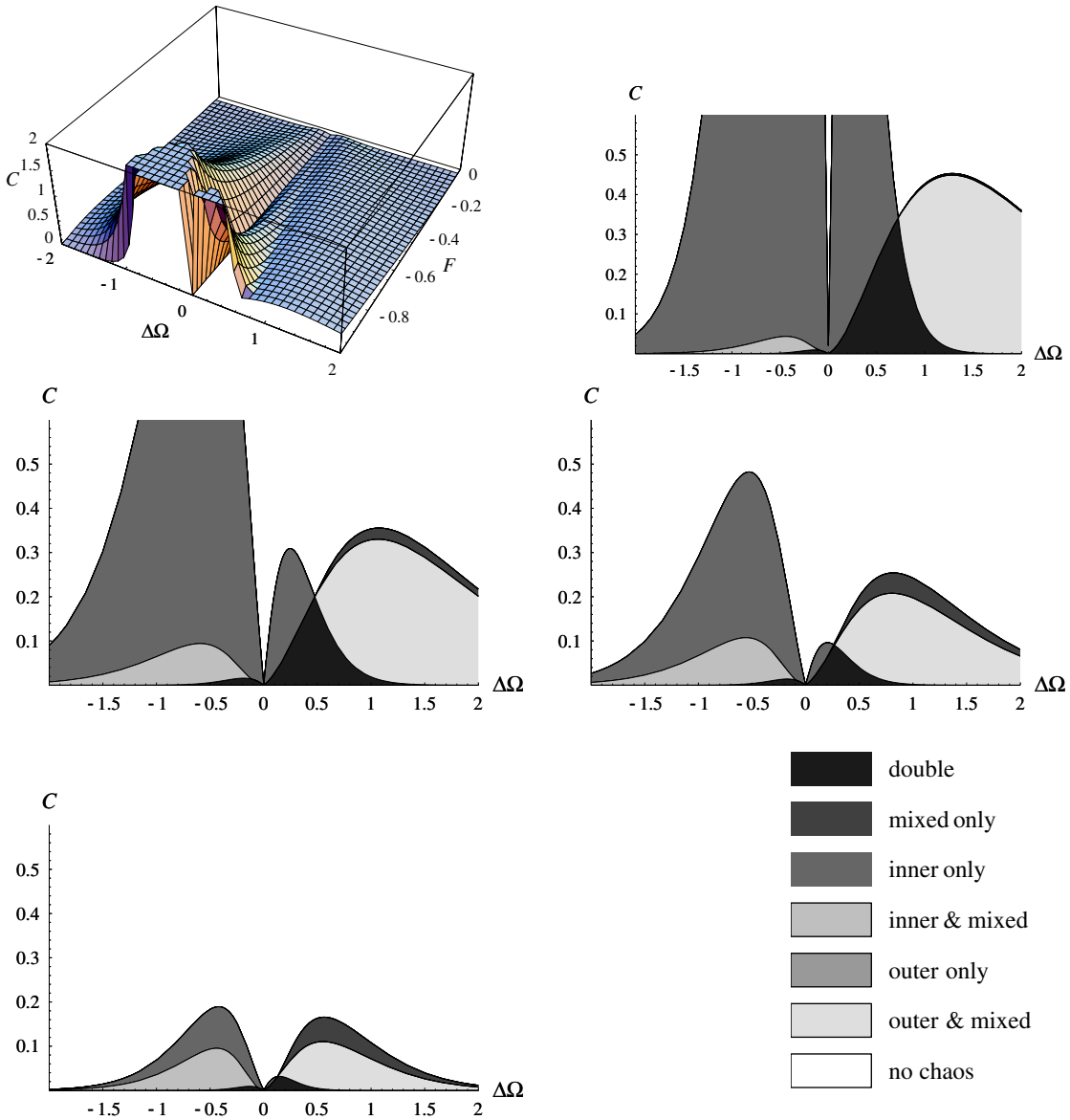


FIGURE 5.2: *At the top: the surface of critical damping which separates parameter space in two regions; above the surface the region which does not allow homoclinic intersection. For parameters chosen below the surface transverse homoclinic intersections occur. The force F was scaled by a factor $4/3\sqrt{1/3\varpi}$. The three other plots are some representative sections of the structure underneath the surface, for forcing strengths $-14/15, -9/15, -5/15$ and $-2/15$ times $4/3\sqrt{1/(3\varpi)}$. In spite of the suggestion provided by the legend, only outer chaos does not occur. Furthermore only inner chaos is the dominant mechanism. These graphs were constructed assuming C_{mixed} separates the parameter ranges for presence and absence of mixed intersections exactly. As section 5.2.2 and figure 5.3 show the situation is somewhat more subtle, becauses there is a sliver of parameter space on top of the surface of critical damping for mixed intersections for which a more detailed analysis is necessary.*

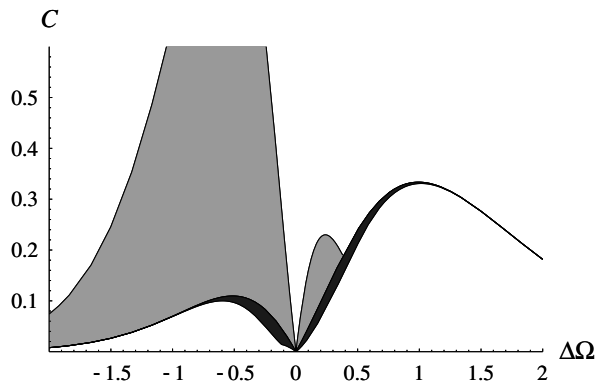


FIGURE 5.3: *A more detailed analysis concerning the phase lag $\Delta\varphi$ is necessary to ascertain the occurrence of mixed homoclinic intersections if the parameters are in the black region (a section for forcing $-9/15$ times $4/3\sqrt{1/(3\varpi)}$ is shown). The gray region bounds the parameter range corresponding to inner chaos only. The white region below the dark shaded area was classified fully and correctly in figure 5.2. The white area on top corresponds to the absence of homoclinic intersections.*

Comparison of C_{mixed} with $C_{\text{no mixed}}$ (cf. figure 5.3) shows that the difference is appreciable only in the range $|\Delta\Omega| < 1$. Furthermore the critical damping surface concerning intersections of the inner manifolds still dominates the left lobe. In conclusion, the analysis in this chapter proves that intersections of inner with outer manifolds may occur in conjunction with intersections of inner with inner, and outer with outer manifolds, since the analysis in section 5.2 is correct for the region $C < C_{\text{mixed}}$. Analysis of $C_{\text{no mixed}}$ furthermore shows that 'inner chaos only' will certainly occur for $C_{\text{no mixed}} < C < C_{\text{in}}$ which is for instance the case if $\Delta\Omega < 0$. Furthermore outer chaos only will not occur at all.

5.3 Catalogue of generic Poincaré sections

The previous section provided a classification of parameter space into six regions of interest. Each region corresponded to regimes of different homoclinic intersections. This section discusses the qualitatively different behaviour in the various cases.

First of all one may try to visualize the intersecting manifolds in three dimensions. Because the Melnikov functions are functions of the variable $\Delta\Omega t_0 + \varphi_0$ an intersection of the manifolds in a plane Σ^{φ_0} is always part of a one-dimensional curve on the homoclinic manifold which spirals from the saddle point to the saddle point. Because the Melnikov functions are periodic (for fixed φ_0) in t_0 any intersection of two manifolds in a Σ^{φ_0} results in countably infinitely many intersections in Σ^{φ_0}. (A different argument to motivate this was provided earlier using the dynamics of the Poincaré-map). Consequently one should visualize the homoclinic manifold(s) as containing a countably infinite set of one dimensional intersection curves (see figure 5.1) spiraling upwards in the φ-direction, such that the result is invariant in translations of φ by 2π. The authors of [23] provide drawings similar to figure 5.1 to illustrate the situation in the case of one single homoclinic orbit in which the top and bottom plane of figure 5.1 which must be identified are glued together to form a torus-like object.

An important result from the previous discussion is that Poincaré sections corresponding to different φ show qualitatively the same tangle, apart from trivial shifts of the lobes and points of intersection as φ is varied.

Qualitative pictures of the homoclinic tangles corresponding to the different types of intersections can be designed, recalling the generic properties

- Once two manifolds intersect they must do so in countably infinitely many points. Successive points of intersection (at least the pip's [16]) form an ordered sequence along the manifold.

- Near the saddle point the amplitude of the lobes increases. This is a consequence of the fact that

the point of intersections accumulate, while the area contained by a lobe is conserved. In fact area preservation in phase space, which is a generic property of hamiltonian flows, is only approximate because of the weak friction which was introduced. The 'amplitude' of the tangles, the maximal distance from the homoclinic manifold, may also be estimated using the fact that $||DH(q(-t))|| \propto e^{-2\nu t}$. The amplitude increases as $e^{2\nu t}$, i.e. as some rational function of the distance to the saddle point.

- Lobes of one manifold may intersect crinkles of other manifolds of different stability type (in points which are not pip's.) One should not confuse the intersection of manifolds with the intersection of orbits in the three dimensional picture; at a point of intersection an orbit in three dimensional phase space on the stable manifold merges with an orbit on the unstable manifold to form a *homoclinic connection*, i.e. an orbit connecting the saddle point with itself.

- The Poincaré map maps two adjoining lobes (necessarily one protruding outwards and one protruding inwards measured along the gradient to the hamiltonian) to the adjoining pair of lobes. This is important because intersections of lobes are mapped to intersections of lobes. If two lobes intersect, their images and preimages also intersect.

The complicated dynamics of lobes intersecting each other, and the evolution of initial conditions in a lobe, i.e. in the region enclosed by the stable and unstable manifolds between two successive points of intersection, is considered in [16]. The principles outlined above were employed to obtain the artists impressions in figures 5.4 and 5.5 which illustrate the tangle structures corresponding to the different types of dynamics distinguished before. Special interest should be directed towards the regions where the crinkles accumulate, which is where sensitive dependence on initial conditions occurs (see [16] and the next section). The seven different types of dynamics are separated into two groups, the ones without mixed intersections in figure 5.4 and the structures with mixed intersections in figure 5.5. Figure 5.4 (top graph) shows the configuration of the manifolds in the case of no zeroes of any Melnikov function (type (1) behaviour). The inner unstable manifold stays (δ-)close to the inner stable manifold, and on the inside. The outer stable manifold spirals away (for negative time), enclosing the unstable outer manifold. According to the analysis concerning mixed Melnikov functions 4.4 the outer unstable manifold bends to follow the inner stable manifold as it reaches a neighbourhood of the saddle point. According to the same analysis for the inner stable manifold, the stable inner manifold traces the outer manifold after passing through a neighbourhood of the saddle point, without ever intersecting any of the other manifolds.

The structure shown in the bottom left drawing in figure 5.4 is entirely different. The case of intersections of only the inner manifolds results in a structure of intersecting lobes which is equivalent to the structure usually seen in the case of one perturbed homoclinic orbit. The only difference, which does not affect the horseshoe map which is associated with this type of behaviour (see next section) is that the outer unstable manifold folds near the saddle point to stay close to the tangle structure. The lobe intersections occur in the region bounded by the inner manifolds. The structure in the right bottom graph of figure 5.4, corresponding to intersections of the outer manifolds only, is practically the same, except that the lobe intersections occur at the other side of the saddle point. These intersections do not occur in a region enclosed by (any of) the invariant manifolds. However, the structure in the left and in the right graph can be transformed into the same one by 'lifting' the outer homoclinic loop over the saddle point, to form a figure ∞-pattern (using an extra spatial dimension).

The dynamics shown in figure 5.5 is essentially different because of the presence of mixed intersections. The top left graph shows the scenario for zeroes of M^{mixed} only. The outer unstable manifold folds in a neighbourhood of the saddle point to intersect the inner stable manifold. Similarly, as calculated in section 4.4, the stable inner manifold folds, and intersects the outer unstable manifold. The lobe intersections occur in the region between the outer unstable and inner stable manifolds. One may wonder at the apparent asymmetry that the inner unstable and outer stable manifolds do not intersect. However, this is a result of the asymmetry introduced by damping. Physically damping requires the friction coefficient to be positive. Intuitively this means that orbits tend to spiral towards the origin in forward time. As a result the stable manifolds tend to spiral outwards, and the inner unstable manifold always contracts inwards. The inner unstable manifold for instance can not fold near the saddle point to trace or intersect an outer manifold without meeting the inner stable manifold along the route first (resulting in extra intersections of

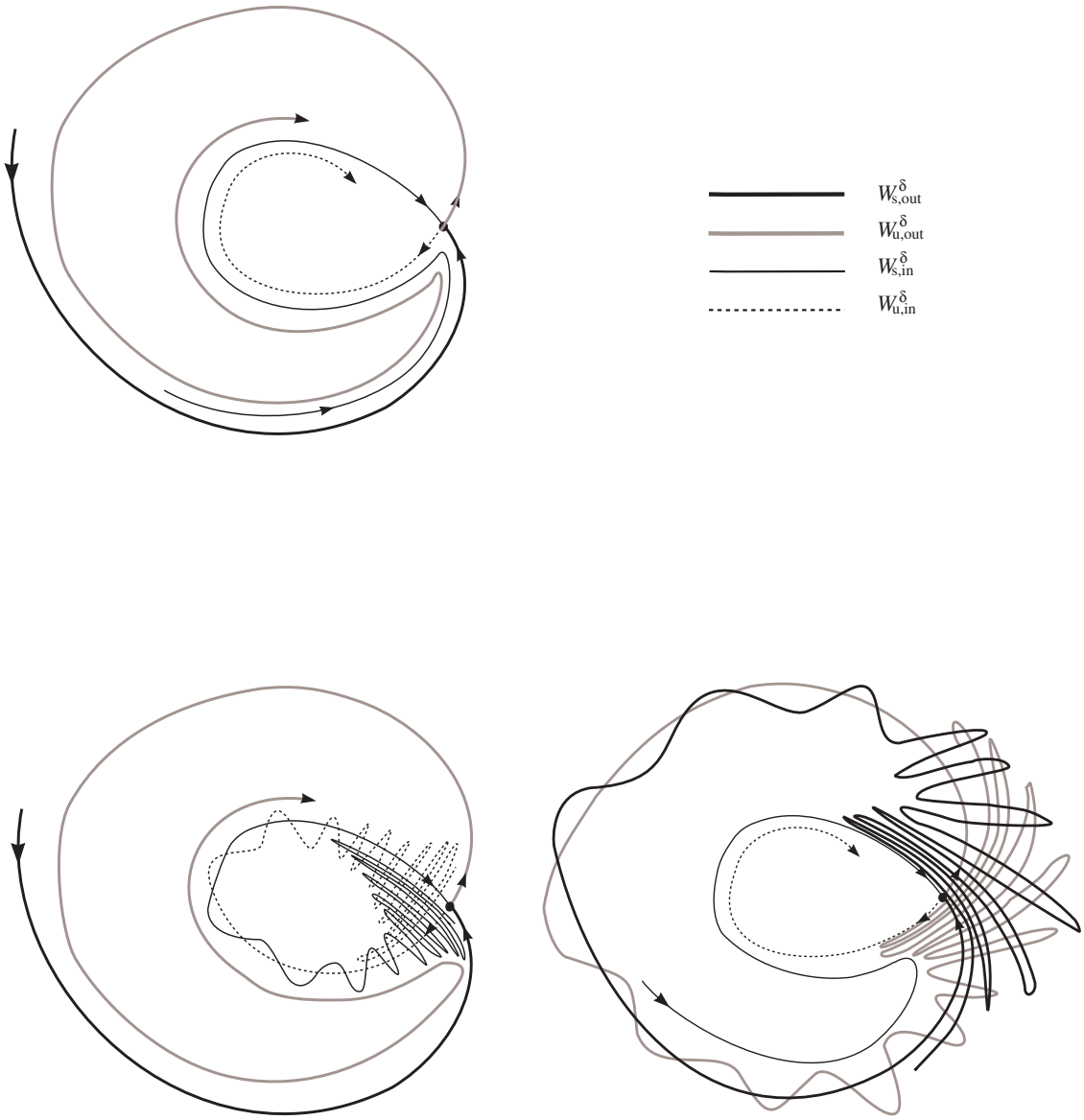


FIGURE 5.4: *Poincaré sections presenting the invariant stable and unstable manifolds in a section* Σ^{φ_0} *show homoclinic tangles, except in the first case, where the manifolds do not intersect. These three cases are the ones without intersections of outer with inner manifolds, respectively no intersections, only inner intersections and only outer intersections. The line-style coding is shown in the top right inset; the black dot represents the saddle point of the Poincaré map and the arrow heads indicate the stability type of the manifolds.*

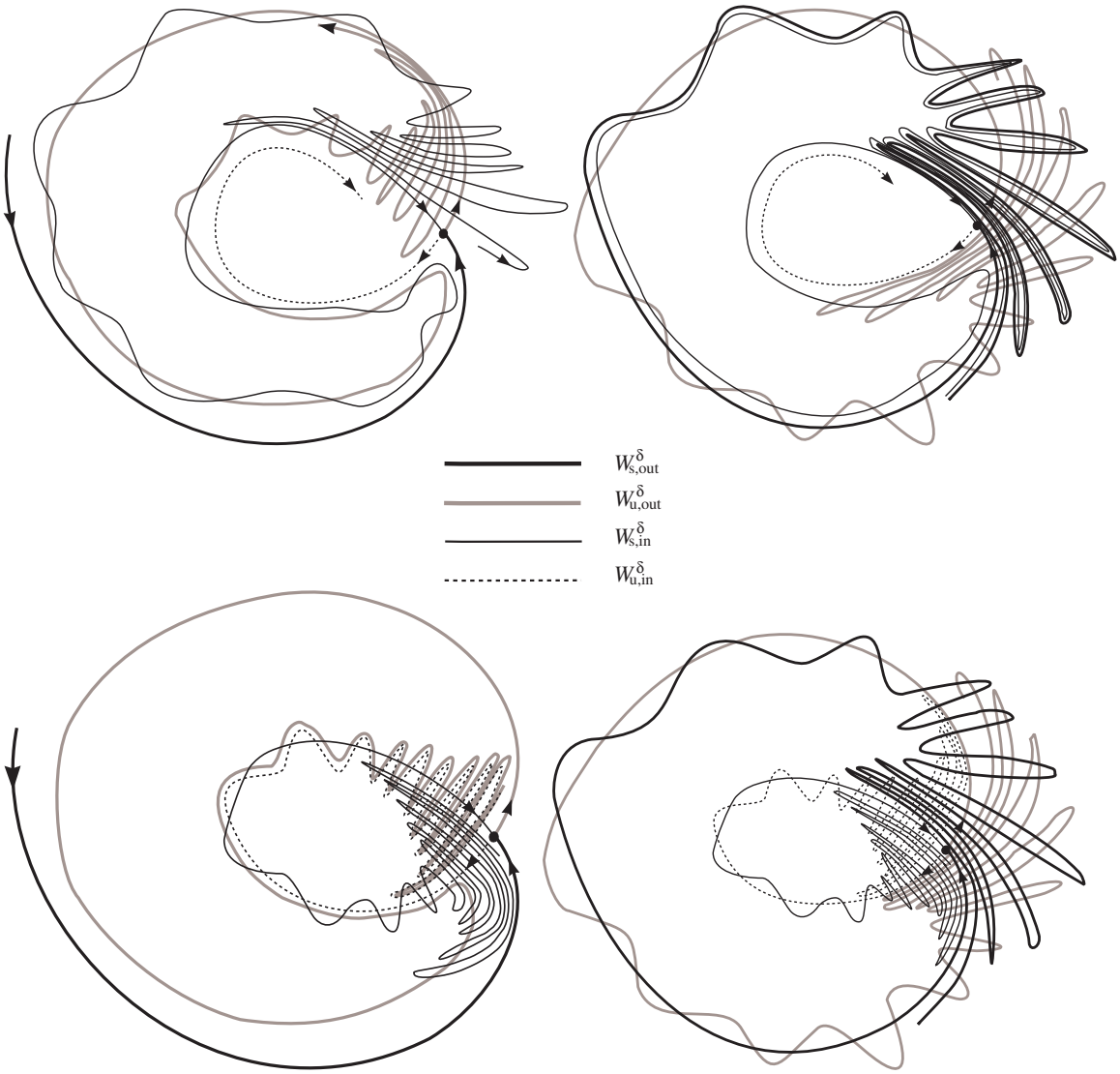


FIGURE 5.5: *The cases which include intersections of inner with outer manifolds are more complicated than the ones shown previously in figure 5.4. A full description of these cases, i.e. only mixed intersections, outer and mixed intersections, as well as inner and mixed respectively outer, inner and mixed intersections, is contained in the text.*

inner manifolds instead of just mixed intersections). Mutatis mutandis the same argument shows that the outer manifolds also have to intersect if the inner unstable and outer stable manifold are to intersect.

The case of only mixed intersections results in nearly the same tangle structure as in the case of a simple homoclinic system. This is apparent from the graph if it is interpreted as a deformed system where the outer unstable and inner stable manifolds play the role of perturbations of a homoclinic loop, and the remaining manifolds are perturbations of originally non-homoclinic manifolds.

The cases where mixed intersections are accompanied by intersections of just the outer or just the inner manifolds differ qualitatively, because lobe intersections occur in several disjoint regions of phase space. In the case of intersections of the outer manifolds with each other and mixed intersections the inner unstable manifold spirals into the region within the inner homoclinic manifold. The inner stable manifold however follows the outer stable manifold at a distance of order δ which is constant in energy measure. The outer stable manifold forms crinkles intersecting the outer unstable manifold near the saddle point (similar to the situation in 5.4, bottom right). The folded stable inner manifold forms the same lobes, though shifted perpendicularly to the outer homoclinic manifold by a shift constant in energy measure . The region outside the homoclinic manifolds, close to the saddle point therefore contains lobe intersections of three manifolds: intersection of the two outer manifolds and intersections of the inner stable manifold with the outer unstable manifold. Furthermore the lobes of the outer unstable manifold, which accumulate near the saddle point intersect the folded inner stable manifold. Each subsequent lobe must intersect the stable inner manifold nearer to the saddle point, where nearer means nearer along the stable inner manifold. Qualitatively the same picture, upon interchange of inner and outer manifolds, is obtained in the case of mixed intersections accompanied by intersections of the two inner manifolds.

Finally the case of both inner homoclinic intersections *and* outer homoclinic intersections should be considered. The analysis of 4.4 concerning the folding of a manifold in a neighbourhood of the saddle point is not valid anymore, because all the manifolds must return countably infinitely many times to any neighbourhood containing the saddle point. Instead a combination of the graphs 5.4 (bottom) and the drawings in figure 5.5 (except the last one) occurs, as in the last plot. In the region contained by the inner homoclinic manifold, the lobes of the tangles from just inner manifolds intersect (as in figure 5.4, bottom left). In the region outside the homoclinic manifolds only tangles of the outer manifold intersect (as in figure 5.4, bottom right). The region cramped between the stable inner and unstable outer manifold (close to the saddle point) contains intersections of lobes from the outer stable and inner unstable manifold. This has no equivalent in any of the other pictures, because (as mentioned before) these intersections can only occur in conjunction with intersections of the inner manifolds combined with intersections of the outer manifolds, due to the positive sign of C. This asymmetry argument seems to lose its validity in the case $C = 0$. However the case $C = 0$ always corresponds to either the dynamics labeled no chaos or to the case of double chaos. Finally the region between the outer stable and inner unstable manifold contains intersections of the tangles of the unstable outer manifold with those of the inner stable manifold. This may be interpreted as a combination of the situation in figures 5.5, top right and bottom left.

It should be noted that these artist impressions are generic; once a type of dynamics is chosen the tangle structure is always (equivalent to) the one shown in figures 5.4, 5.5.

5.4 Horseshoe construction

Several references to the so called Smale-Birkhoff theorem and the Smale horseshoe are scattered throughout the preceding sections. This section will provide an intuitive discussion, some references and applications to the special system considered previously.

The term 'Smale horseshoe' refers to a construction of a discrete dynamical system with a fractal invariant set which was introduced by the mathematician Smale in 1967. Most standard textbooks on dynamical systems discuss this example to illustrate symbolic dynamics. Briefly, a map f from the unit square $[0,1] \times [0,1]$ to the unit square is considered. This map is constructed geometrically by a combination of stretching, compressing and folding. As figure 5.6 illustrates, the unit square is first stretched vertically and compressed horizontally to obtain a long vertical strip. This strip is then folded so that two of the ends of the horseshoe overlap with the square, as in figure 5.6. Only those points which remain in the square for all iterations are considered. Indeed one may reapply the map to see where iterates and

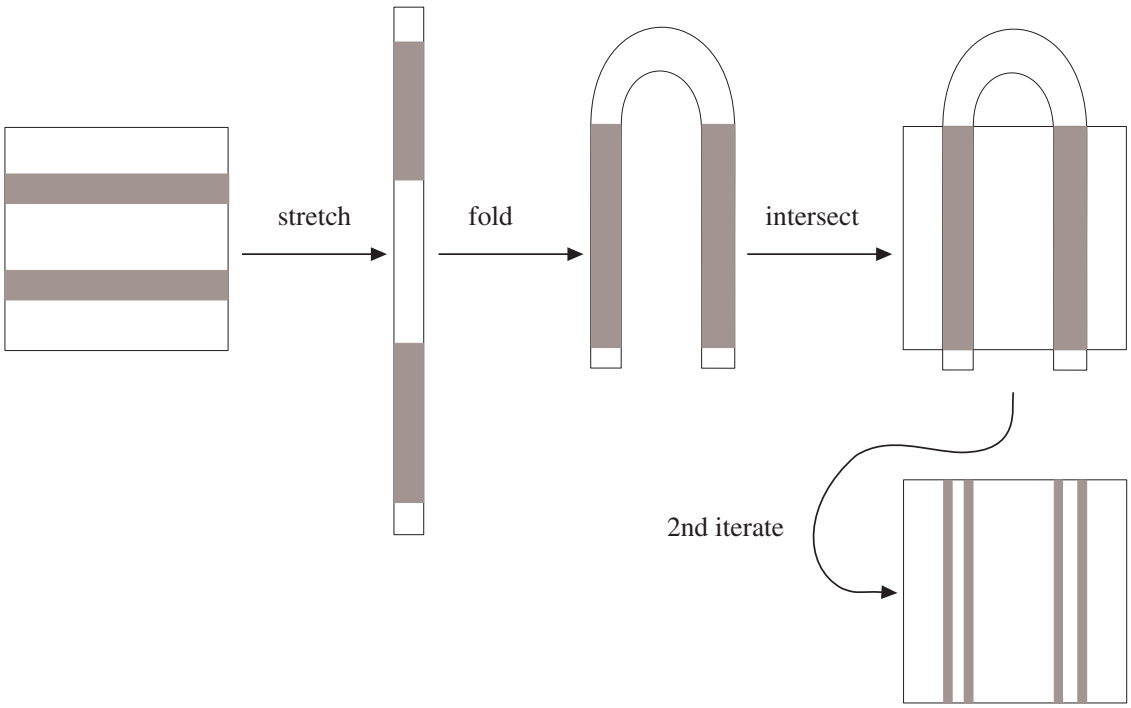


FIGURE 5.6: *Smales horseshoe map can be visualized a a stretching and folding process as shown above. The base square is stretched vertically and compressed horizontally to a long strip. This strip is then bent into a horseshoe shape, and overlapped with the square. Only points which remain in the square for all iterations are considered. The first iterate takes the two horizontal strips (left) to two vertical strips, each of which is split in two on the second iterate. The invariant set has a Cantor set structure.*

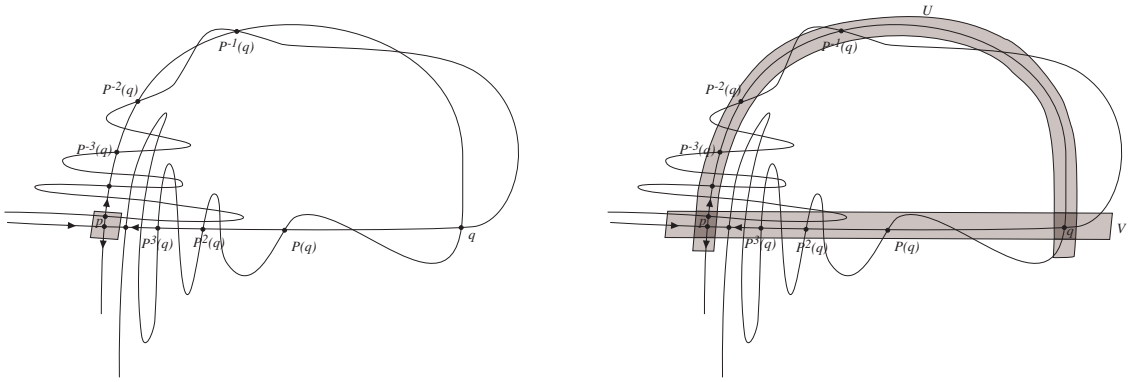


FIGURE 5.7: *The occurrence of a horseshoe map in the case of homoclinic intersections is shown geometrically by considering the forward and backward iterates under the Poincaré map P of some neighbourhood (shaded) of the saddle point p_δ of P. If the manifolds intersect in some q some N_1th forward iterate U of the neighbourhood will intersect some N_2th backward iterate V of the same neighbourhood in a set containing q. The map $P^{N_1+N_2}$ maps V to U; restricting attention to those points returning to V the map $P^{N_1+N_2}$ is just a horseshoe map*

backward iterates of f travel. For instance the shaded horizontal strips in figure 5.6, left, are mapped to two vertical strips (and the rest of the square is not mapped onto the square and therefore ignored) on the first iterate, which are mapped to four vertical strips on the second iterate. The invariant set of this map, i.e. the set of points with the property that all forward and all backward iterates are in the same set, is the Cartesian product of two Cantor sets. One can prove [15] that there exists a bijective map φ from the Cantor-set to the set of double infinite binary sequences. The composite map $\sigma = \varphi \circ f \varphi^{-1}$ is the 'shift on two symbols' that assigns the sequence with elements $b_k = a_{k+1}$ to a sequence a_k with $k \in \mathbb{Z}, a_k \in 0, 1$. This bijection is an important feat, because one can prove through symbolic dynamics on a shift of symbols that a countably infinite number of periodic orbits of all periods exist, that non-periodic orbits exist, and that a dense orbit exists which comes arbitrarily close to any point of the invariant set. 'Initial conditions' in the square can be constructed arbitrarily close to each other (by letting the corresponding symbolic sequences differ only somewhere far away in the tail) which are mapped far away from each other after a larger number of iterates, i.e. as soon as enough shifts have been applied to move the difference in the tail to the 'front' position, indexed 0.

The fractal (in this case: uncountably infinite, measure zero) character of the invariant set, the sensitive dependence on initial conditions, the existence of orbits in the invariant set of any period (or no periodicity) and the existence of a dense orbit are features to which the terms chaotic dynamics is assigned. The invariant set of the horseshoe map is a so called chaotic invariant set.

The essential properties of the horseshoe map aren't due to the fact that the domain is an exact square or the specific shape of the horseshoe, but rather the repeated stretching, compressing and twisting. The Smale-Birkhoff theorem asserts that homoclinic intersections give rise to dynamics in Poincaré sections which is topologically equivalent to the Smale horseshoe, with all the corresponding properties which were proved through symbolic dynamics.. An argument described by for instance Braaksma in [24] visualizes the occurrence of a horseshoe geometrically. As the proof provided by [15] and [16] shows this geometrical construction is the essence of the proof of the Smale-Birkhoff theorem. Consider a rectangle in a plane Σ^{φ_0} containing the saddle point, and (necessarily) intersecting the stable and unstable manifolds, (see figure 5.7). Suppose furthermore that the stable and unstable manifold intersect transversely. We may consider the image under the Poincaré map (abbreviated as P) of this rectangle, as well as all forward and backward iterates. The backward iterates are strips overlapping with part of the stable manifold, and still containing the saddle point (which is stationary) while forward iterates are strips which overlap with the unstable manifold and also contain the saddle point. Because the two manifolds intersect transversely there are integers N_1, N_2 such that the N_1th forward iterate intersects the N_2th backward iterate in some region not containing the saddle point.

The fact that these integers are finite may not be immediately apparent. However, consider a point of intersection q of the stable and unstable manifold. Forward iterates $P^n q$ of q move along the stable manifold towards the saddle point. Because the points of intersection accumulate at the saddle point (even more specifically, there is exponential attraction close to the saddle point) any finite neighbourhood of the saddle point is reached by $P^n q$ within a finite number of iterates $n \leq N_1$. A similar reasoning can be applied to backward iterates (replacing the word stable by unstable and N_1 by N_2).

Apparently the strip V which overlaps the stable manifold is compressed, stretched and folded to a strip U to intersect itself again by the map $f = P^{N_1+N_2}$. This construction shows that $N_1 + N_2$ compositions of the Poincaré map acts on the strip V as the horseshoe map f. Consequently one expects the occurrence of a chaotic invariant set, periodic orbits of all period and all other properties related to Smale's horseshoe.

This intuitive reasoning is contained in the proof of the Smale-Birkhoff homoclinic theorem [15] which asserts that if p is a hyperbolic fixed point of a map from $\mathbb{R}^n \to \mathbb{R}^n$ (the Poincaré map) and there exists a point q not equal to p in which the stable and unstable manifolds of p intersect, then f has a hyperbolically invariant set on which f is equivalent to a subshift[1] on a finite number of symbols (two symbols in the case of Smale's horseshoe).

The construction in figure 5.7 which resulted in Smale's horseshoe can be extended to the structures in figures 5.4, 5.5 to obtain more complicated horseshoe like maps, corresponding to possibly different shift maps.

5.5 Twist-and fold pictures

The general method to construct a horseshoe map is to consider a small rectangle in a graph as shown in figures 5.4, 5.5 containing the saddle point, and examine the image of such a rectangle under forward and backward iterations of the Poincaré map. The occurrence of an intersection of manifolds will result in an intersection of the backward and forward iterates of the rectangle. One should then consider the strip resulting from backward iterates as the base square, and the other strip as the partly overlapping 'horseshoe'. By deformation without changing the topological properties the resulting structures may be molded into figures like 5.6. This construction is shown graphically for the case where the outer manifolds intersect each other, and where the inner stable manifold intersects the outer unstable manifold (as in figure 5.5, top right) in figure 5.8.

The catalogue of occurring horseshoe-like maps in figure 5.9 reveals that horseshoe-like maps with two overlapping sections of base square and twisted square occur in the case of only outer and only inner homoclinic connections, while maps with three such regions are generated by the combinations of mixed and inner, mixed and outer and double chaos.

In the case of outer homoclinic intersections the horseshoe map is twisted in comparison with Smale's horseshoe (which occurs for the case of inner homoclinic intersections). In spite of the twist the symbolic dynamics is still a full shift on two symbols, i.e. the dynamics is essentially the same. This may be deduced from the fact that the bijection with biinfinite binary sequences is obtained through defining the sequence $a = \{a_i\}_{i=-\infty}^{\infty}$ corresponding to a point in the invariant set by

$$a_i = k \qquad \text{if} \qquad P^i(a) \in V_k.$$

This definition used the labeling V_k (in this case $k = 1, 2$ and in some cases 3) to enumerate the vertical strips obtained from one horseshoe iterate of the square from left to right. The enumeration and the bijection can still be used in the case of the twisted horseshoe map (though the actual trajectory of a point may differ, see [17]). This observation allows immediate classification of the map corresponding to only mixed intersections. For this case the shape of the horseshoe map is not immediately evident from the graphical construction which results in either a two legged twisted or non-twisted horseshoe. Because the symbolic dynamics is the same the difference is not important.

The various maps with three intersecting parts of the 'horseshoe' with the base square all correspond to full shifts on three symbols (also see [17], [25]). In these cases the invariant sets are also the cartesian

[1] In contrast to a full shift on n symbols, a subshift is a shift map restricted to a subset of the set of biinfinite sequences on n symbols. A subshifts occurs if the legs of the horseshoe overlap only partly with the base square.

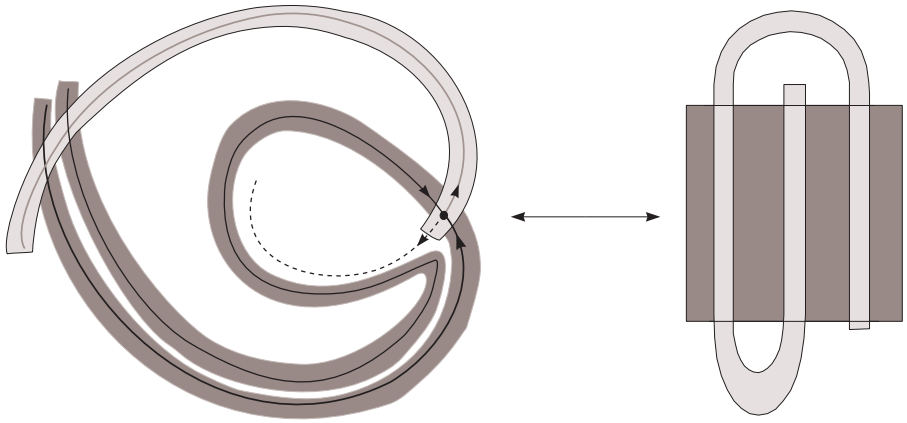


FIGURE 5.8: *The construction of a horseshoe-like map from figure 5.5, top right is shown graphically in the left picture. The dark shaded 'strip is a backward iterate of a neighbourhood of the saddle point; it overlaps the stable manifolds. The lighter shaded strip is the forward iterate of the same rectangle (at least part of it, the uninteresting part of the strip which overlaps with the inner unstable manifold is not shown here). By deforming the strips on the left it turns out that the situation is equivalent to the horseshoe map shown on the right.*

products of two Cantor like sets. These Cantor sets are not equivalent to the one which is obtained from the unit interval by the classical construction. The classical construction consists of removing the middle third segment of $(0,1)$, removing the middle thirds of the remaining segments, et cetera. The Cantor set corresponding to a shift on three symbols can be obtained by first removing the parts $(1/5,2/5)$ and $(3/5,4/5)$ from the unit interval and repeating this process indefinitely with all segments.

In all cases the existence of a countable number of periodic orbits, of any period, the existence of non-periodic orbits and of a dense orbit immediately follow, together with the fractal Cantor-structure of the invariant chaotic set. The number of orbits of a given period does differ from the number corresponding to the dynamics of a shift on two symbols. A more detailed analysis of the horseshoe-like maps, and the question whether they are unique, given a homoclinic tangle, is beyond the scope of this project.

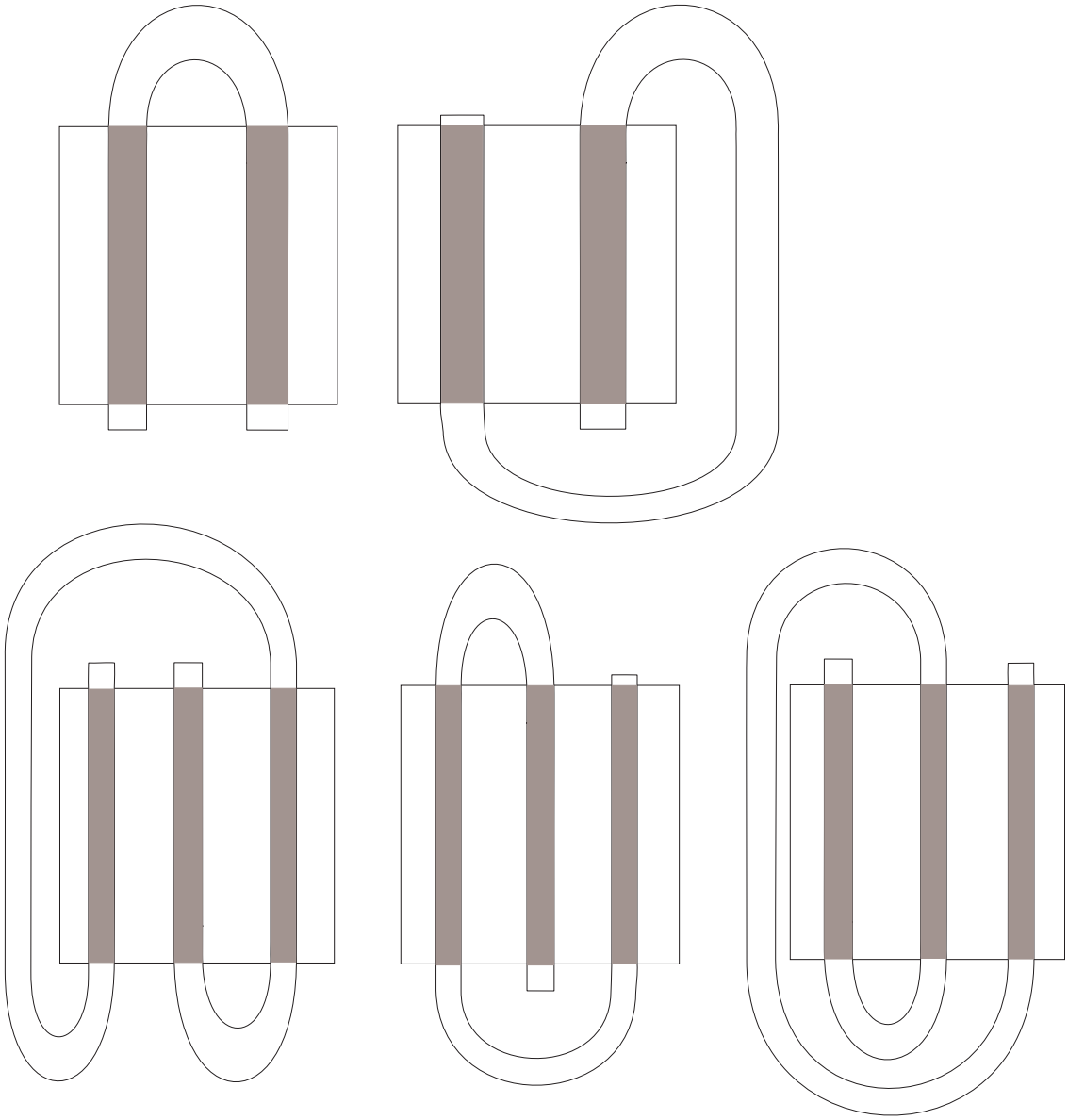


FIGURE 5.9: *The graphical construction of a horseshoe map may be repeated for the different types of dynamics. From left to right the maps for only inner and only outer chaos are shown, (top) as well as the horseshoe maps for inner and mixed, outer and mixed intersections and double chaos. The map for only mixed chaos is not shown here. As discussed in the text it also corresponds to a full shift on two symbols.*

CHAPTER 6

CONCLUSION AND DISCUSSION

In conclusion the method of averaging was used successfully to obtain an asymptotic approximation on a long time scale $1/\epsilon^2$ to the highly simplified model discussed in chapter 1. This approximation of the small amplitude behaviour of the physical model is characterized by steady oscillatory motion on a fast time scale with a slow modulation of the amplitude and phase. The evolution equations for the slow behaviour of amplitude R and phase Φ, were studied in extenso in chapter 3. An important characteristic is the presence of a bended resonance horn. This phenomenon, which is typical for nonlinear oscillators, implies multiple equilibria, thus opening the possibility of chaotic solutions in a perturbed situation. The analysis in terms of Cartesian coordinates derived from R and Φ by interpreting them as polar coordinates, showed the presence of two orbits homoclinic to one saddle point in the unperturbed, frictionless hamiltonian case. It should be noted that the equivalent Cartesian modulation equations could be derived directly from the physical model by combining second order averaging with the van der Pol transformation, instead of with the phase amplitude transformation which is not smooth at the origin. The second order averaging and the analysis in chapter 3 in fact results in an equation generic for all basin geometries. The result is even more widely applicable to any weakly damped, weakly and periodically forced nonlinear oscillator in the case of small amplitude oscillations. In this sense chapters 2 and 3 provide a direct generalization of the analysis in [7]. The presence of homoclinic orbits in the inviscid modulation equations suggests the possibility of chaos in a weakly perturbed case. Therefore an extra, weak and nearly resonant forcing was added, and weak friction was assumed. The extra forcing may be interpreted as for instance the solar tide superimposed on the stronger lunar tide. Using Melnikov's methods it was shown that the perturbed modulation equations may exhibit various types of homoclinic intersections. These types were classified geometrically according to the different possibilities in intersecting two different unstable manifolds with two stable manifolds. From the Smale-Birkhoff theorem it is concluded that the modulation equations possess chaotic invariant sets for some ranges in parameter space. The fact that chaos is present in spite of the additional damping is caused by the presence of *two* forcing frequencies.

Questions which come to mind in this context concern whether the presence of a chaotic invariant set is noticeable in numerical experiments, whether it has any relevance to the behaviour of the full (non-averaged) system and whether it is relevant in the physical situation. Proving the existence of a chaotic invariant set does not at all guarantee that the chaotic set is attracting. To prove this is a formidable task; according to [7] more work is still needed to prove this for the Duffing equation, which is probably the most widely studied nonlinear system. A problematic aspect of numerical simulations is furthermore that the Melnikov method is only valid for weak perturbations. The chaotic invariant set may become attractive, and thus visible in a numerical experiment, only for stronger perturbations. The numerical integration presented in [7], which was not performed with parameters in the regime discussed here, does suggest that the modulation equations possess a chaotic attractor. However, it will probably be extremely difficult to distinguish between the different types of behaviour in figures 5.4 and 5.5.

As mentioned before the behaviour of the modulation equations will only be relevant to the full system if $0 < \epsilon \ll \delta$ and the analysis becomes much more complicated if $\delta = O(\epsilon)$. The model does provide a possible qualitative justification of the occurrence of fundamentally irregular tides. More work and experimental verification will be necessary to validate these claims.

Bibliography

[1] Maas, L. R. M., *On the nonlinear Helmholtz response of almost-enclosed tidal basins with sloping bottom*, J. Fluid. Mech. **349**, pp. 361–380, 1997

[2] Defant, A., *Physical Oceanography (vol. II)*, Pergamon 598 p., Oxford 1961

[3] Rijn, L. C. van, *Principles of fluid flow and surface waves in rivers, estuaries, seas and oceans*, Aqua Publications 335 p., Oldemarkt 1990

[4] Ippen, A. T., *Estuary and coastline hydrodynamics*, McGraw-Hill 744 p., New-York 1966

[5] Ridderinkhof, H., *Tidal and residual flows in the western dutch Wadden Sea. I: Numerical model results*, Neth. J. Sea. Res. **22**, pp. 1–21, 1988

[6] Green, Th., *Liquid oscillations in a basin with varying surface depth*, Phys. Fluids A **4** (3), pp. 630-632, March 1992

[7] Maas, L. R. M. en Doelman, A., *Chaotic Tides*, Universiteit Utrecht Department of Mathematics, Preprint nr. 1068, 27 p., 1998

[8] Sanders, J. A. en Verhulst, F., *Averaging methods in nonlinear dynamical systems*, Springer-Verlag, 247 p., New-York 1985

[9] Verhulst, F., *Nonlinear differential equations and dynamical systems*, Springer-Verlag, 303 p. (2nd, rev. and expanded ed.), Berlin 1996

[10] Nayfeh, A. H., Mook, D. T., *Nonlinear oscillations*, Wiley, 704 p., New York 1979

[11] Yagasaki, K., *Second-order averaging and chaos in quasiperiodically forced weakly nonlinear oscillators*, Physica D **44**, pp. 445–458, 1990

[12] Perko, L. M., *Higher order averaging and related methods for perturbed periodic and quasi periodic systems*, SIAM J. Appl. Math **17**, pp. 698-724, 1969

[13] Mitropolsky, Y. A., *Problems of the Asymptotic Theory of Nonstationary Vibrations*, Israel. Progr. Sc. Transl., 385 p., Jerusalem 1965

[14] Bogoliubov, N. N. and Mitropolsky, Y. A., *Asymptotic methods in the theory of nonlinear oscillations*, transl. from [the] Russian, 537 p., Delhi 1961

[15] Guckenheimer, J. M. and Holmes, P. J., *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Springer-Verlag, 453 p., New York 1983

[16] Wiggins, S., *Introduction to applied nonlinear dynamical systems and chaos*, Springer-Verlag, 672 p., New York 1990

[17] Wiggins, S., *Global bifurcations and chaos: analytical methods*, Springer-Verlag, 494 p., New-York 1988

[18] Arrowsmith, D. K. and Place, C. M., *An introduction to dynamical systems*, Cambridge University Press, 423 p., Cambridge 1990

[19] Wiggins, S. and Holmes, P. J., *Homoclinic orbits in slowly varying oscillators*, SIAM J. Math. Anal. **18**, pp. 612–629, 1987

[20] Greenspan, B. D. and Holmes, P. J., *Homoclinic orbits, subharmonics and global bifurcations in forced oscillators* in *Nonlinear dynamics and turbulence* Ed. Barenblatt, G., Iooss, G. and Joseph, D. D., pp. 172–214, Pitman, London 1982

[21] Prudnikov, A. P., Bryckov, Yu. A. and Marichev O. I., *Integrals and series (vol. 1)*, Gordon and Breach, New York 1986

[22] Duistermaat, J. J. and Eckhaus, W., *Analyse van gewone differentiaalvergelijkingen*, Epsilon, 229 p., Utrecht 1995

[23] Graham, R. H. and Shaw, C. D., *Dynamics: the geometry of behavior (volume III)*, Aerial Press (from the 'the visual mathematics library' series), 126 p., Santa cruz 1984–1988

[24] Braaksma, B., *Gebroken dimensies en fractale verzamelingen* in *Dynamische systemen en chaos: een revolutie vanuit de wiskunde*, Broer, H. W. and Verhulst, F. (red.), Epsilon, 349 p., Utrecht 1990

[25] Ott, E., *Chaos in dynamical systems*, Cambridge University Press, 385 p., Cambridge 1993