Non-Smooth Dynamical Systems: An Overview

Markus Kunze and Tassilo Küpper

Mathematisches Institut der Universität zu Köln, Weyertal 86, D-50931 Köln, Germany, e-mail: mkunze@mi.uni-koeln.de and kuepper@mi.uni-koeln.de

Abstract. We review mathematical results on non-smooth dynamical systems, i.e., systems that incorporate effects of friction and/or impacts.

1 Introduction

Differential equations with non-smooth components occur in various situations. For example they arise in mechanical systems if the effects of dry friction are included into the model, or they occur in the case of impacts. They are present in electrical circuits or in biological systems, if non-smooth characteristics are used to represent switches. In control theory they frequently appear when discontinuous controls are involved.

In the beginning of the theory of non-smooth systems they used to be replaced by smooth approximations whose properties are much better understood. While this approach is indeed appropriate in many situations, in the course of a growing demand concerning a refined understanding of the technical models it is no longer justified to neglect effects giving rise to non-smooth phenomena and to overcome this difficulty by smoothing. Therefore to gain a better understanding of the true effects of the non-smooth parts a direct treatment is required, and correspondingly a more general concept of differential equations is needed. Such an extension has been developed within the frame of differential inclusions that permit typical behaviour such as a lack of uniqueness, non-discrete sets of equilibria, or only temporal sticking to such "rest points". While in this, by now traditional, research area the main focus has been directed towards a theory of differential inclusions that deals with standard questions like existence of solutions, properties of solution sets, etc., see [4,6,11], the aim of our project "Non-Smooth Dynamical Systems" within the DFG priority research program "Ergodentheorie, Analysis und effiziente Simulation dynamischer Systeme" was mainly to analyze the qualitative behaviour of such non-smooth models. As standard methods of dynamical systems theory rely heavily on linearization (and this requires smoothness) explicitly extensions of classical methods had to be found. Our studies were motivated by experimental and numerical investigations [35,36] on the friction oscillator. While those results clearly indicated the presence of attractors, a more reliable picture of the dynamics was not at hand. For such systems Lyapunov exponents provide an useful tool to determine the long-term behaviour. Our initial results based on Lyapunov exponents computed in a strictly formal way suggested that indeed the dynamical behaviour could be characterized by such formally computed quantities. It was our first goal to find out how the notion of Lyapunov exponents could rigorously be extended to non-smooth systems, and in case they do exist, if conclusions can be drawn from the Lyapunov exponents to the dynamics of the system. While still preliminary, mostly formal work was carried out in [41,31,13,14], the new concept has been developed in [18], see also [23] for some preliminary results. In [18] it was also shown that it is possible to verify the hypotheses at least for important prototype models. Further results on numerically computed Lyapunov exponents, in particular for multiple mass friction oscillators, are in [28]. Moreover, implications concerning the asymptotic stability of periodic solution were drawn. In Section 2 we summarize the results on Lyapunov exponents. The importance of the theoretical investigations is underlined by the fact that there are situations where "Lyapunov exponents" can be calculated in a formal way numerically, although they can be shown to exist not at all. The misleading information due to numerical shortcomings indicate the danger of any naive way of arguing only on the basis of formal manipulations without mathematical background. Concerning Lyapunov exponents in general, we took much profit from L. Arnold and his group being as well a member of the priority research program; see especially [2].

Another important tool to prove existence of non-trivial solutions or to study bifurcations are topological methods like the Conley index. In [21] it has been shown how this concept can used for non-smooth systems by extending the classical Conley index to differential inclusions. This approach relies on proper selections of single-valued maps approximating the multivalued right-hand side of the differential inclusion. In this way bifurcation results are obtained which are illustrated by an application to a model that describes the motion of bird wings, see Section 3.

Section 4 explains what are the problems encountered when one tries to use Kolmogorov-Arnold-Moser (KAM) theory for conservative non-smooth oscillators. The related classical question is if all solutions of forced oscillators like $\ddot{x} + f(x) = p(t)$ are bounded in the phase plane. Those classical results always rely on the application of a suitable twist theorem. We then tried to understand the influence of additional discontinuous terms of type $a\mathrm{sgn}(x)$ that model e.g. state-dependent kicks. It turns out that the discontinuity helps to keep the solutions bounded, since for a sufficiently large discontinuity all solutions remain bounded. On the other hand, if a is below a certain threshold (which is sharp), then the solutions are unbounded. From a mathematical point of view, the main problem was how to apply (a variant of) the twist theorem that requires high smoothness assumptions.

For non-smooth dynamical systems, planar models are easy to handle but at the same time do offer many interesting problems. The simple geometric constellation permits special arguments which cannot be applied in higher dimensions. In Section 5 the situation of two linear systems, each given in a half plane, is described. It results in a highly non-linear problem which is investigated in [12] with respect to the onset of periodic orbits. Using geometrical arguments about the form of the Poincaré return map to the discontinuity line, a complete description of the possible bifurcation scenarios could be obtained. Further studies are concerned with a generalized concept of Hopf bifurcation [30] for piecewise linear systems.

Finally in Section 6 we describe recent work on the generalization of Melnikov's method to non-smooth systems. It can be used to analyze the perturbations of planar systems with a homoclinic orbit. In [43] it is shown that for systems where sliding motion is excluded, the "non-smooth" Melnikov function contains two additional terms compared to the usual smooth one which result from impulses when the trajectory crosses the discontinuity line.

The paper is concluded by a short outlook in Section 7. Finally, the items of the bibliography are understood only as a selection. A more exhaustive list of references concerning non-smooth dynamical systems can be found at

$$\texttt{http}:\texttt{//www.mi.uni}-\texttt{koeln.de/mi/Forschung/Kuepper/Forschung1.htm}$$

These references also include a lot of papers from engineering that often focus upon numerical results, a topic which mostly has been omitted in the present paper; see also [20,41,24] for some work of our group in this direction.

2 Lyapunov exponents for non-smooth systems

The idea of Lyapunov exponents is to define characteristic numbers for a dynamical system that allow to classify the behaviour of the system is a concise manner. These numbers should account for exponential convergence or divergence of trajectories that start close to each other. To describe this in more detail, for a general smooth system $\dot{x} = f(t,x)$ consider a reference solution trajectory $x(t)$ emerging from some initial value $x_0 \in \mathbb{R}^n$. Let $\bar{x}(t)$ denote the solution with initial value $\bar{x}_0 \in \mathbb{R}^n$ close to x_0. We are interested in the time evolution of the initial difference $z_0 = \bar{x}_0 - x_0$ and write for this purpose

$$\bar{x}(t) - x(t) \approx \frac{\partial \varphi}{\partial x}(t,x_0)(\bar{x}_0 - x_0) = \frac{\partial \varphi}{\partial x}(t,x_0)z_0, \tag{1}$$

where φ denotes the flow of the system i.e., $x(t) = \varphi(t,x_0)$ and $\bar{x}(t) = \varphi(t,\bar{x}_0)$. Hence the quantity

$$\lambda^+(x_0,z_0) := \limsup_{t\to\infty} \left(\frac{1}{t}\log\left|\frac{\partial \varphi}{\partial x}(t,x_0)z_0\right|\right) \tag{2}$$

somehow describes the evolution of small perturbations of the initial value, and it is called an upper Lyapunov exponent; "upper" referring to the lim sup. These numbers can be used to make reliable predictions only in case that

$\limsup_{t\to\infty}$ does exist as a $\lim_{t\to\infty}$. A criterion for this to hold is provided by the famous ergodic theorem of Oseledets [32] that relies on some *a priori* probability measure μ, related in a natural way to the system and invariant under the flow, i.e., under the flow the μ-measure of sets in phase space remains constant. Additionally, μ is assumed to be ergodic, meaning that every flow-invariant set has μ-measure either zero or one. The theorem reads as follows.

Theorem. *Let μ be an ergodic probability measure which is invariant w.r. to the flow. Then there exists a set $\Gamma \subset \mathbb{R}^n$ of full μ-measure and $k \le n$ distinct numbers $\lambda^{(1)}, \dots, \lambda^{(k)}$, with $\lambda^{(1)} = -\infty$ possible, such that for all initial points $x_0 \in \Gamma$ and all $z_0 \in \mathbb{R}^n$ the quantity $\lambda^+(x_0, z_0)$ from (2) equals one of the $\lambda^{(i)}$, and additionally the* $\limsup_{t\to\infty}$ *is in fact the* $\lim_{t\to\infty}$.

We will only use this part of the theorem, which in its complete version contains much more additional information.

To see the difficulty with the definition of Lyapunov exponents for non-smooth systems, we consider the example

$$\ddot{x} + x + \operatorname{sgn}(\dot{x}) = \gamma \sin(\eta t). \tag{3}$$

This friction oscillator is one of the simplest dry friction models, cf. Fig. 1.

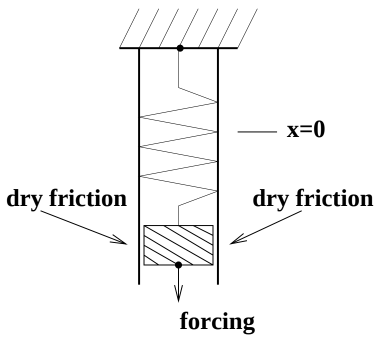

Fig. 1. A pendulum with dry friction

Equation (3) describes a pendulum that consists of a mass being attached to a spring. The mass is sinusoidally forced, it moves in a straight tube and has contact with the wall of the tube. Depending on the size of the dry friction between mass and wall, and depending on the strength of the forcing, the mass moves up or down ("slip phase") or it may stick to the wall. The term $\operatorname{sgn}(\dot{x})$ accounts for the dry friction, by Coulomb's friction law, since the friction force is proportional to the negative of the normal force (constant, since the tube is straight) multiplied by the sign of the velocity. In realistic applications this simple characteristic would have to be replaced by a more refined friction characteristics. Note, too, that in (3) we have set all physical constants equal to one, besides amplitude γ and frequency $(2\pi)/\eta$ of the forcing.

Solution trajectories of (3) may stay either in $\{\dot{x} > 0\}$ or in $\{\dot{x} < 0\}$. In that case we just have a simple oscillator. There is a third possibility that turns (3) non-linear, namely a solution trajectory may stick to $\{\dot{x} = 0\}$ for some time, corresponding to sticking of the mass to the wall. In that case we have $x(t) \equiv x_0$ in some time interval $t \in [t_1, t_2]$. Hence from (3) we find that for sticking to happen we need to have $x_0 + \mathrm{sgn}(0) = \gamma \sin(\eta t)$ for $t \in [t_1, t_2]$, and this is impossible if we simply defined $\mathrm{sgn}(0) := 0$. The proper idea is to take $\mathrm{sgn}(0)$ as the whole interval $[-1, 1]$ that "connects" $+1$ to -1. Therefore we arrive at the condition $x_0 \in [\gamma \sin(\eta t) - 1, \gamma \sin(\eta t) + 1]$ for $t \in [t_1, t_2]$. From this criterion we can decide whether a solution that hits the surface $\{\dot{x} = 0\}$ in phase space at time t_1 will just pass from one to the other side, or whether it has to rest there up to some time $t_2 > t_1$. Note that for amplitudes $\gamma \leq 1$ there are some solutions sticking forever ($t_2 = \infty$), e.g. the one with $x_0 = 0$. In particular, solutions in general do *not* cross the discontinuity surface transversely.

The mathematical theory of differential equations where the values of the right-hand side are intervals (like here $\mathrm{sgn}(0) := [-1, 1]$), or more general sets, is well-established within the frame of differential inclusion, see [4,6]. From these results we know that (3) defines a semi-flow, i.e., a flow for positive times. (Note that we cannot expect solutions to be unique in backward time, since the discontinuity surface can be left in a non-unique way in backward time.) We denote this semi-flow by $\varphi(t, y_0)$ $(t \geq 0)$ where, as (3) is non-autonomous, the initial value $y_0 = (x_0, \dot{x}_0, t_0)$ has three components: a starting time t_0, the position x_0, and the velocity $\dot{x}_0$ at time t_0, respectively. The semi-flow can be shown to be locally Lipschitz continuous w.r. to both t and y_0.

Our general goal is to define and estimate Lyapunov exponents for systems like (3), and it is already here that we encounter the first main problem. As (3) is non-smooth, the semi-flow will not be smooth w.r. to the initial value y_0, and the corresponding $\lambda^+(y_0, z_0)$ from (2) is not well-defined, since it requires $\frac{\partial \varphi}{\partial y}(t, y_0)$ to exist. To see this, imagine a trajectory that touches the discontinuity surface D tangentially ("grazing"), with a nearby initial value $\bar{y}_0$ leading to a trajectory just crossing from, say, $\{\dot{x} > 0\}$ to $\{\dot{x} < 0\}$; see Fig. 2 below.

Thus $\frac{\partial \varphi}{\partial y}(t, y_0)$ is not defined at all y_0, as would be necessary to apply Oseledets' theorem in the above version in order to have a well-defined Lyapunov spectrum. Here it helps that the theorem in fact is much more general as stated. Instead of the linearization it is only necessary to have a cocycle, i.e., a map $T : [0, \infty[\times \mathbb{R}^n \to M(n \times n)$ (the $(n \times n)$-matrices) satisfying

$$T(t+s, y_0) = T(t, \varphi(s, y_0)) \circ T(s, y_0) \quad \text{for} \quad t, s \in [0, \infty[\quad \text{and} \quad y_0 \in \mathbb{R}^n. \quad (4)$$

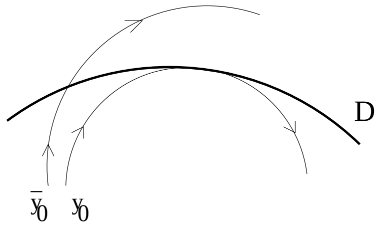

Fig. 2. A trajectory hits some discontinuity surface D tangentially

The corresponding generalization of (2) is

$$\lambda^+(y_0, z_0) := \limsup_{t\to\infty} \left(\frac{1}{t} \log |T(t, y_0)z_0| \right), \tag{5}$$

and the ergodic theorem of Oseledets holds accordingly. Observe that the identity in (4) follows from the flow equation $\varphi(t+s, y_0) = \varphi(t, \varphi(s, y_0))$ in the smooth case by differentiation w.r.t. to the initial value y_0.

Hence our problem has been shifted to defining a suitable cocycle for a non-smooth system as in (3). Of course we will set $T(t, y_0) = \frac{\partial\varphi}{\partial y}(t, y_0)$ whenever this is possible, i.e., as long as the trajectory stays in either $\{\dot{x} > 0\}$ or $\{\dot{x} < 0\}$. In the definition of $T(t, y_0)$ one certainly will have to exclude initial values y_0 like above that lead to grazing. Further, trajectories hitting the discontinuity surface infinitely often in a finite time do pose problems; this occurs e.g. with a gyroscopic pendulum that can have contact to a wall while rotating. However, it can be rigorously proved that it does not happen for examples like (3), and it can also be shown that the set of initial values which will *not* lead to grazing at some later time do form an invariant set G ("good initial values") of full Lebesgue measure in phase space. Therefore it is reasonable to define $T(t, y_0)$ only for $y_0 \in G$, and it is sufficient to clarify what happens when the flow changes from one of the regions $\{\dot{x} > 0\}$, $\{\dot{x} < 0\}$, or the "sticking region", to another, with the trajectory arriving non-tangentially. Therefore we have to derive a formula that describes how the linearization of the semi-flow is transformed while crossing a certain discontinuity surface. So we consider an initial point $y_0 \in G$ that leads to consecutive switching times $0 = t_0(y_0) < t_1(y_0) < t_2(y_0) < \ldots$ from one of the regions to another. Assume that at time $t_{i+1} = t_{i+1}(y_0)$ and at the point $\xi_{i+1} = \xi_{i+1}(y_0)$ the trajectory hits some discontinuity surface $\mathcal{D}$ which can be described as the level set $\{y : h_l(y) = 0\}$ for a smooth function h_l. Moreover, we suppose that just before the hitting the governing equation is $\dot{y} = f_l(y)$ with flow φ_l, while shortly afterwards it becomes $\dot{y} = f_p(y)$ with flow φ_p. We will derive the form of the transition matrix $A_{i+1} = A_{i+1}(y_0)$ of the linearization, i.e., A_{i+1} will satisfy

$$\lim_{t\to t_{i+1}^+} \frac{\partial\varphi}{\partial y}(t,y_0) = A_{i+1}\circ\left(\lim_{t\to t_{i+1}^-} \frac{\partial\varphi}{\partial y}(t,y_0)\right). \tag{6}$$

To get (6), note $\varphi(t,y_0)=\varphi_p(t-t_{i+1},\xi_{i+1})$ for $t\in[t_{i+1},t_{i+2}[$, hence

$$\begin{aligned}
&\frac{\partial\varphi}{\partial y}(t,y_0)\\
&=-\frac{\partial\varphi_p}{\partial t}(t-t_{i+1},\xi_{i+1})\frac{dt_{i+1}}{dy}(y_0)+\frac{\partial\varphi_p}{\partial y}(t-t_{i+1},\xi_{i+1})\frac{d\xi_{i+1}}{dy}(y_0)\\
&=-f_p(\varphi_p(t-t_{i+1},\xi_{i+1}))\frac{dt_{i+1}}{dy}(y_0)+\frac{\partial\varphi_p}{\partial y}(t-t_{i+1},\xi_{i+1})\frac{d\xi_{i+1}}{dy}(y_0).
\end{aligned}$$

Therefore

$$\lim_{t\to t_{i+1}^+} \frac{\partial\varphi}{\partial y}(t,y_0) = -\,f_p(\xi_{i+1})\,\frac{dt_{i+1}}{dy}(y_0)+\frac{d\xi_{i+1}}{dy}(y_0). \tag{7}$$

Denote $\Gamma(t,y)=\varphi_l(t-t_i(y),\xi_i(y))$. Then $\Gamma(t_{i+1}(y),y)=\xi_{i+1}(y)$, thus

$$\frac{d\xi_{i+1}}{dy}(y_0)=f_l(\xi_{i+1})\,\frac{dt_{i+1}}{dy}(y_0)+\frac{\partial\Gamma}{\partial y}(t_{i+1},y_0). \tag{8}$$

By differentiation of $0=h_l(\xi_{i+1}(y))=h_l(\Gamma(t_{i+1}(y),y))$ we additionally obtain

$$\frac{dt_{i+1}}{dy}(y_0)=-\,\frac{1}{\langle\nabla h_l,\,f_l\rangle(\xi_{i+1})}\,\langle\nabla h_l(\xi_{i+1}),\,\frac{\partial\Gamma}{\partial y}(t_{i+1},y_0)\,\cdot\rangle. \tag{9}$$

As $\frac{\partial\varphi}{\partial y}(t,y_0)\to\frac{\partial\Gamma}{\partial y}(t_{i+1},y_0)$ for $t\to t_{i+1}^-$, (6) follows from (7), (8), and (9), with the matrix

$$\begin{aligned}
&A_{i+1}\\
&:=\frac{1}{\langle\nabla h_l,\,f_l\rangle(\xi_{i+1})}\Big(\big[f_p(\xi_{i+1})-f_l(\xi_{i+1})\big]_\alpha\cdot\big[\nabla h_l(\xi_{i+1})\big]_\beta\Big)_{1\le\alpha,\beta\le n}+\mathrm{id}_{\mathbb{R}^n}
\end{aligned} \tag{10}$$

for a general problem in $\mathbb{R}^n$, $[\ldots]_\alpha$ denoting the α's component of a n-vector.

As an example, we consider (3) and a switching from $\{\dot x>0\}$ to $\{\dot x<0\}$ at time $t_{i+1}=t_{i+1}(y_0)$, for a trajectory starting at some $y_0=(x_0,\dot x_0,t_0)$. With $\xi_{i+1}:=(x_{i+1},\dot x_{i+1},t_{i+1})$, $h_l(x,v,t)=v$, $f_p(x,v,t)=(v,-x+\gamma\sin(\eta t)+1,1)$ [the "1" in the third component due to $\dot t=1$ for (3) written as an autonomous system], $f_l(x,v,t)=(v,-x+\gamma\sin(\eta t)-1,1)$, it results that here

$$A_{i+1}=\begin{pmatrix}1&0&0\\0&\theta_{i+1}^+&0\\0&0&1\end{pmatrix},\quad \theta_{i+1}^+=1-\frac{2}{x_{i+1}-\gamma\sin(\eta t_{i+1})+1}. \tag{11}$$

Returning to the general set-up, we can define

$$T(t,y_0)=\begin{cases}\frac{\partial\varphi}{\partial y}(t,y_0) & : \ y_0\in G, \\ & \quad t \text{ is no switching time} \\ A_{i+1}(y_0)\Big(\lim\limits_{s\to t_{i+1}(y_0)^-} T(s,y_0)\Big) & : \ y_0\in G\,,\ t=t_{i+1}(y_0) \\ 0 & : \ y_0\notin G \end{cases}$$

and it can be verified that $T(t,y_0)$ is a cocycle, in the sense that (4) holds. Hence we have shown

Lemma 1. *For systems like (3), a canonical cocycle can be defined. Hence Oseledets' ergodic theorem applies (if there is a suitable invariant ergodic measure), and it ensures that the corresponding Lyapunov spectrum is well-defined.*

See [18] for full details. With Lemma 1 at hand, it will be the next task to estimate the Lyapunov exponents in analytically accessible cases as (3). Here we have the following theorem.

Theorem 2. *For the pendulum with dry friction (3) we have*

$$\lambda^+(y_0,z_0)\le 0$$

for all $y_0,z_0\in\mathbb{R}^3$. *In particular, there are no positive Lyapunov exponent. This holds for every choice of amplitude* $\gamma\ge 0$ *and frequency* $(2\pi)/\eta\ge 0$.

The proof of this result is as follows. By (5) and due to the sublinearity of the exponents, it will be enough to estimate

$$|T(t,y_0)e_j|,\quad y_0\in\mathbb{R}^3,\quad j=1,2,3,$$

with the unit vectors e_j in $\mathbb{R}^3$. Due to the form of T we only need to consider "good" initial values $y_0\in G$. Such y_0 defines a sequence of switching times $0=t_0<t_1<t_2<\dots$ as above, and we have $t_i\to\infty$, since there is no accumulation point of switching times. At each t_{i+1} the cocycle is multiplied by the corresponding transition matrix A_{i+1}, cf. (10), and it can be shown by analyzing all possible cases of switchings that each A_{i+1} will be of the form (11), where possibly θ^+_{i+1} is replaced by

$$\theta^-_{i+1}=1-\frac{2}{\gamma\sin(\eta t_{i+1})+1-x_{i+1}}$$

or by $\theta^0_{i+1}=0$, if the trajectory switches to the sticking region. Then $\theta^+_{i+1}\in[0,1]$ and $\theta^-_{i+1}\in[0,1]$, as is obtained by an analysis of the vector field. Hence all transition matrices like (11) are of norm less or equal to unity. Note that this nicely reflects the dissipative nature of (3), and the largest dissipation occurs when the pendulum sticks to the wall for a certain time.

To estimate $|T(t, y_0)e_j|$, according to the definition of T we have

$$T(t, y_0) = B_{i+1}(t) \circ \Big(\lim_{s \to t_{i+1}^-} T(s, y_0) \Big)$$

for $t \in [t_{i+1}, t_{i+2}[$ with a suitable matrix $B_{i+1}(t)$ arising from the linearization of the semi-flow. This may be iterated to yield for integers i and $t \in [t_{i+1}, t_{i+2}[$

$$T(t, y_0) = B_{i+1}(t) \circ B_i(t_{i+1}) \circ \ldots \circ B_1(t_2) \circ T_1^- \tag{12}$$

with $T_1^- = \lim_{s \to t_1^-} T(s, y_0)$. The proper estimation of (12) is rather technical, but it can be shown, mainly because of $\theta_{i+1}^+ \in [0, 1]$ and $\theta_{i+1}^- \in [0, 1]$, that with each additional multiplication of a matrix $B_{i+1}(t)$ the norm in (12) can increase at most as $\approx (t - t_{i+1})$; see [17]. Hence we obtain the bound

$$|T(t, y_0)e_j| \approx \Big((t - t_{i+1}) + \sum_{k=0}^{i} (t_{k+1} - t_k) \Big) = t,$$

and therefore by (5),

$$\lambda^+(y_0, e_j) \leq \limsup_{t \to \infty} \left(\frac{\log t}{t} \right) = 0.$$

This concludes a short sketch of the proof of Theorem 2.

For the specific example of the pendulum in (3) it turns out that the Lyapunov exponents do reflect the dynamical behaviour of the system, as can be checked in the resonant case $\eta = 1$ where the situation is completely clarified. See [17], also for further references, and [28] for corresponding numerical results. In a general situation there can be derived some conclusions from the Lyapunov spectrum about the stability of the underlying system.

Theorem 3. *Under some reasonable assumption for the non-smooth system: If for a periodic orbit all Lyapunov exponents are negative (besides one being zero), the periodic solution is orbitally asymptotically stable.*

See [28, Kor. 4.2] for the details.

3 Conley index theory for non-smooth systems

In this section we outline extensions to differential inclusions of the Conley index which is a powerful tool to determine the occurrence of bifurcations. The change of some kind of index usually is a sign for bifurcations in a nonlinear dynamical system $\dot{x} = f_\lambda(x)$ depending on some parameter λ. Since we are interested in properties of the flow, the index should be connected to changes of the flow. In Ch. Conley's approach, some index is assigned to special sets (as the equilibria), and this index will account for changes in the phase portrait when λ is varied. The key notions are "isolating neighborhoods" and "isolated invariant sets". An isolating neighborhood is a set such that no boundary point stays in the set forever (in forward and backward time). In a 1D-example, every interval having no equilibrium as an endpoint will serve as an isolating neighborhood. Equivalently, an isolating neighborhood is a set U which contains its maximal invariant subset $\mathrm{inv}(U)$ in the interior. An isolated invariant set I is determined through the requirement that it is the maximal invariant subset of some isolating neighborhood.

To define the Conley index of an isolated invariant set, suppose that for a given isolated invariant set I there exists a special isolating neighborhood B (called an isolating block) such that each boundary point leaves B immediately either in forward or in backward time, and denote b^+ the set of all boundary points that leave B immediately in forward time. Then we can determine the homotopy type of the quotient B/b^+ and define an index as $h(I) = [B/b^+]$, this homotopy type; this only amounts to identifying all points in b^+, they are collapsed to a single distinguished point. It can be shown that in general an isolating block exists, and, moreover, that the index does not depend on the special choice of the isolating block, i.e., if B_1, B_2 are isolated blocks with exit sets b_1^+, b_2^+ and $\mathrm{inv}(B_1) = I = \mathrm{inv}(B_2)$, then $[B_1/b_1^+] = [B_2/b_2^+]$.

The usefulness of this index rests on its properties.

(a) (Existence Principle.) If $I = \mathrm{inv}(U)$ and $h(I) \neq \overline{0}$ (the pointed one-point space), then $\mathrm{inv}(U) \neq \emptyset$, i.e., there exists an initial value in the isolating neighborhood U such that the full forward and backward trajectory through this initial value is contained in the isolating neighborhood.

(b) (Continuation Theorem.) In the flow of a parameter-dependent problem $\dot{x} = f_\lambda(x)$ is continuous in $\lambda \in [a, b]$, and if U is an isolating neighborhood on all λ-levels between a and b, the Conley index of the generated isolated invariant set is independent of λ, i.e., $h(\mathrm{inv}(U); f_\lambda)$ is constant.

As an application of (b), we derive a criterion for bifurcation. Suppose that zero is a fixed-point of all the systems $\dot{x} = f_\lambda(x)$ for $\lambda \in [a, b]$ (i.e., $f_\lambda(0) = 0$) such that $\{0\}$ is an isolated invariant set. Then we have

(c) (Criterion for Bifurcation.) In the above set-up, if $h(\{0\}; f_a) \neq h(\{0\}; f_b)$, then some $\lambda_0 \in [a, b]$ is a bifurcation point for small non-trivial solutions of $\dot{x} = f_\lambda(x)$. This means that for $\varepsilon > 0$ there exists $\lambda \in [a, b]$ and a global non-trivial solution x of $\dot{x} = f_\lambda(x)$ such that $|\lambda - \lambda_0| + |x(t)| \leq \varepsilon$ for all times $t \in \mathbb{R}$.

If λ varies from a to b, a jump of the Conley index causes bifurcation. To prove this, it follows directly from (b) and the definition of a bifurcation point that in case there would be no bifurcation point the function $\lambda \mapsto h(\{0\}; f_\lambda)$ were (locally) constant on $[a, b]$, a contradiction. The Conley index also enjoys an addition property, i.e., the index of a disjoint union of two isolated invariant sets is the "sum" of the respective homotopy types. See [29,37] and others for much more information about classical Conley index theory.

Now we turn to a generalization of this useful apparatus to non-smooth systems. For illustration we consider the example

$$\ddot{x} + \lambda\dot{x} + x^3 + \varphi(x, \dot{x})\,\mathrm{sgn}(\dot{x}) = 0, \tag{13}$$

with some function $\varphi(x, \dot{x})$ and the bifurcation parameter $\lambda \in \mathbb{R}$. Equations of this type are used to model wings of insects; see [38]. Here, as in the last section, sgn(0) should be considered to be the whole interval $[-1, 1]$. Thus if we rewrite (13) with $x' = v$ as the first-order system

$$(x, v)' \;=\; F_\lambda(x, v) = \Big(v,\; -\lambda v - x^3 - \varphi(x, v)\,\mathrm{sgn}(v)\Big), \tag{14}$$

then the right-hand side $F_\lambda(x, v)$ for $v = 0$ is given by

$$F_\lambda(x, 0) = \{0\} \times [-\varphi(x, 0) - x^3, \varphi(x, 0) - x^3]. \tag{15}$$

Since this is a set, we will more exactly write $\dot{y} \in F_\lambda(y)$, but note this reduces to (14), with $y = (x, v)$, for $v \neq 0$. As such differential inclusions $\dot{y} \in F_\lambda(y)$ can have several solutions, we first need to adapt the definition of an isolated invariant set. We introduce

$$\begin{aligned} \mathrm{inv}(U; F_\lambda) := \big\{ y_0 : \text{ there exists } \textit{some} \text{ solution of } \dot{y} \in F_\lambda(y), y(0) = y_0, \\ \text{that remains in } U \text{ for all times} \big\}. \end{aligned}$$

Therefore it is only required that at least one of all possible solutions stayed in U. Observe that in general the system $\dot{y} \in F_\lambda(y)$ will not generate a semi-flow, not even what is called a "multi-valued (semi-) flow".

We call I an isolated invariant set, if $I = \mathrm{inv}(U) \subset U$ for some isolating neighborhood U, and we are going to define a (generalized) Conley index $H(I; F)$ that reduces to the Conley index $h(I; f)$ introduced before, if restricted to usual dynamical systems $\dot{y} = f(y)$ instead of the multi-valued $\dot{y} \in F(y)$.

The key to such extension of Conley index theory is provided by the continuation theorem, cf. (b) above, that may also be seen as follows: if a smooth right-hand side can be deformed continuously to another smooth right-hand side, the Conley index does not change along the deformation. Now it is known that set-valued maps like $F_\lambda(x, v)$ in (14) can be approximated through smooth ones; just think of e.g. approximating

$$\mathrm{sgn}(v) = \begin{cases} -1 & : \quad v < 0 \\ [-1, 1] & : \quad v = 0 \\ 1 & : \quad v > 0 \end{cases} \tag{16}$$

by e.g. $(2/\pi)\arctan(v/\varepsilon)$, $\varepsilon \to 0^+$. Hence given an error bound ε we will find, roughly speaking, a smooth function $f : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$f(y) \in F(y) + \text{ “small error of order } \varepsilon\text{”}, \tag{17}$$

and for this f we already do have a Conley index $h(I; f)$, by the classical theory. Therefore it is tempting to define

$$H(I; F) := h(I; f). \tag{18}$$

Now (18) is indeed justified, since the continuation theorem can be used to prove that this assignment is independent of the choice of the approximation f, if the error bound ε is sufficiently small. In fact, for two approximations f_0 and f_1 it turns out that the deformation $\lambda \mapsto f_\lambda = (1-\lambda)f_0 + \lambda f_1$ of f_0 into f_1 is appropriate for an application of the continuation theorem, and thus $h(I; f_0) = h(I; f_1)$. Hence $H(I; F)$ is well-defined. It also can be shown that an isolated invariant set I for F is an isolated invariant set for the approximation f; see [21] or [17, Sec. 6] for the details.

To summarize, we now have defined a Conley index $H(I; F)$ for problems of the form $\dot{y} \in F(y)$, and it can be seen that all the essential properties of the classical Conley index h are transferred to H via (18). Hence we have

Lemma 4. *The existence principle (a), the continuation theorem (b), and the criterion for bifurcation (c) also hold for the generalized Conley index H.*

As an illustration, we show how the criterion for bifurcation (c) can be used in (13) with some continuous $\varphi(x, v)$ being superlinear in v, i.e., we require $|\varphi(x, v)| \le C|v|^{1+\alpha}$ for some constants $\alpha > 0$ and $C > 0$. Written in the form (14), we first observe that $(x, v) = (0, 0)$ is an equilibrium for all $\lambda \in \mathbb{R}$, since $\varphi(0, 0) = 0$, and therefore $F_\lambda(0, 0) = (0, 0)$, cf. (15). Hence we have to verify that $\{(0, 0)\}$ is an isolated invariant set for F_λ and that $H(\{(0, 0)\}, F_\lambda)$ jumps when λ crosses zero.

We first consider the case $\lambda > 0$. Define the Lyapunov function $V(x, v) = \frac{1}{2}v^2 + \frac{1}{4}x^4$, and $U_r = \{(x, v) \in \mathbb{R}^2 : V(x, v) < r\}$. Then we have the bounds

$$2C|v|^\alpha \le \lambda \quad \text{and} \quad |\varphi(x, v)| \le C|v|^{1+\alpha}, \quad (x, v) \in U_r, \tag{19}$$

if r (depending on λ) is chosen small enough. We claim that $\mathrm{inv}(U_r; F_\lambda) = \{(0, 0)\}$, i.e., $\{(0, 0)\}$ is an isolated invariant set for F_λ with isolating neighborhood U_r. To see this, assume that there is a solution $X = (x, \dot{x}) : \mathbb{R} \to U_r$ of (14) with $X(0) = (x_0, v_0) \neq (0, 0)$. We write (13) in the form

$$\ddot{x} + \lambda\dot{x} + x^3 + \varphi(x, \dot{x})\, w(t) = 0,$$

with some function $w : \mathbb{R} \to [-1, 1]$ that realizes $\mathrm{sgn}(\dot{x}(t))$, i.e., $w(t) \in \mathrm{sgn}(\dot{x}(t))$, the multi-valued signum function from (16). By definition of $w(t)$

and (19) we obtain

$$\begin{aligned}\frac{d}{dt}V(X(t)) &= \dot{x}\,[\ddot{x}+x^3] = \dot{x}\,[-\lambda\dot{x}-\varphi(x,\dot{x})\,w(t)]\\ &= -\lambda\,|\dot{x}|^2-\varphi(x,\dot{x})\,|\dot{x}| \le -\lambda\,|\dot{x}|^2+C\,|\dot{x}|^{2+\alpha}\\ &\le -(\lambda/2)\,|\dot{x}|^2.\end{aligned}$$

It may be shown that this enforces the solution to be identically zero; in particular $(x_0,v_0)=(0,0)$, a contradiction. Hence $\{(0,0)\}$ is an isolated invariant set for F_λ when $\lambda>0$. On the other hand, in case that $\lambda<0$ a similar argument may be used, since then with analogous notation

$$\frac{d}{dt}V(X(t)) = -\lambda\,|\dot{x}|^2-\varphi(x,\dot{x})\,|\dot{x}| \ge -\lambda\,|\dot{x}|^2-C\,|\dot{x}|^{2+\alpha} \ge -(\lambda/2)\,|\dot{x}|^2.$$

Next we will calculate $H(\{(0,0)\};F_\lambda)=\Sigma^0$ in the case $\lambda>0$, with Σ^0 being the pointed zero-sphere. For that, we have to use the definition of H in (18) and approximate the multi-valued map. Let $w_\varepsilon(v)=(2/\pi)\arctan(v/\varepsilon)$ be the standard approximations of $\operatorname{sgn}(v)$, see (16) above; whence

$$f_\varepsilon(x,v)=\Big(v,\,-\lambda v-x^3-\varphi(x,v)\,w_\varepsilon(v)\Big)$$

serves as an approximation of F_λ in the (somewhat rough) sense of (17). Thus

$$H(\{(0,0)\};F_\lambda)=h(\operatorname{inv}(U_r);f_\varepsilon)$$

for $\varepsilon>0$ small by definition (18). Since $|w_\varepsilon(v)|\le 1$, $V(x,v)$ can also be used as Lyapunov function for the system $(x,v)'=f_\varepsilon(x,v)$. As above we obtain $\operatorname{inv}(U_r;f_\varepsilon)=\{(0,0)\}$. Therefore $B=U_r$ is a corresponding isolating block with immediate exit set $b^+=\emptyset$. Thus

$$H(\{(0,0)\};F_\lambda)=h(\operatorname{inv}(U_r);f_\varepsilon)=[B/b^+]=\Sigma^0,$$

because $B=U_r$ can be contracted to a single point.

In case $\lambda<0$, we may proceed similarly. Then $b^+=\partial B$ is the whole boundary of B, and identifying this to a point we obtain the pointed two-sphere in $\mathbb{R}^3$ denoted by Σ^2. Consequently we have shown

$$H(\{(0,0)\};F_\lambda)=\Sigma^0\ (\lambda>0)\quad\text{and}\quad H(\{(0,0)\};F_\lambda)=\Sigma^2\ (\lambda<0).$$

Using the bifurcation criterion (c) this jump of the index implies

Theorem 5. *If $|\varphi(x,v)|\le C|v|^{1+\alpha}$, then $\lambda_0=0$ is a bifurcation point for (13).*

Of course this theorem also holds for other systems, e.g., x^3 can be replaced by x^{2k+1} in (13) for some integer k. In principal what goes on in the example is a Hopf bifurcation of small periodic orbits; see [30].

More sophisticated topological arguments, mainly resting on [40], allow to prove a global bifurcation theorem for problems like (13), see [21] and [17], in the following way: the bifurcation point λ_0 with corresponding solution through the initial value $x_0 = 0 \in \mathbb{R}^n$ lies in some connected component $\mathcal{C}$ of the space $[a,b] \times \mathbb{R}^n$, where in "horizontal" direction the parameter space $\lambda \in [a,b]$ is drawn, and in "vertical" direction the initial values to solutions. Then the component $\mathcal{C}$ has two possibilities, namely either it is unbounded in $[a,b] \times \mathbb{R}^n$, or it has to meet the boundary $\{a,b\} \times \mathbb{R}^n$.

4 Application of KAM theory to non-smooth systems

Contrary to the systems we dealt with in Sects. 2 and 3 which are of a "dissipative" nature, there are also some non-smooth problems that lead to "conservative" equations. An example for this is provided by

$$\ddot{x} + x + a\,\mathrm{sgn}(x) = p(t) \tag{20}$$

with periodic forcing p and a parameter $a > 0$. Equations of this type are models for oscillators with state-dependent kicks, and it may be seen for (20) that the discontinuity line $\{x = 0\}$ is crossed transversely; hence here is no need to generalize the signum function to $\mathrm{sgn}(0) := [-1, 1]$, as was necessary in Sect. 2. It is a classical question about such forced planar oscillators, initiated in [26] some time ago, whether all solutions are bounded in the $(x, \dot{x})$-phase plane, or whether there do exist unbounded solutions. Although the equations under investigation were as simple as $\ddot{x}+2x^3 = p(t)$, the answers even nowadays require the application of tools as advanced as KAM theory. See [17] for a more detailed historical account. The main idea is to find, for a suitably transformed system, invariant curves with arbitrary high energy thus preventing solutions to escape to infinity. This additionally yields for free the existence of infinitely many periodic and quasi-periodic solutions, the former by the Poincaré-Birkhoff theorem.

Usually the application of KAM theory requires high regularity assumptions, and even without going into detail it may be clear that (20) will not give rise to such a smooth system; in action-angle variables (for $p = 0$) the system will only be continuous in the angle for $x = 0$. It turns out, however, that by means of a suitable canonical transformation at a certain stage of the proof the role of time t and the angle variable φ can be interchanged. Under the hypothesis that p is sufficiently smooth, it was proved in [22] that indeed KAM theory applies to give the desired boundedness of all solutions, at least for $a > 0$ sufficiently large. The main technical difficulty is to keep track of the errors through all the transformations. As a special case of a more general oscillator problem with linear growth at infinity (as is satisfied for (20), the result was generalized in [27] to hold even for $|\int_0^{2\pi} p(t)e^{it}\,dt| < 4a$. See [19] for a somewhat streamlined version of the proof in [27]. Since it is known from [15] that the bound for a is sharp, we can summarize the results as

Theorem 6. *(a) For* $|\int_0^{2\pi} p(t)e^{it}\,dt| < 4a$ *all solutions of (20) are bounded in the phase plane. In addition, there exist infinitely many periodic and quasiperiodic solutions.*
(b) For $|\int_0^{2\pi} p(t)e^{it}\,dt| > 4a$ *all solutions of (20) are unbounded.*

The case $|\int_0^{2\pi} p(t)e^{it}\,dt| = 4a$ is degenerate.

5 Non-smooth planar systems

A particular accessible case of non-smooth systems are planar linear systems of the form

$$\dot{q} = Aq + \operatorname{sgn}(w \cdot q)v, \tag{21}$$

with $q = (x, y) \in \mathbb{R}^2$, A a real (2×2)-matrix, and given vectors $v, w \in \mathbb{R}^2$ with $w \neq 0$. Such systems play a role in electrical circuits with a twin triode [1, p. 344] or in control systems with a two-point relay characteristic [25, p. 82]. Equation (21) has a discontinuity line $\{q \in \mathbb{R}^2 : w \cdot q = 0\}$, and the problem is to classify the dynamical behaviour of the system in dependence of the 8 parameters (4 for A, and 2 for w and v, respectively). Following [12,33], we make the following assumptions.

$$\operatorname{trace}(A) \neq 0, \quad \operatorname{trace}(\mathrm{A})^2 < 4\det(A). \tag{22}$$

It is then possible to detect the number of periodic solutions of (21) and their stability. Some of those solutions stick to the discontinuity for some time (sliding motion).

As a first reduction step, the number of parameters can be decreased to three, since by a transformation (21) is seen to be equivalent to

$$\dot{q} = A_\sigma q + \operatorname{sgn}(x)b, \quad A_\sigma = \begin{pmatrix} -1 & 1 \\ -\sigma & 0 \end{pmatrix}, \quad \sigma = \det(A)/\operatorname{trace}(\mathrm{A})^2 > 1/4, \tag{23}$$

where

$$b_1 = -(w \cdot v)/\operatorname{trace}(\mathrm{A}), \quad b_2 = -[(Aw^\perp) \cdot v^\perp]/\operatorname{trace}(\mathrm{A})^2,$$

with $w^\perp = (-w_2, w_1)$. This also rotates the discontinuity to $\{(x, y) \in \mathbb{R}^2 : x = 0\}$. In (23) there are only three parameters left, σ, b_1, and b_2.

Equation (23) has to be considered as the corresponding differential inclusion $\dot{q} \in A_\sigma q + \operatorname{sgn}(x)b$, with $q = (x, y)$ and $\operatorname{sgn}(0) = [-1, 1]$, the latter being necessary since here once more solutions can stick to the discontinuity line. Then an analysis can be carried out for the resulting semiflow to determine the critical points, i.e., such $q_0 = (x_0, y_0)$ with $(0, 0) \in A_\sigma q_0 + \operatorname{sgn}(x_0)b$, in all the different possible cases; see [12]. As the existence of periodic solutions is concerned, $b_1 \geq 0$ is a necessary condition as may be seen from the phase portraits. For $b_1 > 0$ the closed orbits that cross the discontinuity line transversely (no sticking) can be found as fixed points of the return map $\Pi = \Pi^- \circ \Pi^+$, where e.g. Π^+ maps a point $(0, y_0)$ along a trajectory that is contained in $\{(x, y) : x > 0\}$ to the first return point $(0, y_1)$ to the y-axis. Under certain circumstances, it may be verified that Π, considered as a map on the y-axis, is strictly increasing and concave, and hence has at most two fixed points. Thus there are at most two closed orbits that do not stick to the discontinuity line. With some more effort, the precise number of periodic solutions can be detected, and also their stability. Moreover, the existence of periodic solutions with sliding motion and of homoclinic orbits can be studied in detail; see [12] for the complete results. In particular, it turns out that there are at most three periodic solutions for (23).

6 Melnikov's method for non-smooth systems

A further important tool to obtain global analytical results for smooth dynamical systems is Melnikov's method, cf. [3,42]. This method allows to detect subharmonic solutions of arbitrary period and homoclinic orbits in certain systems $\dot{q} = f(q) + \epsilon g(q, t)$ that are small perturbations of systems possessing a homoclinic orbit $q_0(t)$. It turns out that this method carries over to particular non-smooth systems which are of the type $\dot{q} \in f(q) + \epsilon G(q, t)$, with a multi-valued function $G(q, t)$ arising as a model of dry friction as in (3). Thus, the choice of $\epsilon G(q, t)$ means that the friction effect has to be small; see [9,10,5] for analytical and numerical results. A second situation where Melnikov's method can be used in non-smooth systems are planar systems where it is not possible for trajectories to stick to the discontinuity surface for some time. To be definite, we deal with the following example taken from [43]. With $q = (x, y) \in \mathbb{R}^2$ let $f^+(x, y) = (1, -2x)$, $f^-(x, y) = (-y - 1, -x)$, and

$$f(x, y) = f^+(x, y), \quad y > 0, \quad \text{and} \quad f(x, y) = f^-(x, y), \quad y < 0.$$

So we consider

$$\begin{array}{l} \dot{x} = 1 \\ \dot{y} = -2x \end{array}, \quad y > 0 \qquad \text{and} \qquad \begin{array}{l} \dot{x} = -y - 1 \\ \dot{y} = -x \end{array}, \quad y < 0. \tag{24}$$

The system has a homoclinic orbit

$$q_0(t) = \begin{cases} (-e^{t+1}, e^{t+1} - 1) & : \ t \in]-\infty, -1], \\ (t, -t^2 + 1) & : \ t \in [-1, 1], \\ (e^{-t+1}, e^{-t+1} - 1) & : \ t \in [1, \infty[, \end{cases}$$

connecting the hyperbolic fixed point $p_0 = (0, -1)$ to itself. The solution has been adjusted in phase such that $q_0(0) = (0, 1)$, and it transversely intersects the discontinuity line $\{(x, y) \in \mathbb{R}^2 : y = 0\}$ twice, at time $t_d^- = -1$ in $P_d^- = q_0(t_d^-) = (-1, 0)$ and at $t_d^+ = 1$ in $P_d^+ = q_0(t_d^+) = (1, 0)$; see Fig. 3 below.

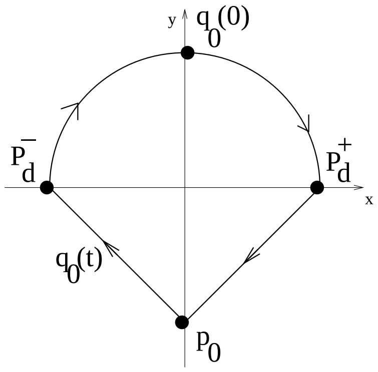

Fig. 3. The homoclinic orbit for (24)

We also note that here

$$\operatorname{trace}(Df^{+}) = \operatorname{trace}(Df^{-}) = 0 \tag{25}$$

in the respective domains of definition. It is for such situations and small perturbations

$$\dot{q} = f(q) + \epsilon g(q,t) \tag{26}$$

that Melnikov's method can be used equally well for non-smooth system to prove e.g. the existence of a homoclinic orbit for ϵ small. Compared to the usual Melnikov function

$$M_0(\theta_0) = \int_{\mathbb{R}} f(q_0(t-\theta_0)) \wedge g(q_0(t-\theta_0),t)\,dt \tag{27}$$

for smooth systems, two additional terms will be picked up through crossing the discontinuity. To see this, we first have to re-develop some parts of the classical theory, cf. [3, Sect. 3.8], always arguing on a quite heuristic rather than a rigorous level. We firstly expect that the fixed point p_0 perturbs to a fixed point p_ϵ of (26) that is of distance $\mathcal{O}(\epsilon)$ of p_0. For the augmented system

$$\dot{q} = f(q) + \epsilon g(q,\theta), \quad \dot{\theta} = 1, \tag{28}$$

this means that there exists a hyperbolic periodic orbit γ_ϵ close to $\gamma_0 = \{(p_0,\theta) : \theta \in S^1)\}$. Fix $\theta_0 \in S^1 \cong [0,2\pi]$ and denote L a line segment in the $\theta = \theta_0$ plane $\Sigma_{\theta_0} = \mathbb{R}^2 \times \{\theta_0\}$ that is perpendicular to the homoclinic orbit $\{q_0(t) : t \in \mathbb{R}\}$ at $q_0(0)$, and hence has the direction of $f^{\perp}(q_0(0))$. Denote p_{ϵ,θ_0} the intersection of γ_ϵ with Σ_{θ_0}, and let $q^{u,s}(t;\theta_0,\epsilon)$ be the unique trajectories of (28) that lie in the unstable ($=u$) resp. stable ($=s$) manifold of p_{ϵ,θ_0} and cross L at shortest distance to $q_0(0)$; see Fig. 4 below.

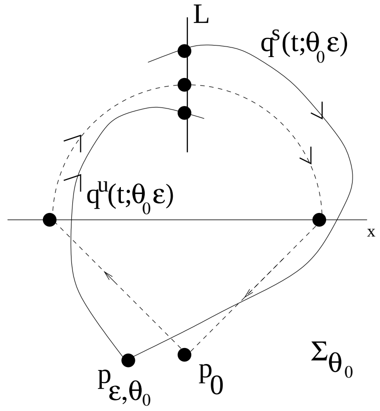

Fig. 4. Unstable and stable manifolds of p_{ϵ,θ_0}

Then we consider the function

$$\Delta_\epsilon(t,\theta_0) = f(q_0(t-\theta_0)) \wedge [q^u(t;\theta_0,\epsilon) - q^s(t;\theta_0,\epsilon)] \tag{29}$$

which for $t = \theta_0$ measures the distance between the unstable and stable manifolds of p_{ϵ,θ_0}. Hence to find a homoclinic orbit for (26) one needs to study the zeroes of $\theta_0 \mapsto \Delta_\epsilon(\theta_0,\theta_0)$; see [3, p. 173], [42, Sect. 4.5]. We have

$$q^{u,s}(t;\theta_0,\epsilon) = q_0(t-\theta_0) + \epsilon q_1^{u,s}(t,\theta_0) + \mathcal{O}(\epsilon^2), \tag{30}$$

with $q_1^{u,s}$ solutions to

$$\dot{q}_1^{u,s}(t,\theta_0) = Df(q_0(t-\theta_0))q_1^{u,s}(t,\theta_0) + g(q_0(t-\theta_0),t). \tag{31}$$

Let $\Delta_\epsilon^{u,s}(t,\theta_0) = \epsilon f(q_0(t-\theta_0)) \wedge q_1^{u,s}(t,\theta_0)$. Since $f(q_0(t))$ is orthogonal to $q_0(t)$, it follows from (29) and (30) that

$$\Delta_\epsilon(t,\theta_0) = \Delta_\epsilon^u(t,\theta_0) - \Delta_\epsilon^s(t,\theta_0) + \mathcal{O}(\epsilon^2). \tag{32}$$

According to the trace condition (25) it then can be shown as in the smooth case that

$$\dot{\Delta}_\epsilon^u(t,\theta_0) = \epsilon f(q_0(t-\theta_0)) \wedge g(q_0(t-\theta_0),t). \tag{33}$$

To integrate this w.r. to t we have to know at which times $\theta_0 + T^{u,s}(\theta_0,\epsilon)$ (on the level Σ_{θ_0}) the trajectories $q^{u,s}(t;\theta_0,\epsilon)$ will cross the discontinuity surface. According to the above construction we expect $\theta_0 + T^{u,s}(\theta_0,\epsilon) \cong \theta_0 + t_d^{\pm} + \mathcal{O}(\epsilon)$, and hence we will omit the error $\mathcal{O}(\epsilon)$, since we expect this to give a further term $\mathcal{O}(\epsilon^2)$ in (32). The intersection points of $q^{u,s}(t;\theta_0,\epsilon)$ with the discontinuity are ϵ-close to P_d^- and P_d^+, respectively. Integration of

(33) over $t \in]-\infty, \theta_0 + t_d^-]$ yields, taking into account that $\Delta_\epsilon^u(-\infty, \theta_0) = 0$ due to $f(p_0) = 0$,

$$-\Delta_\epsilon^{u,-}(\theta_0 + t_d^-, \theta_0) = \epsilon \int_{-\infty}^{\theta_0 + t_d^-} f^-(q_0(t-\theta_0)) \wedge g(q_0(t-\theta_0), t)\, dt;$$

here the "$-$" in $\Delta_\epsilon^{u,-}$ should indicate that one needs to choose f^- in the definition of Δ_ϵ^u, also to determine q_1^u. Similarly, observing $t_d^- < 0$, we have

$$\Delta_\epsilon^{u,+}(\theta_0, \theta_0) - \Delta_\epsilon^{u,+}(\theta_0 + t_d^-, \theta_0) = \epsilon \int_{\theta_0 + t_d^-}^{\theta_0} f^+(q_0(t-\theta_0)) \wedge g(q_0(t-\theta_0), t)\, dt,$$

and analogously, omitting the arguments in the integral,

$$\begin{aligned} \Delta_\epsilon^{s,+}(\theta_0 + t_d^+, \theta_0) - \Delta_\epsilon^{s,+}(\theta_0, \theta_0) &= \epsilon \int_{\theta_0}^{\theta_0 + t_d^+} f^+ \wedge g\, dt, \\ -\Delta_\epsilon^{s,-}(\theta_0 + t_d^+, \theta_0) &= \epsilon \int_{\theta_0 + t_d^+}^{\infty} f^- \wedge g\, dt. \end{aligned}$$

According to (32), and since $q_0(0)$ is in the upper half plane, this leads to

$$\begin{aligned} \Delta_\epsilon(\theta_0, \theta_0) &= \Delta_\epsilon^{u,+}(\theta_0, \theta_0) - \Delta_\epsilon^{s,+}(\theta_0, \theta_0) + \mathcal{O}(\epsilon^2) \\ &= M_0(\theta_0) + \Big(\Delta_\epsilon^{u,+}(\theta_0 + t_d^-, \theta_0) - \Delta_\epsilon^{u,-}(\theta_0 + t_d^-, \theta_0) \\ &\quad + \Delta_\epsilon^{s,-}(\theta_0 + t_d^+, \theta_0) - \Delta_\epsilon^{s,+}(\theta_0 + t_d^+, \theta_0)\Big) + \mathcal{O}(\epsilon^2), \end{aligned}$$

with the standard Melnikov integral $M_0(\theta_0)$ from (27). The Δ-terms are contributions due to the discontinuity, and they are also of order $\mathcal{O}(\epsilon)$. Thus according to the definition of $\Delta_\epsilon^{u,s}$ the relevant "non-smooth" Melnikov function is

$$\begin{aligned} M(\theta_0) = M_0(\theta_0) + \Big(& f^+(q_0(t_d^-)) \wedge q_1^{u,+}(\theta_0 + t_d^-, \theta_0) \\ & - f^-(q_0(t_d^-)) \wedge q_1^{u,-}(\theta_0 + t_d^-, \theta_0) \\ & + f^-(q_0(t_d^+)) \wedge q_1^{s,-}(\theta_0 + t_d^+, \theta_0) \\ & - f^+(q_0(t_d^+)) \wedge q_1^{s,+}(\theta_0 + t_d^+, \theta_0)\Big). \end{aligned}$$

The functions $q_1^{u/s,\pm}$ are determined by (31), e.g. $\varphi(s) = q_1^{u,+}(\theta_0 + s, \theta_0)$ solves the equation $\frac{d}{ds}\varphi = Df^+(q_0(s))\varphi + g(q_0(s), \theta_0 + s)$ for s near $s = t_d^-$. Again the zeroes of $M(\theta_0)$ determine the (transversal) intersection of stable and unstable manifolds, whence a simple zero gives rise to a homoclinic orbit for (26).

Returning to the particular example (24) from the beginning of this section, with $g(q,t) = (0, \sin t)$ the corresponding Melnikov function was calculated in [43], and numerically simple zeroes of M have been found.

7 Further directions

Of course there is still a lot to do for non-smooth systems. Since the models that are mathematically accessible are often by far too simplified, it will be of interest to detect a "small" system within some larger system, say in the spirit of a center manifold, where all the relevant dynamics will take place. The problem is then to investigate how that dynamics of the "small" non-smooth system is related to the full higher-dimensional system. Such ideas will be followed in [34]. There are further tools from smooth dynamical system theory that are useful also for non-smooth systems, like e.g. results on Hopf bifurcation of periodic orbits. This has been studied in [30], and the problem is once more that a non-smooth system in general does not have a linearization. Finally, on the more numerical side, one should try to get a better understanding of the non-smooth dynamics, in particular of real-world systems, through simulation of the corresponding attractors and invariant measures. Here a suitable approach to be taken is through box-division algorithms developed within the priority research program by the group of M. Dellnitz; see [7,8], and also [39] for a first step regarding applications of this algorithm to handle non-smooth problems.

References

1. Andronov A.A., Vitt A.A., Khaikin S.E. (1966) Theory of Oscillators. Dover Publications, New York
2. Arnold L. (1998) Random Dynamical Systems. Springer, Berlin Heidelberg New York
3. Arrowsmith D.K., Place C.M. (1990) An Introduction to Dynamical Systems. Cambridge University Press, Cambridge New York
4. Aubin J.P., Cellina A. (1984) Differential Inclusions. Springer, Berlin Heidelberg New York
5. Awrejcewicz J., Holicke M.M. (1999) Melnikov's method and stick-slip chaotic oscillations in very weakly forced mechanical systems. Int J Bifurcation and Chaos 9: 505-518
6. Deimling K. (1992) Multivalued Differential Equations. de Gruyter, Berlin New York
7. Dellnitz M., Hohmann A. (1997) A subdivision algorithm for the computation of unstable manifolds and global attractors. Numer Mathematik 75: 293-317
8. Dellnitz M., Froyland G., Junge O. (2000) The algorithms behind GAIO-set oriented numerical methods for dynamical systems. In this DANSE book volume
9. Fečkan M. (1996) Bifurcations from homoclinic to periodic solutions in ordinary differential equations with multivalued perturbations. J Differential Equations 130: 415-450
10. Fečkan M. (1999) Chaotic solutions in differential inclusions: chaos in dry friction problems. Trans Amer Math Soc 351: 2861-2873
11. Filippov A.F. (1988) Differential Equations with Discontinuous Right-Hand Sides. Kluwer, Dordrecht Boston London

12. Giannakopoulos F., Kaul A., Pliete K. (1999) Qualitative analysis of a planar system of piecewise linear differential equations with a line of discontinuity. Preprint
13. Hubbuch F. (1995) Die Dynamik des periodisch erregten Reibschwingers. Z Ang Math Mech 75, Supplement 1: S51-S52
14. Hubbuch F., Müller A. (1995) Lyapunov Exponenten in dynamischen Systemen mit Unstetigkeiten. Z Ang Math Mech 75, Supplement 1: S91-S92
15. Kunze M. (1998) Unbounded solutions in non-smooth dynamical systems at resonance. Z Angew Math Mech 78, Supplement 3: S985-S986
16. Kunze M. (1999) Periodic solutions of conservative non-smooth dynamical systems. Z Angew Math Mech 79, Supplement 1: S97-S100
17. Kunze M. (1999) Non-Smooth Dynamical Systems. Habilitation Thesis, Universität Köln
18. Kunze M. (2000) On Lyapunov exponents for non-smooth dynamical systems with an application to a pendulum with dry friction. J Dynamics Differential Equations 12: 31-116
19. Kunze M. (2000) Remarks on boundedness of semilinear oscillators. In: Sanchez L. (Ed.) Proc. Autumn School Nonlinear Analysis and Differential Equations, Lisbon 1998. Birkhäuser, Basel Boston
20. Kunze M., Küpper T. (1997) Qualitative analysis of a non-smooth friction-oscillator model. Z Angew Math Phys 48: 1-15
21. Kunze M., Küpper T., Li J. (2000) On the application of Conley index theory to non-smooth dynamical systems. Differential Integral Equations 13: 479-502
22. Kunze M., Küpper T., You J. (1997) On the application of KAM theory to discontinuous dynamical systems. J Differential Equations 139: 1-21
23. Kunze M., Michaeli B. (1995) On the rigorous applicability of Oseledets' ergodic theorem to obtain Lyapunov exponents for non-smooth dynamical systems. To appear in: Arino O. (Ed.) Proc 2nd Marrakesh International Conference on Differential Equations
24. Kunze M., Neumann J. (1997) Linear complementary problems and the simulation of the motion of rigid body systems subject to Coulomb friction. Z Angew Math Mech 77: 833-838
25. Lefschetz S. (1965) Stability of Nonlinear Control Systems. Academic Press, New York London
26. Littlewood J. (1968) Some Problems in Real and Complex Analysis. Heath. Lexington, Massachusetts
27. Liu B. (1999) Boundedness in nonlinear oscillations at resonance. J Differential Equations 153: 142-174
28. Michaeli B. (1998) Lyapunov-Exponenten bei nichtglatten dynamischen Systemen. Ph D Thesis, Universität Köln
29. Mischaikow K. (1995) Conley Index Theory. In: Johnson R. (Ed.) Dynamical Systems, Montecatini Terme 1994. Lecture Notes in Mathematics Vol 1609. Springer, Berlin Heidelberg New York, 119-207
30. Moritz S. (2000) Hopf-Verzweigung bei unstetigen planaren Systemen. MA Thesis, Universität Köln
31. Müller A. (1994) Lyapunov Exponenten in nicht-glatten dynamischen Systemen. MA Thesis, Universität Köln
32. Oseledets V.I. (1968) A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems. Trans Moscow Math Soc 19: 197-231

33. Pliete K. (1998) Über die Anzahl der geschlossenen Orbits bei unstetigen stückweise linearen dynamischen Systemen in der Ebene. MA Thesis, Universität Köln
34. Pliete K. Ph D Thesis, in preparation
35. Popp K., Stelter P. (1990) Nonlinear oscillations of structures induced by dry friction. In: Schiehlen W. (Ed.) Nonlinear Dynamics in Engineering Systems-IUTAM Symposium Stuttgart 1989. Springer, Berlin Heidelberg New York, 233-240
36. Popp K., Stelter P. (1990) Stick-slip vibrations and chaos. Phil Trans Roy Soc London A 332: 89-105
37. Rybakowski K.P. (1987) The Homotopy Index and Partial Differential Equations. Springer, Berlin Heidelberg New York
38. Scharstein H. (1998) Kräfte- und Leistungsbilanz bei der künstlichen Schlagbewegung einzelner Insektenflügel. In: Nachtigall W., Wisser A. (Eds.) Technische Biologie und Bionik 4, München 1998. Gustaf Fischer Verlag, Stuttgart
39. Voßhage Ch. (2000) Visualisierung von Attraktoren und invarianten Maßen in nichtglatten dynamischen Systemen. MA Thesis, Universität Köln
40. Ward J.R., Jr. (1998) Global bifurcation of periodic solutions to ordinary differential equations. J Differential Equations 142: 1-16
41. Wiederhöft A. (1994) Der periodisch erregte Einmassenreibschwinger. MA Thesis, Universität Köln
42. Wiggins S. (1990) Introduction to Applied Nonlinear Dynamical Systems and Chaos. Springer, Berlin Heidelberg New York
43. Zou Y.-K., Küpper T. (2000) Melnikov method and detection of chaos for nonsmooth systems. Submitted to SIAM J Math Anal