Final Paper Project: Melnikov's Method
EPS 528, Science of Complex Systems

Jonas Katona

November 16, 2022

1 Introduction

1.1 What's chaos?

In the most general sense, we define a chaotic system as one which shows

- sensitivity to initial conditions, i.e., if we start off with two nearby trajectories, these trajectories will eventually separate from each other and show radically different qualitative dynamics after a long enough time, as well as
- dynamics which "look random" but are still deterministic, i.e., there is some set of ODEs, PDEs, difference equations, etc. which can prescribe the dynamics of the system arbitrarily far into the future without any randomness or uncertainly involved.

More rigorously, according to the famed nonlinear dynamicist and mathematician, Robert L. Devaney, to classify a dynamical system as chaotic, the system must exhibit the following properties ([HK03]):[1]

1. *Sensitivity to initial conditions.* Suppose that our system is defined on some subset $U \subseteq \mathbb{R}^n$ and can be evolved forward in time according to some function $\varphi : U \times [0, \infty) \to U$, where we should think of φ as acting on points in our phase space and at a certain time; at some time $t \in [0, \infty)$, φ maps an initial condition to some usually different subset of U. If the system is sensitive to initial conditions, then for any $\delta > 0$ and $\mathbf{x} \in U$, there exists some $\mathbf{y} \in U$ such that $0 < ||\mathbf{x} - \mathbf{y}|| < \delta$ and $||\varphi(\mathbf{x}, t) - \varphi(\mathbf{y}, t)|| > \exp(\lambda t) ||\mathbf{x} - \mathbf{y}||$ for some constant $\lambda > 0$ and all $t \in [0, T]$ up to some time $T > 0$.[2] In other words, regardless of where we start in our system, there will always exist some trajectory arbitrarily close to that starting point, such that if we evolve these two trajectories forward in time, they will diverge exponentially quickly from each other. We can only guarantee this condition up to some time T, because naturally, if the trajectories lie in some bounded region of $\mathbb{R}^n$, they can only go so far from each other.

2. *Topological transitivity.* Using the same definitions as above, we say that the system is topologically transitive if $\varphi(X, t) \cap Y \neq \varnothing$ for any pair of open subsets $X, Y \subseteq U$ and some $t > 0$. In other words, if we take *any* two subsets of our phase space and evolve these forward in time, then these will overlap at some later time.

[1] We will only define things for continuous-time dynamical systems here. The generalization to discrete-time dynamical systems is easy; just consider cases where t can only vary by some set increments.

[2] A $\lambda > 0$ which works is the maximal Lyapunov exponent (which explains our notation here).

3. *Dense periodic orbits.* The system has dense periodic orbits if, for any $\mathbf{x} \in U$ and some $\delta > 0$, there exists some $\mathbf{y} \in U$ such that $0 < ||\mathbf{x} - \mathbf{y}|| < \delta$ and $f(\mathbf{y}, t) = f(\mathbf{y}, t + T)$ for every $t > 0$ and some period $T > 0$. This means that, for the time evolution of the system starting at any point in our phase space, there exists some points arbitrarily close to that point which will result in a periodic orbit. However, for chaotic systems, note that these periodic trajectories lie on some measure zero set, which means that with 100% probability, any random point we pick in the phase space will *not* result in a periodic trajectory if we evolve it forward in time. In fact, chaos cannot happen if our system is periodic.

Note that the above definition is not universal, and there is no universal consensus on a general definition of chaos. In particular, some authors use a weaker definition of chaos which excludes the third one above ([ML01]).

1.2 Where's chaos?

Regardless of how one defines chaos in a precise sense, chaotic systems exist all over nature. In solid mechanics, various pendulums (e.g., forced Duffing oscillator, forced van der Pol oscillator, double pendulum, etc.) and the N-body problem (for $N > 2$) are known to exhibit chaotic behavior for most initial conditions. In fluid mechanics, turbulence and mixing processes in certain, lower-dimensional regimes results in many cases from chaotic dynamics. Cryptography uses chaotic systems to encrypt messages and generate keys, because due to the sensitivity of initial conditions, tracing back a trajectory through phase space to reconstruct the encryption algorithm or decode something is extremely difficult if not impossible without knowing the system that was used. Chaotic systems appear all over chemistry, and not just in fluids: we have pattern formation which exhibits fractal qualities, as well as chemical oscillators which might exhibit chaos. They also appear frequently in population dynamics, as we will see at the end of this report. Finally, the weather itself is known to be an infamously chaotic system, which is not only why weather is so hard to predict, but also a motivating factor to the development of chaos theory in the first place, e.g., the work of Edward Norton Lorenz, a famed mathematician and meteorologist.

Visually, we can see some examples of chaos in physical systems in 1. The top photo shows one nicely-colored sample of trajectory of the Lorenz equations with $\sigma = 10$, $\rho = 28$, and $\beta = \frac{8}{3}$, which has nicely traced out the strange attractor. In particular, the layering gives a semblance of the fractal structure and the "density" of orbits lying in certain regions. The middle one shows the sample paths of the front wheel on the ground for 800 unmanned bicycles who were pushed from the same place at slightly different speeds; the paths end when the bicycle fell over. The drastically different paths taken by varying the speed slightly indicates sensitivity to initial conditions, one of the hallmarks of chaos. Finally, in the bottom photo, we have the simulation for the large eddy of a quickly-moving diffusion flame interacting with evaporating water droplets. Note the fractal-like properties of the diffusion cloud; the relationship between chaos theory for finite-dimensional systems and something like turbulence in infinite-dimensional systems is very, very poorly understood and currently unresolved, since we do not know how fluid should behave under fine scales.

In fact, in a certain sense, one could actually argue that *every* physical system must be chaotic due to the sheer degrees of freedom involved; probabilistically speaking, there is an increasingly high chance that at least one "direction" in a system will be chaotic as we increase the number of variables. One reason why we do not notice such chaos is, for the variables of interest, the effects of such chaos are negligible, e.g., our Lyapunov exponents are very close to zero, or only positive in some dimension or variable which we do not care

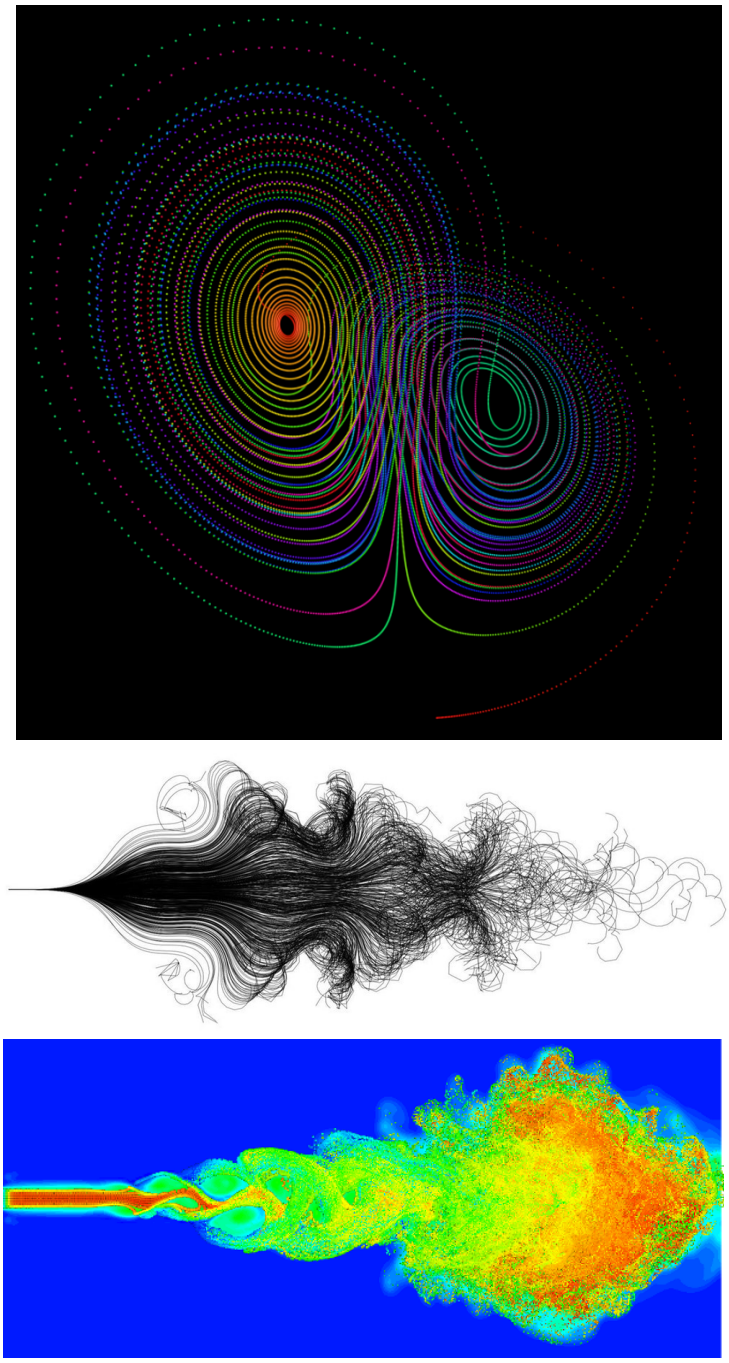

Figure 1: The following photos are taken from http://paulbourke.net/fractals/lorenz/, https://cyclingmagazine.ca/sections/news/paths-unmanned-bicycles-pushed-fall/, and http://www.2decomp.org/dstar.html respectively, and all show examples of how chaos manifests itself beautifully in systems found throughout nature.

about. Another is usually because we just naïvely dismiss the chaos as noise — the power spectrum of noise is indistuingishable from chaos ([Kin06]), and in many cases, one can approximate the complexities of chaotic dynamics with noise over a large enough timescale. In increasing numbers over the past few decades, we are beginning to realize that a lot of what we formerly classified as noise would actually be more appropriately realized as chaos, e.g., short-term and higher-order deviations from the usual sunspot cycle actually represent a low-dimensional, chaotic attractor; see [MMC91], [Let+06], or [HB10] for more information. But of course, the similarities between noise and deterministic chaos means that *with certain caution*, chaotic systems are amenable to and analyzable with statistical techniques.

That being said, there are multiple ways to distinguish between noise and chaos, such as by looking at the maximal Lyapunov exponent (MLE) or the entire Lyapunov spectrum if we want to look at which "directions" are chaotic or experience the greatest exponential divergence of neighboring trajectories ([Kin06; Gao+06]), or by looking at the dimension (correlation, information, etc.) of the attractor underlying a dataset, since a non-integer dimension indicates a fractal or *strange* attractor, which in turn underlies chaotic dynamics ([FOY83]). The main point here is that, while predicting the existence of chaotic dynamics in a system can often be tricky and computationally intense, it can be done. And as we will find, the Melnikov method provides a relatively straightforward way to predict or rule out the existence of chaos in a certain class of systems which is actually quite general, and covers many of the applications we discussed above.

1.3 Why chaos?

But why is chaos as a field so important? Why do we need to understand chaos? What is so difficult about chaos? Well, the dominance of chaotic dynamics in these above fields of study has massive ramifications on our abilities to predict and study these systems. If we want to make predictions or simulations of chaotic systems, then we know that a small error is magnified exponentially over time. Naïve linearization, a frequent, simplifying technique employed to study nonlinear systems, fails to capture the complex dynamics in chaotic systems. To capture the aperiodic, unpredictable behavior of chaotic trajectories, we cannot just throw any numerical scheme at the system, since numerical methods will tend to be more numerically unstable when simulating a chaotic system, i.e., chaotic systems are stiff. Even more directly, there might be certain physical systems, e.g., spacecraft, electric circuits, etc., in which we have direct control over the parameters, and in such systems where predictability and controllability are crucial, it might be advantageous for us to monitor and adjust the parameters to prevent chaotic motion from occurring. After all, chaotic systems are generally harder to control and monitor due to the unpredictable, fickle nature of deterministic chaos.

But once we recognize that a system is or should be chaotic, then there are a class of techniques out there, such as the ones outlined above and others (robust numerical solvers, statistical techniques, Lyapunov metrics, Poincaré maps and dimensionality reduction, etc.) which can be used to carefully deal with chaotic dynamics. But point is: Chaotic systems must be handled with care, and we must be prepared for when they arise. As Lorenz stated in a seminal talk he gave on December 29, 1972 in Washington D.C. at the 139th meeting of the American Association for the Advancement of Science ([Lor95]), "If the flap of a butterfly's wings can be instrumental in generating a tornado, it can equally well be instrumental in preventing a tornado."

1.4 Why Melnikov's method?

In a nutshell, Melnikov's method gives us a quantitative way to predict the occurrence of chaos in non-autonomous, smooth, nonlinear systems under periodic or stochastic perturbation. The method was first used by the famed mathematician, physicist, engineer, and philosopher Henri Poincaré in 1890 ([PM99]), but generalized by Soviet mathematicians Melnikov in 1963 ([Mel63]), with further generalizations in the years to follow. Bear with me for now, and this will be explained in the next two sections — Melnikov's method utilizes something called a Melnikov function, which measures the distance between the stable and unstable manifolds of a homoclinic point (respectively, some heteroclinic points) using an approximate Poincaré map. If we can compute the Melnikov function for a given system, then when this function has a simple zero, then this indicates that the stable and unstable manifolds have crossed tranversely due to the nonautonomous perturbation. This leads to a homoclinic (heteroclinic) tangle, which in turn signals the presence of chaotic dynamics from that point onward in the vicinity of the homoclinic (heteroclinic) orbit(s) that arise from the points mentioned earlier. What is especially nice about Melnikov's method is that, when one can write down some governing model or equations for the system in question, then one can usually derive an inequality in terms of the parameters in the system which gives an approximate condition that must be satisfied for Melnikov chaos to occur in the system. This allows one to predict whether or not the dynamics of the system will be chaotic for a given combination of parameters. Now, of course, there could be some complex cases where a perturbation causes chaos to occur even when this condition is not satisfied, especially if certain parameters are too large. This is also because chaos could arise for reasons other than transverse crossing of the stable and unstable manifolds. Hence, Melnikov's method is slightly more useful in confirming chaotic behavior rather than ruling it out, although in most applications, it appears that Melnikov's method works decently well in ruling out chaotic behavior as well, especially since in planar cases, by the Poincaré–Bendixson theorem, chaotic dynamics cannot occur in phase spaces with dimension below three ([Wig03], along with many other textbooks on nonlinear dynamics and chaos).

For this report, we discuss Melnikov's method only when restricted to smooth, planar systems under periodic or stochastic perturbation. However, more recent papers have provided generalizations to this, such as non-smooth cases ([Kuk07; GHS12]) and in higher dimensions ([CZ18; DG00; Shi91]).

2 Mathematical preliminaries

Before we define what a Melnikov function is, how to use Melnikov's method, and provide a sketch as to how one constructs the Melnikov function for a given system, we should go over some mathematical terminology and concepts which lead up to these things. Most of the mathematics here is taken from [Wig03], which is an excellent resource for the fundamentals of nonlinear dynamics and dynamical systems found throughout scientific, real-world applications.

2.1 Poincaré maps

Let us use the notation introduced on pg. 1. Consider the ODE system

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n \tag{1}$$

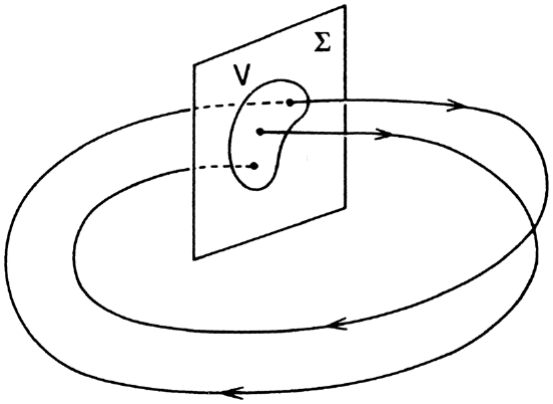

Figure 2: A sketch of a sample Poincaré map and orbit taken from [Wig03], pg. 123, showing Σ. Furthermore, Wiggins has also visualized the orbit of a point with period at least three, if the point is even periodic in the first place.

for some C^r (r-times continuously differentiable) vector field $F : U \to \mathbb{R}^n$ defined on some open set $U \subseteq \mathbb{R}^n$. Suppose that there exists a periodic orbit $\varphi(t, \mathbf{x}_0)$ with period T which solves (1) and passes through some point $\mathbf{x}_0 \in U$. Furthermore, let Σ be an $n-1$ dimensional surface which lies transverse to the vector field F at $\mathbf{x}_0$.[3] We call Σ the *cross-section to the vector field* (1).

By Theorems 7.1.1 in [Wig03], our flow $\varphi(t, \mathbf{x})$ is C^r if our vector field $F : U \to \mathbb{R}^n$ is also C^r. Hence, by definition of continuity, there exists some open set in U which is equal to the pre-image of Σ under F, i.e., there exists some open $V \subseteq U$ such that $V = \varphi^{-1}(\{(\tau, \Sigma\})$ for any $\tau > 0$ as close as we need to T.[4] Thus, trajectories starting in V return to Σ at a time τ close to T. This means that there exists a map $P : V \to \Sigma$ which maps points in V to their points of first return to Σ, which we call a **Poincaré map**. By construction $P : \mathbf{x} \mapsto \varphi(\tau(\mathbf{x}), \mathbf{x})$, where $\tau(\mathbf{x})$ is the time when $\mathbf{x}$ first returns to Σ. In particular, this implies that $\tau(\mathbf{x}_0) = T$ and $P(\mathbf{x}_0) = \mathbf{x}_0$, which means that fixed points of P correspond to periodic orbits of (1) and points of period k (i.e., points $\mathbf{y} \in U$ such that $P^k(\mathbf{y}) = \mathbf{y}$ and $P^i(\mathbf{y}) = V$ for $i = 1, \ldots, k$) are also periodic orbits of (1), but just with a different period. A visual sketch of a Poincaré map can be found in Figure 2.

Hence, we can see how a Poincaré map is essential in analyzing the local behavior near periodic orbits (e.g., stability, long-term behavior, identifying other periodic orbits, etc.), as well as a form of dimensionality reduction through which we can "average out" the periodic components of the system. Ideally, we can choose a Σ such that all local periodic orbits to $\mathbf{x}_0$ (if there are more) will intersect Σ at later times, but perhaps with different periods relative to P; that way, our Poincaré map would be optimally useful. Unfortunately, while a Poincaré map seems very powerful, in general, there is no *analytical* way to construct a Poincaré map associated with any arbitrary periodic orbit, although, there are some computational methods which exist in the literature ([Hen82; Tuc02; KKG11]). That being said, for the particular case that Melnikov's method covers, we *can* construct a Σ and, implicitly, a Poincaré map which can track the positions of the unstable and stable manifolds on Σ.

[3] By transverse, we mean that $F(\mathbf{x}_0) \cdot \mathbf{v} < ||\varphi(t, \mathbf{x}_0)|| \, ||\mathbf{v}||$ for all $\mathbf{v}$ in the tangent space of Σ at $\mathbf{x}_0$, or alternatively, $F(\mathbf{x}_0) \cdot \mathbf{n}(\mathbf{x}_0) \neq 0$, where $\mathbf{n}(\mathbf{x}_0)$ is the normal vector to Σ at $\mathbf{x}_0$. In other words, the vector field should not go *along* Σ at $\mathbf{x}_0$; instead, the vector field arrows should point *through* Σ at $\mathbf{x}_0$.

[4] If this preimage is empty, then the trajectories just did not come back to Σ, but that usually just means we picked a poor cross-section since $\mathbf{x}_0 \in V$ is periodic.

2.2 Invariant manifolds in phase space

Every fixed point for an ODE, e.g., (1), in which case $\mathbf{x}^* \in U$ is a fixed point for (1) if $F(\mathbf{x}^*) = 0$, has associated local, invariant manifolds which lie in some neighborhood of $\mathbf{x}^*$ in phase space. By local, we mean that these manifolds might have a boundary, in the sense that far enough from $\mathbf{x}^*$, these manifolds might end,[5] and by invariant, we mean that under the time evolution of the system, these manifolds remain in the same position in phase space. (Global invariant can exist too, but not always.) The dimensions of each of these added together adds up to the dimension of the phase space, and furthermore, the dimension of each manifold corresponds to the dimension of the corresponding eigenspace associated with that fixed point in the linearized system. By corresponding, we mean that the stable/unstable/center eigenspace for the linearized system actually lies *tangent* to the corresponding stable/unstable/center manifold for the fully nonlinear system, i.e., the vectors which span the stable/unstable/center eigenspace of a fixed point also span the tangent space for the stable/unstable/center manifold at the fixed point.

- The local, invariant stable manifold $W^s_{\text{loc}}(\mathbf{x}^*)$ lies tangent to the stable eigenspace ($\text{Re}\lambda < 0$) and has the same dimension as the stable eigenspace. $W^s_{\text{loc}}(\mathbf{x}^*)$ consists of all points which, under the time evolution of the flow φ, approach the fixed point as $t \to \infty$. In fact, trajectories starting on $W^s_{\text{loc}}(\mathbf{x}^*)$ approach $\mathbf{x}^*$ at an exponential rate asymptotically as $t \to \infty$.

- The local, invariant unstable manifold $W^u_{\text{loc}}(\mathbf{x}^*)$ lies tangent to the unstable eigenspace ($\text{Re}\lambda > 0$) and has the same dimension as the unstable eigenspace. $W^u_{\text{loc}}(\mathbf{x}^*)$ consists of all points which, under the time evolution of the flow φ, approach the fixed point as $t \to -\infty$. Similar to above, trajectories starting on $W^u_{\text{loc}}(\mathbf{x}^*)$ approach $\mathbf{x}^*$ at an exponential rate asymptotically as $t \to -\infty$.

  Note: This does *not* necessarily imply that points on $W^u_{\text{loc}}(\mathbf{x}^*)$ *diverge* from the fixed point as $t \to \infty$, although, this is often the case. This contrasts from what we would expect from the linearization, in which case we cannot have trajectories converge towards or diverge from $\mathbf{x}^*$ in *both* backwards and forwards time. In particular, one crucial thing that linearization does not tell us is how sometimes, the local unstable and stable manifolds *coincide*, which is a case that we will explore in great detail very soon.

- The center manifold $W^c_{\text{loc}}(\mathbf{x}^*)$ lies tangent to the center eigenspace ($\text{Re}\lambda = 0$) and has the same dimension as the center eigenspace. Points lying on $W^c_{\text{loc}}(\mathbf{x}^*)$ may approach the fixed point, diverge from the fixed point, or neither as $t \to \infty$. The behavior on $W^c_{\text{loc}}(\mathbf{x}^*)$ is "higher-order" and might even be described by another system of ODEs which can have its own set of dynamics. In fact, the dynamics on the center manifold could even be chaotic, which demonstrates even further how linearization really tells us nothing about the dynamics on the center manifold, even in the long-term.

In particular, by Theorem 3.2.1 in [Wig03], $W^s_{\text{loc}}(\mathbf{x}^*)$, $W^s_{\text{loc}}(\mathbf{x}^*)$, and $W^c_{\text{loc}}(\mathbf{x}^*)$ are each also C^r, just like F and φ.

Locally, we can represent $W^s_{\text{loc}}(\mathbf{x}^*)$, $W^u_{\text{loc}}(\mathbf{x}^*)$, and $W^c_{\text{loc}}(\mathbf{x}^*)$ as graphs, i.e., express these manifolds as functions of coordinates in phase space, and in some cases, explicit expressions for the manifolds are known. As an example, take the planar vector field $\dot{x} = x$ and

[5] For instance, if we have a local, invariant unstable manifold, then trajectories starting on that manifold might eventually leave the local manifold as $t \to \infty$, but only after a certain point far enough from $\mathbf{x}^*$.

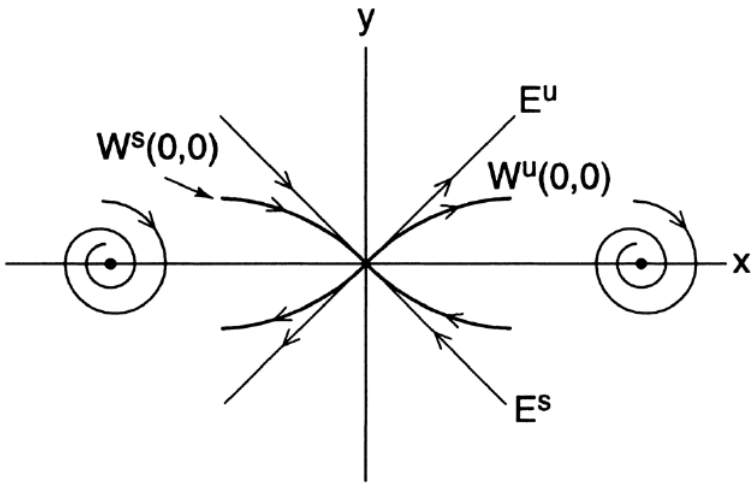

Figure 3: The following example is taken from [Wig03], pg. 41. This phase plane sketch is for the unforced Duffing oscillator, which is described by the ODE system $\dot{x} = y$ and $\dot{y} = x - x^3 - \delta y$ for some $\delta > 0$. The relationship and tangency between the stable (resp. unstable) eigenspace E^s (E^u) and the stable (unstable) manifold $W^s(0,0)$ ($W^u(0,0)$) of the origin is shown in the figure. In fact, with Melnikov's method, we can show that this system experiences chaotic dynamics when subject to *any* $\mathcal{O}(1)$ periodic forcing.

$\dot{y} = -y + x^2$ for $(x, y) \in \mathbb{R}^2$. This system has a fixed point at $(x^*, y^*) = (0, 0)$. Note that on the line $x = 0$, the equations look like $\dot{x} = 0$ and $\dot{y} = -y$, which gives us equations of motion $x(t) = 0$ and $y(t) = y_0 \exp(-t) \xrightarrow{t\to\infty} 0$ for any initial condition $y(0) = 0$ on this line. Thus, $W^s_{\text{loc}}(0,0) = W^s(0,0) = \{(x, y) \in \mathbb{R}^2 : x = 0\}$. Meanwhile, if we start on the parabola $y = x^2$, then our ODE system becomes $\dot{y} = 0$ and $\dot{x} = x$, which gives us equations of motion $y(t) = y_0$ and $x(t) = x_0 \exp(t) \xrightarrow{t\to-\infty} 0$ for any initial condition $(x(0), y(0)) = (x_0, x_0^2)$. Hence, $W^u_{\text{loc}}(0,0) = W^u(0,0) = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$. Another example can be found in Figure 3.

(I am only including this section because Amer asked during my presentation about how to compute invariant manifolds.) An invariant manifold is an integral of motion for the system, i.e, a curve $C(\mathbf{x}) = C_0$ such that $\frac{d}{dt}[C(\mathbf{x})] = 0$, where $C_0 \in \mathbb{R}$ is determined by initial conditions. Generally speaking, we can use the ODE system to derive a PDE which describes the hypersurface in phase space; in fact, from the perspective of PDE theory, the ODE actually describes the characteristics of this PDE and a natural "parameterization" for the surface (where t is obviously the variable which parameterizes the manifold). Suppose that our phase space U is some open subset of $\mathbb{R}^n$ and that our invariant manifold lies on some m-dimensional subset of U for $m < n$. Then, perhaps only locally (the invariant manifold could be something like a circle, so we might only have a local way to represent the function as a graph), we can express the invariant manifold like $\mathbf{y} = h(\mathbf{x})$ for coordinates $\mathbf{x} \in \mathbb{R}^{n-m} \cap V$ and $\mathbf{y} \in \mathbb{R}^m \cap V$ and some local mapping $h : \mathbb{R}^{n-m} \cap V \to \mathbb{R}^m \cap V$ defined on some open $V \subseteq U$. Then, using the fact that this manifold is *invariant* under the flow φ, $\frac{d}{dt}[h(\mathbf{x}) - \mathbf{y}] = 0 \Rightarrow \dot{\mathbf{x}} \cdot \nabla h(\mathbf{x}) - \dot{\mathbf{y}} = 0 \Rightarrow \dot{\mathbf{x}} \cdot \nabla h(\mathbf{x}) = \dot{\mathbf{y}}$. Then, possibly under a change of coordinates, suppose that our ODE system (1) can be written in the form $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{y})$ and $\dot{\mathbf{y}} = g(\mathbf{x}, \mathbf{y})$ for $f : \mathbb{R}^n \cap U \to \mathbb{R}^{n-m}$ and $g : \mathbb{R}^m \cap U \to \mathbb{R}^n$. If so, then our invariant manifold PDE reads like $f(\mathbf{x}, \mathbf{y}) \cdot \nabla h(\mathbf{x}) = g(\mathbf{x}, \mathbf{y}) \Rightarrow \boxed{f(\mathbf{x}, h(\mathbf{x})) \cdot \nabla h(\mathbf{x}) = g(\mathbf{x}, h(\mathbf{x}))}$, which we see is a generally nonlinear, first-order PDE for the graph function h. (There are other methods such as Liapunov-Perron method, but this method is the most straightforward to

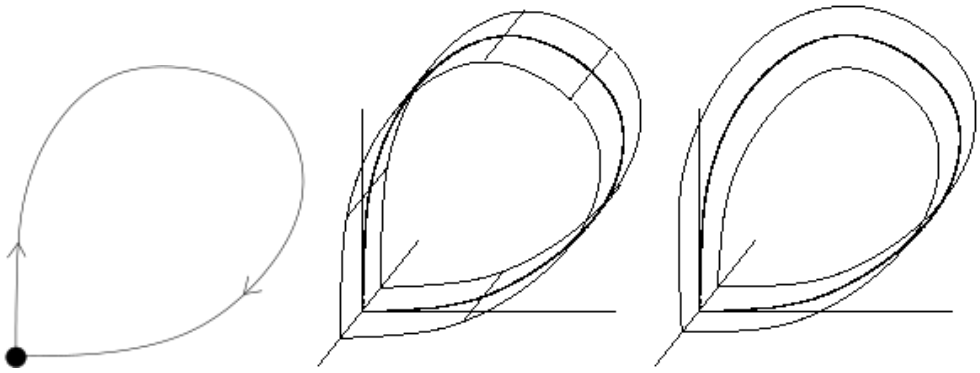

Figure 4: All photos are taken from https://en.wikipedia.org/wiki/Homoclinic_orbit. The left photo is a homoclinic orbit on its own. However, the right two show the homoclinic orbit along with the geometry of the unstable manifold which underlies it. The middle one is called an *oriented* homoclinic orbit and is topologically equivalent a a cylinder, while the right one is called a *twisted* homoclinic orbit and is topologically equivalent to a Möbius strip.

understand.)

In particular, if we wanted to compute the local, invariant stable, unstable, and center manifolds for a fixed point, we need to use two properties for such manifolds: (1) these cross through the fixed point *and* (2) lie tangent to the stable, unstable, and center eigenspaces. These provide us enough initial conditions for us to solve the PDE above uniquely for the stable and unstable manifolds (the center manifold is still not *always* unique), and to do this, we usually make an affine change of coordinates $(\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{z}, \mathbf{w})$ which "centers" and "diagonalizes" our system, such that the stable, unstable, and center eigenspaces are orthogonal to each other in our new coordinates $(\mathbf{z}, \mathbf{w})$ and the fixed point in question lies at the origin in $\mathbb{R}^n$. That way, conditions (1) and (2) can be expressed simply as $h(\mathbf{z}^*) = 0$ and $\nabla_{\mathbf{z}} h(\mathbf{z}^*) = 0$, respectively. From here, the PDE can be solved explicitly using these initial conditions if possible or (usually) via a power series.

To prove that Melnikov's method works, we do not actually need to compute the stable and unstable manifolds for the relevant fixed points as equations of the phase space coordinates only. Instead, to compute the Melnikov function, we use the homoclinic (or heteroclinic) orbit(s) explicitly as functions of time, which in those cases trace out the stable and unstable manifolds.

2.3 Homoclinic and heteroclinic orbits

A solution $\mathrm{x}(t)$ to (1) is called a **homoclinic orbit** ("homo-" means same, "-clinic" means associated with a slope) if $\lim_{x \to \pm\infty} \mathbf{x}(t) = \mathbf{x}^*$. Hence, by definition, we see that a homoclinic orbit lies in the intersection of the local stable and unstable manifolds of $\mathbf{x}^*$. **Note:** The fixed point $\mathbf{x}^* \in U$ may not be isolated, in the sense that a homoclinic orbit might be associated with a line or even surface of fixed points, creating a number of homoclinic orbits that lie right next to each other. A few physical manifestations of homoclinic orbits can be found in Figure 4.

Suppose that (1) has two fixed points $\mathbf{x}_1, \mathbf{x}_2 \in U$. Then, similarly, a solution $\mathrm{x}(t)$ to (1) is called a **heteroclinic orbit** ("hetero-" means other) if $\lim_{t \to -\infty} \mathbf{x}(t) = \mathbf{x}_1$ and $\lim_{t \to \infty} \mathbf{x}(t) = \mathbf{x}_2$ (or vice-versa). Then, this orbit lies in the intersection of the unstable (resp. stable) manifold of $\mathbf{x}_1$ and the stable (unstable) manifold of $\mathbf{x}_2$.

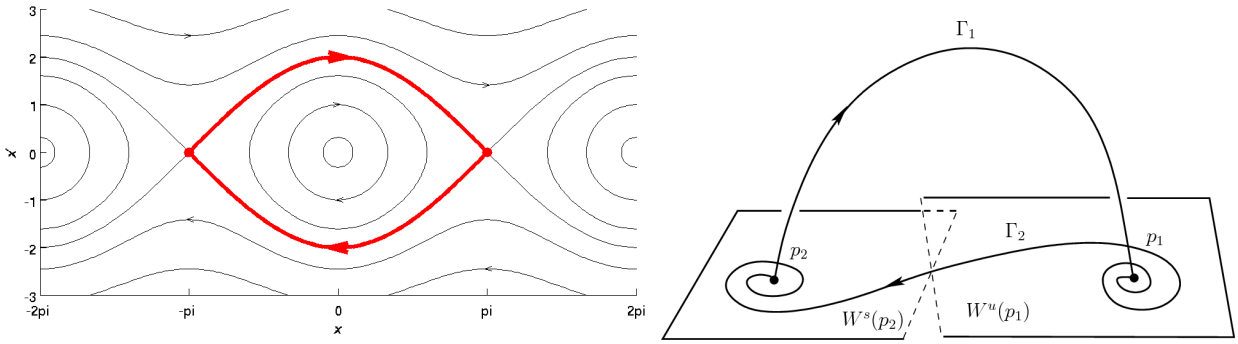

Figure 5: Heteroclinic orbits do not have to be simple. Here are two examples of conjoined heteroclinic orbits, taken from https://en.wikipedia.org/wiki/Heteroclinic_orbit and [KLW14], respectively. Both show two heteroclinic orbits joined in a heteroclinic cycle. The left one is from the phase portrait for the nonlinear pendulum, i.e., the pendulum problem without the $\sin\theta \approx \theta$ approximation, showing heteroclinic orbits between saddle points, while the right one is a schematic for the so-called T-point heteroclinic cycle in $\mathbb{R}^3$, which occurs between two saddle-foci.

Heteroclinic orbits are generally more complicated than homoclinic orbits since we are dealing with two as opposed to one fixed point, which opens up the possibilities for some rather complicated dynamics, of which two examples are in Figure 5. One such possibility is a so-called heteroclinic cycle, which will be relevant when we generalize Melnikov's method to the case of heteroclinic orbits. Suppose that (1) has a collection $\{\mathbf{p}_i\}_{i=0}^n$ of n hyperbolic points, i.e., fixed points with eigenvalues that all have nonzero real parts, where $\mathbf{p}_0 = \mathbf{p}_n$. If there is a heteroclinic orbit that goes between $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ for all $i = 0, 1, \ldots, n-1$, then the fixed points form a *heteroclinic cycle*; we already saw two simple examples of these in 5. In other words, a heteroclinic cycle consists of joined heteroclinic orbits, such that we can go around continuously along each of these orbits and eventually come back around to where we started. The original Melnikov's method applies only to homoclinic cycles, but it turns out that the same results apply if the Melnikov function has a simple zero on *each* heteroclinic orbit within a heteroclinic cycle. We will explore the reasons behind this relationship later on.

2.4 Homoclinic and heteroclinic tangles in the phase plane

We will restrict our discussion from now on for a vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$ and homoclinic orbits, but we will mention later how these results extend to heteroclinic cycles as well. The following concept is difficult to understand, mainly because there is a lot of mathematical machinery, lemmas, and theorems behind these concepts which we are omitting — namely, the lambda lemma — but hopefully we can capture the gist of it. If the hyperbolic fixed point $\mathbf{x}^*$ of a homoclinic orbit for (1) has stable and unstable manifolds $W^s_{\text{loc}}(\mathbf{x}^*)$ and $W^u_{\text{loc}}(\mathbf{x}^*)$ which intersect transversely, we call this a **transverse homoclinic point**. (This contrasts with a homoclinic point, which just lies in the intersection of these manifolds.) Now, suppose that we have the orbit of a transverse homoclinic point $\mathbf{p}$, and from this consider the map which tracks the crossings of these two manifolds. Since the stable and unstable manifolds are invariant under this map, the forward and backward iterates of $\mathbf{p}$ must also lie on both the stable and unstable manifolds of $\mathbf{x}^*$, i.e., $\varphi(t, \mathbf{p})$ still tends to $\mathbf{x}^*$ as $t \to \pm\infty$. Since $\varphi(t, \mathbf{p})$ lies on $W^s_{\text{loc}}(\mathbf{x}^*)$, the orbit of $\mathbf{p}$ approaches $\mathbf{x}^*$ exponentially quickly. However,

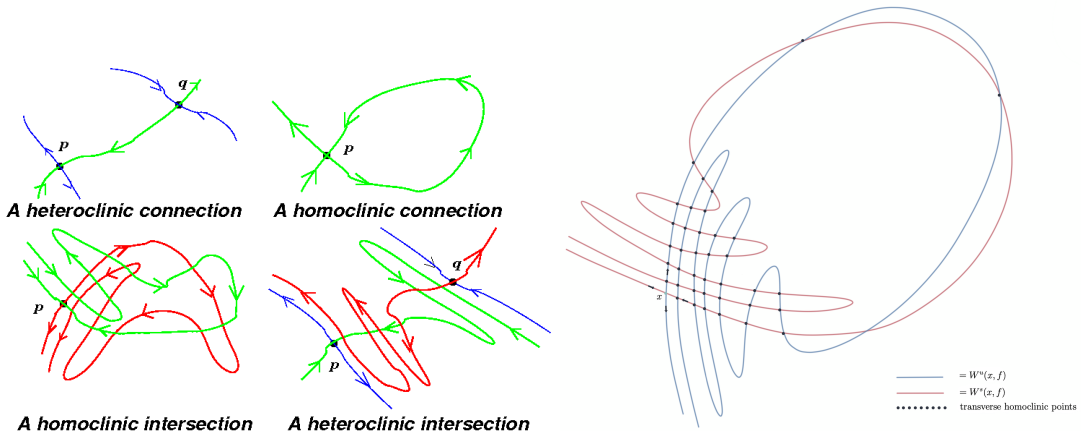

Figure 6: Heuristic sketches showing the qualitative behavior of homoclinic and heteroclinic tangles, from https://en.wikipedia.org/wiki/Homoclinic_connection and https://www.merry.io/dynamical-systems/47-homoclinic-tangles/, respectively.

these crossings continue to happen, as tracked by this map, and hence, the existence of one transverse intersection implies the existence of an infinite number of transverse intersections which accumulate at the fixed point at the origin, as $W^u_{\text{loc}}(\mathbf{x}^*)$ continues to cut transversely across $W^s_{\text{loc}}(\mathbf{x}^*)$ the closer and closer we get to $\mathbf{x}^*$ as $t \to \infty$, and vice-versa for $t \to -\infty$. In fact, the folds of $W^u_{\text{loc}}(\mathbf{x}^*)$ (respectively, $W^s_{\text{loc}}(\mathbf{x}^*)$) continue to get thinner and longer as $t \to \infty$ ($t \to -\infty$), becoming roughly parallel to each other as time increases (decreases).

For iterations large enough in magnitude of this Poincaré map, the orbits of other nearby points go on to form other transverse homoclinic points that do not lie in $\varphi(t, \mathbf{p})$, and these in turn form on similar accumulations of transverse crossings. The union of all transverse homoclinic points created in this process is called a **homoclinic tangle**, and this ultimately is the process by which chaos arises when a time-dependent perturbation causes the stable and unstable manifolds of a homoclinic or heteroclinic orbit to cross. This complex phenomenon is visualized in Figure 6.

Furthermore, the reason why we can generalize these phenomena to heteroclinic cycles is because of the following theorem (pg. 632 of [Wig03]): Suppose that we have what is called a *transversal* heteroclinic cycle, which is defined as a collection $\{\mathbf{p}_i\}_{i=0}^n$ of n hyperbolic points where $\mathbf{p}_0 = \mathbf{p}_n$ such that $W^u_{\text{loc}}(\mathbf{p}_i)$ transversely intersects $W^s_{\text{loc}}(\mathbf{p}_{i+1})$ for all $i = 0, 1, \ldots, n-1$. Then, $W^u_{\text{loc}}(\mathbf{p}_i)$ also transversely intersects $W^s_{\text{loc}}(\mathbf{p}_i)$ for all $i = 0, 1, \ldots, n-1$. Hence, by the above observations, if a time-dependent perturbation causes the unstable manifold of each hyperbolic fixed point in a transversal heteroclinic cycle to cross with the stable manifold for the next equilibrium in the cycle, the unstable manifold for each equilibrium will transversely cross the stable manifold for the same equilibrium, which signals the same conditions for these two manifolds to tangle within each other.

2.4.1 Smale's horseshoe map

Accompanying the formation of a homoclinic or heteroclinic tangle is a so-called **Smale's horseshoe map**. Without going into too many tedious details, we will go through a heuristic, geometric understanding of how Smale's horseshoe works. Originally, the famed mathematician and pioneer of dynamical systems theory, Stephen Smale, formulated this map back in 1967 ([Sma67]) to help understand the heteroclinic tangles which form in the phase space for the forced van der Pol oscillator.

If we take a square and consider what happens to its shape under the horseshoe map,

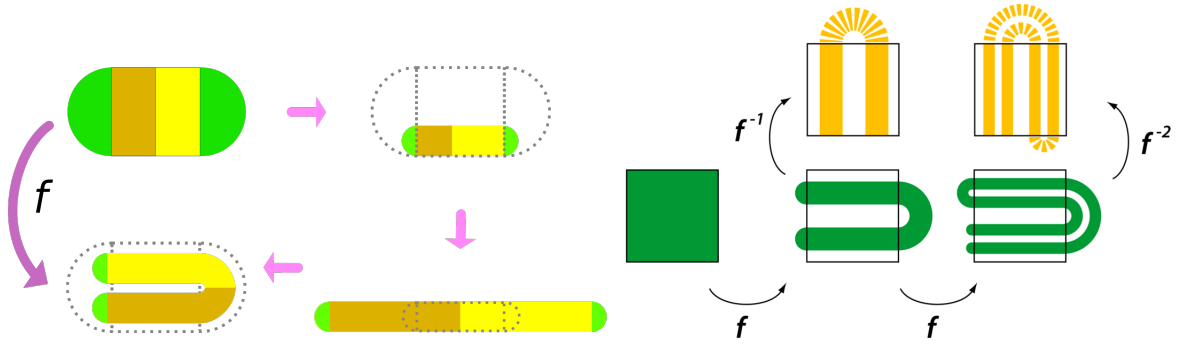

Figure 7: These visuals are taken from https://en.wikipedia.org/wiki/Horseshoe_map. f is the horseshoe map; the left picture shows what happens to different regions within the square after one iteration, while the right shows what happens to the square after one and two iterations both backwards (undoing the map) and forward (doing the map).

then the horseshoe map consist of three geometric transformations: contraction, expansion, and folding, as shown in Figure 7.

- *Contraction.* The square is contracted along the *vertical* direction by a factor of $a < \frac{1}{2}$. If $a \geq \frac{1}{2}$, then the latter two steps would just lead to the entire square being mapped to itself somehow, which is pretty boring; there should be a gap between the two strips which remained in the horseshoe. Furthermore, the semi-circles on both ends are contracted uniformly.

- *Expansion.* The rectangle resulting from contraction is expanded in the *horizontal* direction by a factor of $\frac{1}{a}$, and the caps remain the same size.

- *Folding.* The strip is folded into a horseshoe shape around the left (or right) side and through the square, such that the two caps stick out from the right (resp., left) side.

If we look at what happens to the phase space surrounding a homoclinic or heteroclinic tangle, then we might note that the way the horseshoe map continues to fold on itself looks somewhat similar to how the stable manifold continues to cross an unstable manifold, or vice-versa in backwards time, for a homoclinic or heteroclinic tangle. Our intuition here is not misleading, as what happens to the phase space near a homoclinic or heteroclinic tangle is topologically similar to the horseshoe map. To further establish the connection between the horseshoe map and a homoclinic or heteroclinic tangle, it turns out that the fixed which remain in the square under the horseshoe map form a fractal set which is also a Cantor set, i.e., a set which is closed, every point is a limit point, and is totally disconnected, while similarly, the set of all the transverse homoclinic points for $\mathbf{x}^*$ above is also a Cantor set.

Okay, great: Transversal crossings of the stable and unstable manifolds of the fixed point(s) associated with a homoclinic (heteroclinic) orbit lead to the formation of a homoclinic (heteroclinic) tangle, which in turn causes the phase space to resemble that generated by Smale's horseshoe map repeatedly acting on the square, in which there is a fractal, invariant set. A fractal phase space structure usually is a hallmark of chaotic dynamics. But where precisely do the chaotic dynamics arise? Well, proving that the horseshoe map is chaotic using the three properties mentioned at the beginning of this report is probably impossible, but according to [Sma67], the horseshoe map when restricted to the fractal, invariant set is topologically conjugate to a shift map σ on N symbols, and that is much easier shown to be chaotic. By a map on N symbols, we mean that σ acts on elements of the form $\{\dots s_{-n} \dots s_{-1}.s_0 s_1 \dots s_n \dots\}$ belonging to a set Σ^N constructed by N distinct symbols

(e.g., the whole numbers 0 to 9), such that each s_i for $i \in \mathbb{Z}$ is chosen from these N symbols, and as a shift map, $\sigma(\{\dots s_{-n} \dots s_{-1}.s_0 s_1 \dots s_n \dots\}) = \{\dots s_{-n} \dots s_{-1} s_0.s_1 \dots s_n \dots\}$.

For a proof of the above, see chapters 23 and 24 of [Wig03] for details; we would need to define a metric via which we can define properties such as continuity, distance between elements of this symbol space, etc. But heuristically speaking, we can intuit how the shift map would lead to chaotic dynamics. For simplicity's sake, consider the shift map acting on the digits in base-10, i.e., where $\sigma : [0,1] \to [0,1]$ (so mod 1). We have sensitivity off of initial conditions, because if we have two numbers that differ by at some digit, then each time the shift map acts on both of these numbers, this digit moves up by decimal place. Hence, at each iteration, the shift map magnifies any errors by 10, which gives us a Lyapunov exponent of $\log 10$. For topological transitivity, if we choose any subsets of $\mathbb{R}$, we can choose a number in one of them such that, far enough down the decimal places, there is a sequence in our decimal expansion which gets as close as we want to any other number, and in particular, any number in that other subset. For instance, if we want to get a number in the interval $(0.02, 0.0202)$ which ends up in the interval $(0.202, 0.2021)$ after some numbers of iterations, just pick 0.020120205. Then, $\sigma^4(0.020120205) = 0.20205 \in (0.202, 0.2021)$. Finally, we have dense periodic orbits; for any point $r \in [0,1]$, we can just truncate the decimal expansion for r after k decimal places and then repeat those digits ad infimum, which is within 10^{-k} of r, and we can make 10^{-k} arbitrarily small.

The punch line: The formation of a homoclinic/heteroclinic tangle is what motivates us to define a Melnikov function. The Melnikov function measures when the unstable and stable manifolds of homoclinic or heteroclinic point(s) intersect transversely, and thereafter, given appropriate conditions, these manifolds will form a homoclinic/heteroclinic tangle, from which we get a chaotic, fractal set. Most of the time, Melnikov functions cannot be evaluated explicitly and can only be computed with the help of numerical quadrature, but in cases where closed-form expressions exist, we would usually compute the Melnikov function via calculus of residues.

3 Melnikov functions

Without further ado, we will define the Melnikov function for several cases in increasing generality, alongside the theorems necessary to use these to prove or disprove that homoclinic/heteroclinic chaos exists.

3.1 Hamiltonian case

This is the original theorem brought forth by [PM99] and formalized in more modern terms by [Mel63]. Consider a Hamiltonian flow:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial H}{\partial x} \\ \frac{\partial H}{\partial y} \end{pmatrix} + \varepsilon \begin{pmatrix} g_1(\mathbf{q}, t, \varepsilon) \\ g_2(\mathbf{q}, t, \varepsilon) \end{pmatrix}, \tag{2}$$

where $\mathbf{q} = (x, y)$, $H \in C^3$ is a Hamiltonian for (2), and $\mathbf{g}(\mathbf{q}, t, \varepsilon) = (g_1(\mathbf{q}, t, \varepsilon), g_2(\mathbf{q}, t, \varepsilon))$ is a function with period $\frac{2\pi}{\omega}$.

Suppose that (2) exhibits a homoclinic orbit $\mathbf{q}_0(t)$ associated with a fixed point for the unperturbed system ($\varepsilon = 0$). Then, we define the Melnikov function

$$M(t_0, \phi_0) = \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{q}_0(t), \omega(t - t_0) + \phi_0, 0) \cdot \nabla H(\mathbf{q}_0(t))\, dt. \tag{3}$$

If there is a point (t_0^*, ϕ_0^*) such that

- $M(t_0^*, \phi_0^*) = 0$ and
- $\partial_t M(t_0, \phi_0)|_{(t_0,\phi_0)=(t_0^*,\phi_0^*)} \neq 0$,

then we call (t_0^*, ϕ_0^*) a *simple zero* for (3). When (3) exhibits a simple zero for ε small enough, the unstable and stable manifolds of the fixed point intersect tranversely at $\mathbf{q}_0(-t_0)$, signaling the presence of chaotic dynamics. Moreover, if $M(t_0, \phi_0) \neq 0$ for all $(t_0, \phi_0) \in \mathbb{R} \times S^1$, then the unstable and stable manifolds will not intersect.

The proof of the above is rather long and not super informative, but the point is that (3) roughly measures (i.e., for $\varepsilon < 0$) the distance between the unstable and stable manifolds associated with the fixed point from which the homoclinic orbit $\mathbf{q}_0(t)$ of the unperturbed system arises. We essentially want to check how the time-dependent perturbation $\mathbf{g}$ affects the splitting up the stable and unstable manifolds along this trajectory, because if these manifolds cross transversely, then they will continue to cross with period $\frac{2\pi}{\omega}$, which means that this perturbation non-autonomously "forces" a homoclinic tangle to appear in our system, and chaotic dynamics alongside this. Furthermore, the reason why we require a simple zero is, if we had something like a double zero, then it is possible that the stable and unstable manifolds touch but do not fully go through one another, i.e., they intersect, but not transversely.

(3) tracks the distance between the stable and unstable manifolds with an $\mathcal{O}(\varepsilon)$ error. If we want a more precise way to measure the distance between the stable and unstable manifolds, more recent papers have introduced higher-order corrections to (3) (e.g., [LR04; CW17]).

3.2 Non-Hamiltonian case

There are two Melnikov functions that we will cover here. The first one is more theoretically robust, since it still accurately measures the transverse crossings between the stable and unstable manifolds and predicts chaotic dynamics for $t \to \pm\infty$. Meanwhile, the second one does not necessarily measure the actual distance between the manifolds well, but nevertheless implies that, with a simple zero, the stable and unstable manifolds still must intersect transversely and that the system will exhibit chaotic dynamics on any finite-time interval $(-T, T)$ for $T > 0$. Furthermore, the second one includes extra assumptions on the boundedness of the perturbation g, but in return does not even require g to be periodic.

The first formulation is from Section 4.5 of [GH83] and [Ber88] and proceeds in a similar fashion as in the Hamiltonian case; proofs can be found in these texts. Consider the planar dynamical system

$$\dot{\mathbf{x}} = f(\mathbf{x}) + \varepsilon g(\mathbf{x}, t) \tag{4}$$

for some C^2 vector fields $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ where $\mathbf{x} \in \mathbb{R}^2$, $g(\mathbf{x}, t) = g(\mathbf{x}, t + T)$ for all $\mathbf{x} \in \mathbb{R}^2$ and some $T > 0$, and $0 < \varepsilon \ll 1$. Suppose that when $\varepsilon = 0$, (4) exhibits a homoclinic orbit $\mathbf{q}_0(t)$ associated with a hyperbolic fixed point $\mathbf{x}^* \in \mathbb{R}^2$. Then, we define the Melnikov function

$$M(t_0) = \int_{-\infty}^{\infty} f(\mathbf{q}_0(t)) \wedge g(\mathbf{q}_0(t), t + t_0) \exp\left[\int_0^t \operatorname{tr}(Df(\mathbf{q}_0(s)))\, ds\right] dt, \tag{5}$$

where $\operatorname{tr}(Df(\mathbf{q}_0(s)))$ is the trace of the Jacobian matrix Df evaluated at time s along the homoclinic orbit $\mathbf{q}_0$. In particular, if is a Hamiltonian system, then $\operatorname{tr}(Df(\mathbf{x})) =$

$\frac{\partial}{\partial x}\frac{\partial H}{\partial y} + \frac{\partial}{\partial y}\left(-\frac{\partial H}{\partial x}\right) = \frac{\partial^2 H}{\partial x \partial y} - \frac{\partial^2 H}{\partial y \partial x} = 0$, and hence, (5) recovers the same Melnikov function we have in (3).

We note that (5) inherits the same periodicity in t_0 as the original function g had. In fact, since $g\left(\mathbf{q}_0\left(t\right), t + t_0 + T\right) = g\left(\mathbf{q}_0\left(t\right), t + t_0\right)$, we immediately see that $M\left(t_0 + T\right) = M\left(t_0\right)$.

The second formulation comes more recently from [CM09], in which a proof can be found as well. Consider the same system above, except we further assume that $g\left(\mathbf{q}_0\left(t\right), t\right)$, $\frac{\partial g}{\partial t}\left(\mathbf{q}_0\left(t\right), t\right)$, and $\nabla g\left(\mathbf{q}_0\left(t\right), t\right)$ are bounded for all $t \in \mathbb{R}$. However, in return, g *does not* need to be periodic. Then, if

$$M\left(t_0\right) = \int_{-\infty}^{\infty} f\left(\mathbf{q}_0\left(t - t_0\right)\right) \wedge g\left(\mathbf{q}_0\left(t - t_0\right), t\right) dt \tag{6}$$

has a simple zero at some $t_0 \in \mathbb{R}$, then the flow exhibited by (4) is chaotic for all $t \in (-T, T)$ and *any* $T > 0$. Note again that we *do not* even need to assume that g is periodic; g just needs to satisfy certain continuity and boundedness assumptions.

We also emphasize again that (6) *does not* necessarily track the distance between the stable and unstable manifolds. However, the key point here is that we do not need to do this. In [CM09], Castilho and Marchesin extend this now-nonperiodic system on the time interval $(-T, T)$ to a T-periodic vector field on all of $\mathbb{R}$, on which we can apply the usual Melnikov theory. Then, they prove that simple zeroes for (6) imply simple zeroes for the Melnikov function in this extended, periodic system, and hence, we have chaotic dynamics in both cases.

3.2.1 Example: Gylden's problem

This second extension seems rather natural, and for practical applications, we really only need to simulate or understand a system for finite times, which means that (6) works as a Melnikov function for practical use. It also shows that, interestingly, all we might need for some systems is a small impulse or peak to cause the system to become chaotic. For instance, in [CM09], they apply their new Melnikov function to the following system which arises (after a few changes of coordinates) in modeling the motion of variable-mass bodies under the action of Newtonian attraction forces:

$$\frac{du}{d\tau} = v, \quad \frac{dv}{d\tau} = u - u^3 + \varepsilon u p\left(\frac{p_\theta^4}{16} t\left(\tau\right)\right). \tag{7}$$

This problem is called Gylden's problem, or alternatively, the Gylden–Meshchersky problem; $p : \mathbb{R} \to \mathbb{R}$ is an explicit function of time, while τ is a new, time-like variable defined by the integral $\tau\left(t\right) = \frac{1}{4}\int_0^t \left|u(t)\right|^3 dt$. Hence, we should think of $t = t\left(\tau\right)$ as an inverse for this integral equation, and in particular, along the homoclinic orbit, $t\left(\tau\right) = 4\sinh\left(\tau\right) + \frac{4}{3}\sinh^3\left(\tau\right)$. For (7), the relevant Melnikov function, associated with a homoclinic orbit of the origin $(u^*, v^*) = (0, 0)$, is computed in the paper as

$$M\left(\tau_0\right) = -\int_{-\infty}^{\infty} \sqrt{2}\,\mathrm{sech}\left(\tau\right)\tanh\left(\tau\right) p\left(\frac{p_\theta^4}{16} t\left(\tau + \tau_0\right)\right) d\tau, \tag{8}$$

with $t = t\left(\tau\right)$ defined above.

The Gylden's problem also appears in modeling a test charge under the influence of changing radiation pressure from a star. In the authors' words, "Since stars 'flicker, flare, wax, and wane,' we consider functions $p(t)$ that qualitatively represent some of these behaviors." The functions of interest which they study are the following, letting $p_\theta = 2$ for convenience:

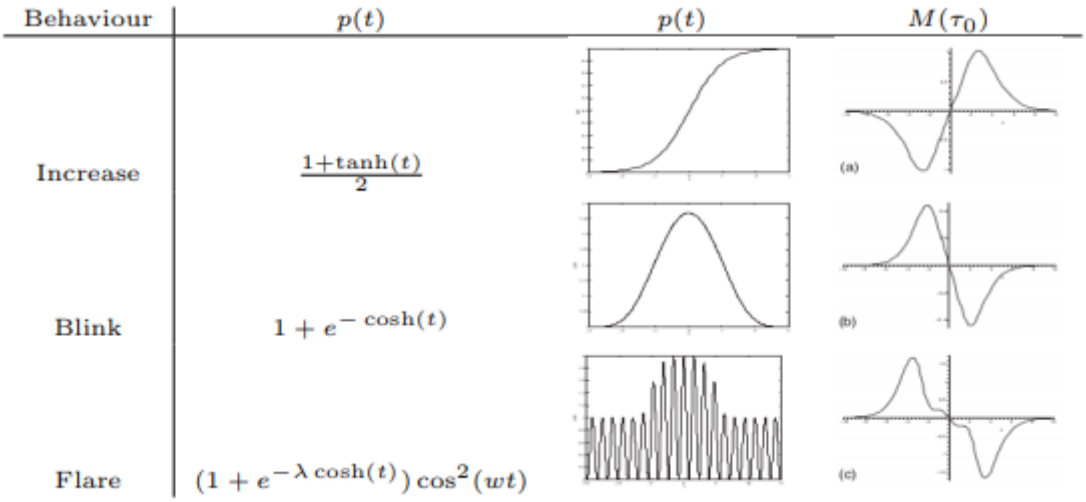

Figure 8: The following diagram is taken from http://depts.washington.edu/amath/wordpress/wp-content/uploads/2014/01/melnikov.pdf, but the plots contained therein come directly from [CM09]. These plots were made using the following parameters: $p_\theta = 2$, $\lambda = 0.01$, and $w = 2$.

1. *Wax or wane.* $p(t) = \frac{1+\tanh(t)}{2}$,
$$M(t_0) = -\sqrt{2}\int_{-\infty}^{\infty} \operatorname{sech}(\tau)\tanh(\tau)\left[1+\tanh\left(t\left(\tau+\tau_0\right)\right)\right]d\tau.$$

2. *Flare.* $p(t) = 1 + e^{-\cosh(t)}$,
$$M(t_0) = -\sqrt{2}\int_{-\infty}^{\infty} \operatorname{sech}(\tau)\tanh(\tau)\left[1+e^{-\cosh(t(\tau+\tau_0))}\right]d\tau.$$

3. *Flicker.* $p(t) = \left(1+e^{-\lambda\cosh(t)}\right)\cos^2(wt)$ $(\lambda > 0)$,[6]
$$M(t_0) = -\sqrt{2}\int_{-\infty}^{\infty} \operatorname{sech}(\tau)\tanh(\tau)\left[1+e^{-\lambda\cosh(t(\tau+\tau_0))}\right]\cos^2\left[wt\left(\tau+\tau_0\right)\right]d\tau.$$

The above functions and their corresponding Melnikov functions are shown qualitatively in Figure 8. Note that each of these perturbations are nonperiodic, yet still induce a simple zero in the Melnikov function (6) at the origin. Consequently, according to the results of this paper, we should expect the system to exhibit chaotic dynamics for these above perturbations regardless of the values we choose for $\lambda > 0$ and $w \in \mathbb{R}$.

3.3 Generalization to heteroclinic orbits

This generalization is trivial, since on pg. 11, we explained how every theorem above also applies to heteroclinic cycles; we just require that the Melnikov functions have simple zeros along each separate heteroclinic orbit. Note that these simple zeroes do not need to occur at the same time t_0, since we only require that transverse crossings occur at *some* time along each orbit separately. Furthermore, we could just align these Melnikov functions to have simple zeroes at the same time, since we could just "start" each heteroclinic orbit along each portion of the cycle at the right times, such that the simple zeros line up at some time t_0.

[6] Yes, even though these symbols look similar, they use w and *not* ω as we might expect.

4 Applications to an SIR Model

SIR (susceptible-infected-recovered) models are used in epidemiology to model the spread of contagious diseases, especially diseases for which general population(s) eventually come to develop a resistance to such diseases, assuming that the population(s) is (are) closed. Even more complicated disease models are often generalizations or are otherwise related to an SIR model. An SIR model consists of three compartments:

- *Susceptible individuals.* (S) This variable represents the number of susceptible individuals who can get sick by coming into contact with an infected individual. If they get sick, then they will be considered infected instead and their numbers will be transferred to I.
- *Infected individuals.* (I) This variable represents the number of infected individuals who can infect individuals in S, but not those in R.
- *Recovered (immune) individuals.* (R) This variable represents the number of recovered individuals who are either immune now to infection or have died.

For most models, we assume that the total population is roughly constant, which is represented by the relationship $S + I + R = 1$. Hence, we can think of S, I, and R as being the proportion of the total population which is susceptible to infection, has been infected, and has recovered/died, respectively. Of course, this assumes either that the birth and death rates are roughly equal, and that any variation due to one of these overcoming the other has a negigible effect on the timescale of the simulation.

This past year, and right in time for the COVID-19 pandemic, mathematician Yanxiang Shi in [Shi20] used Melnikov's method to analyze the effects of periodic and bounded noise on the presence of chaos in a classical SIR model. In fact, this paper follows from an earlier one which only considers periodic forcing by applied mathematicians Paul Glendinning and Louise P. Perry back in 1997 ([GP97]). The model of interest is the following:

$$\begin{aligned} \frac{dS}{dt} &= \mu\left(1 - S\right) - \beta(t) I^2 S, \\ \frac{dI}{dt} &= \beta(t) I^2 S - \left(\gamma + \mu\right) I, \\ \frac{dR}{dt} &= \gamma I - \mu R, \end{aligned} \tag{9}$$

which exhibits a rich array of dynamics. (9) can either be derived heuristically or through an analogous mean-value approach used to derive the ODE system for the SIS model from the agent-based model.

The three variables S, I, and R in (9) are constrained by the relationship mentioned above: $S+I+R = 1$. Hence, as we should expect, (9) directly shows us that $\frac{d}{dt}\left[S + I + R\right] = \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = \mu\left(1 - S\right) - \beta(t) I^2 S + \beta(t) I^2 S - \left(\gamma + \mu\right) I + \gamma I - \mu R = 0 \Rightarrow S + I + R = \text{const}$. Other parameters in the model are the following:

- $\mu > 0$ is the natural mortality rate per capita.
- $\gamma > 0$ is the recovery rate for infected individuals.
- $\beta(t) > 0$ is the rate of transmission per infected, such that $\beta(t) I^2 S$ individuals go from infected to suspectible from t to $t + dt$. $\beta(t)$ is periodic or stochastic with period T_t.

$\beta(t)$ will eventually be related to a forcing function like g as we have in the Melnikov theory in Section 3. The significance in assuming that the rate of transmission periodic is how we might expect the transmission rate of a contagious disease to fluctuate periodically due to seasonal events such as going back to school, the holiday season in November and December, Halloween, Spring Break, etc. For instance, in the past few weeks, we have seen COVID-19 cases skyrocket in the United States due to Thanksgiving and seasonal festivities in time for Christmas, Hanukkah, and New Year. And of course, modeling stochastic noises is important to capture local, random fluctuations which might be present in, say, some dataset.

For most initial conditions when $\beta = \text{const.}$, the model usually predicts that, after an initial transient (which may even look quite aperiodic), the system settles down to fixed points for S, I, and R. Thus, it is quite surprising that adding even a small perturbation around a constant state may produce chaotic dynamics, but as discussed earlier, even a small splitting and transversal crossing of the stable and unstable manifolds along a formerly homoclinic/heteroclinic orbit results in a tangle. In [Shi20], Shi uses the following function: $\beta(t) = 16\mu\left(1 + \varepsilon^2 b_1 + \varepsilon^4 b_2 + \varepsilon^5 b_3 \sin\left(\mu\varepsilon\Omega_1 t\right) + \varepsilon^5 b_4 \xi(t)\right)$, where $\xi(t) = b_4 \sin\left(\Omega_2 t + \Psi\right)$ is a harmonic function with constant amplitude and frequency Ω_2, but with a random phase Ψ. Furthermore:

- $b_1, \ldots, b_4, \Omega_1 \sim \mathcal{O}(1)$ such that $4b_2 - b_1^2 > 0$.
- $\Psi(t) = \sigma B(t) + \Gamma$.
- $B(t)$ is a standard Wiener process, i.e., a continuous-time random walk.
- Γ is a random variable uniformly distributed in $[0, 2\pi)$ which represents a random phase angle.
- $\Omega_1 > 0$ gives the frequency of the periodic excitation.
- $\Omega_2 > 0$ gives the average frequency of the stochastic excitation.
- $\sigma > 0$ gives the strength of the noise.

The Melnikov theory in its original form gives us an easy way to generalize g to a stochastic process as well, as well as the first formulation in the non-Hamiltonian case.[7] In fact, provided that g is bounded, we can just use the Melnikov functions (5) and (6) as is with a stochastic g. That being said, there are some nice statistical properties of $\{\xi(t)\}_{t\in\mathbb{R}}$ which allow us to derive Melnikov functions and criteria for simple zeroes of these, but we will omit these because they go into a lot of probability theory which is beyond the scope of this project. (Although, I did learn about these things in S&DS 600.)

Since we have the conserved quantity $S + I + R = 1$, we can always write out one of the variables in terms of the other two, and thereby remove a degree of freedom from (9). Using this relationship to write S in terms of I and R, Shi proceeds to go through several rescalings and transformations which transform (9) into the following system:[8]

$$\begin{aligned} \frac{dx}{d\tau} &= y + \mathcal{O}(3), \\ \frac{dy}{d\tau} &= b_1^2 - 4b_2 + x^2 + \varepsilon\left(b_1 y + xy - 4b_3 \sin\left(\Omega_1 \tau\right) - 4b_4 \xi\left(\tau\right)\right) + \mathcal{O}\left(\varepsilon^2\right) + \mathcal{O}(3) \end{aligned} \tag{10}$$

[7] We cannot use the second formulation discussed for the non-Hamiltonian case, because the second formulation requires certain assumptions about differentiability, and naturally, a stochastic process is not differentiable in a classical sense.

[8] By $\mathcal{O}(3)$, we mean cubic terms, such that (10) is correct up to higher-order terms that are at least third-order.

for new independent variables x and y, as well as a rescaled time $\tau = \varepsilon t$. (10) is in the standard form for the Takens–Bogdanov bifurcation when $b_3 = b_4 = 0$, i.e., when the nonautonomous terms in the system are zero. This is a two-parameter bifurcation which occurs when the critical equilibrium in question has a zero eigenvalue of multiplicity two. For nearby parameter values, there are two equilibria (a saddle and non-saddle) near this one which undergo a saddle-node/blue-sky bifurcation, i.e., one in which they collide and vanish as the parameters are varied. In particular, the non-saddle equilibrium undergoes a Hopf bifurcation, and hence, a limit cycle appears, but once the SN bifurcation occurs, the limit cycle eventually degenerates into a homoclinic orbit which connects back to the saddle point at the SN bifurcation, and then vanishes in a homoclinic saddle-node bifurcation. By what we know about homoclinic tangles, the presence of a homoclinic orbit usually indicates that a nonautonomous perturbation to the phase portrait *might* lead to chaotic dynamics, and accordingly, the homoclinic saddle-node bifurcation occurs when $\beta = 16\mu$, i.e., when $\varepsilon = 0$, and $\gamma = \mu$. Physically, this occurs when the rate of transmission is some special value and the mortality and recovery rates are equal.

Now that we have (10), we will go through the same process Shi does to derive the Melnikov function. When $\varepsilon = 0$, (10) becomes a Hamiltonian system with Hamiltonian function $H(x, y) = \frac{1}{2}y^2 + (4b_2 - b_1^2)\, x - \frac{1}{3}x^3$. Hence, the Hamiltonian is a first integral for the system, and consequently, all solution curves for x and y satisfy $H(x, y) = \text{const}$. We want to use this to solve for the homoclinic orbit. From what we know about the SN bifurcation, $H(x, y)$ should cross the x-axis *twice* (i.e., there are two fixed points) at the degenerate energy level for $H(x, y) = \text{const}$. where we have the homoclinic orbit. Furthermore, for energies above and below this point, we will have three or one x-axis crossings, respectively or vice-versa. To check when this should occur, we set $y = 0$ and analyze the roots of the polynomial

$$\frac{1}{3}x^3 - \left(4b_2 - b_1^2\right) x + H_0. \tag{11}$$

For there to be repeated roots, the discriminant of (11) must be zero, and for any cubic polynomial $ax^3 + bx^2 + cx + d$, the discriminant is $\Delta_3 = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$.[9] In particular, we have values $a = \frac{1}{3}$, $b = 0$, $c = -\left(4b_2 - b_1^2\right)$, and $d = H_0$, which gives us the following discriminant:

$$\Delta_3 = -\frac{4}{3}\left[-\left(4b_2 - b_1^2\right)\right]^3 - 27\left(\frac{1}{3}\right)^2 H_0^2 = \frac{4}{3}\left(4b_2 - b_1^2\right)^3 - 3H_0^2. \tag{12}$$

Setting (12) to zero, we see that the homoclinic orbit occurs when $H_0 = \frac{2}{3}\left(4b_2 - b_1^2\right)^{3/2}$. Then, the homoclinic orbit $\left(x_h(\tau), y_h(\tau)\right)$ satisfies $\frac{2}{3}\left(4b_2 - b_1^2\right)^{3/2} = \frac{1}{2}y_h^2 + \left(4b_2 - b_1^2\right) x_h - \frac{1}{3}x_h^3$, which we can solve for y_h in terms of x_h and then substitute into (10). This gives us the following first-order ODE

$$\frac{dx_h}{d\tau} = \pm\sqrt{2}\sqrt{\frac{2}{3}\left(4b_2 - b_1^2\right)^{3/2} - \left(4b_2 - b_1^2\right) x_h + \frac{1}{3}x_h^3} \tag{13}$$

at *any* initial condition $x(0) = x_0$ which lies on the homoclinic orbit, where we choose either $+$ or $-$ for $\pm$ depending on if we are on the upper- or lower-half of the homoclinic orbit. Surprisingly enough, by separation of variables, we can solve (13) exactly. Shi uses the initial

[9] See https://brilliant.org/wiki/cubic-discriminant/.

condition $x_0 = -2c$ in the paper, where $c = \sqrt{4b_2 - b_1^2}$, because it leads to some nice cancellations, from which we find that $x_h(\tau) = c - 3c\cosh^{-2}\left(\tau\sqrt{\frac{c}{2}}\right)$. Then, using the fact that $\frac{2}{3}(4b_2 - b_1^2)^{3/2} = \frac{1}{2}y_h^2 + (4b_2 - b_1^2)x_h - \frac{1}{3}x_h^3$ again, $y_h(\tau) = 3c\sqrt{2c}\cosh^{-2}\left(\tau\sqrt{\frac{c}{2}}\right)\tanh\left(\tau\sqrt{\frac{c}{2}}\right)$.

Finally, now that we have solved for the homoclinic orbit as a function of time, we can write out the Melnikov function for (10). When $b_4 = 0$, then there is no stochastic excitation, and (5) becomes

$$M(t_0) = \int_{-\infty}^{\infty} y_h(t)\left[b_1 y_h(t) + x_h(t)\, y_h(t) - 4b_3 \sin\left(\Omega_1(t + t_0)\right)\right] dt. \tag{14}$$

(14) can be integrated explicitly via residue calculus, and a very similar integral can be found in [BN83]. From here, the result is

$$M(t_0) = \frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2} - \frac{24\pi\Omega_1^2 b_3}{\sinh\left(\frac{\Omega_1\pi}{\sqrt{2c}}\right)}\cos(\Omega_1 t_0). \tag{15}$$

Cosine is periodic with period Ω_1, such that $M(t)$ is also periodic with period Ω_1; this is expected from what we noted earlier. Furthermore, $|\cos(\Omega_1 t_0)| < 1$. (15) needs to have simple zeroes for the system to exhibit a homoclinic tangle. Hence, the term $\frac{24\pi\Omega_1^2 b_3}{\sinh\left(\frac{\Omega_1\pi}{\sqrt{2c}}\right)}$ needs to be large enough in magnitude such that the constant term $\frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2}$ does not overpower the oscillating term with cosine in (15); otherwise, the oscillations would not be large enough to hit the x-axis. Hence, chaos exists in this SIR model when $|b_3| > b_c$, where

$$b_c \approx \frac{\frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2}}{24\pi\Omega_1^2}\sinh\left(\frac{\Omega_1\pi}{\sqrt{2c}}\right). \tag{16}$$

Note that (16) is approximate because the Melnikov function (14) is correct up to an $\mathcal{O}(\varepsilon)$ error.

In the case where $b_4 \neq 0$, we do have stochastic excitation, and hence, we can only talk about the Melnikov function in a statistical sense. As noted in [LY96], stochastic noise generally has a tendency to enlargen the range of parameters for which a chaotic attractor or chaotic motion appears, and in cases where we have a parameter which controls the strength of stochastic noises in the system, letting this parameter go to zero in known cases where we have a stochastic Melnikov function in mean-squared sense recovers the classical Melnikov function for the non-stochastic case. For reasons which are not completely elaborated upon in the literature I was able to find so far, the mean representation for the stochastic Melnikov function alone gives too stringent criteria for chaos, but the mean-square representation appears to be more accurate. (I will explain what these are in a moment.) This does make some sense, as if the Melnikov function is on average zero, then this heuristically indicates that the stable and unstable manifolds are crossing transversally "on average," i.e., over many different trials, the distance between these will be zero at t_0. Do not forget that we really only need these manifolds to cross once in a given simulation; having the distance between these manifolds equal zero on average at a given time t_0 is sufficient, but definitely not necessary.

By mean representation of the Melnikov function, we are referring to $\mathbb{E}[M(t_0)]$, where by $\mathbb{E}[\cdot]$, we are referring to the ensemble average of $M(t_0)$ over many test runs of the stochastic

noise at each point $t_0 \in \mathbb{R}$. The random Melnikov function in this case is

$$\begin{aligned} M(t_0) &= \int_{-\infty}^{\infty} y_h(t) \left[b_1 y_h(t) + x_h(t) y_h(t) - 4b_4 \xi(t+t_0) \right] dt \\ &= \frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2} - 4b_4 \int_{-\infty}^{\infty} y_h(t) \xi(t+t_0)\, dt. \end{aligned} \tag{17}$$

The first two terms in (17) represent the mean of the Melnikov process from the periodic forcing, while the last one is related to the stochastic forcing from the bounded noise $\xi(t)$. By ergodicity and stationarity of the bounded noise, the ensemble average of $\int_{-\infty}^{\infty} y_h(t) \xi(t+t_0)\, dt$ goes to zero, and hence, (17) implies that $\mathbb{E}[M(t_0)] = \frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2}$, such that the Melnikov function has a zero in a mean sense only when $b_1 = \frac{5}{7}c = \frac{5}{7}\sqrt{4b_2 - b_1^2} \Rightarrow b_1^2 = \frac{50 b_2}{37}$ or $c = 0 \Rightarrow b_1^2 = 4b_2$, which are very strict conditions and measure zero in our parameter space.

However, we can also try the mean-squared representation, which amounts to showing that the variance, $\mathrm{Var}[M(t_0)] = \mathbb{E}\left[(M(t_0) - \mathbb{E}(M(t_0)))^2 \right]$, is equal to $[\mathbb{E}(M(t_0))]^2$. Using various statistical properties of ξ, Shi derives the following *threshold* condition for the Melnikov function to have zeros in a mean-square sense:

$$\left[\frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2} \right]^2 = \frac{9\pi b_4^2 c}{4}$$

$$\int_{-\infty}^{\infty} \left[\frac{\sigma^2 \omega^4}{\sinh^2\left(\frac{\pi\omega}{4\sqrt{\frac{c}{2}}} \right) \cosh^2\left(\frac{\pi\omega}{4\sqrt{\frac{c}{2}}} \right)} \left(\frac{1}{4(\omega - \Omega_2)^2 + \sigma^4} + \frac{1}{4(\omega + \Omega_2)^2 + \sigma^4} \right) \right] d\omega. \tag{18}$$

While condition (18) superficially appears to be just as strict as the one derived via the mean representation, it in fact is not. Note that b_4 controls the amplitude for the stochastic forcing. Thus, a b_4 that solves (18) gives us a *lower threshold* for the onset of chaos in (10), and therefore in (9) as well.

4.1 An extension for arbitrary periodic forcing

From here, we make a small extension to the results in [GP97; Shi20]. Suppose that, in some application of this SIR model, we have some time series data which represents periodic, autonomous perturbations, σ, to the rate of transmission about $\beta = 16\mu$. Then, as is usual for some time series, we can decompose σ into a finite Fourier series approximation with N modes using a Fast Fourier Transform or something similar, i.e., $\sigma(t) = \sum_{j=0}^{N} a_j \cos(jt) + \sum_{j=1}^{N} b_j \sin(jt)$ for Fourier coefficients $a_0, a_j, b_j \in \mathbb{R}$ for all $j = 1, \ldots, N$. Of course, subtracting out what our Fourier transform takes care of, we might be left with some (hopefully) marginal stochastic noises which can be treated via the mean or mean-squared representations of the Melnikov function found above.

Hence, neglecting the stochastic noise (hopefully, our Fourier transform takes care of most of the noise), we can write our new transmission rate as

$$\begin{aligned} \beta(t) &= 16\mu \left(1 + \varepsilon^2 b_1 + \varepsilon^4 b_2 + \varepsilon^5 \sigma(t) \right) \\ &= 16\mu \left(1 + \varepsilon^2 b_1 + \varepsilon^4 b_2 + \varepsilon^5 \sum_{j=0}^{N} a_j \cos(jt) + \varepsilon^5 \sum_{j=1}^{N} b_j \sin(jt) \right). \end{aligned} \tag{19}$$

Note that changing the forcing term does not affect any of our steps leading up to (14), since these all depend off of the system when $\varepsilon = 0$. The only thing that is affected is the Melnikov function, and using (19) to generalize the steps taken in the paper to derive (10), we find that our Melnikov function this time is

$$\begin{aligned} M(t_0) &= \int_{-\infty}^{\infty} y_h(t)\left[b_1 y_h(t) + x_h(t)\, y_h(t) - 4\sigma(t+t_0)\right] dt \\ &= \int_{-\infty}^{\infty} y_h(t)\left[b_1 y_h(t) + x_h(t)\, y_h(t)\right] dt \\ -4\int_{-\infty}^{\infty} y_h(t) &\left[\sum_{j=0}^{N} a_j \cos\left(j\left(t+t_0\right)\right) + \sum_{j=1}^{N} b_j \sin\left(j\left(t+t_0\right)\right)\right] dt \\ &= \int_{-\infty}^{\infty} y_h(t)\left[b_1 y_h(t) + x_h(t)\, y_h(t)\right] dt \\ -4\sum_{j=0}^{N} a_j \int_{-\infty}^{\infty} y_h(t)\cos\left(j\left(t+t_0\right)\right) dt &- 4\sum_{j=1}^{N} b_j \int_{-\infty}^{\infty} y_h(t)\sin\left(j\left(t+t_0\right)\right) dt. \end{aligned} \tag{20}$$

From here, note the trigonometric identity $\cos(x) = \sin\left(x + \frac{\pi}{2}\right)$ for all $x \in \mathbb{R}$. Then, excluding the $j = 0$ term, the penultimate integral in (20) becomes

$$\begin{aligned} \int_{-\infty}^{\infty} y_h(t)\cos\left(j\left(t+t_0\right)\right) dt &= \int_{-\infty}^{\infty} y_h(t)\sin\left[j\left(t+t_0\right) + \frac{\pi}{2}\right] dt \\ &= \int_{-\infty}^{\infty} y_h(t)\sin\left[j\left(t+t_0+\frac{\pi}{2j}\right)\right] dt. \end{aligned} \tag{21}$$

Hence, by direct analogy with (15), noting that y_h is an odd function which will integrate to zero from $-\infty$ to ∞, and using (21), 20 simplifies to

$$\begin{aligned} M(t_0) &= \frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2} \\ -\sum_{j=1}^{N} \frac{24\pi j^2 a_j b_3}{\sinh\left(\frac{j\pi}{\sqrt{2c}}\right)} \cos\left(j\left(t_0 + \frac{\pi}{2j}\right)\right) &- \sum_{j=1}^{N} \frac{24\pi j^2 b_j b_3}{\sinh\left(\frac{j\pi}{\sqrt{2c}}\right)} \cos\left(jt_0\right). \end{aligned} \tag{22}$$

Using the fact that $\cos\left(x + \frac{\pi}{2}\right) = -\sin(x)$ for all $x \in \mathbb{R}$, (22) further simplifies to

$$\begin{aligned} M(t_0) &= \frac{6}{5}(2c)^{5/2} b_1 - \frac{3}{7}(2c)^{7/2} \\ +24\pi b_3 \sum_{j=1}^{N} \frac{j^2 a_j}{\sinh\left(\frac{j\pi}{\sqrt{2c}}\right)} \sin\left(jt_0\right) &- 24\pi b_3 \sum_{j=1}^{N} \frac{j^2 b_j}{\sinh\left(\frac{j\pi}{\sqrt{2c}}\right)} \cos\left(jt_0\right). \end{aligned} \tag{23}$$

Remember now: The condition for this arbitrary periodic forcing σ to induce chaotic dynamics in our system is achieved if (23) has a simple zero at some $t_0 \in \mathbb{R}$. Note however that (23) is still a finite Fourier series. In [Boy06], atmospheric scientist and applied mathematician John P. Boyd derives an algorithm for calculating the roots of a finite Fourier series. He shows that for a finite Fourier series of the form $f_N(t) = \sum_{j=0}^{N} a_j \cos(jt) + \sum_{j=1}^{N} b_j \sin(jt)$, we can define a $2N \times 2N$ matrix, B, constructed from the coefficients of this Fourier series. Then, if B has eigenvalues given by z_k, the $2N$ roots for $f_N(t)$ are of the form $t_k = \arg(z_k) + 2\pi m - i\log(|z_k|)$, where $m \in \mathbb{Z}$. Thus, if we want to check if (23) has a simple root, it is enough to check that $\log(|z_k|) = 0 \Rightarrow |z_k| = 1$ for some $k = 1, \ldots, 2N$, i.e., at least one of the eigenvalues for B must lie on the unit circle in the complex plane.

References

[BN83] C. Baesens and G. Nicolis. "Complex bifurcations in a periodically forced normal form". In: *Zeitschrift für Physik B Condensed Matter* 52.4 (Dec. 1983), pp. 345–354. DOI: 10.1007/bf01307404. URL: https://doi.org/10.1007/bf01307404.

[Ber88] Andrea Louise Bertozzi. "Heteroclinic orbits and chaotic dynamics in planar fluid flows". In: *SIAM Journal of Mathematical Analysis* 19 (Nov. 1988), pp. 1271–1294.

[Boy06] John P. Boyd. "Computing the zeros, maxima and inflection points of Chebyshev, Legendre and Fourier series: solving transcendental equations by spectral interpolation and polynomial rootfinding". In: *Journal of Engineering Mathematics* 56.3 (Nov. 2006), pp. 203–219. DOI: 10.1007/s10665-006-9087-5. URL: https://doi.org/10.1007/s10665-006-9087-5.

[CZ18] Maciej J. Capiński and Piotr Zgliczyński. "Beyond the Melnikov method II: Multidimensional setting". In: *Journal of Differential Equations* 265.9 (Nov. 2018), pp. 3988–4015. DOI: 10.1016/j.jde.2018.05.028. URL: https://doi.org/10.1016/j.jde.2018.05.028.

[CM09] César Castilho and Marcelo Marchesin. "A practical use of the Melnikov homoclinic method". In: *Journal of Mathematical Physics* 50.11 (Nov. 2009), p. 112704. DOI: 10.1063/1.3251336. URL: https://doi.org/10.1063/1.3251336.

[CW17] Fengjuan Chen and Qiudong Wang. "High-Order Melnikov Method for Time-Periodic Equations". In: *Advanced Nonlinear Studies* 17.4 (Oct. 2017), pp. 793–818. DOI: 10.1515/ans-2017-6017. URL: https://doi.org/10.1515/ans-2017-6017.

[DG00] A. Delshams and P. Gutiérrez. "Splitting Potential and the Poincaré-Melnikov Method for Whiskered Tori in Hamiltonian Systems". In: *Journal of Nonlinear Science* 10.4 (Aug. 2000), pp. 433–476. DOI: 10.1007/s003329910016. URL: https://doi.org/10.1007/s003329910016.

[FOY83] J. Doyne Farmer, Edward Ott, and James A. Yorke. "The dimension of chaotic attractors". In: *Physica D: Nonlinear Phenomena* 7.1-3 (May 1983), pp. 153–180. DOI: 10.1016/0167-2789(83)90125-2. URL: https://doi.org/10.1016/0167-2789(83)90125-2.

[Gao+06] J. B. Gao et al. "Distinguishing chaos from noise by scale-dependent Lyapunov exponent". In: *Phys. Rev. E* 74 (6 Dec. 2006), p. 066204. DOI: 10.1103/PhysRevE.74.066204. URL: https://link.aps.org/doi/10.1103/PhysRevE.74.066204.

[GP97] Paul Glendinning and Louise P. Perry. "Melnikov analysis of chaos in a simple epidemiological model". In: *Journal of Mathematical Biology* 35.3 (Feb. 1997), pp. 359–373. DOI: 10.1007/s002850050056. URL: https://doi.org/10.1007/s002850050056.

[GHS12] A. Granados, S. J. Hogan, and T. M. Seara. "The Melnikov Method and Subharmonic Orbits in a Piecewise-Smooth System". In: *SIAM Journal on Applied Dynamical Systems* 11.3 (Jan. 2012), pp. 801–830. DOI: 10.1137/110850359. URL: https://doi.org/10.1137/110850359.

[GH83] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer New York, 1983. DOI: 10.1007/978-1-4612-1140-2. URL: https://doi.org/10.1007/978-1-4612-1140-2.

[HB10] A. Hanslmeier and R. Brajša. "The chaotic solar cycle". In: *Astronomy and Astrophysics* 509 (Jan. 2010), A5. DOI: 10.1051/0004-6361/200913095. URL: https://doi.org/10.1051/0004-6361/200913095.

[HK03] Boris Hasselblatt and Anatole Katok. *A first course in dynamics : with a panorama of recent developments*. Cambridge New York: Cambridge University Press, 2003. ISBN: 978-0-521-58750-1.

[Hen82] M. Henon. "On the numerical computation of Poincaré maps". In: *Physica D: Nonlinear Phenomena* 5.2-3 (Sept. 1982), pp. 412–414. DOI: 10.1016/0167-2789(82)90034-3. URL: https://doi.org/10.1016/0167-2789(82)90034-3.

[Kin06] W. Kinsner. "Characterizing chaos through Lyapunov metrics". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 36.2 (2006), pp. 141–151. DOI: 10.1109/TSMCC.2006.871132.

[KLW14] J. Knobloch, J. Lamb, and K. Webster. "Using Lin's method to solve Bykov's problems". In: *Journal of Differential Equations* 257 (2014), pp. 2984–3047.

[KKG11] Egemen Kolemen, N. Jeremy Kasdin, and Pini Gurfil. "Multiple Poincaré sections method for finding the quasiperiodic orbits of the restricted three body problem". In: *Celestial Mechanics and Dynamical Astronomy* 112.1 (Oct. 2011), pp. 47–74. DOI: 10.1007/s10569-011-9383-x. URL: https://doi.org/10.1007/s10569-011-9383-x.

[Kuk07] Peter Kukučka. "Melnikov method for discontinuous planar systems". In: *Nonlinear Analysis: Theory, Methods & Applications* 66.12 (June 2007), pp. 2698–2719. DOI: 10.1016/j.na.2006.04.001. URL: https://doi.org/10.1016/j.na.2006.04.001.

[LR04] Stefano Lenci and Giuseppe Rega. "Higher-order Melnikov functions for single-DOF mechanical oscillators: theoretical treatment and applications". In: *Mathematical Problems in Engineering* 2004.2 (2004), pp. 145–168. DOI: 10.1155/s1024123x04310045. URL: https://doi.org/10.1155/s1024123x04310045.

[Let+06] C. Letellier et al. "Evidence for low dimensional chaos in sunspot cycles". In: *Astronomy & Astrophysics* 449.1 (Mar. 2006), pp. 379–387. DOI: 10.1051/0004-6361:20053947. URL: https://doi.org/10.1051/0004-6361:20053947.

[LY96] H. Lin and S. C. S. Yim. "Deterministic and stochastic analyses of chaotic and overturning responses of a slender rocking object". In: *Nonlinear Dynamics* 11.1 (Sept. 1996), pp. 83–106. DOI: 10.1007/bf00045052. URL: https://doi.org/10.1007/bf00045052.

[Lor95] Edward Lorenz. *The essence of chaos*. Seattle: University of Washington Press, 1995, p. 181. ISBN: 978-0295975146.

[ML01] Alfredo Medio and Marji Lines. *Nonlinear dynamics : a primer*. Cambridge New York: Cambridge University Press, 2001. ISBN: 978-0-521-55874-7.

[Mel63] V. K. Mel'nikov. "On the stability of the center for time-periodic perturbations". English. In: *Trans. Mosc. Math. Soc.* 12 (1963), pp. 1–56. ISSN: 0077-1554; 1547-738X/e.

[MMC91] Michael D. Mundt, W. Bruce Maguire II, and Robert R. P. Chase. "Chaos in the sunspot cycle: Analysis and prediction". In: *Journal of Geophysical Research: Space Physics* 96.A2 (1991), pp. 1705–1716. DOI: https://doi.org/10.1029/90JA02150. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/90JA02150. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/90JA02150.

[PM99] H. Poincaré and R. Magini. "Les méthodes nouvelles de la mécanique céleste". In: *Il Nuovo Cimento* 10.1 (July 1899), pp. 128–130. DOI: 10.1007/bf02742713. URL: https://doi.org/10.1007/bf02742713.

[Shi20] Yanxiang Shi. "Melnikov analysis of chaos in a simple SIR model with periodically or stochastically modulated nonlinear incidence rate". In: *Journal of Biological Dynamics* 14.1 (Jan. 2020), pp. 269–288. DOI: 10.1080/17513758.2020.1718222. URL: https://doi.org/10.1080/17513758.2020.1718222.

[Shi91] L. P. Shilnikov. "Multidimensional Hamiltonian chaos". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 1.2 (Aug. 1991), pp. 134–136. DOI: 10.1063/1.165822. URL: https://doi.org/10.1063/1.165822.

[Sma67] S. Smale. "Differentiable dynamical systems". In: *Bulletin of the American Mathematical Society* 73.6 (Nov. 1967), pp. 747–818. DOI: 10.1090/s0002-9904-1967-11798-1. URL: https://doi.org/10.1090/s0002-9904-1967-11798-1.

[Tuc02] Warwick Tucker. "Computing accurate Poincaré maps". In: *Physica D: Nonlinear Phenomena* 171 (2002), pp. 127–137.

[Wig03] Stephen Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*. New York: Springer, 2003. ISBN: 978-0-387-00177-7.