Chapter 2 A Survey of Chaos Theory

Abstract This chapter briefly summarizes chaos theory. The chapter begins with describing chaos as bounded aperiodic random-like deterministic motion, which is sensitive to initial states and thus unpredictable after a certain time of a system. The geometrical structure of chaos is analyzed via the Poincaré map. Three typical routes to chaos are introduced as period-doubling sequence, intermittency, and quasiperiodic torus breakdown. The chapter covers two main numerical approaches to identify chaos, Lyapunov exponents and power spectra. The Melnikov theory is presented to predict the transversal intersection of stable and unstable manifolds of a saddle point. Such an intersection results in complicated dynamical behaviors which are sensitive to initial conditions. Finally, chaos is treated in the context of Hamiltonian systems. KAM theorem is stated without the proof. Two mechanisms of Hamiltonian chaos are illustrated as KAM tori breakup and Arnol'd diffusion. The Melnikov theory is generalized to higher-dimensional Hamiltonian systems.

Keywords chaos, Poincaré map, period-doubling sequence, intermittency, quasiperiodic torus breakdown, Lyapunov exponents, power spectra, Melnikov theory, KAM theorem, KAM tori breakup, Arnol'd diffusion

This chapter briefly summarizes chaos theory, most of which will be applied in the subsequent chapters. The chapter begins with describing chaos as bounded aperiodic random-like deterministic motion, which is sensitive to initial states and thus unpredictable after a certain time of a system. The geometrical structure of chaos is analyzed via the state space as well as the Poincaré map. Three typical routes to chaos are introduced as period-doubling sequence, intermittency, and quasiperiodic torus breakdown. The chapter covers two main numerical approaches to identify chaos, Lyapunov exponents and power spectra. The Melnikov theory is presented to predict the transversal intersection of stable and unstable manifolds of a saddle point. Such an intersection is explained to result in complicated dynamical behaviors which are sensitive to initial conditions. Finally, chaos is treated in the context of Hamiltonian systems. KAM theorem is stated without the proof. Two mechanisms of Hamiltonian chaos are illustrated as KAM tori breakup and Arnol'd diffusion. The Melnikov theory is generalized to higher-dimensional Hamiltonian systems. This chapter is only a brief survey of chaos, and references [1-6] present

Y. Liu et al., *Chaos in Attitude Dynamics of Spacecraft*

a more comprehensive treatment of chaos with the emphasis on engineering applications.

2.1 The Overview of Chaos

2.1.1 Descriptions of Chaos

Motions of many natural or engineering systems, including attitude motion of spacecraft, can be governed by a set of equations derived from the natural laws such as Newton's laws or Euler's equation. The set of equations, defined mathematically as a **dynamical system**, yields the time evolution of the state of a system from the knowledge of its previous history. Therefore, the state at any time can be determined by the governing equations and the initial states. The equations desceibing a dynamical system may be algebraic or differential equations.

In modern science, **chaos** is a term to describe a type of motion, or time evolution resulting from a dynamical system, that appears, on detailed examination, to be completely disordered and extremely complex. The disorder and complicacy are due to the following reasons.

Chaos is a recurrent aperiodic motion. Hence, chaos can be practically defined as a bounded steady-state response that is not an equilibrium state or a periodic motion, or a quasiperiodic motion. For systems with finite degrees of freedom, a bounded response of linear systems must be an equilibrium state, a periodic motion, or a quasiperiodic motion. Hence chaos is a striking feature of a nonlinear system. As a recurrent motion, chaos is bounded so that it will trend to the infinite.

Chaotic motions are also characterized by **sensitivity to initial states**; that is, tiny differences in the initial conditions can be quickly amplified to produce huge differences in the response. Due to such sensitivity, the long-term prediction for chaos is impossible, because all initial conditions have to be prescribed in a certain precision, while, after enough time, the motion depends on the digits in the conditions beyond the precision. That is, chaos is unpredictable after enough time because a small difference in the initial conditions beyond their precision will result in rapidly (usually exponentially) growing perturbation of the motion. This phenomenon is vividly called **butterfly effect**. A disturbance caused by the wings of a butterfly in Shanghai can lead to a rainstorm a few days later in Toronto.

Chaos, as a recurrent aperiodic motion, has no pattern or order to follow, just like a stochastic process. Actually, the spectrum of a chaotic motion has a continuous broadband, which is the same as a true random signal. In contrast, the spectra of periodic or quasiperiodic motions consist of a number of sharp spikes. In addition

to the broadband component, the spectrum of a chaotic motion often contains spikes, which indicate the predominant frequencies of the motion. Chaos is the superposition of an infinite number of unstable periodic motions. Therefore, a chaotic motion may settle for a short time near a periodic motion and then may switch to another periodic motion with a different period. However, chaos usually describes a special type of motion in a deterministic system that is without any random inputs. Hence, the random-likeness of chaotic motion is called **intrinsic stochasticity** or **spontaneous stochasticity**. A true stochastic process is unpredictable at any time, while chaos can be predicted only after a very short time from the beginning.

In short, as a steady-state response of a deterministic system, chaos is sensitive to initial states and thus unpredictable after a certain time, and is recurrent but either periodic or quasiperiodic hence like a random single.

2.1.2 Geometrical Structures of Chaos

The recurrent aperiodicity of chaos can be intuitively illustrated in the phase plane. For a single degree of freedom, two independent parameters are needed to describe the state of motion completely (not only the position, but also the position change). These parameters are usually chosen as the generalized displacement and velocity of the system. When the parameters are used as coordinate axes, the resulting graphical illustration of the motion is called the phase plane representation. Thus each point in the phase plane represents a possible state of the system. The state of the system changes with the time evolution. A typical of representative point in the phase plane, such as the point representing the state of the system at time $t=0$, moves and traces a curve known as a **trajectory** or an **orbit**. The trajectory demonstrates how the motion beginning at a given initial state varies with time. On the phase plane, an equilibrium state is represented by a point on the displacement (usually horizontal) axis. The trajectory of a periodic motion is a closed curve, because the trajectory repeats itself after a period. A chaotic motion is represented by a trajectory that never closes and repeats itself because of the aperiodicity of the motion, and the trajectory is located in a bounded region due to the recurrence of the motion. Therefore, the trajectory of chaos in the phase plane usually occupies a part of the phase space. However, the trajectory of a quasiperiodic motion does not close on itself either, although it looks much more regular than a chaotic trajectory. In addition, it is difficult in practice to differentiate a trajectory of chaos from that of a periodic motion with a sufficient large period. Therefore, new techniques are necessary to describe the recurrent aperiodicity of chaos and distinguish chaos from periodic or quasiperiodic motions.

Consider a set of ordinary differential equations in the vector form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x},t) \quad \boldsymbol{x} \in R^n, \quad t \in R \tag{2.1.1}$$

where $\boldsymbol{f}$ is a vector function defined in R^{n+1} with its value range in R^n, $\boldsymbol{x}$ is a n dimensional vector to specify the state of the system, and t is time. The vector $\boldsymbol{x}$ is called a state vector, and R^n in which $\boldsymbol{x}$ evolves is called a state space. A stage space is called a phase space when half of the state variables are displacements and the other half are velocities. Obviously, the phase space is the generalization of the phase plane. If the vector function does not depend explicitly on time t, the dynamical system governed by Eq. (2.2.1) is call autonomous; otherwise, it is called nonautonomous.

The geometrical structure representing an asymptotically long-time behavior in a state space is called an attractor. Mathematically, an **attractor** is an indecomposable, closed, invariant set that attracts all trajectories starting at points in some neighborhood. Here, an **indecomposable set** is a set that cannot be separated into smaller pieces, and an **invariant set** is a set that trajectories starting in the set remain in it for all time.

An attractor may be a point, called a **point attractor**, which represents an asymptotically stable equilibrium state. An attractor may be a closed curve, called a **periodic attractor**, which represents a periodic motion. An attractor may be a torus, called a **quasiperiodic attractor**, which represents a quasiperiodic motion. If an attractor is not a point attractor, a periodic attractor, or a quasiperiodic attractor, it is call a **strange attractor**. A strange attractor usually represents a chaotic motion, and thus it is also referred as a **chaotic attractor**. A chaotic attractor has typically embedded within it an infinite number of periodic orbits that are unstable. Mathematically, unstable periodic orbits are dense in a chaotic attractor. The orbits pass through any neighborhood, no matter how small it is, of any point on the attractor.

The Ponicaré map will be defined in a state space. It can discretize a trajectory of a dynamical system governed by a set of ordinary-differential equations into a set of points. The **Poincaré map** or the **Poincaré section map**, named after Henri Poincaré, is the intersection of a trajectory, which moves periodically, quasi-periodically, or chaotically, in an n-dimensional state space, with a transversal hypersurface whose dimension is $n - 1$. Here a transversal hypersurface means that, at the intersection point, the normal of the hypersurface is not orthogonal to the tangent of the trajectory. More specifically, one considers a trajectory with initial conditions on the hyperplane and observes the point at which this trajectory returns to the hyperplane. The **Poincaré section** refers to the hyperplane, and the Poincaré map refers to the map of points in the hyperplane induced by the intersections. If the vector function $\boldsymbol{f}$ in Eq. (2.1.1) is periodic in time with period T, then the Poincaré map can be constructed by monitoring stroboscopically the state variables at intervals of the period T. The Poincaré map can be denoted by

$$\boldsymbol{X}_{i+1} = \boldsymbol{P}(\boldsymbol{X}_i) \qquad \boldsymbol{X} \in R^n, \quad i \in Z \tag{2.1.2}$$

The points on the Poincaré section obtained by iterating $\boldsymbol{P}$

$$\{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_i, \cdots\} \qquad (\boldsymbol{X}_i = \boldsymbol{x}(iT),\ i = 1, 2, \cdots) \tag{2.1.3}$$

which is also called sometimes the Poincaré map, can be used to determine if a motion is periodic, quasiperiodic, or chaotic.

For the Poincaré map obtained by sampling the state variables at intervals of T, a periodic motion with the period mT will collect m points on the Poincaré section. Therefore, the Poincaré map (2.1.3) of a periodic motion is a set of finite points. The Poincaré map (2.1.3) of a quasiperiodic motion does not contain finite points. To explore its characteristic, consider the following simple example. A motion given by

$$x_1(t) = A\sin t + B\sin \pi t,\ x_2(t) = C\cos t + D\cos \pi t \tag{2.1.4}$$

is quasiperiodic because it is characterized by the two incommensurate frequencies 1 and π, and called two-period quasiperiodic. A Poincaré map is constructed by sampling the trajectory at intervals 2π starting at $t = 0$. Then the discrete points are collected as

$$x_{1i} = B\sin i\pi^2,\ \ x_{2i} = C + D\cos i\pi^2 \ \ (B, D \neq 0) \tag{2.1.5}$$

Hence all mapping points on the closed curve are defined by

$$\frac{x_{1i}^2}{B^2} + \frac{(x_{2i} - C)^2}{D^2} = 1 \tag{2.1.6}$$

It can be further demonstrated that the points fill densely the closed curve. Therefore, the Poincaré map (2.1.3) of a quasiperiodic motion is a set of infinite points located densely on a loop or a torus. Since a chaotic motion is neither periodic nor quasiperiodic, the Poincaré map (2.1.3) of chaos is a set of infinite points that do not fill any loops or tori.

2.1.3 Routes to Chaos

In addition to its physics and geometry, chaos can also be investigated from the view of its emerging processes with the variation of system parameters. The processes are often referred as **routes to chaos** or **transitions to chaos**. In most systems, chaos occurs only for some range of parameter values. How a regularly behaving system becomes chaotic is a fundamental and significant problem. Theoretically, routes to chaos can reveal the nature and mechanisms of chaos. Practically, routes to chaos can serve as an effective approach to identify chaos, especially to distinguish chaos from truly random motion. Actually, for a system

with bounded irregular motion, if a route to chaos appear with a change of a system parameter, then the motion is quite surely chaotic rather than stochastic. There are several fairly well-understood and (relatively) easily recognizable routes to chaos in a particular transition process in a prescribed system. These routes include period-doubling sequence, intermittency, and quasiperiodic torus breakdown.

The period-doubling cascade, also referred to as period-doubling scenario, is the best understood route to chaos. In the period-doubling cascade, as a system parameter is gradually varied, a periodic motion transitions to a chaotic motion via a sequence of period-doubling bifurcations. This route was discovered in the context of 1-dimensional maps by Feigenbaum in 1978 [7], and it is now known to occur in almost all kinds of systems.Consider a system with parameter μ. In multi-parameter systems, one can vary one of them and fix the others. Suppose the motion with period T for $\mu=\mu_0$. With changing μ, when $\mu = \mu_1$, the period of motion becomes $2T$. Such a sudden change of the motion is called a **period-doubling bifurcation**. Generally speaking, if the motion is with period 2^kT for $\mu = \mu_k$, the period-doubling bifurcation at $\mu = \mu_{k+1}$ turns the motion period into $2^{k+1}T$. As the motion period continues to double, it becomes larger and larger, and finally, infinite, which actually corresponds to aperiodic motion. Observing the Poincaré map, one finds that one point becomes two points, two points become four points, and so on. At last, an infinite point set is created, and chaos appears. Theoretically, in the absence of noise, an infinite number of period-doubling bifurcations occur in the transition to chaos. Practically, as noise is always present, some of the higher period-doubling bifurcations may be suppressed by the noise, resulting in a finite sequence of bifurcations. It should be remarked that

$$\delta = \lim_{n\to\infty} \frac{\mu_n - \mu_{n-1}}{\mu_{n+1} - \mu_n} \tag{2.1.7}$$

is a constant for the infinite sequence of period-doubling bifurcation values $\{\mu_n\}$. In fact, in a certain class of systems, different systems have the same constant regardless of the details of each system. Therefore, δ is called a universal constant. The universality characterizes the period-doubling cascade as a route to chaos.

Intermittency is another frequently observed route to chaos. **Intermittency** is a phenomenon characterized by random alternations between a regular motion and relatively short irregular bursts. The term *intermittency* has been used in the theory of turbulence to denote burst of turbulent motion on the background of laminar flow. During early stages of intermittency, for a certain system parameter value, the motion of the system is predominantly periodic with occasional bursts of chaos. As the parameter value is changed, the chaotic bursts become more frequent, and the time spent in a state of chaos increases and the time spent in periodic motion decreases until, finally, chaos is observed all the time. As a result, the periodic motion becomes chaotic motion. This route was found by Pomeou and Manneville in

1980 [8]. Geometrically, the intermittency route is associated with a periodic attractor in the state space bifurcating into a new, larger chaotic attractor, including previous periodic trajectories as its subset. The trajectory of a system can reside some time in the chaotic part of the attractor, but it is ultimately attracted back to the periodic part. As the system parameter is varied, the relative proportion of the chaotic part increases, ultimately covering the whole attractor.

Quasiperiodic torus breakdown is the third typical way that a system may evolve as its parameter is changed. **Quasiperiodic torus breakdown route** signifies the destruction of the torus and the emergence of a chaotic attractor. The system, if it is not externally driven by a periodic action, may be at equilibrium. As the system parameter is varied, the equilibrium may lose its stability, leading to the emergence of a stable periodic motion. Such a change resulting in a new motion frequency is called the **Hopf bifurcation**. In the state space, a point attractor becomes a periodic attractor. With a further change in the parameter, the periodic attractor undergoes a secondary Hopf bifurcation, resulting in a 2-period quasiperiodic attractor. The trajectories in the state space reside on the surface of a torus. If the two frequencies are incommensurable, the trajectory eventually covers the surface of the torus. For some systems, further changes in the parameter result in the introduction of a third frequency. In the state space, the trajectories live on a 3-dimensional torus. With further parameter changes, the motion of a system becomes chaotic. Some systems may apparently switch directly from two-periodic quasiperiodic motion to chaos. The discovery of the route to chaos started with Ruelle and Takens, who in 1971 proposed an alternative to the Landau-Hopf picture of infinitely increasing number of incommensurable frequencies for the onset mechanism of turbulence [9]. As Ruelle and Takens demonstrated, quasiperiodic motion on a torus with 4 incommensurable frequencies is generally unstable and can be perturbed into a strange attractor corresponding to turbulent motion. In 1978, Hewhouse, Ruelle and Takens proved that a torus with 3 incommensurable frequencies is generally unstable and can be perturbed into chaos [10]. In the same year, Swinney and Gollub experimentally showed that a quasiperiodic motion with 2 incommensurable frequencies directly leads to chaos [11]. In 1982, Feigenbaum, Kadanoff and Shenker revealed the universality in the quasiperiodic route to chaos [12]. In 1983, Grebogi, Ott, and Yorke confirmed that the quasiperiodic torus with 3 incommensurable frequencies is usually stable and thus a periodic motion becomes chaotic after only two bifurcations [13]. That is, the 2-period quasiperiodic motion may lead to chaos directly.

Chaos may suddenly occur in a system. One mechanism to account for a sudden appearance of chaos is crisis, the term introduced by Grebogi, Ott, and Yorke in 1983 [14]. A **crisis** is a sudden qualitative change in which a chaotic attractor disappears or suddenly expands in size as a system parameter is varied. Therefore, if a typical route to chaos is observed, one may conclude that chaos is taking place in a system. On the other hand, even if no typical route to chaos is observed, one cannot exclude the possibility of the appearance of chaos.

2.2 Numerical Identification of Chaos

2.2.1 Introduction

The numerical identification of chaos is an important aspect of nonlinear dynamics. The identification of chaos is some diagnostic tests to determine if chaotic behavior occurs in a specific system. Some numerical characteristics associated with the motion of a system can be used to identify chaos. These characteristics include Lyapunov exponents, fractal dimensions, power spectra, and entropies. If one or more of these characteristics satisfy certain conditions, the motion may be chaotic.

As explained in the previous section, chaos can be described in different aspects. Quantifying these descriptions leads to corresponding numerical characteristics. To specify the sensitivity of chaos to initial states, Lyapunov exponents are introduced. To highlight the recurrent aperiodicity of chaos, various dimensions can be defined. To detect the stochasticity of chaos, power spectra may be used. To reveal the unpredictability of chaos, entropies can be employed. However, only Lyapunov exponents and power spectra will be presented in this section. They will be applied in the following chapters.

2.2.2 Lyapunov Exponents

The extreme sensitivity to initial states makes neighboring trajectories diverge rapidly as the time elapses. Therefore, a numerical approach to identify chaos can be developed based on the quantitative characterization of the divergence among the neighboring trajectories. Lyapunov exponents are a set of numerical characteristics to quantify the divergence of trajectories. In an n–dimensional state space, the displacement between two points on two nearby trajectories has n components in n different directions. Trajectories may diverge in some directions, but they must converge in other directions. Otherwise, the motion will become unbounded. Hence the change rates of the distance along the n directions are different, and each change rate is a Lyapunov exponent. This intuitionistic idea can be presented in a more rigorous way as follows.

Consider a dynamical system governed by

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \quad \boldsymbol{x} \in R^n \tag{2.2.1}$$

Choose two trajectories L_0 and L_1 starting at two close initial conditions $\boldsymbol{x}_0$ and $\boldsymbol{x}_0 + \Delta\boldsymbol{x}_0$, respectively. Define L_0 starting at $\boldsymbol{x}_0$ as the unperturbed trajectory, and L_1 starting at $\boldsymbol{x}_0+\Delta\boldsymbol{x}_0$ as a perturbed trajectory. Denote $\boldsymbol{x}(\boldsymbol{x}_0 + \Delta\boldsymbol{x}_0, t)$ and $\boldsymbol{x}(\boldsymbol{x}_0, t)$ as

the points at time t on the perturbed and unperturbed trajectories respectively. Denote the difference as $\boldsymbol{w}(\boldsymbol{x}_0, t) = \boldsymbol{x}(\boldsymbol{x}_0 + \Delta\boldsymbol{x}_0, t) + \boldsymbol{x}(\boldsymbol{x}_0, t)$. Then for sufficient small $\boldsymbol{w}$, $\boldsymbol{w}$ satisfies the linearized equation of Eq. (2.2.1) at $\boldsymbol{x}_0$, namely,

$$\dot{\boldsymbol{w}} = \mathbf{D}\boldsymbol{f} \cdot \boldsymbol{w} \tag{2.2.2}$$

where $\mathbf{D}\boldsymbol{f}$ is the $n \times n$ Jacobi matrix calculated at $\boldsymbol{x}_0$. Now the averaged rate of exponential expansion or contraction in the direction of $\boldsymbol{w}$ on the trajectory starting at $\boldsymbol{x}_0$ is given by

$$\lambda(\boldsymbol{x}_0, \boldsymbol{w}) = \lim_{\substack{t \to \infty \\ \boldsymbol{w}_0 \to 0}} \frac{1}{t} \ln \frac{\|\boldsymbol{w}\|}{\|\boldsymbol{w}_0\|} \tag{2.2.3}$$

where the symbol $\| \ \|$ denotes a vector norm and $\boldsymbol{w}_0 = \boldsymbol{w}(\boldsymbol{x}_0, 0)$. In the n-dimensional state space, all $\boldsymbol{w}$ form n-dimensional state space moving along the trajectory L_0. Take a set of base vector $\{\boldsymbol{e}_i, i = 1,2,\cdots,n\}$. For every $\boldsymbol{e}_i$, Eq. (2.2.3) yields $\lambda(\boldsymbol{x}_0, \boldsymbol{e}_i)$ ($i = 1, 2, \cdots, n$). Those numbers are ordered such that

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \tag{2.2.4}$$

The number λ_i is called the **Lyapunov exponent**, and the set of n number λ_i is called the **Lyapunov spectrum**.

Roughly speaking, the Lyapunov exponents of a trajectory characterize the mean exponential rates of divergence (in different directions) of other trajectories surrounding it. A Lyapunov exponent may be positive or negative. A positive Lyapunov exponent implies the divergence in the corresponding direction. That is, all trajectories near the trajectory under consideration diverge locally from it along the direction. A negative exponent implies the constriction in the corresponding direction. All trajectories close to the trajectory under consideration locally converge toward it in the direction. Therefore, if all Lyapunov exponents are negative, the motion is in a stable equilibrium.

For a limit cycle of an autonomous system, there is always a zero Lyapunov exponent corresponding to an initial deviation along a tangent to the closed orbit. In addition, for a stable periodic motion, all other Lyapunov exponents are negative. Those negative exponents essentially correspond to perturbations along directions normal to the closed orbit. For an m-torus, m Lyapunov exponents are zero because there are m tangential directions to the torus along which there is neither growth nor decay.

If there are one or more positive Lyapunov exponents, these exponents correspond to the directions along which the initial disturbances become larger and larger. Thus a bounded trajectory with one or more positive Lyapunov exponents represents chaotic motion.

Based on the above-mentioned analysis, the types of attractors and motions can be summarized in Table 2.1 for low dimensional systems.

Table 2.1 Classification of Attractors in Low Dimensional State Spaces

Dimension	Sign of Lyapunov Exponents					Types of Attractors	Types of Motion
3	–	–	–			stable fixed point	equilibrium
	0	–	–			limit cycle	periodic motion
	0	0	–			2-torus	quasiperiodic motion
	+	0	–			strange attractor	chaotic motion
4	–	–	–	–		stable fixed point	equilibrium
	0	–	–	–		limit cycle	periodic motion
	0	0	–	–		2-torus	quasiperiodic motion
	0	0	0	–		3-torus	quasiperiodic motion
	+	0	–	–		strange attractor	chaotic motion
	+	0	0	–		strange attractor on a 3-torus	chaotic motion
	+	+	0	–		strange attractor	chaotic motion
5	–	–	–	–	–	stable fixed point	equilibrium
	0	–	–	–	–	limit cycle	periodic motion
	0	0	–	–	–	2-torus	quasiperiodic motion
	0	0	0	–	–	3-torus	quasiperiodic motion
	0	0	0	0	–	4-torus	quasiperiodic motion
	+	0	–	–	–	strange attractor	chaotic motion
	+	0	0	–	–	strange attractor on a 3-torus	chaotic motion
	+	0	0	0	–	strange attractor on a 4-torus	chaotic motion
	+	+	0	–	–	strange attractor on a 3-torus	chaotic motion
	+	+	0	0	–	strange attractor on a 4-torus	chaotic motion
	+	+	+	0	–	strange attractor	chaotic motion

2.2.3 Power Spectra

The power spectra are a basic tool to analyze random vibrations. In a power spectrum, the square of the Fourier amplitude per unit time is displayed at each frequency. The power spectra also help in distinguishing among periodic, quasiperiodic, and chaotic motions.

For a sample function of a signal $x(t)$, the power spectrum can be defined in two ways. The power spectrum $\Phi_x(\omega)$ is the time average of the square of its Fourier amplitude, namely,

$$\Phi_x(\omega) = \lim_{T\to\infty} \frac{1}{T} \left| \int_0^T x(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t \right|^2 \tag{2.2.5}$$

On the other hand, the power spectrum $\Phi_x(\omega)$ is also the Fourier transform of the autocorrelation function, namely,

$$\Phi_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) \mathrm{e}^{-\mathrm{i}\omega\tau} \mathrm{d}\tau \tag{2.2.6}$$

where the autocorrelation function $R_x(\tau)$ is defined as

$$R_x(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}x(t)x(t+\tau)\,\mathrm{d}t \tag{2.2.7}$$

Based on the Wiener-Khinchin relations of stochastic processes, the above two definitions are equivalent to the condition that $R_x(\tau)$ decays rapidly with time.

In experimental measurements or numerical simulations, researchers often obtain a time series with the same delay interval,

$$x_1, x_2, \cdots, x_N \tag{2.2.8}$$

Adding a periodic condition $x_{N+i}=x_i\,(i=1,2,\cdots)$, the autocorrelation can be calculated as the discrete convolution

$$c_i=\frac{1}{N}\sum_{j=1}^{N}x_j x_{j+i} \tag{2.2.9}$$

Its discrete Fourier transform

$$p_j=\sum_{j=1}^{N}c_i\,\mathrm{e}^{\frac{2\mathrm{i}\pi ki}{N}} \tag{2.2.10}$$

is the discrete power spectrum of the time series (2.2.8).

In practical calculations of discrete power spectra, a more effective approach is to evaluate directly the coefficients of the discrete Fourier transform

$$\begin{aligned} a_j&=\frac{1}{N}\sum_{k=1}^{N}x_k\cos\left(\frac{\pi kj}{N}\right)\\ b_j&=\frac{1}{N}\sum_{k=1}^{N}x_k\sin\left(\frac{\pi kj}{N}\right)\end{aligned} \tag{2.2.11}$$

and then compute

$$\overline{p}_j=a_j^2+b_j^2 \tag{2.2.12}$$

Usually, for many sets of $\{x_i\}$ evaluate the corresponding $\{\overline{p}_j\}$. The average of many resulting $\{\overline{p}_j\}$ will approximate the discrete power spectrum defined by Eq. (2.2.10). In this way, it is unnecessary to calculate the discrete autocorrelation (2.2.9). That is the basic idea of the fast Fourier transform attributed to Cooley and Tukey [15]. Nowadays, there are many commercial software packages available for determining the fast Fourier transform of a given signal.

The spectrum of a periodic motion with period T consists of discrete spikes at the frequency $1/T$ and possibly a certain number of other spikes at m/T for an integer m. The spectrum of a k-period quasiperiodic motion is made up of spikes at

integer multiples of all its frequencies. Theoretically, the spectrum of a quasiperiodic motion can be distinguished from that of a periodic motion, because the peaks of the quasiperiodic spectrum are not spaced at integer multiples of a particular frequency. Practically, due to the impossibility of determining whether a measured value is rational of irrational, a spectrum seeming to be quasi-periodic may actually be periodic with an extremely large period.

Chaos is a random-like motion. The power spectrum of chaos has a continuous, broad-band nature, which is a characteristic exhibited by all chaotic motion. In addition to the broad-band component, it is rather common for a chaotic spectrum to contain spikes indicating the predominant frequencies of the system. In practical simulations, chaotic spectra are much more complicated than regular ones. Typically, they consist of some dominant peaks surrounded by a lot of grass-like components. Although it is uncertain if the grassy portion of the spectrum is truly continuous, the difference in a spectrum between regular and chaotic motion is usually quite striking to provide a feasible means to identify chaos numerically. However, a power spectrum does not distinguish chaos from a truly random motion, which is a limitation of application of power spectra to identification of chaos.

2.3 Melnikov Theory

2.3.1 Introduction

In 1963, Melnikov developed an analytical technique to detect a geometrical structure with the hallmark of chaos [16]. The key issues are the consequence and the prediction of transversal intersection of stable and unstable manifolds. Therefore, the concepts of stable and unstable manifolds will be introduced at the beginning. Then it will be explained that such an intersection implies the resulting sensitive dependence on initial conditions. The Melnikov function will be derived to predict the transversal intersection in a planar integrable system with small periodic perturbations. This analytical prediction approach will be generalized to higher-dimensional systems in the next section. Finally, the relation between the Melnikov analysis and the occurrence of chaos is clarified.

2.3.2 Transversal Homoclinic/Heteroclinic Point

Consider a dynamical system governed by Eq. (2.2.1). The **stable manifold** of a fixed point $\boldsymbol{x}_0$, denoted by $W^{\mathrm{s}}(\boldsymbol{x}_0)$, is the set of all initial conditions such that the trajectories initiated at these points asymptotically approaches the fixed point $\boldsymbol{x}_0$ as $t\rightarrow+\infty$, whereas the **unstable manifold** of a fixed point $\boldsymbol{x}_0$, denoted by $W^{\mathrm{u}}(\boldsymbol{x}_0)$, is the set of all initial conditions such that the trajectories initiated at these points

asymptotically approach the fixed point $\boldsymbol{x}_0$ as $t\to-\infty$. That is,

$$\begin{aligned} W^{\mathrm{s}}(\boldsymbol{x}_0) &= \left\{ \boldsymbol{X}\in \mathrm{R}^n \,\middle|\, \lim_{t\to+\infty} \boldsymbol{x}(\boldsymbol{X},t) = \boldsymbol{x}_0 \right\} \\ W^{\mathrm{u}}(\boldsymbol{x}_0) &= \left\{ \boldsymbol{X}\in \mathrm{R}^n \,\middle|\, \lim_{t\to-\infty} \boldsymbol{x}(\boldsymbol{X},t) = \boldsymbol{x}_0 \right\} \end{aligned} \tag{2.3.1}$$

where $\boldsymbol{X}$ is a point in the state space, and $\boldsymbol{x}(\boldsymbol{X},t)$ is a trajectory stating at $\boldsymbol{X}$. The stable manifold and the unstable manifold share a common feature that a trajectory with a starting point in it remains in the manifold forever. Therefore, both the stable manifold and the unstable manifold are called the **invariant manifold**.

The concept of the invariant manifold of a fixed point can be generalized to that of a periodic orbit. The stable or unstable manifolds of a periodic orbit is the set of all initial conditions which approach the periodic orbit as $t\to+\infty$ or $t\to-\infty$. The invariant manifold of a periodic orbit corresponds to the invariant manifold of a fixed point on the Poincaré map. Consider a fixed point $\bar{\boldsymbol{x}}$ on the Poincaré map $\boldsymbol{P}$

$$\bar{\boldsymbol{x}} = \boldsymbol{P}(\bar{\boldsymbol{x}}) \quad \bar{\boldsymbol{x}}\in \mathrm{R}^{n-1} \tag{2.3.2}$$

Denote the Jacobian matrix of map $\boldsymbol{P}$ as $\mathbf{D}\boldsymbol{P}$. If $n-1$ eigenvalues of $\mathbf{D}\boldsymbol{P}$ are such that their magnitudes are either larger than 1 or smaller than 1, the fixed point is called a **hyperbolic fixed point**, and the corresponding periodic orbit is called a **hyperbolic periodic orbit**. A hyperbolic fixed point is called a **saddle point** (hyperbolic saddle point sometimes) if the magnitudes of some eigenvalues are larger than 1 and those of the rest are smaller than 1. For a saddle point $\boldsymbol{p}_{\mathrm{s}}$, its stable manifold $W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}})$ and its unstable manifold $W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$ are respectively defined as

$$\begin{aligned} W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}}) &= \left\{ \boldsymbol{z} \,\middle|\, \boldsymbol{z}\in \mathrm{R}^{n-1}, \lim_{k\to\infty} \boldsymbol{P}^k(\boldsymbol{z}) = \boldsymbol{p}_{\mathrm{s}} \right\} \\ W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}}) &= \left\{ \boldsymbol{z} \,\middle|\, \boldsymbol{z}\in \mathrm{R}^{n-1}, \lim_{k\to\infty} \boldsymbol{P}^{-k}(\boldsymbol{z}) = \boldsymbol{p}_{\mathrm{s}} \right\} \end{aligned} \tag{2.3.3}$$

where $\boldsymbol{z}$ is a point on the Poincaré section hyperplane.

If the stable manifold $W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}})$ of a saddle point $\boldsymbol{p}_{\mathrm{s}}$ coincides with the unstable manifold $W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$, namely $W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}}) = W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$, the manifold is called a **homoclinic orbit**. The homoclinic orbit is a closed orbit on which all points tend to the same saddle points $\boldsymbol{p}_{\mathrm{s}}$ as $t\to\pm\infty$. For two different saddle points $\boldsymbol{p}_{\mathrm{s}1}$ and $\boldsymbol{p}_{\mathrm{s}2}$, if the stable manifold $W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}1})$ of $\boldsymbol{p}_{\mathrm{s}1}$ coincides with the unstable manifold $W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}2})$ of $\boldsymbol{p}_{\mathrm{s}2}$, the manifold is called a **heteroclinic orbit** on which all points tend to $\boldsymbol{p}_{\mathrm{s}1}$ and $\boldsymbol{p}_{\mathrm{s}2}$ as $t\to+\infty$ and $t\to-\infty$, respectively. If meanwhile the unstable manifold $W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}1})$ of $\boldsymbol{p}_{\mathrm{s}1}$ coincides with the stable manifold $W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}2})$ of $\boldsymbol{p}_{\mathrm{s}2}$ so that there is another heteroclinic orbit, the two heteroclinic orbits form a **heteroclinic cycle**. A heteroclinic cycle may consist of more than two heteroclinic orbits connecting a few saddle points. Fig. 2.1 shows examples of homoclinic orbits, heteroclinic orbits, and heteroclinic cycles on a plane.

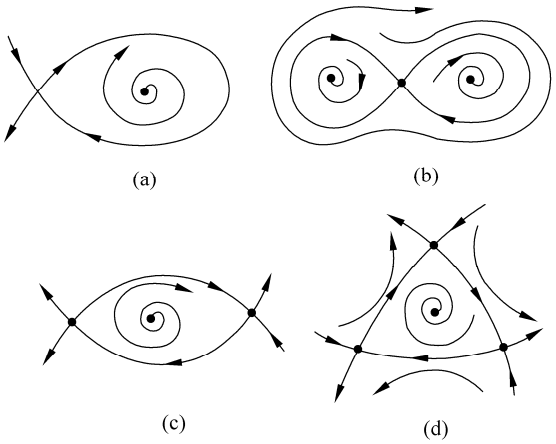

Figure 2.1 Homoclinic orbits, heteroclinic orbits, and heteroclinic cycles (a) a homoclinic orbit (b) two homoclinic orbits (c) heteroclinic cycle consisting of two heteroclinic orbits (d) heteroclinic cycle consisting of three heteroclinic orbits

If the stable manifold and the unstable manifold do not coincide with each other, they may intersect each other. If the stable manifold and the unstable manifold intersect transversally at a point, the point of intersection is called a **transversal homoclinic point**. Here the transversality of an intersection of manifolds means that the union of the tangent spaces of the intersecting manifolds spans the whole space. Intuitively, a transversal intersection means that two intersecting manifolds are not tangent to each other at the point of intersection. If the stable manifold of a saddle point intersects transversally the unstable manifold of another saddle point, the point of intersection is called a **transversal heteroclinic point**.

If there is a transversal homoclinic point $\boldsymbol{q} \in W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}}) \cap W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$. Then $\boldsymbol{q} \in W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}})$ and $\boldsymbol{q} \in W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$. Because both $W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}})$ and $W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$ are invariant manifolds, $\boldsymbol{P}^m(\boldsymbol{q}) \in W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}})$ and $\boldsymbol{P}^m(\boldsymbol{q}) \in W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$ for all integer m. Thus, $\boldsymbol{P}^m(\boldsymbol{q}) \in W^{\mathrm{s}}(\boldsymbol{p}_{\mathrm{s}}) \cap W^{\mathrm{u}}(\boldsymbol{p}_{\mathrm{s}})$. Similar argument is applicable in the case of a transversal heteroclinic point. That is, if the stable manifold and the unstable manifold intersect once, then they will intersect an infinite number of times, as shown in Fig. 2.2.

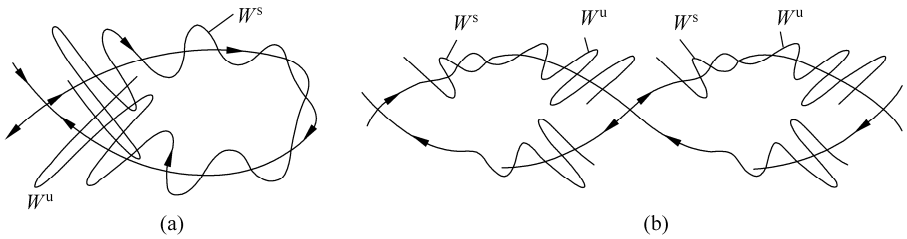

Figure 2.2 Intersections of the stable and unstable manifolds: (a) homoclinic point (b) heteroclinic point

Consider a rectangular area near a homoclinic point in a Poincaré section. Points in the area represent the intersections of trajectories starting at different initial conditions. In the process of mapping, the rectangle moves to the next homoclinic point. Meanwhile, the rectangle contracts in the direction of the stable manifold, stretches in the direction of the unstable manifold, and distorts as it moves. At a later time, the rectangle is transformed into a shape of horseshoe. The horseshoe overlaps the original rectangle to form two new smaller rectangular areas, which are still near the homoclinic point. Therefore, the whole transformation process can be repeated. It can be inferred from Fig. 2.3, two close points in the original rectangle may end up far away from each other and thus the initial difference is amplified. Therefore, such a geometrical structure is highly sensitive to initial conditions, which is a hallmark of chaos. In 1963, Smale proposed the map that contracts, stretches, and folds a rectangle and intersects the image with itself [17]. The map is called the **Smale horseshoe**.

2.3.3 Analytical Prediction

The above-mentioned analysis shows that the occurrence of a transversal homoclinic or heteroclinic point is a possible mechanism resulting in chaos. Melnikov developed an approximate analytical expression of the distance between the stable and unstable manifolds in a planar integrable system with small periodic disturbances. Therefore, a transversal homoclinic or heteroclinic point can be predicted.

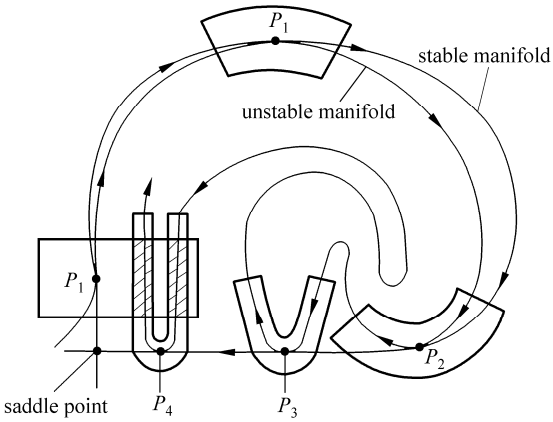

Figure 2.3 Illustration of the Smale horseshoe

Consider a planar nonautonomous system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \varepsilon \boldsymbol{g}(\boldsymbol{x}, t) \qquad \boldsymbol{x} \in \mathrm{R}^2 \tag{2.3.4}$$

where ε is a small parameter and the disturbance $\boldsymbol{g}$ is a periodic function with respect to t. Suppose that when $\varepsilon = 0$, the unperturbed system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathrm{R}^2 \tag{2.3.5}$$

has a saddle point $\boldsymbol{p}_\mathrm{s}$ with a homoclinic orbit $\boldsymbol{x}^\mathrm{h}(t - t_0)$ such that

$$\lim_{t \to \pm\infty} \boldsymbol{x}^\mathrm{h}(t - t_0) = \boldsymbol{p}_\mathrm{s} \tag{2.3.6}$$

in which t_0 is the beginning time that can be an arbitrary real number.

If $\varepsilon \neq 0$ but still sufficiently small, Eq. (2.3.4) exists a unique periodic orbit $\boldsymbol{x}_{\mathrm{s}\varepsilon}(t) = \boldsymbol{p}_0 + \mathrm{O}(\varepsilon)$. Thus its Poincaré map has a unique saddle point $\boldsymbol{p}_{\mathrm{s}\varepsilon} = \boldsymbol{p}_0 + \mathrm{O}(\varepsilon)$. Although the stable and unstable manifolds of $\boldsymbol{p}_{\mathrm{s}\varepsilon}$ no longer coincide, both of them are still sufficiently close to the homoclinic orbit $\boldsymbol{x}^\mathrm{h}(t - t_0)$ for $\varepsilon = 0$. Therefore, the equations of the stable and unstable manifolds can be assumed as

$$\boldsymbol{x}^\mathrm{s}(t, t_0) = \boldsymbol{x}^\mathrm{h}(t - t_0) + \varepsilon \boldsymbol{x}_1^\mathrm{s}(t, t_0) + \mathrm{O}(\varepsilon^2) \tag{2.3.7}$$

$$\boldsymbol{x}^\mathrm{u}(t, t_0) = \boldsymbol{x}^\mathrm{h}(t - t_0) + \varepsilon \boldsymbol{x}_1^\mathrm{u}(t, t_0) + \mathrm{O}(\varepsilon^2) \tag{2.3.8}$$

Equations (2.3.7) and (2.3.8) can also be regarded as the expansions in terms of ε. At time t, the displacement of a point on the stable manifold relative to the point on the unstable manifold is

$$\begin{aligned} \boldsymbol{d}(t, t_0) &= \boldsymbol{x}^\mathrm{s}(t, t_0) - \boldsymbol{x}^\mathrm{u}(t, t_0) \\ &= \varepsilon(\boldsymbol{x}_1^\mathrm{s}(t, t_0) - \boldsymbol{x}_1^\mathrm{u}(t, t_0)) + \mathrm{O}(\varepsilon^2) \end{aligned} \tag{2.3.9}$$

Project $\boldsymbol{d}(t, t_0)$ to the normal $\boldsymbol{N}$ to the a homoclinic orbit of the undisturbed system (2.3.5). Notice that the normal, as shown in Fig. 2.4, is defined by

$$\boldsymbol{N}(t, t_0) = (-f_2(\boldsymbol{x}^\mathrm{h}(t - t_0)), f_1(\boldsymbol{x}^\mathrm{h}(t - t_0))) \tag{2.3.10}$$

where $(f_1, f_2) = \boldsymbol{f}$. One gets

$$d_N(t, \tau) = \boldsymbol{N} \cdot \boldsymbol{d} = \boldsymbol{f} \wedge \boldsymbol{d} = \varepsilon(d_N^\mathrm{s} - d_N^\mathrm{u}) + \mathrm{O}(\varepsilon^2) \tag{2.3.11}$$

where

$$\begin{aligned} d_N^\mathrm{s} &= \boldsymbol{f} \wedge \boldsymbol{x}_1^\mathrm{s} \\ d_N^\mathrm{u} &= \boldsymbol{f} \wedge \boldsymbol{x}_1^\mathrm{u} \end{aligned} \tag{2.3.12}$$

and the wedge product is defined by

$$\boldsymbol{a} \wedge \boldsymbol{b} = a_1 b_2 - a_2 b_1 \tag{2.3.13}$$

for the vectors $\boldsymbol{a} = (a_1, a_2)$ and $\boldsymbol{b} = (b_1, b_2)$. Actually, the wedge product is a vector

cross product in which only the magnitude is taken into consideration.

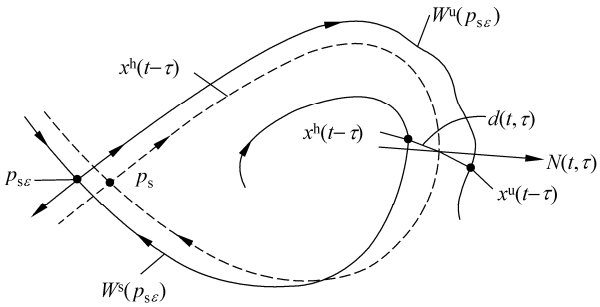

Figure 2.4 Illustration of the derivation of Melnikov's function

Differentiating d_N^s with respect to t yields

$$\dot{d}_N^{s} = \dot{\boldsymbol{f}} \wedge \boldsymbol{x}_1^{s} + \boldsymbol{f} \wedge \dot{\boldsymbol{x}}_1^{s} = \mathbf{D}\boldsymbol{f} \cdot \dot{\boldsymbol{y}}^{h} \wedge \boldsymbol{x}_1^{s} + \boldsymbol{f} \wedge \dot{\boldsymbol{x}}_1^{s} \tag{2.3.14}$$

where the Jacobian $\mathbf{D}\boldsymbol{f}$ is calculated at $\boldsymbol{x}^{h}$. Substituting Eq. (2.3.7) into Eq. (2.3.4) and neglecting ε^2 and higher order terms in the resulting equation lead to

$$\dot{\boldsymbol{x}}_1^{s} = \mathbf{D}\boldsymbol{f} \cdot \boldsymbol{x}_1^{s} + \boldsymbol{g}(\boldsymbol{x}^{h}(t-t_0), t) \tag{2.3.15}$$

Substituting Eqs. (2.3.7) and (2.3.15) into Eq. (2.14) and omitting ε^2 and higher order terms in the resulting equation give

$$\dot{d}_N^{s} = \mathbf{D}\boldsymbol{f} \cdot \boldsymbol{f} \wedge \boldsymbol{x}_1^{s} + \boldsymbol{f} \wedge \mathbf{D}\boldsymbol{f} \cdot \boldsymbol{x}_1^{s} + \boldsymbol{f} \wedge \boldsymbol{g} \tag{2.3.16}$$

Direct computation of the first two terms on the right hand of Eq. (2.3.16) yields

$$\dot{d}_N^{s} = \operatorname{tr}\left(\mathbf{D}\boldsymbol{f}\right) \boldsymbol{f} \wedge \boldsymbol{x}_1^{s} + \boldsymbol{f} \wedge \boldsymbol{g} \tag{2.3.17}$$

Equations (2.3.12) and (2.3.17) mean that

$$\dot{d}_N^{s} = \operatorname{tr}\left(\mathbf{D}\boldsymbol{f}\right) d_N^{s} + \boldsymbol{f} \wedge \boldsymbol{g} \tag{2.3.18}$$

Equation (2.3.18) is a first order linear ordinary differential equation of d_N^s, and it can be integrated from τ to $+\infty$ as

$$d_N^{s}(+\infty, t_0) - d_N^{s}(t_0, t_0) = \int_{\tau}^{+\infty} \boldsymbol{f} \wedge \boldsymbol{g}\, \mathrm{e}^{-\int_0^{t-t_0} \operatorname{tr}(\mathbf{D}\boldsymbol{f}(\boldsymbol{x}^{h}(z)))\,\mathrm{d}z} \mathrm{d}t \tag{2.3.19}$$

Using Eqs. (2.3.12) and (2.3.6) and noticing that $\boldsymbol{p}_s$ is the saddle point, one has

$$d_N^{s}(+\infty, t_0) = \boldsymbol{f}(\boldsymbol{x}^{h}(+\infty - t_0)) \wedge \boldsymbol{x}_1^{s} = \boldsymbol{f}(\boldsymbol{p}_s) \wedge \boldsymbol{x}_1^{s} = 0 \tag{2.3.20}$$

Hence

$$d_N^{\mathrm{s}}(t_0,t_0) = -\int_{\tau}^{+\infty} \boldsymbol{f} \wedge \boldsymbol{g}\, \mathrm{e}^{-\int_0^{t-t_0} \mathrm{tr}(\mathbf{D}\boldsymbol{f}(\boldsymbol{x}^{\mathrm{h}}(z)))\,\mathrm{d}z}\,\mathrm{d}t \tag{2.3.21}$$

A similar procedure yields

$$d_N^{\mathrm{u}}(t_0,t_0) = \int_{-\infty}^{\tau} \boldsymbol{f} \wedge \boldsymbol{g}\, \mathrm{e}^{-\int_0^{t-t_0} \mathrm{tr}(\mathbf{D}\boldsymbol{f}(\boldsymbol{y}^{\mathrm{h}}(z)))\,\mathrm{d}z}\,\mathrm{d}t \tag{2.3.22}$$

Based on Eqs. (2.3.11), (2.3.21), and (2.3.22), if one defines the Melnikov function $\mathcal{M}(\tau)$ as

$$\mathcal{M}(t_0) = \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{x}^{\mathrm{h}}(t-t_0)) \wedge \boldsymbol{g}(\boldsymbol{x}^{\mathrm{h}}(t-t_0),t)\, \mathrm{e}^{-\int_0^{t-t_0} \mathrm{tr}(\mathbf{D}\boldsymbol{f}(\boldsymbol{x}^{\mathrm{h}}(z)))\,\mathrm{d}z}\,\mathrm{d}t \tag{2.3.23}$$

then

$$d_N(t_0,t_0) = -\varepsilon \mathcal{M}(t_0) + \mathrm{O}(\varepsilon^2) \tag{2.3.24}$$

The existence of simple zeros of the Melnikov function $\mathcal{M}(t_0)$ indicates that the displacement $\boldsymbol{d}(t_0, t_0)$ vanishes. At a simple zero t_z, $\mathcal{M}(t_z)\, d_N(t_0,t_0) = 0$, but d $\mathcal{M}(t_z)/\mathrm{d}t_0 \neq 0$. In this case, the stable and unstable manifolds are to intersect transversely to form a transversal homoclinic point.

Equation (2.3.24) can be equivalently written as

$$\mathcal{M}(t_0) = \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{x}^{\mathrm{h}}(t)) \wedge \boldsymbol{g}(\boldsymbol{x}^{\mathrm{h}}(t),t+t_0)\, \mathrm{e}^{-\int_0^{t} \mathrm{tr}(\mathbf{D}\boldsymbol{f}(\boldsymbol{x}^{\mathrm{h}}(z)))\,\mathrm{d}z}\,\mathrm{d}t \tag{2.3.25}$$

If the unperturbed system is Hamiltonian, then $\mathrm{tr}(\mathbf{D}\boldsymbol{f}) = 0$. Equation (2.3.25) becomes

$$\mathcal{M}(t_0) = \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{x}^{\mathrm{h}}(t)) \wedge \boldsymbol{g}(\boldsymbol{x}^{\mathrm{h}}(t),t+t_0)\,\mathrm{d}t \tag{2.3.26}$$

2.3.4 Interruptions

The Melnikov theory is of considerable significance because it can be applied to check in specific systems whether the stable and unstable manifolds intersect transversely or not, by a direct calculation of the approximate distance between these manifolds. The Melnikov function is an explicitly computable function that can be evaluated analytically or numerically. The Melnikov theory has been extended to multi-degree-of-freedom systems [18] and infinite-degree-of-freedom systems [19, 20]. A generalized version for Hamiltonian systems with finite degrees-of-freedom will be presented in the next section, and it will be applied in the subsequent chapters.

It should be kept in mind that the Melnikov theory only predicts transverse intersections of stable and unstable manifolds, or the existence of a homoclinic point. Usually, such an intersection yields an invariant set with sensitivity to initial conditions. Not all invariant sets are an attractor, because they may be without attractability. In dissipative systems, all observable chaos, in laboratory experiments or numerical simulations, should be attractor, with significantly large basin of attraction. The collection of initial conditions under which the motion tends toward a given attractor is called a **basin of attraction**. In a more general sense, the range of values of certain system parameters for which the motion tends toward a prescribed attractor is called a basin of attraction in the parameter space. Therefore, in a practical system, the existence of such an invariant set does not imply that chaotic motion is observed.

In addition, not all attractors sensitive to initial conditions represent chaotic motion, because a periodic motion or even an equilibrium position may depend sensitively on initial conditions when there are two or more attractors in a nonlinear system. In this case, the basin boundary is nonsmooth, intertwined and complicated, actually fractal. The transition from one basin of attraction to another is called a **basin boundary**.

In spite of the limitations, the occurrence of transverse intersections of stable and unstable manifolds is a significant hint to the appearance of chaos. The basin boundary is identical to the stable manifold. A homoclinic point means that the stable and unstable manifolds touch an infinite number of times, which leads to an infinite folding of the stable manifold and hence an infinite folding of the basin boundary and the resulting sensitivity to initial conditions.

2.4 Chaos in Hamiltonian Systems

2.4.1 Hamiltonian Systems, Integrability and KAM Theorem

The Hamiltionian formulation is an effective and powerful approach to model and analyze dynamical problems. For a system with n degrees-of-freedom, its motion can be specified by n generalized coordinates $q_i (i = 1,2,\cdots,n)$ and n generalized momenta p_i. All (q_i, p_i) pairs form a $2n$-dimensional phase space $(\boldsymbol{q}, \boldsymbol{p})$. If all actions on the system are derived from a potential function, W. R. Hamilton (in 1834) proposed the following differential equation of motion

$$\begin{aligned} \dot{q}_i &= \frac{\partial \mathcal{H}}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial \mathcal{H}}{\partial q_i} \end{aligned} \tag{2.4.1}$$

which is called **Hamilton's canonical equations**. In Eq. (2.4.1),

$$\mathcal{H} = \mathcal{H}(\boldsymbol{q}, \boldsymbol{p}, t) \tag{2.4.2}$$

is called the **Hamiltonian**. For a mechanical system, $\mathcal{H}$ is the mechanical energy of the system. If a dynamical system is governed by the canonical equations, it is called a **Hamiltonian system**. Using Eq. (2.4.2), one has

$$\begin{aligned}\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} &= \sum_{i=1}^{n}\left(\frac{\partial \mathcal{H}}{\partial q_i}\dot{q}_i + \frac{\partial \mathcal{H}}{\partial p_i}\dot{p}_i\right) + \frac{\partial \mathcal{H}}{\partial t} \\ &= \sum_{i=1}^{n}\left(\frac{\partial \mathcal{H}}{\partial q_i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial \mathcal{H}}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q_i}\right) + \frac{\partial \mathcal{H}}{\partial t} = \frac{\partial \mathcal{H}}{\partial t}\end{aligned} \tag{2.4.3}$$

Therefore, if the time variable t does not appear in $\mathcal{H} = \mathcal{H}(\boldsymbol{q}, \boldsymbol{p})$ explicitly, then $\mathcal{H}$ is a conserved quantity, a constant during the motion. Such a system is called a **conservative system**. If $\mathcal{H}$ depends on t explicitly, the original $2n$-dimensional phase space can be enlarged into $2(n+1)$-dimensional phase space $(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}})$ by introducing the $n+1$ generalized coordinate $\overline{q}_{n+1} = t$ and generalized momentum $\overline{p}_{n+1} = -\mathcal{H}$. Then the Hamiltonian of the enlarged system is

$$\overline{\mathcal{H}}(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}}) = \mathcal{H}(\boldsymbol{q}, \boldsymbol{p}, q_{n+1}) + p_{n+1} \tag{2.4.4}$$

and the corresponding canonical equations are

$$\begin{aligned}\dot{\overline{q}}_i &= \frac{\partial \overline{\mathcal{H}}}{\partial \overline{p}_i} \\ \dot{\overline{p}}_i &= -\frac{\partial \overline{\mathcal{H}}}{\partial \overline{q}_i}\end{aligned} \quad (i = 1, 2, \cdots, n+1) \tag{2.4.5}$$

The construction of the enlarged system means the equivalence of Eqs. (2.4.1) and (2.4.5), while the enlarged system is conservative as its Hamiltonian is explicitly independent of t. Thus all Hamiltonian systems can be equivalently transformed into a conservative one. In the following, only conservative systems will be considered.

A Hamiltonian system can be substantively simplified via an appropriate transform from one set of variables $(\boldsymbol{q}, \boldsymbol{p})$ to some new set $(\boldsymbol{Q}, \boldsymbol{P})$,

$$\begin{aligned}\boldsymbol{Q} &= \boldsymbol{Q}(\boldsymbol{q}, \boldsymbol{p}) \\ \boldsymbol{P} &= \boldsymbol{P}(\boldsymbol{q}, \boldsymbol{p})\end{aligned} \tag{2.4.6}$$

Equation (2.4.1) is accordingly changed into the form

$$\begin{aligned} \dot{\boldsymbol{Q}} &= \dot{\boldsymbol{Q}}(\boldsymbol{Q}, \boldsymbol{P}) \\ \dot{\boldsymbol{P}} &= \dot{\boldsymbol{P}}(\boldsymbol{Q}, \boldsymbol{P}) \end{aligned} \tag{2.4.7}$$

If the structure of Eq. (2.4.1) is still preserved, that is, Eq. (2.4.7) takes the form

$$\begin{aligned} \dot{Q}_i &= \frac{\partial h}{\partial P_i} \\ \dot{P}_i &= -\frac{\partial h}{\partial Q_i} \end{aligned} \qquad (i = 1, 2, \cdots, n) \tag{2.4.8}$$

where $h = h(\boldsymbol{Q}, \boldsymbol{P})$, then the transformation is called a **canonical transformation**. A canonical transformation changes a set of canonical equations into another set of canonical equations. It can be proved that the inverse of a canonical transformation and the composition of canonical transformations are still canonical transformations.

Suppose there are a series of canonical transformations to change variables $(\boldsymbol{q}, \boldsymbol{p})$ into variables $(\boldsymbol{I}, \boldsymbol{\theta})$, such that the Hamiltonian in the new variables depends only on $\boldsymbol{I}$, and that

$$\mathcal{H} = H(\boldsymbol{I}) \tag{2.4.9}$$

is independent of $\boldsymbol{\theta}$. Then the Hamiltonian equations are

$$\begin{aligned} \dot{I}_i &= -\frac{\partial H}{\partial \theta_i} = 0 \\ \dot{\theta}_i &= \frac{\partial H}{\partial I_i} = \Omega_i(I_1, I_2, \cdots, I_n) \end{aligned} \qquad (i = 1, 2, \cdots, n) \tag{2.4.10}$$

Integration of Eq. (2.4.10) yields

$$\begin{aligned} I_i(t) &= I_i(0) \\ \theta_i(t) &= \Omega_i(I_1(0), I_2(0), \cdots, I_n(0))t + \theta_i(0) \end{aligned} \tag{2.4.11}$$

where $2n$ constants $\boldsymbol{I}(0)$ and $\boldsymbol{\theta}(0)$ can be determined by initial condition $(\boldsymbol{q}(0), \boldsymbol{p}(0))$. Variables $(\boldsymbol{I}, \boldsymbol{\theta})$ are called **action-angle variables**. According to Eq. (2.4.11), the motion of Eq. (2.4.10) can be uniquely specified by n angle variables θ_i. Mathematically, an n-dimensional manifold where the point is specified by n angles is called an ***n*-torus**, denoted as T_n. An 1-torus T_1 is a circle, and 2-torus T_2 is a usual tori, while there is no plot of n-torus T_n for $n \geqslant 3$ in 3-dimensional physical space. Actually, I_i is the n radii of the n-torus. As explained in 2.1.2, if Ω_i are incommensurable, that is, there do not exist not all zero integer k_i such that $\sum_{i=1}^{n} k_i \Omega_i = 0$, then it can be proved that the trajectory winds up on the torus endlessly without closing, and is dense there. The corresponding motion is

quasiperiodic. If $\boldsymbol{\Omega}_i$ is commensurable, then the trajectory closes on the torus, and the motion is periodic.

Hamiltonian systems expressed in action-angle variables are integrable. Generally, a Hamiltonian with n degree-of-freedom is an **integrable system**, if there exist n independent isolating integrals of motion

$$I_i(\boldsymbol{q},\boldsymbol{p}) = C_i \quad (i = 1,2,\cdots,n) \tag{2.4.12}$$

where C_i are constants. Functions I_i are independent if the differentials $\mathrm{d}I_i$ are linearly independent. Since n isolating integrals of motion exist, the $2n$ dimensional phase space is confined in an n-dimensional manifold that is homeomorphic to n-tori. The manifold is an invariant torus, because all trajectories starting on it remain there all the time. Hence integrable systems can never be chaotic, and their motion is periodic or quasiperiodic.

In 1892, Poincaré proved that many dynamical systems, including the three-body problem, are not integrable. A system that has fewer constants of motion than degrees-of-freedom is called a **nonintegrable system**. Integrability is an exceptional property for Hamiltonian systems with degrees-of-freedom larger than 2. In fact, integrable systems are so rare that in general it is impossible to approximate a nonintegrable Hamiltonian system by a series of integrable ones. However, there is no direct criterion to determine the integrability. An integrable system is made slightly nonintegrable by adding a small disturbance. Such a system is called **near integrable system**. In terms of action-angle variables, the Hamiltonian of a near integrable system can be written as

$$\mathcal{H}(\boldsymbol{I},\theta) = \mathcal{H}_0(\boldsymbol{I}) + V(\boldsymbol{I},\theta) \tag{2.4.13}$$

where $\mathcal{H}_0$ is integrable and V is sufficiently small. If no disturbance is present, then $V = 0$ and the system is integrable. If $V \neq 0$, the integrability is usually violated.

In 1954, Kolmogorov described the qualitative picture of near integrable systems [21]. Arnol'd and Moser completely proved the conclusion that is known as famous Kolmogorov-Arnol'd-Moser theorem. In the following, KAM theorem is presented without proof, which is outside the scope of this monograph.

KAM theorem: Suppose that Hamiltonian (2.4.13) satisfies the following conditions:

(i) $\mathcal{H}(\boldsymbol{I},\boldsymbol{\theta})$ is a real analytic function on a region Σ_0: $|\mathrm{Im}\boldsymbol{\theta}| \leqslant t$, $|\boldsymbol{I}-\boldsymbol{I}_0| \leqslant s$;

(ii) $\boldsymbol{\Omega}_j = \dfrac{\partial \mathcal{H}_0}{\partial I_j}$ $(j = 1,2,\cdots,n)$ calculated at I_0 such that $\left|\dfrac{\partial \boldsymbol{\Omega}_j}{\partial I_k}\right| \neq 0$ (nondegeneracy conditions);

(iii) For arbitrary integer vector $\boldsymbol{k} = (k_1, k_2, \cdots, k_n)$ there exist $C(\boldsymbol{\Omega})>0$ and $\mu > n-1$ such that the nonresonance condition holds as follows

$$\left|\sum_{j=1}^{n} k_j \Omega_j\right| \geqslant C\left(\sum_{j=1}^{n}\left|k_j\right|\right)^{-\mu} \tag{2.4.14}$$

Then for any $\varepsilon>0$, there is a $\delta=\delta(\varepsilon, C, \mu, s, t)$ such that, if $|V|<\delta$ in Σ_0, the solution to equation

$$\begin{aligned} \dot{\theta}_i &= \frac{\partial \mathcal{H}}{\partial I_i} \\ \dot{I}_i &= -\frac{\partial \mathcal{H}}{\partial \theta_i} \end{aligned} \quad (i=1,2,\cdots,n) \tag{2.4.15}$$

lies on an n-dimensional invariant torus

$$\begin{aligned} \boldsymbol{I} &= \boldsymbol{I}_0 + \Gamma(\Theta) \\ \theta &= \Theta + \Phi(\Theta) \end{aligned} \tag{2.4.16}$$

where Γ and Φ are real analytic functions with period 2π defined on $|\mathrm{Im}\Theta| \leqslant t/2$. The trajectory on the torus is governed by

$$\begin{aligned} I_i(t) &= I_i(0) \\ \Theta_i(t) &= \left.\frac{\partial \mathcal{H}}{\partial I_i}\right|_{I_j(t)=I_j(0)} t + \Theta_i(0) \end{aligned} \tag{2.4.17}$$

The torus is sufficiently close to the torus of the undisturbed system, that is,

$$|\Gamma|+|\Theta|<\varepsilon \tag{2.4.18}$$

The conditions in the KAM theorem require the disturbance leading to the nonintegrability to be sufficiently small, the Hamiltonian to be analytic function, the system to be nondegenerate, and the undisturbed frequencies to be nonresonant, in which the analytical condition and the nondegenerate condition can be technically weakened. Under these conditions, most nonresonant tori survive but may be slightly deformed. Hence the tori exist in the phase space of the disturbed system, and the trajectories wind densely on them. The number of independent frequencies is equal to the number of degrees-of-freedom of the system. Those tori are called **KAM tori**, **KAM surfaces** or **KAM curves**.

2.4.2 Stochastic Layers and Global Chaos

Now consider the resonance of a near integrable Hamiltonian system. In this case, Ω_i is commensurable. The trajectories of Eq. (2.4.1) are periodic orbits on the n-dimensional invariant torus T_n. Since a conservative Hamiltonian always has

its energy as its constant of motion, its motion in $2n$-dimensional phase space is confined to a $(2n-1)$-dimensional energy surface after given an initial energy. The intersection of the energy surface and the n-dimensional invariant tori yields a $(2n-2)$-dimensional surface Σ that can serve as a cross section to define a Poincaré map $\boldsymbol{P}_0$. The surface Σ cuts the torus T_n on a level curve Γ. Due to the periodicity of the motion, every point of curve Γ is a k-periodic point of $\boldsymbol{P}_0$ for some integer k. A k-periodic point of a map f is defined as the fixed point of f^k but not the fixed point of f^m for any $m<k$. For the disturbed system (2.4.13), the same surface Σ still defines a Poincaré map $\boldsymbol{P}$, and the change of the level curve Γ reflects deformation of the invariant torus T_n.

Before the discussion of Poincaré map $\boldsymbol{P}$, some basic concepts of Hamiltonian maps need to be presented. A **Hamiltonian map** is a map that conserves volume in the phase space. Hence the determinant of its Jacobian is equal to 1. The Poincaré map of a Hamiltonian system is a Hamiltonian map. Consider a 2-dimensional Hamiltonian map $\boldsymbol{M}$, which can serve as a Poincaré map of a Hamiltonian system with 2 degree-of-freedom. The Hamiltonian system has a 4-dimensional phase space, a 3-demensional energy surface, a 2-dimensional Poincaré section Σ, a 2-dimensional invariant torus T_2, and a 1-dimensional level curve Γ. Suppose $\boldsymbol{z}_0$ to be a k-periodic point $\boldsymbol{M}$. Then the Jacobian of $\boldsymbol{M}^k$ calculated at $\boldsymbol{z}_0$ with two eigenvalues λ_1 and λ_2 satisfying $\lambda_1\lambda_2=1$ because $\det \mathrm{D}\boldsymbol{M}^k=1$. Therefore, λ_1 and λ_2 are two real numbers with $0<\lambda_1<1<\lambda_2$ or a pair of complex conjugates with the unit modulus. The periodic point is defined as a hyperbolic point in the first case and an elliptic point in the second case. This conclusion is true for general cases. A periodic point of a Hamiltonian map must be either a hyperbolic point or an elliptic point.

In 1935, based on Poincaré's previous work in 1899, Birkhoff proved the following conclusion: For a sufficiently small disturbance, the level curve Γ breaks up into $2mk$ k-periodic points of Poincaré's map $\boldsymbol{P}$ for some integer m; these periodic points lie near Γ; mk points are hyperbolic and mk points are elliptic. This conclusion is referred as the **Poincare-Birkhoff theorem**. However, the theorem does not specify the value of the integer m.

In the situation described by the Poincaré-Birkhoff theorem, Γ is called a **resonant level curve**, and a certain region around Γ containing the hyperbolic and elliptic points is called a **resonance zone**. Around each elliptic point there is a series of periodic orbits. Any two adjacent hyperbolic points are connected by heteroclinic orbits. If the heteroclinic orbits intersect transversely in a homoclinic point, then, according to the analysis presented in 2.3.2, the transverse heteroclinic point results in an infinitely complicated set of intersections, which are a cause of chaotic behavior. Such a complicated geometrical structure in a Hamiltonian system is called a **stochastic layer**. The regions around the elliptic points bounded by the heteroclinic orbits are called **islands**. Those islands compose an **island chain** if there are several elliptic points. For the situation $mk=3$, Fig. 2.5 shows the breakup of a resonant level curve, and Fig. 2.6 illustrates a stochastic layer.

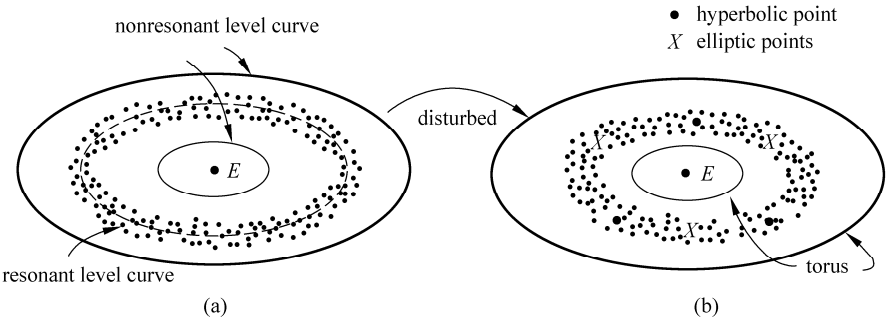

Figure 2.5 A resonant level curve breakup into hyperbolic points (•) and elliptic points (X)

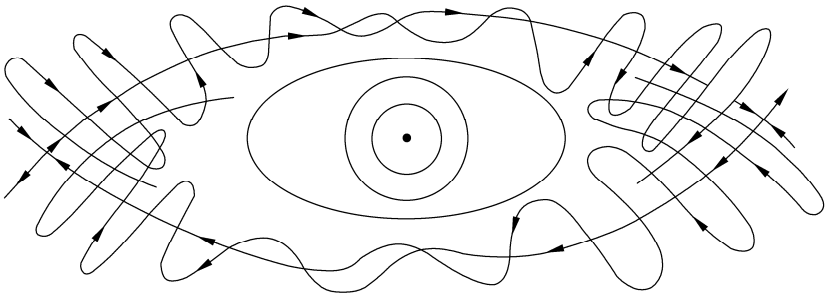

Figure 2.6 A stochastic layer in a Hamiltonian system

The newly formed elliptic points due to the breakup of a resonant level curve are surrounded by smaller level curves. In the resonance, according to the Poincaré-Birkhoff theorem, those curves become a chain of elliptic and hyperbolic points around the earlier elliptic points. This self-similar pattern can repeated infinitely, as depicted in Fig. 2.7, while most of these points on such a small scale that it is difficult to locate them in numerical simulations. Meanwhile, there are nonresonant

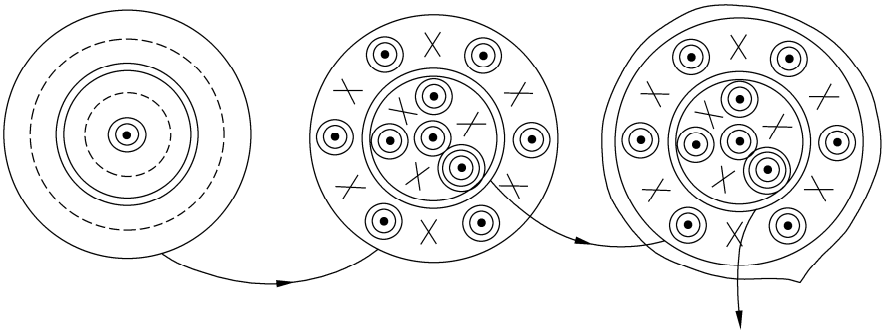

Figure 2.7 Self-similarity near an elliptic point

level curves corresponding to KAM tori, which are preserved, near each elliptic point. Regular behaviors represented by nonresonant level curves coexist with chaos represented by stochastic layers in resonance zones. Thus, in Hamiltonian systems, initial conditions are so crucial that some sets of initial conditions lead to regular motion while others lead to chaos for the same set of system parameters. Therefore there is a complex nested structure of KAM tori surrounded by chains of elliptic and hyperbolic points.

Stochastic layers exist in all nonintegrable Hamiltonian systems. However, for an integrable Hamiltonian system with very small disturbance, the stochastic layers may be so slight that they cannot be found in numerical calculations. Hence only regular motion occurs in the system. With the increase of the disturbance, the system exhibits chaotic behavior manifested in the emergence of observable stochastic layers. According to the KAM theorem, there still exist nonresonant KAM tori that divide the stochastic layers. Irregular motion due to the stochastic layers separated by the KAM tori is called **local chaos**. With further increase of the disturbance, the KAM tori separating the adjacent stochastic layers successively break up and the stochastic layers merge into larger stochastic layers. Thus the thickness of stochastic layers expands with the disturbance. For a sufficiently large disturbance, the resonance zones may overlap so that there is a transverse intersection of stable and unstable manifolds for two hyperbolic points from two different resonance zones. In this case, stochastic layers are no longer confined by KAM tori, and the corresponding behavior is called **global chaos**. In global chaos, there may still exist KAM tori not destroyed by the disturbance. Those tori resemble islands in a chaotic ocean, while there are smaller stochastic layers and KAM tori on the islands. Hence global chaos is a very complicated self-similar structure in the phase space.

2.4.3 Arnol'd Diffusion

The motions of integrable Hamiltonian (2.4.9) are confined in an n-dimensional torus (2.4.11) in a $2n$-dimensional phase space. For a nonintegrable Hamiltonian system, stochastic layers appear near the intersection of $(2n-1)$-dimensional energy surface and $(2n-1)$-dimensional **resonance surface** defined by

$$\sum_{i=1}^{n} k_i \boldsymbol{\Omega}_i(\boldsymbol{J}_0) = 0 \tag{2.4.19}$$

where k_i are integers not all equal to zero. The energy surface and the resonance surface intersect on a surface with the dimension $(2n-1)-1=2n-2$. The stochastic layers around the surface in a space with dimension $(2n-2)+1=2n-1$. Diverse curves form when the resonance surface cuts the energy surface, and these curves

interconnect each other to constitute a complex network spreading all over the energy surface. Such a network is called an **Arnol'd web**.

In an m-dimensional space, an $(m-1)$-dimensional closed surface, such as T_{m-1}, can divide the space into two distinct parts, while the closed surface of less than $m-1$ dimension cannot do so. In a Hamiltonian system with n degrees-of-freedom, KAM tori is n-dimensional. Only if $(2n-1)-1=n$ i.e. $n=2$, a KAM torus divides the $(2n-1)$-dimensional energy surface into two disconnected parts, and thus it can isolate stochastic layers. For $n \geqslant 3$, trajectories in gaps between the tori can escape to other regions of the energy surface. Therefore, all stochastic layers on the energy surface are connected into a single complex network, which is the above-mentioned Arnol'd web. The web permeates the entire energy surface, intersecting or lying infinitesimally close to every point. For an initial condition within the web, the subsequent trajectory will eventually intersect every finite region of the energy surface. Such an irregular motion in a higher degree-of-freedom Hamiltonian system is called the **Arnol'd diffusion**. In 1964, Arnol'd proved that stochastic layers merge into a single web in a specific nonlinear Hamiltonian system [22].

The structures of Arnol'd web depends on the energy surface and the resonance surface, which are both dependent on the integrable Hamiltonian and independent of the disturbance. Thus there are global Arnol'd diffusions for arbitrary small disturbances. In the case when the disturbances are large enough to yield observable stochastic layers, the Arnol'd diffusion links together the chaotic regions on all scales. However, the Arnol'd diffusion is usually very slow. In 1977, N. N. Nehoroshev proved a rigorous but overestimated upper bound on the diffusion rate. For an integrable Hamiltonian system with a disturbance of the order ε, the change of system momentum satisfies

$$\left\| \boldsymbol{p}(t) - \boldsymbol{p}(0) \right\| < \varepsilon^a \qquad t \in \left[0, \frac{1}{\varepsilon} \mathrm{e}^{\varepsilon^{-b}} \right] \tag{2.4.20}$$

where a and b are positive constants determined by the undisturbed integrable Hamiltonian system [23].

2.4.4 Higher-Dimensional Version of Melnikov Theory

The idea of the Melnikov theory can be generalized to higher-dimensional system to develop quantitative methods for handling the Poincare-Birkhoff breakup of resonance zones. In the following, a version [24] proposed by Holmes and Marsden in 1983 is presented without proof. This version will be applied to treat a gyrostat with a rotor in the following chapter.

Consider an $n+1$ degrees-of-freedom integrable Hamiltonian system with small

disturbance

$$\begin{aligned}\mathcal{H}(\boldsymbol{\mu},\theta_1,\cdots,\theta_n,I_1,\cdots,I_n;\varepsilon) &= F(\boldsymbol{\mu})+\sum_{i=1}^{n}G_i(I_i)+\varepsilon\mathcal{H}_1(q,p,\theta_1,\cdots,\theta_n,I_1,\cdots,I_n)\\ &= \mathcal{H}_0(\boldsymbol{\mu},I_1,\cdots,I_n)+\varepsilon\mathcal{H}_1(q,p,\theta_1,\cdots,\theta_n,I_1,\cdots,I_n)\end{aligned} \tag{2.4.21}$$

where μ are a set of m Lie-Poisson variables, ($\theta_1,\cdots,\theta_n$, $I_1,\cdots,I_n$) are action-angle coordinates ($n\geqslant 2$), ε is a small parameter, and $\mathcal{H}_1$ is 2π-periodic in $\theta_1,\cdots,\theta_n$.

For integrable Hamiltonian system with $\varepsilon=0$, $\boldsymbol{\mu}$ is decoupled from action coordinates ($I_1,\cdots,I_n$). Suppose that F contains a homoclinic (or heteroclinic) orbit $\boldsymbol{\mu}^{\mathrm{h}}$ with energy h_0. The coadjoint orbit containing $\boldsymbol{\mu}^{\mathrm{h}}$ is assumed to be two-dimensional. The saddle points for $\boldsymbol{\mu}^{\mathrm{h}}$ are denoted $\boldsymbol{\mu}_{\pm}$, which may be coincident. Suppose for $j=1,\cdots,n$, $\Omega_j(I_j)=G_j'(I_j)>0$. For a given energy

$$\mathcal{H}(\boldsymbol{\mu},\theta_1,\cdots,\theta_n,I_1,\cdots,I_n;\varepsilon)=h \tag{2.4.22}$$

System (2.4.21) has a reduced integrable part with the Hamiltonian

$$L_0(\boldsymbol{\mu},I_1,\cdots,I_{n-1},h)=G_n^{-1}\left(h-F(\boldsymbol{\mu})-\sum_{j=1}^{n-1}G_j(I_j)\right) \tag{2.4.23}$$

The Hamiltonian system L_0 has two ($n-1$)-parameter family of invariant ($n-1$)-dimensional tori $T_{\pm}(h_1,\cdots,h_{n-1})$ defined by

$$\boldsymbol{\mu}=\boldsymbol{\mu}_{\pm},\ G_j(I_j)=h_j,\ \theta_j=\Omega(G_j^{-1}(h_j))\theta_n+\theta_j(0)\pmod{2\pi}\quad(j=1,\cdots,n-1) \tag{2.4.24}$$

where h_j is a constant. Correspondingly, the system for H_0 has two n-parameter family of invariant tori $T_{\pm}(h_1,\cdots,h_n)$. Henceforth the phase constants of integration $\theta_j(0)$ is written as θ_{0j} for $j=1,\cdots,n$. The tori $T_{\pm}(h_1,\cdots,h_{n-1})$ are connected by the n-dimensional homoclinic manifold defined by

$$\boldsymbol{\mu}=\boldsymbol{\mu}^{\mathrm{h}}(\theta_n+\theta_{0n}),\ G_j(I_j)=h_j,\ \theta_j=\Omega(G_j^{-1}(h_j))\theta_n+\theta_{0j}\pmod{2\pi}\quad(j=1,\cdots,n-1) \tag{2.4.25}$$

where the phase constant θ_{0n} associated with the reduced degrees-of-freedom appears explicitly. This manifold consists of the coincident stable and unstable manifolds of the tori $T_{\pm}(h_1,\cdots,h_{n-1})$; i.e. $W^{\mathrm{s}}(T_{\pm}(h_1,\cdots,h_{n-1}))=W^{\mathrm{u}}(T_{\pm}(h_1,\cdots,h_{n-1}))$ given by Eq. (2.4.25).

For $\varepsilon\neq 0$, a system defined by Hamiltonian (2.4.21) possesses a Poincaré map $\boldsymbol{P}_{\varepsilon}$ from a piece of ($\boldsymbol{\mu},\theta_1,\cdots\theta_{n-1}$, $I_1,\cdots,I_{n-1}$) space to itself where θ_n goes through an increment of 2π, starting at some fixed value θ_{n0}. Assume that the constants

$G_i'(I_i) = h_i (i = 1,\cdots,n)$ are chosen so that the disturbed frequencies $\Omega_i(I_i)$ satisfy the nondegeneracy conditions $\Omega_i'(I_i) \neq 0$ and the nonresonance condition (2.4.14) of the KAM theorem. These conditions ensures that the tori $T_{\pm}(h_1,\cdots,h_{n-1})$ perturb to invariant tori $T_{\varepsilon\pm}(h_1,\cdots,h_{n-1})$ for $\boldsymbol{P}_\varepsilon$ with sufficiently small ε. Let $h = h_0+h_1+\cdots+h_n$ where $h_i>0$ $(i = 1,\cdots,n)$ and the undisturbed homoclinic manifold be filled with an n-parameter family of orbits given by

$$(\boldsymbol{\mu},\theta_1,\cdots,\theta_n,I_1,\cdots,I_n) = (\boldsymbol{\mu}^{\mathrm{h}}(t),\Omega_1(I_1)t+\theta_{01},\cdots,\Omega_n(I_n)t+\theta_{0n},I_1,\cdots,I_n) \tag{2.4.26}$$

Pick one such orbit and let $\{\{F, \mathcal{H}_1\}\}$ denote the Lie-Poisson bracket of $F(\boldsymbol{\mu})$ and $\mathcal{H}_1(\boldsymbol{\mu},\theta_1,\cdots\theta_n, I_1,\cdots, I_n)$ evaluated on this orbit. Similarly, let $\{I_j, H_1\} = -\partial H_1/\partial\theta_j$ $(j = 1,\cdots,n-1)$ be evaluated on this orbit. Define the **Melnikov vector**

$$\begin{aligned}\boldsymbol{\mathcal{M}}(\theta_0) = (&\mathcal{M}_1(\theta_{10},\cdots,\theta_{n0},h,h_1,\cdots,h_{n-1}),\cdots,\mathcal{M}_{n-1}(\theta_{10},\cdots,\theta_{n0},h,h_1,\cdots,h_{n-1}),\\ &\mathcal{M}_n(\theta_{10},\cdots,\theta_{n0},h,h_1,\cdots,h_{n-1}))\end{aligned} \tag{2.4.27}$$

by

$$\begin{aligned}\mathcal{M}_j(\theta_{10},\cdots,\theta_{n0},h,h_1,\cdots,h_{n-1}) &= \int_{-\infty}^{+\infty}\{I_j,\mathcal{H}_1\}\,\mathrm{d}t \quad (j = 1,\cdots,n-1)\\ \mathcal{M}_n(\theta_{10},\cdots,\theta_{n0},h,h_1,\cdots,h_{n-1}) &= \frac{1}{\Omega_n}\int_{-\infty}^{+\infty}\{F,\mathcal{H}_1\}\,\mathrm{d}t\end{aligned} \tag{2.4.28}$$

Assume that the multiply 2π-periodic Melnikov vector $\boldsymbol{\mathcal{M}}$ has at least one simple zero; i.e. there is a point $\boldsymbol{\theta}_0 = (\theta_{10},\cdots\theta_{n0})$ for which

$$\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}_0) = 0,\ \det(\mathbf{D}\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}_0)) \neq 0 \tag{2.4.29}$$

where $\mathbf{D}\boldsymbol{\mathcal{M}}$ is the $n\times n$ matrix of partial derivatives of $\mathcal{M}_1,\cdots, \mathcal{M}_n$ with respect to $\theta_{10},\cdots\theta_{n0}$, the initial phases of the orbit. Under these assumptions, Holmes and Marsden demonstrated that, for sufficiently small ε, the disturbed stable and unstable manifolds $W^{\mathrm{s}}(T_{\varepsilon\pm})$ and $W^{\mathrm{u}}(T_{\varepsilon\pm})$ of the disturbed tori $T_{\varepsilon\pm}$ intersect transversely.

References

[1] Moon FC. *Chaotic and Fractal Dynamics*: *An Introduction for Applied Scientists and Engineers*. New York: John Wiley & Sons, 1992

[2] Nayfeh AH, Balachandran B. *Applied Nonlinear Dynamics: Analytical, Computational, and Experiment Methods*. New York: John Wiley & Sons, 1995

[3] Thompson JMT, Stewart HB. *Nonlinear Dynamics and Chaos*: *Geometrical Methods for Engineers and Scientists* (2nd edn.). New York: John Wiley & Sons, 2002

[4] Ott E. *Chaos in Dynamical Systems* (2nd edn.). Cambridge: Cambridge University Press, 2002

[5] Wiggins S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (2nd edn.). Berlin: Springer-Verlag, 2003

[6] Chen LQ, Liu YZ. *Nonlinear Dynamics* (2nd edn.). Beijing: Higher Education Press, 2008 (in Chinese)

[7] Feigenbaum MJ. Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics*, 1978, 19, 25-52

[8] Pomeau Y, Manneville P. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 1980, 74, 189-197

[9] Ruelle D, Takens F. On the nature of turbulence. *Communications in Mathematical Physics*, 1971, 20, 167-192

[10] Newhouse SE, Ruelle D, Takens F. Occurrence of strange axiom A attractors near quasiperiodic flows on T^m, $m \geqslant 3$. *Communications in Mathematical Physics*, 1978, 64, 35-40

[11] Swinney HL, Gollub JP. The transition to turbulence. *Physics Today*, 1978, 31, 41-49

[12] Feigenbaum MJ, Kadanoff LP, Shenker SJ. Quasiperiodicity in dissipative systems: a renormalization group analysis. *Physics D*, 1982, 5, 370-386

[13] Grebogi C, Ott E, Yorke JA. Are three-frequency quasiperiodic orbits to be expected in typical dynamical systems? *Physical Review Letters*, 1983, 51, 339-342

[14] Grebogi C, Ott E, Yorke JA. Crises, sudden changes in chaotic attractors and transient chaos. *Physics D*, 1983, 7, 181-200

[15] Cooley TW, Tukey JW. An algorithm for the machine calculations of complex Fourier series. *Mathematics of Computation*, 1965, 19, 297-301

[16] Melnikov VK. On the stability of the center for time periodic perturbations. *Transactions of Moscow Mathematical Society*, 1963, 12, 1-57

[17] Smale S. *The Mathematics of Time: Essays on Dynamical Systems, Economic Processes, and Related Topics*. Berlin: Springer-Verlag, 1980

[18] Wiggins S. *Global Bifurcations and Chaos: Analytical Methods*. Berlin: Springer-Verlag, 1988

[19] Mclaughlin D W, Shatah J. Melnikov analysis for Pde's. *Lectures in Applied Mathematics*, 1996, 31, 51-100

[20] Guo BL, Gao P, Chen HL. *Near Integrable Infinite-Dimensional Systems*. Beijing: Defense Industry Press, 2004 (in Chinese)

[21] Kolmogorov AN. Preservation of conditionally periodic movements with small change in the Hamiltonian function. *Dokl. Akad. Nauk SSSR*, 1954, 98, 525-530 (in Russian)

[22] Arnol'd VI. Instability of dynamical systems with several degrees of freedom. *Soviet Mathematics*, 1964, 5, 581-585

[23] Nekhoroshev NN. An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. *Russian Mathematical Surveys*, 1977, 32(6), 1-65

[24] Holmes PJ, Marsden JE. Horseshoes and Arnold diffusion for Hamiltonian systems on Lie groups. *Indiana University Mathematics Journal*, 1983, 32, 273-310