1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Part D

Dynamical Systems

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

16

Nonlinear Dynamics: a Brief Introduction

Alessandro Moura and Celso Grebogi

Institute for Complex Systems and Mathematical Biology, University of Aberdeen, UK

16.1   Introduction

Biological systems are governed by complex interactions among their constituent parts, and these interactions result in nonlinear dynamics. This is attested by the ubiquity of feedback and feedforward mechanisms in biology, in scales ranging from whole multicellular organisms to single-cell dynamics. The nonlinearity of biological interactions has a number of crucially important consequences for the dynamics of biological systems, including the emergence of multiple stability and limit cycle behaviour. This chapter is dedicated to another characteristic class of phenomena that can appear in nonlinear dynamical systems: chaos.

Chaotic dynamics is common in nonlinear systems, and is characterised by irregular and unpredictable behaviour of trajectories in phase space, which are impossible to predict for long times [1–5]. Despite this irregularity, chaotic dynamics is governed by a few simple principles which have deep connections with concepts in statistical physics and information theory. In this short introduction to chaos in nonlinear dynamics, our focus will be on these aspects of the theory, which help explain why chaos theory is applicable to such a broad range of phenomena. We will avoid technicalities and try to convey what we consider to be the key ideas in the area, without attempting to be mathematically rigorous, and we will emphasise especially the connection between all these different ideas, and how they all are ultimately a consequence of the sensitivity of the dynamics to initial conditions which is the hallmark of chaos. We direct the readers interested in more in-depth coverage of the topics presented here to references [1–5].

We start by defining precisely the single most characteristic feature of chaos – the sensitivity of trajectories to initial conditions, as measured by the Lyapunov exponent (Section 16.2). We then see how the combination of the exponential separation of close trajectories with bounded motion leads to a dynamics which can be characterised by probabilities, resembling statistical physics (Section 16.3). This leads to the definition of entropy for a chaotic system, which allows us to quantify the information generated by a chaotic system (Section 16.4). Finally, we see how the description of the dynamics of chaotic systems can be drastically

simplified without any loss of information, by means of symbolic dynamics, and we derive some important results on the complex structure of orbits in chaotic systems (Section 16.5).

16.2   Sensitivity to initial conditions and the Lyapunov exponent

Chaotic systems are characterised by a *sensitivity to initial conditions*, which means that very close initial conditions lead to trajectories which quickly separate, and end up going into completely distinct motions. The separation takes place exponentially fast in chaotic systems. More precisely, for a pair of initial conditions that are separated by a distance δ_0 defined in an appropriate metric in the phase space, the distance $\delta(t)$ at future times t increases as $\delta(t) \sim \delta_0 e^{\lambda t}$ for sufficiently large t, where λ is the Lyapunov exponent, first introduced by Oseledec [6]:

$$\lambda = \lim_{\delta_0 \to 0, t \to \infty} \frac{1}{t} \ln \left(\frac{\delta(t)}{\delta_0} \right). \tag{16.1}$$

Now let us interpret δ_0 as our limit of resolution in distinguishing initial conditions – for example, suppose we have a measurement accuracy limit which makes it impossible to distinguish two initial conditions separated by a distance less than δ_0. If the system has a positive Lyapunov exponent λ, within a time of the order of

$$\tau \sim 1/\lambda, \tag{16.2}$$

it will be impossible to predict the state of the system. τ is called the *Lyapunov time*. The existence of a positive Lyapunov exponent imposes severe restrictions on the predictability of chaotic systems, even though they are governed by totally deterministic laws. To see this, suppose we want to improve the prediction time of a certain chaotic system of interest by increasing the accuracy δ_0 of the measurement of the initial conditions. Assume that due to a technological breakthrough our measurement is now ten orders of magnitude more accurate. This would mean that we managed to decrease δ_0: $\delta_0 \to 10^{-10}\delta_0$. Let Δ be the greatest error we consider acceptable in our prediction. So we can define the prediction time T as the time it takes for the initial error δ_0 to increase to Δ; in other words,

$$T = \frac{1}{\lambda} \ln \left(\frac{\Delta}{\delta_0} \right). \tag{16.3}$$

The above expression means that decreasing δ_0 by a factor 10^{-10} would increase the predictability time by an additional $\lambda^{-1} 10 \ln 10$. If, for example, $\Delta/\delta_0 = 10^3$ before the increase in accuracy, this would represent only about a fourfold increase in the prediction time. So we would have to increase the accuracy by ten orders of magnitude to buy a fourfold increase in the prediction time! The problem here is that we would need to increase the accuracy exponentially fast in order to obtain a linear increase in the quality of our prediction as measured by T, which makes increasing T indefinitely a hopeless proposition. This means that in practise, for times much larger than the Lyapunov time, chaotic systems are unpredictable.

16.3   The natural measure

In many cases the trajectories of a dynamical system are confined to a bounded region in phase space; this is the usual situation in models of biological systems. For dissipative systems, volume in phase space is contracted by the motion, which implies that trajectories eventually converge to a set of zero volume (or measure) in phase space, called *attractors*. If the system is chaotic, however, this zero-volume set cannot be a simple fixed point or limit cycle, since the positive Lyapunov exponent of chaotic systems means that nearby orbits separate exponentially. This *stretching* characterised by the Lyapunov exponent is combined with the *folding* caused

by the confined nature of the motion and the contraction caused by the dissipative dynamics, resulting in attractors with very complex geometries – *strange attractors* [7], as they are known. It can be shown that this stretching and folding mechanism generates attractors with fractal properties – that is, nonsmooth manifolds with nontrivial structure in arbitrarily small scales.

How are orbits organised within attractors? The exponential separation of initially close orbits suggests that trajectories starting from a neighbourhood of any given point in the attractor can eventually get arbitrarily close to any other point of the attractor. This property is called *topological transitivity*, and it is a crucial feature of chaotic dynamics. It implies the existence of *dense orbits* in the attractor. An orbit is dense if every neighbourhood of every point in the attractor has a nonempty intersection with the orbit.

Topological transitivity suggests that single trajectories may sample the whole attractor, in much the same way that in classical statistical mechanics the state of a gas is imagined to sample all the phase space region it has access to – the energy shell in the microcanonical ensemble. There we pass from a description of the system based on trajectories to one based on probabilities by invoking the idea that the system samples the accessible phase space such that the average time it spends in a certain volume of the phase space is proportional to that volume, independently of the initial conditions. This (unproved) assumption of classical statistical physics allows us to assign probabilities to given regions of the phase space, and forget about trajectories, simplifying enormously the task of making predictions about the system.

We can follow a similar path in chaotic systems if dynamics is *ergodic* [8], that is, if it is such that almost all trajectories of the system spend the same fraction of time in any given region of phase space, independently of their initial conditions. By 'almost all' we mean all initial conditions except possibly for a set of volume zero (more precisely, of Lebesgue measure zero). To make this concept more precise, let us define a partition of the phase space as a set of N disjoint sets $\{W_i\}_{i=1}^{N}$ such that the attractor is contained in the union of all N sets. Consider a certain initial condition x_0 which generates a certain trajectory, and define as $p_i^{(x_0)}$ the fraction of time the trajectory spends in the set W_i. If the dynamics is ergodic, the $p_i^{(x_0)}$ are independent of x_0 except for a set of zero volume. We can therefore write the fractions as simply p_i. Interpreting p_i as the probability that we find the system in the region W_i of phase space, we make a conceptual leap similar to the one made in statistical mechanics.

But are chaotic systems really ergodic? Ergodicity requires a loss of memory: trajectories should 'forget' about their initial conditions after a sufficiently long time for ergodicity to hold. This is strongly suggested by the existence of positive Lyapunov exponents and exponential separation of trajectories – as we have seen previously, for times much longer than the Lyapunov time trajectories can be expected to be found anywhere, regardless of where they started from. Ergodicity has been proved rigorously in some simple systems, and has been verified numerically to high accuracy in many others, and it is widely believed that it is a general property of chaotic systems.

The combination of topological transitivity (the property that a single orbit can explore the whole of the attractor) and ergodicity (the fact that orbits do this exploration independently of their initial conditions) ensures that one can assign uniquely a probability $p(A)$ to any region A of phase space; the function p is called the *natural measure* of the dynamics [9], and plays a crucial role in dynamical systems theory.

16.4 The Kolmogorov – Sinai entropy

Now that we can assign probabilities to sets of states for chaotic systems, we can make use of all the concepts used in statistical physics. In particular, the entropy $S(\{W_i\})$ of a given partition $\{W_i\}$ is defined as

$$S(\{W_i\}) = S^{(1)} = -\sum_i p_i \ln p_i. \tag{16.4}$$

$p_i = p(W_i)$ is the probability of finding the system in the region W_i of phase space at a certain time t, say at $t = 0$. Now let us define p_{ij} as the probability that the system is observed in region W_i at time $t = 0$, and then at region W_j at a later time $t = \Delta t$. Thus p_{ij} is given by

$$p_{ij} = p(W_{ij}) = p(W_i \cap M_{\Delta t}^{-1} W_j), \tag{16.5}$$

where $M_{\Delta t}^{-1} W_j$ denotes the set of initial conditions such that they are found in W_j at time Δt (M^{-1} is the pre-image of the mapping M which advances the system's state by a time interval Δt). The sets defined by $W_{ij} = W_i \cap M_{\Delta t}^{-1} W_j$ constitute a more refined partition than the original partition $\{W_i\}$, with more regions with smaller measure. We can define an entropy $S^{(2)}(\{W_i\}) = S(\{W_{ij}\})$ for this new partition:

$$S^{(2)} = S(\{W_{ij}\}) = - \sum_{ij} p_{ij} \ln p_{ij}. \tag{16.6}$$

Borrowing from Shannon's information theory [10], $S^{(1)}$ can be interpreted as the amount of information we have about the system, given that we know in which region W_i the system is at $t = 0$. Similarly, $S^{(2)}$ is a measure of the amount of information we have if we know in which of the N regions of the partition the system is at times $t = 0$ and $t = \Delta t$. So the difference $S^{(2)} - S^{(1)} = \Delta S^{(2)}$ is the amount of information we have gained by taking note of where the system goes one time step in the future. We can define the rate of 'information production' $H(\{W_i\})$ by a chaotic system during a time interval Δt, for a given partition $\{W_i\}$, as

$$H(\{W_i\}) = \lim_{n \to \infty} \Delta S^{(n)}. \tag{16.7}$$

The *Kolmogorov – Sinai entropy* H is defined by choosing a partition $\{W_i\}$ which maximises the information gain [11]:

$$H = \sup_{\{W_i\}} H(\{W_i\}). \tag{16.8}$$

When we introduced the concept of the separation of trajectories, the Lyapunov exponent was interpreted as a measure of the amount of time it takes for a chaotic system to become unpredictable. There chaos was seen as a creator of uncertainty. Now we are talking about chaos as a generator of information. How are these two points of view reconciled?

The answer is that the exponential separation of trajectories makes it very hard to infer the future state of the system from the present state (beyond the Lyapunov time), but the same phenomenon enables one to infer the past state from the present one with exponentially increasing accuracy, as shown in the discussion about the ever finer partitions $W_{ij\ldots k}$ above. Suppose our partition $\{W_i\}$ represents the accuracy with which we can measure the state of the system. If we observe the system at time $t = 0$ and see that it is in region W_1, for example, we only know that the initial condition is in that region. But after observing a further time step and seeing that the system is in, say, W_3, we now know that the initial condition must be in $W_1 \cap M_{\Delta t}^{-1} W_3$, which is a smaller region than W_1; so we have increased the accuracy of our estimation of the initial condition. A further observation will increase our accuracy even more, and so on. Because of the exponential separation, clearly the measure of the region where we estimate the initial condition must be decreases exponentially — otherwise the Kolmogorov–Sinai entropy would be zero from Equation (16.8).

16.5 Symbolic dynamics

For a given partition $\{W_i\}_{i=1}^{N}$, any trajectory of the system can be mapped to a bi-infinite sequence of symbols

$$\cdots a_{-2} a_{-1} . a_0 a_1 a_2 \cdots, \tag{16.9}$$

where the symbols $a_i \in \{1, 2, \ldots, N\}$ correspond to the successive positions of the trajectory with respect to the partition $\{W_i\}$. Taking the index $i = 0$ to mean time t, a sequence $\cdots 21.31 \cdots$ would mean that at time $t = 0$ the system is in W_3, at time $t = \Delta t$ it will be in W_1, at time $t = -2\Delta t$ it was in W_2, and so on.

We can imagine that the infinite sequence (16.9) is approached by a limit process where symbols are added successively to the specification of an orbit. So we would start with $a_{-1}.a_0$, and then add one more time step in the past and in the future and make the sequence $a_{-2}a_{-1}.a_0a_1$, and so forth. Let us denote by A_{ij} all the orbits in the system which correspond to the sequence $i.j$, that is, to $a_{-1} = i$ and $a_0 = j$. There are many trajectories satisfying this. Now consider the set A_{kijl}; this is a subset of A_{ij}, of smaller measure. As we increase the number of symbols we prescribe, we narrow down the set of possible trajectories more and more. In the limit of infinite symbols, we could expect that we would narrow it down to a single orbit. This is indeed the case if we choose the appropriate partitions, called *Markov partitions* [2]. Using Markov partitions, we establish a one-to-one mapping between the orbits in the system and the set Λ of all allowed symbol sequences. Advancing one time step Δt corresponds to moving the dot in Equation (16.9) one position to the right:

$$\cdots a_{-2}a_{-1}.a_0a_1a_2 \cdots \longrightarrow \cdots a_{-2}a_{-1}a_0.a_1a_2 \cdots . \tag{16.10}$$

This procedure of mapping the original dynamics of the system to a dynamics in the symbol space Λ is referred to as *symbolic dynamics* [12]. A number of properties of chaotic dynamics can be very clearly seen using symbolic dynamics. For simplicity in the following discussion, we will assume that we have a Markov partition with only two symbols, which we will designate by 0 and 1. We will assume, furthermore, that all possible transitions are allowed – that is, any symbol (0 or 1) can succeed any symbol. This means that all possible combinations of 0 and 1 are allowed sequences, and correspond to unique trajectories. The conclusions which follow do not depend on these assumptions; they are only made to simplify the arguments.

It is clear that because of the one-to-one mapping between Λ and the trajectories, all repeated symbol sequences correspond to periodic orbits – for example, $\cdots 01.0101 \cdots$ corresponds to a period-2 orbit, while $\cdots 001.001 \cdots$ and $\cdots 101.101 \cdots$ are period-3 orbits. It follows immediately from this that there are infinitely many periodic orbits in the system, with orbits of arbitrarily high periods. There are more ways of constructing period-3 symbol sequences than period-2, and in general, it can be easily verified that the number of periodic orbits increases exponentially with the period. These are all general properties of chaotic systems.

What about the sequences which never repeat themselves? These correspond to aperiodic orbits, and represent 'most' of the orbits in the system, even though there are infinitely many periodic orbits. This can be seen by the following argument. First we show that we can map every symbol sequence in Λ to a point in the unit square. To do this, we define the coordinate x on the square to be the real number with a base-2 expansion of $0.a_0a_1a_2 \cdots$. That is,

$$x = \frac{1}{2}a_0 + \frac{1}{2^2}a_1 + \frac{1}{2^3}a_2 + \cdots . \tag{16.11}$$

Similarly, we define the y coordinate by

$$y = \frac{1}{2}a_{-1} + \frac{1}{2^2}a_{-2} + \frac{1}{2^3}a_{-3} + \cdots . \tag{16.12}$$

Clearly $0 \leq x, y \leq 1$. Periodic orbits then correspond to points in the square with rational coordinates, which are associated with repeated digit sequences. So the set of all points with irrational coordinates are the aperiodic orbits. But we know from analysis that the set of points with irrational coordinates has full measure – that is, area 1 – while the set of points with rational coordinates has total area zero. So 'almost all' orbits in the system are aperiodic. The set of periodic orbits, however, is very important nevertheless, since it is dense: there are (infinitely many) periodic orbits arbitrarily close to any other orbit in the system, and any finite-time

trajectory in the system can be approximated with arbitrary precision by a periodic orbit. In all chaotic systems, periodic orbits are the 'skeleton' of the dynamics, and all the important dynamical features of a system can be understood in terms of the periodic orbits.

16.6   Chaos in biology

Chaos theory and dynamical systems theory in general have close ties with biology, and many of the most paradigmatic examples in chaos are inspired from biology. For example, the logistic map is one of the most well-known examples of one-dimensional mappings exhibiting chaotic behaviour, and it was suggested by May as an idealised discrete-time model for population growth in a single-species ecosystem with limited resources [13]. In fact, chaos is ubiquitous in ecology and population dynamics, including in particular epidemiology [14].

Chaos is also relevant for many other areas in biology, from the dynamics of whole organs in multicellular organisms, such as the heart, to that of single cells, such as neurons, and even subcellular processes, such as intracellular calcium oscillations [14], to name just a few. The nonlinearity of the dynamical interactions between the various components of biological systems makes chaos a likely phenomenon, and we expect chaos theory to continue to be relevant for biology in the future.

References

[1] Eckmann, J.-P., Ruelle, D. (1985) *Rev. Mod. Phys.* **57**, 617.

[2] Alligood, K. T., Sauer, T.D., Yorke, J.A. (1996) Chaos: an Introduction to Dynamical Systems, Springer-Verlag, New York, NY.

[3] Devaney, R. L. (2003) An Introduction to Chaotic Dynamical Systems, 2nd edn, Westview Press, Boulder, CO.

[4] Ott, E. (2002) Chaos in Dynamical Systems, Cambridge University Press, New York, NY.

[5] Strogatz, S. (2000) Nonlinear Dynamics and Chaos, Perseus Publishing, Cambridge, MA.

[6] Oseledec, V. I. (1968) *Trans. Moscow Math. Soc.* **19**, 197.

[7] Ruelle, D., Takens, F. (1971) *Comm. Math. Phys.* **20**, 167.

[8] Birkhoff, G. D. (1931) *Proc. Natl Acad. Sci. USA* **17**, 656.

[9] Bowen, R., Ruelle, D. (1975) *Invent. Math.* **79**, 181.

[10] Shannon, C. E. (1948) *Bell Syst. Tech. J.* **27**, 379.

[11] Kolmogorov, A. K. (1958) *Dokl. Akad. Nauk. SSSR* **98**, 525.

[12] Williams, S. (ed.) (2004) *Proc. Symp. Appl. Math.* 60.

[13] May, R. M. (1976) *Nature* **261**, 459.

[14] Murray, J. D. (2001) Mathematical Biology, Springer-Verlag, New York, NY.