Understanding Complex Systems

Springer : COMPLEXITY

Yurii Bolotin
Anatoli Tur
Vladimir Yanovsky

Chaos: Concepts, Control and Constructive Use

Second Edition

Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems – cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Editorial and Programme Advisory Board

Understanding Complex Systems

Founding Editor: J.A. Scott Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition – typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels – and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

More information about this series at http://www.springer.com/series/5394

Yurii Bolotin • Anatoli Tur • Vladimir Yanovsky

Chaos: Concepts, Control and Constructive Use

Second Edition

Springer

Yurii Bolotin
A.I. Akhiezer Institute for Theoretical Physics
Kharkov, Ukraine

Anatoli Tur
Institut de Recherche en Astrophysique et Planetologie
CNRS Université Paul-Sabatier
Toulouse, France

Vladimir Yanovsky
Institute for Single Crystals
National Academy of Sciences of Ukraine
Kharkov, Ukraine

ISSN 1860-0832 ISSN 1860-0840 (electronic)
Understanding Complex Systems
ISBN 978-3-319-42495-8 ISBN 978-3-319-42496-5 (eBook)
DOI 10.1007/978-3-319-42496-5

Library of Congress Control Number: 2016951642

1st edition: © Springer-Verlag Berlin Heidelberg 2009
2nd edition: © Springer International Publishing Switzerland 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

The study of physics has changed in character, mainly due to the trend from analysis of linear systems to analysis of nonlinear systems. Needless to say, this trend began a long time ago, but a qualitative change took place and began to evolve in a more pronounced way only after physicists began to understand the nature of chaos in nonlinear systems. The importance of these systems comes from the fact that most of physical reality is in fact nonlinear. Linearity appears as a result of the simplification of real systems and is often difficult to achieve in experimental studies. In this book, we focus our attention on some general phenomena, naturally linked with nonlinearity, in which chaos plays a constructive role.

The first chapter is a general introduction. The second discusses the concept of chaos. It attempts to describe the meaning of chaos according to our current understanding in physics and mathematics. The content of this chapter is essential to understand the nature of chaos and its appearance in deterministic physical systems. Using the Turing machine, we formulate the concept of complexity according to Kolmogorov. Further, we state the algorithmic theory of Kolmogorov's Martin-Löf randomness, which gives a deep understanding of the nature of deterministic chaos. Readers will not need any advanced knowledge to understand it, and all the necessary facts and definitions will be explained. The third chapter describes the main properties of chaos and its numerous qualitative features. Those presented in this chapter are widely used for the investigation of various nonlinear processes and objects, as well as for theoretical and experimental research. In this regard, it deals with the main tools used to study chaos, such as the Poincaré section, the Lyapunov index, etc. In the fourth chapter, we briefly consider the problem of the restoration of a dynamical system with an attractor, based on the observation of temporal data for some generalized coordinates. We discuss the attractor fractional dimension and the fundamental Takens theorem.

Chapter 5 deals with one of the key fields in nonlinear dynamics, namely, the control of chaos, which was discovered in the pioneering work of Ott, Grebogi, and Yorke "Controlling Chaos" Phys. Rev. Lett. 64: 1196 (1990). If a system is chaotic, then small perturbations increase exponentially in time and completely change the behavior of the system. On the one hand, this feature complicates working with a chaotic system, but on the other hand, it allows purposeful control of the system's behavior using these small perturbations. In the same chapter, we consider different methods for transforming chaos into periodic motion (both discrete and continuous), based on the key concept of reconstructing a global system at the expense of a small perturbation. We describe algorithms for dissipative, reversible, and Hamiltonian systems, as well as for scattering problems. The methods developed are applied to control chaos, both in abstract models and in real operational systems. At the end of this chapter, we discuss the possibility of quantum dynamical control. Chapter 6 is about the synchronization of chaotic systems. Two identical chaotic systems with almost the same initial conditions can diverge exponentially in phase space. This is the main difficulty to be overcome if we wish to solve this problem. We outline the methods for its solution, for both dissipative and Hamiltonian systems. We demonstrate the possibility of solving the synchronization problem using the control methods explained in the previous chapter. We are also interested in the influence of noise on the process of synchronization. Further, we briefly examine the principles of the transfer of coded information based on the effects of chaos synchronization.

In the seventh chapter, we investigate in detail the stochastic resonance effect. This effect makes it possible to set up a stochastic system for maximum amplification of the modulation signal by means of noise intensity variations. Stochastic resonance can be realized in any nonlinear system which has some characteristic time scale, and one scale can be controlled with the help of the noise. We examine various generalizations of the initial-value problem, such as stochastic resonance in chaotic systems. The possible relation of this effect to global changes in the Earth's climate is also discussed in this chapter. The main topic of Chap. 8 is transport phenomena in spatially periodic systems without macroscopic forces (gradients). We discuss in detail the necessary conditions for the existence of a macroscopic current in such a situation. At the same time, we provide a classification of systems (ratchets) in which currents without gradient are possible. As one possible realization of the effect, we consider the generation of regular motion in media with nonlinear friction. In this chapter, we also expound the theory of biological motors (molecular motors), which is a very important application of ratchet theory. The concluding Chaps. 9 and 10 are devoted to the problem of quantum chaos or, more precisely, to quantum manifestations of classical stochasticity. There is no doubt that nonlinear systems manifest chaotic regimes in certain ranges of their parameters. On the other hand, a fundamental description of the dynamics requires the use of quantum mechanics, rather than classical physics. What are the particular features of quantum systems if their classical analogues show chaotic behavior? The answer to this question is far from trivial. The problem is that, on the one hand, the energy spectrum of any quantum system with finite motion is discrete, so its evolution will be quasiperiodic, but, on the other hand, the correspondence principle requires a transition to classical mechanics that involves not only regular modes but also chaotic ones. However, a correct formulation of the quantum chaos problem allows us to avoid any contradictions. Signatures of classical chaos are identified in the statistical properties of the energy spectra, the wave function structure, and the dynamics of the wave packets. The main methods used here are the Gutzwiller trace formula for the density of the quantum spectrum, the spectral method, and the Riemann zeta function. Moreover, as we show in Chap. 9, although tunneling is absent in classical physics, it is the structure of the classical phase space that defines a purely quantum effect for tunneling. The paradigm of chaos-assisted tunneling (CAT) is annular billiards. This theory is the main topic of Chap. 10, which also deals with the presence of chaos in nuclear dynamics.

We wish to express our deepest gratitude to Hermann Haken for his interest in our work and his support of this book. We are very grateful to Christian Caron and Gabriele Hakuba for their invaluable help in editing the book. We would like to thank the Institut de Recherche en Astrophysique et Planétologie (CNRS, Université Paul Sabatier) and particularly Philippe Louarn for his support of our project. We thank Tatiana Tour for her assistance in the preparation of this book.

Kharkov, Ukraine — Yurii Bolotin
Toulouse, France — Anatoli Tur
Kharkov, Ukraine — Vladimir Yanovsky
May 2016

Contents

Chapter 1
Introduction

The notion of chaos is so deeply integrated into all fields of science, culture, and other human activities, that it has become fundamental. It is quite probable that each person has his or her own intuitive concept of what is chaos. The first cosmogonical views which explained the origin of the world contained the notion of chaos. For example, the ancient Greeks believed that chaos appeared before everything and then only afterwards the world appeared.

For example, Hesiod says: "In the beginning there was only chaos. Then out of the void appeared Erebus, the unknowable place where death dwells, and Night. All else was empty, silent, endless, darkness. Then somehow Love was born bringing a start of order. From Love came Light and Day. Once there was Light and Day, Gaea, the earth appeared."

The origin of the word chaos itself, $\chi\alpha o\varsigma$, comes from ancient Greek $\chi\alpha\iota\nu\omega$, which means to open wide.

We can find a similar concept concerning the origins of the world in ancient Chinese myth: "In the beginning, the heavens and earth were still one and all was chaos…"

In ancient Indian literature the origins of the world are also associated with chaos: "A time is envisioned when the world was not, only a watery chaos (the dark, "indistinguishable sea") and a warm cosmic breath, which could give an impetus of life." [1]

An ancient Egyptian origin myth holds that in the beginning, the universe was filled with the primeval waters of chaos, which was the god Nun.

Such unanimity of such distant civilizations (in space as well as in time) on the idea of the origination of the world from chaos is quite striking. It seems that this notion is one of the most important to come from ancient times. What is also striking is the fact that the problem of chaos is still topical after some hundreds of years of its study in mathematics and physics. Even a simple list of fields where chaos plays a fundamental part in modern science would take up most of this book. However, a

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_1

new understanding of the origins of chaos was reached quite recently in connection with the discovery of deterministic chaos.

Before the discovery of this phenomenon, all studies of random processes and of chaos were usually conducted within the frame of classical theory of probability, which requires one to define a set of random events or a set of random process realizations or a set of other statistical ensembles. After that, probability itself is assigned and studied as a measure on this set, which satisfies Kolmogorov's axioms [2]. The discovery of deterministic chaos radically changed this situation.

Chaos was found in dynamical systems, which do not contain elements of randomness at all, i.e., they do not have any statistical ensembles. On the contrary, the dynamic of such systems is completely predictable, the trajectory, assigned precisely by its initial conditions, reproduces itself precisely, but nevertheless its behavior is chaotic.

At first sight, this situation does not correspond to our intuitive understanding, according to which chaotic behavior by its very nature cannot be reproduced. A simplified and pragmatic explanation is frequently used to explain this phenomenon. Dynamical chaos appears in non-linear systems with trajectories utterly sensible to minor modifications of initial conditions. In that case any person calculating a trajectory using a computer observes that the small uncertainty of initial conditions engenders chaotic behavior.

This answer does leave some feeling of dissatisfaction. In fact, we know that, for instance, the number $\sqrt{2}$ exists accurately, without any uncertainty. What would happen if the trajectory began precisely from $\sqrt{2}$? The usual answer to this question is that the behavior of trajectory will become more and more complex, because the number $\sqrt{2}$ is irrational and ultimately will be practically indistinguishable from the chaotic, although remaining determined.

However, two questions persist: what do we mean by "complex" and what does "practically indistinguishable from the chaotic" mean? For example, genetic code is complex but not chaotic, while a coin toss is a simple, but chaotic process. Even from the above we can see that the phenomenon of deterministic chaos requires a deeper understanding of randomness, not based on the notion of a statistical ensemble.

Such a theory was developed by Kolmogorov and his disciples even before the discovery of the phenomenon of deterministic chaos. The main principle of this theory will be stated in the next chapter, where we will introduce all its necessary components: algorithms, Turing machine, Kolmogorov's complexity, etc. It is significant that Kolmogorov came to his theory when discussing in articles [3, 4] the limited nature of Shannon's theory of information [5].

As an example, let us question how to understand what is the genetic information, for example, of a tiger or of Mr. Smith. It seems that, since the notion of information is based on the introduction of probabilities, we have to examine a set of tigers with assigned probability. Only after that, one can calculate Shannon information in the tiger's genes. It is clear that something in these considerations provokes anxiety. Above all, the dissatisfaction is caused by the introduction of the set of Smiths, let's say. Obviously, it is more pleasant to consider that Mr. Smith is unique and has individual genetic information like a particular tiger. The limited nature of probabilistic approach becomes even clearer if you consider how much information is contained in the book "War and Peace" by Leo Tolstoy. Then the problem with the introduction of the set "War and Peace" becomes perfectly evident.

The problem is that we are interested in information for the object that is individual. In other words, we have only one specimen of the object and it is impossible to create another one even mentally. Therefore, the theory of probability and the theory of information must be restated in such a way that the individual object get the character of random one. Kolmogorov states [4]: "1) The main concepts of the theory of information must be and can be founded without using the theory of probability and in such a way that notions of "entropy" and "quantity of information" become applicable for individual objects. 2) Notions of the theory of information introduced in this way can underlie the new concept of random, which corresponds to natural thought, that randomness is the absence of laws."

The algorithmic theory of randomness developed by Kolmogorov and his disciples is a natural mathematical basis for understanding the theory of deterministic chaos; we will examine it in the next chapter. This theory gives natural answers to the questions that were put earlier: when complexity turns into randomness and how algorithmic randomness obeys the theory of probability. From the physical point of view, this means that the distinction between dynamical and statistical laws gets erased.

Note that the reader is not obliged to begin the study of the theory of chaos and its applications with Kolmogorov's algorithmic theory of randomness. In other words, the chapter Paradigm for Chaos is to a great extent independent from the other chapters in this book. However, experience has shown that the study of deterministic chaos can feel unfinished and unsatisfying. Therefore, if at any time the reader feels a desire for a deeper understanding of the nature of deterministic chaos, they can turn to "Paradigm for Chaos."

References

1. H.H. Wilson was the first to make a complete translation of the Rig Veda into English, published in six volumes during the period 1850–1888. Wilson's version was based on the commentary of Sa-yan.a. In 1977, Wilson's edition was enlarged by Nag Sharan Singh (Nag Publishers, Delhi, 2nd ed. 1990).
2. Kolmogorov, A.N.: Foundations of Probability. Chelsea Publishing Company, New York (1950)
3. Kolmogorov, A.N.: Problemy Peredachi Informatzii **1**, 3–11 (1965)
4. Kolmogorov, A.N.: IEEE Trans. Inf. Theory **IT-14**, 662–664 (1968)
5. Shannon, C.E.: A mathematical theory of communication. Pt. I, II. Bell. Syst. Technol. J. **27**(N3), 379–423; N4, 623–656 (1948)

Chapter 2
Paradigm for Chaos

One of the major concepts central to the deeper understanding of contemporary physics is the concept of chaos. It would not be an exaggeration to say that chaos is everywhere in physics. The chaotic behavior of physical systems was considered until lately as the result of unknown factors which influence the system. In other words, chaos was supposed to appear in physical systems because of the interactions with other systems. Earlier, as a rule, it was considered that these actions were complicated and uncontrolled and usually random. Random parameters and fields arose in dynamical systems due to these phenomena. As a result, the variables, which describe the dynamical systems, are random. The development of non-linear physics and the discovery of deterministic chaos led to an important change of point of view on both the apparition of chaos in physical systems and the nature of chaos.

Currently, there are many good books and reviews dealing with the theory and applications of deterministic chaos. In most descriptions of deterministic chaos a pragmatic point of view prevails: chaos appears in dynamical systems with trajectories utterly sensible to minor changes in initial conditions. At that, the individual trajectory is, as a rule, so complex that it is practically impossible to distinguish it from a chaotic one. At the same time, this trajectory is completely determined. This point of view, which is sufficient enough for practical work with non-linear dynamical systems, is the one most commonly used. However, the question of the deeper origins of deterministic chaos is rarely discussed. When and why is the behavior of a determined trajectory not only complex and "similar to random," but really random? In other words, can we apply probability laws to it, despite the fact that at the same time, it is quite determined and unique with the same initial conditions? It is clear that answers to these questions are of fundamental importance even if they will not contribute to perfecting techniques of practical calculations of chaos characteristics.

In this chapter we will state the foundations of the algorithmic theory of randomness of Kolmogorov–Martin-Löf which can provide a deeper understanding of the origins of deterministic chaos. The algorithmic theory of randomness does not

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_2

deal directly with dynamical systems. Instead, this theory examines the work of the universal Turing machine. A Turing machine works according to some determined program and a prints finite or infinite sequence of symbols from a finite alphabet (for example, 0 or 1). We can consider the Turing machine an abstract model for a determined dynamical system or as a model of a computer programmed to solve a motion equation for a dynamical system. We can think of the sequence printed by this Turing machine as the trajectory of the dynamical system.

Now we can examine the same question again but this time within the frame of the mathematical theory of Turing machine: supposing that the sequence (trajectory) is complex, would it be in any sense random? The theory of Kolmogorov–Martin-Löf gives the answer to this question. Kolmogorov formalized the concept of complexity when analyzing the length of programs for the Turing machine. He introduced the concept of complexity, now called Kolmogorov's complexity. Kolmogorov's disciple, Martin-Löf, proved the remarkable theorem. Complex sequences, according to Kolmogorov, are random to the extent that they obey all the theorems of the theory of probability with an accuracy up to set of zero measure. This theorem is astonishing, because its proof concerns not only already-known theorems of probability theory, but also theorems which are not yet proven.

Thus, it was strictly proven that the complexity of determined sequences (trajectories), which is understood as the absence of laws, actually turns into true randomness. As a result, the theory of Kolmogorov–Martin-Löf, whose importance is probably not yet fully appreciated in physics, gives a new understanding of the origins of randomness and of deterministic chaos. This is applicable to individual objects without using statistical ensembles.

2.1 Order and Disorder

In order to discuss these concepts, it is natural to start with the most obvious ones. It seems normal that order is simpler than disorder. Let us imagine an experimenter who works with an instrument and who measures the value of some variables. If his instruments record the value $7, 7, 7, 7, 7, \ldots, 7$ multiple times, a rule becomes obvious and simple, under the condition that the experimenter is sure that it would continue in the same way. Other results can also appear, like $7, 2, 7, 2, \ldots, 7, 2$ or $7, 2, 3, 5, 7, 2, 3, 5, \ldots$, so the rule can be seen without any difficulty if the experimenter is sure to repeat the same results as before. However, there are situations when the rule is more complicated and its finding requires efforts which go beyond the scope of these simple observations. The reasoning above suggests that as a definition of ordered behavior or, in this case, ordered sequences of numbers, one can propose a seemingly simple definition. This naive definition means that we can predict all the terms of the sequences using its limited part only. But this definition is not very useful since it is practically impossible to guess the rules of construction for a complex sequence. For instance, if we took the sequence of the first thousand decimal digits belonging to the fractional part of number π, it would seem random. However, when we investigate the simple rule of its construction (a short program for a computer), we can no longer consider this sequence as being random. Actually, if we have a limited part of the sequence, we can imagine an endless number of rules which give different sequences where the limited part of the beginning is present. All this shows that this attempt to define the concept of order is not at all sufficient for understanding the concept of chaos. Therefore some strict mathematical methods are needed to distinguish well-ordered sequences from chaotic.

Let us now consider a different way of introducing these concepts of order and disorder. For the sake of simplicity, we will consider sequences containing only 0 and 1. Then the main question appears: how to distinguish ordered sequences of 0 and 1 from disordered ones. It goes without saying that the origins of these sequences are not important.

The first attempt to define random consequences using the frequency approach was made by Mises [1] who tried to formulate the essential logical principles of the probability theory. To begin with let us examine the Mises's scheme.

Let us suppose an infinite sequence $x_1, x_2, x_3, \ldots$ made of zeros and ones. According to Misses, above all, the necessary condition of randomness is to be fulfilled, i.e., the limit must exist:

$$P = \lim_{N \to \infty} \frac{1}{N} \sum_{i \leq N} x_i \,. \tag{2.1}$$

It is clear that this condition is not sufficient, as, for example, the sequence $0, 1, 0, 1, 0, 1 \ldots$ obeys condition (2.1), but can in no way be considered random. Therefore, Mises believed that there is another condition for randomness. Let us choose from the infinite sequence of numbers $1, 2, 3, \ldots$ a subset of numbers and let us designate it as $n_1, n_2, n_3, \ldots, n_k, \ldots,$. Following this choice let us consider variables Mises' second idea was that the initial sequence of variables x_i is random if for the chosen subsequence $x_{n_1}, x_{n_2}, \ldots, x_{n_k}, \ldots$ the limit (2.1) remains the same, i.e.,

$$P = \lim_{M \to \infty} \frac{1}{M} \sum_{k \leq M} x_{n_k} \,. \tag{2.2}$$

It is clear that the choice of the subsequence is not arbitrary. For instance, it is impossible to choose all the variables x_i as zeros or ones. That is the reason why the rules of the choice of subsequence are most important in Mises' theory. Mises gave only the general characteristic of these rules and restricted himself to some examples: in particular, prime numbers can be chosen as variable numbers, etc. But Mises could not formulate the mathematical scheme of choice or rules, since the concept of rules and laws of choice were not formulated mathematically in his time. In other words, the concepts of algorithm, recursive functions, and Turing

Machine, which formalized his intuitive ideas of laws and rules of choice were not yet developed.

The next studies of the foundation of the theory of randomness were stopped for a long time because of Kolmogorov's proposition to consider the probability theory as an applied measure theory [2]. The elegance of the axioms of Kolmogorov as well as its great possibilities led to the fact that the main efforts of scientists were concentrated on the development of the probability theory in this direction. The question of foundation was forgotten until the appearance of Kolmogorov's new work where he started to study this problem again. The starting point of his work was his introduction of the new concept of complexity as a measure of chaos. The complexity of finite sequences of 0 and 1, according to Kolmogorov, can be measured by the minimum length of its "description": the rules used to construct this sequence. A good example of sequence description is given in the book "The Good Soldier Švejk" by Jaroslav Hašek [3] when Švejk explains a method to remember the number of the railway engine 4268 which the track supervisor recommends to the engine-driver. "On track no. 16, is engine no. 4268. I know you have a bad memory for figures and if I write any figures down on paper you will lose the paper. Now listen carefully and as you're not good at figures I'll show you that it's very easy to remember any number you like. Look: the engine that you are to take to the depot in Lysa nad Labem is no. 4268. Now pay careful attention. The first figure is four, the second is two, which means that you have to remember 42. That's twice two. That means that in the order of the figures 4 comes first. 4 divided by 2 makes 2 and so again you've got next to each other 4 and 2. Now don't be afraid! What are twice 4? 8, isn't it? Well, then get it into your head that 8 is the last in the series of figures in 4268. And now when you've already got in your head that the first figure is 4, the second 2, and the fourth 8, all that's to be done is to be clever and remember the 6 which comes before 8. And that's frightfully simple. The first figure is 4, the second 2 and 4 and 2 are 6. So now you've got it: the second from the end is 6 and now we shall never forget the order of figures. You now have indelibly fixed in your mind the number 4268. But of course you can also reach the same result by an even simpler method. So he then began to explain to him the simpler method of how to remember the number of the engine 4268. 8 minus 2 is 6. And so now he already knew 68. 6 minus 2 is 4. So now he knew 4 and 68, and only the two had to be inserted, which made $4 - 2 - 6 - 8$. And it isn't too difficult either to do it still another way by means of multiplication and division. In this way the result will be the same."

As we can see, there are plenty of ways to describe a sequence. This is the reason why the main problem consists of how to find a method which would contain all the ways to describe the $0, 1$ sequence, from which we need to pick the smallest one. The theory of algorithms by Turing and Post [4, 5] gives the foundations of the formal description of the construction rules for sequences. Their works laid the basis for many mathematical branches, such as mathematical logic, the theory of recursive functions, cybernetics, and the theory of information. Let us consider these works in more detail.

2.2 Algorithms and Turing Machine

In mathematics, an algorithm is the rule which permits one to find a solution to a mathematical problem (if a solution exists), using only regular procedures, without additional heuristic methods. The classical example is the Euclidian algorithm of division. The word algorithm comes from the name of the great Arabian mathematician Mohamed al-Horezmi, whose treaty in Latin begins with the words "Dixit algorizmi" which means "al-Horezmi said." Turing's reflections on the concept of algorithms led him to introduce a new mathematical concept, which is currently called the Turing machine. Nowadays, by definition, the Turing machine is the set:

$$M = \left(\Sigma, S, P, q_0, q_f, a_0\right) , \tag{2.3}$$

where Σ is an external alphabet, with which you can write down the input and output words (sets of letters which are contained in the external alphabet). S is an internal alphabet which describes the internal states of the Turing machine, q_0 is the initial state, q_f is the final state, a_0 is the empty cell, P is the machine program, i.e., the list of commands. As regarding the commands there are three kinds of words:

1. $$qa \to rb$$

   The meaning of this expression is the following. The Turing machine in the state of q and watching the letter a must pass to the state r and write down letter b on the band.

2. $$qa \to rbR$$

   This expression means that the machine in the state of q and watching the letter a must pass in state r, and write down the letter b and move to the right.

3. $$qa \to rbL$$

   This means that the machine in the state of q and watching the letter a must pass in the state of r, and write down the letter b and move to the left.

R, L, and $\to$ are not part of the alphabet Σ and S. By definition the program is a finite sequence of these commands. It is convenient to see the Turing machine as an endless band which is divided into separate cells, on one of which the Turing machine is fixed (see Fig. 2.1). On the band on each cell only one letter can be written from the alphabet Σ. As an example in Fig. 2.2, we present three commands

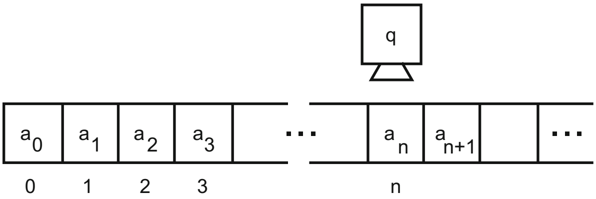


Fig. 2.1 Turing machine. The band of symbols, empty cells, and the reading head of the Turing machine are shown

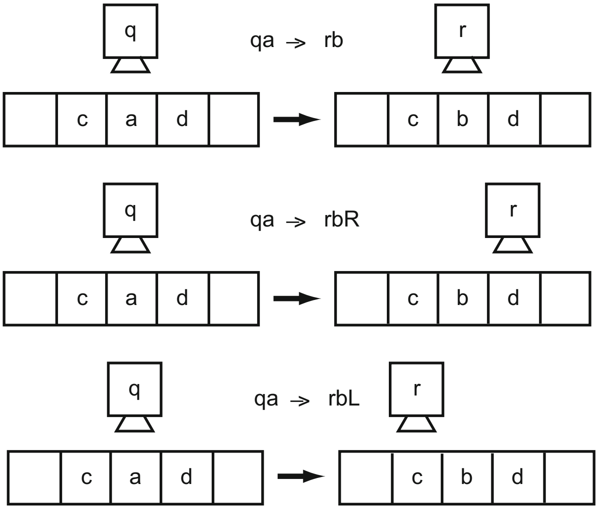


Fig. 2.2 Examples of the execution of commands by the Turing machine

performed by the Turing machine that were described earlier. The state of the Turing machine at any cycle is denoted as $A(q, a)B$. This means, that in the state of q, the machine is fixed on the letter a, on the band on the left of the letter a is the word A and on the right is written the word B.

Let us see now how the Turing machine works. We suppose that there is a machine word or a configuration of a word on the band. The Turing machine's work starts with the initial configuration $(q_0, a_0)B$. After the realization of program P the machine will stop at the final configuration $C(q_f, a_0)D$. The transition between the initial and the final configuration is performed by the command of the P program. Functions, which are calculated by the Turing machine, are called particular recursive functions (the word "particular" refers to the fact that the function is not defined at initial configuration). If the particular recursive function can be defined at any initial configuration it is called recursive. In the case of recursive functions the Turing machine starts to work with any input of integer numbers and always finishes its work giving the value of calculated functions.

The machine can be defined in the initial configuration if it performs its program and stops with the final result $C(q_f, a_0)D$. However, the initial configuration can be undefined for two reasons. First of all, in the process of program realization, the machine can find a configuration to which no commands of this program can be applied. The second reason is that the process of the execution of the program can be endless and the machine might not stop.

The Turing machine that is described gives an algorithm definition in a mathematical sense. It implies that there is an algorithm for a calculation or a process. The machine that is described has modest possibilities. However, it is possible to build a complex Turing machine thanks to the unification of simple ones. At the same time, the calculation possibilities of this machine will grow.

It is important to say that it has been demonstrated that a universal Turing machine can be built. This kind of machine can do whatever any Turing machine M does with an initial configuration. As an input for the universal Turing machine we give the initial configuration and description of Turing machine M. As a result, the initial configuration of treatment will be the same as for the machine M. We have to notice that this machine has impressive capacities. In principle, any modern computer can be coded on a band, and as a result, the universal Turing machine can do anything a modern computer does. The last question that interests us is whether all algorithms, from an intuitive point of view, coincide with the formal definition of the Turing machine. Generally speaking this question is not a mathematical one because there cannot be an algorithm definition in an intuitive sense. Church [6, 7] first answered this question when he proposed a thesis in which he said that every alphabetical, discreet, massive, determined, and closed algorithm can be defined by the Turing machine. We can say that any algorithm in an intuitive sense is given by the Turing machine. We have to emphasize that Church's thesis is not a mathematical assertion, it is more like a statement about energy conservation in physics. However, mathematical experience supports this thesis.

2.3 Complexity and Randomness

Now we have a universal method to describe finite sequences thanks to the Turing machine. Actually, any sequence can be associated with a program P of the Turing machine, thanks to which the Turing machine can write it down. It is clear that for every chosen sequence there are an endless number of programs that can perform it. This is why, according to Kolmogorov, we can define the concept of complexity [8] as related to the Turing machine M. Let us say that the machine M writes down n-value sequences of 0 and 1. By definition, the complexity K_M coincides with the length of the shortest program (in bytes) after the realization of which the machine will write down the given sequence $X = (x_1, x_2, \ldots, x_n)$:

$$K_M = \begin{cases} \min \ell\,(P)\,, & M\,(P) = X, \\ \infty, & M(P) \neq X \end{cases}.$$

This concept of complexity clearly depends on the machine M. However Kolmogorov managed to prove that there is a universal Turing machine for which:

$$K\,(X) \leq K_M\,(X) + C_M\,.$$

The constant C_M does not depend on the sequence and this is why it can be chosen identically for all the sequences. So the complexity K will be minimal and $K(X)$ will be called the complexity of the sequence X according to Kolmogorov. It is possible to prove that there are sequences X (with a n length) for which $K(x) \geq n$. This means that there are no simpler algorithms or ways to describe them than this sequence. Such sequences correspond to our intuitive understanding of random sequences because they do not contain any rules which could simplify them. Nevertheless, for the time being, we do not have any reason to think that probability laws are applicable to these sequences. Thus we have defined the concept of complex finite sequences or a random sequence. In a certain sense this definition can be considered as final. However, it is necessary to extend our definition to infinite sequences. It seems natural to define random infinite sequences of 0 and 1 $X = (x_1, x_2, x_3, \ldots)$ so that

$$K\,(X^n) \geq n + \text{const}\,. \tag{2.4}$$

for any final segment $X^n = (x_1, x_2, x_3, \ldots, x_n)$.

Here the constant depends on the sequence X. However, this definition is not satisfactory. One can prove that sequences for which conditions (2.4) are fulfilled for every n do not exist. We can intuitively understand the reason for these phenomena. Experience shows that in every random sequence, for example, the one you obtain after a coin toss, there are ordered parts of sequence numbers (for example, $1, 1, 1, \ldots, 1$). This means that in every random infinite sequence there is an endless number of ordered segments. Thus complexity behaves like an oscillation function with a growth of n (see Fig. 2.3). In other words, there are many values of n for which:

$$K\,(X^n) < n\,. \tag{2.5}$$

This means that definition (2.4) is not appropriate for random infinite sequences. In order to avoid these difficulties we chose another definition.

The sequence $X = (x_1, x_2, x_3, \ldots)$ is called random according to Kolmogorov if there is a constant C such that for each number n the following condition is satisfied:

$$|K\,(X^n) - n| \leq C\,. \tag{2.6}$$

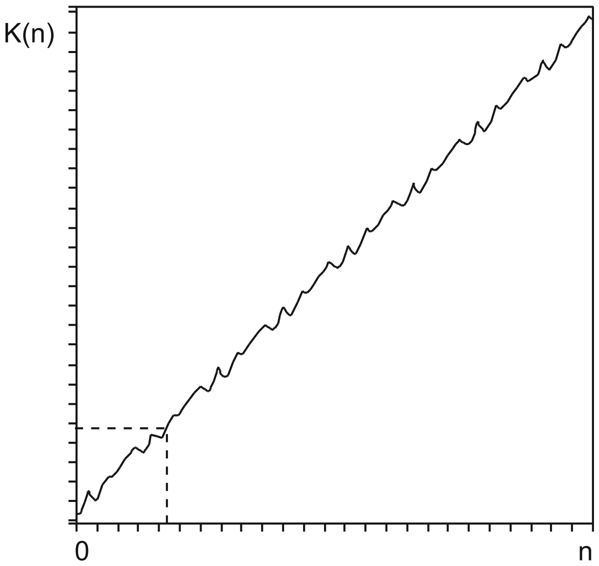


Fig. 2.3 The behavior of complexity with the growth of the number of sequence members. The features of the decrease of complexity on ordered segments of sequences are shown symbolically

From a physical point of view, this means that the decrease of complexity does not go over a certain level and with the growth of n the relative contribution of this decrease to complexity is small.

We will understand random infinite sequences as sequences that fulfill the conditions (2.6). According to Kolmogorov this is a final definition of individual random sequences. Another way to define random sequence is to use monotonous complexity or monotonous entropy (see, for example, [9]) instead of the simple Kolmogorov's complexity. The introduction of monotonous entropy permits us to avoid the difficulty of oscillation in the simple Kolmogorov's complexity.

It seems very natural to think that these sequences are chaotic. However, one question remains: will these sequences be random in the sense that they will obey all theorems of the probability theory? Martin-Löf obtained a positive answer to this question when developing Kolmogorov's ideas [10, 11]. It is not our aim to present the Martin-Löf theory, which is quite complex. However, because of the importance of his results, we are going to explain the main ideas of this theory. Let us consider a set Ω of infinite sequences of 0 and 1. It is clear that the power of set of all sequences has the capacity of a continuum. Let P be a measure for this set. How does the observer exclude all the sequences which have all possible laws from this set? The observer can treat the initial segments of the sequences, find some laws in them as, for example, repetitions, calculate how many bits of regularity he finds, and then exclude these sequences from the admissible set. As a result, the measure of admissible candidates for random sequences will fall. Since these actions are recursive, the observer can charge the Turing machine with these actions. This idea is the base of the "universal test of randomness" or universal sequential test of Martin-Löf. Martin-Löf's test P for randomness is a recursive function F (or the Turing machine) which treats finite sequences of 0 and 1 with a length of m and find how many bytes of regularity there are in it (roughly speaking the common segment). Then for every byte of regularity, recursive function F decreases by at least twice the set of admissible sequences ω and its measure P. The sequences which remain in the admissible set for any length m are considered a sequence which goes through the test P. The union of all P-tests gives a universal sequential test which is a limiting concept. The sifting of the sequences through the Martin-Löf test eliminates all the sequences which have any laws. As the number of regular sequences is much smaller than that of complex ones, as a result we have only complex sequences in the admissible set after sifting. Martin-Löf demonstrated that the complex sequences after Kolmogorov go through the universal test. We have to explain how the random sequences after Martin-Löf satisfy all the theorems of probability theory which can be tested effectively. Let us suppose that we have the sequence ω which does not satisfy one of the theorems of probability. In this case this theorem can be reformulated and added in the new P-test. Now this sequence does not go through the new P-test, i.e., does not go through the universal test and must be rejected as not random.

So it has been proved that random sequences exist. The power of the set of random sequences is continuum. Let us emphasize that now the Martin-Löf random sequences are random in the classical sense. These sequences obey the theorems of the probability theory.

We have managed to establish a relationship between Kolmogorov and Martin-Löf random sequences. These two sets coincide. Now we can affirm that complex sequences or random in the sense of Kolmogorov are random from the probability theory point of view. Let us present another important theorem. Random sequences cannot be calculated with the Turing machine. There are no algorithms for the calculation of random sequences. This is a very important property. In conclusion, let us note that Church formulated the intuitive ideas of Mises using the theory of algorithms [12]. The final theory which developed the frequency approach of the probability of Mises was proposed by Kolmogorov and Loveland [13–15]. Kolmogorov considered the frequency approach very important, as it explains why the abstract probability theory is applicable in the real world, where the number of trials is always finite[4]. Unfortunately the algorithmic notion of randomness according to Mises–Kolmogorov–Loveland does not correspond in full to our intuitive understanding [16]. The random sequence according to Martin-Löf is random according to Mises–Kolmogorov–Loveland. However there are sequences which are random according to Mises–Kolmogorov–Loveland, but are not random according to Martin-Löf.

2.4 Chaos in a Simple Dynamical System

Now let us look at the source of the appearance of chaos in determined systems. Before introducing the motion equation or system evolution we shall discuss its phase space. For the sake of simplicity, let us limit ourselves to one-dimensional models which evolve in the bounded area of the real R^1 line. Then, as a phase space, without the loss of commonness, we can choose the segment [0, 1]. All the points of this segment can be used as values of our system positions during evolution and also as its initial conditions. Now we are reminded of some of simple facts of the number theory. The segment [0, 1] is filled up with real numbers which are separated as rational numbers and irrational ones. The power of the set of rational numbers coincides with the power of integer ones, i.e., a countable set. The power of the set of irrational numbers is a continuum. For the description of numbers in the calculus system with the radix b (integer number) one uses their single valued representability like a series:

$$x = \frac{a_1}{b} + \frac{a_2}{b^2} + \frac{a_3}{b^3} + \cdots = \sum_{i=1}^{\infty} \frac{a_i}{b^i}$$

$$x \in [0, 1] .$$

where $a_i = 0, 1, 2, \ldots, b-1$. Another expression to write down the numbers in the selected calculus system:

$$x = 0, a_1, a_2, a_3, \ldots$$

is named the b form of number presentation. This form is well known by everyone through the decimal form of writing rational numbers. One often uses binary calculus system in which the radix $b = 2$. For instance, in Babylon, the calculus system with the radix $= 60$ was used because ancient mathematicians did not like fractions. From this point of view, the relatively small number 60 with the big number divisor 12 was a very convenient foundation for the calculus system. As a result, we inherited the division of 60 min in an hour and 60 s in a minute from that Babylonian calculus system. Rational numbers written down in the b form can be easily distinguished from the irrational ones. Actually, any rational number is represented as:

$$\frac{p}{q} = 0, a_1, a_2 \cdots a_m a_1 a_2 \cdots a_m a_1 a_2 \cdots a_m \cdots \tag{2.7}$$

In other words, after m figures in the writing of any rational numbers the e is a block of figures fully repeated periodically. Irrational numbers do not have such periodicity. Obviously it does not help too much to recognize numbers. For example, it is hard to use this fact even to prove the irrationality of $\sqrt{2}$. To do this, one would have to have the infinite recording of this number but that is impossible. That is why the well-known proof of the irrationality of $\sqrt{2}$ obtained by the ancient Greeks was based on another principle.

In some cases the demonstration of rationality or irrationality of some concrete numbers can be very difficult. For example, we currently do not know if the Euler constant $c \approx 0.6$ is a rational or irrational number. By definition the Euler constant is given by the following expression:

$$c = \lim_{n\to\infty} \left(\sum_{i=2}^{n} \frac{1}{i} - \ln n \right) .$$

What is important for us is that rational or irrational numbers are dense everywhere on a segment $[0, 1]$ (see, for example, [17]) and that between two different real numbers there is an infinite number of rational and irrational numbers. Let us note that it is not important if these two numbers are rational or not. The theorem is always true. Hence in the small neighborhood of any number there is an infinite number of rational and irrational numbers. The fact that rational numbers are dense everywhere permits us to have a good approximation of irrational numbers by the rational ones. It has been proven that any irrational number x can be approximated by rational numbers p/q with precision so that:

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^2} .$$

As an example we give an approximation of the irrational number π, $\pi \approx 355/113$. It is easy to verify that:

$$\left| \pi - \frac{355}{113} \right| = 2,66 \cdot 10^{-7} .$$

This means that this fraction coincides with the number π up to the sixth order and corresponds to the inequality that was given before.

Let us transfer some properties of chaotic sequences onto the points of our phase space. It is easy to understand that if we attach 0 with comma (0,) to any sequence, we obtain one-to-one correspondence between the sequences and the point of our phase space. At the same time, the coordinates of our phase space are written down in the calculus system on radix 2. Since the continuum of points exists and there is no algorithm to calculate their coordinates, the coordinates of these points are random.

One might think that these points correspond to all irrational numbers. However, this is not true. For example, the number e is irrational but is not complex according to Kolmogorov, since there is a simple algorithm to calculate it with the expression:

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n} \right)^n .$$

This example shows that not all the sequences which at first sight are complex or chaotic are such in reality, because they can contain some hidden algorithms which were used for their construction. Now let us consider the non-linear dynamical system in discrete time:

$$x_{n+1} = 2x_n \mod 1 . \tag{2.8}$$

Here x mod 1 is a fractional part of the number x. If we know the coordinates x_n at the moment n, it is easy to have its coordinates x_{n+1} at the moment $n + 1$. If we substitute x_{n+1} in the right part, we find the same way the coordinates of the system at the moment $n + 2$. It is hard to imagine a simpler and more determined system. The exact solution of this equation:

$$x_n = 2^n x_0 \mod 1 \tag{2.9}$$

gives a guarantee of the existence and uniqueness of the solution. Thus, the mapping (2.9) has all the features of a strictly determined system. Let us consider one of the trajectories of the system with initial conditions X which belong to the set of "random" points. The dynamic of the system (2.8) means in reality the shift of the comma of one position to the right and the rejection of the integer part of the number at each step. This is why the whole trajectory is actually number x, which is random, hence the trajectory of the motion of the system can be shown to be random.

This example shows that determination is not in contradiction with randomness. Mapping (2.8) is determined and does not contain any random parameters. Moreover, it has an exact solution (2.9). The trajectories with the same initial conditions repeat exactly. However, the behavior of the system, or of the trajectory, is random. In this sense we can speak about a deterministic chaos.

There is a simple way to test it. Let us divide the segment [0, 1] into two segments [0, 1/2] and [1/2, 1]. Now the question is to know in which segment the solution is. The answer to this question depends on whether 0 or 1 is in the corresponding place in the presentation of number x. At the same time, an expert to whom we can present the data about particle position in the first or second segment will not be able to find any difference between this data and the data of a coin toss when 0 or 1 is associated with heads and the other figure is associated with tails. In both cases, he will find that the probability to find the particle in the left segment will be 1/2—the same as for the right segment. In this regard, we can say that the dynamic system (2.8) is the model of coin toss and describes the classical example of the probability process.

Such indeed is the meaning implied when one speaks of continuous phase space as the reason for the chaotic behavior of the system. One might think that this result appeared after a too simple partition of the phase space in cells. However, this is not true. It is possible to partition the segment [0, 1] into more cells and study the transition between them (here we enter into another mathematical branch, the so-called symbolic dynamics [18–20]). The transitions between these cells are described by the Markov process, which are classical examples of probability processes [21, 22]. We can even work with infinitely small cells but if we do not take into account some mathematical difficulties we will not yield anything more than randomness x which has already been proven.

Let us go back to the reasons for chaotic behavior of the trajectory of dynamic systems. First of all it is the continuality of phase space. However, if the chaotic behavior of the system is related to the uncalculated initial data, then why do we observe chaotic behavior in one system but not in others? We can try to answer this question. As a matter of fact if we look attentively at the system (2.8) we can remark that the behavior of this system in time depends more and more on the distant figures of the development of the number x_0. So if we know only a finite number of symbols in the development of the initial data x_0, for example, m, we can describe our system only on a finite interval m. As a result, the dynamical system is sensitive in an exponential way with respect to the uncertainty of the initial data. In such dynamical systems, the potential randomness which is contained in the phase space continuity transforms itself into an actual randomness. At the same time we obtain the first criterion of stochastic behavior of a determined system: stochastic behavior of the trajectory is possible in systems which have an exponential sensitivity to the uncertainty of initial data. This criterion can be presented differently. Chaotic behavior is achieved in nonlinear systems with an exponential divergence in neighboring trajectories. We can easily understand this if we consider uncertainty as a module of the difference between the two possible initial conditions. For our system, the distance $|\varDelta x_0|$ between trajectories which are close at the initial time will grow with the time as $|\varDelta x_0|\, e^{n \ln 2}$.

Another important observation can be made from the study of this simple dynamical system, concerning periodical trajectories or orbits. Let us consider the positions of the periodical trajectories in the phase space. Taking into account the fact that all rational numbers have the form (2.7) it is easy to see that all trajectories with initial conditions x_0, which coincide with the rational numbers, will be periodical. Hence, the periodical orbits are a countable set and are dense everywhere in the phase space. Obviously, the trajectories which were initially close to the periodical orbits will go far away exponentially fast. This is why we call these periodical orbits unstable. Thus, periodical orbits are everywhere dense in the phase space of the dynamical system (2.8). As we will see later, this feature will always be observed in the dynamical systems with chaotic behavior.

Let us pay attention to one important property. If we choose a small neighborhood of initial conditions ω and launch the trajectories out of their neighborhood so, at the moment n, ω will occupy a certain neighborhood ω_n. Our system has the following property, which is easy to test: for any neighborhood ω one can find the time n when we have $\omega_n \cap \omega \neq \varnothing$. Dynamic systems which have this property are called transitive.

Our simplest system which shows the chaotic behavior is transitive. Let us note that the choice of a one-dimensional system is not important. All these properties are exactly applicable in multidimensional systems. For example, a bidimensional system $(x_1, \ldots, x_n, \ldots)$ and $(y_1, \ldots, y_n, \ldots)$ can be easily reduced to a one-dimensional case if we write down these sequences in the form of one sequence $(x_1, y_1, x_2, y_2, \ldots, x_n, y_n, \ldots)$. All these qualitative features are similar in the multidimensional case.

References

1. Mises, R.V.: Math. Z. **5**, 52–99 (1919)
2. Kolmogorov, A.N.: Foundations of Probability. Chelsea Publishing Company, New York (1950)
3. Hašek, J.: The Good Soldier Švejk. Penguin Books, New York (1974)
4. Turing, A.M.: Proc. Lond. Math. Soc. (Ser. 2) **42**, 230–265 (1936)
5. Post, E.L.: J. Symb. Log. **1**, 103–105 (1936)
6. Church, A.: Am. J. Math. **58**, 345–363 (1936)
7. Church, A.: Introduction to Mathematical Logic. Princeton University Press, Princeton (1956)
8. Kolmogorov, A.N.: Probl. Inf. Transm. **3**, 3–7 (1969)
9. Zvonkin, A.K., Levin, L.A.: Usp. Math. Nauk **25**(6), 85–127 (1970)
10. Martin-Löf, P.: Inf. Control **9**, 602–619 (1966)
11. Martin-Löf, P.: Algorithms and Random Sequences. Erlangen University Press, Erlangen (1966)
12. Church, A.: Bull. Am. Math. Soc. **46**(N2), 130 (1940)
13. Kolmogorov, A.N.: Sankhya A **25**(N4), 369 (1963)
14. Loveland, D.: Ztrschr. Math. Log. und Grundl. Math. Bd. **12**(4), 279 (1966)
15. Loveland, D.: Trans. Am. Math. Soc. **125**(N3), 497 (1966)
16. Schen, A.: Semiotics and Information, vol. 18, p. 14. VINITI, Moskva (1982) (in Russian)
17. Brannan, D.A.: A First Course in Mathematical Analysis, p. 459. Cambridge University Press, Cambridge (2006)
18. Dieudonné, J.: Foundations of Modern Analysis. Academic, New York/London (1960)
19. Alekseyev, V.M.: Symbolic Dynamics. 11th Summer Mathematical School. Mathematics Institute of the Ukrainian Academy of Science, Kiev (1976) (in Russian)
20. Bowen, R.: Lecture Notes in Mathematics, vol. 470. Springer, Berlin/Heidelberg/New York (1975)
21. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. I. Wiley, New York (1957)
22. Dynkin, E.B.: Markov Processes. Academic, New York (1965)

Chapter 3
Main Features of Chaotic Systems

When investigating non-linear dynamical systems it is important to determine their character of motion: whether the behavior of the system is regular or chaotic. Methods of determining the type of motion and introducing quantitative characteristics of the chaoticity measure are based on different fundamental features of chaotic regimes. The following will discuss the basic signatures, or manifestations, of chaotic regimes in non-linear systems.

3.1 Poincaré Sections

Researchers of dynamic chaos quite often use the method proposed by Jules Henri Poincaré (1854–1912), now known as the Poincaré section. Poincaré used it to analyze the evolution of trajectories in a multi-body problem in the presence of gravitational interaction [1–4]. This method is important because it allows one to establish a connection between continuous dynamical systems and discrete mappings. In many cases it is possible to transfer the well-established facts obtained by the analysis of mappings onto continuous systems. The Poincaré section is also one of the ways to illustrate the behavior of dynamical systems with a phase space which has a dimensionality of $D > 2$. This method permits the detection of phase space regions with both the regular behavior of trajectories and those with chaotic behavior. The principle of the method is very simple.

Let us imagine the trajectories of a continuous dynamical system in three-dimensional phase space (see Fig. 3.1). We place a plane there in such way that the trajectories of the system intersect it transversally, i.e., not touching it. Now we will follow the intersection points of a chosen trajectory with the plane. Usually one accounts only for those intersection points that pass in a particular direction: for example, from the bottom up, as in Fig. 3.1. Thus, the temporal evolution of the trajectory uniquely determines the occurrence of the intersection points in discrete

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_3

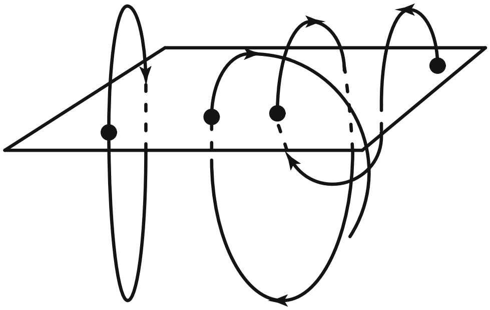


Fig. 3.1 Poincaré sections. *Circles* show the intersections with the chosen plane

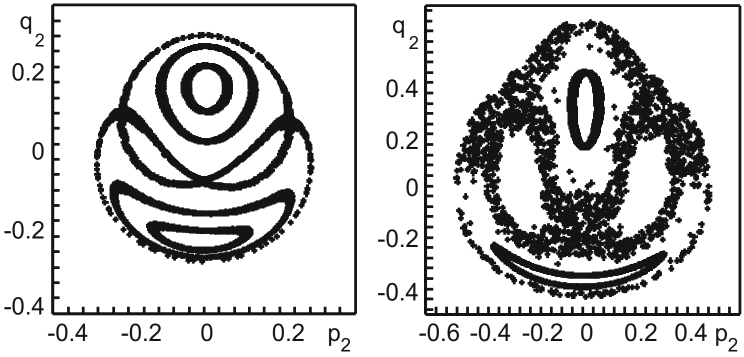


Fig. 3.2 An example of Poincaré section of Hamiltonian system with the Hénon–Heiles Hamiltonian for two different energy levels. On the *left*—$H = 1/24$, on the *right*—$H = 1/8$. Quasi-periodic orbits are well visible. On the *right* one can see a region with chaotic trajectory behavior

time and also determines the Poincaré self-mapping of the two-dimensional plane. Periodic cycles of the Poincaré mapping also correspond to periodic orbits. Chaotic orbits will "fill" whole regions of the chosen plane or in the Poincaré mapping.

Example 3.1 Figure 3.2 presents the Poincaré sections for the well-known Hamiltonian system with the Hénon–Heiles Hamiltonian [5]

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} + \frac{q_1^2}{2} + \frac{q_2^2}{2} + q_1^2 q_2 - \frac{q_2^3}{3}.$$

At a fixed energy level three-dimensional manifold of constant energy is achieved. The intersections of the trajectories on the manifold with the plane (q_2, p_2) give the desired Poincaré section. These two Poincaré sections differ in the selection of different energy levels in the system. It is evident that at energy value $H = 1/24$ the Poincaré section corresponds to regular motion. The entire plane is occupied by invariant "curves," which correspond to quasi-periodic, or to periodic with a long period trajectories of the original system. Periodic trajectories with a short period correspond to cycles in the Poincaré section that consist of small number of points. These are usually not presented because of their low illustrative potential. However, such trajectories are always present. The word “curves” is not to be taken literally, because they consist of finite number of points and therefore are not real curves in that sense.

At $H = 1/8$ in the phase space there is a region of chaotic motion. In Fig. 3.2 on the right one can see two invariant curves. Other points, filling whole region of the plane, are generated by one trajectory, which can be called chaotic.

Of course, in spite of its apparent simplicity, the construction of an explicit Poincaré mapping for a concrete dynamical system is actually a very complicated and analytically unsolvable problem [6]. Usually this method is efficiently used only in the numerical analysis of dynamical systems, where the choice of a Poincaré section is relatively easy to make. Nevertheless, the very possibility of establishing a connection between continuous dynamical systems and discrete mappings is itself of vital importance.

3.2 Spectral Density and Correlation Functions

The visual characteristic which permits the evaluation of the character of motion in dynamical systems is spectral density or spectrum. This characteristic is often measured and used in experimental research. There are important reasons for this, related to its simplicity of measurement. Moreover, its origin has its roots in optics and radio-physics. Spectral density allows for easy differentiation between periodic, quasi-periodic, and chaotic regimes. Let us introduce the spectral density $S(\omega)$ for an arbitrary dynamical system. A trajectory of the system is defined as $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$. We will choose one of the components $x_i(t)$ and meanwhile disregard the indices.

The concept of spectrum or spectral density is based on the idea that periodic functions are designed simply while chaotic functions are complicated. However, this requires clarification. In what sense are the periodic functions simple and how can this simplicity be characterized? This is an important distinction because within a period, a function can become very complicated. In order to understand the relative simplicity of periodic functions let us discuss their Fourier transformation. Let $x(t)$ be a periodic function with period T. Such functions are decomposed in the Fourier series

$$x(t) = \sum_{n=-\infty}^{+\infty} X(\omega_n)\, e^{i\omega_n t}\,, \tag{3.1}$$

where $\omega_n = 2\pi n/T$, and the coefficients $X(\omega_n)$ are determined, provided $\int\limits_0^T |x(t)|dt < 0$, by integrals

$$X(\omega_n) = \frac{1}{T} \int\limits_{-T/2}^{T/2} x(t)\, e^{-i\omega_n t} dt\,. \tag{3.2}$$

The coefficients of the decomposition $X(\omega_n)$, of course, contain all information about the function $x(t)$. Quantity $x(t)$ is a real function, and $X(\omega_n)$, in general case, complex. Therefore one should separate the information about the amplitude of the oscillations from the information about the phases of those oscillations. It is easy to do introducing the amplitude spectrum and the phase spectrum. To that end, we represent the complex function $X(\omega_n)$ in the form

$$X(\omega_n) = |X(\omega_n)|\, e^{i\theta(\omega_n)}\,.$$

Here $|X(\omega_n)|$ denotes the module of the complex function $X(\omega_n)$, which is equal by definition $|X(\omega_n)| = (X(\omega_n) \cdot X^*(\omega_n))^{1/2}$. Asterisk $*$ denotes complex conjugation. Function $\theta(\omega_n)$ is the phase of the function $X(\omega_n)$. Now the decomposition (3.1) can be written as

$$x(t) = \sum_{n=-\infty}^{+\infty} |X(\omega_n)|\, e^{i(\omega_n + \theta(\omega_n))}\,. \tag{3.3}$$

Function $|X(\omega_n)|$ is called the amplitude spectrum, and $\theta(\omega_n)$ the phase spectrum. For periodic functions, the amplitude spectrum has line structure. In other words, the amplitude spectrum for periodic functions is non-zero only at discrete frequency values $\omega = \omega_n$. It is essentially the property that characterizes the simplicity of periodic functions structure. Let us generalize these considerations on generic functions; this can be done in multiple ways, such as a method often used in physics, the tendency $T \to \infty$. Its meaning is easily understood if one considers the example of oscillatory motions in a box with periodic boundary conditions, and the tendency of its dimensions to infinity. Returning to definition (3.1), (3.2), let us note that

$$\frac{2\pi}{T} = \omega_{n+1} - \omega_n \equiv \Delta\omega\,.$$

At tendency $T \to \infty$, $\Delta\omega \to 0$. For arbitrary function $x(t)$ let us rewrite the expression (3.1) in slightly modified form

$$x(t) = \sum_{n=-\infty}^{+\infty} \frac{1}{2\pi} \int\limits_{-T/2}^{T/2} x(t')\; e^{-i\omega_n t'} dt' e^{i\omega_n t} \Delta\omega \tag{3.4}$$

Defining

$$X_T(\omega) = \frac{1}{2\pi} \int_{-T/2}^{T/2} x(t)\ e^{-i\omega_n t} dt\,,$$

it is easy to understand that the sum at $T \to \infty$ and, therefore, $\Delta\omega \to 0$ transforms to an integral and the expression (3.4) can be rewritten as

$$x(t) = \int_{-\infty}^{+\infty} X(\omega)\ e^{i\omega t} d\omega\,. \tag{3.5}$$

where $X(\omega) = \lim\limits_{T\to\infty} X_T(\omega)$. Therefore, a periodic function can be represented as a Fourier integral under the condition $\int\limits_{-\infty}^{+\infty} |x(t)|\, dt < \infty$. Now we introduce amplitude spectral density essentially repeating the proceeding considerations

$$x(t) = \int_{-\infty}^{+\infty} |X(\omega)|\ e^{i(\omega t + \theta(\omega))} d\omega\,.$$

Hence it is easy to see that the quantity $|X(\omega)|d\omega$ characterizes the amplitude of oscillations on the interval $[\omega, \omega + d\omega]$, which forms the quantity $x(t)$. Therefore $|X(\omega)|$ can be called amplitude spectral density. Of course, if the magnitude $x(t)$ is neither periodic nor quasi-periodic, then its spectral density is the continuous function of frequency and, as is customary to say, has continuous spectrum. Instead, physicists often use the square of the spectral density

$$S(\omega) = |X(\omega)|^2\,.$$

This quantity also makes simple physical sense. Its elegant interpretation is based on state filling numbers from quantum mechanics or quantum field theory. Avoiding a digression into that field, a different explanation will follow. Beforehand, however, it is necessary to discuss another characteristic, which distinguishes chaotic trajectories (or signals) from regular ones. This characteristic is used in statistical physics as a measure of statistical independence of two quantities in different temporal or spatial points. It is a correlation function. In the simplest variations, the correlation function is defined as

$$B_{ij}(\tau) = \langle x_i(t+\tau)x_j(t)\rangle - \langle x_i\rangle\langle x_j\rangle\,.$$

Angle brackets denote time averaging, which is defined in the usual way as

$$B_{ij}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int\limits_{-\frac{T}{2}}^{\frac{T}{2}} x_i(t+\tau)x_j(t)dt\,.$$

Henceforth we will assume for simplicity that the averaged values $\langle x_i \rangle = 0$. Strictly speaking, the quantities $B_{ij}(\tau)$ form a correlation tensor, with indexes $i, j = 1, \ldots, n$. As points move one away from another, it is expected for chaotic signals that the statistical connections between them decrease. In terms of the correlation function it means that the average $\langle x_i(t+\tau)x_j(t)\rangle$ for very distant and statistically unconnected points tends to $\langle x_i \rangle \langle x_j \rangle$. In physics, this is termed the decay of correlations. In other words, the correlation function for statistically unconnected time points turns to zero. It allows chaotic trajectories to be determined by the drop of the correlation function with growth of τ. Essentially, this reflects the intuitively clear property of chaos as the loss of the dependence of quantity values $x(t+\tau)$ on their values $x(t)$ in the preceding moment in time. Now we calculate the correlation function for the periodic $x_i(t+T) = x_i(t)$. Omitting simple transformations, we give the result

$$B_{ij}(\tau) = \sum_{n=-\infty}^{\infty} X_i(\omega_n)X_j^*(\omega_n)e^{(i\omega_n\tau)}\,.$$

Analogous to the examples above, for aperiodic functions, sum transforms into integral

$$B_{ij}(\tau) = \int\limits_{-\infty}^{\infty} X_i(\omega)X_j^*(\omega)e^{(i\omega\tau)}d\omega.$$

For $i = j$ the correlation functions are also called autocorrelation functions. Taking a Fourier transform of the autocorrelation function on τ it is easy to obtain that

$$B_{ii}(\tau) = \int\limits_{-\infty}^{\infty} |X_i(\omega)|^2 e^{i\omega\tau}d\omega\,. \tag{3.6}$$

This can be understood recalling that the Fourier transform of a convolution coincides with the product of their Fourier transforms. Now it is clear that the autocorrelation functions are closely connected with the spectral density $S(\omega)$.

$$S(\omega) = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} B(\tau)e^{-i\omega\tau}d\tau\,.$$

This connection also determines the physical sense of spectral density $S(\omega)$. Indeed, if quantity $x(t)$ is the velocity of motion, then $B_{ii}(0)$ is proportional to kinetic energy and, as follows from the relation (3.6), quantity $|X_i(\omega)|^2$ represents the energy density contained in oscillations with the frequencies that belong to the interval $[\omega, \omega + d\omega]$. Therefore, such spectral density is called the energy spectrum and it characterizes the quantity of energy concentrated in oscillations with the frequency ω. Thus, we can use spectral densities and correlation functions as criteria for chaoticity. Spectral densities for chaotic trajectories or signals have continuous spectra, and for periodic or quasi-periodic—linear ones (see Figs. 3.3 and 3.4). Generally speaking, correlation functions for chaotic signals fall down exponentially with growth of τ.

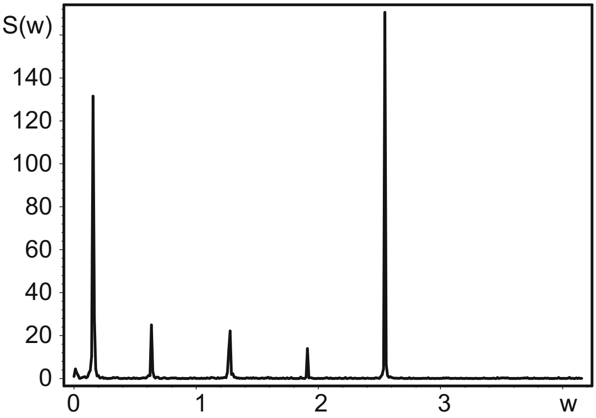


Fig. 3.3 An example of typical line spectrum

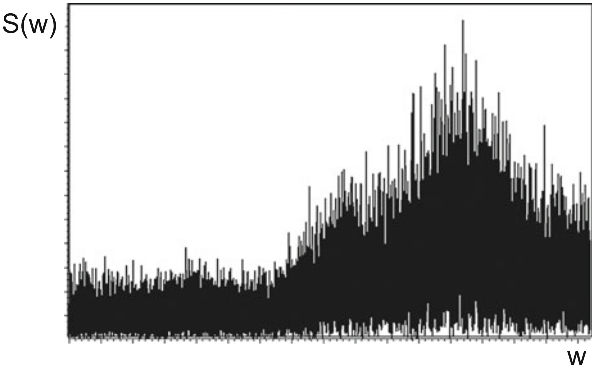


Fig. 3.4 A typical example of the continuous spectrum of chaotic signal

3.3 Lyapunov Exponent

One of the commonly used methods of determining the degree of chaoticity is based on the properties of exponential divergence in chaotic trajectories. This essential quality has already been discussed. The exponential sensibility of dynamical systems with chaotic behavior found realization in the determination of the Lyapunov exponent. This is one of the most popular characteristics or criteria of dynamical chaos. For one-dimensional systems with discrete time, or cascades, the Lyapunov exponent was actually introduced before as a characteristic of the divergence of two trajectories close at $t = 0$. For continuous one-dimensional systems it is analogously defined. Let the initial points of two trajectories at $t = 0$ be situated in distance $|\delta x(0)|$ one from another. If the distance between these points $|\delta x(t)|$ changes with time (at $t \gg 1$) as

$$|\delta x(t)| \sim |\delta x(0)| \, e^{\lambda t} ,$$

then λ is called the Lyapunov exponent. It is clear that in order to find λ we can use the relation

$$\lambda = \frac{1}{t} \ln \frac{|\delta x(t)|}{|\delta x(0)|} , \quad t \to \infty . \tag{3.7}$$

In accordance with above discussed sensitivity of dynamical systems, at $\lambda > 0$ errors in determination of initial conditions will lead to the impossibility of predicting system behavior. For generic one-dimensional mappings

$$x_{n+1} = f(x_n) , \tag{3.8}$$

the Lyapunov exponent can also be easily introduced. Let the distance between two points equal δx_n on the nth time step, i.e., let the points correspond to positions x_n and $x_n + \delta x_n$. It is easy to calculate the distance between them in the moment $n + 1$. Indeed,

$$x_{n+1} + \delta x_{n+1} = f(x_n + \delta x_n) . \tag{3.9}$$

Assuming them close in the moment of time n, we decompose the right-hand side in the Taylor series over δx_n

$$x_{n+1} + \delta x_{n+1} = f(x_n) + \delta x_n \cdot f'(x_n) + \cdots .$$

Taking into account the mapping (3.8), and neglecting the non-linear corrections, we obtain

$$\delta x_{n+1} = \delta x_n \cdot f'(x_n) . \tag{3.10}$$

This linear relation clearly states that variation of distance in one iteration step is determined by the derivative value in the preceding point. Using this simple relation it is possible to express the distance δx_0 and values of $f'(x)$ along the trajectory

$$\delta x_{n+1} = \delta x_0 \prod_{k=0}^{n} f'(x_k) \ . \tag{3.11}$$

Then the Lyapunov exponent, which characterizes the divergence of the trajectories $|\delta x_n| = |\delta x_0| \ e^{\lambda n}$, is defined as

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{k=o}^{n-1} \ln \left| f'(x_k) \right| \ . \tag{3.12}$$

There are several modifications of this simple formula as well as numerical methods for determining λ [6, 7]. However, we will discuss below how one can use the Lyapunov exponent for dynamical systems in multidimensional phase space. Here the situation is less inconsequential than in one-dimensional cases. If we are interested in the deviation of trajectories from a chosen one in two-dimensional space, we consider a neighborhood in the phase space which contains a point of that trajectory. All other points in the neighborhood correspond to other trajectories, which are close to the chosen one. Let the neighborhood be a circle of radius ε. It is clear that if all the points in the neighborhood are shifted according to dynamical law, then after the time period Δt the neighborhood will be shifted in the phase space and most importantly, it will change its shape. In other words, some trajectories will approach the chosen one while others will move away from it. These changes in distance during the time period Δt will depend on the trajectory selected (see Fig. 3.5). We will demonstrate how to determine the Lyapunov exponent for mappings in two-dimensional phase space. Generally, such mappings have the following form:

$$\begin{aligned} x_{n+1} &= f(x_n, y_n) \\ y_{n+1} &= g(x_n, y_n) \end{aligned} \ . \tag{3.13}$$

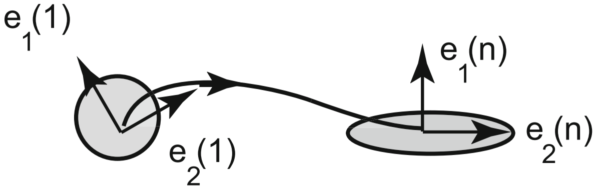


Fig. 3.5 Here the trajectory, eigenvectors and the initial neighborhood surrounding it are shown. The positions of the neighborhood and the eigenvectors at the moment of time n qualitatively demonstrate their temporal behavior

The position of trajectory in the moment n is conveniently characterized by the vector or the column

$$z_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} .$$

Let us now consider a vector connecting two close trajectories z_n and $z_n + \delta z_n$, where

$$\delta z_n = \begin{pmatrix} \delta x_n \\ \delta y_n \end{pmatrix} .$$

Using the dynamical law (3.13) and the smallness of δx_n and δy_n, we can establish, as in the one-dimensional example, how δz_n is changed in one time step. It is easy to obtain that

$$\delta z_{n+1} = A_n \cdot \delta z_n , \tag{3.14}$$

where A_n is the square 2 × 2 matrix of the following form:

$$A_n = \begin{pmatrix} \frac{\partial f(x_n y_n)}{\partial x_n} & \frac{\partial f(x_n y_n)}{\partial y_n} \\ \frac{\partial g(x_n y_n)}{\partial x_n} & \frac{\partial g(x_n y_n)}{\partial y_n} \end{pmatrix} . \tag{3.15}$$

This is usually called the Jacobian matrix. From Eq. (3.14) it is easy to determine the deviation for n steps

$$\delta z_n = A_{n-1} A_{n-2} \cdots A_0 \cdot \delta z_0 \equiv Q_n \delta z_0 . \tag{3.16}$$

Thus the deviation for n time steps is determined by the initial deviation multiplied by the matrix Q_n, depending on n and on the derivatives along the trajectory of the system. As a natural basis convenient for the decomposition of any vector, eigenvectors of the matrix Q_n are usually used:

$$Q_n \; e_i\,(n) = \nu_i\,(n) \;\; e_i\,(n) \; , \tag{3.17}$$

where index i labels the eigenvectors, and in two-dimensional cases $i = 1, 2$. The eigenvectors depend on n and can be chosen to be orthonormal, i.e., $e_1 \cdot e_2 = 0$, $e_1 \cdot e_1 = e_2 \cdot e_2 = 1$. Quantities $\nu_i\,(n)$ are called the eigen-numbers of the matrix Q_n and they also depend on time n. The conditions of solvability for the system of linear equations (3.17) lead to the characteristic equation for eigenvalues

$$\det |Q_n - \nu\,(n) \;\; \mathrm{I}| = 0 . \tag{3.18}$$

Here for convenience we introduce the unity matrix

$$\mathrm{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

Solutions of Eq. (3.18) determine the eigenvalues of the matrix Q_n. Any vector or deviation can be now conveniently written in the basis of eigenvectors $e_i(n)$. Initial deviation in that basis takes the form

$$\delta z_0 = \sum_i a_i \, e_i(n) \ , \tag{3.19}$$

where a_i are coordinates of the initial deviation in the basis (e_1, e_2). Then the deviation on the nth time step is easily obtained from (3.16) in the new basis

$$\delta z_n = Q_n \sum_i a_i \, e_i(n) = \sum a_i \, \nu_i(n) \ e_i(n) \ . \tag{3.20}$$

Thus, $\nu_i(n)$ characterize the variation of the deviation coordinates (in the new basis) along ith direction, determined by the eigenvector $e_i(n)$.

Let us now introduce the Lyapunov exponents along those directions, i.e.,

$$\nu_i(n) \sim e^{\lambda_i n}$$

at sufficiently large n. Then they are found to be

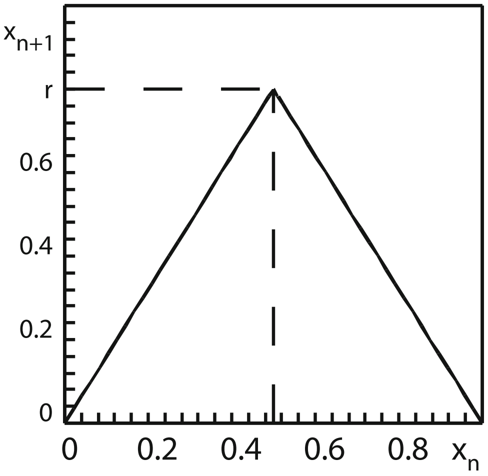
$$\lambda_i = \lim_{n \to \infty} \frac{\ln \nu_i(n)}{n} .$$

As in two-dimensional cases, there are two directions determined by e_i, and two Lyapunov exponents appear; these can be positive as well as negative. Let us discuss the limitations on these exponents that arise in dynamical systems. For that purpose we consider a small rectangular neighborhood with sides that coincide with the vectors $a_1 e_1$ and $a_2 e_2$ [see (3.19)]. The initial area of the rectangle is $\omega_0 = a_1 a_2$. Now let us consider how its area changes after n iterations according to dynamical law (3.13) or (3.14). Using (3.20), we obtain

$$\omega_n = a_1 \nu_1 \cdot a_2 \nu_2 = e^{n(\lambda_1 + \lambda_2)} \omega_0 .$$

It is easy to see that for conservative mappings, which preserve the phase volume, $\lambda_1 + \lambda_2 = 0$. In such systems, one of the Lyapunov exponents is positive and the other is negative: $\lambda_2 = -\lambda_1$. Thus, the presence of chaos in such systems is determined by the maximum positive Lyapunov exponent. We should note that there is a divergence of close trajectories in one direction and a convergence in the other.

Fig. 3.6 Graph view of tent mapping



For dissipative dynamical mappings $\lambda_1 + \lambda_2 < 0$. Therefore, at least one Lyapunov exponent is negative.

In phase spaces of higher dimensions the situation is analogous to the two-dimensional phase space example. It is likewise possible to determine the Lyapunov exponents (their number coincides with the phase space dimensionality), and the maximum positive Lyapunov exponent will also give a quantitative characteristic of the chaoticity measure in multidimensional cases.

Example 3.2 Calculating the Lyapunov exponent for the tent mapping

$$x_{n+1} = \begin{cases} 2r x_n, & \text{if } \ 0 \le x_n \le \frac{1}{2} \\ 2r\,(1 - x_n)\,, & \text{if } \ \frac{1}{2} < x_n \le 1 \end{cases} . \tag{3.21}$$

This is a continuous non-linear mapping with the phase space [0, 1]. A graph view is presented in Fig. 3.6.

It is easy to see that $|f'(x)| = 2r$ and it does not depend on the position of x_n. Accordingly, using the definition (3.12), we obtain

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln 2r = \lim_{n\to\infty} \frac{n \ \ln 2r}{n} = \ln 2r\,.$$

Thus the Lyapunov exponent for that mapping is $\lambda = \ln 2r$. Therefore, chaotic behavior in that mapping will be observed at $\lambda > 0$ or at $2r > 1$. The transition to a chaotic regime bears a resemblance to the characteristic features of phase transitions (see Fig. 3.7). So at $r < r_c \equiv \frac{1}{2}$ the chaotic phase is absent, while at $r > r_c$ it appears.

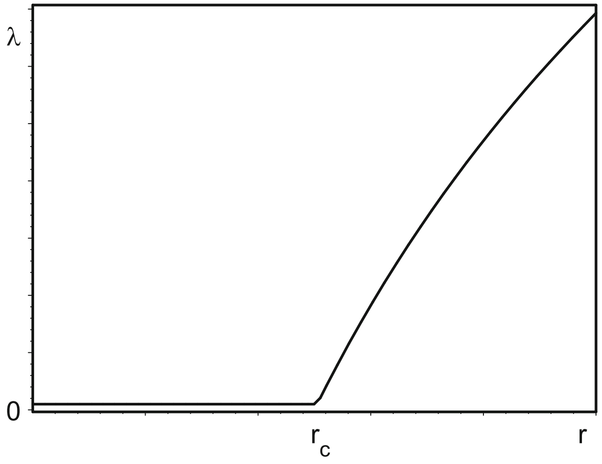


Fig. 3.7 Dependence of the order parameter for the tent mapping on the mapping parameter

If we consider the positive Lyapunov exponent as a parameter of order, then near the phase transition $r \gtrsim r_c$ it changes according to the power law

$$\lambda \sim (r - r_c)^{\beta}$$

the same way as in phase transitions. The critical index is $\beta = 1$.

It should be noted that value of the Lyapunov exponent is conserved at variable transformations in dynamical systems. This can be ascertained if one considers the mapping

$$x_{n+1} = f\,(x_n) \tag{3.22}$$

and the mapping

$$y_{n+1} = g\,(y_n)\ , \tag{3.23}$$

which are obtained by the transition to new variables $y_n = h\,(x_n)$ with function $h'\,(x) \neq 0$ in the phase space of the mapping (3.22). The Lyapunov exponent for the mapping (3.22) is determined by the following:

$$\lambda_f = \frac{1}{n}\sum_{i=0}^{n-1} \ln\left|f'\,(x_i)\right|\ . \tag{3.24}$$

For the mapping (3.23)

$$\lambda_g = \frac{1}{n} \sum_{i=0}^{n-1} \ln \left| g'(y_i) \right| . \tag{3.25}$$

Let us make a transformation of variables $y_i = h(x_i)$ in the relation (3.25). We account that

$$\frac{dy_{n+1}}{dy_n} = g'(y_n)$$
$$\frac{dx_{n+1}}{dx_n} = f'(x_n) ,$$

and $dy_n = h'(x_n)\, dx_n$. Using these relations we obtain

$$g'(y_n) = \frac{h'(x_{n+1})}{h'(x_n)} f'(x_n) .$$

Then the Lyapunov exponent (3.25) is transformed to the form

$$\lambda_g = \frac{1}{n} \sum_{i=0}^{n-1} \left[\ln \left| f'(x_i) \right| + \ln \left| \frac{h'(x_{i+1})}{h'(x_i)} \right| \right] .$$

Now it follows that

$$\lambda_g = \lambda_f + \frac{1}{n} \ln \left| \frac{h'(x_n)}{h'(x_0)} \right| .$$

At $n \to \infty$ the latter contribution disappears. Thus, the value of the Lyapunov exponent is conserved at geometrical transformations of variables.

In conclusion, let us cite without proof several useful properties of the Lyapunov exponent. First of all, considering all the Lyapunov exponents (and not only the maximum positive one), it is possible to analyze the global spectrum of the Lyapunov exponents. In systems with n degrees of freedom their number is exactly n. Values of the exponents can be ordered

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n .$$

Naturally, among these there is an exponent that characterizes the deviation along the trajectory. This exponent is always equal to zero if the trajectory does not terminate in a singular point. For Hamiltonian systems with even-dimensional phase space the Lyapunov exponents have additional symmetry. For any Lyapunov exponent λ_i there always exists an exponent $\lambda_j = -\lambda_i$. Therefore, in Hamiltonian systems at least two Lyapunov exponents turn to zero. Of course for integrable Hamiltonian systems all Lyapunov exponents are zeros.

For any dynamical system, the sum of all the Lyapunov exponents coincides with the divergence of the vector field averaged along the trajectory. This can be understood from the very meaning of the Lyapunov exponents, which characterize the divergence of trajectories along some directions, and therefore, the volume variation

$$\langle \mathrm{div}\mathbf{V} \rangle = \sum_{i=1}^{n} \lambda_i .$$

This means that for dissipative systems their sum is negative, and for conservative ones, it equals zero.

The Lyapunov exponents are important not only as a chaoticity criterion for the systems, but they can also serve as a useful tool to analyze the types of limit regimes or attractors. Omitting the one-dimensional case, let us consider as an example the attractors of two-dimensional systems. In such systems there exist only stable points and limit cycles. The Lyapunov exponents in the former case $(\lambda_1 \lambda_2) = (-,-)$ are both negative, and in the latter case $(\lambda_1 \lambda_2) = (-,0)$. In three-dimensional systems there are many more types of attractors.

- Stable node, or focus: $(\lambda_1, \lambda_2, \lambda_3) = (-,-,-)$
- Stable limit cycles: $(\lambda_1, \lambda_2, \lambda_3) = (-,-,0)$
- Stable torus: $(\lambda_1, \lambda_2, \lambda_3) = (-,0,0)$
- Strange attractor: $(\lambda_1, \lambda_2, \lambda_3) = (-,0,+)$

The latter limit regime will be discussed in the following: sections.

3.4 Invariant Measure

In dynamical systems with chaotic behavior one can try to develop a statistical theory, an important element of which is the notion of invariant density. Let us introduce a function which characterizes the density of initial conditions, or more exactly, probability density of initial conditions $P_0(x)$. The probability dw of an initial condition to fall in the interval $[x_0, x_0 + dx_0]$ is by definition $dw = P_0(x_0)\, dx_0$. If dynamics of our system is defined by the mapping

$$x_{n+1} = f(x_n) , \tag{3.26}$$

then it is possible to determine the time transformation law for the above density. In order to do that, one should use conservation probability during the evolution of our system. Then in one time step the probability transformation is determined by the equation

$$P_1(x) = \int\limits_{\Omega} P_0(y)\,\delta(x - f(y))\,dy\,. \tag{3.27}$$

It is clear that this equation will preserve its form for any time stage of the evolution. Therefore let us formulate this in a more general form

$$P_{n+1}(x) = \int\limits_{\Omega} P_n(y)\ \delta(x - f(y))\,dy\,, \tag{3.28}$$

where Ω is the phase space of the dynamical system or of the mapping (3.26), and $P_n(x)$ is the probability density on the nth time step. Equation (3.28), which describes the temporal evolution of the probability density, is called the Frobenius–Perron equation. This equation can be written in another form. Using the well-known property of δ-function

$$\delta(x - f(g)) = \sum_i \frac{\delta(y - y_i)}{|f'(y_i)|}\,, \tag{3.29}$$

where y_i is the solution of the equation $x - f(y_i) = 0$. The summation is performed over all the roots. Using the relation (3.29), we integrate in Eq. (3.28)

$$P_{n+1}(x) = \sum_i \frac{P_n(y_i(x))}{|f'(y_i(x))|}\,. \tag{3.30}$$

In fact, Eq. (3.28) is transformed into a functional equation. Another useful form of the Frobenius–Perron equation is obtained with the help of Fourier transform of δ-function and has the following form:

$$P_{n+1}(x) = \sum_i P_n(y_i)\ e^{-\lambda_i(y_i)}\,, \tag{3.31}$$

where $\lambda(y_i) = \ln|f'(y_i)|$ is the characteristic quantity, or local Lyapunov exponent. This equation demonstrates the role of chaoticity in the probability density evolution.

In the evolutionary process the distribution function $P_n(x)$ can come to a stationary state, the role of which is similar to equilibrium states in statistical physics. Then the dependence on time n disappears and $P_n(x) = P(x)$. This stationary probability density is called invariant and it determined by the equation

$$P(x) = \int\limits_{\Omega} P(y)\ \delta(x - f(y))\ dy\,. \tag{3.32}$$

In other words, the invariant distribution function does not change at phase space transformations by the phase flow (or by the mapping) (3.26). Naturally, it satisfies (as well as $P_n(x)$) the normalization condition

$$\int_{\Omega} P(x)\ dx = 1\,.$$

It should be stressed that the introduction of the invariant density is based on the relaxation of the initial density in relation to the invariant one. There arises an important question: in what systems, or more precisely, in systems with what properties will such relaxation take place? Discussion of this question requires that we touch on the question of the foundations of statistical mechanics, ergodicity, and measure theory [8]. We will return to them in another section. Besides this, introduction of the distribution density is not justified for all dynamical laws. There are situations when it is impossible to extract the density from the measure, and in such dynamical systems singular or multi-fractal measures appear. The examples will be considered further.

Example 3.3 Let us now consider a simple example of invariant density calculation for tent mapping. The functional Frobenius–Perron equation (3.30) in that case takes the form

$$P(x) = \frac{1}{2r}\left(P\left(\frac{x}{2r}\right) + P\left(1 - \frac{x}{2r}\right)\right)\,.$$

It is easy to find a particular solution of that equation $P(x) = 1$. This solution is realized at $r = 1$ and it corresponds to uniform distribution. At other values of $r > 1$ the situation, as well as the invariant density form, is less insignificant.

For dynamical systems under conditions of ergodicity the Lyapunov exponent can be calculated from the invariant density distribution in the following way:

$$\lambda = \int_{\Omega} \rho(x)\ \ln\left|f'(x)\right| dx \equiv \left\langle \ln\left|f'\right|\right\rangle\,.$$

The invariant density permits us to calculate also the mean-square fluctuations λ

$$\Delta\lambda^2 \simeq \left\langle\left(\ln\left|f'\right| - \left\langle \ln\left|f'\right|\right\rangle\right)^2\right\rangle\,.$$

In some cases this is sufficiently efficient.

3.5 Entropy of Dynamic Systems

At first glance, the relation of entropy to chaos does not require discussion. However, several different definitions in physics are associated with entropy. In fact, Kolmogorov introduced a new value called entropy of dynamical systems [9, 10]. For many reasons, this characteristic of dynamical systems is very important. However, first we have to define the entropy of a dynamical system. Let us clarify what is understood by a dynamical system in ergodic theory [5] since the definition of the entropy first appeared there. The dynamic system (M, μ, U_t) includes the phase space M, which is measurable and has the measure μ, invariant with respect to the mapping (or one-parameter family of mappings U_t) $U : M \to M$. The invariance of the measure means that $\mu(A) = \mu(U_t A)$ for all the measurable subsets A of the space M. It is generally considered that $\mu(M) = 1$. Strictly speaking, the definition is much more general, but this simplified variant will suffice for understanding. Let us now define the conditional entropy of partitions. Partition α of phase space M should be understood as a finite or countable set of the non-intersecting subsetsA_i, with $M = \bigcup_i A_i$ and $A_i \bigcap A_j = \emptyset$ at $i \neq j$. It is clear that other partitions M can exist, e.g., $\beta = \{B_i\}$. In the set of partitions, one can introduce natural ordering. We understand $\alpha \leq \beta$ so that for every $B \subset \beta$, there is $A \subset \alpha$ with $B \subset A$. In addition, we introduce the product of partitions $\alpha \vee \beta$. This is a new partition with various sets $A_i \bigcap B_j$ as its elements.

Now let us introduce the conditional entropy of the partition α with respect to the other partition β as

$$H(\alpha|\beta) = -\sum_i \mu(B_i) \sum_j \mu(A_j|B_i) \ln \left(\mu(A_j|B_i)\right)$$

where $\mu(A_j|B_i) = \mu(A_j \bigcap B_i)/\mu(B_i)$. The entropy is a positive number, which follows from positive definiteness of the measure and from the properties of the function $-x \ln x$ (see Fig. 3.8).

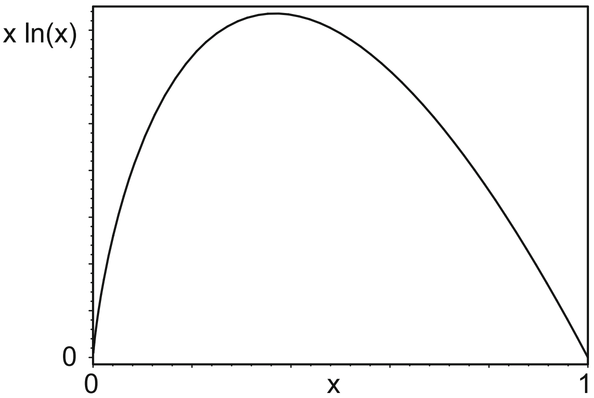


Fig. 3.8 The plot of the function $-x \ln x$ in the domain [0, 1]. It should be noted that this function in zero is by definition zero ($0 \ln 0 = 0$)

If as a partition β we choose a partition which consists of one element M, then we obtain the entropy definition for the partition α

$$H(\alpha) = -\sum_j \mu(A_j) \ln \left(\mu(A_j)\right).$$

Conditional entropy and partition entropy have a number of properties which can be checked.

- $H(\alpha|\beta) \geq 0$ $H(\alpha) \geq 0$
- $H(\alpha \vee \beta|\beta) = H(\alpha|\beta)$
- $H(\alpha|\beta) = 0$ if $\alpha \leq \beta$ and only in this case
- $H(\alpha \vee \beta) = H(\alpha|\beta) + H(\beta)$
- If $\alpha \leq \beta$, then $H(\alpha) \leq H(\beta)$
- If $\alpha \leq \gamma$, then $H(\alpha|\beta) \leq H(\gamma|\beta)$
- $H(\alpha \vee \gamma|\beta) \leq H(\alpha|\beta) + H(\gamma|\beta)$
- If $\beta \leq \gamma$, then $H(\alpha|\gamma) \leq H(\alpha|\beta)$
- $H(\alpha \vee \gamma|\beta) \leq H(\alpha|\beta) + H(\gamma \vee \beta)$
- $H(\alpha \vee \beta \vee \gamma) + H(\gamma) \leq H(\alpha \vee \gamma) + H(\beta \vee \gamma)$

Example 3.4 We consider the entropy of the partition α into a finite number of sets N of the same measure $\mu(A_j) = 1/N$. We obtain that $H(\alpha) = \ln N$. In the general case, when measures A_j do not coincide, $H(\alpha) \leq \ln N$.

Next we use the presence of mapping and define the entropy for one step over time:

$$\begin{aligned} h(\alpha, U) &= \lim_{n\to\infty} \frac{1}{n} H(\alpha \vee U\alpha \vee U^2\alpha \vee \ldots \vee U^n\alpha) \\ &= \lim_{n\to\infty} \frac{1}{n} H(U^n\alpha|\alpha \vee U\alpha \vee U^2\alpha \vee \ldots \vee U^{n-1}\alpha). \end{aligned}$$

Then the entropy of the dynamic system (M, μ, U) is the value

$$h(U) = \sup_{\alpha} h(\alpha, U),$$

where the upper boundary is taken for all partitions α of the phase space M. Though this definition is rigorous, it is complicated to use it to calculate the entropy of concrete dynamical systems. The properties discovered by Kolmogorov are often applied to calculate entropy. To formulate them, one needs the concept of the generating partition of the phase space with respect to mapping U. In a sense, that is the completeness property of the partitions $\bigvee_{-\infty}^{\infty} U^i\alpha$. The exact definition of the generating partition for the mapping U is the measurable, countable partition α, so that partition $\bigvee_{-\infty}^{\infty} U^i\alpha$ built on it generates all the σ algebra of mappings M. For the generating partitions α of the mapping U it is proved (e.g., see [5]) that

$$h(U) = h(\alpha, U).$$

From this we can understand the physical meaning of the entropy of dynamical systems. Let us recall that in statistical physics entropy means the logarithm of state numbers with the degree of freedom tending to infinity. From this viewpoint, the entropy of dynamical systems can be understood as logarithm of system trajectory type numbers with time tending to infinity. From the geometric point of view entropy can be interpreted in a different way. Let us consider a dynamical system with two-dimensional phase space and with the invariant measure coinciding with the area. We choose some element A of the partitions α, bounded with, for example, a circle. It is clear that the mapping U will transform it into UA of partition $U\alpha$. Due to the invariance of the measure, $\mu(A) = \mu(UA)$, but the form of the boundary changes. For mappings with positive entropy, the boundary becomes complicated and sinuous. Now, if we roughen the boundary, then the area of the roughened region increases (see Fig. 3.9).

In discrete, as well as in continuous time, this change is described with the function

$$S \sim e^{h \cdot t}$$

If entropy is equal to zero, then growth can be absent or take place not in an exponential way. Taking into account Bouen's results [11], we can use another definition of topological entropy of dynamical systems. Let the phase space be a compact metric space with the metric ρ, and the mapping $U : M \to M$ is continuous. The subset $K \subset M$ is called (n, ε) separated, provided that for any $x, y \in K$, $x \neq y$ one can find i $0 \leq i \leq n$, so that $\rho(U^i x, U^i y) > \varepsilon$. Let us denote with $N_n(\varepsilon)$ the largest number of elements in (n, ε) separated set M. Then we define the topological entropy as follows:

$$h(U) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \ln N_n(\varepsilon)$$

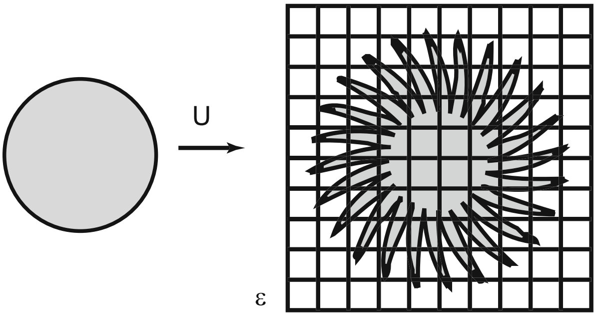


Fig. 3.9 This figure symbolically shows the evolution of the circle under the influence of mapping. The area is covered by the grid of ε scale. It is easy to note the increase of the area while roughening, if by the roughened area, we mean the area of the cells containing the points of the circle image

This definition clearly reflects the fact that the topological entropy coincides with the asymptotic of the number of various trajectories of mapping U, which differ from the given arbitrary accuracy. Roughly speaking, the fewer there are different types of trajectories, the smaller the entropy of the dynamical system, and the simpler the system. One can conclude from this geometric meaning that a relation can exist between the entropy of dynamic systems and the positive Lyapunov indexes

$$h = \sum_i \lambda_i$$

here the summation is made over i, for which $\lambda_i > 0$. For some systems it is proved.

Entropy of dynamical systems is important for two main reasons. First, dynamical systems with positive entropy have strong statistical properties. Second, the same dynamical systems (or, to be more precise, isomorphic systems) have the same topological entropy. This means that all dynamical systems can be divided into classes with the same entropy. However, one should remember that entropy is not the only topological invariant of the dynamical systems.

Theorem 3.5 *Entropy is an invariant of the dynamical system.*

To calculate the entropy of concrete dynamical systems is a rather complicated problem. There is a set of examples of dynamical systems, for which topological entropy is found.

Example 3.6 The entropy for mapping of torus into torus is known. The simplest way to specify the mapping of two-dimensional torus is using an integer numbers matrix.

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Its determinant is equal to 1. It is proven that if λ is the eigenvalue of this matrix in modulus less than one, then

$$h = -\ln \lambda .$$

Actually, this example shows the existence of the dynamical systems with the positive entropy. To calculate the topological entropy, one can use invariant kneading.

3.6 Kneading Invariants

The theory of kneading invariants was established by a remarkable mathematician Milnor [12]. For simplicity, let us consider it for the case of one-dimensional mappings. Let the dynamics of our system be defined by the mapping of

$$x_{n+1} = f(x_n).$$

Mapping f is piecewise monotonous. Let us introduce the position function of the initial point x, taking two values depending on the behavior of the function f^k in the point k of the orbit.

$$\Theta_k(x) = \begin{cases} +1, & \text{if } f^k \text{ increasing at } x, \\ -1, & \text{if } f^k \text{ decreasing at } x. \end{cases} \tag{3.33}$$

Here we use the designation for k multiple composition of functions $f^k = f^k(x) = \underbrace{f \circ f \circ \cdots \circ f}_{k}$, where $f \circ f \equiv f(f(x))$, we consider $f^0(x) = x$ by definition. Obviously, we can define this function in extremum points $x = x_{\mathrm{ext}}$, as, for instance, zero. Using this function of every trajectory of the mapping, one can put one-to-one corresponding sequence:

$$\Theta_f(x) = (\Theta_0(x), \Theta_1(x), \Theta_2(x), \ldots,).$$

In fact, this sequence describes the change of orientation when the motion is along the trajectory. At first, such a sequence ± 1 seems quite trivial. However, recall that every real number can be written as a sequence in a two-symbol alphabet. We have already made use of this when studying chaos in simple dynamic systems. The comparison of each trajectory of such a sequence allows us to transfer the properties of the sequences on the trajectories. For example, it is possible to introduce a lexicographical ordering of the trajectories. The sequences written in the alphabet $A = (a_1, a_2, \ldots, a_m)$ satisfy the lexicographical ordering $\Theta < \Theta'$, if $\Theta_i = \Theta'_i$ with $i < n$ and the symbol Θ_n appears in the alphabet A earlier than Θ'_n. It turns out that one can prove that the natural order of real numbers and the lexicographical order of the sequences are related.

Theorem 3.7 *Mapping $x \to \Theta_f(x)$ is monotonous.*

The character of the monotonicity depends on the types of extrema of the function f. Let us now introduce a kneading series. For every point x, it compares the function

$$Q(x) = \Theta_0(x) + \Theta_1(x)t + \cdots + \Theta_k(x)t^k + \cdots$$

The property of the monotonicity of the sequences Θ_f is applied to these functions. So the kneading series are the monotonous functions of x. The type of monotonicity (increase or decrease) depends on the type of the extremum. In the extremum points the kneading series can have discontinuities. Due to their monotonicity, the kneading series possess natural ordering, which allows us to consider them as a kind of natural universal coordinates. Their universality manifests itself in invariance (or independence) at any continuous coordinate changes. This means that their nature is purely topological. Even more important is that the dynamics of the mapping is defined by a finite number of values of the kneading series. Indeed, if the mapping

f has one extremum at $x = x_{\mathrm{ext}}$, then we can introduce:

$$Q_- = \lim_{x \to x_{\mathrm{ext}} - 0} Q(x)$$

$$Q_+ = \lim_{x \to x_{\mathrm{ext}} + 0} Q(x).$$

These are invariants of the continuous mappings: $Q_+ = -Q_-$. For two mappings f, g one can prove that if $Q_-(f) = Q_-(g)$, then there is such h that $hf = gh$. So, there is an isomorphism between the mappings. The series Q_- have an important property, which distinguishes them from an arbitrary series. This property is called ν permissibility. The formal power series is called ν permissible if for every $n \geq 0$, $|\sigma^n(\Theta)| \geq \nu$ or $\sigma^n(\Theta) = 0$. Here the designation σ is introduced for the shift along the sequence, and $|a| = a_0 \cdot a$, where the point $\cdot$ designates the term by term multiplication of the sequence $a = a_0, a_1, a_2, \ldots$ by a number. Naturally, the permissibility property is also defined for the sequences. For instance, the sequence $a = a_0, a_1, a_2, \ldots$ is permissible if $a_0 = 1$, $a_i = \pm 1$ and for any i and $n \geq 0$, $|\sigma^n a| \geq a$ is true. The importance of this class of sequences becomes obvious if we remember that periodic trajectories generate periodic sequences $\Theta_f(x)$ (can also prove the converse assertion). So, if we know the classification of the permissible periodic sequences, then we can understand much in the organization of the periodic trajectories of mapping f. These periodic sequences have been studied in several works. Let us set a periodic sequence:

$$\theta = \theta_1, \theta_2, \ldots, \theta_k, \theta_1, \theta_2, \ldots, \theta_k,$$

with the minimum period k. For every sequence like this we can define an antiperiodic sequence:

$$\theta^{(1)} = \theta_1, \theta_2, \ldots, \theta_k, -\theta_1, -\theta_2, \ldots, -\theta_k, \theta_1, \theta_2, \ldots, \theta_k, -\theta_1, -\theta_2, \ldots, -\theta_k,$$

with the period k. Note that the period of this sequence is $2k$, but in the definition of antiperiodic sequence there is just one k. This is done to facilitate further formulation. The operation of antisymmetrization of the sequences can be applied many times, and then it is convenient to introduce n-multiple to the antisymmetrization as $a^{(n)} = (a^{(n-1)})^{(1)}$. The value $n = 0$ is defined as $a^{(0)} = a$. It is proven that the sequence θ is permissible if $\theta^{(1)}$ is permissible but only in this case. The period of sequence $\theta^{(n)}$ is equal to $k2^n$.

Theorem 3.8 *If the mapping f has a cycle with the period k, then some point x exists so that one of the series $\Theta_f(x)$, $Q_-(x)$, or $\Theta_+(x)$ is permissible periodic or antiperiodic with the period k. If for some point x one of the series $\Theta_f(x)$, $Q_-(x)$, or $\Theta_+(x)$ is permissible periodic or antiperiodic with the period k, then the mapping f has a cycle with the period k.*

All the permissible periodic series are ordered, so the cycles of the mapping are ordered too. It is proven that if a mapping with one maximum has a cycle with the period n, then there are cycles with periods $\lhd n$, located in the following order:

$$1 \lhd 2 \lhd 2^2 \lhd 2^3 \lhd \cdots \lhd 2 \cdot 7 \lhd 2 \cdot 5 \lhd 2 \cdot 3 \lhd \cdots \lhd 7 \lhd 5 \lhd 3.$$

This surprising result was proved by the remarkable mathematician Sharkovsky [13–15] and this series is therefore called the Sharkovsky series.

Another remarkable property of a kneading series is the possibility of defining topological entropy h_{top}. It appears that the closest to the origin of the coordinate positive root t_* of kneading series $Q_+(t) - Q_-(t) = 0$ is related to the topological entropy by a simple equation:

$$h_{\mathrm{top}} = -\ln(t_*).$$

The algorithm of the topological entropy computation with the use of kneading series proves to be quite efficient and easily realized by computer. We should note that similar relations are also known for the mappings which have several extrema and discontinuities points; only the piecewise monotonicity of such mappings is important.

References

1. Poincaré, H.: Les Méthodes Nouvelles de la Mécanique Céleste, vol. I. Gauthier-Villars, Paris (1892)
2. Poincaré, H.: Les Méthodes Nouvelles de la Mécanique Céleste, vol. 3. Gauthier-Villars, Paris (1899)
3. Poincaré, H.: Sur les propriétés des fonctions définies par les équations aux différences partielles, Oeuvres, pp. XCIX–CX. Gauthier-Villars, Paris (1929)
4. Birkhoff, G.D.: Dynamical Systems. American Mathematical Society Colloquium Publications, vol. 9. American Mathematical Society, Providence (1927), reprinted 1966
5. Hénon, M., Heiles, C.: Astron. J. **69**, 73–79 (1964)
6. Lichtenberg, A.J., Lieberman, M.A.: Regular and Stochastic Motion. Springer, New York (1983)
7. Bergé, P., Pomeau, Y., Vidal, C.: Order Within Chaos. Wiley, New York (1984)
8. Sinai, Ya.G.: Topics in Ergodic Theory. Princeton University Press, Princeton (1994)
9. Sinai, Ya.G.: Introduction to Ergodic Theory. Princeton University Press, Princeton (1977)
10. Kolmogorov, A.N.: Dokl. AN SSSR **115**, 861–864 (1958)
11. Zvonkin, A.K., Levin, L.A.: Usp. Math. Nauk **25**(6), 85–127 (1970)
12. Milnor, J., Thurston, W.: On Iterated Maps of the Interval Dynamical Systems (College Park, MD, 1986–1987). Lecture Notes in Mathematics, vol. 1342, pp. 465–563. Springer, Berlin (1988)
13. Sharkovskij, A.N.: Ukr. Mat. Zh. **16**, 51–71 (1964)
14. Sharkovskij, A.N.: Ukr. Mat. Zh. **17**, 104–111 (1965)
15. Sharkovskij, A.N., Kolyada, S.F., Spivak, A.G., Fedorenko, V.V.: Dinamika Odnomernyh Otobrazhenij. Naukova Dumka, Kiev (1989)

Chapter 4
Reconstruction of Dynamical Systems

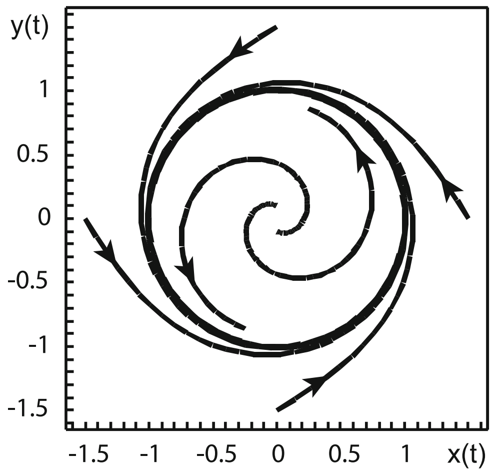
The concept of attractors plays an important role in physical research. This is because of their prevalence. Aside from that, when investigating properties of dynamical systems, we are actually studying attractors. In a certain sense, attractors realize the dynamical variant of the statistical principle of shortened description. Indeed, it is not necessary to study the behavior of a system in detail, starting from initial conditions, if after a time the system will reach a stationary regime corresponding to the motion on the attractor. It is these limit regimes which must first of all be investigated. The existence and prevalence of such regimes is also important to reconstruct or restore the properties of dynamical systems from experimental data. Modern progress in that direction is to a great extent connected with the conception of attractors. In fact, one of the central problems in physics is the creation of models for the description of real physical phenomena. In many cases, those phenomena are so complicated that it is not always clear what physical principles lie at their root. This is especially clearly seen in attempts to describe various phenomena in biological systems. On the other hand, experimental investigations do not and cannot give exhaustive or complete information on real systems. Therefore it becomes important to study the following problem.

4.1 What Is Reconstruction?

Let us consider a classical physical system about which we know very little or nothing at all. Usually, it is called a black-box in order to stress our ignorance about the system. The temporal behavior of this black-box is determined by the variation of its generalized coordinates, which are all the parameters that determine the black-box state. There can be many such parameters—the exact number is also unknown. It is evident that experimental monitoring of all those generalized coordinates is a hopeless task. It was earlier believed that only in such hopeless

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_4

Fig. 4.1 Phase portrait with limit cycle



cases could reliable information on the classical (non-quantum) system be obtained. A more realistic task is the experimental measurement of the evolution of one parameter over time. A question arises: what can we know about the properties of the black-box from such incomplete experimental information? Of course, we must nevertheless rely on some assumptions. Thus, we will assume that the system is non-linear and dissipative—this will be the most complex and interesting example. We can therefore expect that the black-box is in a stationary regime corresponding to some attractor. Our measurements can confirm or refute this, if, for example, our black-box explodes. The very presence of an attractor is already an important fact that allows us to advance further. Let us begin with a simple and well-known example: a very simple attractor known as the limit cycle (see Fig. 4.1).

Example 4.1 In this example we will try to understand a basic idea that can also be applied in more complicated cases. Let the limit cycle be determined by the parametric equations

$$\begin{aligned} x &= a\sin(\omega t) \\ y &= a\cos(\omega t)\,. \end{aligned} \tag{4.1}$$

This means that it coincides with a circle of the radius a centered in the origin. From the gedanken experiment we extract only the temporal variation of the coordinate $x = x(t) = a\sin(\omega t)$. However, in the process of evolution the system visited all the attractor points. Therefore, this coordinate contains information about the entire attractor. The question is only how to extract it. Let us construct a vector $\mathbf{x}$ according to the following rule:

$$\mathbf{x} = (x_1(t),\, x_2(t)) = (x(t),\, x(t+\tau))\,. \tag{4.2}$$

Let us assume now that τ is small and transform the vector to the following form:

$$\mathbf{x} = \left(x(t),\ \frac{\partial x(t)}{\partial t}\tau \right) . \tag{4.3}$$

Using the explicit form of $x = x(t)$, it is easy to rewrite it in the form

$$\mathbf{x} = (x_1(t),\ x_2(t)) = (x(t),\ y(t)\tau \cdot \omega) .$$

After the choice $\tau = \frac{1}{\omega}$ we obtain the vector exactly running through the original limit cycle.

In some sense, we reconstructed the whole attractor from only one known coordinate. In the above procedure only the choice of τ required some additional information about the attractor. All the other steps were executed using only the assumed experimental data on the temporal variation of the first coordinate. The choice of τ could be grounded independently by the following "principle." Characteristic scales on both axes of the attractor coincide or are very close. This concept is also very often used in more complex cases. Of course we knew in advance the dimension of the space where the limit cycle was situated, and limited ourselves to the two-dimensional vector. The presented considerations are qualitative. At first sight they appeared to be efficient because of the specific form of the attractor, but this not quite so. To illustrate, let us consider a system of non-linear equations.

$$\begin{aligned}
\frac{dx_1}{dt} &= A(x_1, x_2, x_3) \\
\frac{dx_2}{dt} &= B(x_1, x_2, x_3) \\
\frac{dx_3}{dt} &= C(x_1, x_2, x_3) .
\end{aligned}$$

Time-differentiating the first equation and using the two other equations to eliminate the derivatives $\frac{dx_2}{dt}$ and $\frac{dx_3}{dt}$, we obtain the following equations:

$$\begin{aligned}
\frac{dx_1}{dt} &= A(x_1, x_2, x_3) \\
\frac{d^2x_1}{dt^2} &= D\left(x_1, x_2, x_3, \frac{dx_1}{dt} \right) .
\end{aligned}$$

Let us assume that we able to solve that system of equations with respect to x_2 and x_3.

$$x_2 = G\left(x_1, \frac{dx_1}{dt}, \frac{d^2x_1}{dt^2}\right)$$

$$x_3 = R\left(x_1, \frac{dx_1}{dt}, \frac{d^2x_1}{dt^2}\right) .$$

From those relations, it follows that in principle, it is possible to determine the temporal variation of the coordinates x_2 and x_3 for given the variation $x_1(t)$, $\frac{dx_1(t)}{dt}$, and $\frac{d^2x_1}{dt^2}$. In other words, from data for $x_1(t)$ and, accordingly, $x_1(t+\tau)$, $x_1(t+2\tau)$, we can reconstruct the coordinates x_2 and x_3. Therefore, the possibility of reconstructing the attractor from information on the temporal variation of one coordinate exists in more complex cases as well.

The idea of reconstructing the phase portrait in multidimensional dynamical systems from the temporal dependence of one coordinate was first expressed in [1]. The method of introducing a vector to determine the system's position in n-dimensional phase space is absolutely analogous

$$\boldsymbol{\xi} = (x(t),\, x(t+\tau), \ldots, x(t+(n-1)\cdot\tau)) \ , \tag{4.4}$$

where $\tau, 2\tau, \ldots, (n-1)\,\tau$ are called time delays. This vector evolves in time and draws the trajectory in n-dimensional phase space. Then, the problem is in determining the dimension n of the phase space.

4.2 Embedding Dimension

Before discussing embedding dimension let us recall what is understood as manifold embedding in space. Let the space dimension or X manifold be smaller than the space dimension Y. If the map $f : X \to Y$ gives one-to-one correspondence between points X and $f(X)$, then the Jacobian rank $\frac{\partial f}{\partial x}$ is everywhere equal to the manifold dimension X $\dim X$, and is called the embedding of X in Y. For instance, the embedding of a circle into two-dimensional space is shown in Fig. 4.2 at the left. Mapping $f : X \to Y$ which does not need one-to-one correspondence is called immersion. To put this another way, immersion is an embedding but is only local. Figure 4.2 at right shows an example of the immersion of a circle into R^2

In the previous example the dimension of the space where the attractor was embedded was known. There were two important topological characteristics: the dimension of the attractor and the dimension of the space (manifold) to which the attractor belongs. Of course, these characteristics cannot take arbitrary values independently. For instance, the attractor's dimensions cannot exceed the dimensions of the embedding space. The dimensions of the space where the attractor is embedded

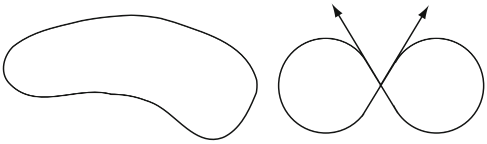


Fig. 4.2 At *left* is the embedding of the circle $S^1 \to R^2$. At the *right* is only its immersion into R^2. In this last case each vector of velocity on the circle corresponds to the unique velocity vector on the figure-eight. But there is no one-to-one correspondence between the points of the image and of the prototype

are important since they determine the minimum number of non-linear equations which describe the attractor in this phase space. Let us assume that the attractor belongs to a D-dimensional manifold. Then we can use Whitney's theorem (see, for example, [2–5]), providing that any D-dimensional manifold can be embedded into Euclidean space R^{2D+1}.

There is a simple way to understand Whitney's theorem. To obtain the one-to-one correspondence of image and prototype we need to set D coordinates. In order to have the same correspondence between the tangent vectors we also have to set another coordinate D. And finally, we have to add another coordinate to avoid self-intersection as shown in Fig. 4.2. Thus D-dimensional manifold can be embedded into space with the dimension $n = 2D + 1$. Now, we have to be sure that there are non-intersection points. Let us consider the condition providing the absence of intersection points of two hypersurfaces in space R^n. Let us define the hypersurface of D dimension by the equations

$$f_i\,(x_1, x_2, \ldots, x_n) = 0\,.$$

The number of these equations is $n - D$. Therefore $i = 1, 2, \ldots n - D$. Let us define the second hypersurface of D dimension by the equations

$$g_i\,(x_1, x_2, \ldots, x_n) = 0\,.$$

Where $i = 1, 2, \ldots n - D$. The choice of the same dimension of hypersurfaces is suitable for the analysis of intersection points as well. The total number of equations is $2n - 2D$, while the number of unknown coordinates is n. It is clear that if $2D = n$, then the number of equations is equal to the number of unknown coordinates. That means that in general cases, these equations have a finite number of solutions which determine the coordinates of a finite number of intersection points for these hypersurfaces. If $2D < n - 1$, then the number of equations is bigger than the number of unknown coordinates. Hence there is no solution of this kind of equation in general cases. It means that if

$$n > 2D + 1\,,$$

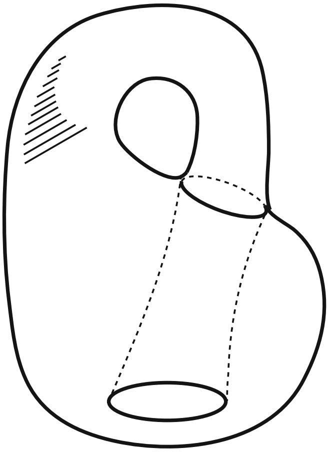


Fig. 4.3 This figure shows one side surface of the Klein bottle

there are no intersection of these hypersurfaces. Now it is evident that this dimension limitation corresponds with Whitney's theorem.

Whitney also proved the enhanced version of the embedding theorem [2]. According to this theorem every paracompact D, or dimensional Hausdorff manifold, is embedded into R^{2D}space.

Example 4.2 Notice that there are some bidimensional manifolds which cannot be embedded into three-dimensional Euclidean space, R^3. This follows from Whitney's theorem. The best-known example of such a manifold is the Klein bottle. Figure 4.3 shows its model in R^3 with self-intersections. However, without self-intersections Klein bottle can be embedded into four-dimensional Euclidean space. This corresponds fully to Whitney enhanced theorem.

Theoretical progress followed this direction. In 1981, Takens [6] was able to prove the theorem, which immediately attracted much attention. Its content can be explained in the following way. Let a system of equations, in phase space of which an attractor exists, be the following:

$$\frac{dx_i}{dt} = F_i(\mathbf{x}) ,$$

where $i = 1, \ldots, n$. Then in $2D + 1$-dimensional space

$$\boldsymbol{\xi} = (x(t),\ x(t+\tau), \ldots, x(t+2D\cdot\tau)) \ , \tag{4.5}$$

there is another attractor, all the metric properties of which coincide with the ones of the original attractor. The dimension of that space is $D_E \geqslant (2D+1)$ and it is called the embedding dimension (see also [7]). The sign of inequality means that if the attractor can be embedded into a space of some dimension, then it can be embedded into spaces of higher dimensions.

Generalization of Whitney's theorem for a case with a fractal image which is consequently not a manifold is done in [7]. In this example it was proved that the manifold A can be embedded into dimension space n providing

$$n \geqslant 2D_k + 1\,.$$

Where D_k is the capacity of compact set A. Farther on we discuss this and other dimensions in detail. In fact, instead of $x(t)$ we can take any coordinate from the original system. In spite of the limitations and idealized character of that theorem, it was immediately utilized in the reconstruction of attractors and their properties in many physical systems. First, we suppose that temporal series are measured with absolute precision on the infinite time interval. In addition, we suppose that there are no noises at all which can influence the system. In spite of this extreme idealization Takens' theorem is broadly used to restitute attractors in phase space as well to find its dimension, Lyapunov Index, topological entropy, and other important characteristics of dynamical systems. A generalization of this theorem for systems with external forces, including stochastic ones, was also obtained in [8]. However, in order to really utilize that theorem or its generalizations one needs to learn to determine either the attractor dimension or the embedding space dimension from one-dimensional signal data.

4.3 Attractor Dimension

Let us turn now to a discussion of the dimension of attractors. In the case of an attractor like the one in the example above, there is no difficulty in determining its dimension. The limit cycle itself represents a manifold which is topologically equivalent to a circle S^1. However, it is well known that attractors can be of different natures, and they can also be non-manifold. This is one of the impressive discoveries at the heart of modern theory on dynamical systems. In other words, there are stationary regimes distinct from constants (where the attractor is a stable fixed point) and from periodic ones (where the attractor is a limit cycle or a limit torus). The appearance of such new stationary regimes strongly affected views on many physical processes. A widely known Ruel and Takens scenario [9] of turbulence origination is also based on the concept of strange attractors. They are called strange due to their unusual geometrical properties. Stationary regimes of dynamical system motion on strange attractors are chaotic. The strange attractors are already not manifolds and they are arranged in quite a complicated way. In modern terms, we can say that in some directions they appear to be fractals like the Cantor set. It is clear that the dimensions of such objects deserve more detailed discussion.

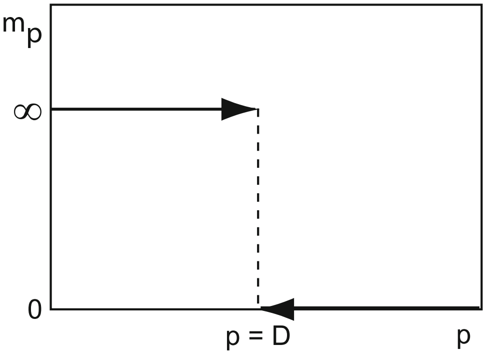


Fig. 4.4 Shows symbolically the dependence of the Hausdorff measure m_p on value p

The Hausdorff dimension is named after the man who invented one of the main ways of determining dimensions for fractal sets [10]. First, to introduce the Hausdorff measure. Let us cover our object X in spheres A_i of radius $\delta(A_i)$ that do not exceed radius ε, so that every point of X will belong to one of the coverage spheres. Now we consider the sum

$$m_p^{\varepsilon} = \inf \sum_{i=1}^{\infty} \delta\ (A_i)^p\ ,$$

where inf is the lower boundary over all coverages and $p \in \mathbb{R}$ is arbitrary real parameter. Then the following:

$$m_p\,(X) = \sup_{\varepsilon > 0} m_p^{\varepsilon}\,(X)\ ,$$

is called the p-dimensional Hausdorff measure. This measure as the function of p has a remarkable property. It looks like a special step: at small p it goes to infinity, and at large ones it is 0 (see Fig. 4.4).

Number D, for which $m_p\,(X) = 0$ for $p > D$ and $m_p\,(X) = \infty$ for $p < D$, is called the Hausdorff dimension of the set X. It means:

$$\dim_H X = D\,.$$

It is important to note that the Hausdorff dimension is not necessarily an integer and therefore it is a very useful tool for the analysis of the structure and complexity of different sets.

If we introduce $N\,(\varepsilon)$ as the minimum number of p-dimensional cubes of side ε needed to cover the set, then their capacity can be defined as,

$$\dim_K X = \lim_{\varepsilon \to 0} \left(-\frac{\log N\,(\varepsilon)}{\log \varepsilon} \right)\ , \tag{4.6}$$

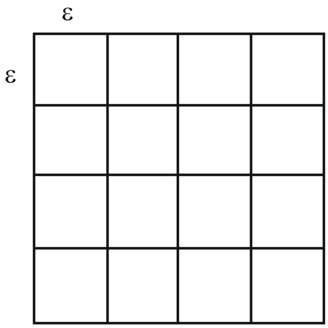


Fig. 4.5 Coverage of a square by square neighborhoods of size ε, the number of which is $N(\varepsilon) = \varepsilon^{-2}$

where ε tends to zero. The capacity of a set was originally defined by Kolmogorov [11] (also called the box-counting dimensional). The Hausdorff dimension and the capacity require only a metric (i.e., a concept of distance) for their definition, and consequently we refer to them as "metric dimensions." The value of capacity coincides with fractal dimension.

Let us consider some examples of sets which have different dimensions. We start from an ordinary square Q. It is evident that its usual (topological) dimension equals 2. Let us now determine its Hausdorff dimension. First of all we cover the square Q with the same quadratic neighborhood of dimensionless size ε. In this case it is easy to calculate the sum m_ε^p

$$m_q^\varepsilon(Q) = \sum \varepsilon^p = N(\varepsilon)\, \varepsilon^p .$$

Where $N(\varepsilon)$ is the number of square neighborhoods of size ε, covering the square Q (see Fig. 4.5). The number of such squares (see Fig. 4.5) is

$$N(\varepsilon) = \varepsilon^{-2} .$$

After substitution into the previous equation we obtain

$$m_q^\varepsilon(Q) = \varepsilon^{p-2} .$$

To estimate the Hausdorff measure we need to use the limit $\varepsilon \to 0$. It is clear that when $p > 2$ the sum m_ε^p is proportional to the positive power of ε and correspondingly this sum limit is equal to zero. Hence the Hausdorff measure with $p > 2$ is equal to zero. When $p < 2$ the sum $m_\varepsilon^p(Q)$ is proportional to negative power of ε and the limit $m_\varepsilon^p(Q)$ is going to infinity. Thus we obtain by definition

$$\dim_H Q = 2 .$$

It is easy to see that in this case (as well as for other "normal" sets) the Hausdorff dimension coincides with the usual (topological) dimension. Let us now consider a

Fig. 4.6 Iterative method for Cantor set construction

more exotic example in a Cantor set K. This set is built by the iterative procedure, which implies dividing the interval, obtained in the preceding step of the iteration, by 3 and consecutively removing the central part. Several steps of this process are shown in Fig. 4.6. We can see that after n steps of construction the set consists of $N_n = 2^n$ intervals whose length is $\varepsilon_n = 3^{-n}$. For the limit case the Cantor set consists of points with the coordinates

$$x = \frac{a_1}{3} + \frac{a_2}{3^2} + \cdots \cdot \frac{a_n}{3^n} + \cdots ,$$

where a_n coincides either with 0 or with 2. Let us now determine its Hausdorff measure. For that we cover it on n-step of construction with intervals of $\varepsilon_n = 3^{-n}$ length and we calculate the sum $m_\varepsilon^p\,(K)$. For a set like this, it is easy to find m_p^ε (K) at each iteration step

$$m_p^{\varepsilon_n}(K) = \Sigma\varepsilon_n^p = N_n\varepsilon_n^p \,.$$

The number of segments which covers it on n-step is also easy to calculate as $N_n\,(\varepsilon_n) = 2^n$. Now we can write $N_n\,(\varepsilon_n)$ as a function of ε_n using some simple transformations

$$N_n\,(\varepsilon_n) = 2^n = e^{-n\ln 2} = e^{-n\frac{\ln 2}{\ln 3}\ln 3} = \left(e^{-n\ln 3}\right)^{\frac{\ln 2}{\ln 3}} = (\varepsilon_n)^{\frac{\ln 2}{\ln 3}}$$

and we obtain

$$N_n\,(\varepsilon_n) = (\varepsilon_n)^{\frac{\ln 2}{\ln 3}} \,.$$

As a result, for the sum $m_\varepsilon^p\,(K)$ we have

$$m_p^n\,(K) = N_n\varepsilon_n^p = (\varepsilon_n)^{p-\frac{\ln 2}{\ln 3}} \,.$$

When analyzing the limit of this sum with $\varepsilon_n \to 0$ in the same way as we do it for the square Q we can find the Hausdorff dimension

$$\dim_H K = \frac{\ln 2}{\ln 3} \,.$$

The main result of these simple calculations is that we have a real example of a set (formerly such sets were called perfect), which has non-integer Hausdorff dimensions. Aside from this, because points in the set can be separated from one another (by the empty set), its usual dimension (topological) is $\mathrm{Dim} K = 0$. The appearance of such sets in mathematics was formerly seen as exotic. However, research has moved on, and we now consider square sets exotic. Reasons for this are mathematical and physical in nature. A simple example of objects with fractional dimension is represented by plots of everywhere non-differentiable functions. The first example of a function like this was presented by Weierstrass. A natural question arises: among continuous functions, are the smooth (i.e., those that have derivatives) or the everywhere non-differentiable functions more frequent? Banach proved [12], using the category approach [13], that in terms of categories almost all continuous functions are everywhere non-differentiable, i.e., the differentiable functions have measure zero. In that sense, the differentiable continuous functions are exotic. Taking into account the evident connection between plots of functions and observations of natural phenomena and processes, we should expect wider prevalence of the objects and processes which are characterized by non-integer Hausdorff dimensions.

Mathematical researches on dimensionality theory lay outside of the interests of physics for a long time. One of the first works in that domain which remained long unnoticed was Richardson's paper, devoted to the investigation of coastline length in Britain (see [14]). Interest in perfect sets in physics appeared and arose to a great extent due to the works of Mandelbrot [14]. His main idea was tied to the extraction of objects (sets), which have self-similarity, i.e., those in which a part of the set in some sense is similar to the whole set. In other words, such sets are scale-invariant. This means that when we extract a part of the set and change the scale in the appropriate manner, we get a new set that either coincides exactly with the original whole set or coincides statistically with the given probability measure. Such self-similarity can appear as a result of an iterative construction procedure, as well as due to other, physical reasons. In particular, the self-similarity of the Cantor set (discussed above) is evident and is the result of the simple iterative scheme used in its construction. In some sense, self-similar sets are analogous to auto-model solutions of different equations, which are also connected to the presence of higher symmetries compared to all other solutions.

The importance of the self-similarity concept is manifested by the appearance of simpler and more efficient methods for the description of self-similar set properties. Indeed, returning to expression (4.6) for self-similar sets, one can see that if a part is similar to the whole, then the expression (4.6) is satisfied not only in the limit, and also in finite, but sufficiently small coverage scales. Then it can be simplified and rewritten in the following form:

$$N(\varepsilon) \sim \varepsilon^{-D_F}, \qquad (4.7)$$

where $N(\varepsilon)$ is the number of neighborhoods with characteristic size ε, containing points of the considered, and D_F coincides with the capacity and the Hausdorff dimension for such sets. Mandelbrot proposed naming D_F the fractal dimension or cellular dimension of the set. From the definition (4.7) it clearly follows that D_F gives a quantitative characteristic of the self-similarity, i.e., it points out how to change the scale so that the chosen part will coincide with the whole set. While using (4.7) as a definition of the fractal dimension, it was seen that it is quite easy to determine it from physical experiments and a huge number of physical objects and processes have now been discovered which have non-integer fractal dimensions. Thereby fractal objects and processes have become an essential part of physical objects and processes.

The next step in the generalization of scaling led to the so-called self-affine fractals and, respectively, to other characteristics of these objects. When talking about the scale invariance of sets, we tend to assume that the space where the considered set is embedded is similar to a Euclidean space, where all the coordinates are equivalent and that the scaling acts on all coordinates in the same way. However, this is far from being always true from both physical and mathematical points of view. For example, in space-time, scalings on spatial and temporal coordinates can be independent. For objects like trajectories in space-time, the similarity coefficients on the time and space coordinates are not necessarily the same.

Another possibility for generalization comes from physical concepts about the beginnings and growth of fractal clusters. Indeed, in this case there is a minimum scale—the size of particles. Therefore the tending of coverage size to zero is insignificant. However, the cluster size in the process of growth does not have an upper limit, and we can increase the size of the cell or coverage up to ∞. This allows us to introduce the global cluster dimension at $\varepsilon \to \infty$. Then the usual definition can be understood as the determination of a local cellular dimension.

Entropic or informational dimension is introduced using approaches originating from information theory or statistical physics. By covering the fractal set with neighborhoods of size ℓ we can introduce probability to find the points of the set in any ith neighborhood. This probability equals $p_i(\ell) = N_i/N$, where N_i is the number of points in the fractal set that fell into the ith neighborhood, and N is the total number of points in the set. Having introduced those probabilities we can define the entropy of the fractal set

$$I(\ell) = -\sum_i p_i \ln p_i \,.$$

It is clear that the entropy value depends on the scale of the neighborhood. Let us now define the informational dimension as the velocity of entropy or information variation at scale alteration

$$D_I = -\lim_{\ell \to 0} \frac{I(\ell)}{\ln \ell} \,.$$

With the uniform filling of space on all scales, the probabilities p_i are proportional to the volume of the neighborhood ℓ^d, where d is the space dimension. Let us calculate the value of the informational dimension for such a case

$$D_I = \lim_{\ell \to 0} \frac{N(\ell) d\ell^d \ln \ell}{\ln \ell} = \lim_{\ell \to 0} \frac{\ell^{-d} d\ell^d \ln \ell}{\ln \ell} = d\,.$$

Thus, for non-fractal sets the informational dimension coincides with the topological dimension of the set. For fractal sets, however, the informational dimension will have a different value this will generally be different even from the fractal dimension.

Using the probability approach, we can introduce one more dimension called the correlational dimension. The idea behind its introduction is connected to the behavior of correlations during a decrease in the distance between points. Let us define the correlation integral as

$$C(\ell) = \lim_{N \to \infty} \sum_{i,j} \frac{N_{ij}(\ell)}{N^2}\,,$$

where $N_{ij}(\ell)$ is the number of pairs of points, which lie in distance less than ℓ. Then the correlational dimension is defined by the following relation:

$$D_c = \lim_{\ell \to 0} \frac{\ln C(\ell)}{\ln \ell}\,.$$

The value of this dimension for objects studied in physics can also differ from both fractal and informational dimensions. These characteristics are important due to the distinctions in their definitions as well as to the fact that they characterize the degree of scale inhomogeneity for the objects under consideration.

We can now return to the determination of the attractor dimension. Strictly speaking, we should treat its dimension as a Hausdorff. However, its determination contains limit procedures that do not allow us to reconstruct it from a finite (even if it is a very large) data set. Therefore, in order to determine the attractor dimension, one usually uses the correlational dimension D_c, which gives the lower boundary of the attractor dimension D_A

$$D_A \geqslant D_c\,.$$

Here, under the attractor dimension, we shall understand the Hausdorff dimension of the attractor. However, in this way we can define the attractor dimension in its proper phase space. Our task is more complex because of the uncertainty in the very phase space of the system. Therefore, the determination of the attractor dimension is also connected with finding the embedding dimension.

4.4 Finding Embedding Dimension

One can determine the embedding and attractor dimensions simultaneously using the following strategy. Let us find the correlational dimension of the reconstructed attractor in the space (4.4) at small values of $n = n_0$. For that we construct the dependence of a correlation integral on resolution ε in logarithmic scales. Figure 4.7 represents the typical plot. Saturation at $\varepsilon \geqslant \varepsilon_{\max}$ is due to the fact that when ε attains the attractor dimension, all point pairs fall within this neighborhood and consequently $C \to 1$ (see Fig. 4.7). On the other end of the linearity interval $\varepsilon \leqslant \varepsilon_{\min}$, the neighborhoods ε become so small that the attractor structure is indistinguishable. In this domain the dependence correlation integral C on scale ε becomes complicated and non-linear (see Fig. 4.7). Values $\varepsilon_{\min}$ and $\varepsilon_{\max}$ depend on experimental data and on the physical system, as well. The correlation dimension at chosen value n_0 is determined by the tangent of the slope angle on the linear part of this plot. Let us note that experimental data does not always give such a clear picture. Sometimes it is more complicated due to the particularities of the physical system, measuring instruments, etc.

Then we increase the value of n by one and again we determine the correlation dimension of the attractor in a space of higher dimension. A typical example of the dependence of $\ln C$ on $\ln \varepsilon$ at different values n is shown in Fig. 4.8. Repeating this procedure, we can construct from the obtained data the dependence of the correlation dimension on the phase space dimension number n. It is clear that after the quantity n reaches value D_E the attractor dimension stops changing (see Fig. 4.8). In other words, such dependence reaches saturation at $n = D_E$ and the correlation dimension value at that point determines attractor dimension (see Fig. 4.9). Thus, analyzing how the curve tends to saturation we determine both the embedding dimension and the attractor dimension. Knowing the embedding dimension, we can, using the given temporal dependence of one coordinate, reconstruct the attractor

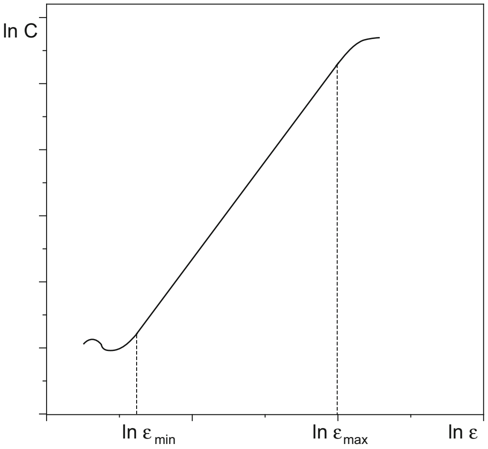


Fig. 4.7 The dependence of the correlation integral logarithm on scale logarithm ε. Typical deviations from linearity are shown symbolically for large and small ε. The correlation dimension is defined by the slope of the linear part only

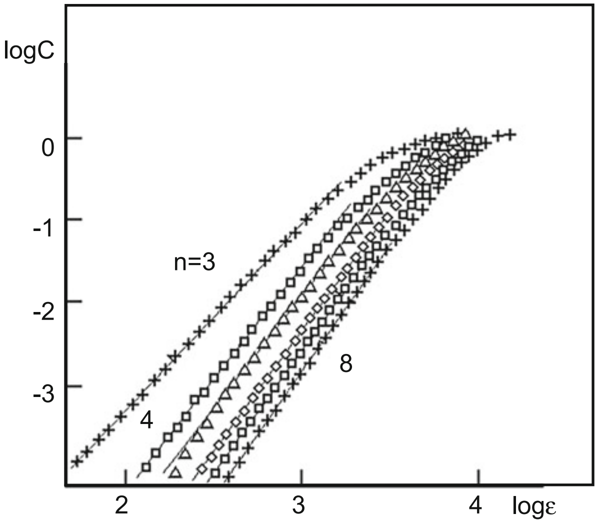


Fig. 4.8 The plot shows $\log C$ vs $\log \varepsilon$ for different values of the embedding dimension in the Rayleigh–Benard experiment [15, 16]. It is clear that when n increases from 3 to 8 the linear part slope stabilizes

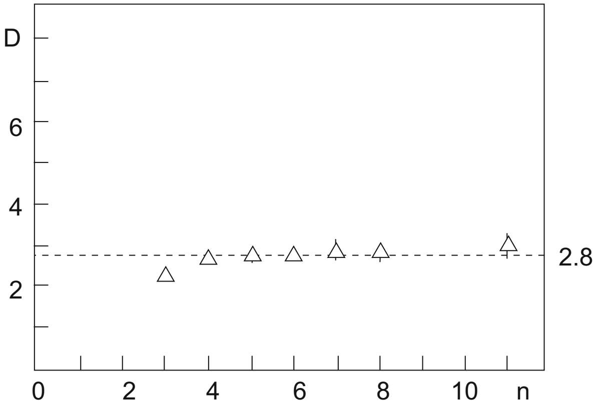


Fig. 4.9 The typical stabilization of the correlation dimension at $D_c \approx 2.8$ with the increasing of the embedding dimension. The data corresponds to Fig. 4.8

from that data and we can study its properties in the phase space of dimension D_E. Of course, there is a whole range of difficulties in performing such a strategic task which are discussed in details in the literature. Most of them are connected with the choice of the delay time τ, scale intervals ε, and the number of "experimental" points N.

Generally speaking, the reconstruction of systems using experimental data requires an enormous quantity of experimental points. Let us return to the question of the reconstruction of attractor dimension with the help of correlation dimension. Let the number of experimental points be $N_{\exp}$. The dimension of neighborhoods changes from 0 to the attractor dimension L when calculating the correlation interval. Let designate the number of point pairs within the neighborhood with radius ε as $N(\varepsilon)$. It is a priori clear that this number changes from 0 to $\frac{N_{\exp}^2}{2}$. All the points are in the neighborhood of L dimension, while at smaller neighborhood dimensions, the number of point pairs is proportional to the neighborhood volume

$$N(\varepsilon) \approx \frac{N_{\exp}^2}{2}\varepsilon^D .$$

It is obvious that to achieve good statistical results this number must be big enough, i.e., $N(\varepsilon) \gg 1$. Hence the following condition has to be satisfied:

$$\frac{N_{\exp}^2}{2}\varepsilon^D \gg 1 .$$

It is easy to write this inequality as

$$N_{\exp} > \varepsilon^{-\frac{D}{2}} .$$

This inequality shows how much experimental data $N_{\exp}$ is needed to reconstruct the D-dimension attractor with resolution ε. With increases in dimension and accuracy resolution we need more experimental data. This inequality can be presented as

$$D < \frac{2\ln N_{\exp}}{\ln\left(\frac{1}{\varepsilon}\right)} .$$

For instance, if $N_{\exp} = 10^4$ and $\varepsilon = 0.1$, then the reconstructed attractor dimension is $D < 2 \cdot \frac{4}{1} = 8$.

The choice of time delay is not easy. Obviously we do not have to choose an excessive time delay since the correlation between consecutive values of measurements disappears. This naturally leads to inaccurate reconstruction. The value τ cannot be too small, either, since in this case the differences in observed values are practically absent for some steps of observations. Hence $x(t) \approx x(t+\tau) \approx x(t+2\tau) \approx \cdots \approx x(t+(n-1)\tau)$ and data are concentrated near the hyperdiagonal of phase space. That gives a strong anisotropy of reconstructed attractor. There are some formal ways to choose τ [17, 18]. The trial-and-error method seems to be more appropriate to choose τ, since small values of τ give strong compression toward the phase space diagonal, while at large τ the thin attractor structure spreads. In practical situations τ is optimal if $n\tau$ has the same order as the characteristic period of a dynamical system.

There are also numerous modifications of methods proposed for attractor reconstruction (see, for example, the review [19]), the development and perfection of which is ongoing.

4.5 Global Reconstruction of Dynamical Systems

Lately, a more general problem has arisen and developed, which can be called the global reconstruction of non-linear dynamical systems. Its goal is to reconstruct the form of the dynamical system (i.e., the system of non-linear equations), which generates the known temporal dependence of one of the generalized coordinates. Such a task is mathematically incorrect, as many different dynamical systems can generate the same attractor. Nevertheless, interest in this attractive problem is fairly active. There are several methods of such global reconstruction, and all of them are based on essential assumptions and hypotheses about the expected form of dynamical systems. For the moment, they are applied to test examples of well-studied dynamical systems. In the simplest cases, the reconstruction can be divided into several stages. The first is the local attractor reconstruction in the phase space. This stage has already been discussed. At the second stage, one makes assumptions about the form of the non-linear system. It is essential that such assumptions must restrict the infinite-dimensional space of right-hand sides in first order systems to a finite-dimensional space. Usually, this is achieved by assuming that the right-hand sides are polynomials of phase space coordinates. Of course this is also complemented by restrictions on the order of these polynomials. Then, the problem is reduced to determining the finite set of unknown parameters (the coefficients of the polynomials) for the system using the initial data. The third stage namely consists of the selection of these coefficients. The simplest way to do it is to compare the obtained solution of the system of equations with the "experimental" data and to fit the coefficients by the method of least squares [20, 21]. Various more sophisticated variants of the global reconstruction of dynamical systems can be found in the works referenced [22–25].

References

1. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Phys. Rev. Lett. **45**, 712–715 (1980)
2. Whitney, H.: Ann. Math. **45**, 247–293 (1949)
3. Hirsch, M.: Ann. Math. **74**, 494–497 (1961)
4. Dubrovin, B.A., Novikov, S.P., Fomenko, A.T.: Modern Geometry. Nauka, Moscow (1979)
5. Hirsch, M.W.: Differential Topology. Springer, New York (1976)
6. Takens, F.: Dynamical Systems and Turbulence. Lecture Notes in Mathematics, vol. 898, pp. 366–381. Springer, New York (1981)
7. Mañé, R.: On the dimension of the compact invariant sets of certain nonlinear maps. In: Dynamical Systems and Turbulence, Warwick. Lecture Notes in Mathematics, pp. 230–242. Springer, New York (1981)

8. Stark, J., Broomhead, D.S., Davies, M.E., Huke, J.: Takens embedding theorems for forced and stochastic systems, nonlinear analysis, theory, methods and applications. In: Proceedings of 2nd Congress of Nonlinear Analysis, vol. 30, issue 8, pp. 5303–5314. Elsevier Science Ltd, Amsterdam (1997)
9. Ruel, D., Takens, F.: On the Nature of Turbulence. Communications in Mathematical Physics, vol. 20, pp. 167–192. Springer, Berlin (1971)
10. Hausdorff, F.: Math. Ann. **79**, 157–179 (1919)
11. Kolmogroov, A.N.: Dokl. Akad. Nauk SSSR **119**, 861–864 (1958)
12. Banach, S.: Colloq. Math. **1**, 103–108 (1948)
13. Baire, R.: Leçons Sur Les Fonctions Discontinues. Gauthier-Villars, Paris (1904)
14. Mandelbrot, B.: The Fractal Geometry of Nature. Freeman, San Francisco (1982)
15. Malraison, B., Atten, P., Berge, P., Dubois, M.: J. Phys. (Paris) Lett. **44**, 897–902 (1983)
16. Dubois, M.: Lecture Notes in Physics, vol. 164, pp. 177–191. Springer, Berlin (1982)
17. Framer, A.M., Swinney, H.L.: Phys. Rev. A **33**, 1131–1140 (1986)
18. Liebert, W., Schuster, H.G.: Phys. Lett. A **142**, 107–112 (1989)
19. Breeden, J.L., Packard, N.H.: J. Bifurcation Chaos **4**(2), 311–326 (1994)
20. Cremers, J., Hubler, A.: Z. Naturforschung A **42**, 797–802 (1987)
21. Crutchfield, J.P., McNamara, B.S.: Complex Syst. **1**, 417–452 (1987)
22. Breeden, J.L., Hubler, A.: Phys. Rev. A **42**(10), 5817–5826 (1990)
23. Brown, R., Rulkov, N.F., Tracy, E.R.: Phys. Rev. E **49**(5), 3784–3800 (1994)
24. Borland, L., Stochastic nonlinear modeling. In: Chaotic, Fractal, and Nonlinear Signal Processing. AIP Conference Proceeding Mystic, CT, vol. 375, pp. 719–725. AIP Press, New York (1995)
25. Bunner, M.J., Popp, M., Meyer, Th., Kittel, A., Rau, U., Parisi, J.: Phys. Lett. A **211**, 345–349 (1996)

Chapter 5
Controlling Chaos

Chaotic dynamics is one of the most general ways of non-linear systems evolution. Chaotic regimes are abundantly present in nature as well as in devices created by human hands. However, it is difficult to unambiguously answer whether chaos is useful or harmful. Chaos is beneficial when it increases the chemical reaction rate by intensifying mixing, providing a powerful mechanism for heat and mass transfer. However, in many situations chaos is an undesirable phenomenon which can, for example, lead to additional mechanical fatigue of the elements of construction due to their irregular vibrations. The possibility of non-resonant energy absorption in a chaotic regime can lead the system parameters beyond safe levels. Therefore, it is clear that the ability to control chaos, i.e., to enforce or suppress it, has great practical importance. Earlier, when chaos was still unusual, the problem of its amplification was at the center of attention. However, at the beginning of the 1990s, the pendulum swung in the other direction. Considerable theoretical and experimental efforts were made to convert chaos into periodic motion. A new and intensely developing domain of non-linear dynamics—controlled chaos—originated from the pioneering work [1] of the same name. From this point on the term “controlled chaos” entered into the vocabulary of physicists which deal with non-linear dynamical systems.

5.1 Statement of the Problem

All of the numerous definitions of chaos are based on the extreme sensitivity of chaotic dynamics to initial conditions. In the real world any system is disturbed by external noise, and in computer simulations a small perturbation appears due to numerical round-off. If a system is chaotic, then a small perturbation grows exponentially in time, drastically changing the behavior of the system. This peculiarity of chaotic systems considerably complicates analysis, driving experimenters

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_5

to despair. However, the reason for despair may become the origin of hope. After all, if the system is so sensitive to small perturbations, can we use them to control it? This idea is the basis for the Ott–Grebogy–Yorke (OGY) control method [1].

It is interesting to note that in the pioneering work [1] there was already a deep understanding of the prospective aim of chaos control. "One may want a system to be used for different purposes or under different conditions at different times. Thus, depending on the use, different requirements are made of the system. If the system is chaotic, this type of multiple-use situation might be accommodated without alteration of the gross system configuration. In contrast, in the absence of chaos, completely separate systems might be required for each use. Such multipurpose flexibility is essential to higher life forms, and we, therefore, speculate that chaos may be a necessary ingredient in their regulation by the brain."

The OGY method and its numerous variations are based on the fundamental concept of global reconstruction of the system due to small perturbation. They are used for chaos control both in abstract models and in different concrete systems beginning with the simplest pendulum [2] and ending with the such complex biological systems as the heart [3] and brain [4].

Practically any dynamical system can be an object for control. At the present time the best results are achieved in the domain of control for systems with dynamics which are chaotic over time. They are described by systems of ordinary differential equations and are finite-dimensional inherently. The dynamics of such a system can also be described using mappings in terms of discrete time. Such transitions can be performed using the well-known technique of Poincaré sections (see Sect. 3.1). Systems with spatially chaotic dynamics are also described by systems of ordinary differential equations. The realization of control in such systems requires only minor modification of the methods used for systems chaotic in time. Infinite-dimensional systems with dynamics which are chaotic both in time and in space are described by equations in partial derivatives. It is the control of such systems that now presents the main difficulty, though even in this domain considerable progress has been achieved.

One of the most attractive features of the developed control methods is that it is not necessary to have any analytical model of the controlled system. For input information describing the system dynamics we can use masses of experimental data of any dimension. This feature of the control methods opens up the possibility for its application to practically any system if its dynamics contain a chaotic component.

5.2 Discrete Parametric Control and Its Strategy

In any chaos control method the principal role is played by the immediate objects of stabilization, which are the unstable periodic orbits. One of the fundamental properties of chaos is the simultaneous coexistence of many different motions in the system. In particular, chaotic motion on the strange attractor coexists with an infinite number of unstable periodic orbits. The motion of the system along a chaotic orbit can be understood as the continuous switching between neighboring unstable periodic orbits [5]. One can imagine the periodic orbits as forming the skeleton of the chaotic attractor, supporting dynamics in the phase space. The same fundamental role is played by periodic orbits in quantum mechanics as well. We will cite just the famous trace formula [6], which allows the quantum spectrum to be linked in semiclassical approach with a sum over classical periodic orbits.

Although the existence of unstable periodic orbits does not apparently manifest in free chaotic evolution, these objects play a principal role from the point of view of chaos control. In order to demonstrate this, we shall discuss the strategy first realized in the OGY chaos control method. Let us consider a set of unstable periodic orbits embedded in the chaotic attractor of a dynamical system. For each of such orbits we shall ask the question: does motion on that trajectory optimize some system characteristic? Then we choose one of the optimal trajectories. Assuming that motion on the attractor is ergodic, we can always wait until the moment of time when the chaotic trajectory during its random walks approaches the chosen unstable periodic orbit. When it happens, we apply a small programmed perturbation in order to direct the trajectory towards the target orbit. If there is noise present in the system, we shall repeat that procedure in order to keep the periodic orbit.

To realize that strategy we need to do the following:

- to identify the unstable periodic orbits on the attractor;
- to study the attractor structure in the neighborhood of such orbits;
- to choose the system parameters appropriate for realization of control and to study the system response on their perturbation.

Perhaps the most difficult step in this strategy is the localization of the unstable periodic orbits. A whole row of papers [7–10] is devoted to this problem. However, the method which is the most closely related with the main idea of chaos control is the one based on the transformation of unstable objects into stable ones [11]. The essence of the method is in utilizing the universal set of linear transformations, which allow the transformation of unstable periodic orbits into the stable ones, localized in the same points of the phase space as the required unstable orbits. Stable periodic orbits obtained as the result of such transformation can be found with the help of standard iterative procedures. Analysis of the attractor structure in the neighborhood of unstable fixed points or unstable periodic orbits (which can be considered as the combinations of unstable fixed points) does not present any problem. To do this, one should observe the motion of the point representing the current state of the system in the Poincaré section surface. From time to time this point will approach the unstable fixed point along the direction which is called the stable one, and then it will move off along the unstable direction. These two directions form a geometric structure which is called the saddle. Knowledge of those directions (the eigenvectors of the Jacobi matrix) and the velocities of approaching and moving off along those directions (the eigenvalues of the Jacobi matrix) represent all the necessary information about the local structure of the attractor in the neighborhood of the fixed point, which is needed for the realization of the discrete parametric control method.

The identification of optimal system parameters is a relatively difficult task. We note that a number of modifications of the OGY method [12] permit this difficulty to be avoided. But as soon as the choice is made, there remains only to determine the positions of the unstable periodic orbits for a few parameter values close to the nominal one. This is all the necessary information needed for the realization of the discrete parametric control of chaos by the OGY method.

For a better understanding of the fundamental ideas lying at the core of this method, disregarding the difficulties connected with multidimensionality, let us perform chaos control [13] in one-dimensional logistic mapping

$$X_{n+1} = f(X_n, p) = pX_n(1 - X_n)\,, \tag{5.1}$$

where X is limited in the interval [0, 1], and p is the unique parameter of the mapping. It is well known [14] that one of the mechanisms of transition to chaos in that mapping is period doubling. As p grows, a sequence of period doubling bifurcations takes place at which the orbits with consecutive period doubling became stable. Period doubling bifurcations cascade ends at $p = p_\infty \approx 3.57$, after which chaos begins.

Let us assume that we want to avoid chaos at $p = 3.8$. More specifically, we want the trajectory with randomly chosen initial conditions to be as close as possible to some unstable periodic orbit, assuming that this orbit optimizes some system characteristic. Thus we will consider that we can only fine tune p near the value $p_0 = 3.8$, i.e., let us limit the range of variation for the parameter p by the interval $(p_0 - \delta, p_0 + \delta;\ \delta \ll 1)$.

In view of the fact that the motion is ergodic, a trajectory with arbitrary initial condition X_0 with unit probability will sooner or later appear in a neighborhood of the chosen periodic orbit. However, because of its chaotic nature (exponential divergence) the trajectory will quickly deviate from the periodic orbit. Our task is to program the parameter variation so that the trajectory will stay in the neighborhood of the periodic orbit during the control time. We stress that according to the very formulation of the problem we can use only a small perturbation of the parameter. Otherwise, chaos itself can be excluded, for example, changing the parameter p from 3.8 to 2.

Let us consider the orbit with period i:

$$X_0(1) \to X_0(2) \to \cdots X_0(i) \to X_0(i+1) = X_0(1)\,.$$

If in the moment of time n the chaotic trajectory appeared in the neighborhood of the mth component of the periodic orbit, then the linearized dynamics in the neighborhood of that component reads as the following:

$$
\begin{aligned}
X_{n+1} - X_0(m+1) &= \left.\frac{\partial f(X,p)}{\partial X}\right|_{X=X_0(m),p=p_0} \delta X_n \\
&\quad + \left.\frac{\partial f(X,p)}{\partial p}\right|_{X=X_0(m),p=p_0} \delta p_n \\
&= p_0[1-2X_0(m)]\delta X_n + X_0(m)[1-X_0(m)]\delta p_n\,,
\end{aligned}
$$

here $\delta X_n \equiv X_n - X_0(m)$, $\delta p_n = p_n - p_0$. Requiring that $X_{n+1} = X_0(m+1)$, we obtain the parameter perturbation, needed for $n+1$ iteration to get on the periodic orbit

$$
\delta p_n = p_0 \frac{[2X_0(m)-1]\,\delta X_n}{X_0(m)\,[1-X_0(m)]}\,. \tag{5.2}
$$

Relation (5.2) takes place if only the trajectory X_n appears in a small neighborhood of the chosen periodic orbit, i.e., when $\delta X_n \ll 1$ and, therefore, the perturbation δp_n is small. Otherwise the system evolves according to the initial parameter value p_0.

The procedure described above is convenient because it allows us to stabilize different periodic orbits in different times. Let us assume that we stabilized a chaotic trajectory in the neighborhood of some periodic orbit, for example, one of period 2. Then we decided to stabilize the period-1 orbit, i.e., an unstable fixed point, assuming that it is the orbit that optimizes some system characteristic in the present time. Let us switch off the control. After that, the trajectory starts to deviate exponentially quickly from the period-2 orbit. Due to ergodicity, after some time the trajectory will appear in the small neighborhood of the chosen fixed point. In that moment of time, we shall switch on the control, but for the unstable fixed point [calculated according to (5.2)], and we will stabilize the chaotic trajectory in its neighborhood. The result is described in Fig. 5.1a.

In the presence of external noise the controlled trajectory can be accidentally kicked out from the periodic orbit neighborhood. If this happens, we should switch off the control and let the system evolve freely until the chaotic trajectory returns to the neighborhood of the target periodic orbit, and the control can be resumed within the given range of parameter variation. For additive Gaussian noise it is easy to check that the average length of the controlled phase grows with the decreasing of noise amplitude. This situation is illustrated in Fig. 5.1b. The noise is modeled by additive term in the logistic mapping of the form $\eta\sigma$, where η is the noise amplitude, and σ is the Gaussian distributed random variable with zero average value and unit dispersion.

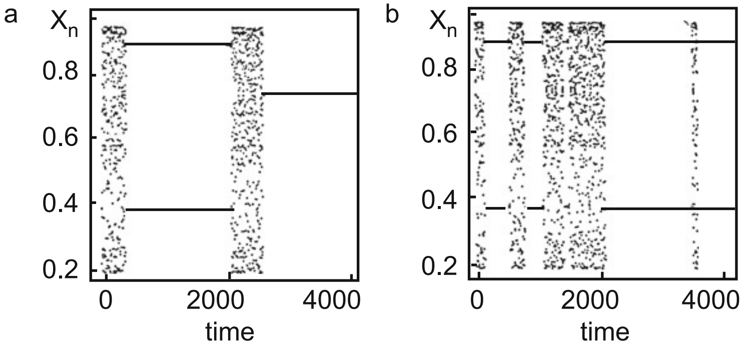


Fig. 5.1 (**a**) Control of unstable periodic orbit with period 2 in the logistic mapping in absence of noise; (**b**) control of the same orbit in presence of additive Gaussian noise $\left(\eta = 2.6 \times 10^{-4}\right)$ [13]

5.3 Main Equations for Chaos Control

Having illustrated the general strategy of the parametric control in the one-dimensional example, we now turn to multidimensional system control. For simplicity, we will consider the case of control for an unstable fixed point in two-dimensional phase space. This example contains all the basic features of multidimensionality.

We consider a three-dimensional continuous system with a two-dimensional Poincaré section, dynamics of which are described by the following mapping:

$$\mathbf{Z}_{n+1} = \mathbf{F}\left(\mathbf{Z}_n, p\right) , \tag{5.3}$$

where p is some parameter tunable in a small interval $\delta \ll 1$,

$$|p - p_0| < \delta , \tag{5.4}$$

around some initial value p_0.

The key difference between one-dimensional and two-dimensional (multidimensional) cases is the fact that in the latter, any unstable fixed point is connected with some geometrical structure, namely for each fixed point (or for every component of periodic orbit) there exist stable and unstable directions, which we mentioned before. The control strategy, accounting for the complication of geometry, consists of the following. Any time when the point $\mathbf{Z}_n$ of intersection of the orbit with the Poincaré section surface appears sufficiently close to the fixed point $\mathbf{Z}_F(p_0) = \mathbf{F}(\mathbf{Z}_F(p_0), p_0)$, the controlling parameter p acquires the new value p_n, such that after consecutive iteration, the point $\mathbf{Z}_{n+1} = \mathbf{F}(\mathbf{Z}_n, p_n)$ gets on the local stable manifold of the fixed point $\mathbf{Z}_F(p_n)$.

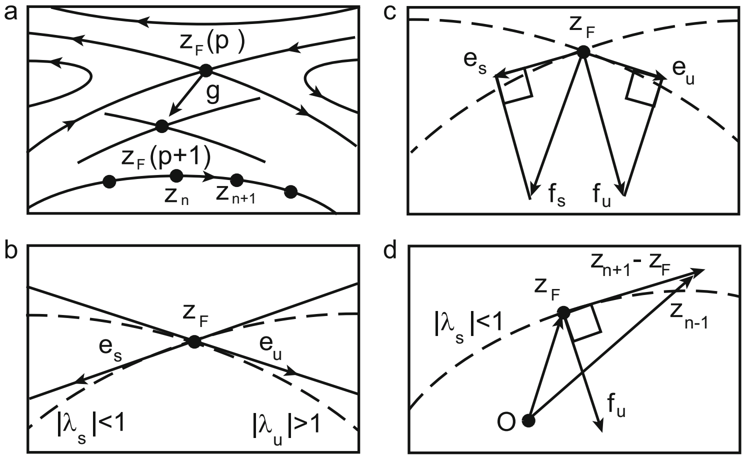


Fig. 5.2 (**a**) Shift of the unstable fixed point with variation of the parameters; (**b**) eigenvectors of the Jacobi matrix ($\mathbf{e}_u, \mathbf{e}_s$) for the fixed point $\mathbf{Z}_F$; (**c**) auxiliary basis ($\mathbf{f}_u, \mathbf{f}_s$); (**d**) iteration of the mapping in the neighborhood of the fixed point necessary to realize the control [15]

Let us realize this strategy [1, 15]. Shift of the fixed point due to variation of the parameter ($p_0 \to p_n = p_0 + \delta p_n$) equals

$$\mathbf{Z}_F(p_n) = \mathbf{Z}_F(p_0) + \mathbf{g}\,\delta p_n\,, \tag{5.5}$$

where the shift vector of the fixed point is $\mathbf{g} = \left.\frac{d\mathbf{Z}_F(p)}{dp}\right|_{p=p_0}$ (see Fig. 5.2a).

Linearized dynamics in the neighborhood of the fixed point $\mathbf{Z}_F(p_0)$ looks like the following:

$$\mathbf{Z}_{n+1} - \mathbf{Z}_F(p_n) \cong \hat{A}(p_0)\,(\mathbf{Z}_n - \mathbf{Z}_F(p_n))\ , \tag{5.6}$$

where $A_{ij} = \left.\frac{\partial F_i}{\partial Z_j}\right|_{\mathbf{Z}=\mathbf{Z}_F(p_0),p=p_0}$ is the Jacobi matrix.

The Jacobi matrix $\hat{A}$ is characterized by its eigenvectors $\mathbf{e}_u, \mathbf{e}_s$ and eigenvalues λ_u, λ_s,

$$\hat{A}\mathbf{e}_u = \lambda_u \mathbf{e}_u$$

$$\hat{A}\mathbf{e}_s = \lambda_s \mathbf{e}_s\,,$$

where the indices u and s correspond, respectively, to unstable and stable directions of $\mathbf{Z}_F(p_0)$ (see Fig. 5.2b): $|\lambda_s| < 1 < |\lambda_u|$. These eigenvectors are normalized, but they are not orthogonal:

$$\mathbf{e}_s^T\mathbf{e}_s = \mathbf{e}_u^T\mathbf{e}_u = 1, \quad \mathbf{e}_s^T\mathbf{e}_u \neq 0\,, \tag{5.7}$$

where the symbol T denotes the transposition operation. The Jacobi matrix can be presented in the form:

$$\hat{A} = [\mathbf{e}_u \mathbf{e}_s] \begin{pmatrix} \lambda_u & 0 \\ 0 & \lambda_s \end{pmatrix} [\mathbf{e}_u \mathbf{e}_s]^{-1} \ .$$

Because of non-orthogonality of the vectors $\mathbf{e}_u$ and $\mathbf{e}_s$ it is convenient for formulation of the control to introduce a new "orthogonal" basis $\{\mathbf{f}_u, \mathbf{f}_s\}$ (see Fig. 5.2c):

$$\mathbf{f}_s^T \mathbf{e}_s = \mathbf{f}_u^T \mathbf{e}_u = 1; \quad \mathbf{f}_u^T \mathbf{e}_s = \mathbf{f}_s^T \mathbf{e}_u = 0 \,. \tag{5.8}$$

Those are connected to bases by the simple relation

$$\begin{pmatrix} f_{u1} & f_{u2} \\ f_{s1} & f_{s2} \end{pmatrix} = \begin{pmatrix} e_{u1} & e_{s1} \\ e_{u2} & e_{s2} \end{pmatrix}^{-1} \ .$$

From the latter we obtain components for the new basis:

$$\begin{aligned} f_{u1} &= e_{s2}/\Delta, \quad f_{u2} = -e_{s1}/\Delta, \\ f_{s1} &= -e_{u2}/\Delta, \quad f_{s2} = e_{u1}/\Delta; \\ \Delta &\equiv e_{u1}e_{s2} - e_{s1}e_{u2} \,. \end{aligned}$$

The Jacobi matrix can be expressed also in the mixed e,f-basis:

$$\hat{A} = \lambda_u \mathbf{e}_u \cdot \mathbf{f}_u^T + \lambda_s \mathbf{e}_s \mathbf{f}_s^T \,.$$

Projecting this relation on the direction $\mathbf{f}_u$, we obtain a useful result

$$\mathbf{f}_u^T \hat{A} = \lambda_u \mathbf{f}_u^T \,. \tag{5.9}$$

We can now formulate the control condition—getting $\mathbf{Z}_{n+1}$ on the local stable manifold (see Fig. 5.2d) $\mathbf{Z}_F(p_0)$—in the following form:

$$\mathbf{f}_u^T \delta \mathbf{Z}_{n+1} = 0; \quad \delta \mathbf{Z}_{n+1} = \mathbf{Z}_{n+1} - \mathbf{Z}_F(p_0) \,. \tag{5.10}$$

Substituting (5.5) into (5.6) and using (5.9) together with the control condition (5.10), we get the following:

$$\delta p_n = \frac{\lambda_u}{\lambda_u - 1} \frac{\mathbf{f}_u^T \delta \mathbf{Z}_n}{\mathbf{f}_u^T \mathbf{g}} \,. \tag{5.11}$$

Relation (5.11) is the basic formula of the discrete parametric OGY control.

This result can also be represented in an alternative form. Simultaneously accounting the transition $\mathbf{Z}_n \to \mathbf{Z}_{n+1}$ and variation of the parameter $p_0 \to p_0 + \delta p_n$, we can present the dynamics in the neighborhood of the fixed point $\mathbf{Z}_F(p_0)$ in the following form:

$$\delta \mathbf{Z}_{n+1} \simeq \hat{A}(p_0)\delta \mathbf{Z}_n + \mathbf{B}\delta p_n; \quad \mathbf{B} = \left.\frac{\partial \mathbf{F}}{\partial p}\right|_{\mathbf{Z}=\mathbf{Z}_F(p_0),p=p_0}. \tag{5.12}$$

Projecting (5.12) on the direction $\mathbf{f}_u$ and utilizing the control condition (5.10), we get

$$\delta p_n = -\lambda_u \frac{\mathbf{f}_u^T \delta \mathbf{Z}_n}{\mathbf{f}_u^T \mathbf{B}}. \tag{5.13}$$

Vectors $\mathbf{B}$ and $\mathbf{g}$ are connected with the relation

$$\mathbf{B} = \left(1 - \hat{A}\right)\mathbf{g}.$$

The main result of the OGY control method can be presented in the form

$$\delta p_n = C\mathbf{f}_u^T \delta \mathbf{Z}_n, \tag{5.14}$$

which we can interpret in the following way. Deviation of the parameter from its initial value δp_n, necessary to perform the control, is proportional to the projection of the vector $\delta \mathbf{Z}_n$ onto the stable direction $\mathbf{f}_u$. The proportionality coefficient C is calculated from the fixed point shift $\mathbf{g}$ projection onto the same direction and from the unstable eigenvalue λ_u.

Let us now turn to the geometrical interpretation of the obtained result. Figure 5.3 represents the point $\mathbf{Z}_n$, approaching the unstable fixed point $Z_F(p_0)$ along the stable direction $\mathbf{e}_s$. In absence of control in the consecutive moments of time, the point will move off the $Z_F(p_0)$ along the unstable direction $\mathbf{e}_u$. Let us now introduce into the system such parameter perturbation δp_n, that the point $\mathbf{Z}_n$, determining the system position, will appear to lie between the new and old stable directions (Fig. 5.3b). Motion along the new stable direction $\mathbf{e}_s'$ with consecutive moving off the new unstable fixed point $\mathbf{Z}_F(p_0 + \delta p_n)$ along the unstable direction $\mathbf{e}_u'$ will be at the same time a motion towards the old stable fixed point $\mathbf{Z}_F(p_0)$. Therefore, if we properly choose δp_n [according to the OGY formula (5.11)], we can then make it so that the point $\mathbf{Z}_{n+1}$ will get precisely onto the stable manifold $\mathbf{Z}_F(p_0)$. After that we return the parameter to its initial value p_0, and the point describing the system position, remaining on the stable manifold, will approach $\mathbf{Z}_F(p_0)$ (Fig. 5.3c). A schematic three-dimensional analogue of the two-dimensional geometry of the OGY control is presented in Fig. 5.3d.

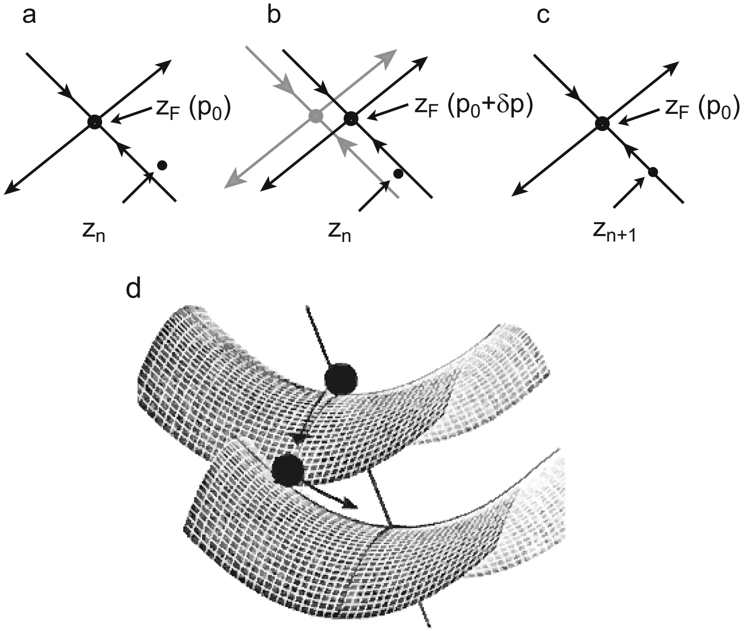


Fig. 5.3 (**a**) Point $\mathbf{Z}_n$ in the neighborhood of unstable fixed point $Z_F\,(p_0)$; (**b**) shift of the fixed point with variation of the parameter $p_0 \to p_0 + \delta p_0$; (**c**) final stage of the OGY control; (**d**) schematic three-dimensional analogue of two-dimensional control geometry [1, 16]

As an example we consider the result of stabilization for the period-1 orbit in the Hénon mapping [17]

$$\begin{aligned} X_{n+1} &= p - X_n^2 + 0.3Y_n \\ Y_{n+1} &= X_n\,. \end{aligned} \tag{5.15}$$

Starting from some randomly chosen initial condition on the attractor, the image point undergoes chaotic walks until at $n \sim 75$ it appears in the given neighborhood of the chosen unstable fixed point (it is marked by a cross in Fig. 5.4a). In that moment we turn on the control algorithm. The result of the control is presented in Fig. 5.4b. Figure 5.4c shows deviations of the parameter p from its nominal value ($p_0 = 1$), necessary to realize the control. In the absence of noise, the parameter deviations are non-zero only on the initial stage of control. Figure 5.4c presents deviation of the orbit δr_n from the unstable fixed point in logarithmic scale. After the control is turned off at ($n \sim 150$) the chaotic motion restores sufficiently quickly.

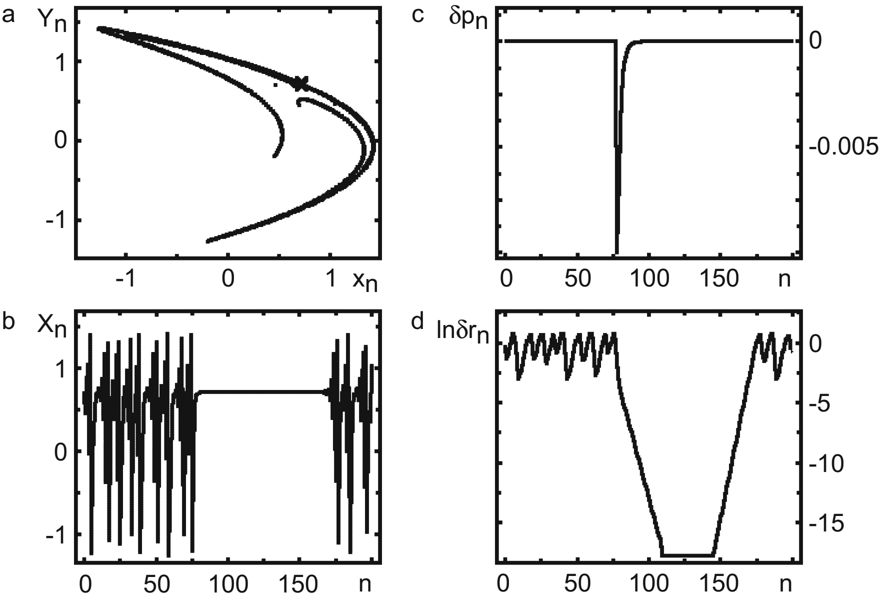


Fig. 5.4 OGY control for the Hénon mapping (without noise)

A logarithmic scale used in Fig. 5.4c sharply marks out all the characteristic stages of the control procedure:

1. chaotic motion up to the control turning on,
2. exponentially fast approach of the controlled trajectory to the unstable fixed point,
3. keeping in the neighborhood of the unstable fixed point with accuracy determined by numerical calculation errors,
4. exponential divergence of trajectories after the control is turned off,
5. restitution of free chaotic motion.

Let us consider in the same example the influence of noise on the described control mechanism. For that purpose, we add in the right-hand sides of the Hénon mapping (5.15) the terms $\varepsilon\delta X_n$ and $\varepsilon\delta Y_n$. Independent random variables δX_n and δY_n are Gaussian-distributed with zero mean values and unit dispersion. Figure 5.5 presents the result of stabilization for the unstable fixed point of the Hénon mapping for $\varepsilon = 0.014$. Even with the presence of noise, the OGY algorithm realizes the stabilization, but with a shortened control interval. In that case, the quantity δp_n is non-zero for whole duration of the control.

In conclusion, let us formulate the main advantages of the OGY discrete parametric control method:

The method requires minimum computational effort.

Realization of the control needs only small variations of the system parameters.

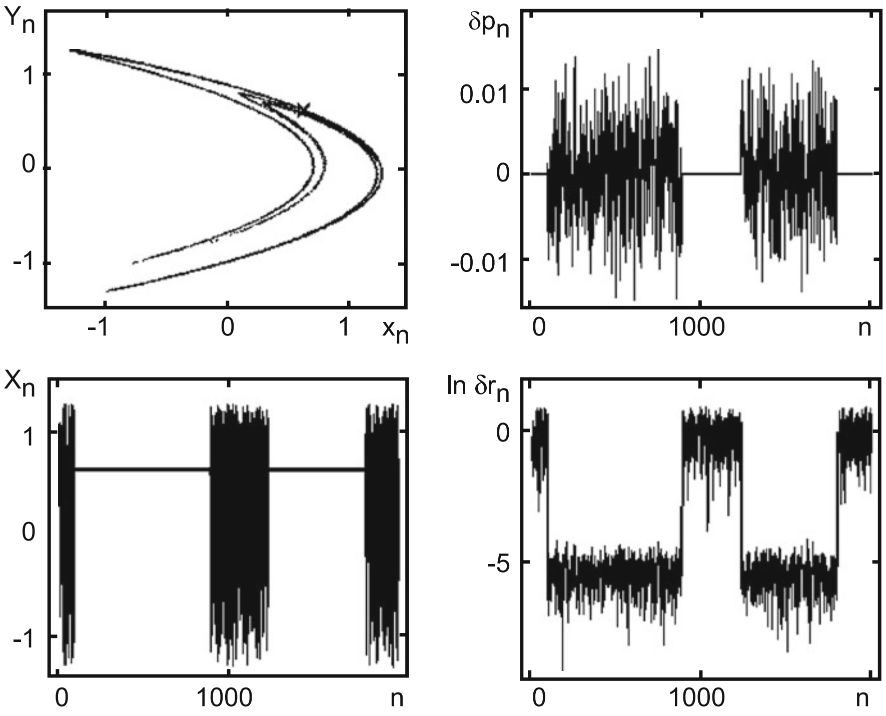


Fig. 5.5 OGY control for the Hénon mapping (with noise)

The control does not change the structure of the unperturbed system phase space.

Different unstable periodic orbits can be stabilized in common region of the parameter space.

The method can be applied to any non-linear system if its evolution allows description in terms of mappings.

The method does not require a priori model description of the dynamics. (The latter remark requires an explanation which will be given in the next section.)

5.4 Control of Chaos Without Motion Equations

The OGY method is based on a chaos control strategy that does not require a priori knowledge of equations of motion for the controlled object. As we have seen, realization of the method only requires knowledge of the local structure of the phase space in neighborhood of the target periodic orbit (or fixed point), i.e., the Jacobi matrix $\hat{A}$ and vector $\mathbf{g}(\mathbf{B})$, which enter into the relation (5.11) and (5.13). It can be shown [1, 18–20] that quantities can be reconstructed without an exact model (or equations of motion) of the controlled system.

This feature makes the OGY method particularly attractive for chaos control in real experiments. Indeed, with rare exceptions, experimentors do not have adequate models of the phenomena under investigation. To begin with, we will make an optimistic assumption that we know a sufficiently long segment of the dynamical system trajectory on the attractor (further on, we will weaken this assumption) and then we show how to reconstruct the information that interests us. Let the trajectory be given in the form of sufficiently long series of intersections $\mathbf{Z}_1, \mathbf{Z}_2 \ldots \mathbf{Z}_n$ with the Poincaré section surface. If two consecutive intersections, for example, $\mathbf{Z}_i$ and $\mathbf{Z}_{i+1}$, appear sufficiently close to each other ($(\mathbf{Z}_{i+1} - \mathbf{Z}_i)^2 \ll l^2$, where l is the characteristic size of the region of finite motion), then, generally speaking, the fixed point must be somewhere nearby. Having fixed the first pair, we will discover other analogous pairs in small neighborhood of the first "almost return." Because of ergodicity of motion on the strange attractor, there will be relatively many such pairs, if the trajectory is known for a sufficiently long time interval. We can try to reproduce the sequence of intersections with the help of linear mapping:

$$\mathbf{Z}_{n+1} = \hat{A}\mathbf{Z}_n + \mathbf{C}\,. \tag{5.16}$$

As noise is always present in the record of a real trajectory, in order to reproduce the matrix $\hat{A}$ and vector $\mathbf{C}$ we should use as many pairs as possible, adjusting the data with the method of least squares. Matrix $\hat{A}$, thus obtained, serves as an approximation of the Jacobi matrix, eigenvectors, and eigenvalues of which are required for the OGY control realization. The corresponding fixed point is approximated by the relation

$$\mathbf{Z}_F = (1 - \hat{A})^{-1}\mathbf{C}\,. \tag{5.17}$$

In order to find the approximate expression for the vector $\mathbf{g}$ one should change the parameter $p \to p + \Delta p$, reproduce the time series (trajectory) with that new value, redefine the fixed point $\mathbf{Z}_F(p + \Delta p)$, and find $\mathbf{g}$ as

$$\mathbf{g} = \frac{\mathbf{Z}_F(p + \Delta p) - \mathbf{Z}_F(p)}{\Delta p}\,. \tag{5.18}$$

To determine the quantities necessary for the stabilization of the period-2 orbit, one should perform an analogous procedure, but for closely intersecting pairs $\mathbf{Z}_n$ and $\mathbf{Z}_{n+2}$, and likewise for higher period orbits.

Let us illustrate the above-described procedure in the example of a non-linear pendulum subject to simplest periodic perturbation [2]. The non-linear pendulum, which for centuries represented the paradigm of periodic motion, is now often used to demonstrate the features of chaotic dynamics. The equation of motion for this system reads

$$\frac{d^2\theta}{dt^2} + k\frac{d\theta}{dt} + \sin\theta = f\cos\Omega t\,, \tag{5.19}$$

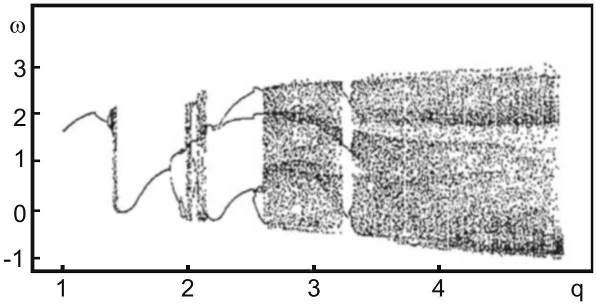


Fig. 5.6 Bifurcation diagram for forced oscillations of a non-linear pendulum [2]

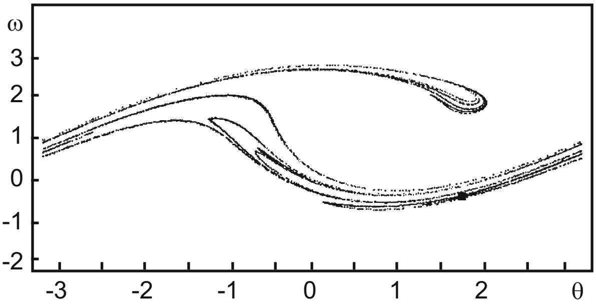


Fig. 5.7 Stroboscopic section for the non-linear pendulum; the *dark square* marks the unstable fixed point [2]

where θ is the angle of deviation from the vertical line, k is the friction coefficient, f and Ω are, respectively, the amplitude and the frequency of the driving force. Depending on the parameter values (k, f, Ω) the pendulum demonstrates different types of dynamical behavior. The bifurcation diagram (Fig. 5.6), which shows the angular velocity $\omega = d\theta/dt$ as a function of the parameter $q = 1/k$, reflects the graduate transition to chaos as the friction coefficient decreases. Further on, we will use the parameters set $(q = 3.9, f = 1.5, \Omega = 2/3)$, at which the pendulum dynamic is chaotic.

For now, let us assume that the equation of motion for the pendulum is unknown to us, but, observing the system experimentally, we can determine the quantities (θ_n, ω_n) in some discrete moments of time $t_n = nT \quad (T = 2\pi/\Omega)$. Laying these points on a plane (θ, ω), we get the so-called stroboscopic section—an analogue of the Poincaré section. This section is presented in Fig. 5.7. For the realization of the OGY control we must extract from the stroboscopic section the following information: the coordinates of the unstable fixed point (θ_F, ω_F); the dependence of the position of that point on the controlling parameter (if that parameter is $-q$), $(\partial\theta_F/\partial q,\ \partial\omega_F/\partial q)$; the Jacobi matrix in the neighborhood of the fixed point, its eigenvectors $\mathbf{e}$ and eigenvalues λ, the orthogonal basis $\mathbf{f}$. Using the relations (5.16) and (5.17) for the set of points in the neighborhood $(1.5, 0.4)$ (the dark square on

Fig. 5.7), we find for the fixed point that

$$Z_F = \begin{pmatrix} \theta_F \\ \omega_F \end{pmatrix} = \begin{pmatrix} 1.523 \\ -0.415 \end{pmatrix}$$

and for the Jacobi matrix

$$\hat{A} = \begin{pmatrix} -3.42 & -5.79 \\ -1.52 & -2.48 \end{pmatrix} .$$

Eigenvalues and normalized eigenvectors for that matrix read

$$\lambda_u = -5.85, \quad \mathbf{e}_u = (e_{u1}, e_{u2}) = (0.92, 0.40) ,$$
$$\lambda_s = +0.050, \quad \mathbf{e}_s = (e_{s1}, e_{s2}) = (0.86, -0.52) .$$

Using the relations (5.8) we can also find the orthogonal basis components necessary for control realization,

$$\mathbf{f}_u = (f_{u1}, f_{u2}) = (0.63, 1.04) ,$$
$$\mathbf{f}_s = (f_{s1}, f_{s2}) = (0.49, -1.12) \; .$$

At last, we can determine how the variation of the friction coefficient affects the position of the fixed point. For small changes of the parameter q

$$\begin{pmatrix} \theta'_F \\ \omega'_F \end{pmatrix} \approx \begin{pmatrix} \theta_F \\ \omega_F \end{pmatrix} + \delta q \begin{pmatrix} \frac{\partial \theta_F}{\partial q} \\ \frac{\partial \omega_F}{\partial q} \end{pmatrix} = \mathbf{Z}_F + \mathbf{g}\delta q \, .$$

To determine vector **g** we shall follow the variation of the fixed point position with changes of parameter q. Having constructed the graphical dependencies $\theta(q)$ and $\omega(q)$, we can determine the components of the vector **g**. We should note that the OGY control mechanism is not very sensitive to that parameter, therefore, in order to determine the vector **g** we can restrict ourself with a small number of dimensions. Now we have all the components necessary for the realization of the OGY control with the help of relation (5.11).

We should stress that we got all the necessary information only from the experimentally observable quantities $\theta(t)$, $\omega(t)$. The result of control for the period-1 unstable orbit (fixed point) is presented in Fig. 5.8. The control was turned on in the vicinity of the 1000th period of the external perturbation and was turned off near the 3000th period. About ten cycles were required to get the control. Only small variations of the controlling parameter $|\delta q| < 0.1$ were allowed during the control process. Large parameter changes could transfer the system into another dynamical regime (see the bifurcation diagram in Fig. 5.6). As we can see, during the control

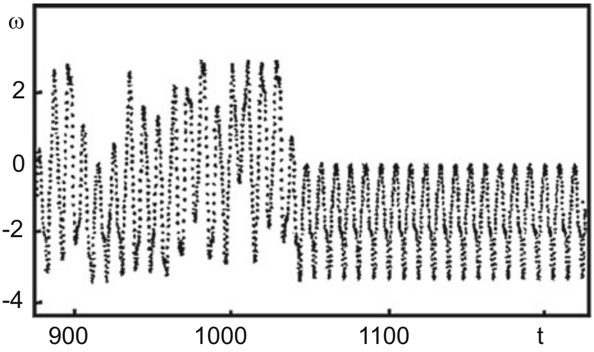


Fig. 5.8 Control of unstable fixed point for non-linear pendulum [2]

time, we were able to keep the chaotic trajectory in the vicinity of the target periodic orbit.

We should, however, note that it is difficult to measure experimentally the full N-dimensional vector of the system state in a given moment of time, but it is exactly this information which is required for the above procedure. As a limiting case, consider the situation where only one scalar system characteristic $f(t)$ is available for measurement. As Chap. 4 showed, it is possible to reconstruct the full dynamics of N-dimensional system from a single scalar characteristic.

5.5 Targeting Procedure in Dissipative Systems

In the control scheme considered above, with a limited interval of the controlling parameter variation ($|p_n - p_0| \leqslant \delta$) the control is turned on only after the trajectory being stabilized gets into ε-neighborhood ($\varepsilon \sim \delta$) of some component of the target periodic orbit. The efficiency of this control scheme is determined to a great extent by the time it takes to get into the required region or, as we shall say, control setup time.

Average time $\langle \tau \rangle$ required to get in the ε-vicinity of some point during chaotic motion on the strange attractor [1]

$$\langle \tau \rangle \sim \varepsilon^{-D}, \tag{5.20}$$

where D is the fractal dimension of the attractor (see [21, 22]). Therefore, if we do not make special efforts, the decrease of the allowed region of parameter variation will result in power-law growth of the control setup time. However, right after the appearance of the OGY control method, a procedure was proposed [23, 24], named “targeting” by its authors, which by special small change of the controlling parameter permitted the transformation of the control setup time growth law from the power one to the essentially slower law—a logarithmic one. The procedure uses the exponential sensitivity of the chaotic trajectory to the initial conditions.

Let us discuss the targeting procedure in the simplest setup, when the attractor dimension is close to one [23, 24]. Generalizations on cases of higher dimension can be found in [25]. Suppose we have a time-continuous dynamical system, described by the equations of motion $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$. According to the usual scheme with the help of Poincaré sections we transit from the continuous equations of motions to the time-discrete reversible mapping

$$\mathbf{Z}_{n+1} = \mathbf{F}(\mathbf{Z}_n, p)\,. \tag{5.21}$$

Let us remember that if the equations of motion are not given, the Poincaré section can be reconstructed from experimental data. Suppose we want, starting from the point $\mathbf{Z}_s$, to get a small vicinity (with linear dimensions ε_t) of the point $\mathbf{Z}_t$. From now on, we will call this point the target. As usual, we assume that the system parameter p is subject only to small tuning on each iteration step:

$$p_n = p_0 + \delta p_n; \quad -\delta < \delta p_n < \delta; \quad |\delta| << p_0\,.$$

On the first iteration we include a small variation of the parameter $-\delta_1 < \delta p_1 < \delta_1$. Iterating (5.21) with values of p from the interval $[p_0 - \delta_1, p_0 + \delta_1]$, we get some segment ξ, passing through the point $\mathbf{F}(\mathbf{Z}_s, p_0)$. Let us denote length of that segment as $\delta\xi$. After that, we return back to the initial value of the parameter p_0. As the system is chaotic, the segment length will exponentially grow with each consecutive iteration. At long last, say, after η_1 iterations, it will reach size of the system. Without restricting the generalization, we can consider the linear dimensions of the attractor to be of the order of one. Then

$$\eta_1 \sim \lambda_1^{-1} \ln \delta\xi^{-1}\,. \tag{5.22}$$

Here λ_1 is the positive Lyapunov exponent of the mapping F. Similarly, if we will iterate vicinity of the target ε_t back in time, we find that the number of iterations required for that region to stretch up to the attractor dimensions equals

$$\eta_2 \sim |\lambda_2|^{-1} \ln \varepsilon_t^{-1}\,, \tag{5.23}$$

where λ_2 is the negative Lyapunov exponent of the mapping $\mathbf{F}^{-1}$. As both objects (the segment and the target vicinity) are stretched up to the attractor dimensions, we can find their intersection point. Iterating it η_1 times back in time, we find the point on the segment $\delta\xi$, which after $\eta_1 + \eta_2$ iterations maps into the target vicinity with linear dimensions ε_t. The total time required for this is

$$\tau = \lambda_1^{-1} \ln \delta\xi^{-1} + |\lambda_2|^{-1} \ln \varepsilon_t^{-1}\,. \tag{5.24}$$

Setting $\delta\xi \sim \varepsilon_t$, we obtain

$$\tau \sim \ln \varepsilon_t^{-1}\,, \tag{5.25}$$

contrary to the power-law growth without the targeting procedure.

The logarithmic behavior of the control setup time was confirmed in the following numerical experiment [23]. The source and the target were randomly chosen on the attractor of the Hénon mapping of the following form:

$$\begin{aligned} X_{n+1} &= p - X_n^2 + 0.3Y_n \\ Y_{n+1} &= X_n\,. \end{aligned}$$

Then the target size was fixed at ε_t, and the above targeting algorithm was applied for each source-target pair using p as the controlling parameter ($p_0 = 1.4$). The calculated time required to get the target was averaged over an ensemble of the source-target pairs at fixed target size ε_t. The results of the numerical experiment are presented in Fig. 5.9. The solid line with slope $\lambda_1^{-1} + |\lambda_2|^{-1}$, predicted by (5.24), agrees with the obtained data. The dashed line corresponds to the power law (5.20) with $D \cong 1.26$ (the fractal dimension of the Hénon attractor). The variation of the parameter in realizing the targeting procedure did not exceed 0.1 % of its initial value.

Peculiarities of the targeting procedure in Hamiltonian systems will be considered in the next section.

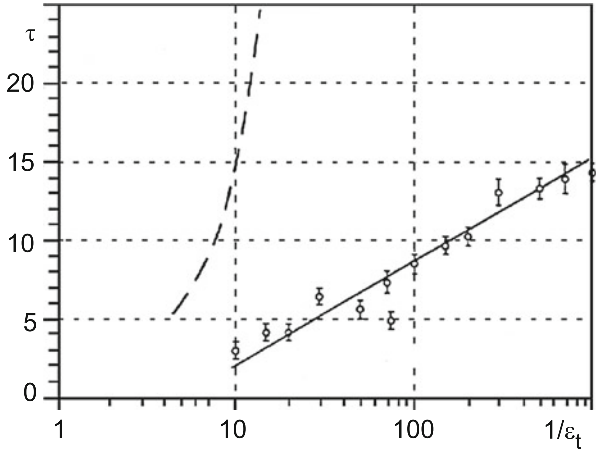


Fig. 5.9 Average time required to get the target of definite size (ξ_t); *solid line* shows the time defined by the relation (5.6), *dotted line* is the time in the absence of targeting [power law (5.1)] [23]

5.6 Chaos Control in Hamiltonian Systems

In this section we will consider an OGY method generalization that allows the realization of chaos control in Hamiltonian systems [26]. There are several reasons that make this generalization a non-trivial task.

Because of the phase volume conservation in Hamiltonian systems, some unstable periodic orbit components have the Jacobi matrix with complex eigenvalues. It makes it impossible to use the formulae (5.11), (5.13), expressed in terms of real eigenvalues immediately for control. We can utilize the unmodified OGY algorithm if we apply the controlling perturbation only over the period, i.e., on each *m*th step, if the periodic orbit has period *m*. However, the stabilized chaotic orbit, affected by noise, can deviate from the target orbit before the next perturbation will be applied, and control over the trajectory will be lost. Therefore, for a real system, where noise is always present, an efficient control algorithm must allow control on each time step. The initial control algorithm needs to be slightly modified. Let us do it for the two-dimensional mapping $\mathbf{Z}_{n+1} = \mathbf{F}(\mathbf{Z}_n, p)$ with the usual limitation, imposed on the smallness of the parameter p perturbation. The linearized dynamics in the vicinity of the period-*m* orbit ($\mathbf{Z}_{01} \to \mathbf{Z}_{02} \to \ldots \mathbf{Z}_{0m} \to \mathbf{Z}_{0(m+1)} = \mathbf{Z}_{01}$) reads:

$$\mathbf{Z}_{n+1} - \mathbf{Z}_{0(n+1)}(p_0) = \hat{A}(\mathbf{Z}_n - \mathbf{Z}_{0n}(p_0)) + \mathbf{B}\delta p_n\,. \tag{5.26}$$

Here we will not, as we did before, express the matrix $\hat{A}$ in terms of its eigenvectors and eigenvalues, as they can be complex in some points of the periodic orbit. Instead, we shall use stable and unstable directions, connected with each periodic orbit component. If $m \neq 1$, then these directions do not necessarily coincide with the eigenvectors of the Jacobi matrix at the same point. The algorithm for determining the stable and unstable directions for periodic orbit components in two-dimensional mappings can be found in [26].

Suppose $\mathbf{e}_{s(n)}$ and $\mathbf{e}_{u(n)}$ are, respectively, stable and unstable directions in the point of the periodic orbit Z_{0n} , and $\mathbf{f}_{s(n)}$, $\mathbf{f}_{u(n)}$ are two vectors satisfying the conditions

$$\begin{aligned} \mathbf{f}_{u(n)}^T \mathbf{e}_{u(n)} &= \mathbf{f}_{s(n)}^T \mathbf{e}_{s(n)} = 1 \\ \mathbf{f}_{u(n)}^T \mathbf{e}_{s(n)} &= \mathbf{f}_{s(n)}^T \mathbf{e}_{u(n)} = 0\,. \end{aligned} \tag{5.27}$$

For the stabilization of an unstable periodic orbit we require that the point, which, as the result of evolution appeared in small vicinity of some periodic orbit component $\mathbf{Z}_{0n}$, will, on the next $(n+1)$ iteration, get on the stable direction of the component $\mathbf{Z}_{0(n+1)}$. It means that

$$\mathbf{f}_{u(n+1)}^T \left(\mathbf{Z}_{n+1} - \mathbf{Z}_{0(n+1)}(p_0)\right) = 0\,. \tag{5.28}$$

Projecting the relation (5.26) on the direction $\mathbf{f}^T_{u(n+1)}$ and using the condition (5.28), we get [26]:

$$\delta p_n = -\frac{\mathbf{f}^T_{u(n+1)}\left[\hat{A}\delta\mathbf{Z}_n(p_0)\right]}{\mathbf{f}^T_{u(n+1)}\mathbf{B}}; \quad \delta\mathbf{Z}_n(p_0) = \mathbf{Z}_n - \mathbf{Z}_{0n}(p_0)\,. \tag{5.29}$$

This formula represents an analogue of the relation (5.11) for the OGY chaos control method in Hamiltonian systems. So for the case of the unstable fixed point stabilization $\mathbf{f}^T_{u(n+1)} = \mathbf{f}^T_u$, $\mathbf{f}^T_u\hat{A} = \lambda_u f_u^T$ the relation (5.29) also transforms into (5.11). It is important to note that the parameter perturbation (5.29) is applied to the system on each time step, which minimizes the influence of external noise.

The obtained algorithm was applied in [27] for chaos control in a version of the already considered standard mapping

$$\begin{aligned} X_{n+1} &= (X_n + Y_n) \bmod 2\pi \;\; - \pi \\ Y_{n+1} &= Y_n + p\sin(X_n + Y_n)\,, \end{aligned} \tag{5.30}$$

using p as the controlling parameter. Figure 5.10 shows the results of control for the period-10 unstable periodic orbit. Anomalously long control setup times—about 10^4 iterations—are striking. This is one more difficulty in the realization of OGY control in Hamiltonian systems. In dissipative chaotic systems the average control setup time $\langle\tau\rangle$ is always finite. It is connected with the exponential decay of the distribution function $P(\tau)$ on long times [28]

$$P(\tau) \sim \exp\left[-\tau/\left\langle\tau\right\rangle\right]\,. \tag{5.31}$$

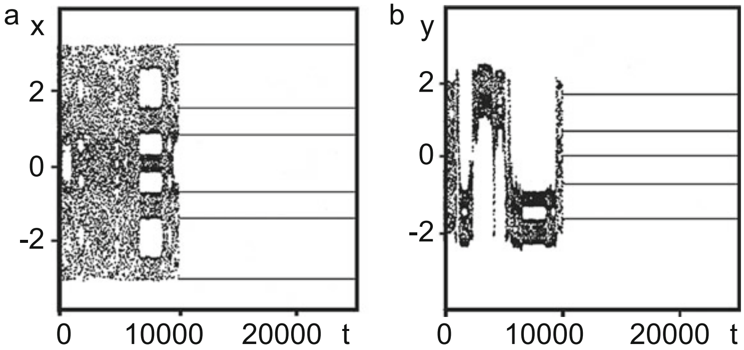


Fig. 5.10 OGY control for the period-10 unstable periodic orbit in the standard mapping (5.30). Only some of the lines corresponding to the periodic orbit are shown. Other lines, when projected on the corresponding planes, have coordinates that coincide with the plotted ones [26]

In Hamiltonian systems, the corresponding distribution function decays considerably slower on long times [29]

$$P(\tau) \sim \tau^{-\alpha}; \quad 1 < \alpha < 2\,. \tag{5.32}$$

This leads to the fact that the average control setup time in Hamiltonian systems

$$\langle \tau \rangle \sim \int \tau^{1-\alpha} d\tau\,, \tag{5.33}$$

tends to infinity. The physical reason for such distribution function behavior is the sticking effect of the trajectory to the invariant tori surviving in the phase space. Therefore, efficient control in Hamiltonian systems can be realized only under conditions of considerable abridgement of the control setup time.

Let us briefly cite one of ways to solve that problem, proposed in the paper [27]. For explanation of the method the authors used the following analogy. Suppose in some mountainous country you must get from one valley to another. If you are not acquainted with the landscape and try to achieve the goal by the random walking, then the march will take a considerable time. The required time can be remarkably shortened if you use the passes connecting the neighboring valleys. Therefore, the authors named their method the pass targeting method.

Let us explain it using the example of two-dimensional Hamiltonian mapping. The phase space structure of a system corresponding to such a mapping in the region of transition from absolute regularity to complete chaos represents a chain of resonance overlaps [30]. This picture is schematically illustrated in Fig. 5.11, where two overlapping resonances are shown. Each resonance is associated with an orbit of a certain period. For example, the unstable fixed point (the saddle point P) is associated with the period-1 resonance (lower hatched region in Fig. 5.11), and the unstable period-2 orbit (the saddle points P_1 and P_2)—with the period-2 resonance (upper hatched region in Fig. 5.11). To transition from one resonance to another, it is necessary to intersect the region of neighboring resonance overlap. This is the pass

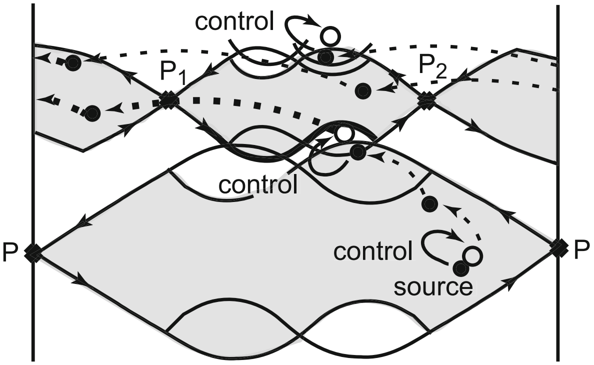


Fig. 5.11 Targeting procedure in a Hamiltonian system [standard mapping (5.30)] in the case of two overlapping resonances [27]

used in the above analogy. Thus, the targeting procedure in Hamiltonian systems must have a multistage character. As the intermediate target on each stage one can choose the neighboring resonance overlap region. The authors of the method checked its efficiency on the standard mapping (5.30). The average time to reach the target, separated from the source by seven resonances for 50 sets of initial conditions, chosen in the chaotic region using the targeting procedure, was within the limit of $125-132$ iterations. The uncontrolled transport time for the same source-target combination was from 1119 to 3.77 million iterations.

As a realistic example of the targeting procedure realization in Hamiltonian systems we shall briefly mention the so-called restricted three-body problem [31]: the description of the motion of a light body in the gravitational field of two other bodies, significantly exceeding it in mass. The heavier bodies turn around the common center of gravity under action of mutual attractive forces. Such a model can be used to describe spaceship dynamics in the Earth–Moon gravitational field. The solution obtained in framework of such a model is used for the zero approximation. Subsequent approximations account for the influence of the Sun and other planets.

Let our goal be to transfer the spaceship from a near-earth orbit to a circumlunar one. The straightforward way to achieve that goal is to accelerate the spaceship in order to let it leave the near-earth orbit and then to slow it down for the capture by the Moon's gravity field.

A very different approach [32] is based on the existence of a chaotic sea between the Earth and the Moon (due to the stochasticity of the reduced three-body problem). In that case, a small quantity of rocket fuel can be used to transfer the spaceship from the near-earth orbit into the chaotic sea. Then the spaceship can reach the vicinity of the circumlunar orbit without any fuel losses. However it will take very long time—about 27 years. Using the above targeting procedure in a Hamiltonian system, this time can be shortened to 293 days with multiplied fuel savings [27].

5.7 Stabilization of the Chaotic Scattering

In the present section we will, following [33], consider one more example of controlled Hamiltonian dynamics, but now for cases of infinite motion—chaotic scattering. This represents a type of scattering at which arbitrarily small changes of input variables can result in considerable output changes. In other words, as in any chaotic process, chaotic scattering is characterized by an anomalous sensitivity to initial conditions.

We begin by formulating the problem. An arbitrary particle impacting the scatterer will, generally speaking, stay only a finite time in the scattering region. However, in many important applications (chemical and nuclear reactions, channeling relativistic particles in crystals) it is necessary to keep the particle in the scattering region for longer. Therefore, we naturally come to the following: how can we keep a particle inside the scattering region as long as needed, using only small variations in the system parameters? This task is equivalent to the problem of unstable periodic orbit stabilization inside the scattering region.

Below we will briefly discuss this problem in application to the non-hyperbolic chaotic scattering in Hamiltonian systems. The term “hyperbolic scattering” means scattering in a case when all the periodic orbits are unstable and the invariant tori are absent in the scattering region. At the same time the term “non-hyperbolic chaotic scattering” describes the situation when the surviving invariant tori coexist with the chaotic invariant sets.

Control of non-hyperbolic chaotic scattering has two characteristic features. First let us remember that for strange attractors, the probability of finding a particle in a small vicinity of the target periodic orbit equals unity. However, in the case of chaotic scattering the invariant chaotic set is not an attractor. Therefore, in order to get a finite probability of finding a particle in the vicinity of the target orbit, we should prepare the ensemble of initial conditions, corresponding to motion towards the chaotic set.

Another peculiarity is immediately connected to the non-hyperbolic character of the scattering. If the target unstable periodic orbit is situated far from the invariant tori present in the scattering region, the latter will only slightly affect the average control setup time. However, if the orbits situated near the surviving tori are stabilized, the sticking effect mentioned in the previous section may appear significantly stronger than in the first case.

Let us study the possibility of controlling the chaotic scattering in a simple model, describing the one-dimensional dynamics of a particle driven by δ-like pulses [34]. As the controlling parameters in this model we can use the intensity of the pulses and the time interval between two consecutive collisions. The Hamiltonian of the model reads

$$H(x,p,t) = \frac{p^2}{2m} + T_0 G(x) \sum_{i=-\infty}^{\infty} \delta(t - T_i)\,, \tag{5.34}$$

where T_0 is a constant, The sequence $\{T_i\}$ determines the moments of the pulses, and $T_0 G(x)$ is the pulse amplitude at point x. Suppose $\{x_n, p_n\}$ are the dynamical variables of the particle before the nth pulse. Then, immediately before the $(n+1)$th pulse, those dynamical variables are defined by the following Hamiltonian (area-preserving) mapping:

$$\begin{aligned} p_{n+1} &= p_n - T_0 \frac{dG(x_n)}{dx_n} \\ x_{n+1} &= x_n + T_n P_{n+1}\,, \end{aligned} \tag{5.35}$$

where T_n is the time interval between nth and $n+1$th pulse.

In order to make the model (5.34) describe the scattering dynamics, we should take the function $G(x)$ such that the derivative $dG(x)/dx$ turns to zero over long distances. Let us choose $G(x)$ in the form

$$G(x) = D(1 - e^{-\alpha x})^2\,, \tag{5.36}$$

where D, α are free parameters. After the following scaling transformation:

$$p_n \to p_n/(\alpha T_0), \quad x_n \to x_n/a$$

the mapping (5.35) takes the form

$$\begin{aligned} p_{n+1} &= p_n - d\left(e^{-x_n} - e^{-2x_n}\right) \\ x_{n+1} &= x_n + \frac{T_n}{T_0} p_{n+1}\,, \end{aligned} \tag{5.37}$$

where $d = 2\alpha^2 T_0^2 D$.

As was shown in [33], the mapping (5.37) demonstrates different types of dynamical behavior depending on the values of the parameters d and T_n. In particular, for $T_n = T_0$ the mapping reproduces both hyperbolic and non-hyperbolic scattering at different values of d. In the case $0 < d < d_c \approx 4.58$ the scattering is non-hyperbolic, because the phase space contains the invariant tori. Figure 5.12

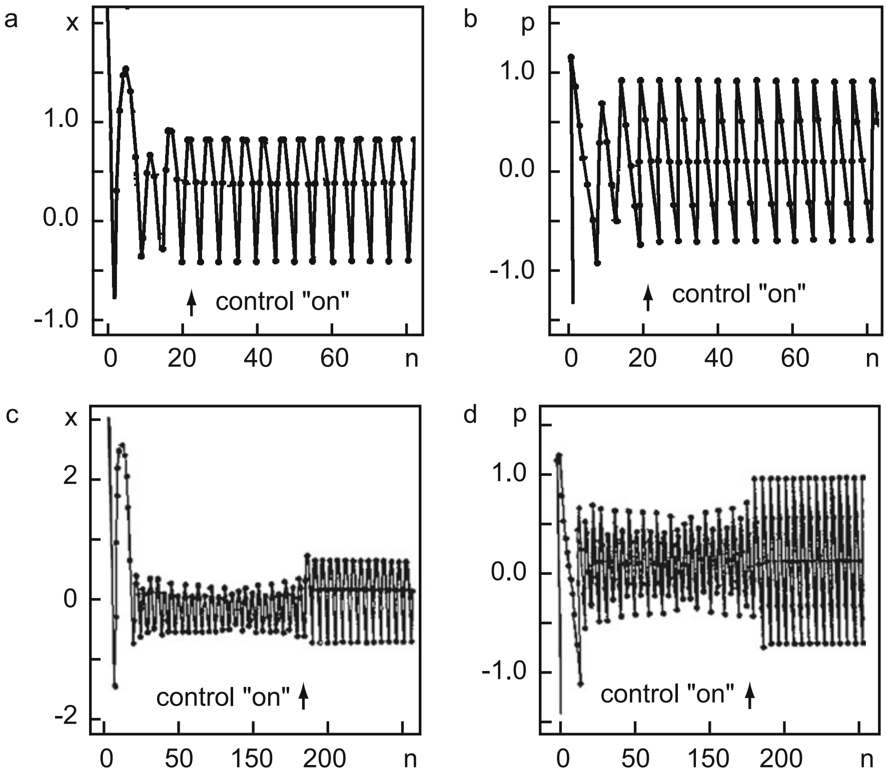


Fig. 5.12 Two examples of the OGY control for the chaotic scattering [33] in the model (5.37). (**a**), (**b**): $X_0 = 8,\ P_0 = -4.398$; (**c**), (**d**): $X_0 = 8,\ P_0 = -9.072$

presents the results of control for period-5 unstable periodic orbit ($d = 1.8$) with the algorithm (5.29) for two sets of initial conditions. The relatively longer period of control setup in the second case is connected with influence of the surviving invariant tori, mentioned above.

5.8 Control of High-Periodic Orbits in Reversible Mapping

In the present section we will demonstrate the efficiency of the discrete parametric control method for the stabilization of high-period orbits in reversible mappings, which we introduced in Chap. 4. As was mentioned above, the specific feature of these systems is that the basic elements of Hamiltonian systems (e.g., resonances) and those of the dissipative systems (e.g., attractors) can coexist in their phase space [35, 36]. The coexistence of those elements broadens the circle of physical phenomena which can be realized in reversible systems compared with Hamiltonian or dissipative ones.

Let us consider a simple reversible system—two-dimensional two-parametric (a, ε) mapping, describing the discrete dynamics of a linear oscillator subject to δ-like pulses with the stiffness coefficient proportional to the velocity:

$$\mathbf{r}_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \mathbf{F}(\mathbf{r}_n) = \begin{pmatrix} x_n + y_{n+1} \bmod 2 \\ y_n - \varepsilon(a - y_n)x_n \end{pmatrix} . \tag{5.38}$$

The phase space for this mapping is the cylinder $x \in (-1, 1)$, $y \in R$; the values $x = -1$ and $x = 1$ are identified. The variable x_n plays the role of the angular coordinate. The mapping (5.38) has fixed points $P_k^s = \left(x_k^s, y_k^s\right)$, where $x_k^s = 0$ and $y_k^s = 2k$ ($k = \pm 1, \pm 2, \ldots ; a \neq 2k$). For fixed values of ε and a the solutions of the characteristic equation

$$\lambda^2 + \lambda \mathrm{SpA} + \det A = 0\,, \tag{5.39}$$

determine the type of the fixed points. Here $A(\mathbf{r}_k) = (\partial \mathbf{F}/\partial \mathbf{r})_{\mathbf{r}=\mathbf{r}_k}$ is the Jacobi matrix of the mapping (5.38). The characteristic equation (5.39) is obtained as the result of linearization of (5.38) in vicinity of the fixed point. It easy to see that

$$\det A = 1, \qquad \mathrm{SpA} = 2 - \varepsilon\left(a - y_k^s\right) . \tag{5.40}$$

A compact classification of fixed points depending on the SpA and $\det A$ values is presented in Fig. 5.13. The condition $\det A = 1$ means that there are only hyperbolic (saddles) or elliptic (centers) fixed points, that is, precisely those phase space elements which exist in Hamiltonian systems. The phase portrait of the mapping (5.38) contains one more important element absent in Hamiltonian

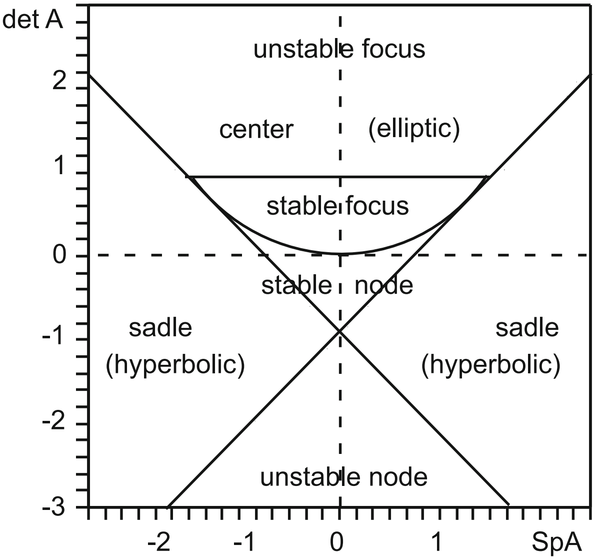


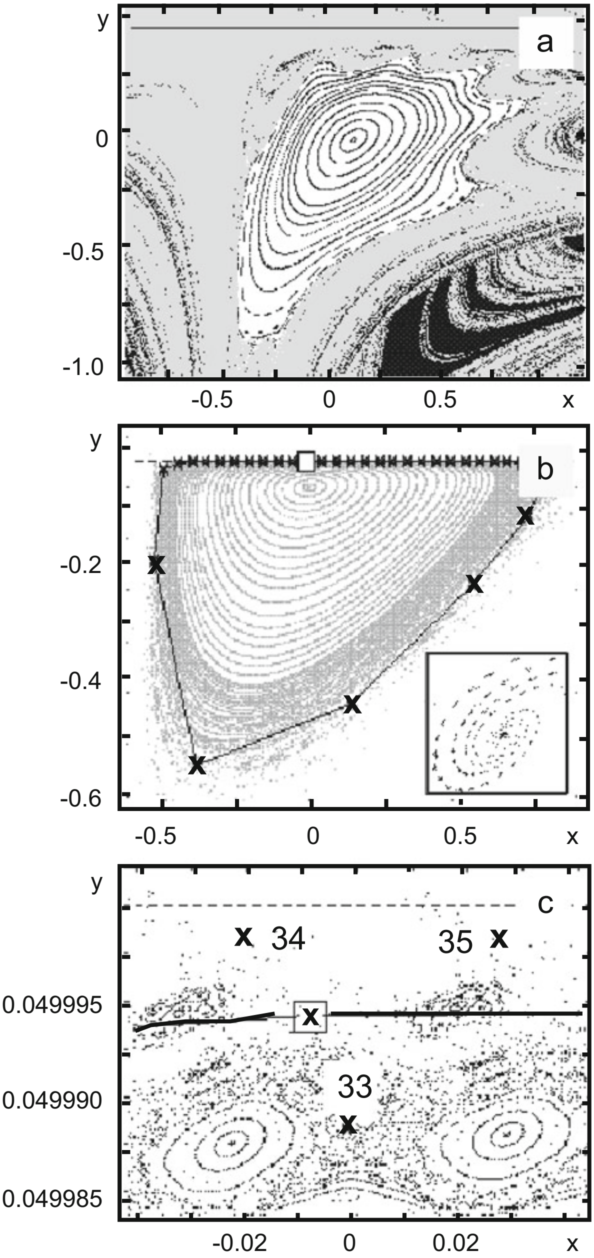
Fig. 5.13 Compact classification of fixed points depending on the SpA and det$A$ values

systems—the invariant set consisting of the family of singular solutions

$$y_n = a, \quad x_{n+1} = (x_n + a) \mod 2\,. \tag{5.41}$$

For the fixed value of a each one of the solutions (5.41) (they differ one from another in the choice of the initial condition x_0) represents a periodic or quasi-periodic trajectory for rational and irrational values of a, respectively. For $\varepsilon < 1$ the invariant set (5.41) attracts nearby trajectories with the increment $\gamma \cong \frac{\varepsilon^2}{6}$ and therefore it can be considered an attractor [36]. The region of attraction to the attractor has a complicated fractal structure. Along with the regular component, the phase space of the mapping (5.38) also contains a chaotic one. The scenario of the transition to chaos in reversible systems is distinct from those that are observed both in dissipative and Hamiltonian systems. On the one hand, it is connected to the absence of a strange attractor, and on the other hand, to the fact that the trajectories are attracted by the attractor at $y_n = a$ for any arbitrarily small ε value, that does not allow to realize in full measure the resonance overlapping scenario [37], characteristic for Hamiltonian systems. Interaction of the attractor with the periodic trajectories, surrounding the elliptic fixed point, determines the specifics of transition to chaos in the considered mapping. Figure 5.14a shows a fragment of the considered mapping with a stability island in vicinity of the point $(x, y) = (0, 0)$. As the island and the attractor come together, i.e., at the decreasing of the parameter a (Fig. 5.14b), the destruction of high-order resonance separatrices and the formation of the stochastic layer takes place (Fig. 5.14c). Numerical calculations [38] show

Fig. 5.14 (**a**) A fragment of phase space of the mapping (5.38) with a stability island in vicinity of the point $(x, y) = (0, 0)$; (**b**) deformation of the stability island at its approach to the attractor; (**c**) destruction of high-order resonances separatrices and formation of the stochastic layer for the unstable periodic orbits with $k = 34$ [a fragment of the phase space corresponding to the *white square* on (**b**)]



that at $k \sim 30$ (k is the order of resonance or of the orbit) for $a = 0.05$ widths of the resonances and distances between them become of the same order. According to Chirikov's criterion of non-linear resonances, this means that in the region of higher k values the transition to global stochasticity must be observed. However, unlike Hamiltonian systems, where the resonance width is determined only by the non-integrable perturbation amplitude, in reversible mapping (5.38) the reason of the transition to chaos is the approach (interaction) of Hamiltonian and dissipative phase space elements: namely of the stability island and of the attractor.

Even this simple analysis allows the dynamical system (5.38) to be related to the class of the so-called complex systems [39], which are characterized by the following main features:

1. a complex system is structurally inhomogeneous;
2. individual components of a complex system can be both regular and chaotic;
3. a complex system has a space-time scale hierarchy.

Because of this structural complexity, we can expect that even a weak perturbation applied to the system results in transitions between the different components. Therefore, it seems natural to use the parametric control method to stabilize chaotic regimes in reversible mappings like (5.38) [38, 40].

Before discussing the control problem, we need to find an adequate method for localizing the unstable high-period orbits that interest us. The traditional methods based on the Newton–Rafson procedure are not efficient in cases of unstable orbits because they require highly precise initial conditions needed to perform the iteration procedure. An alternative method [11], which was already mentioned above, implies the preliminary linear transformation of coordinates, which transforms the unstable periodic orbits into stable ones, preserving their position in space. After that, the position of the stable periodic orbits (in new coordinates) can be determined with the help of simple iteration procedures. For the considered two-dimensional mapping, the coordinate transformation has the following form:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \Lambda_i \left[\mathbf{F}^k(\mathbf{r}_n) - \mathbf{r}_n\right] ,$$

where k is the period of the considered orbit ($\mathbf{r} \rightarrow \mathbf{r}_2 \rightarrow \dots \mathbf{r}_k \rightarrow \mathbf{r}_{k+1} = \mathbf{r}_1$), Λ_i is one of $\alpha_2 = 8$ $(i = 1, 2, \dots, 8)$ reversible 2×2 matrices. In D-dimensional space $\alpha_D = D!2^D$. The concrete form of Λ_i is determined by type of the corresponding unstable point. The inset in Fig. 5.14b demonstrates an example of the transformation that transfers the saddle point into a stable focus. As a control object we take the unstable periodic orbit of the mapping (5.38) with $k = 34$, lying at $a = 0.05$ in the global stochasticity region (see Fig. 5.14c). For stabilization we will use the main formula of the discrete parametric control (5.11), taking as p one of the parameters a or ε. Figure 5.15 shows in action the basic mechanism of the used control method. We took four trial points (black squares) in the vicinity of a randomly chosen saddle point, belonging to the period-34 unstable orbit. The trajectories of the four trial points are shown after three consecutive iterations. After the third iteration all four trial points are already lined up along the stable

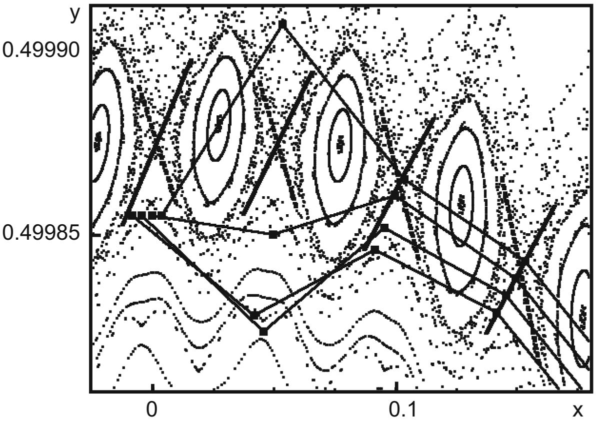


Fig. 5.15 OGY control mechanism: temporal evolution of four trial points

direction. After consecutive iterations they stay on the stable direction approaching the saddle point after each iteration. Figure 5.16 shows the behavior of the deviation $\left|\mathbf{r}_n - \mathbf{r}_n^*\right|$ of the system position $\mathbf{r}_n$ from the periodic orbit $\mathbf{r}_n^*$. We use a logarithmic scale in order to follow all the control stages: the chaotic oscillations preceding the control setup, the exponentially fast approach to the target period orbit, the stable motion along the periodic orbit $\left|\mathbf{r}_n - \mathbf{r}_n^*\right| \sim 10^{-15}$, the exponentially fast deviation from the target orbit after turning off the control, and the restitution of the chaotic oscillations. As in the previously considered cases of the OGY control of dissipative and Hamiltonian systems, the analogous reversible system control method appears to be relatively steady with respect to external noise. Figure 5.16b gives the result of the control with the inclusion of the term $s\boldsymbol{\xi}_n$ on the right-hand side of the mapping (5.38). The components $\xi_{x,y;n}$ represent independent Gaussian random variables with zero mean and unit dispersion. The action of noise considerably lowers the control efficiency, but even in this case the method allows us to keep the chaotic trajectory in the vicinity of the unstable periodic orbit during the time period of almost the same order of magnitude as in the absence of noise. At first glance it seems that the results of the high-period orbits control in the reversible mapping are quite similar to the corresponding results for the Hamiltonian systems. However, more careful consideration [40] shows that the coexistence of attractor and stability islands, which is a characteristic feature of reversible systems, substantially complicates the situation. As it was mentioned many times previously, the control is turned on only when the trajectory $\mathbf{r}_n$ gets into a region sufficiently close to the target periodic orbit. Let us call it the capture region. The capture region size and its shape are determined by the maximum admissible value of the controlling parameter deviation $\delta p_{\max}$ from its nominal value and by local characteristics of the periodic orbit. The basic formula of OGY control (5.11) can be presented in the form

$$\delta p_n = M_i \delta x_n + N_i \delta y_n; \quad i = (n) \bmod k\,.$$

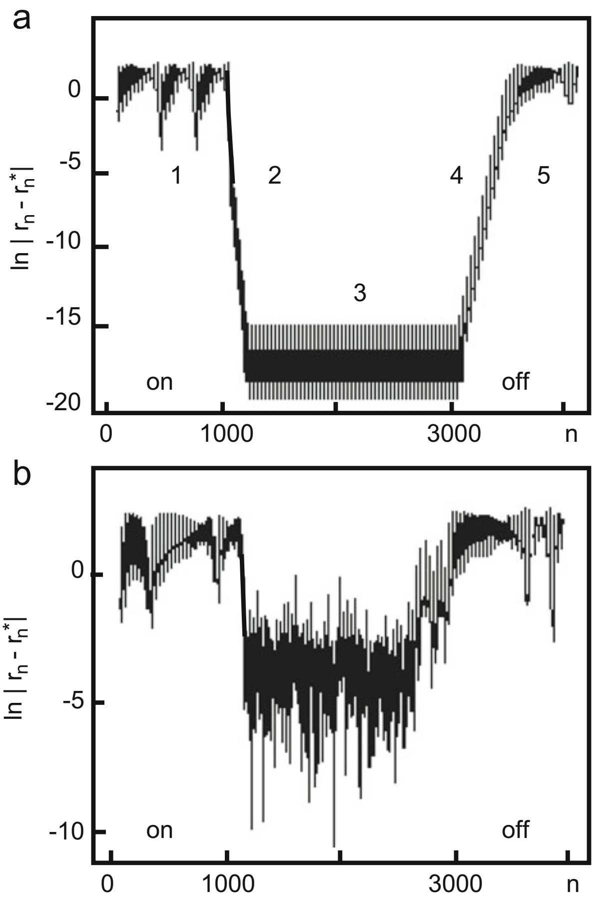


Fig. 5.16 Result of the OGY control of the mapping (5.38): (**a**) without noise; (**b**) with Gaussian noise

The coefficients M_i and N_i can be obtained from (5.11) in explicit form. The capture region for any ith point of the periodic orbit is determined by the condition

$$|M_i\delta x + N_i\delta y| < \delta p_{\max} \,.$$

It is evident that the capture region size determines both the control setup time and the critical amplitude of noise destructing the control. As numerical calculations show, areas of the capture regions of the considered period-34 unstable orbit differ in several orders of magnitude. Such situation is typical for generic periodic orbits in complex (in the sense of the above definition of complexity) dynamical systems. Accounting for this, for orbits with considerably different capture regions it may be convenient to introduce the concept of local and global control [40]. In the case of local control, the condition $|\delta p_n| < \delta p_{\max}$ is satisfied only for some points of the periodic orbit, whereas in the case of global control, it is satisfied for all points. It is evident that there is no difference between local and global controls for the fixed points and for the periodic orbits with capture regions approximately equal for all points of the orbit. On the contrary, for unstable periodic orbits with substantially different capture regions, global control takes place only in cases when the local control condition is satisfied for the points with the minimal capture area. From the point of view of control realization, those points can be called the dangerous ones. For the considered period-34 orbit the dangerous points constitute less than 30 % of the total number of points forming the periodic orbit. The strategy relying on the local control setup for the dangerous points will automatically lead to a global control setup as well, and it will allow to substantially lower computational efforts.

5.9 Controlling Chaos in Time Dependant Irregular Environment

The above considered schemes of the chaos control are immediately applicable to the systems where the noise is relatively small, i.e., it does not interfere with the structure of the initial phase space. Let us now turn to a principally different situation, when the system is in contact with a time dependent environment (a medium). As the environment we shall understand some large dynamical system, whose evolution does not depend on the controlled system, but strongly affects the latter.

Our goal is to adapt the OGY control technique for cases where the medium changes irregularly and short-term predictions of the evolution of the medium are possible. The effectiveness of the modified technique [41] will be demonstrated on the following problem: to control and prevent ship upset due to a beam sea (waves running at right angles to the boat's course). Here the ocean waves can be understood as the medium. The control algorithm should admit considerable irregular variations in wave amplitudes and phases.

For a description of the ship driven by a beam sea we shall use the non-linear oscillator model

$$\ddot{x} + \nu\dot{x} + \omega^2(x - \alpha x^3) = W(t)\,, \tag{5.42}$$

where x is the angle of deviation of the ship mast from the vertical, ν is the friction coefficient, ω is the frequency of small oscillations near the potential minimum, α is the non-linearity parameter, and $W(t)$ is the term describing the action of the ocean waves on the ship. In the absence of waves ($W(t) = 0$) at small shifts x the oscillations dampen and the ship returns to the vertical position. For large shifts, the gravitational force exceeds the hydrostatic extrusion and x has a tendency to the attractor situated at $|x| = \infty$. When this happens, we can say that the ship upsets.

Suppose that the irregular wave term $W(t)$ has the form

$$W(t) = f(t)\,[1 + \varepsilon_a g(t)]\sin\phi(t) \equiv F(t)\sin\phi(t)\,, \tag{5.43}$$

where $F(t)$ is the wave amplitude, $f(t)$ is its slowly varying component, $g(t)$ is the fast irregular component, and $\varphi(t)$ is the phase whose evolution is determined by the relation

$$\phi(t) = \varOmega + \varepsilon_p h(t)\,, \tag{5.44}$$

where $h(t)$ is also an irregular function of time. As the irregular functions $g(t)$, $h(t)$ we will use the solutions for well-known chaotic systems: the Duffing oscillator [42] and the Rössler system [43]. Under the normalization condition for the functions $g(t), h(t)$ the quantities $\varepsilon_a,\ \varepsilon_p$ serve as the relative measures of amplitude and phase irregularity. The use of low-dimensional chaotic systems to generate the random functions $g(t),\ h(t)$ is dictated only by considerations of convenience and it does not lead to essential differences from the uses of other random functions or chaotic systems of higher dimensions. For numerical calculations in the model (5.37) we will use the following parameters: $\nu = 0.5, \alpha = 1, \omega = \varOmega = 1$.

In the case of purely sinusoidal waves ($\varepsilon_a = \varepsilon_p = 0, f(t) = f_0$) for $0 < f_0 < 0.7$ the ship dynamics is strictly regular: it has periodic oscillations with the period $T = 2\pi/\varOmega$. At a further increase of the wave amplitude, the period doubling bifurcations cascade takes place, resulting in the chaotic dynamics of the ship. At $f_0 \approx 0.726$ the boundary of the chaotic attractor is destroyed and almost all the initial conditions get on the attractor $|x| = \infty$, i.e., in the absence of control the ship capsizes at $f > f_0$. As was shown in the paper [41] the use of a slightly modified OGY control procedure allows us to avoid the upset both for purely sinusoidal waves with the amplitude considerably exceeding critical levels and in the case of relatively strong amplitude and phase irregularity $\left(\varepsilon_a \neq 0, \varepsilon_p \neq 0\right)$

The equation of motion for the variable x after turning on the controlling perturbation $C(t)$ has the form

$$\ddot{x} + \nu\dot{x} + \omega^2(x - \alpha x^3) = W(t) + C(t)\,. \tag{5.45}$$

To realize the discrete control in a standard way we transition from the ordinary differential equation (5.45) to a mapping in the Poincaré section plane, defining the latter by the conditions $W(t_n) = 0,\ dW/dt > 0$. We will assume that $C(t)$ does not change between two consecutive intersections of the Poincaré section. In the considered problem the perturbation $C(t)$ can be realized, for example, due to a shift of the ballast with respect to the ship's axis in the moment $t = t_n$. As always, we assume the smallness of the perturbation to be $C(t) \ll W(t)$. To that end, we limit the perturbation by the condition $-C_0 \leqslant C \leqslant C_0$.

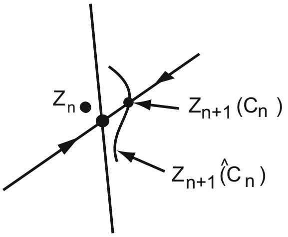


Fig. 5.17 Schematic representation of control in random environment [41]

Let Z_n $(Z = (x, \dot{x}))$ be an unstable fixed point of the Poincaré mapping (see Fig. 5.17) in the moment $t = t_n$ at $\varepsilon_a = \varepsilon_p = 0$, $f(t) = f(t_n)$. Setting $\varepsilon_a \neq 0$, $\varepsilon_p \neq 0$, we introduce irregularity into the wave. Suppose now, that as a result of observations, we can make sufficiently accurate predictions about the behavior of $W(t)$ on the interval $t_n \leqslant t \leqslant t_{n+1}$. Integrating the equation of motion (5.45) with the predicted value $W(t)$ and different values of $\hat{C}$ from the interval $[-C_0, C_0]$, we obtain the system's position in the phase space at the moment $t = t_{n+1}$. To make a decision (on the ballast shift) we will use that value $\hat{C} = C_n$, at which the point Z at the moment t_{n+1} gets on the stable direction of the unstable fixed point.

Figure 5.18 presents the control results in the presence of both amplitude ($\varepsilon_a = 0.15$) and phase ($\varepsilon_p = 0.1$) irregularities for systems where $f(t)$ is the function of time linearly growing from the value $f(0) = 0.7$ to the value $f(300) = 1$. The use of the considered control scheme allows the ship's stability to improve considerably.

5.10 Continuous Control with Feedback

Having devoted sufficient attention to the numerous merits of the OGY method, we will now point out its limitations. The OGY chaos control method is immediately applicable to dynamics described by mappings. By controlling the chaos observable in experimentation, the method reduces the real dynamics to the mapping generated by the Poincaré section, which also determines the discrete character of the controlling parameter variation. Suppose τ is the time interval between consecutive changes to the parameter and λ is the maximal Lyapunov exponent for the target unstable periodic orbit. Then evidently the OGY method is efficient only for those orbits that satisfy the condition

$$\lambda \ll 1/\tau\,. \tag{5.46}$$

The discrete character of the controlling parameter variation also worsens the stability of the OGY method with respect to noise. For relatively rare parameter changes there is a high probability of control failure. Those native disadvantages of discrete control make continuous control realization more attractive. As before, we require the smallness of the controlling perturbation variation because we intend to

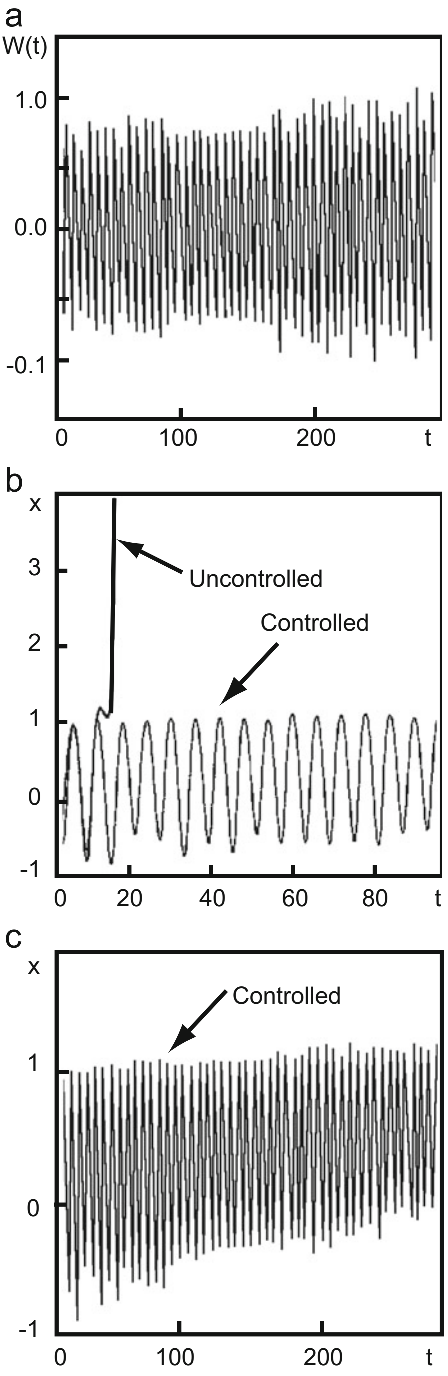


Fig. 5.18 An example of control realization in random environment: (**a**) the perturbation $W(t)$ with parameters $\varepsilon_a = 0.15,\ \varepsilon_p = 0.1$; (**b**) controlled and uncontrolled orbits; (**c**) more extensive segment of the controlled orbit [41]

stabilize the chaotic trajectory in the vicinity of a periodic orbit of the unperturbed system. This goal can be achieved only with a feedback control scheme. The two first continuous control feedback schemes were proposed and realized in the work [44]. Both schemes were based on special constructions of time-continuous perturbation which, without changing the target unstable periodic orbits, under certain conditions stabilize them. The combination of the feedback and the periodic external force lies at the core of the first scheme. The second one does not require any external force, but uses the self-controlled feedback.

We begin with the first scheme: continuous control with external force. Suppose we have a dissipative dynamical system described by some set of ordinary differential equations. Suppose also that the input of the system is available for external force application and we can measure some scalar characteristic on the output. Those assumptions are satisfied by the following model:

$$\begin{aligned} d\mathbf{x}/dt &= Q(\mathbf{x}, y) \\ dy/dt &= P(\mathbf{x}, y) + F(t) \,. \end{aligned} \tag{5.47}$$

Here y is the variable registered on the output, and $\mathbf{x}$ are all other dynamical variables of the system, that are either unavailable for measurement or do not make interest for the observer. We assume for simplicity that the input signal $F(t)$ perturbs only that equation which corresponds to the variable registered on the output. We will also consider that the dynamical system (5.47) in absence of the external force ($F(t) = 0$) has a strange attractor. When working with a real system, exact knowledge of the model (5.47) is not necessary. Using the time delay method described in Chap. 4 we can reconstruct full system dynamics from the observable scalar characteristics. Using this method we can reconstruct various periodic orbits $y = y_i(t)$, $y_i(t + T_i) = y_i(t)$, where T_i is the period of ith unstable periodic orbit. Let us choose from these obtained orbits one which we want to stabilize. Later, we will need an additional oscillator generating a signal proportional to $y_i(t)$. The difference $D(t)$ between $y_i(t)$ and the output signal $y(t)$ will be used as the controlling perturbation

$$F(t) = K\,[y_i(t) - y(t)] = KD(t)\,. \tag{5.48}$$

Here K is the experimentally tunable weight of the perturbation. The perturbation i applied on the system input as the negative feedback ($K > 0$). The flow-chart of the continuous control with external force is represented in Fig. 5.19. For many physical

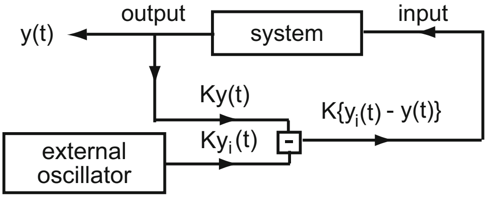


Fig. 5.19 Block-diagram of the continuous control with external force

systems, its experimental realization does not present any difficulty. An important feature of the perturbation choice in the form (5.48) consists of the fact that the perturbed system preserves the initial periodic orbits: $y(t) = y_i(t)$ is a solution of (5.47) with $F(t) = 0$.

The stabilization of the unstable periodic orbit by this control method is achieved by varying the weight factor K. When stabilization is achieved, the output signal $y(t)$ is very close to $y_i(t)$ and therefore, as in the OGY method, only small perturbation is used on the control time interval.

The experimental realization of the considered continuous control version can be divided into two stages. At the first, preliminary, stage we shall study the signal at the unperturbed system output and construct the oscillator generating the signal proportional to $y_i(t)$. At the second stage, the control is carried out by the scheme presented in Fig. 5.19.

Let us demonstrate the efficiency of the continuous control with external force using an example of the Rössler system [43]

$$\begin{aligned}
\frac{dx}{dt} &= y - z \\
\frac{dy}{dt} &= x + 0.2y + F(t) \\
\frac{dz}{dt} &= 0.2 + z(x - 5.7)\,.
\end{aligned} \tag{5.49}$$

We have chosen $y(t)$ as the scalar signal measured on the system output. The result of control does not depend on the choice of perturbed variable. Figure 5.20 presents the results of the stabilization of the period-5 unstable orbit. The beginning of the curve F corresponds to the moment perturbation is turned on. As expected, after a small transition period, the perturbation becomes small and the system comes to the periodic regime corresponding to the target orbit. The same figure presents the results of the stabilization of the period-2 unstable orbit for the Lorenz system [45]

$$\begin{aligned}
\frac{dx}{dt} &= 10(x - y) \\
\frac{dy}{dt} &= -xz + 28x - y + F(t) \\
\frac{dz}{dt} &= xy - \frac{8}{3}z\,.
\end{aligned} \tag{5.50}$$

The perturbation amplitude in the control regime depends on two factors: the precision of the unstable periodic orbit $y_i(t)$ reconstruction and the noise intensity. In an ideal case of the system moving along the orbit at zero noise level, stabilization can be achieved with a negligibly small level of the external oscillator signal (see Fig. 5.20).

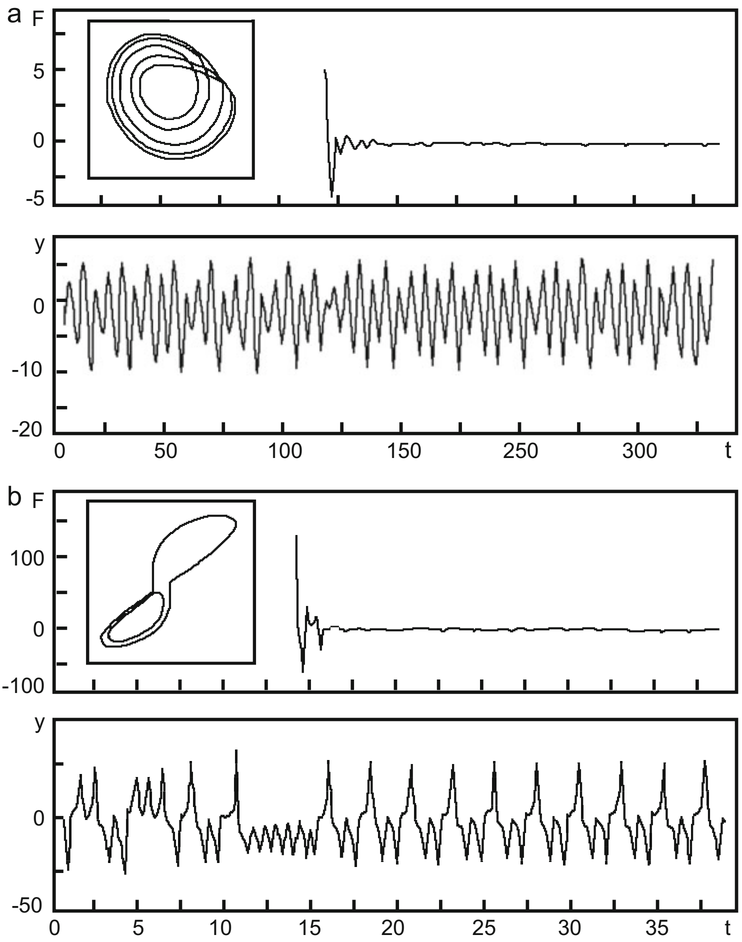


Fig. 5.20 Results of the continuous control with external force: (**a**) output signal $y(t)$ and external force $F(t)$ for the Rössler system (5.49) at $K = 0.4$; (**b**) the same quantities for the Lorenz system (5.50) [44]

Let us now dwell on the influence of noise determining the perturbation amplitude in the control regime. We will again use the Rössler system and introduce on the right-hand sides of equations (5.49) the additional terms $\varepsilon\xi_x(t)$, $\varepsilon\xi_y(t)$, $\varepsilon\xi_z(t)$. Random functions ξ_x, ξ_y, ξ_z are independent from one another and they have zero mean values and unit dispersions. Figure 5.21 presents the results of control for the period-1 orbit in the Rössler system for two different noise levels: $\varepsilon = 0.1$, $\varepsilon = 0.5$. Because the control is continuous, even for high noise levels on sufficiently long time segments there is no stabilization failure, as can be observed in the discrete control. Increase in noise levels leads only to growth in the controlling perturbation amplitude and to some "smearing" of the periodic orbit.

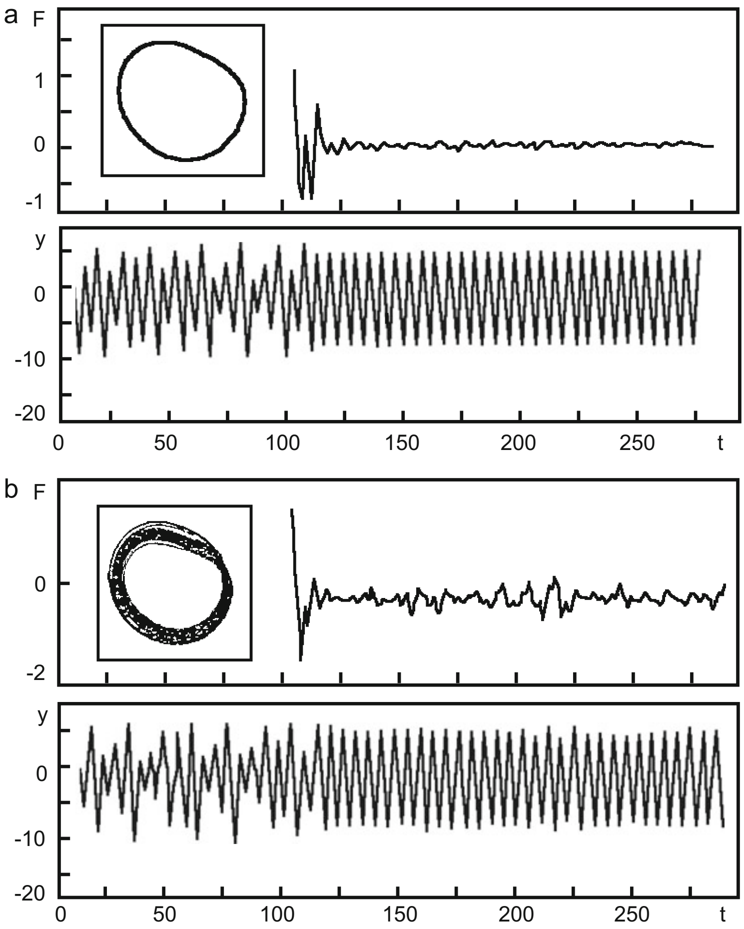


Fig. 5.21 Results of continuous control for the period-1 orbit in the Rössler system (5.49) for two different noise levels at $K = 0.4$: (**a**) $\varepsilon = 0.1$, (**b**) $\varepsilon = 0.5$ [44]

We should note one more important distinction between continuous and discrete control. The former starts to work only if the system is close to the target orbit, as it is based on the linearization of the deviation from it. In the continuous control method there is no need to wait for the approach of the system to the target orbit. The perturbation can be turned on at any time. Thus the Rössler system is efficiently synchronized with the external oscillator even if the initial conditions are far from the periodic orbit. Although, in that case, the initial perturbations increase. However, we should not expect an analogous situation for more complex systems where the stabilized orbits belong to different basins of initial conditions. Such multi-stability substantially complicates the achievement of the goal. A large initial perturbation can also be undesirable for the experiment, the control of which is planned. In many cases, both problems can be solved by forced limitation of the perturbation. Introducing some non-linear element in the feedback chain allows $F(t)$ to reach saturation for large deviation values $D(t)$:

$$\begin{cases} -F_0, & KD(t) < -F_0, \\ KD(t), & -F_0 < KD(t) < F_0, \\ F_0, & KD(t) > F_0. \end{cases} . \tag{5.51}$$

Here $F_0 > 0$ is the saturating perturbation value. Although the perturbations (5.48) and (5.51) work identically in the vicinity of the stabilized unstable periodic orbit, they lead to distinct transition processes. In the case of (5.51) the perturbation is always small (at small F_0), including the transition process; however, the latter considerably increases in average. The system "waits" until the chaotic trajectory approaches the target orbit sufficiently closely, and only after that synchronizes it with the external oscillator. As in the discrete control method the average duration of the transition process grows quickly with decrease of F_0.

In order to analyze the local stability of the system in the control regime it is useful to calculate the maximal Lyapunov exponent. To do that we use the example of the Rössler system (5.49), linearized in small deviations from the target periodic orbit. The dependence of the maximal Lyapunov exponent λ on the parameter K for period-1 and period-2 orbits is presented in Fig. 5.22. Negative values of the Lyapunov exponent $\lambda(K)$ determine the interval K, corresponding to the stabilized unstable periodic orbits. For the Rössler system the period-1 orbit is stabilized on the finite interval $[K_{\min}, K_{\max}]$. Values of $K_{\min}$ and $K_{\max}$ determine the stabilization threshold: $\lambda(K_{\min}) = \lambda(K_{\max}) = 0$. The period-2 orbit has infinite stabilization interval. The Lyapunov exponent $\lambda(K)$ for both orbits has a minimum at some value $K = K_{\mathrm{op}}$, providing the optimal control. We should note that the control interval size $K_{\max} - K_{\min}$ depends on the choice of controlled variable. So, for example, for the Rössler system, the control of the y variable is the most efficient, because this choice leads to the maximal interval corresponding to stabilization. Some systems can have several stabilization intervals for the same variable. Thus the Lorenz system in the

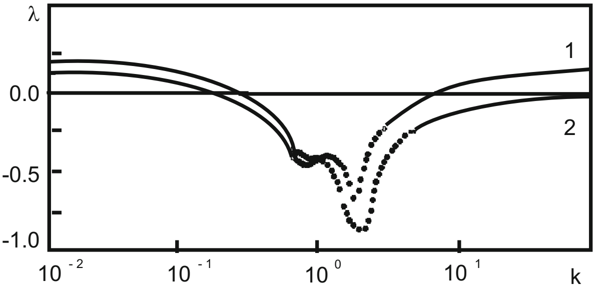


Fig. 5.22 Dependence of the maximal Lyapunov exponent λ on the parameter K for period-1 and period-2 orbits in the Rössler system [44]

case of z variable control has two isolated stabilization intervals. The presence of the threshold $K_{\min}$ is well understood: the perturbation must be sufficiently strong in order to compensate for the divergence of trajectories close to the unstable orbit, i.e., to invert the λ sign. However, large values of K worsen the control. This is connected with the fact that in the considered realization of continuous control the perturbation acts immediately only on one of the system variables. For large K those perturbations change so quickly in time that the other variables do not have time to follow those changes. The analysis shows that in the multi-parametric control version, when perturbation is introduced in each of the equations of motion, the monotonous decrease of $\lambda(K)$ is observed and the second threshold for $K_{\max}$ is absent.

The latter observation leads to the following question (particularly important for experimental realization of the continuous control): in what chaotic systems is single-parametric control efficient? The answer is based on the assumption that stabilization is possible only in cases where perturbation has a number of degrees of freedom sufficient to suppress the exponential divergence in all available directions. In other words, the minimal number of the controlled variables must be equal to the number of the positive Lyapunov exponents in the controlled system. The chaotic systems, where two or more Lyapunov exponents are positive, are called hyper-chaotic. No version of single-parametric control makes possible the stabilization of hyper-chaotic systems. At the same time, however, the multi-parametric control is efficient for such systems.

The complexity of experimental realizations of the above control method is due to the presence of the special external oscillator. An alternative continuous control method—continuous control with delayed feedback—is free of that weak point. The method replaces the external signal $y_i(t)$ in (5.48) with the delayed output signal. In other words, we will use the controlling perturbation in the form

$$F(t) = K\left[y(t-\tau) - y(t)\right] = KD(t)\,, \tag{5.52}$$

where τ is the delay time. If this time coincides with the period of ith periodic orbit $\tau = T_i$, then the solution of the system (5.47) will also correspond to that periodic orbit, i.e., $y(t) = y_i(t)$. It means that the perturbation of the form (5.52), as well as (5.48), does not change the periodic orbits in the system. Choosing the appropriate weight K of the feedback, we can achieve the stabilization of the system. The block-diagram corresponding to this version of the continuous control method is presented in Fig. 5.23.

The results of the period-3 orbit in the Rössler system and period-1 orbit in the non-autonomous Duffing oscillator

$$\frac{dx}{dt} = y,\ \frac{dy}{dt} = x - x^3 - dy + f\cos\omega t + F(t)\,, \tag{5.53}$$

are presented in Fig. 5.24. The situation is very similar to the case considered above of continuous control with external force. However, now the experimental realization is much simpler, as it does not require any external periodic perturbation.

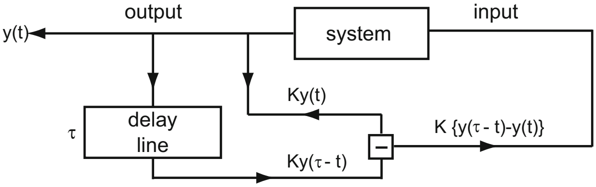


Fig. 5.23 Block-diagram of the continuous control with delayed feedback

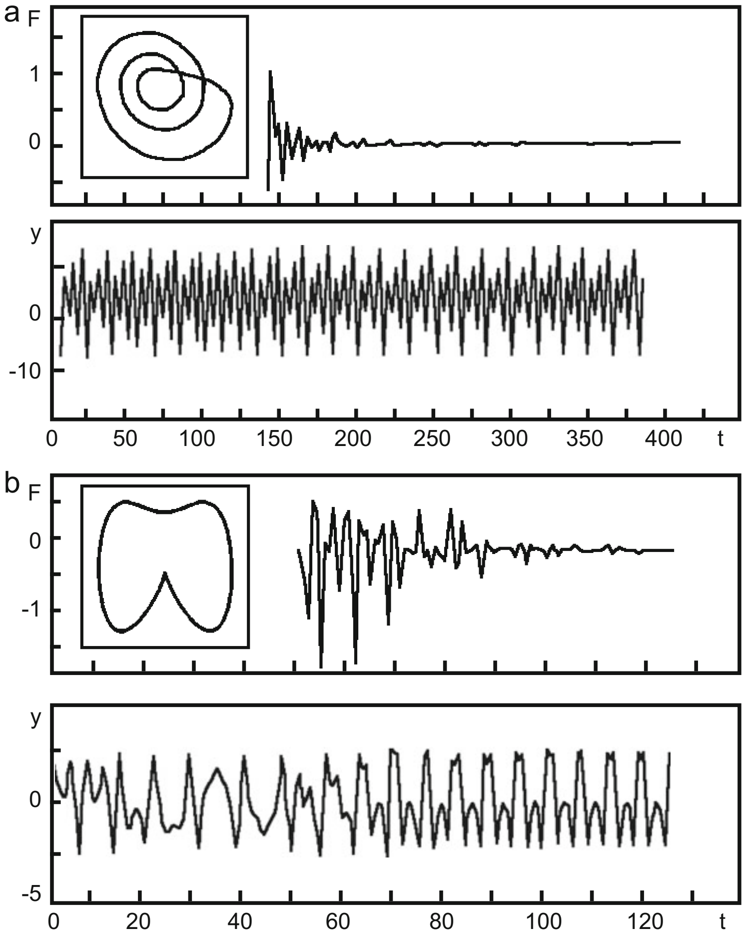


Fig. 5.24 Stabilization of unstable periodic orbits using the continuous control with delayed feedback: (**a**) a period-3 unstable orbit for the Rössler system ($K = 0.2,\ \tau = 17.5$); (**b**) period-1 unstable orbit for the non-autonomous Duffing oscillator ($f = 2.5,\ \omega = 1,\ d = 0.2,\ K = 0.4,\ \tau = 2\pi/\omega$) [44]

The difference between the delayed output signal and the proper output signal is used as the controlling perturbation. This feedback works as the self-control. Only a simple delay chain is needed for its experimental realization. In order to achieve the target unstable periodic orbit stabilization two parameters must be available for tuning in the experiment process: the delay time τ and the feedback weight K. The feedback signal amplitude can be considered as a criterion of the unstable periodic orbit stabilization. When the system is in the control regime the feedback amplitude is extremely small (see Fig. 5.24).

We should note that at the core of both the systems considered there is the same mechanism—the extension of the initial system's dimensions. In the first case, the dimensions increase due to the introduction of the external signal, and in the second one, due to the delay. The perturbation does not change the projection of the periodic orbit on the initial space of lower dimension. Additional degrees of freedom only change the Lyapunov exponents of the controlled system. We will explain this statement based on the example of the logistic mapping which we have already addressed many times. The unperturbed ($F_n = 0$) logistic mapping

$$X_{n+1} = 4X_n(1 - X_n) + F_n \tag{5.54}$$

has an unstable fixed point $X_n = 3/4$ with eigenvalue $\lambda = -2$. The perturbation in the delay form

$$F_n = K(X_{n-1} - X_n) \tag{5.55}$$

does not change the X coordinate of the fixed point, but increases the mapping dimension up to two. Analysis of that mapping shows that modules of the two eigenvalues of the Jacobi matrix for that point in the interval $K = [-1, -05]$ are less than unity. Therefore, for that value K the one-dimensional fixed point transforms into a two-dimensional stable point.

This scheme also suffers from the multi-stability problem related to the existence of two (or more) stable solutions with different basins of initial conditions. As in the case of control with external force, the multi-stability problem can be solved by introducing a limitation on the type (5.51) perturbation magnitude. Making use of this limitation, the asymptotic behavior of the system becomes single-valued for all K.

Figure 5.25 shows the dependence of the maximal Lyapunov exponent for period-1 ($\tau = 5.9$) and period-2 ($\tau = 11.75$) unstable orbits of the Rössler system. We can see that as in the case of the control with external force, each of the unstable orbits can be stabilized on the finite interval of K. However, those intervals are considerably narrower than in the former case. This means that the delayed control

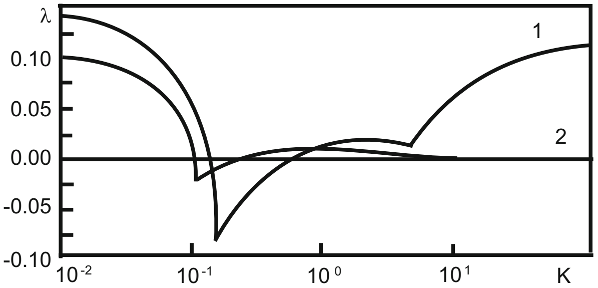


Fig. 5.25 Dependence of maximal Lyapunov exponent λ on K for period-1 ($\tau = 5.9$) and period-2 ($\tau = 11.75$) unstable orbits of the Rössler system in the case of continuous control with delayed feedback [44]

is more sensitive to the agreement of parameters, because the controlling external force always tries to attract the trajectory to the target periodic orbit. In the case of the control with delay, the perturbation brings the trajectory together with delayed one, which does not exactly coincide with the target orbit.

We now apply the continuous control scheme for stabilization of aperiodic (chaotic) orbits [46]. The considered scheme, using only a small perturbation of special form, allows us to synchronize the current behavior of the system with its past, previously recorded. As a result, we obtain the ability to predict long time segments of chaotic behavior. Essentially, the modern continuous control scheme is the combination of two different approaches to the chaos control problem: the OGY method, based on utilization for control of only a small perturbation with feedback, and the synchronization method (to be considered below) for two strongly connected chaotic systems. As the result of this synthesis we can synchronize aperiodic orbits due to a small perturbation with feedback.

As before, we assume that the controlled object is described by the system of the form (5.47) with all the above assumptions. The realization of the method splits into two stages. At the first stage, some time segment $y_{\mathrm{ap}}(t)$ must be extracted and recorded. At the second stage, we apply to the system the feedback perturbation of the form

$$F(t) = K\left[y_{\mathrm{ap}}(t) - y(t)\right] . \tag{5.56}$$

As well as above the perturbation represents a positive feedback, therefore $K > 0$. The block-diagram of experimental realization of the aperiodic orbits control method is presented in Fig. 5.26. One of the important features of the perturbation is the fact that it turns to zero when the output signal coincides with the one recorded in the system memory: $F(t) = 0$ for $y(t) = y_{\mathrm{ap}}(t)$. Therefore, the perturbation

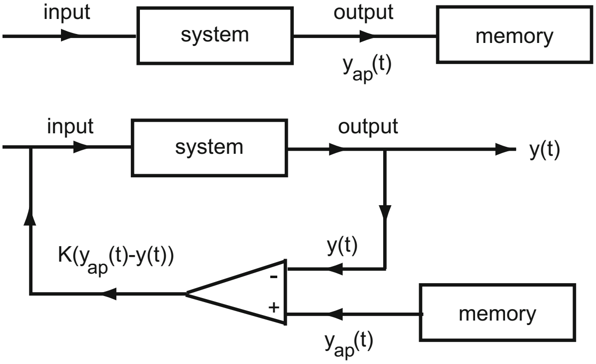


Fig. 5.26 Block-diagram of continuous control for aperiodic orbits

does not change the unperturbed system solution for the time interval corresponding to the recorded signal $y_{\mathrm{ap}}(t)$. The perturbation, as in the case of unstable periodic orbits, works as the self-control because it always brings the current trajectory $y(t)$ to the target aperiodic orbit $y_{\mathrm{ap}}(t)$. The synchronization can be achieved for a sufficiently large weight K. In the synchronization regime $\big(y_{\mathrm{ap}}(t) \approx y(t)\big)$ the perturbation becomes very small (to the degree of $\big(y_{\mathrm{ap}}(t) - y(t)\big)$ quantity).

The results of this synchronization for the Rössler, Lorenz, and Duffing systems are presented in Fig. 5.27. For all three systems, relatively soon after the perturbation turning on the current trajectory synchronizes with y_{ap}, i.e., $\Delta y \equiv y_{\mathrm{ap}}(t) - y(t) \to 0$ relative to the degree of noise, and of the constancy of the system characteristics. Synchronization was achieved irrespective of the initial conditions (if they were chosen from a common basin).

The non-autonomous system, considered above as a control object and represented in Fig. 5.26, can be transformed into a more complex autonomous system containing two connected subsystems. Indeed, the memory unit used for the input signal generation in the first case, can be replaced by an additional identical chaotic system, which, starting from appropriate initial conditions, generates the aperiodic signal identical to the one recorded in memory.

As a result, the two-stage experiment is replaced by the single-stage one presented in Fig. 5.28. The original problem is therefore reduced to the synchronization of two connected identical chaotic systems, which will be considered later.

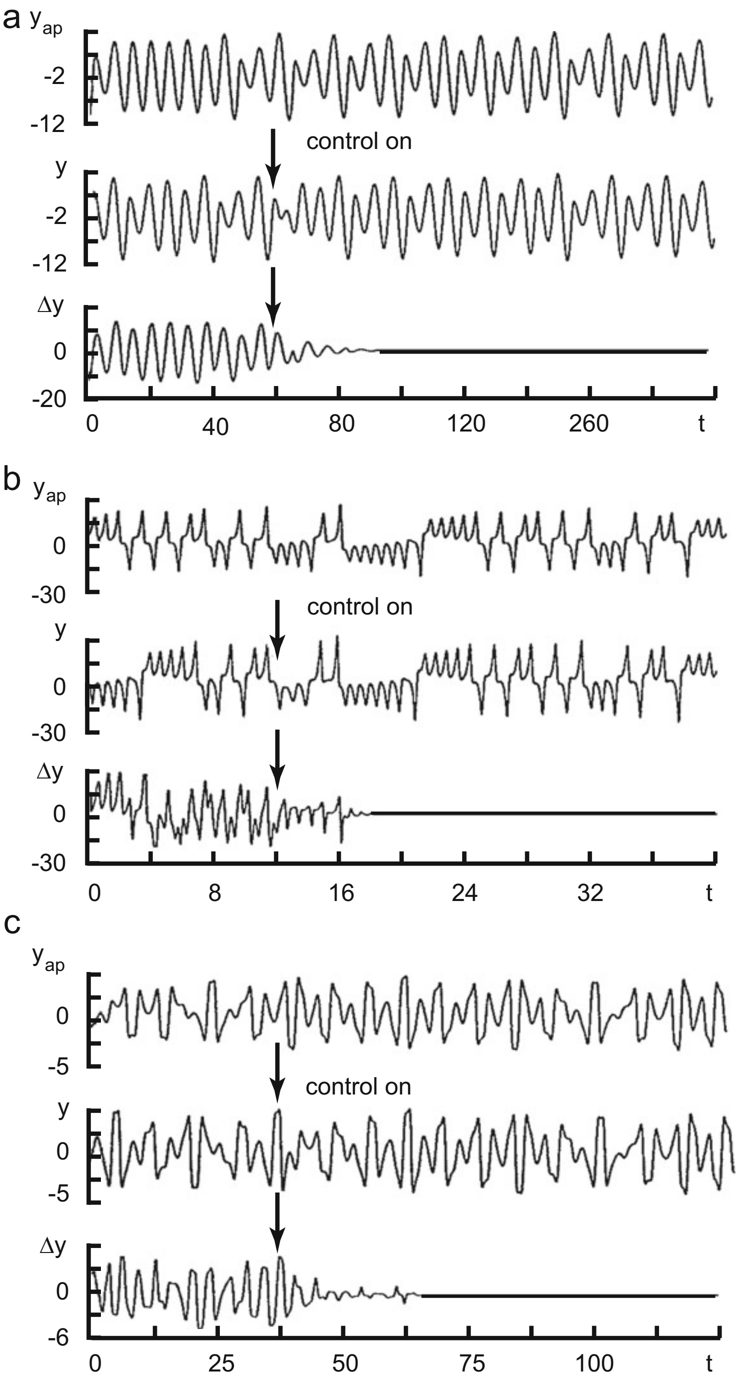


Fig. 5.27 Results of the aperiodic orbits control: (**a**) the Rössler system, (**b**) the Lorenz system, (**c**) the Duffing oscillator [46]

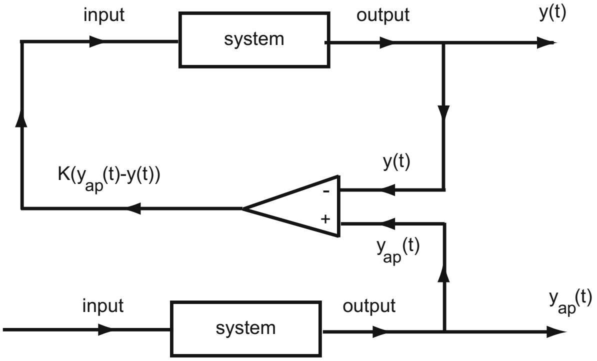


Fig. 5.28 Autonomous block-diagram for the aperiodic orbits control

References

1. Ott, E., Grebogi, C., Yorke, J.: Phys. Rev. Lett. **64**, 1196–1199 (1990)
2. Baker, G.: Am. J. Phys. **63**, 832–838 (1995)
3. Garfinkel, A., Spano, M.L., Ditto, W.L., Weiss, J.N.: Controlling Cardiac Chaos Sci. **257**, 1230–1235 (1992)
4. Schiff, S.J., Jerger, K., Duong, D.H., Chang, T., Spanno, M.L., Ditto, W.L.: Nature **370**, 615–620 (1994)
5. Berry, M.V.: Semiclassical Mechanics of Regular and Irregular Motion. Les Houches Lecture Series, vol. 36, pp. 171–270. North-Holland, Amsterdam (1983)
6. Gutzwiller, M.C.: Chaos in Classical and Quantum Mechanics. Springer, New York (1990)
7. Biham, O., Wenzel, W.: Phys. Rev. A **42**, 4639–4646 (1990)
8. Grassberger, P., Kantz, H., Moening, U.: J. Phys. A **22**, 5217–5230 (1989)
9. Hansen, K.: Phys. Rev. E **52**, 2388–2391 (1995)
10. Cvitanovic, P., Gunaratne, G.: Phys. Rev. A **38**, 1503–1520 (1988)
11. Schmelcher, P., Diakonos, F.: Phys. Rev. Lett. **78**, 4733–4736 (1997)
12. Boccaleti, S., Grebogi, C., Lai, Y.-C., Mancini, H., Maza, D.: Phys. Rep. **329**, 103–197 (2000)
13. Lai, Y.-C.: Comput. Phys. **8**, 62–67 (1994)
14. Feigenbaum, M.: J. Stat. Phys. **19**, 25–52 (1978)
15. Ditto, W., Spano, M.L., Linder, J.F.: Physica D **86**, 198–211 (1995)
16. Ditto, W.L., Showalter, K.: Chaos **7**, 509–511 (1997)
17. Hénon, M.: Commun. Math. Phys. **50**, 69–77 (1976)
18. Ott, E., Grebogi, C., Yorke, J.: In: Campbell, D.K. (ed.) Chaos: Soviet-American Perspectives on Nonlinear Science. American Institute of Physics, New York (1990)
19. Shinbrot, T., Grebogi, C., Ott, E., Jorke, J.A.: Nature **363**, 411–417 (1993)
20. Ditto, W.L., Rauseo, S.N., Spano, M.L.: Phys. Rev. Lett. **65**, 3211–3214 (1990)
21. Mandelbrot, B.: The Fractal Geometry of Nature. Freeman, San Francisco (1982)
22. Feder, J.: Fractals. Plenum Press, New York (1988)
23. Shinbrot, T., Ott, E., Grebogi, C., Jorke, J.A.: Phys. Rev. Lett. **65**, 3215–3218 (1990)
24. Shinbrot, T., Ditto, W., Grebogi, C., Ott, E., Spano, M., Yorke, J.A.: Phys. Rev. Lett. **68**, 2863–2866 (1992)
25. Kostelich, E., Grebogi, C., Ott, E., Jorke, J.A.: Phys. Rev. E **47**, 305–310 (1993)
26. Lai, Y.-C., Ding, M., Grebogi, C.: Phys. Rev. E **47**, 86–92 (1993)
27. Schroer, C., Ott, E.: Chaos **7**(4), 512–519 (1997)
28. Grebogi, C., Ott, E., Yorke, J.: Phys. Rev. Lett. **57**, 1284–1287 (1986)
29. Karney, C.: Physica D **8**, 360–380 (1983)

30. Lichtenberg, A.J., Lieberman, M.A.: Regular and Stochastic Motion. Springer, New York (1983)
31. Siegel, C.L., Mozer, J.: Lectures on Celestial Mechanics. Springer, Berlin (1971)
32. Bolt, E., Meiss, J.: Phys. Lett. A **204**, 373–378 (1995)
33. Lai, Y.-C., Tel, T., Grebogi, C.: Phys. Rev. E **48**, 709–717 (1993)
34. Gaspard, P., Rice, S.: J. Phys. Chem. **90**, 2225–2241 (1989)
35. Politi, A., Oppo, G.L., Badii, R.: Phys. Rev. A **33**, 486–490 (1986)
36. Gonchar, V.Yu., Ostapchyk, P.N., Tur, A.V., Yanovsky, V.V.: Phys. Lett. A **152**(5–6), 287–292 (1991)
37. Chirikov, B.V.: At. Energiya **6**, 630–638 (1959)
38. Bolotin, Yu.L., Gonchar, V.Yu., Tur, A.V., Yanovsky, V.V.: Phys. Rev. Lett. **82**, 2504–2507 (1999)
39. Poon, L., Grebogi, C.: Phys. Rev. Lett. **75**, 4023–4026 (1995)
40. Bolotin, Yu.L., Gonchar, V.Yu., Krokhin, A.A., Hernandes-Tejeda, P.H., Tur, A.V., Yanovsky, V.V.: Phys. Rev. E **64**, 0262XX1–0262XX9 (2001)
41. Ding, M., Ott, E., Grebogi, C.: Physica D **74**, 386–394 (1994)
42. Wiggins, S.: Introduction to Applied Nonlinear Dynamical Systems and Chaos. Springer, New York (1990)
43. Rössler, O.: Phys. Lett. A **57**, 397–398 (1976)
44. Pyragos, K.: Phys. Lett. A **170**, 421–428 (1992)
45. Lorenz, E.: J. Atmos. Sci. **20**, 130–141 (1963)
46. Pyragos, K.: Phys. Lett. A **181**, 203–210 (1993)

Chapter 6
Synchronization of Chaotic Systems

The synchronization of stable oscillations is a well-known non-linear phenomenon frequently found in nature and widely used in technology [1–5]. Under synchronization, one usually understands the ability of coupled oscillators to switch from an independent oscillation regime, characterized by beats, to a stable coupled oscillation regime with identical or rational frequencies, when the coupling constant increases.

The statement of the problem of chaotic oscillation synchronization may appear paradoxical in contrast to stable oscillations. Two identical autonomous chaotic systems with almost the same initial conditions diverge exponentially quickly in the phase space. This is the main difficulty, at first sight making it impossible to create synchronized chaotic systems which will function in reality. Nevertheless, there are several reasons which make the realization of chaotic synchronization a very promising goal.

The noise-like behavior of chaotic systems suggests that they can be useful for secure communications. Even a fleeting glance at the Fourier spectrum of a chaotic system confirms this: no dominating peaks, no dominating frequencies, a normal broadband spectrum. Any attempt to use a chaotic signal for communication purposes makes it necessary for the recipient to have a duplicate of the signal used in the transmitter (i.e., the synchronized signal). In practice, synchronization is needed for many communication systems, not necessarily just chaotic ones. Unfortunately, existing synchronization methods are not suitable for chaotic systems, and therefore this purpose requires the development of new ones.

Chaos is widely used in cybernetic, synergetic, and biological applications [5–7]. If we have a system composed of several chaotic subsystems, then it is clear that their efficient joint functioning is possible only after the synchronization problem is solved.

In spatially extended systems, we often face the transition from homogeneous spatial motion to one changing in space (including also chaotic changes). For example, in the Belousoff–Zhabotinski reaction, dynamics can be chaotic but

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_6

spatially homogeneous. This means that different spatial parts are synchronized with each other, i.e., they perform the same motions in the same moment of time, even if those motions are chaotic. But under other conditions the homogeneity loses stability and the system becomes inhomogeneous. Such spatial homogeneity ↔ inhomogeneity transitions are typical for extended systems, and synchronization must play a key role there.

The interest in the chaotic synchronization problem goes far beyond the limits of the natural sciences. It seems natural that the efficiency of an advertisement is determined by ability of the advertising objects to synchronize. The same can also be said about the unified perception of the mass culture.

6.1 Statement of Problem

The first works on synchronization of coupled chaotic systems were written by Yamada and Fujisaka [8]. They used local analysis (special Lyapunov exponents) to investigate changes in the dynamical systems when the coupling constant increased. Afraimovich et al. [9] introduced the basic notions now used in the description of the chaotic synchronization process. A principally important role in the development of the chaotic synchronization theory was played by the paper [10], where a new geometrical point of view on the synchronization phenomenon was developed.

Let us formulate the synchronization problem for a dynamical system described by a system of ordinary differential equations [10]. A generalization for the case of mappings requires only minimal changes. Consider an n-dimensional dynamical system

$$\dot{u} = f(u)\,. \tag{6.1}$$

Let us divide the system arbitrarily into two subsystems $u = (v, w)$

$$\begin{aligned} \dot{v} &= g(v, w)\,, \\ \dot{w} &= h(v, w)\,, \end{aligned} \tag{6.2}$$

where

$$\begin{aligned} v &= (u_1 \dots u_m)\,; \quad w = (u_{m+1} \dots u_n)\,, \\ g &= (f_1(u) \dots f_m(u))\,; \quad h = (f_{m+1}(u) \dots f_n(u))\,. \end{aligned} \tag{6.3}$$

Now we create a new subsystem w', identical to w, and we make the change $v' \to v$ in the function h, attaching to (6.2) the equation for the new subsystem

$$\begin{aligned} \dot{v} &= g(v, w)\,, \\ \dot{w} &= h(v, w)\,, \\ \dot{w}' &= h(v, w')\,. \end{aligned} \tag{6.4}$$

The coordinates $v = (v_1 \ldots v_m)$ are called forcing variables, and $w' = \left(w'_{m+1} \quad w'_n\right)$ are the forced variables. Consider the difference $\Delta w = w' - w$. The subsystem components w and w' will be considered synchronized if $\Delta w \to 0$ at $t \to \infty$. In the limit $\Delta w \to 0$ the equation for variations $\Delta w \equiv \xi$ reads the following:

$$\dot{\xi}_i = [D_w h(v(t), w(t)]_{ij}\, \xi_j\,, \tag{6.5}$$

where $D_w h$ is the Jacobian for the w subsystem with respect to variable w only. It is clear that if $\xi(t) \to 0$ at $t \to \infty$, then the trajectories of one of the subsystems converge to the same values of the other one. In other words, the subsystems are synchronized. The necessary condition of this subsystem synchronization is the negativity of the Lyapunov exponents of the equation system (6.5). It can be shown [11] that these Lyapunov exponents are negative when the Lyapunov exponents of subsystem w are negative. This condition is necessary but insufficient for synchronization. One should separately consider the question of the initial set of conditions w', which can be synchronized with w.

6.2 Geometry and Dynamics of the Synchronization Process

Let us begin the description of the synchronization process with the example of one well-known dynamical Lorenz system. We will also consider general cases and types of synchronization below. Assuming that we have two identical chaotic Lorenz systems, already considered in the previous chapter, can we synchronize these two chaotic systems by transmitting some signal from the first system to the second one? Let this signal be x component of the first Lorenz system. Throughout the second system, we replace x component with the signal from the first system. Such an operation is commonly called a complete replacement [12]. Thus, we get a system of five connected equations:

$$\begin{aligned} \dot{x}_1 &= -\sigma(y_1 - x_1)\,, \\ \dot{y}_1 &= -x_1 z_1 + r x_1 - y_1, \quad \dot{y}_2 = -x_1 z_2 + r x_1 - y_2\,, \\ \dot{z}_1 &= x_1 y_1 - b z_1, \quad \dot{z}_2 = x_1 y_2 - b z_2\,. \end{aligned} \tag{6.6}$$

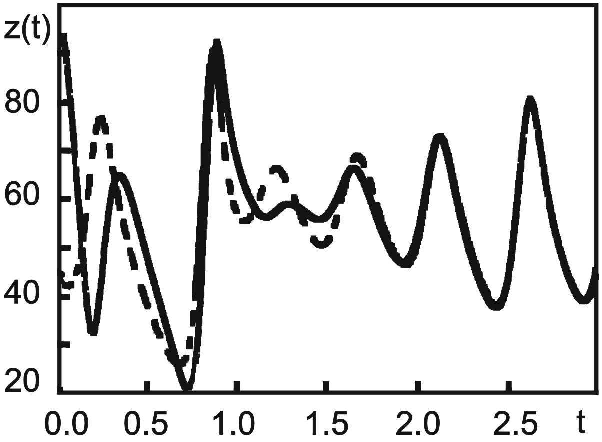


Fig. 6.1 Time dependence of the $z(t)$ coordinate for the driving (*dashed line*) and the driven (*solid line*) Lorenz systems [13]

The variable x_1 can be considered the driving force for the second system. If we start in (6.6) with arbitrary initial conditions, then, analyzing the numerical solution of the system, we will see that y_2 converges to y_1, and z_2 to z_1, after several oscillations and in the long-time asymptotic $y_2 = y_1, \; z_2 = z_1$ (see Fig. 6.1). Hence we have two synchronized chaotic systems. Usually, this situation is called identical synchronization since both subsystems are identical and have equal components.

The equations $y_1 = y_2$ and $z_1 = z_2$ determine a hyperplane in the original five-dimensional phase space $(x_2 \to x_1)$. The limitation of motion by the hyperplane is the geometrical image of the identical synchronization. Therefore, this hyperplane is sometimes [12] called the synchronization manifold.

In the example of two synchronized Lorenz systems considered above, we saw that the differences $|y_1 - y_2| \to 0$ and $|z_1 - z_2| \to 0$ at $t \to \infty$. This is possible only if the synchronization manifold is stable. In order to make sure of this, we transform to the new coordinates

$$\begin{aligned} x_1 &= x_1\,, \\ y_\perp &= y_1 - y_2; \quad y_\parallel = y_1 + y_2\,, \\ z_\perp &= \; z_1 - z_2; \quad z_\parallel = z_1 + z_2. \end{aligned} \tag{6.7}$$

In the new variables the three coordinates $\left(x_1, y_\parallel, z_\parallel\right)$ belong to the synchronization manifold, and the two others $(y_\perp, z_\perp)$ to the transversal. The synchronization condition is satisfied by the tending to zero of the variables $y_\perp$ and $z_\perp$ at $t \to \infty$. In other words, the point $(0,0)$ in the transversal manifold must be stable. The system dynamics in the vicinity of that point is described by the equation

$$\begin{pmatrix} \dot{y}_\perp \\ \dot{z}_\perp \end{pmatrix} = \begin{pmatrix} -1 & -x_1 \\ x_1 & -b \end{pmatrix} \begin{pmatrix} y_\perp \\ z_\perp \end{pmatrix} . \tag{6.8}$$

The general condition of stability is to have negative Lyapunov exponents for Eq. (6.8). This condition is equivalent to the negativity of Lyapunov exponents for the variables y_2, z_2 for the system (6.6) since the Jacobi matrices for this subsystems are identical. Therefore, we can consider the driven system (y_2, z_2) to be a separate dynamical system, driven by the driving signal x_1 and we can calculate the Lyapunov exponents for that subsystem in the usual way. Those Lyapunov exponents will depend on x_1 and therefore they will be called conditional Lyapunov exponents [13]. The values for the conditional Lyapunov exponents for a given dynamical system will depend on the choice of driving coordinate.

This complete replacement scheme can be slightly modernized [14]. The modernization procedure entails introducing the driving coordinate only in some, but not in all, driven system equations. The choice of the equations, where the replacement is performed, is dictated by two factors. First, whether the replacement leads to stable synchronization. Second, whether it is possible to realize the corresponding replacement in a real physical device which we want to construct. Let us consider the following example of partial replacement, based on the Lorenz system

$$\begin{aligned} \dot{x}_1 &= \sigma\,(y_1 - x_1)\,, & \dot{x}_2 &= \sigma(\,y_1 - x_2)\,, \\ \dot{y}_1 &= rx_1 - y_1 - x_1 z_1, & \dot{y}_2 &= rx_2 - y_2 - x_2 z_2\,, \\ \dot{z}_1 &= x_1 y_1 - b z_1, & \dot{z}_2 &= x_2 y_2 - b z_2\,. \end{aligned} \tag{6.9}$$

In (6.9) the replacement was made only in the second equation. This replacement will lead to a new Jacobi matrix defining the stability condition. Now it is a 3 × 3 matrix with zeroes in the positions of the partial replacement

$$\begin{pmatrix} \dot{x}_\perp \\ \dot{y}_\perp \\ \dot{z}_\perp \end{pmatrix} \approx \begin{pmatrix} -\sigma & 0 & 0 \\ r - z_2 & -1 & x_2 \\ y_2 & x_2 & -b \end{pmatrix} \begin{pmatrix} x_\perp \\ y_\perp \\ z_\perp \end{pmatrix} . \tag{6.10}$$

Generally speaking, in such cases the stability conditions differ from complete replacements. Sometimes they can appear to be more preferable.

In some cases, it may be useful to send the driving signal only at random moments of time. In this synchronization version (which is called "random synchronization" [15]), the driven system is subject to influence only in random moments, and in the intervals between them, it evolves freely. It is interesting to note that in this approach it is sometimes possible to achieve the stability of the synchronized state even in cases when continuous driving does not work.

From a more general point of view, the synchronization of chaotic systems can be considered in terms of negative feedback, which we used earlier in the example of continuous control. Introducing a damped term into the equations for the driving system, we get the following:

$$\dot{\mathbf{x}}_1 = \mathbf{F}(\mathbf{x}_1), \qquad \dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}_2) + \alpha \hat{E}(\mathbf{x}_1 - \mathbf{x}_2)\,, \tag{6.11}$$

where matrix $\hat{E}$ determines the linear combinations of the $\mathbf{x}$-components, which form the feedback loop, α is the coupling constant. For example, for the Rössler system

$$\begin{aligned} \dot{x}_1 &= -(y_1 + z_1), \quad \dot{x}_2 = -(y_2 + z_2) + \alpha(x_1 - x_2)\,, \\ \dot{y}_1 &= x_1 + ay_1, \quad \dot{y}_2 = x_2 + ay_2\,, \\ \dot{z}_1 &= b + z_1(x_1 - c); \quad \dot{z}_2 = b + z_2(x_2 - c)\,. \end{aligned} \tag{6.12}$$

In this case

$$\hat{E} = \begin{pmatrix} 1\,0\,0 \\ 0\,0\,0 \\ 0\,0\,0 \end{pmatrix}. \tag{6.13}$$

(Equations of motion for the transversal manifold coordinates)

This gives us a new equation of motion for the transversal manifold coordinates

$$\begin{pmatrix} \dot{x}_\perp \\ \dot{y}_\perp \\ \dot{z}_\perp \end{pmatrix} = \begin{pmatrix} -\alpha & -1 & -1 \\ 1 & \alpha & 0 \\ z & 0 & x - c \end{pmatrix} \begin{pmatrix} x_\perp \\ y_\perp \\ z_\perp \end{pmatrix}. \tag{6.14}$$

By calculating the conditional Lyapunov exponents for the matrix in (6.14), we can see whether the transversal perturbations are damped and therefore if the synchronization manifold is stable. In practice, it is sufficient to find only the maximal transversal Lyapunov exponent $\lambda^\perp_{\max}$. Its negativity guarantees the stability of the synchronization process. Figure 6.2 shows the dependence of the maximal transversal Lyapunov exponent on the coupling constant α for the Rössler system. Introduction of feedback initially leads to a decrease in the Lyapunov exponent. Therefore, in some intermediate region of the coupling constant values, the two Rössler systems can be synchronized. However, with further increases of the coupling constant, $\lambda^\perp_{\max}$ becomes positive and synchronization is impossible. It is easy to see that for extremely large values of $\alpha \quad x_2 \to x_1$ and the feedback introduced in (6.12) becomes equivalent to the full replacement considered above. Then the sign of quantity $\lambda^\perp_{\max}\,(\alpha \to \infty)$ determines the possibility of system synchronization in the case of full replacement.

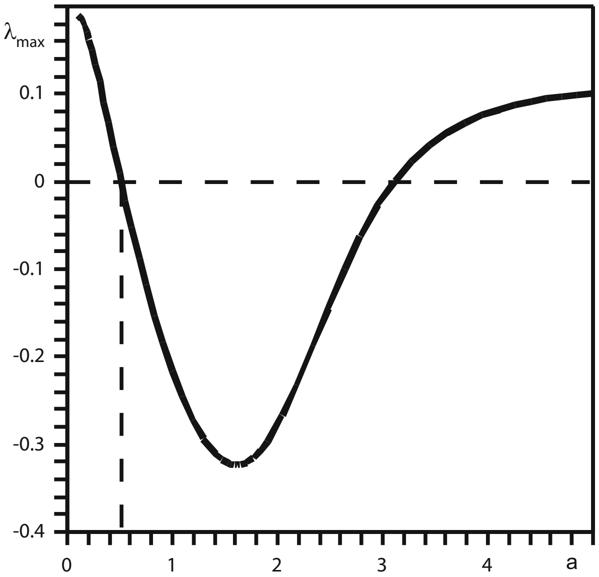


Fig. 6.2 The maximal Lyapunov exponent $\lambda_{\max}^{\perp}$ as function of the coupling constant α in the Rössler system [12]

6.3 General Definition of Dynamical System Synchronization

In the last decade many new types of chaotic synchronization appeared: apart from those mentioned in the preceding sections, there are phase synchronization, delayed synchronization, generalized synchronization, and others. As almost always happens in the first stages of investigation of any newly discovered phenomenon, there are no strict universal definitions. Such definitions are replaced by a “list”: when the researches face a new effect in a discovered phenomenon, they just extend the list. This situation is clearly unsatisfactory and at some stage this list must be replaced by a strict definition, encompassing all known effects connected with the phenomenon, as well as those to be discovered in future.

In the present section, following [16], we will make an attempt to give such a definition for finite-dimensional systems. Although we discuss explicitly the case of synchronization for two time-continuous dynamical systems, the results can be generalized for N systems, both continuous and discrete in time.

In order to construct the definition, let us assume that some large stationary dynamical system can be divided into two subsystems

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}_1(\mathbf{x},\mathbf{y};t)\,,\\ \dot{\mathbf{y}} &= \mathbf{f}_2(\mathbf{x},\mathbf{y};t)\,.\end{aligned} \tag{6.15}$$

The vectors $\mathbf{x}$ and $\mathbf{y}$ can have different dimensions. The phase space and the vector field of the big system are direct products of the phase spaces and vector fields of the subsystems. The list of phenomena described by (6.15) is inexhaustible.

Generally speaking, under synchronization we understand the time-correlated behavior of two different processes. The Oxford English Dictionary defines synchronization as "to agree in time" and "to happen at the same time." This intuitive definition means that there are ways of measuring the characteristics of subsystems as well as the criterion of concordance in time of these measured data. If these conditions are satisfied, we can say that the systems are synchronized. Further on, we will attempt to formalize each of these intuitive concepts. Let $\varphi(\mathbf{z}_0)$ be a trajectory of the original system, given by (6.15) with the initial condition $\mathbf{z}_0 = [\mathbf{x}_0, \mathbf{y}_0]$. Respectively, the curves $\varphi_x(\mathbf{z}_0)$ and $\varphi_y(\mathbf{z}_0)$ are obtained by inclusion of $\mathbf{y}$ and $\mathbf{x}$ components, e.g., by projecting. The functions $\varphi_x(\mathbf{z}_0)$ and $\varphi_y(\mathbf{z}_0)$ may be considered as the trajectories of the first and of the second subsystem, respectively. The set of trajectories of each subsystem can be used to construct subsystems characteristics $g(\mathbf{x})$ or $g(\mathbf{y})$. The measurable characteristic can either depend on time explicitly [for example, the first subsystem coordinate at time moment t, $\mathbf{x}(t) = \mathbf{g}(\mathbf{x})$], or represent a time average [for example, the Lyapunov exponent $\lambda = g(\mathbf{x})$].

Let us now give the following definition of synchronization: two subsystems (6.15) are synchronized on the trajectory $\varphi(\mathbf{z}_0)$ with respect to properties $\mathbf{g}_x$ and $\mathbf{g}_y$, if there is a time independent comparison function $\mathbf{h}$, for which

$$\|\mathbf{h}\left[\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})\right]\| = 0 . \tag{6.16}$$

We would like to emphasize that this definition must be satisfied for all trajectories. The given definition is convenient because it a priori does not depend on the measured characteristics, nor on comparison function.

The most frequently used types of comparison functions are

$$\begin{aligned}
\mathbf{h}\left[\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})\right] &\equiv \mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y}) \ , \\
\mathbf{h}\left[\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})\right] &\equiv \lim_{t\to\infty} \left[\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\right] \ , \\
\mathbf{h}\left[\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})\right] &\equiv \lim_{T\to\infty} \frac{1}{T} \int_t^{t+T} \left[\mathbf{g}(\mathbf{x}(s)) - \mathbf{g}(\mathbf{y}(s))\right] ds .
\end{aligned} \tag{6.17}$$

This definition is quite useful because the most important characteristic of finite motion is the frequency spectrum. The measured frequencies $\omega_x = g(\mathbf{x})$ and $\omega_y = g(\mathbf{y})$ represent peaks in the power spectrum. To study frequency synchronization we usually take the comparison function in the form:

$$h\left[g(\mathbf{x}), g(\mathbf{y})\right] = n_x\omega_x - n_y\omega_y = 0 . \tag{6.18}$$

In case of identical synchronization the second equation (6.17) is necessary to compare the trajectory of one system with another one, i.e., $\mathbf{g}(\mathbf{x}) = \mathbf{x}(t)$, $\mathbf{g}(\mathbf{y}) = \mathbf{y}(t)$.

This definition also covers the so-called delayed synchronization, when some measured characteristics are delayed with respect to others during the same time period τ. In that case, we can take $\mathbf{g}(\mathbf{x}) = \mathbf{x}(t)$ and $\mathbf{g}(\mathbf{y}) = \mathbf{y}(t+\tau)$, up to use the first relation in (6.17) as the comparison function.

Therefore, the definition (6.16) includes all the examples of finite-dimensional dynamical system synchronization listed above.

6.4 Chaotic Synchronization of Hamiltonian Systems

Up to now we considered chaotic synchronization only for dissipative systems. In the present section we show [17] that using the same approach as for dissipative systems, we can synchronize two Hamiltonian systems. At first glance, it seems that any attempt to synchronize two chaotic Hamiltonian systems is doomed to failure. Indeed, as was shown above, the necessary condition of any synchronization is the local synchronization, provided by the negativity of all Lyapunov exponents for a driven subsystem (recall that we called them the conditional Lyapunov exponents, because they depend also on the driving subsystem coordinates). However, if the system preserves the phase volume, as we have seen in Chap. 3 and it would seem that the synchronization is impossible. However, it does not follow that is the sum of the Lyapunov exponents is equal to zero, for a subsystem the sum of the conditional Lyapunov exponents also equals zero; a subsystem of a phase volume preserving system does not necessarily preserve the phase volume, and therefore a Hamiltonian system can be synchronized.

Let us consider as an example the so-called standard mapping, which we dealt with in the previous chapter, in the following form:

$$\begin{aligned} I_{n+1} &= I_n + k\sin\theta_n\,, \\ \theta_{n+1} &= \theta_n + I_n + k\sin\theta_n, \ \bmod 2\pi; \quad k>0\,. \end{aligned} \tag{6.19}$$

We will further drop mod2π. On the variable I the mapping has period 2π, therefore it is sufficient to study it in the square $[0,2\pi]\times[0,2\pi]$ with identifying the opposite sides. The mapping has a well-known physical interpretation [18]—the frictionless pendulum driven by periodic pulses. In this interpretation I_n, θ_n represents the angular momentum and angular coordinate immediately before nth pulse.

Following the standard synchronization procedure, we make a duplicate of the original system

$$\begin{aligned} J_{n+1} &= J_n + k\sin\phi_n\,, \\ \phi_{n+1} &= \phi_n + J_n + k\sin\phi_n\,. \end{aligned} \tag{6.20}$$

Let us chose the angular momentum of the first system I as the driving variable. Then the full system will be described by the system of the connected equations

$$\begin{aligned} I_{n+1} &= I_n + k\sin\theta_n\,, \\ \theta_{n+1} &= \theta_n + I_n + k\sin\theta_n\,, \\ J_{n+1} &= I_n + k\sin\phi_n\,, \\ \phi_{n+1} &= \phi_n + I_n + k\sin\phi_n\,. \end{aligned} \tag{6.21}$$

The subsystems will be synchronized provided the condition

$$\lim_{n\to\infty} |\theta_n - \phi_n| = 0\,. \tag{6.22}$$

Difference between the driving and the driven angular variables is

$$\theta_{n+1} - \phi_{n+1} = \theta_n - \phi_n + k(\sin\theta_n - \sin\phi_n)\,. \tag{6.23}$$

Linearization of (6.23) at small deviations of φ_n from the driving angular variable θ_n gives

$$\Delta_{n+1} = \Delta_n(1 + k\cos\theta_n)\,, \tag{6.24}$$

where $\Delta_n = \theta_n - \varphi_n$. Equation (6.24) has a solution

$$\Delta_n = \prod_{j=0}^{n-1}(1 + \cos\theta_j)\Delta_0\,. \tag{6.25}$$

Local synchronization takes place if this product at $n \to \infty$ tends to zero. It is equivalent to the requirement that the conditional Lyapunov exponent on the angular variable

$$\lambda_\theta = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\ln\left|1 + k\cos\theta_j\right| \tag{6.26}$$

is negative. The sum entering (6.26) represents the time average of the function $g(\theta) = \ln|1 + k\cos\theta|$. This time averaging can be formally represented as a mean value of that function with respect to the invariant measure $\rho(\theta)$ (see Chap. 3). The latter determines the iteration density for the mapping $\theta_{n+1} = f(\theta_n)$ and is defined in the following way:

$$\rho(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\delta[\theta - f^i(\theta_0)]\,. \tag{6.27}$$

It allows us to replace the time average $\bar{g}(\theta)$ by the average over the invariant measure

$$\bar{g}(\theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\theta_i) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g\left[f^i(\theta_0)\right] = \int d\theta \rho(\theta) g(\theta)\,. \tag{6.28}$$

Let us use this expression to transform the relation (6.26). In a rough approximation for chaotic orbits in the standard mapping (6.19) the invariant measure can be considered homogeneous on the interval $[0, 2\pi]$, i.e., $\rho(\theta) = 1/2\pi$ and for λ_θ we obtain

$$\lambda_\theta = \frac{1}{2\pi} \int_0^{2\pi} \ln|1 + k\cos\theta|\, d\theta\,. \tag{6.29}$$

The integral (6.29) can be calculated analytically,

$$\lambda_\theta = \begin{cases} \ln\left(\frac{1+\sqrt{1-k^2}}{2}\right), & 0 \le k \le 1 \\ \ln\left(\frac{k}{2}\right), & k \ge 1 \end{cases}. \tag{6.30}$$

Figure 6.3 presents the conditional Lyapunov exponent λ_θ as function of k. Quantity λ_θ is negative for $k < 2$. As is well known, the Chirikov criterion of non-linear resonance overlap determines the start of the transition to global stochasticity in the standard mapping at $k \approx 1$. From there it follows that in the global stochasticity region $1 < k < 2$ it is possible to synchronize the Hamiltonian system (6.19), if we choose the angular momentum I as the driving variable. It is interesting to note that the minimal value of the conditional Lyapunov exponent $(\lambda_\theta)_{\min} = -\ln 2$ is

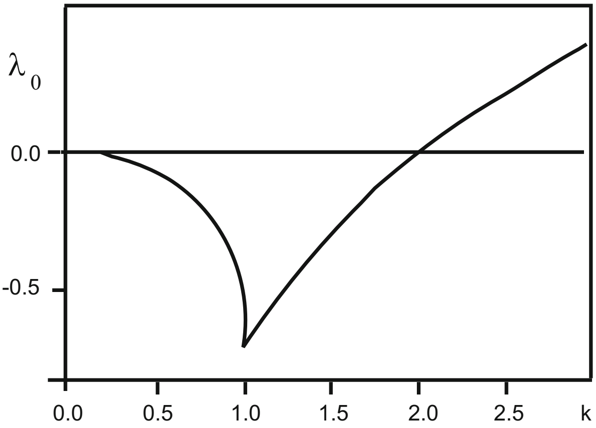


Fig. 6.3 Conditional Lyapunov exponent for the standard mapping as function of k [17]

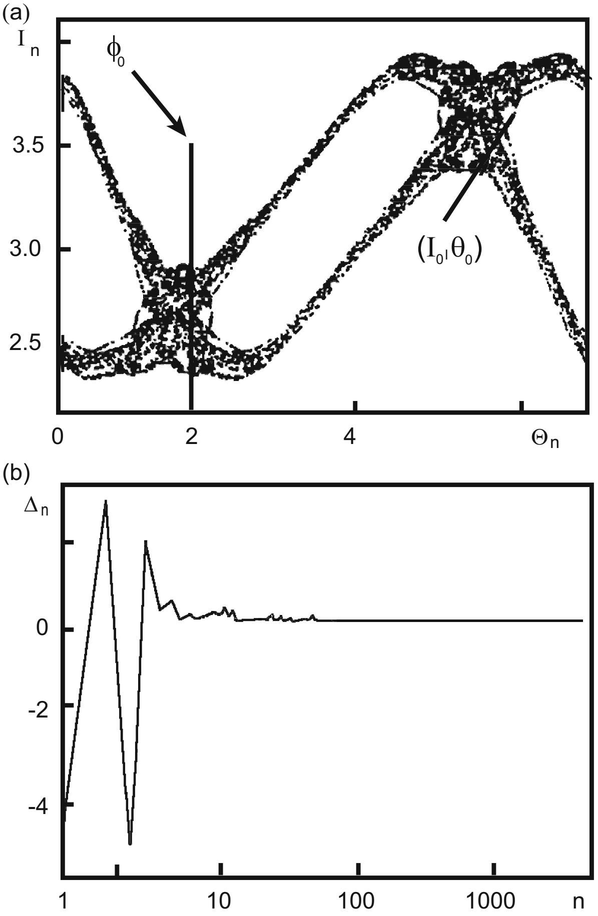


Fig. 6.4 (**a**) A chaotic trajectory for the driving system (standard mapping). *Arrows* point to the initial conditions for the two subsystems; (**b**) difference of angular coordinates for the driving and the driven subsystems as function of time (or iteration number) [17]

achieved at $k = 1$. It means that this value of k corresponds to the minimal time needed to achieve synchronization.

Figure 6.4a presents a chaotic trajectory of the driving system (I, θ) and shows the initial conditions for the two subsystems. In Fig. 6.4b the difference of the angular coordinates Δ_n is plotted as a function of the iteration number n. Complete synchronization is achieved at $n \sim 100$. If we take the angular coordinate θ as the driving variable, then it can be shown that the conditional Lyapunov exponent equals to zero in that case. It means that synchronization is impossible, because each subsystem preserves the phase volume separately.

6.5 Realization of Chaotic Synchronization Using Control Methods

In this section, taking after [19], we will try to answer the following problem. Suppose that we have two almost identical chaotic systems. So, can we, using the OGY parametric control method considered in the previous chapter, achieve synchronization of chaotic trajectories? In other words, if the original OGY method was used to stabilize unstable periodic orbits, can we modify it in order to stabilize a chaotic trajectory of one system in a relatively small vicinity of the chaotic trajectory of another system? A positive answer to this question was already obtained by using of continuous control methods. Now we consider this question as applied to discrete parametric OGY control.

Suppose we have two chaotic systems A and B, and let some system parameter (say, of system B) is available for alteration. Let us also assume that some system variables of both systems can be measured. Based on those measurements we can change a moment of time when the measured variables are close to each other. Having calculated the required parameter perturbation using the OGY method we can synchronize the systems in a short time period. Due to the inevitable presence of noise there is a finite probability of losing the synchronization. However, because of ergodicity, after some time the system's trajectories will again appear close in the phase space, and we will be able to synchronize them anew.

Let us realize this scheme for the case of two almost identical chaotic systems, which can be described by the following two-dimensional mappings:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{F}(\mathbf{x}_n, p_0) \quad [A]\,, \\ \mathbf{y}_{n+1} &= \mathbf{F}(\mathbf{y}_{n+1}, p) \quad [B]\,, \end{aligned} \tag{6.31}$$

where $\mathbf{x}_n, \mathbf{y}_n \in \mathrm{R}^2$, $\mathbf{F}$ is an analytic function of its variables, p_0 is a fixed parameter for the system A, and p is an externally fitted parameter of the system B. As in the OGY control case, we require a small variation region of the parameter p $|p - p_0| < \delta$. Suppose that the systems start from different initial conditions. Generally speaking, the chaotic trajectories describing the evolution of each system are absolutely uncorrelated. However, due to ergodicity of motion, with unit probability they will appear arbitrary close to each other at some later moment n_c. Without control, the trajectories begin to diverge exponentially for $n > n_c$. Our goal is to program the parameter p variation in such way that $|\mathbf{y}_n - \mathbf{x}_n| \to 0$ for $n \geqslant n_c$. Linearized dynamics in vicinity of the target trajectory $\{x_n\}$

$$\mathbf{y}_{n+1} - \mathbf{x}_{n+1}(p_0) = \hat{A}\,[\mathbf{y}_n - \mathbf{x}_n(p_0)] + \mathbf{B}\delta p_n \tag{6.32}$$

(see definitions in Sect. 5.3 of Chap. 5). As we have already pointed out in consideration of chaos control in Hamiltonian systems, due to the conservation of phase volume, the Jacobi matrix can have complex eigenvalues in this case. That is why it is convenient for the description of linearized dynamics to transit from eigenvectors to stable and unstable directions at every point of the chaotic orbit. Let $\mathbf{e}_{s(n)}$ and $\mathbf{e}_{u(n)}$ be unit vectors in those directions, and $\left\{\mathbf{f}_{s(n)}, \mathbf{f}_{u(n)}\right\}$ is the corresponding "orthogonal" basis, defined by the relations (5.15) in Chap. 5. Then, in this basis the condition under which the vector $\mathbf{y}_{n+1}$ gets onto the stable direction of the point $\mathbf{x}_{n+1}(p_0)$, which is required to realize synchronization, reads as the following:

$$[\mathbf{y}_{n+1} - \mathbf{x}_{n+1}(p_0)] \cdot \mathbf{f}_{u(n+1)} = 0 . \tag{6.33}$$

Using (6.32) and (6.33), we get the parameter perturbation $\delta p_n = p_n - p_0$, necessary to satisfy that condition:

$$\delta p_n = \frac{\left\{\hat{A} \cdot [\mathbf{y}_n - \mathbf{x}_n(p_0)]\right\} \cdot \mathbf{f}_{u(n+1)}}{-\mathbf{B} \cdot \mathbf{f}_{u(n+1)}} . \tag{6.34}$$

If $(\Delta p)_n$ calculated according to (6.34) appears greater than δ, we set $\delta p_n = 0$.

Let us check the efficiency of the functioning of this scheme in a Hénon mapping ((5.15), Chap. 5). Let us fix the value of $p = p_0 = 1.4$ for one of the systems, and for the other, we will consider it as a fitting parameter, changing according to (6.34) in a small interval $[1.39, 1.41]$. Let the two systems start in the moment $t = 0$ from different initial conditions: $(x_1, y_1) = (0.5, -0.8)$ and $(x_2, y_2) = (0.001, 0.001)$. Then the two systems move along completely uncorrelated chaotic trajectories. At some moment, the systems appear sufficiently close one to another. The required proximity of the trajectories is determined by the magnitude of the parameter δ. When that happens, we switch on the synchronization mechanism, i.e., the perturbation of the parameter p according to (6.34). Figure 6.5a shows time sequences for the two chaotic trajectories (crosses and squares) before and after the synchronization mechanism is switched on. It is clear that after the control is switched on (approximately the 2500th iteration) the crosses and the squares overlap, though the trajectories remain chaotic. Figure 6.5b presents the time dependence of $\Delta x(t) = x_2(t) - x_1(t)$, tending to zero after the synchronization mechanism is switched on. The time needed to achieve the synchronization, as well as the control setup time, dramatically grows with the decrease of δ. Unfortunately, a direct application of the targeting methods considered in the previous chapter, allowing us to shorten the control setup time considerably, is impossible: in the case of control the target unstable periodic orbit is fixed, and in the case of synchronization the target is not only unfixed, but it also moves chaotically, which is why the problem becomes extremely complicated.

The following problem [20] is very close in formulation to the problems of periodic control, where stabilization is achieved due to the purposeful alteration of its parameters. Suppose

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}) \tag{6.35}$$

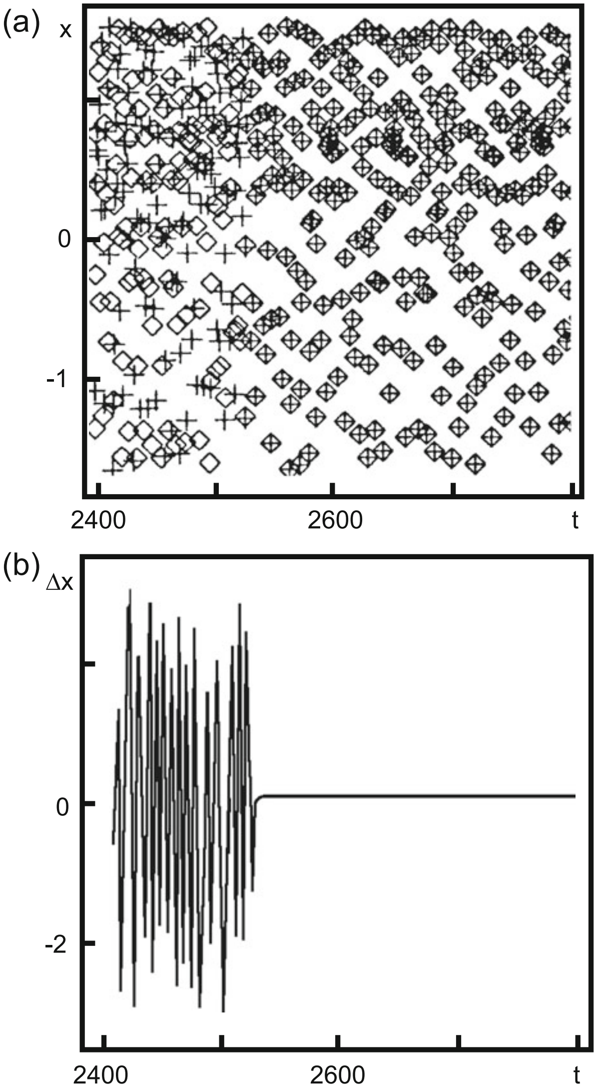


Fig. 6.5 Synchronization of two Hénon mappings: (**a**) two chaotic trajectories before and after the control switch on; (**b**) time dependence of $\Delta x = x_2 - x_1$, corresponding to (**a**) [19]

is an experimental realization of a dynamical system, whose parameters $p \in \mathrm{R}^m$ are known. Let us consider that we know the time dependence of some scalar observable quantity $s = h(\mathbf{x})$ and function $\mathbf{f}$, describing the model dynamics. Suppose, then, that we can construct the system

$$\dot{\mathbf{y}} = \mathbf{g}\left(s, \mathbf{y}, \mathbf{q}\right) , \tag{6.36}$$

which will be synchronized with the first one ($\mathbf{y} \to \mathbf{x},\ t \to \infty$), if $\mathbf{q} = \mathbf{p}$. If the functional form of the vector field $\mathbf{f}$ is known, then for the construction of the required subsystem we can use the decomposition methods considered in Sect. 6.1. The answer that we are interested in is the following: can we construct an ordinary differential equations system for parameters $\mathbf{q}$,

$$\dot{\mathbf{q}} = \mathbf{u}\,(s, \mathbf{y}, \mathbf{q}) \tag{6.37}$$

such that $(\mathbf{y}, \mathbf{q}) \to (\mathbf{x}, \mathbf{p})$ if $t \to \infty$. Let us show on a concrete example that, generally speaking, there is a positive answer to that question. To that end, we again address the Lorentz system

$$\begin{aligned} \dot{x}_1 &= \sigma\,(x_2 - x_1)\ , \\ \dot{x}_2 &= p_1 x_1 - p_2 x_2 - x_1 x_3 + p_3\,, \\ \dot{x}_3 &= x_1 x_2 - b x_3\,, \end{aligned} \tag{6.38}$$

with $p_1 = 28$, $p_2 = 1$, $p_3 = 0$, $b = 8/3$. We will assume that the observable variable is $s = h\,(\mathbf{x}) = x_2$. We will use it as the driving variable,

$$\begin{aligned} \dot{y}_1 &= \sigma\,(s - y_1)\ , \\ \dot{y}_2 &= q_1 y_1 - q_2 y_2 - y_1 y_3 + q_3\,, \\ \dot{y}_3 &= y_1 y_2 - b y_3\,. \end{aligned} \tag{6.39}$$

Suppose that the parameters q variation process is described by the following system of equations:

$$\begin{aligned} \dot{q}_1 &= u_1\,(s, \mathbf{y}, \mathbf{q}) = [s - h\,(\mathbf{y})]\,y_1 = (x_2 - y_2)\,y_1\,, \\ \dot{q}_2 &= u_2\,(s, \mathbf{y}, \mathbf{q}) = [s - h\,(\mathbf{y})]\,y_2 = -\,(x_2 - y_2)\,y_2\,, \\ \dot{q}_3 &= u_3\,(s, \mathbf{y}, \mathbf{q}) = [s - h\,(\mathbf{y})] = (x_2 - y_2)\ . \end{aligned} \tag{6.40}$$

In order to show that $(\mathbf{y}, \mathbf{q}) = (\mathbf{x}, \mathbf{p})$ are the stable solutions of (6.39), (6.40), it is necessary to study dynamics of the differences $\mathbf{e} \equiv \mathbf{y} - \mathbf{x}$ and $\mathbf{f} = \mathbf{q} - \mathbf{p}$. Those differences obey the following system of equations:

$$\begin{aligned} \dot{e}_1 &= -\sigma e_1\,, \\ \dot{e}_2 &= q_1 y_1 - p_1 x_1 - q_2 y_2 + p_2 x_2 - y_1 y_3 + x_1 x_3 + f_3\,, \\ \dot{e}_3 &= y_1 y_2 - x_1 x_2 - b e_3 \\ \dot{f}_1 &= -e_2 y_1, \quad \dot{f}_2 = e_2 y_2, \quad \dot{f}_3 = -e_2\,. \end{aligned} \tag{6.41}$$

where the parameters $\mathbf{p}$ are assumed to be constant. From the first equation it follows that $e_1 \to 0$, i.e., $y_1 \to x_1$. In the limit $t \to \infty$ the system (6.41) can be

$$\begin{aligned} \dot{e}_2 &= f_1 y_1 - f_2 y_2 - p_2 e_2 - y_1 e_3 + f_3\,, \\ \dot{e}_3 &= y_1 e_2 - b e_3\,, \\ \dot{f}_1 &= -e_2 y_1, \quad \dot{f}_2 = e_2 y_2, \quad \dot{f}_3 = -e_2\,. \end{aligned} \tag{6.42}$$

In order to study the global stability of the system we will use the Lyapunov functions method [21], whose main principle is the following. Suppose that on a plane (the method works in a space of any dimension, but we restrict ourselves to the plane) there is a vector field with a fixed point $(\bar{x}, \bar{y})$, and we want to know whether it is stable. In accordance with obvious ideas about stability, it will suffice to find some vicinity U of the fixed point, such that the trajectory starting in U remains inside it at all the consecutive time moments. This condition can be satisfied if the vector field on the boundary of U is directed either inside the region towards $(\bar{x}, \bar{y})$, or is tangential to the boundary (see Fig. 6.6a). The Lyapunov functions method allows us to answer the question of whether the considered vector field has such a geometry.

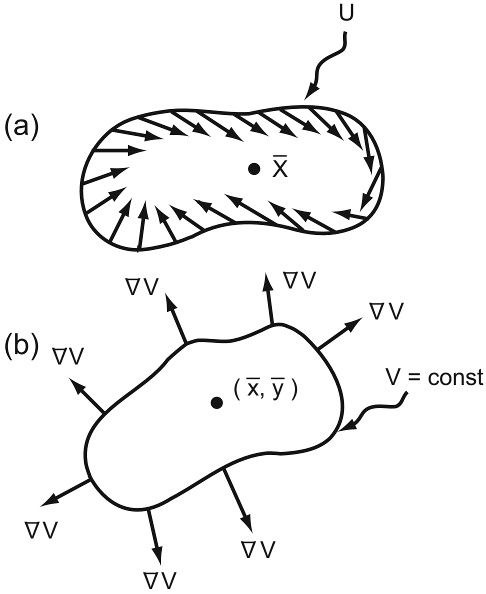


Fig. 6.6 (**a**) Vector field on the boundary of U. (**b**) Gradient of V in different points of the boundary

Suppose that the considered vector field is defined by the equations

$$\frac{dx}{dt} = f(x,y)\,,$$
$$\frac{dy}{dt} = g(x,y)\,. \tag{6.43}$$

Let $V(x,y)$ be some scalar function on R^2, at least once differentiable. As well, $V(\bar{x},\bar{y}) = 0$ and the set of points, satisfying the condition $V(x,y) = C$, form closed curves, surrounding the point $(\bar{x},\bar{y})$ for different values of C, while $V\,(x,y) > 0$ (see Fig. 6.6b). It is easy to understand that if the vector field has the above geometry, then

$$\nabla V(x,y) \cdot (dx/dt, dy/dt) = \dot{V} \leqslant 0\,. \tag{6.44}$$

Thus, if it is possible to construct a function with given properties (the Lyapunov function), satisfying the condition (6.44), then the considered fixed point is globally stable.

Let us now return to considering the stability of the system (6.42). For the Lyapunov function we choose the following:

$$V = e_2^2 + e_3^2 + f_1^2 + f_2^2 + f_3^2\,. \tag{6.45}$$

Using Eq. (6.42), we get

$$\dot{V} = -2\left(p_2 e_2^2 + b e_3^2\right)\,. \tag{6.46}$$

For $p_2 > 0$ $(b = 8/3)$ that derivative is negative, and, therefore, according to (6.44), the driven system parameters $\mathbf{q}$ on large time scales tend to values of the initial system parameters $\mathbf{p}$. Figure 6.7a illustrates this process. For initial conditions we have chosen the following $\mathbf{x} = (0.1, 0.1, 0.1)$, $\mathbf{y} = (-0.1, 0.1, 0)$, $\mathbf{q} = (10, 10, 10)$. The points on the figure denote the values of the parameters $p_1/10 = 2.8$, $p_2 = 1$, $p_3 = 0$ (the first coefficient is divided by ten for convenience). In this case, we assume that all other coefficients coincide exactly. On the figure, one can see quite rapid $(\mathbf{q} \to \mathbf{p})$ convergence. Figure 6.7b shows the same process, but for a case when the driving system parameter $\sigma = 10$ is replaced by the value $\sigma = 10.1$ for the driven system. In this case, there is no exact convergence, but oscillations of the parameters $\mathbf{q}$ around the exact values are observed.

6.6 Synchronization Induced by Noise

In this section we will consider one more example of the constructive role of chaos—the synchronization of chaotic systems with help of additive noise [22]. The effect that we intend to study consists of the fact that the introduction of noise (with

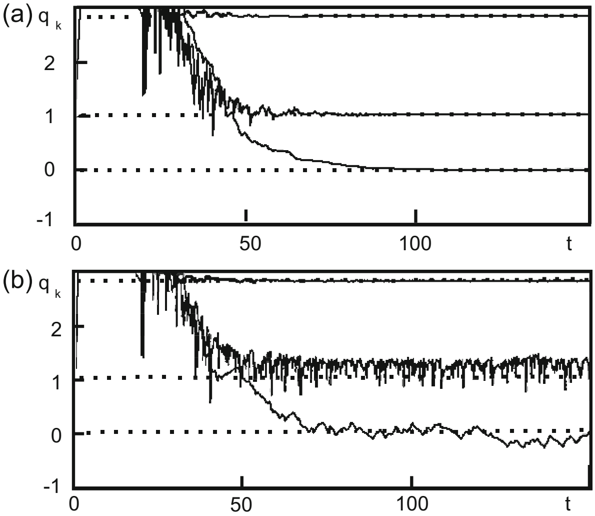


Fig. 6.7 (**a**) Process of $\overline{q} \rightarrow \overline{p}$ convergence for coinciding values of other parameters. (**b**) The same process for values $\sigma = 10$ (driving system) and $\sigma = 10.1$ (driven system) [20]

sufficiently high intensity) in independent copies of systems makes them collapse in the same trajectory, independently of the initial conditions of each copy. This synchronization of chaotic systems represents one more example that contradicts intuitive ideas of the destructive role of noise. We want to clarify the essence of the effect and to analyze the structural stability of the phenomenon.

Noise-induced synchronization has a short but interesting history. The ordering effects of noise in chaotic systems were first considered in the paper [23], the authors of which came to the conclusion that noise can make a system less chaotic. Later, in [24] the noise-induced chaos-regularity transition was demonstrated. Noise-induced synchronization was considered for the first time in [25]. The authors showed that particles in external potential, subject to random forces, tend to collapse on the same trajectory. Among the further papers written on that topic we would emphasize the one [26] which evoked violent polemics. The authors of the paper analyzed the logistic mapping

$$x_{n+1} = 4x_n(1 - x_n) + \xi_n \,, \tag{6.47}$$

where ξ_n is the noise term with homogeneous distribution on the interval $[-W, W]$.

They showed that if W is sufficiently large (i.e., for high noise intensities), two different trajectories starting from distinct initial conditions but subject to identical noise (the same sequences of random numbers) will at last end at the same trajectory. The authors showed that the same situation also takes place for the Lorenz system. This result provoked a harsh criticism [27], connected with the fact that the two systems can be synchronized only in the case when the maximal Lyapunov exponent is negative. For the logistic mapping in presence of noise, the maximal Lyapunov exponent is positive and therefore the observed synchronization is the result of a loss in calculation accuracy. It was also noted [28] that the noise used for the simulation (6.47) is not symmetrical in reality. A non-zero mean value $\langle \xi_n \rangle$ appears because the requirement $x_n \in [0, 1]$ forces us to exclude those random numbers that induce any violation of that condition. The introduction of noise with a non-zero mean value implies that the authors of [26] essentially changed properties of the original deterministic mapping. As a result of a whole row of works it was, however, shown that some chaotic mappings can nevertheless be synchronized by additive noise with zero mean. The mechanism leading to synchronization was explained in [29]; its essence is the following. As we have already mentioned, synchronization can be achieved only in the case of negative Lyapunov exponents. In presence of noise, due to the reconstruction of the distribution function, the system appears to spend more time in the regions of stability, where the Lyapunov exponents are negative, and it provides the global negativity of the Lyapunov exponents. Let us analyze this reasoning in more detail.

Let us consider the mapping

$$x_{n+1} = F(x_n) = f(x_n) + \varepsilon \xi_n \,, \tag{6.48}$$

where $\{\xi_n\}$ is the set of uncorrelated Gaussian variables with zero mean value and unit dispersion. For an example of a concrete realization of (6.48) we choose the following:

$$f(x) = \exp\left[-\left(\frac{x - 0.5}{\omega}\right)^2\right] . \tag{6.49}$$

The investigation of the relative behavior of the trajectories, described by (6.48) and starting from different initial conditions, is equivalent to an analysis of such behavior in two identical systems of the form (6.48) subject to the same noise, under which we understand using the same sequence of random numbers $\{\xi_n\}$. Figure 6.8 shows the bifurcation diagram for that mapping in absence of noise. The chaoticity regions are well visible on the diagram. In those regions the maximal Lyapunov exponent is positive. So, for example, for $\omega = 0.3$ (this case will be analyzed further) $\lambda \approx 0.53$. In Fig. 6.9 one can see that at a sufficient noise level ε, for most values of ω this Lyapunov exponent becomes negative. So for $\omega = 0.3$ and $\varepsilon = 0.2$ we find that $\lambda = -0.17$.

The positivity of the Lyapunov exponent in a noiseless case means that the trajectories starting from different initial conditions are excited by the determined part $f(x_n)$, and by the same random sequence of numbers $\{\xi_n\}$, will not coincide at any arbitrarily large n. In this case, the synchronization diagram ($x^{(2)}$ as a function of $x^{(1)}$) represents a wide and almost uniform distribution (Fig. 6.10a). However, at $\varepsilon \geqslant 0.2$, when the maximal Lyapunov exponent becomes negative, we observe almost complete synchronization (Fig. 6.10b). The noise intensity is not high enough to

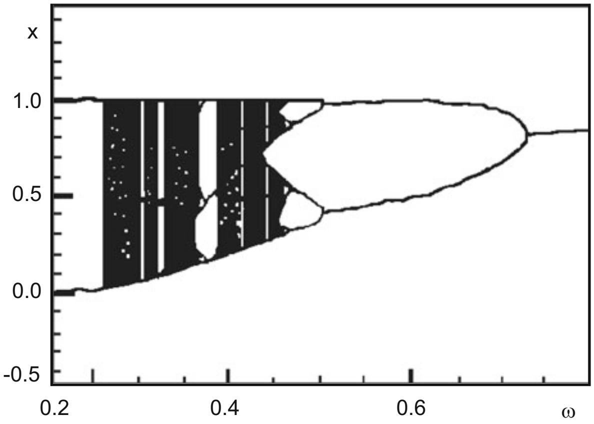


Fig. 6.8 Bifurcation diagram for the mapping (6.48) in absence of noise [22]

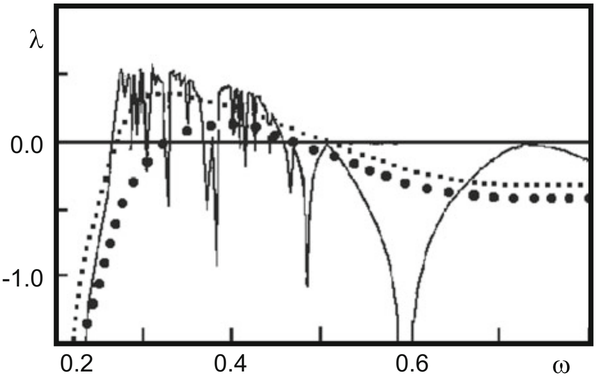


Fig. 6.9 Lyapunov exponent for the mapping (6.49): $\varepsilon = 0$ (*solid line*), $\varepsilon = 0.1$ (*dashed line*), $\varepsilon = 0.2$ (*dash-dot line*) [22]

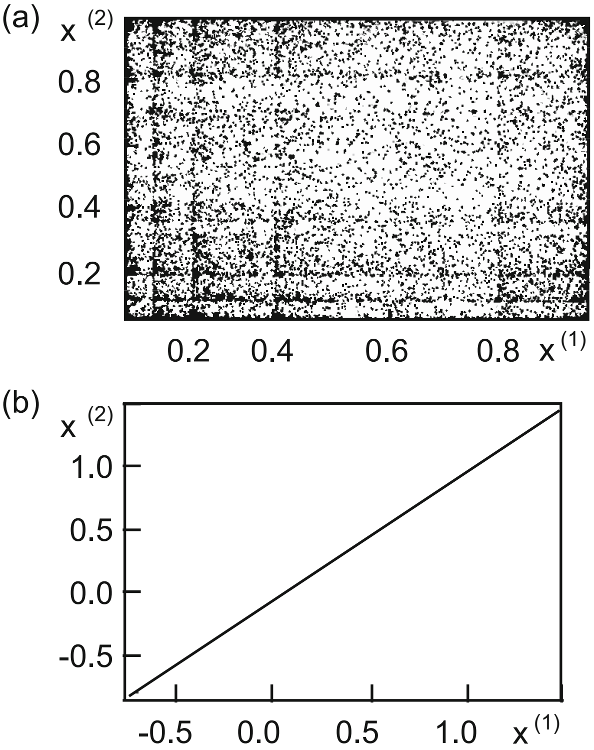
neglect the deterministic term in (6.48). Therefore, the synchronization mechanism that we want to understand is far from trivial.

The Lyapunov exponent determining the synchronization condition for the mapping (6.48) can be represented in the form

$$\lambda = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \ln \left| F'(x_i) \right| . \tag{6.50}$$

This expression represents the mean value of the logarithm of the absolute value of the derivative F' (slope), calculated along the trajectory $\{x_i\}$. The slopes in the interval $[-1, 1]$ give negative contribution in λ, leading to the synchronization. Larger or smaller slopes give positive contribution in λ and generate a divergence

Fig. 6.10 The synchronization diagram ($x^{(2)}$ as a function of $x^{(1)}$) for the case $\omega = 0.3$. (**a**) $\varepsilon = 0$, (**b**) $\varepsilon = 0.2$ [22]



of the trajectories. At first sight it seems, as $F' = f'$, that the presence of noise does not modify the Lyapunov exponent. However, this is not so. The modification of the Lyapunov exponent due to noise is connected with noise-induced modification of the trajectory, along which the averaging (6.50) takes place. In order to understand this, we will use the expression for the Lyapunov exponent in terms of the stationary distribution function $P_{\mathrm{st}}(x)$,

$$\lambda = \left\langle \log \left| F'(x) \right| \right\rangle = \left\langle \log \left| f'(x) \right| \right\rangle \equiv \int P_{\mathrm{st}}(x) \log \left| f'(x) \right| dx \,. \tag{6.51}$$

We see that with the inclusion of any perturbation there are two mechanisms leading to the modification of the Lyapunov exponent: the change of $|f'(x)|$ and the reconstruction of the distribution function. At the inclusion of the additive noise, the latter mechanism works. In Fig. 6.11, one can see the reconstruction of the stationary distribution function for the mapping (6.48). We can conclude that synchronization will be a common feature of those mappings [for example, (6.48)], for which, with the inclusion of noise, the regions with $|f'(x)| < 1$ have sufficient statistical weight.

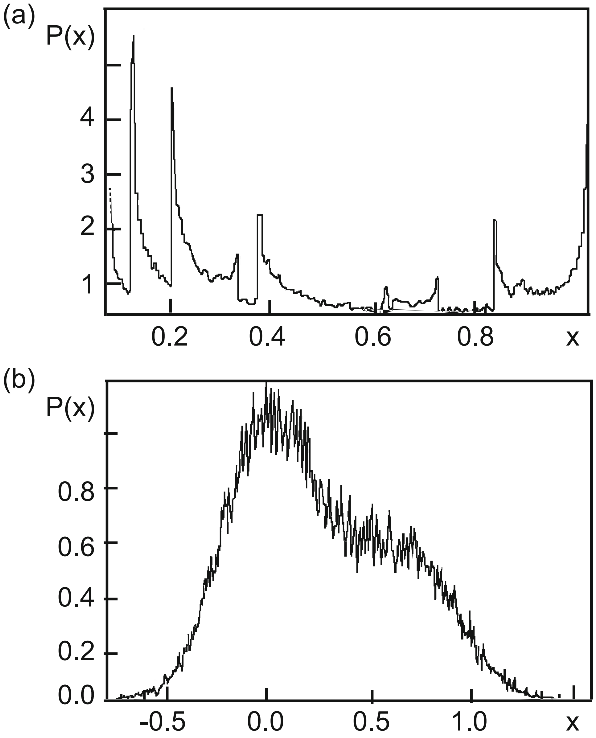


Fig. 6.11 Distribution function for the mapping (6.48) in the case $\omega = 0.3$. (**a**) $\varepsilon = 0$, (**b**) $\varepsilon = 0.2$ [22]

Let us consider one more example—noise-induced synchronization in the Lorenz system with additive noise, introduced into the equation for the coordinate y,

$$\begin{aligned} \dot{x} &= p(y - x)\,, \\ \dot{y} &= -xz + rx - y + \varepsilon\xi\,, \\ \dot{z} &= xy - bz\,. \end{aligned} \tag{6.52}$$

Here $\xi(t)$—the white noise—is the Gaussian random process with zero mean: $\langle\xi(t)\rangle = 0$; $\langle\xi(t)\xi(t')\rangle = \delta(t - t')$. As we have already seen in the previous chapter, for the parameter values $p = 0,\ b = 8/3,\ r = 28$ and in the absence of noise ($\varepsilon = 0$), the system (6.52) is chaotic (the maximal Lyapunov exponent is $\lambda \approx 0.9 > 0$). Therefore, the trajectories starting from different initial conditions are absolutely uncorrelated (see Fig. 6.12a). The same situation also takes place at low noise intensities. However, at a noise intensity that provides a negative maximal Lyapunov exponent (for $\varepsilon = 40$, $\lambda \approx -0.2$), almost complete synchronization of all three coordinates is observed (see Fig. 6.12b for the coordinate z). We stress that,

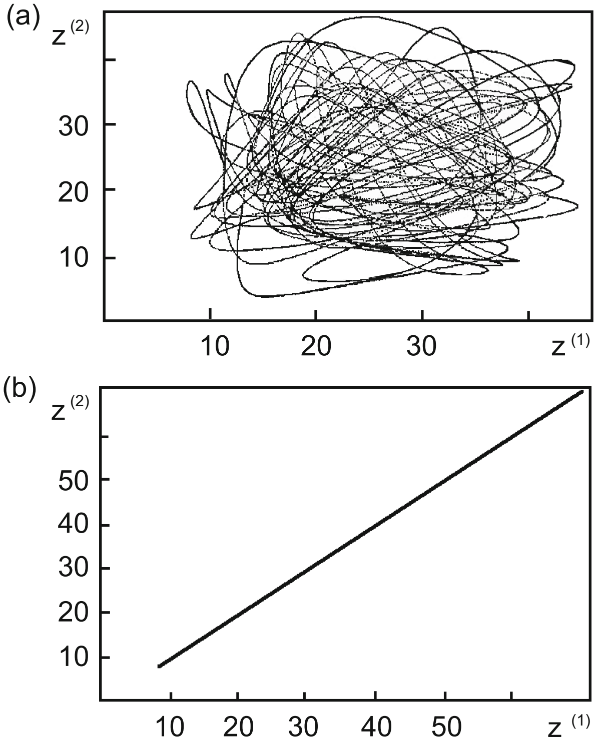


Fig. 6.12 Noise-induced synchronization for the Lorenz system [22]. (**a**) $\varepsilon = 0$, (**b**) $\varepsilon = 40$

although the noise intensity is relatively high, the structure of the strange attractor preserves the "butterfly" topology, characteristic for the deterministic case. This fact stresses again that in the considered examples we are not dealing with trivial synchronization, which takes place in the case when the deterministic terms in the mapping (or equations of motion) can be neglected.

A natural question arises about the structural stability of the considered phenomenon. Unlike the two identical Lorenz systems (with the same coefficients p, b, r) two real systems never have identical sets of parameters. Therefore, if we intend to use noise-induced synchronization, for communication purposes, for example, we should preliminarily estimate the permissible difference between the parameters of the transmitter and the receiver. In order to solve this problem, we will numerically analyze the dynamics of two Lorenz systems with slightly different parameters (p_1, b_1, r_1) and (p_2, b_2, r_2), but subject to the same noise factor $\varepsilon\xi$. In order to estimate the effects of variation on each of the parameters, we will vary them independently. The result of the procedure is presented in Fig. 6.13. On that figure we plot the part of the full observation time (in percent), during which the systems were synchronized with an accuracy up to 10 %. This means that the trajectories of the two systems were considered synchronized if the relative difference of their coordinates was less than

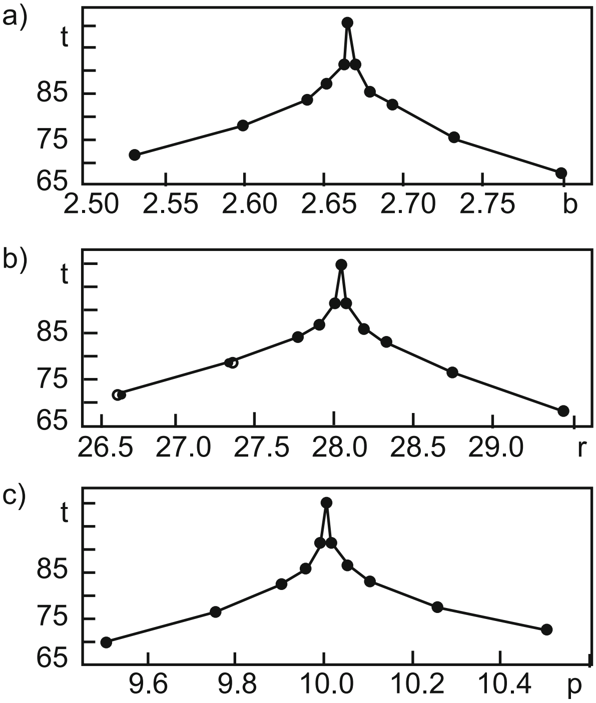


Fig. 6.13 The synchronization time for the Lorenz system (in percents with respect to the total observation time) as a function of the parameters [22]: (**a**) *b*, (**b**) *r*, (**c**) *p*

10 %. From Fig. 6.13 one can conclude that, with a parameter variation of an order of 1 %, during 85 % of total observation time, the systems remained synchronized.

6.7 Synchronization of Space-Temporal Chaos

Most physical phenomena in domains where we deal with extended physical objects (hydrodynamics, electromagnetism, plasma physics, chemical dynamics, biological physics, and many others) can be described only with the help of partial derivative equations. Only with some simplifying assumptions do those equations reduce to a system of connected ordinary differential equations or connected grid mappings. All of the examples of chaotic systems synchronization that we have considered belong to finite-dimensional (moreover, low-dimensional) systems. The behavior of spatially extended non-linear systems is considerably complicated by space-temporal chaos (turbulence), which is characteristic for most of them. In these cases, chaotic behavior is observed both in time and in space. A natural question arises: how efficient will the above low-dimensional systems synchronization methods be for space-temporal chaos? We will not dwell on this question in detail, instead redirecting the reader to the reviews [12, 30]. We will only consider the possibility of space-temporal chaos synchronization [31] on an example of an autocatalytic model, demonstrating chaos [32],

$$\begin{aligned}\frac{\partial u_1}{\partial t} &= -u_1 v_1^2 + a(1-u_1) + D_u \nabla^2 u_1 \,, \\ \frac{\partial v_1}{\partial t} &= u_1 v_1^2 - (a+b)v_1 + D_v \nabla^2 v_1 \,,\end{aligned} \tag{6.53}$$

where u_1 and v_1 are reactive and activator concentrations, respectively, a, b are reaction parameters, and D_u, D_v are diffusion constants. We will consider the system (6.53) as driving in relation to the analogous system

$$\begin{aligned}\frac{\partial u_2}{\partial t} &= -u_2 v_2^2 + a(1-u_2) + D_u \nabla^2 u_2 \,, \\ \frac{\partial v_2}{\partial t} &= u_2 v_2^2 - (a+b)v_2 + D_v \nabla^2 v_2 + f(x,t) \,.\end{aligned} \tag{6.54}$$

Suppose $v_2(t-0)$ is value of v_2 immediately before time moment t_2. The driving function $f(x,t)$ acts on the system in the following way. Let L be the linear dimension of the chemical reactor, $L = NX$, $t = kT$, $T > 0$, $X > 0$, N, k are integer numbers. In every moment of time $t = kT$ in N spatial points $x = 0, X, 2X, \ldots, (N-1)X$ the driving function transits simultaneously

$$v_2(kT-0) \to v_2(kT) = v_2(kT-0) + \varepsilon \left[v_1(kT) - v_2(kT-0) \right] \,. \tag{6.55}$$

In the time moments $t \neq kT$ the systems (6.53) and (6.54) are not connected and evolve independently. We note that for $X = T = 0$, $\varepsilon = 1$ such driving reduces to the full replacement considered above. Motivation to select driving in the form (6.55) is determined by two reasons. On one hand we intend to achieve synchronization by controlling only a finite number N of spatial points, and on the other hand, we want to use time-discrete perturbation to do this.

The results of numerical simulation of evolution described by (6.53), (6.54) are presented in Fig. 6.14. For integration, the Euler scheme was implemented with $M = 256$ spatial nodes and time step equal to $\Delta t = 0.05$. The following parameter values were chosen:

$$a = 0.028,\ b = 0.053,\ D_v = 1.0 \times 10^{-5},\ D_u = 2D_v,\ L = 2.5 \,.$$

Figure 6.14a demonstrates the space-temporal evolution $u_1(x,t)$, described by (6.53), with initial conditions $u(x) = 1$, $v(x) = 0$.

To simulate the partial derivative equation systems (6.53), (6.54) with the condition (6.55) the following parameters values were taken: $\varepsilon = 0.2$, $T = 20\Delta t$,

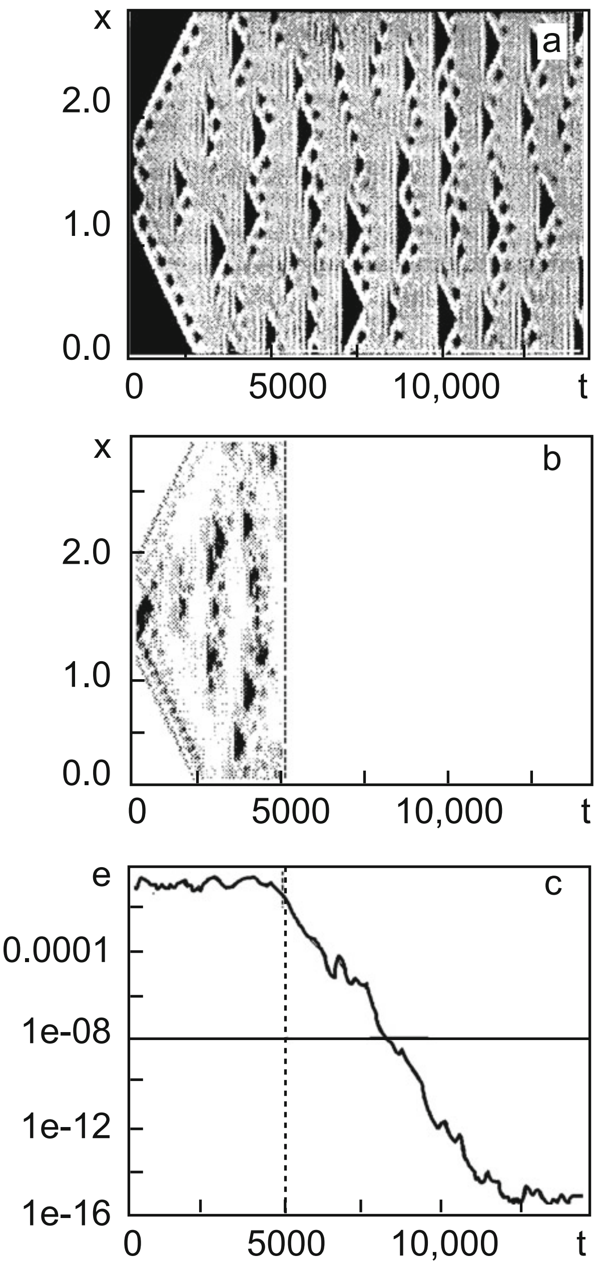


Fig. 6.14 The results of numerical simulation of evolution described by the systems (6.53), (6.54): (**a**) space-temporal dependence $u_1(x,t)$; (**b**) difference $|u_1 - u_2|$; (**c**) global synchronization error $e(t)$(6.56) [30]

$X = (8/256)\,L$. In other words, the perturbation acted on 32 of 256 spatial nodes. It appeared that there is a critical value X_{cr}, such that for all $X < X_{\mathrm{cr}}$ the systems (6.53) and (6.54) can be synchronized. For the chosen parameter set $X_{\mathrm{cr}} = (14/256)\,L$, and this number does not change with an increase of M. This important example shows that an infinite-dimensional system can be synchronized by the perturbation of a finite number of points, i.e., synchronization is achieved with help of the driving signal in the form of N-dimensional vector.

Suppose the driving function is turned on at $t = 5000$. Figure 6.14b presents the difference $|u_1 - u_2|$ (the turn-on moment is denoted by the dashed line). Those regions of (x, t) space, where that function is large, i.e., the desynchronization regions, are painted in black. One can see that such regions are present only before the moment the driving signal is turned on, $t < 5000$. In order to make the effect clearer, we introduce the global synchronization error $e(t)$,

$$e = \sqrt{\frac{1}{L}\int_0^L \left[(u_1 - u_2)^2 + (v_1 - v_2)^2\right] dx}\,. \tag{6.56}$$

As one can see from Fig. 6.14c, that error tends to zero after the synchronization mechanism is turned on (6.55).

6.8 Additive Noise and Non-identity Systems Influence on Synchronization Effects

In the present section we intend to make a quantitative investigation of the transition of the initial idealized problem formulation (identical system synchronization in absence of noise) to a realistic one, accounting for the obligatory presence of internal noise and deviation in the system parameters [33]. The latter implies that the free dynamics of the driving and of the driven systems will differ for the same initial conditions. In the transition from idealization to reality we face the problem of the experimentally measurable time series synchronization. Under the driving system we will understand an experimentally observable system, whose dynamics are known only in the sense that the time series of the system's characteristics measurements are given. The driven system represents a model that can be constructed based on the temporal measurements made with the driving system. Suppose that the unknown dynamics of the driving system in some "work phase space" is represented by the equation

$$\dot{\mathbf{x}} = \mathbf{G}(\mathbf{x}) \tag{6.57}$$

and the model dynamics in the same space is

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})\,. \tag{6.58}$$

We assume that the corresponding embedding theorems (see Chap. 4) provide existence of (6.57) in the work phase space. Figure 6.15 represents an example of synchronization in the model (6.58) with the time series obtained from (6.57). Let $\mathbf{x}(t)$ be some trajectory, measured with help of some "experimental setup" (6.57). We now use that trajectory and the model (6.58) in order to generate two new trajectories. The trajectory $\mathbf{w}(t)$ is obtained by forward time integration of (6.58) using the first point of the trajectory $\mathbf{x}(t)$ as the initial condition. The trajectory $\mathbf{y}(t)$ results from the synchronization process: the substitution of the measured time series for one coordinate into the model equation (6.58). The lower curve in Fig. 6.15 represents the square of the distance between the driving and the driven trajectories $|\mathbf{z}|^2 = |\mathbf{x} - \mathbf{y}|^2$. The upper curve is the distance between the driving trajectory and the free one in the model system $|\mathbf{z}|^2 = |\mathbf{x} - \mathbf{w}|^2$. The degree of smallness of the lower curve with respect to the upper one determines the quality of the synchronization.

In the total absence of noise and model errors (i.e., for $\mathbf{F} = \mathbf{G}$) we expect exact synchronization $|\mathbf{z}|^2 = 0$. For physical devices and model equations, this will never happen, as in the driving signal there is always a noise component and model errors are inevitable. Therefore, a physical device and a model can be synchronized only approximately. As there are no two exactly identical devices, this also concerns the synchronization of two experimental setups. It is natural to expect that with an increase of the noise level or of the magnitude of model errors, the lower curve amplitude in Fig. 6.15 will grow. It is the character of that growth which determines the quantitative measure of the influence of noise and model errors on the synchronization process.

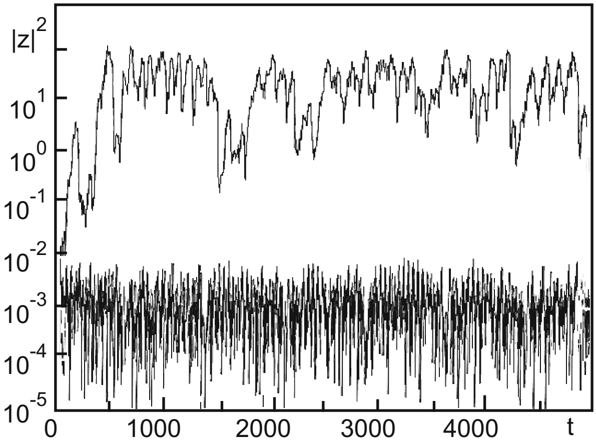


Fig. 6.15 An example of synchronization for the model (6.58). *Lower curve*: squared difference between the driving and the driven trajectories $|z|^2 = |x - y|^2$; *upper curve*: distance between the driving and the free trajectory of the model system $|\overline{z}|^2 = |\overline{x} - \overline{w}|^2$ [33]

Let us now use the following quantity as the driving signal:

$$\mathbf{x} + \sigma\mathbf{u}\,. \tag{6.59}$$

Here $\mathbf{x}$ is the time series (6.57), $\sigma\mathbf{u}$ is the additive noise term, associated with errors in the driving signal measurements, σ is the noise level, and $\mathbf{u}$ is a random Gaussian vector with unit dispersion of the components and zero mean value. Errors may be induced by random deviations of the device parameters from nominal values and by background noise, measured together with the signal. To synchronize the device (6.57) and the model (6.58) we use negative feedback (6.59)

$$\dot{\mathbf{y}} = \mathbf{F}(\mathbf{y}) - \hat{E}[\mathbf{y} - (\mathbf{x} + \sigma\mathbf{u})]\,. \tag{6.60}$$

The matrix $\hat{E}$ determines the connection between $\mathbf{y}$ and the experimentally measured time series. Further, we assume that the matrix has a unique non-zero element lying on the diagonal and let this element be $E_{ii} = \varepsilon$, if the ith component of $\mathbf{x} + \sigma\mathbf{u}$ is used as the driving signal. Inside some region of values ε, determining the negativity of the maximal conditional Lyapunov exponent, the feedback (6.60) must lead to synchronization between $\mathbf{x}$ and $\mathbf{y}$, and all deviations are connected either to model errors or to the presence of noise. Assuming the smallness of $|\mathbf{z}|$—the deviation of the model dynamics from the device dynamics—the linearized time evolution $\mathbf{z}$ is described by the equation

$$\dot{\mathbf{z}} = \left[\mathbf{DF}(\mathbf{x}) - \hat{E}\right]\mathbf{z} + \sigma\hat{E}\mathbf{u} + \Delta\mathbf{G}\,(x)\,, \tag{6.61}$$

where $\Delta\mathbf{G} = \mathbf{F} - \mathbf{G}$, $(\mathbf{DF})_{ij} = \frac{\partial F_i}{\partial x_j}$. The quantity $\Delta\mathbf{G}$ has two potential sources. The first source generating $\Delta\mathbf{G}$ is the error arising from the modeling of an unknown vector field $\mathbf{G}$. In any real situation $\mathbf{F}$ and $\mathbf{G}$ never coincide. The second source is connected to the fact that the driving signal dynamics differ from the dynamics reproduced by the time series used to construct the model. In order to separate these two sources we assume that the time series used to construct the model comes from vector field $\mathbf{G}$, while the driving signal is generated by the field $\mathbf{G}'$. We will consider that distinction of those two fields is connected to the variation of some parameter set $\mathbf{p}$ of the driving system, i.e.,

$$\mathbf{G} \cong \mathbf{G}' + \left(\partial\mathbf{G}'/\partial\mathbf{p}\right)\cdot\delta\mathbf{p} \tag{6.62}$$

then

$$\Delta\mathbf{G}\,(\mathbf{x}) \cong \Delta\mathbf{G}'\,(\mathbf{x}) + \left(\partial/\partial\mathbf{p}\left(\Delta\mathbf{G}'\,(\mathbf{x})\right)\right)\cdot\delta\mathbf{p}\,, \tag{6.63}$$

where $\Delta\mathbf{G}' \equiv \mathbf{F} - \mathbf{G}'$. Equation (6.61) [accounting for (6.63)] is an evolutionary equation for the connected device-model system in the vicinity of synchronized motion. In the absence of noise ($\sigma = 0$) and for ideal model dynamics ($\Delta\mathbf{G} = 0$)

$$\dot{\mathbf{z}} = \left[\mathbf{DF}\,(\mathbf{x}) - \hat{E}\right]\mathbf{z}\,. \tag{6.64}$$

The formal solution of that homogeneous linear equation reads

$$\mathbf{z}(t) = \exp\left[\int_{t_0}^{t}\left[\mathbf{DF}(\tau) - \hat{E}\right]d\tau\right]\mathbf{z}(t_0) \equiv \hat{U}(t,t_0)\mathbf{z}(t_0)\,, \tag{6.65}$$

where $\mathbf{DF}(\tau) = \mathbf{DF}\,[\mathbf{x}(\tau)]$. The evolution operator $\hat{U}(t,t_0)$ maps the initial condition $\mathbf{z}(t_0)$ forward in time accounting for the connection, but in the absence of noise and modeling errors. In order to obtain the general solution of Eq. (6.61) one should add its particular solution to the general solution of the homogeneous equation (6.65). To obtain the particular solution we make the variables transformation $\mathbf{z}(t) = \hat{U}(t,t_0)\mathbf{w}(t)$. Substitution in (6.61) gives

$$\frac{dw}{dt} = U^{-1}(t,t_0)\left[\Delta G(t) + \sigma E \cdot u(t)\right]\,. \tag{6.66}$$

Solving this equation taking into account (6.65), we obtain the general solution of Eq. (6.61) in the form

$$\mathbf{z}(t) = \hat{U}(t,t_0)\cdot\mathbf{z}(t_0) + \int_{t_0}^{t}\hat{U}(t,\tau)\cdot\left[\Delta G(\tau) + \sigma\hat{E}\cdot\mathbf{u}(\tau)\right]d\tau\,. \tag{6.67}$$

This equation describes the time evolution of the difference between the trajectory given by Eq. (6.58) and the "exact" system trajectory. Such a solution has a place only under conditions close to the synchronization regime. Because of the stability of synchronized motion, we can neglect the first term in (6.67), as it tends exponentially quickly to zero with increasing time. The second term in (6.67) describes complicated non-local dependence of $\mathbf{z}(t)$ on model errors and noise: the degree of synchronization at moment t is determined by model and noise fluctuations in all preceding moments.

Returning to Fig. 6.15 we note that, although the time dependence of $|\mathbf{z}|^2$ is very complex, its mean value is practically constant. This mean can be used to characterize the degree of synchronization between the exact driving signal and the one generated by the model. We define that characteristic by the following time average:

$$\left[\left\langle|\mathbf{z}|^2\right\rangle\right]^{1/2} = \left[\lim_{t\to\infty}\frac{1}{t-t_0}\int_{t_0}^{t}|\mathbf{z}(\tau)|^2\,d\tau\right]^{1/2}\,. \tag{6.68}$$

This expression can be represented in the form

$$\left[\left\langle |\mathbf{z}|^2 \right\rangle\right]^{1/2} = \left[A^2 + (\sigma B)^2\right]^{1/2} , \tag{6.69}$$

where A is some complicated function of the model errors, and the quantity B is determined by the statistical properties of the noise. We stress that neither A nor B depend on the noise level σ. The dependence (6.69) is confirmed by numerical experiments [33].

We finish this section by discussing the connection between the obtained results and their possible applications. On of them is the identification of chaotic sources. Suppose that the only available information about some non-linear system is the preliminarily measured time dependence $\mathbf{x}(t)$. At some later time moment we get a new time dependence $\mathbf{x}'(t)$, and we want to know whether both signals come from the same system. In order to answer this question we should construct a model approximately reproducing the series $\mathbf{x}(t)$ and try to synchronize it with an analogous model for $\mathbf{x}'(t)$. If synchronization is possible, then there is a high probability that $\mathbf{x}(t)$ and $\mathbf{x}'(t)$ have a common source. Noise and errors in model construction will obviously affect the synchronization quality. Therefore, if we want to use synchronization as a system identification method we must know to how estimate the influence of noise and model errors.

An interesting application of the obtained results is connected with the realization of the so-called non-destructive control methods. Let us consider some device to be placed in a difficult-to-access work space (for example, a sensor in a nuclear reactor). Before use, the device is subjected to a calibrating signal and the corresponding time dependence is recorded. After that, a device model is constructed and one determines the degree of synchronization between the model and the recorded time dependence. After some time we again act on the device with a calibrating signal and record the new time series. Then we try to synchronize that series with the old model. As the device was under the strong influence of the environment, its dynamics changed. This will lead in turn to changes in the degree of synchronization. Observing these changes, we can make conclusions about the need to repair or replace the device. In order to make correct conclusions one needs the above quantitative estimates of the influence of dynamics changes on synchronization.

6.9 Synchronization of Chaotic Systems and Transmission of Information

The possibility of synchronizing chaotic systems opens wide possibilities for the application of chaos for information transmission purposes. Any new information transmission scheme must satisfy some fairly evident requirements:

- Competitiveness (realization simplicity and at least partial superiority over existing analogues).
- High performance.
- Reliability and stability with respect to noise of different types: self-noise and external noise.
- Guarantee of a given security level.
- Simultaneous access to multiple users.

Of course, originally every new scheme is oriented to achieve success in one of the above points, but then one should show that the proposed scheme to some extent satisfies all other requirements. We will choose for the central requirement the achievement of a security level which exceeds available analogues. Our choice is dictated by the fact that it is connected with the use of new physics—the synchronization of chaotic systems.

Codes appeared in antiquity. Caesar had his own secret alphabet. In Middle Ages Bacon, Viet, and Cardano worked at inventing secret ciphers. Edgar Allen Poe and Sir Arthur Conan Doyle did a great deal to popularize the deciphering process. During the Second World War, the unraveling of the enemy's ciphers played an important role in the outcome of particular episodes. Finally, Shannon demonstrated that it is possible to construct a cryptogram which cannot be deciphered if the method of its composition is unknown.

Random variables have many advantages in the transmission of secure information. First, a random signal can be unrecognizable on a background of natural noise. Second, even if the signal could be detected, the unpredictability of its variation will not furnish any direct clues to the information contained in it. Also, a broadband chaotic signal is harder to jam. However, the legal recipient should be able to decode the information. In principle, a secret communication system of this type could use two identical chaotic oscillators: one as a transmitter and another as a receiver. The chaotic oscillations of the transmitter would be used for coding and those of the receiver for decoding. The idea is simple but difficult to realize, because any small difference in the initial conditions and parameters of the chaotic system will lead to totally different output signals.

Different ways to overcome this principal difficulty were investigated. At the last it appeared that the most likely direction was the chaotic synchronization which has been considered in the present chapter. Using synchronized chaos for secret communications was the topic of a series of papers published in the 1990s (see [34–36]).

The principal scheme for the transmission of coded information based on the chaos synchronization effect is presented in Fig. 6.16. The transmitter adds to the informational (for example, sound) signal i the chaotic x-component, generated by the driving system. The addition should be understood in a broad sense. This includes: (1) the transmission of the proper sum of chaotic $x(t)$ and informational $i(t)$ signals; (2) the transmission of the product $x(t)i(t)$; and (3) the transmission of the combination $x(t)[1 + i(t)]$. The sum signal is detected by the receiver. The

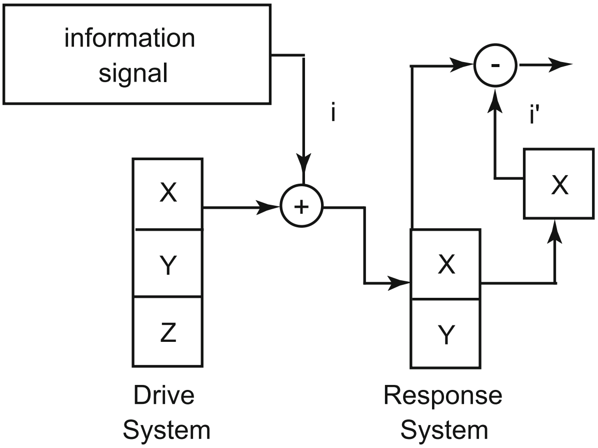


Fig. 6.16 Principal scheme of the coded information transmission, based on the chaos synchronization effect

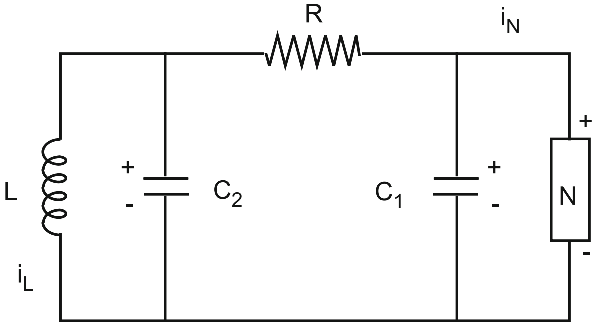


Fig. 6.17 Analog realization of the Van der Pol–Duffing oscillator model

synchronized signal, generated in the receiver, is subtracted from the received message. The difference approximately equals to the coded informative signal.

It is evident that the principal ability of such a scheme to work is based on roughness of the synchronization process: addition of a weak informative signal to a chaotic one does not affect its ability to synchronize the receiver and the transmitter. Let us analyze in more detail the function of this scheme [36] on the example of a physically interesting model—the Van der Pol–Duffing oscillator. Its analog realization is presented in Fig. 6.17. Recall that under an analog setup we understand a system where every instantaneous value of the quantity entering into the input relations corresponds to an instantaneous value of another quantity, often different from the original one in its physical nature. Every elementary mathematical operation on the machine's quantities corresponds to some physical law. This law establishes the dependence between the physical quantities on the input and output of the deciding setup: for example, Ohm's law—to division, Kirchhoff's law—to addition, and the Lorenz force—to vector product.

We introduce a cubically non-linear element N into the chain (Fig. 6.17), which gives the following relation:

$$I(V) = aV + bV^3; \quad a < 0,\ b > 0 \tag{6.70}$$

between the current I and applied voltage V. Applying Kirchhoff's laws to different parts of the chain and rescaling the the variables, we obtain the following set of dynamical equations:

$$\begin{aligned} \dot{x} &= -\gamma(x^3 - \alpha x - y)\,, \\ \dot{y} &= x - y - z\,, \\ \dot{z} &= \beta y\,. \end{aligned} \tag{6.71}$$

Here x, y, z are rescaled voltage on C_1, C_2 and current through L, respectively; α, β, γ are rescaled chain parameters. Numerical simulation of Eqs. (6.71) with fixed α, γ demonstrates the transition to chaos by the scenario of period doubling with the decrease of β. In particular, for $\gamma = 100$, $\alpha = 0.35$, $\beta = 300$ a chaotic attractor is observed in the phase space. We will consider the system (6.71) as the driving one, and the coordinate x as the full replacement variable. Then, the equations of motion for the driven system (its coordinates are stroked)

$$\begin{aligned} x' &= x\,, \\ \dot{y}' &= x - y' - z'\,, \\ \dot{z}' &= \beta y'\,. \end{aligned} \tag{6.72}$$

Let us now show that the subsystem (6.72), which we have chosen for the driven one, is globally stable. For this we use the Lyapunov function method. Denoting $y - y' = y^*$, $z - z' = z^*$, from (6.71), (6.72) we get

$$\begin{pmatrix} \dot{y}^* \\ \dot{z}^* \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ \beta & 0 \end{pmatrix} \begin{pmatrix} y^* \\ z^* \end{pmatrix}\,. \tag{6.73}$$

For the Lyapunov function we take the following:

$$L = \frac{1}{2}\left[(\beta y^* + z^*)^2 + \beta y^{*2} + (1+\beta) z^{*2}\right]. \tag{6.74}$$

Using the equations of motion (6.73), we find

$$\begin{aligned} \dot{L} &= (\beta y^* + z^*)\,(\beta \dot{y}^* + \dot{z}^*) + \beta y^* \dot{y}^* + (1+\beta)\, z^* \dot{z}^* \\ &= -\beta\left(y^{*2} + z^{*2}\right) \leqslant 0, \quad (\beta > 0)\,. \end{aligned} \tag{6.75}$$

Therefore the subsystem (6.72) is globally stable, i.e., for $t \to \infty$

$$\left|y - y'\right| \to 0, \quad \left|z - z'\right| \to 0\,. \tag{6.76}$$

There is an interesting possibility to obtain a cascade of the driven subsystems [37]. Suppose that the driving system is represented by (6.72), and the first driven system is represented in terms of variables y', z', excited by $x(t)$. In addition, we can imagine that we have a system containing the variable x'', excited by the variable y'. The total cascade of the systems looks like the following.

The driving system

$$\begin{aligned} \dot{x} &= -\gamma(x^3 - \alpha x - y)\,, \\ \dot{y} &= x - y - z\,, \\ \dot{z} &= \beta y\,. \end{aligned} \tag{6.77}$$

The first driven system

$$\begin{aligned} \dot{y}' &= x - y' - z'\,, \\ \dot{z}' &= \beta y'\,. \end{aligned} \tag{6.78}$$

The second driven system

$$\dot{x}'' = -\gamma\left[\left(x''\right)^3 - \alpha\left(x''\right) - y'\right]. \tag{6.79}$$

If all the systems are synchronized, the signal $x''(t)$ is identical to the driving signal $x(t)$.

Let us now focus our attention on using the constructed cascade system for the transmission of secret information. In accordance with the above principal scheme we use the $x(t)$ signal as the one of mask noise, and $s(t)$ as the information medium. Let the receiver detect the transmitted signal $r(t) = x(t) + s(t)$. As an analysis of the system of equations (6.77) shows (6.78), (6.79) [36], if the power level of the informative signal is considerably lower than the noise medium power level $|x(t)| >> |s(t)|$, then $|x(t) - x''(t)| << |s(t)|$. This, in turn, means that the signal $s^{(1)}$ obtained as the result of the operation

$$s^{(1)} = r(t) - x''(t) = x(t) + s(t) - x''(t) \approx s(t) \tag{6.80}$$

will be close to the initial informative signal $s(t)$. Authors of [36] numerically solved the system of equations (6.77), ((6.78), (6.79) with parameters $\alpha = 0.35$, $\beta = 300$, $\gamma = 100$). The information medium signal $s(t)$ was chosen in the three following forms:

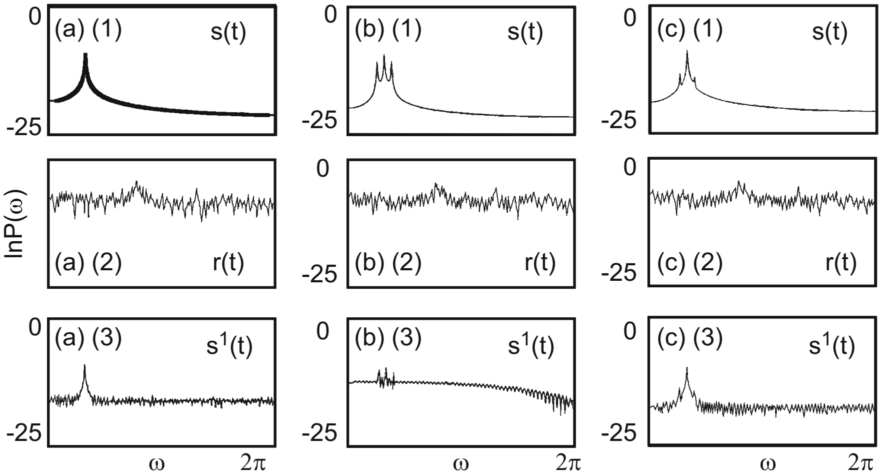


Fig. 6.18 Power spectra for the informative signal $s(t)$, transmitted signal $r(t)$, and reconstructed signal $s^{(1)}(t)$ for monochromatic (**a**), amplitude-modulated (**b**), and frequency-modulated (**c**) signals $s(t)$ [36].

Monochromatic signal: $s(t) = F\sin(\omega t), F = 0.02, \omega = 1.$

Amplitude-modulated signal: $s(t) = F\sin(\omega t)[1 + f\sin(\Omega t)], F = 0.02, \omega = 1, f = 1, \Omega = 0.2.$

Frequency-modulated signal: $s(t) = F\sin[\omega t + f\sin(\Omega t)], F = 0.02, \omega = 1, f = 0.2, \Omega = 0.2.$

The informative signal $s^{(1)}(t)$ was reconstructed from numerical calculation results according to (6.80). Figure 6.18 presents power spectra for the informational signal $s(t)$, transmitted signal $r(t) = x(t) + s(t)$, and reconstructed signal $s^{(1)}(t)$ for all three listed cases. If the informational signal power level is considerably lower than for the chaotic medium, the frequency components of the informational signal in the transmission are not detectable, at least visually. The spectrum quality of the reconstructed signal is comparable to the received one.

References

1. Andronov, A., Vitt, A., Khaikin, S.: Theory of Oscillations. Pergamon Press, Oxford (1966)
2. Blekhman, I.I.: Synchronization in Science and Technology. ASME Press, New York (1988)
3. Lindsey, W.: Synchronization Systems in Communications and Control. Prentice-Hall, New Jersey (1972)
4. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization. In: A Universal Concept in Nonlinear Sciences. Cambridge University Press, Cambridge (2001)
5. Fradkov, A.L.: Cybernetical Physics: From Control of Chaos to Quantum Control. Springer, Berlin (2007)

6. Haken, H.: Brain Dynamics: An Introduction to Models and Simulations. Springer Series in Synergetics. Springer, Berlin/Heidelberg (2008)
7. Haken, H.: Information and Self Organisation: A Macroscopic Approach to Complex Systems. Springer Series in Synergetics. Springer, Berlin/Heidelberg (2006)
8. Yamada, T., Fujisaka, H.: Prog. Theor. Phys. **70**, 1240–1248 (1983)
9. Aframovich, V.S., Verichev, N.N., Rabinovich, M.I.: Izv. Vuzov, Radiofizika **29**, 795–803 (1986)
10. Kolmogorov, A.N.: IEEE Trans. Inf. Theory **IT-14**, 662–664 (1968)
11. Yamada, T., Fujisaka, H.: Prog. Theor. Phys. **72**, 885–894 (1984)
12. Pecora, L.M., Carroll, T.L., Johnson, G.A., Mar, D.J.: Chaos **7**(4), 520–543 (1997)
13. Pecora, L.M., Carroll, T.L.: Phys. Rev. A **44**, 2374–2383 (1991)
14. Guemez, J., Matas, M.A., et al.: Phys. Rev. E **52**, R2145–R2148 (1995)
15. Amritkar, R., Gupte, N.: Phys. Rev. E **47**, 3889–3895 (1993)
16. Brown, R., Kocarev, L.: arXiv:chao-din/9811013 (1999)
17. Heagy, J.F., Carroll, T.L.: Chaos **4**, 385–390 (1994)
18. Chirikov, B.V.: Phys. Rep. **52**, 265–379 (1979)
19. Lai, Y.-C., Grebogi, C.: Phys. Rev. E **47**, 2357–2360 (1993)
20. Parlitz, U.: Phys. Rev. Lett. **76**, 1232–1235 (1996)
21. Verhulst, F.: Nonlinear Differential Equations and Dynamical Systems. Springer, Berlin (1996)
22. Toral, R., Mirasso, C., Hernandez-Garsia, E., Piro, O.: Chaos **11**, 665–673 (2001)
23. Matsumoto, K., Tsuda, I.: J. Stat. Phys. **31**, 87–106 (1983)
24. Yu, L., Ott, E., Chen, Q.: Phys. Rev. Lett. **65**, 2935–2938 (1990)
25. Fahy, S., Hamman, D.: Phys. Rev. Lett. **69**, 761–764 (1992)
26. Maritan, A., Banavar, J.: Phys. Rev. Lett. **72**, 1451–1454 (1994)
27. Pikovsky, A.: Phys. Rev. Lett. **73**, 2931–2932 (1994)
28. Herzel, H., Freund, J.: Phys. Rev. E **52**, 3238–3241 (1995)
29. Lai, C.H., Zhou, C.: Europhys. Lett. **43**, 376–380 (1988)
30. Kosarev, L., Tazev, Z., Stojanovski, T., Parlitz, U.: Chaos **7**(4), 635–643 (1997)
31. Kosarev, L., Tazev, Z., Parlitz, U.: Phys. Rev. Lett. **79**, 51–54 (1997)
32. DeWitt, A., Dewel, G., Borckmans, P.: Phys. Rev. E **48**, R4191–R4194 (1993)
33. Brown, R., Rulkov, N., Tufillaro, N.: Phys. Rev. E **50**, 4488–4508 (1994)
34. Cuomo, K.M., Oppenheim, A.V.: Phys. Rev. Lett. **71**, 65–68 (1993)
35. Kosarev, L., Halle, K.S., Eckert, K., et al.: Int. J. Bifurcation Chaos **2**, 709–713 (1992)
36. Murali, K., Lakshmanan, M.: Phys. Rev. E **48**, R1624–R1626 (1993)
37. Carroll, T.L., Pecora, L.M.: Physica D **67**, 126–140 (1993)

Chapter 7
Stochastic Resonance

There are well-known and strictly regulated algorithms for the solution of linear problems. The physical meaning of the solution for any linear problem is clear on an intuitive level. The particularity of the linear system does not play an essential role. However, if we want to deal with real situations, we must take into account two new elements—non-linearity and noise. Non-linearity leads to incredible complications in solving technique. The combination of non-linearity with noise complicates the situation even more. In attempts to predict the behavior of such systems, the most refined intuition fails. The stochastic resonance effect represents an example of the paradoxical behavior of non-linear systems under influence of noise. The term "stochastic resonance" unites a group of phenomena for which the growth of disorder (noise amplitude) upon input into a non-linear system leads under certain conditions to an increase of order on the output. Quantitatively, the effect manifests in the fact that such integral system characteristics as gain constant, noise-to-signal ratio have a clearly marked maximum at some optimal noise level. At the same time, the system entropy reaches a minimum, giving evidence of noise-induced order growth.

7.1 Qualitative Description of the Effect

The concept of stochastic resonance was first proposed in papers [1, 2] and independently in [3, 4]. The authors of the works studied the problem of the alternation of ice ages on Earth. Analysis showed that ice ages alternate with the period of an order of 100,000 years. This result seemed curious. The only quantity of this time scale in the dynamics of the Earth is the period of the oscillations of the Earth's orbital eccentricity, connected to planetary gravitational perturbations. However, changes in the flow of solar energy coming to the Earth, connected to those oscillations, equals only about 0.1 % of it. At first glance it seemed that this

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_7

was absolutely insufficient for such radical climate changes, and one should seek principally new mechanisms to amplify the weak periodic oscillations. One of the possible solutions to the problem was found in accounting for the joint action of the two mechanisms: the simultaneous action of periodic perturbation (oscillations of the Earth's orbital eccentricity) and noise (climatic oscillations) at certain conditions (stochastic resonance) led to a qualitative explanation of the observed climatic changes.

The first works on stochastic resonance necessarily included the following fundamental elements: non-linearity, bistability, external periodic signal and noise. Later, it became clear that the three latter elements are not necessary attributes of the effect. So it appeared that there was no need to consider only bistable systems and stochastic resonance can be presented as purely a threshold effect. It is also possible to set up the problem in the absence of an external periodic signal. In many non-linear systems coherent motion is stimulated by the internal dynamics of the non-linear system rather than by external force. However paradoxical it may be, it is also possible to set up the stochastic resonance problem without external noise [5]. So for chaotic systems noise can be generated by their own chaotic dynamics, and the role of noise intensity is played by the system parameters, determining the measure of chaoticity. There are also less radical deviations from the initial formulation of the problem. For example, the typical signals apprehended by biological systems have a complex spectral composition, and the monochromatic signal seems an excessive idealization for them. Colored noise, i.e. noise with a finite correlation time and limited spectrum, can be more appropriate than white noise for the reality being studied.

The physical mechanism of the stochastic resonance effect is demonstrated in the simplest way on its canonical model: a Brownian particle in a symmetric bistable potential

$$V(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4; \quad (a, b > 0)\,, \tag{7.1}$$

subjected to a weak periodic signal $F \cos \omega t$. The potential minima are situated in the points $x_{\min} = \pm\sqrt{a/b}$ and separated by a potential barrier at $x = 0$ of height $\Delta V = a^2/4b$. As is well known, the fluctuating forces cause rare (at low temperature) random transitions over the potential barrier. The rate of the transitions is determined by the famed Kramers formula [6]

$$W_k \sim e^{-\Delta V/D}\,, \tag{7.2}$$

where D is the intensity of the fluctuations, and the pre-exponential factor depends on the potential geometry. Now suppose that the particle is subjected to an additional deterministic force—a weak periodic perturbation of frequency ω. The term "weak" means that the force itself cannot result in the barrier being overcome. In the presence of periodic perturbation, the initial potential parity will be broken (see Fig. 7.1), which will lead to dependence of the frequency of the transition rates W_k

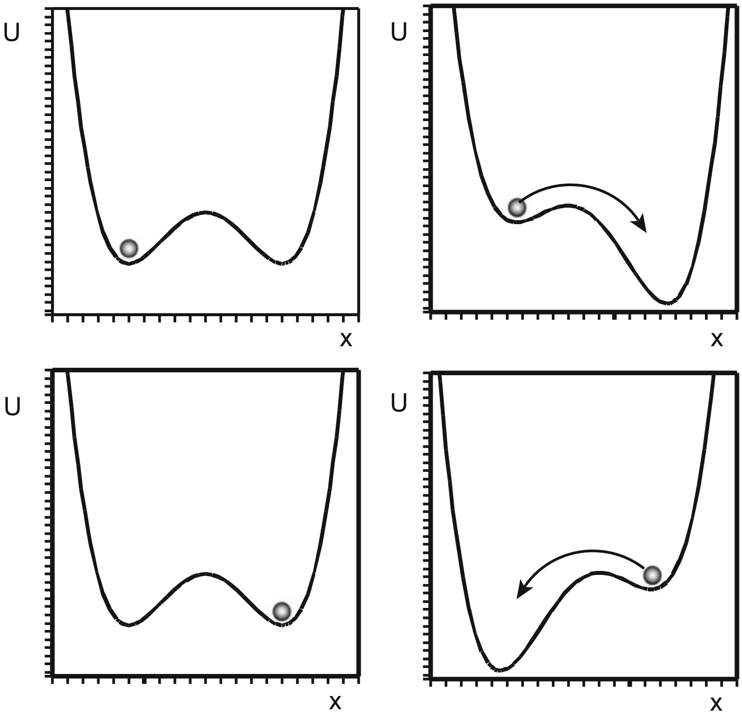


Fig. 7.1 Stochastic resonance in a bistable symmetric well with periodic perturbation

on time. Varying the intensity of the noise, it is possible to achieve a situation when the Kramers rate will be close to the frequency of the potential barrier modulation. This can be achieved provided the condition

$$W_k^{-1}(D) \equiv t_k(D) \approx \frac{\pi}{\omega}\,, \tag{7.3}$$

is fulfilled. Analogous considerations can be made also for the more general case of two meta-stable states 1, 2, where the height of the barrier between them changes due to the periodic perturbation of frequency $\omega = 2\pi/T$. Suppose that the particle performs transitions between these states in average times $T^+(1 \to 2)$ and $T^-(2 \to 1)$. It is natural to assume that the system is optimally adjusted to external perturbation under the condition

$$\frac{2\pi}{\omega} \approx T^+ + T^-\,. \tag{7.4}$$

In the symmetric case $T^+ = T^- = t_K$ we return to the relation (7.3).

For negligibly low noise levels ($D \to 0$) the Kramers transition rate practically equals to zero and no coherence is present. For very high noise levels the coherence is also absent as system response becomes practically random. Between these two extreme cases there is an optimal intensity defined by the relation (7.3), which optimizes coherence. For this situation it seems natural to use the term “resonance,” although it is evident that this phenomenon differs considerably from familiar deterministic resonances, but statistically the term has a well-defined meaning and in the last few years it has become quite widely accepted. The effect consists of the possibility of adjusting a stochastic system by variation in noise intensity to achieve a regime of maximal amplification of the modulation signal. Stochastic resonance can be realized in any non-linear system where several characteristic time scales exist, one of which can be controlled with the help of noise [7, 8].

Systems demonstrating stochastic resonance are in some sense intermediate between regular and irregular: they are described by a random process where jumps do not obey a deterministic law, but nevertheless in a resonance regime they have some degree of regularity. In particular, the coordinate correlation function does not tend to zero over long times.

7.2 The Interaction Between the Particle and Its Surrounding Environment: Langevin’s Equation

For a quantitative description of the stochastic resonance effect, it is necessary to obtain a dynamical equation describing the interaction between a Brownian particle and its environment [9–12]. The equation must contain only the coordinates of the particle but to account for the interaction with the environment phenomenologically.

Suppose a Brownian particle of mass M moves in the potential $V(R)$ (for example, in a bistable potential, considered in the previous section). The Lagrange function L for such motion reads

$$L_0(R, \dot{R}) = \frac{1}{2}M\dot{R}^2 - V(R)\,. \tag{7.5}$$

For the thermal reservoir (environment) we take the simplest model considering it to be a set of a large number of non-interacting harmonic oscillators with coordinates q_i, masses m_i, and frequencies ω_i

$$L_{hb}(q_i, \dot{q}_i) = \sum_i \frac{m_i}{2}(\dot{q}_i^2 - \omega_i^2 q_i^2)\,. \tag{7.6}$$

The interaction between the particle and the thermal reservoir we assume to be separable and linear on the oscillatory coordinates. As a result, we obtain for the full Lagrangian

$$L(R,\dot{R};q_i,\dot{q}_i) = L_0(R,\dot{R}) + L_{hb}(q_i,\dot{q}_i) + \sum_i f_i(R)q_i\,. \tag{7.7}$$

Corresponding Euler–Lagrange equations take the form

$$\begin{aligned} M\ddot{R} &= -\frac{dV(R)}{dR} + \sum_i q_i \frac{df_i(R)}{dR}\,, \\ m_i\ddot{q}_i &= -m_i\omega_i^2 q_i + f_i(R)\,. \end{aligned} \tag{7.8}$$

In order to obtain the equation for the coordinates R it is necessary to exclude from those equations the coordinates q_i of the thermal reservoir. The latter equal to

$$q_i(t) = q_i^0(t) + \int\limits_0^t ds \frac{f_i(R(s))}{m_i\omega_i} \sin\omega_i(t-s)\,, \tag{7.9}$$

where $q_i^0(t)$ are the solutions of the free equation for the reservoir coordinates,

$$q_i^0(t) = q_{i0}\cos\omega_i(t-t_0) + (p_{i0}/m_i\omega_i)\sin\omega_i(t-t_0)\,, \tag{7.10}$$

where q_{i0} and p_{i0} are coordinates and momenta of the oscillators in the initial time moment t_0. Substituting (7.10) in the former of Eq. (7.8), we obtain the result

$$M\ddot{R} = \tilde{F}(R) + F_{\text{frict}}(R,\dot{R}) + F_L(R,t)\,. \tag{7.11}$$

Here we introduced the renormalized potential force $\tilde{F}(R)$, the friction force $F_{\text{frict}}(R,\dot{R})$, which depends on coordinates and velocities, and the random (Langevin) force $F_L(R,t)$. Those forces equal to

$$\tilde{F}(R) = -\frac{d\tilde{V}(R)}{dR}; \quad \tilde{V}(R) = V(R) - \sum_i \frac{1}{2m_i\omega_i^2}\left[f_i(R)\right]^2\,, \tag{7.12}$$

$$F_{\text{frict}}(R,\dot{R}) = -\sum_i \frac{1}{m_i\omega_i^2}\int\limits_{t_0}^t ds \frac{df_i(R(t))}{dR}\cos\omega_i(t-s)\frac{df_i(R(s))}{dR}\dot{R}(S)\,, \tag{7.13}$$

$$F_L(R,t) = \sum_i q_i^0(t)\frac{df_i(R)}{dR}\,. \tag{7.14}$$

The friction force (7.13) represents a delayed force of the form

$$F_{\text{frict}}(R, \dot{R}) = -\int_{t_0}^{t} ds\gamma(t, s)\dot{R}(s)\,, \tag{7.15}$$

where the integral kernel (assuming the equality of all masses $m_i = m$ and all form-factors $f_i(R) = f(R)$) equals to

$$\gamma(t, s) = \frac{df(R(t))}{dR}\frac{df(R(s))}{ds}\sum_i \frac{1}{m\omega_i^2}\cos\omega_i(t-s)\,. \tag{7.16}$$

The sum of the oscillating terms in (7.16) approximately equals to zero everywhere except the region $|t-s| \leqslant \varepsilon$. Therefore we can approximately write

$$\sum_i \frac{1}{m\omega_i^2}\cos\omega_i(t-s) \approx 2\gamma_0\delta_\varepsilon(t-s)\,, \tag{7.17}$$

where $\delta_\varepsilon(t-s)$ is the so-called smeared δ-function

$$\delta_\varepsilon(t-s) = \begin{cases} 1/\varepsilon, & \text{if } -\varepsilon/2 \le (t-s) \le \varepsilon/2, \\ 0, & \text{otherwise} \end{cases}\,. \tag{7.18}$$

Factor 2 is introduced in (7.17) for further convenience. From (7.17) it follows that

$$2\gamma_0 = \int_{-\infty}^{+\infty} dt\frac{1}{m\omega_i^2}\cos\omega_i t\,. \tag{7.19}$$

Using (7.17) for the kernel $\gamma(t, s)$ we get the expression

$$\gamma(t, s) = 2\gamma(R)\delta_\varepsilon(t-s); \quad \gamma(R) = \gamma_0\left[\frac{df(R)}{dR}\right]^2\,. \tag{7.20}$$

We made the assumption that $R(s) = R(t)$ for $|s-t| \leqslant \varepsilon$. Substituting (7.20) into (7.15) we finally get that, given the above assumptions, the friction force is a local function of the coordinates

$$F_{\text{frict}} = -\gamma(R)\dot{R}\,. \tag{7.21}$$

Now we move on to the transformation of the expression (7.14) for random force.

In assuming that the form factors are equal for all oscillators we get

$$F_L(R,t) = \frac{df(R)}{dR}\xi(t)\,, \tag{7.22}$$

where

$$\xi(t) = \sum_i q_i^0(t)\,. \tag{7.23}$$

The initial conditions for the oscillators representing the thermal reservoir q_{i0}, p_{i0} should be naturally considered as independent random variables with the following statistical properties

$$\begin{gathered} \langle q_{i0}\rangle = 0;\ \langle p_{i0}\rangle = 0\,, \\ \left\langle q_{i0}q_{j0}\right\rangle = \delta_{ij}\left\langle q_{i0}^2\right\rangle;\ \left\langle p_{i0}p_{j0}\right\rangle = \delta_{ij}\left\langle p_{i0}^2\right\rangle;\ \left\langle q_{i0}p_{j0}\right\rangle = 0\,. \end{gathered} \tag{7.24}$$

Using (7.10) and (7.24) for the random variable $\xi(t)$ determining the Langevin force, we get

$$\langle\xi(t)\rangle = 0; \quad \left\langle\xi(t)\xi(t')\right\rangle \approx \sum_i \frac{\langle\varepsilon_{i0}\rangle}{m\omega_i^2}\cos\omega_i(t-t')\,, \tag{7.25}$$

where $\langle\varepsilon_{i0}\rangle$ is the average energy of the ith oscillator

$$\langle\varepsilon_{i0}\rangle = \frac{p_{i0}^2}{2m_i} + \frac{1}{2}m_i\omega_i^2\left\langle q_{i0}^2\right\rangle\,. \tag{7.26}$$

If the thermal reservoir is in thermal equilibrium at temperature T, then

$$\langle\varepsilon_{i0}\rangle = \frac{1}{2}k_BT \tag{7.27}$$

and

$$\left\langle\xi(t)\xi(t')\right\rangle \approx \frac{1}{2}k_BT\sum_i \frac{1}{m\omega_i^2}\cos\omega_i(t-t')\,. \tag{7.28}$$

Taking into account (7.17), the sum in the right-hand side of (7.28) can be replaced by $2\gamma_0\delta_\varepsilon(t-t')$ and for the correlation function of the quantity $\xi(t)$ we get

$$\left\langle\xi(t)\xi(t')\right\rangle = \gamma_0k_BT\delta_\varepsilon(t-t')\,. \tag{7.29}$$

It is convenient to introduce the normalized random variable $\Gamma(t)$,

$$\Gamma(t) = \left(1/\sqrt{\gamma_0 k_B T}\right)\xi(t)\,, \tag{7.30}$$

which, according to the central limit theorem, has a Gaussian distribution, and its mean value and correlation function equal to:

$$\langle \Gamma(t)\rangle = 0; \quad \left\langle \Gamma(t)\Gamma(t')\right\rangle = \delta_\varepsilon(t-t')\,. \tag{7.31}$$

Substituting (7.30) into (7.22) we get the resulting expression for the Langevin force

$$F_L(R,t) = \sqrt{D(R)}\,\Gamma(t); \quad D(R) = \left[\frac{df(R)}{dR}\right]^2 \gamma_0 k_B T\,. \tag{7.32}$$

Comparing (7.20) with (7.32) we find that the intensity of the Langevin force is connected to the friction coefficient and the reservoir temperature by the relation known as a particular case of the fluctuation-dissipation theorem [13–15]

$$D(R) = \gamma(R) k_B T\,. \tag{7.33}$$

The origin of this connection is a consequence of the fact that both the friction force and the Langevin force are generated by the same interaction of the Brownian particle with the thermal reservoir. Thus, we have shown that Brownian particle dynamics can be phenomenologically described by a stochastic differential equation (equation with random force)

$$M\ddot{R} = F(R) - \gamma(R)\dot{R} + F\cos\omega t + \sqrt{D(R)}\,\Gamma(t)\,. \tag{7.34}$$

In Eq. (7.34) we included the external periodic force, which is independent of interaction with the thermal reservoir. An equation of the form (7.34) (as well as the method of dynamics description by the direct probabilistic method) was first proposed by Langevin [16] and is named after him. Let us consider the Langevin equation (7.34) from a more general point of view. It represents a particular case of a stochastic differential equation of the form

$$\dot{x}(t) = G\left(x(t), t, \xi(t)\right)\,, \tag{7.35}$$

where the variable $\xi(t)$ describes some stochastic (random) process. It can be presented as a family of functions $\xi_u(t)$ depending on the results u of some experiment S. Therefore, the stochastic differential equation (7.35) is a family of ordinary differential equations, differing for every result (realization) of u

$$\dot{x}_u(t) = G\left(x_u(t), t, \xi_u(t)\right)\,. \tag{7.36}$$

The family of solutions of those equations $x_u(t)$ for different u represents the stochastic process $x(t)$. We can say that each realization $\xi_u(t)$ of the random process $\xi(t)$ corresponds to a realization $x_u(t)$ of the random process $x(t)$. Thus the solution $x(t)$ becomes a functional of the process $\xi(t)$.

7.3 The Two-State Model

We now turn to an investigation of stochastic resonance in the simplest analytically solvable model—the two-state model [17]. This model represents a discrete analogue of a continuous bistable system where the state variable takes only two discrete values $x_\pm$ with probabilities $n_\pm$. The rates of transition from one state to another $W_\pm$ are assumed to be known. It is obvious that they can be obtained only from models accounting for the internal system dynamics, which itself determines the rates of transitions between local minima. If the distribution function $P(x)$ of the continuous analogue of the two-state model is known, then

$$n_- = 1 - n_+ = \int_{-\infty}^{x_{\max}} P(x)dx\,, \tag{7.37}$$

where $x_{\max}$ is the position of the maximum in the two-well potential. The master equation can be written in the form

$$\frac{dn_+}{dt} = -\frac{dn_-}{dt} = W_-(t)n_- - W_+(t)n_+ = W_-(t) - [W_-(t) + W_+(t)]\,n_+\,. \tag{7.38}$$

Transition probabilities depend on time due to the presence of periodic perturbation. The master equation (7.38) is applicable only in the so-called adiabatic limit when the perturbation period is much longer than the characteristic relaxation times. For the two-well problem, under “relaxation time” we understand the time of thermal equilibrium to be established in a separate well.

The solution of the linear differential equation with periodic coefficients (7.38) has the form

$$\begin{aligned} n_+(t) &= g(t)\left[n_+(t_0) + \int_{t_0}^{t} W_+(\tau)g^{-1}(\tau)d\tau\right]\,, \\ g(t) &= \exp\left(-\int_{t_0}^{t}[W_+(\tau) + W_-(\tau)]\,d\tau\right)\,, \end{aligned} \tag{7.39}$$

where $n_+(t_0)$ is an as yet undefined initial condition.

Let us now assume that the transition rates have the form

$$W_\pm(t) = f(\mu \pm \eta_0 \cos \omega t)\,, \tag{7.40}$$

where η_0 is the dimensionless parameter determining the perturbation intensity, and $f(\mu, \eta_0 = 0)$ turns into the Kramers rate (7.2) and no longer depends on time. In other words, we assume that the effect of perturbation reduces to only modulations of the height of the potential barrier determining the Kramers rate. In the case of weak periodic perturbation we can decompose the transition rates (7.40) over the small parameter $\eta_0 \cos \omega t$:

$$W_{\pm}(t) = \frac{1}{2}\left(\alpha_0 \mp \alpha_1 \eta_0 \cos \omega t + \alpha_2 \eta_0^2 \cos^2 \omega t \mp \cdots\right) ,$$
$$W_{+}(t) + W_{-}(t) = \alpha_0 + \alpha_2 \eta_0^2 \cos^2 \omega t , \tag{7.41}$$

where

$$\frac{1}{2}\alpha_0 = f(\mu), \quad \frac{1}{2}\alpha_n = \frac{(-1)^n}{n!}\frac{d^n f}{d\eta^n} . \tag{7.42}$$

The expression (7.39) now can be integrated and in the first order in the small parameter $\eta_0 \cos \omega t$ we get

$$n_{+}\left(t|\, x_0, t_0\right) = \frac{1}{2}\left[e^{-\alpha_0(t-t_0)}\left[2\delta_{x_0 c} - 1 - \frac{\alpha_1 \eta_0 \cos(\omega t_0 - \varphi)}{(\alpha_0^2 + \omega^2)^{1/2}}\right] + 1 + \frac{\alpha_1 \eta_0 \cos(\omega t - \varphi)}{(\alpha_0^2 + \omega^2)^{1/2}}\right] , \tag{7.43}$$

where $\varphi = \arctan(\omega/\alpha_0)$. The Kronecker symbol $\delta_{x_0 c}$ equals unity if in the moment of time t_0 the particle was in a state $x_+ = c$, and equals zero if it was in $x_- = -c$. The quantity $n_+\left(t|\, x_0, t_0\right)$ is the conditional probability of the particle at time moment t to be in the state c, if in the time moment t_0 it was in the state x_0 (c or $-c$).

The obtained probabilities allow us to calculate all the statistical characteristics of the process. In particular, reduction to the "continuous" bistable system is done with the help of the relations

$$P\left(x, t|\, x_0, t_0\right) = n_{+}\left(t|\, x_0, t_0\right)\delta(x-c) + n_{-}\left(t|\, x_0, t_0\right)\delta(x+c) . \tag{7.44}$$

For example, for the average value of coordinate x we get

$$\langle x(t)|\, x_0, t_0\rangle = \int x P\left(x, t|\, x_0, t_0\right) dx = c n_{+}\left(t|\, x_0, t_0\right) - c n_{-}\left(t|\, x_0, t_0\right) . \tag{7.45}$$

In the absence of periodic modulation in the state of equilibrium $n_+ = n_- = 1/2$ that average value equals zero.

Analogously, we can find the autocorrelation function

$$\langle x(t+\tau)x(t)|\,x_0,t_0\rangle = \int xyP\left(x,t+\tau|\,y,t\right)P\left(y,t|\,x_0,t_0\right)dxdy\,, \tag{7.46}$$

which we will further need to find the power spectrum. Using the expressions (7.44) for conditional probabilities, we get

$$\begin{aligned}
&\langle x(t+\tau)x(t)|\,x_0,t_0\rangle \\
&\quad = c^2 n_+\left(t+\tau|\,c,t\right)n_+\left(t|\,x_0,t_0\right) - c^2 n_+\left(t+\tau|-c,t\right)n_-\left(t|\,x_0,t_0\right) \\
&\qquad - c^2 n_-\left(t+\tau|\,c,t\right)n_+\left(t|\,x_0,t_0\right) + c^2 n_-\left(t+\tau|-c,t\right)n_-\left(t|\,x_0,t_0\right) \\
&\quad = c^2\{2[n_+(t+\tau|\,c,t) + n_+(t+\tau|-c,t) - 1]n_+(t|\,x_0,t_0) \\
&\qquad - 2n_+(t+\tau|-c,t) + 1\}\,.
\end{aligned} \tag{7.47}$$

The asymptotic limit of the autocorrelation function at $t_0 \to \infty$ presents us with some interest. In that limit, using (7.43), we get

$$\begin{aligned}
\langle x(t)x(t+\tau)\rangle &= \lim_{t_0\to\infty}\langle x(t)x(t+\tau)|x_0t_0\rangle \\
&= c^2 e^{-\alpha_0|\tau|}\left[1 - \frac{\alpha_1^2\eta_0^2\cos^2(\omega t-\varphi)}{\alpha_0^2+\omega^2}\right] \\
&\quad + \frac{c^2\alpha_1^2\eta_0^2\left\{\cos\omega\tau + \cos\left[\omega(2t+\tau)+2\varphi\right]\right\}}{2\left(\alpha_0^2+\omega^2\right)}\,.
\end{aligned} \tag{7.48}$$

Due to the presence of periodic perturbation, the autocorrelation function depends not only on the time shift τ, but also periodically on time. To calculate the characteristics of stochastic resonance we should average over the perturbation period. This procedure is equivalent to averaging over an ensemble of random initial phases of perturbation and it corresponds to experimental methods of measuring the statistical characteristics obtained from the correlation functions. Considering t as a random variable, uniformly distributed in the interval $[0, 2\pi]$, we get

$$\begin{aligned}
\langle x(t)x(t+\tau)\rangle_t &= \frac{\omega}{2\pi}\int_0^{2\pi/\omega}\langle x(t)x(t+\tau)\rangle dt \\
&= c^2 e^{-\alpha_0|\tau|}\left[1 - \frac{\alpha_1^2\eta_0^2}{2\left(\alpha_0^2+\omega^2\right)}\right] + \frac{c^2\alpha_1^2\eta_0^2\cos\omega\tau}{2\left(\alpha_0^2+\omega^2\right)}\,.
\end{aligned} \tag{7.49}$$

Let us recall that (see Sect. 3.2) the power spectrum (spectral density) is the Fourier transform of the autocorrelation function [18], averaged over the perturbation period

$$\begin{aligned}\langle S(\Omega)\rangle_t &= \int_{-\infty}^{\infty} \langle\langle x(t)x(t+\tau)\rangle\rangle_t e^{-i\Omega\tau} d\tau \\ &= \left[1 - \frac{\alpha_1^2\eta_0^2}{2\left(\alpha_0^2+\omega^2\right)}\right]\left[\frac{2c^2\alpha_0}{\alpha_0^2+\Omega^2}\right] \\ &\quad + \frac{\pi c^2\alpha_1^2\eta_0^2}{2\left(\alpha_0^2+\omega^2\right)}\left[\delta\left(\Omega-\omega\right)+\delta\left(\Omega+\omega\right)\right] . \end{aligned} \tag{7.50}$$

Further we will use the power spectrum $S(\Omega)$, defined only for positive Ω,

$$\begin{aligned} S(\Omega) &= \langle S(\Omega)\rangle_t + \langle S(-\Omega)\rangle_t \\ &= \left[1 - \frac{\alpha_1^2\eta_0^2}{2\left(\alpha_0^2+\omega^2\right)}\right]\left[\frac{4c^2\alpha_0}{\alpha_0^2+\Omega^2}\right] + \frac{\pi c^2\alpha_1^2\eta_0^2}{\left(\alpha_0^2+\omega^2\right)}\delta(\Omega-\omega) \\ &= S_N(\Omega) + \frac{\pi c^2\alpha_1^2\eta_0^2}{\left(\alpha_0^2+\omega^2\right)}\delta(\Omega-\omega) . \end{aligned} \tag{7.51}$$

The power spectrum naturally divides into two parts: the one describing the periodic component of the output signal on the perturbation frequency (proportional to δ-function) and the noise component $S_N(\Omega)$. The noise spectrum represents a product of only the Lorenz factor $\alpha_0/(\alpha_0^2+\Omega^2)$ and the correction factor, describing the influence of the signal on noise. At low signal amplitudes the correction factor is close to unity. This factor describes the energy pumping from the background of noise to the periodic component. It is interesting to note that the total power on the system output does not depend on the frequency nor on the amplitude of the signal: contributions from the correction factor and from the periodic component exactly compensate each other, as

$$\int_0^{\infty} d\Omega \frac{\alpha_0}{\alpha_0^2+\Omega^2} = \frac{\pi}{2} . \tag{7.52}$$

This exact compensation represents a characteristic feature of the two-state model. It comes from the Persival theorem: the time integral of the signal squared equals the power spectrum integrated over all frequencies. The time integral in the two-state model for any time interval T equals c^2T and does not depend on perturbation frequency or amplitude. Therefore, the power spectrum integrated over all frequencies must also remain constant.

Let us now return to the continuous analogue of the discrete two-state model—the double symmetric well (7.1). In presence of periodic perturbation the potential energy of the system takes the form

$$U(x,t) = -\frac{a}{2}x^2 + \frac{b}{4}x^4 - Fx\cos\omega t\,. \tag{7.53}$$

For further convenience we present the potential energy in the form

$$V(x,t) = V_0\left[-2(x/c)^2 + (x/c)^4\right] - V_1(x/c)\cos\omega t\,, \tag{7.54}$$

here $c = \pm\sqrt{a/b}$ are the positions of the potential minima at $F = 0$, $V_0 = a^2/4b$ is the potential barrier height, and $V_1 = Fc$ is the amplitude of the modulation of barrier height.

As was shown in the previous section, the time evolution of the particle in a potential field interacting with the equilibrium thermal reservoir can be described by the Langevin equation (7.34). We will further consider the so-called over-damped case—a case involving strong friction, when the inertia (mass) term in the equation of motion can be neglected. In this approximation, assuming the coordinate independence of the frequency coefficient (and therefore of the Langevin force intensity), the Langevin equations can be presented in the form

$$\dot{x} = -\frac{\partial V(x,t)}{\partial x} + \sqrt{D}\Gamma(t)\,. \tag{7.55}$$

The statistical properties of the random force $\Gamma(t)$ are determined by the relation (7.31).

As we have already mentioned, in the absence of modulation ($F = 0$) the average rate of the over-barrier transitions—the Kramers rate—is determined by the relation

$$W_k = \frac{\left[|V''(0)|\,V''(c)\right]^{1/2}}{2\pi}e^{-V_0/D} = \frac{a}{\sqrt{2}\pi}e^{-V_0/D}\,. \tag{7.56}$$

As the Kramers rate depends only on the barrier height and the potential curvature in its extremes, the exact form of the potential is irrelevant. Therefore, the results obtained below are qualitatively applicable to a wide class of bistable systems.

Using the expression (7.40) for transition rates in the presence of periodic perturbation, we obtain

$$W_{\pm}(t) = \frac{a}{\sqrt{2}\pi}\exp\left[-\left(V_0 \pm V_1\cos\omega t\right)/D\right]\,. \tag{7.57}$$

We recall that the Kramers rate (7.56) is obtained under the assumption that the particle is in equilibrium with the thermal reservoir. In order to satisfy that condition in the presence of time-dependent perturbation, it is necessary that the perturbation frequency be much smaller than the characteristic speed of the thermal equilibrium setup in the well. The latter is determined by the quantity $V''(\pm c) = 2a$. Therefore, the adiabatic approximation applicability condition is given by the inequality $\omega \ll 2a$.

As one of the main characteristics of stochastic resonance we will use the signal-to-noise ratio (*SNR*), under which we will understand the ratio of spectral densities for signal and noise on the signal frequency, i.e.

$$\mathrm{SNR} = \left[\lim_{\Delta\Omega\to 0}\int_{\omega+\Delta\Omega}^{\omega-\Delta\Omega} S(\Omega)d\Omega\right]/S_N(\omega) = \frac{S(\omega)}{S_N(\omega)}\,. \tag{7.58}$$

From the relation (7.51), neglecting the influence of the signal on the background of noise, we get

$$\mathrm{SNR} = \frac{\pi}{4}\frac{\alpha_1^2}{\alpha_0}\eta_0^2\,. \tag{7.59}$$

The coefficients α_0 and α_1 can be found with the help of the relations (7.42), in which

$$f(\mu + \eta_0\cos\omega t) = \frac{a}{\sqrt{2}\pi}e^{-(\mu+\eta_0\cos\omega t)}\,, \tag{7.60}$$

where $\mu = V_0/D$, a $\eta_0 = V_1/D = Fc/D$. As a result, in the considered approximation for *SNR* we finally get

$$\mathrm{SNR} \approx \frac{a}{\sqrt{2}}\left(\frac{Fc}{D}\right)^2 e^{-V_0/D}\,. \tag{7.61}$$

For $D \ll V_0$ the exponent tends to zero faster than the denominator and SNR → 0. For large D, the growth of the denominator again assures the tending of *SNR* to zero. In the intermediate region at $D \sim V_0/2$ the approximate expression for *SNR* (7.61) has a unique maximum (see Fig. 7.2).

The two-state model also allows us to find contributions for higher approximations in parameter $\eta_0 = Fc/D$, i.e. to calculate higher harmonics of stochastic resonance. The power spectrum $S(\Omega)$, taking into account these contributions, can be represented as a superimposition of the noise background $S_N(\Omega)$ and δ-peaks, centered at $\Omega = (2n + 1)\omega$. The generation of only odd harmonics of the input signal frequency is a consequence of the symmetry of the considered non-linear

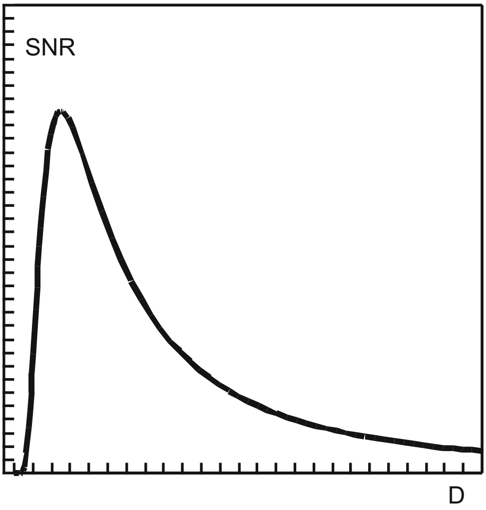


Fig. 7.2 Dependence of the signal-to-noise ratio (*SNR*) on the noise level D in the two-state model

system [19]. We give the expressions for *SNR* on the third and fourth harmonics [20]:

$$\mathrm{SNR}_3 = \frac{\pi}{72}\omega z \left(\frac{Fc}{D}\right)^6 \frac{z^2 + 1/16}{4z^2 + 1}\,,$$

$$\mathrm{SNR}_5 = \frac{\pi}{10^2 \cdot 2^{13}}\omega z \left(\frac{Fc}{D}\right)^{10} \frac{\left(64/3z^2 - 1\right)^2 + (14z)^2}{(4z^2 + 1)\,(4z^2 + 9)}\,, \tag{7.62}$$

where $z \equiv W_k/\omega$. The maxima of these curves are placed in the points $D = V_0/2k$ (k is the harmonic number).

Besides *SNR*, the average value of coordinate x (more precisely, the asymptotic limit of that average at $t_0 \to -\infty$), defined by the relation (7.45), also presents some interest. Using (7.43), we get

$$\langle x(t)\rangle = A(D)\cos\left[\omega t + \phi(D)\right]\,, \tag{7.63}$$

where the amplitude $A(D)$ and phase shift $\phi(D)$ are determined by the expressions

$$A(D) = \frac{Fc^2}{D}\frac{2W_k}{\left(4W_k^2 + \omega^2\right)^{1/2}}\,, \tag{7.64}$$

$$\phi(D) = -\arctan\frac{\omega}{2W_k}\,. \tag{7.65}$$

From the response amplitude on the system output we will determine the power amplification coefficient η,

$$\eta = \frac{A^2(D)}{F^2} = \frac{4W_k^2 c^4}{D^2\left(4W_k^2 + \omega^2\right)}\,. \tag{7.66}$$

From (7.56) and (7.66) it follows that the amplification coefficient η as function of the noise intensity D has a unique maximum.

7.4 Stochastic Resonance in Chaotic Systems

The coexistence of several attractors is typical for the phase space of chaotic systems. Those attractors undergo an infinite number of bifurcations with variations in the system parameters. As a result, such systems are very sensitive to external perturbations. External perturbations in such systems may generate a row of interesting effects connected to the interaction of the attractors, including noise-induced transitions. Therefore, chaotic deterministic systems open new possibilities to set up a stochastic resonance problem. In particular, one may consider the problem of the interaction of two chaotic attractors subject to the influence of external noise and/or some control parameter variation. This interaction is also characterized by some switching frequency, depending on noise intensity and parameter value. Therefore, we can expect the appearance of resonance effects, and as a consequence the possibility of observing a peculiar stochastic resonance in the presence of additional modulation.

Following [21], let us consider, as an example, the discrete system

$$x_{n+1} = (a-1)x_n - ax_n^3\,. \tag{7.67}$$

For $0 < a < 2$ there is a unique stable fixed point at $x_1 = 0$. At $a = 2$ a bifurcation takes place, as a result of which in the region $2 < a < 3$ there are two stable fixed points at $x_{2,3} = \pm c$, $c = \left[(a-2)/a\right]^{1/2}$ and one unstable point in the origin. In the region $3 \leqslant a < 3.3$ a period doubling bifurcations cascade takes place, after which for $a \geqslant 3.3$ the mapping (7.67) demonstrates chaotic behavior.

If $3.3 < a \leqslant 3.6$, there are two disconnected symmetric attractors. Their attraction basins are separated by the separatrix $x = 0$. The stationary probability density $P(x)$ for this case is presented in Fig. 7.3a. At $a \cong a* = 3.598$ the attractors merge and a new chaotic attractor appears with the probability density shown in Fig. 7.3b for $a > a*$. The bifurcation of the merging attractors is accompanied by the alternation phenomenon of the chaos–chaos type [22]: the trajectory lives for a long time in the basin of one of the attractors, and then makes a random transition into the other attractor's basin. The average residence time τ_1 for each of the attractors

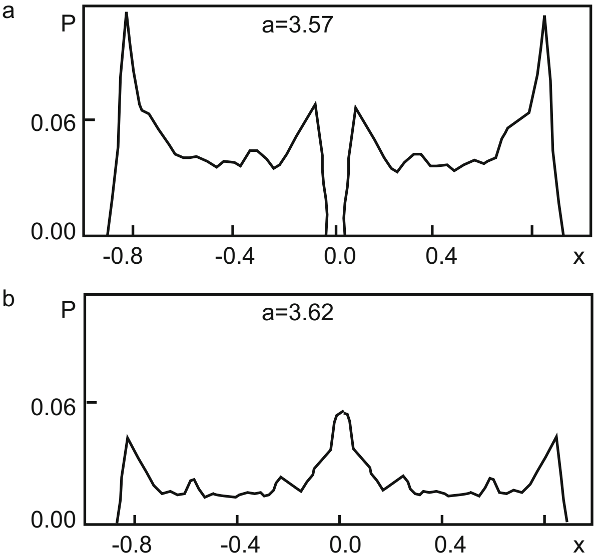


Fig. 7.3 Stationary probability density $P(x)$ for the system (7.67) for two parameter a values: (a) $a < a^*$, (b) $a > a^*$ [21]

obeys the universal critical relation

$$\tau_1 \sim (a - a*)^{-\gamma}\,; \quad \gamma = 0.5\,. \tag{7.68}$$

The alternation effect of the chaos–chaos type can be achieved also as a result of the action of additive noise. In this case, the dependence (7.68) is preserved, but the critical index γ becomes a function of the noise intensity, $\gamma = \gamma(D)$.

Let us introduce periodic modulation and additive noise into the mapping (7.67)

$$x_{n+1} = (a-1)x_n - ax_n^3 + \varepsilon \sin(2\pi f_0 n) + \xi(n)\,, \tag{7.69}$$

where ε and f_0 are amplitude and frequency of modulation, and the statistical properties of the noise are the following

$$\langle \xi(n) \rangle = 0, \quad \langle \xi(n)\xi(n+k) \rangle = 2D\delta(k)\,. \tag{7.70}$$

Let us study the system (7.69) in the two-state approximation, replacing the $x(n)$ coordinate by $+1$, if $x(n) > 0$ and by -1, if $x(n) < 0$. In the approximation $\dot{x} = x_{n+1} - x_n$, we can transform the discrete model (7.67) into the differential equation

$$\dot{x} = (a-2)x - ax^3 \tag{7.71}$$

and introduce the potential $U(x)$:

$$U(x) = -\frac{a-2}{2}x^2 + \frac{a}{4}x^4 \,. \tag{7.72}$$

This allows us to determine the Kramers rate

$$W_k = -\frac{a-2}{\pi\sqrt{2}} \exp\left[-\frac{(a-2)^2}{4aD}\right] \tag{7.73}$$

and to obtain the expression for *SNR* in an adiabatic approximation

$$\mathrm{SNR} = -\frac{(a-2)^2\varepsilon^2}{aD^2} \exp\left[-\frac{(a-2)^2}{4aD}\right] . \tag{7.74}$$

Let us consider the dynamics of (7.69) at $a = 3.4$, which corresponds to the case of the coexistence of two disconnected attractors. The addition of noise (at $\varepsilon = 0$) smoothes the probability density and induces transitions between the attractors. The basic characteristics of the dynamics in the absence of modulation (probability density $P(x)$, power spectrum $S(f)$, residence time distribution function for the attractor $p(n)$ and average frequency of transitions between the attractors f_s) as functions of the noise amplitude D are presented in Fig. 7.4. They reflect the typical features of the bistable system in presence of noise.

Figure 7.5 presents the results of the numerical analysis of the mapping (7.69) with the inclusion of periodic perturbation with $\varepsilon = 0.05$ and $f_0 = 0.125$. A sharp peak appears in the power spectrum at the frequency f_0. Peaks also appear in the function of the distribution of residence time on the background decay, and they are centered on the times divisible by an even number of the perturbation semi-periods. And finally, SNR(D) demonstrates a clear maximum at a certain noise intensity. The dependence SNR(D) agrees with theoretical predictions (7.74). It may be said that the replacement of potential wells with isolated attractors preserves all the features of stochastic resonance.

Now we turn to a case of the absence of external noise ($D = 0$). As we have already said (see Fig. 7.3), at $a > a*$ random transitions between the attractors occur due to the internal deterministic dynamics of the system. We can assume that in this case the synchronization between those transition and the periodic perturbation frequency also leads to some analogue of the usual stochastic resonance with external noise. A numerical calculation of the SNR(a) dependence confirms this assumption. Figure 7.6 shows the SNR(a) dependencies in the two-state approximation (Fig. 7.6a) and for full dynamics (Fig. 7.6b), described by (7.69). Both in the first and second cases we observe clear maxima for the *SNR* curves at parameter a values corresponding to a ratio of frequencies f_s:f_0 = 1:3, 1:1, 4:3.

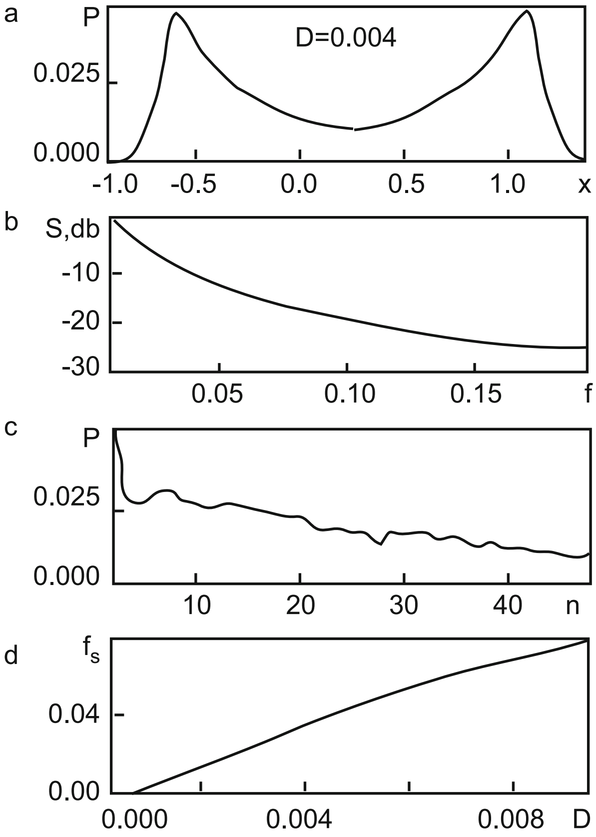


Fig. 7.4 Basic dynamical characteristics (7.69) in absence of modulation: (**a**) probability density $P(x)$, (**b**) power spectrum $S(f)$, (**c**) the distribution function of the residence times for the attractor $p(n)$, (**d**) average frequency of transitions between the attractors f, as functions of the noise amplitude D; $a = 3.4$ [21]

This result is understandable from the point of view of general stochastic resonance philosophy. As we have already mentioned, stochastic resonance represents a generic phenomenon for non-linear systems with several time scales. The dependence of one of the scales on external perturbation allows us to assure certain resonance conditions. In the original setup of a bistable system perturbed by periodic signal and noise, one utilizes the dependence of the Kramers rate of transitions between the potential minima on the noise amplitude. In order to obtain analogous results in the case of a chaotic system with several attractors in a chaos–chaos alternation regime, one can use the dependence of the average frequency of the transitions between the attractors on the controlling parameter a.

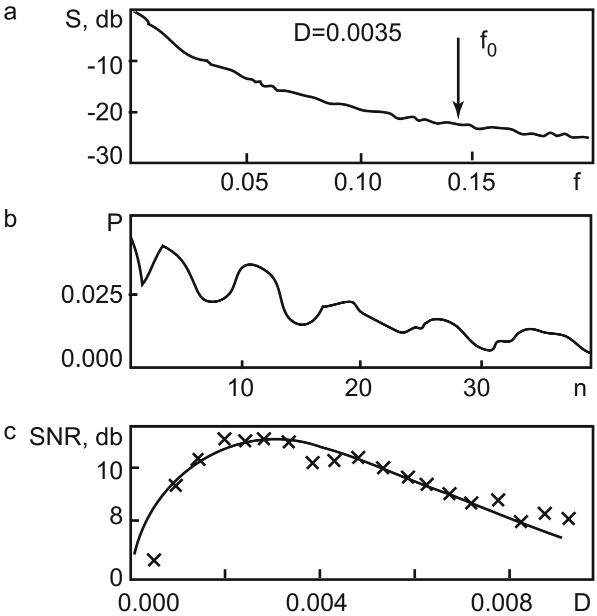


Fig. 7.5 Results of numerical analysis of the mapping (7.69) with inclusion of periodic perturbation with $\varepsilon = 0.05$ and $f_0 = 0.125$: (**a**) power spectrum, (**b**) residence time distribution function, (**c**) signal-to-noise ratio [21]

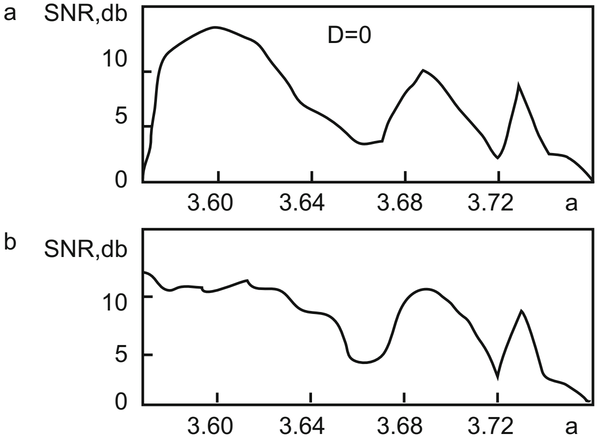


Fig. 7.6 Results of SNR(a) calculation for the system (7.69) in absence of noise ($\varepsilon = 0.05$, $f_0 = 0.125$, $D = 0$): (**a**) in the two-state approximation, (**b**) precise dynamics [21]

7.5 Stochastic Resonance and Global Change in the Earth's Climate

Now we discuss in more detail how to use the stochastic resonance effect for a qualitative explanation of the alternation of ice ages on Earth [1–4].

The chronology of ice ages (the global volume of ice on Earth) can be reconstructed from the ratio of the isotopes ^{16}O and ^{18}O in organic sediments [23]. Almost all of the oxygen in water is made up of the ^{16}O isotope and only fractions of a percent belong to the heavier ^{18}O. As the evaporation of heavier isotopes is less probable, precipitations on land (they are mainly determined by the evaporation of the oceans) are ^{18}O isotope depleted. During ice ages, the continental glacial cover increases at the cost of the ocean (in the last Ice Age 18,000 years ago, the ocean level was almost 100 m lower than in the present, and up to 5 % of all total water volume was on land in form of ice) and they are enriched in ^{18}O isotope. The ratio which interests us can be determined analyzing the isotope composition of calcium carbonate $CaCO_3$, which shells of sea animals are made of. These shells accumulate on the ocean floor in form of the sedimentary layers. The more the ratio $^{18}O/^{16}O$ is in those sediments, the larger the continental ice volume was at the moment of shell formation.

The isotope composition time dependence [2], constructed based on these measurements, clearly demonstrates the periodicity of the variation in global ice quantity on the planet: the ice ages came every 100,000 years. Of course, the time dependence presented in Fig. 7.7 is non-trivial: the dominating 100,000-year cycle interferes with additional smaller oscillations. What external effects could result in such periodic dependence? In the first half of the twentieth century, a Yugoslavian astronomer, M. Milankovich, developed a theory connecting the global changes of Earth climate to variations in insolation (the quantity of solar energy reaching Earth). Even if we assume that solar radiation is constant, global insolation will still depend on geometrical factors describing the Earth orbit. In order to consider the dynamics of insolation, one should study the time dependence of the following three parameters: the slope of the Earth's axis in relation to the orbital plane,

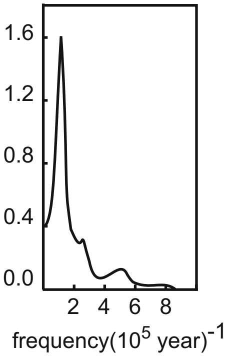


Fig. 7.7 The power spectrum of climatic changes for the last 700,000 years [2]

orbital eccentricity, and the precession of the Earth's orbit. Gravitational interaction with the Moon and other planets lead to the time dependence of those parameters. Measurements and calculations showed that during the last million years these dependencies have an almost periodic character. The slope of the Earth's axis changes between 22.1° and 24.5° in a period of about 40,000 years (at present, it is 23.5°). The eccentricity of the Earth's orbit oscillates between 0.005 and 0.06 (being 0.017 at present) with a period of 100,000 years (the very time scale that interests us). And finally, the period of the precession of the Earth's axis is 26,000 years. What is the role of these factors in the Earth climate dynamics? An increase in the Earth's slope increases the amplitude of seasonal oscillations. The precession weakly affects the insolation and mostly determines the perihelion passing time. The latter smoothes the seasonal contrasts in one hemisphere and amplifies them in the other. Therefore, the first two factors do not affect the total insolation, but just redistribute it along latitudes and in seasons. Only the variation of eccentricity changes the total annual insolation. However, the insolation oscillations connected with that effect do not exceed 0.3 %, which leads to average temperature changes of not more than a few tenths of a degree, while during an ice age, the average annual temperature decreases in the order of 10°C. So how can variations in the parameters of the Earth's orbit cause global climate changes? The answer to the question is given by the following statement: a simultaneous account of a small external periodic force with a period of 10^5 years (modeling the oscillations of the eccentricity of the Earth's orbit) and random noise effects (modeling climate fluctuations at shorter time scales, connected with random processes in the atmosphere and in oceanic currents) in the dynamics of climate changes allows us to satisfactorily reproduce the observed periodicity of ice ages.

In order to prove the above made statement we consider a simple model allowing us to account for the influence of insolation variation on the average temperature of the Earth T. The model represents the heat-balance equation for the radiation coming to Earth R_{in} and emitted by it R_{out}

$$C\frac{dT}{dt} = R_{\text{in}}(T) - R_{\text{out}}(T)\,, \tag{7.75}$$

where C is the Earth's thermal capacity. For the quantities R_{in} and R_{out} we use the following parametrization

$$\begin{aligned} R_{\text{in}}(T) &= Q\mu\,, \\ R_{\text{out}}(T) &= \alpha(T)Q\mu + \varepsilon(T)\,. \end{aligned} \tag{7.76}$$

Here Q is the solar radiation reaching the earth, averaged over a long time period, μ is the dimensionless parameter allowing us to introduce explicit time variation in the incident flow, $\alpha(T)$ is the average albedo of the Earth surface (albedo is the photometric quantity determining the ability of a matte surface to reflect the incident radiation, i.e. ratio of the radiation reflected by the surface to the incident), $\varepsilon(T)$ is the long-wave surface radiation of the heated Earth $\left(\varepsilon(T) \sim T^4\right)$.

Let us rewrite (7.75) in the form

$$\frac{dT}{dt} = F(T); \quad F(T) \equiv \left(R_{\mathrm{in}}(T) - R_{\mathrm{out}}(T)\right)/C\,. \tag{7.77}$$

Solutions of the equation $F(T) = 0$ represent physically observable equilibrium states of the considered model (7.75). They are usually called "climates" [2]. The properties of climatic stability are determined by the pseudo-potential Φ,

$$\Phi = -\int F(T)dT\,. \tag{7.78}$$

It is evident that

$$F(T) = -\frac{\partial \Phi}{\partial T} \tag{7.79}$$

and therefore the extrema of function Φ correspond to the above notions of climate. The climate will be stable if it corresponds to a minimum of Φ. In this case it is physically observable. The model (7.75) or (7.77) must reproduce the two following basic observable facts of the Earth's climate dynamics:

1. Local climatic changes are limited by a temperature scale in the order of a few degrees.
2. At time scales of an order of 10^5 years, substantially larger average temperature changes occur (in the order of 10°), resulting in drastic changes in the planet's climate.

For a description of such dynamics, it is natural to use the pseudo-potential Φ with two stable climates T_1 and T_3 (the minima), separated by the temperature interval of an order of 10°, and one unstable climate T_2 (the maximum) between them (see Fig. 7.8). One of the minima (T_1) corresponds to the ice age climate, the second (T_3)—to the present climate. The appearance of an unstable climate (T_2) in the intermediate region is easily understood from simple physical considerations. Let the unstable state correspond to some quantity of planetary ice cover. If, due to local fluctuations, the temperature decreases, the ice surface will increase, which will lead to an increase of the Earth's local albedo and further temperature decrease. An analogous situation also takes place with local temperature increases in the vicinity of the same point.

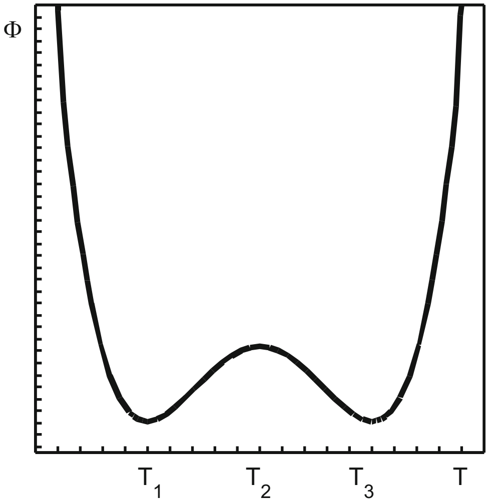


Fig. 7.8 Pseudo-potential Φ with two stable (T_1, T_3) and one unstable T_2 climates

Let us now introduce into Eq. (7.75) the time dependent factor $\mu(t)$, accounting for the variations of insolation, connected to the oscillations of the eccentricity of the Earth's orbit

$$\mu(t) = 1 + 0.0005 \cos \omega t; \quad \omega = \frac{2\pi}{10^5 \text{ years}}. \tag{7.80}$$

The transition $F(T) \to F(T, t)$ corresponds to the introduction of the time-dependent potential $\Phi(T, t)$. A time dependence (7.80) of such low amplitude leads only to temperature oscillations in vicinity of the states (climates) T_1 and T_3, and it cannot explain the alternation of ice ages.

Let us now take into account short time scale climate fluctuations, including into Eq. (7.77) white noise, and transforming it into the stochastic differential equation

$$\frac{dT}{dt} = F(T, t) + \sigma \xi(t). \tag{7.81}$$

In this formulation the problem of the Earth's climate changes coincides exactly with the above considered problem of particle dynamics in the symmetric double well under the simultaneous influence of weak periodic perturbation and noise. Figure 7.9 shows the numerical solution of Eq. (7.81) with parameters $T_3 - T_1 = 10\,\text{K}$ and a white noise dispersion of $0.15\,\text{K}^2/\text{year}$. The figure clearly shows the stochastic resonance effect, manifesting in the periodic transitions $T_1 \leftrightarrow T_3$, accompanied by small oscillations around the stable states.

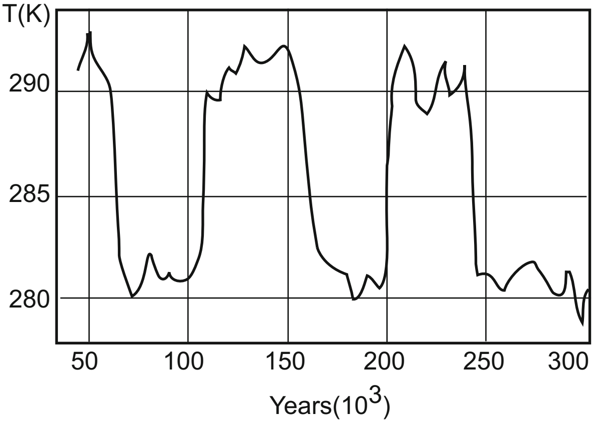


Fig. 7.9 Results of numerical solving of the stochastic differential equation (7.81) [2]

References

1. Benzi, R., Sutera, A., Vulpiani, A.: J. Phys. A **14**, L453–L457 (1981)
2. Benzi, R., Parisi, G., Sutera, A., Vulpiani, A.: Tellus **34**, 10–16 (1982)
3. Nicolis, C., Nicolis, G.: Tellus **33**, 225–237 (1981)
4. Nicolis, C.: Tellus **34**, 1–9 (1982)
5. Gang, H., Ditzinger, T., Ning, C.Z., Haken, H.: Phys. Rev. Lett. **71**, 807–810 (1993)
6. Kramers, H.: Physica (Utrecht) **7**, 284–312 (1940)
7. Anishenko, V.C., Neiman, A.B., Moss, F., Shimansky-Gaier, L.: UFN **169**, 7–39 (1999)
8. Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F.: Rev. Mod. Phys. **70**, 223–287 (1998)
9. Gross, D.: Lect. Notes Phys. **117**, 81 (1980)
10. Frobrich, P., Lipperheide, R.: Theory of Nuclear Reactions. Oxford University Press, Oxford (1996)
11. Frobrich, P., Gontchar, I.: Phys. Rep. **292**, 131–237 (1998)
12. Gardiner, C.W.: Handbook of Stochastic Methods. Springer Series in Synergetics, vol. 13, 2nd edn. Springer, Berlin/Heidelberg (1985)
13. Nyquist, H.: Phys. Rev. **32**, 110–113 (1928)
14. Callen, H., Welton, T.: Phys. Rev. **83**, 34–40 (1951)
15. Kubo, R.: J. Phys. Soc. Jpn. **12**, 570–586 (1957)
16. Langevin, P.: C. R. Acad. Sci. Paris **146**, 530–533 (1908)
17. McNamara, B., Wiesenfeld, K.: Phys. Rev. A **39**, 4854–4869 (1989)
18. Landau, L.D., Lifshitz, E.M.: Statistical Physics. Pergamon Press, Oxford (1980)
19. Jung, P., Hänggi, P.: Europhys. Lett. **8**, 505–510 (1989)
20. Zhiglo, A.V.: Probl. Atom. Sci. Technol. **6**, 251–254 (2001)
21. Anishchenko, V.S., Neiman, A.V., Safanova, M.A.: J. Stat. Phys. **70**, 183–196 (1993)
22. Anishenko, V.S.: JETF Lett. **10**, 629–633 (1984)
23. Covey, C.: Sci. Am. **250**, 42–50 (1984)

Chapter 8
The Appearance of Regular Fluxes Without Gradients

The central topic of the present section is transport phenomena in spatially periodic systems in the absence of displacing macroscopic forces (gradients). A complete solution of the problem must include an understanding of the effect's essence, the establishment of the conditions at which a gradient-free current is possible, and a quantitative investigation of the models and realistic systems where the effect can be observed.

8.1 Introduction

Mass, energy, and charge transport is the basis of very diverse phenomena in nature. On the macroscopic level one should distinguish between convective and diffusive transport. The characteristic feature of the former is directed motion, generated by gradients of different fields: force, temperature, concentration, etc. At the same time, directed motion is absent in diffusive transport, which is generated by random collisions. The ambitious task to obtain useful work (or directed motion) from fluctuations, in spite of seemingly overwhelming difficulties, has long attracted the attention of enthusiasts. According to statistics, for every proposal of a perpetuum mobile of first kind there are eight projects of the second kind. The principal difficulty is in the fact that useful work cannot be extracted from thermal equilibrium fluctuations. Such a machine would violate the second law of thermodynamics.

There are several equivalent formulations of the second law. The very term as well as the first formulation is credited to Clausius [1]: a process where heat spontaneously transits from the colder bodies to the warmer ones is impossible. The term "spontaneously" should be understood in the sense that the transition cannot be realized with the help of any setups without some other changes in nature. In other words, it is impossible to realize a process, the only outcome of which would be heat transfer from a colder body to a warmer one. If, in violation of the Klausius

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_8

formulation, such a process appeared to be possible, it would allow one to obtain two reservoirs with different temperatures by the simple division of one reservoir in two parts and the transferring of heat from one part to another. In its turn it would allow one to perform the Carno cycle and to obtain mechanical work with the help of a periodically acting machine at the cost of internal energy—in the end, at the cost of one thermal reservoir. As this is impossible, in nature it is impossible to realize a process, the only result of which would be load lifting (i.e., mechanical work) executed due to the cooling of the thermal reservoir. The latter represents the formulation of the second law of thermodynamics by W. Thomson. It is easy to see that the two formulations are interrelated. The mechanical work obtained by cooling of the colder reservoir could be used to heat the warmer one (e.g., by friction), which would violate the Klausius principle.

Having come to terms with the fact that a perpetuum mobile of the second kind is unrealizable, we can pose the problem of how to minimally violate the functioning conditions of the second law in order to make gradient-free current possible. To that end, let us look at the problem from a somewhat different point of view. A perpetuum mobile of the second kind represents one of the many idealized objects convenient for theorists to deal with but absolutely unavailable to experimenters. Let's divide the physicists interested in this problem on those who gives priority to symmetry approach and those who "honestly" (analytically or numerically) solve the equations of motion. In the 1950s, the physical elite (the physicists who dealt with elementary particles) turned to symmetry. That transition was to a great extent necessitated by the absence of corresponding equations of motion. It is well known that symmetries are equivalent to the conservation laws. Not calling into question the great progress achieved in this way, let us pose a naive question: which is better, symmetries (conservation laws) or equations of motion? The objective answer is: equations of motion are better, for two reasons. First, equations of motion account for symmetry automatically, while symmetry does not contain any dynamics. Second, a real physical situation always corresponds to broken symmetry, and a breakdown in symmetry is more easily inserted into equations of motion. We note that the computer revolution gave rise to the possibility of substantially advancing solutions of realistic problems with broken symmetry. In essence, the newest history of the considered problem began from the attempt [2] to consider a realistic physical situation instead of an idealized Smoluchowski–Feynman gedanken experiment. But we will start from the latter.

In the Feynman lectures [3] the problem is discussed with the help of a mechanical ratchet model. The model was first invented and analyzed by Smoluchowski [4] in the golden age of the Brownian motion theory. Smoluchowski showed that in absence of a thinking creature (like the Maxwell demon) the intriguing possibility to extract useful work from equilibrium fluctuations cannot be realized.

The device (see Fig. 8.1) is very simple: on one end of the axis there is a rotator, on the other—a ratchet, which, due to the pawl, can rotate only in one direction. If the rotator is surrounded with a gas, then the collisions of gas molecules with the rotator blades will make it rock in random way. At the first sight it seems that, due to

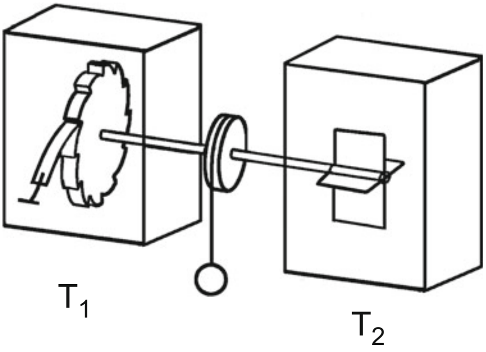


Fig. 8.1 A ratchet with a pawl

presence of the ratchet and the pawl on the other end of the axis, sufficiently strong unidirectional fluctuations will lift the pawl and the ratchet will rotate.

However, analysis shows that in order to obtain mechanical work, the rotator must be inside a thermal reservoir at temperature T_1 higher than that surrounding the ratchet T_2. Aside from that, calculations of the device's efficiency made with some simplifying assumptions show that it is equal to the efficiency of the Carno cycle.

Let us now consider how the setup works without a load and at equal temperatures of the two reservoirs. Let T be that temperature and ε the energy required to lift the pawl up the tooth, overcoming the action of the spring that pulls the pawl down. At low temperatures, the speed of fluctuations providing the rotator with sufficient energy in order to turn the ratchet on one tooth is proportional to the factor $\exp(-\varepsilon/k_BT)$. But the pawl is also placed into the reservoir with temperature T and therefore it can also be lifted by the fluctuations of the reservoir. Moreover, that inverse motion occurs with the same speed. Therefore, if both reservoirs are at the same temperature, no directed ratchet motion appears.

Suppose further that the reservoirs have different temperatures $T_1 > T_2$, i.e. that the pawl is colder than the rotator. Now the jump speeds are no longer equal and their difference, generally speaking, can be used to lift the load attached by the thread which is wound on the axis of our setup (see Fig. 8.1). Evidently, the value of the load's weight (it is more convenient to refer to its rotating moment L) is such that both speeds are equal and the ratchet will not rotate directionally. This value is easy to calculate [5].

If $L\theta$ is the potential energy of the load acquired by it during the rotation of the ratchet on one tooth, then $\varepsilon + L\theta$ is the total energy necessary for this rotation. This energy mostly comes from the rotator, so the "forward" rotation speed is proportional to $\exp[-(\varepsilon + L\theta)/k_BT_1]$. For rotation in the inverse direction, it is necessary to lift the pawl, and the energy needed for that equals ε. Feynman supposed that the energy is taken from the thermal reservoir where the ratchet is placed, and therefore that speed is proportional to $\exp(-\varepsilon/k_BT_2)$. There exists a moment L_0 for which those speeds are equal

$$\frac{L_0\theta + \varepsilon}{\varepsilon} = \frac{T_1}{T_2}\,. \tag{8.1}$$

Let us now turn to the calculation of the energy transferred between the reservoirs, the ratchet, and the rotator. We have seen that after the "forward" jump the system acquires the energy $\varepsilon + L\theta$ from the first reservoir. After the jump, the energy ε will dissipate. Feynman assumed that it is completely dissipated in the second reservoir. At backward motion, energy ε is taken from reservoir 2 and after the jump the energy $\varepsilon + L\theta$ dissipates. A further assumption states that it dissipates in the reservoir 1.

Let us now take the L value a little less than L_0, such that the wheel will slowly move "forward," lifting the load. With the above assumptions on energy transfer it is not difficult to calculate the efficiency of our setup. If the ratchet makes N_+ steps forward and N_- steps back, then the total work equals $(N_+ - N_-)\, L\theta$, and the heat quantity taken from the reservoir 1 is $(N_+ - N_-)\,(\varepsilon + L\theta)$. Therefore the coefficient of efficiency for the considered setup is

$$\eta = \frac{L\theta}{L\theta + \varepsilon}\,, \tag{8.2}$$

and in the limit $L \to L_0$ the setup efficiency is reduced to one of the Carno cycle η_c,

$$\eta \to \eta_c = 1 - \frac{T_1}{T_2}\,. \tag{8.3}$$

We note that, as was shown in [5], the Feynman analysis contains certain inexactitude, which we will briefly discuss. As it is well known, efficiency (8.3) is achieved when the Carno cycle works in a reversible way. If, at some time interval, the setup received from reservoir 1 the heat quantity Q_1 and transmitted to reservoir 2 the heat quantity Q_2, then the work executed is $W = Q_1 - Q_2$. Thus, the reservoir 1 entropy decreased in quantity Q_1/T_1, and in the second reservoir it increased in quantity Q_2/T_2. Reversibility implies the constancy of entropy, i.e.

$$\varDelta S = \frac{Q_1}{T_1} - \frac{Q_2}{T_2} = 0\,. \tag{8.4}$$

From there, the efficiency coefficient immediately follows

$$\frac{W}{Q_1} = \eta_c = 1 - \frac{T_1}{T_2}\,. \tag{8.5}$$

Therefore the Feynman analysis assumes that the ratchet works reversibly. However, this is not so [5]. In contrast to the Carno machine, the Feynman ratchet (at different temperatures) always works in conditions of non-equilibrium. Different parts of the system simultaneously contact the reservoirs with different temperatures. Therefore the system can never be in thermodynamic equilibrium. We stress that we speak not about the trivial irreversibility due to the heat flow along the setup axis (in an idealized experiment it can be considered equal to zero due to a choice of material with zero thermal conductivity). Of course, that inexactitude does not affect the important statement about the absence of systematic rotation at $T_1 = T_2$.

The above analysis of ratchet "dynamics" makes more clear the following statement: if we exclude transfer processes, then directed transport in spatially periodic systems contacting the unique source of dissipation and noise—a thermal reservoir—is forbidden by the second law of thermodynamics. Therefore, to solve the problem the system must be transferred into a state far from thermal equilibrium with the help of deterministic or stochastic perturbation. We will mostly consider the case of periodic perturbation and a limited set of stochastic processes. The case of non-displacing perturbation, i.e. when the space, time, and ensemble averages are zeros, presents the most interest. Those perturbations can be either externally applied, or of internal origin, or due to a second thermal reservoir with another temperature, or an additional non-thermal reservoir. We focus on small systems, where one should not only account for thermal noise, but it even can play dominant role. The physical source of thermal noise is the thermal environment of the system. As an unavoidable consequence of this, there are dissipative effects.

Besides the breaking of thermal equilibrium, there is a second necessary condition of the directed transport—the breaking of spatial reflective symmetry. There are a number of different possibilities of doing this, and we will refer to the ratchet or equivalently about the Brownian motor every time one of the conditions or a combination of both is satisfied.

8.2 Dynamical Model of the Ratchet

Let us turn to a simple model [6], describing the one-dimensional dynamics of the Brownian particle with coordinate $x(t)$ and mass m in the potential $V(x)$. We will assume that the particle is in contact with a thermal equilibrium reservoir. As was shown in the previous chapter, the dynamics of this particle can be described by the Langevin equation

$$m\ddot{x}(t) + V'(x) = -\gamma\dot{x}(t) + \xi(t)\,. \tag{8.6}$$

Here $V(x)$ is the periodic potential with period L,

$$V(x+L) = V(x)\,, \tag{8.7}$$

playing the role of ratchet in the Smoluchowski–Feynman model, and therefore it has broken spatial symmetry. Under the latter we understand a potential property where there is no Δx, for which

$$V(-x) = V(x+\Delta x)\,, \tag{8.8}$$

for all x. In other words, the potential cannot be made an even function of coordinate by any choice of the origin. A typical example of ratchet potential $V(x) = V_0[\sin(2\pi x/L) + 0.25\sin(4\pi x/L)]$ is presented in Fig. 8.2.

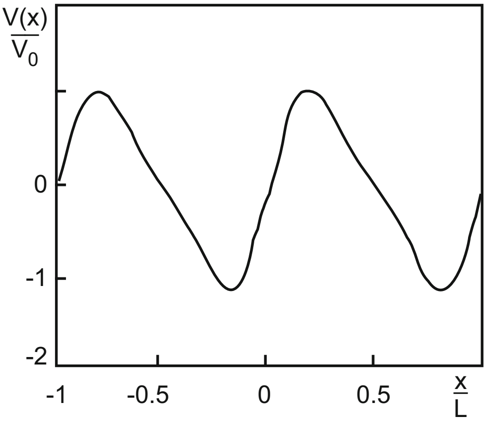


Fig. 8.2 The ratchet potential $V(x) = V_0[\sin(2\pi x/L) + 0.25\sin(4\pi x/L)$ [6]

The right-hand side of Eq. (8.6) accounts for the effects of interaction with the thermal reservoir. We recall that both the friction force $-\gamma\dot{x}(t)$ and the random fluctuating force $\xi(t)$, which we assume to be Gaussian, are due to the same mechanism—the interaction of the Brownian particle with the degrees of freedom of the thermal reservoir. The statistical properties of the random force are completely determined by the setting of its mean value and correlation function

$$\left\langle \xi(t)\xi(t')\right\rangle = 2\gamma k_B T\delta(t-t')\,, \tag{8.9}$$

where the averaging $\langle\ldots\rangle$ is made over the random process realizations $\xi(t)$.

It is important to note that in the Brownian world (where thermal fluctuations are of the order of the potential barrier heights), the renormalized dimensionless mass usually appears to be much less than unity, while all other dimensionless parameters remain at unity. This means that the inertia term $m\ddot{x}$ can be neglected in a good approximation. Such so-called over-damped dynamics will mostly interest us further. In that limit the simplest version of the Smoluchowski–Feynman model is described by the equation

$$\gamma\dot{x}(t) = -V'[x(t)] + \xi(t)\,. \tag{8.10}$$

The Smoluchowski–Feynman ratchet model, formulated with the help of the Langevin equation (8.6) [or (8.10)] has a row of important advantages before the original mechanic model. Dealing with the Langevin equation (or with the equivalent Fokker–Planck equation) we should not refer to the second law of thermodynamics as to some a priori statement. Besides that, any modification of the original model is treated using the Langevin equation in the standard way.

A natural further step is to consider the statistical ensemble of random variables $x(t)$, representing the independent realizations of the Langevin equation (8.10). The corresponding probability density $P(x,t)$,

$$P(x,t) \equiv \langle\delta(x-x(t))\rangle\;, \tag{8.11}$$

as is clearly seen (see, for example, [7]), obeys the Fokker–Planck equation which in the particular case of over-damped motion is called the Smoluchowski equation

$$\frac{\partial}{\partial t}P(x,t) = \frac{\partial}{\partial x}\left\{\frac{V'(x)}{\gamma}P(x,t)\right\} + \frac{k_BT}{\gamma}\frac{\partial^2}{\partial x^2}P(x,t)\,. \tag{8.12}$$

The quantity which interests us is the particle current, defined as the average over the ensemble of the particle's velocities

$$j(t) \equiv \langle \dot{x}(t)\rangle\ . \tag{8.13}$$

In the considered model, this quantity can be easily calculated. Averaging (8.10) over the ensemble and accounting that $\langle \xi(t)\rangle = 0$, we get $j = -\langle V'[x(t)]\rangle/\gamma$. As the ensemble average is by definition averaging with the probability density $P(x,t)$, then

$$j = -\int\limits_{-\infty}^{\infty} dx\frac{V'(x)}{\gamma}P(x,t)\,. \tag{8.14}$$

The expression for the current j can be obtained independently from the model dynamics using the continuity equation

$$\frac{\partial}{\partial t}P(x,t) + \frac{\partial}{\partial x}J(x,t) = 0\,, \tag{8.15}$$

where $J(x,t)$ is the probability current density,

$$J(x,t) \equiv \langle \dot{x}(t)\delta(x - x(t))\rangle\ . \tag{8.16}$$

Integrating (8.16) over x, we find a connection between the particle current and the probability current density

$$j = \int\limits_{-\infty}^{\infty} dxJ(x,t)\,. \tag{8.17}$$

Integrating (8.17) by parts and using the continuity equation (8.15), the latter relation can be rewritten in the form

$$j = \frac{d}{dt}\int\limits_{-\infty}^{\infty} dxxP(x,t)\,. \tag{8.18}$$

This relation can be considered as an alternative definition of particle current.

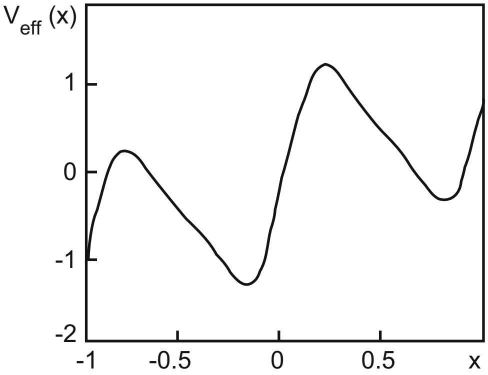


Fig. 8.3 Tilting ratchet potential $V(x) = \sin(2\pi x) + 0.25\sin(4\pi x) - Fx$, $(F = -1)$ [6]

Comparing the continuity equation (8.15) with the Fokker–Planck equation, we get an explicit expression for $J(x,t)$ in the considered model

$$J(x,t) = -\left\{\frac{V'(x)}{\gamma} + \frac{k_BT}{\gamma}\frac{\partial}{\partial x}\right\} P(x,t)\,. \tag{8.19}$$

Substituting (8.19) into (8.17) and taking into account that $P(x,t) \to 0$ for $x \to \pm\infty$, we return to (8.14).

Let us now obtain the expression for stationary particle current in the periodic potential (8.7), making first some generalizations of the initial model (8.10). Let us include into the model (8.10) a uniform stationary load F, equivalent to the plummet in the mechanical Smoluchowski–Feynman model. On the potential level it implies a transition to the sloping potential

$$\bar{V}(x) = V(x) - Fx\,. \tag{8.20}$$

Figure 8.3 gives an example of the sloping potential $(F = -1)$ with broken reflective symmetry

$$\bar{V}(x) = \sin(2\pi x) + \sin(4\pi x) - Fx\,. \tag{8.21}$$

The Langevin equation in the presence of additional uniform static force F reads

$$\gamma\dot{x}(t) = -V'[x(t)] + F + \xi(t)\,. \tag{8.22}$$

The corresponding Smoluchowski equation is

$$\frac{\partial P(x,t)}{\partial t} = \frac{1}{\gamma}\left(V'(x) - F + k_BT\frac{\partial}{\partial x}\right) P(x,t) = \frac{\partial J(x,t)}{\partial x}\,. \tag{8.23}$$

The stationary solution (8.22), corresponding to the constant density of the probability current J, satisfies the equation

$$\gamma J = (F - V'(x))P(x) - k_B T \frac{dP(x)}{dt} \, . \tag{8.24}$$

The solution of this equation can be presented in the following form [7]

$$P(x) = e^{-\bar{V}(x)/k_B T} \left[N - \gamma (J/k_B T) \int_0^x e^{\bar{V}(x')/k_B T} dx' \right] , \tag{8.25}$$

where N is the integration constant. If we require that $P(x)$ is bounded for large x, then we can prove that the function $P(x)$ is periodic. For that we calculate the integral

$$\int_0^{nL+x} e^{\bar{V}(x')/k_B T} dx' = \int_0^L e^{\bar{V}(x')/k_B T} dx' + \cdots \int_{(n-1)L}^{nL} e^{\bar{V}(x')/k_B T} dx' + \int_{nL}^{nL+x} e^{\bar{V}(x')/k_B T} dx' \, ,$$

performing the integration and accounting that $\bar{V}(x + nL) = \bar{V}(x) - nLF$, we get

$$\int_0^{nL+x} e^{\bar{V}(x')/k_B T} dx' = I \frac{1 - e^{-nLF/k_B T}}{e^{-LF/k_B T}} + e^{-nLF/k_B T} \int_0^x e^{\bar{V}(x')/k_B T} dx' \, , \tag{8.26}$$

where $I = \int_0^L e^{\bar{V}(x)/k_B T} dx$. From (8.25) taking into account (8.26), we get

$$\begin{aligned} P(x + nL) = {} & e^{-\bar{V}(x)/k_B T} \left[N - \frac{\gamma J I}{k_B T (1 - e^{-LF/k_B T})} \right] e^{nLF/k_B T} \\ & + e^{-\bar{V}(x)/k_B T} \left[\frac{\gamma J I}{k_B T \left(1 - e^{-LF/k_B T}\right)} - \gamma \frac{J}{k_b T} \int_0^x e^{\bar{V}(x')/k_B T} dx' \right] . \end{aligned} \tag{8.27}$$

For $F > 0$ ($F < 0$) that expression can be bounded in the limit $n \to +\infty$ ($n \to -\infty$) only provided the first bracket (8.27) will be equal to zero, i.e.

$$\gamma J I = k_B T N (1 - e^{-LF/k_B T}) \, . \tag{8.28}$$

Then from (8.25) and (8.27) we get

$$P(x+L) = P(x) \tag{8.29}$$

which proves the above statement about the periodicity of the distribution function $P(x)$. We normalize the distribution function on the periodicity interval

$$\int_0^L P(x)dx = N \int_0^L e^{\bar{V}(x)/k_BT} dx - \gamma \frac{J}{k_BT} \int_0^L e^{-\bar{V}(x)/k_BT} \left(\int_0^x e^{\bar{V}(x')/k_BT} dx' \right) dx \,. \tag{8.30}$$

Finally, the particle current j can be found by the Langevin equation averaging over the ensemble of the random variable $\xi(t)$ realizations, i.e. by averaging with the distribution function $P(x)$,

$$\begin{aligned} j = \langle \dot{x} \rangle &= \gamma^{-1} \left\langle F - V'(x) + \xi(t) \right\rangle = \gamma^{-1} \left\langle F - V'(x) \right\rangle \\ &= \gamma^{-1} \int_0^L \left[F - V'(x) \right] P(x) dx = \gamma^{-1} \int_0^L \left(\gamma J + k_B T dP/dx \right) dx = LJ \,. \end{aligned} \tag{8.31}$$

Deriving (8.31) we used the relations (8.22), (8.24) and the periodicity condition for the distribution function $P(x)$ (8.29). Finding from (8.28) the expression for the probability current density J and substituting it into (8.31), for the particle current j we finally get

$$j = \frac{Lk_BT}{\gamma} \frac{N}{I} \left(1 - e^{-LF/k_BT} \right) \,. \tag{8.32}$$

The normalization constant N can be found by exclusion of J from the relations (8.28) and (8.30).

A principally important result immediately follows from the relation (8.32): the particle current j for $F = 0$ is zero for any periodic potential. The reason for this is that of the two necessary conditions for the appearance of current in the absence of displacing macroscopic forces (gradients)—the thermal disequilibrium and spatial reflective symmetry breaking—we satisfied only the latter. We note that this result was obtained from solution of the Langevin equation without reference to the second law of thermodynamics.

8.3 Ratchet Effect: An Example of Real Realization

In order to explain the ratchet effect—the appearance of gradient-free currents in spatially periodic asymmetric systems far from thermodynamic equilibrium—we use the so-called diffusive ratchet [8], which is a generalization of the

Smoluchowski–Feynman model (8.22), where the temperature of Gaussian white noise $\xi(t)$ is subject to time-dependent periodic perturbation with period τ, i.e.

$$\begin{aligned}\left\langle \xi(t)\xi(t')\right\rangle &= 2\gamma k_B T(t)\delta(t-t')\\ T(t+\tau) &= T(t)\,. \end{aligned} \tag{8.33}$$

Due to the time dependence, the noise $\xi(t)$ is no longer stationary. However, stationariness can be restored if we rewrite (8.21) in the form

$$\gamma\dot{x}(t) = -V'[x(t)] + F + g(t)\hat{\xi}(t)\,, \tag{8.34}$$

where $\hat{\xi}(t)$ is Gaussian white noise

$$\left\langle \hat{\xi}(t)\hat{\xi}(t')\right\rangle = 2\delta(t-t'); \quad g(t) \equiv [\gamma k_B T(t)]^{1/2}\ . \tag{8.35}$$

Such a model is commonly [6] called diffusive ratchet. For numerical calculations we use the following time dependence

$$T(t) = \bar{T}\,[1 + A\ sign\,\{\sin(2\pi t/\tau)\}]\ . \tag{8.36}$$

Considering that the tending of the period of temperature oscillations to infinity implies temperature constancy in any finite time interval, the equality to zero of the particle current becomes evident in the light of the above results. In the transition to an extremely small τ $\left(\tau \ll \gamma^{-1}\right)$, i.e. fast temperature oscillations, it is reasonable to assume that the system does not have time to follow the temperature changes and behaves as in the presence of an average constant temperature

$$T = 1/\tau\int\limits_0^\tau dtT(\tau)\,. \tag{8.37}$$

What is the situation in the intermediate region? Because of temperature oscillations Eq. (8.34) does not allow time-independent solutions. However, time dependence is absent in the particle current asymptotes averaged over the oscillation period; this is the quantity of interest in the case of periodically time-dependent perturbations. Figure 8.4 shows the numerical solution of Eq. (8.34) for long times of the period τ averaged particle current J as function of the force F. The ratchet potential was taken in the form (8.21). The result is significantly different from the one expected in the case of time-independent temperature. Indeed, as according to (8.32) at $F = 0$ for constant temperature the current is absent, then at $F \neq 0$, i.e. for the sloping potential the particles in average should always (at any fixed T) move in the direction of the slope determined by the F sign. At first glance, it seems that with periodic temperature variations the particles must still move on average in the direction determined by the F sign. Despite expectations, we see in Fig. 8.4 that

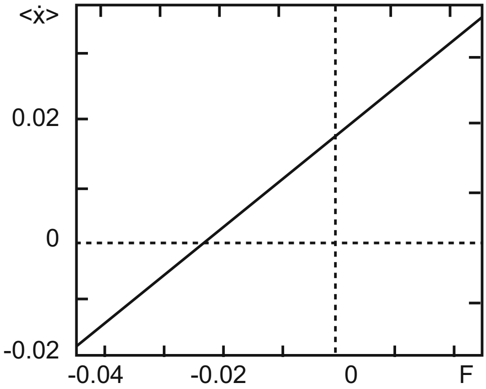


Fig. 8.4 Numerical solution of Eq. (8.34) for long times of the period τ averaged particle current J as function of the force F [6]

there is whole interval of negative F values where the particles on average overcome the static load, move up the potential slope, and execute work against the force F. It is the transformation of the random fluctuation energy into useful work that is called the "ratchet effect." For the names of the experimental setups or theoretical models where this effect is realized, the following synonyms are used: "thermal ratchet," "Brownian motor," "Brownian rectifier," "stochastic ratchet," or simply "ratchet." For a qualitative detection of the ratchet effect it is sufficient to consider the case $F = 0$. Then the ratchet effect is equivalent to appearing of finite particle current at zero static load,

$$j \neq 0 \quad \text{on} \quad F = 0\,. \tag{8.38}$$

We stress that the effect of the appearance of directed motion, obtained in the diffusive ratchet model, does not contradict the second law of thermodynamics, as we can consider the time depending temperature $T(t)$ as being generated by several thermal reservoirs with different temperatures. The specific example of temperature dependence (8.35) models a situation with two thermal equilibrium reservoirs with different temperatures. The fact that such system executes a work is not a miracle but it is also far from non-trivial.

In order to understand the physical mechanism leading to the appearance of the particle current at $F = 0$ for a diffusive ratchet we use a version of dichotomic temperature modulation (8.35). During the first time interval $t \in [\tau/2, \tau]$ the temperature is kept at constant value $\bar{T}(1-A)$, which (with the appropriate choice of A) is much less than the potential barrier ΔV between the neighboring local minima. Therefore, at the end of that time interval, the particles will mostly gather in the vicinity of the local minimum, as it is shown in Fig. 8.5. Then the temperature jumps to the value $\bar{T}(1+A)$, which is considerably higher than ΔV (with appropriate choice of A) and it will remain constant during the next half-period $[\tau, 3/2\tau]$. Because at that time interval, the particles practically do not feel the potential, which is small compared to the perturbing thermal noise, the distribution function will evolve

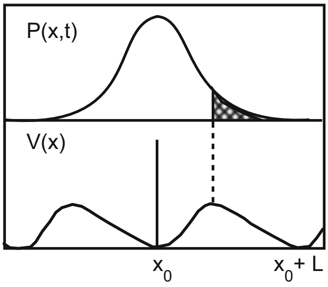


Fig. 8.5 Mechanism of diffusive ratchet function [6]

practically in the same way as at free diffusion (upper part of Fig. 8.5). Finally, the temperature jumps again to the value $\bar{T}(1 - A)$ and the particles will roll down to the nearest potential $V(x)$ minimum. But due to the asymmetry of the potential $V(x)$ the initial population of one minimum will be redistributed asymmetrically and a summary mean displacement will appear after one period τ.

In the case when the potential has one minimum and one maximum on each period L (see Fig. 8.5), and the local minimum is closer to the local maximum from the right, then a positive particle current $j > 0$ will appear, otherwise a negative current will appear. For potentials of more complex geometry, the current direction determination can be no longer evident.

It is expected that the qualitatively analogous behavior will be observed also for more complicated modulation $T(t)$ provided it is sufficiently slow. The effect is relatively rough with respect to the potential shape and it conserves for random (non-deterministic) variations of $T(t)$ [9, 10], i.e. for seldom random switches between the two temperature values and for dynamics in discrete space of states [11]. The ratchet effect also takes place in the case of finite inertia, and with colored (non-white) noise.

Let us now look at the ratchet effect from a somewhat more general point of view. At the first stage, we made sure that in thermal equilibrium there was no preferential motion in the random dynamics regardless of the spatial symmetry breaking. This result is a direct consequence of the second law of thermodynamics, though it was obtained without direct reference to it. At the second stage we considered the diffusive ratchet—a system with broken thermal equilibrium, for which the second law of thermodynamics does not apply. In the absence of that and other restricting reasons, and in the presence of broken spatial symmetry, the appearance of the directed motion of particles does not seem so amazing. Moreover, it can be made natural in light of the so-called Curie principle [6, 12]: if some phenomenon is not forbidden by a symmetry, then it must take place. In other words, the Curie principle postulates the absence of accidental symmetry in a common situation. Accidental symmetry can appear as an exceptional coincidence but not as a typical situation. Any accidental symmetry is structurally unstable and an arbitrarily small perturbation destroys it, while broken symmetry is structurally stable.

8.4 Principal Types of Ratchets

For the basis of ratchet classification we take [6] the minimal generalization of the Smoluchowski–Feynman model (8.22)

$$\gamma\dot{x}(t) = -V'\left[x(t),f(t)\right] + y(t) + F + \xi(t)\,. \tag{8.39}$$

We will assume that

$$V\left[x+L,f(t)\right] = V\left[x,f(t)\right] \tag{8.40}$$

for all time moments t. The functions $y(t)$, $f(t)$ are either periodic or random functions of time (with zero mean). In cases when they represent a random process, we consider it to be statistically independent from both thermal fluctuations $\xi(t)$ and the system state $x(t)$. As before we will neglect the effects of inertia and model the thermal fluctuations by uncorrelated white noise. Generalizing (8.8), we will call the potential $V\left[x,f(t)\right]$ a spatially asymmetric one if there is no Δx such that

$$V[-x,f(t)] = V\left[x+\Delta x,f(t)\right]\ . \tag{8.41}$$

The ratchets described by the model (8.39) can be divided into two classes. The first includes pulsating ratchets, for which $y(t) \equiv 0$, and the second tilting ratchets, for which $f(t) \equiv 0$. A very important subclass of the pulsating ratchets is the so-called fluctuating potential ratchet, for which

$$V\left[x,f(t)\right] = V(x)\left[1+f(t)\right]\ . \tag{8.42}$$

This subclass contains an interesting particular case of on–off ratchets, for which the function $f(x)$ takes only two values ± 1. The state -1 corresponds to the turned off potential. Evidently, the potential in the right-hand side of (8.42) satisfies the condition (8.41) only if the same condition is satisfied by the potential $V(x)$. For the considered class of ratchets the current is exactly zero for spatially symmetric potentials (independently of $f(t)$ properties), and for the spatially asymmetric potentials, the current magnitude is determined by the spatial symmetry breaking degree. We note that the term fluctuating includes both random and periodic functions $f(t)$.

The second subclass of pulsing ratchets includes the traveling potential ratchets, for which

$$V\left[x,f(t)\right] = V\left[x-f(t)\right]\ . \tag{8.43}$$

Those ratchets are interesting because the potential (8.43) never satisfies the symmetry condition (8.41), independently of the fact whether the potential $V(x)$ is spatially symmetric or not. Therefore even symmetric $V(x)$ can be used for current generation.

The third subclass of the pulsing ratchets is described by the model (8.39) with $f(t) \equiv 0,\ y(t) \equiv 0$, but with spatial or time dependence of temperature in the relation

$$\left\langle \xi(t)\xi(t') \right\rangle = 2\gamma T(t \,\mathrm{or}\, x)\delta(t-t')\,. \tag{8.44}$$

Such ratchets are called diffusive. In cases of spatial dependence for the temperature, $T(x)$ is assumed to be a periodic function with the same period L as in the potential $V(x)$. In cases of time dependence, the function $T(t)$ is assumed to be either periodic or random. In the strictest sense, diffusive ratchets are not pulsing ones, for which $f(t) \neq 0$, but they are shown [6] to reduce to the latter.

From the point of view of Einstein relations, a change of the character of the diffusion can be equally due to change in both the temperature and friction coefficient. However, such modification of the Smoluchowski–Feynman model does not break the detailed balance symmetry and therefore there is no ratchet effect in that case [6].

We will briefly discuss the tilting ratchet ($f(t) = 0,\ y(t) \neq 0$). In this case $V[x,f(t)] = V(x)$. In the case of the potential $V(x)$ with broken spatial symmetry we can restrict ourselves to the case of symmetric function $y(t)$. Under symmetric function we will understand a periodic one $y(t)$, for which there is such Δt, that

$$-y(t) = y(t + \Delta t) \tag{8.45}$$

for all t. If $y(t)$ represents a random process, then we will call it symmetric if the statistical properties of the processes $y(t)$ and $-y(t)$ coincide. In the case when the driving force $y(t)$ is a random process, the ratchet is called a fluctuating force ratchet, and in the case of periodic driving, swinging [13]. If, in the case of periodic excitement, we drop thermal noise into the Langevin equation, then we obtain a so-called deterministic ratchet [14, 15].

In the case of symmetric potential $V(x)$ for the occurrence of the finite current, generally speaking, is sufficient to break the symmetry of $y(t)$. The term asymmetric tilting ratchet is usually applied in a case when $y(t)$ is a non-symmetric function, regardless of the fact whether it is periodic or random, and whether the potential $V(x)$ is symmetric or not.

Of course, the presented classification does not exhaust all conceivable ratchets. Let us note in particular a curious class of ratchet not included in the considered classification—the so-called supersymmetric ratchets [6, 16]. In such systems, the ratchet effect is absent even at deviation from the thermal equilibrium and under broken spatial (or time) symmetry. In particular, in such systems the average stationary current is zero for any choice of friction coefficient and time dependencies of temperature and of external forces $f(t), y(t)$. The term "super-symmetry" is due to the fact that the unperturbed system $f(t) = y(t) = F = 0$ (8.39) at a certain choice of the potential $V(x)$ can be described in terms of supersymmetric quantum mechanics [17]. The usefulness of this connection lies in the possibility to transform the Fokker–Planck equation into the Schrödinger equation in imaginary time, and further to use the powerful arsenal of supersymmetric quantum mechanics.

To conclude this section, we will discuss what physical situations are covered by the model (8.39) and the ratchet classification based on it. They are so diverse that their systematic enumeration is senseless. Therefore we will limit ourselves to a few remarks.

The stochastic process in (8.39) has the space of states in form of the real axis and for simplicity is often called a Brownian particle. In some cases $x(t)$ actually represents the position of a real physical particle, in other cases, very diverse collective degrees of freedom or slow variable states. It can be a chemical reaction coordinate, fission degree of freedom of a nucleus, ratchet position with respect to the pawl, or a Josephson phase in super-conducting contacts. The conditions of potential periodicity are critical and can be connected either with real spacial periodicity or to be due to the phase origin of the corresponding variable. Neglecting the inertia term is typical for mesoscopic systems. However, there are times when it is important to take into account inertia in order to describe the experimental situation correctly.

One more important feature of the model (8.39) is the method of description of the interaction with the thermal environment. The version of such interaction used with the help of local friction force and white Gaussian noise represents just the simplest possibility. Its conformity to a concrete physical system must be checked in every specific case. In particular, the memory-free friction mechanism and uncorrelated fluctuations are consequences of the assumption about the possibility of dividing into characteristic time scales the fast degrees of freedom of the thermal reservoir and relatively slow system dynamics $x(t)$. The x-independent system-reservoir coupling constant γ is natural if the periodic potential and the thermal environment have a different physical origin, but x-dependent friction models are also considered.

Applied perturbation may either act immediately on the variable of state x or change the periodic potential. In the former case, perturbation acts as a normal force, i.e. the system receives (or loses) energy in displacement on one spatial period. The experimental realization of such situation, according to the above classification, will represent an asymmetric tilting ratchet. In the latter case, we are dealing with pulsing ratchets. For example, the electric field can change the internal charge distribution of a neutral Brownian particle or of the periodic substrate which it interacts with.

The ratchet effect has a rich history. Some of its aspects were already present in the works of Archimedes, Maxwell, and Curie, but only after the generalization of the Smoluchowski model by Feynman did the problem of thermal fluctuation rectification appear at the center of physicists' attention. The newest history of the effect was initiated by attempts to solve the intracellular transport problem [2]. Between the "ancient" and "newest" epochs of the effect's investigation there was a very important and fruitful period. More than 30 years ago, the ratchet effect was experimentally observed in the form of the so-called photo-galvanic effect [18, 19]. After that the directed transport induced by non-displacing time dependent perturbations in periodic structures with broken symmetry was the subject of several hundred experimental and theoretical works during the 1970s. The complete theory was created by Sturman, and Fradkin [20] and Belinicher and Sturman [21]. In those works the authors developed a theory of generation of current in crystals without the symmetry center under the influence of homogeneous illumination, i.e. the ratchet theory was built for a specific case.

Electrical current in a medium is usually generated either by applied fields (electrical and magnetic) or by spatial inhomogeneities (temperature and illumination gradients). Along with that, in thermodynamically non-equilibrium conditions currents of different natures are possible, connected to the absence of symmetry center in the medium. One such effect is the photo-galvanic effect, the appearance of direct electric current in homogeneous crystals without a symmetry center under the influence of homogeneous illumination. In experiments performed on the ferroelectric $LiNbO_3$ and $BaTiO_3$ crystals, the constant photocurrents $j \sim 10^{-10}\,\mathrm{A/sm^2}$ and photo-tension considerably exceeding the forbidden gap width were observed. The influence of transition processes caused by the crystal heating by light or by relaxation processes was excluded due to long observation of the currents.

Later, it became clear that the photo-galvanic effect is possible without exception in all media, which are not symmetric with respect to spatial inversion. Besides crystals without a symmetry center, these also include isotropic media—gases and liquids—containing molecules with natural optical activity. Non-equilibrium may originate not only in light, but also sound, particle flows, etc. It also refers to non-stationary phenomena not supported by external sources. Any process of relaxation to the thermodynamic equilibrium in media without a symmetry center must be accompanied by current.

Let us give a simple example of particle flow appearing in a medium without a symmetry center in the presence of inhomogeneity. We consider a gas of non-interacting particle scattering on randomly situated and identically oriented wedges. Evidently, such a medium does not have a symmetry center. In the absence of external actions, as the result of impacts the isotropic particle velocities distribution sets up, because at elastic scattering from any convex body the spherically symmetric particle distribution remains spherically symmetric. Let the non-equilibrium source be the alternating field $e\mathbf{E}\cos\omega t$. Its action supports the non-equilibrium distribution of anisotropic particle velocities, because it increases the fraction of particles along and against the field $\mathbf{E}$. As it can be seen from Fig. 8.6, the scattering of the particles moving vertically after the scattering on the wedge leads to a collective flow directed to the left. Another simple mechanism of the diffusive flow anisotropy appearing with fluctuations of the diffusion coefficient is discussed in the paper [22].

It interesting to note that it is already more than half a century that an effect is known which is related to the photo-galvanic effect in its nature. In 1956 Lee and Yang [23] supposed that in weak interactions parity is not conserved, i.e. our space, being isotropic, does have a symmetry center. As a result of that the electrons appeared at β-decay, $n \to p + e + \nu$, in the presence of magnetic field $\mathbf{H}$ must have

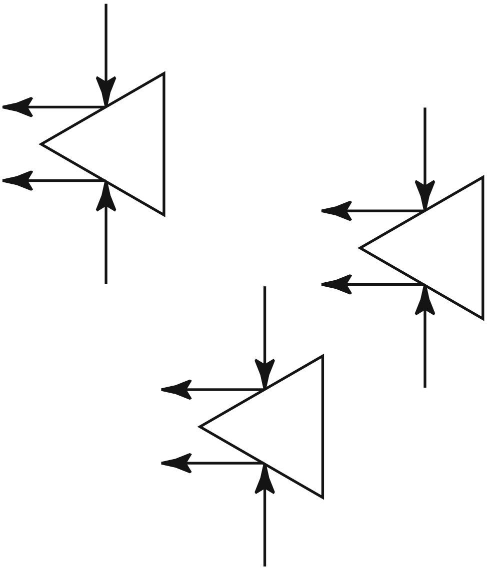


Fig. 8.6 The simplest example of particle flow appearing in a medium without a symmetry center in the presence of non-equilibrium

asymmetric distribution, i.e. in the direction $\eta \mathbf{H}$ (η is a pseudoscalar) a current must appear. Current magnitude is determined by the degree of parity non-conservation in weak interactions.

8.5 Nonlinear Friction as the Mechanism of Directed Motion Generation

Let us consider one more example that is beyond the simple classification considered in the previous section: the motion of an object under the action of a random force $F(t)$ with zero mean in a medium with friction [24]. We will assume that the friction coefficient depends on the direction of the object's motion. Such dependence can be induced in the simplest case by the particle shape or its conformations. Such conformation transitions in the considered model are not connected to the presence of an internal "program" of transitions, but are induced by the influence of the external medium. We will use the term "umbrella" for objects with this peculiarity (see Fig. 8.7). For an umbrella under the action of forces with zero mean one can pose a question about the existence of a non-zero current even in absence of potential: the reflective symmetry is already broken on the level of the friction coefficient. This model can be called a non-potential ratchet with nonlinear friction.

The umbrella's equation of motion under action of the external random force $F(t)$ can be written in the form

$$m\ddot{x} = F(t) - \alpha(\dot{x})\dot{x}\,, \tag{8.46}$$

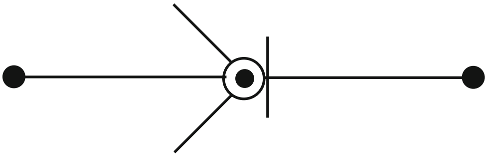


Fig. 8.7 Umbrella—an object composed from an axis, two movable blades and stopper

where α is the direction-dependent friction coefficient. The simplest model of such dependence can be obtained if we consider that when moving in the positive direction of the x axis the umbrella closes and the friction coefficient is α_1. When the umbrella moves in the opposite direction, it opens and the friction coefficient grows up to α_2. Such dependence can be described by the expression

$$\alpha = \frac{1}{2}(\alpha_1 + \alpha_2) + \frac{1}{2}(\alpha_1 - \alpha_2)\frac{\dot{x}}{|\dot{x}|} \tag{8.47}$$

Writing the original Eq. (8.46) in terms of velocity $\dot{x} = v$, we get

$$m\dot{v} = F(t) - \frac{1}{2}\left((\alpha_1 + \alpha_2)\, v + (\alpha_1 - \alpha_2)\,|v|\right) . \tag{8.48}$$

The simplifying assumption in this model is that the umbrella opens and closes instantly. Of course, that approximation is reasonable, if the characteristic time scales of the random force T and of the umbrella τ satisfy the inequality $T \gg \tau$. More detailed models can be built using the continuous velocity dependencies of the friction coefficient, for example

$$\alpha(v) = \alpha_1 + \frac{\alpha_2 - \alpha_1}{1 + \exp(\beta v)} , \tag{8.49}$$

where β is the characteristic scale of transition between the two friction coefficient values.

Let us now discuss the mechanism of the appearance of directed motion under the action of a zero mean random force. Its nature is in the breakdown of the equations of motion symmetry with respect to the replacement $v \to -v$. In order to estimate the appearing directed velocity we use the following simple considerations. We average Eq. (8.48) over the random force. The stationary state is determined by the equality

$$\langle v \rangle = \frac{\alpha_2 - \alpha_1}{\alpha_2 + \alpha_1} \langle |v| \rangle . \tag{8.50}$$

The average value of the velocity module we estimate through its quadratic moment as

$$\langle |v| \rangle \simeq \left\langle v^2 \right\rangle^{1/2} . \tag{8.51}$$

From the energy balance in the stationary state we estimate the pair correlator of the velocity

$$\langle vF \rangle = \frac{1}{2}(\alpha_1 + \alpha_2)\left\langle v^2 \right\rangle . \tag{8.52}$$

Now we express the correlator in the left-hand side of (8.52) through the correlation characteristics of the force. It can be easily done from the estimate

$$v \simeq \frac{1}{m}\int F(\tau)d\tau . \tag{8.53}$$

Then

$$\langle vF \rangle = \frac{1}{m}\int \langle F(t)F(\tau)\rangle \, d\tau \approx \frac{1}{m}\left\langle F^2 \right\rangle \tau_c . \tag{8.54}$$

Using (8.50)–(8.53) we finally get

$$\langle v \rangle \simeq \frac{(\alpha_2 - \alpha_1)}{(\alpha_2 + \alpha_1)}\left(\frac{2\tau_c \left\langle F^2 \right\rangle}{m(\alpha_2 + \alpha_1)}\right)^{1/2} . \tag{8.55}$$

Thus a directed motion under action of fluctuational random force with zero mean appears. The magnitude and direction of the velocity are determined by the difference $(\alpha_2 - \alpha_1)$. Of course such simple estimates demonstrate only the possibility of the directed motion appearing and they require more accurate substantiation. In order to do that we will now obtain the kinetic equation for the umbrella's velocity distribution function.

In order to derive the kinetic equation we take the usual definition of the distribution function

$$f(V,t) = \langle \delta(V - v(t))\rangle . \tag{8.56}$$

Here $v(t)$ is the solution of the dynamical equation with the random force and the averaging is performed over all its realizations. Time differentiating $f(V,t)$ and using the equations of motion, we get

$$\frac{\partial f}{\partial t} = -\frac{1}{m}\frac{\partial}{\partial V}\left(\langle F(t)\delta(V - v(t))\rangle - \alpha(V)Vf\right) \tag{8.57}$$

or in more convenient form

$$\frac{\partial f}{\partial t} - \frac{\partial}{\partial V}\left(\frac{\alpha(V)}{m}Vf\right) = -\frac{1}{m}\frac{\partial}{\partial V}\langle F(t)\delta(V - v(t))\rangle . \tag{8.58}$$

Thus the problem is reduced to averaging of the term in the right-hand side and its expression through the distribution function. Using standard methods (see, for example, [7, 25–28]), it is easy to obtain the closed equation for the distribution function in the case of Gaussian δ-correlated force. That equation reads

$$\frac{\partial f}{\partial t} - \frac{\partial}{\partial V}\left(\frac{\alpha(V)}{m} Vf\right) = \frac{\left\langle F^2 \right\rangle \tau_c}{2m^2} \frac{\partial^2 f}{\partial V^2} \tag{8.59}$$

where we used the notion (8.53). The stationary solution of that equation corresponding to absence of flow in the velocity space is easy to obtain for the two-level model $\alpha(V)$ (8.47)

$$\begin{aligned} f(V) &= C \exp\left\{ -\frac{2m}{\langle F^2 \rangle \tau_c} \int V\alpha(V) dV \right\} \\ &= C \exp\left\{ -\frac{mV^2}{\langle F^2 \rangle \tau_c} \left(\alpha_2 + (\alpha_1 - \alpha_2)\, \Theta(V)\right) \right\}\ . \end{aligned} \tag{8.60}$$

The constant C is determined from the normalization condition for the distribution function

$$C = \left(\frac{m\alpha_1 \alpha_2}{\pi \langle F^2 \rangle \tau_c} \right)^{1/2} \frac{1}{\sqrt{\alpha_1} + \sqrt{\alpha_2}}\,. \tag{8.61}$$

Using the found equilibrium distribution function we obtain the average velocity of the umbrella motion in the form

$$\langle V \rangle = \int_{-\infty}^{\infty} Vf(V)dV = \left(\frac{\left\langle F^2 \right\rangle \tau_c}{\pi m \alpha_1 \alpha_2} \right)^{1/2} \frac{\alpha_2 - \alpha_1}{\sqrt{\alpha_2} + \sqrt{\alpha_1}}\,. \tag{8.62}$$

The velocity value differs from the above estimate (8.55) by a more complicated dependence on the two friction components.

Let us now estimate the efficiency of this method of generating the directed motion with respect to its average energy

$$\xi \equiv \frac{\langle V \rangle^2}{\langle V^2 \rangle} = \frac{2\left(\alpha_2 - \alpha_1\right)^2}{\pi \left(\sqrt{\alpha_1} + \sqrt{\alpha_2}\right)\left(\alpha_1^{3/2} + \alpha_2^{3/2}\right)};\ \text{for}\ \alpha_2 \gg \alpha_1 = \frac{2}{\pi}\,. \tag{8.63}$$

The maximal efficiency coefficient achievable for the described motion is

$$\eta = \frac{\langle V \rangle^2}{\langle V^2 \rangle + \langle V \rangle^2} = \frac{\xi}{1+\xi}; \quad \text{for}\ \alpha_2 \gg \alpha_1 = \frac{2}{\pi + 2}\,. \tag{8.64}$$

As we can see, the efficiency of the directed motion generation is relatively high. One can make sure of this, having considered the umbrella's motion in a constant gravitational field. The constant gravity force acts in the negative direction of the x axis. Then the kinetic equation describing the umbrella travel takes the following form

$$\frac{\partial f}{\partial t} - \frac{\partial}{\partial V}\left(gf + \frac{\alpha(V)}{m} Vf \right) = \frac{\left\langle F^2 \right\rangle \tau_c}{2m^2} \frac{\partial^2}{\partial V^2} \,. \tag{8.65}$$

The stationary solution of that equation like in the previous case is easily found

$$\begin{aligned} f(V) &= C \exp\left\{ -\frac{2m}{\langle F^2 \rangle \tau_c} \int (gm + V\alpha(V))\, dV \right\} \\ &= C \exp\left\{ -\frac{2gm^2 V}{\langle F^2 \rangle \tau_c} - \frac{mV^2}{\langle F^2 \rangle \tau_c} (\alpha_2 + (\alpha_1 - \alpha_2)\, \Theta(V)) \right\} \,. \end{aligned} \tag{8.66}$$

The normalization constant C is determined from the normalization condition for the distribution function f,

$$\frac{C}{q}(I_1 + I_2) = 1 \,. \tag{8.67}$$

Here $q = 2gm^2 / \left\langle F^2 \right\rangle \tau_c > 0$, a I_i $(i = 1, 2)$ are defined as

$$I_i = \sqrt{\pi} \beta_i e^{\beta_i^2} \operatorname{erfc}\left((-1)^{i+1} \beta_i \right) \,. \tag{8.68}$$

The parameters β_i $(i = 1, 2)$ are equal to

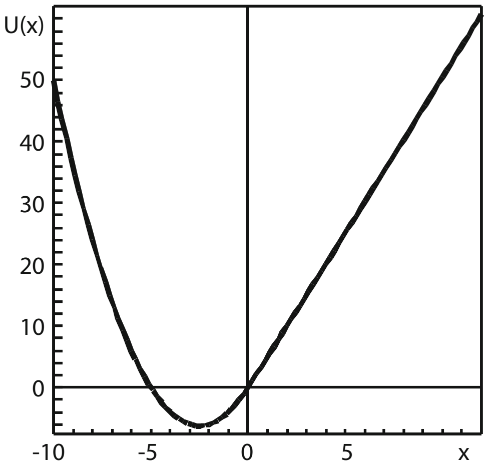
$$\beta_i = \frac{m^{3/2} g}{\sqrt{\alpha_i \langle F^2 \rangle \tau_c}} \,. \tag{8.69}$$

The criterion for the appearance of the directed motion in the positive direction of the x axis can be written in the form

$$\langle V \rangle = \frac{-2}{q(I_1 + I_2)} \left(\beta_1^2 I_1 + \beta_2^2 I_2 + \beta_2^2 - \beta_1^2 \right) > 0 \,. \tag{8.70}$$

It is easy to make sure that there is a region of the parameters where that condition is satisfied. The condition can be qualitatively understood from a simple physical interpretation. The Langevin equation (8.48), which is used to derive the kinetic equation, is equivalent to the over-damped case of the usual Langevin equation. One can easily make sure of that having made the replacement $V \to x,\ m \to \alpha$. Therefore we can describe the dynamics of the system under consideration using

Fig. 8.8 The effective potential



the over-damped Langevin equation philosophy. In essence, in this interpretation the particle performs finite motion in the effective potential

$$U(x) = mgx + \frac{x^2}{2}\left(\alpha_2 + (\alpha_1 - \alpha_2)\,\Theta(x)\right) . \tag{8.71}$$

Obviously, the global minimum of that potential is situated in the negative x region (Fig. 8.8) and therefore the average value of the system's position at low energies is negative. Returning to the initial variables, we find that the umbrella has negative average velocity, i.e. it falls under the action of the force of gravity. However, at high energies the situation changes. The average position of the particle may become positive due to the different asymptotic behavior of the potential ($\alpha_1 < \alpha_2$). It is easy to show that if the energy level is

$$E \geqslant \frac{4(mg)^2\,(\alpha_1 + \alpha_2)}{(\alpha_1 - \alpha_2)^2} , \tag{8.72}$$

then the average position is positive. In the over-damped case the energy level is determined by the level of the pair correlator for the external random force. In our case it would mean that provided the condition

$$\frac{\left\langle F^2 \right\rangle \tau_c}{2m} \geqslant \frac{4(mg)^2\,(\alpha_1 + \alpha_2)}{(\alpha_1 - \alpha_2)^2} , \tag{8.73}$$

the umbrella will move in the positive direction against the gravity force.

The obtained criterion is rather rough as it does not account for distinctions in the characteristic times of motion in the regions $x > 0$ and $x < 0$. It is easy to understand that accounting for that effect will lead to a motion regime against

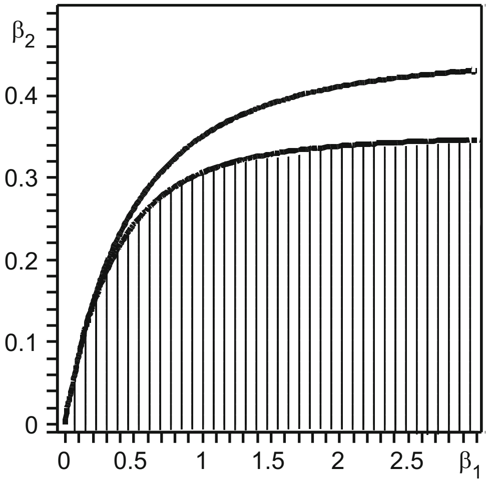


Fig. 8.9 Distinctions between the exact and the rough criteria

gravity force at greater β_2, than the inequality (8.73) gives. This is confirmed by the exact criterion which follows from the condition (8.70). Figure 8.9 shows the distinctions between the exact and the rough criteria. Thus the efficiency of this motion method is sufficient to overcome the counteraction of constant forces. In a certain sense, such mechanisms for directed motion generation can be observed for many biological objects both of micro and macro-sizes. In the macro case the role of random forces is played by periodic forces due to the retraction and extrusion of a medium (a liquid). This method is used by many inhabitants of an aquatic environment, such as jellyfish. It should be stressed that there is no need to exceed the forces on the liquid retraction stage over the extrusion stage. The period average of those forces may be and must be zero. Molecules with special asymmetry can travel in random external fields using analogous mechanisms.

8.6 Change of Current Direction in the Deterministic Ratchet

The ratchet problem is closely related to the problem of deterministic particle dynamics in a periodic spatially asymmetric potential (the ratchet potential). According to the classification considered in Sect. 8.4, this system belongs to the class of deterministic ratchets ($f = 0$, $\xi = 0$). If the inertia term is taken into account, then the particle dynamics in certain regions of parameter space can be chaotic, and this leads in turn to the modification of the transport properties of the system. In particular, it appears that there are strict correlations between the structure of the system's bifurcation diagram and the direction of the induced current [15].

Let us consider the problem of motion of the particle of mass m under the action of a periodic time-dependent force with zero mean in a spatially asymmetric potential. The equation of motion in the simplest periodic time dependence takes the form

$$m\ddot{x} + \gamma\dot{x} = -\frac{dV(x)}{dx} + F_0 \cos \Omega t\,, \tag{8.74}$$

where F_0 and Ω are amplitude and frequency of the external force, respectively. We choose the ratchet potential in the form

$$V(x) = V_1 - V_0 \sin\left[\frac{2\pi(x-x_0)}{L}\right] - \frac{V_0}{4}\sin\left[\frac{4\pi(x-x_0)}{L}\right], \tag{8.75}$$

where L is the spatial period of the potential, V_0, V_1 are some constants. The shift on x_0 provides the minimum position in the origin.

Let us introduce the following dimensionless variables

$$x' = x/L,\ t' = \omega_0 t,\ \omega = \Omega/\omega_0,\ a = F_0/mL\omega_0^2,\ b = \gamma/m\omega_0\,. \tag{8.76}$$

Here $\omega_0 = 4\pi^2 V_0\delta/mL^2$ is the frequency of linearized motion in the vicinity of the potential minimum, $\delta = \sin 2\pi\left|x_0'\right| + \sin 4\pi\left|x_0'\right|$. The equation of motion in the dimensionless variable (we will further drop the primes) reads

$$\ddot{x} + b\dot{x} = -\frac{dV(x)}{dx} + a\cos\omega t\,, \tag{8.77}$$

where the dimensionless potential is

$$V(x) = C - \frac{1}{4\pi^2\delta}\left[\sin 2\pi(x-x_0) + 0.25\sin 4\pi(x-x_0)\right]\,. \tag{8.78}$$

The constant C is chosen from the condition $V(0) = 0$. The parameter values that satisfy the above conditions are the following:

$$x_0 \simeq -0.19,\ \delta \simeq 1.6,\ C \simeq 0.0173\,.$$

Equation (8.77) contains three dimensionless parameters a, b, ω, and each of them is determined in terms of the initial physical parameters. This is a nonlinear differential equation with explicit time dependence and thus it allows both regular and chaotic regimes. In cases where the inertia term $m\ddot{x}$ is neglected, chaotic regimes are absent.

We now turn to the numerical investigation of current appearing in the system (8.77), i.e. in the deterministic ratchet. We use the definition of current as the time average of particle velocity averaged over the ensemble of initial conditions. This definition includes two different averages. The first one is the averaging over the M initial conditions that we will assume to be uniformly distributed on some interval around the origin with zero initial velocity. For the fixed time moment t_j we get for the average velocity v_j the following expression

$$v_j = \frac{1}{M} \sum_{j=1}^{M} \dot{x}_i(t_j)\,. \tag{8.79}$$

Having time averaged that quantity [over the discrete time used for numerical solving of Eq. (8.77)], we find for the current

$$j = \frac{1}{N} \sum_{j=1}^{N} v_j\,. \tag{8.80}$$

Here N is the finite set of different times t_j.

For the fixed set of parameters the current j is uniquely determined by the expression (8.80). We intend to follow how it changes with the variation of parameters. In particular, let us consider a case when the parameter a changes at fixed values of b and ω. Using the definition of the parameter ω_0 we can present a in the form

$$a = \frac{1}{4\pi^2\delta} \frac{F_0}{(V_0/L)}\,. \tag{8.81}$$

Up to a constant factor, parameter a represents the ratio of external force amplitude to the average force due to the potential $V(x)$. As we have already mentioned, the system's (8.77) dynamics include both regular and chaotic regimes. One can make sure of that having considered either the stroboscopic Poincaré section or the bifurcation diagram. Figure 8.10a gives the bifurcation diagram for $b = 0.1,\ \omega = 0.67$ and for the small interval $a = [0.072, 0.086]$. From the diagram one can see that the transition to chaos takes place according to a somewhat modified period doubling scenario. In particular, after the bifurcation at the critical value $a_c \simeq 0.0809$ a window appears with the period-4 periodic orbits. Figure 8.10b presents the current j (8.80) as a function of the parameter a in the same region. We see that value a_c, at which the chaos-regularity bifurcation takes place, exactly coincides with that a value at which the induced current direction change occurs.

Figure 8.10b shows only a small region of the parameter a variation where the first change of the current direction takes place. With further variation a the current direction changes multiple times [14].

In order to better understand the nature of the change in current direction let us study the structure of individual orbits below and above the critical point (point of bifurcation). Figure 8.11 presents the time dependencies of the particle coordinate for the values $a = 0.074$ and $a = 0.081$, corresponding to the two periodic windows on the bifurcation diagram (Fig. 8.10). The first trajectory (period-2 orbit) corresponds to current in the positive direction of the x axis, the second one (period-4 orbit) corresponds to current in the negative direction, in which the ratchet potential slope is greater. In the latter case, the advancing mechanism is interesting:

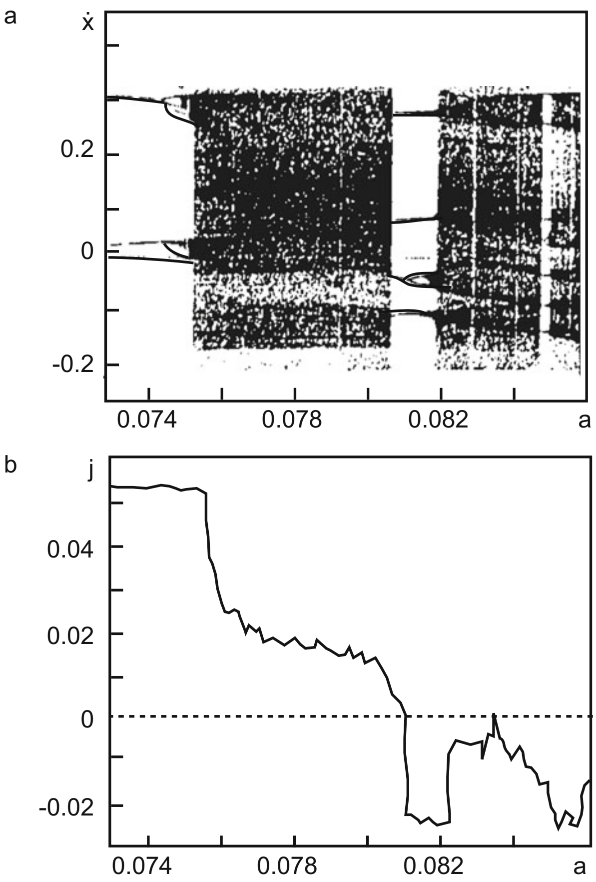


Fig. 8.10 (**a**) The bifurcation diagram for Eq. (6.75) solutions $\dot{x}(a)$ for the interval $a = [0.072, 0.086]$, $b = 0.1$, $w = 0.671$. (**b**) Current j as function of the parameter a [15]

in order to make one step to the left, the particle makes a step to the right and then two steps to the left. As a result, current in the negative direction appears.

Let us now consider a typical trajectory in the region immediately below a_c (Fig. 8.12). The particle starting with zero velocity from one of minima of the ratchet potential chaotically travels in another minimum to the left or to the right. For some time, the particle is trapped by the potential minima and there undergoes oscillating motion. Then there occurs a transfer to the running mode, corresponding to motion in the negative direction. In terms of velocity, these running modes correspond to the above periodic. The phenomenology of such dynamics can be described in the following way. For values of $a > a_c$ the system attractor represents the period-4. In the region of values of parameter a slightly lower than a_c, the attractor becomes chaotic. However, irregardless of that, there are relatively long time intervals during which the trajectory is close to the periodic orbit from the region $a > a_c$. Those regular (almost periodic) intervals suddenly give way to the finite duration intervals at which the system trajectory behaves chaotically. In other words, in this case we

Fig. 8.11 The particle trajectories realizing the currents in different directions (for the same parameter values as in Fig. 8.10): (**a**) $a = 0.074$, current in the positive direction; (**b**) $a = 0.081$, current in the negative direction [15]

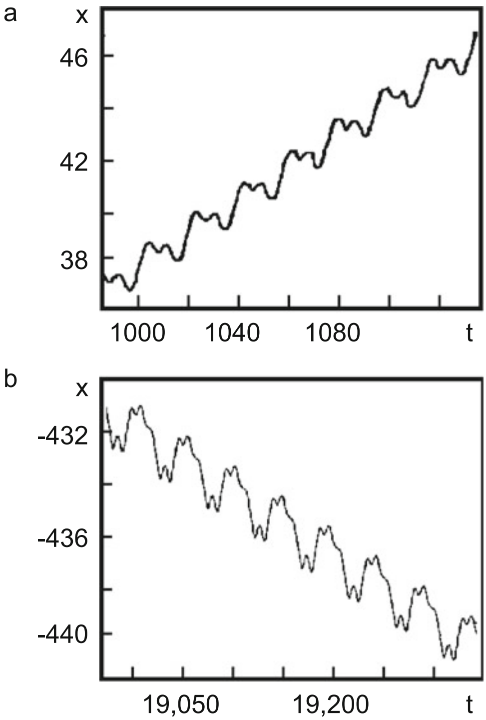
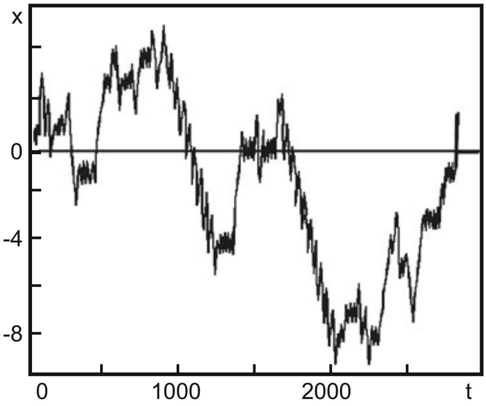


Fig. 8.12 The intermittency effect for Eq. (8.77) solutions in the immediate vicinity of the regularity-chaos bifurcation a_c ($b = 0.1$, $w = 0.67$) [15]



are dealing with an intermittency picture typical for nonlinear dynamics [29]. As a approaches to a_c the regular motion intervals continuously grow and at last for $a \geqslant a_c$ motion becomes purely regular.

8.7 Bio or Molecular Motors

In the last section of this chapter we will discuss one of the applications of the "flux without gradients" concept, explained above—-so-called biological motors [30, 31]. Technical terminology is distinctive for this branch of research that lies on the boundary between physics and modern biology. For example, the terms "channels" and "pumps" denote protein aggregates, that provide transport for correspondingly, passive and active ions through biological membranes. The term "molecular motor" or "bio-motor" refers to proteins or protein complexes that transform chemical energy into mechanical work. More concretely, we will understand a molecular motor as a macroscopic object that carries out directional motion along one-dimensional periodic structures. Why do bio-motors play so important part in maintenance of vital functions of living matter? The most primitive cells do not have a nucleus. Their entrails consist of unstructured broth. Such cells are very small and intracellular transport in them can be provided through thermal diffusion. In contrast, cells that form any multicellular organisms are not only more organized, but also larger. Due to this, passive diffusive transport becomes insufficient: when cell scale grows by 20 times, diffusion slows down 400 times. The distinctive feature of the structure of such cells is the existence of the nucleus, responsible for the storage and duplication of genetic information, and a net of filaments that connect the different parts of the cell. These filaments radially disperse from the nucleus to the periphery of the cell. For our purposes, we need only to know that filaments are periodic and fairly rigid structures with a period of the order of 10 nm. They have moreover polarity, so that one can define a "positive" and a "negative" extremity. Let us note that bio-motors of definite type always move in the same direction, which is determined by motor type and filament polarity. Apart from some additional aims that are beyond our scope, the filament system realizes metabolic processes between different parts of the cell. Now we will try to apply the concept of noise-induced transport considered earlier in this chapter to explain the functioning of bio-motors. Let us consider an isothermal reaction in the presence of a catalyst. In the simplest case, this reaction could be described by one reaction coordinate that cyclically passes through a set of chemical states. An adequate model is a Brownian particle under the action of thermal fluctuations in periodic potential. The local minimum represents some chosen chemical state, while passing through the chemical cycle is modeled as a displacement of the reaction coordinate on one spatial period. The full cycle in one direction means that all existing molecules were transformed into reaction products as a result of the catalytic reaction. Passing through the cycle in reverse corresponds to the reverse reaction. With reference to the case of interest, the situation looks as follows [31]. In the first step, the "filling" of bio-motor M is carried out, when the organic complex adenosine triphosphate (ATP) joins it. In all living organisms this complex acts the part of a universal accumulator of energy. Bio-motors obtain energy from the degradation of ATP. The energy (about 12 kT) is stored in a phosphate bond and is released when this bond is broken, to form adenosine diphosphate (ADP) and inorganic phosphate (P). The motor continuously breaks ATP in never-ending cycle: M → M × ATP → M × ADP × P → M × ADP → M. Although this cycle is most common, in general, different motors could function with the use of different cycles. Biologists carry out careful experiments to clarify the details of every cycle. Notice that we discussed only the problem of the fuel for the bio-motor, but did not yet touch upon the main problem: by what mechanism is released energy transformed into directional motion. At first glance, the solution to the problem could be connected to the existence of a temperature gradient along the filament, on which bio-motors move. However, all temperature heterogeneity in the cell on a scale of a few tens of nanometers decay on time scales of microseconds. This tiny scale is much smaller than the characteristic times of the chemical reactions that carry out motor filling considered above. So, we are again faced with the problem of the generation of directional motion without temperature or force field gradients.

Thus, let us consider [30] the molecular motor as a Brownian particle in local thermal equilibrium with fixed temperature T. The action of the motor is generated by generalized forces of two types. The first type f_{ext} is connected to the mechanical interaction of the motor with the surrounding environment and, in particular, involves the forces of viscous friction. The second type $\Delta\mu$ is generated by the difference of chemical potential, which is equal to free-energy per consumed "fuel" molecule. The chemical potential difference $\Delta\mu$ for the process ATP → ADP + P given by

$$\Delta\mu = \mu(\text{ATP}) - \mu(\text{ADP}) - \mu(\text{P})\,. \tag{8.82}$$

At chemical equilibrium $\Delta\mu = 0$, whereas it is positive when ATP in excess and negative when ADP is in excess.

The action of generalized forces leads to fuel consumption and the motion of the motor. It is useful to introduce generalized currents to describe these effects: the average rate of consumption of fuel molecules r (i.e., the average number of ATP molecules hydrolyzed per unit time, per motor) and the average velocity of the motor's displacement v. The dependencies $v(f_{\text{ext}}, \Delta\mu)$ and $r(f_{\text{ext}}, \Delta\mu)$ are in general nonlinear, since bio-motors often function far from thermal equilibrium ($\Delta\mu \simeq 10\,\text{kT}$). First let us consider the linear regime($\Delta\mu \ll kT$). In this regime linear response theory [30, 32] allows us to write:

$$\begin{aligned} v &= \lambda_{11} f_{\text{ext}} + \lambda_{12}\Delta\mu \\ r &= \lambda_{21} f_{\text{ext}} + \lambda_{22}\Delta\mu\,. \end{aligned} \tag{8.83}$$

Here λ_{11} is the mobility coefficient, λ_{12} and λ_{21} are the coefficients of mechanochemical coupling coefficient, λ_{22} is a generalized mobility relating ATP consumption and chemical potential difference. According to the Onsager principle of the symmetry of kinetic coefficients $\lambda_{12} = \lambda_{21}$. Whenever $f_{\text{ext}} v < 0$, some work is performed by the motor; whenever $r\Delta\mu < 0$, chemical energy is generated. A given motor/filament system can work in eight different regimes. However, only four regimes are of interest: when mechanical work and chemical energy have different signs. If both $f_{\text{ext}}v$ and $r\Delta\mu$ are positive, there is no energy output from the system; all work performed at the system is simply dissipated in the thermal bath. Cases for which both $f_{\text{ext}}v$ and $r\Delta\mu$ are negative are forbidden by the second law of thermodynamics. We shall now discuss a concrete model for the motion of bio-motors. We restrict ourselves to the consideration of the so-called two-state model [33]. In this model, energy consumption by a bio-motor leads to coordinated transitions between states 1 and 2. These transitions could be described in terms of chemical kinetics. For each of the states the one-dimensional potential $V_i(x)$ $(i = 1, 2)$ is introduced (for systems with a variable number of particles it is more exact to talk about free-energy), where x is a coordinate of the motor's center of mass. It is supposed that potential is periodic and spatially asymmetric (the symmetry of the potential reflects the symmetry of the filament). Although the exact form of the potential is unknown, experiments indicate that its period is about 10 nm [34]. According to the classification represented in Sect. 8.4, this is a pulsating ratchet. The mechanism of current generation within it is quite simple. Let us consider a case when $V_1(x)$ has a saw-tooth potential with spatial asymmetry, and $V_2(x) = \text{const}$. When $k_BT << \Delta V$ (ΔV is high in potential), the dynamics of the motor splits into two stages: diffusion in potential $V_1(x)$, leading to the narrowing of the diffusion bell, and transition to an exited state due to the breaking of the phosphate bond, when dynamics of the diffusion bell are reduced to free symmetric broadening. Due to the spatial asymmetry of the potential, the bio-motor could transit to the neighboring right minimum with the probability proportional to the stroked area of the probability density P. The value of the induced current depends on the relation between times τ_1 and τ_2 of work of potentials V_1 and V_2. Optimal time could be defined as a narrow diffusive peak formation time, which is approximately equal to the time of the rolling down from the top of the inclined plane to the potential minimum. This time is defined by parameters of the potential and mass of the bio-motor. We can also evaluate the upper boundary of the time τ_2: the diffusive bell has not to become blurred so much that the reverse current appears. More complicated models of molecular motors are discussed in reviews [6, 30] (Fig. 8.13).

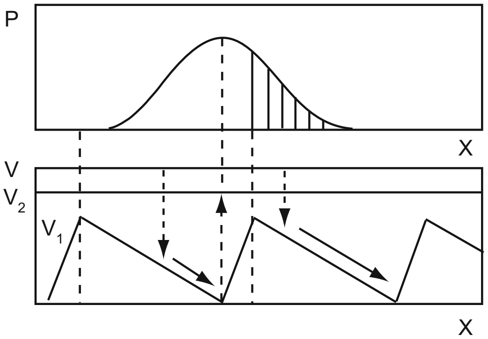


Fig. 8.13 The basic working mechanism of a pulsating ratchet [30]

References

1. Clausius, R.: Ann. Phys. **79**, 368–397 (1850)
2. Magnasco, M.: Phys. Rev. Lett. **71**, 1477–1481 (1993)
3. Feynman, R., Leighton, R., Sands, M.: The Feynman Lectures on Physics, vol. 1, Chap. 46. Addison Wesley, Reading, MA (1963)
4. Smoluchowski, M.V.: Phys. Z. **13**, 1069–1080 (1912)
5. Parrondo, J.M., Espanol, P.: Am. J. Phys. **64**, 1125–1134 (1996)
6. Reimann, P.: Phys. Rep. **361**, 57–271 (2002)
7. Risken, H.: The Fokker-Planck Equation. Springer, Berlin (1984)
8. Reimann, P., Bartussek, R., Haussler, R., Hänggi, P.: Phys. Lett. A **215**, 26–31 (1996)
9. Luczka, J., Czernik, T., Hänggi, P.: Phys. Rev. E **56**, 3968–3975 (1997)
10. Li, Y.-X.: Physica A **238**, 245–251 (1997)
11. Sokolov, M.I., Blumen, A.: J. Phys. A **30**, 3021–3027 (1997)
12. Curie, P.: J. Phys. (Paris) **T.III**, 393–402 (1894)
13. Bartussek, R., Hänggi, P., Kissner, J.G.: Europhys. Lett. **28**, 459–463 (1994)
14. Jung, P., Kissner, J.G., Hänggi, P.: Phys. Rev. Lett. **76**, 3436–3439 (1996)
15. Mateos, J.L.: Phys. Rev. Lett. **84**, 258–261 (2000)
16. Bender, C.M., Coper, F., Freedman, B.: Nucl. Phys. B **219**, 61–80 (1983)
17. Kane, G., Shifman, M.: The Supersymmetric World. World Scientific, Singapore (2000)
18. Chen, F.: J. Appl. Phys. **40**, 3389–3401 (1969)
19. Volk, T.R., Grekov, A.A., Kosonogov, I.A., Fridkin, V.M.: FTT **14**, 3216–3218 (1972)
20. Sturman, B.I., Fradkin, V.M.: The Photovoltaic and Photorefractive Effect in Noncentrosymmetric Materials. Gordon and Breach, Philadelphia (1992)
21. Belinicher, V.I., Sturman, B.I.: UFN **130**, 415–458 (1980)
22. Dubinko, V.I., Tur, A.V., Turkin, A.A., Yanovsky, V.V.: Radiat. Eff. Defects Solids **112**, 233–243 (1990)
23. Lee, T.D., Yang, C.N.: Phys. Rev. **104**, 254–258 (1956)
24. Bolotin, Yu.L., Tur, A.V., Yanovsky, V.V.: JTP **72**, 9–12 (2002)
25. Chandrasekhar, S.: Rev. Mod. Phys. **15**(1), 1–90 (1943)
26. Haken, H.: Synergetics. Springer, Berlin/Heidelberg/New York (1978)
27. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry. North Holland, Amsterdam (1981)
28. Klimontovich, Yu.L.: Statistical Theory of Open Systems, I. Kluwer Academic Publisher, Dordrecht (1995)
29. Ott, E.: Chaos in Dynamical Systems. Cambridge University Press, Cambridge (1992)
30. Julicher, F., Ajdari, A., Prost, J.: Rev. Mod. Phys. **69**, 1269–1281 (1997)
31. Magnasco, M.: Phys. Rev. Lett. **72**, 2656–2659 (1994)
32. Hill, T.L.: Prog. Biophys. Mol. Biol. **28**, 267–340 (1974)
33. Prost, J., Chauwin, J., Peliti, L., Ajdari, A.: Phys. Rev. Lett. **72**, 2652–2655 (1994)
34. Nishizaka, T., Miyata, H., Yoshikawa, H., Ishiwata, S., Kinosita, K.: Nature **377**, 251–254 (1995)

Chapter 9
Quantum Manifestations of Classical Stochasticity

After almost a hundred years of development, quantum mechanics has become a universal picture of the world. All its predictions are correct for every observable scale of energy. However, this does not mean that from time to time, quantum mechanics does not face some new challenges. A serious conceptual problem, defined in the second part of the last century, took the name of quantum chaos. The point is that on the one hand, the energy spectrum of every quantum system with finite motion is discrete, and thus its evolution is quasiperiodic; but on the other hand, the correspondence principle requires the transition to classical mechanics that demonstrates not only regular modes but also the chaotic ones. To solve this problem, we have to answer first this question: how should we understand the statement that one theory is a limiting case of another? [1].

9.1 Formulation of the Problem

As a rule, a more general theory G is associated with a special theory S with a dimensionless parameter δ such as

$$G \to S \quad \text{if} \quad \delta \to 0.$$

For example, if we understand G as the special theory of relativity, and S as classical mechanics, then $\delta = v^2/c^2$. In the simplest case, we can represent the general theory as a Taylor series with the parameter δ.

Actually, such a simple situation is a very rare exception. In the most general (and the most interesting) case, $\lim\limits_{\delta \to 0} G$ is singular, and the transition $G \to S$ is far from being trivial. So, for example, the transition from the Navier–Stokes equations (viscous fluid) to the Euler equations (ideal fluid) is singular: dissipation does not turn to zero smoothly at zero viscosity. Therefore the "no-man's-land" between

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_9

the two theories is difficult to study and contains new physics, such as turbulence and critical behavior at phase transitions. In a similar area, we have to consider the influence of classical stochasticity on semiclassical behavior. In our case, G means quantum mechanics, S classical mechanics, and δ is some dimensionless combination of physical values with Planck constant $\hbar$ in the numerator. According to Berry [1], the $\hbar \to 0$ limit swarms with nonanalyticities.

Let us now discuss another principal difficulty. Since in classical mechanics, chaos is realized only on large time scales (required for complete mixing, i.e., to realize the limiting tendency to zero of the correlation function), any meaningful consideration of semiclassical limits must simultaneously take into account both limit transitions $t \to \infty$ and $\hbar \to 0$ [2]. A natural question arises as to whether the two nontrivial limits $t \to \infty$ and $\hbar \to 0$ commute? The answer is negative: the long-time semiclassical evolution fundamentally differs from the long-time classical evolution. In the common situation, the classical long-time evolution is chaotic, while in semiclassics, the temporal asymptotics are not, so any chaos is just a transition process. Therefore, in an attempt to construct the quantum theory of dynamical chaos, we are immediately confronted by a number of evident and very deep contradictions between the well-established principles of classical chaos and the fundamental principles of quantum mechanics. What is the reason for those contradictions?

As is known, the energy spectrum of any quantum system in finite motion is always discrete. This is a consequence of the fundamental principles of quantum mechanics: the discrete nature of the phase space, or more formally, the noncommutativity of quantum phase space. Indeed, according to the uncertainty principle, an individual quantum state cannot occupy the phase volume $V_1 \leq \hbar^N$, where N is the dimensionality of the configuration space. Therefore, a motion limited within region V will contain V/V_1 eigenstates. According to existing ergodic theory, such motion is considered regular, unlike chaotic motion with a continuous spectrum and exponential instability. The latter statement can be verified using the notion of algorithmic complexity [3], which can be defined as

$$C = \frac{N_{\text{in}}}{N_{\text{out}}}, \tag{9.1}$$

where N_{in} and N_{out} are the input and output lengths of the program (in bits) respectively. This quantity can be determined for any moment of time; however, the distinction between regular and chaotic motion is manifested only in the limit $t \to \infty$. If the motion is chaotic, then $C \to \text{const} > 0$; if it is regular, then $C \to 0$. To understand the reason for this, we should note that the output length of the program gives information about the trajectory at an arbitrary moment of time and increases proportionally to t. In addition, the input data sequence consists of two main parts. The first is the algorithm for the solution of equations of motion, its length not depending on t. The second is the definition of the initial conditions with the precision needed to reproduce the required final result. For chaotic systems, where errors grow exponentially, this part is proportional to t, and therefore, its contribution dominates in the input program length. Therefore, in that case, the algorithmic complexity (9.1) will tend to a constant. For nonchaotic systems, the input program part that is associated with the initial conditions grows more slowly (for example as $\ln t$ when errors grow linearly), and the limiting value of the algorithmic complexity is $C = 0$.

All the experiments carried out thus far show that this rule is strictly implemented for classical chaotic systems. For quantum systems and for the classical analogues that are chaotic, as well as for those that are regular, only zero algorithmic complexity was observed. This result can be briefly formulated in accordance with Bohr's complementarity: that classical evolution is deterministic, but random; quantum evolution is not deterministic and not random. In other words, the problem is that the discrete nature of the spectrum never implies chaos in any quantum system with finite motion. Meanwhile, the correspondence principle requires the presence of chaos in the semiclassical limit.

If we assume that chaos never appears in quantum mechanics, then it would be logical to give up the study of this question. However, this would mean that we avoid the challenge presented by nature concerning the limit of small $\hbar$ and large t, which is equivalent to ignoring other singular phenomena, such as turbulence or phase transitions. The alternative point of view is not to wait for the complete solution of the problem (or rather for its correct formulation), but to study its limited option: to examine the special features of quantum systems whose classical analogues are chaotic, or in other words, to search for quantum manifestations of classical stochasticity (QMCS). We will focus on this approach.

9.2 Semiclassical Quantization

Deterministic chaos is a common feature of Hamiltonian systems with the number of integrals of motion less than the number of degrees of freedom. The absence of the full set of integrals of motion (the full set includes a number of integrals that is equal to the number of degrees of freedom for the quantized system) makes it impossible to realize the traditional procedure of quantization of multidimensional systems. Let us consider this statement in detail. As is known [4], in the one-dimensional case, it is always possible to introduce canonically conjugate "action-angle" variables such that the Hamiltonian becomes a function of the action variable only. The standard definition of the action variable uses the integral along the periodic orbit

$$I = \frac{1}{2\pi} \oint p(x)dx, \tag{9.2}$$

where $p(x)$ is the particle's momentum. In the context of the semiclassical approach, we can construct an approximate solution of the Schrödinger equation in the terms of the integral along the classical trajectory [5]

$$\psi(x) = \frac{1}{\sqrt{p}} \exp\left(\frac{i}{\hbar}\int_{x_0}^{x} p(x')dx'\right). \tag{9.3}$$

This solution makes sense only in the case in which the phase grows in multiples of 2π along the periodic orbit. This limitation immediately leads to the semiclassical quantization condition

$$I = \frac{1}{2\pi}\oint pdx = \left(n + \frac{\mu}{4}\right)\hbar, \tag{9.4}$$

where n is a nonnegative integer and μ is the so-called Maslov index, which is equal to the number of points along the periodic orbit where the semiclassical approximation is violated (in the one-dimensional case, this occurs at the turning points and $\mu = 2$). The semiclassical energy eigenvalues E_n are obtained by the computation of the Hamiltonian $H(I)$ for quantized values of the action variable

$$E_n = H\left(I = \left(n + \frac{\mu}{4}\right)\hbar\right). \tag{9.5}$$

For multidimensional systems, such a procedure can be performed only if the number of integrals of motion is equal to the number of degrees of freedom, i.e., for integrable systems [6]. In this case, the procedure is called Einstein–Brillouin–Keller quantization. For simplicity, let us remain restricted to the two-dimensional case. If the system is integrable, then there exist two pairs of canonically conjugate action-angle variables (I_1, θ_1) and (I_2, θ_2) with the classical Hamiltonian depending only on the action variables

$$H = H(I_1, I_2). \tag{9.6}$$

The classical finite motion is periodic in every angle variable with the frequencies

$$\Omega_i = \frac{\partial H}{\partial I_i}, \quad (i = 1, 2). \tag{9.7}$$

In the general case, the frequencies Ω_i are incommensurable values, and the motion in four-dimensional phase space is quasiperiodic. Phase trajectories lie on invariant tori that are defined by integrals of the motion I_i. The semiclassical wave functions can be constructed in a form that is analogous to (9.3), but the turning points must be replaced by caustics. The uniqueness of the wave function requires the following quantization conditions:

$$I_i = \left(n_i + \frac{\mu_i}{4}\right)\hbar. \tag{9.8}$$

As in the one-dimensional case, the energy eigenvalues can be obtained by the substitution of (9.8) into the Hamiltonian (9.6). Einstein understood already in 1917 that this method could be applied only to quantum integrable systems with trajectories lying on tori. For nonintegrable (i.e., chaotic) systems, there were no consistent quantization methods for half of the last century. But how can we perform the quantization in a nonintegrable situation?

Progress on the problem of quantization of chaotic systems was made using Feynman's formulation of quantum mechanics [7]. The first indication of the applicability of path integrals to chaotic systems was given by Selberg [8], who constructed the dynamics of a particle on a Riemannian surface with negative curvature in terms of path integrals. This is certainly a chaotic system, although that term did not exist at the time.

Gutzwiller first applied successfully an analogous approach to the quantization of chaotic systems. In 1982, he showed that a semiclassical quantization in the form of path integrals allows us to obtain the spectrum of a chaotic system [9].

The periodic orbits play the main role in Gutzwiller's method of quantization of nonintegrable systems. The final purpose of the method is the evaluation of the density of levels

$$\rho(E) = \sum_n \delta(E - E_n)$$

in terms of the solutions of the classical equations of motion.

Using the expression

$$\delta(E - E_n) = \frac{1}{\pi} \mathrm{Im} \lim_{\varepsilon \to 0+} \frac{1}{E_n - E - i\varepsilon},$$

we obtain

$$\rho(E) = -\frac{1}{\pi} \mathrm{Im}\, Sp \left(\frac{1}{E - \hat{H}} \right). \tag{9.9}$$

The operator under the trace is the Green function,

$$G(q_A, q_B, E) = \sum_n \frac{\Psi_n^*(q_A)\Psi_n(q_B)}{E - E_n} = \sum_n \Psi_n^*(q_A) \frac{1}{E - \hat{H}} \Psi_n(q_B),$$

$$\sum_n \Psi_n^*(q_A)\Psi(q_B) = \delta(q_A - q_B) = \int \delta(q_A - q)\delta(q_B - q)dq, \tag{9.10}$$

and therefore,

$$G(q_{A,} q_B, E) = \left\langle q_A \left| \frac{1}{E - \hat{H}} \right| q_B \right\rangle. \tag{9.11}$$

Thus (9.9) can be rewritten in the following form:

$$\rho(E) = -\frac{1}{\pi}\operatorname{Im} SpG. \tag{9.12}$$

This further procedure implies the construction of the Green's function of semiclassical approximation (and then its Fourier transformation) and calculation of the trace. Gutzwiller showed that this procedure results in the following expression for the levels' densities:

$$\rho(E) \simeq \bar{\rho}(E) + \sum_{p}\sum_{k=1}^{\infty} A_{p,k}(E)\cos\left[k\left(I_p(E) - \frac{\pi}{2}\mu_p\right)\right]. \tag{9.13}$$

Here $\bar{\rho}(E)$ is the smoothed density of levels that could be obtained using the Thomas–Fermi approximation or Weyl's formula for billiards. The sum marked by the index p is evaluated over all "primitive" periodic orbits, and the sum of k over k-reiteration of these orbits. The phase for every periodic orbit consists of the action along this orbit $I_p(E)$ and the Maslov index μ_p. The amplitude $A_{p,k}$ is determined by the expression

$$A_{p,k}(E) = \frac{T_p(E)}{\pi k\sqrt{\det(\hat{M} - \hat{I})}}, \tag{9.14}$$

where $T_p(E) = \partial I/\partial E$ is the orbital period and $\hat{M}$ is the monodromy matrix, which is well known from classical analysis of motion stability. Equation (9.13) is called the Gutzwiller trace formula; it expresses the density of the quantum spectrum using the values that are calculable in the context of classical mechanics (Feynman called it one of the main achievements of theoretical physics of the twentieth century). At the same time, this expression should be understood as a universal semiclassical quantization condition that is correct both for integrable and nonintegrable systems: highly excited (semiclassical) energy levels are the points where the right-hand side of the trace formula has poles.

Although many important results were obtained with the trace formula, not all its analytical features have been made clear until now. This is mostly due to the difficulties in the corresponding classical calculations. First of all, we encounter the problem of the periodic orbit evaluation: their number grows exponentially with the growth of the period, and all of them are by definition unstable. At the same time, there is a problem of adequate description of the orbits' contributions (i.e., the summation problem). Finally, the generalization of Gutzwiller's formula in the considered multiwell case is nontrivial. Thus, the problem of the numerical integration of the Schrödinger equation remains relevant for calculation of the semiclassical part of the spectrum of quantum systems that are chaotic in the classical limit.

9.3 Specifics of Classical Dynamics in Multiwell Potentials: Mixed State

Before moving on to the analysis of the QMCS for the specific physical effects, we will focus on the features of the statistical properties of the energy spectra, the structure of stationary wave functions, and dynamics of wave packets for quantum systems. As the main object of study, we will choose a simple system that generates classical deterministic chaos: a two-dimensional time-independent Hamiltonian system. We pay special attention to the Hamiltonian systems with a potential energy surface having several local minima, i.e., multiwell potentials. Despite the huge number of papers concerning chaotic dynamics, Hamiltonian systems with multiwell potentials have been somewhat neglected. Nevertheless, a Hamiltonian system with a multiwell potential energy surface represents a realistic model, describing the dynamics of the transition between different equilibrium states, including such important cases as chemical and nuclear reactions, nuclear fission, phase transitions, string landscape, and many others. Being substantially nonlinear, such systems represent a potentially important object, both for the study of classic chaos and for QMCS.

Chaotic modes in Hamiltonian systems are usually realized only at certain values of system parameters. We can define such regions of space using so-called stochasticity criteria. If we understand stochastization as the process whereby statistic properties appear in strictly deterministic systems due to local instability, we have a tempting opportunity to identify the values of parameters at which local instability occurs, with the boundary of transition to chaos. Unfortunately, the real situation is much more complicated. The problem is that all the existing stochasticity criteria based on the study of local instability have one fault. The loss of stability in regular motion does not always lead to chaos. In spite of this fault, criteria of this kind, together with the numerical experiment, significantly facilitate the analysis of multidimensional motion, restricting the region of parameters in which the chaotic modes are realized. In particular, for two-dimensional Hamiltonian systems, at least in the one-well case, the so-called criterion of negative curvature [10, 11] is convenient, and it connects the occurrence of local instability with the particle coming into the region of negative Gaussian curvature of the potential energy surface.

With this approach, the motion near the potential minimum is regular where the curvature $K(x, y)$ is positive and separated from the line of zero curvature $K(x.y) = 0$. In the case of energy increasing, the particle spends some time in the region of negative curvature, where close trajectories initially diverge exponentially, which, at length, after a long time, leads to the motion's chaotization. According to this scenario, the critical energy of the transition to chaos E_{cr} coincides with the minimal energy on the zero-curvature line,

$$E_{\mathrm{cr}} = U_{\min}(K = 0).$$

Let us consider the characteristics of classical finite motion in multiwell potentials. They are more complicated than in single-well potentials, and allow the existence of several critical energies of transition to chaos, even for a fixed set of potential parameters. This results in the so-called mixed state in such potentials [12]. At the same energy there are different dynamical modes in different wells, either regular or chaotic, in various local minima. It should be noted that the mixed state is a general feature of Hamiltonians with nontrivial potential energy surface. As a first example, let us demonstrate the existence of a mixed state for the Hamiltonian of nuclear quadrupole oscillations [13]. It can be shown that using only the transformation properties of the interaction, the deformation potential of surface quadrupole oscillations of the nuclei takes the following form [14]:

$$U\left(a_0, a_2\right) = \sum_{m,n} C_{mn} \left(a_0^2 + 2a_2^2\right)^m a_0^n \left(6a_2^2 - a_0^2\right)^n , \tag{9.15}$$

where a_0 and a_2 are internal coordinates of the nuclear surface during the quadrupole oscillations:

$$R(\theta, \phi) = R_0\{1 + a_0 Y_{2,0}(\theta, \phi) + a_2[Y_{2,2}(\theta, \phi) + Y_{2,-2}(\theta, \phi)]\}. \tag{9.16}$$

The constants C_{mn} can be considered phenomenological parameters, which within a particular model or approximation can be directly related to the effective interaction of the nucleons in the nucleus [15]. As in Eq. (9.16), only the transformation properties of interaction were used. This expression describes the potential energy of quadrupole oscillations of a liquid drop of any nature containing the specific character of the interaction only in the coefficients C_{mn}.

Restricting ourselves to the terms of fourth order in the deformation, and assuming the equality of masses for the two independent directions, we obtain the following C_{3v}-symmetric Hamiltonian:

$$H = \frac{p_x^2 + p_y^2}{2m} + U_{\mathrm{QO}}\left(x, y; a, b, c\right), \tag{9.17}$$

where

$$\begin{gathered} U_{\mathrm{QO}}\left(x, y; a, b, c\right) = \frac{a}{2}\left(x^2 + y^2\right) + b\left(x^2 y - \frac{1}{3}y^3\right) + c\left(x^2 + y^2\right)^2 , \\ x = \sqrt{2}a_2, \quad y = a_0, \quad \frac{a}{2} = C_{10}, \quad b = 3C_{01}, \quad c = C_{20}. \end{gathered} \tag{9.18}$$

The symmetry C_{3v} of the potential surface becomes obvious in polar coordinates:

$$x = \beta \sin\gamma, \quad y = \beta\cos\gamma.$$

In these coordinates,

$$U(\beta, \gamma; a, b, c) = \frac{1}{2}a\beta^2 - \frac{1}{3}b\beta^3 \cos 3\gamma + c\beta^4. \tag{9.19}$$

Let us introduce the dimensionless variables $(\bar{x}, \bar{y}; \bar{p}_x, \bar{p}_y; \bar{E})$,

$$\begin{gathered}(x, y) = l_0 (\bar{x}, \bar{y}), \quad (p_x, p_y) = p_0 (\bar{p}_x, \bar{p}_y), \quad E = \varepsilon_0 \bar{E}, \\ l_0 = \frac{b}{c}, \ \ p_0 = \sqrt{m\frac{b^4}{c^3}}, \ \ \varepsilon_0 = \frac{b^4}{c^3}.\end{gathered} \tag{9.20}$$

In the variables $(\bar{x}, \bar{y}, \bar{p}_x, \bar{p}_y; \bar{E})$ (later we will drop the bar line), the Hamiltonian (9.17) has the form

$$\begin{gathered}H = \frac{p_x^2 + p_y^2}{2} + U_{\mathrm{QO}}(x, y; W), \\ U_{\mathrm{QO}}(x, y, W) = \frac{1}{2W}(x^2 + y^2) + \left(x^2 y - \frac{1}{3}y^3\right) + (x^2 + y^2)^2.\end{gathered} \tag{9.21}$$

The Hamiltonian (9.21) and the corresponding equations of motion depend only on $W \equiv b^2/ac$, which is the unique dimensionless parameter that can be constructed from the parameters a, b, and c, such that it completely determines the potential energy surface (Fig. 9.1).

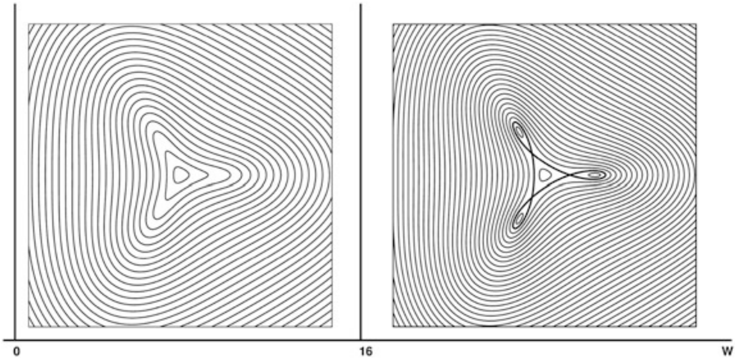


Fig. 9.1 The level lines of the $U_{\mathrm{QO}}(x, y; W)$ potential for $W = 13$ (*left*) and $W = 18$ (*right*)

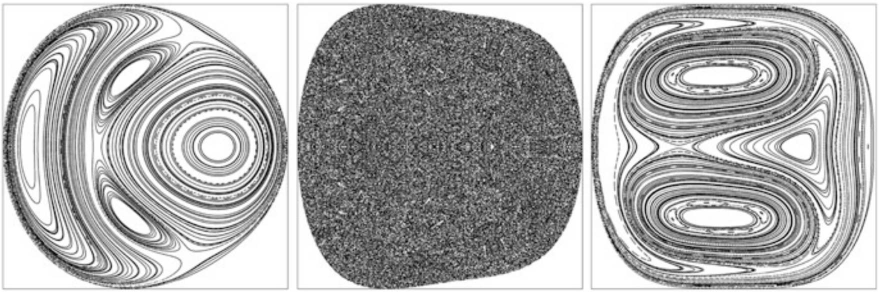


Fig. 9.2 Poincaré sections for R-C-R transition in the quadrupole oscillation potential $U_{QO}(x, y; W = 13)$

The potential U_{QO} (x, y, W) is a generalization of the well-known Henon–Heiles potential [16]

$$U(x, y) = \frac{1}{2}(x^2 + y^2) + x^2y - \frac{1}{3}y^3,$$

which became a traditional object for examination of new ideas and methods in investigations of stochasticity in Hamiltonian systems. It is essential that, in contrast to the Henon–Heiles potential, the motion in (9.21) is finite for all energies, ensuring the existence of stationary states in the quantum case.

The interval $0 < W < 16$ includes potentials with a single extremum: the minimum at the origin that corresponds to the spherically symmetric shape of the nucleus. Numerical calculations [17] of equations of motion in this area (i.e., the area of single-well potentials) indicate the regularity–chaos–regularity (R-C-R) transition: a gradual transition from regular to chaotic motion when energy increases, followed by restoration of regular motion for high energies (Fig. 9.2). For a stochastization scenario, based on the criterion of negative curvature, the R-C-R transition is connected to the finitude of a region of negative Gaussian curvature of the potential surface $U_{QO}(x, y; W = 13)$.

For $W > 16$, the potential energy surface has seven critical points: four minima (one central and three peripheral) and three saddles, separating the peripheral minima from the central one (Fig. 9.1b). Let us consider in detail the case $W = 18$, when the potential U_{QO} (x, y, W) has four minima with the same value $E_{\min} = 0$. In the area W > 16 (multiwell potentials), we encounter a much more complicated situation than in the one-well case. In Fig. 9.3, Poincaré sections are presented for different energies. They demonstrate the evolution of dynamics in central and peripheral minima ($W = 18$). At low energies, the motion is clearly quasiperiodic for both minima. Let us turn our attention to the difference in the topology of the Poincaré sections. At the central minimum, the Poincaré section structure is complicated and has fixed points of different types, while in peripheral minima, the Poincaré sections have trivial structure with only one elliptical fixed point. When the

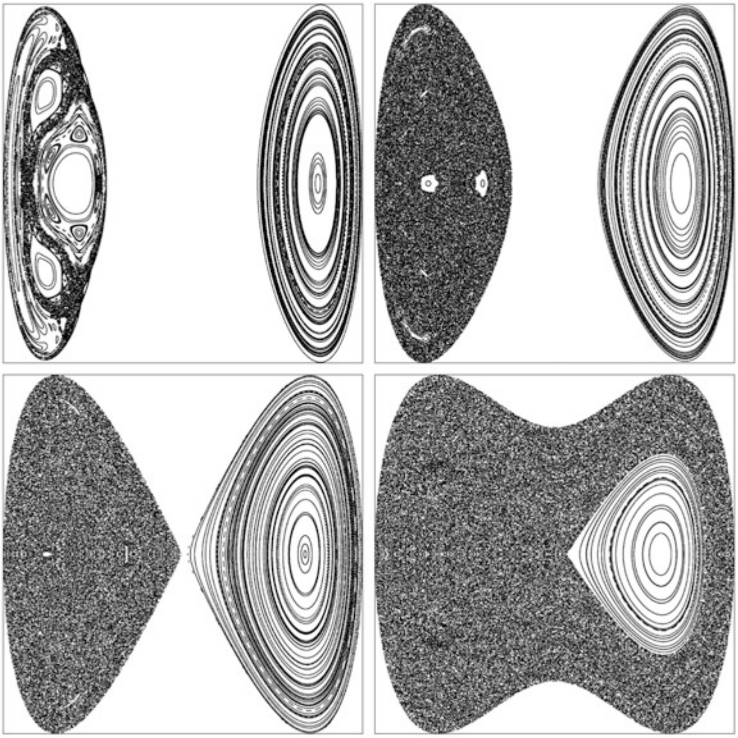


Fig. 9.3 Poincaré section for multiwell quadrupole oscillation potential $(W = 18)$ at different values of energy: $E/E_S = [0.5, 0.8.1, 2]$

energy increases, a gradual transition to chaos is observed, but the changes in the character of motion are totally different in different minima. In the central minimum already at energy equal to half the saddle energy E_S, a sizeable part of the trajectories is chaotic, and at the saddle energy there are almost no regular trajectories at all. Under the same conditions, in the peripheral minimum, the motion remains quasiperiodic. Moreover, even at energies higher than the saddle energy, there is a substantial part of the phase space that is occupied by the quasiperiodic motion. This type of dynamics, when chaoticity measured at fixed energy significantly differs in different local minima, is the common-case situation in multiwell potentials and is called the mixed state [12, 17].

The mixed state, shown above for the potential of quadrupole oscillations, is the representative state for a wide class of two-dimensional potentials with several local minima. According to catastrophe theory [18], a rather wide class of polynomial potentials with several local minima is covered by the germs of the lowest umbilical catastrophes of D_4^-, D_5, D_7 types, subject to certain perturbations. We note that the Henon–Heiles potential coincides with the elliptic umbilic D_4^-.

Figure 9.4 represents the level lines and Poincaré sections at different energies for multiwell potentials from a family of umbilical catastrophes D_5 and D_7:

$$\begin{aligned} U_{D_5} &= 2y^2 - x^2 + xy^2 + \frac{1}{4}y^4 + 1, \\ U_{D_7} &= \sqrt{2}y^2 + \frac{3}{8}x^2 + xy^2 - \frac{1}{2}x^4 + \frac{1}{6}x^6. \end{aligned} \tag{9.22}$$

The mixed state is observed for all the potentials considered here of umbilical catastrophes in the interval of energies $E_{\mathrm{cr}} < E < E_S$ (here, E_{cr} is the critical energy of the transition to chaos at the local minimum, where chaos is observed at energies smaller than the saddle energy).

9.4 The Spectral Method

Calculation of the semiclassical part of the spectrum for systems with a multiwell potential energy surface requires appropriate numerical methods. The matrix diagonalization method is effective in the one-well case. However, this numerical procedure is not efficient at the transition to potential energy surface of complicated topology (multiwell potentials). In particular, the diagonalization of the quadrupole oscillations Hamiltonian (9.21) with $W > 16$ in the harmonic oscillator basis requires too large a number of basis functions, which is beyond the limits of today's computing resources. In this case, the so-called spectral method for solution of the Schrödinger equation represents an attractive alternative to diagonalization. This method was originally proposed in [19] applied to one- and two-dimensional potential systems, but it can be easily generalized to solve the stationary Schrödinger equation in arbitrary dimensions:

$$\left[-\frac{\hbar^2}{2}\sum_{i=1}^{D}\partial_i^2 + U(x_1,\ldots,x_D)\right]\psi_n(x_1,\ldots,x_D) = E_n\psi_n(x_1,\ldots,x_D), \tag{9.23}$$

where D is the dimension of the configuration space of the system. Let us assume that the potential $U(x_1,\ldots,x_D)$ allows only finite motion for all energies (as is the case for quadrupole oscillation potentials). Therefore, the task is to obtain the discrete energy spectrum E_n and the stationary wave functions $\psi_n(x_1,\ldots,x_D)$.

Let us consider the time-dependent solution $\psi(x_1,\ldots,x_D;t)$ of the corresponding nonstationary Schrödinger equation

$$\left[-\frac{\hbar^2}{2}\sum_{i=1}^{D}\partial_i^2 + U(x_1,\ldots,x_D)\right]\psi(x_1,\ldots,x_D;t) = i\hbar\partial_t\psi(x_1,\ldots,x_D;t) \tag{9.24}$$

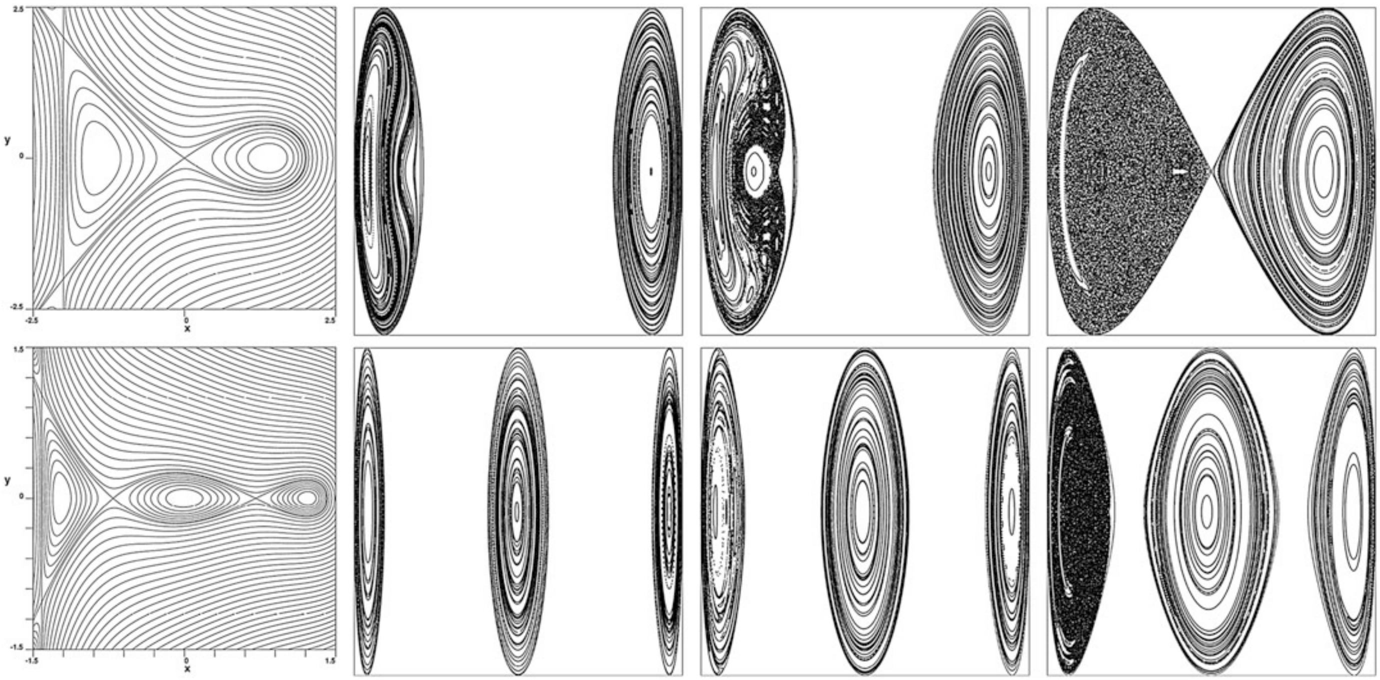


Fig. 9.4 Level lines and Poincaré sections for D_5 and D_7 potentials

with some, generally speaking, arbitrary initial condition $\psi_0(x_1,\dots,x_D) = \psi(x_1,\dots,x_D;t=0)$. Using the decomposition

$$\psi_0(x_1,\dots,x_D) = \sum_{n=1}^{\infty} a_n\psi_n(x_1,\dots,x_D), \tag{9.25}$$

we obtain

$$\psi(x_1,\dots,x_D;t) = \sum_{n=1}^{\infty} a_n\psi_n(x_1,\dots,x_D)e^{-i\frac{E_nt}{\hbar}}. \tag{9.26}$$

Here and below we assume that the wave functions $\psi_n(x_1,\dots,x_D)$ form an orthonormal basis:

$$\int dx_1\dots dx_D\bar{\psi}_i(x_1,\dots,x_D)\psi_k(x_1,\dots,x_D) = \delta_{ik}. \tag{9.27}$$

Let us consider a correlator of the following form:

$$P(t) = \int dx_1\dots dx_D\bar{\psi}_0(x_1,\dots,x_D)\psi(x_1,\dots,x_D;t) = \sum_{n=1}^{\infty}|a_n|^2\,e^{-i\frac{E_nt}{\hbar}}. \tag{9.28}$$

The wave function of the initial state $\psi_0(x_1,\dots,x_D)$ is also assumed to be normalized:

$$\int dx_1\dots dx_D\bar{\psi}_0(x_1,\dots,x_D)\psi_0(x_1,\dots,x_D) = \sum_{n=1}^{\infty}|a_n|^2 = 1. \tag{9.29}$$

Thus $P(0) = 1$. The Fourier transform of the correlator (9.28) contains information about the energy spectrum E_n (see Fig. 9.5),

$$P(E) = \int_{-\infty}^{+\infty} dte^{i\frac{Et}{\hbar}}P(t) = \hbar\sum_{n=1}^{\infty}|a_n|^2\,\delta(E-E_n), \tag{9.30}$$

while information about the stationary wave functions $\psi_n(x_1,\dots,x_D)$ can be extracted in turn from the Fourier transform of the function $\psi(x_1,\dots,x_D;t)$:

$$\begin{aligned}\psi(x_1,\dots,x_D;E) &= \int_{-\infty}^{+\infty} dte^{i\frac{Et}{\hbar}}\psi(x_1,\dots,x_D;t)\\ &= \hbar\sum_{n=1}^{\infty}a_n\psi_n(x_1,\dots,x_D)\delta(E-E_n).\end{aligned} \tag{9.31}$$

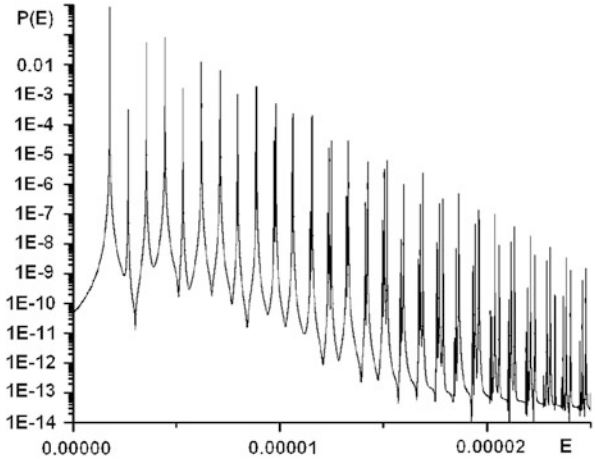


Fig. 9.5 $P(E)$ correlator for the Hamiltonian of the quadrupole oscillations (9.21) ($W = 18$)

Naturally, in practice, one never has to find all the E_n and $\psi_n(x_1,\ldots,x_D)$. The usual task is to obtain all energy levels E_n within some range $E_1 < E < E_2$ with some accuracy δ_E and to calculate the corresponding stationary wave functions $\psi_n(x_1,\ldots,x_D)$ on some coordinate grid $x_i^{(k)} = x_{i0}^{(k)} + k\Delta x_i$, $k = 0,\ldots,N_i$, with finite accuracy as well.

We now show how this concept of the spectral method can be realized as an efficient computational procedure. Evidently, in practical calculations, one always deals with a finite number of values known with finite accuracy. In the case in which we are interested, this means that the most fundamental value $\psi(x_1,\ldots,x_D;t)$ will be calculated on a finite set of points both in time and in spatial coordinates. Therefore, the correlator $P(t)$ will be known only at the points $t_k = k\Delta t, k = 1,\ldots,M$, and the Fourier transforms (9.30) and (9.31) will take the form

$$P_T(E) = \frac{1}{T}\int_0^T dt e^{i\frac{Et}{\hbar}} P(t) = \sum_{n=1}^{\infty} |a_n|^2\, \delta_T(E-E_n) \tag{9.32}$$

$$\begin{aligned}\psi_T(x_1,\ldots,x_D;E) &= \frac{1}{T}\int_0^T dt e^{i\frac{Et}{\hbar}} \psi(x_1,\ldots,x_D;t)\\ &= \sum_{n=1}^{\infty} a_n\psi_n(x_1,\ldots,x_D)\delta_T(E-E_n),\end{aligned} \tag{9.33}$$

where $T = M\Delta t$, and the finite analogue of the δ-function reads

$$\delta_T(E) = \frac{1}{T}\int_0^T dte^{i\frac{Et}{\hbar}} = \frac{e^{i\frac{ET}{\hbar}} - 1}{i\frac{ET}{\hbar}} = f_T\left(\frac{ET}{h}\right),$$
$$f_T\,(x) = \frac{\sin \pi x}{\pi x}e^{i\pi x}. \tag{9.34}$$

Unlike the usual δ-function, here $\delta_T(0) = 1$. The plot for $P(E)$ $(P_T(E))$ in Fig. 9.5 displays a set of sharp local maxima at $E = E_n$, where E_n are the energy eigenvalues. Once the eigenvalues are known, the corresponding eigenfunctions can be computed by substituting $E = E_n$ into Eq. (9.32):

$$\psi_n(x_1,\ldots,x_D) \approx \psi_T(x_1,\ldots,x_D;E_n).$$

This procedure is very efficient when implemented using the fast Fourier transform algorithm.

9.5 Statistical Properties of Energy Spectra

The energy spectra represent historically the first object for the investigation of quantum signatures of chaos in Hamiltonian systems. Substantial progress in the detection of QMCS in the 1980s was associated with this object. Up to that time, the connection between spectral properties of complex systems and ensembles of random matrices had been established, which created the background for understanding the statistical properties of the energy spectra. This stimulated the transition from the study of the behavior of separate levels to the statistical characteristics of the energy spectrum as a whole. During research, it became clear that only the local (not averaged) statistical properties of energy spectra are of interest from the point of view of QMCS. Why should we take special interest in the local characteristics of a spectrum? The fact is that global characteristics, such as the numbers of states $N(E)$ or the smoothed density of levels $\rho(E)$, are too rough. At the same time, such local characteristics as the distribution function $P(S)$ of the nearest-neighbor spacing $S = E_{i+1} - E_i$ between levels are very sensitive to the properties of the potential or to the shape of a billiards boundary.

In the 1960s, Wigner [20], Porter [21], and Dyson [22] built the statistical theory of the complex systems spectra based on the following hypothesis: the distribution of energy levels in complex systems is equivalent to the distribution of the set of eigenvalues of random matrices of a certain symmetry. The ultimate result for $P(S)$ obtained in that theory has the following form:

$$P(S) = aS^{\alpha}e^{-bS^2}, \tag{9.35}$$

where a and b are slowly varying functions of energy, and the critical exponent α defines the behavior of the distribution function in the limit $S \to 0$, depending on the symmetry properties of the random matrices. In particular, $\alpha = 1$ for a Gaussian orthogonal ensemble of a random matrix.

The predictions of the statistical theory of energy spectra (mainly for a Gaussian orthogonal ensemble of random matrices) were carefully compared to all available data on nuclear spectra. No considerable contradictions between the theory and experiment were found. In particular, the random matrices ensembles perfectly reproduced such an important spectral characteristic as spectral rigidity, describing the small fluctuations of energy levels around the averaged values in a given interval. A measure of rigidity is the statistic Δ_3 of Dyson [22] and Mehta [23]:

$$\Delta_3(L,x) = \frac{1}{L} \min_{A,B} \int_x^{x+L} [n(\varepsilon) - A\varepsilon - B]\, d\varepsilon, \tag{9.36}$$

where $n(E)$ is a staircase state number function,

$$n(E) = \int_{-\infty}^{E} dE\rho(E), \tag{9.37}$$

which determines the least-squares deviation of the staircase representing the cumulative density $n(\varepsilon)$ from the best straight line fitting it in any interval $[x, x+L]$. An example of the maximum correlated spectrum is the equidistant spectrum of the harmonic oscillator, for which $\Delta_3 = 1/12$. The opposite case is the Poisson spectrum, for which $\Delta_3 = L/15$, reflecting the strong fluctuations near the mean level density.

Analogous comparisons were made also for atomic spectra. Good agreement with the theory was found for them as well, although for much weaker statistics.

A completely different approach to the problem of statistical properties of energy spectra was developed on the basis of the nonlinear theory of dynamical systems. As numerical simulations show [24–27]—and they are confirmed by serious theoretical considerations [9, 28, 29]—the main universal property of systems that have a regular type of dynamics in the classical limit is the level clustering phenomenon, while for systems that are chaotic in the classical limit, level repulsion is observed. Sometimes, this statement is called the hypothesis of the universal character of energy spectra fluctuations [24].

In the case of regular motion, every classical trajectory lies on a surface topologically equivalent to a torus. Each such torus corresponds to a certain set of integrals of motions (or quantum numbers), with the total number equal to the number of degrees of freedom. Different eigenfunctions of integrable quantum systems correspond to different sets of quantum numbers, and therefore lie on different tori. Their eigenvalues are not correlated (i.e., their wave functions do not overlap), which leads to $P(S)$ of Poisson type:

$$P_P(S) \propto e^{-S}. \tag{9.38}$$

This distribution corresponds to clustering levels (i.e., a large probability of small spacings between neighboring levels).

In the transition to chaos, some tori are destroyed, leading to the formation of chaotic areas in the phase space. The classical trajectories in such areas diffuse between the different (already destroyed) tori. This means in quantum terms the occurrence of overlap between the wave functions with different quantum numbers, giving the interaction between levels. Every interaction, in turn, causes repulsion between levels. We can explain this phenomenon [30] by considering the Hamiltonian H, defined at some fixed basis with its matrix elements. We can consider the repulsion of levels as a result of the fact that the subspace with a degenerate spectrum has a smaller dimension than the overall space of the matrix elements. In this way, the degeneration or a weaker effect of clustering has low probability. In the limit of a completely chaotic system, (i.e., all tori are destroyed), all modes of an initially integrable system mix with each other, so that repulsion exists between every pair of levels, changing $P(S)$ from Poisson to Wigner:

$$P_W(S) = \frac{\pi}{2} S e^{-\frac{\pi}{4}S^2}. \tag{9.39}$$

The situation becomes much more complicated for generic Hamiltonian systems, where the phase space contains both regular and chaotic components. How is it reflected in the energy spectra's statistical properties?

Berry and Robnik [31], and independently, Bogomolny [32], based on semiclassical arguments, showed that $P(S)$ in the case of the dynamics of mixed type represents the superposition of Poisson and Wigner distributions with weights equal to some relative measures of phase space occupied by regular and chaotic motion, respectively:

$$P_{\mathrm{BRB}}(s) \propto \rho^2 e^{-\rho S} \mathrm{erf}\left(\frac{\sqrt{\pi}}{2}(1-\rho)S\right) + \left[\frac{\pi}{2}(1-\rho)^2 S + 2\rho\right](1-\rho) e^{-\rho S - \frac{\pi}{4}(1-\rho)^2 S^2}, \tag{9.40}$$

where ρ is the relative phase volume occupied by the regular trajectories in the mixed system. The limit $\rho \to 1$ corresponds to a regular system, and $\rho \to 0$ to a completely chaotic one.

Among the systems subject to detailed numerical analysis of the spectral properties, two-dimensional billiards has the central place. A two-dimensional billiard is a point particle freely moving in the plane inside some region of arbitrary shape and elastically reflecting from the boundary. These systems have attracted the active interest of researchers for the following reasons:

1. Simplicity of classical dynamics.
2. Stochasticity criteria for billiards can be formulated in geometrical terms.
3. Homogeneity of the phase space.
4. Availability of efficient methods for solving the Schrödinger equation for billiards.

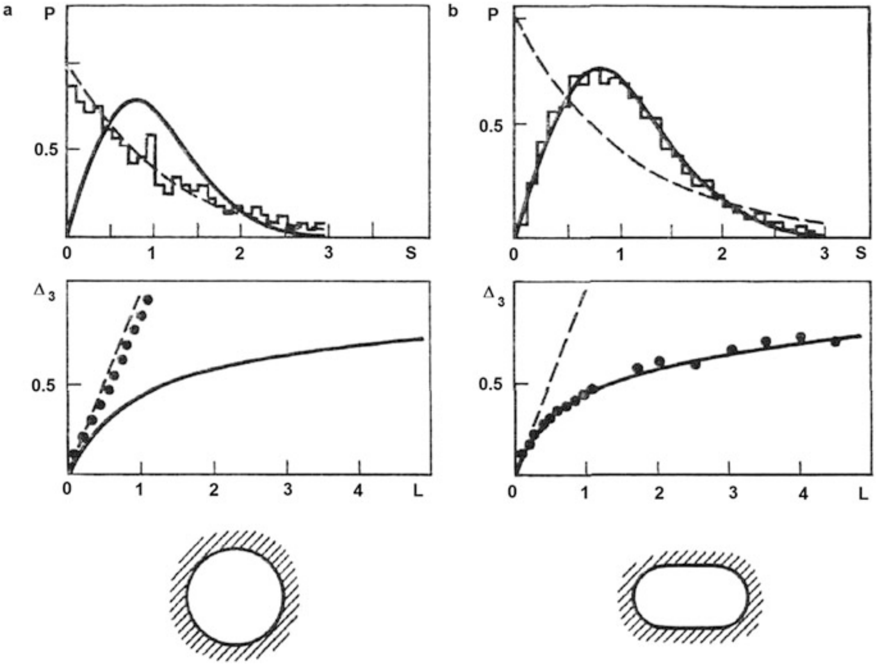


Fig. 9.6 Statistical characteristics of energy spectra for circular (**a**) and "stadium" (**b**) billiards

5. The smooth component of level density is well known due to the Weyl formula.
6. Billiard dynamics reflects the real situation in many physical systems (e.g., quantum dots, Josephson's junctions, and nuclear billiards).
7. At present, there is efficient experimental realization both for classical and quantum dynamics in microwave and optical two-dimensional billiards.

For billiards with a fixed boundary shape, one of two limiting cases is realized: exact integrability and absolute chaos. Therefore, in circular billiards (Fig. 9.6a), the angular momentum is the second (after energy) integral of motion, and this system is integrable. Billiards of "stadium" type (Fig. 9.6b) is one of the simplest stochastic systems. In Fig. 9.6, the statistical characteristics of energy spectra (spacing distribution function $P(S)$ and rigidity Δ_3) are presented for both systems. In complete agreement with the hypothesis of the universal character of the spectral fluctuations, the function $P(S)$ for a circular billiard is perfectly approximated by the Poisson distribution, and the rigidity Δ_3 is a linear function of the length of the considered interval. In the nonintegrable case, the level repulsion effect is distinctly manifested, leading to a Wigner distribution, and the dispersion grows much more slowly because of the higher rigidity of the considered spectrum.

Now let us consider the possibility of realizing quantum dynamics in microwave billiards [33, 34].

By microwave billiards, we mean the electromagnetic field in a cavity of arbitrary form. The electromagnetic radiation in the billiards cavity is described by the Maxwell equations:

$$\begin{aligned}
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t},\\
\nabla \times \mathbf{H} &= \frac{\partial \mathbf{D}}{\partial t},\\
\nabla \mathbf{D} &= 0,\\
\nabla \mathbf{B} &= 0,
\end{aligned} \tag{9.41}$$

where the electric induction $\mathbf{D}$ and the magnetic induction $\mathbf{B}$ are linked by the intensity of the electric $\mathbf{E}$ and magnetic $\mathbf{H}$ fields in the following correlations:

$$\begin{aligned}
\mathbf{D} &= \varepsilon_0 \mathbf{E},\\
\mathbf{B} &= \mu_0 \mathbf{H}.
\end{aligned} \tag{9.42}$$

Here ε_0 and μ_0 are the dielectric and magnetic permeability of a vacuum. Assuming the periodic dependence of the electromagnetic fields with time, we can obtain Helmholtz equations for the fields $\mathbf{E}$ and $\mathbf{B}$, using standard procedures:

$$\begin{aligned}
\left(\Delta + k^2\right) \mathbf{E} &= 0,\\
\left(\Delta + k^2\right) \mathbf{B} &= 0,
\end{aligned} \tag{9.43}$$

where $k = \omega / c$ is the wave vector and ω is the angular frequency of radiation. Equation (9.43) should be amplified with the boundary conditions

$$\mathbf{n} \times \mathbf{E} = 0, \quad \mathbf{n}\mathbf{B} = 0, \tag{9.44}$$

where $\mathbf{n}$ is a unit vector of a normal to the surface. If we consider only cavities, limited by cylindrical surfaces (with various cross sections), then the boundary conditions on the cylindrical surface S (the z-axis is chosen parallel to the cylinder axis) take the form

$$E_z|_S = 0, \quad \nabla_\perp B_z|_S = 0. \tag{9.45}$$

Here $\nabla_\perp$ designates the normal derivative. There are two possibilities for meeting the boundary conditions (9.44). One of them represents the so-called transverse magnetic modes

$$\begin{aligned}
E_z(x,y,z) &= E(x,y)\cos\left(\frac{n\pi z}{d}\right), \quad n = 0, 1, 2, \ldots,\\
B_z(x,y,z) &= 0,
\end{aligned} \tag{9.46}$$

where d is the height of the cylinder, and $E(x, y)$ satisfies the two-dimensional Helmholtz equation

$$\left[\Delta + k^2 - \left(\frac{n\pi}{d}\right)^2\right] E(x, y) = 0, \tag{9.47}$$

with boundary conditions of Dirichlet type on the cylinder's surface:

$$E(x, y)|_S = 0. \tag{9.48}$$

For the transverse electric mode, we obtain, analogously,

$$\begin{aligned} E_z(x, y, z) &= 0, \\ B_z(x, y, z) &= B(x, y) \sin\left(\frac{n\pi z}{d}\right), \qquad n = 1, 2, 3, \dots, \end{aligned} \tag{9.49}$$

where $B(x, y)$ is the solution of the two-dimensional Helmholtz equation

$$\left[\Delta + k^2 - \left(\frac{n\pi}{d}\right)^2\right] B(x, y) = 0, \tag{9.50}$$

with a boundary condition of Neumann type

$$\nabla_\perp B(x, y)|_S = 0. \tag{9.51}$$

For wave vectors $k < \pi/d$, only the TM mode is realized with $n = 0$, and Eq. (9.49) is reduced to

$$\left(\Delta + k^2\right) E(x, y) = 0. \tag{9.52}$$

This equation is fully equivalent to the two-dimensional Schrödinger equation, boundary conditions included, for the particle in the area S, bounded by infinitely high potential walls. This means that from the formal point of view, the system representing the resonator of cylindrical geometry with electromagnetic radiation of transverse magnetic type is equivalent to the billiards in the area equal to the base of the cylinder. In this analogy, the function $E(x, y)$ corresponds to the wave function of the particle $\psi(x, y)$, and k^2 corresponds to its energy. Now the difficult problem to measure the wave function and to find the spectrum in two-dimensional billiards of arbitrary form is substituted by a much simpler problem that consists in finding the electric field intensity component and determining the spectrum of the eigenmodes in the resonator.

Now let us discuss the efficiency of this way of studying the spectrum. In fact, it is defined by the accuracy of the realization of the boundary conditions (9.44). The latter hold only for ideally conductive walls. In real conditions, we always deal with finite conductivity of walls. As a result, the radiation penetrates into the walls of the cavity to the depth

$$\delta = \sqrt{\frac{2}{\mu_0 \omega \sigma}}, \tag{9.53}$$

where σ is the conductivity of the wall. For good conductors, $\sigma \sim 10^7\,\Omega^{-1}\,\mathrm{m}^{-1}$, and at typical microwave frequencies $\omega \sim 10\,\mathrm{GHz}$, the typical penetration depth is $\delta \sim 1\,\mathrm{mm}$. The dissipation of the radiation in the walls leads to exponential electromagnetic energy attenuation,

$$W(t) = W_0 e^{-t/\tau}, \tag{9.54}$$

and so in the final width of the line,

$$\Delta\omega = \frac{1}{\tau}. \tag{9.55}$$

Attenuation at the frequency ω_0 is defined by the Q factor of the resonator $Q = \omega_0/\Delta\omega$,

$$\tau = \frac{Q}{\omega_0}. \tag{9.56}$$

For typical conductors, $Q \sim 10^3 - 10^4$. We can show that the maximum number of solvable resonances is $N_{\max} \sim 1/3Q$. This means that using the ordinary conductive walls of the cavity, one can solve a number of resonances (levels) of order 1000. The only possibility for increasing the resolution is using resonators having a much higher value of Q. This possibility was realized in the resonators limited by superconductive walls with Q up to 10^7 [34]. The super-sharp resonances allow the resolution to be made several orders of magnitude higher.

Let us now turn to the statistical properties of the spectra of two-dimensional Hamiltonian systems, using as an example potentials of the quadrupole oscillations $U_{\mathrm{QO}}(x,y)$ and umbilical catastrophes $U_{D_5}(x,y)$. As a first example, we consider the transformation of the energy spectrum in the R-C-R transition for the one-well case ($W < 16$). At the fixed topology of the potential surface ($W = \mathrm{const}$), the unique free parameter of the Hamiltonian is the scaled Planck's constant $\bar{\hbar}$. In the study of the concrete energetic interval (R_1, C or R_2), corresponding to a definite type of classical motion, the choice of $\bar{\hbar}$ is dictated by the possibility of achieving the necessary statistical reliability (i.e., the required number of levels in the considered energy interval) with conservation of the precision of spectrum calculation (with respect to the limited possibility of the diagonalization for matrices of large dimensions). The numerical results are presented in Fig. 9.7. Both the distribution function $P(S)$ and rigidity Δ_3 correspond well to the predictions of the Gaussian orthogonal ensemble of random matrices for the chaotic (C) area. The

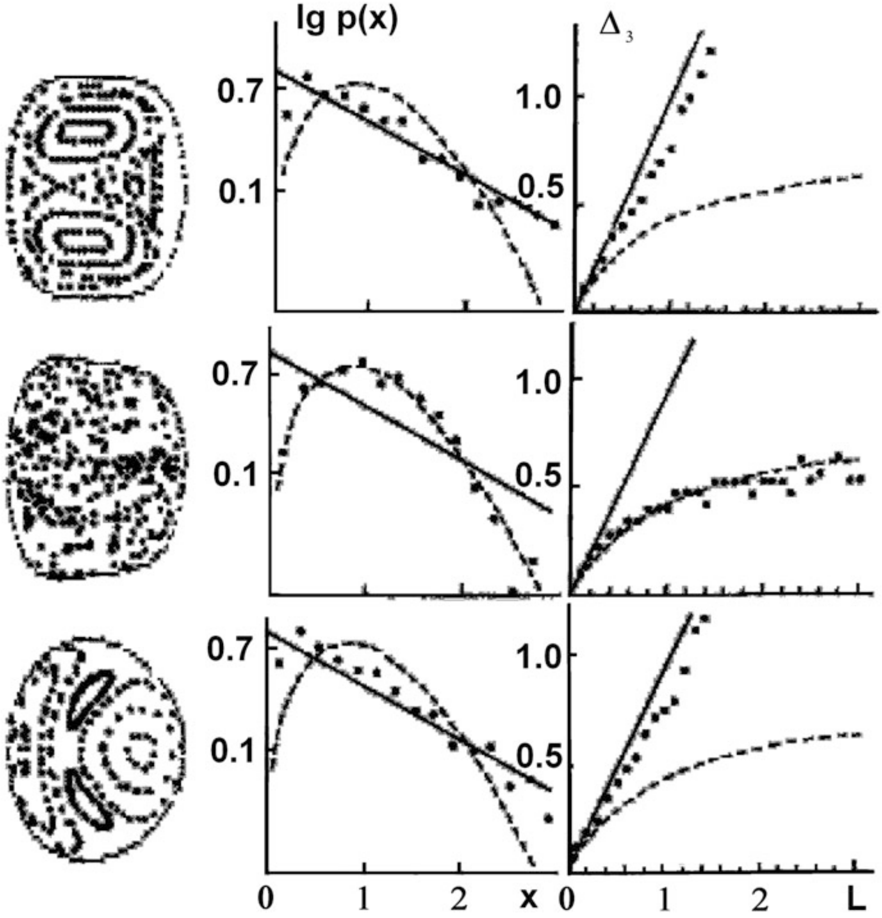


Fig. 9.7 Correlation between the character of classical motion and the statistical properties of energy spectra in R-C-R transition for the Hamiltonian of quadrupole oscillations (9.21) for $W = 13$. On the *left*: Poincaré sections, in the *middle*: distribution function $P(S)$, on the *right*: the Δ_3-statistics. From the bottom up: the first regular range R_1, the chaotic range C, and the second regular range R_2

logarithmic scale for $P(S)$ is suitable for tracing this correspondence at large S. For the regular areas (R_1 and R_2), the distribution function must be represented by a straight line (the logarithm of the Poisson distribution).

The results demonstrate the agreement with the hypothesis of the universal character of the fluctuations of energy spectra, though small-sized deviations are observed for small distances between levels. The tendency of repulsion increasing in the regular area, apparently, is associated with a small admixture of chaotic components.

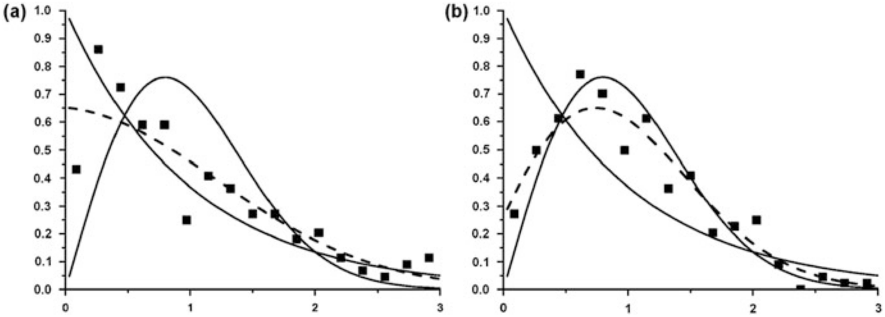


Fig. 9.8 Spacing distribution $P(S)$ for the D_5 potential $E < E_{\rm cr}$ (**a**) and $E > E_{\rm cr}$ (**b**). The *points* represent the numerical data; *solid lines* are the Poisson and Wigner distributions; *dashed line* is the best approximation by the Berry–Robnik–Bogomolny distribution

The spectral ranges of multiwell potentials realizing the mixed state open new possibilities for the study of intermediate statistics. At those energies, the chaotic and regular components are separated not only in phase space as in the usual case of motion of mixed type, but even in configuration space. A priori, the form of spacing distribution function in the mixed state is not necessarily reduced to a definitely weighted superposition of Poisson and Wigner distributions. In that case, we deal not with statistics of the admixture of two level series with different spacing distributions, but with statistics of spectrum series, where each level cannot be attributed to definite (Poisson or Wigner) statistics. The statistical properties of these systems were little studied until recently, although they are the systems corresponding to the most common situation.

Let us consider the simplest potential realizing the mixed state: the two-well potential of lower umbilical catastrophe D_5 (9.22). In order to describe the statistical properties of the corresponding energy spectrum, let us try to use the Berry–Robnik–Bogomolny distribution (9.39), where the parameter ρ equals the relative phase volume occupied by regular trajectories. In the case of the D_5 potential, $\rho \sim 1$ for $E < E_{\rm cr}$ and $\rho \sim 1/3$ for $E > E_{\rm cr}$, which qualitatively (although very roughly) agrees with the numerical data for energy level statistics (Fig. 9.8).

The Berry–Robnik–Bogomolny distribution function by its construction describes the energy level fluctuations in the absence of interaction between the regular and chaotic components. Therefore, the best agreement between the Berry–Robnik–Bogomolny spectral fluctuation theory and experimental data should be expected in the mixed-state spectra, where the interaction of chaotic and regular states with each other is additionally suppressed by the energy barrier separating the corresponding local minima. In practice, however, such an agreement is never observed for several reasons.

The first reason is not related directly to the mixed state properties and consists of technical difficulties of calculating long spectral series in potentials with several local minima. In practice, one usually should content oneself with a series of several hundred levels at most, which is definitely insufficient for obtaining accurate distribution functions for spectral fluctuations.

The second reason for disagreement between the Berry–Robnik–Bogomolny distribution and actual numerical data is common to all potential systems and consists in the fact that in smooth potential systems, the classical chaoticity measure depends significantly on the energy. Indeed, the upper and lower levels in sufficiently long spectral series will correspond to the different relations of regular and chaotic components, while all known theoretical distribution functions assume the relation to be constant. This forces us to consider spectral series that are narrow in energy, which makes statistics levels even poorer.

Finally, the specific feature of the D_5 potential, as for the D_7 and Henon–Heiles potentials, is that in all the aforementioned potentials, the discrete energy spectrum, strictly speaking, is absent. The reason for this is that all these models allow the infinite motion for all energies. Nevertheless, many workers have discussed spectral fluctuations in such systems, implying the spectrum of quasistationary states, localized in the corresponding potential well and extremely slowly decaying by tunneling into the continuous spectrum. Every numerical calculation of such states practically contains the implicit reformulation of the original model, in order to make it more correct from the quantum-mechanical point of view. The quasistationarity of such states becomes especially obvious near the saddle energies, which are the most interesting energy regions for the consideration of the mixed state.

9.6 Quantum Chaos and the Riemann Zeta Function

Discussing the universal statistical properties of quantum spectra, let us turn to a very useful and fruitful analogy between theoretical and numerical correlations linking the nontrivial zeros of the Riemann zeta function and prime numbers, on the one hand, with semiclassical correlations linking the quantum energy levels and classical periodic orbits on the other hand. As we have seen above, the energetic spectra in the semiclassical limit have unique distributions, depending only on the general properties of classical dynamics. These distributions coincide with the statistics of certain ensembles of random matrices. Therefore, the attempt to explain this wonderful fact in terms of the periodic orbits, which, according to the Gutzwiller trace formula, let us construct the semiclassical spectrum of an arbitrary system (integrable as well as nonintegrable), seems natural. From this point of view, the Riemann zeta function plays an important role because, as has been discovered [35], its nontrivial zeros have a statistical distribution notably close to the distribution of eigenvalues of Gaussian unitary ensembles of random matrices, and deviations from these distributions coincide with the deviation from the universality in semiclassical distributions. There is also quite pragmatic interest in the aforementioned analogy. Formulas of number theory are precise and not approximate, and the transition from the periodic orbits to prime numbers allows one to reduce computer calculations as well as to make substantial analytic progress.

The Riemann zeta function, a function of a complex variable s, can be defined for $\mathrm{Re} s > 1$ using the Euler product:

$$\varsigma(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \tag{9.57}$$

where p runs through the prime numbers.

The fundamental theorem of arithmetic states that every integer greater than 1 either is prime itself or is the product of prime numbers, and that this product is unique. Using this theorem the Euler representation can be transformed into a Dirichlet series.

$$\varsigma(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{9.58}$$

which also converges for $\mathrm{Re} s > 1$.

Euler introduced the Riemann zeta function as a function of a real variable in 1737 (as a sum). He also pointed out the possibility of its representation as a product of prime numbers. In relation to the distribution of prime numbers, it was considered by Dirichlet and by Chebyshev. However, the most profound properties of the Riemann zeta function were discovered when it was considered as a function of a complex variable. First, this was done by Riemann in 1876. Riemann showed that there is a representation $\varsigma(s)$, converging everywhere in the complex plane except for the point $s = 1$, where it has a simple pole.

The interest of mathematicians in the Riemann zeta function is due to the fact that it is a useful tool for the study of prime numbers, which is indicated by Eq. (9.57). This correlation allows us to construct the function $\pi(x)$ of the number of prime numbers less than x. This function can be constructed in terms of the zeros of the Riemann zeta function and thus shows that the distribution of prime numbers is associated with the distribution of zeros $\varsigma(s)$. One hundred years ago, it was hard to assume that physicists would be interested in the reverse problem: to find the distribution of zeros of the Riemann zeta function knowing the distribution of prime numbers.

The Riemann zeta function has an infinite number of zeros at the negative even integers, known as the trivial zeros. They are not of much interest. There is also an infinite number of complex, or nontrivial, zeros, and they are the object of our study. It is known that the nontrivial zeros of the Riemann zeta function have real part greater than zero and less than one. The last of the Riemann's five hypotheses about the zeta function is that the nontrivial zeros lie in the complex plane on the line with real part $1/2$, i.e., the nontrivial solutions of the equation

$$\varsigma\left(\frac{1}{2} - iE\right) = 0 \tag{9.59}$$

are real.

A refutation or proof of the Riemann hypothesis is one of the central open problems of pure mathematics. The "experimentalists" began their own studies without waiting for any progress from the "theoreticians." By 1987, an area of up to $E = 10^9$ had already been studied. The Riemann hypothesis was proved to 10^{-8} accuracy. In the 1990s, a computational boom began due to physicists' interest in the problem [36]. Berry wrote about these works:

> Odlyzko's computations were of epic proportions: 1000 h of Gray X-MP time, generating 2000 Mb of data. By contrast, evaluation of the semiclassical formula takes only a few seconds, illustrating the power and usefulness of asymptotics.

We should note that there is an old idea of Pólya that to prove the Riemann hypothesis, it is desirable to show that the nontrivial zeros of the Riemann zeta function are eigenvalues of some Hermitian operator. Physicists are trying to consider (so far in vain) the even more powerful possibility that the nontrivial zeros of the Riemann zeta function might be quantum levels of a real dynamical system.

Let us now proceed to the equation that represents the foundation of our analogy, connecting the Riemann zeta function and semiclassical quantum spectra. To obtain the equation, let us use a theorem. Let a function $f(z)$ be analytic everywhere in the closed region F, except for a finite number of special isolated points z_k, which lie inside F and all are poles, and let $f(z)$ not vanish at any point of the boundary L of the region F. Then the difference between the full number of zeros (N) and the full number of poles (P) of rgw function $f(z)$ in the region F is defined by the following expression:

$$N - P = \frac{1}{2\pi i}\int_L \frac{f'(z)}{f(z)}dz. \tag{9.60}$$

For the case in which we are interested, $f = \varsigma(z)$, the result can be represented as

$$N_R(E) = \bar{N}_R(E) - \frac{1}{\pi}\mathrm{Im}\log\varsigma\left(\frac{1}{2} - iE\right), \tag{9.61}$$

where

$$\bar{N}_R(E) = 1 - \frac{E}{2\pi}\log\pi - \frac{1}{\pi}\mathrm{Im}\log\Gamma\left(\frac{1}{4} - i\frac{E}{2}\right). \tag{9.62}$$

Here Γ is the gamma function.

Now let us take some liberty to substitute $\varsigma\left(\frac{1}{2} - iE\right)$ by the Euler product (9.57). Strictly speaking, one should not do that, because the product converges only if $\mathrm{Im}E > 1/2$, but we are interested in the values of E on the real axis. After the substitution, we have

$$N_R(E) = \bar{N}_R(E) - \frac{1}{\pi}\sum_p\sum_{k=1}^{\infty}\exp\left(-\frac{1}{2}k\log p\right)\frac{1}{k}\sin\left(Ek\log p\right). \tag{9.63}$$

Differentiating with respect to E and using in $\bar{N}_R(E)$ Stirling's asymptotic formula for the Γ-function at high E, we obtain the main result, the formula for the density of the "levels" (i.e., the nontrivial zeros of the Riemann zeta function):

$$\begin{aligned} d_R(E) &\equiv \sum_n \delta\left(E - E_n\right) \\ d_R(E) &= \bar{d}_R(E) - 2\sum_p \sum_{k=1}^{\infty} \frac{\log p}{2\pi} \exp\left(-\frac{1}{2}k\log p\right) \cos\left(Ek\log p\right) \\ \bar{d}_R(E) &\simeq \frac{1}{2\pi}\log\frac{E}{2\pi}. \end{aligned} \tag{9.64}$$

Let us now pay attention to the wonderful property of Eq. (9.63) with the Gutzwiller trace formula (9.13). The similarity becomes obvious for systems scaled so that $\hbar = 1$, and the orbits have the Maslov index $\mu_p = 0$ and action $I_p(E) = E\log p$, so the period is $T_p = \log p$. Therefore, if there is a Hamiltonian system with quantum levels coinciding with the nontrivial zeros of the Riemann zeta function, then it should have periodic orbits with periods equal to the logarithms of prime numbers. Therefore, this profound analogy allows us to associate the spectrum structure of the Riemann zeta function's nontrivial zeros with the quantum spectrum of the hypothetical dynamical system.

The existence of such systems "prove" the Riemann hypothesis at once. This is an interesting direction, but not the most promising for quantum chaos, which tends to penetrate deeply into the spectral structure of the zeros of the Riemann zeta function, for a better understanding of the direct analogue of the quantum spectrum. The attractiveness of such an analogy is linked also to the possibility of achieving significant statistics in the case of the Riemann zeta function, allowing us to penetrate reliably into the "semiclassical region" ($E \to \infty$). The distribution shown in Fig. 9.9 includes 79 million zeros in a neighborhood of the zero with ordinal number 10^{20}, and ideally reproduces the Wigner distribution $P_W(S)$ for the chaotic system.

9.7 Signatures of Quantum Chaos in Wave Function Structure

As we have seen in the previous section, the statistical properties of spectra strictly correlate with the type of classical motion. It is natural to try to discover analogous correlations in the structure of wave functions, i.e., to assume that the form of the wave function for a semiclassical quantum state depends on whether it is associated with classical regular or chaotic motion. It is important to note that in the analysis of QMCS at the level of energy spectra, the principal role was played by the statistical characteristics, i.e., quantum chaos was treated as a property of a group of states.

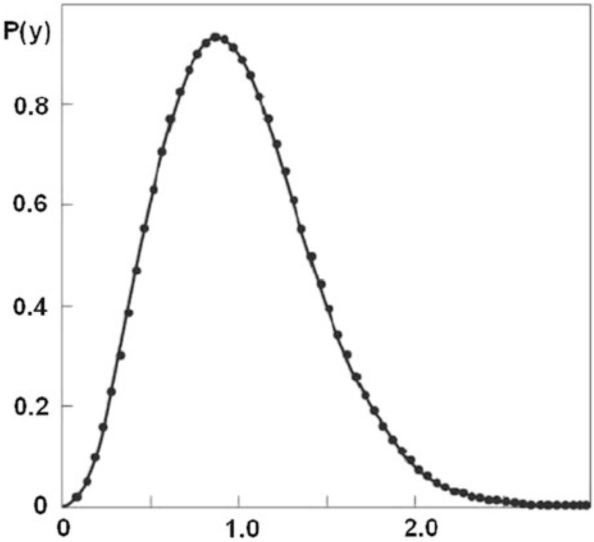


Fig. 9.9 Distribution function of distances between the nontrivial zeros of the Riemann zeta function [36] $P(y)$

The choice of a stationary wave function of the quantum system, which is chaotic in the classical limit, as the basic object of study attributes the phenomenon of quantum chaos to an individual state. The joint study of both possibilities is not a contradiction, but it rather reflects the universality of the phenomenon.

Unlike the spectrum, the form of wave functions depends on the basis in which they are determined. The following three representations are most often used for QMCS studies:

1. The so-called H_0 representation is the representation of eigenfunctions $\{\phi_n\}$ of the integrable part H_0 of the total Hamiltonian $H = H_0 + V$. The main objects of investigation in this case are the coefficients of the expansion C_{mn} of the stationary wave functions ψ_m in the basis $\{\phi_n\}$. The H_0 representation is naturally realized from numerical calculations, as the diagonalization of the Hamiltonian H is realized most often in this representation.
2. The coordinate representation $\Psi(q)$ is the most convenient from the visual point of view, and it allows direct comparison with classical motion along the trajectories in coordinate space. Indeed, according to Schnirelman's theorem [37], the mean probability density $|\Psi(q)|^2$ in the semiclassical limit coincides with the projection of the microcanonical distribution on the coordinate space:

$$\langle|\Psi(q)|\rangle \to \rho\,(q) = \frac{\int dp'\delta(E - H(p',q))}{\int dp'dq'\delta(E - H(p',q'))}. \tag{9.65}$$

In fact, this theorem interpolates the correspondence principle on wave functions. What can we expect from the wave function structure on the basis of this theorem? This is the concentration of the probability density in the areas of the coordinate space covered by (quasi)periodic trajectories for the regular wave functions, in contrast to an almost uniformly blurred probability density for the chaotic wave functions.

3. Representation by means of Wigner functions [38, 39], which have a set of properties in common with the classical distribution function in the phase space, which facilitates the search for QMCS.

As early as 1977, Berry [40] assumed that the form of the wave function for a semiclassical regular quantum state (associated with the classical motion on an N-dimensional torus in $2N$-dimensional phase space) is very different from the form of the wave function for an irregular state (associated with the stochastic classical motion on all or part of the $(2N-1)$-dimensional energy surface in all or part of phase space). For regular wave functions, the average probability density in configuration space is determined by the projection of the corresponding quantized invariant tori onto the configuration space, which means the global order. The local structure is implied by the fact that the wave function is locally a superposition of a finite number of plane waves with the same wave number, as determined by the classical momentum. In the opposite case for the average of chaotic wave functions (over small intervals of energy and coordinates), the probability density in the semiclassical limit $\hbar \to 0$ coincides with the projection of the classical microcanonical distribution onto the coordinate space [see (9.65)]. Its local structure is described by the superposition of an infinite number of plane waves with random phases and the same wave numbers. The random phases might be justified by classical ergodicity. This assumption leads immediately to the Gaussian distribution for the amplitude of probability. This structure of the wave function is in good agreement with the picture of the classical phase space. The classical trajectory homogeneously fills the isoenergetic surface in the case of chaotic motion. In contrast, from the consideration of the regular quantum state as an analogue of the classical motion on a torus, a conclusion can be drawn about the singularity (in the limit $\hbar \to 0$) of the wave function near the caustics (i.e., the boundaries of the region of the classically allowed motion in a coordinate space).

Berry's hypothesis was subjected to the most complete test for billiards of different types and, in particular, for stadium-type billiards [41]. The results have proved Berry's hypothesis, but the probability density distribution appears to be not as homogeneous as could have been expected because of the classical motion ergodicity. Indeed, the observable inhomogeneities, called scars of the wave functions, are the traces of the classical unstable periodic orbits [42].

The R-C-R transition considered in Sect. 9.3 represents an attractive possibility to check Berry's hypothesis for potential systems. Let us start from the H_0 representation, or more exactly the representation of linear combinations of wave functions $|NL\rangle$ of a two-dimensional harmonic oscillator with the equal frequencies

$$\psi_k = \sum_{N,L} C_{NLj}^{(k)} |NLj\rangle, \tag{9.66}$$

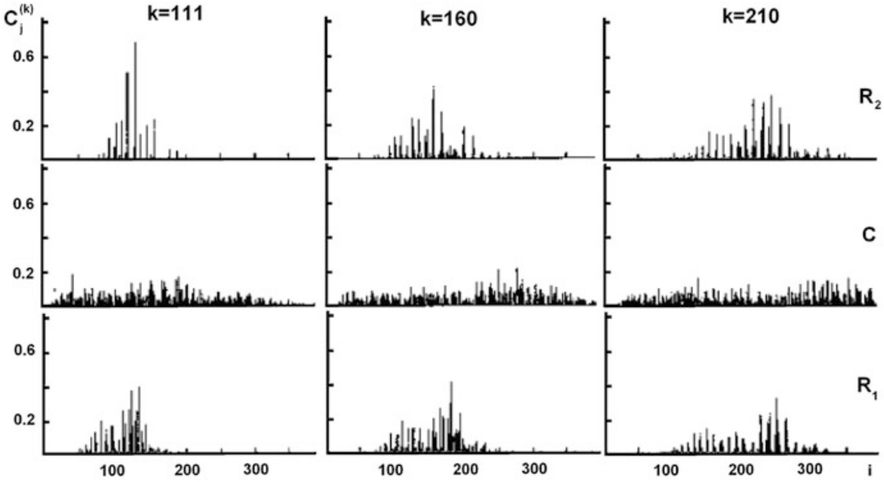


Fig. 9.10 Distribution of coefficients $C_{NLj}^{(k)}$ on the number $i = \{NLj\}$ of the basis states

where

$$
\begin{aligned}
|NLj\rangle &= \frac{P_{L,j}}{\sqrt{2}}\left(|NL\rangle + j|N,-L\rangle\right), \quad j = \pm 1, \\
&\quad N = 0,1,2\ldots, \quad L = N, N-2, \ldots 1 \text{ or } 0, \quad P_{L,j} = j^{\mathrm{Mod}(L,3)}, \\
\left\langle NLj \mid N'L'j' \right\rangle &= \frac{1}{\sqrt{2}} 2^{\delta_{L0}} \delta_{NN'} \delta_{LL'}. \qquad (9.67)
\end{aligned}
$$

If we introduce the notion of distributivity of the wave function in this basis, then the criterion of stochasticity formulated by Nordholm and Rice [43] states that the average degree of distributivity of the wave functions increases with the degree of stochastization in the system. It is clear that this criterion is a direct analogue of Berry's hypothesis for H_0 representation if we interpret the number of the basis state $i = \{NLj\}$ as a discrete coordinate. Figure 9.10 confirms qualitatively this criterion. It can be seen from this figure that the states that correspond to regular motion (areas R_1 and R_2) are distributed in a relatively small number of the basis states. At the same time, states corresponding to the chaotic motion (area C) are distributed in a considerably larger number of basis states. In the latter case, the contributions from a large number of basis states in the expansion (9.66) interfere; this results in a complicated spatial structure of the wave function $\psi_k(x,y)$.

The correlations between the structure of the wave function and the type of classical motion are demonstrated also in Fig. 9.11, where the probability density $|\psi_k(x,y)|^2$ is represented for the states numbered 111, 160, and 210 [17]. The squared module of the wave function reproduces rather well the transition from the function with well-defined structure (area R_1) to the irregular distribution (area C), and further structure restoration in the second irregular area (R_2).

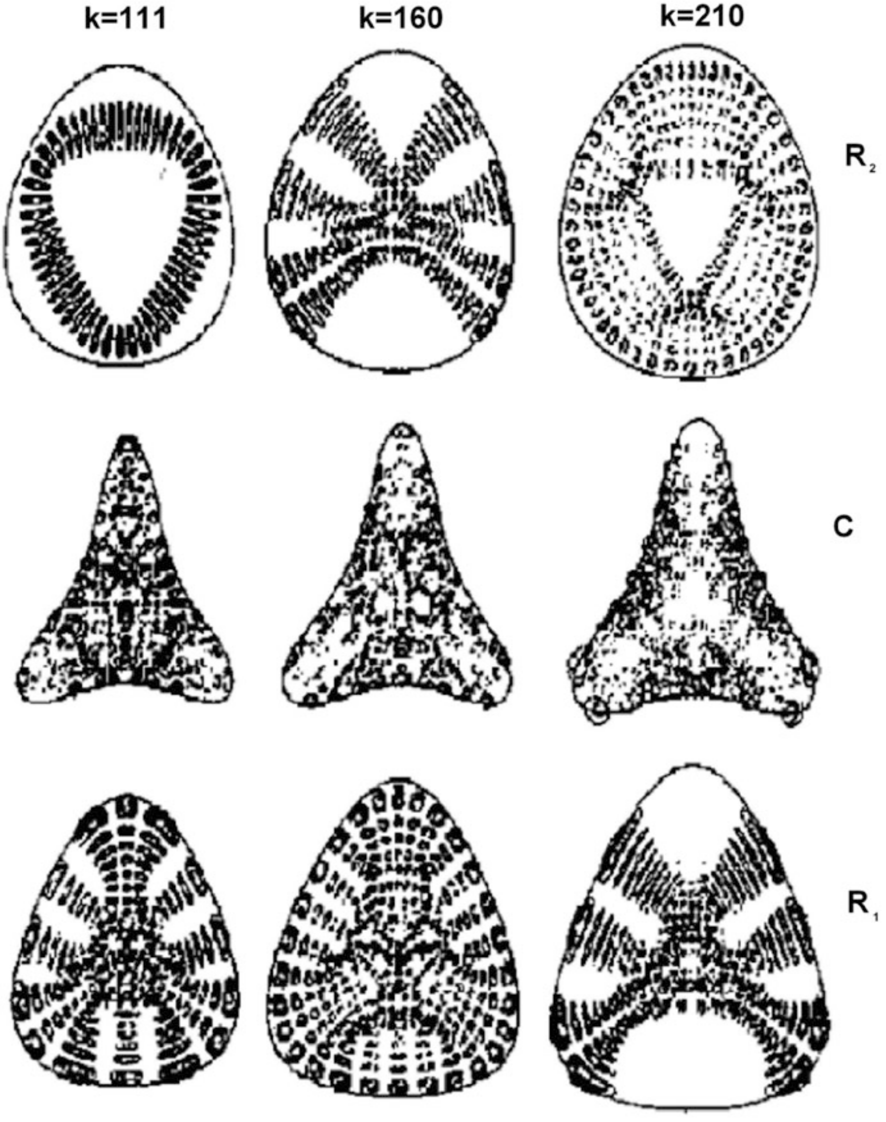


Fig. 9.11 Isolines of the probability density $|\psi_k(x,y)|^2$

For the chosen technique, in which the transition is traced for the wave function with fixed number k (due to the transformation of effective Planck constant $\bar{\hbar}$), the changes in the wave function are associated only with the R-C-R transition.

The specific nature of the stationary wave functions, corresponding to a certain type of classical motion, appears also in the level lines structure, in particular in the manifolds $\Psi(q) = 0$ structure. Depending on the dimensionality of the space

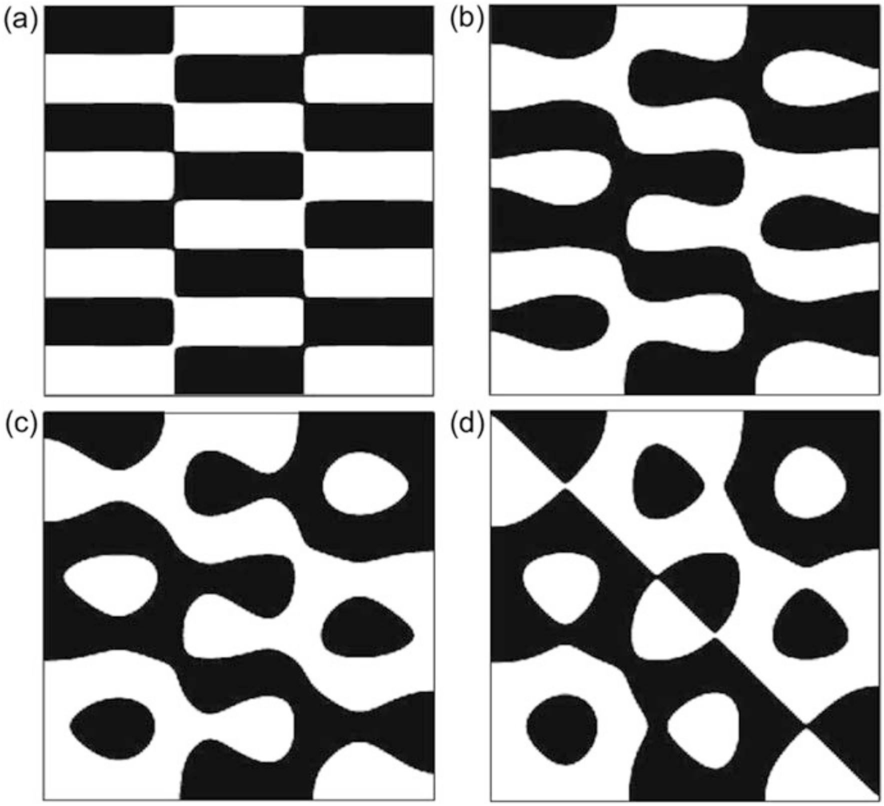


Fig. 9.12 Structure of the nodal domains of the stationary wave functions of the rectangular billiard $\psi\ (x, y) = \sin 3\pi x \cos 8\pi y + \varepsilon \sin 8\pi x \cos 3\pi y$ for $\varepsilon = 0.01$ (**a**), $\varepsilon = 0.1$ (**b**), $\varepsilon = 0.6$ (**c**), and $\varepsilon = 0.99$ (**d**)

configuration, these manifolds are called nodal points ($N = 1$), nodal lines ($N = 2$), or nodal surfaces ($N > 2$).

In 1979, Stratt et al. [44] proposed the criterion of quantum chaos, directly related to the nodal geometry. For the regular states, the system of nodal lines is a grid of quasiorthogonal curves (or is very close to them); but for chaotic spaces, this representation is absent. For two-dimensional systems, it is convenient to describe not the nodal lines themselves, but rather the constant sign areas of the wave function, the so-called nodal domains, so that the boundaries between the domains coincide with the nodal lines. Figure 9.12 shows the evolution of the nodal structure of the wave functions of rectangular billiards, which confirms the proposed criterion.

The change in the nodal structure can be also traced during R-C-R transition in the Hamiltonian of quadrupole oscillations (Fig. 9.13). This structure substantially changes during the R-C-R transition. The spatial structure of the nodal curves for

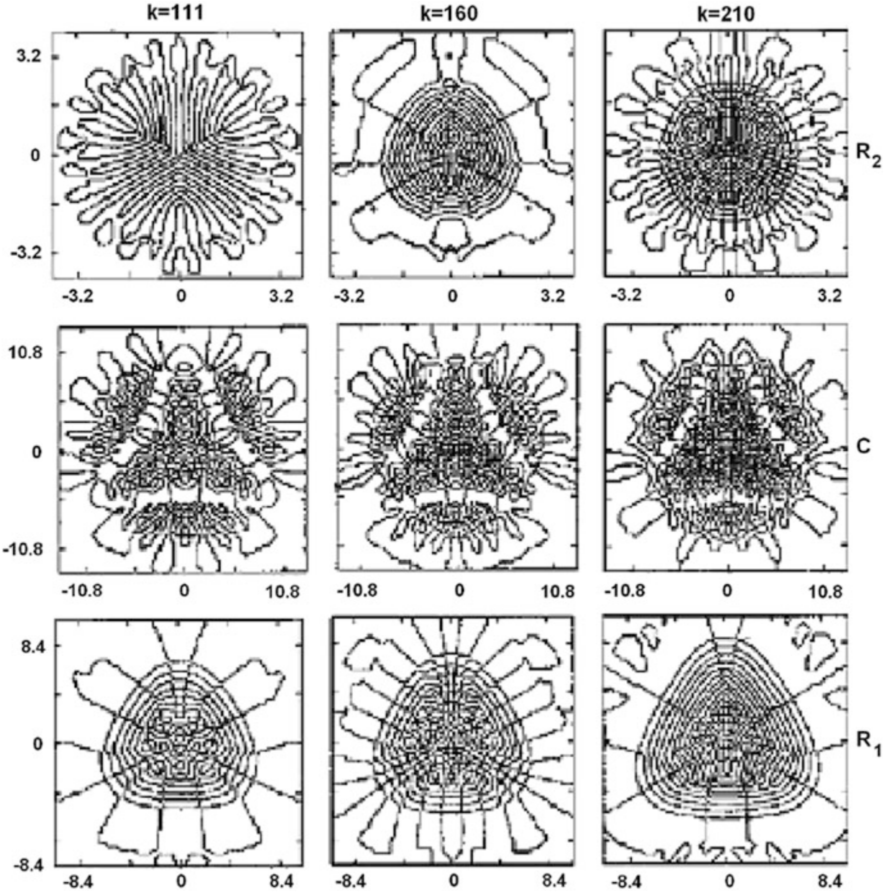


Fig. 9.13 Nodal curves of the wave functions $\psi_k(x, y)$

states from areas R_1 and R_2 of the regular classical motion is much simpler than the same structure for the states from the chaotic area C.

The new approach to QMCS study for the structure of the wave functions can be realized in the potentials with two or more local minima [45]. The main advantage of this approach is the possibility to detect QMCS in the different parts of the same wave function, but not in different wave functions. The efficiency of this method is shown in Fig. 9.14 for two potentials: surface quadrupole oscillations and the umbilical catastrophe D_5. Comparing the structure of eigenfunctions in the central and peripheral minima of the potentials of quadrupole oscillations (Fig. 9.14c), or in the left and right minima of the potential D_5 (Fig. 9.14d), it becomes obvious

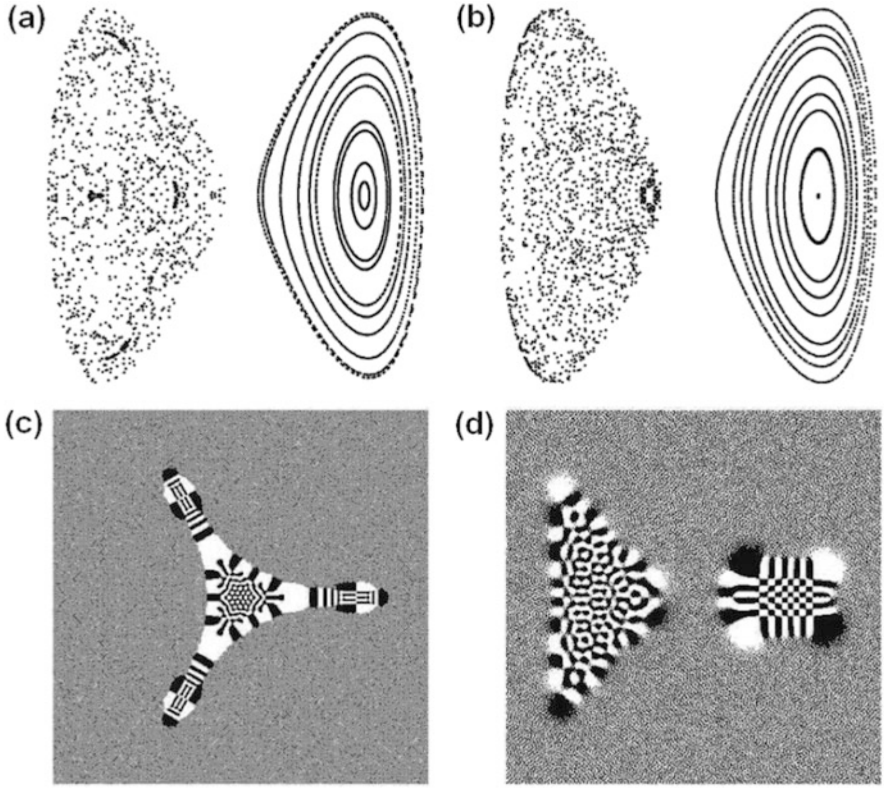


Fig. 9.14 Mixed state in the potentials of quadrupole oscillations (**a**), (**c**) and D_5 (**b**), (**d**): (**a**), (**b**) are the Poincaré cross section, while (**c**), (**d**) are the nodal domains of eigenfunctions

that the nodal structures of the regular and chaotic parts of the eigenfunction clearly differ:

1. Within the classically allowed domain, the nodal domains of the regular part of the wave function form a clearly distinguishable checkerboard pattern [9]; nothing of the kind is observed in the chaotic part.
2. Nodal lines of the regular part show intersections or very narrow quasi-intersections; in the chaotic part, there is a distinctive trend to avoid the quasi-intersections.
3. While crossing the classical turning line $U(x, y) = E_n$, the nodal line structure of the regular part switches immediately to straight nodal lines, going to infinity, which makes the turning point line itself easily localisable in the nodal domains' structure; in the chaotic part, there exists an intermediate area around the turning point line where some of the nodal lines pinch off, making the transition to the classically forbidden domain more graduated and not so apparent in the nodal structure.

References

1. Berry, M.: In: Giannoni, M.-J., Voros, A., Zinn-Justin, J. (eds.) Chaos and Quantum Physics. Les Houches, Session LII, pp. 251–304. Elsevier Science Publisher (1991)
2. Chirikov, B.V.: Open Syst. Inform. Dyn. **4**, 241–280 (1997)
3. Kolmogorov, A.N.: Russ. Math. Surv. **38**(4), 27–36 (1983)
4. Landau, L.D., Lifshitz, E.M.: Mechanics, 3rd ed. Pergamon Press, London (1976)
5. Landau, L.D., Lifshitz, E.M.: Quantum Mechanics. Pergamon Press, London (1982)
6. Arnol'd, V.I.: Mathematical Methods of Classical Mechanics. Springer, New York (1989)
7. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)
8. Selberg, A.: J. Indian Math. Soc. **20**, 47–87 (1956)
9. Gutzwiller, M.: Chaos in Classical and Quantum Mechanics. Springer, Berlin (1991)
10. Toda, M.: Phys. Lett. **48**, 335–240 (1974)
11. Brumer, P.: Chem. Phys. **47**, 201–269 (1981)
12. Bolotin, Yu.L., Gonchar, V.YU., Inopin, E.V.: Yad. Phys. **45**, 350–356 (1987)
13. Eizenberg, J., Greiner, W.: Microscopic Theory of the Nucleus. North-Holland, Amsterdam/London (1976)
14. Mosel, V., Greiner, W.: Z. Phys. **217**, 256 (1968)
15. Bartz, B.I., Bolotin, Yu.L., Inopin, E.V., et al.: Metod Hartri-Foka v Teorii Yadra. Naukova Dumka, Kiev (1982)
16. Henon, M., Heiles, C.: Astron. J. **69**, 73–79 (1964)
17. Berezovoj, V.P., Bolotin, Yu.L, Gonchar, V.Yu., Granovsky, M.Ya.: Part. Nucl. **34**(2), 388 (2003)
18. Poston, T., Stewart, I.: Catastrophe Theory and Its Applications. Dover, New York (1998)
19. Feit, M.D., Fleck, J.A. Jr., Steiger, A.: J. Comput. Phys. **47**, 412–433 (1982)
20. Wigner, E.P.: Ann. Math. **62**, 548–564 (1955)
21. Porter, C.E.: In: Porter C.E. (ed.) Statistical Theories of Spectra: Fluctuations. Academic, New-York/London (1965)
22. Dyson, F.J.: J. Math. Phys. **3**(1), 140–156 (1962)
23. Mehta, M.I.: In: Porter C.E. (ed.) Random Matrices and Statistical Theory of Energy Levels. Academic, New-York/London (1967)
24. Bohigas, O., Giannoni, M., Schmit, C.: Phys. Rev. Lett. **52**, 1–4 (1983)
25. Seligman, T.H., Nishioka, H. (eds.): Quantum Chaos and Statistical Nuclear Physics. Springer, Heidelberg (1986)
26. Seligman, T., Verbaarshot, J., Zirnbauer, M.: Phys. Rev. Lett. **53**, 215–218 (1984)
27. Bohigas, O., Hag, R.V., Pandey, A.: Phys. Rev. Lett. **54**, 645–649 (1985)
28. Haake, F.: Quantum Signatures of Chaos. Springer, Berlin (1991)
29. Chirikov, B.: Phys. Rep. **52**, 263–379 (1979)
30. Brody, T.A., Flores, J., French, J.B., Mello, P.A., Pandey, A., Wong, S.S.: Rev. Mod. Phys. **53**, 385–479 (1991)
31. Berry, M.V., Robnik, M.: J. Phys. A **17**(12), 2413–2421 (1984)
32. Bogomolny, E.B.: JETF Lett. **41**(2), 55–58 (1985)
33. Stöckmann, H.J.: Quantum Chaos: An Introduction. Cambridge University Press, Cambridge (1999)
34. Dietz, B., Richter, A.: Chaos **25**, 097601 (2015)
35. Brack, M., Bhaduri, R.K: Semiclassical Physics. Wesley, New York (1997)
36. Odlyzko, A.M.: AT and T Bell Lab, preprint (1989)
37. Schnirelman, A.I.: Usp. Mat. Nauk **29**, 181–182 (1974)
38. Berry, M.V.: Philos. Trans. R. Soc. A **287**, 237–371 (1977)
39. Hutchinson, J.S., Wyatt, R.E.: Chem. Phys. Lett. **72**(2), 378–384 (1980)
40. Berry, M.V.: J. Phys. A **10**(12), 2083 (1977)

41. McDonald, S.W., Kaufman, A.N.: Phys. Rev. Lett. **42**(18), 1189–1191 (1979)
42. Heller, E.J.: Phys. Rev. Lett. **53**(16), 1515–1518 (1984)
43. Nordholm, K.S.J., Rice, S.A.: J. Chem. Phys. **61**(1), 203–223 (1974)
44. Stratt, R.M., Handy, C.N., Miller, W.H.: J. Chem. Phys. **71**(8), 3311–3322 (1979)
45. Berezovoj, V.P., Bolotin, Yu.L., Cherkaskiy, V.A.: Phys. Lett. A **323**, 218–223 (2004)

Chapter 10
Tunneling and Chaos

Since the early days of quantum mechanics, tunneling has been considered one of its symbols: the ability of quantum particles to penetrate through energy barriers is one of the most impressive applications of quantum theory. This phenomenon has numerous applications in atomic, molecular, and nuclear physics as well as in solid state physics. Almost a century has elapsed since its discovery, and the various aspects of tunneling have been studied, but only in the last decades has it become clear that despite its significantly quantum nature, tunneling is largely, if not entirely, determined by the structure of the classical phase space of the system. The transition of classical dynamics of integrability to chaos substantially modifies the tunneling process. The purpose of this chapter is to make a clear, at first glance, paradoxical statement: tunneling is absent in classical physics, but the structure of the classical phase space defines a purely quantum effect of the tunneling!

10.1 Tunneling in One-Dimensional Systems

As is well known, a quantum system can be completely described by its wave function $|\psi\rangle$, which represents a vector in Hilbert space in mathematical language. In particular, an isolated spinless particle can be represented by a wave function in the coordinate $\langle x|\psi\rangle$ or momentum $\langle p|\psi\rangle$ spaces. The same object in classical mechanics is described in the phase space by the coordinates x, p. The question is this: how can the classical time evolution described by Hamiltonian equations in terms of x, p be reflected in the quantum reality using a single vector $|\psi\rangle$ satisfying the Schrödinger equation? It seems natural to associate the classical particle with the quantum state $|\psi\rangle$, which is localized optimally in the vicinity of the particle's position in the classical phase space at any given time t. But quantum mechanics imposes a fundamental restriction on the limit of the localization, expressed by the

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5_10

uncertainty relation

$$\Delta x \cdot \Delta p \geq \hbar/2.$$

Similar relations also exist for other canonically conjugate variables. Hence, the maximum that we can hope for is to localize the particle in the domain of the finite dimensions (Δx, Δp) around the position of the particle (x, p), with Δx, Δp being significantly less than the scales characterizing the classical trajectory. This should satisfy our desire to construct a quantum state that imitates the classical motion, provided that $|\psi\rangle$ follows the classical time evolution of x, p and Δx, Δp remain small over time. Thus classical bodies are linked to the center of the gravity trajectory even if they have finite dimensions. The quantum state, which possesses such properties, at least in finite times, is called a wave packet (WP). Indeed, the WP is the tunneling object. Let us consider [1] a quantum system with Hamiltonian H, which performs one-dimensional finite motion. Its discrete spectrum E_i and stationary wave functions ψ_i are solutions of the equation

$$H\psi_i = E_i\psi_i.$$

The arbitrary WP $\Psi(t)$, initially localized in a domain R, can be represented as a superposition

$$\Psi(t = 0) = \sum_i c_i\psi_i.$$

The property of the initial localization of the WP means that the integral of the square modulus of the wave function at time $t = 0$ is close to the normalization of the wave function

$$\int_R |\Psi(t = 0)|^2 dx \simeq 1.$$

The temporal evolution of the WP is determined by the time-dependent stationary wave functions:

$$\Psi(x, t) = \sum_i c_i \exp\left(-iE_i t/\hbar\right) \psi_i(x).$$

The probability $p_R(t)$ of remaining in the domain R at later moments t is a linear combination of periodic (trigonometric) functions with frequencies $|E_i - E_k|/\hbar$, where the indices i, k comply with all the components entering the WP. The typical time required for the WP to return to the initial state is of order the least common multiple of the inverse frequencies. Since in the general case, the frequencies are incommensurable, the WP will broaden with time to all available phase space.

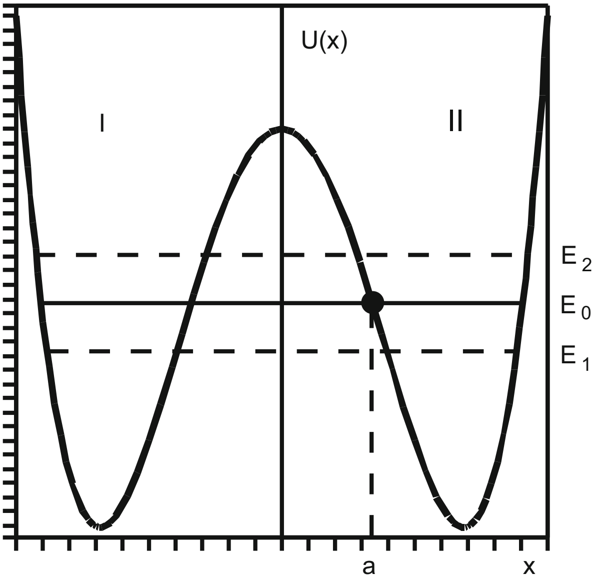


Fig. 10.1 Level splitting in symmetric one-dimensional potential double well

Packet components corresponding to the high frequencies (big$|E_i - E_k|$) will expand faster than the low-frequency components. The simplest WP is a packet consisting of only two components of the wave functions ψ_1 and ψ_2 and energies E_1 and E_2. This situation can adequately reflect the situation in the spectrum for which $|E_1 - E_2| \ll |E_i - E_1| \sim |E_i - E_2|$ for all other states ψ_i. In this case,

$$p_R(t) = p_R(0) - 4A\sin^2\left(|E_1 - E_2|\, t/\hbar\right), \tag{10.1}$$

where

$$A = c_1 c_2 \int\limits_R \psi_1 \psi_2 dx$$

($c_{1,2}$ and $\psi_{1,2}$ are assumed to be real). This approximation can be used to describe the tunneling process in a one-dimensional symmetric double well, as shown in Fig. 10.1.

If the potential barrier between the wells *I* and *II* is impenetrable for a particle, then there are energy levels corresponding to the motion of particles in each well, and the same for the two wells. Consideration of the tunneling (the possibility of transition through the potential barrier) leads to the splitting of each of these levels into two close levels. The wave functions corresponding to these almost degenerate levels (due to the smallness of the interactions between the wells) describe the motion of a particle simultaneously in both wells. Let $\psi_0(x)$ be the semiclassical wave function describing the motion with energy E_0 in an isolated well I, that is, exponentially damped on both sides of the boundaries of this well. When the tunneling is taken into account, the function $\psi_0(x)$ ceases to be a stationary wave function of the overall system, and the level E_0 splits into the levels E_1 and E_2. The correct wave functions corresponding to these levels are symmetric and antisymmetric combinations of the functions $\psi_0(x)$ and $\psi_0(-x)$:

$$\psi_1(x) = \frac{1}{\sqrt{2}}\left[\psi_0(x) + \psi_0(-x)\right],$$

$$\psi_2(x) = \frac{1}{\sqrt{2}}\left[\psi_0(x) - \psi_0(-x)\right]. \tag{10.2}$$

Using the semiclassical approximation for $\psi_0(x)$ [2], one can show that

$$\Delta E \equiv E_2 - E_1 = \frac{\omega\hbar}{\pi}\exp\left(-\frac{1}{\hbar}\int\limits_{-a}^{a}|p|\,dx\right). \tag{10.3}$$

Here ω is the cyclic frequency of the classical periodic motion in a well at the energy E_0, $p = \sqrt{2m(E_0 - U(x))}$ is the imaginary subbarrier momentum of the particle of mass m, and a is the turning point corresponding to the energy E_0. Using the functions $\psi_1(x)$ and $\psi_2(x)$, we can construct in a two-level approximation the WPs $\psi^I(x) = \psi_0(x)$ and $\psi^{II}(x) = \psi_0(-x)$ localized at the initial time in wells I and II respectively,

$$\psi^I(x) = \frac{1}{\sqrt{2}}\left[\psi_1(x) + \psi_2(x)\right],$$

$$\psi^{II}(x) = \frac{1}{\sqrt{2}}\left[\psi_1(x) - \psi_2(x)\right]. \tag{10.4}$$

Each of the WPs tunnels from well to well with period T equal to

$$T = \frac{2\pi\hbar}{\Delta E}. \tag{10.5}$$

Tunneling from well to well, without changing the shape (coherent tunneling), is a feature of the symmetric potential. In the case of an arbitrary potential, the form of WP changes in the tunneling process: the WP form is sensitive to the details of the form of the potential barrier.

10.2 Dynamical Tunneling

Now we generalize the problem of one-dimensional tunneling discussed above to the case of higher dimensions. The degree of complexity of classical dynamics in multiple dimensions is qualitatively greater than that in one dimension. This fact is clearly manifested in the tunneling effect. This not only applies to the analysis of the usual problem of penetration through the barrier but also leads to the emergence of entirely new tunneling scenarios that have no analogues in one dimension. The features of multidimensional tunneling that interest us can be demonstrated on systems with one and a half and two degrees of freedom, and we restrict ourselves to them.

The simplest example of the fundamentally multidimensional effect associated with tunneling is dynamical tunneling [3]. The concept of dynamical tunneling occurs in systems whose phase space contains domains, the transition between which is prohibited at the level of classical mechanics, but the prohibition is not caused by the potential barrier. Of course, this effect is possible only in systems with more than one degree of freedom, where in addition to the energy, the integrals of motion are sources of the corresponding prohibitions. The new type of tunneling is more complex than traditional (potential barrier) tunneling. The complication is due to the fact that a simple consideration of the potential surface does not detect the prohibition conditions. Instead of static potential surfaces, we have to consider the behavior (dynamics) of the trajectories. Let us return to the above semiclassical solution of the problem of a one-dimensional symmetric double well in order to understand the nature of dynamical tunneling. In this case, when we quantize the system, considering each well separately, we obtain a spectrum consisting of the strictly degenerate doublets. Taking into account only the interaction between the wells due to the overlapping of the exponentially small tails of the wave functions, we get the right result: almost degenerate pairs of levels with the known splitting (10.3).

A situation similar to a one-dimensional symmetric double well may exist in multidimensional potential without the energy barrier. Let us examine [3–5] a dynamical system with reflective symmetry of the phase space T. Suppose that there are two unconnected areas in the phase space, A_1 and A_2, each of which is invariant with respect to the classical dynamics, with mapping to each other by the symmetry operations $A_2 = TA_1$. Consider for concreteness the case that the classical motion in $A_{1,2}$ is regular, that is, when these areas represent islands of stability immersed in a chaotic sea. The additional assumption is made that in the classical limit, there is a set of states $\psi(\mathbf{q})$ that are mainly localized in these areas. Using the standard procedure of quantization of integrable systems, we can independently quantize the motion in each of the areas and construct the degenerate wave functions (sometimes called quasimodes) $\psi^{(1)}(\mathbf{q})$ and $\psi^{(2)}(\mathbf{q}) = \psi^{(1)}(T\mathbf{q})$. In taking into account the interaction between the quasimodes, $\psi^{(1,2)}$ should be replaced by their symmetric and antisymmetric combinations

$$\psi^{(\pm)} = \frac{1}{\sqrt{2}}\left(\psi^{(1)}(\mathbf{q}) \pm \psi^{(2)}(\mathbf{q})\right). \tag{10.6}$$

The energy degeneracy between these functions is removed by tunneling processes. But in contrast to the one-dimensional case, the invariant tori in $A_{1,2}$ are not necessarily separated by an energy barrier in the configuration space. The transitions $A_1 \leftrightarrow A_2$ may be prohibited at the classical level of preservation of the integrals of motion other than energy. An example of this type of quantum doublet was first demonstrated in [3]. The authors of that work studied the spectrum and wave functions of the Hamiltonian

$$H = \frac{p_x^2 + p_y^2}{2} + V(x,y); \qquad V(x,y) = \frac{1}{2}\omega_x^2 x^2 + \frac{1}{2}\omega_y^2 y^2 + \lambda x^2 y. \tag{10.7}$$

The parameters used for the calculations are $\omega_x = 1.1$, $\omega_y = 1$, and $\lambda = -0.11$. In the spectrum obtained by the diagonalization of the Hamiltonian (10.7) in the basis of the harmonic oscillator, the pair of states $\psi^{(\pm)}(x,y)$ was found, like the ones shown in Fig. 10.2e, f, with abnormally low splitting of $\Delta E = 0.0001$ and with the energy $E = 13.59$. The sum and difference of these functions give the wave functions $\psi^{(1)}$ and $\psi^{(2)}$, which are similar to those shown in Fig. 10.2c, d. These wave functions are mainly concentrated in the vicinity of the classical periodic orbits shown in Fig. 10.2a, b. It is important to note that despite the absence of the potential barrier, the classical trajectories starting in the vicinity of the phase space of one of the periodic orbits never fall into a neighborhood of another. This

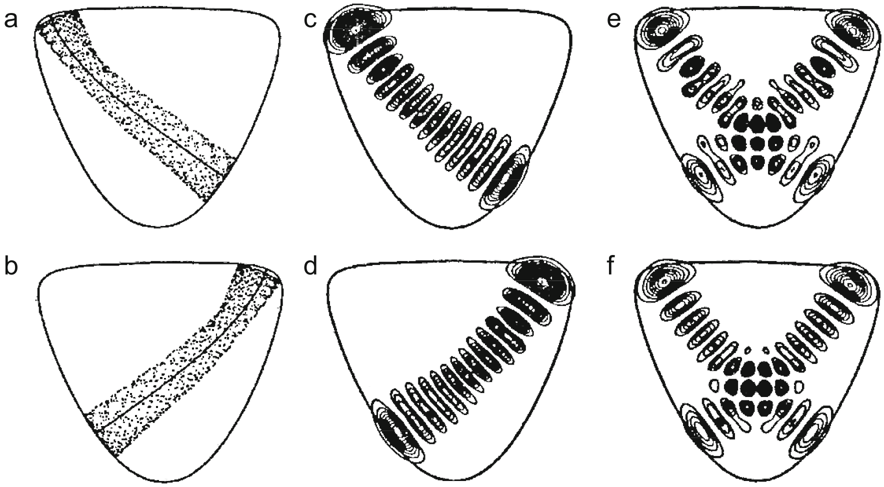


Fig. 10.2 Quantum doublets generated by dynamical tunneling

fact is a consequence of the presence of the additional (other than energy) integral of motion. So the classical trajectories can be “captured” on one or the other side of the dynamical barrier exactly like classical motion in a symmetric double well at energies below the barrier limited by one of the potential wells. The quantum-mechanical situation is fundamentally different. The quasiperiodic orbits, which are similar to the orbits shown in Fig. 10.2a, b with independent quantization, give a pair of degenerate wave functions, as shown in Fig. 10.2c, d. When taking into account the interaction (tunneling), the wave functions $\psi^{(1,2)}$(quasimodes) are not correct eigenfunctions. They should be replaced by the symmetric and antisymmetric combinations $\psi^{(\pm)}(6)$ (Fig. 10.2e, f). These pairs are split in frequency by a factor of 10^{-4}, while the fundamental frequency of the oscillations in the potential well is of order unity. The system initially prepared in a state $\psi^{(1)}$ representing the linear combination of the states $\psi^{(\pm)}$, $\psi^{(+)} \pm \psi^{(-)}$, as in the one-dimensional case, may make transitions between this state and the state $\psi^{(2)}$ with a frequency of order $\Delta E/\hbar$. Naturally, the question arises what factors determine the oscillation frequency of the WP, and in particular, if in the phase space of the system there are two symmetric stability islands separated by a chaotic sea, then how this sea affects the (dynamical) tunneling between the islands.

10.3 Dynamical Tunneling: Anharmonic Oscillator with a Periodic Perturbation

A preliminary response to the above question was obtained by Lin and Ballentine [6, 7]. The authors examined a model of a periodically excited double well in which the degree of stochasticity can be adjusted by changing the amplitude of the exciting force. The model Hamiltonian is

$$\begin{aligned} H &= \frac{P^2}{2M} + Bx^4 - Dx^2 + \lambda x \cos \omega_0 t, \\ M &= 1,\ B = 0.5,\ D = 10. \end{aligned} \tag{10.8}$$

The classical equations of motion are

$$\begin{aligned} \dot{x} &= P/M, \\ \dot{p} &= -4Bx^3 + 2Dx - \lambda \cos \omega_0 t. \end{aligned}$$

To solve the Schrödinger equation, we use the basis $\{|n\rangle\}$ of the harmonic oscillator with the Hamiltonian $\hat{H}$ equal to

$$\begin{aligned} \hat{H} &= \hat{p}^2/2M + \frac{1}{2}M\omega^2x^2, \\ \hat{H}|n\rangle &= \left(n+\frac{1}{2}\right)\hbar\omega\,|n\rangle\,. \end{aligned} \tag{10.9}$$

The Schrödinger equation in this representation is

$$i\hbar\frac{d}{dt}\langle m|\psi(t)\rangle = \sum_{n=0}^{\infty} H_{mn}\langle n|\psi(t)\rangle. \tag{10.10}$$

To calculate the matrix elements, we have to use the formula

$$\int_{-\infty}^{\infty} e^{-(x-y)^2}H_m(x)H_n(x)dx = 2^n\sqrt{\pi}m!y^{n-m}L_m^{n-m}\left(-2y^2\right), \quad m \le n.$$

Here, H_n are the Hermite polynomials, and L_m^n are Laguerre polynomials. The Schrödinger equation was solved for a given set $\langle n \mid \psi(0)\rangle$ with $0 < n < n_{\max}$. An analysis of the Schrödinger equation (10.10) can be conveniently performed using a transition to the Husimi functions [8] $\rho(x,p)$, defined as

$$\rho(x,p) \equiv (2\pi\hbar)^{-1}\left|\left\langle\phi_{x,p} \mid \psi(t)\right\rangle\right|^2, \tag{10.11}$$

where $\left|\phi_{x,p}\right\rangle$ is the minimum uncertainty wave packet [9], having average position x and average momentum p, or the coherent state of the harmonic oscillator [10]. In the basis $\{|n\rangle\}$,

$$\langle\phi_{x,p}|n\rangle = \exp\left(-\frac{1}{2}|\alpha|^2\right)\alpha^{*n}/\sqrt{n!},$$

where $\alpha = (x + igp/\hbar)/\sqrt{2g}$, $g = \hbar/M\omega$. The Husimi function provides a quantum analogue of the classical distribution function in the phase space.

We are using as parameters the driving force values $\lambda = 10$ and $\omega_0 = 6.07$. These parameters provide the following structure of the phase space (see Fig. 10.3): two islands of stability are centered at values of $x \simeq -1.5$, $p = 0$ and $x \simeq 4.15$, $p = 0$. We choose as the initial state the WP $|\psi(0)\rangle = \left|\psi_{x',p'}\right\rangle$, which is centered in the left island of stability in Fig. 10.3. It can be expected that the WP, which is initially localized in the classically regular area, will remain localized in this area or tunnel slowly therefrom into the chaotic region. Contrary to these expectations, it turned out that there are coherent tunneling oscillations between unrelated regular islands.

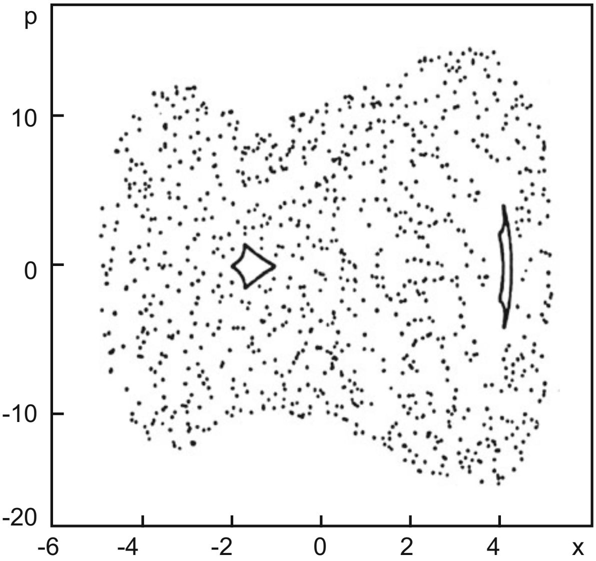


Fig. 10.3 Stroboscopic cross section of the phase trajectories of the Hamiltonian (10.8) [6]

Figure 10.4 shows the time evolution of the Husimi function obtained by numerical integration of the Schrödinger equation in the basis of the harmonic oscillator ($n_{\max} = 115$). Figure 10.4 presents the WP of minimum uncertainty centered in the left island of the stability. With $t = 115\ \tau_0\ \ (\tau_0 = 2\pi/\omega_0)$, there is a probability close to unity of finding a particle in the next potential well, and in $t = 230\tau_0$, the initial state will be restored with a fairly high degree of accuracy. The characteristic time of tunneling can be identified by a half-period of these oscillations. What is the nature of these tunneling oscillations?

As it turned out [11], the effect can be understood in terms very similar to those we used above in the description of dynamical tunneling. Due to the fact that the considered Hamiltonian is periodically dependent on time,

$$\hat{H}(t) = \hat{H}\left(t + \tau_0\right),$$

an arbitrary state can be written as a linear combination of the so-called quasienergy states [12]:

$$\psi(x,t) = \sum_{k=1}^{\infty} a_k \psi_k(x,t) = \sum_{k=1}^{\infty} a_k u_k(x,t) e^{-i\varepsilon_k t/\hbar},$$

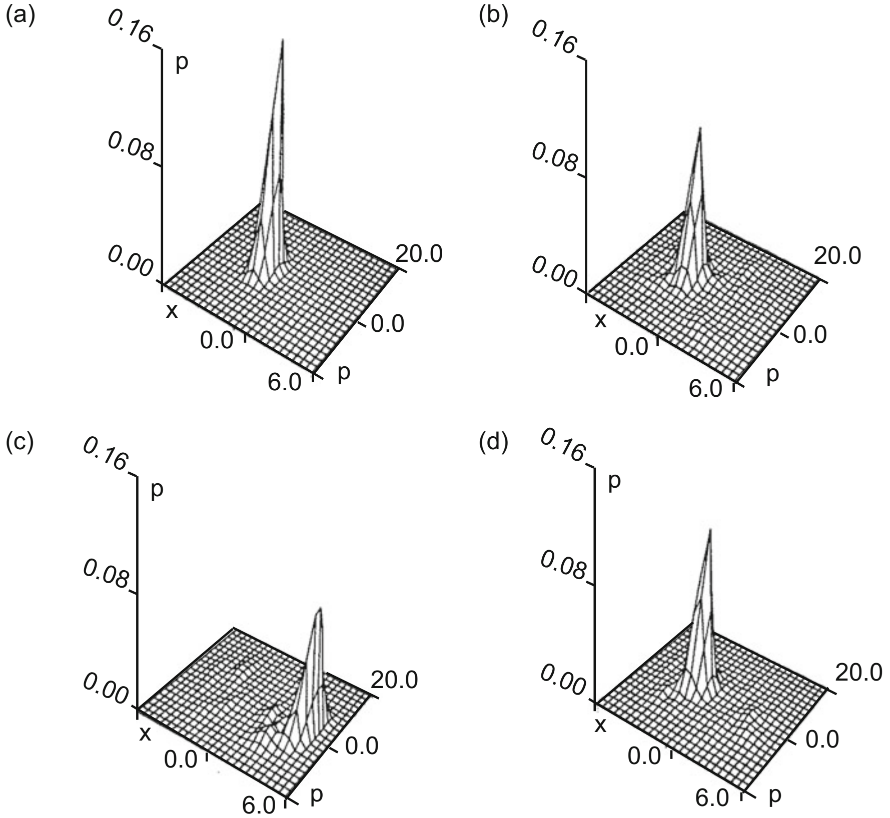


Fig. 10.4 The time-dependence of Husimi functions: (**a**) $t = 0$, (**b**) $t = 5\tau_0$, (**c**) $t = 115\tau_0$, (**d**) $t = 230\tau_0$ [6]

where ε_k is the quasienergy, and the function $u_k(x,t)$ satisfies the equation

$$\left(\hat{H}(t) - i\hbar\frac{\partial}{\partial t}\right) u_k(x,t) = \varepsilon_k u_k(x,t),$$

and furthermore,

$$u_k(x,t) = u_k(x,t+\tau_0).$$

The Hamiltonian (10.8) has a discrete symmetry:

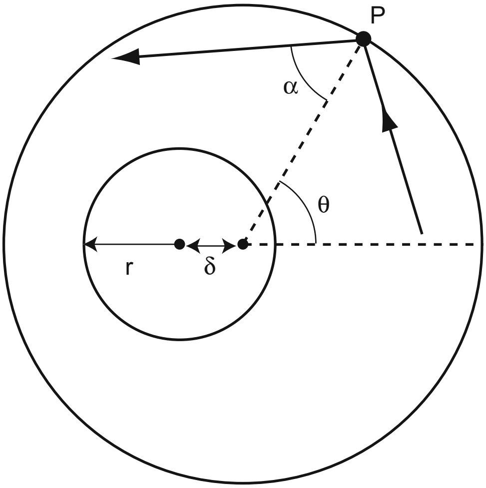
$$H(x,p,t) = H(-x,-p,t+\tau_0/2).$$

This means that the islands of stability in the classical system appear as symmetrically related pairs. For a quantum system, this leads to the fact that the quasienergy states are even or odd with respect to the replacement $x \to -x$, $t \to t + \tau_0/2$. Consequently the quasienergy states can be grouped in doublets. Each doublet includes even and odd members. If this doublet is localized within a pair of classically regular islands, then it is possible to form from them the WP, which is localized in only one of these islands. This doublet is coherently tunneled (forward and backward) between the regular islands, showing the phenomenon of quantum beating. The tunneling time is inversely proportional to the splitting of the quasienergy in the doublet. The doublets that are localized in such a way are called tunneling doublets. If a tunneling doublet exists, then the WP, initially localized in one of the regular islands, will have coherent tunneling oscillations. But this effect does not occur for doublets that are not localized in the regular islands. Unfortunately, the only way to be sure whether this doublet is localized is to perform a numerical simulation.

10.4 Annular Billiards as a Paradigm for Chaos-Assisted Tunneling

In this section we consider a dynamical system that makes it possible to study qualitatively and quantitatively the relation between the energy splitting that determines the tunneling velocity in the two-level approximation and the structure of the classical phase space [13]. The system is composed of an external circle of radius R ($R = 1$) and an internal circle (disk) of radius $r < R$ (see Fig. 10.5). A point particle moves uniformly in a straight line in the area between the two circles and has elastic mirror reflections on the boundary. In this setting, we are dealing with a two-parameter family of billiards, each of which is defined by a pair of real

Fig. 10.5 The geometry of a circular billiard



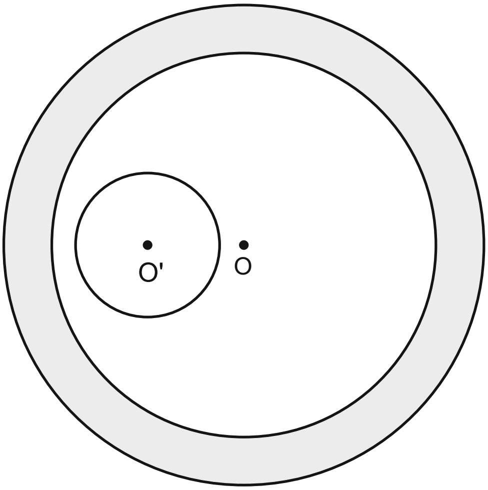


Fig. 10.6 The area (*gray*) of the whispering gallery trajectories

numbers (r, δ), where δ is the eccentricity (shift) of the disk center relative to the center of the external circle.

All the possible trajectories in billiards can be divided into two fundamentally different classes:

1. trajectories that never collide with the internal disk.
2. trajectories that collide at least once with the internal disk.

The trajectories that belong to the first class (and which play a crucial role in further discussion) are commonly called whispering gallery trajectories. They always exist, except in the case of $r + \delta = R$, and are within the shaded symmetric circular area bounded by the external circle and caustic corresponding to the whispering gallery trajectories that are tangent to the internal disk (see Fig. 10.6). For a one-parameter family of billiards $r + \delta = \text{const}$ $(0 < \text{const} < R)$, this area remains the same: if δ varies but $r + \delta$ is kept constant, then all whispering gallery trajectories remain undisturbed.

As the dynamic variables describing the evolution of the system, we can choose the values (see Fig. 10.5) $S = \sin\alpha$, where α is the angle of reflection at the external circle in the point P, and $L = \theta/2\pi R$, where θ is the arc length in P, counted from the intersection point of the line connecting the centers of the circles with the external circle that is closer to the center of the external circle than to the internal one. It is obvious that $|S| \leq 1$ and $|L| \leq 1/2$.

The structure of the phase space in the domain $r + \delta < |S| < 1$ is trivial. A torus in the phase space corresponds to each whispering gallery trajectory in configuration space. The cross section of the torus in the (L, S) plane is a horizontal line $S = \text{const}$; on collision with the external circle, the impact parameter S does not change. Every torus with a given value S can be associated with another torus $-S$, derived from the initial whispering gallery trajectory by changing the rotational direction.

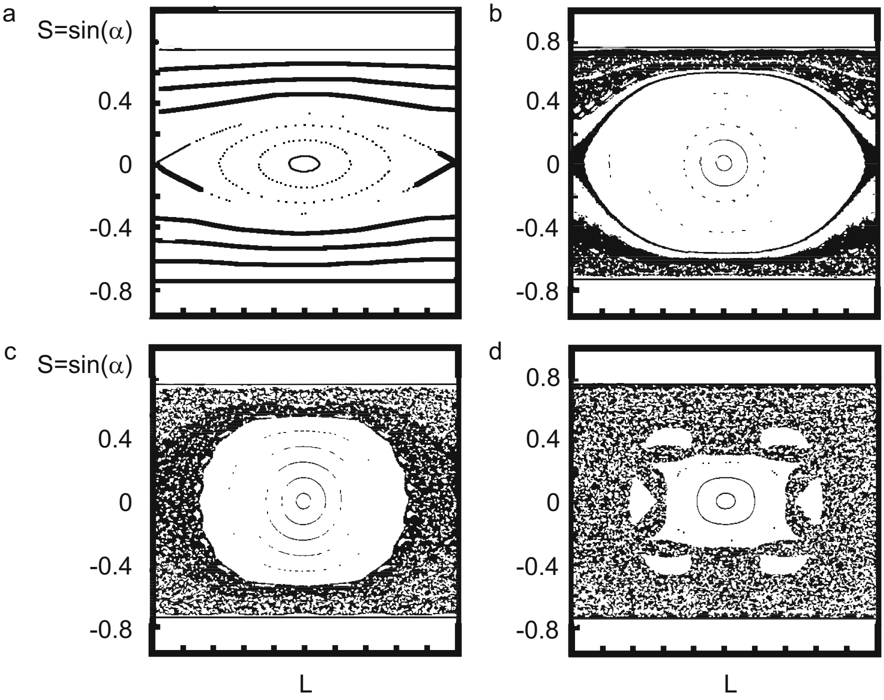


Fig. 10.7 Poincaré sections for $r + \delta = \text{const}$ and different eccentricities δ: (**a**) $\delta = 0.01$, (**b**) $\delta = 0.065$, (**c**) $\delta = 0.10$, (**d**) $\delta = 0.25$ [13]

The rays with a smaller impact parameter $S < r + \delta$ collide with the internal circle, and since the angular momentum in such collisions is not conserved, the motion is no longer integrable (the number of integrals of motion is less than the number of degrees of freedom). This leads to a number of phenomena associated with the concept of a mixed phase space of the nonintegrable system: the islands of stability, immersed in the chaotic sea.

Figure 10.7 shows some Poincaré sections for $r + \delta = 3/4$ and the different eccentricity values δ. For small values of δ (Fig. 10.7a), we can observe around a fixed point at the coordinate origin the presence of libration (or vibration) tori, corresponding to the trajectories that successively collide with internal and external circles. If we reverse this motion, the resulting trajectory will be based on the original torus. Outside the separatrix in the region $|S| \leq r + \delta$, there are rotational tori that like whispering gallery tori, appear in pairs corresponding to the direct and inverse rotation. But in this case (in contrast to the whispering gallery trajectories), the circular motion occurs with the collisions with the internal circle. The rotational tori (unlike the whispering gallery tori) deform when δ is changing, even if $r + \delta = \text{const}$. When δ is increasing (Fig. 10.7b), the region occupied by the libration tori is compressed: part of the vibration tori is destroyed and is replaced by the chaotic layers. The same holds for part of the rotational tori. In the end, the chaotic sea that appears connects the regions $S = \pm(r + \delta)$ of congruent whispering gallery tori (Fig. 10.7c). When δ is further increased, the region covered by libration tori keeps shrinking (see Fig. 10.7d), until it disappears for $\delta = r$.

The quantum mechanics of annular billiards [5, 13] with boundary conditions of Dirichlet type is described by the Helmholtz equation:

$$(\Delta + k^2)\psi(q) = 0, \tag{10.12}$$

and the requirement that the wave functions vanish at the internal and external boundaries of the billiards. The wave number k is associated with the energy by the relation $E = k^2$. The boundary conditions lead to the quantization of energy (wave number). For the case $\delta = 0$, due to the rotational symmetry, the orbital angular momentum $k \to k_{nm}$ is maintained (as well as the energy), where n is the quantum number of the angular momentum and m is the radial quantum number ($n = 0, 1 \ldots; \; m = 1, 2 \ldots$). Let us recall that the angular momentum quantum number n in the semiclassical limit is associated with the classical impact parameter S by the relation $S = n/k$. The stationary wave function of the circular billiards with $\delta = 0$ is coupled in energetically degenerate doublets, consisting of the components of the angular momentum with n and $-n$. In a system with eccentricity ($\delta \neq 0$), the degeneracy is removed by a violation of rotational invariance. But the doublets are perturbed to different degrees, depending on the relative values of n and $k(r+\delta)$. The symmetry-breaking strongly affects the doublets with a small value of the angular momentum $|n| < k(r + \delta)$, corresponding to the classical motion in which there are collisions with the internal circle. For doublets with small n, the degeneration is removed effectively with the increase of δ. The chaotic eigenstates appear and expand on the states between $-k(r + \delta)$ and $k(r + \delta)$. The higher doublets with the angular momentum $|n| > k(r + \delta)$ change only slightly under the violation of the symmetry (this is understandable, since the trajectories with such impact parameters do not collide with the internal circle, so its movement affects them slightly). The doublets are preserved, and the degeneration is slightly removed. The states are basically symmetric and antisymmetric combinations of n and $-n$ components of the angular momentum:

$$\left|\alpha^{(\pm)}\right\rangle \approx \frac{1}{\sqrt{2}} \left(|n\rangle \pm |-n\rangle\right). \tag{10.13}$$

As explained above, the energy splitting between $\left|\alpha^{(+)}\right\rangle$ and $\left|\alpha^{(-)}\right\rangle$ leads to the tunneling oscillations between the quasimodes $|\pm n\rangle$, linked with whispering gallery tori $\pm S = \pm n/k$ $(S > r + \delta)$. A quantum particle prepared in the state $|n\rangle$ will change its rotation from clockwise to counterclockwise and vice versa with the frequency $2\pi\hbar/\Delta E_n$. This tunneling process is a clear example of dynamical tunneling. It occurs in the phase space and not in the configuration space. Whispering gallery tori $\pm S$ are identical in configuration space. Also, in the tunneling process, the overcoming of the potential barrier in the configuration space does not occur. Indeed, the energy does not play any role in tunneling, because the energy is associated only with the absolute value of the momentum and not with its direction. The tunneling process violates the dynamic law of conservation of angular momentum for the rays with large impact parameters (for small impact parameters it is not conserved because of the internal circle).

We want to find out whether the classical dynamics corresponding to the states $|S| < r + \delta$ (chaotic sea) affect the splitting of quasidoublets, which are constructed on the whispering gallery tori $|S| > r + \delta$. In particular, it is interesting to consider the quasidoublets corresponding to the one-parameter family $r + \delta = \text{const}$. Indeed, when the eccentricity changes, the chaos measure also changes in the intermediate domains (between tori S and $-S$; see Fig. 10.7), but quantized whispering gallery tori remain unperturbed.

We now proceed directly to an analysis of the splitting magnitudes of quasidoublets defining the tunneling speed of wave packets. Why do we expect an increase in this speed? The semiclassical argumentation indicates that the probability distribution associated with the quantized torus decays exponentially outside the torus. A small overlap in the classically forbidden domain of the decaying distributions centered on two congruent quantized tori $(n, -n)$ leads to the tunneling splitting. If there is no chaotic domain between the tori, this overlap is small. However, if the chaotic area between the tori exists, the wave functions corresponding to the tori link first with the chaotic state. Due to the ergodic nature of the chaotic wave functions (the homogeneous distribution of the probability density), the connection between the two tori seems to be more effective than in the case of a regular intermediate state. Hence we can expect that tunneling will be reinforced by the presence of a chaotic region.

To make this argumentation more convincing, it should be confirmed by numerical calculations. Bohigas et al. [13] investigated numerically the transformation of a set consisting of the five tunneling quasidoublets of a one-parameter family of circular billiards with $r + \delta = 0.75$. The observed general tendency represented a dramatic increase in the splitting of quasidoublets (several orders of magnitude) with the increase of the proportion of chaotic phase space (growth of δ). Such a tendency suggests the idea [13] of calling the observed effect chaos-assisted tunneling (CAT). An important feature of the CAT effect is that splitting of quasidoublets increases with an increase in δ. That is why the following observations are important for understanding the physics of this effect:

1. The dependence of the splitting of the eccentricity δ is determined by two independent characteristics of the quasidoublet: the position in the spectrum and the position in the phase space.
2. Quasidoublets that are higher on the spectrum have less splitting.
3. Quasidoublets that are nearest to a chaotic sea have maximum splitting.

It is important to note that the numerical results show significant fluctuations of splittings as a function of eccentricity. Each of the fluctuations is associated with quasicrossing "outsider" chaotic level and considered regular (tunnel) quasidoublet. Because there is no dynamic partitioning of the chaotic region in an unrelated symmetric field, as is the case in the regular part of the phase space, we have no reason to expect chaotic doublets. Every chaotic state has a fixed parity. At the quasicrossings, the local domain of the spectrum can be described by a three-level model [4] with the Hamiltonian

$$\hat{H} = \begin{pmatrix} E+\varepsilon & 0 & 0 \\ 0 & E-\varepsilon & v \\ 0 & v & E^c \end{pmatrix}.$$

Here, $E \pm \varepsilon$ are the energies of a regular quasidoublet consisting of the symmetric and antisymmetric combinations of the respective quasimodes. In accordance with this model, the even chaotic state $|C\rangle$ of the energy E^c is associated with the state $|+\rangle$, having energy $E-\varepsilon$ as a constant of the interactions v. In practice, one of the constants v and ε dominates over the other. If ε dominates, then we return to the two-level model that has already been considered. Therefore, we assume that the coupling between the states $|C\rangle$ and $|+\rangle$ is dominant, and we set $\varepsilon = 0$. In this case, the splitting of the quasidoublet $\varDelta E$ is determined by the shift E_+ due to the interaction v. The diagonalization of $\hat{H}$ leads to

$$\varDelta E = \begin{cases} \frac{v^2}{E-E^c} & E-E^c \gg v \\ |v| & E-E^c \ll v \end{cases}$$

Thus when changing δ, we see peaks in the splitting of the quasidoublets of height $|v|$ at the intersection of $E^c(\delta)$ with $E_+(\delta)$.

Let us sum up. Unlike the integrable systems with mixed dynamics with the phase space that contains both regular and chaotic areas, a new mechanism of tunneling is demonstrated. The splitting of the doublets, which defines the tunneling velocity in the mixed systems in the two-level approximation, is usually several orders greater than is the case of similar but integrable systems. In contrast to to the direct process whereby the particle tunnels directly from one state to another, CAT corresponds to the following three-step process:

1. tunneling from the periodic orbit to the nearest point of the chaotic sea;
2. the classical spread in the chaotic domain of phase space to the vicinity of another periodic orbit;
3. tunneling from a chaotic sea to another periodic orbit.

In other words, the process of splitting of doublets owes its existence to the reflection symmetry and is not direct but happens through the compound process of destruction of the wave function, piece by piece, close to one regular domain, the chaotic transport in the neighboring symmetric regular area, and restoration of the initial state. Schematically, this three-step process is shown in Fig. 10.8.

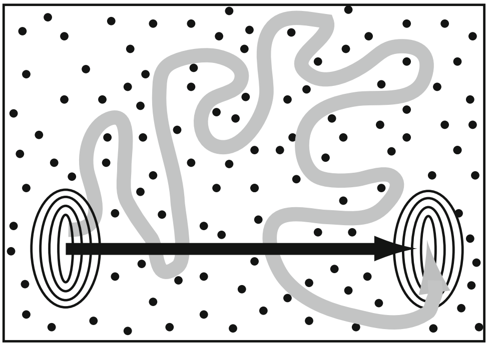


Fig. 10.8 Schematic representation of direct tunneling (*black trajectory*) and the CAT process (*gray trajectory*)

Note that the CAT process is formally of higher order of perturbation theory than direct tunneling. However, the corresponding matrix elements for the CAT are much greater than for direct tunneling. Intuitively, this can be understood as follows: The tunneling from the periodic orbits into a chaotic sea is typically accompanied by significantly smaller violations of classical mechanics than in the case of direct tunneling and therefore has exponentially greater amplitude. More precisely, this means that since a large part of the distance (in a chaotic sea) is a classically allowed transition, we can expect that these indirect trajectories will make a greater contribution to the tunnel flow than direct ones. In the case of direct tunneling, the entire subbarrier trajectory is a forbidden process. The first experimental confirmation of CAT in a microwave version of circular billiards was obtained in [14]. Later, experiments were performed to detect the dynamic tunneling of cold (10 μK) cesium atoms in an amplitude-modulated light wave [15]. The arguments in favor of the CAT being responsible for the effect of the transition from superdeformed states of nuclei to states with a normal deformation were given in [16].

Let us briefly consider the experiments using the microwave analogue of annular billiards [14]. This experimental technique makes it possible to work with the spectra in which the size of the splitting of some quasidoublets is several orders of magnitude smaller than the average distance between the levels. Figure 10.9 shows fragments of the spectrum of the circular microwave billiards with superconducting walls in a neighborhood of $f = 9\,\text{GHz}$ for various values of eccentricity δ [14].

Among a number of the singlet levels, the quasidoublet is clearly visible, and its position in the spectrum changes only slightly with the increase of δ. It becomes possible to measure the quasidoublet splitting of the family of 30 wave functions $(n, m = 1)$. In order to detect the effect of CAT, the quasidoublet splitting should be analyzed as a function of the position of the quasidoublet in the phase space. This can be done using the following considerations. The wave number of states forming a quasidoublet is very close to the concentric case $\delta = 0$. As we pointed out, this happens because the trajectory with a big impact parameter does not collide with the internal circle, so its movement (change δ) weakly perturbs the corresponding

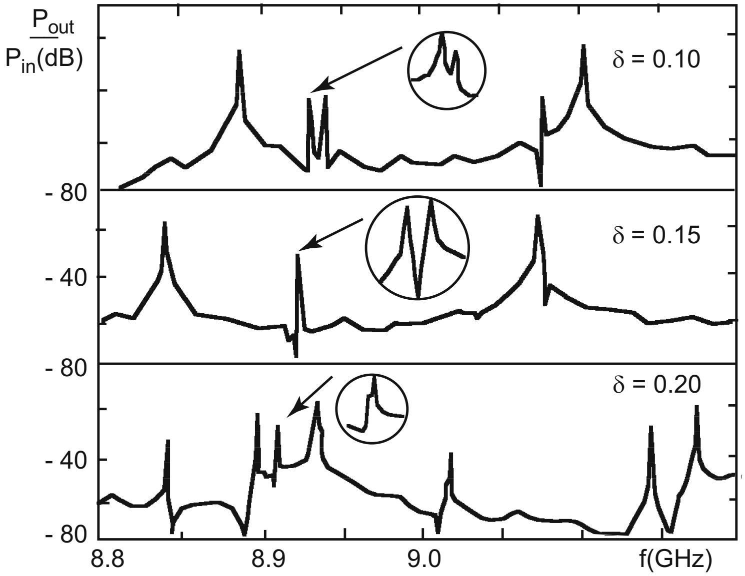


Fig. 10.9 The frequency spectrum of microwave circular billiards with superconductive walls in the vicinity of the frequency $f = 9$ GHz for different parameter values δ. The fragments of the spectrum stretched 50 times by the frequency are shown in *small circles* [14]

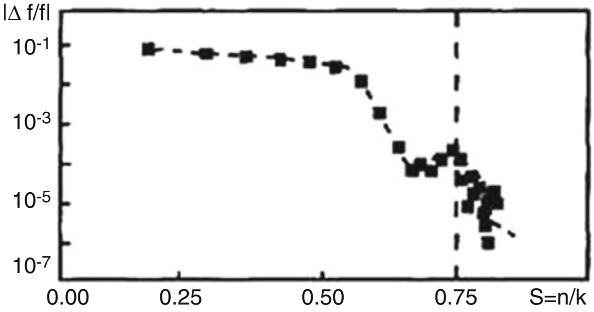


Fig. 10.10 The dependence of the relative doublet splitting $|\Delta f/f|$ on the position of the doublet in the phase space for the value $\delta = 0.20$ [14]

wave functions. This makes it possible to express in terms $S = n/k$ the location of the whispering gallery tori in the phase space. The result of the measurement of the relative doublet splitting $|\Delta f/f|$ for the value $\delta = 0.20$ is given in Fig. 10.10.

In addition to the expected smooth transition from states with big splitting inside the chaotic sea ($S < 0.75$) to states with low splitting in a regular area, the local maximum can be clearly seen in the vicinity of the chaotic sea coast ($S = 0.75$), which can be considered a direct experimental confirmation of the CAT effect.

10.5 Chaotic Nuclear Dynamics

Recent progress in the understanding of chaotic aspects in nonlinear dynamical systems has attracted interest to nuclear dynamics. Indeed, experimental data and modern theory provide a very useful realistic model that allows one to study classical dynamical chaos and the quantum manifestations of classical stochasticity (QMCS).

The chaos concept, despite strong resistance, was introduced in nuclear theory in the 1980s [17–27]. This concept has allowed a fresh look at the peculiarities of nuclear structure [17, 20–23, 28, 29] and nuclear reactions [18, 23, 27], and has helped also to solve a number of old contradictions in nuclear theory [22, 30]. Substantial progress was achieved in the description of concrete nuclear effects [19, 20, 31–33]. Finally, the direct observation of chaotic modes in the simulation of reactions with heavy ions [18] has proven the correctness of the new approach.

Baranger [24] clearly formulated two different approaches to the study of the chaotic aspects of nuclear dynamics.

> Philosophy I. Nuclei are complicated, and chaos comes out of this complication. We expect to find chaos almost everywhere in nuclear physics. Interesting information is in few places without chaos.
>
> Philosophy II. Chaos is a property of simple systems. The interesting new information may be found in simple areas of nuclear physics that are chaotic.

Currently, Philosophy II finds more and more supporters. Let us illustrate this approach by a few examples.

We address first one of the oldest paradoxes in nuclear theory: the possibility to apply two absolutely opposite physical models—the liquid drop model with strong internucleon interaction and the shell model of no interacting particles—to describe the same object (an atomic nucleus). In order to resolve this paradox in the frame of deterministic chaos (Philosophy II), we use the results of [24, 30]:

1. If the nucleons' motion in the nucleus is regular (integrable dynamics), we could expect the manifestations of strong shell effects, which are well described by the model of independent particles in the potential well.
2. If the nucleons' dynamics are dominated by the chaotic component, then one can expect that the liquid drop model or the Thomas–Fermi approximation would be more adequate.

This hypothesis is based on understanding the mechanism according to which the shell effects are destroyed in the process of the regularity–chaos transition. In the early 1900s, Henri Poincaré pointed out that the main problem of dynamics was to study the perturbations of conditionally periodic motion in a system defined by the Hamiltonian

$$H = H_0(I) + \varepsilon V(I, \theta), \tag{10.14}$$

where $H_0(I)$ is the Hamiltonian of the integrable problem, which depends solely on the action variables I, and $\varepsilon V(I,\theta)$ is the nonintegrable perturbation. The solution of this problem is essentially based on the so-called KAM theorem [34]. This theorem guarantees that the classical integrable system can preserve regular behavior, even at sufficiently strong nonintegrable perturbation. We consider the problem of the shell structure destruction in the quantum spectrum, where the theorem plays the principal role. The residual nucleon–nucleon interaction is a nonintegrable addition to the self-consistent field, obtained for instance in the Hartree–Fock approximation [35]. Therefore we can try to relate the destruction of the shells to the deviation from integrability, or conversely, to relate the growth of the shell effects to the approach of the system to the integrable situation. The possibility for the shell structure to exist at sufficiently strong residual interaction is due to the rigidity of the KAM tori. This mechanism of shell structure evolution seems quite natural, taking into account that the semiclassical quantization procedure [36] is based on similar assumptions.

Let us confirm those general considerations by numerical analysis of the shell structure evolution in the potential of quadrupole oscillations U_{QO} (9.18). As we have seen in Chap. 9, the topology of the potential family $U_{\mathrm{QO}}(x,y;a,b,c)$ is determined by only one parameter $W = b^2/ac$. The region $0 < W < 16$ includes the potentials with a unique extremum—a minimum at the origin, corresponding to a spherically symmetric ground state of the nucleus. We shall restrict our consideration to that area. Recall that for $0 < W < 4$, the motion is regular at every energy, and for $4 < W < 16$, we deal with the R-C-R transition.

Based on these assumptions, let us consider the Hamiltonian of the two-dimensional harmonic oscillator in the role of $H_0 = H(b=0, c=0)$. The degenerate equidistant spectrum of this Hamiltonian and the eigenstates $|N,L\rangle$ are well known. Evidently, the eigenstates of the exact Hamiltonian $(c, b \neq 0)$ are not already the eigenstates of the operators $\hat{N}$ and $\hat{L}$. Nevertheless, numerical calculations show that one can use the quantum numbers N and L to classify the quantum states, even at sufficiently strong nonlinearity.

The limiting value of the nonlinearity, up to which such a classification is still reasonable (i.e., the shell structure still exists), is related to the quasicrossing of the neighboring levels. By the quasicrossing we understand the approach of the neighboring levels up to a distance of order the numerical accuracy. As one can see from Fig. 10.11a, in the potential with $W = 13$ approaching the critical energy of the transition to chaos determined by the negative curvature criterion (see Chap. 9), the shell structure is destroyed, observed as the emergence of multiple quasicrossings. At the same time, for the potential with $W = 3.9$ (Fig. 10.11b), there is no local instability generated by the negative curvature of the potential energy surface: the motion remains regular at every energy, so the quasicrossings are absent, even for nonlinearity stronger than that in Fig. 10.11a.

The evolution of the shell structure during the R-C-R transition can be followed on the level of wave functions, using the information contained in the coefficients $C_{NLj}^{(k)}$ of decomposition of the wave functions in terms of the harmonic oscillator basis (9.67). For that purpose, similar to the usual thermodynamic entropy, we

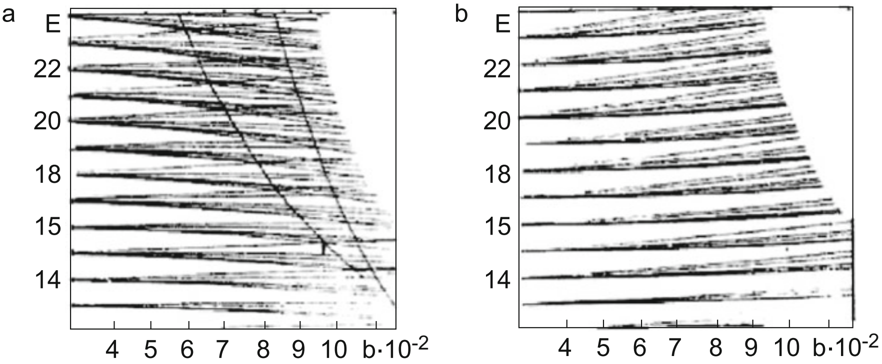


Fig. 10.11 (**a**) Energy spectrum of the quadrupole oscillations Hamiltonian as a function of the parameter b for $W = 13$. The *points* show the quasicrossings. The *solid line* shows the dependence of the critical energy of transition to chaos on the parameter b ($c = 10^{-4}$). The *dashed line* marks the beginning of the quasicrossing region. (**b**) The same for $W = 3.9$. The quasicrossings are absent

introduce [37, 38] the entropy of an arbitrary state k of the exact Hamiltonian H,

$$S(k) = -\sum_{NLj} \left| C_{NLj}^{(k)} \right|^2 \ln \left| C_{NLj}^{(k)} \right|^2 . \tag{10.15}$$

In the regions R_1 and R_2, corresponding to regular motion, the entropy changes correlate with the shell-to-shell transition (see Fig. 10.12). In the chaotic region C, two effects are observed. Firstly, the violation of the quasiperiodic dependence of the entropy on energy reveals the destruction of the shell structure. Secondly, the entropy grows on average monotonically and then reaches a plateau corresponding to the entropy of the random sequence, at energies essentially exceeding the critical energy of the transition to chaos.

The dependence of the shell structure on the type of classical motion is manifest also in such a characteristic of the nucleus as mass. Strutinsky [39] proposed a method to take into account the influence of the shell structure on the nuclear mass. The method is based on the following representation of the total energy of the nucleus as the sum of two terms:

$$U(Z,N,x) = \bar{U}(Z,N,x) + \tilde{U}(Z,N,x), \tag{10.16}$$

where Z and N are the proton and neutron numbers respectively, and x represents a set of the parameters that define the shape of the nucleus. The first term $\bar{U}$ describes the bulk properties of the nucleus, and contains all the contributions that vary smoothly with proton and neutron numbers. The second term $\tilde{U}$ describes the shell effects, related to the shape-dependent microscopic fluctuations. As we have seen

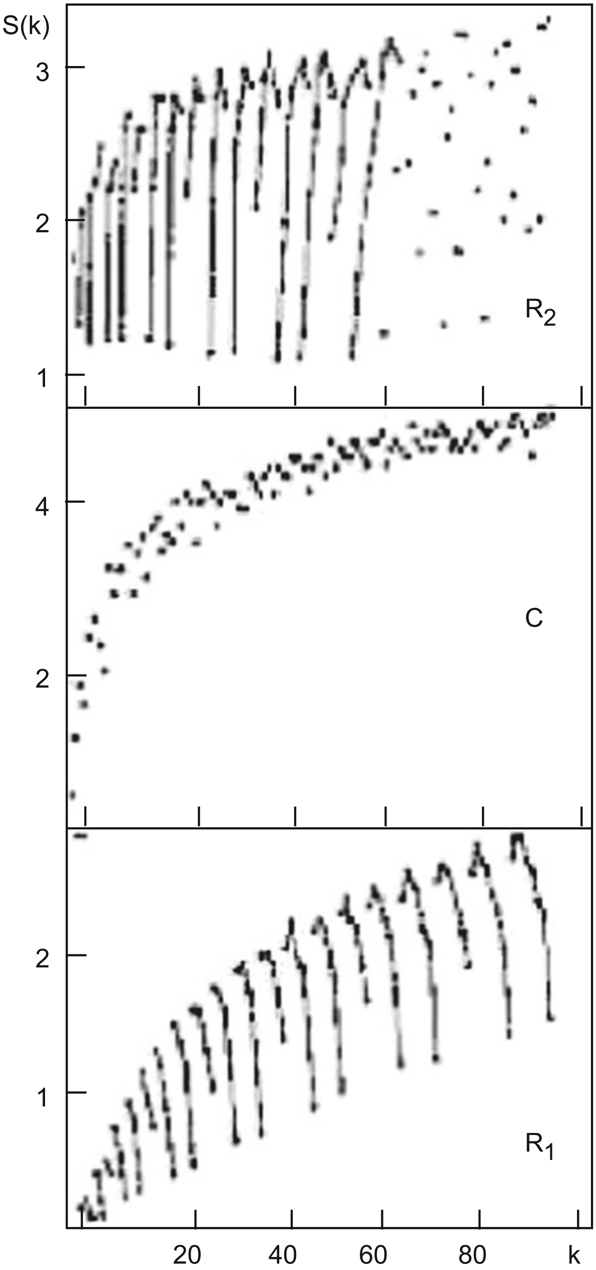


Fig. 10.12 Entropy S_k as a function of state number k for the $U_{\mathrm{QO}}(x, y, W = 13)$ potential. The *solid lines* connecting the points correspond to the shell classification according to N

above, the evolution of the shell structure essentially depends on the type of classical motion. Bohigas and Leboeuf [40] proposed to split the shell correction $\tilde{U}$ into two parts, $\tilde{U} = \tilde{U}_{\mathrm{reg}} + \tilde{U}_{\mathrm{ch}}$. The two contributions $\tilde{U}_{\mathrm{reg}}$ and $\tilde{U}_{\mathrm{ch}}$ correspond respectively to the regular and chaotic components of the nucleon motion. According to [40], the presently calculated masses can correctly reproduce only $\tilde{U}_{\mathrm{reg}}$; however, the final result for the fluctuations produced by the chaotic part of the motion is in fact of a much more general validity, and may be interpreted as arising from the residual interactions. This approach, as was mentioned above, is equivalent to interpretation of the residual nucleon–nucleon interaction as a nonintegrable addition to the self-consistent field.

From a semiclassical point of view, the Gutzwiller trace formula (see Sect. 9.2) allows one to interpret the shell effects as modulations in the single-particle spectrum produced by the periodic orbits of the corresponding classical dynamics. The trace formula (9.13) allows us to calculate the oscillatory part $\tilde{\rho}(E,x)$ of the total level density $\rho(E,x) = \bar{\rho}(E,x) + \tilde{\rho}(E,x)$. The shell correction to the nuclear mass is computed by inserting the oscillatory part of the density of states into the expression for the energy:

$$\tilde{U}\,(x,A,T) = \int dEE\tilde{\rho}(E,x)f\,(E,T)\,, \tag{10.17}$$

where $f\,(E,T)$ is the Fermi function. In a semiclassical expansion, the leading-order term in the integral comes from the energy dependence of the action. Having set $T = 0$, one obtains

$$\tilde{U}(x,A) = 2\hbar^2 \sum_p \sum_{k=1}^{\infty} \frac{A_{p,k}}{k^2 T_p^2} \cos\left[k\left(I_p - \frac{\pi}{2}\mu_p\right)\right]. \tag{10.18}$$

According to (10.18), the presence of the fluctuations in the total energy is therefore a very general phenomenon that occurs for an arbitrary Hamiltonian, irrespective of the nature of the corresponding classical dynamics. However, their importance (i.e., their amplitude) strongly depends on the properties of the dynamics, particularly on its chaotic or regular type [41, 42].

Let us give one more example to show how the atomic nucleus can be a useful tool for QMCS investigations. One of the alternative approaches [43] to describing quantum chaos is based on the traditional method of time series analysis. Let us consider the energy spectrum as a discrete signal, and a sequence of N energy levels as a time series. We shall characterize the spectra fluctuations by the statistical characteristic δ_n defined as

$$\delta_n = \sum_{i=1}^{n} (s_i - \langle s \rangle) = \sum_{i=1}^{n} w_i, \tag{10.19}$$

where the index n runs from 1 to $N-1$. The quantities w_i characterize the fluctuation of the ith spacing between the nearest neighbors with respect to its average value $\langle s \rangle = 1$. The energy spectrum is assumed to be already unfolded in the standard way, i.e., the initial spectrum was mapped to an auxiliary one ($E_i \to \varepsilon_i$) with unit average value of the level density. Then $s_i = \varepsilon_{i+1} - \varepsilon_i$, $i = 1 \ldots N-1$.

Let us consider how the spectral characteristic δ_n changes during the transition from regular systems to chaotic ones. For that purpose, we calculate the power spectrum $S(k)$ for the discrete series defined as

$$S(k) = \left|\hat{\delta}_k\right|^2, \tag{10.20}$$

where δ_k is the Fourier transform of δ_n,

$$\hat{\delta}_k = \frac{1}{\sqrt{N}} \sum_n \delta_n \exp\left(-\frac{2\pi i k n}{N}\right), \tag{10.21}$$

and N is the time series length.

As an object for detailed investigation of the spectrum, Relano et al. [43] chose the atomic nucleus at high excitation energy, where the density level is very high. They obtained the energy spectrum by diagonalization of the nuclear Hamiltonian with a realistic interaction that reproduces the experimental data quite well. Diagonalization was performed for the different values of angular momentum J, parity π, and isospin T. Among the obtained spectra, they selected a series of 256 consecutive levels with equal J^{π}, T from the regions of high-level density. In order to reduce statistical errors and to clarify the general trend, the obtained power spectrum $S(k)$ was averaged over 25 level series.

Figure 10.13 represents the results of the averaged power spectrum for the stable Mg^{24} nucleus (the spectrum was obtained by diagonalization of a 2000 × 2000 matrix) and unstable Na^{34} nucleus (5000 × 5000 matrix). One can see clearly that the power density of the quantity δ_n is close to a power law, presumably following a simple functional form:

$$\langle S(k) \rangle \sim \frac{1}{k^{\alpha}}. \tag{10.22}$$

Least-squares fitting gives the results $\alpha = 1.11 \pm 0.03$ for Na^{34} and $\alpha = 1.06 \pm 0.05$ for Mg^{24}. A natural question arises: can there be some universal relations between the peculiarities of the quantum spectrum corresponding to the type of classical motion and the power spectrum of the δ_n fluctuations?

Perhaps the simplest way to answer that question is to compare the power density for the spectrum corresponding to the Poisson statistics to that of the random matrices ensemble, generating the Wigner distribution. The answer obtained in [43] was quite amazing. Ignoring the very high frequency region, where some deviations

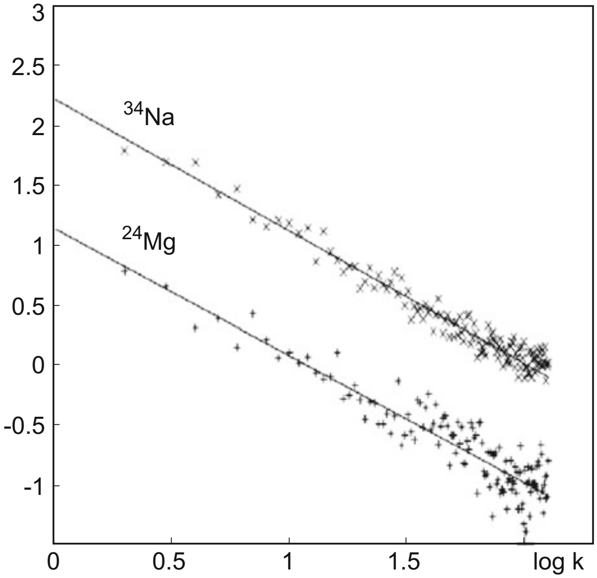


Fig. 10.13 Average power spectrum of the δ_n-function for Mg^{24} from the high-level density region. The plots are displaced to avoid overlapping [43]

were observed probably due to the finiteness of the considered matrices, it was found that $\alpha \simeq 1.99$ for the Poisson spectrum with uncertainty of order of 2 %. In contrast, $\alpha \simeq 1.08$ was found for the Gaussian orthogonal ensemble of the high-dimension matrices, which is usually considered a paradigm of a chaotic quantum spectrum. Therefore, generalizing the obtained results, one can conclude the following [43]: the power spectrum $\langle S(k) \rangle$ behaves as $1/k^{\alpha}$ both for regular and chaotic energy spectra, but the level correlation decreases from the maximum value of $\alpha = 2$ for the regular uncorrelated spectrum to the minimum value of $\alpha = 1$ for chaotic quantum systems.

More generally, this result can be formulated as a hypothesis: the energy spectra are characterized by $1/f$-noise. Recall that the $1/f$-noise (i.e., the flicker-noise) was discovered in the 1920s as a universal satellite of all irreversible stationary processes: its contribution cancels out when the irreversible flows disappear and the system reaches thermodynamic equilibrium.

The proposed hypothesis [43] has a number of attractive features. The considered property characterizes immediately the chaotic spectrum without any reference to properties of other systems (i.e., random matrix ensembles). It is universal for all types of chaotic systems, regardless of whether they are invariant with respect to time reversal, or whether their spin is integer or half-integer. Besides that, the $1/f$ characteristic brings together the quantum chaotic systems with a very wide set of systems from very different branches of science. The flicker-noise is omnipresent.

Therefore, the energy spectra of chaotic systems demonstrate the same type of fluctuations as many others. However, there are no reasons to believe that the $1/f$-noise in the spectral fluctuations of quantum systems implies $1/f$-noise in their classical analogues.

The common conception of the possible stochastization mechanism in quadrupole nuclear oscillations of high amplitude is confirmed by direct observations of chaotic regimes in numerical simulations of reactions with heavy ions [18].

The time-dependent Hartree–Fock theory [35] constitutes a well-defined starting point for the study of such processes. The time-dependent Hartree–Fock equations can be obtained from the variation of the many-body action S,

$$S = \int_{t_1}^{t_2} dt \left\langle \Psi(t) \left| i\frac{\partial}{\partial t} - H(t) \right| \Psi(t) \right\rangle . \tag{10.23}$$

In this expression, H is the many-body Hamiltonian, and the A-nucleon wave function $\Psi(t)$ is chosen to be of determinant form, constructed from time-dependent single-particle states $\psi_\lambda(\mathbf{r}, t)$

$$\Psi(\mathbf{r}_1, \mathbf{r}_2 \ldots \mathbf{r}_A; t) = \frac{1}{\sqrt{A!}} \det |\psi_\lambda(\mathbf{r}, t)| . \tag{10.24}$$

The variation of Eq. (10.23) with respect to the single-particle states ψ_λ and ψ_λ^* yields the equations of motion

$$i\frac{\partial}{\partial t}\psi_\lambda(\mathbf{r}, t) = \frac{\delta \langle H \rangle}{\delta \psi_\lambda^*(\mathbf{r}, t)} \equiv h(\mathbf{r}, t)\, \psi_\lambda(\mathbf{r}, t) , \tag{10.25}$$

and a similar equation for $\psi_\lambda^*(\mathbf{r}, t)$.

The classical nature of these equations can be put into better perspective via the definition of the classical field coordinates $\phi_\lambda(\mathbf{r}, t)$, and conjugate moments $\pi_\lambda(\mathbf{r}, t)$,

$$\begin{aligned} \varphi_\lambda &= \left(\psi_\lambda + \psi_\lambda^*\right)/\sqrt{2}, \\ \pi_\lambda &= \left(\psi_\lambda - \psi_\lambda^*\right)/\sqrt{2}. \end{aligned} \tag{10.26}$$

Then the result is Hamilton's equations:

$$\begin{aligned} \frac{\partial \varphi_\lambda(\mathbf{r}, t)}{\partial t} &= \frac{\delta \langle H \rangle}{\delta \pi_\lambda(\mathbf{r}, t)}, \\ \frac{\partial \pi_\lambda(\mathbf{r}, t)}{\partial t} &= -\frac{\delta \langle H \rangle}{\delta \varphi_\lambda(\mathbf{r}, t)}. \end{aligned} \tag{10.27}$$

The time-dependent Hartree–Fock equation (10.25) and its complex conjugate are solved on a three-dimensional space-time lattice with special initial wave functions [44]:

$$\begin{aligned}\varphi_\lambda(\mathbf{r},t) &\to \sqrt{2}\cos(\mathbf{k}_\lambda\mathbf{r}-\varepsilon_\lambda t)\,\chi_\lambda(\mathbf{r}),\\ \pi_\lambda(\mathbf{r},t) &\to \sqrt{2}\sin(\mathbf{k}_\lambda\mathbf{r}-\varepsilon_\lambda t)\,\chi_\lambda(\mathbf{r}),\end{aligned} \tag{10.28}$$

where ε_λ is the solution of the static Hartree–Fock equations $h\chi_\lambda(\mathbf{r}) = \varepsilon_\lambda\chi_\lambda(\mathbf{r})$, $\lambda = 1\cdots A$, and k_λ is the parameter of the initial boost.

The time-dependent Hartree–Fock calculations for head-on collisions of $\mathrm{He}^4 + \mathrm{C}^{14}$, $\mathrm{C}^{12} + \mathrm{C}^{12}\left(0^+\right)$ and $\mathrm{He}^4 + \mathrm{Ne}^{20}$ were performed by Umar et al. [18] at bombarding energies near the Coulomb barrier. The results are interpreted in terms of their classical behavior. The initial energy and the separation of the centers of the ions are the parameters labeling the initial state. After initial contact, the compound nuclear system (O^{18} or Mg^{24}) relaxes into a configuration undergoing quasiperiodic or chaotic motion. The analysis of nuclear density multipole moments $\left\{M_{LI}(t), \dot{M}_{LI}(t)\right\}$ has been applied for classifying those solutions using Poincaré sections. The definitions of the moments are as follows:

$$\begin{aligned}M_{LI}(t) &= \int d^3r r^L Y_{LM}(\hat{r})\,\rho_I(\mathbf{r},t),\\ M_{LI}(\omega) &= \int dt \exp(-i\omega t)\,M_{LI}(t),\end{aligned} \tag{10.29}$$

where the isoscalar ($I = 0$) and isovector ($I = 1$) densities are

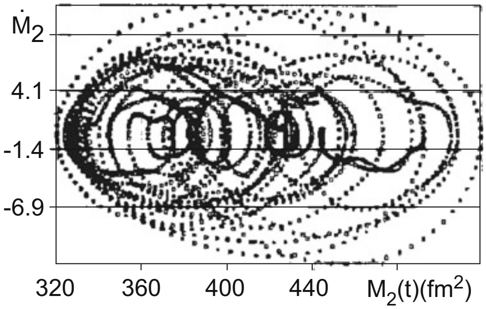
$$\rho_I(\mathbf{r},t) = \begin{cases}\rho_p(\mathbf{r},t) + \rho_n(\mathbf{r},t) & (I=0),\\ \rho_p(\mathbf{r},t) - \rho_n(\mathbf{r},t) & (I=1).\end{cases} \tag{10.30}$$

The proton ρ_p and neutron ρ_n densities in terms of the field coordinates ϕ_λ and moments π_λ are

$$\rho_q(\mathbf{r},t) = \frac{1}{2}\sum_\lambda\left[\left|\pi_{\lambda,q}(\mathbf{r},t)\right|^2 + \left|\varphi_{\lambda,q}(\mathbf{r},t)\right|^2\right], \quad q = p, n. \tag{10.31}$$

The isoscalar quadrupole mode ($L = 2$, $I = 0$), shown in Fig. 10.14 for the Mg^{24} nuclear system, seems to fill most of the available phase space in the Poincaré section $\left\{M_{LI}(t), \dot{M}_{LI}(t)\right\}$.

Fig. 10.14 Poincaré section $\{M_{LI}(t), \dot{M}_{LI}(t)\}$ for isoscalar quadrupole mode for the Mg^{24} system [18]



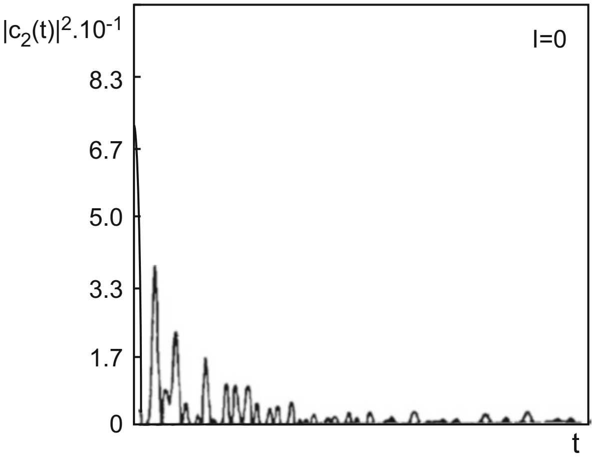


Fig. 10.15 The autocorrelation function $C_{20}(t)$ as a function of time for the isoscalar quadrupole mode in the Mg^{24} system [18]

The corresponding autocorrelation function (Fig. 10.15)

$$C_{20}(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \exp(i\omega t) \, |M_{20}(\omega)|^2 \tag{10.32}$$

damps rapidly. All this favors the view that the corresponding motion is closer to stochastic, rather than quasiperiodic.

In conclusion, let us describe briefly one more nuclear effect, which demonstrates that the nucleus is a laboratory for research of general physical effects such as tunneling, chaos, and phase transitions. Recently, it became clear [45, 46] that superdeformed nuclei offer a new way of understanding nuclear structure. On the one hand, nuclei provide interesting data for the study of these general phenomena; on the other hand, we obtain new information about the nuclear structure.

A superdeformed nucleus is a nucleus that is very far from spherical, forming an ellipsoid with axes in ratios of approximately 2:1:1. Only some nuclei can exist

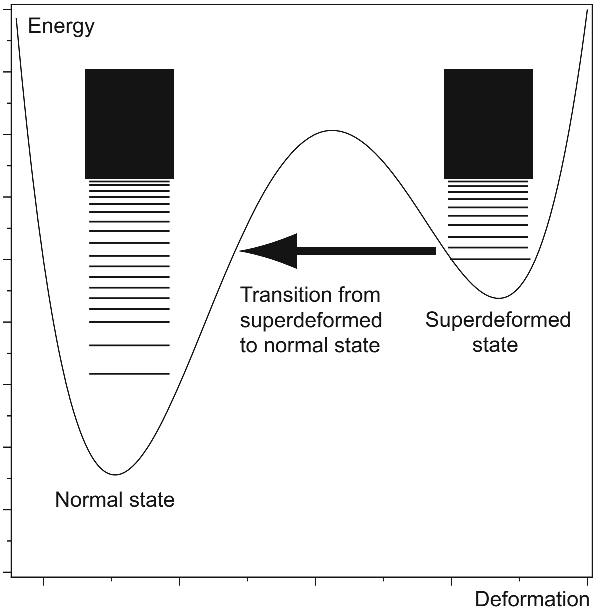


Fig. 10.16 Decay from superdeformed state to normal state

in superdeformed states. These superdeformed nuclei are produced in reactions with heavy ions. Initially accelerated heavy ions, in colliding with nuclei of the target material, produce highly excited and rapidly rotating compound nuclei. These nuclei release part of the excitation energy by the emission of light particles (neutrons, protons, alpha particles) and photons. In all observations of rapidly rotating superdeformed nuclei, rotation breaks off (already at low momentum) when the nucleus suddenly changes its shape and decays to the state that corresponds to lower deformations. Three stages of the transition from the superdeformed state to the normal deformed one are presented in Fig. 10.16 [47]: feeding of superdeformed bands, ordered rotation, and decay from superdeformed to the normal state.

The lowest states in a superdeformed local minimum correspond to high excitations in the main minimum. As a result, the lowest superdeformed states (cold states) are characterized by good quantum numbers and symmetries, while their decays to the main minimum are controlled by the strict rules of selection. In contrast, the normal deformed states at the same energy could correspond to chaotic motion in the semiclassical limit. Statistical analysis confirms this idea: the energy spectrum in this energy region of the main minimum demonstrates all the signs of chaos [47]. This means that in the considered region of nuclei excitation for which the deformation potential allows the existence of the second minimum, a mixed state occurs (see Sect. 9.3). The question is whether the chaos-assisted tunneling mechanism can be used for the description of the transition from the superdeformed state to the normal deformed one. We should stress that we can consider dynamical tunneling because the transition takes place in multidimensional space: the nuclear shape is characterized by at least two parameters. Aberg [16] estimated in perturbation theory the tunneling probability in two limiting cases. The first limit was the case of no mixing between the normal deformed states (completely regular normal deformed system), while the second was that the tunneling strength is spread out over all normal deformed states (i.e., quantum chaos). For the ratio of the probability of tunneling T^{chaotic} and T^{regular}, Aberg found [16]

$$\frac{T^{\text{chaotic}}}{T^{\text{regular}}} \sim 10^4\text{–}10^6. \tag{10.33}$$

Therefore, we expect the tunneling probability to be enhanced by a factor of 10^4–10^6 if the normal deformed states are chaotic. In other words, the tunneling process connected to the decay out of superdeformed states is strongly enhanced by the chaotic properties of the normal deformed states.

References

1. Razavy, M., Pimpale, A.: Phys. Rep. **168** 305–370 (1988)
2. Landau, L.D., Lifshitz, E.M.: Quantum Mechanics. Pergamon Press, London (1982)
3. Davis, M.J, Heller, E.J.: J. Chem. Phys. **72**, 246–254 (1981)
4. Bohigas, O., Tomsovic, S., Ulmo, D.: Phys. Rep. **223**, 43–133 (1993)
5. Doron, E., Frischat, S.: Phys. Rev. **E57**, 1421–1443 (1998)
6. Lin, W.A., Ballentine, L.E.: Phys. Rev. Lett. **65**, 2927–2930 (1990)
7. Lin, W.A., Ballentine, L.E.: Phys. Rev. **A45**, 3637–3645 (1992)
8. Husimi, K.: Proc. Phys. Math. Soc. Jpn. **22**, 264–314 (1946)
9. Schiff, L.: Quantum Mechanics. McGraw-Hill, New York (1955)
10. Glauber, R.: Phys. Rev. **131**, 2766–2788 (1963)
11. Peres, A.: Phys. Rev. Lett. **67**, 2931–2934 (1991)
12. Zel'dovich, Ya.B.: Sov. Phys. JETP **24**, 1006–1008 (1967)
13. Bohigas, O., Boose, D., de Carvalho, R., Marvulle, V.: Nucl. Phys. A **560**, 197–210 (1993)
14. Dembowski, C., et al.: Phys. Rev. Lett. **84**, 867–870 (2000)
15. Steck, D., Oskay, W., Raizen, M.: Science **293**, 274–278 (2001)
16. Aberg, S.: Phys. Rev. Lett. **82**, 299–302 (1999)
17. Williams, R.D., Koonin, S.E.: Nucl. Phys. A **391**, 72–92 (1982)
18. Umar, A.S., Staer, M.R., Gusson, R.Y., Reinhard, P.-G., Bromley, D.A.: Phys. Rev. C **32**, 172–183 (1985)
19. Bolotin, Yu.L., Krivoshei, I.V.: Yad. Fiz. **42**, 53–56 (1985)
20. Arvieu, R., Brut, F., Carbonell, J.: Phys. Rev. A **35**, 2389–2408 (1987)
21. Bolotin, Yu.L., Gonchar, V.Yu., Inopin, E.V., Levenko, V.V., Tarasov, V.N., Chekanov, N.A.: Fiz. Elem. Chastits At. Yadra **20**, 878–929 (1989)
22. Swiateski, W.: Nucl. Phys. A **488**, 375–394 (1989)
23. Bohigas, O., Weidenmuller, H.: Ann. Rev. Nucl. Part. Sci. **38**, 421–453 (1988)
24. Baranger, M.: Order, Chaos and Atomic Nucleus, preprint (1989)
25. Rotter, I.: Rep. Prog. Phys. **54**, 635–682 (1991)
26. Alhassid, Y., Whelan, N.: Phys. Rev. C **43**, 2637–2647 (1991)
27. Zelevinsky, V.G.: Nucl. Phys. A **570**, 411–421 (1994)

28. Rotter, I.: Fiz. Elem. Chastits At. Yadra **19**, 274–303 (1988)
29. Bolotin, Yu.L., Gonchar, V.Yu., Tarasov, V.N., Chekanov, N.A.: Yad. Fiz. **52**, 669–678 (1990)
30. Bjornholm, S.: Nucl. Phys. A **447**, 117–144 (1985)
31. Blumel, R., Smilansky, U.: Phys. Rev. Lett. **60**, 477–480 (1988)
32. Manfredi, V.R., Salasnich, L.: Int. J. Mod. Phys. E **4**, 625–636 (1995)
33. Manfredi, V.R., Rosa-Clot, M., Salasnich, L., Taddei, S.: Int. J. Mod. Phys. E **5**, 521–530 (1996)
34. Arnol'd, V.I.: Mathematical Methods of Classical Mechanics. Springer, New York (1989)
35. Barts, B.I., Bolotin, Yu.L., Inopin, E.V., Gonchar, V.Yu.: The Hartree-Fock Method in Nuclear Theory. Naukova Dumka, Kiev (1982)
36. Keller, J., Rubinov, S.: Ann. Phys. **9**, 24–75 (1960)
37. Yonezava, F.: J. Non-Cryst. Solids **35**, 29–40 (1980)
38. Reichl, L.E.: Europhys. Lett. **6**, 669–675 (1988)
39. Strutinsky, V.M.: Nucl. Phys. A **95**, 420–442 (1967)
40. Bohigas, O., Leboeuf, P.: Phys. Rev. Lett. **88**, 092502 (2002)
41. Leboeuf, P., Monastra, A.G.: Ann. Phys. **297**, 127–156 (2002)
42. Brack, M., Bhaduri, R.K.: Semiclassical Physics. Addison-Wesley, Massachusetts (1997)
43. Relaño, A., Gómez, J., Molina, R.A., Retamosa, J., Faleiro, E.: Phys. Rev. Lett. **89**, 244102 (2002)
44. Davies, K.T.R., Koonin, S.E.: Phys. Rev. C **23**, 2042–2061 (1981)
45. Aberg, S., Flocard, H., Nazarewicz, W.: Annu. Rev. Nucl. Part. Sci. **40**, 439–528 (1990)
46. Baktash, C., Haas, B., Nazarewicz, W.: Annu. Rev. Nucl. Part. Sci. **45**, 485–541 (1995)
47. Aberg, S.: Z. Phys. A **358**, 269–272 (1997)

Index

Y. Bolotin et al., *Chaos: Concepts, Control and Constructive Use*,
Understanding Complex Systems, DOI 10.1007/978-3-319-42496-5