A UNIVERSAL INSTABILITY OF MANY-DIMENSIONAL OSCILLATOR SYSTEMS

Boris V. CHIRIKOV

Institute of Nuclear Physics, 630090 Novosibirsk, U.S.S.R.

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM

PHYSICS REPORTS (Review Section of Physics Letters) 52, No. 5 (1979) 263–379. NORTH-HOLLAND PUBLISHING COMPANY

A UNIVERSAL INSTABILITY OF MANY-DIMENSIONAL OSCILLATOR SYSTEMS

Boris V. CHIRIKOV

Institute of Nuclear Physics, 630090 Novosibirsk, U.S.S.R.

Received January 1979

Contents:

1. Introduction 265
2. What are nonlinear oscillations? A few simple examples 268
   2.1. Pendulum 268
   2.2. Canonical transformations and the superconvergence 270
   2.3. The cubic force 273
   2.4. The separatrix 274
3. Nonlinear resonance 275
   3.1. Parametric resonance 276
   3.2. A universal description of a nonlinear resonance 278
   3.3. A resonance of many-dimensional oscillations 280
4. Interaction of resonances and the stochastic instability 285
   4.1. Two resonances; the overlap criterion 286
   4.2. Many resonances; mappings 290
   4.3. Motion of a charged particle in a magnetic bottle 292
   4.4. Motion in a vicinity of the separatrix 300
   4.5. Many-dimensional oscillations 304
   4.6. The Kolmogorov–Arnold–Moser theory (KAM theory) 306
5. Stochastic oscillations of the pendulum 310
   5.1. The border of stability 311
   5.2. The Krylov–Kolmogorov–Sinai entropy (KS-entropy) 318
   5.3. Local instability and stochasticity 323
   5.4. Diffusion 326
   5.5. Islets of stability 330
6. Stochastic layer 336
   6.1. The whisker mapping 337
   6.2. Width of stochastic layer 339
   6.3. The KS-entropy in the stochastic layer 343
   6.4. Again about the border of stability 345
7. The Arnold diffusion 346
   7.1. Arnold's example 347
   7.2. Evaluation of the diffusion rate 348
   7.3. A more general case 352
   7.4. Nekhoroshev's theory 357
   7.5. Another model 360
   7.6. Numerical experiments 363
   7.7. How does the Arnold diffusion "work"? 369
8. Concluding remarks, problems 371
Acknowledgements 372
Appendix: the Melnikov–Arnold (MA) integral 373
References 376

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 52 No. 5 (1979) 263–379.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 45.00, postage included.

Abstract:

The purpose of this review article is to demonstrate via a few simple models the mechanism for a very general, universal instability – the Arnold diffusion – which occurs in the oscillating systems having more than two degrees of freedom. A peculiar feature of this instability results in an irregular, or stochastic, motion of the system as if the latter were influenced by a random perturbation even though, in fact, the motion is governed by purely dynamical equations. The instability takes place generally for very special initial conditions (inside the so-called stochastic layers) which are, however, everywhere dense in the phase space of the system.

The basic and simplest one of the models considered is that of a pendulum under an external periodic perturbation. This model represents the behavior of nonlinear oscillations near a resonance, including the phenomenon of the stochastic instability within the stochastic layer of resonance. All models are treated both analytically and numerically. Some general regulations concerning the stochastic instability are presented, including a general, semi-quantitative method – the overlap criterion – to estimate the conditions for this stochastic instability as well as its main characteristics.

1. Introduction

Enormous progress has been made during the last few decades in the general theory of dynamical systems, including, as a particular case, the theory of nonlinear oscillations for Hamiltonian systems of classical mechanics (see, e.g., refs. [1–3, 11–15, 96, 110]). Unfortunately for a physicist, most of the theories developed are purely mathematical. This hampers their application to particular problems for, at least, two reasons. First, mathematical models are frequently not adequate for real physical systems, and it is quite unclear how far the conclusions of a mathematical theory can be extended beyond the rigid frame of its stringent premises. Secondly, often it is just hard to understand the language of contemporary mathematics unless you have pierced its "jungles".

Without any doubt, the mathematical theories developed such as the KAM-theory of stability (section 4.6), the theories of C- and K-systems and the whole modern ergodic theory, the differentiable dynamics and others (section 5.3) do serve already now and will do so in the future as firm beacons in a still dense mist of extremely diverse phenomena and processes in nonlinear mechanics, amidst numerous reefs of specific cases and surprise complications. Yet, he who desires to reach a cherished islet of stability in the violent (and stochastic!) sea of nonlinear oscillations should not rely upon the beacons only. One must learn how to find one's way using a more varied arsenal of means and methods besides a stringent and scrupulous mathematical analysis even if the former would provide only a temporary and preliminary solution of a problem. Academician A.N. Kolmogorov has mentioned on an occasion that it is not so much important to be rigorous as to be right. A way to be convinced (and to convince the others!) of the rightness of a solution without a rigorous theory is a tried method of the science – the experiment. In the field of non-linear oscillations the so-called numerical experiment, or numerical simulation, i.e. numerical integration of the motion equations of a dynamical system by computer, has been spreading recently more and more. In the present paper we widely use results of various numerical experiments. With a few exceptions only *numerical* experiments will be mentioned, so we shall say often just "experiment", "empirical", etc. without the risk of confusing the reader.

The main purpose of the present review article is a detailed and, as far as possible, graphic consideration, via simple models, of a basic phenomenon discovered recently – the so-called stochastic instability of nonlinear oscillations (see, e.g., the review articles [9, 32, 76]) and, especially, its most ingenious manifestation – the Arnold diffusion [5, 24, 43]. The latter may be called by right a universal instability of many-dimensional nonlinear oscillations since it occurs in any system (except the completely integrable ones, see below and section 8) and for arbitrarily small perturbation.

The selection of the models is determined by two reasons. First, we have done our best to find as simple as possible, yet non-trivial, models, that is such whose motion is still not completely solved. Sections 2 and 3 which are devoted, mainly, to questions of methodics are an exception. Those questions are thought to be necessary to understand the main content of the paper. Secondly, the models should bear an obvious relation to real physical systems the study of which is of some interest today. This sort of relation is illustrated in section 4.3 where the problem of a charged particle motion in a magnetic bottle is considered. This problem is of importance in various applications but still remains incompletely solved in spite of more than twenty years' efforts. The many-dimensional oscillator described in sections 7.5 and 7.6 may be also considered as a model for the motion of a charged particle in a storage ring. The latter problem is of interest for the colliding beam technique.

The class of problems under consideration is, of course, a small part not only of general theory of dynamical systems but even of the classical nonlinear oscillation theory. In particular, we confine ourselves to Hamiltonian systems only excluding, thus, all questions related to the influence of dissipation, hence, also any relaxation oscillations and the like*. Nevertheless, the remaining set (of oscillatory processes) is not empty, as mathematicians use to say. The role of Hamiltonian, or nearly Hamiltonian, systems lies beyond some of their particular, albeit important, applications such as, for example, the motion of charged (especially heavy) particles in accelerators and storage rings, or the motion of planets in the Solar system as well as that of man-made celestial bodies: sputniks, space ships, etc. What is still more important, perhaps, is that the fundamental laws of motion for "elementary" particles may be expressed in terms of Hamiltonian equations whereas phenomena like dissipation turn out to be specific consequences of the former.

Another essential restriction of the models in question is a limitation of the oscillation nonlinearity from below, that is we assume the nonlinearity to be not too small. The case of a small nonlinearity turns out to be more complicated, strange though it may seem, if, perhaps, a more interesting one, as is demonstrated by a curious example in ref. [40]. As far as a theoretical analysis and estimates are concerned we have also to confine ourselves to the usual case of a small perturbation acting upon a system whose motion is known.

In this article we are not going to present or to represent any general theory of nonlinear oscillations. On the contrary, important (and rather little-known) phenomena and ideas will be considered via simple examples, graphically, if you like, in accordance with a wise dictum that one example may turn out to be sometimes more useful than a dozen of theories. The author, however, allows himself on occasion some natural generalizations at the level of rigor common in physics (see sections 3.3; 4.5; 7.3).

The basic model we are going to consider is that of a simple pendulum under an external periodic perturbation (section 5). Free oscillations of the pendulum simulate the fairly well known phenomenon of a single (isolated) resonance (section 3). The latter may be considered as an elementary phenomenon of nonlinear dynamics, a "cell", a large number of which forms all the diversity of nonlinear oscillatory processes. A significant feature of the nonlinear resonance is the oscillation boundedness and smallness under a small perturbation as distinct from the linear resonance for which there is no such boundedness. The oscillations are bounded due to the dependence of their frequency on the energy. Such a dependence is, thus, an important property and the first sign of an oscillator nonlinearity (section 2). Boundedness of oscillations is a kind of stabilization of a resonant perturbation by the nonlinearity. Various applications of the nonlinear stabilization to provide, for example,

*Concerning the influence of a weak dissipation see the end of section 5.5.

the motion stability of charged particles in accelerators, is a topical task today. The phenomenon of nonlinear stabilization as well as nonlinear resonance are rather well known and have attracted the attention of many researchers.

Less known is that in the way of those applications the stochastic instability, which destroys nonlinear stabilization, arises.* Stochastic instability results from an interaction of several nonlinear resonances (section 4). This interaction is considered via the model of a pendulum acted upon by a special perturbation – a sequence of short periodical "kicks" (section 5). The motion of the latter model may be described by a simple – standard as we call it – mapping (5.1), i.e. by means of difference equations, that simplifies considerably both the theoretical analysis of motion and numerical experiments.

A simple criterion related to the overlap of neighbouring resonances forms the core of our theoretical analysis (section 4.1). We shall repeatedly come back to this criterion to compare results of numerical experiments with analytical estimates (see section 6.4). The overlap criterion is substantiated below by plausible (at least, for a physicist!) considerations and estimates as well as by the results of various numerical experiments. This sort of approach allows us to resolve problems unreachable for a more stringent mathematical analysis (see section 6.2). A retribution for such a progress is inability of the simple criterion to discern the so-called completely integrable systems – the specific and, in a sense, exceptional cases of absolutely stable nonlinear oscillations. An excellent example of the latter is the so-called Toda lattice [108]. The only excuse the overlap criterion may offer is that today any other theory is also uncapable to answer the question whether a given system is completely integrable even though a large experience in "designing" such systems has been gained by now [68].

Numerical experiments reveal that under the overlap of resonances the oscillations of a system become irregular, or stochastic, as if the latter were influenced by a random perturbation even though, in fact, the motion is governed by purely dynamical equations. Thus, we have come across an example of arising a "random" process in a dynamical system, and what's more, in a very simple one having 1.5 degree of freedom altogether (3-dimensional phase space). Along with other examples [11, 38, 34] the latter result leads to a new understanding of the nature of statistical laws in classical mechanics.

In section 6 the oscillations of a pendulum (the latter being just as well a nonlinear resonance) near the separatrix are considered. Motion in this area proves to be described also by a standard mapping the analysis of which leads to the conclusion that in a sufficiently close vicinity of separatrix the so-called stochastic layer always exists. The motion inside the layer is unstable and in the nature of a diffusion which is confined within the layer. The stochastic layer plays an important role in modern oscillation theory, being an "embryo" of an instability, this is just the place from which the instability is spreading, as the perturbation grows, over all, or almost all, of the phase space of a system. The standard mapping, to properties of which a considerable part of the present paper is devoted, plays, thus, a role much more important than just an example of the resonance interaction.

The significance of the stochastic layer comes to light completely in the many-dimensional system some models of which are considered in section 7. In the case of many-dimensional oscillations the set of stochastic layers at nonlinear resonances forms a united network, a "web", motion trajectories inside the latter penetrating nearly all the phase space. A universal instability does set in, the

*This instability may be useful, however, in realization of the so-called Fermi stochastic acceleration [132] (see also [133]) as discussed in ref. [29] and as was actually done for plasma heating [129]. A similar method for the electron heating by an r.f. wave in a mirror machine, also relied upon the stochastic instability, was investigated in a number of works (see, e.g., refs. [134, 136, 33]).

instability which has been discovered by Arnold [5, 67] and which has been called later the Arnold diffusion [43].

In the present paper we do not go beyond the framework of classical mechanics. The study of nonlinear oscillations in quantum mechanics is still in its very outset. For some interesting peculiarities of a quantum nonlinear oscillator already considered see, e.g., refs. [20, 144].

My last remark concerns the references which are not claimed to be complete but pursue, mainly, the aim to help the reader to make familiar himself with original papers on the problems in question. The author offers in advance his apologies for possible omissions and accidental mistakes as well as for somewhat "stochastic" numbering of the reference.

2. What are nonlinear oscillations? A few simple examples

In this section we start with a few simple examples of one degree-of-freedom nonlinear oscillations. We will expose the most important property of the nonlinear oscillator, the so-called *non-isochronicity*, i.e. the dependence of the free oscillation period on the amplitude, or the energy. In section 2.2 the main ideas and a practical application of a modern perturbation theory, based on the canonical change of dynamical variables, will be presented. This technique will be used below. In section 2.4 we shall consider a special trajectory of a nonlinear oscillator, the so-called *separatrix* which will play an important role in what follows.

2.1. Pendulum

Oscillations of a pendulum give us one of the oldest examples of oscillations, in general, and, in particular, nonlinear oscillations. The centuries-old history of studying this apparently simple system has shown a surprising diversity of its motions. For instance, a new phenomenon of the dynamical stability was discovered not so long ago [16]. In the present paper we shall see, in particular, that the study of pendulum properties is still quite far from being completed (see section 5). The simple pendulum model will play a special role for us. We shall see that one of the basic phenomenon of nonlinear oscillations – the nonlinear resonance – can be described by this model under fairly general conditions (section 3).

Let us first consider free oscillations of a pendulum. Let the Hamiltonian of the system have the form:

$$H(p, \varphi) = \frac{p^2}{2M} - U_0 \cos \varphi \tag{2.1}$$

where φ is the angle of pendulum displacement from the lower (stable) position of equilibrium; $p = M\dot{\varphi}$ stands for the pendulum angular momentum; $M = ml^2$ is the moment of inertia. The frequency of small oscillations (ω_0) is related to the amplitude of the potential energy U_0 by $\omega_0^2 = U_0/M$. Note that for a free rotation p is the action variable. In what follows let $M = 1$.

It is well known that the pendulum equations of motion can be integrated in terms of elliptic functions (see, e.g., ref. [4]). There are two kinds of motion: the oscillations and the rotation.

Solving equation $H(p, \varphi) = \text{const}$ for $p = \dot{\varphi}$ we find:

$$\dot{\varphi} = \pm\sqrt{2(H + U_0 \cos \varphi)}. \tag{2.2}$$

Straightforward integration of this equation for $H < U_0$ (oscillations) gives (see, e.g., ref. [58]):

$$\dot{\varphi} = 2\omega_0 \sin(\varphi_0/2)\,\mathrm{cn}(\omega_0 t) \tag{2.3}$$

where $\mathrm{cn}(u)$ is the Jacobian elliptic cosine; φ_0 stands for the amplitude of pendulum displacement, and the time $t = 0$ corresponds to crossing the point $\varphi = 0$ with a positive velocity. Term by term integrating the Fourier series for $\mathrm{cn}(u)$ gives:

$$\varphi(t) = 4\omega \sum_{n=1}^{\infty} \frac{\sin(\omega_n t)}{\omega_n \cosh(K'\omega_n/\omega_0)}. \tag{2.4}$$

Here $K' = K(k')$ is the complete elliptic integral of the first kind; $k' = \sqrt{1-k^2}$; $k = \sin(\varphi_0/2)$. The period of oscillations is $T = 2\pi/\omega$ where

$$\omega(H) = \pi\omega_0/2K \tag{2.5}$$

and the oscillation spectrum is: $\omega_n = (2n-1)\omega$.

In the case of rotation ($H > U_0$), the solution has the form:

$$\pm\varphi(t) = 2\,\mathrm{am}(\omega_r t) = 2\omega t + 4\omega \sum_{n=1}^{\infty} \frac{\sin(\omega_r t)}{\omega_n \cosh(K'\omega_n/\omega_r)}. \tag{2.6}$$

Here $\mathrm{am}(u)$ is the Jacobian elliptic amplitude;

$$k = \sqrt{2U_0/(H+U_0)} = \omega_0/\omega_r; \qquad \omega_r = \sqrt{\tfrac{1}{2}(H+U_0)},$$

and we have introduced half the frequency of the mean rotation: $\omega(H) = \pi\omega_r/2K$ to make the expressions (2.4) and (2.6) as similar to each other as possible; we shall use this in section 2.4. The rotation frequency spectrum is: $\omega_n = 2n\omega$.

Relation (2.5) shows that the pendulum frequency does depend on the oscillation amplitude, or the energy so that the pendulum is generally a non-isochronous oscillator. The latter property is of paramount importance for the problem of motion stability as we shall see below.

Let us consider the case of small pendulum oscillations and relate the argument $k = \sin(\varphi_0/2) \ll 1$ of the elliptic integral to the pendulum energy E reckoned from the minimum of the potential energy: $-U_0$ (2.1). We have: $E = 2U_0k^2$. Expanding the expression $[K(k)]^{-1}$ as a power series in k^2 we find:

$$\omega \approx 1 - \tfrac{1}{8}E - \tfrac{5}{256}E^2 - \tfrac{11}{2048}E^3 \tag{2.7}$$

where we have put $U_0 = 1$ ($\omega_0 = 1$). We shall need such a high accuracy in the next section to compare eq. (2.7) with the result of the perturbation theory.

Let us introduce the dimensionless parameter of the nonlinearity:

$$\alpha = \frac{I}{\omega}\cdot\frac{\mathrm{d}\omega}{\mathrm{d}I}. \tag{2.8}$$

We define α via the action I since we shall use below, as a rule, the action-angle variables (I, θ). To the accuracy $\omega \approx 1 - E/8$ the action $I \approx E/\omega_0$ as for a harmonic oscillator. Hence: $\alpha \approx -I/8$. For $E \ll 1$ the nonlinearity of small oscillations ($E, I \ll 1$) is small ($\alpha \ll 1$).

Another characteristic property of nonlinear oscillations is the *anharmonicity*, i.e. the presence of higher harmonics of the basic frequency ω (2.5). For small oscillations the amplitudes of higher harmonics are small as one can see from eq. (2.4) with $K' \approx \ln(4/k) = \frac{1}{2}\ln(32/E)$. Hence the small oscillations are nearly harmonic.

In the opposite case of a fast rotation ($H \approx E_0 \gg U_0$), the pendulum motion has a quite different character. The amplitudes of higher harmonics are also small (see eq. (2.6); $k = \omega_0/\omega_r \ll 1$), so that the motion is a nearly uniform rotation. However, the frequency of rotation does depend on the energy. Since in this case $p \approx I$ (the action) the nonlinearity parameter $\alpha \approx 1$, i.e. the nonlinearity is not small even though the anharmonicity may be arbitrarily small as $\omega_0/\omega_r \to 0$.

2.2. Canonical transformations and the superconvergence

In spite of great success in the search of completely integrable nonlinear systems [68] the latter are nevertheless exceptional, in my opinion. Anyway the problems we are going to consider do not belong to integrable ones. Hence for their analytical treatment one needs some approximate method, for instance, a version of perturbation theory. The latter is applicable, of course, only if there is a small parameter. According to this we assume that the Hamiltonian of our system may be divided into two parts:

$$H(I, \theta) = H_0(I) + \epsilon\, V(I, \theta) \tag{2.9}$$

the first of which describes an "unperturbed" system and has an integral of motion I (the action). The main "property" of the unperturbed system is our complete knowledge of its motion. Our problem is, however, to study the motion of a "perturbed" system with the Hamiltonian $H(I, \theta)$, and we assume the "perturbation" $\epsilon V(I, \theta)$ to be small ($\epsilon \ll 1$). A characteristic feature of the perturbation is the dependence of the latter on the phase θ of the unperturbed motion. This very dependence leads to a change in the unperturbed action I. Note that the division of the Hamiltonian into an unperturbed part and perturbation is somewhat arbitrary and depends, particularly, on our mathematical skill.

Assume we don't know any elliptic functions. Let us try to find the pendulum motion approximately, taking as an unperturbed system just the harmonic oscillator (the latter is known to everybody!):

$$H_0 = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega_0^2\varphi^2 = \omega_0 I. \tag{2.10}$$

The aim of these exercises is to demonstrate via a simple example the modern perturbation theory for Hamiltonian systems.

It has been known for a long time (see, e.g., ref. [17]) that successive canonical changes in the dynamical variables (or canonical transformations) in a Hamiltonian without making any use of the equations of motion is an efficient method for analytically treating Hamiltonian systems. The basic idea is to find such new variables $(\bar{I}, \bar{\theta})$ in which the perturbation, that is a part of Hamiltonian depending on $\bar{\theta}$, would become zero, would disappear. In other words, the task is to "kill" the perturbation by a change of variables. Then the new $\bar{I}$ is an integral of motion.

To find such an $\bar{I}$ is the same as to solve the problem completely, therefore, it is in general impossible to do in an explicit form. In classical perturbation theory [17] successive canonical transformations: $(I, \theta) \to (I_1, \theta_1) \to \cdots (I_n, \theta_n) \to \cdots (\bar{I}, \bar{\theta})$ are chosen in such a way to lower the order of perturbation by one power of the small parameter in every step: $\epsilon V \to \epsilon^2 V_1 \to \cdots \epsilon^{n+1} V_n \to 0$. However, such sequencies (the so-called asymptotic series, see, e.g., ref. [4]) are diverging, as a rule, and this restricts applications of perturbation theory to a fairly short time interval of motion.

A decisive success in this approach was achieved by a new perturbation theory due to Kolmogorov [1]. He observed that successive canonical transformations may be chosen in such a way that every

next perturbation becomes of the order of the square of the preceding one: $\epsilon V \to \epsilon^2 V_1 \to \epsilon^4 V_2 \to \cdots \epsilon^{(2^n)} V_n \to 0$.

Let us see how does this technique work for the simple example of a pendulum.* We write Hamiltonian (2.1) in the form ($M = \omega_0 = 1$):

$$H(p, \varphi) \approx \frac{p^2}{2} + \frac{\varphi^2}{2} - \frac{\varphi^4}{4!} + \frac{\varphi^6}{6!} - \frac{\varphi^8}{8!} \tag{2.11}$$

take the two first terms as the unperturbed Hamiltonian (2.10) and change the variables in H_0 to (I, θ). Then: $H_0 = I$; $\varphi = \sqrt{2I} \cos \theta$. Substituting the latter expressions into eq. (2.11) we obtain:

$$H(I, \theta) \approx I - \frac{4I^2}{4!} \cos^4 \theta + \frac{8I^3}{6!} \cos^6 \theta - \frac{16I^4}{8!} \cos^8 \theta. \tag{2.12}$$

The small parameter here is the action itself: $\epsilon = I$; $H_0 = \epsilon$; $V \sim H_0$.

We restrict ourselves to the problem of evaluating the pendulum frequency as a function of its energy: $\omega(H)$. The first correction to the unperturbed frequency $\omega_0 = 1$ can be obtained directly from eq. (2.12). To do this we divide the perturbation into the mean and oscillating parts with respect to the phase θ:

$$\epsilon V(I, \theta) = -\frac{I^2}{16} + \frac{I^3}{9 \times 32} - \frac{I^4}{9 \times 2^{10}} - \frac{I^2}{6}(\cos^4 \theta - \tfrac{3}{8}) + \frac{I^3}{90}(\cos^6 \theta - \tfrac{5}{16}) - \frac{I^4}{28 \times 90}(\cos^8 \theta - \tfrac{35}{128}). \tag{2.13}$$

We use, further, the so-called *averaging method* (see, e.g., ref. [4]), that is we simply neglect the oscillating part of perturbation. Then the averaged Hamiltonian takes the form:

$$\langle H(I, \theta) \rangle \approx I - \frac{I^2}{16} + \frac{I^3}{288} - \frac{I^4}{9216}. \tag{2.14}$$

A reason for such an averaging is an intrinsically true idea that the oscillating part of perturbation causes only "small" vibrations. The latter can be neglected, however, only to a certain accuracy. As we shall see below the last two terms in eq. (2.14) exceed the accuracy of averaging and are therefore false.

To treat the problem more accurately we make a canonical transformation to new variables $(I, \theta \to I_1, \theta_1)$ not to neglect the oscillating part of perturbation at will but to "kill" it legally.

Let the generating function of the canonical transformation be of the form

$$F(I_1, \theta) = I_1 \theta + \Phi(I_1, \theta). \tag{2.15}$$

This form provides a small difference between old and new variables while the function Φ is small. Indeed:

$$I = \partial F / \partial \theta = I_1 + \Phi_\theta; \qquad \theta_1 = \partial F / \partial I_1 = \theta + \Phi_{I_1}. \tag{2.16}$$

We substitute now the first expression (2.16) into eq. (2.12) and choose $\Phi(I_1, \theta)$ in such a way as to "kill" the term $\sim I^2 \sim \epsilon V$, which is linear in perturbation parameter ϵ, so that the perturbation would be $\sim \epsilon^2 V \sim I^3$. To achieve this it is sufficient to take

$$\Phi_\theta = \tfrac{1}{6} I_1^2 (\cos^4 \theta - \tfrac{3}{8}) \tag{2.17}$$

*For another example of using the superconvergence technique see section 5.1.

as one can see from eq. (2.13). It is essential that the latter expression has zero mean in θ as we have taken the separation of this mean beforehand (the term $-I^2/16$). Otherwise integration of eq. (2.17) would give a so-called secular term which grows indefinitely in time, and the difference $(\theta_1 - \theta)$ would not be small (see eq. (2.16)). Precisely due to this same reason one cannot "kill" the terms $\sim I^3$ by the same transformation (2.16) since the substitution of eq. (2.17) into the term $-\frac{1}{6}I^2(\cos^4\theta - \frac{3}{8})$ gives the mean $\sim I^3$. Now we see the accuracy of the averaging method naïvely applied to the original Hamiltonian (2.12). This accuracy is of the order of the mean perturbation after carrying-out the canonical transformation. Since for the Hamiltonian (2.12) the mean is $\sim I^3$ (see eq. (2.18) below), only the two first terms in eq. (2.14) may be used. To achieve a higher accuracy one has to perform a new canonical transformation, having splitted off the mean perturbation beforehand.

Now let us write down carefully the new Hamiltonian after the transformation of variables. We shall denote this Hamiltonian by $H_1(I_1, \theta_1)$. Since the original Hamiltonian (2.12) was taken to the accuracy $\sim I^4$ it is sufficient to calculate the new one to the same accuracy:

$$H_1(I_1, \theta_1) = I_1 - \frac{I_1^2}{16} + \frac{I_1^3}{9\times 32} - \frac{I_1^4}{9\times 2^{10}} - I_1^3\left(\frac{f_4}{48} - \frac{f_6}{90} + \frac{f_4^2}{18}\right) + I_1^4\left(\frac{f_4}{9\times 64} - \frac{f_8}{28\times 90} - \frac{f_4^2}{9\times 64} - \frac{f_4^3}{6^3} + \frac{f_4 f_6}{180}\right). \tag{2.18}$$

Here we have introduced the general notation:

$$f_n(\theta) = \cos^n\theta - \langle\cos^n\theta\rangle; \qquad \langle\cos^{2k}\theta\rangle = \frac{C_{2k}^k}{4^k} \tag{2.19}$$

where $C_{2k}^k = (2k)!/(k!)^2$ stands for the combination of $2k$ things k at a time. For odd n the mean $\langle\cos^n\theta\rangle = 0$. The perturbation, i.e. the terms depending on θ, have become smaller but, unfortunately, more complicated. And there's more to come, one should change θ for θ_1 using eq. (2.16). Since $\theta_1 - \theta = \Phi_{I_1} \sim I_1$ (2.17) it is sufficient to change $\theta \to \theta_1$ only in the terms $\sim I^3$. Then additional terms appear in the Hamiltonian (2.18):

$$H_1(I_1, \theta) \to H_1(I_1, \theta_1) + \frac{I_1^4}{18}\left[\frac{f_4'}{8} - \frac{f_6'}{15} + \frac{(f_4^2)'}{3}\right]F_4 \tag{2.20}$$

where $f_n' \equiv \mathrm{d}f_n/\mathrm{d}\theta$ and $\mathrm{d}F_n/\mathrm{d}\theta = f_n$ ($\langle F_n\rangle = 0$).

Splitting off the mean perturbation in eq. (2.20) we may perform a new canonical transformation $(I_1, \theta_1) \to (I_2, \theta_2)$. In the classical perturbation theory we would choose the transformation in such a way as to "kill" the oscillating terms $\sim I_1^3$. It turns out, however, that one can "kill" by this same transformation the terms $\sim I_1^4$ as well. Indeed, the new $\Phi^{(1)} \sim I_1^3$. Substitution of the latter into eq. (2.18) leads to the mean perturbation $\sim I_1^2\Phi_\theta^{(1)} \sim I_1^5$. Hence in the new Hamiltonian $H_2(I_2, \theta_2)$ the perturbation is $V_2 \sim I_2^5$. Repeating this reasoning we find that the perturbation in $H_3(I_3, \theta_3)$ is $V_3 \sim I_3^9$. In general, let the nth step of the successive approximations be $V_n \sim I_n^{k_n}$, then the next step is $V_{n+1} \sim I_{n+1}^{k_n - 1}\Phi_\theta^{(n)} \sim I_{n+1}^{2k_n - 1}$, hence: $k_{n+1} = 2k_n - 1$, whence: $k_n = 2^n + 1$. That is just the superconvergence of successive approximations according to Kolmogorov. The superconvergence allows us to construct the converging series. This enables us, in turn, to solve a number of important problems in the theory of dynamical systems (section 4.6). For practical applications which use a few first approximations the convergence is not essential as was noted correctly in ref. [4]. Nevertheless I think that Kolmogorov's method turns out to be more efficient in the latter case as well, since it demands less canonical transformations for a given accuracy of solution. In any event, this method is not well known so far, and its possibilities haven't been studied enough. Note that Bogolyubov has generalized this method to non-Hamiltonian systems [18, 63].

We come back now to the pendulum and find out both terms of the mean perturbation – the "classical" one $\sim I_1^3$, and the "extra" one $\sim I_1^4$. They are determined by the mean values:

$$\langle f_4^2\rangle = \tfrac{17}{128}; \qquad \langle f_4^3\rangle = \tfrac{3}{128}; \qquad \langle f_4 f_6\rangle = \tfrac{33}{256}.$$

When averaging with F_4 in eq. (2.20) one can integrate by parts: $\langle f_n' F_m\rangle = -\langle f_n f_m\rangle$. Finally, the averaged Hamiltonian takes the form:

$$\langle H_1\rangle = I_1 - \frac{I_1^2}{16} - \frac{I_1^3}{256} - \frac{5I_1^4}{2^{13}}\,. \tag{2.21}$$

Whence we get the oscillation frequency:

$$\omega \approx 1 - \frac{I_1}{8} - \frac{3I_1^2}{256} - \frac{5I_1^3}{2^{11}}\,. \tag{2.22}$$

Solving eq. (2.21) for I_1 to the accuracy of E^3 ($E = \langle H_1\rangle$) we obtain:

$$I_1 \approx E + \frac{E^2}{16} + \frac{3E^3}{256}. \tag{2.23}$$

Substituting this relation into eq. (2.22) we arrive at the exact expansion of ω in E (see eq. (2.7)).

2.3. The cubic force

Let us consider a special nonlinear oscillator with the Hamiltonian:

$$H(P, X) = \tfrac{1}{2}P^2 + \tfrac{1}{4}X^4. \tag{2.24}$$

We shall use this model in section 7. Like a pendulum the latter system has an exact solution in terms of elliptic functions (see, e.g., ref. [58]):

$$\begin{aligned}\frac{X(t)}{a} = \mathrm{cn}(at) &= \frac{\pi\sqrt{2}}{K(1/\sqrt{2})}\sum_{n=1}^{\infty}\frac{\cos[(2n-1)\omega t]}{\cosh[\pi(n-\frac{1}{2})]}\\ &\approx 0.9550\cos(\omega t) + \frac{\cos(3\omega t)}{23} + \frac{\cos(5\omega t)}{23^2} + \cdots.\end{aligned} \tag{2.25}$$

Here a is the amplitude of oscillations; $K(1/\sqrt{2}) \approx 1.8541$. An interesting feature of the system under consideration is the very small contribution of higher harmonics in spite of a large nonlinearity. The frequency of oscillations

$$\omega = \frac{\pi a}{2K(1/\sqrt{2})} = \beta a = \sqrt{2}\beta H^{1/4}; \qquad \beta \approx 0.8472 \tag{2.26}$$

is proportional to the amplitude. The action variable can be found from the relation $\omega = \partial H/\partial I$, whence:

$$I = \int_0^H \frac{\mathrm{d}H}{\omega(H)} = \frac{2\sqrt{2}}{3\beta}H^{3/4} = \frac{a^3}{3\beta}; \qquad H = AI^{4/3}; \qquad A = \left(\frac{3\beta}{2\sqrt{2}}\right)^{4/3}. \tag{2.27}$$

The nonlinearity parameter (2.8) is equal to:

$$\alpha = \frac{I}{\omega}\frac{\mathrm{d}\omega}{\mathrm{d}I} = \frac{I}{\omega}\frac{\mathrm{d}\omega/\mathrm{d}a}{\mathrm{d}I/\mathrm{d}a} = \tfrac{1}{3}; \qquad \frac{\mathrm{d}\omega}{\mathrm{d}I} = \frac{\beta^2}{a^2}. \tag{2.28}$$

2.4. The separatrix

We come back now to the pendulum and consider the border trajectory between rotation and oscillations. This trajectory is called the *separatrix*, or separating trajectory. It separates qualitatively different kinds of motion. One can readily imagine how unstable must be the motion in a close vicinity of the separatrix since a small perturbation can fling over the pendulum from oscillation to rotation and vice versa. That is precisely why we are especially interested in studying the motion near the separatrix. We shall see that just here an instability of nonlinear oscillations appears and is spreading around.

The separatrix corresponds to the pendulum energy $H = U_0$ (see eq. (2.1)) whence its phase trajectory is ($M = 1$; $-\pi \leqslant \varphi \leqslant \pi$):

$$p_{\mathrm{sx}} = \pm 2\omega_0 \cos(\varphi_0/2). \tag{2.29}$$

Different signs correspond with two *branches* of the separatrix. The separatrix forms in the phase plane a distinctive "cross" (fig. 2.1): $p = \pm\omega_0(\pi - \varphi)$. The "intersection" point of separatrix branches determines the position of unstable equilibrium. The latter should be considered as a separate trajectory since without any perturbation the pendulum remains in this position for an indefinitely long time. Whence it is clear that the motion along the separatrix is an asymptotic one arriving at, or departing from, the point of unstable equilibrium. The equation (2.29) can be integrated [5] to give:

$$\varphi_{\mathrm{sx}}(t) = 4\arctan(\mathrm{e}^{\omega_0 t}) - \pi \tag{2.30}$$

where time is counted from the instant of passing the position of stable equilibrium ($\varphi = 0$). The last expression clearly shows the asymptotic nature of the motion along the separatrix: $\varphi_{\mathrm{sx}} \to \pm\pi$ for $t \to \pm\infty$.

According to Arnold's graphic terminology [5] separatrix branches are called "whiskers" which are "fastened" to a "*whiskered torus*". In the case of the one degree-of-freedom conservative system under consideration, the whiskered torus has zero dimensionality and degenerates into a point – the position of unstable equilibrium ($\varphi = \pm\pi$). One distinguishes the *arriving* and *departing* whiskers (see fig. 2.1) on either of the two separatrix branches. They exchange positions under time reversal. For the system under consideration both whiskers on the same branch of separatrix coincide, yet under a perturbation they usually split (section 6.1). According to Poincaré's terminology [6] the whiskers are called doubly-asymptotic trajectories since they are approaching some trajectories for both directions of time ($t \to \pm\infty$). For our system this limiting trajectory – the unstable fixed point $\varphi = \pm\pi$ – is the same in both limits $t \to \pm\infty$. In this case the whiskers are called *homoclinic* trajectories.

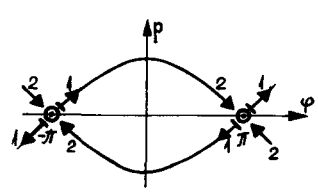

Fig. 2.1. Pendulum separatrix: O – a whiskered torus (point); 1 – departing whiskers; 2 – arriving whiskers.

Let us consider the pendulum motion near the separatrix. We will describe the distance from the separatrix by the relative energy:

$$w = \frac{H - U_0}{U_0} \approx \frac{p^2}{2U_0} + \frac{(\pi - \varphi)^2}{2} \ll 1. \tag{2.31}$$

Using the expressions in section 2.1 we find for both oscillation ($\omega < 0$) and rotation ($\omega > 0$): $k \approx \sqrt{|w|/2}$; $K(k) \approx \frac{1}{2}\ln(32/|w|)$; $K' \approx \pi/2$, and $\omega_r \approx \omega_0$. Hence for both kinds of motion, i.e. at both sides of the separatrix, the solution has the form (see eqs. (2.4), (2.6)):

$$\varphi(t) \approx 4 \sum_n \frac{\sin(n\omega t)}{n \cosh(\pi n\omega/2\omega_0)} \tag{2.32}$$

for rotation only even and for oscillation only odd harmonics being present. The frequency:

$$\omega(w) \approx \frac{\pi\omega_0}{\ln(32/|w|)} \tag{2.33}$$

is decreasing indefinitely on approaching separatrix ($|w| \to 0$). From the spectrum (2.32) one can conlude that the motion *near* the separatrix is approximately the same as *on* the separatrix except that the motion is of a finite period. It is clear also that the nonlinearity is growing indefinitely when approaching the separatrix (2.33).

3. Nonlinear resonance

Free oscillations in a conservative system with one degree of freedom are always stable (if the motion is finite, of course) and therefore are of less interest for us. What happens if one switches on an external perturbation? The Hamiltonian may be written in this case as:

$$H(I, \theta, t) = H_0(I) + \epsilon V(I, \theta, t). \tag{3.1}$$

The external perturbation is described here via an explicit dependence of the Hamiltonian on time. We will assume below that the perturbation is periodic in time with a period T, and the basic frequency $\Omega = 2\pi/T$. The frequencies of the perturbation spectrum are $n\Omega$ in this case. An almost periodic perturbation with an arbitrary discrete spectrum (Ω_n) does not lead to any qualitatively new effects. A perturbation with a continuous spectrum but restricted in time (a perturbation pulse) is of less interest since it causes only a small ($\epsilon \ll 1$) change in the oscillation energy. Finally, a stationary perturbation with a continuous spectrum, for example, an irregular sequence of pulses, causes a diffusion-like process in the system. The theory of such processes, which are very important for applications, comprises now a vast section of the theory of both linear and nonlinear oscillations. The latter problem is, however, beyond the framework of the present paper. So, we assume the perturbation to be periodically dependent on the phase

$$\tau = \Omega t + \tau_0 \tag{3.2}$$

where τ_0 is the initial phase. We expand the perturbation in a double Fourier series:

$$H(I, \theta, \tau) = H_0(I) + \epsilon \sum_{m,n} V_{mn}(I)\, e^{i(m\theta + n\tau)}. \tag{3.3}$$

The effect of a perturbation Fourier-component is the stronger the slower the time variation of its phase

$$\psi_{mn} = m\theta + n\tau. \tag{3.4}$$

In the limiting case of constant phase we come to a resonance:

$$m\omega(I) + n\Omega = 0. \tag{3.5}$$

This relation determines a set of resonant values of the frequency ω and, respectively, those of the energy (or I). Resonant values ω_{mn} form, generally, an everywhere dense set. Nevertheless, we will begin with the opposite limiting case when the sum in eq. (3.3) consists of a single term (or two complex conjugate terms, to be precise). In other words, we shall consider in this section the properties of a single (isolated) resonance. Interaction of many resonances will be discussed in the next section 4. We will start with a simple example of a parametric resonance (section 3.1), the example which has been studied many times in various applications, for instance, in the theory of accelerators [21]. One may find a similar approach to the theory of nonlinear resonance in papers by Ford and his co-workers [34, 40].

3.1. Parametric resonance

We have chosen this particular kind of nonlinear resonance because it is the simplest one for analytical treatment. Consider small oscillations of a pendulum with the Hamiltonian (see eq. (2.11)):

$$H \approx \tfrac{1}{2}(p^2 + \omega_0^2\varphi^2) - \varphi^4/24. \tag{3.6}$$

Let a parametric perturbation be of the form:

$$\omega_0^2(\tau) = 1 + \epsilon \cos \tau. \tag{3.7}$$

Changing the variables for I, θ we find:

$$H(I, \theta, \tau) = I + \epsilon I \cos^2 \theta \cos \tau - \frac{I^2}{6}\cos^4 \theta \to I + \frac{\epsilon I}{4}\cos(2\theta - \tau) - \frac{I^2}{16}. \tag{3.8}$$

One arrives at the last expression after the averaging, that is neglecting rapidly oscillating terms and retaining only the resonant one with the phase: $\psi = 2\theta - \tau$. Let us try first to "kill" the perturbation by a canonical transformation of variables (section 2.2). We choose the generating function of the form: $F(I_1, \theta) = I_1\theta + \epsilon\Phi \sin(2\theta - \tau)$. Unlike section 2 the generating function depends now on time, hence, the Hamiltonian will change under canonical transformation by the quantity $\partial F/\partial t$. Substituting $I = I_1 + 2\epsilon\Phi \cos(2\theta - \tau)$ and taking into account $\partial F/\partial t$ we demand that all terms of the order of ϵ cancel out. This gives:

$$\Phi = \frac{I_1/4}{\Omega - 2(1 - I_1/8)}. \tag{3.9}$$

We see that the perturbation can be "killed" only far off the resonance $\Omega = 2(1 - I_1/8)$. This justifies, by the way, neglecting non-resonant (rapidly oscillating) terms in the Hamiltonian. We are interested now, however, just in a resonance $\Omega \approx 2$ when such a "killing" of the perturbation is impossible due to a small resonant denominator.

Having left the hope to "kill" the perturbation let us try to get rid of the explicit time dependence in

the Hamiltonian (3.8) which is just the cause of the energy variation. To do this we introduce a new (resonance) phase $\psi = 2\theta - \tau$ and choose the generating function $F(I, \psi, \tau) = -I(\psi + \tau)/2$ so that $-\partial F/\partial I = \theta$. The new momentum becomes: $J = -\partial F/\partial\psi = I/2$. Note that the factor (1/2) provides preservation of the phase volume (area) under the canonical transformation: $\mathrm{d}I\,\mathrm{d}\theta = \mathrm{d}J\,\mathrm{d}\psi$.

In the new variables the Hamiltonian (3.8) takes the form:

$$H_r(J, \psi) = J\,\Delta - \tfrac{1}{4}J^2 + \tfrac{1}{2}\epsilon J \cos\psi = \text{const}; \qquad \Delta = 2 - \Omega. \tag{3.10}$$

This new Hamiltonian does not depend on the time explicitly and therefore remains constant. This means that in the new variables the system becomes conservative and, hence, integrable. We shall call H_r the *resonance Hamiltonian.* Let us mention that the quantity H_r might be called also the *quasi-energy* since it is the classical counterpart of the quantum quasi-energy introduced by Zeldovich and Ritus [19] (see also ref. [20]).

First, let initially $J = 0$, then $H_r = 0$ as well. Since we are not interested in the trivial solution $J \equiv 0$ we may divide the equation $H_r = 0$ by J. Then:

$$J = 4(\Delta + \tfrac{1}{2}\epsilon \cos\psi). \tag{3.11}$$

At the exact resonance for a very small ("zero") amplitude $\Delta = 0$, and the action reaches the maximal value $I_{\max} = 4\epsilon$. Hence the position of equilibrium $J = 0$ is unstable under the parametric perturbation.

The full frequency band of this instability starting from $J \approx 0$ may be found using the condition that for a certain ψ value $J = 0$. This gives the unstable band: $|\Delta| \leqslant \epsilon/2$.

It is essential, however, that for a nonlinear oscillator the energy (or the action) of oscillations is always restricted and small for small ϵ. The largest rise of the action corresponds to $\Delta = \epsilon/2$ and is equal to $I_m = 8\epsilon$. For a linear oscillator the term $\sim J^2$ in the Hamiltonian (3.10) would be absent, and we would arrive at the equation: $J(\Delta + \frac{1}{2}\epsilon \cos\psi) = \text{const}$. The latter determines the stop-band: $|\Delta| \leqslant \epsilon/2$ where the oscillation energy is growing indefinitely. This is a very important difference in behaviour of linear and nonlinear oscillators at a resonance. One may say that the nonlinearity stabilizes the resonant perturbation. The mechanism of this *nonlinear stabilization* is related just to the dependence of oscillation frequency on the energy because a variation of energy under perturbation leads to a variation of the frequency, and the system gets out of resonance.

On the other hand, the effect of a small resonant perturbation ($\Delta I = I_{\max} \sim \epsilon$) is much stronger than that of a non-resonant one ($\Delta I \sim \epsilon I \ll \epsilon$ since $I \to 0$, see eq. (3.8)). This shows the importance of resonant phenomena for nonlinear as well as for linear oscillations.

Let us consider now the case of a large detuning: $\Delta \gg \epsilon/2$. The resonance corresponds then to a certain finite amplitude of the oscillations ($\dot{\psi} \approx \Delta - J_r/2 = 0$) whence the resonant value of momentum is $J_r = 2\Delta$. We perform a new canonical transformation to introduce the momentum $p = J - J_r$ and to retain the same phase ψ. The necessary generating function can be easily constructed: $F(J, \psi) = -(J - J_r)\psi$. Substitution of $J = 2\Delta + p$ into eq. (3.10) gives the resonance Hamiltonian:

$$H_r(p, \psi) = \Delta^2 - \tfrac{1}{4}p^2 + \tfrac{1}{2}\epsilon(2\Delta + p)\cos\psi. \tag{3.12}$$

If $|p| \ll |\Delta|$ we can neglect p in the last term. Ignoring the constant term Δ^2 we arrive at the Hamiltonian of a pendulum with a "mass" $M = -2$ and $U_0 = \epsilon \cdot \Delta$. A negative value of the mass means simply that the position of stable equilibrium corresponds not to the minimum of potential energy ($\psi = \pi$) but to its maximum ($\psi = 0$). The region of a nonlinear resonance in phase space corresponds to the region of pendulum oscillations, i.e. a restricted variation of the phase ψ. This

region is situated inside the separatrix (section 2.4). The maximal p value and the frequency of small pendulum oscillations are equal to:

$$p_{\max} = 2\sqrt{2\epsilon\Delta}; \qquad \Omega_\Phi = \sqrt{\epsilon\Delta/2}. \tag{3.13}$$

Using the terminology borrowed of the theory of charged particle accelerators we shall call Ω_Φ the frequency of *phase oscillations* (the oscillations of the resonance phase ψ). The pendulum model is valid for $p_{\max} \ll J_r$. Besides one must demand the frequency of phase oscillations to be sufficiently small: $\Omega_\Phi \ll \omega \approx 1$. Otherwise the terms neglected in the Hamiltonian (like $\cos(2\theta + \tau)$) are no longer rapidly oscillating (as compared to the resonant term $\cos(2\theta - \tau)$) and therefore cannot be ignored even in the first approximation. Since the nonlinearity parameter (2.8) for small pendulum oscillations is equal to $\alpha = I/8 = J/4$ the two conditions for the validity of the pendulum model can be written as a double inequality:

$$\epsilon \ll \alpha \ll 1/\epsilon. \tag{3.14}$$

We shall call this inequality the condition of *moderate non-linearity.*

3.2. A universal description of a nonlinear resonance

We consider now a wider class of resonances. We single out an arbitrary term of the Fourier series (3.3) and write the Hamiltonian in the form:

$$H(I, \theta, \tau) = H_0(I) + \epsilon V_{mn}(I)\cos(m\theta - n\tau). \tag{3.15}$$

We have passed here to real functions and have changed the sign of n to emphasize the resonant nature of the retained term. We introduce a new (resonance) phase $\psi = m\theta - n\tau$ (3.4) and choose a generating function of the form:

$$F(I, \psi, \tau) = -(I - I_r)\left(\frac{\psi + n\tau}{m}\right) \tag{3.16}$$

where I_r is a certain constant (see below). The new momentum is equal to:

$$p = (I - I_r)/m. \tag{3.17}$$

Expanding the unperturbed Hamiltonian $H_0(I)$ up to terms $\sim p^2$ and taking into account that $\partial F/\partial t = -np\Omega$ we get:

$$H_r(p, \psi) \approx p^2/2M + \epsilon V_{mn}(I_r)\cos\psi. \tag{3.18}$$

We have ignored here the constant term $H_0(I_r)$ as well as the next terms of the expansion of $V_{mn}(I)$ in p. The quantity I_r is chosen so that the terms linear in p cancel: $m\omega_0(I_r) - n\Omega = 0$, that is I_r corresponds to the exact resonance. Thus we have arrived again, in a more general case, at the pendulum model. The pendulum mass is determined by the nonlinearity of the unperturbed oscillations:

$$M^{-1} = m^2\, \mathrm{d}\omega_0/\mathrm{d}I\Big|_{I=I_r} \equiv m^2\omega_0'. \tag{3.19}$$

The frequency of small phase oscillations is equal to:

$$\Omega_\Phi = \sqrt{\epsilon V_{mn}/M} \sim \sqrt{\epsilon\alpha}\, m\omega_0. \tag{3.20}$$

The last estimate assumes that the small parameter ϵ is adjusted for a given resonance in such a way that $V_{mn} \sim I_r\, \omega_0(I_r)$.

Let us write the separatrix equation in the original variables I, θ, τ. Note that τ may be considered as an additional coordinate of the system phase space, the coordinate with a given variation in time: $\tau = \Omega t + \tau_0$. If so, the phase space becomes 3-dimensional. One says therefore that the system has 1.5 degrees of freedom. From eqs. (2.29), (3.17) and (3.18) we get for the separatrix:

$$I_{sx} = I_r \pm (\Delta I)_r \sin\left(\frac{m\theta - n\tau}{2}\right), \qquad (\Delta I)_r = mp_r = 2m\sqrt{\epsilon M V_{mn}} \sim \sqrt{\frac{\epsilon}{\alpha}}\, I_r. \tag{3.21}$$

The amplitude of phase oscillations $(\Delta I)_r$ may be called the *resonance half-width* (in I). The quantity p_r is the oscillation amplitude of p (see eq. (3.17)). The resonance half-width in frequency is equal to:

$$(\Delta\omega)_r = \omega_0'(\Delta I)_r = \frac{2}{m}\sqrt{\frac{\epsilon V_{mn}}{M}} = \frac{2\Omega_\Phi}{m} \sim \sqrt{\epsilon\alpha}\,\omega_0. \tag{3.22}$$

Note that again the effect of a resonant perturbation (3.22) ($\sim\sqrt{\epsilon}$) is stronger than that of a non-resonant one ($\sim\epsilon$) (3.15).

The unstable position of pendulum equilibrium ($\psi = 0$; $\omega_0' V_{mn} > 0$) turns in the original variables into a closed spiral: $\theta = n\tau/m$ which is topologically equivalent to a ring. This curve is just a whiskered torus of dimensionality 1 (section 2.4) to which the whiskers (3.21) are fastened.

The universal description of a resonance in the pendulum approximation (3.18) is restricted, first, by the condition of smallness for the terms ignored in the expansion: $mp_r \ll I_r$, or according to eq. (3.21): $\epsilon \ll \alpha$. Secondly, one must demand the frequency of phase oscillations to be small enough for one to be able to neglect non-resonant terms in the Hamiltonian as rapidly oscillating. In section 3.1 we assumed the condition: $\Omega_\Phi \ll \omega_0$. Now, in a more general case (3.3), the corresponding condition must be considerably strengthened since the frequencies of non-resonant terms ($\omega_{mn} = m\omega + n\Omega$) are essentially lower than ω_0. The former are so-called small denominators (see section 4.6). We may write the condition for the pendulum approximation as:

$$\epsilon \ll \alpha \ll \frac{1}{\epsilon}\left(\frac{\omega_{mn}}{m\omega_0}\right)^2. \tag{3.23}$$

This is another form of the moderate nonlinearity condition (3.14).

The resonance equation $m\omega_0 = n\Omega$ for the harmonic (m, n) holds for harmonics (lm, ln) as well ($l \neq 0$ is any integer). Hence, the pendulum approximation is restricted also by the condition of smallness for higher harmonics: $V_{lm,ln} \ll V_{mn}$ $(l > 1)$.

Let us draw our attention to the curious point that the case of small nonlinearity turns out to be more complicated and more varied (section 3.1) as contrasted with a universal description under the moderate nonlinearity. A difficulty of the problem for small nonlinearity was excellently displayed by an example of an interesting system, studied in ref. [40], whose motion is unstable for any perturbation no matter how small it may be (see also the end of section 4.1).

The pendulum Hamiltonian (3.18) has been applied to many particular problems, for example, in the theory of synchrotron oscillations in accelerators [21]. The possibility of a universal description of a

nonlinear resonance within the framework of such a model was mentioned in ref. [8] and discussed thoroughly in ref. [9].

3.3. A resonance of many-dimensional oscillations

Let us consider a resonance of many-dimensional non-linear oscillations. We assume the Hamiltonian of the form:

$$H(I, \theta, \tau) = H_0(I) + \epsilon \sum_{m,n} V_{mn}(I)\, e^{i(m,\theta + n\tau)}. \tag{3.24}$$

Here I, θ, m are N-dimensional vectors, for instance: $I = (I_1, \ldots, I_N)$; N is the number of degrees of freedom; n is an integer; $m, \theta = m_i\theta_i$ is the scalar product*; $\tau = \Omega t + \tau_0$.

The unperturbed system $H_0(I)$ is completely integrable which means it possesses a full set of N integrals of motion: $I = \text{const}$. The motion of such a system is quasi-periodic with N basic frequencies: $\omega_k = \partial H_0/\partial I_k$, or more briefly: $\omega = \partial H_0/\partial I$. Nonlinearity of the system is described by the matrix: $\partial\omega/\partial I \equiv \partial\omega_i/\partial I_k = \partial^2 H_0/\partial I_i\, \partial I_k$. The system phase space is the direct product of N-dimensional space of the actions by N-dimensional space of the phases. An unperturbed phase trajectory covers an N-dimensional torus: $I = \text{const}$; $\theta = t\,\omega(I) + \theta^0$ where θ^0 is the initial phase vector. It is convenient to use also the action space (I-space) and the frequency space (ω-space) related to it. In both those spaces a torus is represented by a point.

The resonance condition has the form:

$$m,\omega(I) + n\Omega = 0 \tag{3.25}$$

which determines a *resonance surface* both in I- and ω-spaces. The resonances $n = 0$ are called *coupling resonances*. They bring about an exchange of energy among different degrees of freedom of the unperturbed system. Resonance structure is most graphically seen in ω-space where every resonance surface is simply a plane perpendicular to the vector m. Indeed, for any shift $\delta\omega$ along a resonance plane one gets from eq. (3.25): $m, \delta\omega = 0$.

Let us consider again a single resonance. The perturbation depends then only on a single phase combination:

$$\psi_1 = m, \theta - n\tau \tag{3.26}$$

of N linearly independent ones. Hence there are $N - 1$ integrals of motion. One may gain a graphic idea about those *resonance integrals* from the equations of motion:

$$\dot{I} = -\epsilon\, \partial V/\partial\theta = -i\epsilon m V_{mn}\, e^{i\psi_1}. \tag{3.27}$$

The last expression shows that the direction (or, better to say, the line) of vector $\dot{I}$ is fixed and parallel with that of the vector m. Note that for a coupling resonance ($n = 0$) both vectors ($\dot{I}; m$) are perpendicular to the vector ω and hence are tangential to the energy surface $H_0 = \text{const}$. This means, in turn, that a resonant perturbation preserves the unperturbed energy H_0. One may argue also another way: since at a single resonance the perturbation $V(\theta) = \text{const}$ ($\psi_1 = \text{const}$) the unperturbed energy must be constant also owing to $H = \text{const}$ (for coupling resonances only, of course).

*Summation over repeated subscripts is understood throughout the paper.

Possessing $N-1$ integrals of motion a many-dimensional system degenerates in the case of a single resonance into a system with only one degree of freedom.

We apply now a canonical transformation of variables choosing N linearly independent phase combinations as new coordinates:

$$\psi_k = \mu_{kl}\theta_l + \nu_k\tau \tag{3.28}$$

where μ_{kl} is some matrix, and the ν_k form a vector ($k, l = 1, \ldots, N$). Let ψ_1 be the resonance phase (3.26), and the others are arbitrary so far. We take a generating function of the form:

$$F(p, \theta, t) = (I_i^r + p_k\mu_{ki})\theta_i + p_k\nu_k\tau \tag{3.29}$$

where p is a new momentum vector (unknown at the moment), and the point I^r belongs to the resonance surface (3.25). The function (3.29) provides new phases of the form (3.28). New momenta can be obtained from the relation $I = \partial F/\partial\theta$, whence:

$$I_i = I_i^r + p_k\mu_{ki}; \qquad p_k = (I_i - I_i^r)\mu_{ik}^{-1}. \tag{3.30}$$

Here μ_{ik}^{-1} is the inverse of the matrix $\mu_{ik}(\mu_{ik}\mu_{kl}^{-1} = \delta_{il})$. The last expression (3.30) shows that new momenta p describe the deviation of the action vector I from a certain point I^r on the resonance surface (comp. with eq. (3.17) in section 3.2).

For a single resonance $N-1$ new momenta are integrals of motion: $p_k = \text{const}$ ($k \geqslant 2$). The integrals may be written also in the form (see eq. (3.30)):

$$I_i\mu_{ik}^{-1} = \text{const}; \qquad k = 2, \ldots, N. \tag{3.31}$$

Such resonance integrals are well known, for example, in the theory of accelerators [21] (see also ref. [34]).

Substituting the first relation (3.30) and also (3.26) into the Hamiltonian (3.24), and expanding $H_0(I)$ in p up to quadratic terms we arrive at the resonance Hamiltonian:

$$H_r(p, \psi) \approx p_k(\mu_{ki}\omega_i + \nu_k\Omega) + \frac{p_k p_l}{2M_{kl}} + \epsilon V_{mn}(I^r)\cos\psi_1 \tag{3.32}$$

where we have dropped the constant $H_0(I^r)$. The "mass" tensor is determined by

$$\frac{1}{M_{kl}} = \mu_{ki}\frac{\partial\omega_i}{\partial I_j}\mu_{lj}. \tag{3.33}$$

In eq. (3.32) we have left only one (real) perturbation term, since we consider a single resonance so far, have neglected all terms but the first in the expansion of $V_{mn}(I)$, and have introduced the term $\partial F/\partial t = p_k\nu_k\Omega$. Note that all quantities in eq. (3.32) depending on I are taken at the point $I = I^r$.

We may simplify the Hamiltonian (3.32) choosing all $\nu_k = 0$ except $\nu_1 = n$, and all the vectors μ_{ki} but one, say, μ_{2i} orthogonal to the vector $\omega^r = \omega(I^r)$. Let the vector μ_{2i} be parallel to ω, and denote its modulus by μ_2. The sum of eq. (3.32) linear in p_k can be reduced then using, in particular, the resonance condition (3.25) to a single term $p_2\mu_2|\omega^r|$ where $|\omega|$ is the modulus of ω. We shall use this "*orthogonal*" *metric* below.*

*Unless the resonance is a pure coupling one ($n = 0$) this method can be taken at all points of a resonance surface but one where the vectors ω^r and m are parallel. Near that point another metric proposed by Ford [139] needs to be introduced, namely, all the vectors μ_{ki} but $\mu_{1i} = m_i$ are set to be orthogonal to the vector ω^r. This metric is more preferable for $n \neq 0$ since it eliminates all the terms of Hamiltonian (3.32) linear in p_k including p_2. We retain, however, the previous option because we shall need it for coupling resonances in section 7.

If we set all resonance integrals $p_k = 0$ $(k > 1)$ what means the point I^r belongs to the trajectory of motion (see eq. (3.30)) we arrive at the pendulum model:

$$H_r = \frac{p_1^2}{2M} + \epsilon V_{mn}(I^r)\cos\psi_1; \qquad \frac{1}{M} = m_i \frac{\partial\omega_i}{\partial I_k} m_k \tag{3.34}$$

where the pendulum mass $M \equiv M_{11}$. Hence the analysis of section 3.2 is applicable, in particular, expressions (3.20) for the phase frequency and (3.21) for the separatrix configuration can be used, the quantities I and m being understood, however, as N-dimensional vectors now, and also $\omega_0 \to |\omega^r|$.

For what follows the structure of a resonance in both I- and ω-spaces is of importance. The direction and amplitude of phase oscillations are found from equation (3.21) to be determined by the *vector of phase oscillations*:

$$\begin{aligned} &(\Delta I)^r = m p_r; \qquad (\Delta\omega)^r = (\partial\omega/\partial I, m) p_r \\ &p_r = 2\sqrt{\epsilon M V_{mn}} = 2M\Omega_\Phi \end{aligned} \tag{3.35}$$

where the vectors $(\Delta I)^r$; $(\Delta\omega)^r$ correspond to I- and ω-spaces respectively. The vector $(\Delta\omega)^r$ is, generally, not parallel to the normal $m/|m|$ to the resonance plane. Projection of $(\Delta\omega)^r$ on that normal determines the resonance half-width in frequency:

$$(\Delta\omega)_r = \frac{m, (\Delta\omega)^r}{|m|} = \frac{2}{|m|}\sqrt{\frac{\epsilon V_{mn}}{M}} = \frac{2\Omega_\Phi}{|m|}. \tag{3.36}$$

Although the $N-1$ new momenta are time-independent this is not the case for the new frequencies $\dot\psi_k$ (and neither for the old ones (ω_k), of course). One readily deduces from eq. (3.32)

$$\begin{aligned} &\dot\psi_k = \frac{p_l}{M_{kl}} \to \frac{p_1}{M_{k1}} = \frac{M}{M_{k1}}\dot\psi_1; \qquad k \neq 2 \\ &\dot\psi = \mu_2|\omega^r| + \frac{p_l}{M_{2l}} \to \mu_2|\omega^r| + \frac{M}{M_{21}}\dot\psi_1 \end{aligned} \tag{3.37}$$

in the orthogonal metric (see above), the last expressions corresponding to the case $p_k = 0$ $(k > 1)$. Whence:

$$\psi_k = \frac{M}{M_{k1}}\psi_1 + \psi_k^0 \quad (k \neq 2); \qquad \psi_2 = \mu_2|\omega^r|t + \frac{M}{M_{21}}\psi_1 + \psi_2^0 \tag{3.38}$$

where ψ_k^0 are the initial values. We see that the phase oscillations in p_1, ψ_1 influence the motion of all other phases as well (but not the other p_k). Let us find also the motion in old phases θ_l. Solving eq. (3.28) for θ_l and using eqs. (3.38), (3.33) we get:

$$\theta_l = \mu_{lk}^{-1}(\psi_k - \nu_k \cdot \tau) = t\omega_l^r + \frac{\partial\omega_l}{\partial I_j} m_j M \psi_1 + \theta_l^0. \tag{3.39}$$

The simplest way to deduce the term $t\omega_l^r$ in the last relation is to observe that the former represents the unperturbed motion ($\epsilon = 0$, and, hence, $\psi_1 = \text{const}$) at a resonance surface $\omega(I^r) = \omega^r$.* Relation

*A more formal way is to verify the vector equality: $-n\Omega\mu_{i1}^{-1} + \mu_2|\omega^r|\mu_{i2}^{-1} = \omega_i^r$ $(i = 1, 2, \ldots, N)$ that may be done immediately if one multiplies the both sides by vectors μ_{il} $(i = 1, 2, \ldots, N)$.

(3.39) is obvious also from eq. (3.35) owing to $\Delta\omega = \dot{\theta} - \omega^r = (\partial\omega/\partial I, m)p_1$ and $p_1 = M\dot{\psi}_1$ (see eq. (3.37)).

The nonlinear stabilization of a resonant perturbation mentioned in section 3.1 depends on the mass M in the pendulum model of resonance (3.35). Unlike a system with 1.5 degrees of freedom considered in section 3.2 it may happen for a many-dimensional oscillator that the mass $M = \infty$, or the effective nonlinearity $M^{-1} = 0$. If so, there is no longer a nonlinear stabilization. To understand this special case another way note that the direction of phase oscillations in ω-space is along the vector $(\partial\omega/\partial I, m)$ (see eq. (3.35)), whereas the normal to a resonance plane is parallel to the vector m. Under the condition $M^{-1} = m$, $(\partial\omega/\partial I, m) = 0$ a trajectory goes along the resonance plane. Hence the system does not get out of resonance, so there is no nonlinear stabilization, nonlinearity does not "work", the nonlinear system behaves like an isochronous one. We shall call this special case of many-dimensional nonlinear oscillations *quasi-isochronous*. Note that the last conception relates to the unperturbed Hamiltonian H_0. To be precise, the condition $M^{-1} = 0$ is necessary but not sufficient for the quasi-isochronicity since nonlinear stabilization may happen due to some higher terms of the expansion of H_0 in p_1.

The phenomenon of quasi-isochronicity was apparently first mentioned in ref. [25]. A general study of the structure of nonlinear resonances was made by Nekhoroshev [24]. In particular, he introduced an important notion of the *steep* Hamiltonian. The exact definition of this notion is somewhat complicated (see ref. [24]) but essentially it means that the energy surfaces of a steep Hamiltonian are everywhere convex. Hence the plane tangential to an energy surface has the single common point with the latter (the point of tangency) whereas for a *non-steep* Hamiltonian that plane and the energy surface may intersect on a certain subsurface. If the convexity of the energy surface is provided already by the quadratic terms of the H_0 expansion near the point of tangency the Hamiltonian H_0 is with Nekhoroshev called *quasi-convex*. The latter property is a bit stronger than the condition $M^{-1} \neq 0$ for the pendulum model to be applicable. The difference is that for the pendulum model the convexity by quadratic terms along a resonance vector m, or, generally, along all *integer* vectors, is sufficient whereas the quasi-convexity means the "full" convexity, i.e. that along *any* direction in the tangent plane.* This slight difference is very important in the case of so-called *multiple resonances*.

A k-multiple resonance takes place when k resonance conditions of the type (3.25) with linearly independent vectors m^i $(i = 1, \ldots, k)$ are fulfilled simultaneously, i.e. for the same vector I^r, or ω^r. This happens on the intersections of k resonance surfaces (3.25).

The resonance Hamiltonian may be deduced now in a way similar to that for the *simple resonance* (3.25) $(k = 1)$. We may even write down the exact resonance Hamiltonian, that is without expansion in p_k:

$$H_r(p, \psi) = T(p) + U(p, \psi)$$
$$T = H_0(I_j^r + p_i\mu_{ij}) - H_0(I^r) + \Omega p_i\nu_i; \qquad U = \epsilon \sum_{l=1}^{k} V_{mn} \cos \psi_l. \tag{3.40}$$

Here T, U are the "kinetic" and "potential" energy of the system, respectively. Let the matrix μ_{ij} of the canonical transformation be of a special form to provide that every p_i $(i = 1, \ldots, k)$ describes motion along a particular vector m^i of the resonance basis, i.e. $\mu_{ij} = m_j^i$. Since the resonance

*To avoid misunderstanding note that in a general case $n \neq 0$ in eq. (3.25) we mean the plane tangential to an energy surface of the new unperturbed Hamiltonian (after the canonical transformation (3.29)): $H_0(I) + \Omega n_k p_k(I)$. The last terms linear in I (see eq. (3.30)) do not influence the convexity of the energy surface in I-space. It is more convenient, perhaps, to consider all those geometrical properties in the p-space of the new momenta (see below).

Hamiltonian depends on $k < N$ new phases ψ_l only there are $N-k$ resonance integrals, say, $p_l = \text{const}$ ($l = k+1, \ldots, N$) which may be set zero. Hence the Hamiltonian (3.40) describes a system with k degrees of freedom. Unlike a simple resonance the latter system is no longer integrable and may exhibit a very complicated, in particular, a stochastic, motion as we shall see later on (section 4.5). Nevertheless, and this is of a great importance, the motion under consideration is always bounded for a steep unperturbed Hamiltonian H_0 and a small enough perturbation. Indeed, if the surfaces $H_0(I) = \text{const}$ are convex, so do the surfaces $T(p) = \text{const}$ since the relation $I(p)$ is linear as is the additional term $\Omega p_i \nu_i$ in eq. (3.40). Therefore the only solution of the equation $T(p) = 0$ is the trivial one: $p = 0$ (the point of the tangency discussed above). Hence for $U \to 0$ the energy conservation $H_r = \text{const}$ bounds oscillations in p and, therefore, also in I. In the opposite case of a quasi-isochronous Hamiltonian the equation $T(p) = 0$ determines a certain subspace of the motion space p_l ($l \leq k$) and nothing prevents a trajectory from following this subspace indefinitely far away.

If we retain in the expansion of $T(p)$ (3.40) only quadratic terms as has been done for a simple resonance the oscillations at a multiple resonance are bounded if the kinetic energy $T(p) \approx p_i p_j / M_{ij}$ (comp. with eq. (3.32)) is sign-definite. This is just the quasi-convexity of Nekhoroshev. To guarantee boundedness of an arbitrary motion near a multiple resonance by the quadratic terms only if it is necessary, however, for the kinetic energy to be completely sign-definite, i.e. for all p values and not only for some of them as under the condition $M^{-1} \neq 0$ at a simple resonance. A mathematical theory of motion near a multiple resonance was developed by Melnikov [114] and by Moser [115].

As an example of many-dimensional oscillations let us consider the system of two coupled oscillators with a cubic nonlinearity (section 2.3) [41]:

$$H = \tfrac{1}{2}(P_1^2 + P_2^2) + \tfrac{1}{4}(X_1^4 + X_2^4) - \mu X_1 X_2 \tag{3.41}$$

where μ is a small coupling parameter. We introduce the action-angle variables and take into account that the unperturbed system ($\mu = 0$) is nearly harmonic (section 2.3): $X_i \approx a_i \cos\theta_i$ ($i = 1, 2$). Then:

$$H(I, \theta) \approx A(I_1^{4/3} + I_2^{4/3}) - \tfrac{1}{2}\mu a_1 a_2 \cos(\theta_1 - \theta_2) \tag{3.42}$$

where $A = (3\beta/2\sqrt{2})^{4/3}$ (see eq. (2.27)), and we have left only the perturbation term responsible for the coupling resonance $\omega_1^r = \omega_2^r$ which we are just going to study.

The unperturbed Hamiltonian is quasi-convex since both second derivatives $\partial^2 H_0 / \partial I_i^2 = \beta^2 / a_i^2 > 0$ (section 2.3). Since the resonance phase is $\psi_1 = \theta_1 - \theta_2$ it is natural to choose the second phase as $\psi_2 = \theta_1 + \theta_2$. This option leads to an orthogonal metric:

$$\begin{gathered} \mu_{ik} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}; \qquad \mu_{ik}^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \\ p_1 = \tfrac{1}{2}(I_1 - I_2) = p; \qquad p_2 = \tfrac{1}{2}(I_1 + I_2) - I_0 = 0. \end{gathered} \tag{3.43}$$

We set $I^r = (I_0, I_0)$ and determine I_0 from the condition $p_2 = 0$. The resonance Hamiltonian has the form (3.34) with parameters:

$$M^{-1} = 2\beta^2 / a^2; \qquad |\epsilon V_{mn}| = \tfrac{1}{2}\mu a^2 \tag{3.44}$$

where we set approximately: $a_1 \approx a_2 = a(I_0)$. The vector of phase oscillations $(\Delta\omega)^r$ is perpendicular to the resonance line $\omega_1 = \omega_2$ and equal to (see eq. (3.35)):

$$(\Delta\omega)^r = m\Omega_\mu; \qquad |(\Delta\omega)^r| = \sqrt{2}\Omega_\mu = (\Delta\omega)_r; \qquad \Omega_\mu = \beta\sqrt{\mu}. \tag{3.45}$$

Here Ω_μ is the frequency of small phase oscillations (3.20) and $(\Delta\omega)_r$ is the resonance half-width in frequency.

The position of stable equilibrium for the pendulum describing the coupling resonance under consideration is determined by the condition $\theta_1 = \theta_2$ ($\psi_1 = 0$, see eq. (3.42)). This is a periodic trajectory corresponding to in-phase oscillations for both degrees of freedom. The position of the unstable equilibrium of the pendulum ($\theta_1 - \theta_2 = \pi$) determines the Arnold whiskered torus which is also a periodic trajectory but corresponding to the out-of-phase oscillations.

4. Interaction of resonances and the stochastic instability

The interaction of resonances is understood as the simultaneous influence of several resonances, i.e. several perturbation harmonics of the Hamiltonian (3.3) (or (3.24) in the many-dimensional case), upon a system. Each perturbation harmonic determines its "own" resonance in a particular domain of the phase space. What is the motion under the simultaneous influence of many resonances like? Until recently, this fundamental question was left unanswered. It seems, at the first glance, that while a system is near one particular resonance the other perturbation terms are nonresonant, and can be "killed", therefore, by a change of variables (see sections 2.2 and 3.1). Indeed, it has been known for ages (see, e.g., ref. [17]) that one can construct a sequence of variable transformations, generally speaking, an infinite one, which does "kill" the non-resonant terms. In the limit one arrives, hence, at an integrable system whose motion is stable. However, this is but a formal mathematical trick. Real stability depends on whether that infinite sequence of variables is convergent, or, in other words, whether there exists that limit in which the motion is formally stable. Until recently, it was considered that such a limit, generally, does not exist, i.e. a generic (typical) nonlinear oscillator is unstable [6, 27]. These sort of ideas was based, particularly, on the famous Poincaré theorem [6] that a generic Hamiltonian system has no analytical integral of motion except the energy.* Today we know that such an understanding of the Poincaré theorem is in error (section 4.6). Not a long time ago a new theory of the dynamical system stability has been developed by Kolmogorov [1], Arnold [2] and Moser [3] (the KAM theory) who managed to construct convergent perturbation series and to show that, generally, there exists a certain critical value of a small perturbation which determines the border of motion stability. The basic ideas of the KAM theory will be outlined in section 4.6. The main result of this theory – the existence of a stability border for non-linear oscillations – is of fundamental importance for the general theory of dynamical systems.

One may attack the problem along a quite different line. Namely, one may try to investigate qualitatively, I would say even graphically, particular mechanisms responsible for the destruction of integrals of motion under a sufficiently strong perturbation and for the resulting instability. The present section is devoted to the description of just such an approach to the problem. It turns out that a fairly general mechanism for arising instability of the motion is the so-called overlap of nonlinear resonances [8, 9]. Study of this mechanism allows to formulate a very efficient, though not rigorous, criterion of stability. Using such a criterion one was able to obtain estimates of the stability border for a number of problems [29–37, 59, 92–94, 131].

The instability of nonlinear oscillations arising under the overlap of resonances has a rather peculiar nature resulting in an irregular, or stochastic, motion of the system. This kind of mechanical

*And, of course, generally all the other additive integrals: momentum, angular momentum and center-of-mass integral.

motion has been the object of a long search in numerous attempts of foundation of the statistical mechanics. This is just where the ergodic theory of dynamical systems [28, 2] has come out. A rather surprise result of the study of nonlinear oscillations has proved to be the possibility of statistical behaviour in extremely simple dynamical systems [10, 38, 39, 34]. The latter aspect of resonance interaction will be discussed in section 5.

4.1. Two resonances; the overlap criterion

The simplest case of resonance interaction is the interaction of only two resonances. Apparently the first example of such a system was studied in detail both analytically and numerically by Walker and Ford [34]. We shall consider below another example, a simpler one for theoretical analysis. We shall need the latter example in section 7.5.

Let an oscillator with a cubic nonlinearity (section 2.3) be acted upon by a driving force of two frequencies Ω_1, Ω_2. The Hamiltonian has the form:

$$\begin{aligned} &H(P, X, t) = \tfrac{1}{2}P^2 + \tfrac{1}{4}X^4 - X(f_1 \cos \tau_1 + f_2 \cos \tau_2) \\ &\tau_1 = \Omega_1 t + \tau_{10}; \qquad \tau_2 = \Omega_2 t + \tau_{20} \end{aligned} \tag{4.1}$$

where τ_{10}, τ_{20} are the initial phases and f_1, f_2 stand for small perturbation amplitudes. We change the variables to I, θ; assume that $X \approx a \cos \theta$ (section 2.3) and retain two resonant terms only:

$$H(I, \theta, \tau_1, \tau_2) \approx AI^{4/3} - \tfrac{1}{2}a[f_1 \cos(\theta - \tau_1) + f_2 \cos(\theta - \tau_2)]. \tag{4.2}$$

Each of two perturbation terms is responsible for its own resonance in a different domain of phase space: $\omega_{1,2}(I) = \Omega_{1,2}$. If there is a single resonance, i.e. either $f_1 = 0$ or $f_2 = 0$, the system is integrable, as we know, and the motion is stable and locked within a resonance domain. We may find out the latter domain using the technique of section 3.2. In our case $M^{-1} = \beta^2/a^2$ and $\epsilon V_{mn} = af_0/2$, the amplitude f_0 being a small perturbation parameter. The resonance half-width in frequency is equal to (see eq. (3.22)):

$$(\Delta\omega)_r = \beta\sqrt{2f_0/a}. \tag{4.3}$$

Let the domains of two resonances be situated sufficiently far from each other, i.e. the difference $|\Omega_1 - \Omega_2|$ is large enough. Then, taking into account the smallness of the perturbation it is reasonable to expect the motion of the system to be still locked within the domain of one resonance or the other depending on initial conditions. As to the second (non-resonant) perturbation term for given initial conditions it distorts only slightly the trajectory as compared to the event of a single resonance. This expectation is well confirmed by numerical experiments (see, e.g., ref. [34] and section 5.1). On the other hand, it is clear that if resonance domains are approaching each other due to a decrease in the frequency difference $|\Omega_1 - \Omega_2|$, for example, one cannot neglect the influence of the second resonance after all. It is obvious that if resonances are close enough a trajectory is no longer locked within one of the resonances, and the system passes, generally, from one resonance to another. This kind of motion could hardly be classified as unstable (at least for two resonances; see, however, section 4.2) if it were not of a qualitatively different nature. Numerical experiments show that the motion in question becomes irregular as if the system were influenced by some random forces even though, in fact, no such a force is present (see eq. (4.1)). That is why this kind of motion was called *stochastic oscillations*, or *stochastic instability*. Its main features will be considered in section 5. Let us mention

incidently that another peculiarity of such a motion is a strong local instability. The latter means that the trajectories started initially close together are separating exponentially with time at the average (section 5.2).

A plausible condition for the occurrence of the stochastic instability seems to be the approach of resonances down to the distance of the order of a resonance size. Such an approach was naturally called the *resonance overlap*. To be precise, the overlap of resonances begins when their separatrices touch each other. The possibility for a system to move from one resonance to another under the above condition is quite obvious. The problem is another one: how to calculate the condition of separatrix touching taking into account a deformation of the separatrix by a neighbouring resonance? The simplest method, a quite rough one, is to use the unperturbed resonance parameters, i.e. to consider each of the resonances as if another one were absent. It is clear that one may expect to get in this way only an order-of-magnitude estimate. This simple procedure is just what one means when talking about the overlap criterion. This criterion, thus, turns out to be a quite rough one, yet it is fairly efficient since the above procedure may be easily performed even in the case of a rather complicated system (section 4.5).

The overlap criterion has also a more serious defect. The point is that the number of resonances, say, for the system (4.2) is equal to 2 only in the first approximation. This is due to the fact that the phase θ depends linearly on time only in zeroth approximation. Affected by a perturbation the phase dependence on time becomes very complicated resulting in new resonances. Generally, resonances $m\omega + n_1\Omega_1 + n_2\Omega_2 = 0$ happen with any integers m, n_1, n_2 that is the set of resonances turns out to be everywhere dense (section 4.6). It is clear, therefore, that absence of the overlap of first order resonances is only a necessary but not sufficient condition of motion stability. One may say also that the overlap criterion gives only an upper limit (in the perturbation strength) for stability. What is more, a fundamental question arises whether there exists any region of stability? And can it happen that such an everywhere dense set of resonances leads to an instability of any motion unless the system is completely integrable (as in the event of a single resonance, for example)? It is obvious that those questions cannot be answered only by visual considerations of the type we used above. To answer them one needs either experiments, including numerical ones, or a rigorous theory. The first way has the common restrictions for experiment, for example, finite time interval over which one can be convinced (and convince the others!) of motion stability (section 5.1). The other way is limited by simplifications of theoretical models (section 4.6). Let us just mention now that only for systems with two degrees of freedom one manages to prove the absolute, or eternal, stability of motion. For $N > 2$ the motion is always unstable in some sense (sections 4.6 and 7.7).

Coming back to the Hamiltonian (4.2) let us apply the overlap criterion. We assume $\Delta\Omega = |\Omega_1 - \Omega_2| \ll \Omega_1$ and $f_1 = f_2 = f_0$. The condition for the unperturbed resonance separatrices to touch has a simple form:

$$(\Delta\omega)_r \approx \tfrac{1}{2}\Delta\Omega. \tag{4.4}$$

From eqs. (4.4) and (4.3) we find the (theoretical) critical value of f_0 as:

$$f_T \approx \omega(\Delta\Omega)^2/8\beta^3 \tag{4.5}$$

where $\omega \approx \Omega_1 \approx \Omega_2$.

Let us compare this estimate with the results of numerical experiments [41]. In those the rate and character of the variation in the Hamiltonian were used as an experimental indication of instability. Since one expected a stochastic behavior under resonance overlap the "diffusion rate" was computed:

$$D_n = \overline{(\Delta\bar{H})^2/\Delta t}. \tag{4.6}$$

Here $\bar{H}$ is the value of Hamiltonian averaged over a period of $\Delta t_n = 10^n$ time units of the system (4.1). The averaging was intended to lower the influence of bounded energy oscillations and to pick out the accumulating changes. The second averaging (upper bar in eq. (4.6)) was done in the following way. The total motion time t ($=10^6$ typically) was subdivided into a number of equal intervals Δt_n. The averaging was made then over all pair combinations of those intervals, $\Delta\bar{H}$ meaning the difference in $\bar{H}$ between the intervals of a given pair and Δt being the time difference between the centers of these intervals. This averaging procedure aimed to increase the time scale for which diffusion was described by the rate (4.6). This facilitated the separation of diffusion processes from side effects. For the above procedure the mean value of Δt in eq. (4.6) is about half the total motion time and is independent of the length of interval Δt_n. Typically two interval lengths were used $\Delta t_4 = 10^4$ and $\Delta t_5 = 10^5$ to see how strong are non-diffusion processes. For bounded oscillations of the energy, for example, the measured "diffusion rate" (4.6) drops proportional to $(\Delta t_n)^2$ due to the averaging over the interval Δt_n.

In fig. 4.1 a dependence of the diffusion rate on the perturbation strength is presented for $\Omega_1 = 0.217$; $\Omega_2 = 0.251$. The initial conditions of motion correspond to the value of frequency $\omega = 0.234$ ($a = 0.276$) which lies just half-way between Ω_1 and Ω_2. We see that within a fairly narrow interval of $f_0 = (2.8\text{–}2.6) \times 10^{-5}$ the diffusion rate D_5 drops by about 6 orders of magnitudes. Is that not the genuine border of stability! We may assume the position of the border to be

$$f_E \approx 2.55 \times 10^{-5}; \qquad f_T \approx 5.76 \times 10^{-5}. \tag{4.7}$$

Here f_E is the experimental value (from the data of fig. 4.1) and f_T stands for that according to expression (4.5). The agreement between f_E and f_T may be considered as satisfactory taking account of the overlap criterion roughness discussed above. Note that the experimental value is well below as compared to the theoretical one ($f_T/f_E \approx 2.26$). This is also in accordance with the above-mentioned considerations. We shall come back to this question in section 5.1 (see also the next section).

Above the stability border (4.7), say, for $f_0 > 2.7 \times 10^{-5}$ both rates D_4, D_5 are approximately equal,

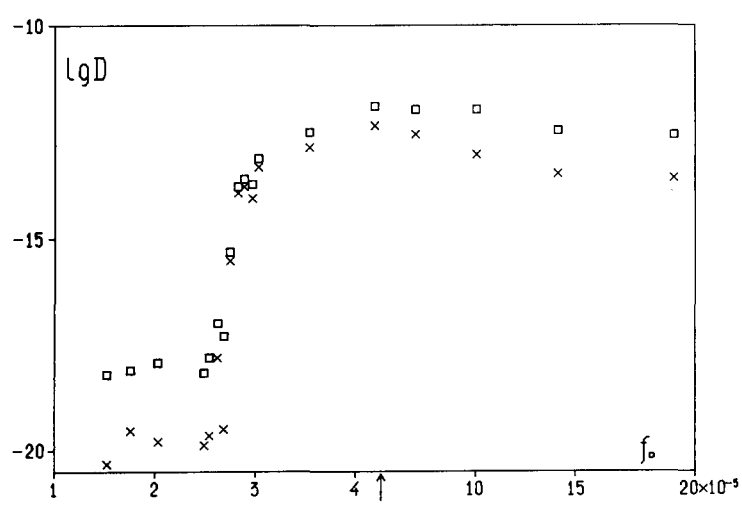

Fig. 4.1. Stochastic instability under the overlap of two resonances; diffusion rate D_n versus perturbation strength f_0 (see the text): □ – D_4; × – D_5; the scale of f_0-axis is different on both sides of the arrow.

the latter being an indication of the diffusion nature of motion. In the paper [34] an irregular character of motion under the resonance overlap is seen quite clearly on phase space pictures of the motion and differs sharply from a regular behavior of trajectories in the absence of the overlap. Below the stability border ($f_0 < 2.6 \times 10^{-5}$) the rates D_4, D_5 differ by 2 orders of magnitude, so that the energy variation in this region is of a non-diffusive nature obviously, being related, mainly, to quasi-periodic oscillations under the influence of the perturbation as well as to computation errors.

It is interesting to note that under resonance overlap the diffusion rate does not increase with f_0 and even falls slightly, both rates going apart again. The latter indicates a reduction of the diffusion contribution as f_0 grows. This, at the first glance, rather strange behavior can be understood if one considers the limiting case $(\Delta\omega)_r \gg |\Delta\Omega|$. The system behavior in that case is especially clear for $\Delta\Omega \to 0$. Then both resonances become one with the double amplitude, and the system is integrable again. One can conclude that a system is most unstable (in the case of two resonances) just around the overlap border:* $(\Delta\omega)_r \geqslant (\Delta\Omega)/2$.

The diffusion due to the overlap of the resonances cannot give rise to a large change of the unperturbed integrals of motion since the area of stochasticity is restricted to the "fused" domains of two resonances. That bounded area is clearly seen in phase space pictures of ref. [34].

Let us consider another quite simple example of resonance overlap for a system with the Hamiltonian:

$$H = \tfrac{1}{2}(P_1^2 + P_2^2) + \tfrac{1}{4}(X_1^4 + X_2^4) - \mu X_1 X_2 - X_1 f_0(\cos\tau_1 + \cos\tau_2). \tag{4.8}$$

We shall use the results in section 7.6. The system under consideration consists of two coupled oscillators with a cubic nonlinearity (section 2.3) one of which is acted upon by a driving force of two frequencies (4.1). The configuration of resonance bands for this system is outlined in fig. 4.2. The coupling resonance $\omega_1^r = \omega_2^r$ intersects both driving resonances: $\omega_1^r = \Omega_1$; $\omega_1^r = \Omega_2$. Hence there always exists two stochastic domains at those intersections, the domains formed by multiple (double) resonances: 1) $\omega_1^r = \omega_2^r = \Omega_1$; 2) $\omega_1^r = \omega_2^r = \Omega_2$ (section 3.3). Under a certain condition, however, a new possibility arises – the transition of the system from one driving resonance into another through the coupling resonance. This transition happens under the "closure" of the chain of vectors $(\Delta\omega)^r$ along which the phase oscillations proceed. In fig. 4.2 the critical situation is shown when the closure just begins according to the condition:

$$(\Delta\omega)_1^{(1)} + (\Delta\omega)_1^{(2)} + 2(\Delta\omega)_1^{(12)} = |\Omega_1 - \Omega_2|. \tag{4.9}$$

Here the two first terms represent components of the phase oscillation vectors for the two driving resonances along the ω_1-axis and $(\Delta\omega)_1^{(12)}$ is the same component for the coupling resonance. Using

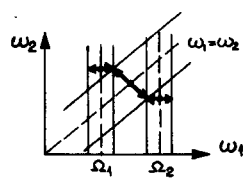

Fig. 4.2. Overlap of three resonances: arrows indicate phase oscillation vectors $(\Delta\omega)^r$ for each resonance.

*Interesting numerical data related to this question see in ref. [135].

expressions (4.3) and (3.45) we arrive at the condition for the overlap (touching) of three resonances:

$$\sqrt{\mu} + \sqrt{2f_0/a} = \Delta\Omega/2\beta \tag{4.10}$$

where we regard for simplicity the parameters of both driving resonances as equal. For a given $f_0 < f_E$ (see eq. (4.7)) the relation (4.9) determines the critical value of μ above which all three resonance domains are linked together by a united (stochastic) motion of the system.

A distinct feature of all the examples considered is the existence of a critical perturbation below which nonlinear oscillations are stable. This is related to the separation of resonance domains under a sufficiently small perturbation. As was mentioned already above the separation does not happen at intersections of resonance surfaces so that in a vicinity of the intersection, i.e. at a multiple resonance, a stochastic domain persists for arbitrarily small perturbation. In the case of a nonlinear unperturbed system ($\alpha \neq 0$, see eq. (2.8)) the dimensions of that stochastic domain are decreasing in proportion to the width of the nonlinear resonance, that is as $\sqrt{\epsilon/\alpha}$ (see eq. (3.21)). However, if $\alpha \sim \epsilon$, i.e. the nonlinearity of oscillations is determined also by the perturbation, and linear frequencies of a system satisfy, at least, two resonance conditions then only the rate of instability is decreasing under $\epsilon \to 0$ but not the dimensions of the stochastic domain. This interesting phenomenon has been discovered by Ford and Lunsford [40].

4.2. Many resonances; mappings

The overlap of two, or a few, resonances results only in a *confined instability* as we shall say. The latter term refers to the fact that a stochasticity domain formed by a few overlapping resonances is finite but bounded in phase space.* Hence the energy variation or the energy exchange among different degrees of freedom in a conservative system is limited and small under a small perturbation. But that's quite another thing if there are many resonances. Then a trajectory of motion may go within a set of overlapping resonances far away from the initial position. An instability at large, or the *gross instability* occurs. In this section we shall consider the limiting case of an infinite number of resonances. A convenient model is a system whose motion is described by difference equations, or by a mapping. Let us consider, as an example so far, a simple 2-dimensional mapping $I, \theta \to \bar{I}, \bar{\theta}$ where

$$\bar{I} = I + K f(\theta); \qquad \bar{\theta} = \theta + \bar{I}. \tag{4.11}$$

Here K stands for a constant parameter and $f(\theta)$ is some function. The mapping (4.11) is a canonical one generated by the function:

$$F(\bar{I}, \theta) = \bar{I}\theta + \tfrac{1}{2}\bar{I}^2 + K\,V(\theta); \qquad f = -\mathrm{d}V/\mathrm{d}\theta. \tag{4.12}$$

For $K \ll 1$ the difference equations (4.11) can be replaced approximately by differential ones

$$\dot{I} \approx (K/T)\, f(\theta); \qquad \dot{\theta} = I/T \tag{4.13}$$

and we arrive at a conservative system, T being the time period corresponding to one iteration of the mapping. To write down the difference equations in the form of exact differential ones we may use a δ-function:

$$\dot{I} = K\Omega\, f(\theta)\, \delta_{2\pi}(\tau); \qquad \dot{\theta} = I/T. \tag{4.14}$$

*The confined instability should not be confused with the local instability, the latter term referring to the behavior of infinitely close trajectories that is to the instability in an infinitely small domain of the phase space (see section 5.2).

Here $\tau = \Omega t + \tau_0$; $\Omega = 2\pi/T$ and the δ-function of period 2π is given by the Fourier expansion:

$$\delta_{2\pi}(\tau) = \frac{1}{2\pi}\left(1 + 2\sum_{n=1}^{\infty} \cos(n\tau)\right). \tag{4.15}$$

It is convenient to introduce a new momentum $J = I/T$, then the equations of motion become:

$$\dot{J} = (K\Omega^2/2\pi)\, f(\theta)\, \delta_{2\pi}(\tau); \qquad \dot{\theta} = J \tag{4.16}$$

and the Hamiltonian:

$$H(J, \theta, \tau) = \tfrac{1}{2}J^2 + (K\Omega^2/2\pi)\, V(\theta)\, \delta_{2\pi}(\tau). \tag{4.17}$$

Note that even though the variables J, θ are canonical the transformation $I, \theta \to J, \theta$ is not since the Jacobian $\partial(J, \theta)/\partial(I, \theta) = 1/T \neq 1$.

Equations (4.16) are equivalent to the original mapping (4.11), so we see that the mapping describes an external perturbation of period T. This perturbation represents a sequence of short "kicks" (4.15) and thus results in an infinite set of resonances (see eq. (4.17) and below).

The described scheme for analysis of nonlinear resonances was used in refs. [9, 46] (see also ref. [43]).

Let $V(\theta) = \cos\theta$; $\Omega = 1$ $(T = 2\pi)$. We arrive then at a pendulum which is acted upon by a periodic perturbation:

$$H(J, \theta, \tau) = \tfrac{1}{2}J^2 + k\sum_{n=-\infty}^{\infty} \cos(\theta - nt) \tag{4.18}$$

where $k = K/(2\pi)^2$ is a new parameter. Assuming the latter to be small we obtain the set of first approximation resonances: $J_r = n$, or $I_r = 2\pi n$. All these resonances are identical except a shift in J. This is obvious also from the original mapping (4.11) since a shift in I by an integer multiple of 2π leaves the mapping unchanged. It may be noted that the approximate replacement of the difference equations (4.11) by differential ones (4.13) is equivalent to taking account of the only resonance $J_r = 0$.

The separatrix half-width of every resonance in J, or in the unperturbed frequency $\omega = \dot{\theta} = J$, is equal to:

$$(\Delta\omega)_r = (\Delta J)_r = 2\sqrt{k}. \tag{4.19}$$

The spacing between resonances: $\Delta\omega_r = \Delta J_r = 1$. From the condition of touching separatrices $(\Delta\omega)_r = \frac{1}{2}$ we get the critical perturbation strength:

$$k_T = \tfrac{1}{16}; \qquad K_T = \tfrac{1}{4}\pi^2 \approx 2.5. \tag{4.20}$$

Although this estimate gives the correct order of magnitude it materially overstates the critical strength of perturbation as well as in the case of two resonances (section 4.1). Numerical experiments give the value: $K_E \approx 1$ (section 5.1). A reason for such a disagreement and improvements of the theoretical estimate will be discussed in sections 5.1 and 6.4. Note that the ratio $K_T/K_E \approx 2.47$ is close to the analogous quantity $f_T/f_E \approx 2.26$ for two resonances (section 4.1).

Since in the model under consideration all the resonances overlap simultaneously a gross instability occurs resulting in an arbitrarily large variation of J. In absence of overlap the change in J is limited by the width of the separatrix: $|\Delta J| \leqslant 2\sqrt{k}$ (see eq. (4.19)).

In section 3 we have seen that a main feature of a single resonance is the nonlinear stabilization of resonant perturbation. The mechanism of resonance overlap destroys that stabilization and lets a

system "wander" over resonances. Arising thus, stochastic instability is the main obstacle in the way of application of nonlinear oscillations for the suppression of resonant perturbation. That has been found out already in ref. [10].

A vast number of papers has been devoted to the study of nonlinear mappings. One may gain an idea of some modern trends in this field from the materials of the Toulouse conference on "Point Mappings and its Applications" [44] (see also ref. [54]). Some times mappings have been used as a simple approximate method for describing motion of a continuous system, for example, in many problems concerning acceleration of charged particles [10, 29, 35, 45–47]. These works are close to the present paper as to methods and results. In refs. [29, 46] similar criteria of stability have been deduced using different methods.

In a number of papers the mappings obtained by the so-called Poincaré surface of section method were studied [17, 38, 116, 54]. The description of such Poincaré mappings is given in the next section via an example of a conservative system having two degrees of freedom. This application of mappings is of a special interest since it shows that mappings are an adequate method of motion description for systems of a fairly broad class even though, at the first glance, this method seems to be artificial and formal. What is more, mappings are very useful for computations (see sections 5, 6) and also for displaying results of numerical simulations for a continuous motion [52, 38, 40]. Mappings are simpler also for a theoretical analysis. To obtain the mapping describing a continuous dynamical system it suffices to integrate the equations of motion over a restricted (and usually small) time interval only. In many events this turns out to be much simpler than to analyse the original equations of motion.

Mappings considered in this section seem to be too special, in particular, because of their linearity in momentum. However, we shall see right now that such mappings may describe a real physical system. An interesting example is the problem of a charged particle motion in a "magnetic bottle".

4.3. Motion of a charged particle in a magnetic bottle

In a sufficiently strong magnetic field a charged particle gyrates around and along a magnetic line approximately and simultaneously drifts over a surface of constant magnetic flux.* We shall restrict ourselves to the case of axially-symmetric and constant magnetic field only. Then the drift motion around the symmetry axis is inessential and may be removed, so we arrive at a conservative system having two degrees of freedom.

A charged particle gyrating in a magnetic field possesses an orbital magnetic moment:

$$\mu = v_\perp^2/2B \tag{4.21}$$

where $v_\perp$ is particle velocity component perpendicular to the magnetic vector $\boldsymbol{B}$, and the particle mass is assumed to be unity.** The direction of the magnetic moment vector $\boldsymbol{\mu}$ is opposite to that of vector $\boldsymbol{B}$. Therefore in a non-uniform magnetic field a particle is pushed along a magnetic line into the area of weak field. If the field strength B has a minimum along a magnetic line the charged particle is confined in some space domain.

The Russian term for such a configuration of magnetic field is the magnetic trap, or, in a more

*A detailed description of the charged particle dynamics in a magnetic field may be found, for example, in refs. [48, 56, 118]. We use an abbreviated term "magnetic line" instead of the full expression "magnetic line of force", the more so that the force in the magnetic field is never directed along the line of force.

**All the expressions of this section hold for arbitrary particle velocities if $m/\sqrt{1-v^2/c^2} = 1$ is set.

accurate translation from the Russian, the trap with magnetic stoppers. A Soviet physicist G.I. Budker has invented this term in the early fifties together with the method for plasma confinement. About the same time an American physicist R.F. Post has independently proposed a similar method for precisely the same purpose but, naturally, under a different name – the magnetic mirror machine, or, in a more popular version, the magnetic bottle (apparently with two throats). We observe that the Russian stoppers nicely complement the American bottle to achieve an efficient confinement!

Let us consider particle motion in a magnetic bottle ignoring, first, the curvature of magnetic lines. The nonrelativistic Hamiltonian in terms of cylindrical coordinates z, r, φ has the form:

$$H(p_z, p_r, z, r) = \frac{p_z^2}{2} + \frac{p_r^2}{2} + \frac{(p_\varphi - (e/c)rA_\varphi)^2}{2r^2} \tag{4.22}$$

where the momenta are $p_z = \dot{z}$; $p_r = \dot{r}$, and A_φ is the only component of the vector potential in an axially-symmetric magnetic field; the symmetry axis is parallel to the z-axis. As well as the energy, the angular momentum

$$p_\varphi = r^2\dot{\varphi} + (e/c)rA_\varphi = \text{const} \tag{4.23}$$

is an exact integral of motion, so that the system is reduced to a conservative one having two degrees of freedom. We introduce a new coordinate $x = r - r_C$ where r_C describes the position of the guiding (gyration) center and expand the last part of the Hamiltonian (4.22) up to terms $\sim x^2$:

$$H(p_z, p_x, z, x) \approx \tfrac{1}{2}p_z^2 + \tfrac{1}{2}p_x^2 + \tfrac{1}{2}\omega^2(z) \cdot x^2. \tag{4.24}$$

Here $p_x = \dot{x}$; $\omega = eB(z)/c$ is the Larmor frequency. Owing to dependence $\omega(z)$ a coupling between z- and x-oscillations occurs which leads, generally, to coupling resonances and an energy exchange between both degrees of freedom.

Consider the transverse part of the Hamiltonian (4.24):

$$H_x(p_x, x, z) = \tfrac{1}{2}p_x^2 + \tfrac{1}{2}\omega^2(z)\, x^2 \approx J\,\omega(z). \tag{4.25}$$

For a given longitudinal motion $z(t)$ we have a harmonic oscillator with a time varying frequency. Under a sufficiently slow variation of the frequency $\omega(t)$ the action of the oscillator:

$$J = \tfrac{1}{2}\omega R_\perp^2 = v_\perp^2/2\omega = (c/e)\mu \tag{4.26}$$

is known to be an adiabatic, i.e. approximate, integral of motion. Here $R_\perp = v_\perp/\omega$ is the particle Larmor gyration radius, or gyroradius, which is also the amplitude of transverse oscillations. Then the last expression for the Hamiltonian (4.25) holds approximately, and the full Hamiltonian (4.24) may be written as:

$$H_z(p_z, z) \approx \tfrac{1}{2}p_z^2 + \mu B(z) \tag{4.27}$$

with $\mu \approx$ const to describe the longitudinal motion.

In a magnetic bottle the "potential" energy of longitudinal motion $\mu B(z)$ has a minimum somewhere inside the bottle, so the particle motion is of the kind of "bouncing" (from the "magnetic stoppers" at both "ends" of a magnetic line), or of oscillations. In what follows we shall restrict ourselves to the simplest case of harmonic oscillations:

$$B(z) = B_0(1 + b^2z^2); \qquad z = a\sin(\Omega t); \qquad \Omega = bv_{\perp 0} \tag{4.28}$$

where the subscript zero denotes the value of a quantity in the midplane (the plane of symmetry) of

the magnetic field: $z = 0$. The longitudinal bounce amplitude a is constant and, hence, the motion along a magnetic line is bounded in so far as $\mu = \text{const}$. Any variation of μ results in a change (and, particularly, increase) of amplitude a which leads, in turn, to a leakage of particles from the magnetic bottle. So the problem arises to evaluate or, at least, to estimate the rate of change in μ, or, due to eq. (4.26), in the action J. This is, therefore, a particular case of the general problem of the adiabatic invariant conservation. A vast number of papers has been devoted to the latter problem. As far as the magnetic confinement of charged particles is concerned the problem has been solved, in principle, by Arnold [2] who rigorously proved that for a sufficiently small adiabaticity parameter (4.36) the particle motion in an *axially-symmetric* magnetic bottle is stable in the sense that the particle is confined in the bottle for ever. A problem still not solved completely today is an efficient estimate for the stability conditions. A version of a partial solution for the latter problem is described below.

Let us mention, first of all, a curious peculiarity of the Hamiltonian (4.24) with a "parabolic" magnetic field (4.28): the oscillations for both degrees of freedom are harmonic but not isochronic. Indeed, the bounce frequency Ω depends on μ according to eq. (4.27) and, hence, on the action J. The average Larmor frequency also depends on μ:

$$\langle\omega\rangle = \omega_0(1 + b^2\langle z^2\rangle) = \tfrac{1}{2}\omega_0(1 + v^2/2\mu B_0). \tag{4.29}$$

Here v is the full speed of a particle, and we have used relation $\mu = v_{\perp 0}^2/2B_0 = v^2/2B(a)$. Resonances between bouncing and gyration of a particle depend on the frequency ratio $\langle\omega\rangle/\Omega$ which changes with μ according to above consideration. We have, thus, a typical problem of a nonlinear resonance interaction. In that formulation the problem was considered in ref. [8] using the Hamiltonian (4.24). Later on it was found (see ref. [51]) that the change in μ is mainly determined by just the curvature of the magnetic lines.

Instead of using the exact Hamiltonian (taking account of magnetic line curvature) we are going to start here from the exact expression for the rate of change in μ as reported in ref. [50]:

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} = \frac{\rho v_\perp}{B}\left(v^2 - \frac{v_\perp^2}{2}\right)\sin\Phi - \frac{v_\parallel v_\perp^2}{2B^2}\frac{\partial B}{\partial s}\cos(2\Phi). \tag{4.30}$$

Here s denotes a coordinate along a magnetic line; $v_\parallel = \dot{s}$ is a velocity component; ρ stands for magnetic line curvature and Φ is the perturbation phase the definition of which is clear from fig. 4.3. All quantities are taken at the particle trajectory. We have slightly changed the notation of some quantities in eq. (4.30) as compared to ref. [50].

For a parabolic magnetic field (4.28) in a close vicinity of the symmetry axis ($b^2r^2 \ll 1$) the magnetic line curvature is given by the relation:

$$\rho = -rb^2(1 - 2b^2z^2)/(1 + b^2z^2)^2 \tag{4.31}$$

as is easily verified. We change the perturbation phase Φ for the Larmor phase θ. From fig. 4.3 we get:

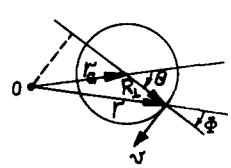

Fig. 4.3. Projection of a gyration trajectory of radius $R_\perp$ with the guiding center at point r_C onto the plane perpendicular to the symmetry axis 0: r, v particle position and velocity; Φ, θ perturbation and Larmor phases, respectively.

$r \sin \Phi = r_C \sin \theta$ assuming that the projection of trajectory onto the plane perpendicular to the symmetry axis is a circle. Effects of distortion of the circle will be discussed below.* We shall see also that while evaluating the change in μ it is sufficient to take only the first harmonic of particle gyration into account, i.e. only the first term in eq. (4.30) which is related just to the magnetic line curvature. The second term corresponding to the simplified Hamiltonian (4.24) gives an exponentially small correction as we shall see. To the same accuracy $\mu = \text{const}$, and all variables in the first term may be taken not at the particle trajectory but at the guiding center r_C. Finally, $(rv_\perp)_C \approx \text{const}$ owing to $v_\perp \sim \sqrt{B}$ ($\mu \approx \text{const}$) and $B_C r_C^2 \approx \text{const}$ (the particle moves over a surface of constant magnetic flux). We get:

$$\frac{d\mu}{dt} \approx -\frac{r_{C0} v_{\perp 0} b^2}{B_0} \frac{(1-2b^2z^2)(v^2-\mu B)}{(1+b^2z^2)^3} \sin \theta. \tag{4.32}$$

Let us apply now Poincaré's surface of section method. To do this we record the state of the system only at moments of crossing the surface (plane) $z = 0$, that is every half-period of particle bouncing. Except for the azimuthal displacement (in φ) this state is completely described by two variables, for example, θ_0 and μ_0. Two other dynamical variables (besides θ_0, μ_0, unknown φ_0 and $z = 0$) of the total number 6 (3 degrees of freedom) may be found from two integrals of motion (energy and p_φ (4.23)). The equations of motion determine a certain correspondence between values μ_0, θ_0 at one crossing the midplane $z = 0$ and the values $\bar{\mu}_0$, $\bar{\theta}_0$ at the next crossing, or a mapping of the plane (μ_0, θ_0) on itself: μ_0, $\theta_0 \to \mu_0$, θ_0. That sort of mapping is customary when one displays the results of numerical experiments [52]. However, for a theoretical analysis the variable μ_0 is not very convenient. As was found out already in the first numerical experiments concerning the motion of a charged particle in a magnetic bottle [52] the main change in μ happens just near the midplane while over the rest of the trajectory $\mu \approx \text{const}$ to the accuracy of small oscillations (see also ref. [50]). Therefore, it is natural to choose just this nearly constant value of μ between successive crossings of the midplane as a dynamical variable. In what follows we shall denote the value of the latter variable simply by μ for the period of motion before crossing the midplane with the phase θ_0 and by $\bar{\mu}$ for that after this crossing, or before the next one with the phase $\bar{\theta}_0$.

We turn to the evaluation of an explicit form for the mapping $\mu, \theta_0 \to \bar{\mu}, \bar{\theta}_0$. We integrate the equations of motion over half the bounce period to obtain the change in θ_0 and μ. The phase change is equal approximately to:

$$\bar{\theta}_0 = \theta_0 + D(\bar{\mu}); \qquad D(\bar{\mu}) \approx \frac{\pi\langle\omega\rangle}{\Omega} = \frac{\pi\omega_0}{2b\sqrt{2B_0}}\left(\frac{1}{\sqrt{\bar{\mu}}} + \frac{v^2}{2B_0\bar{\mu}^{3/2}}\right). \tag{4.33}$$

This change depends on the values of the frequencies between the two crossings, that is on the value of $\bar{\mu}$. The dependence on $\bar{\mu}$ indicates non-isochronicity and, hence, nonlinearity of the oscillations. Note that the change in θ_0 has been evaluated in zero approximation, i.e. for the unperturbed frequencies $\langle\omega\rangle$ and Ω. It is true that the perturbation affects both frequencies, but by a negligibly (exponentially) small amount (see below).

To get the change $\Delta\mu = \bar{\mu} - \mu$ we integrate eq. (4.32) having substituted into its right side the expression $z(t) = a \sin(\Omega t)$ (4.28) deduced for $\mu = \text{const}$. According to the results of numerical experiments we expect the main change in μ to be near $t = 0$. Therefore, we may assume ap-

*There is a trivial distortion due to the inclination of vector $\boldsymbol{B}$ to the symmetry axis. In what follows we assume this inclination angle to be small.

proximately $z(t) \approx a\Omega t$ and extend the integration interval to infinity:

$$\Delta\mu \approx -\frac{r_{C0}}{aB_0} \int_{-\infty}^{\infty} \mathrm{d}\tau \sin\theta \frac{1-2\tau^2}{(1+\tau^2)^3}(v^2 - \mu B). \tag{4.34}$$

We shall evaluate the integral in the complex plane of a new variable $\tau = ab\Omega t$. The integrand has two poles at the points $\tau_p = \pm \mathrm{i}$. Note that at the pole $B = B_0(1+\tau^2) = 0$. Under the assumed approximation $\mu = \text{const}$ the term μB may be removed, therefore, from the integral. We substitute $\sin\theta = \mathrm{Im}(\mathrm{e}^{\mathrm{i}\theta})$ and close the integration contour in the upper τ halfplane. The value of $\Delta\mu$ is mainly determined by the exponential factor $\exp(\mathrm{i}\theta(\tau_p))$. Let us get it. Since $\omega \approx \omega_0(1+\tau^2)$ we have:

$$\theta = \theta_0 + \int_0^t \omega \, \mathrm{d}t \approx \theta_0 + \frac{1}{\epsilon}\left(\tau + \frac{\tau^3}{3}\right) \tag{4.35}$$

where

$$\epsilon = ab\,(\Omega/\omega_0) = (bR_{\perp 0})(v_\parallel/v_\perp)_0 \tag{4.36}$$

is the adiabaticity parameter. This is just the quantity determining the conditions for and the accuracy of μ conservation since the argument of the exponential factor $\mathrm{i}\theta(\tau_p) = \mathrm{i}\theta(\mathrm{i}) \approx -2/3\epsilon$. For $\epsilon \ll 1$ the change in μ is exponentially small. Hence it is clear that higher gyration harmonics $m > 1$ would give additional terms $\sim \exp(-2m/3\epsilon)$ which can be neglected as has been done above. Expression (4.36) shows that the adiabaticity parameter is of the order of the ratio of the bounce frequency to the gyration frequency, or of the ratio of the Larmor radius to the scale of the magnetic field variation in space ($\sim 1/b$). The parameter (4.36) was derived in ref. [50].* Applicability of the parameter is limited by the expansion $\sin(\Omega t) \approx \Omega t = \tau/ab \ll 1$, which we have used above. For $\tau = \tau_p = \mathrm{i}$ we arrive at the condition:

$$\lambda \equiv ab = (v_\parallel/v_\perp)_0 \gg 1 \tag{4.37}$$

which is valid only for particles moving at a small angle to the magnetic line. In ref. [53] a generalization of the adiabaticity parameter to arbitrary values of λ has been worked out (see below).

It is precisely the condition (4.37) which provides a sharp change in μ near the midplane since the pole is, under this condition, very close to the real t-axis: $|\Omega t_p| = 1/\lambda \ll 1$. The value of $|t_p|$ gives the order of the time interval during which the change in μ occurs.

Let us come back to the evaluation of the integral (4.34). Since the pole is a multiple one we need to differentiate the integrand to get the residue. For $\epsilon \ll 1$ it suffices to differentiate the function $\exp(\mathrm{i}\theta)$ only since a large factor $1/\epsilon$ appears (see eq. (4.35)). For the same reason it is enough to take account only of that part of the integrand in eq. (4.34) which has a pole of the highest (third) multiplicity. We *get***;

$$\Delta\mu = \kappa(\mu)\sin\theta_0; \qquad \kappa(\mu) \approx -\frac{3\pi}{4}\frac{r_{C0}v^2}{\epsilon a B_0}\mathrm{e}^{-2/3\epsilon}. \tag{4.38}$$

Now we can, at last, write down the mapping:

*In an implicit form this parameter is contained also in the more general expressions of ref. [51].

**A more detailed evaluation of a similar integral is carried out in the appendix.

$$\bar{\mu} = \mu + \kappa(\mu) \sin \theta_0; \qquad \bar{\theta}_0 = \theta_0 + D(\bar{\mu}) \tag{4.39}$$

where $D(\mu)$ is given by eq. (4.33). The mapping (4.39) is not a canonical one as a result of neglecting the influence of perturbation on the change in frequencies and, hence, in the phase θ_0 (see above). To obtain a canonical mapping we could take account of this influence. Instead, we shall do it in another, simpler way. Using the smallness of $\Delta\mu$ we linearize the mapping (4.39) in μ. Choose a resonant value of $\mu = \mu_r$, such that $D(\mu_r) = 2\pi n$ (n is an integer), and expand

$$D(\mu) \approx D(\mu_r) + D'_\mu(\mu_r)(\mu - \mu_r) = 2\pi n + I \tag{4.40}$$

where D'_μ denotes the derivative in respect to μ, and we have introduced a new variable (momentum) I. Dropping the constant $2\pi n$ we may write the linearized mapping in the form:

$$\begin{aligned} &\bar{I} = I + K \sin \theta_0; \qquad \bar{\theta}_0 = \theta_0 + \bar{I} \\ &K = \kappa(\mu_r)\, D'_\mu(\mu_r) = \frac{3\pi^2}{8} \frac{r_C}{bR_\perp^2} \left(\frac{v}{v_\parallel}\right)^2 \left(1 + \frac{3v^2}{v_\perp^2}\right) e^{-2/3\epsilon} \end{aligned} \tag{4.41}$$

where all quantities are taken at $z = 0$ and for $\mu = \mu_r$. This mapping is just of the type introduced in section 4.2. We shall consider this mapping in detail in sections 5 and 6.

In the framework of the assumed approximation the expression for the mapping parameter K (4.41) resolves the problem of the stability of particle motion in a magnetic bottle.

Note that the quantity $K \sim \exp(-1/\epsilon)$ (4.41) determining the stability of the motion depends exponentially on the inverse of the small perturbation parameter $\epsilon \to 0$ and, hence, cannot be represented by any power series in ϵ. The latter gives only small quasi-periodic corrections to the original expression (4.21) for the particle magnetic moment, corrections which never give rise to any instability, that is, to a large change in μ [119]. Success of the approach described in this section should be attributed to a different method for the construction of successive approximations, the method based on a different kind of the small perturbation parameter $K \sim \exp(-1/\epsilon)$. This special form of perturbation parameter is a consequence of taking account of very weak (and often ignored) resonances between longitudinal bounce oscillations and the Larmor gyration of a particle in a magnetic bottle (see ref. [8]). For a further discussion of this kind of approximation see section 4.4.

The condition for stability of the motion is: $K < 1$ (section 5.1). For $K > 1$ the stochastic instability results in a diffusion of particles over an integral surface, i.e. over a subsurface of the intersection of surfaces $E = \text{const}$ and $p_\varphi = \text{const}$ (4.23). If a particle orbit encircles the field symmetry axis ($r = 0$) the integral surface gets closed, and the particle motion is bounded in spite of stochastic instability. This is the so-called absolute confinement due to exact integrals of motion. For particles whose orbits do not encircle the symmetry axis there is no absolute confinement, they may be kept inside a magnetic bottle only under conditions providing a sufficiently small variation of the magnetic moment μ.

Integral surfaces in an axially-symmetric field of a magnetic dipole were studied in detail by Störmer as early as 1907 in connection with the analysis of cosmic ray motion in the Earth's magnetic field (see ref. [57]). Similar calculations for a magnetic bottle are described, e.g., in ref. [56]. The condition of absolute confinement may be written in the form (see ref. [55]):

$$r_C < R_S; \qquad R_S^2 = R_\perp^2 - R^2/k. \tag{4.42}$$

Here $R = v/\omega_0$ stands for the so-called "full" Larmor radius and $k = B_{max}/B_0$ denotes the mirror ratio

which characterizes the depth of the potential well (4.27) in a magnetic bottle. The quantity R_S may be called the *Störmer radius*. The domain of phase space in which a particle is confined due to exact integrals of motion is usually called the *Störmer zone*.

Expression (4.41) for the stability parameter K is valid, particularly, under the condition $\lambda \gg 1$. For arbitrary pitch angles β of a particle velocity to the magnetic line we need to integrate eq. (4.32) with $z(t) = a\sin(\Omega t)$. The poles t_p are determined by the equation $\omega(t) = 0$ whence:

$$\sin(\Omega t_p) = \pm i/\lambda; \qquad \tau_p = \Omega t_p = \pm i \operatorname{arsinh}(1/\lambda). \tag{4.43}$$

The argument of the exponential factor becomes:

$$i\theta(t_p) = -\frac{\omega_0}{2\Omega}\left[(2+\lambda^2)\operatorname{arsinh}\left(\frac{1}{\lambda}\right) - \sqrt{1+\lambda^2}\right] \equiv -\frac{1}{\epsilon_K} \tag{4.44}$$

where ϵ_K is a new adiabaticity parameter deduced by Krushkal [53].* For $\lambda \gg 1$ ϵ_K is reduced to the old parameter: $\epsilon_K \to \frac{3}{2}\epsilon$.

One cannot integrate eq. (4.32) in infinite limits as eq. (4.34) since there are two infinite rows of poles (4.43) now. Instead, we integrate over half a period of bounce oscillations and then close the contour in the upper half-plane along two vertical lines $\mathrm{Re}(\tau) = \pm\pi/2$ as shown in fig. 4.4. Values of the integral over each of the two straight lines are, generally, not equal owing to a difference in phase θ. Yet, at a resonance ($\mu = \mu_r$; $\langle\omega\rangle = 2n\Omega$, see eq. (4.33)) this difference is equal to $2\pi n$, and the sum of the integrals over the two straight lines vanishes. The rest of the integration procedure is similar to that for $\lambda \gg 1$ as described above. Finally we get:

$$\Delta\mu \approx -\frac{3\pi}{4}\frac{r_{C0}\omega_0 v_{\perp 0}}{B_0} e^{-1/\epsilon_K} \sin\theta_0 \tag{4.45}$$

that coincides exactly, for the assumed shape of magnetic field, with a more general result by Krushkal [53] obtained with the saddle-point method of integration.** Expression (4.45) differs from (4.38) by the factor $(v_\parallel/v)^2$ and also, of course, by another adiabaticity parameter. Factor K of the mapping (4.41) linearized in μ becomes:

$$K = \frac{3\pi^2}{8}\frac{r_C}{bR_\perp^2}\left(1 + \frac{3v^2}{v_\perp^2}\right) e^{-1/\epsilon_K}. \tag{4.46}$$

One may still improve the accuracy of the expression for K evaluating the integral for $\Delta\mu$ "exactly", i.e. taking account of all the terms and not only of the leading one (in ϵ) as has been done

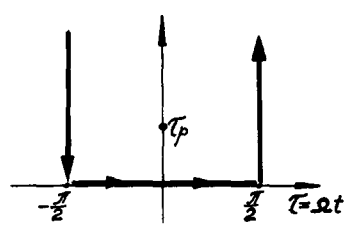

Fig. 4.4. Integration contour for eq. (4.32), arbitrary λ.

*The same expression has been derived recently in ref. [137].

**The last relation is somewhat different as compared with a more accurate one derived in ref. [137]. Nevertheless, a simpler eq. (4.45) does fit quite well the numerical data in ref. [137].

above. We write the word "exactly" in quotation-marks since that "exact" value of integral gives all the same an approximate value of $\Delta\mu$ because we make use of an approximate relation $z(t) \approx a\sin(\Omega t)$. The latter takes no account of higher gyration harmonics. It is true that the higher harmonics lead by themselves to exponentially small corrections. However, in the second and higher approximations in ϵ they may contribute to the first harmonic again and, hence, may change eq. (4.32) by an additional term $\sim\epsilon^2$. Therefore, we can evaluate only the correction $\sim\epsilon$ without taking account of exponentially small terms. Having performed somewhat cumbersome but elementary calculations we arrive at an improved parameter K_ϵ:

$$K_\epsilon = K(1+\delta); \quad \delta = \frac{bRv_\perp}{6v}\frac{10+8\lambda^2+\lambda^4}{(1+\lambda^2)^{3/2}} \tag{4.47}$$

where K is given by eq. (4.46).

Let us compare the estimates obtained with results of numerical computations in ref. [55].* In the latter work a large number of muon trajectories were computed to find out the optimal conditions for muon confinement in a magnetic bottle with the parabolic field (4.28). Parameters of 6 unstable trajectories, i.e. those leaving the bottle after some time, near the stability border are given in table 4.1.

Prior to the comparison some corrections need to be introduced. Namely, it would be more accurate not to take the quantities b and ω_0 (or R) just at the symmetry axis where they are immediately given but to average them over a Larmor circle. For parabolic field (4.28) the corrections have the form:

$$R_\perp \to R_\perp(1+\tfrac{1}{2}b^2\langle r^2\rangle); \qquad b \to b\,(1+\tfrac{1}{2}b^2\langle r^2\rangle) \tag{4.48}$$

where $\langle r^2\rangle = r_C^2 + R_\perp^2$ is the mean square of a particle distance to the symmetry axis (see fig. 4.3) and $R_\perp$, b on the right-hand sides of the equations are taken at the axis.

The quantities $R_\perp$ and r_C in table 4.1 have been calculated for a given energy and the pitch angle β of a particle, and for magnetic field strength $B_0 = 42.2$ kG at the center of the bottle and parameter $b = 7.71\times10^{-3}\,\mathrm{cm}^{-1}$ at the bottle axis. Every particle started its motion in the midplane $z=0$ at a radius $r_0 = R_\perp + r_C = 60$ cm and in a direction normal to vector r_0. Krushkal's adiabaticity parameter has been calculated according to relation (4.44) and also with the corrections (4.48). Then the stability

Table 4.1
The border of particle motion stability in a magnetic bottle

E MeV	$R_\perp$ cm	r_C cm	R_S cm	β	ϵ_K	N	K_0	K	$\frac{K}{K_0}$	δ	K_ϵ	$\frac{K_\epsilon}{K_0}$	K_1
450	40.7	19.3	16	63°	0.314	370	1.60	1.20	0.75	0.45	1.74	1.09	1.15
400	36.6	23.4	15	64°	0.271	340	1.62	1.05	0.65	0.41	1.48	0.91	0.87
350	32.5	27.5	14	64°	0.239	310	1.64	1.01	0.62	0.36	1.38	0.84	0.74
300	28.1	31.9	11	62°	0.219	270	1.68	1.14	0.68	0.31	1.49	0.89	0.82
250	24.1	35.9	7	59°	0.208	220	1.73	1.40	0.81	0.25	1.75	1.01	1.02
200	20.2	39.8	4	56°	0.192	180	1.79	1.50	0.84	0.20	1.81	1.01	1.02
	Mean values								0.72			0.96	0.94

*For another example of that comparison see ref. [138].

parameter K was evaluated, using eq. (4.46). It can be seen from the table that $K > 1$ for all trajectories. This may be related to the fact that the stability border has been determined computationally for a finite time of a particle confinement in the bottle. Namely, a trajectory was considered to be stable if the confinement time exceeded half the muon life time at a given energy. The quantity N in the table gives the number of crossings of the midplane during this time interval. As will be shown in section 5.1 the K values in this case are somewhat greater than for $N = \infty$ varying with N according to the approximate relation: $K_0 \approx 1 + (100/N)^{0.4}$. The values of K_0 are also given in the table. It is seen now that eq. (4.46) underestimates K (mean value of the ratio is $\langle K/K_0 \rangle = 0.72$). The agreement becomes much better if one takes account of the correction δ (4.47): $\langle K_\epsilon/K_0 \rangle = 0.96$. The expected accuracy of the estimate (4.47) is $\sim \epsilon_K^2 \sim 0.1$ as discussed above. Note that exponential factor $\exp(-1/\epsilon_K)$ is varying over the data of table 4.1 from 1/183 to 1/24, that is by more than a factor 7. The biggest value of the exponential factor is still small enough to allow the neglect of higher harmonic terms as discussed above. It is interesting to note also that analytical estimates are applicable not only to trajectories which don't encircle the field symmetry axis ($r_C > R_\perp$, the last 3 events in table 4.1) but also to those encircling the axis ($r_C < R_\perp$, the first 3 events) provided, of course, that the latter trajectories do not get into a Störmer zone of absolute confinement (4.42) (cf. values of r_C and R_S in table 4.1).

The numerical data in table 4.1 may be processed another way. Namely, for a given K_ϵ one may find the border of the so-called eternal stability K_1, that is of indefinitely long confinement of a particle in a magnetic bottle, according to the relation (see section 5.1):

$$K_1 \approx K_\epsilon - (100/N)^{0.4}. \tag{4.49}$$

The values of K_1 are given in the last column of table 4.1. The mean value $\langle K_1 \rangle = 0.94$ is in a good agreement with the value $K_1 = 1$ obtained directly by the mapping (4.41) in section 5.1.

Recently an interest in the problem of a single particle motion in a magnetic bottle has revived again in connection with controlled nuclear fusion research. Stability of a single particle motion is suspected now to be the basic limitation for some thermonuclear devices even in the case of a dense plasma (see, e.g., ref. [87]). Unfortunately, magnetic fields in plasma devices are, as a rule, neither parabolic nor axially-symmetric, and still more so taking account of the self-fields in a dense plasma. This poses a problem of particle motion stability in a magnetic field of much more general shape as compared to the simplest case considered in this section.

4.4. Motion in a vicinity of the separatrix

A more important occasion to apply a mapping of the type (4.11) is the motion near the separatrix of a nonlinear resonance (sections 3.2 and 3.3). This problem will be considered in detail in section 6. Here we will show how to construct a mapping in such a case.

Let us consider the pendulum, as a model of nonlinear resonance, under a periodic parametric perturbation described by the Hamiltonian (cf. section 3.1):

$$\begin{aligned} &H(p, \varphi, \tau) = H_0(p, \varphi) + \epsilon V(\varphi, \tau), \qquad H_0 = \tfrac{1}{2}p^2 - \omega_0^2 \cos\varphi \\ &\epsilon V = \epsilon\omega_0^2 \cos\varphi \cos\tau = \tfrac{1}{2}\epsilon\omega_0^2[\cos(\varphi - \tau) + \cos(\varphi + \tau)]. \end{aligned} \tag{4.50}$$

Here $\epsilon \ll 1$ is a small perturbation parameter; $\tau = \Omega t + \tau_0$; Ω, τ_0 are perturbation frequency and initial phase, respectively. For small oscillations ($|\varphi| \ll 1$) and the frequency ratio $\Omega/\omega_0 = 2$ a well known

parametric resonance occurs (section 3.1). Far off the resonance the perturbation results in only a small energy modulation which can be neglected in first approximation. Even for $\epsilon \sim 1$ the modulation is small if the perturbation frequency is high enough ($\Omega \gg \omega_0$). As we know, the basic idea of the averaging method (section 2.2) is just to neglect those rapidly oscillating (non-resonant) perturbation terms. We shall see right now that near the separatrix the averaging method does not work any more.

Let us construct a mapping describing the motion of system (4.50) near the separatrix. We may do it in a way similar to that in section 4.3. We need to calculate the change in the pendulum energy (ΔH) over half a period of rotation of the pendulum. This may be done via integration of the equation:

$$\mathrm{d}H/\mathrm{d}t = \partial H/\partial t = \epsilon\, \partial V/\partial t \tag{4.51}$$

over a period of motion. Near the separatrix one can put approximately: $\varphi(t) \approx \varphi_{sx}(t)$ (2.30), that is substitute the motion law *at* the separatrix for that *near* the separatrix, and then extend the integration time interval to infinity. This is a basic approximation for the problem under consideration which we shall apply repeatedly in what follows. Keeping only one perturbation term ($\cos(\varphi - \tau)$) so far we get from eq. (4.51):

$$\begin{aligned}\Delta H &\approx \int_{-\infty}^{\infty} \mathrm{d}t\, \frac{\epsilon\omega_0^2\Omega}{2} \sin(\varphi_{sx} - \tau) \\ &= -\tfrac{1}{2}\epsilon\omega_0\Omega\, A_2(\Omega/\omega_0) \sin\tau_0 \approx -4\pi\epsilon\Omega^2 \exp(-\pi\Omega/2\omega_0) \sin\tau_0.\end{aligned} \tag{4.52}$$

Here $A_m(\lambda)$ is the Melnikov–Arnold integral (the MA integral) evaluated in the appendix and the last expression gives its approximate value for $\Omega/\omega_0 \gg 1$.

This kind of integral describing the perturbation of the separatrix was used, apparently, first by Melnikov [12] and also by Arnold [5]. A qualitative study of the separatrix perturbation was carried out already by Poincaré [6] (see section 6.1 for detail).

Instead of calculating the change in the full Hamiltonian (4.50) one may find that in the unperturbed energy H_0. We have:

$$\mathrm{d}H_0/\mathrm{d}t = [H, H_0] = -\epsilon p\, \partial V/\partial\varphi \tag{4.53}$$

where [,] is the Poisson bracket. The last expression has a simple physical meaning of a driving force power since the speed $\dot\varphi = p$. Substituting approximately: $p(t) \approx p_{sx}(t) = 2\omega_0 \cos(\varphi/2)$ (2.29) and integrating in infinite limits we arrive at:

$$\begin{aligned}\Delta H_0 &\approx \int_{-\infty}^{\infty} \mathrm{d}t\, \epsilon\omega_0^3 \sin(\varphi_{sx} - \tau) \cos(\tfrac{1}{2}\varphi_{sx}) \\ &= -\tfrac{1}{2}\epsilon\omega_0^2 (A_1 + A_3) \sin\tau_0 = -\tfrac{1}{2}\epsilon\omega_0\Omega\, A_2 \sin\tau_0\end{aligned} \tag{4.54}$$

that is exactly at eq. (4.52). The last expression of (4.54) has been deduced using the recurrence relation (A.8) for the MA integral.

The equality $\Delta H = \Delta H_0$ is related to the special definition of the MA integral (see appendix). Namely, we ignore the oscillating part of the integral, that is the part periodic in φ and τ. This part is of minor importance for us since its oscillations are bounded and have nothing to do with any instability of motion. That oscillating part can be transformed away, or "killed", by a canonical change of variables either (section 2.2). In the particular case under consideration the difference

$H - H_0 = \epsilon V(\varphi, \tau)$ is just such a periodical function in both φ and τ (4.50) whose variation is bounded and small due to $\epsilon \ll 1$. So we simply ignore this oscillating part and concentrate our efforts on the calculation of aperiodic changes (ΔH) in the Hamiltonian due to the perturbation, the changes which may cumulate over many pendulum oscillations and, hence, give rise to instability.

We have assumed above, while evaluating integrals (4.52) and (4.54), that $\dot{\varphi}_{sx} > 0$. For $\dot{\varphi}_{sx} < 0$ the perturbation term $\cos(\varphi - \tau)$ used gives a negligible contribution as soon as $\Omega \gg \omega_0$ (see appendix). However, the second perturbation term $\cos(\varphi + \tau)$ contributes exactly as much as the first term for $\dot{\varphi}_{sx} > 0$ as is easily verified. Thus, in the case of the symmetric perturbation (4.50) the energy change is given by eq. (4.52), or (4.54), for every half-period of oscillations and for either direction of rotation provided $\Omega \gg \omega_0$, the condition we assume to be valid henceforth.

The energy change (4.54) depends on the perturbation phase value τ_0 at the surface $\varphi = 0$. The change in τ_0 itself is determined approximately by the frequency of pendulum oscillations $\omega(w)$ (section 2.4):

$$\bar{\tau}_0 \approx \tau_0 + \pi\Omega/\omega(\bar{w}) \approx \tau_0 + \lambda \ln(32/|\bar{w}|) \tag{4.55}$$

where $\lambda = \Omega/\omega_0$ and $\bar{w} = (\bar{H}_0/\omega_0^2) - 1$ is the relative oscillation energy past the crossing surface $\varphi = 0$ with phase τ_0 (cf. section 4.3). Note that for $\lambda \gg 1$ the change in w occurs over a relatively short time interval $\sim 1/\Omega \ll 1/\omega_0 < 1/\omega$ about the surface $\varphi = 0$.

Combining eqs. (4.54) and (4.55) we arrive at the mapping describing the pendulum oscillations near the separatrix under a periodic parametric perturbation:

$$\bar{w} = w + W \sin \tau_0; \qquad \bar{\tau}_0 = \tau_0 + \lambda \ln(32/|\bar{w}|) \tag{4.56}$$

where the amplitude of the change in w:

$$W = -4\pi\epsilon\lambda^2 \mathrm{e}^{-\pi\lambda/2}. \tag{4.57}$$

Note that the mapping (4.56) is canonical.

For $\lambda \gg 1$ the change of w is small, and we may linearize the mapping (4.56) in w (but not in τ_0!) about a resonant value w_r to get a new mapping:

$$\bar{I} = I + K \sin \theta; \qquad \bar{\theta} = \theta + \bar{I} \tag{4.58}$$

where the new momentum:

$$I = -\frac{\lambda}{w_r}(w - w_r) \tag{4.59}$$

and the parameter

$$K = -\frac{\lambda W}{w_r} = \frac{4\pi\epsilon\lambda^3}{w_r} \mathrm{e}^{-\pi\lambda/2}. \tag{4.60}$$

We have changed the phase notation ($\tau_0 \to \theta$) to emphasize identity of the latter mapping (4.58) with the mapping (4.11) in section 4.2 ($f(\theta) = \sin \theta$). In particular, both mappings are equivalent to a continuous system represented by the Hamiltonian (4.18). Note that the mapping (4.41) in section 4.3 describing approximately the motion of a charged particle in a magnetic bottle is also the same. We see that a number of dynamical problems (surely not all though) can be reduced approximately to the mapping (4.58), so the latter merits to be studied in detail as will be done in sections 5 and 6. We shall call eq. (4.58) the *standard mapping*.

The mapping (4.56) is also of importance since it describes the motion near the separatrix which is the ultimate origin of stochastic instability as we shall see in section 6. This is clear also from eqs. (4.58) and (4.60) since no matter how small the perturbation, i.e. ϵ and $1/\lambda$, there exists a domain (a layer) sufficiently close to the unperturbed separatrix ($w_r \to 0$) where $K > 1$ and, hence, the motion is unstable (sections 4.2 and 5.1). Graphically speaking, the mapping (4.56) describes a perturbation of the Arnold whiskers (section 2.4), so we have some reason to call eq. (4.56) the *whisker mapping*.

Resonant values w_r of the whisker mapping (4.56) are determined by the condition: $\lambda \ln(32/|w_r|) = 2\pi n$ with integer n, whence:

$$|w_r| = 32\,\mathrm{e}^{-2\pi n/\lambda}. \tag{4.61}$$

This set of resonances corresponds approximately to the resonances $I_r = 2\pi m$ of the standard mapping (4.58) as one can see from eqs. (4.59) and (4.61).

It is worth noting that the standard mapping can be immediately applied to describe its own motion near a resonance separatrix (with different values of the parameters, of course). Indeed, the standard mapping is equivalent to the continuous system (4.18) but the latter differs from the system (4.50) of this section only by high frequency terms which make a negligible contribution to the energy change (see eq. (4.52)). Therefore, the standard mapping (4.58) describes not only the system (4.50) for which it has been deduced above but also the system (4.18) with $-\omega_0^2 = k$; $\Omega = 1$; $\lambda = 1/\sqrt{k} = 2\pi/\sqrt{K}$; $\epsilon = 2$, and, hence, the standard mapping itself (4.11). It hardly needs to be mentioned that the physical meanings of the two standard mappings in this chain ((4.11) and (4.58)) (and still more their parameters) are completely different. The original standard mapping (4.11) describes main resonances between the *original oscillations* and an external perturbation whereas the second standard mapping (4.58) does so for resonances between the same perturbation and *phase oscillations* at a main resonance. These resonances may be called *second-level resonances* as distinct from the main, or original, resonances of the first level. Resonances between the perturbation and phase oscillations at a second-level resonance are then third-level resonances (and are described also by a standard mapping) and so on. An infinite recursive hierarchy of different-level resonances appears which brings about a very intricate structure of the motion (see ref. [71] and section 5.5). Nevertheless, and this is also a striking feature of the motion, the resonance behavior can be described recursively at every level by the same standard mapping. This remains so even if the original system is a continuous one like that considered in this section (4.50).

Taking account of higher (or, perhaps, better to say, deeper) level resonances, at least, those of the second level, is of a great importance for the problem of motion stability. Suppose the Hamiltonian (4.50) describes a nonlinear resonance in the pendulum approximation (section 3.2). Then the frequency of small phase oscillations (of the first level) $\omega_0 \sim \sqrt{\mu}$, μ being another small perturbation parameter (besides ϵ). The parameter (4.60) of the second level mapping (4.58) is then $K \sim \exp(-C/\sqrt{\mu})$, C being a constant. It is customary to say now that effects of such a (exponentially) small order are unreachable for any asymptotic perturbation theory. This term means that one is not concerned about either the convergence of the appropriate series or about its remainder term being satisfied by the construction of a formal series in a small perturbation parameter (see the beginning of section 4). It is true that a function like $\exp(-C/\sqrt{\mu})$ cannot be expanded as a power series in μ. But it would be misleading to conclude that we have got the relation (4.60) by means of some powerful non-asymptotic theory. Far from it! All we have done was the consistent consideration of any kind of resonances encountered including those of deeper levels as was explained above. We have used nothing more than the simple averaging method which is by the way a common manner to construct an asymptotic series. So the essential point is not the type of a perturbation series (asymptotic or

convergent) but rather the nature of the perturbation parameter to expand in. Particularly, for the second-level resonances the small perturbation parameter (4.57) $W \sim \exp(-C/\sqrt{\mu})$ is a non-analytic function of μ (at $\mu = 0$) from the beginning. Similar problems have been solved earlier by Melnikov [12] and by Arnold [5] precisely in this way.

The moral of the above reasoning is that the consistent resonance approach combined with even the simple averaging method permits us to resolve problems which are widely believed to be beyond such an approach. This approach may, therefore, be called the *resonant perturbation theory* as a particular case of asymptotic perturbation theory.

4.5. Many-dimensional oscillations

In this section we will give some very rough estimates for the condition of stochastic instability in a many-dimensional oscillator system with the main purpose to compare the overlap criterion to the KAM theory (section 4.6). Characteristics of a single resonance of many-dimensional oscillations are given in section 3.3. We shall consider first, and in more detail, the case of an autonomous, or closed, system, i.e. that of coupling resonances: $m, \omega = 0$. For the reader's convenience we again write down the Hamiltonian (3.24) ($n = 0$):

$$H(I, \theta) = H_0(I) + \epsilon \sum_m V_m \, e^{im.\theta} \tag{4.62}$$

and recall that quantities I, θ, m, ω are N-dimensional vectors. We assume, further, for the sake of simplicity that $H_0(I) = |I|^2/2$, then $\omega = I$, and the nonlinearity matrix $\partial\omega/\partial I$ becomes unity. The condition for stochastic instability depends essentially on the behavior of the Fourier amplitudes for large $|m|$. We shall consider two cases.

In the first we set:

$$V_m \sim V_1 \, e^{-\sigma S}; \qquad S = \sum_{k=1}^{N} |m_k| \tag{4.63}$$

where V_1 gives the order of perturbation lower harmonics. This case corresponds to a perturbation $V(\theta)$ which is analytical as function of phases (see, e.g., [2]).* We will be interested below in small values of the parameter σ when the perturbation possesses many harmonics.

The second kind of perturbation called *smooth*, or differentiable, corresponds to the estimate:

$$V_m \sim V_1/S^{l+2} \tag{4.64}$$

parameter l characterizing the *smoothness* of function $V(\theta)$. For an integer l it is the highest order of still continuous partial derivatives with respect to the components of the vector θ (the dependence on I is assumed to be analytic). One says in this case that the function $V(\theta)$ belongs to the class C^l. The relation (4.64) is readily understood for a function of one variable ($N = 1$). According to (4.64) the $(l+2)$nd derivative has Fourier amplitudes independent of harmonic number $S = m_1$, and is, hence, a kind of δ-function or a sum of such functions. Therefore, the $(l+1)$st derivative has a jump discontinuity but the lth and all the preceding derivatives are already continuous. For a function of many variables the estimate (4.64) implies that Fourier amplitudes of the $(l+2)$nd derivative

*It is, of course, a highly specific case; a little more general one would correspond to $\sigma S \to \sigma_k m_k$.

$$\frac{\partial^{l+2}}{\partial\theta_1^{n_1}\ldots\partial\theta_N^{n_N}}V(\theta); \qquad n_1+\cdots+n_N=l+2$$

depend only on ratios of harmonic numbers and not on their moduli.

Under the above assumptions concerning H_0 the half-width of a single resonance m, $\omega=0$ is given by the relation (see eq. (3.22)):

$$(\Delta\omega)_r = 2\sqrt{\epsilon V_m}. \tag{4.65}$$

The volume of a resonance, or, better to say, of a *resonance layer*, in ω-space is of the order:

$$v_r \sim (\Delta\omega)_r\,|\omega|^{N-1} \sim \sqrt{\epsilon}\,V_m^{1/2}\,|\omega|^{N-1}.$$

Consider first a smooth perturbation (4.64). For a given S there are $n_S \sim S^{N-1}$ different resonance layers of total volume $v_S \sim v_r\,n_S$. The total volume of ω-space occupied by all the resonance layers is $v \sim \Sigma_S\,v_S$, and the filling factor, or the relative volume:

$$\mu \sim \frac{v}{|\omega|^N} \sim \sqrt{\epsilon}\frac{V_1^{1/2}}{|\omega|}\sum_{S=1}^{\infty}\frac{1}{S^{l/2+2-N}}. \tag{4.66}$$

When $\mu \gg 1$, all of ω-space is filled up with resonances, and the stochastic instability ensures; if $\mu \ll 1$ instability occurs in a neighbourhood of resonance layer intersections only whereas for most initial conditions the motion is stable.

The sum in eq. (4.66) converges under the condition:

$$l > l_c = 2N-2 \tag{4.67}$$

where l_c is the critical value of perturbation smoothness. If $l \leqslant l_c$ resonances are overlapping for an arbitrarily small perturbation $\epsilon \to 0$, that is the motion is always unstable.

In the case of an analytical perturbation (4.63) we get in a similar way:

$$\mu \sim \sqrt{\epsilon}\frac{V_1^{1/2}}{|\omega|}\sum_{S=1}^{\infty}S^{N-1}\,\mathrm{e}^{-\sigma S/2} \sim \sqrt{\epsilon}\frac{V_1^{1/2}}{|\omega|}\left(\frac{N}{\sigma}\right)^N. \tag{4.68}$$

The last estimate is valid for $N \gg 1$ and $\sigma \ll 1$. For analytical perturbations there always exists a critical perturbation strength ϵ_c below which the resonances do not overlap. The value ϵ_c corresponds to $\mu \sim 1$, whence:

$$\epsilon_c \sim \frac{H_0}{V_1}\left(\frac{\sigma}{N}\right)^{2N} \tag{4.69}$$

where we have used the relation $H_0 = |I|^2/2 = |\omega|^2/2$ assumed above.

It remains now to estimate the ratio H_0/V_1. For this we normalize the small parameter ϵ in such a way that:

$$H_0^2 = \langle V^2\rangle = \sum_m |V_m|^2 \tag{4.70}$$

where averaging over phases θ is understood. For the spectrum (4.63):

$$\langle V^2\rangle \sim \sum_{S=1}^{\infty}V_1^2\,S^{N-1}\,\mathrm{e}^{-2\sigma S} \sim V_1^2\left(\frac{N}{\sigma}\right)^N$$

whence: $H_0/V_1 \sim (N/\sigma)^{N/2}$ and:

$$\epsilon_c \sim (\sigma/N)^{3N/2}. \tag{4.71}$$

It is the final estimate for the critical perturbation strength as a function of the analyticity parameter σ and the number of degrees of freedom N. One should bear in mind that the estimate is very rough: not only a numerical factor, which may happen to be fairly large, has been dropped from the estimate but also no account of a non-uniformity of resonance distribution over ω-space has been taken. The latter effect leads to some increase in ϵ_c.

Let us consider in conclusion a particular case of time-dependent Hamiltonian corresponding to a mapping (section 4.2). It implies that all Fourier harmonics of the "external" frequency Ω are of equal amplitudes. The only difference from the above estimates is that the number of resonance layers $m, \omega + n\Omega = 0$ for a given S is now $n_S \sim S^N$ rather than S^{N-1}, so the critical smoothness becomes:

$$l_c = 2N. \tag{4.72}$$

In particular, for $N = 1$ (the mapping (4.11) in section 4.2, for example) $l_c = 2$ as was obtained in ref. [35].

4.6. The Kolmogorov–Arnold–Moser theory (KAM theory)

Rigorous inequalities determining the position of the stability border for a system with the type (4.62) Hamiltonian have been obtained in the KAM theory. As we mentioned already, the problem of nonlinear oscillation stability has been successfully solved thanks to two fundamental ideas by Kolmogorov [1]. We have spoken a lot about the first of them in section 2.2 – a new perturbation theory which ensures the superconvergence of successive approximations. The second idea concerns the very formulation of the problem. To understand it let us introduce a change of variables "killing" the perturbation $\sim\epsilon$. Choosing a generating function as ($m \neq 0$):

$$F(I^{(1)}, \theta) = \theta, I^{(1)} + \epsilon\Phi_m \, e^{im,\theta} \tag{4.73}$$

and substituting $I = I^{(1)} + i\epsilon m\Phi_m \, e^{im,\theta}$ into eq. (4.62) we get:

$$\Phi_m = \frac{iV_m}{m, \omega(I^{(1)})}. \tag{4.74}$$

A distinct feature of this relation is the appearance of "small denominators" $\omega_m = m, \omega$ vanishing on any resonance surface. Since the latter are generally everywhere dense in phase space one may readily imagine the scale of difficulties to be encountered. The difficulties are well known in celestial mechanics. The first task is not to get just on a resonance surface. For the standard formulation of the stability problem, i.e. for given initial conditions, already $I^{(1)}$ in eq. (4.74) may happen to lie on a resonance surface. Even if this is not the case it may be so after the next canonical transformation, i.e. $m, \omega(I^{(2)}) = 0$. Since an infinite number of transformations are generally required it is completely unclear how one could ensure missing all the resonance surfaces in all the approximations of perturbation theory to say nothing of a possibility for $I^{(n)}$ to come so close to a resonance surface to spoil the convergence of a perturbation series. To cope with this difficulty Kolmogorov has changed the formulation of the problem: instead of studying the behavior of a trajectory with fixed initial conditions he is studying trajectories with a fixed set of frequencies: $\omega = \omega^* = \text{const}$ in all approximations. Owing to a change in frequencies from one approximation to another, due to the

average part of perturbation ($V_0^{(n)} = \langle V^{(n)}(\theta^{(n)})\rangle$) the initial conditions have to be corrected in every approximation to compensate this change. For the sequence of corrections to be convergent the nonlinearity of the system must be not too small – a condition similar to one of the inequalities for moderate nonlinearity (3.14): $\alpha \gtrsim \epsilon$. One may say, alternatively, that in Kolmogorov's approach the behavior of a phase space torus with fixed frequencies is under study. A stable torus gets only slightly distorted under perturbation, an unstable one gets destroyed and converted into something extremely intricate and incomprehensible (see section 6).

The second task is to estimate small denominators from below to ensure the convergence of the perturbation series. This problem is quite similar to that of evaluating the condition for resonance overlap (section 4.5). Indeed, we may exclude some "dangerous" neighbourhood of resonance surfaces (the resonance layers) and thus bound small denominators from below. Since the frequencies are fixed the small denominators remain bounded in all approximations. A difference from the problem of resonance overlap is that the resonance width is determined by the dynamics of a system, whereas the width of the "dangerous" layer that we exclude from initial conditions may be chosen arbitrarily and, hence, in an optimal way. The layer width has to decrease with resonance harmonic numbers to ensure that the total volume of excluded frequency domains remains finite (and small). The estimate (4.66) shows that a power decrease law may be chosen. Since the estimates used for small denominators are not always optimal we will consider the question in more detail.

Let ω be within a resonance layer excluded so that

$$|\omega_m| = |m, \omega| \leq \delta/S^k \equiv |m|\,(\Delta\omega)_d \tag{4.75}$$

where $(\Delta\omega)_d$ is the half-width of the layer and δ is some constant. We find first the relation between $|m|$ and S:

$$S^2 = |m|^2 + \sum_{i\neq j} |m_i|\,|m_j| = |m|^2\Big(1 + \sum_{i\neq j} |m_i|\,|m_j| \Big/ |m|^2\Big). \tag{4.76}$$

The expression in brackets is maximal for $|m_1| = |m_2| = \cdots = |m_N| = S/N$, whence:

$$|m| \leq S \leq |m|\sqrt{N}. \tag{4.77}$$

Eq. (4.75) gives then:

$$(\Delta\omega)_d \leq \sqrt{N}\,\delta/S^{k+1}. \tag{4.78}$$

Substituting $(\Delta\omega)_d$ for $(\Delta\omega)_r$ in the estimates of the preceding section we find the relative volume of the excluded frequency domains:

$$\mu \leq \frac{\delta}{|\omega|}\sqrt{N} \sum_{S=1}^{\infty} \frac{1}{S^{k+2-N}}. \tag{4.79}$$

Convergence of the sum determines the minimal power index $k_{\min}$ in the decrease law (4.75):

$$k > k_{\min} = N - 1. \tag{4.80}$$

Estimates of small denominators were considered in many papers. For $N = 2$ it is the problem of accuracy in approximation of a real number by rational ones with a given denominator (see, e.g., ref. [42]). A general estimate with power index (4.80) has been given by Moser [60]. Recently Rüssmann has reconsidered the problem in detail [117].

For an analytical perturbation the convergence of the perturbation series is almost "obvious" since

small denominators (4.75) are decreasing much slower than Fourier amplitudes (4.63). However, a detailed description of the proof for this "obvious" fact comprises more than a half of a bulky book [61]. A brief and lucid presentation of the proof is outlined, though, in an excellent review article by the author of the proof Arnold [2].

The technique for constructing convergent perturbation series has been improved considerably by Moser who managed to extend it onto a smooth perturbation. In the first paper [3] a mapping similar to those of section 4.2 was considered, and a critical value of the perturbation smoothness $l_c = 333$ (!) has been obtained. But even for this, rather modest nowadays, achievement a special technique for the approximation of a smooth perturbation by analytic functions was required. It is interesting to remark that the idea of such an approach making use of finite trigonometric polynomials for the approximation had been mentioned as early as in 1955 by Bogolyubov and Mitropolsky (see [4], §13) even though later on Mitropolsky and Samoilenko utilized a different technique (smoothing operator) [63].

The best estimate for the mapping has been achieved so far by Rüssmann [64]: $l_c = 6$. In the book [3] it is asserted (without proof) that the latter estimate can be lowered down to $l_c = 4$ and is conjectured that in fact $l_c = 3$. On the other hand, for $l = 2$ the motion is unstable as has been shown in ref. [65]*. Thus rigorous estimates obtained turn out to be fairly efficient, at least, in regard to a sufficient smoothness of perturbation. The overlap criterion gives for the mapping: $l_c = 2$ [35] (see also section 4.5).

Quite unexpectedly numerical experiments have revealed that the motion described by a mapping proved to be much more stable, at least, in regard to the gross instability, as compared to all above estimates including the lower one due to resonance overlap. In the work [35], for example, a mapping of the type (4.11) with a smoothness parameter $l = 1$ and $K = 1.145$ was studied. Nevertheless, an instability observed was remaining perfectly confined within a relatively small interval $\Delta I \sim 1$ as long as 3×10^6 iterations (during all the computation time). This curious phenomenon has been apparently encountered first in numerical experiments by Hine [84] but still has not received a satisfactory explanation. There is only a suspicion that it may be due to a very slow diffusion if sufficiently high harmonic resonances are involved.

Coming back to a continuous system with the Hamiltonian (4.62) we observe, first of all, that to ensure the convergence of perturbation series it is not enough for Fourier amplitudes V_m to decrease faster than small denominators in eq. (4.74). It would lead to the value $l_c = N - 3$ in plain contradiction to the lower estimate due to resonance overlap (4.67). Even for the series

$$|I - I_1| = \epsilon \left| \sum_m m\, \Phi_m\, e^{im.\theta} \right| \leqslant \epsilon \sum_S \frac{S\, S^{N-1}}{S^{l+2-k}}$$

of the first canonical transformation of variables to be convergent we have to require $l > 2N - 2$ ($k > N - 1$, see eq. (4.80)). In subsequent approximations the number of continuous derivatives will progressively decrease due to small denominators, and after a finite number of successive approximations the series are no longer convergent. "Smoothing" a smooth perturbation to an analytic one allows us to cope with this difficulty, and Moser finds [60, 66] (see also ref. [3]):

$$l_c = 2N + 2. \tag{4.81}$$

*Let us draw attention that all values for l mentioned are greater by one as compared to those in the book [3] because in this paper the parameter l is a characteristic of the Hamiltonian (or generating function) whereas in ref. [3] l is that of mapping itself, that is to say, of the "force".

Comparison with the overlap criterion ($l_c = 2N - 2$) shows that the sufficient estimate (4.81) by Moser is fairly efficient.

The more powerful technique for constructing convergent series in the case of a smooth perturbation may be applied back to the analytical case. To do so we observe that an anytical perturbation of the kind (4.63) can be limited from above by a smooth one:

$$V_1 \mathrm{e}^{-\sigma S} = \frac{V_1}{S^{l+2}} S^{l+2} \mathrm{e}^{-\sigma S} \leqslant \frac{V_1}{S^{l+2}} \left(\frac{l+2}{\mathrm{e}\sigma}\right)^{l+2}. \tag{4.82}$$

But according to Moser's theorem [60, 66] the last perturbation does not destroy the stability of the motion if $l > 2N + 2$ and $\epsilon V_1 [(l+2)/\mathrm{e}\sigma]^{l+2} < \epsilon_0 H_0$ where ϵ_0 is some (unknown) constant independent of σ. Hence the stability of the motion for the original system (4.62) with analytical perturbation (4.63) is assured when

$$\epsilon V_1 \leqslant (\epsilon V_1)_c \sim (\sigma/N)^{l+2} \epsilon_0 H_0; \qquad l > 2N + 2. \tag{4.83}$$

Applying the normalization (4.70) we get:

$$\epsilon_c \sim \epsilon_0 (\sigma/N)^p; \qquad p > \tfrac{3}{2}N + 4 \tag{4.84}$$

which is rather close to the estimate (4.71) from the resonance overlap. We emphasize that the sufficient estimate (4.84) is related to the *root-mean-square* perturbation owing to the specific normalization of ϵ (4.70). That normalization is rather convenient since it does not depend on phase relations between perturbation Fourier harmonics.

Let us consider now in more detail what does the border of stability (4.84) really mean. Literally it implies that some (non-resonant, or "good") invariant tori of the unperturbed system are only slightly distorted remaining topologically tori and invariant. A "good" torus is related to those frequency vectors ω for which small denominators (4.75) are not too small. The volume of the complementary area occupied by "bad" (resonant) tori whose conservation cannot be ensured by the theory is small under sufficiently small perturbation (4.79). However, "bad" tori fill up the everywhere dense set of resonance layers. The problem of the behavior for such a system becomes, therefore, improper, or ill-set, since it is actually impossible to localize initial conditions within a "good" domain. Nevertheless, it turns out that for a system with only two degrees of freedom the stability of the motion can be assured under a sufficiently small perturbation. This is related to the special topological structure of the phase space for such a system. Namely, one says that 2-dimensional tori divide a 3-dimensional energy surface. It implies that any transition from one torus to another is possible only through all the intermediate tori. In other words, the set of tori on an energy surface can be ordered as a one-dimensional set. It is clearly related to the fact that dimensionality of the energy surface exceeds that of a torus exactly by one. It's said also that tori are put inside one another on an energy surface. This specific structure reminds, as an American physicist J. Ford has mentioned once, the Russian doll "Matreshka". One may also graphically imagine this structure on the frequency plane of the system. Projection of an energy surface onto this plane is a line every point of which corresponds to a certain different ratio of the frequencies. It is clear, therefore, that a system confined within a resonance layer can move along the energy line only by a distance of the order of the resonance layer width. An exception is a system whose energy lines go just along the resonance lines. They are so-called non-steep, or quasi-isochronous, systems (see section 3.3).

From the above consideration it must be quite clear that no matter what may happen with a two degrees-of-freedom system inside a resonance layer it cannot penetrate the nearest invariant tori and

is, thus, confined between sufficiently (although not everywhere) dense invariant tori. The maximal distance between adjacent invariant tori is of the order of resonance layer width, i.e. $\sim\sqrt{\epsilon}\to 0$ for $\epsilon\to 0$. From the physical point of view this is just stability of motion. One may say, alternatively, that an instability, if any, is perfectly confined to an arbitrarily small domain for $\epsilon\to 0$.

The conservation of tori implies also conservation of the full set (N) of unperturbed integrals of motion, which are close to those of the unperturbed system. However, the integrals of motion are, or may be, destroyed inside the everywhere dense set of resonance layers. Hence their dependence on dynamical variables is everywhere discontinuous. This clarifies the meaning of the old Poincaré theorem [6] that a generic Hamiltonian system has no analytical integrals of motion except the energy (see footnote on page 285).

Similar considerations can be applied also to a system having one degree of freedom but acted upon by a periodic perturbation, that is to say, having 1.5 degrees of freedom (see section 3.2). Again, 2-dimensional tori divide 3-dimensional phase space (there is no energy conservation now!), or, more graphically, resonant values of the only variable frequency ω are separated by non-resonant ones.

For the number of degrees of freedom $N>2$ the n-dimensional invariant tori do not divide $(2N-1)$-dimensional energy surface any longer since the dimensionality difference $(2N-1)-N = N-1>1$. Another manifestation of the same property is an amalgamation of isolated resonance layers into a united network over an energy surface. Indeed, in the N-dimensional ω-space an intersection of the $(N-1)$-dimensional resonance surface and an energy surface (also $(N-1)$-dimensional) is of dimensionality $N-2>0$ for $N>2$. Hence those intersections do also generally intersect each other forming an everywhere dense resonance network, or a "web", entangling all the energy surface. Invariant tori occupy most of the space off the "web". However, for initial conditions inside the "web" a system can move over all the energy surface remaining all the time inside one resonance layer or another. Such a general conjecture has been put forward by Arnold [67] subsequent to his construction of the first example of a system moving along a resonance layer [5]. We shall consider this problem in detail in section 7. Let us just mention now that this peculiar instability, called the Arnold diffusion, is really a universal one since it occurs for any, arbitrarily small, perturbation. This is a distinction between Arnold diffusion and stochastic instability which requires the overlap of resonances, i.e. a sufficiently strong perturbation.

5. Stochastic oscillations of the pendulum

In this section we shall consider in detail the motion of a dynamical system represented by the mapping:

$$\bar{I} = I + K\sin\theta; \qquad \bar{\theta} = \theta + \bar{I} \tag{5.1}$$

which we call the standard mapping for brevity. In the preceding section we have seen that this mapping describes approximately some real physical systems, for example, a charged particle in a magnetic bottle (section 4.3). Of a greater importance is that the standard mapping describes the motion near the separatrix of a fairly general nonlinear resonance (section 4.4). So I believe the standard mapping deserves thorough study even though this mapping is but a very particular case in the general theory of nonlinear mappings. In connection with various applications the standard mapping has been studied in many works (see, e.g., [45–47]). Nevertheless, our knowledge of its properties is still far from being complete. Even such a fundamental characteristic as the stability border (the critical value of the

perturbation parameter $K = K_1$) is known only approximately from numerical experiments. Analytical estimates for K_1 are still fairly rough. These questions are considered in section 5.1.

The problem of the nature of the motion and of the related structure of the phase plane turns out to be still more complicated. It is especially true near the stability border ($K \sim K_1$) where the phase plane structure is extremely intricate (section 5.5). However, far inside the unstable region ($K \gg K_1$) a simple statistical description proves to be applicable (section 5.4). It is found, in this connection, that the instability occuring for $K > K_1$ results in an irregular ("random") motion of the system which has been called the stochastic motion. The basic property of such a motion is a strong local instability implying an exponential divergence of adjacent trajectories (section 5.2). Statistical properties of the motion are subsequent just upon that instability (section 5.3).

A plain interpretation of the mapping (5.1) is the motion of a pendulum acted upon by the successive short "kicks" (see section 4.2). Under some conditions these regular (periodic) "kicks" give rise to "random" (irregular) oscillations of the pendulum. This graphic picture determined the title of the present section. Let us remark that stochastic oscillations have proved to be a new and somewhat surprising kind of pendulum motion, the study of which is now just beginning.

5.1. The border of stability

In this section we shall consider the so-called gross instability of motion (see section 4.2) implying arbitrarily large excursions in the unperturbed motion integral I of system (5.1). We have learned already in the preceding section 4 that the mechanism of the instability is related to the transitions of the system from one resonance to another. The structure of the phase plane of the standard mapping is periodic not only in θ but also in I (both are of period 2π, see eq. (5.1)). The phase plane is, thus, topologically equivalent to a torus and may be displayed as a square with opposite sides identified. Hence it suffices to consider the motion over this square and, particularly, over a momentum period $\Delta I = 2\pi$, that is over an interval between two adjacent integer resonances: $I_r = 2\pi n$. Therefore, an equivalent mapping

$$\bar{P} = \left\{P + \frac{K}{2\pi q}\sin(2\pi X)\right\}; \qquad \bar{X} = X + q\bar{P} \tag{5.2}$$

may be used to study the standard mapping (5.1) where curly brackets denote the fractional part. The new mapping (5.2) is deduced from the standard one (5.1) via a change of variables: $\theta = 2\pi X$; $I = 2\pi q P$ and via the "closure" of P over the interval (0, 1), i.e. by identification of the values of P and $P + m$ for any integer m, the number q being an integer as well. So the phase space of the mapping (5.2) is a unit torus, or a unit square (see, e.g., fig. 5.1), and we shall call it the *reduced mapping*. Since integer resonances of the standard mapping ($I_r = 2\pi n$) are identical with the resonances $P_r = n/q$ for the reduced mapping there are exactly q integer resonances within the unit square of system (5.2) ($P_r = 0$ and 1 correspond to the same resonance)*.

For a numerical determination of the stability border the value of $q = 2$ is most appropriate (two integer resonances: $P_r = 0(1)$; 1/2). The critical perturbation strength K_1 was determined as follows. For different values of K and various initial conditions (P_0, X_0) the motion time N (the number of iterations of the reduced mapping) during which the system has come from the initial interval about $P = P_0 \approx 0$ ($\sim 10^{-3}$) into an interval about $P \approx \frac{1}{2}$ ($|P - \frac{1}{2}| < 0.125$) was computed. The quantity N can

*Let us emphasize again that we are talking about integer resonances of the standard mapping using the reduced mapping to study the former.

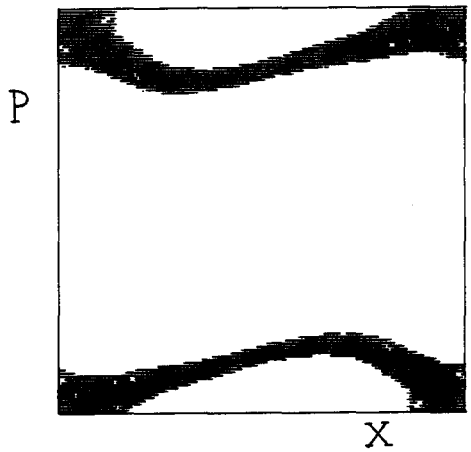

Fig. 5.1. Phase map of system (5.2): $K = 0.96$; $q = 2$; $P_0 = 0.219 \times 10^{-3}$; $X_0 = 0.502 \times 10^{-3}$; $t = 10^6$; resolution: 128×128 bins.

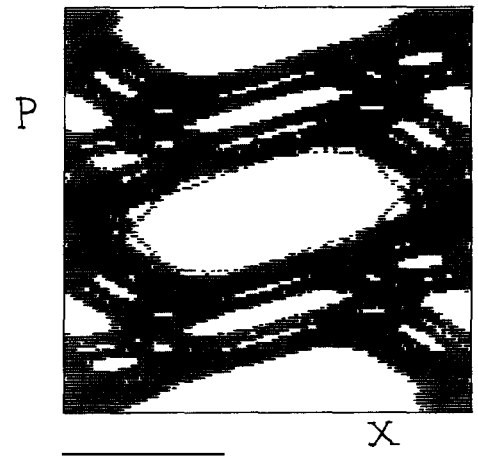

Fig. 5.2. The same as in fig. 5.1 except $K = 1.13$; $t = 10^5$.

also be called the *transition time* between adjacent integer resonances. If during the full motion time t (the total number of iterations computed) the system has not got into a specified interval around $P = \frac{1}{2}$ the motion is considered to be stable (over the time interval t). It is worth noting that the use of mappings rather than differential equations in numerical experiments permits us to study a fairly long-term motion, the computational errors reducing actually to round-off errors only. A typical motion time for the reduced mapping was about 10^6 iterations, the maximal one having reached 10^7.

In fig. 5.1 an example of the phase map for stable motion over the time interval $t = 10^6$ ($K = 0.96$) is shown. A trajectory with initial conditions $P_0 = 0.219 \times 10^{-3}$; $X_0 = 0.502 \times 10^{-3}$ fills up the hatched domains of the phase square. Both domains correspond to the same resonance $P_r = 0$. A single bin of the picture is of dimensions $1/128 \times 1/128$. For aesthetic reasons the adjacent bins crossed by a trajectory are drawn as a solid line. As is seen from fig. 5.1 the trajectory has failed during 10^6 iterations to get into domain of the neighbouring integer resonance. Hence the motion is bounded and stable in the above-mentioned sense, there is no gross instability. This implies, in particular, that resonances do not overlap the whole interval $\Delta I = 2\pi$, or $\Delta P = \frac{1}{2}$. On the other hand, the fact that a trajectory in fig. 5.1 does not look like a curve but fills up a certain finite area indicates some localized, or confined, instability of motion. We shall consider in detail that confined instability in section 6. Now let us just note that the unstable area forms something like a layer along the separatrix of resonance $P_r = 0$.

In fig. 5.2 an example of the phase map for unstable motion is shown. Now resonances do overlap the whole interval of P. It implies, in particular, that a trajectory of the standard mapping can cross one resonance after another, and the variation of I becomes indefinitely large, that is, a gross instability sets in. As is seen from fig. 5.2 the instability takes place, however, only for some special initial conditions inside the hatched domain. At the same time there are also many domains of stability which the unstable trajectory fails to enter. The two biggest domains correspond to the central part of integer resonances: $P_r = \frac{1}{2}; 0$ (the latter is split in two halves in fig. 5.2). Besides, there are many smaller stability domains to which we shall return in section 5.5. Thus far we can conclude that even under the resonance overlap the unstable domain comprises only a part of the whole phase space

including, in the first place, vicinities of resonance separatrices. This is precisely the reason to choose the initial conditions for computation near the separatrix of a resonance ($P_0, X_0 \ll 1$).

Numerical data concerning the dependence of the transition time versus perturbation parameter K are presented in fig. 5.3. A distinctive feature of this dependence is a sharp increase of N at some K. We have interpolated the data using the function (see below for explanation):

$$N = A/(K - K_E)^B \tag{5.3}$$

where A, B and K_E are constants to be determined via interpolation. The least square fit of the data provides:

$$A = 103; \qquad B = 2.55; \qquad K_1 \approx K_E = 0.989 \approx 1. \tag{5.4}$$

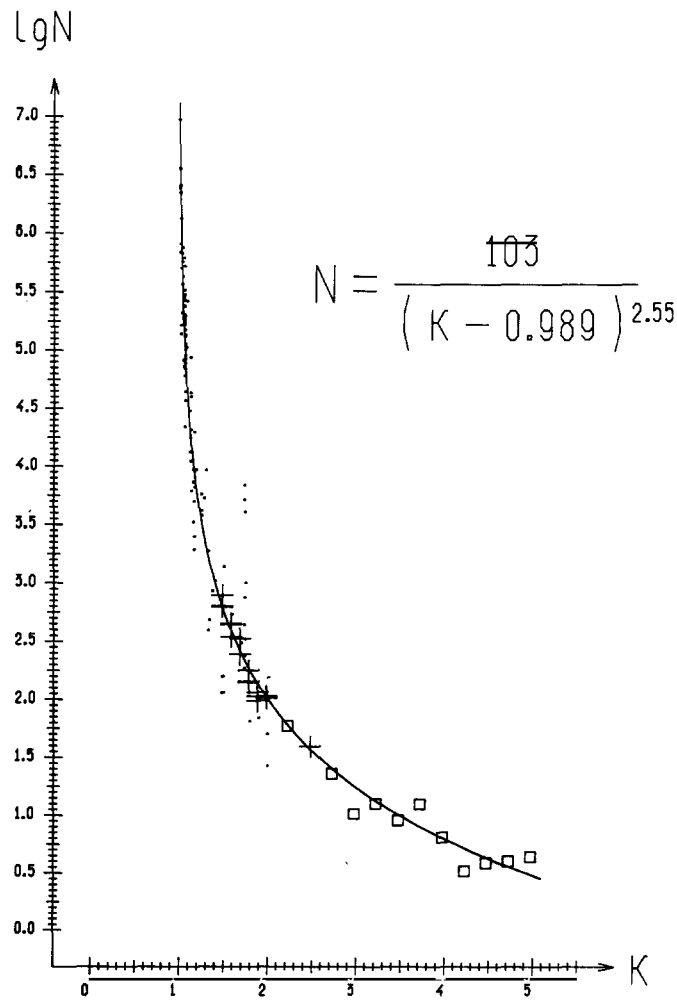

Fig. 5.3. Transition time N versus perturbation parameter K for reduced mapping (5.2): ● – a single trajectory; + – the average over 100 trajectories; □ – from the diffusion rate (5.7). Solid curve gives least square fit of the data according to the relation written in the figure.

Eq. (5.3) shows that the quantity K_E can be considered as an empirical value for the stability border K_1.

Another peculiar feature of dependence $N(K)$ is extremely big fluctuations. Around $K \approx 1.8$ the fluctuations are as large as about two orders of magnitude! We could add only that for some initial conditions (with $K > K_1$) the motion proves to be even just stable, that is $N > t$. The latter phenomenon is related, apparently, to entering very small stability domains. This question was investigated more thoroughly for a similar mapping in the work [26]. To suppress the fluctuations we have used averaging N over 100 trajectories with slightly different initial conditions ($P_0, X_0 \ll 1$, see fig. 5.3).

For large $K \gg K_1$ the described method of measuring N does no longer make any sense since the resonances get completely destroyed, and the variation of P becomes "random", or stochastic (sections 5.2 and 5.4). In other words, there are no more resonances to transit between. On the other hand, this very stochasticity of motion suggests another method to apply in this region. Namely, as will be shown in section 5.4 the stochastic motion can be described by a simple diffusion-like law:

$$\langle(\Delta I)^2\rangle = \tfrac{1}{2} t K^2 \tag{5.5}$$

where ΔI is the change in momentum during time t, and the averaging is over an ensemble of trajectories with different initial conditions. But near the stability border ($K \gg K_1$) one can write:

$$\langle(\Delta I)^2\rangle = \frac{(2\pi)^2}{N} t \tag{5.6}$$

assuming a "random" transition between adjacent resonances with equal probability in both senses ($t \gg N$). Hence the average transition time N may be calculated also from the relation:

$$N = 4\pi^2 t / \langle(\Delta I)^2\rangle. \tag{5.7}$$

This method fits any K, yet it can be used actually only for sufficiently large K (small N, see fig. 5.3) since a substantially longer computation time is required in this case as compared to the first method ($t \gg N$). Values of N obtained by the second method are marked in fig. 5.3 by a square □. An averaging over 100 trajectories has been made for each point ($P_0, X_0 \ll 1$, $t = 10^3$). One can see that both methods fit each other quite well (region $K \sim 2.5$).

Comparing eqs. (5.5) and (5.6) we get for $K \gg K_1$:

$$N = 8\pi^2 / K^2. \tag{5.8}$$

On the other hand, the data in fig. 5.3 do reveal a critical value of $K = K_1$, such that $N \to \infty$ for $K \to K_1$. Therefore, just the expression (5.3) has been chosen to interpolate the experimental data.

The existence of eternal (for any $t \to \infty$) stability border is a corollary of the rigorous KAM theory (section 4.6). The question arises, however, to what accuracy this border can be determined from the data in fig. 5.3 for a finite motion time $t \leqslant 10^7$? Big fluctuations of N make it difficult to get reliable estimate for the errors. One can hope, however, that they must not be too large due to a steep change of N near $K = K_1$. Measurements of the critical strength of perturbation by another method (see section 6.2) suggest an error in K_1 of the order of a few per cent.

Let us compare now the experimental data obtained with an estimate by the overlap criterion (4.20): $K_T = \pi^2/4 \approx 2.5$. The latter overestimates materially the critical value as compared to the experimental one (5.4). Examination of fig. 5.2 reveals, at least, three reasons for that discrepancy:

1. The overlap is provided not only by integer resonances, as was assumed in evaluating the estimate (4.20), but also by higher harmonic resonances. In fig. 5.2 two half-integer resonances (harmonic number $m = 2$) are clearly seen via their stability domains. Inside each half-integer resonance there are two stability domains along the X-axis, one of which being split in the figure in two parts by the line $X = 0 = 1$. Four resonances with $m = 3$ (3 stability domains each) are also seen, and two $m = 4$ resonances seem to exist on both sides of the integer resonance $P_r = 0(1)$. It is clear that higher harmonic resonances facilitate the overlap.

2. Resonance separatrices are distorted ("twisted") due to a resonance interaction that has not been taken into account.

3. Separatrices are actually layers (see also fig. 5.1) rather than curves (the stochastic layers, section 6) that also facilitates the resonance overlap.

Below we shall attempt to improve the simple estimate (4.20) taking account of some higher harmonic resonances. We shall not consider separatrix "torsion" which is of a less influence on the resonance overlap as is obvious from fig. 5.2. Account of a finite width of separatrix stochastic layers will be taken in section 6.4.

To study the influence of higher harmonic resonances we transfer from the standard mapping to an equivalent continuous system with Hamiltonian (see section 4.2):

$$H(J, \theta, t) = \frac{J^2}{2} + k \sum_{n=-\infty}^{\infty} \cos(\theta - nt); \qquad J = \frac{I}{2\pi}; \qquad k = \frac{K}{(2\pi)^2}. \tag{5.9}$$

Assuming k to be small we take $H_0 = J^2/2$ as the unperturbed Hamiltonian and introduce a canonical transformation $J, \theta \to J_1, \theta_1$ such as to "kill" the perturbation $\sim k$ (see sections 2.2 and 4.6). We are looking for a generating function of the usual form:

$$F(J_1, \theta) = \theta J_1 + k\, \Phi(J_1, \theta, t)$$

whence

$$J = J_1 + k\, \Phi_\theta; \qquad \theta_1 = \theta + k\, \Phi_{J_1}; \qquad H_1 = H + k\, \Phi_t. \tag{5.10}$$

Substituting these relations into eq. (5.9) we arrive at the condition for "killing" the perturbation $\sim k$ in H_1 in the form of a partial differential equation for the function $\Phi(J_1, \theta, t)$:

$$J_1\, \Phi_\theta + \Phi_t + V(\theta, t) = 0; \qquad V = \sum_{n=-\infty}^{\infty} \cos(\theta - nt). \tag{5.11}$$

A solution of this equation may be looked for in the form

$$\Phi(J_1, \theta, t) = \sum_n a_n(J_1) \sin(\theta - nt)$$

suggested by the type of perturbation V. Substituting Φ into eq. (5.11) we get: $a_n = 1/(n - J_1)$, whence:

$$\Phi = \sum_n \frac{\sin(\theta - nt)}{n - J_1}; \qquad \Phi_\theta = \sum_n \frac{\cos(\theta - nt)}{n - J_1}; \qquad \Phi_{J_1} = \sum_n \frac{\sin(\theta - nt)}{(n - J_1)^2}. \tag{5.12}$$

The new Hamiltonian is:

$$H_1 = \tfrac{1}{2} J_1^2 + k^2\, V_1(J_1, \theta_1, t) \tag{5.13}$$

where the new perturbation

$$V_1 = \frac{\Phi_\theta^2}{2} = \frac{1}{2}\sum_{m,n} \frac{\cos(\theta - mt)\cos(\theta - nt)}{(m - J_1)(n - J_1)} \tag{5.14}$$

and $\theta(J_1, \theta_1, t)$ is determined by the second relation of eq. (5.10).

The perturbation (5.14) has terms resulting in half-integer resonances: $J_1 = J_{1r} = (2p+1)/2$ with any integer p. Corresponding resonance phases are $2\theta_1 - (m+n)t$, $m + n = 2p + 1$ where we have set approximately $\theta \approx \theta_1$ since the difference $|\theta - \theta_1| \sim k$ is small (see eq. (5.10) and below). Note that the new Hamiltonian is inapplicable near integer resonances ($J_{1r} = 2p$) owing to small denominators in eq. (5.14). Characteristics of a half-integer resonance $J_{1r} = p + \frac{1}{2}$ are determined by the sum:

$$S_2 = \sum_{m+n=2p+1} \frac{1}{(m - J_{1r})(n - J_{1r})} = -\sum_n \frac{1}{(n - p - \frac{1}{2})^2} = -\pi^2. \tag{5.15}$$

Since this sum is independent of p all half-integer resonances are alike except a shift in J. This is obvious also just from the periodicity of the phase space structure in I discussed at the beginning of this section: there is only one half-integer resonance per unit square of the reduced mapping (5.2). Substituting $J_1 = J_{1r} = \frac{1}{2}$ into eq. (5.14) and neglecting non-resonant terms we deduce from eq. (5.13) the Hamiltonian for a half-integer resonance as

$$H_1^{(2)} \approx \tfrac{1}{2}J_1^2 - (\tfrac{1}{2}\pi k)^2 \cos(2\theta - t). \tag{5.16}$$

Applying the technique of section 3.2 we find the separatrix to be described by:

$$J_s^{(2)} = \tfrac{1}{2} \pm \pi k \cos(\theta - \tfrac{1}{2}). \tag{5.17}$$

Substituting now $\theta(\theta_1)$ according to eq. (5.10) into the perturbation (5.14) we may represent H_1 as a series in the small parameter k. However, it makes sense to retain only terms $\sim k^3$ in addition since following a new canonical transformation according to Kolmogorov the perturbation will be $\sim k^4$ (section 2.2). To get terms $\sim k^3$ it suffices to set in eq. (5.10)

$$\theta \approx \theta_1 - k\,\Phi_{J_1}(\theta_1) = \theta_1 - k \sum_n \frac{\sin(\theta - nt)}{(n - J_1)^2}.$$

After substitution of this expression into eq. (5.14) we get terms $\sim k^3$ as

$$k^2 V_1^{(3)} = \frac{k^3}{2} \sum_{m,n,l} \frac{\sin(2\theta_1 - (m+n)t)\sin(\theta_1 - lt)}{(m - J_1)(n - J_1)(l - J_1)^2}. \tag{5.18}$$

Some terms here result in third harmonic resonances with phases $3\theta - (m+n+l)t$; $m + n + l = p$ and $J_1 = J_{1r} = p/3$ ($p \neq 3q$ for any integer q to avoid integer resonances). All the resonances are alike again except a shift in J_1. Indeed, the denominator in eq. (5.18) may be written as: $(3m - p)(3n - p) \times (3m + 3n - 2p)^2$. It suffices to consider only $p = 1$; 2 within the interval of periodicity $\Delta J_1 = 1$ (see below). But quantities $(3m - 1)$ and $(3m - 2)$ are exchanged under $m \to 1 - m$, the sum (5.18) being unchanged. The same is true also for n. Thus, each of third harmonic resonances is determined by the sum:

$$S_3 = \sum_{m+n+l=1} \frac{81}{(3m-1)(3n-1)(3l-1)^2} = \sum_{m,n} \frac{81}{(3m-1)(3n-1)(3m+2n-2)^2} \approx -86.4. \tag{5.19}$$

The Hamiltonian describing a third harmonic resonance has the form:

$$H_1^{(3)} \approx \tfrac{1}{2}J_1^2 - \tfrac{1}{4}S_3 k^3 \cos(3\theta - t) \tag{5.20}$$

and the resonance separatrix is

$$\begin{aligned} J_s^{(3)} &= \tfrac{1}{3} \pm (\Delta J)_3 \sin(\tfrac{3}{2}\theta - \tfrac{1}{2}t) \\ (\Delta J)_3 &= |S_3|^{1/2} k^{3/2} \approx 9.30 k^{3/2}. \end{aligned} \tag{5.21}$$

Resonances of the first three harmonics are outlined in fig. 5.4 at $t = 0$. One can see that the maxima of all the resonances coincide in θ (cf. fig. 5.2). Therefore, the overlap condition corresponds to the maximal width of resonance separatrices:

$$(\Delta J)_1 = 2\sqrt{k}; \qquad (\Delta J)_2 = \pi k; \qquad (\Delta J)_3 = |S_3|^{1/2} k^{3/2}. \tag{5.22}$$

Now we can improve the theoretical estimate for critical perturbation. First, we take account of half-integer resonances only. Then the perturbation is critical if a half-integer resonance separatrix just touches separatrices of two adjacent integer resonances, overlapping the gap between them: $(\Delta J)_1 + (\Delta J)_2 = \frac{1}{2}$, whence:

$$\sqrt{k_{12}} = \frac{\sqrt{1 + \pi/2} - 1}{\pi} \approx 0.192; \qquad K_{12} \approx 1.46 \tag{5.23}$$

where the subscript indicates the resonance pair determining critical perturbation. The value (5.23) is considerably lower than the old one due to integer resonances: $K_{11} \approx 2.5$, yet the former still much exceeds the experimental $K_1 \approx 1$.

Taking account of third harmonic resonances we need to consider the two pairs of adjacent resonances: $(0, \frac{1}{3})$ and $(\frac{1}{3}, \frac{1}{2})$. In the first case touching condition is: $(\Delta J)_1 + (\Delta J)_3 = \frac{1}{3}$, whence:

$$\sqrt{k_{13}} \approx 0.151; \qquad K_{13} \approx 0.90. \tag{5.24}$$

For the second pair: $(\Delta J)_3 + (\Delta J)_2 = \frac{1}{6}$ which gives

$$\sqrt{k_{23}} \approx 0.185; \qquad K_{23} \approx 1.35. \tag{5.25}$$

The latter value of K is decisive, but it improves only slightly the estimate (5.23).*

Further progress in theoretical estimates may follow two lines:

1. Taking account of still higher harmonic resonances: $m > 3$. I wonder if one of the readers would like to try this.

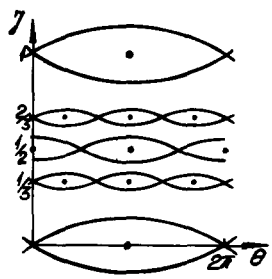

Fig. 5.4. Diagram of separatrices for resonances: $J_r = 0$; 1/3; 1/2; 2/3; 1.

*As was shown by Cary [140] the estimate (5.25) can be improved by taking account of the frequency shift due to $\langle V_1(\theta)\rangle \neq 0$ (see eq. (5.14)); he arrives at $\sqrt{k_{23}} = 0.180$; $K_{23} = 1.28$.

2. Taking account of a finite width of the stochastic layer. It will be considered in section 6.4.

To the best of my knowledge the only work where the stability border for the standard mapping has been determined numerically was that by Lieberman and Lichtenberg [46]. They have obtained also a theoretical estimate by a different method as compared to the present paper. Here are their results:

$$K_E \approx 0.80; \qquad K_T = 2. \tag{5.26}$$

The theoretical estimate is somewhat better than ours (4.20) due to integer resonances. The experimental value is also close to that in eq. (5.4). In any event, the value $K_E = 1$ does not contradict the data in ref. [46] (see fig. 10 there).

A critical perturbation for the standard mapping can be extracted also from the numerical data obtained by Channell [100]. In this work motion stability for a system given by the mapping

$$\bar{I} = I + \epsilon \sin\theta; \qquad \bar{\theta} = \theta + T\omega(\bar{I})$$
$$\omega(I) = \alpha + \beta m I^{m-1} \tag{5.27}$$

was studied. Linearizing in I we arrive at the standard mapping with the parameter

$$K = \epsilon T \omega'(I) = \epsilon\beta m(m-1)I^{m-2}; \qquad \langle K_1 \rangle = 0.97. \tag{5.28}$$

The last value gives the averaged critical perturbation strength recalculated from the numerical data in ref. [100]. In numerical experiments of ref. [100] the critical value of $I = I_1$ was determined, such that for $I > I_1$, the motion became stochastic judging by the phase space pictures displayed.

As we know the mapping describes an infinite set of resonances (5.9). Let us consider the opposite limiting case of the two resonance interaction (cf. section 4.1). A model of the latter system may be described by a Hamiltonian of the type (5.9) if we keep only two terms in the perturbation:

$$H(J, \theta, t) = \tfrac{1}{2}J^2 + k[\cos\theta + \cos(\theta - t)]. \tag{5.29}$$

The first approximation resonances (integer) remain unchanged under such a "truncation" of the perturbation. Consideration of higher harmonic resonances in the next approximation is similar to that for the standard mapping as has been done above. It turns out that the sums S_2 and S_3, having now just a few terms, change only slightly owing to rapid decreasing of the terms. So now we get: $S_2 = -8$ instead of $-\pi^2$, and $S_3 = -60.8$ instead of -86.4. Hence the critical perturbation strength for the two resonance interaction proves to be only a bit larger than in the case of an infinite set of resonances. From experimental data on the overlap of the two resonances given in section 4.1 one can deduce that $K_T/K_E \approx 2.26$ where $K_T = \pi^2/4$ is related to the overlap of integer resonances only. Hence for the two resonances the $K_1 = K_E \approx 1.09$ as compared to $K_1 \approx 1$ for the mapping.

5.2. The Krylov–Kolmogorov–Sinai entropy (KS-entropy)

Let us consider now the nature of the instability, arising in system (5.1) described by the standard mapping for $K > K_1$. In this section we consider the so-called local instability, a basic property of the unstable motion. The local instability characterizes the behavior of close trajectories, strictly speaking, infinitely close. Let the vector l with components ξ, η describe the mutual position of two close points in the phase plane of standard mapping:

$$\xi = \theta' - \theta; \qquad \eta = I' - I. \tag{5.30}$$

The variation of the vector l is described by the linearized mapping which is deduced from eq. (5.1) by differentiation with respect to I and θ:

$$\bar{\eta} = \eta + \xi K \cos\theta; \qquad \bar{\xi} = \xi + \bar{\eta} \tag{5.31}$$

where time-dependence of $\theta(t)$ is determined by the original mapping (5.1). Linearized mapping (5.31) is called the *tangent mapping*. Vector l is also called the tangent*. Although the tangent mapping is linear its coefficient depends on time (via $\theta(t)$). Therefore, its analytical treatment is almost as difficult as that of the original nonlinear standard mapping (5.1). To get some ideas about the nature of motion in the system (5.31) let us consider first another canonical mapping:

$$\bar{p} = p + k x; \qquad \bar{x} = \{x + \bar{p}\}. \tag{5.32}$$

Owing to the fractional part taken ($\{\ldots\}$) the dependence of the perturbation on x is not linear, of course, but has a jump discontinuity at $x = 0$ (a "saw-toothed" perturbation). Nevertheless, the tangent mapping in this case is plainly linear:

$$\bar{\eta} = \eta + k\xi; \qquad \bar{\xi} = \xi + \bar{\eta}. \tag{5.33}$$

Therefore, this mapping is much simpler to study since its complete solution is well known in an explicit form that is not the case for the mapping (5.32). The eigenvalues ($\lambda_\pm$) and the eigenvectors ($e_\pm$) of mapping (5.33) are obtained by a routine procedure and may be written as:

$$\lambda_\pm = 1 + \tfrac{1}{2}k \pm \sqrt{k(1 + \tfrac{1}{4}k)}; \qquad \eta_\pm/\xi_\pm = k/(\lambda_\pm - 1) \tag{5.34}$$

where $\xi_\pm$, $\eta_\pm$ are the components of eigenvectors. Let u, v be projections of the vector l onto eigenvectors ($l = ue_+ + ve_-$). Then, the solution to eq. (5.33) is well known to have the form

$$u = u_0 \lambda_+^t; \qquad v = v_0 \lambda_-^t \tag{5.35}$$

where u_0, v_0 correspond to the initial vector l_0, and t is a discrete time (number of iterations). For

$$-4 < k < 0 \tag{5.36}$$

the eigenvalues are complex numbers, and $|\lambda_\pm| = 1$ (5.34). That means, as is also well known, that the oscillations of l are bounded. One speaks in this event about the local stability of motion for the original (nonlinear) system (5.32) since the variation of the distance between close trajectories is bounded.

We are more interested, however, in the opposite case of parameter k to be outside the stable interval (5.36). Let $|\lambda_+| > 1$ and $|\lambda_-| < 1$ in the latter case ($\lambda_+\lambda_- = 1$). The motion (5.35) is then doubly-asymptotic according to Poincaré [6]: $|u| \to \infty$, $v \to 0$ ($t \to +\infty$); $u \to 0$, $|v| \to \infty$ ($t \to -\infty$) (see fig. 5.5). The asymptotes are directed along the dilation (e_+) and contraction (e_-) eigenvectors (for $t \to +\infty$), and are exchanged under time reversal. The motion of the original system (5.32) is now locally unstable since its close trajectories are diverging exponentially. Note that the divergence of trajectories takes place in *both* directions of time ($t \to \pm\infty$), the fact which is of great importance to understand properly the implications of local instability in the statistical behavior of a dynamical system [43].

*The meaning of this term is related to the fact that the vector l characterizes an infinitely small section of a certain curve in the phase plane. The curve is transformed according to the original mapping (5.1) but its small section does so according to linearized mapping (5.31), which, thus, describes the behavior of a beam of close trajectories. Let us mention also that the *completely* linear mapping (5.31) should not be confused with the *partly* (in I only) linearized standard mapping (5.1).

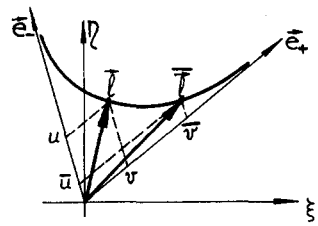

Fig. 5.5. A trajectory of tangent mapping (5.33) for system (5.32) (solid curve).

For $t \gg 1$ the vector l nearly coincides with dilation eigenvector e_+ (5.35) by direction and varies according to the law:

$$l(t) \approx l_0 |\lambda_+|^t = l_0 \, e^{t \cdot \ln|\lambda_+|}. \tag{5.37}$$

The quantity

$$h = \ln|\lambda_+| \tag{5.38}$$

describing the rate of divergence for close trajectories, is called the *Krylov–Kolmogorov–Sinai entropy*, or, in short, the *KS-entropy*. This quantity was widely used by Krylov [7] in his studies on the statistical properties of mechanical systems, he never used the term entropy though. Later on the same quantity has been independently introduced in the ergodic theory by Kolmogorov [13] via a quite different consideration based on the information theory, and was called by him the *entropy per unit time*. The relation between the Kolmogorov entropy and the local instability of motion has been found out by Sinai [14]. One should not confuse the KS-entropy with the entropy in statistical mechanics. Although a certain link does exist between them [9], these are quite different quantities. In statistical mechanics the entropy describes an instantaneous state of a system via the distribution function and may be defined as a dimensionless quantity. The KS-entropy has the dimensions of a frequency and describes the rate of local instability, or of the mixing process in a dynamical system (see section 5.3).

Now we turn back to the tangent mapping (5.31) for system (5.1). Here we cannot find the solution for an arbitrary trajectory $\theta(t)$ of the standard mapping (5.1). Yet, one may study a particular case of periodic motion. The simplest one is the motion of period 1 (fixed points of standard mapping): $\theta = 0$; π ($I = 2\pi n$). In the latter case eq. (5.31) is reduced to eq. (5.33) with $k = \pm K$. Then, as follows from eq. (5.36), one of the fixed points ($\theta = 0$) is always unstable (for $K > 0$, the separatrix of an integer resonance) whereas the other ($\theta = \pi$, the center of a resonance) becomes unstable only for $K > 4$. The latter condition is considered by some authors as a certain stability border [46, 92]. Indeed, under $K > 4$ the largest domains of stability inside integer resonances do disappear (see fig. 5.2). One should bear in mind, however, that the gross instability sets in under a considerably weaker perturbation ($K > 1$). Besides, some stability domains, of smaller dimensions though, persist for $K > 4$ as well (fig. 5.9).

Results closer to the border of gross instability can be obtained from the instability condition for trajectories of longer periods [71, 72]. It can be shown, for example, that for the standard mapping a solution of period 2 ($I = \pi$; $\theta = 0$; π – two centers of a half-integer resonance) becomes unstable if $|K| > 2$ (see section 5.5). The latter value is somewhat greater than that due to the overlap criterion, taking account of half-integer resonances (5.23), since the resonance centers may remain stable even if the separatrices do touch each other (fig. 5.2). It is not excluded, however, that considering solutions

of a sufficiently long period one would be able to get a stability condition close enough to the actual border of gross instability.* Another approach to using the local instability for the prediction of the onset of stochasticity was considered by Toda [120]. For subsequent discussions and developments of this method see refs. [111, 121, 122].

According to Sinai [14] the KS-entropy of a mapping is given generally by the relation:

$$h = \langle \ln(\bar{l}/l) \rangle \tag{5.39}$$

where the averaging is performed over a motion trajectory or, due to ergodicity, over the stochastic component, i.e. over a compact phase space area of the unstable motion. The stochastic component for the standard mapping is generally of an extremely intricate structure as is seen, e.g., in fig. 5.2. Numerical experiments show, however, that for $K \gg 1$ this component occupies actually the whole phase plane (see, e.g., fig. 5.9). Since asymptotically $\bar{l}/l \to |\lambda_+|$ where λ_+ is given by eq. (5.34) with $k = K \cos\theta$ the averaging in eq. (5.39) can be made just over the phase:

$$h \approx \langle \ln|\lambda_+(\theta)| \rangle. \tag{5.40}$$

For $K \gg 1$ the quantity $k = K \cos\theta$ is large except in narrow (in phase) stable phase intervals (5.36). Hence $|\lambda_+| \approx |k| = K\,|\cos\theta|$ (5.34), and eq. (5.40) is reduced to a fairly simple expression:

$$h_T \approx \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \ln|\cos\theta| + \ln K = \ln \frac{K}{2}. \tag{5.41}$$

The numerical determination of the KS-entropy was carried out by simultaneously iterating both tangent (5.31) and standard (original) (5.1) mappings. The mappings are taken in the reduced form (5.2) ($q = 1$) to avoid a rapid diffusion due to the stochastic instability. The quantity l in (5.39) was kept constant: $l = \sqrt{\xi^2 + \eta^2} \equiv 1$ during computation by the reducing vector $\bar{l}$, after every iteration, to that of the unit length without any change of its direction: $\xi = \bar{\xi}/\bar{l}$; $\eta = \bar{\eta}/\bar{l}$ (cf. ref. [39]). The entropy was computed then according to Sinai's formula (5.39):

$$h = \frac{1}{t} \sum_n^t \ln \bar{l}_n \tag{5.42}$$

where n stands for an iteration serial number and t is the total number of iterations. The method described is now under study in collaboration with G. Gasati, J. Ford and F.M. Izraelev. We attempt, in particular, to apply this method to much more complicated, many-dimensional systems to locate the onset of the stochastic instability. The method is similar to one used by Froeschlé [70]. In the latter a use of the tangent mapping is also made, the eigenvalues being computed in every iteration. This requires a lot of computation as compared to our version of the method but yields much more information.

Computation results for the KS-entropy of the standard mapping are given in table 5.1 (the quantity h_1). Values of h_1 were averaged over a number of trajectories computed, the motion time being $t = 10^5$ iterations per trajectory. The value of h_1 was almost independent of the initial direction of vector l_0 and, for $K > 1$, depended only slightly on initial conditions for the original mapping (5.1). For $K = 1.3$, for example, 6 different trajectories have yielded: $h_1 = 0.220$–0.237. The dependence on the motion time is illustrated by following data: $h_1 = 1.126$; 1.162; 1.164; 1.165 for $t = 10^3$; 10^4; 10^5; 10^6,

*Recently Greene has arrived in this way at the stability border for the standard mapping at $K = 0.971635\ldots$ [146].

Table 5.1
The KS-entropy of the standard mapping

K	h_2 [39] 2 trajectories	h_1 linear map	h_T (5.41)	$\frac{h_1}{h_T}$	$h_s = h_1 T_a$
1000	6.206	6.213	6.215	0.9997	1.11
200	4.603	4.601	4.605	0.9991	1.17
100.2	3.914	3.918	3.914	1.0010	1.12
50	3.227	3.222	3.219	1.0009	1.02
25	2.537	2.522	2.526	0.9984	0.89
6.21	1.157	1.164	1.133	1.027	0.78
4	—	0.833	0.693	1.202	0.82
3	—	0.672	0.405	1.659	0.90
1.3	—	0.227	—	—	0.80
1	—	0.132	—	—	0.64
0.5	—	6.86×10^{-2}	—	—	0.77
0.2	—	2.00×10^{-2}	—	—	0.65

respectively ($K = 6.21$). For $K = 1.3$ the value of h_1 obtained is well in agreement with the result by Froeschlé [70]: $h_1 \approx 0.22$ (see his fig. 1; note that in ref. [70] $\log \lambda$ is decimal!). For comparison the values of KS-entropy obtained by means of another method [39] are also given in table 5.1. In this method the vector l was determined by the two close trajectories (of the original mapping (5.1)) computed simultaneously. In the work [39] the initial $l_0 = 10^{-7}$ and the full motion time $t = 10^4$ were used.

According to the data in table 5.1 the quantity h_1 is fitted fairly well for $K > 4$ by the simple relation (5.41). The mean ratio $(h_1/h_T) = 0.9998$ over the interval $K = 25$–1000. This shows, in particular, that the stochastic component of the motion really fills up homogeneously all the phase square of the reduced system (see also section 5.5). At $K = 3$ a better agreement with the measured value of the entropy is given by a more complicated expression (5.40): $h_T = 0.613$. For still lower K the latter is getting invalid as well due to the large domains of stability formed, in the first place, inside the integer resonances (fig. 5.2).

For $K < 1$ the stochastic component occupies only narrow layers along the resonance separatrices. This case will be considered in detail in section 6. Let us just mention now that the quantity $h_s = h_1 T_a$ is approximately constant over this region according to the data in table 5.1. Since T_a is the mean half-period of oscillations inside a stochastic layer the product $h_1 T_a$ gives KS-entropy per this half-period near a resonance separatrix. The variations of h_s within the interval $K = 0.2$–1.3 amount to about 20 per cent only even though the value of h_1 changes there by a factor of 10. The values of T_a in this interval of K were calculated numerically and proved to be in a good agreement with the analytical relation (6.18) (see table 6.1). Curious enough, that if one calculates h_s for all the values of K in table 5.1 it turns out that h_s changes less than by a factor of 2 for h_1 varying by 300 times! This is especially surprising if one recalls that for $K > 4$ the stochastic layers are no longer of any sense because there are no more stability domains inside the integer resonances (fig. 5.9).

Certainly, all the values of h_1 given are related to the stochastic component which occupies only a part of the phase square, and just a small part for $K < 1$. In the domain of stability $h \equiv 0$. However, the measurement of h_1 by the method described above always yields some finite values. For $K = 1.3$; $t = 10^5$ and the initial conditions: $\theta_0 = \pi$; $I_0 = 1$, for example, the measured "entropy" $h_1 = 7.53 \times 10^{-5}$ (cf. the entropy of stochastic component $h_1 = 0.227$, table 5.1). The finite value of h_1 is related to the

fact that close trajectories are diverging even in the stable domain due to a difference in frequency which depends in the nonlinear system on initial conditions. The divergence of close trajectories in the stability domain is known to be linear with time* (see, e.g., ref. [76]). Hence the measured value of "entropy" h_1 will decrease roughly in inverse proportion to the motion time: $h_1 \sim (\ln t)/t$ (5.39). This peculiarity permits us to discern a stable motion from a stochastic one. For the standard mapping, e.g., at $K = 0.5$ and initial conditions: $\theta_0 = 0.3$; $I_0 = 0$ the measured values of "entropy" turned out to be: 1.14×10^{-4}; 1.18×10^{-5} for $t = 10^5$; 10^6, respectively ($h_1 = 6.86 \times 10^{-2}$ inside the stochastic layer, table 5.1).

The linear divergence of close trajectories was frequently used as an empirical criterion of the stability of motion for nonlinear oscillator systems. In particular, just by means of that method Ford, Stoddard and Turner discovered [95] the absolute stability, or the complete integrability, of the Toda lattice – a highly nonlinear many-dimensional system [108].

5.3. Local instability and stochasticity

A strong local instability of motion, i.e. a rapid (exponential) divergence of close trajectories, is not only of an intrinsic importance as one of the basic characteristics of motion but it has also important consequences. It turns out that such an instability results in an irregular motion of the system which may be called, in a sense, random or stochastic. The irregularity of motion, lack of the time correlations leads, on the one hand, to an inexhaustible diversity and intricacy of the dynamical picture of motion, of its particular trajectories, in a stochastic system. On the other hand, however, this same irregularity as if levels all the trajectories on an average and permits, thus, a fairly simple statistical description of such a system in terms of the average quantities and probabilities.

It is worth noting that such a motion has been searching for since long ago with the purpose of foundation of the statistical mechanics. However, subsequent upon the first poor attempts of the classical ergodic theory to derive the statistical properties from the dynamical laws an erroneous idea has taken root that the intricacy of motion is determined ostensibly by the complexity of a system and, in the first place, by the number of its degrees of freedom. Meanwhile, it turns out that the strong local instability of motion is capable of giving rise to an extremely intricate motion even in a system as simple as the system (5.1) having only 1.5 degrees of freedom altogether. An idea of that intricacy can be gained, for example, from fig. 5.2.

Statistical properties of a mapping similar to the standard one were studied in some detail in the work [39]. Let us mention also the classical example by Sinai [11] who has proved rigorously the stochasticity of motion in a conservative (closed) system having as few as two degrees of freedom. Such a system may be, for example, just a ball sliding over a billiard-table and bouncing from its closed rightangular rim and from some *everywhere convex* boundary inside, say, from a fixed cylinder. What may be still simpler? In the latter example the stochasticity of motion is subsequent upon bouncing from a convex boundary that leads to a scattering of trajectories and, hence, to the local instability of motion**.

Now why does the local instability result just in a statistical behavior? The intrinsic development of

*Now it is just high time to clarify the notion of trajectory. The latter is understood not so much as a path the system follows but rather as a process of motion along a motionless path, or an orbit. For the problem under consideration, for example, there is a divergence of adjacent trajectories linear in time but none of the orbits.

**This sort of "billiard problems" was discussed by many authors including Hadamard [143] and Birkhoff [17].

that instability is regular, isn't it? Just look at the regular trajectories (5.35) of the tangent mapping (5.33) describing the local instability of a stochastic system (5.32). The point turns out to be in fact that a domain of the stochastic motion is always bounded in phase space. Thus, in Sinai's example the motion of a ball is bounded by the external rim of the billiard-table as well as by energy conservation. In the case of standard mapping the phase interval $(0, 2\pi)$ is limited owing to the periodicity of perturbation ($K \sin \theta$), and so does the system (5.32) as well. In other words the stochastic motion is always oscillatory (in a broad sense of this word), i.e. a recurrent motion, at least, in some of the dynamical variables.

Under the local instability close trajectories diverge after all up to a distance comparable with the size of the whole domain of motion. It is just the moment when the mixing begins to result in the dying-out of the time correlations with initial conditions, that is the system forgets its initial state. As soon as the divergence of trajectories is exponential the system forgets the initial conditions very quickly and for ever. In other words, the "stochastic memory" is short ($\sim 1/h$) and it gets lost abruptly.

One may consider the mechanism of mixing in another way. The results of Sinai's work [75] permit us to conclude that a stochastic system possesses an everywhere dense set of periodic trajectories, the number $\nu(T)$ of the trajectories with period $\leq T$ fitting the asymptotic ($T \to \infty$) estimate:

$$\nu(T) \sim e^{hT}. \tag{5.43}$$

The meaning of this estimate is quite simple: any trajectory of a stochastic system can be "closed" into a periodic one by a slight change of the initial conditions, or by a displacement in phase space, due to a rapid divergence of adjacent trajectories. Since the mechanism of local instability is immediately applicable only to very close trajectories, a finite "deformation" of the trajectory necessary to "close" it into a periodic one has to be done via a series of successive transitions between the adjacent trajectories. Such a *transition chain* has been invented by Arnold and applied by him to a different problem [5] (see also section 7.1). The required displacement is the smaller (exponentially smaller) the longer is the time of motion around a "closed" trajectory. This leads to an exponential increase in the number of periodic trajectories with their period (5.43). Since the periodic trajectories are unstable, the other, non-periodic, trajectories as if scatter by the former. Such a very vivid picture of the mixing in a stochastic system is due to Ford [76]. This sort of phase motion resembles the motion of a small shot on the Galton Board [123]. Its simplest version is just a wooden plank fixed vertically with many pins stuck squarely into it. A shot is falling through the "forest" of pins scattering by them "randomly". This classical device for a visual demonstration of random processes is a typical system with an exponential local instability of motion (compare with Sinai's "scattering billiard-table" described above).

Using the above mentioned method of "trajectory deformation" by means of a transition chain one can proceed much further. For example, the trajectories linking any two prescribed domains of finite dimensions can be found provided the time of motion is long enough. Thus, we arrive at the topological Markovian chain, that is to say, the motion of such a system will be qualitatively similar to a Markovian random process. The latter approach has been widely used by Alekseev in his theory of quasi-random dynamical systems [77]. To apply a quantitative, probabilistic description of the stochastic motion the latter needs to conserve some measure, e.g. the phase space volume for Hamiltonian systems. Then a topological Markovian chain can be converted into a standard probabilistic Markovian chain, that is the system behaves completely as a "random" one.

Finally, one can find such trajectories of a stochastic system which go over any prescribed

sequence of the (finite) domains in phase space, including infinite and non-periodic ones [77, 78]. The presence of such trajectories naturally follows from the statistical description and, in particular, from the existence of the transition probabilities for a Markovian chain. All this emphasizes again the wealth and diversity of the dynamical picture for a stochastic motion.

Let us mention still another approach to the understanding of statistical properties of the dynamical motion [79]. In principle, any two trajectories, arbitrarily close to each other, respresent completely different dynamical motion of a system. Indeed, any trajectory is determined by the initial conditions, i.e. by a set of real numbers. In the exact formulation of the classical mechanics each of those real numbers contains, borrowing the term of the information theory, infinite information on the future (and the past) motion of the system. Equally, the difference in the information for two distinct trajectories is also infinite for any finite distance between them. However, for the stable motion, which has been the main, not to say the only, subject of study in mechanics until recently, all that potential diversity of motions is not realized, and the behavior of a system is monotonous (quasi-periodic!) and dreary with only a weak and smooth dependence on the initial conditions. Yet, under the local instability (especially exponential one) all those hidden microscopic details of the initial conditions lead, in the course of time, to an inexhaustible diversity of motions even in the system as simple as one of the type (5.1).

Apparently the first* who clearly understood a link between the instability of motion and the statistical behavior of mechanical systems was Poincaré [81], he even made some estimates for molecular collisions in a gas. Smoluchowski [89] expressed the condition for applicability of the probabilistic conception to a physical system by the phrase: "little causes, big effects". Mises [83] was of a similar view. The first mathematical theory rigorously linking the instability and ergodicity of motion has been developed by Hedlund and Hopf for a somewhat specific case – the geodesic flow on a surface of everywhere negative curvature [80]. This line of research has been pursued by Anosov who introduced an important general notion of the C-system and thoroughly studied the peculiarities of its motion [15]. Roughly speaking, a distinctive feature of the C-system is a *homogeneous* exponential local instability of motion, that is the instability whose rate is limited from below and, moreover, homogeneously in the initial conditions for both the main phase space of a system and the tangent space of linearized equations of motion. The term C-system is related just to the demand for a system to fit that special condition. For example, the mapping (5.32) is a C-system (off the interval (5.36)) since the eigenvalues $\lambda_{\pm}$ depend only on k but not on the initial conditions. Yet, the standard mapping (5.1) we are interested in is not, apparently, a C-system for any K because it seems always to have some domains of stable motion (section 5.5).

Somewhat weaker conditions are determined a central in the modern ergodic theory notion of the K-system introduced by Kolmogorov [13] (under another name of the quasi-regular system). Again oversimplifying the situation, one may say that a distinctive feature of the K-system is a positive (non-zero) KS-entropy of motion for any initial conditions, that is the non-zero *mean* instability rate. Owing to the presence of stability domains the standard mapping is not a K-system either.

Both K- and C-systems (with an invariant measure) possess the full set of statistical properties known so far: ergodicity, mixing, positive KS-entropy and the continuous spectrum. We are not going to discuss all these properties** but will, instead, consider in the next section a more graphic process of diffusion for the standard mapping.

*See, however, [123] and above.

**A formal presentation of the modern ergodic theory may be found in the excellent book by Arnold and Avez [2]; an adapted representation "for pedestrians" is contained, e.g. in the review article [9]. Recently, a new and the strongest statistical property of dynamical motion has been discovered, namely, the so-called Bernoulli property (see, e.g., ref. [145]).

Let us touch upon, in conclusion, an interesting question: to what extent the stochastic motion of a dynamical system can be considered as a "genuine" random process? Some researchers flatly object to such a possibility repudiating it by definition, saying: if there are exact equations of motion no true randomness is possible. Indeed, despite a strong local instability, resulting in a fast mixing of trajectories, some correlations between an instant state of the system and the initial conditions always remain, if negligibly small, yet of principal importance, the correlations which permit to restore, in principle, all the past motion. Many people will apparently agree that our intuition is against such correlations in the "genuine" random process. Recall, for example, the irregularity postulate for a random sequence introduced by Mises [83]. But may intuition mislead us? Could it be that in Nature there are no such "genuine" random processes as we fancy them? In my opinion, presented in more detail in ref. [79] it is not excluded that no "more random" processes than, for instance, the motion of a K-system do really exist.

5.4. Diffusion

The diffusion is a distinctive random process. So if we consider motion of the standard system (5.1) for $K > 1$ at least, as being similar to a random one, a diffusion in momentum I must occur. The diffusion will develop just in I since the motion is unbounded in this direction. As to the phase θ its variations are limited by the interval $(0, 2\pi)$, and for $K > 1$ all this interval is passed through in a few iterations.

Variations of I are determined by the sequence of phase values $\theta(t)$:

$$\Delta I = K \sum_t \sin \theta(t) \tag{5.44}$$

where t is an integer time, or the serial number of the iteration. The function $\theta(t)$ depends, in turn, on the behavior of the reduced system (5.2) within its unit phase square, or that of the standard system within the phase square $2\pi \times 2\pi$ (see the beginning of section 5.1). If the stochastic component of motion were to occupy all this square the distribution of θ over the interval $(0, 2\pi)$ would be homogeneous due to ergodicity of the motion. The numerical experiments already mentioned in section 5.2 suggest that it is nearly so for the sufficiently large $K \gg 1$. Moreover, since the rise time of local instability and, hence, also the correlation relaxation time is $\sim 1/h \sim 1/\ln(K/2)$ (see eqs. (5.37) and (5.41)) even the values of θ for adjacent iterations are almost independent. In short, we can assume approximately that the phase values θ are random and independent for both the successive iterations and different initial conditions. Let us call this approximation the *limiting stochasticity*, it is also the simplest one. Then from eq. (5.44) it follows immediately that:

$$\overline{(\Delta I)} = \langle(\Delta I)\rangle = 0; \qquad \overline{(\Delta I)^2} = \langle(\Delta I)^2\rangle = \tfrac{1}{2} t K^2. \tag{5.45}$$

Here, as usual, the bar and the angle brackets denote the time mean and the phase space mean, respectively. Both are equal due to the ergodicity, the phase space mean being reduced in the case under consideration to the averaging over θ (5.44). Relations (5.45) show that the motion in question is a diffusionlike process with the rate:

$$D_{\mathrm{T}} = \langle(\Delta I)^2\rangle / t = \tfrac{1}{2} K^2. \tag{5.46}$$

This rate was measured numerically by averaging over 100 trajectories with various initial conditions for every of 180 values of K within the interval (10–1000). The least square fit of the data

with a power dependence $D(K)$ gives:

$$D_E = K^{1.982}/1.866. \tag{5.47}$$

That is very close to the asymptotic ($K \gg 1$) theoretical relation (5.46), especially as to the value of the power index. Nevertheless, it should be noted that the root-mean-square deviation of D_T from D_E over the K interval given is equal to $\sqrt{\langle (D_T/D_E - 1)^2 \rangle} \approx 0.056$ and exceeds considerably the bare statistical error expected: $\sqrt{2/(100 \times 180)} \approx 0.011$. This discrepancy shows that the simplest assumption of the limiting stochasticity holds only approximately even at an average, with a fairly good accuracy though. Let us consider this question in more detail.

Fig. 5.3 reveals some periodicity in the dispersion of experimental points. Such a periodical variation of D_E with K goes as far as up to $K \sim 100$. The variation turns out to be of a period close to 2π. It is clearly seen in fig. 5.6 where the dependence of normalized diffusion rate D_E/D_T versus the fractional part $\{K/2\pi\}$ is plotted. A possible explanation of this curious phenomenon will be discussed in section 5.5.

A more delicate test concerning statistical properties of the standard mapping is the examination of distribution function $f(\Delta I, t)$ where ΔI is a change in momentum over the motion time t. Relations (5.45) imply the normal (Gaussian) distribution:

$$f(\Delta I, t) = \frac{\exp(-(\Delta I)^2/2\langle(\Delta I)^2\rangle)}{\sqrt{2\pi\langle(\Delta I)^2\rangle}}; \qquad \langle(\Delta I)^2\rangle = \tfrac{1}{2} t K^2. \tag{5.48}$$

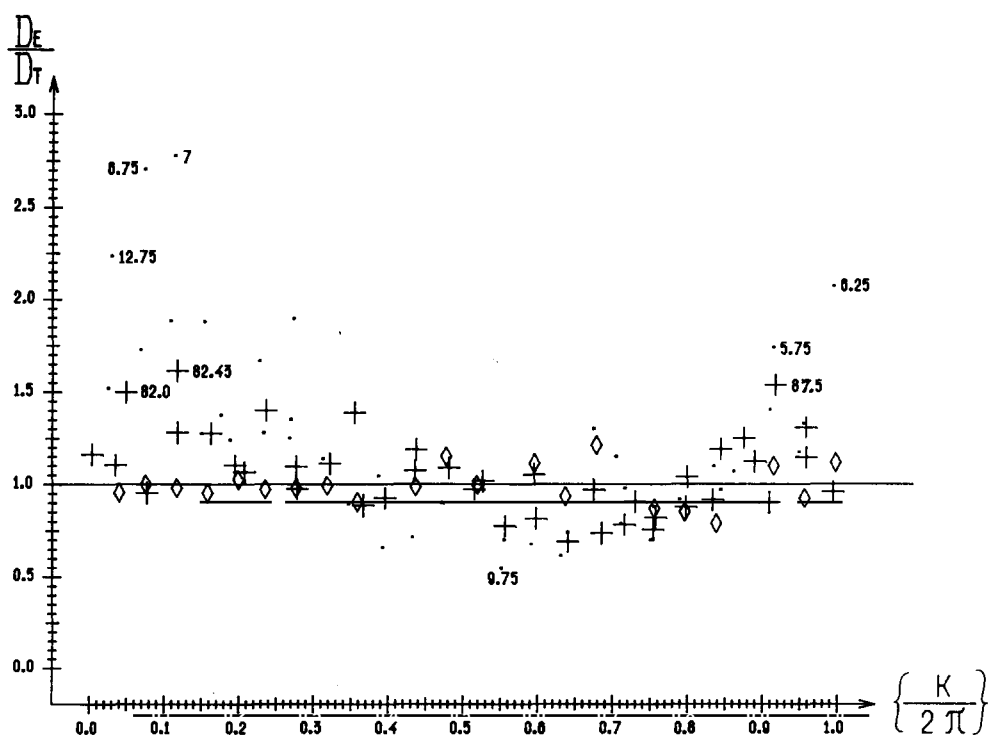

Fig. 5.6. Periodical variation of the diffusion rate with perturbation parameter K: ● – $K = 4$–20; + – $K = 80$ – 100; ◇ – $K = 980$ – 1000; $\{K/2\pi\}$ – fractional part of $K/2\pi$; statistical accuracy of D_E/D_T measurement about ±15 per cent; numbers at some points give K values.

For comparison with the numerical results it is convenient to introduce the normalized distribution function:

$$f_n(|\Delta I|) = 2f\sqrt{\pi t K^2} = e^{-E}; \qquad E = (\Delta I)^2/(t\,K^2). \tag{5.49}$$

In fig. 5.7 an example of the distribution function obtained by two methods is given for $K = 5$. According to the first method the distribution was plotted using the computation data of 10^5 trajectories with random (over the phase square $2\pi \times 2\pi$) initial conditions and for the motion time $t = 100$ iterations (symbols ∇ in fig. 5.7). The second method utilized the data of a single trajectory computed over the motion time $t_m = 10^7$. The trajectory was subdivided into time intervals of the length $t = 10^2$; 10^3; 10^4, and the distribution of values $|\Delta I|$ over each interval was calculated. In both methods the width of a distribution bin $(\delta I)/K\sqrt{t} = 1/10$ was used, and the quantity E was calculated at the center of a bin.

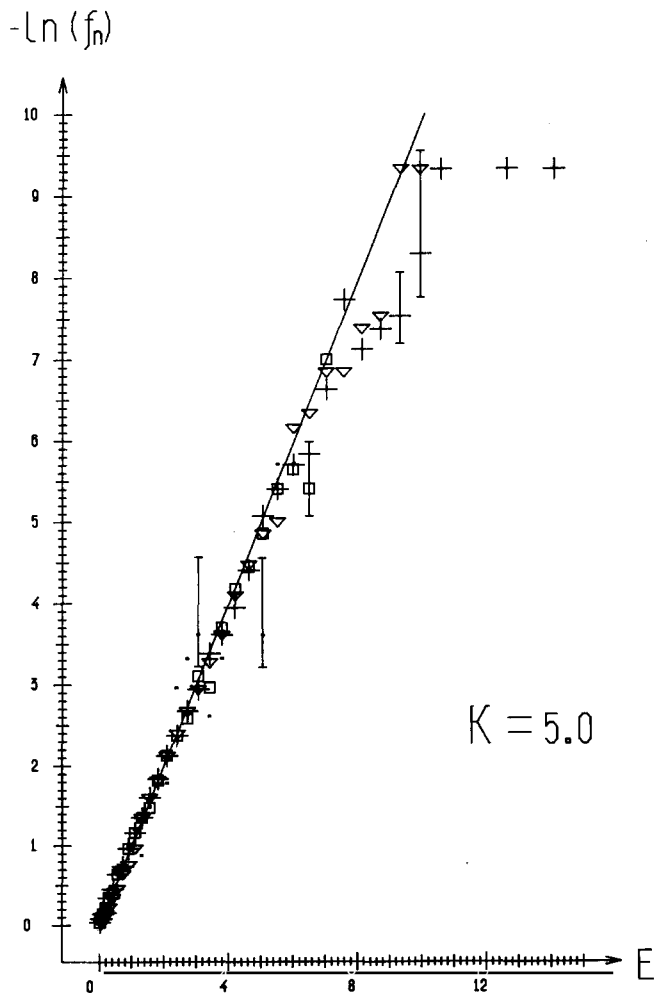

Fig. 5.7. Normalized distribution function (5.49) for $K = 5.0$; $\nabla - 10^5$ trajectories with random initial conditions, motion time $t = 100$; a single trajectory over 10^7 iterations: $+$ – motion intervals $t = 10^2$, $\square - t = 10^3$, $\bullet - t = 10^4$ (see the text); straight line is theoretical dependence (5.49); all errors are barely statistical.

In the whole range of the distribution function variation (of about 4 orders of magnitude) a satisfactory agreement with the theoretical distribution (5.49) (solid line in the figure) is generally observed. The statistical errors shown for some points give an indication of the accuracy in measuring f_n. Within the errors the normalized distribution function does not depend on the motion time t. It is worth noting, however, that the distribution "tail" goes, at the average, above the theoretical dependence, that is the actual probability of large fluctuations in the diffusion exceeds the expected one.

The summary data concerning distribution function are given in table 5.2. In the two cases marked by asterisk the data were obtained from a single trajectory and in the rest it was done from 3×10^5 trajectories with random initial conditions as has been explained above. In all cases the motion time was $t = 100$ iterations.

The parameters of the distribution function were calculated as follows. The least square fit of the computed f_n was made using the function

$$-\ln f_n = A + B\,E. \tag{5.50}$$

The diffusion rate D_f normalized to the theoretical value (5.46) and the share of diffusing component w_D were calculated from the relations:

$$f_n = w_D\sqrt{\frac{D_T}{D_f}}\exp\left(-\frac{D_T E}{D_f}\right); \qquad \frac{D_f}{D_T} = \frac{1}{B}; \qquad w_D = \frac{e^{-A}}{\sqrt{B}}. \tag{5.51}$$

The two fits were made: 1) weighted in proportion to f_n; 2) unweighted (or equally weighted). The first fit stresses an initial interval of the distribution function ($E \leqslant 1$) where f_n is large. A discrepancy between the two fits indicates a deviation of the distribution from the normal one. Note that in all cases in table 5.2 the D_f values of the weighted fit are smaller if only a bit. This indicates a slower decrease of the distribution "tail" as compared to eq. (5.51). In other words, the probability for the system to gain an energy much in excess of the mean value is increased in regard to the expected distribution (cf. fig. 5.7). The quantity w_D in table 5.2 corresponds to the weighted fit as a more accurate one. The share w_D differs from unity for smaller K. This is related apparently to the presence of stability domains inside which there is no diffusion at all.

Table 5.2
Parameters of the distribution function

K	$\left\{\frac{K}{2\pi}\right\}$	D_f/D_T unweighted	D_f/D_T weighted	$\frac{D_E}{D_T}$	w_D
988.457	0.32	1.023	0.9997	—	0.9991
96.016	0.28	1.085	1.078	—	1.0004
*96.016	0.28	1.161	1.091		1.0004
92	0.64	0.850	0.842	0.675	1.001
82.434	0.12	1.198	1.174	1.600	1.0003
5.0	0.80	1.073	1.028	0.710	0.971
* 5.0	0.80	1.196	1.034	0.710	0.992
4.9	0.78	1.453	1.082	—	0.994
4.25	0.68	1.113	1.059	1.303	0.998
4.0	0.64	1.013	0.865	0.741	0.972
3.8	0.60	1.093	0.753	—	0.971
3.75	0.60	0.814	0.673	—	0.985

Summarizing one can say that the results of numerical experiments confirm the diffusionlike nature of motion for the standard mapping (5.1) under $K > 4$. The motion can be described approximately by the simple relations (5.46) and (5.48). Apparently the first such result for a similar system has been obtained by Hine [84]. Remarkably, even for a fairly large $K \sim 100$ the clear distinctions of the motion from a purely random one are observed. We shall consider this question in some detail in the next section.

For a smaller $K < 4$, near the stability border $K = 1$, the simple diffusion picture of motion is obviously not the case, mainly, due to a strong correlation in the successive values of phase θ. The correlation arises because the motion goes almost along a resonance separatrix. Nevertheless, the transitions between resonances can be considered apparently as random since the system moves inside a stochastic layer. Then for a time interval which is much in excess of the transition time between the adjacent resonances the motion can be described, as before, by a simple diffusion law. The diffusion rate in this region can be found from eqs. (5.6), (5.3) and (5.4):

$$D = (2\pi)^2/N = 0.38(K-1)^{2.55}. \tag{5.52}$$

An approximate applicability of the diffusion law for $K \geqslant 1$, and, hence, the idea of a random transition between the adjacent resonances, are confirmed by the data in fig. 5.3 (section 5.1). As is seen from this figure the results of the direct measurement of transition time N agree satisfactorily with those from the diffusion law (5.7) (cf. eq. (5.52)).

It should be emphasized, however, that the behavior of a system in the *intermediate zone* ($K \sim 1$), i.e. about the stability border, is extremely complicated. This problem was studied in some detail in refs. [26, 39]. Here we just mention a curious event which has been encountered during numerical experiments with the standard mapping. In fig. 5.8 the distribution function for $K = 3$ is presented. A striking peculiarity of motion in this case is the presence of the two quite different diffusion components. For the first one of share $w_D^{(1)} \approx 0.83$ the diffusion rate is about normal: $D_f^{(1)}/D_T = 0.683$ whereas for the second of $w_D^{(2)} \approx 0.06$ the rate materially in excess: $D_f^{(2)}/D_T = 5.71$. There is also a third component of $w_D^{(3)} \approx 0.11$ with the diffusion rate nearly zero (the points with $f_n > 1$ in fig. 5.8). For the many trajectories with various initial conditions (points in fig. 5.8) this component is quite comprehensible, and relates to the existence of stable domains. Remarkably, it is present also in the case of a single trajectory which has no possibility to enter a stable domain (daggers in fig. 5.8). It may be that the latter event is due to a very slow diffusion in an intermediate zone around the stability domains (see section 5.5). It is interesting that the peculiar structure of motion described is observed only in a narrow interval of $K \approx 2.95$–3.15.

5.5. Islets of stability

A distinct peculiarity of the standard mapping is the existence of domains with a regular motion for arbitrarily large $K \to \infty$, that is far inside the stochastic region. This problem was studied in detail in the work [39] (see also appendix in ref. [88]). The occurrence of such domains is related to the fact that for the standard mapping there always exists the stable phase interval (5.36) within which the eigenvalues of tangent mapping (5.31) are complex-conjugate and $|\lambda_\pm| = 1$ (5.34). The simplest regular motion of this type is the motion with constant phase $\theta = \theta_1$. Together with the stability condition (5.36) the parameters of such a trajectory can be found from the relations:

$$K \sin\theta_1 = 2\pi n; \qquad -4 < K\cos\theta_1 < 0 \tag{5.53}$$

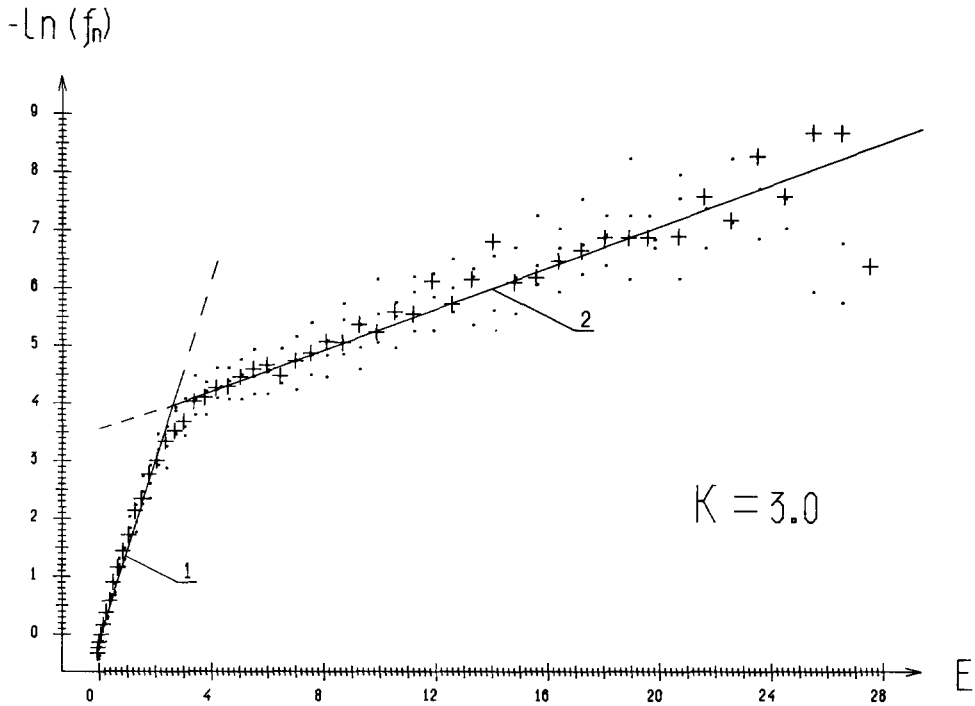

Fig. 5.8. The same as in fig. 5.7 except $K = 3$; + – a single trajectory over $t_m = 10^7$, $t = 100$; ● – random initial conditions, $t = 100$ (3 *series*); straight lines represent the least square fit according to eq. (5.50): $A = 0.01$, $B = 1.465$ (1); $A = 3.62$, $B = 0.175$ (2).

where n is any integer, zero included. Both conditions are compatible only for the values of K within the intervals:

$$(2\pi n)^2 < K^2 < (2\pi n)^2 + 16; \qquad \{K/2\pi\} < |n|(\sqrt{1+(2/\pi n)^2} - 1). \tag{5.54}$$

Remarkably, these special values of K may be arbitrarily large. For $|n| \to \infty$ the values of $K \approx 2\pi n$ increase indefinitely, and the interval width decreases as $\Delta K \approx 4/\pi|n| \approx 8/|K|$ but remains finite. If $n = 0$ the value of $I = 2\pi m = \text{const}$ (see eq. (5.1)) with any integer m, i.e. the motion is of period $T = 1$ (a fixed point of mapping (5.1)). Around each of those fixed points there is a stability domain – the central part of an integer resonance (section 5.2). This stability domain is destroyed for $K > 4$ (5.54), yet, for $K > 2\pi$ a new stability domain ($n = 1$) arises and so on. The character of motion in the domains related to $n \neq 0$ is quite different as compared to those with $n = 0$, namely:

$$I = I_0 + 2\pi n t; \qquad I_0 = 2\pi m, \tag{5.55}$$

i.e. I is changing monotonically with time. The relation holds exactly for the central trajectory with a constant phase $\theta = \theta_1$, and does so at an average for neighbouring trajectories. Recently such an "accelerator" regime of motion was studied for a similar mapping in ref. [85]. Generally, this sort of motion is known since long ago, and is of basic importance for the performance of an electron accelerator – the microtron (see, e.g., [86]). Perhaps, it is not appropriate to call the motion (5.55) a stable one since the variation of I is unbounded, it would be better to speak about a regular motion. Actually we mean the stable oscillations of phase θ around the fixed value θ_1. Besides, if we change over from the standard (5.1) to reduced (5.2) mapping the trajectories (5.55) become periodic ($T = 1$).

Bearing all this in mind we retain the former term – the domains, or *islets of stability*. The latter term emphasizes small dimensions of the islets for $K \gg 1$.

The size of an islet of stability can be estimated for $K \gg 1$ in the following way [39]. The size in θ is determined by the stable phase interval (5.53): $\Delta\theta \sim 4/K$. The size ΔI must be of the same order not to drive the phase θ out of the interval (5.53). Since the structure of the phase plane for the standard mapping is of period 2π in I the relative area of a stability domain

$$\sigma \sim (\Delta I)(\Delta\theta)/(2\pi)^2 \sim 1/K^2. \tag{5.56}$$

Let us consider also trajectories of period $T = 2$ (in phase θ). We introduce notation: $k_i = K \cos \theta_i$. For the problem in question this quantity has only two values k_1, k_2 corresponding to the periodical oscillations of the phase: $\theta_1 \rightleftarrows \theta_2$. Multiplying the two matrices of tangent mapping (5.31) with $k = k_1$, k_2 we get matrix for the period $T = 2$. The stability condition ($|\lambda_\pm| = 1$) is deduced similarly to that for the condition (5.36) and has the form:

$$-4 < k_1 k_2 + 2(k_1 + k_2) < 0. \tag{5.57}$$

Note that in the event $k_1 = k_2$ this condition is reduced to the former (5.53) for $T = 1$, and it remains so for any T. Indeed, if $k_1 = k_2 = \cdots = k_T$ the matrix of the tangent mapping over period T is the same as for the motion with a constant phase θ.

From the periodicity condition (in θ) we have (see second eq. (5.1)): $I + \bar{I} = 2\pi n$ with any integer n, or more briefly: $I + \bar{I} = 0 \pmod{2\pi}$. If the trajectory is periodic also in I then $\sin\theta_1 + \sin\theta_2 = 0$, and, hence, there are only two possibilities as to the phase values:

$$\theta_2 + \theta_1 = 0 \pmod{2\pi}; \qquad k_1 = k_2 \tag{5.58a}$$

$$\theta_2 - \theta_1 = I_2 = \pi \pmod{2\pi}; \qquad k_1 = -k_2. \tag{5.58b}$$

Like θ the momentum I has also only the two values, and $I_1 + I_2 = 0 \pmod{2\pi}$. Since $I_2 - I_1 = 2I_2 = K \sin\theta_1 \pmod{2\pi}$ and $I_2 = \theta_2 - \theta_1 = -2\theta_1 \pmod{2\pi}$ (in case (5.58a)) we arrive at a transcendental equation for the phase θ_1:

$$K \sin\theta_1 = 2\pi n - 4\theta_1. \tag{5.59}$$

Together with the stability condition (5.53) it determines the intervals of K values within which the periodic motion under consideration exists surrounded by some stable area. The size of this area is estimated by the former expression (5.56) since the stability condition remains as before (5.53).

In the second case (5.58b) $k_1 + k_2 = 0$, and stability condition (5.57) yields:

$$|K \cos\theta_1| < 2. \tag{5.60}$$

In combination with the relation $2I_2 = K \sin\theta_2$ and eq. (5.58b) it leads to the inequalities similar to those (5.54):

$$(2\pi n)^2 < K^2 < (2\pi n)^2 + 4; \qquad \{K/2\pi\} < |n|(\sqrt{1 + 1/\pi^2 n^2} - 1). \tag{5.61}$$

Besides the purely periodic trajectories considered there are also "accelerator" trajectories of period $T = 2$ when only motion in θ is periodic whereas the variation of I is unbounded (cf. eq. (5.55)). In this case the condition $I + \bar{I} = 2\pi n$ implies $K\ (\sin\theta_1 + \sin\theta_2) = 2\pi l$ with an integer $l \neq 0$, and, hence, there are much more possibilities for the values of θ_1, θ_2 as compared to eqs. (5.58), the more the larger K is.

The analysis of periodic trajectories becomes progressively complicated as the period grows. Some rough estimates are given in ref. [39] (see also ref. [43]).

An example of the islets of stability around a trajectory of period $T = 2$ is presented in fig. 5.9 for the reduced mapping (5.2) with $q = 2$ and $K = 5$. The periodic trajectory inside the islets is of the type (5.58a) with $\theta_1 = -\theta_2 \approx 2.02$, and it corresponds to $n = 1$ in eq. (5.59). This θ_1 value lies near the center of stable interval (5.53): $K \cos \theta_1 \approx -2.17$. The relative area of the stable component measured directly on fig. 5.9 and also by another method (see below) amounts to about 1.5 per cent. Besides the four comparatively large islets of stability the many (34) little spots missed by a trajectory are seen in fig. 5.9. These are not, however, necessarily related to the additional islets of stability. On the contrary, they are an inevitable corollary of the random motion. According to the Poisson distribution the expected number of randomly missed bins for the total amount of bins $N_0 = 128 \times 128$ and during $t = 10^5$ iterations is equal to $N_m = N_0 \exp(-t/N_0) \approx 37$.

The islets of stability are apparently the cause of a periodical variation of the diffusion rate with K as has been described in section 5.4 (see fig. 5.6). The influence of "accelerator" domains (5.55) is especially substantial. Even though a stochastic trajectory cannot enter those domains it does approach their very border and stays there for a long time due to a very slow diffusion near the stability border (see fig. 5.3). As a result the oddly fast variation of the momentum may occur which increases the mean diffusion rate. According to eq. (5.54) this must take place for $\{K/2\pi\} \approx 0$. The latter peculiarity is clearly seen in fig. 5.6, indeed.

A fundamental problem concerning the islets of stability is their total area. Although the size of the islets related to short-period trajectories is small and is rapidly decreasing as K grows (see eq. (5.56) and fig. 5.9) we have actually no knowledge as to the long-period trajectories, nor have we even a rough estimate for the corresponding domains of stability. Meanwhile, the number of all periodic trajectories in a stochastic system is exponentially large according to Sinai's estimate (5.43). It is true, that most of them are unstable in all likelihood, yet some may happen to be stable as we know from the above consideration of the short-period trajectories. So a misgiving that the total stable area may be much larger than it seems to be from a simple estimate like (5.56) or from a phase map as in fig. 5.9 is not groundless. Similar indications follow also from abnormalities of the diffusion process even for the K values as large as 100 (section 5.4, table 5.2). May it turn out that the measure of the whole

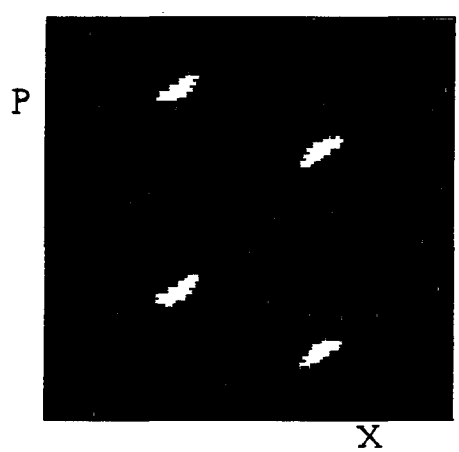

Fig. 5.9. Phase map of system (5.2) for a single trajectory: $K = 5$; $q = 2$; motion time $t = 10^5$; resolution: 128×128 bins.

stable component of the motion is close to unity, and the measure of the complementary stochastic component is nearly (or even equal exactly to) zero?

To clarify this question the measurements of the relative shares for both the stochastic and stable components have been carried out in the case of the reduced mapping (5.2) with $q = 1$. Two methods were used. In the first one the phase square was subdivided into 100×100 bins, and the number N_s of the bins crossed by a trajectory during t_s iterations was counted. The relative area of the stochastic component was assumed to be $\sigma_s = N_s/10^4$. The motion time t_s was being increased until σ_s reached some limit. The values of σ_s for $K = 1$–5 are given in table 5.3. It is seen that for $K \geqslant 5$ the stochastic component occupies almost all bins of the phase square, that is the trajectory crosses nearly all the bins provided the motion time is long enough. However, the size of a bin is really not so small, and a doubt arises if the structure of the stochastic component may be so intricate that the latter penetrates all the bins in spite of a small (zero?) total area? Why may it not consist, for example, of very thin but densely located layers?

To make clear the situation a different method was used. For each of the N_r initial conditions chosen randomly over the phase square a trajectory was computed during time interval t_r and the value of KS-entropy h_1 was measured by the method described in section 5.2. All the entropy values obtained for a given K were sorted into the 20 intervals, and the histogram n_i was plotted, n_i being the number of values within an interval of the serial number $i = 1, \ldots, 20$ proportional to the entropy value h_1. An example of the histogram is given in fig. 5.10 for $K = 2$; $N_r = 100$; $t_r = 10^4$. The two components of motion are clearly seen, namely, the stochastic peak (77 per cent) and stable trajectories (19 per cent) with $h_1 \approx 0$. There are also 4 trajectories with an intermediate entropy as a result of the fluctuations of local instability, in all likelihood. The relative area of the stable component was assumed to be $\sigma_r = n_1/N_r$. The values of σ_r are also given in table 5.3. They agree satisfactorily with the values of σ_s ($\sigma_s + \sigma_r \approx 1$). We see that the stable component of the motion nearly vanishes for $K > 5$.

The decisive advantage of the last method for measuring the share of stable component is

Table 5.3
Measure of stochastic component for the reduced mapping (5.2)

K	σ_s	σ_r
8.888	—	$<10^{-4}$
7.701	—	10^{-4}
6.59	—	0.0099
6.28	—	0.0025
6.21	—	0.004
5	0.98	0.014
4	0.92	0.08
3	0.89	0.11
2	0.79	0.19
1	0.44	0.52
0.5	—	0.96

σ_s, σ_r – relative area of stochastic and stable components of motion, respectively.

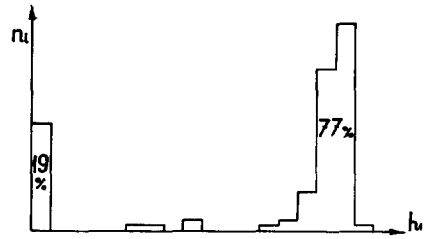

Fig. 5.10. Histogram of the KS-entropy for 100 trajectories; motion time $t = 10^4$; $K = 2$.

related to the fact that the set of all possible initial conditions is incomparably more rich than those 10^4 bins of the phase square used in the first method. In some sense the transition to the random initial conditions is equivalent to the increase of the amount of bins in the first method up to the total number of all different initial conditions ($\sim 10^{24}$ for our computations). This enormous number allows us to hope, in my opinion, that the results of the σ_r measurement really confirm a negligible area occupied by the stable component of the motion for $K \gg 1$. Nevertheless, the problem in question cannot be considered as completely solved. One of the reasons is related to the principal limitation of the numerical experiments, namely, the limitation connected with the discreteness of the computer representation for any quantity. In other words, any number in the computer is always integer in essence. Particularly, we have always to deal with a discrete phase space in the numerical experiments. No matter how small such a "quantum" of the space may be it certainly influences the motion under an exponential local instability. Only a few papers have been devoted so far to this specific problem in the numerical experiments, a paper by Rannou [91] among them* (see also ref. [43]).

Summarizing the above considerations we can say that for a sufficiently large K the standard mapping becomes "almost" a K-system, i.e. its motion allows a simple statistical description (section 5.4). The word "almost" means here the exclusion of the small but finite domains of regular motion – the islets of stability. Therefore, such a system is, strictly speaking, neither a K-system nor, still more, a C-system in the sense of the modern ergodic theory. This is just a reason for various "abnormalities" of its motion and, particularly, of the diffusion process (section 5.4).

What is more serious is that a system with the islets of stability is structurally unstable. This means that the structure of its motion may qualitatively change under the influence of an additional arbitrarily small perturbation. A weak dissipation is an example of such a perturbation. This question was studied in ref. [39] for a mapping similar to the standard mapping. Watching motion of the system by a display we always observed a "degeneration" of stochastic motion into a periodical one due to the "capture" of a stochastic trajectory into some stability domains or other. The "lifetime" of stochastic component grows, naturally, as dissipation decreases, yet the capture onto a periodic orbit does occur after all. It is worth noting that for a system with 2.5 degrees of freedom of the type of two coupled oscillators under an external periodic perturbation we failed to observe the capture for a sufficiently strong coupling [90].

Although the standard mapping is certainly not a structurally stable system in the common sense of the term, i.e. for the arbitrary perturbation, it seems to be structurally stable if we restrict the permissible class of perturbations by the canonical perturbations only. All our numerical experiments indicate that. Since the simple standard mapping describes real physical systems always only

*In a number of works a similar problem for the so-called pseudorandom number generators – usually one-dimensional and not area-preserving mappings – was studied, see, e.g., [124].

approximately it were of great importance to develop a rigorous theory of such a *restricted structural stability*.

The appearance of the islets of stability is related, as has been shown above, to the existence of stable phase intervals (5.53). The latter appear, in turn, for a perturbation $V(\theta)$ sufficiently smooth and periodic in phase θ. It suffices for the derivative $V''(\theta)$ to be continuous, or for the $V(\theta)$ to have the smoothness parameter $l \geqslant 2$ (section 4.5). Subject to this condition and due to the periodicity of $V(\theta)$ the derivative $V''(\theta)$ necessarily crosses zero that leads to the formation of stable phase intervals (5.53). It is interesting to mention that the condition for the presence of the islets of stability as $K \to \infty$ ($l \geqslant 2$) almost coincides with the condition for the stability of the motion as $K \to 0$ ($l > l_c = 2$, see the end of section 4.5).

Thus, the structure of motion with the islets of stability is a typical one for the nonlinear oscillations. We call this structure the *divided phase space*, divided into a regular and a stochastic component of motion, the regular component possessing the full set of the integrals of motion and being stratified into the invariant tori supporting a quasi-periodic motion. Note that the confinement of the stochastic motion to some domain of the phase space may be interpreted as the influence of some weak conservation law. It restricts the motion not to a certain subspace of smaller dimensionality, as the usual conservation law does, but only to a part of the space with the full dimensionality.

The structure of the divided phase space is always extremely intricate since the border between the regular and stochastic components has to be always a stochastic curve (surface). That intricate structure is graphically displayed, for example, in refs. [82, 47, 39] (see also fig. 5.2). Some researchers [76, 78] call this structure a pathology. As to me, I admire the wealth and diversity of motion in even extremely simple dynamical systems.

6. Stochastic layer

In this section the behavior of the standard mapping (5.1) for a small perturbation $K < 1$ will be considered. For most initial conditions the motion in this case is just the stable oscillations. However, as we shall see, there are domains where the motion is stochastic for an arbitrarily small K. Those domains are the so-called stochastic layers in a vicinity of the separatrix of a nonlinear resonance. Quite naturally, the motion near a separatrix must be extremely unstable since the latter separates the qualitatively different kinds of motion (oscillation and rotation), transitions between them being possible under a very weak perturbation. Hence the motion in this area is extremely intricate, or "pathological" [76]. In particular, instead of the single unperturbed separatrix, as is the case, for example, in the free oscillations of a pendulum (section 2.4), the two separatrices, or, in Arnold's language, the two whiskers interwoven in an intricate manner appear (whiskers left "uncombed"). According to a presently popular quotation from Poincaré one is struck by the complexity of the interwoven separatrix figure that he (Poincaré) himself was not even attempting to draw [6]. This has been done by Melnikov [12], and since then his drawing has gone over many articles and books on the nonlinear oscillations. Much later the whisker structure has been computed in a number of cases (see, e.g., refs. [47, 82]) and turned out to have a striking resemblence to Melnikov's sketches.

Some qualitative ideas concerning that homoclinic structure were applied to prove the non-integrability of certain dynamical equations (see, e.g., refs. [27, 107]). The qualitative theory of homoclinic structure is being developed by many researchers (see, e.g., refs. [96–98]), and it has formed by now a new section in the general theory of dynamical systems – the so-called differentiable

dynamics [74]. In particular, Smale and Shilnikov have proved the existence of the quasi-random (in a certain sense) trajectories in the homoclinic structure. This follows also from Alekseev [77]. Some interesting applications of the differentiable dynamics to the problems in the general theory of relativity has been published recently in ref. [125].

The first quantitative results in the theory of homoclinic structure have been obtained by Melnikov [12] who estimated the order of magnitude for the separatrix splitting. To the best of my knowledge, presently there is no rigorous estimate as to the full size of the domain of instability around the separatrix. One has been managed to solve this problem only in the framework of a semi-empirical criterion for the resonance overlap. It was done first in ref. [99], and later on, in a more explicit and general form, in ref. [9] (see also refs. [62, 43]). More accurate estimates for the width of a stochastic layer in the problem of the destruction of a magnetic surface have been obtained recently in ref. [69].

The behavior of a dynamical system in the stochastic layer of a nonlinear resonance was observed apparently first in the numerical experiments by Hénon and Heiles [38], and later on, by many others as well (see, e.g., refs. [99, 34, 43, 47]), in all cases for a mapping. By now the picture of irregularly scattered points which all belong to a single trajectory is already customary. The examples of relatively broad stochastic layers may be seen in figs. 5.1, 5.2.

Below we will describe the results of the numerical and analytical studies which rely upon the properties of the standard mapping (section 5). Our main task is to consider generic properties of the motion in a vicinity of the separatrix. It turns out that the motion here is much simpler and not as "pathological" after all as it may seem to be from a more scrupulous mathematical analysis.

6.1. The whisker mapping

As has been shown in section 4.4 the motion in a close vicinity of the pendulum separatrix may be described approximately by the mapping:

$$\bar{w} = w + W \sin \tau; \qquad \bar{\tau} = \tau + \lambda \ln(32/|\bar{w}|); \qquad \lambda = \Omega/\omega_0. \tag{6.1}$$

Here w is a relative energy shift from the unperturbed separatrix (section 2.4); ω_0, Ω stand for the frequency of small pendulum oscillations and of the perturbation, respectively; τ is the perturbation phase at the time when the pendulum crosses the position of stable equilibrium (we have dropped here the subscript of τ as compared to eq. (4.56)). A small parameter determining the approximation to which eq. (6.1) holds is the frequency ratio: $\omega_0/\Omega = 1/\lambda \ll 1$. In other words, we are going to consider the influence of a high-frequency perturbation on the pendulum oscillations near separatrix. A particular expression for the amplitude of perturbation in eq. (6.1):

$$W_{\mathrm{T}} = -4\pi\epsilon\lambda^2 \mathrm{e}^{-\pi\lambda/2} \tag{6.2}$$

was obtained in section 4.4 for the parametric perturbation (4.50). For a different kind of perturbation the description remains qualitatively the same as we shall see in the next section 7.

As was mentioned already there both the whisker mapping (6.1) and the relation (6.2) can be applied also to describe the motion of the standard system (5.1) near the separatrix of an integer resonance, using the parameters:

$$-\omega_0^2 = k; \qquad \Omega = 1; \qquad \lambda = 2\pi/\sqrt{K} = 1/\sqrt{k}; \qquad \epsilon = 2. \tag{6.3}$$

The whisker mapping (6.1) is the basis of our consideration of the motion in a stochastic layer. Therefore, we will check, first of all, to what accuracy does it describe the actual motion near the

separatrix. Below we shall compare the first of eqs. (6.1) with the results of numerical experiments; the checking of the second equation is postponed until the next section.

The numerical experiments were carried out with the standard mapping (5.1) for the K values in the interval (0.05–1). The initial conditions were chosen in such a way to provide the system to be inside the stochastic layer of the resonance $I_r = 0$; usually: $I(0)$, $\theta(0) \sim 10^{-3}$–10^{-7}. Note that in the computation one cannot put just $I(0) = \theta(0) = 0$ since it is an unstable one though yet still the equilibrium.

The successive values of w_i were determined at the time of the maximal approach of pendulum to the position of unstable equilibrium $\theta = 0$, that is when either $I = 0$ (pendulum oscillations) or $\theta = 0$ (the rotation). Since the motion of system (5.1) was governed by a discrete mapping the value w_i at the specified instant of time was calculated from the values of θ_i or I_i obtained by the linear interpolation of these variables between the nearest integer instants of time. The values of the perturbation phase τ_i were calculated in the same way. Assuming that the adjacent values w_i are related as follows (see eq. (6.1)):

$$\Delta w_i = w_{i+1} - w_i = W_E \sin(\tau_i + \beta) \tag{6.4}$$

one can calculate the mean values for both the amplitude W_E and a possible additional phase shift β. This was done during the computation by the three methods:

1) for each pair of successive eqs. (6.4) $(i, i+1)$ the values of W_E, β were computed and subsequently averaged over all i; $\overline{W_E} = W_E^{(1)}$;

2) the maximal value $\overline{|\Delta w_i|_{max}}$ $(= W_E^{(2)}$ for the relation (6.4)) was computed;

3) the mean square $\overline{(\Delta w_i)^2}$ $(= (W_E^2)^2/2$ for eq. (6.4) and for the homogeneous distribution of τ_i over the interval $(0, 2\pi))$ was computed.

The results of the numerical experiments indicate, first of all, that within statistical errors the phase shift $\bar{\beta} = 0$ for all values of K investigated. A typical accuracy in the determination of the mean value $\bar{\beta}$ is $\sim 3 \times 10^{-3}$ but the standard deviation for a single β value amounts up to ~ 0.3. This indicates a more complicated dependence than eq. (6.4). It is also confirmed by the data of measurement of W_E. The values of $W_E^{(1)} \approx W_E^{(3)}$, yet the values $W_E^{(2)}$ are much larger as a rule. The ratio $W_E^{(2)}/W_E^{(1)}$ grows with K from ~ 1.5 (at $K \sim 0.1$) up to ~ 2.5 for $K \sim 1$. We shall use $W_E^{(1)} \approx W_E^{(3)} \equiv W_E$ in what follows.

The ratio W_E/W_T does not depend on K, and is equal at the average to:

$$r = \langle W_E/W_T \rangle = 2.15 \pm 0.04. \tag{6.5}$$

This mean value has been obtained by 52 points with the standard deviation of 28 per cent for a single value. Thus, the experimental (actual) perturbation near separatrix of an integer resonance of the standard mapping is more than by a factor of 2 stronger as compared to the theoretical prediction (6.2). The latter has been evaluated (section 4.4) in the first approximation of the resonant perturbation theory, and it is suitable, thus, only to estimate the order of magnitude. The discrepancy (6.5) is apparently due to higher approximations. Now recall that the small parameter of the problem under consideration is the frequency ratio $1/\lambda = \sqrt{k} = \sqrt{K}/2\pi$ (6.3) which is fairly small even for the maximal value of $K = 1$ considered. Therefore, so strong influence of the higher approximations is unusual and should be ascribed to a peculiarity of the MA integral.

To elucidate the question consider the Hamiltonian:

$$H(p, q, t) = \frac{p^2}{2} + \frac{\cos\theta}{\lambda^2} + \frac{\epsilon}{2\lambda^2}\cos(\theta \pm t) \tag{6.6}$$

which describes a parametric perturbation of a pendulum and which is equivalent to the Hamiltonian (4.50) with $\Omega = 1$; $\lambda = 1/\omega_0$. Introducing a canonical transformation of variables we can decrease the order of non-resonant perturbation: $\epsilon/\lambda^2 \to \epsilon^2/\lambda^4$ but subsequent upon this same transformation the second harmonic in θ appears in the perturbation. Hence the parameter m of the MA integral increases by 2 times ($m = 2 \to 4$) that leads to an additional factor $\sim\lambda^2$ (see eq. (A.11) in the appendix). Thus, in the second approximation only an additional factor $\sim\epsilon$ is left in the perturbation. If $\epsilon \ll 1$ the effects of higher approximations can be neglected. But the standard mapping is approximately equivalent to the system (6.6) with $\epsilon = 2$; $\lambda = 1/\sqrt{k}$ (4.18). In this case the effects of the second approximation are of the same order as of the first approximation. A similar situation remains in the next approximation as well. Therefore, the exact expression for the amplitude W_T is given by a series without any small parameter. As to the convergence of this series it seems to be ensured by the gamma function in the expression for the MA integral (A.11). Yet, the evaluation of this series is not an easy task. In what follows we shall use the empirical value (6.5) as the correction due to the higher approximations.

The mapping (6.1) describes, in particular, a perturbation of the separatrix itself. Without the perturbation ($\epsilon = 0$; $W = 0$) the separatrix is a single curve described in terms of the variables w and τ by the condition: $w = \bar{w} = 0$. Under the perturbation it splits into the departing ($w = 0$; $\bar{w} \neq 0$) and the arriving ($w \neq 0$; $\bar{w} = 0$) whiskers. The maximal "distance" between the whiskers ($2|W|$) characterizes the scale of the separatrix splitting. It agrees in the order of magnitude with the estimates due to Melnikov [12].

6.2. Width of stochastic layer

In section 4.4 it was shown that linearization of the whisker mapping (6.1) in w leads to the standard mapping whose parameter may be written in the form:

$$K_s = -\lambda\, W/w_r. \tag{6.7}$$

According to the results of the numerical experiments described in section 5.1 the region of stochasticity for the standard mapping is determined by the condition: $|K| > 1$. If so, it follows from eq. (6.7) that the stochastic instability near the separatrix takes place within a layer:

$$|w| \leqslant w_s; \qquad w_s = \lambda |W|. \tag{6.8}$$

This expression is approximate because the parameter K_s in eq. (6.7) is related to a resonant value of $w = w_r$ (4.61). Let us consider this question in more detail.

First, we rewrite the whisker mapping in the form:

$$\bar{s} = s + (\sin\tau)/\lambda; \qquad \bar{\tau} = \tau - \lambda \ln|\bar{s}| + G \tag{6.9}$$

where we have introduced a new dimensionless variable $s = w/w_s$ describing the position of a system near the separatrix in the scale of the half-width w_s of a stochastic layer. The constant $G = \lambda \ln(32/w_s)$ affects only the resonant values of $s = s_r$. These are determined from the condition $G - \lambda \ln|s_r| = 2\pi n$ and are equal to:

$$|s_r| = \frac{32}{w_s} e^{-2\pi n/\lambda}. \tag{6.10}$$

It follows also immediately from eq. (4.61). Let us recall that the resonances of whisker mapping (6.9)

are the second level resonances of the original system (5.1) (see section 4.4), i.e. the resonances with the phase oscillations at a resonance (of the first level) in the system (5.1).

We set $s = s_r + \delta s$ and expand the logarithm in eq. (6.9):

$$\ln|s_r + \delta s| = \ln|s_r| + \delta s/s_r + \tfrac{1}{2}(\delta s/s_r)^2 + \cdots.$$

The standard mapping is deduced by neglecting the last term which estimates, thus, the order of the error. It is small for $|\delta s/s_r| \ll 1$. The quantity δs is determined, first of all, by the mapping itself: $|\delta s| \geqslant 1/\lambda$ that leads to the condition: $\lambda\,|s_r| \gg 1$. Introducing the new momentum $P \doteq -\lambda(\delta s)/s_r$ we get the standard mapping from eq. (6.9):

$$\bar{P} = P + K_s \sin\tau; \qquad \bar{\tau} = \tau + \bar{P}; \qquad K_s = -1/s_r. \tag{6.11}$$

This mapping describes the motion of system (6.9) in some neighbourhood δs around an integer resonance $s = s_r$. In regard to the resonance overlap and stochastic instability a minimal neighbourhood must comprise the change in P by 2π (cf. section 5.1), whence: $|\delta s/s_r| \geqslant \pi/\lambda$. Therefore, the conditions for the applicability of the standard mapping to a local description of system (6.9) may be written in the form:

$$\lambda \gg \pi; \qquad |s_r| \gg 1/\lambda. \tag{6.12}$$

Now we can improve the estimate (6.8) for the width of a stochastic layer. The former depends on a particular disposition of the resonant values s_r. For instance, if one of them has $|s_r| = 1$ the half-width s_T of the layer will exceed the value (6.8) by a half-width of the resonance with $|s_r| = 1$ since the system will move along the separatrix of this resonance independent of its overlap with the adjacent resonance of a larger $|s_r|$. On the other hand, the overlap with the adjacent resonance of a smaller $|s_r|$ is ensured because it occurs for $|s| < 1$. Obviously, the maximal s_T corresponds to such a disposition of the marginal resonance that its lower (in $|s|$) edge corresponds to $|s| = 1$. Then s_T exceeds the value (6.8) by the full width of the resonance separatrix. Assuming the half-width of an integer resonance for system (6.11) to be $(\Delta P)_r \approx 2\sqrt{K_s} \approx \lambda(\Delta s)_r/s_r$, or $(\Delta s)_r \approx 2/\lambda$ $(K_s, |s_r| \approx 1)$ (see section 4.2) we may write for the half-width of a stochastic layer:

$$s_T(\zeta) \approx 1 + \frac{4\zeta}{\lambda} \to 1 + \frac{2\sqrt{K}}{\pi}\zeta; \qquad 0 \leqslant \zeta \leqslant 1. \tag{6.13}$$

The last expression corresponds to a resonance separatrix of the standard mapping for which $\lambda = 2\pi/\sqrt{K}$ (section 4.2). The evaluation of ζ is a more difficult problem since it depends generally on the overlap of fractional resonances in system (6.11). However, the related uncertainty in the width of the stochastic layer is insignificant for large λ.

In table 6.1 the results of the measurement concerning the width of the stochastic layer for an integer resonance of the standard mapping are presented for various values of the parameter K. The quantity $s_E = |w|_{max}/\lambda W_r$ was calculated for the maximal value $|w|_{max}$ reached in computation for a given K. The quantity W_r was assumed to be equal to the theoretical value (6.2) (with parameters (6.3)) multiplied by the mean empirical correction r (6.5). According to eq. (6.13) the values of s_E must lie within the interval:

$$1 \leqslant s_E \leqslant s_{max} = s_T(1) = 1 + 4/\lambda. \tag{6.14}$$

The last quantity is also given in table 6.1. It is seen that in many cases the value of s_E gets out of the

Table 6.1
Width of stochastic layer

K	λ	N	s_E	s_1	s_{max}	T_E	T_a	w_s
1	6.28	22580	4.04		1.64	4.43	4.83	0.69
			(0.69)	(0.75)				
0.9	6.62	20936	2.24		1.60	4.47	5.49	0.47
			(1.10)	(1.26)				
0.8	7.02	15662	2.00		1.57	6.38	6.33	0.30
			(0.85)	(0.96)				
0.75	7.26	139170	1.29	1.39	1.55	7.18	6.85	0.24
0.7	7.51	12732	1.17	1.40	1.53	7.85	7.44	0.17
0.6	8.11	10700	1.17	1.43	1.49	9.34	8.96	0.85×10^{-1}
0.5	8.89	8432	0.93	1.10	1.45	11.86	11.15	0.33×10^{-1}
0.4	9.93	6510	0.82	0.96	1.40	15.36	14.54	0.88×10^{-2}
0.3	11.47	4679	0.77	0.92	1.35	21.37	20.41	0.12×10^{-2}
0.2	14.05	3062	0.81	1.02	1.28	32.64	32.69	0.39×10^{-4}
0.15	16.22	2225	0.86	1.15	1.25	44.92	45.45	0.19×10^{-5}
0.1	19.87	1336	0.64	0.82	1.20	74.81	71.85	0.12×10^{-7}
0.05	28.10	6350	0.69	0.80	1.14	157.5	154.8	0.81×10^{-13}

interval (6.14) and, moreover, in both sides. Lower values of $s_E < 1$ could be explained partly by the insufficient computation time so that the system has not enough time to reach the edges of the stochastic layer where the diffusion is very slow (fig. 5.3). Using the empirical dependence (5.3) with the parameters (5.4) we can correct s_E in a way similar to that in section 4.3. In the present case $K_\epsilon \to 1/s_E$; $K_1 \to 1/s_1$ where s_1 is the corrected value of s_E. The quantity N in eq. (4.49) means now the number of crossings of the surface $\theta = \pi$ to which the whisker mapping (6.1) relates. The values of N and s_1 are also given in table 6.1. It is seen that after the correction the number of events with $s_1 < 1$ has been reduced, yet they still persist. It is related, perhaps, to big fluctuations of the slow diffusion (see section 5.1)).

All events with $s_E > s_{max}$ correspond to $K \approx 1$. Such a big width of the stochastic layer is related apparently to an overlap of the layer with some nearest fractional resonances. In any event, for $K = 1$ the resonances do overlap the whole interval of periodicity $\Delta I = 2\pi$ (section 5.1) so that w will grow indefinitely with time. Note that it is true only for the outer part of a stochastic layer ($w > 0$). Inside a resonance ($w < 0$) the stability domains persist so that the negative values of w are bounded. These values are also presented in table 6.1 for the first three cases (in brackets). For $K = 1$ the corresponding $s_1 < 1$ that is related apparently to a rather long "wandering" of the system amid the adjacent fractional resonances (in a region of $w > 0$).

Thus, the actual width of a stochastic layer is described satisfactorily by the eq. (6.13) based ultimately on the properties of the standard mapping (section 5). The accuracy of the analytical estimate increases as λ grows. The agreement between the theory and numerical experiments may be really considered as satisfactory if one takes into account that the layer width (w_s) changes over the data of table 6.1 by 13 orders of magnitude! The width of stochastic layer for small K is so little that the two last cases in table 6.1 had to be computed with the double precision. The width of a stochastic layer may be immediately found also from the computation of the whisker mapping (6.9). It is true, that the mapping (6.9) describes the motion near a separatrix only approximately. On the other hand, it requires much less computation as compared to the original standard mapping (5.1). As a result we have managed to compute mapping (6.9) with $\lambda = 700$ (as compared to the maximal value of $\lambda = 28.1$

in the previous method, see table 6.1) and for 6×10^6 iterations (cf. the values of N in table 6.1). We have obtained that the quantity s_1 lies within the interval (1.004–1.013) that is compatible with the value of $s_{\max}=1.006$.

We may use this result to check the critical value of $K=K_1$ for the standard mapping found in section 5.1: $K_1=0.989$. Applying the relation $K_1\approx 1/|s_r|$ (6.11), where s_r are the two nearest to s_1 resonant values we have got from the computation data:

$$0.98\lesssim K_1\lesssim 1. \tag{6.15}$$

Let us consider still another method to determine the width of a stochastic layer. The method is based upon the measurement of the mean rotation period (or a mean oscillation half-period) T in a stochastic layer. For the free oscillations of a pendulum near the separatrix (see section 2.4):

$$T(s)\approx\frac{1}{\omega_0}\ln\left(\frac{32}{|w|}\right)\to\frac{1}{\sqrt{K}}\ln\left(\frac{32}{|s|\,w_s}\right). \tag{6.16}$$

The last expression is again related to the case in which the pendulum represents an integer nonlinear resonance of the standard mapping (5.1), period T being measured in the number of iterations. From the numerical experiments one can easily find the mean value $T_E=t_N/N$ where t_N is the time interval of motion corresponding to exactly N periods, or to the $N+1$ crossings of the surface $\theta=\pi$.

To deduce the theoretical mean (T_a) one needs to average eq. (6.16) over time, which $s(t)$ depends on, or, due to the ergodicity, over the stochastic component in the layer. Neglecting stability domains inside the layer, which are of importance near the layer edges only, we can average eq. (6.16) over the whole layer. Recall now that the original mapping, whose properties we are studying, is the standard one (5.1). Since the transformation of variables $(I,\theta)\to(s,\tau)$ by transition from the standard to whisker mapping is not canonical we need the Jacobian of this transformation. The latter may be represented as a sequence of successive transformations: (I,θ) (5.1)$\to(J,\theta)$ (4.18)$\to(I_s,\varphi_s)\to(w,\tau)$ (6.1)$\to(s,\tau)$ (6.9) where I_s, φ_s are the action-angle variables of the unperturbed system (4.18) with the Hamiltonian:

$$H_0=\tfrac{1}{2}J^2+k\cos\theta=k(1+w) \tag{6.17}$$

describing the oscillations of frequency $\omega(w)$ (see eq. (2.33)) near the separatrix. Jacobians of all the transformations but one (I_s, $\varphi_s\to w,\tau$) in the above chain are constants independent of the dynamical variables. So we need to evaluate only the Jacobian $\partial(w,\tau)/\partial(I_s,\varphi_s)=(\partial w/\partial I_s)\ (\partial\tau/\partial\varphi_s)$. The last expression is due to the fact that w depends only on I_s but not on φ_s. The deviative $\partial w/\partial I_s$ may be found from the relation $\partial H_0/\partial I_s=\omega(w)$, whence: $\partial w/\partial I_s=\omega/k$ (see eq. (6.17)). To evaluate $\partial\tau/\partial\varphi_s$ we write: $\tau=\Omega t^0$ and $\omega t^0=\varphi_s=\text{const}$ – the value of φ_s at the surface $\theta=\pi$; t^0 is the time at which the system crosses this surface. Eliminating t^0 we get: $\tau=-\Omega\varphi_s/\omega$, and $\partial\tau/\partial\varphi_s=-\Omega/\omega$, whence the Jacobian we are interested in $\partial(w,\tau)/\partial(I_s,\varphi_s)=-\Omega/k$ does not depend on the dynamical variables, and so does the full Jacobian $\partial(s,\tau)/\partial(I,\theta)$. Hence we can evaluate any average simply over the phase plane (τ,s) of the whisker mapping. Since $T(s)$ does not depend on τ we get:

$$\begin{aligned}T_a&\approx\int_0^1 ds\,T(s)=\frac{1}{\sqrt{K}}\ln\left(\frac{32e}{w_s}\right)\approx\frac{\pi^2}{K}-\frac{1}{\sqrt{K}}\ln\left(\frac{2r\pi^4}{eK^{3/2}}\right)\\ w_s&=32\exp(-1-\sqrt{K}\,T_a).\end{aligned} \tag{6.18}$$

We have used here the relations (6.2), (6.3) and (6.8) to evaluate w_s as well as the correction (6.5). The last expression in eq. (6.18) permits us to calculate the width of a stochastic layer from the mean period of motion.

The comparison of the theoretical (T_a) and measured (T_E) values of the mean period is given in table 6.1. A good agreement between them over the whole interval of K variation provides an additional confirmation for the theoretical estimates of the width of a stochastic layer. It is also a check of the second equation in the whisker mapping (6.9).

Thus, we come to the conclusion that the mapping (6.9) satisfactorily describes the motion inside a stochastic layer of a nonlinear resonance. An example of the phase map for such a motion is given in fig. 6.1 with $\lambda = 8.89$ ($K = 0.5$). The layer width measured from the figure is $|s|_{max} = 1.16$. This value should be compared with the quantity s_1 in table 6.1 because the phase map in fig. 6.1 has been computed for a rather large $N = 10^6$.

The most important peculiarity of the whisker mapping is a strictly bounded domain of the stochastic motion, that is a *finite* width of the stochastic layer. This width drops exponentially as the perturbation parameter K of the original system decreases. In order of magnitude the width of a stochastic layer w_s proves to be larger by a factor of λ than the separatrix splitting (W, see eq. (6.8)). The structure of a stochastic layer is fairly complicated especially at its edges due to innumerable islets of stability (fig. 6.1). This structure may be studied in detail using the standard mapping (6.11). Roughly speaking, the layer consists of the two parts – the central one ($|s| \leqslant \frac{1}{4}$) where a fast diffusion takes place, and there are practically no islets of stability and the peripheral one ($\frac{1}{4} \leqslant |s| \leqslant 1$) with a slow diffusion and a substantial share of the stable component (fig. 6.1). Let us mention that a similar structure of the phase plane has been described for the problem of the Fermi acceleration [46], the latter being qualitatively similar with the problem of the motion inside a stochastic layer.

6.3. The KS-entropy in the stochastic layer

Some data concerning the KS-entropy of the standard mapping (5.1) for $K < 1$, i.e. inside the stochastic layer, were presented in table 5.1 (section 5.2). As was mentioned there the product of the entropy h_1 (per iteration of mapping (5.1)) by the mean motion period T_a in the stochastic layer

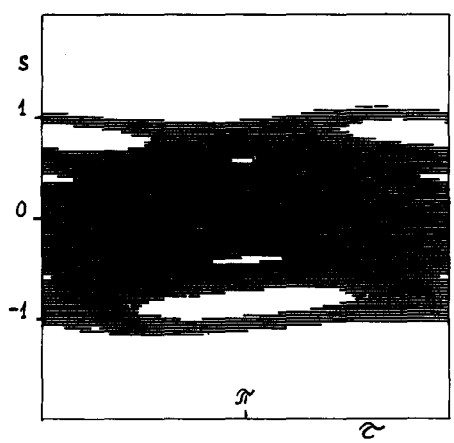

Fig. 6.1. Structure of a stochastic layer as described by the whisker mapping (6.9) with $\lambda = 8.89$ ($K = 0.5$); $s > 0$ – outer half of the layer (rotation); $s < 0$ – the inner half (oscillations); layer width $|s|_{max} = 1.16$; motion time 10^6 iterations.

remains approximately constant over a broad range of K variation. This product $h_1 T_a = h_s$ is nothing but the entropy per iteration of the whisker mapping (6.9). Therefore we will begin just with the consideration of the KS-entropy for this mapping. Since its local properties are described again by some standard mapping (6.11) with the parameter $K_s(s)$ depending only on s but not on λ the entropy of the whisker mapping (6.9) is really some universal constant. It can be found by the averaging of the entropy of standard mapping (6.11) over s.

Assume the dependence for the entropy of the standard mapping on $K_s \approx 1/s$ in the form:

$$h(s) = \begin{cases} \ln(1/2s); & s < s_b \\ \gamma/s; & s > s_b \end{cases} \tag{6.19}$$

where s_b, γ are some constants. The first dependence corresponds to eq. (5.41), and judging by the data of table 5.1, describes the KS-entropy of the standard mapping quite well for $K \geqslant 4$. It corresponds to the central part of the stochastic layer ($|s| \leqslant 1/4$). The second of eqs. (6.19) can be deduced from the relations: $h_s \approx 0.8$ (table 5.1) and $K T_a \approx 5$ (see eq. (6.18) and table 6.1). Numerical values for these nearly constant quantities are taken in the domain of $K_s \sim 1$ which is most important for the averaging over s. Then the value of $\gamma = h_s/K_s T_a \approx 0.16$. This value may be obtained also in a different way. In accordance with a rough structure of the stochastic layer described at the end of the preceding section we attribute the two dependences of $h(s)$ (6.19) to the central and peripheral parts of the layer, respectively, and assume $s_b = \frac{1}{4}$. The quantity γ can be found then from the condition of the equality of both relations for $h(s)$ at $s = s_b$. We get:

$$\gamma = s_b \ln(1/2s_b) = 0.173; \qquad s_b = \tfrac{1}{4} \tag{6.20}$$

that is close to the above value. In what follows we assume the value (6.20).

Now we can evaluate the KS-entropy of the whisker mapping (6.9). Using eqs. (6.19) and (6.20) we have:

$$h_W = \int_0^1 \mathrm{d}s\, h(s) = s_b \ln(\mathrm{e}/2s_b) + \gamma \ln(1/s_b) = 0.423 + 0.240 = 0.663. \tag{6.21}$$

The value of h_W has been obtained also from numerical experiments with the whisker mapping (6.9) employing the method of tangent mapping described in section 5.2. The h_W value averaged over 28 trajectories with various initial conditions and for different values of parameter λ in the interval (3–9) is equal to:

$$\langle h_W \rangle = 0.666 \pm 0.013 \tag{6.22}$$

the standard deviation of a single value being ± 0.069. Within the specified interval of λ the value of $\langle h_W \rangle$ does not depend on λ to the accuracy determined by fluctuations. For $\lambda < 3$ the mean $\langle h_W \rangle$ decreases, apparently, due to an increase of the layer width (6.13), and for $\lambda > 9$ it increases, probably, due to an insufficient time of motion ($t = 10^5$ iterations) to penetrate into the layer peripheral domain of a slow diffusion ($|s| \approx 1$).

The results of measurement concerning the KS-entropy of the standard mapping (5.1) are presented in table 6.2. These data expand those in table 5.1 over the region of small K values. The entropy was measured employing the tangent mapping (section 5.2) over the motion time $t = 10^5$. In the middle of table 6.2 the measured values of h_s are in a good agreement with the theoretical prediction (h_W, eq.

Table 6.2
The KS-entropy in the stochastic layer

K	h_1 linear map	$h_s = h_1 T_a$	h_s/h_W (6.21)
0.15	0.0231	1.050	1.58
	(0.00643)	(0.292)	(0.44)
0.2	0.0295	0.964	1.45
	(0.0200)	(0.654)	(0.99)
0.3	0.0355	0.725	1.09
0.5	0.0686	0.765	1.15
0.7	0.0920	0.684	1.03
1	0.132	0.638	0.96
1.3	0.227	0.799	1.21
2	0.425	0.896	1.35
3	0.672	0.896	1.35
4	0.833	0.824	1.24

(6.21)). For very small K the value of h_s grows, apparently, due to an insufficient motion time (see above). The values of h_s for $t = 10^6$ are given in brackets. In the case of $K = 0.2$ h_s value has "descended" down to the theoretical one but for $K = 0.15$ it did so still much lower. This is caused, perhaps, by a "sticking" of the system in a peripheral part of the layer. For $K > 1$ the ratio h_s/h_W grows up appreciably, probably, due to the overlapping of different stochastic layers. Summarizing, we can conclude that the idea of a constant KS-entropy in the stochastic layer (per motion period T_a) permits us to describe satisfactorily and, what is both important and pleasant, very simply the instability rate inside the stochastic layer up to $K \sim 1$ and even, strange though it may seem, for fairly large K, with less accuracy though (see table 5.1).

6.4. Again about the border of stability

Now we can turn back to the evaluation of the border of gross instability for the standard mapping (section 5.1). The best estimate deduced from the overlap condition for the resonances of the first three harmonics gives: $K_T \approx 1.35$ (5.25). Taking account of the stochastic layer around separatrix permits us to improve this estimate.

Below we will confine ourselves to a simplified scheme for the overlap taking account of the integer and half-integer resonances only and neglecting the stochastic layers of the latter. The relation between the dimensionless energy w and displacement δI from the unperturbed separatrix can be found from the Hamiltonian (6.17). Since $w = \delta H_0/k = J(\delta J)/k = I(\delta I)/K$ we get: $(\delta I)/I = Kw/I^2 \to w/4$, the latter expression corresponding to the maximal width of separatrix at $\theta = \pi$ $(I \to I_m = 2\sqrt{K})$ which determines the overlap condition. The edge of the stochastic layer is related to $w = s_T w_s$, s_T being given by eq. (6.13). It is convenient to describe the influence of stochastic layer on the resonance overlap by a factor giving an effective width of a resonance:

$$l(K) = 1 + \frac{\delta I}{I_m} = 1 + \frac{s_T w_s}{4} = 1 + r s_T \frac{16\pi^4}{K^{3/2}} e^{-\pi^2/\sqrt{K}} \tag{6.23}$$

where $r = 2.15$ is the empirical correction (6.5) due to higher approximations.

Taking account of the factor (6.23) we can write down the condition for the touching between an integer and the adjacent half-integer resonances in the form (cf. section 5.1):

$$2\sqrt{K}\, l(K) + \tfrac{1}{2}K = \pi. \tag{6.24}$$

For $l = 1$ the critical value of $K_T = 1.46$ (5.23). Solving eq. (6.24) with $l(K)$ given by eq. (6.23) we get the critical value of K_T within the interval:

$$K_T = 1.019\text{–}1.098. \tag{6.25}$$

The uncertainty of K_T is related to the uncertainty of the layer width s_T (6.13) due to the marginal resonance. The ends of interval (6.25) correspond to the limiting values of the unknown parameter $\zeta = 1; 0$, respectively. Note that if we had ignored the empirical correction (6.5), and had set just $r = 1$ in eq. (6.23) we would get $K_T = 1.14\text{–}1.21$ instead of eq. (6.25). The value of ζ may be found from the whisker mapping (6.9) with $\lambda = 2\pi$. The numerical experiments over the motion time $t = 10^6$ reveal $|s|_{max} = 1.25$, whence $\zeta = 0.39$. Substituting this value into eq. (6.23) and solving eq. (6.24) we get: $K_T = 1.062$. This value as well as that from eq. (6.25) is still larger than the actual critical value $K_E \approx 0.99$ (section 5.1) but are already fairly close to the latter. This confirms once more that a quite simple picture of the resonance overlap is not too far from the truth.

7. The Arnold diffusion

The motion inside the stochastic layer of a resonance may be considered in a sense as some manifestation of a universal instability since those stochastic layers always exist almost in any nonlinear oscillator system. If, however, the number of degrees of freedom ≤ 2 such a universal instability is of no importance because for a sufficiently small perturbation the unstable motion is confined within a very (exponentially) small domain, and the stochastic layers of different resonances are separated from each other by stable invariant tori (section 4.6). Therefore, the change in the integrals of the unperturbed motion due to that instability is also exponentially small even inside a stochastic layer.

But it's quite another thing when a nonlinear oscillator system having more than two degrees of freedom is involved. In such a case the stochastic layers of different resonances do intersect. Since the motion inside a layer is stochastic and, particularly, ergodic, it will inevitably spread out over the whole system of intersecting layers (section 4.6). Such an instability of many-dimensional nonlinear oscillations has been predicted by Arnold and demonstrated by him via a fairly simple example [5] (see section 7.1). Arnold conjectured also that the mechanism of this instability is a generic one for many-dimensional nonlinear oscillations [67]. An analysis of the instability by means of the overlap criterion has revealed that it is a stochastic instability. This is just a reason to call it the *Arnold diffusion* [43, 41].

The Arnold diffusion apparently was actually observed already in the experiments with electrons in a magnetic bottle [101] (an analysis of these experiments is given in refs. [43, 138]) and also in experiments on the interaction of colliding electron–positron beams [102]. Recently a number of experiments on the electron motion in a model of the geomagnetic field has been carried out [126]* close in results to those in ref. [101]. In the numerical experiments with a many-dimensional nonlinear

*For analysis of these experiments and their relation to the Arnold diffusion see refs. [138, 142].

mapping [35] a very weak instability was observed which is also related apparently to the Arnold diffusion. A more definite analysis as to the nature of a similar weak instability observed in the numerical experiments with a simple model is presented in ref. [41]. The distinctive peculiarities of this instability agree, qualitatively and quantitatively (at least, in order of magnitude), with the inferences of an analytical theory for the Arnold diffusion. One can believe, therefore, that the existence of the Arnold diffusion is proved. Let us make clear that in the numerical experiments of ref. [41] just the *diffusion* nature of the developing instability has been established; a proof, and moreover a rigorous one, of the *very existence* of instability has been given already in the first paper by Arnold [5].

7.1. Arnold's example

In ref. [5] a simple example of a system having 2.5 degrees of freedom (5-dimensional phase space, see (7.1)) was considered. The motion of this system has proved to be always unstable for certain initial conditions. The system is described by the Hamiltonian:

$$\begin{aligned} &H(I_1, I_2, \theta_1, \theta_2, t) = \tfrac{1}{2}(I_1^2 + I_2^2) + \epsilon(\cos\theta_1 - 1)(1 + \mu B(\theta_2, t)) \\ &B = \sin\theta_2 + \cos t. \end{aligned} \tag{7.1}$$

For $\epsilon = \mu = 0$ the system has two integrals of motion I_1, $I_2 = \text{const}$ which determine the invariant tori supporting a quasi-periodical motion of two frequencies $\omega_{1,2} = I_{1,2}$.

For $\mu = 0$; $\epsilon \neq 0$ there are still two integrals: $I_2 = \text{const}$ and:

$$H_1 = \tfrac{1}{2}I_1^2 + \epsilon(\cos\theta_1 - 1) = \text{const}. \tag{7.2}$$

The latter integral describes the nonlinear resonance $\omega_1^r = 0$, function $H_1(I_1, \theta_1)$ being the resonance Hamiltonian (cf. sections 3.3 and 3.2). The separatrix of the resonance corresponds to the value of $H_1 = 0$ (section 2.4). Since for $\mu = 0$ the system is still completely integrable we shall take it for the unperturbed one with the Hamiltonian:

$$H_0 = H_1 + H_2; \qquad H_2 = \tfrac{1}{2}I_2^2. \tag{7.3}$$

The original (perturbed) system (7.1) may be represented then by the Hamiltonian:

$$\begin{aligned} &H = H_0(I_1, I_2, \theta_1) + \mu V(\theta_1, \theta_2, t) \\ &V = \epsilon B(\theta_2, t)(\cos\theta_1 - 1). \end{aligned} \tag{7.4}$$

The periodic (in θ_2, t) perturbation V affects the phase oscillations at a nonlinear resonance $\omega_1^r = 0$ (7.2) and leads, in particular, to the formation of a stochastic layer around the separatrix of this resonance (section 6). In the present case, however, the perturbation changes not only I_1 but also I_2 owing to the dependence of V on θ_2 (7.4). Therefore, a motion *along* the stochastic layer (in I_2) takes place. By virtue of stochasticity of the motion inside the layer the variation of I_2 will be also stochastic, giving rise to a diffusion in I_2. As a result, I_2 will change indefinitely, that is we have a gross instability.

The resonance $\omega_1^r = 0$ is not the only one even in the first approximation. The full set of resonances in this approximation is outlined in fig. 7.1; it includes 6 resonances:

$$\omega_1^r = 0; \qquad \omega_2^r = 0; \qquad \omega_1^r \pm \omega_2^r = 0; \qquad \omega_1^r \pm 1 = 0.$$

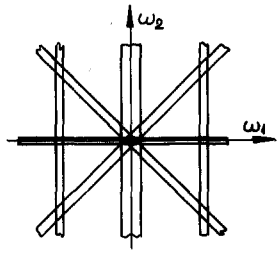

Fig. 7.1. First approximation resonances for Arnold's example (7.1).

The resonance lines intersect at 7 points: $\omega_1 = \omega_2 = 0$ and $\omega_1 = 0$; $\omega_2 = 0$; ± 1. Hence the diffusion spreads over all this resonance set. However, for $\mu \ll 1$ the diffusion rate is negligible along all resonances but $\omega_1^r = 0$ (see section 7.2).

The position of the unstable equilibrium for system (7.2) ($I_1 = \theta_1 = 0$) corresponds in the full (5-dimensional) phase space of the original system (7.1) to a 2-dimensional torus the motion on which is determined by the variation of phase θ_2 and time t. This is just a whiskered torus according to Arnold. The three-dimensional whiskers "fastened" to this whiskered torus correspond to the separatrix of resonance $\omega_1^r = 0$ (cf. section 2.4). For $\mu = 0$ the whiskers of the tori with different values of I_2 are isolated from each other because $I_2 = \text{const}$ (invariant). However, if $\mu \neq 0$ the variation of I_2 results in an intersection of the whiskers belonging to adjacent (in I_2) whiskered tori. A fine point of Arnold's elegant proof [5] is the transition from an elementary fact of the intersection of whiskers of close tori to the existence of trajectories connecting some neighbourhoods of the whiskered tori which are arbitrarily far (in I_2) from each other. To prove this far from trivial fact Arnold employs, essentially, the local instability of motion in a small vicinity of whiskers, he does not speak about that explicitly though. A basic idea how to "construct" such unstable trajectories under the condition of the local instability of motion was outlined in section 5.3, it may be done just by means of Arnold's transition chain.

7.2. Evaluation of the diffusion rate

We consider now Arnold's example, employing the regularities of the stochastic motion described in sections 4 through 6. Let initial conditions lie inside the stochastic layer of the resonance $\omega_1^r = 0$ (7.2). We want to find a change in the full Hamiltonian (energy) (7.1) over a half-period of the phase oscillations. We have (cf. section 4.4):

$$\dot{H} = \mu\, \partial V/\partial t = \epsilon\mu \sin t + \tfrac{1}{2}\epsilon\mu(\sin(\theta_1 - t) - \sin(\theta_1 + t)). \tag{7.5}$$

Here $\theta_1(t)$ may be expressed approximately via the asymptotic motion on the separatrix (2.30):

$$\begin{aligned} \theta_1 &= \varphi_{sx}(t) + \pi = 4\arctan(\exp\{\sqrt{\epsilon}(t - t^0)\}) \\ \dot{\theta}_1 &= 2\sqrt{\epsilon}\,\sin(\tfrac{1}{2}\theta_1) \end{aligned} \tag{7.6}$$

where t^0 is the time when $\theta_1 = \pi$. Note that in the relation for $\dot{\theta}_1$ (7.6) only the positive sign needs to be used assuming that one branch of the separatrix corresponds to $\theta_1 > 0$ whereas the other does so to $\theta_1 < 0$. The latter definition of asymptotic motion turns out to be more convenient as compared to the former (2.29), and we will employ it henceforth.

On the right side of eq. (7.5) we ignore the first term, which brings about only small oscillations of H, and arrive, by means of the MA integral (see appendix, $\varphi = \theta_1 - \pi$), at:

$$\Delta H \approx \frac{\sqrt{\epsilon}\mu}{2} \int_{-\infty}^{\infty} \mathrm{d}t \sqrt{\epsilon} \sin(\theta_1 - t) = \frac{\sqrt{\epsilon}\mu}{2} A_2\left(\frac{1}{\sqrt{\epsilon}}\right) \sin t^0. \tag{7.7}$$

For a *symmetric perturbation* (having the terms for both phases $\theta_1 \pm t$ which are equal in modulus) the expression (7.7) holds for both signs of $\dot{\theta}_1$ (see the appendix and section 4.4). Note that since the difference $H - H_0 = \mu V(\theta_1, \theta_2, t)$ (7.4) is a periodic function of the arguments it does not contribute to the MA integral (see the appendix and section 4.4). Therefore, $\Delta H_0 = \Delta H$ (7.7).

Inside the stochastic layer (in its central part, to be precise) the successive values of t^0 may be taken as random to a good accuracy, that is the limiting stochasticity can be assumed (sections 5 and 6). However, it would be erroneous to conclude that *all* the values of t^0 in eq. (7.7) are random and independent of each other. We would get thus a wrong relation for the diffusion rate (cf. eq. (7.20)):

$$D_{\mathrm{H}} \equiv \frac{\overline{(\Delta H)^2}}{T_{\mathrm{a}}} = \frac{\epsilon\mu^2}{4T_{\mathrm{a}}} A_2^2\left(\frac{1}{\sqrt{\epsilon}}\right) \overline{\sin^2 t^0} = \frac{\epsilon\mu^2}{8T_{\mathrm{a}}} A_2^2\left(\frac{1}{\sqrt{\epsilon}}\right) \tag{7.8}$$

where T_{a} is the mean half-period of phase oscillations in the stochastic layer (6.18). Actually, some correlations between different valules of t^0 arise when the system approaches the edges of the layer.

To clarify this question let us consider the variation of $H_2 = I_2^2/2$ (7.3). We have:

$$\dot{H}_2 = I_2 \dot{I}_2 = \epsilon\mu\omega \cos\theta_2 - \tfrac{1}{2}\epsilon\mu\omega(\cos(\theta_1 - \theta_2) + \cos(\theta_1 + \theta_2)) \tag{7.9}$$

where $\omega = \dot{\theta}_2 = I_2 \approx$ const owing to a small change in I_2. Similarly to the evaluation of ΔH (7.7) we can neglect the first term in eq. (7.9). Since the rest of the perturbation is symmetric ($\theta_1 \pm \theta_2$) we have for both signs of $\dot{\theta}_1$:

$$\Delta H_2 \approx \tfrac{1}{2}\sqrt{\epsilon}\mu\omega A_2(\omega/\sqrt{\epsilon}) \cos\theta_2^0 \tag{7.10}$$

where $\theta_2^0 = \theta_2(t^0)$. Again, if we would consider the successive phase values θ_2^0 as random and independent the diffusion rate in H_2 were:

$$D_2 \equiv \frac{\overline{(\Delta H_2)^2}}{T_{\mathrm{a}}} = \frac{\epsilon\mu^2\omega^2}{4T_{\mathrm{a}}} A_2^2\left(\frac{\omega}{\sqrt{\epsilon}}\right) \overline{\cos^2\theta_2^0} = \frac{\epsilon\mu^2\omega^2}{8T_{\mathrm{a}}} A_2^2\left(\frac{\omega}{\sqrt{\epsilon}}\right). \tag{7.11}$$

Both relations ((7.8) and (7.11)) undoubtedly hold during sufficiently short intervals of time while the system is wandering in the central part of stochastic layer, that is while the variations ΔH; ΔH_2 are less than the width of the layer. We are interested, however, in the opposite limiting case of a long-range ("far") diffusion. In this case both diffusion rates (D_{H}, D_2) will vary from the maximal values (7.8), (7.11) within the central part of the layer down to zero at the layer edges. Moreover, being averaged over a sufficiently long period of time both rates must be equal: $\bar{D}_{\mathrm{H}} = \bar{D}_2$, since a substantial variation of H is possible only via the variation of H_2 whereas the energy H_1 (7.2) is to remain constant to the accuracy of the order of layer width. This leads to a relation between the phases:

$$\begin{aligned} &\overline{\sin^2 t^0} = v^2 \overline{\cos^2\theta_2^0} \\ &v = \frac{\omega A_2(\omega/\sqrt{\epsilon})}{A_2(1/\sqrt{\epsilon})} \approx \omega^2 \exp\left\{(1-\omega)\cdot\frac{\pi}{2\sqrt{\epsilon}}\right\}. \end{aligned} \tag{7.12}$$

This relation shows that for $v \neq 1$ both sequences of phases $(t^0; \theta_2^0)$ cannot be random simultaneously. It is clear, therefore, that without a detailed analysis of the phase correlation near the edges of the stochastic layer the value of $\overline{\sin^2 t^0}$ (or $\overline{\cos^2 \theta_2^0}$) and, hence, the diffusion rate D_H cannot be evaluated. On the other hand, any analysis of the phase relation in the peripheral part of the stochastic layer seems to be too complicated for a physicist. Therefore, we have to introduce some additional assumptions which seem to be, however, quite natural.

We note, first of all, that a slow diffusion in the peripheral part of the stochastic layer reduces both $\overline{\sin^2 t^0}$ and $\overline{\cos^2 \theta_2^0}$. The reduction factor (R_T) can be estimated roughly as the ratio of the width of the central part to the full width of the layer: $R_T \approx \frac{1}{4}$ (section 6.2). A more accurate estimate for R_T can be obtained by averaging the ratio of the actual diffusion rate (5.52) to the limiting one (5.46). Assuming this ratio as $R_T(s) \approx (K-1)^2/K^2 \approx (1-s)^2$ (see section 6.2) we get:

$$R_T = \int_0^1 \mathrm{d}s\, R_T(s) \approx \tfrac{1}{3} \tag{7.13}$$

which is in a reasonable agreement with the above rough estimate: $R_T \approx \frac{1}{4}$.

We take into account, further, that the phase correlation related to the limitation of free diffusion across the stochastic layer diminish still more the value of $\overline{\sin^2 t^0}$ (and/or of $\overline{\cos^2 \theta_2^0}$). Let $v^2 \ll 1$ $(\omega > 1)$, for example. It is plausible to assume that the *larger mean* is approximately random one (with the factor R_T, of course, i.e. $\overline{\cos^2 \theta_2^0} \approx R_T/2$) whereas the smaller mean is just suppressed by the correlation. For $v^2 \gg 1$ $(\omega < 1)$ it will be vice versa, namely: $\overline{\sin^2 t^0} \approx R_T/2$. In other words, we assume that in spite of some correlation present the phase of the *smaller perturbation* (e.g., θ_2^0 for $v^2 \ll 1$, see also eq. (7.14)) is still partly random in the sense that the sum squared:

$$\left(\sum_i^t \cos \theta_{2i}^0\right)^2 \longrightarrow \tfrac{1}{2} R_T\, t; \qquad t \to \infty$$

grows in proportion to the time of motion but the growth rate is reduced by the factor $R_T < 1$ due to a slow diffusion in the peripheral part of the stochastic layer. We shall call this type of motion the *reduced stochasticity* as distinct from the limiting stochasticity (section 5.4) with $R_T = 1$. Below we shall consider this assumption from another point of view.

It turns out that the Arnold diffusion can be also described by a mapping. To construct this mapping we find first the change in the unperturbed resonance Hamiltonian H_1 over a half-period of the unperturbed phase oscillations. The simplest way to do this is to make use of the relations: $H_0 = H_1 + H_2$ (7.3) and $\Delta H_0 = \Delta H$ (7.7), whence:

$$\Delta H_1 \approx \tfrac{1}{2}\sqrt{\epsilon}\,\mu\, A_2(1/\sqrt{\epsilon})\,(\sin t^0 - v \cos \theta_2^0). \tag{7.14}$$

The perturbation ΔH_1 depends on the two phases $(t^0; \theta_2^0)$, their variation being determined by the relations (cf. sections 4.3 and 4.4):

$$\overline{t^0} \approx t^0 + T(\bar{w}); \qquad \overline{\theta_2^0} \approx \theta_2^0 + \omega\, T(\bar{w}). \tag{7.15}$$

Here $T(w)$ is a half-period of the unperturbed phase oscillations which depends on the relative energy $w = H_1/\epsilon$ (section 2.4). Since the variation of ω is small $(|\bar{\omega} - \omega| \ll \omega)$ we can introduce a new phase $\tau^0 = \theta_2^0/\omega$ which obeys the equation: $\overline{\tau^0} \approx \tau^0 + T(\bar{w})$. Hence $\tau^0 - t = \beta = \text{const}$ and $\theta_2^0 \approx \omega(t^0 + \beta)$. We have excluded, thus, one of phases. Combining eqs. (7.14) and (7.15) we arrive at a mapping:

$$\bar{w} = w + \frac{\mu}{2\sqrt{\epsilon}} A_2\left(\frac{1}{\sqrt{\epsilon}}\right)(\sin t^0 - v\cos(\omega(t^0+\beta)))$$
$$\bar{t}^0 = t^0 + T(\bar{w}) \tag{7.16}$$

which describes the phase correlation during the Arnold diffusion in the system under consideration. This mapping resembles the whisker mapping (6.1) in some respects. The most important distinction is a quasi-periodical character of the perturbation in eq. (7.16) for an irrational ω.

Employing the mapping (7.16) we may analyse again the phase correlation. Since the latter is related to the boundedness (by stochastic layer) of the w variation it seems quite plausible to suggest that the phase values of the *larger term* in the first of eqs. (7.16) will be correlated because just this term does determine the motion across the stochastic layer. To this accuracy the mapping (7.16) is reduced to the whisker mapping (6.1), and we can immediately apply all the inferences of section 6 concerning the motion across the stochastic layer. In particular, the layer width is given by the expression (see eq. (6.8)):

$$w_s \approx \begin{cases} w_0; & v \ll 1 \quad (\omega > 1) \\ w_0 v\omega; & v \gg 1 \quad (\omega < 1) \end{cases}$$
$$w_0 = \frac{\mu}{2\epsilon} A_2\left(\frac{1}{\sqrt{\epsilon}}\right) \approx \frac{4\pi\mu}{\epsilon^{3/2}} \exp\left\{-\frac{\pi}{2\sqrt{\epsilon}}\right\}. \tag{7.17}$$

The mean half-period of phase oscillations in the stochastic layer can be found then from eq. (6.18):

$$T_a = \frac{1}{\sqrt{\epsilon}} \ln\left(\frac{32e}{w_s}\right) = \begin{cases} T_0; & v \ll 1 \\ T_0 - \ln(\omega v); & v \gg 1 \end{cases} \tag{7.18}$$
$$T_0 = \frac{1}{\sqrt{\epsilon}} \ln\left(\frac{32e}{w_0}\right).$$

The mean diffusion rate which describes a long-term motion inside the stochastic layer can be evaluated now from eq. (7.8) taking account of the phase relation (7.12): $\overline{\sin^2 t^0} = v^2 \cos^2 \theta_2^0$ and the inference of the reduced stochasticity, namely:

$$\overline{\sin^2 t^0} \approx \begin{cases} \frac{1}{2} R_T v^2; & v \ll 1 \\ \frac{1}{2} R_T; & v \gg 1. \end{cases} \tag{7.19}$$

For the mean rate of the Arnold diffusion in Arnold's example we get finally:

$$\overline{D_H} \approx \begin{cases} D_0 v^2; & v \ll 1 \\ \dfrac{D_0}{1 - \ln(\omega v)/T_0}; & v \gg 1 \end{cases}$$
$$D_0 \approx \frac{\epsilon\mu^2 R_T}{8T_0} A_2^2\left(\frac{1}{\sqrt{\epsilon}}\right) \approx \frac{8\pi^2\mu^2 R_T}{T_0} \exp\left(-\frac{\pi}{\sqrt{\epsilon}}\right). \tag{7.20}$$

Unlike the motion *across* the stochastic layer the diffusion *along* the layer is determined by the smaller perturbation term in eq. (7.16).

Let us call the resonance along whose stochastic layer the diffusion goes the *guiding resonance.* It

may be called also the actual resonance since the system is moving just within a domain of this resonance, of its stochastic layer, to be precise. The other perturbation terms of Hamiltonian (7.4) are related to the resonances which we shall call *driving resonances*. They may be called also virtual resonances since under different initial conditions some of them may become the actual, or guiding, resonances.

Eq. (7.20) shows that the rate of Arnold diffusion is determined, mainly, by the "strength" of the guiding resonance, i.e. by the strength of the perturbation responsible for guiding resonance. The diffusion rate (7.20) depends exponentially on the perturbation parameter ϵ related to the guiding resonance. Therefore, the diffusion rate along other resonances is negligible for $\mu \ll 1$ (7.1).

7.3. A more general case

Let us consider now a more general problem concerning the Arnold diffusion. Let the Hamiltonian of a many-dimensional system be:

$$H(I, \theta) = H_0(I) + \epsilon V_g \cos(m^g, \theta) + \epsilon V$$
$$V = \sum_m V_m \cos(m, \theta) \tag{7.21}$$

where ϵ is small perturbation parameter; quantities I, θ, ω, m are N-dimensional vectors, and m^g stands for the vector of guiding resonance: $m^g, \omega^r = 0$; $\omega^r = \omega(I^r) = \partial H_0/\partial I$ (section 3.3). We confine ourselves, thus, to a conservative system, or to coupling resonances*. The Arnold diffusion in this system is due to the influence of perturbation V on the phase oscillations near the separatrix of a guiding resonance described by the Hamiltonian (7.21). We make the canonical transformation of variables $I, \theta \to p, \psi$ to introduce the so-called orthogonal metric (see section 3.3). Recall that it means a special form of the transformation matrix μ_{ki} (3.28), namely:

$$\mu_{1i} = m_i^g; \quad \mu_{2i} = \omega_i^r/|\omega^r|; \quad \mu_{ki} = e_i^k \ (k = 3, \ldots, N) \tag{7.22}$$

where all vectors e^k are unit and orthogonal to the vector ω^r and to each other. Besides, we choose now e^k to be orthogonal with the normal $n^r = m^g, \partial\omega/\partial I$ to the guiding resonance surface (in I-space, see below for explanation). The vectors (7.22) determine the directions of new momenta, and are all (except for the first one) mutually orthogonal and of unit length. The vector μ_{1i} is orthogonal to μ_{2i} but generally not to some e^k. Indeed, the angle α between the vectors n^r and m^g may be found from the relation

$$|n^r||m^g| \cos \alpha = n^r, m^g = m^g, \partial\omega/\partial I, m^g = 1/M_g$$

and is not equal generally to zero. It is essential also that $\alpha \neq \pi/2$ or the pendulum approximation for a guiding resonance breaks down ($M_g = \infty$, see section 3.3). In this case the vectors (7.22) would be no longer independent since the vector μ_{1i} being orthogonal to vector n^r, might be resolved into the vectors e^k.

Let us explain now the special option for vectors e^k. The point is that the Arnold diffusion spreads approximately over the intersection of an energy surface with the guiding resonance surface. Indeed, on the one hand, the system keeps within a stochastic layer, that is near the resonance surface m^g, $\omega(I) = 0$. On the other hand, the total energy (H) is conserved exactly that implies an approximate

*A particular case of the time-dependent Hamiltonian will be considered in section 7.5.

conservation of the unperturbed energy H_0 as $\epsilon \to 0$. The subspace of intersection of the two mentioned surfaces will be called the *diffusion surface*, and the plane tangential to this surface at the point $I = I^r$ the *diffusion plane*. This $(N-2)$-dimensional plane is built just upon the $(N-2)$ vectors e^k. These vectors determine, thus, a local space of the Arnold diffusion. Let us denote, in this connection, the projection of the new momentum vector p onto the diffusion plane by q. It is a $(N-2)$-dimensional vector with components q_k $(k = 3, \ldots, N)$.

A typical disposition of vectors (7.22) is shown in fig. 7.2 for $N = 3$ – the minimal number of degrees of freedom when the Arnold diffusion is still possible. The diffusion plane degenerates in this case into a line (one-dimensional diffusion).

In the new variables it suffices for the problem under consideration to keep only the following terms in the Hamiltonian (7.21) (see section 3.3):

$$\begin{aligned} &H(p, \psi) \approx H_0(I^r) + H_1(p_1, \psi_1) + \epsilon V(p, \psi) \\ &H_1 = p_1^2/2M_g + \epsilon V_g \cos \psi_1; \qquad 1/M_g = m^g, \partial\omega/\partial I, m^g \end{aligned} \tag{7.23}$$

where the perturbation V is defined by eq. (7.21), and the driving resonance phases (m, θ) are to be expressed in terms of the new phases ψ (see below). Without perturbation ($V = 0$) the quantities p_2, q_k are integrals of motion which we set to be zero. The quantity H_1 is also conserved for $V = 0$, which follows immediately from $H = \text{const}$ and $p_2 = q_k = 0$ (7.23). So we have the full set of N integrals of motion.

Under the influence of the perturbation ($V \neq 0$) all the integrals are varying somehow. To calculate their variations we transform, first of all, the driving resonance phases (see section 3.3):

$$\begin{aligned} &\varphi_m \equiv m, \theta = \nu, \psi \approx \xi_m \psi_1 + \omega_m t + \beta_m \\ &\xi_m = M_g/M_{(m)} = M_g \sum_{n=1}^{N} \nu_n/M_{n1}; \quad \omega_m = m, \omega^r = |\omega^r|\, \nu_2 \\ &M_{11} = M_g; \qquad \frac{1}{M_{21}} = \frac{\omega^r}{|\omega^r|}, \frac{\partial\omega}{\partial I}, m^g; \qquad \frac{1}{M_{k1}} = e^k, \frac{\partial\omega}{\partial I}, m^g; \qquad \frac{1}{M_{(m)}} = m, \frac{\partial\omega}{\partial I}, m^g \end{aligned} \tag{7.24}$$

where β_m stands for some constant phases. The first representation for φ_m follows from the transformation $\theta \to \psi$, the vector ν having generally real components as distinct from the integer vector m (see below). The second (approximate) representation for φ_m can be derived if one changes the function $\theta(t)$ for the unperturbed one ($V = 0$; eq. (3.39)); $\xi_m = M_g/M_{(m)}$ in this case. If we use, however, the unperturbed function $\psi(t)$ (see eq. (3.38)), $M_2 = 1$, we arrive at the second expression for $\xi_m(\nu)$, and also at the relation $\omega_m = |\omega^r|\, \nu_2$ (7.24).

Now we can find the variation of the unperturbed integrals:

$$\dot{p} = -\epsilon \frac{\partial V}{\partial \psi} = \epsilon \sum_m \nu(m)\, V_m \sin \varphi_m. \tag{7.25}$$

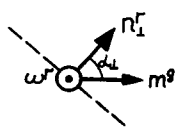

Fig. 7.2. Kinematics of the Arnold diffusion for conservative system (7.21) with $N = 3$. Vector ω^r is perpendicular to the plane of the figure which is tangential to energy surface $H_0 = \text{const}$. Vector $n_\perp^r$ is a projection of the normal to the guiding resonance surface onto the energy surface. Dotted line indicates the diffusion plane (line in this case).

This equation holds for all components of the vector p except p_1 which is not an unperturbed integral. Instead of $\dot{p}_1$ we evaluate the variation of integral H_1. We have (cf. section 4.4, eq. (4.53)):

$$\dot{H}_1 = \dot{\psi}_1 \left(-\epsilon \frac{\partial V}{\partial \psi_1}\right) = \epsilon \sum_m \nu_1(m)\, V_m\, \dot{\psi}_1 \sin \varphi_m. \tag{7.26}$$

To integrate eqs. (7.25) and (7.26) we substitute, as usual, the function $\psi_1(t)$ at the separatrix for that near the separatrix of the guiding resonance, and set $\dot{\psi}_1 \approx 2\Omega_g \sin(\psi_1/2)$ where $\Omega_g = \sqrt{\epsilon V_g / M_g}$ is the frequency of small phase oscillations, and both ψ_1, $\dot{\psi}_1$ have signs alike. We need now to evaluate the two integrals:

$$I_1 = \int_{-\infty}^{\infty} \mathrm{d}t \sin \varphi_m(t); \qquad I_2 = \int_{-\infty}^{\infty} \mathrm{d}t\, \dot{\psi}_1 \sin \varphi_m(t). \tag{7.27}$$

The first one is immediately expressed via the MA integral (see the appendix):

$$\begin{aligned} I_1 &= \frac{1}{\Omega_g} A_{2|\xi_m|}(\lambda_m) \sin \varphi_m^0 \\ \lambda_m &= -\frac{\omega_m}{\Omega_g} \operatorname{sign}(\xi_m\, \psi_1); \qquad \varphi_m^0 = \xi_m \psi_1^0 + \omega_m t^0 + \beta_m \end{aligned} \tag{7.28}$$

where $\varphi_m^0 = \varphi_m(t^0)$, and $\psi_1^0 = \psi_1(t^0) = \pm\pi$. To get I_2 we keep only that term of the integrand which has the maximal modulus of the factor at ψ_1 since it makes the main contribution to the integral:

$$2 \sin(\tfrac{1}{2}\psi_1) \sin \varphi_m \to -\operatorname{sign}(\xi_m) \cos(\operatorname{sign}(\xi_m)\, (|\xi_m| + \tfrac{1}{2})\psi_1 + \omega_m t + \beta_m).$$

We get:

$$\begin{aligned} I_2 &\approx -\operatorname{sign}(\xi_m)\, A_{2|\xi_m|+1}(\lambda_m) \cos(\varphi_m^0 + \tfrac{1}{2}\pi \operatorname{sign}(\xi_m\, \psi_1)) \\ &\approx -\frac{\omega_m}{\xi_m \Omega_g} A_{2|\xi_m|}(\lambda_m) \sin \varphi_m^0 = -\frac{\omega_m}{\xi_m} I_1. \end{aligned} \tag{7.29}$$

Employing the integrals I_1, I_2 we find from eqs. (7.25) and (7.26):

$$\begin{aligned} \Delta p &= \frac{\epsilon}{\Omega_g} \sum_m \nu(m)\, Q_m \sin \varphi_m^0 \\ \Delta H_1 &= -\frac{\epsilon |\omega^r|}{\Omega_g} \sum_m \frac{\nu_1(m)\, \nu_2(m)}{\xi_m} Q_m \sin \varphi_m^0 \\ Q_m &= V_m\, A_{2|\xi_m|}(\lambda_m) \end{aligned} \tag{7.30}$$

where we have used also the relation $\omega_m = |\omega^r|\, \nu_2$. The expression for Δp holds again for all the components except Δp_1 instead of which ΔH_1 should be taken.

The Arnold diffusion for system (7.21) may be described also by a mapping similar to that (7.16) for Arnold's example. First we find the change in phases φ_m^0 over a period of the unperturbed motion. Let $T(w)$ denote, as usual, a half-period of oscillations or a period of rotation for the phase ψ_1 as a function of the dimensionless energy $w = (H_1/\epsilon V_g) - 1$. The change in t^0 is $\Delta t^0 = T$. For the phase rotation ($w > 0$) $\Delta\psi_1^0 = \pm 2\pi$ in dependence of the direction of rotation, and this quantity must not be neglected unless ξ_m is an integer. For the phase oscillation ($w < 0$) every driving resonance "works"

only during one of the two successive half-periods since during the other one $\lambda_m < 0$ and its contribution is neglegible. In this case $\Delta\psi_1^0 = 0$ (phase motion is periodic) but $\Delta t^0 = T_1 + T_2$ where T_1, T_2 are the successive half-periods of oscillations. Generally $T_1 \neq T_2$ owing to a change in w at every half-period under the influence of some driving resonances. In order not to complicate the general relations to be derived we shall consider the two limiting cases:

1) A symmetric perturbation (see section 7.2), that is all the driving resonances are broken up into pairs, both components of each pair contributing equally to the variation of the integrals of motion during both half-periods of the phase oscillation, respectively. This is just the case for Arnold's example. Then $\Delta t^0 = T$, that is the perturbation is of the period T.

2) The influence of driving resonances during one half-period of the phase oscillation is much stronger than during the other one. Then $T_1 \approx T_2$ that may be considered as equivalent to the former relation $\Delta t^0 \approx T$ with the frequencies ω_m twice as much.

Under the above restrictions the change in a resonance phase may be written in the form:

$$\Delta\varphi_m^0 = \omega_m T + C_m$$

where C_m is some constant, for example, $C_m = \pm 2\pi\xi_m$ for rotation. Employing the relation (7.30) for ΔH_1 we arrive at a mapping describing the Arnold diffusion in a many-dimensional system:

$$\begin{aligned} \bar{w} &= w - \frac{|\omega^r|}{\Omega_g} \sum_m W_m \sin\varphi_m^0 \\ \overline{\varphi_m^0} &= \varphi_m^0 + \omega_m T(\bar{w}) + C_m \\ W_m &= \frac{\nu_1(m)\,\nu_2(m)}{\xi_m}\,\frac{Q_m}{V_g}. \end{aligned} \tag{7.31}$$

Assume like in section 7.2 that one of the terms in this sum is much larger than the others. Owing to an exponential dependence of the quantities W_m on the parameters a substantial difference in their values seems to be a typical case. The dynamics of the motion across the stochastic layer of a guiding resonance, that is the variation of w, is determined approximately just by this largest term whereas the other terms bring about the Arnold diffusion over the layer (cf. section 7.2). In other words, the diffusion is driven really by all the driving resonances except the strongest one which is mostly responsible for the formation of the stochastic layer as well as for its properties (width, KS-entropy etc.). Therefore, it would be more appropriate, perhaps, to retain the term driving resonance for all the resonances but the strongest one (and the guiding resonance, of course) which we shall call the *layer resonance.* Let us mark the quantities related to the layer resonance by subscript L and introduce the ratios:

$$v_m = W_m/W_L \ll 1 \ (v_L = 1); \qquad r_m = \omega_m/\omega_L; \qquad s = w/w_s$$

where

$$w_s = \frac{|\omega^r|}{\Omega_g}\lambda\, W_L \tag{7.32}$$

is the width of stochastic layer, and $\lambda = -\omega_L/\Omega_g > 0$. Comparing the relation for $\Delta\varphi_L^0 = \omega_L T + C_L$ (7.31) with that for $\Delta\varphi_m^0 = \omega_m T + C_m = r_m(\omega_L T + C_L) + C_m - r_m C_L$ we can express all phases φ_m^0 via the phase φ_L^0 which we prefer to denote by τ:

$$\varphi_m^0 = r_m \tau + b_m t + d_m. \tag{7.33}$$

Here $b_m = C_m - r_m C_L$ and d_m are some constants; time t is measured in the number of iterations of the mapping (7.31). Further (see eq. (2.33)):

$$\omega_L T(w) = -\lambda \ln(32/|w|) = \lambda \ln|s| - G; \quad G = \lambda \ln(32/w_s)$$

and we may write mapping (7.31) in the form:

$$\begin{aligned} \bar{s} &= s - \frac{\sin\tau}{\lambda} - \frac{1}{\lambda}\sum_m{}' v_m \sin(r_m\tau + b_m t + d_m) \\ \bar{\tau} &= \tau + \lambda \ln|\bar{s}| - G \end{aligned} \tag{7.34}$$

where prime at the symbol of the sum means the exclusion of the largest term ($\sin\tau$). This mapping differs from a similar one (7.16), a larger number of terms apart, in an explicit dependence of the perturbation on time unless all ξ_m are integers.

Let us define the diffusion tensor as:

$$D_{ij} = \overline{\Delta q_i\, \Delta q_j}/T_a. \tag{7.35}$$

Here $i, j = 3, \ldots, N$ in correspondence with the definition of vector q which is confined to the diffusion plane, and T_a is the mean half-period of the phase oscillation inside the stochastic layer of a guiding resonance. According to eq. (6.18):

$$T_a = \frac{1}{\Omega_g} \ln\left(\frac{32e}{w_s}\right). \tag{7.36}$$

To evaluate the means $\Delta q_i\, \Delta q_j$ we assume the reduced stochasticity to hold for the system under consideration as we did so already for Arnold's example (section 7.2). Recall that it means we take for granted that the boundedness of w (or s) variation is provided by the correlation of the phase τ related to the largest perturbation term only. As to the other phases the relation

$$\overline{\sin^2 \varphi_m^0} \approx \tfrac{1}{2} R_T$$

is implied where the reduction factor R_T (7.13) takes account of a slow diffusion in the peripheral part of the stochastic layer.

A new feature of the problem under consideration as compared to Arnold's example (section 7.2) is a possible interference of several driving resonances. The interference terms appear in evaluating the means $\overline{\Delta q_i\, \Delta q_j}$ and have a form like:

$$\begin{aligned} &\sin(\tau r_m + t b_m + d_m)\sin(\tau r_{m'} + t b_{m'} + d_{m'}) \\ &\quad \to \cos(\tau(r_m \pm r_{m'}) + t(b_m \pm b_{m'}) + d_m \pm d_{m'}). \end{aligned}$$

The interference is important if $r_m = r_{m'}$ and $b_m = b_{m'}$ for a pair of the driving resonances. In this case the interfering terms should be summed up beforehand and then be treated as a single term. If $r_m \neq r_{m'}$ or $b_m \neq b_{m'}$ the corresponding interference term is averaged away for a sufficiently long time of motion. For a small difference $|r_m - r_{m'}|$ a large G in eq. (7.34) helps since it makes the value of τ large. A big G implies a big λ ($G \sim \lambda^2$), or a small perturbation ($G \sim 1/\epsilon$), as well as a very narrow stochastic layer. Yet, all this is not the case for a small $|b_m - b_{m'}|$. Besides, the quantity $(b_m \pm b_{m'})/2\pi$ must be irrational since t is integer.

Now the tensor (7.35) is immediately evaluated from eqs. (7.30), and has the form:

$$D_{ij} \approx \frac{\epsilon^2 R_T}{2T_a\Omega_g^2} \sum_m{}' \rho_i\rho_j V_m^2 A_{2|\xi_m|}^2(\lambda_m)$$
$$\approx \frac{2\pi^2\epsilon^2 R_T}{T_a|\omega^r|^2} \sum_m{}' \frac{\rho_i\rho_j}{\nu_2^2} \left|\frac{2\omega_m}{\Omega_g}\right|^{4|\xi_m|} \frac{V_m^2}{\Gamma^2(2|\xi_m|)} \exp\left(-\frac{\pi|\omega_m|}{\Omega_g}\right). \tag{7.37}$$

Here $\Gamma(x)$ is the gamma function, and ρ $(\rho_3, \ldots, \rho_N)$ stands for the projection of the vector ν onto the diffusion plane. The last expression in eq. (7.37) makes use of the asymptotic representation (A.11) for the MA integral. The relation (7.37) holds under the condition $|v_m| \ll 1$, yet it persists apparently in order of magnitude for $|v_m| \sim 1$ as well (see section 7.6).

The tensor (7.37) is symmetric, and can be diagonalized, so the rate of the Arnold diffusion depends generally on $N-2$ coefficients. For this, however, the sum in eq. (7.37) has to contain at least the $N-2$ terms with linearly independent vectors $\rho(m)$, or the diffusion will be confined to some subsurface of the diffusion surface. The minimal number of resonances which provides the Arnold diffusion in a system of the type (7.21) is equal to 3: guiding, layer, and driving resonances. To ensure the diffusion over the whole diffusion surface the N resonances are required, that corresponds to the complete set of linearly independent vectors ν (or m, including m^g).

7.4. Nekhoroshev's theory

The estimates for the rate of the Arnold diffusion derived in the preceding section are not rigorous, of course, if only because we have failed to analyze the dynamics of the resonance phases and had to assume some suppositions instead. A rigorous estimate from *above* for the rate of the Arnold diffusion has been given by Nekhoroshev [24]. His theory is based upon the separation of the perturbation into resonant and non-resonant parts, and "killing" the latter by means of successive canonical transformations of variables. Below we are going to explain the basic ideas of this theory as well as to represent its main results.

Let us start again from Hamiltonian of the type (7.21) where the terms with amplitudes V_g (not necessarily a single one as in the preceding section) describe the resonant perturbation, and those with V_m do so for the non-resonant one. The resonant terms are defined by the condition:

$$|m^g, \omega| < \eta \tag{7.38}$$

where η is a small quantity to be optimized. Note that for a sufficiently small η the maximal number of linearly independent vectors m^g is equal to $N-1$. If $\eta = 0$ (a multiple resonance, see section 3.3) all vectors m^g belong to the plane tangential to an energy surface at some point $I = I^0$. No matter what interaction may occur between the resonances in question the system never leaves this plane under the influence of these resonances. For a steep unperturbed Hamiltonian H_0 any plane tangential to the energy surface cannot intersect the latter (see section 3.3). Therefore, the resonant perturbation can result only in a bounded and small change in the unperturbed integrals no matter how long the time of motion is, i.e. $|\Delta I| < (\Delta I)_\eta \sim \epsilon^p$ where the constant $p < 1$ depends on the geometry of the energy surface (7.39). This result persists under a sufficiently small η in eq. (7.38), the allowed change $(\Delta I)_\eta$ grows with η though. Thus, the resonant perturbation proves to be not dangerous for the stability of motion unless the unperturbed motion is quasi-isochronous, or the Hamiltonian is not steep.

The non-resonant perturbation terms ($|m, \omega| \geq \eta$) can be "killed", in principle, by a canonical change of variables as this is done in the KAM theory (section 4.6). However, the present problem is a different and more difficult one. In the KAM theory the vector ω is chosen in the special way that

permits to "kill" the perturbation completely, yet the inference of the eternal stability of motion holds only for special initial conditions. Nekhoroshev repudiates the latter restriction, and investigates the motion for arbitrary initial conditions but in this case the vector ω does not generally satisfy the conditions of the KAM theory. Consequently the non-resonant perturbation defies to be "killed" completely even under an arbitrarily small perturbation ($\epsilon \to 0$), and one (Nekhoroshev) succeeds only in estimating the effect of the perturbation from *above*. This sets an *upper* limit for the rate of the Arnold diffusion.

It seems, at the first glance, that the condition $|m, \omega| > \eta$ for non-resonant terms is just one we need to ensure the convergence of the perturbation series. However, the vector ω is no longer constant now as it was in the KAM theory. Although the resonant terms lead only to a bounded variation of the vector I, as has been explained above, they bring about the phase oscillations of vector ω. These oscillations are generally in a resonance with some perturbation harmonics (m, θ) which are called "non-resonant" only provisionally, in the sense that they are off any resonance for the unperturbed, constant frequencies ω. The condition $|m, \omega| > \eta$ permits, therefore, only to bound the amplitude of the terms responsible for the resonances of "non-resonant" perturbation with the phase oscillations. The strength of these resonances grows as η decreases, so we have the contradictory trends in the behavior of resonant and non-resonant components of the perturbation in respect to the variation of η. Optimizing somehow the value of η Nekhoroshev arrives at an asymptotic ($\epsilon \to 0$) estimate for the mean rate of the instability. Here we represent his estimate in a simplified form (for rigorous formulation see ref. [24])*:

$$\overline{|\dot{I}|} \leqslant |\omega|\,|I|\,\epsilon^{1+p} \exp(-1/\epsilon^{q})$$

$$q(N) = \frac{2}{12\zeta + 3N + 14}; \qquad p = \frac{q}{\alpha} \tag{7.39}$$

$$\zeta(H_0) \geqslant \tfrac{1}{2}N(N-1); \qquad \alpha(H_0) \geqslant 1.$$

The equality in the two last expressions is reached for, and only for, a quasi-convex H_0 (see section 3.3). Just as the relation (7.37) for the diffusion rate, Nekhoroshev's estimate predicts an exponential dependence of the instability rate on the perturbation parameter ϵ. However, it leaps to the eye that the power q in eq. (7.39) is rather small: $q(3) = 1/29.5$ even for the minimal $N = 3$. At larger N the quantity $q(N) \sim 1/3N^2$ drops rapidly. Meanwhile, in the relation (7.37) the corresponding power, at least, at the first glance, $q = \frac{1}{2}$ ($\Omega_g \sim \sqrt{\epsilon}$), and does not depend on N at all. Let us consider this question in more detail.

It is clear, first of all, that the exponent in eq. (7.37) depends drastically on the quantities $\omega_m = m, \omega^r$. But this is just those small denominators which spoil the convergence of perturbation series (section 4.6). It is well known that the value of these denominators rapidly drops as $|m|$ grows (section 4.6). On the other hand, the amplitude of perturbation harmonics V_m also decreases with $|m|$. Let the perturbation $V(\theta)$ be an analytical function to which the estimate (4.63) is applicable. Owing to the inequalities (4.77) we may employ the quantity $|m|$ instead of S that is more convenient in the present case. Hence, we can use the estimate (4.75) with $k > N - 1$ (4.80) having exchanged also S for $|m|$.

As follows from eq. (7.24) $\xi_m = M_g/M_{(m)} \sim |m|\xi_g$ where the constant $\xi_g \sim 1/|m^g|$ is determined by the guiding resonance. On the other hand, employing the representation of ξ_m via ν_i (see eq. (7.24)) we

*In the last paper [141] Nekhoroshev has announced an improved estimate with $q(N) = 1/(3\zeta + N + 4)$.

find that $\nu_i \sim |m|$ except ν_2 which is much smaller. The reason for ν_2 to be distinguished from the other components of the vector ν is related simply to the fact that just the driving resonances with a small ν_2 make the principal contribution to the Arnold diffusion since $\omega_m = |\omega^r|\nu_2$ (see eq. (7.24)). Finally, estimating the gamma function according to Stirling's formula we may represent eq. (7.37) in order of magnitude as:

$$-\ln D \sim \frac{\pi\delta}{\Omega_g|m|^k} + 4|m|\xi_g \ln(\Omega_g F|m|^{k+1}) - (1+3k)\ln|m| - 6\ln\Omega_g + U;$$
$$F = \frac{\xi_g}{e\delta}\exp(\sigma/2\xi_g). \tag{7.40}$$

Here U is some nearly constant quantity including also a doubly logarithmic dependence on $|m|$ and Ω_g due to eq. (7.36). For a given perturbation $\sqrt{\epsilon} \sim \Omega_g$ the function (7.40) has the minimum at certain $|m| = m_0$ related to the maximal diffusion rate. As we shall see a bit later the quantity m_0 increases as $\Omega_g \to 0$. Hence for a sufficiently small Ω_g (large m_0) we can keep in the right-hand side of eq. (7.40) the two first terms only. Differentiating eq. (7.40) with respect to $|m|$ we arrive at the equation for m_0:

$$y\ln y = bk\,e^k; \qquad y = F\Omega_g(em_0)^{k+1}; \qquad b = \tfrac{1}{4}\pi\exp(\sigma/2\xi_g). \tag{7.41}$$

If $b = 1$ an obvious solution to this equation were $y = e^k$. For $b \sim 1$ we are looking for the solution of the form $y = e^{k+k_1}$. Substituting it into eq. (7.41) and taking logarithm we find:

$$y = b^{k/(k+1)}e^k = (\tfrac{1}{4}\pi)^{k/(k+1)}e^{k(1+\sigma/2\xi_g(k+1))} \tag{7.42}$$

subject to the condition $(\ln b)/k \approx \sigma/2k\xi_g \ll 1$. Whence:

$$m_0 = \frac{1}{e}\left(\frac{y}{F\Omega_g}\right)^{1/(k+1)} \sim \Omega_g^{-1/(k+1)}. \tag{7.43}$$

We see that m_0 increases, indeed, as Ω_g goes down. Therefore, by substitution of m_0 into eq. (7.40) we can again keep only the two first terms as $\Omega_g \to 0$. Subsequent to some elementary manipulations with the above relations the asymptotic ($\Omega_g \to 0$) estimate for the rate of the Arnold diffusion can be reduced to the form [43]:

$$D \sim D_u\exp(-A\Omega_g^{-1/(k+1)}); \qquad A = 4\xi_g\left(\frac{\pi\delta}{4\xi_g b^{1/(k+1)}}\right)^{1/(k+1)}\ln b \tag{7.44}$$

where D_u is some quantity of the proper dimensions which may be considered as a constant to the accuracy of the estimate. For a large $\sigma \gg 2\xi_g$ and for $k \gg 1$ the quantity $A \approx 2\sigma$. However, for $\sigma \to 0$ the expression for A in eq. (7.44) makes no sense any longer because $\ln b < 0$. This is related, first of all, to the violation of the approximate eq. (A.11) employed above. Indeed, according to the Appendix the accuracy of eq. (A.11) $\sim \gamma^2 = (m/\lambda)^2$, hence, eq. (A.11) holds for $\gamma \ll 1$. In the notations of this section:

$$\gamma = \frac{2|\xi_m|\Omega_g}{|\omega_m|} \approx \frac{2\xi_g}{\delta}\Omega_g|m|^{k+1}; \qquad \gamma_0 \approx \frac{\pi}{2b^{1/(k+1)}} \tag{7.45}$$

where $\gamma_0 = \gamma(m_0)$, i.e. at the maximum of the diffusion rate. Thus, we can apply the estimate (7.44) for $b^{1/(k+1)} \gg 1$ or for $\sigma \gg 2\xi_g(k+1)$, only. As $\sigma \to 0$, and, in particular, for $\sigma = 0$, which corresponds to a smooth perturbation, we can employ eqs. (7.37) and (7.40) in the region $|m| < m_0$ ($\gamma < \gamma_0$) only, that is

somewhat below the maximum of the diffusion rate. If $\gamma/\gamma_0 = (|m|/m_0)^{k+1} \ll 1$ the factor A in eq. (7.44) can be represented in the form:

$$A \approx \frac{2\pi\xi_g}{\gamma}\left(\frac{\gamma\delta}{2\xi_g}\right)^{1/(k+1)}\left(1+\frac{2\gamma}{\pi}\ln b\right) \approx \frac{2\pi\xi_g}{\gamma}. \tag{7.46}$$

The last fairly simple expression is related to the case $k \gg 1$.

Another, more serious reason for γ to be bounded is related to the limitation of small denominators $\omega_m = (m, \omega)$ in Nekhoroshev's theory. Indeed, if $|m, \omega| < \eta$ the corresponding perturbation term becomes resonant according to Nekhoroshev that implies a strictly bounded motion due to this term: $|\Delta I| < (\Delta I)_\eta$. If the size of this bounded domain is much less than the width of the guiding resonance layer ($(\Delta I)_\eta \ll (\Delta I)_r^g$) the corresponding driving resonance (or resonances) will only slightly distort the unperturbed phase oscillations without any long-range diffusion (cf. the effect of a multiple resonance discussed in section 3.3). So the lower limit for ω_m, or the upper limit for $|m|$ of a driving resonance can be assumed as determined roughly by the condition $(\Delta I)_\eta \sim (\Delta I)_r^g$. For a quasi-convex Hamiltonian (section 3.3) the size $(\Delta I)_\eta$ is proportional to the angle β_m between the vector m and the plane orthogonal to the vector ω^r. This angle can be found from the relation: $m, \omega^r = \omega_m \approx |m|\,|\omega^r|\,\beta_m$. Similarly, the width $(\Delta I)_r^g$ is proportional to the angle swing β_g of the vector ω due to phase oscillations at the guiding resonance where (see eq. (3.36)):

$$\beta_g \sim (\Delta\omega)_r^g/|\omega^r| \sim \Omega_g/|m^g|\,|\omega^r|.$$

Hence the above condition $(\Delta I)_\eta \sim (\Delta I)_r^g$ may be written also in the form:

$$\beta_g/\beta_m \sim |m|\Omega_g/|m^g|\omega_m \sim \gamma \sim 1 \tag{7.47}$$

where we have used eq. (7.45) and the estimate $|\xi_m| \sim |m|/|m^g|$ (see above).

Estimate (7.44) for the rate of the Arnold diffusion in a many-dimensional system is similar in functional dependence on the perturbation parameter ϵ to that due to Nekhoroshev (7.39), the power index q in eq. (7.44) depending now on the number of degrees of freedom:

$$q(N) \approx 1/2N$$

since $\Omega_g \sim \sqrt{\epsilon}$ (cf. eq. (7.39)). We have set here $k \approx N-1$ instead of $k > N-1$ (see above) because the consideration in section 4.6 shows that it suffices for k to exceed $(N-1)$ only slightly.

7.5. Another model

In this section we shall consider a simple model close to Arnold's example. The Arnold diffusion in this model was studied at length both numerically and analytically [41]. We have already treated parts of this model in previous sections. Let the unperturbed system be described by the Hamiltonian (3.41). This system consists of the two identical nonlinear oscillators coupled by a linear term $\sim\mu$. This particular model describes qualitatively some real physical systems of practical interest, for example, the motion of a charged particle in the focusing magnetic field of an accelerator or a storage ring. A most important difference of the model from real systems is a large nonlinearity of the oscillators (nonlinearity parameter $\alpha = \frac{1}{3}$ (2.28)). This is done to simplify the analytical analysis of the Arnold diffusion.

We choose the initial conditions of the motion near the main coupling resonance of the system (3.41): $\omega_1^r = \omega_2^r = \omega_0$. In the new variables $(p; \psi)$ (see section 3.3, eqs. (3.43) and around):

$$I_1 = I_0 + p_1 + p_2; \qquad I_2 = I_0 - p_1 + p_2; \qquad I_0 = \tfrac{1}{2}(I_1 + I_2) = \text{const}$$
$$\theta_1 = \tfrac{1}{2}(\psi_1 + \psi_2); \qquad \theta_2 = \tfrac{1}{2}(\psi_2 - \psi_1). \tag{7.48}$$

The Hamiltonian (3.41) may be written approximately in the form:

$$H_\mu \approx H_0(I_0) + 2\omega_0 p_2 + H_1(p_1, \psi_1), \qquad H_1 \approx \frac{\beta^2}{a^2} p_1^2 - \frac{\mu a^2}{2} \cos \psi_1 \tag{7.49}$$

where $H_0(I_0)$ is some constant, and a denotes the oscillation amplitude which we assume to be about the same for both oscillators due to the resonance condition $\omega_1 \approx \omega_2$ and $\mu \ll 1$; $\beta \approx 0.8472$ (section 2.3).

Now we introduce an external perturbation of the two frequencies precisely as was done in section 4.1 for $\mu = 0$ (cf. eq. (4.2)):

$$H(p, \psi, t) = H_\mu(p, \psi) + V(p, \psi, t)$$
$$V = -\tfrac{1}{2} a f_0 (\cos(\tfrac{1}{2}(\psi_1 + \psi_2) - \tau_+) + \cos(\tfrac{1}{2}(\psi_1 + \psi_2) - \tau_-)) \tag{7.50}$$

where we have retained the low frequency terms only and changed θ_1 for ψ_1, ψ_2 (7.48); $\tau_\pm = \Omega_\pm \cdot t + \tau_0^\pm$.

Let the initial conditions be close to the separatrix of coupling resonance $\omega_1^r = \omega_2^r$. First we evaluate the changes in both H and H_1 over a period of the unperturbed motion near the separatrix employing the technique of sections 4.4; 7.2; 7.3. We choose the perturbation frequencies as follows:

$$\Omega_+ - \omega_0 \approx \omega_0 - \Omega_- \equiv (\delta\omega) > 0. \tag{7.51}$$

This corresponds to a symmetric disposition of the two driving resonances in respect to the initial position of the system in the frequency plane (see fig. 4.2). Then, in dependence of the sign of $\dot{\psi}_1 \approx 2\Omega_\mu \sin(\psi_1/2)$ (or that of ψ_1) either one or the other driving resonance for which $\lambda > 0$ "works"; $\Omega_\mu = \beta\sqrt{\mu}$ is the frequency of small phase oscillations (3.45). Let t^0 be, as usual, the time when $\psi_1(t^0) = \pm\pi$, and $\psi_2 = 2\omega_0 t + \chi$ (see eq. (7.49)). We have:

$$\begin{aligned}
\dot{H} &= \partial V/\partial t \approx -\tfrac{1}{2} a f_0 \omega_0 \sin(\tfrac{1}{2}\psi_1 \mp t\,(\delta\omega) + \tfrac{1}{2}\chi - \tau_0^\pm) \\
&= \mp \tfrac{1}{2} a f_0 \omega_0 \sin(\tfrac{1}{2}|\psi_1| - t\,(\delta\omega) \pm \tfrac{1}{2}\chi \mp \tau_0^\pm) \\
&\to \mp \tfrac{1}{2} a f_0 \omega_0 \cos(\tfrac{1}{2}|\psi_1| - \tfrac{1}{2}\pi - (t - t^0)\,(\delta\omega)) \sin(\tfrac{1}{2}\pi - t^0\,(\delta\omega) \pm \tfrac{1}{2}\chi \mp \tau_0^\pm) \\
\dot{H}_1 &= -\dot{\psi}_1\, \partial V/\partial \psi_1 \to \tfrac{1}{4} a f_0 \Omega_\mu \cos(\psi_1 \mp t\,(\delta\omega) + \tfrac{1}{2}\chi - \tau_0^\pm) \\
&= \tfrac{1}{4} a f_0 \Omega_\mu \cos(|\psi_1| - \pi - (t - t^0)\,(\delta\omega)) \cos(\pi - t^0\,(\delta\omega) \pm \tfrac{1}{2}\chi \mp \tau_0^\pm).
\end{aligned}$$

Here the arrows indicate that only the main terms, which make the principal contribution to the MA integral to be evaluated, are retained. The alternating signs coincide with those of $\dot{\psi}_1$ (or ψ_1) and $(\delta\omega) \ll \omega_0$ is assumed so that $\Omega_+ \approx \Omega_- \approx \omega_0$. Applying the MA integral we get:

$$\Delta H = \mp \frac{a f_0 \omega_0}{2\Omega_\mu} A_1(\lambda) \cos \varphi, \qquad \Delta H_1 = -\frac{a f_0}{4} A_2(\lambda) \cos \varphi \tag{7.52}$$
$$\lambda = \delta\omega/\Omega_\mu; \qquad \varphi = t^0\,(\delta\omega) \mp \tfrac{1}{2}\chi \pm \tau_0^\pm.$$

We see that both changes, ΔH and ΔH_1, depend on the same phase φ. Hence, as long as the phase ψ_1 is rotating the sign in eq. (7.52) for ΔH remains unchanged, and the variation of H is bounded and

small due to the confinement of H_1 within the stochastic layer. The Arnold diffusion is possible, therefore, in the case of the ψ_1 oscillation only.

Since the change ΔH_1 depends on the single phase φ the mapping which describes the motion inside the stochastic layer of the coupling resonance under consideration will be almost identical to the mapping (6.9) in section 6.2 for the simplest example of the dynamics inside the stochastic layer. In any event, we can apply literally all the technique of section 4.4 to get:

$$\bar{s} = s - (\cos\varphi)/\lambda; \quad \bar{\varphi} = \varphi - \lambda \ln|\bar{s}| + G \pm \tau_0 \tag{7.53}$$

where $s = w/w_s$; $w = (2H_1/\mu a^2) - 1$; $G = \lambda \ln(32/w_s)$ and τ_0 is related to the constants χ, $\tau_0^{\pm}$. The width w_s of the stochastic layer can be evaluated employing eqs. (6.8); (7.52), and is equal for $\lambda \gg 1$ to:

$$w_s = \frac{4\pi f_0 \lambda^2}{\mu a} e^{-\pi\lambda/2}. \tag{7.54}$$

It is instructive to compare the mapping (7.53) with that in Arnold's example (7.16) or with a more general mapping (7.34) for a conservative system. In both latter cases the minimal number of phases in the mapping equals two that corresponds to three resonances altoghether (see the end of section 7.3). In the present case the motion is determined by the single phase (7.53), yet the total number of resonances is still three. The presence of the two driving resonances is important. With a single driving resonance not only the diffusion rate drops drastically since for one sign of $\dot{\psi}_1$ the parameter $\lambda < 0$ but also any long-range diffusion becomes impossible at all. This is related to the fact that the function $A_1(\lambda)$ changes sign if λ does so (see eq. (A.7) in the appendix), hence, the variation of ΔH goes in proportion to that of ΔH_1 even in the case of the ψ_1 oscillation (see eq. (7.52)), and both quantities H, H_1 remain strictly bounded.

A principal distinction of the mapping (7.53) from the whisker mapping (6.9) similar to the former in appearance is that the present mapping (7.53) describes, nevertheless, a many-dimensional system (7.50) in which the Arnold diffusion is possible. The variation of the total energy for this system is determined by the sum (see eq. (7.52)):

$$H(t) = H(0) + \frac{a f_0 \omega_0}{2\Omega_\mu} A_1(\lambda) \sum_i^t (\mp \cos\varphi_i) \tag{7.55}$$

where successive phase values φ_i are given by the mapping (7.53). If we assume that, taking account of sign alternation in eq. (7.55), i.e. in the case of the ψ_1 oscillations at the guiding resonance, the reduced stochasticity (see section 7.2) takes place then the diffusion rate in H is immediately derived from eq. (7.55) and is given for $\lambda \gg 1$ by the expression:

$$D_T = \frac{\overline{(\Delta H)^2}}{T_a} = \frac{(\pi a f_0 \omega_0)^2}{\Omega_\mu} \frac{R_T}{L(s)} e^{-\pi\lambda}. \tag{7.56}$$

Here R_T is the reduction factor (7.13), and $L(s)$ stands for the logarithm which determines the mean period of motion in the stochastic layer (see eq. (6.18)):

$$L(s) = \ln\left(\frac{32e}{|w|}\right) = \frac{\pi\lambda}{2} + \ln\left(\frac{8e}{\pi} \frac{\mu a}{|s| f_0 \lambda^2}\right). \tag{7.57}$$

In what follows we shall need also the diffusion rate across the stochastic layer, i.e. that in H_1, since we are interested, mainly, in a long-range Arnold diffusion when the time of motion is much longer than the time interval required for the diffusion to cross all the stochastic layer. Comparing H

and H_1 in eqs. (7.52) we get ($\lambda \gg 1$):

$$D_s = \overline{(\Delta H_1)^2}/T_a = D_T\,(\delta\omega/\omega_0)^2 \tag{7.58}$$

that is the diffusion across the layer is, unfortunately, much slower than that along the layer as soon as $(\delta\omega) \ll \omega_0$. The energy half-width for the part s of the stochastic layer is equal to: $\delta H_1 = s w_s \mu a^2/2$. Let us introduce the quantity:

$$\delta_s^2 = \frac{D_s t}{\overline{(\delta H_1)^2}} = t\,\frac{\beta^3}{4}\,\frac{R_T}{s^2 L(s)}\,\frac{\mu^{3/2}}{(\delta\omega)^2} \tag{7.59}$$

where t is the time of motion. For $\delta_s \gg 1$ the diffusion has enough time to "fill up" the stochastic layer and to spread out far along the layer. If, however, $\delta_s \lesssim 1$ a finite width of the stochastic layer is insignificant, and we have a short-range ("near"), or local diffusion.

7.6. Numerical experiments

The simplest numerical experiments were carried out with the mapping (7.53) which approximately describes the Arnold diffusion in system (7.50). As was explained in the preceding section the diffusion rate depends on the behavior of the sum $\Sigma(\pm\cos\varphi_i)$ (7.55). A sequence of phases φ_i is determined by the mapping (7.53), hence, the sum $\Sigma\cos\varphi_i$ is bounded. The numerical experiments have been intended, thus, to find out the influence of sign alternation, or equivalently, of the phase shift by π, on the phase dynamics. That phase shift cannot affect, of course, the reduction of the diffusion rate at the edges of the stochastic layer (factor R_T), yet the shift may destroy such specific phase correlations which result in a strict boundedness of the sum $\Sigma\cos\varphi_i$. Therefore, the quantity:

$$R_E = \frac{2}{n}\overline{\left(\sum_{i=1}^{n}(\pm\cos\varphi_i)\right)^2} \tag{7.60}$$

was computed in numerical experiments. A typical time of motion for one trajectory of the system (7.53) was $t = 10^6$ iterations. The mean value of R_E (7.60) was calculated over 10 sections of a trajectory with $n = 10^5$. The values obtained happened to be within the interval (0.25–1.09). The mean value over all (9) trajectories

$$\langle R_E\rangle \approx 0.52. \tag{7.61}$$

Within (fairly large) fluctuations no dependence on the mapping parameters λ = 3–9 and G = 9–99 was observed. The value (7.61) is close to the expected one ($R_T \approx \frac{1}{3}$, eq. (7.13)).

In the other series of experiments the mapping (7.34) in the simplest case of the two phases (τ; $\varphi = r\tau + bt + d$) was employed. Recall that Arnold's example may be reduced to precisely that mapping (7.16). The dependence of the quantities

$$R_E = \frac{2}{n}\overline{\left(\sum_{i=1}^{n}\sin\tau_i\right)^2}; \qquad R'_E = \frac{2}{n}\overline{\left(\sum_{i=1}^{n}\sin\varphi_i\right)^2} \tag{7.62}$$

on v was studied under the same conditions as in the previous series of experiments. The results are presented in table 7.1 where the values of R_E, R'_E are given averaged over several ($\sim$10) trajectories for every value of v. The mapping (7.34) has already quite a lot of parameters, so any systematic study of their influence has not been done. However, it was observed that the results are sensitive to the value of G. For $G \lesssim 10$ the big fluctuations of R_E, R'_E began apparently due to the correlation between

Table 7.1

v	0.11	0.33	0.5	1
R'_E	0.427	0.233	0.169	0.130
R_E	0.00712	0.0233	0.0388	0.137

the phases τ and φ (see section 7.3). Any significant dependence on the other parameters has not been observed.

Data in table 7.1 clearly indicate the difference in the values $R_E \ll R'_E$ up to $v \sim 1$. This confirms the reduced stochasticity assumption accepted in sections 7.2, 7.3 and 7.5 as to the fact that the correlations are significant only for the phase of the largest term in the sum of eq. (7.34). For $v \ll 1$ the value of R'_E for the "free" phase φ is close to the value (7.61) obtained from a different kind of experiment. A discrepancy between them is certainly within the big fluctuations. The main origin of the latter is a prolonged "sticking" of the trajectory somewhere in the peripheral part of the stochastic layer (cf. section 5.1).

Most thorough numerical experiments related to the Arnold diffusion were carried out with the model described in section 7.5. The preliminary results are presented in ref. [41]. A detailed description of these studies will be given in a separate paper to appear [41]. Below we are going to consider some of these results related to the mechanism of the Arnold diffusion.

The model parameters chosen for a series of numerical experiments were as follows. The perturbation periods (7.50): $T_+ = 25$; $T_- = 29$ time units of the system in question. Since $T_\pm$ are integer the perturbation is of the period $25 \times 29 = 725$. This permitted to tabulate in advance the time dependence of the perturbation that materially cut down the computation time. The perturbation frequencies: $\Omega_+ = 0.2513$; $\Omega_- = 0.2167$. The initial conditions were taken for a system to be about half-way between the two driving resonances (see fig. 4.2). That is the frequency $\omega_0 \approx (\Omega_+ + \Omega_-)/2 = 0.234$ which corresponds to the oscillation amplitude $a \approx 0.27$, the detune being $(\delta\omega) \approx 0.0173$. The initial conditions were chosen, further, in such a way to put the system inside the stochastic layer of the coupling resonance, namely: $\psi_1 = \theta_1 - \theta_2 \approx \pi$ (7.49), that means the out-of-phase oscillations. Unlike the preliminary experiments a very close disposition of the driving resonances (small detune $\delta\omega$) was accepted to advance into the region of a smaller perturbation (f_0, μ). Both perturbation parameters f_0, μ were decreased in proportion to keep the ratio $f_0/\mu = 0.01$ constant. Such a small ratio was accepted not only to suppress the effects of higher approximations (see section 6.1) but, mainly, to shift the overlap of the resonances to as large μ as possible.

According to the data in section 4.1 (eq. (4.7)) the critical amplitude of the force at which the overlap of the two driving resonances starts is equal to $f_E \approx 2.55 \times 10^{-5}$. For $f_0/\mu = 0.01$ it leads to the condition $\mu < 2.55 \times 10^{-3}$, or $1/\sqrt{\mu} > 20$, to get rid of the overlap. A more stringent condition to meet is related to the overlap of the three resonances including the coupling one (section 4.1). Applying the relation (4.10) we get: $1/\sqrt{\mu} > 62$. This condition, however, takes account of the unperturbed resonance width. For the two driving resonances a similar condition would lead to the critical amplitude $f_T \approx 5.76 \times 10^{-5}$ (4.7) and, respectively, to the unequality: $1/\sqrt{\mu} > 13$. If we assume that the correction factor due to the resonances of higher harmonics is the same in both cases, i.e. for the overlap of both 3 and 2 resonances, the above border will draw up to $1/\sqrt{\mu} > 95$.

The main quantity to be measured was the diffusion rate in the energy. The technique for its measurement is described in section 4.1. The dependence of the diffusion rate on the coupling parameter $1/\sqrt{\mu}$ as well as some other auxiliary quantities are presented in table 7.2. The data for

every value of μ were obtained from a single trajectory over the time of motion $t = 10^6$ except the cases marked by asterisk for which the mean values over a number of trajectories are given, those numbers being indicated in brackets next to the asterisk.

The basic problem in that sort of experiment is the reliability of the data obtained, that is an effective exclusion of various side processes such as the bounded energy oscillations, computation errors and the like. Checking the results was carried out in different ways. First, the diffusion rate was computed for the two averaging intervals of time (D_4, D_5) which differ by an order of magnitude as described in section 4.1. An approximate equality $D_4 \approx D_5$ indicates that the motion is really diffusionlike. If, however, $D_4 \gg D_5$ some processes of a different nature are obviously at hand. The data in table 7.2 show that up to $1/\sqrt{\mu} \approx 200$ both diffusion rates are fairly close, so we have a certain diffusion process, indeed. It is worth noting, however, that even in this region $D_4 > D_5$ always that indicates a certain, insignificant though, contribution of some non-diffusion processes. It will be more correct, therefore, to take $D_5 = D_E$ as the experimental diffusion rate. For $1/\sqrt{\mu} > 200$ both rates

Table 7.2
Arnold diffusion

$\frac{1}{\sqrt{\mu}}$	$-\lg D_4$ / $-\lg D_5$	F_{21}	δ_s	R_4 / R_5	R_N
23.93	11.77	16.1	411	0.173	
*(11)	11.96	83.5		0.110	0.04
43.93	12.25	103	171	1.179	
	12.49	1100		0.677	0.75
63.93	13.74	14.5	98	0.418	
	13.79	115		0.372	0.79
83.93	14.64	13.6	64.5	0.438	
*(11)	14.71	42.1		0.372	1.20
103.93	15.32	12.9	46.5	0.637	
	15.52	16.4		0.407	1.48
123.93	16.44	9.1	35	0.307	
	16.47	20.9		0.287	1.07
143.93	17.06	8.5	27.5	0.440	
*(6)	17.26	8.5		0.278	0.95
163.93	17.72	8.0	22	0.532	
*(11)	17.97	21.7		0.305	0.88
183.93	18.44	3.0	18	0.0544	
	18.83	17.4		0.224	0.50
203.93	18.69	1.7	15.5	1.577	
	18.97	40.3		0.828	1.35
223.93	19.01	—	13	3.788	
	19.59			0.997	1.13
243.93	19.25	—	11	10.72	
	20.10			1.507	1.13
263.93	19.28	—	9.5	48.43	
	20.58			2.406	1.16
283.93	19.25			243.3	
	21.28	—	8.5	2.268	0.67
303.93	19.46	—	7.5	701.2	
	21.21			12.31	2.18
323.93	19.58	—	6.5	2421	
	21.91			11.22	1.15

disagree a lot, actually D_4 ceases to decrease with μ and is determined apparently by the computation errors. The rate D_5 is less sensitive to the errors due to a better averaging, yet under the lack of checking any inferences about the nature of the diffusion in this region are unreliable and, at least, preliminary.

The second checking method employed the "switching-off" one of the two driving resonances. The ratio of the diffusion rate for the two driving resonances to the double rate for one resonance is also given in table 7.2 (quantity F_{21}) for both D_4, D_5, respectively. The fact that always $F_{21} \gg 1$ (this is especially true for D_5) indicates some highly specific mechanism of the diffusion which requires necessarily the two driving resonances. These raise our confidence in the interpretation of the experimental data as the Arnold diffusion the theory of which naturally explains also this specific peculiarity of the motion.

In table 7.2 one more auxiliary quantity δ_s (7.59) is given. Roughly speaking, it shows by how many times the range of the free diffusion would exceed during time t the width of the stochastic layer if the latter did not confine it. The values of δ_s in table 7.2 are calculated for $t = 5 \times 10^5$ which roughly corresponds to the mean time of diffusion according to the particular procedure of measurement for the diffusion rate (see section 4.1). As a width of the stochastic layer the width of its central part has been taken ($s = \frac{1}{4}$) since in the peripheral part of the layer the diffusion rate is already considerably slower. We see that, at least, in the region up to $1/\sqrt{\mu} \approx 200$ the quantity $\delta_s \gg 1$, so it is really the case of a long-range Arnold diffusion we are just interested in. Note that δ_s varies roughly in proportion to F_{21} (for D_5). This related to the fact that the "switching-off" one of the driving resonances influences the diffusion rate the stronger the longer is the range of diffusion since in a short-range diffusion the two driving resonances have no advantage as compared to a single one.

Still another checking method was based upon the measurement of the diffusion rate in dependence on the initial conditions of motion. A distinctive feature of the Arnold diffusion is the localization of the diffusion within the narrow stochastic layers. Usually, the initial conditions were taken in the form: $P_1 = P_2 = 0$; $X_1 = -X_2 = a$ (the out-of-phase oscillations, see section 7.5 and eq. (3.41)). This corresponds to $\psi_1 = \pi$, that is to the unstable equilibrium at the coupling resonance. If we started with the in-phase oscillation instead ($X_1 = X_2$) the system were in the stable center of the coupling resonance. To get on the separatrix some detune in frequency (amplitude) between the two oscillators is necessary. Let us set:

$$X_1 = a + d; \qquad X_2 = a - d. \tag{7.63}$$

In fig. 7.3 an example of the dependence of diffusion rate on the initial conditions is given for

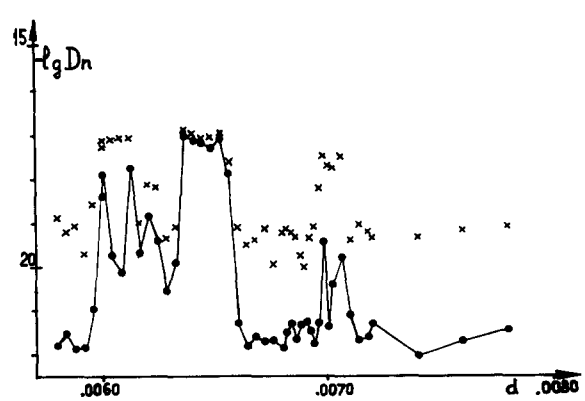

Fig. 7.3. Rate of the Arnold diffusion versus initial conditions for system (7.50): × – values of $-\lg D_4$; ● –the same for $-\lg D_5$; $1/\sqrt{\mu} = 143.93$.

$1/\sqrt{\mu} = 143.93$. A stochastic layer of the width $\Delta d_E \approx 1.8 \times 10^{-4}$ with the center at $d_E \approx 0.65 \times 10^{-2}$ is clearly seen. On the outer side of the layer the diffusion rate D_5 drops by 4 orders of magnitude, and the values D_4, D_5 differ by 2 orders that indicates a negligible background to the Arnold diffusion. The background is larger by an order of magnitude on the inner side, yet it is still negligible as compared to the diffusion inside the layer. At some distance on both sides of the layer the diffusion rate grows and depends on d in a complicated manner. The values D_4, D_5 differ there, at least, by an order of magnitude that indicates a confinement of the diffusion. In all likelihood, the increase, mainly, of D_4 is related to some resonances outside the stochastic layer.

The shift d (7.63) in the oscillation amplitude is related to the frequency: $\omega_{1,2} = \omega_0 \pm \beta d$ (see eq. (2.26)). Employing eq. (3.45) for the width of the separatrix we can predict the theoretical position of the layer center:

$$d_T = \sqrt{\mu} \approx 0.70 \times 10^{-2}; \qquad \Delta d_T = \tfrac{1}{2} w_s\, d_T \approx 1.38 \times 10^{-4}. \tag{7.64}$$

The theoretical relation for the width of the layer (Δd_T) is derived from eq. (7.49) for H_1 (at $\psi_1 = 0$) via the quantity w_s (7.54) (cf. section 6.4). The numerical values are given for $1/\sqrt{\mu} = 143.93$. The layer turns out to be shifted somewhat inside the resonance ($d_E/d_T \approx 0.93$) apparently due to a distortion of the unperturbed separatrix by the driving resonances. The measured width of the layer also agrees quite well with the expected one: $\Delta d_E/\Delta d_T \approx 1.30$. Thus, the checking by the initial conditions confirms also that we do observe the Arnold diffusion.

In the two last columns of table 7.2 the main results of the numerical experiments on the Arnold diffusion are given. The quantity R was calculated by comparison of the experimental values for the diffusion rate (D_4 or D_5) with the theoretical relation (7.56), in which we substituted the value $s = 1$ since during a sufficiently long time interval the diffusion spread over the whole stochastic layer. The dependence of R versus $1/\sqrt{\mu}$ is plotted also in fig. 7.4. The two different regions with the border at $1/\sqrt{\mu} \approx 200$ between them are clearly seen in the figure. For $1/\sqrt{\mu} < 200$ the quantity R remains approximately constant except the two leftmost points affected undoubtedly by the resonance overlap (see above). The remaining 7 points lead to the mean values:

$$\langle R_5 \rangle = 0.321; \qquad \langle R_4 \rangle = 0.474 \tag{7.65}$$

to be compared with the values $R_E = 0.52$ (7.61) from the mapping (7.53). Owing to large fluctuations of R it is difficult to judge if there is any other origin for the discrepancy. In the region under consideration ($63.93 \leq 1/\sqrt{\mu} \leq 183.93$) the values of R_5 differ less than by a factor 2 whereas the diffusion rate drops over this region more than by 5 orders of magnitude! We can regard, therefore, that the theory satisfactorily describes the Arnold diffusion in this particular region. It is important that the theory does not contain any arbitrary parameters to be fitted; the mean value of R_5 (7.65) is close to the expected $R_T = 0.33$ (7.13).

Apart from the mean diffusion rate the energy distribution was studied in the following way. For each of 10 trajectories with the same μ and the time of motion $t = 10^6$ but with different initial conditions the 100 means $\bar{H}_i$ over successive intervals $\Delta t = 10^4$ as well as 99 differences $\Delta\bar{H} = \bar{H}_{i+1} - \bar{H}_i$ characterizing the change in energy over each interval $\Delta t = 10^4$ were computed. The total amount (990) of the quantities $\Delta\bar{H}$ was distributed over 20 bins according to their $\Delta\bar{H}$ values, and the histogram $n_i(|\Delta\bar{H}|)$ was plotted in terms of coordinates $(\Delta\bar{H})^2$; $\ln(n_i)$. An example of the histogram is presented in fig. 7.5 for $1/\sqrt{\mu} = 83.93$. At large $|\Delta\bar{H}|$ the distribution function has the Gaussian "tail": $\ln(n_i) \sim -(\Delta\bar{H})^2$, yet at small $|\Delta\bar{H}|$ a substantial deviation from the Gaussian distribution is observed. It is natural to attribute the Gaussian "tail" of the distribution function to the most rapid diffusion in

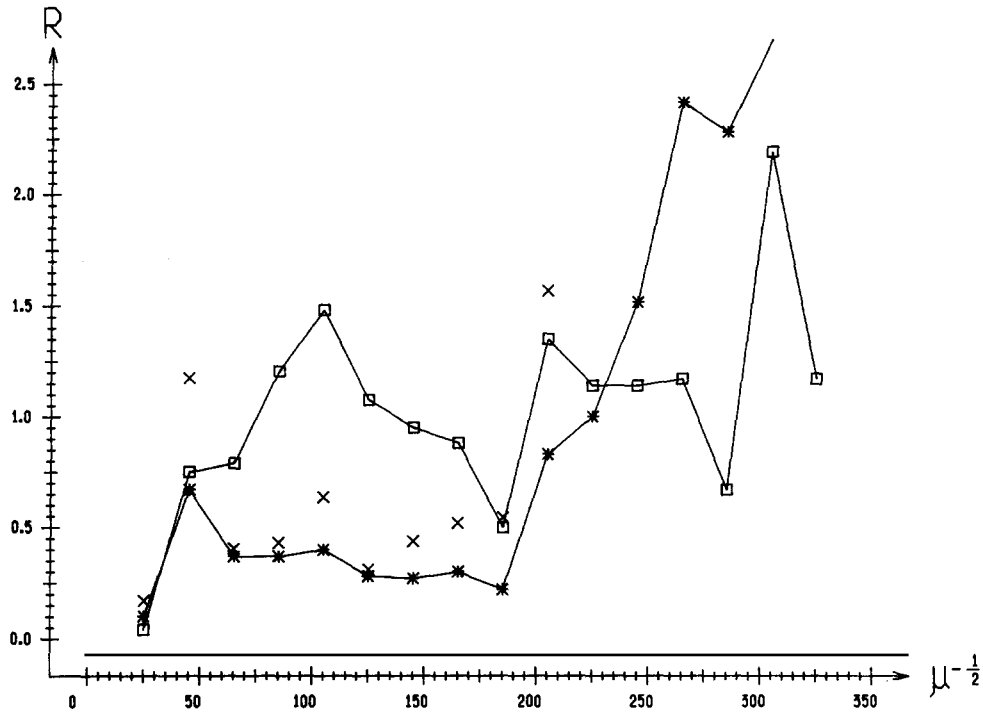

Fig. 7.4. Rate of the Arnold diffusion, comparison with theory: $*-R_5$; $\times-R_4$; $\square-R_N$.

the central part of the stochastic layer whereas a slow diffusion in the peripheral part increases the distribution function at small $|\Delta\bar{H}|$. If so, the diffusion rate for the "tail" must be described by the theoretical relation (7.56) with $R^{\text{tail}} \approx 1$ and $s \approx \frac{1}{4}$. From the distribution in fig. 7.5 we can immediately determine only the rate D_4^{tail} (at the "tail"). If we assume that $D_4^{\text{tail}}/D_4 \approx D_5^{\text{tail}}/D_5$ the rate D_5^{tail} can be also recalculated. We get: $R_5^{\text{tail}} = 0.937$, that is very close to unity, indeed. It should be mentioned, however, that the same recalculation for $1/\sqrt{\mu} = 163.93$ gives an appreciably worse agreement with the theory: $R_5^{\text{tail}} = 0.463$.

We now turn back to fig. 7.4 and consider the region of small μ. For $1/\sqrt{\mu} > 200$ the theoretical

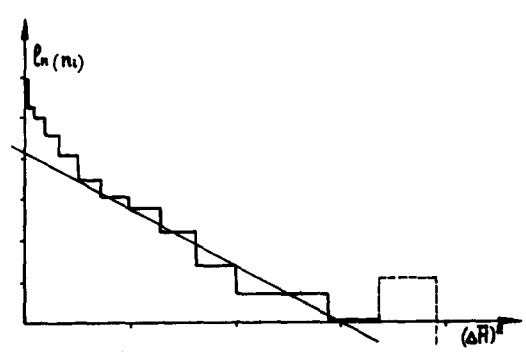

Fig. 7.5. Fluctuations of the Arnold diffusion; straight line fits the distribution "tail", ignoring the right-most interval which is obviously off the regularity owing to a poor statistic.

relation (7.56) is certainly inapplicable. It is true, as was mentioned above, that just in this region both diffusion rates (D_4, D_5) differ considerably, so one has some reason to believe that a deviation from the theory is simply due to the computation errors. However, there is another, more interesting explanation. The point is that the theoretical relation (7.56) takes into account only three resonances of the first approximation – the coupling and two driving ones. Meanwhile, for a sufficiently small perturbation the resonances of higher harmonics come into play, and we get into the *Nekhoroshev region* where the diffusion rate drops with the perturbation substantially slower (section 7.4). Let us try to compare the numerical data with the estimates of the type (7.39) or (7.44). The estimate (7.44) may be represented also in the form:

$$D_N = D_u \exp(-A/\mu^q); \qquad \ln\ln(D_u/D_N) = \ln A + 2q \ln(1/\sqrt{\mu}) \tag{7.66}$$

where D_u, A, q are the unknown parameters to be fitted to the numerical data. It has turned out that the data on D_5 do fit the dependence (7.66) not only for $1/\sqrt{\mu} > 200$ but also, much to our surprise, for $1/\sqrt{\mu} < 200$! The least square fit gives for the parameters of eq. (7.66) ($1/\sqrt{\mu} \geqslant 43.93$):

$$q = 0.178; \qquad A = 5.53; \qquad D_u = 7.5 \times 10^{-4}. \tag{7.67}$$

The values of $R_N = D_5/D_N$ for these values of the parameters are given in the last column of table 7.2 as well as in fig. 7.4. Except the largest perturbation ($1/\sqrt{\mu} = 23.93$) the ratio R_N changes by a factor of 4 with the root-mean-square deviation of ±27 per cent whereas the variation of the diffusion rate D_5 over this range of μ comprises almost 10 orders of magnitude! There is nothing but astonishment for such a simple relation (7.66) to describe quite acceptably all this enormous range of the Arnold diffusion. The empirical parameter $1/q \approx 5.60$ may be compared to the quantity $2(k+1) = 2N$ (7.48). By the sense of parameter k the quantity N is equal to the number of basic frequencies which combinations determine the resonances of higher harmonics (section 4.5). In the present case $N = 4(\omega_1, \omega_2, \Omega_+, \Omega_-)$. Whence: $1/q = 8$ that is more or less close to the experimental value. The other two parameters (A, D_u) remain, however, barely empirical. Here is a significant distinction of the simple (and more general?) estimate (7.66) from a more complicated and restricted relation (7.56) which gives, however, the absolute estimate for the rate of the Arnold diffusion.

7.7. How does the Arnold diffusion "work"?

It seems at first glance that the Arnold diffusion is of no importance for real physical systems since it occurs only for very special initial conditions – inside the exponentially narrow stochastic layers. On the other hand, the full set of those layers is everywhere dense in the phase space. Subject to such controversial conditions the problem of the dynamical motion becomes actually improper since one cannot resolve where a real system is in fact, whether it is in a stability domain or in a stochastic layer. A method for the regularization of this problem is described in ref. [43]. It consists of the introduction of some additional, or external (in respect to the system) diffusion with the rate $D_0 \to 0$ independent of initial conditions. Such a diffusion is always present in any real system. For instance, in the case of a charged particle moving in accelerators it is the Coulomb scattering in the residual gas. Taking account of the external diffusion the exact initial conditions become insignificant since the external diffusion brings the system into one of the stochastic layers from time to time. The resulting motion depends then on the ratio D_A/D_0 where D_A is the rate of the Arnold diffusion. It depends, mainly, on the guiding resonance. For the analytical perturbation (section 4.5):

$$\Omega_g \sim \sqrt{\epsilon V_g} \sim \exp(-\sigma|m^g|/2).$$

Now, if one makes use of estimate (7.44) the rate of the Arnold diffusion, as a function of the harmonic number $|m^g|$ for guiding resonance, drops according to the *doubly exponential law* [43]

$$D_A \sim D_u \cdot \exp(-A \cdot \epsilon^{-1/(2k+2)} \cdot \exp\{\sigma|m^g|/(2k+2)\}). \tag{7.68}$$

It shows, in particular, that for any given D_0 there exists a *finite* set of the guiding resonances with $D_A \gg D_0$ which we shall call the *working resonances*. The guiding resonances of higher harmonics are insignificant in the presence of the external diffusion owing to $D_A \ll D_0$. The motion of a system under these conditions proceeds, roughly speaking, in the following way. During some time T_0 the system enters, due to the external diffusion, the stochastic layer of one of the working resonances, and then rapidly moves, due to the Arnold diffusion ($D_A \gg D_0$), along a set of stochastic layers over the whole phase space (or over an energy surface for a conservative system), the total diffusion time being determined, mainly, by the time of the initial diffusion T_0. This time is much less than the time of the purely external diffusion (for $D_A = 0$) since the mean spacing between the working resonances is typically small ($\Delta_0 \ll |\omega|$ for some $|m^g| = m_0 \gg 1$). Very roughly, the Arnold diffusion reduces the total diffusion rate by a factor of $(|\omega|/\Delta_0)^2$. This factor is, by the way, the larger the slower is the rate of external diffusion D_0 since the resonances of higher harmonics and, hence, of a smaller spacing between them begin to work. Very roughly, $\Delta_0(m_0) \sim |\omega|/N_r$ where $N_r \sim m_0^N$ is the total number of the resonances with the harmonic numbers up to $\sim m_0$, whence:

$$(|\omega|/\Delta_0)^2 \sim 1/m_0^{2N}. \tag{7.69}$$

The quantity $m_0(D_0)$ is determined by the estimate (7.68) from the condition $D_A(m_0) \sim D_0$.

Owing to an extremely slow rate the Arnold diffusion may play a role only in the systems with a negligible energy dissipation. The motion of a charged particle in a magnetic field gives an example of such a system. It is especially true in case of heavy particles, for instance, protons for which the energy losses due to the electromagnetic radiation can be completely neglected in any reasonable conditions. As was mentioned already above the Arnold diffusion has been observed apparently in the experiments with electrons in a magnetic bottle [101] (see also ref. [126]). It was conjectured also that such a diffusion could be responsible for a slow "inflation" of the proton colliding beams [103]. In any event, the special experiments with an additional nonlinear lens indicated a strong influence of the oscillation nonlinearity on the lifetime of a proton beam [104]. Recently some interesting implications of a slow instability of the proton motion in the geomagnetic field as to the dynamics of the radiation belts were discussed [126]. The instability has been studied in the model experiments with electrons and was, in all likelihood, the Arnold diffusion again.

Another system very close to a conservative one is the Solar system. The Arnold diffusion may result here in the formation of so-called Kirkwood's gaps in the family of the minor planets-asteroids. The gaps mean dips in the distribution of asteroids according to their rotation frequency around the Sun, and they occur near some lower harmonic resonances with Jupiter's motion. The related estimates are given in ref. [43]. For a different explanation of the gaps see, e.g., [127, 128] and also [49] (the book by Siegel and Moser, §34). Still another, and quite unexpected conjecture is due to Shuryak [20].

It should be emphasized, however, that the quantitative theory of the Arnold diffusion, if only a semi-empirical one, has not yet constructed so far for any real physical system apart from a special model (section 7.5).

Since the Arnold diffusion is extremely slow many other factors omitted in the first analysis may come into play. One of them is related to the quantum peculiarities of motion. The discussion of quantum effects is beyond the framework of this paper which is devoted to the classical nonlinear oscillations. Nevertheless, it will be not superfluous, I believe, just to mention that some surprise restrictions of the Arnold diffusion appear in the quantum mechanics. Owing just to an extremely slow rate of this diffusion the restrictions may be of importance even in a system which would seem to be perfectly classical. First interesting estimates of this sort were made by Shuryak [20].

8. Concluding remarks, problems

The stability analysis of nonlinear oscillations presented in this paper relies essentially upon the simple criterion of the resonance overlap (section 4.1). Subject to that overlap the influence of a regular, particularly, periodic perturbation is equivalent to that of a "random" perturbation. Such a queer, at the first glance, "regeneration" of the perturbation is confirmed, nevertheless, by all the numerical experiments some of which have been described above. The overlap criterion is fairly simple and efficient, yet there are still a lot of related problems to be solved. Here are some of them.

1) The analysis of a basic phenomenon of the nonlinear oscillation – the stochastic instability – was made above via a set, or, better to say, a chain of successively simplified models. As a rule, all those models are not structurally stable. Therefore, the problem of the reliable esteem as to the influence of various approximations on the evolution of a system, and especially on a long-term evolution arises. In this respect it would be of importance to develop a theory of the restricted structural stability (see section 5.5).

2) An important example of this problem is the impact of a specific ("integer") representation of any quantity in computer on the dynamics of motion (section 5.5). We mean that any quantity in computer is represented by a finite number of digits. This particular problem is the more important that the numerical experiments by computer have served already a good deal and will do so as a basic method of investigation in the field of nonlinear mechanics.

3) As we have seen even in extremely simple cases the stochastic motion possesses a regular component of the motion which has a fairly complicated structure (section 5.5). Even though the numerical experiments justify the neglecting this component far enough inside the stochastic region a more thorough study of the problem is desirable. It should be emphasized that so far we have even no reliable estimate for the measure of the stochastic component, could this measure turn out to be very small (section 5.5)?

4) For the simple mappings considered the inference of the "random" motion means simply that certain phase values are random and statistically independent of each other – the limiting stochasticity (section 5.4) or its version – the reduced stochasticity (section 7.2). The kinetic equation which describes that random motion is reduced, thus, to a simple version of the diffusion equation (section 5.4). How to derive the kinetic equation in a general case irreducible to a mapping? Some aspects of this problem were considered and have been solved, e.g., in refs. [30, 9, 43, 88, 105], yet a general method for deriving of the kinetic equation based on the stochasticity of motion is still to be developed.

5) The above analysis of nonlinear resonances was essentially restricted to the case of a *moderate*, in particular, not too weak, nonlinearity. Since in many applications the nonlinearity may very well be weak the problem of stability for the weakly nonlinear oscillations arises.

6) A most difficult problem is related apparently to the so-called completely integrable nonlinear systems (see, e.g., a recent review article by Zakharov [68]). These systems possess a hidden symmetry which ensures the complete set of the integrals of motion. A classical example of such a system is the so-called Toda lattice [108] which consists of an arbitrary number of equal masses coupled by the nonlinear forces of a special type. The absolute stability of oscillations in this system had been discovered by Ford and co-workers [95], and subsequently Henon has found the analytical expression for all the integrals of motion [106]. Another interesting example of a whole family of the completely integrable nonlinear systems has been constructed by McMillan [109] in the form of special nonlinear mappings. Gardner et al. [110] have marked the beginning of the intensive studies (and/or "construction") of completely integrable systems. It should be recalled, however, that the first system with a hidden symmetry, which is well-known by now, was a classical system of the two bodies interacting via Newton's law of gravitation, the system which marked the beginning of the classical mechanics itself.*

The simple methods employed in this paper are not capable to discern a completely integrable system. We may comfort ourselves, of course, that such systems are exceptional, or non-generic, in a sense which is rigorously formalized, by the way, in the modern theory of dynamical systems [73] (see also the book [2], §5). Nevertheless, the problem of the "extra" integrals remains.

7) There is a more constructive problem concerning the behavior of the system close to a completely integrable one. Some preliminary studies [37, 111] indicate that even a very weak perturbation sharply destroys the integrability of such a system, so the domain of the Hilbert space with an improved stability around a completely integrable system seems to be fairly small. Is it always like this? What is the mechanism responsible for the rapid destruction of the stability of motion?

Acknowledgements

I wish to express my sincere gratitude to G.V. Gadiyak, F.M. Izraelev, E. Keil, A.M. Sessler, V.A. Tayursky and G.M. Zaslavsky whose fruitful collaboration at different stages of the research has made it possible to achieve many of the results reported in the present article. Almost all I have learned in modern mathematical theories of dynamical systems is due to D.V. Anosov, V.I. Arnold, V.K. Melnikov and Ya.G. Sinai whose permanent and tireless discussions and explanations on this subject as well as valuable advice are highly appreciated. All possible misconceptions in this review are, of course, my own. I am very grateful to many scientists including V.M. Alekseev, E. Courant, M.G.N. Hine, M.D. Kruskal, L.J. Laslett, N.N. Nekhoroshev, R.Z. Sagdeev, E.V. Shuryak, S.M. Ulam, N.J. Zabusky, V.E. Zakharov for illuminating discussions of separate questions as well as for helpful suggestions. The final version of the review was considerably improved thanks to important corrections and many useful comments suggested by the participants of the Berkeley study group, including J. Cary, R.H. Cohen, A.N. Kaufman, G.R. Smith, A. Weinstein, and also by G. Casati and D. ter Haar who had studied closely the preliminary version of the manuscript. Sincere thanks are also due to L.F. Khailo and G.B. Minchenkov for their great assistance in computation. It is a pleasure to acknowledge my debt to J. Ford for urging me to write a review article and for his ready response and permanent encouragement without which this paper would have never appeared.

*It is worthwhile to mention that a motivation for ref. [110] was to comprehend the surprising results of the numerical experiments by Fermi, Pasta and Ulam [130] who had observed a puzzling stability of some many-dimensional nonlinear oscillations. These authors had put forward a hypothesis similar to what has been actually proved in the KAM theory.

The initial impetus to this research has been given by my Teacher, the late director of the Institute of Nuclear Physics academician Gersh Itskovich Budker whose great driving spirit and most ingenious personality deeply influenced everyone who was as fortunate as to have worked close by...

Appendix: the Melnikov–Arnold (MA) integral

Consider the improper integral*:

$$A_m(\lambda) = \int_{-\infty}^{\infty} \mathrm{d}t \cos(\tfrac{1}{2}m\varphi(t) - \lambda t) \tag{A.1}$$

where the function

$$\varphi(t) = 4 \arctan(\mathrm{e}^t) - \pi \tag{A.2}$$

describes the motion along the pendulum separatrix (2.30) with $\omega_0 = 1$ and $\dot{\varphi} > 0$; m and λ are parameters. Since cosine's argument is antisymmetric ($\varphi(-t) = -\varphi(t)$) the MA integral may be written in the form:

$$A_m(\lambda) = \int_{-\infty}^{\infty} \mathrm{d}t \exp\{\mathrm{i}(\tfrac{1}{2}m\varphi - \lambda t)\}. \tag{A.3}$$

Let $x = 2\arctan(\mathrm{e}^t)$, whence:

$$\mathrm{e}^t = \tan\left(\frac{x}{2}\right) = \frac{\mathrm{e}^{\mathrm{i}x} - 1}{\mathrm{i}(\mathrm{e}^{\mathrm{i}x} + 1)}; \qquad \mathrm{e}^{\mathrm{i}\varphi/2} = -\mathrm{i}\mathrm{e}^{\mathrm{i}x} = \frac{1 + \mathrm{i}\mathrm{e}^t}{\mathrm{i} + \mathrm{e}^t}$$

and

$$A_m(\lambda) = \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{-\mathrm{i}\lambda t} \left(\frac{1 + \mathrm{i}\mathrm{e}^t}{\mathrm{i} + \mathrm{e}^t}\right)^m. \tag{A.4}$$

For integer m the MA integral can be expressed exactly via the integrand residues. If $m > 0$ the integrand poles are at the points:

$$\mathrm{e}^{t_p} = -\mathrm{i}; \qquad t_p = -\tfrac{1}{2}\mathrm{i}\pi - 2\pi \mathrm{i} n \tag{A.5}$$

where n is any integer including zero.

For $\lambda > 0$ we close the integration contour in the lower half-plane of the complex t. This operation needs to be clarified. We may close the contour, for example, along the two lines $\mathrm{Re}(t) = \pm T$. The integral over these lines does not equal zero, generally, but is proportional for $T \to \infty$ to the expression: $\sin(\tfrac{1}{2}m\pi - \lambda T)$, that is the MA integral oscillates as T grows. It is clear also from the asymptotic behavior of the indefinite integral:

$$\int \mathrm{d}t \exp\{\mathrm{i}(\tfrac{1}{2}m\varphi - \lambda t)\} \longrightarrow \frac{\mathrm{i}}{\lambda} \exp\{-\mathrm{i}\lambda t \pm \tfrac{1}{2}m\pi\} \text{ as } t \to \pm\infty.$$

*Similar integrals were evaluated also in ref. [69].

For the problem of the stability of motion these periodic oscillations are of minor importance (section 4.4), so we may just drop them from the expression for the MA integral and retain the aperiodical part only.

Substituting a new variable $z = t - t_p$ into eq. (A.5) and evaluating the residues we get:

$$A_m(\lambda) = -\frac{2\pi i}{i^m (m-1)!} e^{-\pi\lambda/2} \lim_{z\to 0} \frac{d^{m-1}}{dz^{m-1}} \left(e^{-i\lambda z} \left(\frac{1+e^z}{1-e^z} z \right)^m \right) \sum_{n=0}^{\infty} e^{-2\pi\lambda n}$$
$$= \frac{2\pi}{(m-1)!} \frac{e^{\pi\lambda/2}}{\sinh(\pi\lambda)} (2\lambda)^{m-1} (1 + f_m(\lambda)) \tag{A.6}$$

where we have introduced the new functions $f_m(\lambda)$ (see below). Note that the integration is being done in the negative direction around the contour.

For $\lambda < 0$ we close the contour in the upper half-plane and integrate around in the positive direction. The nearest pole to the real t-axis is now at the point $t_0 = 3\pi i/2$ but the relation $e^t = -ie^z$ remains the same as for $\lambda > 0$. Therefore, we may write immediately:

$$A_m(\lambda < 0) = (-1)^m A_m(|\lambda|) e^{-\pi|\lambda|}. \tag{A.7}$$

For large $|\lambda|$ the values of the MA integral is considerably smaller in this case.

To derive the functions $f_m(\lambda)$ it is convenient to apply the recurrence relation for $A_m(\lambda)$ found by O.V. Zhirov:

$$A_{m+1} = \frac{2\lambda}{m} A_m - A_{m-1}. \tag{A.8}$$

It can be deduced if one integrates eq. (A.4) by parts along the real t-axis and neglects the term $-(2/\lambda) \cdot \sin(\frac{1}{2}m\pi - \lambda T)$. But this is just the quantity we have ignored already on the grounds discussed above. From eqs. (A.8) and (A.6) we arrive at the recurrence relation:

$$f_{m+1} = f_m - (1 + f_{m-1})\, m(m-1)/4\lambda^2. \tag{A.9}$$

A few first f_m are:

$$f_1 = f_2 = 0; \qquad f_3 = -1/2\lambda^2; \qquad f_4 = -2/\lambda^2;$$
$$f_5 = -5/\lambda^2 + 3/2\lambda^4; \qquad f_6 = -10/\lambda^2 + 23/2\lambda^4.$$

For $|\lambda| \gg m$ the functions f_m describe small corrections to the leading term of the MA integral:

$$A_m(\lambda) \approx \frac{4\pi(2\lambda)^{m-1}}{(m-1)!} e^{-\pi\lambda/2}; \qquad \lambda \gg m. \tag{A.10}$$

Let us check the relations obtained using the exact expansion for pendulum oscillations near the separatrix (section 2.1):

$$\frac{\dot\varphi}{2\omega_0} = \cos\left(\frac{\varphi}{2}\right) = \frac{2\omega}{\omega_0} \sum_n \frac{\cos(\lambda_n \omega_0 t)}{\cosh(\pi\lambda/2)}; \qquad \lambda_n = \frac{n\omega}{\omega_0}.$$

On the other hand, we can express approximately a Fourier coefficient of this series by the MA integral:

$$F_n \approx \frac{2\omega}{\pi} \int_{-\infty}^{\infty} \mathrm{d}t \cos\left(\frac{\varphi_{sx}}{2}\right) \cos(\lambda_n \omega_0 t) = \frac{\omega}{\pi\omega_0}(A_1(\lambda) + A_1(-\lambda)) = \frac{2\omega}{\omega_0 \cosh(\pi\lambda/2)}.$$

For the real (non-integer) values of m the MA integral can be evaluated approximately if $|\lambda| \gg m$. We integrate around the cut, coming along the imaginary t-axis up (or down) to the singularity nearest to the real t-axis. For $\lambda > 0$ it is a pole at $t_0 = -\frac{1}{2}\pi\mathrm{i}$ (A.5). We introduce a new variable y according to the relation: $t = t_0 - \mathrm{i}y/\lambda$, whence:

$$\mathrm{e}^{-\mathrm{i}\lambda t} = \mathrm{e}^{-\pi\lambda/2}\,\mathrm{e}^{-y}; \qquad \frac{1+\mathrm{i}\mathrm{e}^t}{\mathrm{i}+\mathrm{e}^t} \approx -\frac{2\lambda}{y}; \qquad y \ll \lambda.$$

The MA integral (A.4) takes the form:

$$A_m(\lambda) \approx -\frac{\mathrm{i}}{\lambda}\mathrm{e}^{-\pi\lambda/2}(2\lambda)^m \int_C \mathrm{d}y\,(-y)^{-m}\,\mathrm{e}^{-y}.$$

The integration contours in the t- and y-planes are shown in fig. A.1. The last integral can be expressed in terms of the gamma function (see, e.g., ref. [58]), and we get:

$$A_m(\lambda) \approx \frac{4\pi(2\lambda)^{m-1}}{\Gamma(m)}\mathrm{e}^{-\pi\lambda/2}. \tag{A.11}$$

For integer m this relation coincides with the approximate formula (A.10) and has, thus, the accuracy $\sim (m/\lambda)^2$.

For $\lambda < 0$ and real $m > 0$ the nearest singularity to the real t-axis is a branch point (zero) of the integrand at $t_0 = \pi\mathrm{i}/2$ $(1 + \mathrm{i}\mathrm{e}^{t_0} = 0)$. Like the previous case we put: $t = t_0 + \mathrm{i}y/|\lambda|$ to find:

$$\mathrm{e}^{-\mathrm{i}\lambda t} = \mathrm{e}^{-\pi|\lambda|/2}\,\mathrm{e}^{-y}; \qquad \frac{1+\mathrm{i}\mathrm{e}^t}{\mathrm{i}+\mathrm{e}^t} \approx -\frac{y}{2|\lambda|}; \qquad y \ll |\lambda|.$$

Substituting these relations into eq. (A.4) and integrating around the cut (fig. A.1) we get:

$$A_m(\lambda < 0) \approx -\int_C \frac{\mathrm{i}\,\mathrm{d}y}{|\lambda|}\frac{\mathrm{e}^{-\pi|\lambda|/2}}{(2|\lambda|)^m}\mathrm{e}^{-y}\,(-y)^m = -\frac{4\,\mathrm{e}^{-\pi|\lambda|/2}}{(2|\lambda|)^{m+1}}\Gamma(m+1)\sin(\pi m).$$

This expression becomes inapplicable near an integer m where eq. (A.7) should be used. If $|\lambda| \gg m$ the values of the MA integral are always larger for positive than for negative λ.

We have assumed $m > 0$ and $\dot{\varphi} > 0$ until now. As is seen from the original representation of the MA

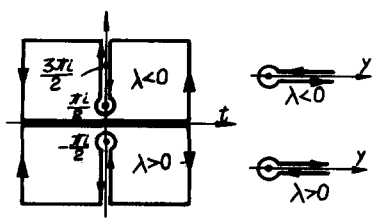

Fig. A.1. Integration contours for evaluation of the Melnikov–Arnold integral.

integral (A.1) only relative signs of $m\dot{\varphi}$ and λ are essential, the bigger value of the integral corresponding to the case: $m\dot{\varphi}\lambda > 0$.

Let us mention in conclusion that the main contribution to integral is gained over the interval of $t \sim 1/|\lambda|$ about $t = 0$.

References

[1] A.N. Kolmogorov, Dokl. Akad. Nauk SSSR 98 (1954) 527; General Theory of Dynamical Systems and Classical Mechanics, Proc. Intern. Congress of Mathematicians, Amsterdam, 1958, Vol. I, p. 315.

[2] V.I. Arnold, Usp. Mat. Nauk 18, No 6 (1963) 91;
V.I. Arnold and A. Avez, Ergodic Problems of Classical Mechanics (Benjamin, Inc., New York, 1968).

[3] J. Moser, Nachr. Akad. Wiss., Göttingen, Math. Phys. Kl., No 1 (1962) 1; Stable and Random Motions in Dynamical Systems, Annals of Mathematics Studies No 77 (University Press, Princeton, 1973).

[4] N.N. Bogoliubov and Yu.A. Mitropolsky, Asymptotic Methods in the Theory of Nonlinear Oscillations (Hindustan Publ. Corp., Dehli, 1961).

[5] V.I. Arnold, Dokl. Akad. Nauk SSSR 156 (1964) 9.

[6] H. Poincaré, Les méthodes nouvelles de la mécanique celeste, Vols. 1-3 (1892, 1893, 1899; Engl. Transl. NASA, Washington, 1967).

[7] N.S. Krylov, Studies on the Foundation of Statistical Physics (Akad. Nauk SSSR, M.-L., 1950) (in Russian).

[8] B.V. Chirikov, Atomnaya energiya 6 (1959) 630.

[9] G.M. Zaslavsky and B.V. Chirikov, Soviet Physics Uspekhi 14 (1972) 549.

[10] F.K. Goward, in: Lectures on the Theory and Design of an Alternating Gradient Photon Synchrotron (CERN, Geneva, 1953) p. 19;
M.G.N. Hine, ibid, p. 69.

[11] Ya.G. Sinai, Dokl. Akad. Nauk SSSR 153 (1963) 1261; Usp. Mat. Nauk 25, No 2 (1970) 141.

[12] V.K. Melnikov, Dokl. Akad. Nauk SSSR 144 (1962) 747; 148 (1963) 1257; Trudy Moskovskogo mat. obschestva 12 (1963) 3.

[13] A.N. Kolmogorov, Dokl. Akad. Nauk SSSR 119 (1958) 861; 124 (1959) 754.

[14] Ya.G. Sinai, Proc. Intern. Congress of Mathematicians, 1962; Izv. Akad. Nauk SSSR, mat. 30 (1966) 15.

[15] D.V. Anosov, Dokl. Akad. Nauk SSSR 145 (1962) 707; 151 (1963) 1250; Geodesic Flows on Closed Riemannien Manifolds of Negative Curvature, Trudy MIAN im. V.A. Steklova 90 (1967) (in Russian).

[16] Harold Jeffreys and Bertha Swirles, Methods of Mathematical Physics (University Press, Cambridge, 1950) p. 488;
P.L. Kapitsa, Zh. Eksp. Teor. Fiz. 21 (1950) 588;
N.N. Bogoliubov, Trudy Inst. stroit. mekh. Akad. Nauk UkrSSR, 1950, No 14, p. 9.

[17] G.D. Birkhoff, Dynamical Systems (New York, 1927).

[18] N.N. Bogoliubov, On Quasi-Periodic Solutions in Nonlinear Problems of Mechanics, First Math. Summer School, Kanev, 1963 (Naukova Dumka, Kiev, 1964) Vol. 1, p. 11 (in Russian).

[19] Ya.B. Zeldovich, Zh. Eksp. Teor. Fiz. 51 (1966) 1492;
V.I. Ritus, ibid, p. 1544.

[20] E.V. Shuryak, Zh. Eksp. Teor. Fiz. 71 (1976) 2039.

[21] A. Schoch, CERN Report 57-21 (1958);
A.A. Kolomensky and A.N. Lebedev, Theory of Cyclic Accelerators (Wiley, New York, 1966).

[22] H. Goldstein, Classical Mechanics (Addison-Wesley Press, Inc., Cambridge, 1950).

[23] V.I. Arnold, Mathematical Methods in Classical Mechanics (Nauka, Moskva, 1974) (in Russian).

[24] N.N. Nekhoroshev, Functs. Analis i ego prilozheniya 5 (1971) 82; Method of Successive Canonical Transformations of Variables, Appendix to Russian Transl.: J. Moser, Lectures on Hamiltonian Systems (Mir, Moskva, 1973).

[25] M.M. Khanaev, Dokl. Akad. Nauk SSSR 193 (1970) 46;
A.P. Markeev, Priklad. mat. i mekh. 37 (1973) 753.

[26] G.V. Gadiyak and F.M. Izraelev, Dokl. Akad. Nauk SSSR 218 (1974) 1302.

[27] C.L. Siegel, Math. Ann. 128 (1954) 144.

[28] P.R. Halmos, Lectures on Ergodic Theory (Chelsea, New York, 1958).

[29] G.M. Zaslavsky and B.V. Chirikov, Dokl. Akad. Nauk SSSR 159 (1964) 306.

[30] M.N. Rosenbluth, R.Z. Sagdeev, J.B. Taylor and G.M. Zaslavsky, Nuclear Fusion 6 (1966) 297.

[31] B.V. Chirikov, Dokl. Akad. Nauk SSSR 174 (1967) 1313.

[32] A.J. Lichtenberg and E.F. Jaeger, Physics of Fluids 13 (1970) 392;
E.F. Jaeger and A.J. Lichtenberg, Ann. Phys. 71 (1972) 319.

[33] G.S. Smith and A.N. Kaufman, Phys. Rev. Lett. 34 (1975) 1613; Diffusion due to a single wave in a magnetized plasma, Proc. Nobel Symposium on Nonlinear Effects in Plasmas, Sweden, 1976.

[34] G.H. Walker and J. Ford, Phys. Rev. 188 (1969) 416.
[35] B.V. Chirikov, E. Keil and A.M. Sessler, J. Statist. Phys. 3 (1971) 307.
[36] F.M. Izraelev and B.V. Chirikov, Dokl. Akad. Nauk SSSR 166 (1966) 57.
[37] B.V. Chirikov, F.M. Izraelev and V.A. Tayursky, Computer Phys. Comm. 5 (1973) 11.
[38] M. Hénon and C. Heiles, Astron. J. 69 (1964) 73.
[39] B.V. Chirikov and F.M. Izraelev, Some Numerical Experiments with a Nonlinear Mapping: Stochastic Component, Colloques Internationaux du C.N.R.S. Transformations Ponctuelles et leurs Applications (Toulouse, 1973) p. 409.
[40] J. Ford and G.H. Lunsford, Phys. Rev. A1 (1970) 59.
[41] G.V. Gadiyak, F.M. Izraelev and B.V. Chirikov, Numerical Experiments on a Universal Instability in Nonlinear Oscillator Systems (the Arnold Diffusion), Proc. 7th Intern. Conf. on Nonlinear Oscillations (Berlin, 1975), Vol. II, 1, p. 315; The Arnold diffusion in a System of Three Resonances (to be published).
[42] A.Ya. Khinchin, Continued Fractions (Fizmatgiz, Moskva, 1961) (in Russian).
[43] B.V. Chirikov, Research Concerning the Theory of Nonlinear Resonance and Stochasticity, Preprint 267 (Institute of Nuclear Physics, Novosibirsk, 1969) (Engl. Trans., CERN Trans. 71–40, 1971).
[44] Colloques Internationaux du C.N.R.S. Transformations Ponctuelles et leurs Applications (Toulouse, 1973).
[45] V.N. Melekhin, Zh. Exper. Teor. Fiz. 61 (1971) 1319; 68 (1975) 1601.
[46] M.A. Lieberman and A.J. Lichtenberg, Phys. Rev. A5 (1972) 1852.
[47] L.J. Laslett, Stochasticity, 9th Intern. Conf. on High Energy Accel., SLAC, Stanford, 1974.
[48] B. Lehnert, Dynamics of Charged Particles (North-Holland Publishing Company, Amsterdam, 1964);
T.G. Northrop, The Adiabatic Motion of Charged Particles (Interscience Publishers, Inc., New York, 1963).
[49] J. Moser, Lectures on Hamiltonian Systems, Memoirs A.M.S., No 81 (1968);
C.L. Siegel and J. Moser, Lectures on Celestial Mechanics, Grundlehren Bd. 187, 1971.
[50] R.J. Hastie, G.D. Hobbs and J.B. Taylor, Non-Adiabatic Behaviour of Particles in Inhomogeneous Magnetic Fields, Proc. 3d Intern. Conf. on Plasma Physics and Controlled Nuclear Fusion Research, Novosibirsk, 1968 (Int. Atom. Energy Agency, Vienna, 1969) Vol. 1, p. 389.
[51] A.M. Dykhne and A.V. Chaplik, Zh. Exper. Teor. Fiz. 40 (1961) 666.
[52] A. Garren et al., Individual Particle Motion and the Effect of Scattering in an Axially Symmetric Magnetic Field, Proc. 2nd Intern. Conf. on Peaceful Uses of Atomic Energy 31 (1958) 65.
[53] E.M. Krushkal, Zh. Tekhn. Fiz. 42 (1972) 2288.
[54] Yu.I. Neimark, The Method of Point Mappings in the Theory of Nonlinear Oscillations (Nauka, Moskva, 1972) (in Russian).
[55] O.V. Serdyuk, Zh. Tekhn. Fiz. 46 (1976) 2136.
[56] L.A. Artsimovich, Controlled Thermonuclear Reactions (Oliver and Boyd, Edinburgh, 1964) §7.3.
[57] C. Störmer, The Polar Aurora (Oxford University Press, London and New York, 1955).
[58] E.T. Whittaker and G.N. Watson, A Course of Modern Analysis (Cambridge Univ. Press, 1946);
I.S. Gradstein and I.M. Ryzhik, Handbook of Integrals (Fizmatgiz, Moskva, 1962) (in Russian).
[59] Ya.S. Derbenev, S.I. Mishnev and A.N. Skrinsky, Atomnaya Energiya 20 (1966) 217.
[60] J. Moser, Annali della Scuola Normale Superiore di Pisa, Serie III, Vol. XX, Fasc. III (1966) 499.
[61] E.A. Grebennikov and Yu.A. Ryabov, New Qualitative Methods in Celestial Mechanics (Nauka, Moskva, 1971) (in Russian).
[62] G.M. Zaslavsky and N.N. Filonenko, Zh. Exper. Teor. Fiz. 54 (1968) 1590.
[63] N.N. Bogoliubov, Yu.A. Mitropolsky and A.M. Samoilenko, The Method of Rapid Convergence in Nonlinear Mechanics (Naukova dumka, Kiev, 1969) (in Russian).
[64] H. Rüssmann, Nachr. Akad. Wiss. in Göttingen, II. Math. Phys. Kl. (1970) No 5, 67.
[65] F. Takens, Indag. Math. 33 (1971) 379.
[66] J. Moser, On the Construction of Almost Periodic Solutions for Ordinary Differential Equations, Proc. Intern. Conf. on Functional Analysis and Related Topics, Tokyo (1969) p. 60.
[67] V.I. Arnold, Stability Problem and Ergodic Properties of Classical Dynamical Systems, Proc. Intern. Congress of Mathematicians, Moscow, 1966.
[68] V.E. Zakharov, The Method of Reverse Scattering Problem, Lecture Notes in Mathematics (Springer) in print.
[69] A.B. Rechester and T.H. Stix, Phys. Rev. Lett. 36 (1976) 587.
[70] C. Froeschlé, Astron. & Astrophys. 9 (1970) 15.
[71] J.M. Greene, J. Math. Phys. 9 (1968) 760.
[72] A. Renieri, Non-Linear Beam-Beam Effect Computer Simulation, 9th Intern. Conf. on High Energy Accelerators, SLAC, Stanford, 1974.
[73] S. Smale, Bull. Amer. Math. Soc. 73 (1967) 747.
[74] Z. Nitecki, Differentiable Dynamics (The MIT Press, Cambridge, Massachusetts and London, England, 1971).
[75] Ya.G. Sinai, Izv. Akad. Nauk SSSR, mat. 30 (1966) 1275.
[76] J. Ford, The Transition from Analytic Dynamics to Statistical Mechanics, Advances in Chemical Physics, XXIV (John Wiley & Sons, Inc., 1973); Stochastic Behavior in Non-linear Oscillator Systems, in: Lectures in Statistical Physics, Lecture Notes in Physics, Vol. 28 (Springer, 1974); The Statistical Mechanics of Classical Analytic Dynamics, in: Fundamental Problems in Statistical Mechanics, Vol. 3, ed. E.D.G. Cohen (North-Holland Publishing Company, Amsterdam, 1975).

[77] V.M. Alekseev, Mat. Sbornik 76 (1968) 72; 77 (1968) 545; 78 (1969) 3.
[78] R.C. Churchill and D.L. Rod, J. Diff. Equations 21 (1976) 39; 66.
[79] B.V. Chirikov, Wiss. Zs. Humboldt-Universität zu Berlin, Ges.-Sprachw. R., XXIV (1975) 215.
[80] G.A. Hedlund, Bull. Amer. Math. Soc. 45, No. 4 (1939) 241;
E. Hopf, Math. Ann. 117 (1940) 590.
[81] H. Poincaré, J. de Phys., Ser. 4, 5 (1906) 369; Science et méthode (Flammarion, Paris, 1908).
[82] I. Gumowski and C. Mira, Proc. 8th Intern. Conf. on High Energy Accelerators, CERN, Geneva, 1971, p. 374; Zagadnienia Drgan Nilinowych (Nonlinear Vibration Problems) 14 (1973) 417.
[83] R.V. Mises, Vorlesungen aus dem Gebiete der angewandten Mathematik, Bd. 1 – Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik (Lpz.-W., 1931).
[84] M.G.N. Hine, private communication, 1967.
[85] L.D. Pustylnikov, Dokl. Akad. Nauk. SSSR 202 (1972) 287.
[86] S.P. Kapitsa and V.N. Melekhin, The Microtron (Nauka, Moskva, 1969) (in Russian).
[87] B. McNamara, D.V. Anderson, J.K. Boyd, J.A. Byers, R.H. Cohen, T.A. Cutler, L.S. Hall, R.F. Post and M.E. Rensink, Theory of Mirror Machines at High Beta, Proc. 6th Intern. Conf. on Plasma Physics and Controlled Nuclear Fusion Research, Berchtesgaden, FRG, 1976 (Int. Atom. Energy Agency, Vienna, 1977) Vol. III, p. 161.
[88] G.M. Zaslavsky, Statistical Irreversibility in Nonlinear Systems (Nauka, Moskva, 1970) (in Russian).
[89] M.V. Smoluchowski, Naturwissenschaften 6 (1921) 253.
[90] F.M. Izraelev and B.V. Chirikov, Some Numerical Experiments with a Simplest Model of Turbulence, Proc. of the Meeting on Programming and Mathematical Models for Solving the Physical Problems (JINR, Dubna, 1974) (in Russian).
[91] F. Rannou, Astron. & Astrophys. 31 (1974) 289.
[92] M.N. Rosenbluth, Phys. Rev. Lett. 29 (1972) 408.
[93] J. Finn, Nuclear Fusion 15 (1975) 845.
[94] G.N. Kulipanov, S.I. Mishnev and A.N. Skrinsky, Study of the Stochastic Instability for a Beam under Periodic Crossing a Resonance of Betatron Oscillations, Proc. 7th Intern. Conf. on High Energy Accelerators, Erevan, 1970, Vol. 2, p. 300; Beam Behavior in a Storage Ring under Simultaneous Influence of Two Betatron Oscillation Resonances, ibid., p. 353.
[95] J. Ford, S.D. Stoddard and J.S. Turner, Prog. Theor. Phys. 50 (1973) 1547.
[96] S. Smale, Diffeomorphisms With Many Periodic Points, Differential and Combinatorial Topology, Princeton Mathematical Series, No 27 (University Press, Princeton, 1965).
[97] L.P. Shilnikov, Dokl. Akad. Nauk SSSR 172 (1967) 298; 180 (1968) 286; Mat. sbornik 74 (1967) 378.
[98] Yu.I. Neimark, Dokl. Akad. Nauk SSSR 172 (1967) 1021.
[99] N.N. Filonenko, R.Z. Sagdeev and G.M. Zaslavsky, Nuclear Fusion 7 (1967) 253.
[100] P. Channell, An Illustrative Example of the Chirikov Criterion for Stochastic Instability, Lawrence Berkeley Laboratory Internal Report PEP-218 (1976).
[101] V.G. Ponomarenko, L.Ya. Trainin, V.I. Yurchenko and A.N. Yasnetsky, Zh. Exper. Teor. Fiz. 55 (1968) 3;
A.N. Dubinina and L.S. Krasitskaya, Zh. Exper. Teor. Fiz. (Pisma) 5 (1967) 230.
[102] V.L. Auslender et al., Atomnaya Energiya 22 (1967) 179.
[103] E. Keil, Incoherent Space Charge Phenomena in the ISR, Proc. 8th Intern. Conf. on High Energy Accel., CERN, Geneva, 1971, p. 372.
[104] E. Keil and G. Leroy, Effects of a Non-linear Lens on the Stored Proton Beam in the ISR, Nuclear Science NS-22 (1975) 1370.
[105] Ya.S. Derbenev and S.A. Kheifets, Derivation of Kinetic Equation for a Model System with Discrete Spectrum without the Correlation Damping Hypothesis, Teor. Mat. Fiz. 31 (1977) 220.
[106] M. Hénon, Phys. Rev. B9 (1974) 1921.
[107] A.J. Dragh and J.M. Finn, J. Geophys. Res. 81 (1976) 2327.
[108] M. Toda, Studies of a Non-linear Lattice, Physics Reports 18C (1975) 1.
[109] E.M. McMillian, A Problem in the Stability of Periodic Systems, in: Topics in Modern Physics (Colorado Associated University Press, 1971).
[110] C.S. Gardner, J.M. Greene, M.D. Kruskal and R.M. Miura, Phys. Rev. Lett. 19 (1967) 1095.
[111] G. Casati and J. Ford, Phys. Rev. A12 (1975) 1702.
[112] E. Courant, Nuclear Science NS-12 (1965) 550.
[113] L.C. Teng, Nuclear Science NS-20 (1973) 843.
[114] V.K. Melnikov, Dokl. Akad. Nauk SSSR 165 (1965) 1245.
[115] J. Moser, SIAM Review 8 (1966) 145; Math. Annalen 169 (1967) 136.
[116] J. Ford, Empirical Determination of Integrability for Non-Linear Oscillator Systems Using Area-Preserving Mappings, Colloques Internationaux du C.N.R.S. Transformations Ponctuelles et leur Applications, Toulouse, 1973, p. 123.
[117] H. Rüssmann, On Optimal Estimates for the Solutions of Linear Partial Differential Equations of First Order with Constant Coefficients on the Torus, in: Dynamical Systems. Theory and Applications, Lecture Notes in Physics, Vol. 38 (Springer, 1975).
[118] R.J. Hastie, J.B. Taylor and F.A. Haas, Ann. Phys. 41 (1967) 302.
[119] M.D. Kruskal, J. Math. Phys. 3 (1962) 806.

[120] M. Toda, Phys. Lett. A48 (1974) 335.
[121] G. Casati, Lett. Nuovo Cimento 14 (1975) 311.
[122] P. Brumer and J.W. Duff, J. Chem. Phys. 65 (1976) 3566.
[123] F. Galton, Natural Inheritance (London, 1889).
[124] D.I. Golenko, Simulation and Statistical Analysis of Pseudorandom Numbers by Computer (Nauka, Moskva, 1965);
M.V. Antipov, F.M. Izraelev and B.V. Chirikov, in: Computing Systems, No 30 (Nauka, Novosibirsk, 1968) 77;
M.V. Antipov, Analysis of Generating Factors for the Multiplicative Pseudorandom Number Generator, Preprint No 42–71 (Institute of Nuclear Physics, Novosibirsk, 1971) (in Russian).
[125] O.I. Bogoyavlensky and S.P. Novikov, Usp. Mat. Nauk 31, No 5 (1976) 33.
[126] V.D. Ilyin and A.N. Ilyina, Zh. Exper. Teor. Fiz. 72 (1977) 983;
V.D. Ilyin, Geomagnetism i aeronomiya 8 (1968) 564;
V.D. Ilyin, A.N. Ilyina and S.N. Kuznetsov, Kosmicheskie issledovaniva 11 (1973) 559;
V.D. Ilyin and S.N. Kuznetsov, Non-adiabatic Effects of Particle Motion in a Static Dipole Field and in Time Varying Fields, Proc. 7th Leningrad Intern. Seminar, Leningrad, 1975, p. 269 (in Russian).
[127] A.D. Bryuno, Mat. sbornik 83 (1970) 273.
[128] H. Scholl and C. Froeschlé, Astron. and Astrophys. 42 (1975) 457.
[129] V.I. Volosov and A.V. Komin, Zh. Tekhn. Fiz. 38 (1968) 846;
V.N. Bocharov, V.I. Volosov, A.V. Komin, V.M. Panasyuk and Yu.N. Yudin, Plasma Containment in a Stellarator for Different Free Path Lengths, Proc. 3d Intern. Conf. on Plasma Physics and Controlled Nuclear Fusion Research, Novosibirsk, 1968 (Int. Atom. Energy Agency, Vienna, 1969) Vol. 1, p. 561.
[130] E. Fermi, J. Pasta and S. Ulam, Studies of Nonlinear Problems I, Los Alamos Scientific Report LA-1940, 1955; in: Collected Works of Enrico Fermi, Vol. II (University of Chicago Press, 1965) p. 978.
[131] D.D. Ryutov and G.V. Stupakov, Zh. Exper. Teor. Fiz. (Pisma) 26 (1977) 186.
[132] E. Fermi, Phys. Rev. 75 (1949) 1169.
[133] E.L. Burstein, V.I. Veksler and A.A. Kolomensky, The Stochastic Method for Particle Acceleration, in: Some Problems in the Theory of Cyclic Accelerators (Akad. Nauk SSSR, Moskva, 1955) (in Russian).
[134] E.F. Jaeger, A.J. Lichtenberg and M.A. Lieberman, Plasma Physics 14 (1972) 1073;
M.A. Lieberman, A.J. Lichtenberg, Plasma Physics 15 (1973) 125.
[135] G.R. Smith, Numerical Study of Particle Motion in Two Waves, Preprint UCRL-80121 (Lawrence Livermore Laboratory, Livermore, 1977); Proc. Int. Conf. on Stochastic Behavior in Classical and Quantum Hamiltonian Systems (Como, Italy, 1977) to appear.
[136] A.V. Timofeev, Nuclear Fusion 14 (1974) 165; Fizika Plasmy 1 (1975) 88.
[137] R.H. Cohen, G. Rowlands and J.H. Foote, Nonadiabaticity in Mirror Machines, Preprint UCRL-78889 (1977).
[138] B.V. Chirikov, The Problem of Motion Stability for a Charged Particle in a Magnetic Trap, Fizika Plasmy 4:3 (1978) 521.
[139] J. Ford, private communication.
[140] J. Cary, private communication.
[141] N.N. Nekhoroshev, Usp. Mat. Nauk 32:6 (1977) 5.
[142] V.D. Ilyin and A.N. Ilyina, Zh. Eksp. Teor. Fiz. 75 (1978) 518.
[143] M.J. Hadamard, Sur le billard non euclidien, Proc. de la Soc. Sciences physiques et naturelles de Bordeaux (1898).
[144] G. Casati, B.V. Chirikov, J. Ford and F.M. Izraelev, Stochastic Behavior of a Quantum Pendulum under a Periodic Perturbation, Proc. Intern. Conf. on Stochastic Behavior in Classical and Quantum Hamiltonian Systems, Como, 1977 (to appear).
[145] D.S. Ornstein, Ergodic Theory, Randomness, and Dynamical Systems (New Haven and London, Yale University Press, 1974).
[146] J.M. Greene, A Method for Determining a Stochastic Transition (to be published).