Russian Mathematical Surveys

AN EXPONENTIAL ESTIMATE OF THE TIME OF STABILITY OF NEARLY-INTEGRABLE HAMILTONIAN SYSTEMS

To cite this article: N N Nekhoroshev 1977 *Russ. Math. Surv.* **32** 1

View the article online for updates and enhancements.

You may also like

- PARTIAL DIFFERENTIAL INEQUALITIES
J L Lions
- Basic wavelet theory
I Ya Novikov and S B Stechkin
- Spectral theory of random matrices
V L Girko

This content was downloaded from IP address 18.117.183.150 on 23/04/2024 at 09:13

Russian Math. Surveys 32:6 (1977), 1–65
From Uspekhi Mat. Nauk 32:6 (1977), 5–66

AN EXPONENTIAL ESTIMATE OF THE TIME OF STABILITY OF NEARLY-INTEGRABLE HAMILTONIAN SYSTEMS

N. N. Nekhoroshev

Contents

The first three sections are written informally. The first part of the Introduction (§1) is a survey of investigations of nearly-integrable systems that are based on the methods of perturbation theory. In the second half of §1 we discuss in detail the results of this article. In §3 we explain the proof of our main theorem, in which we establish an exponential estimate of the time of stability; §§4–9 are devoted to the statement and proof of this theorem. The proofs of certain lemmas used to prove the main theorem, will be given elsewhere.[1] These lemmas are stated in §10. Their proofs, together with the contents of §§5–9, constitute the complete

[1] It is hoped that this article will be published in the Proceedings of the Petrovskii Seminar, No.5, under the title, "An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. II".

proof of the main theorem.

§11 is written informally. In it we use the methods and ideas of the proof of the main theorem to make certain deductions about the behaviour of nearly-integrable Hamiltonian systems. In the last section, §12, we derive as a consequence of the main theorem, an exponential estimate of the time of stability of a planetary system.

§1. Introduction

1.1 Nearly-integrable Hamiltonian systems. Perpetual stability and stability during finite intervals of time. In this article we investigate the behaviour of the variables I in the Hamiltonian system of canonical equations

$$\dot{I} = -\frac{\partial H}{\partial \varphi}, \quad \dot{\varphi} = \frac{\partial H}{\partial I}$$

with the Hamiltonian

$$H = H_0(I) + \varepsilon H_1(I, \varphi), \tag{1.1}$$

where $\varepsilon \ll 1$ is a small parameter, the perturbation $\varepsilon H_1(I, \varphi)$ is 2π-periodic in $\varphi = \varphi_1, \dots, \varphi_s$, and I is an s-dimensional vector, $I = I_1, \dots, I_s$.

The φ_1 are called angular variables, and the I_i action variables. A system with a Hamiltonian depending only on the action variables is said to be integrable, and the system (1.1) is said to be nearly-integrable. The system (1.1) is also called a perturbation of the system with the Hamiltonian H_0, which is called the unperturbed system.

It follows from Hamilton's equations that the variables I_i are first integrals of the unperturbed system. It is also clear that the rate of change of these variables along the solutions of the perturbed system is of order ε.

But it turns out that during large intervals of time (considerably greater than $1/\varepsilon$) the values of the action variables differ only slightly from the initial values.

In this context a number of questions arise; we list the most important ones. Does the point $I(t)$ for all or for part of the solutions $I(t)$, $\varphi(t)$, perpetually lie close to its initial position, that is, for all these solutions, is $|I(t) - I(0)| < \alpha(\varepsilon)$ for all real t, where $\alpha(\varepsilon) \to 0$ as $\varepsilon \to 0$? We call such solutions perpetually stable. If perpetually stable solutions exist, how many of them are there? If there exist solutions that are not perpetually stable, what are the estimates of their "times of stability". In the main we shall be interested in lower estimates of the times of stability, that is, in functions $T(\varepsilon)$ such that $I(t)$ is close to $I(0)$ during the time interval $[0, T(\varepsilon)]$ for all sufficiently small ε:

$$|I(t) - I(0)| \ll 1 \quad \text{for} \quad t \in [0, T(\varepsilon)].$$

Throughout this article, we mean by stability (during finite or infinite intervals of time) stability in this sense, and not the usual Lyapunov stability.

Some of the above questions go back even to Laplace and Lagrange, who attempted to understand rather than explain the stability of the solar system. Other questions were posed and investigated by Poincaré [2], Birkhoff [3], and Siegel [4] and [5]. More recently, major contributions to the solution of these problems have been made by Kolmogorov [6], [7], Arnol'd [8]–[11], and Moser [12], [13].

1.2. Kolmogorov tori. Perpetual stability of two-frequency systems. Arnol'd diffusion. Arnol'd [8], [10] has proved, in particular, the following assertion. Let H_0 be a function of "general form". Then, in the phase space of a system with the Hamiltonian (1.1) there exists a set consisting of s-dimensional invariant tori close to the tori given by the equations $I_1 = \text{const}, \dots, I_s = \text{const}$. This set, which is called a Kolmogorov set, is closed and nowhere dense, but it forms a "large part" of the phase space, more precisely, the measure of the complement of a Kolmogorov set tends to 0, as $\varepsilon \to 0$.

As $\varepsilon \to 0$, the Kolmogorov tori of the system (1.1) are transformed into the invariant tori $\{I, \varphi | I = \text{const}\}$ of the unperturbed system. Hence, those solutions whose trajectories lie on a Kolmogorov set (this is the "majority" of them) are perpetually stable.

A sufficient condition for the existence of a Kolmogorov set is that at least one of the determinants Δ_1, Δ_2, where

$$\Delta_1 = \det\left(\frac{\partial^2 H_0}{\partial I^2}\right), \tag{1.2}$$

$$\Delta_2 = \det\begin{pmatrix} \dfrac{\partial^2 H_0}{\partial I^2} & \dfrac{\partial H_0}{\partial I} \\ \dfrac{\partial H_0}{\partial I} & 0 \end{pmatrix}, \tag{1.3}$$

does not vanish identically. Here Δ_1 is the Hessian of H_0, that is, the determinant of the matrix of order s whose elements are the second derivatives $\dfrac{\partial^2 H_0}{\partial I_i \partial I_j}$; the matrix in (1.3) is symmetric and of order $(s+1)$, and $\dfrac{\partial H_0}{\partial I}$ is the vector whose components are the first derivatives of H_0.

As a consequence of these results, for systems of two degrees of freedom ($s = 2$) the following stronger assertion has been proved: If Δ_2 does not vanish, then all the solutions of the system are perpetually stable. A detailed explanation of the meaning of the conditions $\Delta_1 \not\equiv 0$ and $\Delta_2 \not\equiv 0$, and of the nature of the perpetual stability, are given in the book [27], 365–383.

The results mentioned above were first obtained under the assumption that H is analytic. Moser proved that this condition can be replaced by the existence of a sufficiently many (namely, 333) derivatives ([12], [13]). By the efforts of Moser and Rüssman [31] and [32], the number of derivatives has been reduced to 6.

Zehnder has proved that the existence of a Kolmogorov set follows from a general theorem of the type of an implicit function theorem for non-smooth functionals (see [33] and [34]).

In all these investigations, the behaviour of solutions in the complement of the set of Kolmogorov tori for $s \geqslant 3$ has remained an open question. Arnol'd has constructed examples of systems such that part of their solutions $I(t)$ is arbitrarily far from $I(0)$ (see [11]); this effect is known as "Arnol'd diffusion" ([14]).

1.3. An exponential estimate of the time of stability. The interval of time during which the point $I(t)$ for the unstable solutions constructed in [11] is a short distance from $I(0)$ grows exponentially, as ε decreases linearly. In this connection, Arnol'd [15] has conjectured that for systems with a Hamiltonian of general form, the "time of constraint" of $I(t)$ close to $I(0)$ grows faster than any power of $1/\varepsilon$ for all initial conditions. This conjecture has been verified. In this article we establish an exponential estimate for the time of stability.

The precise statement of the exponential estimate forms the content of Theorem 4.4, which is the main result of the article. We mention here a theorem which is not completely accurate and differs from Theorem 4.4 in certain details.

1.4 THEOREM. *Suppose that H_0 satisfies certain "steepness conditions", which will be given in* §1.7. *Then there are positive constants a, b, and ε_0 with the following property.*

Let $0 < \varepsilon < \varepsilon_0$. Then for every solution $I(t)$, $\varphi(t)$ of the system with the Hamiltonian $H_0(I) + \varepsilon H_1(I, \varphi)$,

$$|I(t)-I(0)| < \varepsilon^b \tag{1.4}$$

for all $t \in [0, T]$, where

$$T = \frac{1}{\varepsilon}\exp\left(\frac{1}{\varepsilon^a}\right). \tag{1.5}$$

The constants a and b depend only on H_0; ε_0 depends only on H_0 and on the two parameters[1] in H_1 that estimate, respectively, the largest quantity and the rate of decrease of the coefficients in the Fourier series for H_1 in terms of φ. The values of a and b are given in §1.10 and are discussed in §§1.10, 1.11, and 2.1B.

The estimate (1.5) is proved under the assumption that H is analytic. If we only require H to be smooth, then the estimate is not exponential, but a power estimate whose exponent is the larger the more derivatives H has.

[1] In Theorem 4.4 the role of the second parameter is played by the width ρ of the complex neighbourhood of the real plane of the variables φ in which H_1 is analytic. As is well known, this width determines the rate of decrease of the Fourier coefficients.

The exponential estimate is also valid for systems with a Hamiltonian of a slightly more general form than (1.1). We describe these systems in §§1.5 and 1.6. Generalizations of the estimate to systems essentially different from (1.1) are discussed in §2.2.

1.5. The exponential estimate for systems with parameters. In §§4–9 we prove the exponential estimate for systems with the Hamiltonian

$$H = H_0(I) + \varepsilon H_1(I,\ p,\ \varphi,\ q), \tag{1.6}$$

where p and q are $(n-s)$-dimensional vectors ($n \geqslant s$), the variables I and p are the canonical conjugates of φ and q, respectively, and H_1 is 2π-periodic in $\varphi = \varphi_1, \ldots, \varphi_s$. When $n = s$, the Hamiltonian (1.6) is the same as (1.1).

This generalization allows us to apply our main theorem to a planetary system like our solar system (for details see §1.18). We remark that the generalization of the estimates to the system (1.6) hardly complicates the calculations of their proof, and does not make these estimates worse.

We also remark that the assertion about the validity of the estimates (1.4) and (1.5) for the system (1.6) trivially implies the equivalent assertion: These estimates hold for non-autonomous systems for which the perturbation depends on the "slow time" εt, that is, for systems with the Hamiltonian

$$H = H_0(I) + \varepsilon H_1(I,\ \varphi,\ \varepsilon t).$$

We may call the variables p and q in (1.6) parameters. The dependence of a perturbation on the parameters and the slow time does not worsen the estimates (1.4) and (1.5) in the main theorem.

The Kolmogorov–Arnol'd–Moser theory (see §1.2) does not carry over to the systems considered in this subsection, apart from those cases when the Hamiltonian has the form (1.1) or can be reduced to this form.

1.6. Systems with a perturbation depending periodically on time. We consider a system with the Hamiltonian

$$H = H_0(I) + \varepsilon H_1(I,\ \varphi,\ t), \quad \text{where } H_1(I,\ \varphi,\ t+2\pi) = H_1(I,\ \varphi,\ t). \tag{1.7}$$

Conditions on H_0 other than the steepness condition ensure an exponential estimate of the time of stability of the solutions of this system. These conditions (called conditions of P-steepness) are given in §1.9B. They are equivalent to the steepness conditions on a function $\mathscr{H}$ of $(s+1)$ variables $I_1, \ldots, I_s$, T that is defined in terms of H_0 by $\mathscr{H}(I,\ T) = H_0(I) + T$.

1.7. Steep functions. Now we turn to the formulation of conditions on the unperturbed Hamiltonian H_0 that ensure an exponential estimate of the time of stability of the system (1.6). They generalize the condition $\left|\dfrac{dH_0(I)}{dI}\right| \geqslant g > 0$, which in an obvious way guarantees perpetual stability with respect to I of the systems (1.1) with a single frequency, $s = 1$.

Let H_0 be an arbitrary function defined in some domain G ($\sim$ to an open set) of the Euclidean space $\mathbf{E}^s$. By affine subspaces of $\mathbf{E}^s$ we mean planes of any dimension. Let I be any point of G, and let λ be a plane containing I, $\dim \lambda \neq 0$. We denote by grad $(H_0|_\lambda)$ the gradient of the restriction of H_0 to λ, and by $m_{I,\lambda}(\eta)$ the smallest value of the length of this gradient on the sphere with centre at I and of radius η:

$$m_{I,\lambda}(\eta) = \min_{\{I' \in \lambda : |I'-I| = \eta\}} |\operatorname{grad}(H_0|_\lambda)|_{I'}|.$$

1.7.A. DEFINITION. We say that a function H_0 is *steep at I on the plane* λ if we can find constants $C > 0$, $\delta > 0$, and $\alpha \geqslant 0$ such that

$$\max_{0 \leqslant \eta \leqslant \xi} m_{I,\lambda}(\eta) > C\xi^\alpha \tag{1.8}$$

for all $\xi \in (0, \delta]$. We call C and δ the *coefficients* and α the *index of steepness.*

In the following two definitions, we only impose, in essence, conditions for the uniformity of the estimate (1.8).

We denote by $\Lambda^r(I)$ the set of all r-dimensional planes in $\mathbf{E}^s$ that contain $I \in \mathbf{E}^s$.

1.7.B. DEFINITION. We say that a function H_0 of s variables, $s \geqslant 1$, is *steep at the point I* if the following conditions hold: Firstly, $|\operatorname{grad} H_0|_I| \geqslant g$, where $g > 0$. Secondly, when $s \geqslant 2$, for every $r = 1, \ldots, s-1$ there are numbers $C_r > 0$, $\delta_r > 0$, and $\alpha_r \geqslant 1$ such that H_0 is steep at I on every plane $\lambda \in \Lambda^r(I)$ perpendicular to grad $H_0|_I$ with coefficients C_r and δ_r, and index α_r. We call the numbers $g, C_1, \ldots, C_{s-1}, \delta_1, \ldots, \delta_{s-1}$ the *steepness coefficients*, and $\alpha_1, \ldots, \alpha_{s-1}$ the *steepness indices* of H_0 at I.

The following condition ensures the uniformity of condition 1.7.B at all points where the function is defined.

1.7.C. DEFINITION. We say that a function H_0 of s variables is *steep in a domain G with coefficients $g, C_1, \ldots, C_{s-1}, \delta_1, \ldots, \delta_{s-1}$ and indices $\alpha_1, \ldots, \alpha_{s-1}$* if H_0 is steep at every point $I \in G$ with these coefficients and indices.

Conditions 1.7.C ensure an exponential estimate of the time of stability for all solutions $I(t)$, $\varphi(t)$ of the system (1.1) with initial conditions in the domain of $G \times T^s$ outside a small neighbourhood of the boundary of this domain; here T^s is the s-dimensional torus, and $\varphi \in T^s$.

1.8. Remarks on the steepness conditions. The geometrical meaning of steepness. We explain the fundamental definition of these three (Definition 1.7.A). Suppose that H_0 is steep at I on the plane λ with the coefficients C and δ, and index α. Let γ be any curve on λ that joins I to any other point at a distance d from I, where $d < \delta$. Then on this curve we can find a point $\tilde{I}$ such that the length of grad $(H_0|_\lambda)$ at this point is bounded below by a power estimate:

$$|\operatorname{grad}(H_0|_\lambda)|_{\bar{I}}| > Cd^\alpha.$$

Precisely this property of steep functions will be used in the proof of our main theorem.

We can give an intuitive interpretation of this property. We consider the graph of $H_0|_\lambda$. Then a traveller moving over the surface of the graph from a point over I to the other point necessarily has to surmount the steep slopes of this surface.

We remark that if the plane λ is not perpendicular to grad $H_0|_\lambda$, then grad $(H_0|_\lambda)_I \neq 0$, and the constants C, δ and α mentioned above can always be found; moreover, we can take α to be 0. If $\lambda \perp$ grad $H_0|_I$, as in Definition 1.7.B, then grad $(H_0|_\lambda)_I = 0$. In this case H_0 cannot be steep; for example, if we can find a curve γ passing through I and lying on λ such that at all points of it grad $(H_0|_\lambda)$ vanishes (for the details see §1.15). If H_0 is steep at I on λ, then $\alpha \geqslant 1$ always, and the steepness index cannot be smaller (this is a consequence of the smoothness of H_0).

We also note the relative independence of the steepness conditions at a point for different r, where $r = \dim \lambda$. The following fact points to this independence. There are functions of three variables some of which satisfy these conditions for $r = 1$ but not for $r = 2$, and vice versa for the others. A function of the first type is, for example, $(I_1 - I_2^2)^2 + I_3^2$ at the points where $I_1 - I_2^2 = 0$ and $I_3 \neq 0$, and of the second type, $I_1 + I_2^2 - I_3^2$ at all points.

Conditions close to the steepness conditions were introduced by Glimm in [16].

These conditions refer to the following situation: the system investigated is not (1.1), but a neighbourhood of a position of equilibrium of an arbitrary Hamiltonian system (see §2.2.B). In this case Glimm proved the so-called formal stability of the system under the assumption that a certain function defined by the system satisfies conditions close to the steepness conditions. Glimm did not clarify whether functions of "general position" satisfy his conditions. These functions satisfy the steepness conditions, as is proved in [17] (see also §1.13 above).

1.9. Variants of the steepness conditions. 1.9.A. Conditions of S-steepness. The condition $\left|\frac{dH_0(I)}{dI}\right| \geqslant g > 0$, which ensures the perpetual stability of the system (1.1) for $s = 1$, can be weakened. We can allow $\frac{dH_0}{dI}$ to vanish, but require it to satisfy a power estimate. A generalization of this condition to multi-frequency systems are the conditions of S-steepness stated below. They too allow grad H_0 to vanish. Nevertheless, they ensure the exponential estimate (1.5) of the time of stability of the system (1.6), although with a worse value of a. This is a simple consequence of our main theorem.

DEFINITION. We say that a function H_0 of s variables is *symmetrically steep* (or *S-steep*) *at* I if the following holds: For every $r = 1, \ldots, s$ there are numbers $C_r > 0$, $\delta_r > 0$, and $\alpha_r \geqslant 0$ such that H_0 is steep at I on every plane $\lambda \in \Lambda^r(I)$ with coefficients C_r and δ_r and indices α_r.

The requirement that these conditions hold uniformly at all points where H_0 is defined ensures an exponential estimate. This requirement is formulated by analogy with the conditions of steepness in a domain (see § 1.7.C).

1.9.B. Conditions of *P*-steepness. The requirement that the following conditions hold uniformly at all points where H_0 is defined, guarantees an exponential estimate of the time of stability of the systems in § 1.6 with a perturbation depending periodically on the time.

DEFINITION. We say that a function H_0 of s variables is *P-steep at* I_0 with coefficients $C_r > 0$ and $\delta_r > 0$, and indices $\alpha_r \geqslant 1$ $(r = 1, \ldots, s)$ if the following condition holds:

We denote by $\mathscr{K}$ the function $\mathscr{K}(I) = H_0(I) - \langle \omega(I_0), I \rangle$, where $\omega(I_0) = \operatorname{grad} H_0|_{I_0}$. ($\mathscr{K}$ is characterized by the fact that it differs from H_0 by a linear function, and that its gradient at I_0 is zero.) Then $\mathscr{K}$ must be S-steep at I_0 with coefficients C_r and δ_r, and indices α_r.

1.10. Steepness indices and the time of stability of a system. The constants a and b defining the estimates (1.4) and (1.5) of the main theorem depend only on the steepness indices $\alpha_1, \ldots, \alpha_{s-1}$ (defined in 1.7.C) of the unperturbed Hamiltonian H_0 and can be expressed in terms of them as follows:

$$a = \frac{2}{12\zeta + 3s + 14}, \qquad b = \frac{3a}{2\alpha_{s-1}}, \tag{1.9}$$

where

$$\zeta = [\alpha_1(\alpha_2 \ldots (\alpha_{s-3}(\alpha_{s-2} \cdot s + s - 2) + s - 3) + \ldots + 2) + 1] - 1$$

for $s > 2$, and $\zeta = 1$ for $s = 2$; s is the number of frequencies: $I = I_1, \ldots, I_s$.

For a we can obtain a better value:

$$a = \frac{1}{3\zeta + s + 4} - \sigma,$$

where $\sigma > 0$ is arbitrarily small. In a subsequent article we shall explain how this can be done, but it seems that this value could be improved.

Since $\alpha_r \geqslant 1$, we have $\zeta > s(s-1)/2$, where equality holds if $\alpha_r = 1$ for all r. Thus, in all systems with the same number s of frequencies, the best estimate of the time of stability strongly deteriorates as s increases.

In the next two subsections we describe two important classes of steep functions.

1.11. Quasiconvex functions. DEFINITION. A function H_0 is said to be *quasiconvex* in a domain G, $G \subset \mathbf{E}^s$, if for every $I \in G$:

a) $\operatorname{grad} H_0|_I \neq 0$;

b) the system

$$
\left\{
\begin{aligned}
&\sum_{i=1}^{s} \frac{\partial H_0(I)}{\partial I_i}\eta_i = 0,\\
&\sum_{i,j=1}^{s} \frac{\partial^2 H_0(I)}{\partial I_i\,\partial I_j}\eta_i\eta_j = 0
\end{aligned}
\right.
\tag{1.10}
$$

does not have real solutions $\eta = \eta_1, \ldots, \eta_s$ (apart from the trivial solution $\eta = 0$).

Condition b) can be restated as follows: the restriction of the second order term $\sum_{i,j=1}^{s} \frac{\partial^2 H_0(I)}{\partial I_i \partial I_j}$ of the Taylor series of H_0 about I to the hypersurface $\sum_{i=1}^{s} \frac{\partial H_0(I)}{\partial I_i}\eta_i = 0$ tangential to the level surface of H_0 at I has a fixed sign.

The level surfaces of these functions are convex, and this explains the term "quasiconvex function".

The steepness of quasiconvex functions in some neighbourhood of any point I in the domain of definition is obvious: for every subspace λ containing I and lying in the tangent hyperplane $\sum_{i=1}^{s} \frac{\partial H_0(I)}{\partial I_i}\eta_i = 0$, the graph of $H_0|_\lambda$ in a neighbourhood of I is close to an elliptic paraboloid (a dish). Hence we also see that all the steepness indices α_i ($i = 1, \ldots, s-1$) are equal to 1. Note that for every non-quasiconvex function with non-zero gradient among the steepness indices there is at least one that is greater than 1. Hence, the quasiconvex functions are "steepest".

It is not difficult to verify that for functions of two variables ($s = 2$), quasiconvexity is equivalent to the non-vanishing of the determinant (1.3). Hence, a function in "general position" of two variables is quasiconvex. For functions of three variables, quasiconvexity is equivalent to the condition that the determinant (1.3) is negative.

REMARK. As we have already remarked in § 1.10, the constant a defining the estimate (1.5) of the time of stability depends only on the steepness indices of H_0. The smaller these indices, the larger the estimate. The author conjectures that, in fact, if we compare systems (1.6) with the same number s of frequencies, then those for which H_0 has smaller steepness indices are in a certain sense significantly more stable than the systems with larger indices. In particular, systems with quasiconvex unperturbed Hamiltonians H_0 are the most stable.

For systems with three frequencies ($s = 3$), the validity of this conjecture would imply that the stability of a system essentially depends on the sign of the determinant (1.3): if it is negative, then the diffusion is much slower than if it is positive. It would be interesting to verify this dependence somehow, for example, on a computer.

REMARK. For periodic systems (1.7), the best estimate of the time of stability of the solutions is obtained for systems with a convex H_0, that is, a function for which the second order term in the Taylor series has fixed sign.

1.12. Conditions on a 3-jet function. We consider functions satisfying at every point

$$I = I_1, \ldots, I_s$$

of their domain the conditions:

a) $\operatorname{grad} H_0|_I \neq 0$;

b) the system

$$(1.11) \qquad \begin{cases} \displaystyle\sum_{i=1}^{s} \frac{\partial H_0(I)}{\partial I_i} \eta_i = 0, \\ \displaystyle\sum_{i,j=1}^{s} \frac{\partial^2 H_0(I)}{\partial I_i \partial I_j} \eta_i \eta_j = 0, \\ \displaystyle\sum_{i,j,k}^{s} \frac{\partial^3 H_0(I)}{\partial I_i \partial I_j \partial I_k} \eta_i \eta_j \eta_k = 0 \end{cases}$$

has no real solutions, apart from the trivial solution $\eta = \eta_1, \ldots, \eta_s = 0$.

It is not difficult to prove that in some neighbourhood of every point of their domain such functions satisfy the steepness conditions.

Obviously, the conditions (1.11) are weaker than (1.10); quasiconvex functions satisfy (1.11).

We also note that a function of three variables in "general position" satisfies (1.11).

1.13. The infinite degeneracy of non-jet functions. The question arises whether functions of "general position" of four or more variables are steep in a neighbourhood of a point. It turns out that they are. Moreover, the following theorem is proved in [17].

1.13.A. THEOREM. *Suppose that the gradient at I of a function H_0 of s variables ($s \geqslant 2$) is non-zero. If H_0 is not steep in any neighbourhood of I, then it is infinitely degenerate: the coefficients in the Taylor series of H_0*

about this point satisfy infinitely many independent algebraic equations.

By slightly modifying the proof of this theorem, we can obtain a similar assertion for the P-steep and S-steep functions defined in §1.9. These differ from Theorem 1.13.A by the absence of the condition $\operatorname{grad} H_0|_I \neq 0$.

1.13.B. THEOREM. *If H_0 is not P-steep (S-steep) in any neighbourhood of I, then it is infinitely degenerate in the sense of Theorem* 1.13.A.

1.14. Algebraic steepness criteria. **1.14.A.** *The set of jets of non-steep functions with a non-zero gradient.* The infinite degeneracy of non-steep functions follows from the following fact.

We denote by $J^r(s)$ the space of r-jets of functions of s variables at an arbitrary point I, that is, the space of vectors formed from the lowest coefficients, up to order r inclusive, in the Taylor series of the function about this point. Then we can find in every space $J^r(s)$ (r, s = 2, 3, 4, ...) a semi-algebraic set $\Sigma^r(s)$ such that the following conditions hold: firstly the r-jets of all non-steep functions with non-zero gradient lie in $\Sigma^r(s)$, more precisely, every function H_0 whose r-jets lie outside $\Sigma^r(s)$, either has $\operatorname{grad} H_0|_I = 0$, or is steep in some neighbourhood of I; secondly, for every $s = 2, 3, 4, \ldots$, the codimension of $\Sigma^r(s)$ in $J^r(s)$ tends to infinity as $r \to \infty$.

The second property of the sets $\Sigma^r(s)$ is a consequence of the estimate,[1] which can be proved on the basis of results in [17]:

$$(1.12)\qquad \operatorname{codim} \Sigma^r(s) \geqslant \begin{cases} \max\left[0,\ r-1-\frac{s(s-2)}{4}\right] & \text{for even } s, \\ \max\left[0,\ r-1-\frac{(s-1)^2}{4}\right] & \text{for odd } s. \end{cases}$$

We denote by $r_m(s)$ the smallest value of r such that functions of s variables with a non-zero gradient at I whose r-jets are of "general position" are steep in a neighbourhood of I. It follows from (1.12) that

$$r_m(s) \leqslant \begin{cases} \frac{s(s-2)}{4} + 2 & \text{for even } s. \\ \frac{(s-1)^2}{4} + 2 & \text{for odd } s. \end{cases}$$

In particular, $r_m(2) \leqslant 2$ and $r_m(3) \leqslant 3$; we have already mentioned these facts in §§1.11 and 1.12, when we asserted that functions of "general position" of two or three variables satisfy (1.10) and (1.11), respectively. Note that, in fact, $r_m(2) = 2$ and $r_m(3) = 3$.

1.14.B. *The constructive nature of the conditions distinguishing the jets of non-steep functions.* We prove (1.12) elsewhere.[2] We shall also describe there the algebraic conditions that define a certain subset $\sigma^r(s)$ of $J^r(s)$

[1] We assume that the codimension of $\Sigma^r(s)$ in $J^r(s)$ is exactly equal to the right-hand side of (1.12).
[2] Similar, but worse estimates are established in [17].

whose closure coincides with $\Sigma^r(s)$. Unfortunately, these conditions have an implicit character, that is, they have the form of (1.10) or (1.11), and not of (1.2) or (1.3). (More precisely, they have the form of a collection of systems of polynomial equalities and inequalities in the lower order Taylor coefficients and certain other variables (parameters).) The solubility of at least one system of the collection for fixed values of the Taylor coefficients means that the vector formed from these values belongs to $\sigma^r(s)$.) In spite of the implicit character of these conditions, to *a certain extent they answer the question about the effective verification of the steepness of the function along a jet.*

1.14.C. REMARK. The conditions defining $\sigma^r(s)$ can be regarded as a generalization of the steepness conditions (1.10) and (1.11). For these conditions are close in form to (1.10) and (1.11). Moreover, the conditions defining $\sigma^2(s) = \Sigma^2(s)$ for $s = 2, 3, 4, \ldots$, are equivalent to the conditions for quasiconvexity (1.10), and the conditions defining $\sigma^3(s) = \Sigma^3(s)$ for $s = 2$ and 3, are only slightly different from (1.11).

1.14.D. REMARK. The steepness indices α_i of functions with a non-zero gradient at I whose r-jets at this point lie outside $\Sigma^r(s)$ are bounded above by quantities depending only on s and r.

1.15. Examples of non-steep functions. The simplest examples of non-steep functions are linear functions of two and more variables:

$$H_0 = \sum_{i=1}^{s} a_i I_i + b.$$

Suppose that the domain G of H_0 contains a finite or infinite segment of a straight line for which the restriction of H_0 to it is constant. Then H_0 is not steep at any point of this segment, and so not in any domain intersecting it.

We consider a more general situation: there are a plane λ, $\dim \lambda \neq 0$, and a curve γ lying on this plane, such that the gradient of the restriction of H_0 to λ vanishes at all points of γ:

$$\text{(1.13)} \qquad \operatorname{grad}\ (H_0|_\lambda)|_I = 0 \quad \text{for all} \quad I \in \gamma.$$

Then H_0 is not steep at any point of γ.

Note that these examples of non-steep functions are also examples of non-P-steep and non-S-steep functions.

1.16. The importance of the steepness conditions for an estimate of the time of stability better than an a priori estimate. Necessary conditions for such an estimate. It follows straightaway from Hamilton's equations that for all solutions $I(t)$, $\varphi(t)$ of a system with the Hamiltonian

$$\text{(1.14)} \qquad H = H_0(I) + \varepsilon H_1(I, \varphi)$$

the point $I(t)$ is close to $I(0)$ during an interval of time much less than

$1/\varepsilon$. We call this estimate ($\ll 1/\varepsilon$) a lower a priori estimate for the time of stability of the solutions of the system. The steepness conditions on the unperturbed Hamiltonian H_0 guarantee an estimate better than an a priori one (an exponential estimate if H is analytic, and a power estimate if H has a certain number of derivatives). But, can this estimate hold for a system with an arbitrary H_0? It turns out that the answer is no. We can claim that a fairly large class $\mathfrak{M}$ of non-steep functions has the following property. Let $H_0 \in \mathfrak{M}$. Then a system with the Hamiltonian (1.14) and with a suitable perturbation εH_1, $H_1 = H_1(H_0)$, has for any $\varepsilon > 0$ solutions $I_\varepsilon(t)$, $\varphi_\varepsilon(t)$ such that $I_\varepsilon(t)$ leaves its initial position $I_\varepsilon(0)$ with a speed of order ε during a time interval $1/\varepsilon$. Thus, the time of stability of these solutions is much less than $1/\varepsilon$ and coincides with the a priori estimate.

Here is a simplified form of this theorem.

1.16.A. DEFINITION. A plane λ obtained by a translation from the linear hull of vectors with integer components is said to be *rational*. (As above, by a plane we mean an affine subspace of any dimension.)

1.16.B. NOTATION. We denote by $\mathfrak{M}$ the class of functions whose domains lie in $\mathbf{E}^s$, $I \in \mathbf{E}^s$, and which have the following properties For every $H_0 \in \mathfrak{M}$ we can find a rational plane λ in $\mathbf{E}^s$ and a curve γ, $\gamma \subset \lambda$, such that, firstly, (1.13) holds, and secondly, there is an equation $\dot{I} = V(I)$ with a smooth right-hand side for which γ is one of its phase curves.

1.16.C. THEOREM. *For every* $H_0 \in \mathfrak{M}$ *we can find an* $H_1 = H_1(I, \varphi)$ *and a one-to-one mapping* ξ: $[0, 1] \to \mathbf{E}^s$ *with the following property. For every* $\varepsilon > 0$ *there is a solution* $I_\varepsilon(t)$, $\varphi_\varepsilon(t)$ *of the system* (1.14) *that is defined on* $[0, 1/\varepsilon]$ *and such that* $I_\varepsilon(t) = \xi(\varepsilon t)$.

1.16.D. REMARK. If H_0 and the field V are analytic, then for H_1 we can take an analytic function such that the Hamiltonian (1.14) is analytic. This follows from the proof of the theorem.

The next result follows from Theorem 1.16.C.

1.16.E. COROLLARY. *In the class of conditions on the unperturbed Hamiltonian* H_0, *the condition* $H_0 \notin \mathfrak{M}$ *is necessary for an estimate of the time of stability of solutions of* (1.14) *to be better than an a priori estimate.*

1.16.F. REMARK. An important difference between functions belonging to $\mathfrak{M}$ and the non-steep functions considered at the end of §1.15 is that for functions from $\mathfrak{M}$ the corresponding plane λ is rational.

1.16.G. REMARK. The trajectories of $I_\varepsilon(t)$ coincide with the curve γ ($\{I \mid I = \xi(t), t \in [0, 1]\} \equiv \gamma$). For a suitable perturbation the curve γ becomes a "channel of superconductivity" along which $I(t)$ moves with speed ε.

For systems with two frequencies ($s = 2$), a necessary condition for an estimate of the time of stability to be better than an a priori estimate is that $\mathbf{E}^2$ does not contain straight lines with rational slopes on which $H_0 = H_0(I_1, I_2)$ is constant.[1] Below we give an example of a system for which H_0 has such a straight line and which has "fast" solutions $I_\varepsilon(t)$, $\varphi_\varepsilon(t)$.

1.17. Example of a system with a fast evolution. The system with the Hamiltonian

$$H = \frac{1}{2}(I_1^2 - I_2^2) + \varepsilon \sin(\varphi_1 - \varphi_2) \tag{1.15}$$

has the fast solution $I_1 = -\varepsilon t$, $I_2 = \varepsilon t$, $\varphi_1 = -\frac{1}{2}\varepsilon t^2$, $\varphi_2 = -\frac{1}{2}\varepsilon t^2$.

REMARK. For this system $H_0 = \frac{1}{2}(I_1^2 - I_2^2)$, consequently, the determinant (1.2) does not vanish anywhere. Thus, despite the existence of a measurably large Kolmogorov set of invariant tori, in the space of action variables I we can have the "channels of superconductivity" described in § 1.16.G. In this case the straight line $I_1 + I_2 = 0$ is such a channel.

Examples of systems with fast evolution similar to (1.15) have been given by Moser [18], Khapaev [19], and others.

1.18. Non-linear weakly-connected oscillators. Stability of a planetary system during an exponentially large interval of time. By non-linear weakly-connected oscillators we mean a system with a Hamiltonian (1.16) such that $H_0 = \sum_{i=1}^{s} f_i(I_i)$ and the second derivatives of the f_i and the gradient of H_0 do not vanish. If, moreover, these second derivatives have the same sign, then H_0 is convex, and so quasiconvex, hence, steep.

An example of this is a planetary system: $s + 1$ points are attracted to one another by Newton's law, and the mass of one point (the sun) is much greater than that of the others (the planets). In fact, it is well known that in some domain of the phase space the Hamiltonian of the system can be reduced to the form (1.6), and

$$H_0 = -\sum_{i=1}^{s} \frac{C_i^{(0)}}{I_i^2},$$

where the $C_i^{(0)} > 0$ are constants.

Suppose that the initial positions and the velocities of the bodies are roughly as in the solar system. We prove in § 12 that in this case the system does not collapse during a time T that is exponentially large in comparison

[1] Note that if H_0 is constant on a rational line, then the ratio of the frequencies $\dot{\varphi}_i(I) = \frac{\partial H_0(I)}{\partial I_i}$ $(i = 1, 2)$ of the unperturbed motion on this line is also constant and rational. The rationality follows from the rationality of the line. The frequencies $\dot{\varphi}_i(I)$ are the frequencies of the motion on a torus with "angular coordinates" $\varphi = \varphi_1, \varphi_2$ for a given value of I. Similar relations between the frequencies $\dot{\varphi}_i$ hold also for systems with more than two frequencies $(s > 2)$, with $H_0 \in \mathfrak{M}$ provided that the corresponding plane λ is rational.

with $1/\varepsilon$ where ε is the ratio of the masses of the planets to the mass of the sun. More precisely, suppose that the motion of the bodies in this system ensures that initially: a) the planets are moving along ellipses that are not too prolate; b) the lengths of the semi-major axes of these ellipses are sufficiently different from one another; c) the angles between the planes of the orbits are not too large, and the planets are moving in the same direction. Then, during time T, it is impossible for the bodies to collide or for a body to leave the system. Moreover, during the whole of this time, conditions a), b), and c) hold, and the lengths of the semi-major axes of the orbits hardly change.

We remark that this assertion does not follow directly from the main theorem, since when at least two bodies approach one another, the perturbation $H - H_0$ becomes small. We need to use another fact (the existence of the integral of the moment of momentum) in the same way as in a similar situation in [9]. The fact is that for fixed values of the lengths a_i of the semi-major axes, the length G of the moment of momentum vector attains its maximum close to the points of the phase space corresponding to the motion of the planets in circular orbits in the same plane and in the same direction. Hence, for any pairwise distinct $\alpha_i > 0$, we can find a number γ such that the conditions $a_i = \alpha_i (i = 1, \ldots, s)$ and $G \geqslant \gamma$ characterize a set in the phase space to each point of which there corresponds a position of the bodies of the system such that the distances between them are not too small. (The conditions $a_i = \alpha_i$ and $G \geqslant \gamma$ are sharper than a), b), and c).) Therefore, to prove that a close approach is impossible during the time in question, it is sufficient to prove that during this time the changes in the a_i are small. But it is well known that $a_i = C_i^{(1)} I_i^2$, where the $C_i^{(1)} > 0$ are constants. Hence, a large deviation of the a_i from their initial values would imply a large deviation of the action variables I_i from their initial values. But this would contradict the main theorem, since it would involve a small perturbation.

We digress slightly in the next two subsections, and consider non-Hamiltonian perturbations.

1.19. Stability of the solar system. It is estimated that our solar system has existed for such an immense period as 4×10^9 times the age of the earth. This period could be compared with the estimate (1.5), or with other more exact estimates, but we do not do this. Even if this period could be explained only on the basis of the slowness of Arnol'd diffusion, it remains quite obscure how at some instant the planets began to move in almost circular orbits, and almost in the same plane. Hence, in an investigation of the stability of the solar system we need to take account of non-Hamiltonian perturbations. These are insignificantly small in comparison with Hamiltonian perturbations, but the effect of the latter over a large interval of time is also very small. Apparently, it is cancelled by the action of non-Hamiltonian perturbations, which tend to send the system into a stable state.

One of the explanations of this property of non-Hamiltonian perturbations is that to a large extent they cancel the energy $h_i = -C_i^{(0)}C_i^{(1)}\frac{1}{a_i}$ of the planets rather than the length G, so that G tends to its maximum value for the available lengths of the a_i. But then, by §1.18, the system tends to a motion in orbits with zero eccentricity, and the planes of the orbits do not deviate from one another.

Possibly the influence of tides and the resistance of the surrounding medium play a fundamental role. Simple calculations show for example, that a necessary relation between the decrease of the a_i and G holds if we take into account only the resistance of the surrounding medium and assume that the particles of the medium are at rest relative to the centre of mass of the system.

1.20. The main theorem and non-Hamiltonian perturbations. Actually, it does not follow from the main theorem as stated in the text that if the non-Hamiltonian perturbations are very small, then the time of stability of the system becomes very large. But it follows from an analysis of the proof of this theorem that the exponential estimate (1.5) holds also for systems with very small non-Hamiltonian perturbations. A rough statement of this generalization is as follows.

THEOREM. *We denote by* ε_H *the magnitude of the Hamiltonian perturbation, and by* ε_N *that of the non-Hamiltonian perturbation. We write*

$$\varepsilon = \max\left[2\varepsilon_H, \left(\log\frac{1}{(2\varepsilon_N)^{1-\delta}}\right)^{-1/a}\right],$$

where $\delta = \delta(\varepsilon) > 0$ *is a known function that tends to* 0 *as* $\varepsilon \to 0$. *Suppose that* $0 < \varepsilon < \varepsilon_0$. *Then* (1.4) *and* (1.5) *hold for all solutions of the system. Here* a *and* ε_0 *are the same as in the main theorem.*

This assertion can be stated as follows: *The speed of diffusion, that is, the mean speed of displacement of the action variables, over a sufficiently large interval of time, is bounded by*

$$\max\left\{2\varepsilon_H \exp\left[-\left(\frac{1}{2\varepsilon_H}\right)^a\right], (2\varepsilon_N)^{1-\delta}\right\}.$$

Thus, a "superposition of estimates" holds approximately: the estimate of the speed of diffusion produced by the combined action of both forms of perturbations differs insignificantly from the sum of our estimate of the speed of Arnol'd diffusion (that is, the diffusion arising only from Hamiltonian perturbations) and the a priori estimate of the speed of diffusion produced by the non-Hamiltonian perturbations.

The author thanks V.I. Arnol'd for introducing him to the problem, and for teaching him the technique of constructing changes of variables by the

method of successive approximations, which enabled him to obtain coordinates in the phase space that are convenient for the investigation of the systems in question. The author also thanks A.D. Bryuno for some useful remarks.

§2. Unsolved problems. Conjectures. Generalizations

2.1. The principal questions. 2.1.A. *An exponential estimate from above for the time of stability.* Although we have constructed examples of solutions in which $I(t)$ moves away from $I(0)$ exponentially slowly (see §§1.2 and 1.3), we have not proved that such solutions exist for any system with a Hamiltonian of "general position". A proof would allow us to establish once and for all the exponential character of the dependence of the time of stability of the system on the magnitude of a perturbation. There is no estimate of the measure of unstable solutions.

The possible existence of unstable solutions, that is, of Arnol'd diffusion, is clarified in §11.4.

2.1.B. *The dependence of the time of stability of a system on the steepness indices.* We feel that this dependence, which is indicated but not proved, by our results is the most important consequence of practical interest.

This dependence was first put forward as a conjecture in §1.11, where it was suggested that it might be verified on a computer. It would be more interesting to prove the result analytically. The author feels that fairly precise upper and lower estimates of the minimum time of stability (over all solutions) could be obtained by means of the steepness indices. This would allow us to give a rigorous proof of the dependence mentioned above.

2.1.C. *The direction of diffusion.* It would also be interesting to investigate the direction in which unstable solutions move, and the preferred direction of diffusion. In a certain sense the steepness of H_0 can differ in different directions, and this must affect the direction of diffusion. Besides, other factors influence the direction of diffusion, namely, the location in the space of action variables of the resonance zones, which will be defined in §3.

2.1.D. *Stability and resonances.* Another problem is to explain the observed stability of the so-called resonance relations (for the details see, for example, [20]). To all appearances non-Hamiltonian perturbations impart this stability to real systems. What conditions ensure it?

The same question arises when an unperturbed system is of general form and not Hamiltonian, as in this article (see [37]). In the case of two frequencies, the stability of resonances in systems of this form has been discussed by Neishtadt [35], [36].

2.2. Generalization of the exponential estimate to Hamiltonian systems of other types and to canonical mappings. We mentioned in §1.6 that the exponential estimate of the time of stability of a system in the action variables holds not only for the systems (1.6), but also for the systems (1.7) with a perturbation depending periodically on time. Here we consider two further types of Hamiltonian systems, specifically, a neighbourhood of the equilibrium position of the Hamiltonian systems, and two types of canonical mappings. We state propositions on an exponential estimate of the stability of these systems and mappings.

We do not prove these assertions formally. But the following fact points to their validity. There are a number of theorems that hold with insignificant variations both for the systems (1.1) and (1.7) as well as for the systems and mappings discussed in this section. Among these theorems is one about Kolmogorov tori, which is close in its direction to an exponential estimate (this theorem is referred to in § 1.2; see also [27]). The author is convinced that the exponential estimate is a theorem of this type.

We mention that the systems and mappings below have been investigated by several authors such as Siegel [4] and [5], Moser [21], [22], [23], [12], [13], and [18], Arnol'd [10], Glimm [16], Contopoulos [24], Hénon [25], Bryuno [26], and others. They have obtained deep and important results, some of which are used later in this section.

2.2.A. *The mapping of a 2s-dimensional ring.* We consider the mapping

$$A\colon (I, \varphi) \mapsto (I', \varphi'), \quad \text{where} \quad I' = I - \frac{\partial S}{\partial \varphi}, \quad \varphi' = \varphi + \frac{\partial S}{\partial I'},$$

and the function S, the so-called generating function, has the form $S(I', \varphi) = S_0(I') + \varepsilon S_1(I', \varphi)$, $\varepsilon \ll 1$. Here $I \in G \subset \mathbf{E}^s$, and $\varphi \in T^s$; $\mathbf{E}^s$ and T^s are, respectively, the s-dimensional Euclidean space and torus.

THEOREM. *Let S_0 be P-steep in G* (for the definition, see § 1.9.B). *Then there are constants $a > 0$ and $b > 0$, depending only on the index of P-steepness of S_0 in G, and having the following property*:

Let $I^{(0)}$, $\varphi^{(0)}$ be an arbitrary point of the domain $G \times T^s$ that does not lie in a small neighbourhood of the boundary of this domain. Then

$$|I^{(m)} - I^{(0)}| < \varepsilon^b$$

for all integers $m \in [0, T]$, where

$$T = \frac{1}{\varepsilon} \exp\left(\frac{1}{\varepsilon^a}\right),$$

and $I^{(m)}$, $\varphi^{(m)}$ denotes the point $A^m(I^{(0)}, \varphi^{(0)})$.

2.2.B. *A neighbourhood of an equilibrium position of a conservative system.* We consider a neighbourhood of a singular point of an arbitrary autonomous Hamiltonian system with s degrees of freedom. We assume that the eigenvalues of the linearized system are purely imaginary, and denote them by $\pm i\omega_1, \ldots, \pm i\omega_s$. Suppose that the eigenfrequencies ω_j do not satisfy any resonance relation of order l or lower order, that is,

$$\sum_{j=1}^{s} k_j\omega_j \neq 0$$

for all vectors $k = k_1, \ldots, k_s$ with integer components such that $0 < \sum_{j=1}^{s} |k_j| \leqslant l$, where l is some natural number.

It is well known (see, for example, [27]) that in this case in a neighbourhood of a singular point we can make a canonical change of variables such that the new coordinates p, q at this point are zero, and in these coordinates the Hamiltonian H of the system has the form

$$H = H_0(I) + H^{(l+1)}(p, q). \tag{2.1}$$

Here H_0 is a polynomial in $I = I_1, \ldots, I_s$ of degree not higher than the integer part $[l/2]$ of $l/2$, $I_j = (p_j^2 + q_j^2)/2$, and the Taylor series of $H^{(l+1)}$ in terms of p and q about the origin does not contain terms of order less than $l+1$.

We investigate a system whose Hamiltonian has the form (2.1).

THEOREM. *Suppose that $H_0 = H_0(I)$ is steep in a neighbourhood of $I = 0$. Then there are constants $a > 0$ and $b > 0$ depending only on the steepness indices α_i of H_0, with the following property.*

Let

$$l \geqslant \frac{2}{b} + 1 \qquad (b = b(\alpha_1, \ldots, \alpha_{s-1})). \tag{2.2}$$

Let $p(t)$, $q(t)$ be an arbitrary solution of the system with an initial condition $p(0)$, $q(0)$ sufficiently close to the origin. We put $\varepsilon = |I(0)|^{(l-1)/2}$, where $I_j(t) = (p_j^2(t) + q_j^2(t))/2$. Then, $|I(t) - I(0)| < \varepsilon^b$ for all $t \in [0, T]$, where $T = \frac{1}{\varepsilon} \exp \frac{1}{\varepsilon^a}$.

From this theorem follows the validity of the estimate of the time during which the trajectory of the solution lies in a small neighbourhood of the singular point; this estimate depends exponentially on the initial distance to the singular point. A sufficient condition for the stability of the singular point is the steepness of H_0 and the absence of resonances of order l and lower orders, where l and the α_j satisfy (2.2).

2.2.C. *A neighbourhood of an equilibrium position of a system depending periodically on time.* We consider a neighbourhood of a stationary singular point of a system whose Hamiltonian is 2π-periodic in t. (The Hamiltonian of a conservative system can be reduced to this form in a neighbourhood of a periodic trajectory.) We assume that all the multipliers, that is, the eigenvalues $\lambda_1^{\pm}, \ldots, \lambda_s^{\pm 1}$ of the substitution of the fundamental matrix of solutions of the linearized system for the period 2π, have modulus 1. Suppose that these numbers do not satisfy any resonance relation

$$\lambda_1^{k_1} \ldots \lambda_s^{k_s} = 1 \tag{2.3}$$

of order l or lower order, that is, for all $k = k_1, \ldots, k_s$ such that $0 < \Sigma \mid k_j \mid \leqslant l$. Then in a neighbourhood of a singular point the Hamiltonian of the system can be reduced by a canonical transformation 2π-periodic in t, to the form (2.1), the only difference being that $H^{(l+1)}$ is 2π-periodic in t: $H^{(l+1)}(p, q, t + 2\pi) \equiv H^{(l+1)}(p, q, t)$.

The analogue of our main theorem for systems with a Hamiltonian of this type differs from Theorem 2.2.B only in that we impose on H_0 the conditions for P-steepness (§1.9.B) instead of the conditions for steepness.

2.2.D. *A neighbourhood of a fixed point of a canonical mapping.* We assume that all the eigenvalues $\lambda_1^{\pm 1}, \ldots, \lambda_s^{\pm 1}$ of the linearization of the mapping A have modulus 1 at a fixed point. Suppose that they do not satisfy any resonance relation (2.3) of order l or lower order. Then in a neighbourhood of the fixed point we can construct coordinates p, q that vanish at this point, and in which the mapping A (with accuracy up to the terms of order l in its Taylor series about $p = q = 0$) has the form

$$(I, \varphi) \mapsto \left(I, \varphi + \frac{\partial S_0}{\partial I}\right).$$

Here $I = I_1, \ldots, I_s$, $\varphi = \varphi_1, \ldots, \varphi_s$, where I_j and φ_j are defined by $p_j = \sqrt{(2I_j)} \sin \varphi_j$, $q_j = \sqrt{(2I_j)} \cos \varphi_j$, and $S_0(I)$ is a polynomial in $I_1, \ldots, I_s$ of degree not higher than $[l/2]$.

We investigate a mapping of this type.

THEOREM. *Suppose that S_0 is P-steep in some neighbourhood of the origin. Then there are constants $a > 0$ and $b > 0$ depending only on the indices of P-steepness of S_0 and having the following property. Let $l \geqslant \frac{b}{2} + 1$. Then for every point $p^{(0)}, q^{(0)}$ in some neighbourhood of the fixed point $p = q = 0$ we have $\mid I^{(m)} - I^{(0)} \mid < \varepsilon^b$ for all integers $m \in [0, T]$, where $T = \frac{1}{\varepsilon} \exp \frac{1}{\varepsilon^a}$. Here $I^{(m)}$ denotes $[(p_j^{(m)})^2 + (q_j^{(m)})^2]/2$, where $p^{(m)}, q^{(m)} = A^{(m)}(p^{(0)}, q^{(0)})$ and $\varepsilon = \mid I^{(0)} \mid^{(l-1)/2}$.*

§3. The main ideas of the proof of the exponential estimate

In this section we discuss certain ideas and the approach to the proof of the main theorem (Theorem 4.4). For simplicity we give the arguments for systems without parameters, that is, for systems with a Hamiltonian

$$H = H_0(I) + \varepsilon H_1(I, \varphi). \tag{3.1}$$

3.1. The analytic and geometric parts of the proof. The proof of the main theorem can be divided into two parts. The first, analytic part, relies on the methods of perturbation theory developed by Kolmogorov and Arnol'd (see [6]–[9]); in this part we use in a straightforward way the technique developed in [8]. The purpose of the main calculations in the first part is to eliminate the so-called "non-resonance terms". The second part is based on an investigation of the geometry of the "resonance zones" and the "planes of fast drift" described below.

We remark that the general scheme of the proof of the main theorem is close to the general scheme of the proof given by Glimm in [16] (see §1.8).

We explain the analytical part of the proof in §3.2.

3.2. Almost integrals. In this subsection we explain the main reason for the stability of the projections $I(t)$ of the solutions of the system (3.1) on the space of the action variables I. It turns out that the domain G ($I \in G$) of H_0 is "permeated" by surfaces of different dimensions such that the average speed of the projections $I(t)$ across these surfaces is much less than the perturbation ε. By the mean speed we mean the ratio of the distance between the initial and final points to the elapsed time, which we assume to be sufficiently large.

These surfaces are planes in the space of the variables I, more precisely, the intersections of planes with certain subdomains of G. Here and elsewhere, by planes we mean affine subspaces of any dimension.

For every sufficiently small subdomain U of G we can find linear functions such that the intersections of the level surfaces of the restrictions of these functions to U are surfaces with the properties described above. We call these functions *almost integrals.*

3.2.A. *Description of the intersections of the level surfaces of almost integrals.* Let U be an arbitrary subdomain of G. We describe the planes λ containing the intersections of the level surfaces of almost integrals defined on the whole of this subdomain (provided that these almost integrals exist). All these planes are parallel to one another; they are obtained by a translation from the linear hull of the resonance vectors for U of lower orders, as defined below.

We denote by $\mathbf{Z}^s$ the lattice of s-dimensional vectors $k = k_1, \ldots, k_s$ with integer components. Let $\nu > 0$ be an arbitrary number.

DEFINITION. A vector $k \in \mathbf{Z}^s$ is called a *ν-resonance vector* for a set U if $|\langle k, \omega(I)\rangle| < \nu$ for some $I \in U$. Here $\omega(I) = \operatorname{grad} H_0|_I$ and $\langle k, \omega\rangle$ is the scalar product:

$$\langle k, \omega\rangle = \sum_{i=1}^{s} k_i\omega_i.$$

By the order of a vector $k \in \mathbf{Z}^s$ we mean $|k| = \sum_{j=1}^{s} |k_j|$. We denote by λ_0 the linear hull of all ν-resonance vectors for U of order not greater than some number N. Let $r \leqslant s$, where $r = \dim \lambda_0$.

Then in U there exist $s - r$ linearly independent almost integrals. The *intersections of the level surfaces of these integrals are obtained from the plane λ_0 by a translation.*

The properties of almost integrals are determined by the parameters ε, ν, and N. In particular, the smaller ε and the larger ν, the more they become similar to true first integrals.

3.2.B. *Construction of almost integrals.* Let $H = \sum_k h_k(I)e^{ik\varphi}$ be the expansion of the Hamiltonian H in a Fourier series. The existence of almost integrals is proved by reducing H to a special form by a canonical change of variables close to the identity. This special form is such that H hardly differs from the function whose Fourier series contains only those harmonics with the resonance numbers k. The reduction is carried out in Lemma 10.3.

This lemma, which is about the "elimination of non-resonance harmonics", is roughly as follows: In $U \times T^s$ we can make the change of variables $I, \varphi \to J, \psi$ such that the Hamiltonian has the form

$$H = \bar{H}(J, \psi) + R(J, \psi), \tag{3.2}$$

where the number k of each harmonic in the Fourier series for $\bar{H}$ belongs to λ_0:

$$\bar{H} = \sum_{k \in \lambda_0 \cap \mathbf{Z}^s} \bar{h}_k(J)\, e^{ik\psi},$$

and the effect $|\operatorname{grad} R|$ of the remainder R is very small for large N:

$$|\operatorname{grad} R| < \varepsilon e^{-N^{1-\varkappa}}; \tag{3.3}$$

here $\varkappa > 0$ can be taken to be arbitrarily small. The deformation of coordinates $|(J, \psi) - (I, \varphi)|$ under the change of variables is bounded above by a product of powers of ε, ν, and N.

To prove the existence of almost integrals it remains to use the following simple fact, which is a consequence of Hamilton's equations. We consider the system with an arbitrary Hamiltonian of the form $\mathcal{H} = \sum h_k(J)e^{ik\psi}$. Then for every solution of this system the vector J defined by the harmonic $h_k(J)e^{ik\psi}$ is increasing in the direction of k.

It follows from this proposition that the projections $J(t)$ of the solutions of the system with the Hamiltonian $\bar{H}$ lie precisely on the planes obtained from λ_0 by a translation. Hence, for the solutions $J(t)$, $\psi(t)$ of the system with the Hamiltonian (3.2) the length of the component of $\dot{J}(t)$ perpendicular to λ_0 is not greater than $|\operatorname{grad} R|$. If we take N sufficiently large, then by (3.3), $|\operatorname{grad} R|$ is much less than ε. We return to the variables I and φ. If we take ε, ν, and N so that the deformation of coordinates is sufficiently small, then we obtain the almost integrals described above.

3.2.C. REMARK (Drift and vibration.) The geometrical interpretation of the reduction of the Hamiltonian to the form (3.2) is as follows. Let

$I(t)$, $\varphi(t)$ be an arbitrary solution of the system (3.1) whose trajectory lies in $U \times T^s$. Then we can separate the motion of $I(t)$ into two motions. One is that of the point $J(t) = J(I(t), \varphi(t))$, while the other is the one we have eliminated by the change of coordinates. The second is a small amplitude oscillation (a "vibration"); it is not essential for the stability of the system. On the other hand, the motion of $J(t)$ is simpler than that of $I(t)$, since the speed of $J(t)$ perpendicular to λ_0 is very small.

We can say that the motion of $J(t)$ is the average motion or "drift" of $I(t)$. Here the drift across planes λ obtained from λ_0 by a translation is very slow, and the speed of drift along them is of order ε. For this reason we call the planes λ "planes of fast drift".

3.2.D. REMARK. We have constructed the new variables J, ψ in terms of I and φ by making a large but finite number ($\sim N$) of successive substitutions. These are analogous to those made by Kolmogorov and Arnol'd in proving the existence of a measurably large set of invariant tori (see §1.2). Note that they considered an infinite and not a finite sequence of these substitutions.

Now we explain the second, geometrical, part of the proof of the main theorem.

3.3. Traps. Both the number of linearly independent almost integrals and the location of their level surfaces in the space of action variables depend on the subdomain of G in question. In particular, there are subdomains in which we can distinguish a complete set of s linearly independent almost integrals. In these subdomains the planes λ degenerate to points; hence, the speed of drift of the projections $I(t)$ of the solutions $I(t)$, $\varphi(t)$ in them is practically zero for sufficiently large N.

Where the number of almost integrals is less than s, $I(t)$ can drift with a speed of order ε (that is, a relatively large speed) along planes λ of non-zero dimension. But where H_0 is steep, the almost integrals from the given and neighbouring subdomains, trap $I(t)$ in a set whose diameter is small together with ε; $I(t)$ is in a trap of small size, through which it can move with speed ε.

To construct the traps we need the following concept.

3.3.A. Definition of resonance surfaces. Let $k \in \mathbf{Z}^s$ be an arbitrary non-zero vector. We call the set of points $I \in G$ such that $\langle k, \omega(I) \rangle = 0$ the *1-fold resonance surface* defined by k. Let $k^1, \dots, k^r$ be r linearly independent vectors from $\mathbf{Z}^s$. The set of $I \in G$ such that $\langle k^j, \omega(I) \rangle = 0$ for all $j = 1, \dots, r$ is called the *r-fold resonance surface* defined by $k^1, \dots, k^r$.

Note that the surface defined by $k^1, \dots, k^r$ is the same as the intersection of the r 1-fold resonance surfaces $\mathcal{R}(k^j)$ ($j = 1, \dots, r$), where $\mathcal{R}(k^j)$ is defined by k^j.

We fix some $N > 1$ and consider only those resonance surfaces defined by vectors whose order is not greater than N. It is useful to visualize the structure of the set of resonance points, that is, the union of all these surfaces (see Fig. 1). Note that the structure of this set is determined by the structure of the set of resonance frequencies:

$$\{\omega \in \mathbf{R}^s \mid \exists k \in \mathbf{Z}^s \quad \text{such that} \quad 0 < |k| \leqslant N \quad \text{and} \quad \langle k, \omega \rangle = 0\}.$$

3.3.B. Resonance zones and blocks. We are interested not so much in the resonance surfaces as in their neighbourhoods, which we call resonance *zones.* In the proof of the main theorem a zone is a neighbourhood of a special form of a surface $\mathcal{R}$, and it is characterized by a certain parameter ν, called the *resonance width* of the zone. The zone contracts to $\mathcal{R}$ as $\nu \to 0$.

In addition to zones, with each resonance surface $\mathcal{R}$ we can associate a set, called a *block*, which is the difference between the resonance zone of $\mathcal{R}$ and the union of the resonance zones of all surfaces $\mathcal{R}'$ not containing $\mathcal{R}$, $\mathcal{R}' \not\supseteq \mathcal{R}$ (Fig. 1).

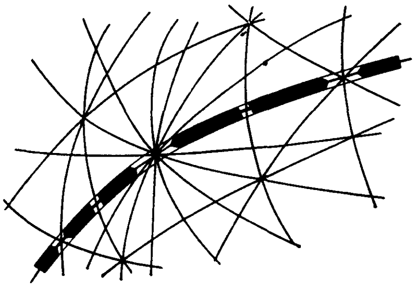

Fig. 1. The intersections of the resonance surfaces with the level surface of a function H_0 of 3 variables ($s = 3$). The intersection with the surface H_0 = const of the block corresponding to the 1-fold surface $\mathcal{R}$ is shown in black.

By the *multiplicity* of a zone and a block we mean the multiplicity of the corresponding surface $\mathcal{R}$, and the vectors *defining* the zone and the block are those defining this surface.

We call the complement in G of the union of all resonance zones the *0-fold* or *non-resonance* block; we assume that its defining vector is the zero vector.

We discuss which vectors $k \in \mathbf{Z}^s$ are resonance vectors for these blocks and zones. Here and elsewhere we are only interested in those k whose order is not greater than N, $|k| \leqslant N$.

Both on a surface and on a zone the exact resonance relations $\langle k, \omega(I)\rangle = 0$ can hold with very different k. In contrast to this, for a sufficiently small $\nu > 0$, all the ν-resonance vectors for a block belong to the linear hull λ_0 of the vectors defining this block. Hence, we can apply the arguments of §3.3 to a block and find that in a block the speed of drift of $I(t)$ across the planes λ obtained from λ_0 by a translation is practically zero for sufficiently large N.

3.3.C. *Boundedness of fast drift.* With each resonance surface we associate a block and consider the set of all such blocks, adding to it the non-resonance block. We take the blocks so that all the 1-fold zones involved in the construction of any set of blocks have the same resonance width. In this case the blocks of a set cover the whole of G.

We assume that $I(t)$ can move in each block only along the planes λ. The block is part of the corresponding zone. It turns out that the steepness conditions on H_0 ensure that $I(t)$ can leave a resonance zone by a displacement along a plane λ in any direction (Fig. 2).

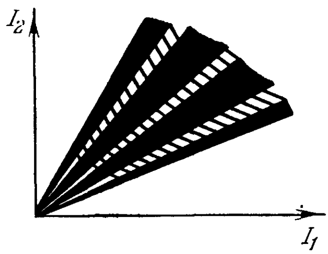

Fig. 2. Three neighbouring 1-fold resonance zones of a planetary system with two planets ($s = 2, H_0 = -\frac{C_1}{I_1^2} - \frac{C_2}{I_2^2}$, see § 1.18). The lines in each zone are the intersections of the lines λ of fast drift with this zone.

We can choose the resonance width of the zones used to construct the set of blocks so that if a point leaves each zone close to the corresponding block, then it passes into a block of lower multiplicity than that of this zone. For this reason, as $I(t)$ moves from its initial position $I(0)$, it passes from one block into a block of lower multiplicity until it reaches a non-resonance block. In a non-resonance block the planes λ degenerate to points. By our assumption, the speed in this block is zero, and so it is impossible for $I(t)$ to move further away from $I(0)$. The point $I(t)$ is in a trap formed from the intersections of the resonance zones with the planes λ (Fig.3).

3.3.D. *Intersection of a plane of fast drift and a resonance zone. The role of the steepness condition.* Now we explain how the steepness conditions ensure that a point can leave a resonance zone by a displacement along the planes λ.

First we consider the intersection of a plane λ and a resonance surface. Let $\mathscr{R}$ be an arbitrary resonance surface, λ an arbitrary plane of fast drift in the block corresponding to $\mathscr{R}$, and $k^1, \ldots, k^r$ the vectors defining $\mathscr{R}$. The plane λ is obtained by a translation from the linear hull of these vectors, and $\mathscr{R}$ is the set of points where $\omega(I) = \text{grad } H_0|_I$ is perpendicular to them. Hence, grad $H_0|_I$ is perpendicular to λ if and only if $I \in \mathscr{R}$. But the projection of grad H_0 on λ is grad $(H_0|_\lambda)_I$ for all $I \in \lambda$.

Hence, $I \in \mathscr{R} \cap \lambda$ *if and only if* $\operatorname{grad}(H_0|_\lambda)_I = 0$.

Actually, in one aspect we can strengthen this assertion: The larger $|\operatorname{grad}(H_0|_\lambda)_I|$, the greater the distance from I to $\mathscr{R}$. *The power character of the lower estimate of the length of the gradient (valid for steep functions) requires an upper power estimate for the diameter of the intersection of* λ *and a resonance zone, depending on the resonance width of this zone.*

3.3.E. REMARK. If the intersection of $\mathscr{R}$ and λ is not pointwise, then there can be disastrous consequences. For a suitable perturbation εH_1 every smooth curve γ lying in $\mathscr{R} \cap \lambda$ becomes a "channel of superconductivity" along which $I(t)$ moves with speed ε. These channels were described in §1.16, where they were characterized by another condition, namely, that γ lies on a rational plane λ, and $\operatorname{grad}(H_0|_\lambda)_I = 0$ for all $I \in \gamma$. By the arguments at the beginning of §3.3.D, this condition is, in fact, equivalent to $\gamma \subset \mathscr{R} \cap \lambda$.

3.3.F. REMARK. Every plane λ of fast drift touches the level surfaces of H_0 at each point of $\mathscr{R} \cap \lambda$ and nowhere else.

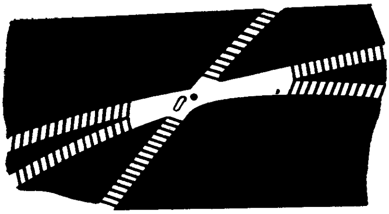

Fig. 3. The part of the level surface of $H_0 = H_0(I_1, I_2, I_3)$ close to the point 0 of the intersection of this surface with a 2-fold resonance line. The black part, without gaps, is the non-resonance domain, and the short lines are the intersections of 1-fold zones with the lines λ of fast drift. In the white part the planes of fast drift are 2-dimensional, and the 1-dimensional and 2-dimensional planes are parallel to the tangent plane to the level surface of H_0 at 0. It is clear from the figure that if we assume $I(t)$ to move along only planes of fast drift, then it is perpetually close to its initial position.

3.4. Derivation of the exponential estimate. Here we explain how to obtain the relations from which the exponential estimate follows.

In order to be sure that the construction from the intersections of the resonance zones with the planes λ (roughly described in §3.3.C) actually produces traps, we need to take the resonance width of these zones sufficiently small. The 1-fold zones have the smallest width, and it is sufficient to take it not greater than some positive power of $1/N$, which we denote by $\nu_0 = \nu_0(N)$. (The precise definitions of $\nu_0(N)$ and other functions considered in this subsection will be given elsewhere.) All the notation used in the proof of the main theorem is the same as here, except that instead of ε we use M, where M has a slightly different meaning. The power character of the estimate of the width follows from that of the estimates in the steepness conditions. We take the resonance width of the 1-fold zones in the construction to be as large as possible, that is, $\nu_0(N)$.

We have assumed that the motion in a block is just along the planes λ. In fact, this is not so, and in the above construction of traps we need to replace the planes λ by neighbourhoods of them. If the thickness of these neighbourhoods is not greater than some $b_1 = b_1(N)$, a power of $1/N$, then this modification to our construction does not change anything fundamental and, as before, it gives traps of diameters that are small with $1/N$.

To justify the claim that in each block $I(t)$ moves in a b_1-neighbourhood of the plane λ we only need to use the lemma about the elimination of the non-resonance harmonics (see §3.2.B). For this it is sufficient that the following three conditions hold. Firstly, the lemma about the elimination is applicable to every block; for this it is sufficient to take ν in this lemma not greater than ν_0, the smallest resonance width of the zones in the construction of the blocks. Secondly, the deformation $|(I, \varphi) - (J, \psi)|$ does not exceed $b_1/2$; this is needed to prevent $I(t)$ from leaving a $b_1/2$-neighbourhood of λ by a "vibration". Thirdly, $I(t)$ does not leave this neighbourhood by a drift; for this we need to restrict the time of motion T of this point.

As we noted in §3.2.B, the deformation of coordinates has an upper estimate by a power product of ε, ν, and N, which we denote by $b = b(\varepsilon, \nu, N)$, and the speed of drift across the planes λ does not exceed $\varepsilon \exp(-N^{1-\varkappa})$, where $\varkappa > 0$ is arbitrarily small. Hence, for these three conditions to hold it is sufficient that

$$\nu \leqslant \nu_0(N),$$

$$b(\varepsilon, \nu, N) \leqslant \frac{1}{2} b_1(N),$$

$$T \cdot \varepsilon \exp(-N^{1-\varkappa}) \leqslant \frac{1}{2} b_1(N).$$

A value of T satisfying these inequalities is an estimate of the time of stability of the system. Our problem is to obtain the largest possible value of T. It follows from the third inequality that for this we must have N as large as possible. If we substitute the explicit values of $\nu_0(N)$, $b(\varepsilon, \nu, N)$, and $b_1(N)$ in the first two inequalities, then it is clear that the largest N is obtained by replacing these inequalities by equalities, and using them to express ν and N in terms of ε. These equalities are of power character, so that ν and N are positive powers of ε and $1/\varepsilon$, respectively.

It follows from the third inequality that we can take $T = \frac{1}{\varepsilon} \exp N^{1-2\varkappa}$. Since $N = N(\varepsilon)$ is a power of $1/\varepsilon$, we obtain an exponential estimate of the time of stability T:

$$T = \frac{1}{\varepsilon} \exp \frac{1}{\varepsilon^a}.$$

The main theorem is formulated in §4 and proved in §§5–9. The proof is based on the lemma about eliminating non-resonance harmonics, and on certain technical lemmas in §10.

§4. Steepness conditions. Precise statement of the main theorem

4.1. Steepness conditions. We denote by $\mathbf{E}^s$ the Euclidean space of vectors $I = I_1, \ldots, I_s$ with the norm $|I| = \sqrt{(\sum_{j=1}^{s} I_j^2)}$, and by $\Lambda^r(I)$ the set of all r-dimensional affine subspaces (planes) λ passing through the point $I \in \mathbf{E}^s$, that is, $I \in \lambda$.

Let H_0 be an arbitrary function defined in a domain $G \subset \mathbf{E}^s$ (here and in what follows a domain is an open set). Let $I \in G$ and $\lambda \in \Lambda^r(I)$ be arbitrary. We denote by $m_{I,\lambda}(\eta)$ the lower bound of the lengths of grad $(H_0|_\lambda)$ on the intersection of G with the sphere on λ of radius η and with centre at I:

$$m_{I,\lambda}(\eta) = \inf_{\{I' \in \lambda \cap G:\ |I' - I| = \eta\}} |\operatorname{grad}(H_0|_\lambda)|_{I'}|,$$

where $H_0|_\lambda$ is the restriction of H_0 to λ.

4.1.A. DEFINITION. We say that H_0 is *steep at a point $I \in G$ on the plane* $\lambda \in \Lambda^r(I)$ if we can find numbers $C > 0$, $\delta > 0$, and $\alpha \geqslant 1$ such that $m_{I,\lambda}(\eta)$ is defined on $[0, \delta)$, and

$$\sup_{0 \leqslant \eta \leqslant \xi} m_{I,\lambda}(\eta) > C\xi^\alpha$$

for all $\xi \in (0, \delta)$. We call C and δ the *coefficients*, and α the *index* of steepness.

4.1.B. DEFINITION. We say that H_0 is *steep at a point I* if the following two conditions hold. Firstly, grad $H_0|_I \geqslant g$, where $g > 0$; secondly, if the number s of variables is greater than 1, then for every $r = 1, \ldots, s - 1$, there are constants $C_r > 0$, $\delta_r > 0$, and $\alpha_r \geqslant 1$ such that H_0 is steep at I on every plane $\lambda \in \Lambda^r(I)$ perpendicular to grad $H_0|_I$ with C_r and δ_r as the steepness coefficients, and α_r the steepness index.

We call the numbers g, $C_1, \ldots, C_{s-1}$ and $\delta_1, \ldots, \delta_{s-1}$ the *coefficients* and $\alpha_1, \ldots, \alpha_{s-1}$ the *indices* of steepness of H_0 at I.

4.1.C. DEFINITION. We say that H_0 is *steep in a domain G with coefficients g, $C_1, \ldots, C_{s-1}$, $\delta_1, \ldots, \delta_{s-1}$ and indices $\alpha_1, \ldots, \alpha_{s-1}$* if H_0 is steep at every point $I \in G$ with these steepness coefficients and indices.

For the formulation of the main theorem we also need the following function ζ of all the steepness indices, except α_{s-1}:

$$(4.1)\quad \zeta = [\alpha_1(\alpha_2 \cdot \ldots (\alpha_{s-3}(\alpha_{s-2} \cdot s + s - 2) + s - 3) + \ldots + 2) + 1] - 1$$

for $s > 2$. If $s = 2$, then we put $\zeta = 1$.

4.2. Notation. Let X be a subset of a metric space. We denote by $X + \varepsilon$ an ε-neighbourhood of X, and by $X - \varepsilon$ the set of points that are contained in X together with their ε-neighbourhoods.

Let (a_{ij}) be a complex square matrix of order s. We denote by $\|(a_{ij})\|$ the norm of the linear operator in the Hermitian space C^s whose matrix has the form (a_{ij}), that is,

$$\|(a_{ij})\| = \max_{|I|=1} \sqrt{\sum_{i=1}^{s} \Big|\sum_{j=1}^{s} a_{ij} I_j\Big|^2}, \quad \text{where} \quad |I| = \sqrt{\sum_{i=1}^{s} |I_i|^2}.$$

4.3. Description of the function H. Canonical equations. We denote by H a function of the form

$$H = H_0(I) + H_1(I, p, \varphi, q)$$

that is 2π-periodic in $\varphi_1, \ldots, \varphi_s = \varphi$, where I is an s-dimensional and p and q are $(n-s)$-dimensional vectors. We assume that $n \geqslant s \geqslant 2$ and that the conditions 4.3.A and 4.3.B hold.

4.3.A. H is defined and analytic on the complex set F:

$$F = \{I, p, \varphi, q\colon \operatorname{Re} I \in G, \ \operatorname{Re} p, q \in D, \ |\operatorname{Im} I, p, \varphi, q| \leqslant \rho\},$$

where G and D are arbitrary domains in $\mathbf{E}^s$ and $\mathbf{E}^{2(n-s)}$, respectively, and $\rho > 0$. (Here and in what follows, $|\operatorname{Im} I, p, \varphi, q| \leqslant \rho$ means

$$|\operatorname{Im} I_i| \leqslant \rho, \ |\operatorname{Im} \varphi_i| \leqslant \rho \quad (i = 1, \ldots, s),$$
$$|\operatorname{Im} p_j| \leqslant \rho, \ |\operatorname{Im} q_j| \leqslant \rho \quad (j = 1, \ldots, n-s).$$

Apart from this norm and the norm of the integer vectors k (introduced below in the proof of the main theorem), all the other norms in the finite-dimensional spaces are the usual Euclidean or Hermitian norms.

For real values of the arguments, H is a real-valued function.

4.3.B. The Hessian of H_0 is uniformly bounded on F, that is, if

$$m = \sup_F \left\| \frac{\partial^2 H_0(I)}{\partial I^2} \right\|, \tag{4.2}$$

then $m < \infty$, where $\dfrac{\partial^2 H_0}{\partial I^2} = \left(\dfrac{\partial^2 H_0}{\partial I_i \partial I_j}\right)$ is the matrix of the second order derivatives of H_0.

4.3.C. By a system of canonical equations with Hamiltonian H we mean a system:

$$\begin{cases} \dot{I} = -H_\varphi, & \dot{p} = -H_q, \\ \dot{\varphi} = H_I, & \dot{q} = H_p. \end{cases}$$

4.4. The main theorem. *Suppose that* $H = H_0(I) + H_1(I, p, \varphi, q)$ *is defined on the set* F, *where* H *and* F *are as in* §4.3.A. *We assume that* H_0 *is steep in* G *with coefficients* g, C_r, *and* δ_r, *and indices* α_r $(r = 1, \ldots, s-1)$. *Then there is a positive constant* $M_0 = M_0(H_0, \rho)$ *not depending on any characteristics of* H_1 *other than* ρ, *and with the following property. Let*

$$M = \sup_F |\operatorname{grad} H_1|, \tag{4.3}$$

and suppose that $0 < M < M_0$. *Let* $I(t)$, $p(t)$, $\varphi(t)$, $q(t)$ *be an arbitrary (real) solution of the system of canonical equations with Hamiltonian* H, *such that*

$$I(0) \in G - d \text{ and } p(t),\ q(t) \in D - d \text{ for all } t \in [0, C],$$

where $C > 0$ *is arbitrary. Then*

$$|I(t) - I(0)| < d/2 \text{ for all } t \in [0, \min [C, T]].$$

Here

$$d = M^b, \qquad b = \frac{3}{(12\zeta + 3s + 14)\alpha_{s-1}}, \tag{4.4}$$

$$T = \frac{1}{M} \exp\left(\frac{1}{M}\right)^a, \quad a = \frac{2}{12\zeta + 3s + 14}, \tag{4.5}$$

and $\zeta = \zeta(\alpha_1, \ldots, \alpha_{s-2})$ *is defined by* (4.1).

4.5. REMARKS. 4.5.A. The assertion of the theorem is completely unchanged if we take M to be $\sup_F |H_1|$ instead of $\sup_F |\operatorname{grad} H_1|$ (see (4.3)).

4.5.B. M_0 depends only on the steepness coefficients and indices $g, C_1, \ldots, C_{s-1}, \delta_1, \ldots, \delta_{s-1}, \alpha_1, \ldots, \alpha_{s-1}$ and on m and ρ.

Part of the assertions forming the proof of the main theorem (Propositions 8.6 and 9.2) is true not only for the systems described above, but also for a wider class of systems including, for example, Hamiltonian systems in a neighbourhood of a singular point when the eigenvalues of the linearized system are purely imaginary (see §2.2.B). We call them *systems with frequencies*; they are defined below.

§5. Forbidden motions

5.1. Forbidden motions of frequency systems. Let $\mathbf{Z}^s$ be the lattice of s-dimensional vectors k with integer components.

5.1.A. DEFINITION. By the *order* of a vector $k = k_1, \ldots, k_s \in \mathbf{Z}^s$ we mean $|k| = \sum_{i=1}^{s} |k_i|$.

5.1.B. DEFINITION. Let $\omega\colon Y \to \mathbf{C}^s$ be a mapping from an arbitrary set Y into the complex vector space. Then $k \in \mathbf{Z}^s$ is called a *ν-resonance vector* for Y relative to ω if we can find a $y \in Y$ such that $|\langle k, \omega(y)\rangle| < \nu$, where $\langle k, \omega\rangle = \sum_{i=1}^{s} k_i \omega_i$ is the scalar product of k and ω, and ν is a number.

Let X be a vector field on a real manifold V and $\mathcal{J}: V \to G$ a mapping, where G is a subset of $\mathbf{E}^s$, which we call the space of action variables. Let ω: $G \to \mathbf{R}^s$ be a mapping, where $\mathbf{R}^s$ is an s-dimensional linear space, which we call the frequency space.

5.1.C. DEFINITION. We call the triple $(X, \mathcal{J}, \omega)$ a *frequency system.*

5.1.D. DEFINITION. We say that $(X, \mathcal{J}, \omega)$ satisfies the condition of *forbidden motions with the parameters N, ν, b,* and T if the following condition holds:

Let U be an arbitrary subset of $G - 2b$. Let λ_0 denote the linear hull of all ν-resonance vectors k for U (relative to ω) for which $|k| \leqslant N$. Let U be such that $\dim \lambda_0 < s$. Let $v(t)$ be an arbitrary solution of the system of differential equations $\dot{v} = X(v)$ defined by a vector field X, such that

$$\mathcal{J}(v(t)) \in U \quad \text{and} \quad v(t) \in V - 2b \quad \text{for all} \quad t \in [0, C],$$

where $c > 0$ is arbitrary. Then

$$\rho(\mathcal{J}(v(t)), \lambda) < b \quad \text{for all} \quad t \in [0, \min[C, T]],$$

where λ is the affine subspace in $\mathbf{E}$ obtained from λ_0 by a translation and containing the point $\mathcal{J}(v(0))$, and $\rho(\mathcal{J}(v(t)), \lambda)$ is the distance from $\mathcal{J}(v(t))$ to λ.

5.2. The condition for forbidden motions holds for a system with Hamiltonian H. Let $H = H_0(I) + H_1(I, p, \varphi, q)$ be analytic on the complex set F, where H and F are as in §4.3. We recall that

$$F = \{\operatorname{Re} I \in G,\ \operatorname{Re} p,\ q \in D, |\operatorname{Im} I,\ p,\ \varphi,\ q| \leqslant \rho\},$$

where G and D are domains (open sets).

We denote by V the real set

$$V = \{I \in G,\ p,\ q \in D,\ \varphi \in T^s\}, \tag{5.1}$$

and by X the vector field on V defined by H. Let $\mathcal{J}$: $V \to G$ be a mapping such that $\mathcal{J}(I, p, \varphi, q) = I$, and ω: $G \to \mathbf{R}^s$ a mapping such that $\omega(I) = \dfrac{\partial H_0}{\partial I}$.

5.2.A. DEFINITION. We say that the so constructed frequency system $(X, \mathcal{J}, \omega)$ is *obtained from the system with the Hamiltonian* $H = H_0 + H_1$.

The following proposition is a consequence of Lemma 10.3 on the elimination of non-resonance harmonics.

5.2.B. PROPOSITION. *We consider the frequency system* $(X, \mathcal{J}, \omega)$

constructed from a system with the Hamiltonian H. Let m and M be as in (4.2) *and* (4.3), *respectively. Let N, ν, b, and T be positive numbers such that we can find a* $\varkappa \in (0, 1)$ *for which* (10.1)–(10.4) *hold* (*these relations are cumbersome, but they are only needed to make sure that the conditions of Lemma* 10.3 *hold, therefore we do not write them out*). *Suppose also that*

$$4mbN \leqslant \nu, \tag{5.2}$$

$$2b < \rho, \tag{5.3}$$

$$T \leqslant \frac{\rho}{16\pi s} b \frac{1}{M} \exp(\rho N^{1-\varkappa}). \tag{5.4}$$

Only N, ν, b, T, ϰ, M, m, ρ, n, and s occur in all these equalities and inequalities. Then $(X, \mathcal{J}, \omega)$ *satisfies the condition for forbidden motions with the parameters N, ν, b, and T.*

5.3. PROOF OF PROPOSITION 5.2.B. Let $U \in G - 2b$, and let ΔF be the closed set

$$\Delta F = \{I, p, \varphi, q\colon \operatorname{Re} I \in \overline{U+b},\ \operatorname{Re} p, q \in D - b,\ |\operatorname{Im} I, p, q| \leqslant b,\ |\operatorname{Im} \varphi| \leqslant \rho\}.$$

(The bar denotes closure.) It follows from (5.3) that $\Delta F \subset F$.

By A_I we denote the set

$$A_I = \{I : \operatorname{Re} I \in \overline{U+b},\ |\operatorname{Im} I| \leqslant b\}.$$

Let U be such that $\dim \lambda_0 < s$, where λ_0 is defined in §5.1.D.

5.3.A. ASSERTION. *Every* $(\nu/2)$*-resonance vector* $k \in \mathbf{Z}^s$ *for* A_I *such that* $|k| \leqslant N$ *belongs to* λ_0.

In fact, for such a k we can find an $I \in A_I$ such that $|\langle k, \omega(I)\rangle| < \nu/2$. It follows from the definition of A_I that we can find an $I' \in U$ such that the segment joining I and I' lies entirely in A_I and its length is less than $2b$. For such a point I' we obtain

$$|\langle k, \omega(I')\rangle| \leqslant |\langle k, \omega(I)\rangle| + m|k|\,|I - I'| < \frac{\nu}{2} + 2mbN.$$

From this and (5.2) we find that $|\langle k, \omega(I')\rangle| < \nu$. It follows from the definition of λ_0 that $k \in \lambda_0$, and so the assertion is proved.

Now we apply Lemma 10.3 (about the elimination of non-resonance harmonics), with λ_0 in place of λ, to H restricted to ΔF. By this lemma and (5.3) we can make on the real set

$$\{I, p, \varphi, q : I \in \bar{U}, p, q \in D - 2b, \varphi \in T^s\} \tag{5.5}$$

an analytic canonical change of variables $I, p, \varphi, q \to J, P, \psi, Q$ such that on this set

$$|J(I, p, \varphi, q) - I| < \frac{b}{15}. \tag{5.6}$$

In the variables J, P, ψ, Q the Hamiltonian H has the form $H = \bar{H} + R$, where the expansion of $\bar{H}$ contains only those harmonics whose number k belongs to $\lambda_0 \cap \mathbf{Z}^s$:

$$\bar{H} = \sum_{k \in \lambda_0 \cap \mathbf{Z}^s} \bar{h}_k (J, P, Q)\, e^{ik\psi}, \tag{5.7}$$

and

$$|\operatorname{grad} R| < 8\pi s \frac{1}{\rho} M \exp(-\rho N^{1-\varkappa}). \tag{5.8}$$

Now we need the following lemma.

5.3.B. LEMMA (ON THE CONTRIBUTION OF A HARMONIC). *Let $\mathscr{H} = \Sigma h_k(J, P, Q)e^{ik\psi}$ be an arbitrary Hamiltonian, and let $J(t)$, $P(t)$, $\psi(t)$, $Q(t)$ be an arbitrary solution of the system with this Hamiltonian. Then the direction in which the vector J defined by the harmonic $h_k(J, P, Q)e^{ik\psi}$ is increasing is the direction of k.*

The lemma is proved by calculations based on Hamilton's equations[1]

$$\dot{J}_j = -\frac{\partial}{\partial \psi_j}(h_k(J, P, Q)\, e^{ik\psi} + \ldots) = \\ = k_j(-ih_k(J, P, Q)\, e^{ik\psi}) + \ldots \qquad (j = 1, \ldots, s).$$

5.3.C. COMPLETION OF THE PROOF OF PROPOSITION 5.2.B. Let $v(t) = I(t), p(t), \varphi(t), q(t)$ be any solution of the system with Hamiltonian H such that $I(t) \in U$ and $v(t) \in V - 2b$ for all $t \in [0, C]$, where $C > 0$ and the set V is defined by (5.1). It is not difficult to see that the trajectory of this solution for $t \in [0, C]$ belongs to the set (5.5). We denote by $J(t)$, $P(t)$, $\psi(t)$, $Q(t)$ the same solution in the coordinates J, P, ψ, Q.

It follows from (5.7), (5.8), and Lemma 5.3.B that the length of the component of $\dot{J}(t)$ perpendicular to λ_0 is less than w for all $t \in [0, C]$, where w denotes the right-hand side of (5.8). Let λ' denote the plane in $\mathbf{E}^s$ that is obtained from λ_0 by a translation and contains $J(0)$. Then the distance $\rho(J(t), \lambda')$ from $J(t)$ to λ' is less than tw for $t \in (0, C]$. It follows from (5.4) that $\rho(J(t), \lambda') < b/2$ for $t \in [0, \min[C, T]]$. By using (5.6) and the fact that $\lambda' \| \lambda$ we obtain

$$\rho(I(t), \lambda) \leqslant \rho(I(t), J(t)) + \rho(J(t), \lambda') + \rho(\lambda', \lambda) < \frac{b}{15} + \frac{b}{2} + \frac{b}{15} < b$$

for all these t, as required.

§6. Resonances. Resonance zones and blocks

Thus, the behaviour of $I(t)$ depends strongly on what and how many vectors k, $|k| \leqslant N$, are resonance vectors for some neighbourhood of this point. In this context the following concepts are useful. Let $N > 0$ be an arbitrary number.

[1] This simple calculation is the only place in the proof of the main theorem where the Hamiltonian property of the system is used in an essential way.

6.1. DEFINITION. Let $k^1, \ldots, k^r$ be arbitrary linearly independent vectors of $\mathbf{Z}^s$ of order not greater than N, $|k^i| \leqslant N$. We denote by Z_N the set of all $k \in \mathbf{Z}^s$ belonging to the linear hull of $k^1, \ldots, k^r$ and such that $|k| \leqslant N$. We call Z_N a *resonance of multiplicity* r *and order* N. We call the k^i $(i = 1, \ldots, r)$ the *defining* vectors of Z_N. The set $\{0\}$ consisting of the zero vector only is called an *0-fold resonance*; we take its defining vector to be $k = 0$.

Note that for every $N \geqslant 1$ there is only one s-fold resonance. (For $N = 0$ there is only one resonance, the 0-fold one.)

We consider the s-dimensional vector space $\mathbf{R}^s$ (the frequency space); let $w \in \mathbf{R}^s$.

6.2. DEFINITION. Let Z_N be an arbitrary r-fold resonance, where $r \neq 0$, and let $\nu > 0$ be arbitrary. We denote by $O(Z_N, \nu)$ the set:

$$O(Z_N, \nu) = \{w \in \mathbf{R}^s: \text{we can find } r \text{ linearly independent vectors } k^i \in Z_N \text{ such that } |\langle k^i, w\rangle| < \nu \ (i = 1, \ldots, r)\}$$

which we call a *zone of influence* of Z_N of *width* ν or simply a Z_N-resonance zone. By the *multiplicity* of $O(Z_N, \nu)$ we mean the multiplicity r of Z_N. For every $\nu > 0$ the *zone of multiplicity* 0 is the whole of $\mathbf{R}^s$.

We fix $N > 0$ and denote by $\mathfrak{Z}_N^r$ the set of all r-fold resonances, and by $\mathfrak{Z}_N$ the set of all resonances:

$$\mathfrak{Z}_N = \bigcup_{r=0}^{s} \mathfrak{Z}_N^r.$$

Let N and $\nu_1, \ldots, \nu_s$ be arbitrary positive numbers. Let O_r denote the union of all r-fold resonance zones of width ν_r:

$$O_r = \bigcup_{Z_N \in \mathfrak{Z}_N^r} O(Z_N, \nu_r). \tag{6.1}$$

We consider the mapping from $\mathfrak{Z}_N$ into a subset of $\mathbf{R}^s$ that associates with each resonance $Z_N \in \mathfrak{Z}_N$ the subset $B = B(Z_N)$:

$$\begin{cases} B = O(Z_N, \nu_r) & \text{if} \quad r = s, \\ B = O(Z_N, \nu_r) \setminus O_{r+1} & \text{if} \quad r = 1, 2, \ldots, s-1, \\ B = \mathbb{R}^s \setminus O_1 & \text{if} \quad r = 0, \end{cases} \tag{6.2}$$

where r is the multiplicity of Z_N.

6.3. DEFINITION. We call this mapping a *partition of* $\mathbf{R}^s$ *into blocks with the parameters* $N, \nu_1, \ldots, \nu_s$. The set $B(Z_N)$ is called the *block* of this partition corresponding to Z_N.

(Roughly speaking, a block $B(Z_N)$ is the difference of the zone of influence of Z_N and the union of the zones of influence of resonances not contained in Z_N.)

6.4. PROPOSITION. *For any positive N, $\nu_1, \ldots, \nu_s$, the union of all blocks of multiplicity less than r in the partition with the parameters N, $\nu_1, \ldots, \nu_s$ is $\mathbf{R}^s \setminus \bigcap_{i=1}^{r} O_i$. The union of blocks of all multiplicities is $\mathbf{R}^s$.*

PROOF. It follows from (6.2) and (6.1) that the set

$$\bigcup_{Z_N \in \mathfrak{Z}_N^r} B(Z_N)$$

coincides with O_r for $r = s$, and with $O_r \setminus O_{r+1}$ for $r = 1, \ldots, s-1$. Therefore,

$$\bigcup_{i=0}^{r-1} \bigcup_{Z_N \in \mathfrak{Z}_N^i} B(Z_N) = \left[\bigcup_{i=1}^{r-1} (O_i \setminus O_{i+1})\right] \cup (\mathbf{R}^s \setminus O_1) = \mathbf{R}^s \setminus \bigcap_{i=1}^{r} O_i.$$

This proves Proposition 6.4.

6.5. Zones of influence, and the partitioning into blocks on an arbitrary set. Let ω: $G \to \mathbf{R}^s$ be a mapping of an arbitrary set G into the frequency space $\mathbf{R}^s$. We fix G and ω.

Let Z_N be an arbitrary resonance, and let N, ν, and $\nu_1, \ldots, \nu_s$ be arbitrary positive numbers. Suppose that $B(Z_N)$ is a block of the partition of $\mathbf{R}^s$ with the parameters N, $\nu_1, \ldots, \nu_s$.

6.5.A. DEFINITION. We call the inverse image $\omega^{-1}(O(Z_N, \nu))$ of $O(Z_N, \nu)$ a *zone of influence of width ν* of Z_N on the set G (relative to ω). The mapping that associates with $Z_N \in \mathfrak{Z}_N$ the subset $\omega^{-1}(B(Z_N))$ of G is called a *partition of G* (relative to ω) *into blocks with the parameters* N, $\nu_1, \ldots, \nu_s$, and $\omega^{-1}(B(Z_N))$ is called the *block* in G corresponding to Z_N. By the *multiplicity* of this block we mean that of Z_N.

In what follows, for brevity we denote the sets $\omega^{-1}(O(Z_N, \nu))$ and $\omega^{-1}(B(Z_N)$ by $O(Z_N, \nu)$ and $B(Z_N)$, respectively; this should not cause confusion since we shall not use the corresponding sets in the frequency space any more.

The next assertion follows from Proposition 6.4.

6.5.B. COROLLARY. *We consider a partition of a set G into blocks, with arbitrary positive parameters N, $\nu_1, \ldots, \nu_s$. Then the union of all blocks of multiplicity less than r is $G \setminus \bigcap_{i=1}^{r} O_i$, where O_i is the union of all i-fold resonance zones on G of width ν_i. The union of all blocks coincides with G.*

§7. Dependence of the diameters of discs of fast drift on the steepness of the unperturbed Hamiltonian

Let Z_N be an arbitrary resonance, and let I be a point in $\mathbf{E}^s$. We denote by $\lambda(Z_N, I)$ the affine subspace of $\mathbf{E}^s$ that is obtained by a translation from the linear hull of the vectors $k^i \in Z_N$, and contains I.

7.1. DEFINITION. We call $\lambda(Z_N, I)$ the *plane of fast drift* for Z_N that passes through I.

Note that $\dim \lambda(Z_N, I) = r(Z_N)$, the multiplicity of Z_N.

Let ω: $G \to \mathbf{R}^s$ be a mapping of some domain $G \subset \mathbf{E}^s$ into $\mathbf{R}^s$.

Let $Z_N \neq 0$ be an arbitrary resonance, ν and b arbitrary positive numbers, and $I \in G$. We consider the intersection $O(Z_N, \nu) \cap (\lambda(Z_N, I) + b)$ of the resonance zone $O(Z_N, \nu)$ on G (relative to ω) defined in §6.5.A, and a b-neighbourhood of the plane $\lambda(Z_N, I)$.

7.2. DEFINITION. We denote by $D = D(Z_N, \nu, b, I)$ the connected component of this intersection that contains I and call it a *disc of fast drift* on G relative to ω, or simply a *disc*, for the resonance Z_N and the point I. We call ν the *width* of the resonance of the disc and b the *thickness* of the disc.

We call that part of the boundary of the disc contained in the boundary of $O(Z_N, \nu)$ the *lateral surface* of the disc, and the part contained in the boundary of the b-neighbourhood of $\lambda(Z_N, I)$ the *base* of the disc.

Let H_0 be an arbitrary twice differentiable function defined in some domain $G \subset \mathbf{E}^s$, and let $m > 0$ be arbitrary.

7.3. DEFINITION. We say that H_0 is *generalized steep in G with coefficients g, m, $C_1, \ldots, C_{s-1}$, $\delta_1, \ldots, \delta_{s-1}$, and indices $\alpha_1, \ldots, \alpha_{s-1}$* if the following two conditions hold:

A. H_0 is steep in G with steepness coefficients g, $C_1, \ldots, C_{s-1}$, $\delta_1, \ldots, \delta_{s-1}$, and steepness indices $\alpha_1, \ldots, \alpha_{s-1}$ (see Definition 4.1.C).

B. In G

$$\left\| \frac{\partial^2 H_0}{\partial I^2} \right\| \leqslant m$$

(see (4.2)).

7.4. PROPOSITION. *Let H_0 be generalized steep in some domain $G \subset \mathbf{E}^s$ with coefficients g, m, $C_1, \ldots, C_{s-1}$, $\delta_1, \ldots, \delta_{s-1}$ and indices $\alpha_1, \ldots, \alpha_{s-1}$. Let ω: $G \to \mathbf{R}^s$ be the mapping defined by* $\omega(I) = \operatorname{grad} H_0|_I$.

Let Z_N be an arbitrary r-fold resonance, where $r = 1, \ldots, s-1$. Let D be a disc on G of thickness b and resonance width ν (relative to ω) for Z_N and any $I \in G - 2d$, where N, b, ν, and d satisfy the conditions

$$(7.1) \qquad d \leqslant \min [g/3m, \ \delta_r],$$

$$(7.2) \qquad \varepsilon < g/4,$$

$$(7.3) \qquad b < \min [d/4, \ \varepsilon/4m],$$

$$(7.4) \qquad rN^{r-1}\nu \leqslant \varepsilon;$$

where ε *is the quantity*

$$\varepsilon = \frac{C_r}{5}\left(\frac{d}{2}\right)^{\alpha_r}. \tag{7.5}$$

Then the diameter of D *is not greater than* d.

The proof relies on the technical Lemmas 10.6 and 10.5; it is by contradiction.

Suppose that the diameter of D is greater than d. Since D is an open linearly connected set, there is a curve γ lying in D and having the properties 10.6.A–10.6.C of Lemma 10.6 with $\lambda = \lambda(Z_N, I)$. The validity of 10.6.D in Lemma 10.6 follows from (7.1)–(7.3). From this lemma we obtain that there is a point $\tilde{I} \in \gamma$ such that

$$|\operatorname{grad}(H_0|_{\tilde{\lambda}})|_{\tilde{I}} > \varepsilon, \tag{7.6}$$

where $\tilde{\lambda} = \lambda(Z_N, \tilde{I})$ is the plane of fast drift for Z_N through $\tilde{I}$.

On the other hand, it follows from the definition of $O(Z_N, \nu)$ that for any $I \in O(Z_N, \nu)$ we can find r linearly independent vectors $k^i \in Z_N$ ($i = 1, \dots, r$) such that $|\langle k^i, \omega(I)\rangle| < \nu$. By Lemma 10.5, $|\operatorname{pr}_{\lambda_0} \omega(I)| < rN^{r-1}\nu$ for all $I \in O(Z_N, \nu)$, where λ_0 is the linear hull of the k^i ($i = 1, \dots, r$), and $\operatorname{pr}_{\lambda_0} \omega(I)$ is the orthogonal projection of $\omega(I)$ on λ_0. By taking into account (7.4) and the fact that $\gamma \subset D \subset O(Z_N, \nu)$, we obtain

$$|\operatorname{pr}_{\lambda_0} \omega(I)| < \varepsilon \tag{7.7}$$

for all $I \in \gamma$. But $\operatorname{grad}(H_0|_{\tilde{\lambda}})|_{\tilde{I}} \equiv \operatorname{pr}_{\lambda_0} \omega(\tilde{I})$, hence (7.6) and (7.7) contradict one another. This proves Proposition 7.4.

§8. Condition for the non-overlapping of resonances

Let I be an arbitrary point of some domain G and b a positive number.

8.1. DEFINITION. The intersection of G with the ball of radius b and centre at I is called the *disc* on G *for the* 0-*fold resonance* $Z_N = \{0\}$ and I; b is called the *thickness* of the disc. By the *base* of this disc we mean the whole of its boundary, and we regard the *lateral surface* as being the empty set (see Definition 7.2).

8.2. The rigging of blocks by discs of fast drift. Let ω: $G \to \mathbf{R}^s$ be a fixed mapping. We consider a partition of G (relative to ω) into blocks with fixed parameters $N, \nu_1, \dots, \nu_s$ (see Definition 6.5.A). As in §6, $B(Z_N)$ denotes the block of this partition that corresponds to the resonance Z_N.

Let Z_N be arbitrary, I any point of $B(Z_N)$, and b a positive number. We denote by r the multiplicity of Z_N, and by $D = D(Z_N, I, b)$ the disc for Z_N and I of thickness b and (for $r \neq 0$) resonance width ν_r (see Definitions 7.2 and 8.1).

DEFINITION. We call $D = D(Z_N, I, b)$ a *disc of thickness b of the rigging of the block $B(Z_N)$ at I.*

We consider the partition of G relative to ω with the parameters $N, \nu_1, \ldots, \nu_s$ such that $\nu_{s-1} = \nu_s$. Let $b > 0$ and $\nu_0 > 0$ be arbitrary numbers.

8.3. DEFINITION. We say that ω satisfies the condition for the *non-overlapping of resonances with parameters $N, \nu_0, \nu_1, \ldots, \nu_{s-1}, b$* if the following conditions hold:

A. The s-fold block is empty.

B. Let Z_N be an arbitrary resonance of multiplicity r $(r = 0, 1, \ldots, s-1)$. Let D be any disc of thickness b of the rigging of $B(Z_N)$ such that $\bar{D}$ is contained in G. Then $\bar{D}$ does not intersect $O(Z_N', \nu_r)$ for any $Z_N' \not\subseteq Z_N$:

$$\bar{D} \cap O(Z_N', \nu_r) = \varnothing \quad \text{for all} \quad Z_N' \not\subseteq Z_N.$$

8.4. PROPOSITION. *Suppose that a continuous mapping ω satisfies the condition for the non-overlapping of resonances with parameters $N, \nu_0, \nu_1, \ldots, \nu_{s-1}, b$. We consider a partition of G relative to ω with the parameters $N, \nu_1, \ldots, \nu_{s-1}, \nu_s = \nu_{s-1}$. Then the following assertions hold.*

A. *Blocks of the same multiplicity in this partition are disjoint.*

B. *Let D be an arbitrary disc of thickness b of the rigging of an arbitrary block $B(Z_N)$ of multiplicity $r = 1, \ldots, s-1$ such that $\bar{D} \subset G$. Then the lateral surface of D is contained in the union of blocks of multiplicity less than r.*

C. *Let D be an arbitrary disc of thickness b of the rigging of an arbitrary block $B(Z_N)$, and let $\bar{D} \subset G$. Then the set of all ν_r-resonance vectors $k \in \mathbf{Z}^s$ for $\bar{D}$ whose order is not greater than N (see Definitions* 5.1.B *and* 5.1.A) *is contained in Z_N, where r is the multiplicity of $Z_N (r = 0, 1, \ldots, s-1)$.*

A follows from condition 8.3.B and Definitions 6.5.A and 6.1.

Let us prove B. Since a resonance zone is an open set, the boundary $\partial O(Z_N, \nu_r)$ of the zone $O(Z_N, \nu_r)$ does not intersect it. Hence and from condition 8.3.B, the lateral surface of D can only intersect $O(Z_N', \nu_r)$ when $Z_N' \subset Z_N$. Consequently, this surface is contained in $G \setminus O_r$, where O_r is defined in Corollary 6.5.B. By this corollary, the union of blocks of multiplicity less than r contains $G \setminus O_r$, and this proves B.

Proof of C. Suppose that we can find an $I \in \bar{D}$ such that $|\langle k, \omega(I)\rangle| < \nu_r$ for some $k \in \mathbf{Z}^s$, $|k| \leqslant N$. If $k = 0$, then $k \in Z_N$ always. Suppose that $k \neq 0$. Then $I \in O(Z_N', \nu_r)$, where Z_N' is a 1-fold resonance defined by k (see Definition 6.1). From condition 8.3.B we obtain $Z_N' \subseteq Z_N$. Hence, in this case, $k \in Z_N$, as required.

Proposition 8.4 is now completely proved.

8.5. DEFINITION. We say that the mapping ω: $G \to \mathbf{R}^s$ satisfies *the* *extended condition for the non-overlapping of resonances with the parameters* $N, \nu_0, \nu_1, \ldots, \nu_{s-1}, b, d_0, d_1, \ldots, d_{s-1}$, if the following two conditions hold.

We denote by G' the interior of the set $G - 2(d_0 + d_1 + \ldots + d_{s-1})$ and by ω' the restriction of ω to G'. Then:

A. ω' satisfies the condition for the non-overlapping of resonances with the parameters $N, \nu_0, \nu_1, \ldots, \nu_{s-1}, b$.

B. We consider a partition of G' into blocks relative to ω' with the parameters $N, \nu_1, \ldots, \nu_{s-1}, \nu_s = \nu_{s-1}$. For any block $B(Z_N)$ of this partition, the diameter of every disc of thickness b of the rigging of $B(Z_N)$ is not greater than d_r, where r is the multiplicity of $B(Z_N)$ $(r = 0, 1, \ldots, s-1)$.

8.6. PROPOSITION. *Let H_0 be generalized steep in G with coefficients $g, m, C_1, \ldots, C_{s-1}, \delta_1, \ldots, \delta_{s-1}$, and indices $\alpha_1, \ldots, \alpha_{s-1}$ (see Definition 7.3). We denote by ω: $G \to \mathbf{R}^s$ the mapping defined by $\omega(I) = \operatorname{grad} H_0|_I$. Then ω satisfies the extended condition for the non-overlapping of resonances with the parameters $N, \nu_0, \nu_1, \ldots, \nu_{s-1}, b, d_0, d_1, \ldots, d_{s-1}$, provided that for all $r = 1, \ldots, s-1$,*

$$d_r < \min\left[\frac{g}{3m}, \delta_r\right], \tag{8.1}$$

$$\varepsilon_r < \frac{g}{2}, \tag{8.2}$$

$$b < \min\left[\frac{d_r}{4}, \frac{\varepsilon_r}{4m}\right], \tag{8.3}$$

$$b < d_0, \tag{8.4}$$

$$\nu_{s-1} \leqslant \frac{g}{2sN^{s-1}}, \tag{8.5}$$

$$\nu_r \leqslant \frac{\varepsilon_r}{rN^{r-1}}, \tag{8.6}$$

$$d_{r-1} \leqslant \frac{1}{2mN}\nu_r, \tag{8.7}$$

$$\nu_{r-1} \leqslant \frac{\nu_r}{2}, \tag{8.8}$$

where

$$\varepsilon_r = \frac{C_r}{5}\left(\frac{d_r}{2}\right)^{\alpha_r}. \tag{8.9}$$

8.7. PROOF OF PROPOSITION 8.6. To prove this it is sufficient to show that ω' satisfies conditions 8.5.B, 8.3.A, and 8.3.B. We do this below.

8.7.A. **Estimate of diameters of discs.** For $r = 1, \ldots, s-1$, condition 8.5.B follows from Proposition 7.4 and the inequalities (8.1)–(8.3) and (8.6). For $r = 0$, this condition follows from (8.4) and Definition 8.1.

In the following lemma and its two corollaries, zones and blocks refer to the whole set G.

8.7.B. LEMMA. *Under the conditions of Proposition 8.6, the s-fold zone*

$O(Z_N, 2\nu_{s-1})$ *of resonance width* $2\nu_{s-1}$ *is empty.*

This lemma follows from the technical Lemma 10.5 and (8.5). We prove Lemma 8.7.B by contradiction. Suppose that the zone in question is not empty; then we can find an $I \in G$ and s linearly independent vectors $k^i \in \mathbf{Z}^s$, $|k^i| \leqslant N$, such that $|\langle k^i, \omega(I)\rangle| < 2\nu_{s-1} (i = 1, \ldots, s)$. It follows from Lemma 10.5 that $|\omega(I)| < 2sN^{s-1}\nu_{s-1}$. From (8.5) we obtain $|\omega(I)| < g$. But this contradicts the condition in Proposition 8.6 that $|\omega(I)| \geqslant g$ in G (see Definition 7.3), and so the lemma is proved.

8.7.C. COROLLARY. *Under the same conditions, an s-fold block is empty.*

For it follows from Definition 6.5.A, firstly, that $O(Z_N, \nu) \subseteq O(Z_N, \nu')$ for every Z_N and $0 < \nu < \nu'$, and, secondly, that $B(Z_N) \subseteq O(Z_N, \nu_r)$, where $N, \nu_1, \ldots, \nu_s$ are the parameters of the partition, and r is the multiplicity of Z_N. Hence, in our case,

$$B(Z_N) \subseteq O(Z_N, \nu_{s-1}) \subseteq O(Z_N, 2\nu_{s-1}) = \varnothing,$$

where Z_N is an s-fold resonance.

8.7.D. COROLLARY. *Under the same conditions, the mapping* ω *satisfies condition* 8.3.B *when* $r = s - 1$.

PROOF. Let Z_N be an $(s - 1)$-fold resonance, and Z'_N a resonance such that $Z'_N \not\subseteq Z_N$. Then if follows from Definition 6.5.A that

$$\overline{O(Z_N, \nu_{s-1})} \cap O(Z'_N, \nu_{s-1}) \subseteq O(Z''_N, 2\nu_{s-1}),$$

where Z''_N is an s-fold resonance. It follows from Lemma 8.7.B that the intersection on the left-hand side is empty, and from Definition 7.2 that $\bar{D} \subseteq \overline{O(Z_N, \nu_{s-1})}$, where $\bar{D}$ is the closure of the disc of the rigging of the block $B(Z_N)$. Hence, $\bar{D} \cap O(Z'_N, \nu_{s-1}) = \varnothing$, as required.

It follows from Corollary 8.7.C that ω' satisfies condition 8.3.A, and from Corollary 8.7.D that it satisfies condition 8.3.B when $r = s - 1$.

8.7.E. PROOF THAT CONDITION 8.3.B HOLDS FOR $r = 0, 1, \ldots, s-2$. Let Z_N be an arbitrary resonance of multiplicity r, and D a disc of thickness b of the rigging of the block $B(Z_N)$ of the partition of G'. Let I be any point of $\bar{D}$, and k any vector from $\mathbf{Z}^s$ such that $|k| \leqslant N$ and $k \notin Z_N$. By §8.7.A, we can find an $I_0 \in B(Z_N)$ such that $|I - I_0| \leqslant d_r$, where r is the multiplicity of $B(Z_N)$. It follows from Definitions 6.5.A and 6.3 that $|\langle k, \omega(I_0)\rangle| \geqslant \nu_{r+1}$.

The segment joining I and I_0 in G' is contained in G. By using condition 7.3.B and (8.7) we obtain $|\langle k, \omega(I)\rangle| \geqslant |\langle k, \omega(I_0)\rangle| - m\,|k|\,|I_0 - I| \geqslant \nu_{r+1} - mNd_r \geqslant \nu_{r+1}/2$.

By (8.8), $|\langle k, \omega(I)\rangle| \geqslant \nu_r$ for every $I \in \bar{D}$ and $k \notin Z_N$, $|k| \leqslant N$. Hence and from Definitions 6.5.A and 6.2, if $\bar{D} \cap O(Z'_N, \nu_r) \neq \varnothing$, then we can find a system of vectors defining Z'_N (see Definition 6.1) each of which belongs to Z_N. Hence $Z'_N \subseteq Z_N$, as required.

Proposition 8.6. is now completely proved.

§9. Traps in frequency systems. Completion of the proof of the main theorem

We consider an arbitrary frequency system $(X, \mathcal{J}, \omega)$ (see Definition 5.1.C). Let d and T be positive numbers.

9.1. DEFINITION. We say that $(X, \mathcal{J}, \omega)$ is *d-stable during time T* if the following condition holds:

For every solution $v(t)$ of the system such that

$$\mathcal{J}(v(0)) \in G - d \text{ and } v(t) \in V - d/2 \text{ for all } t \in [0, C],$$

where $C > 0$ is any number, we have

$$|\mathcal{J}(v(t)) - \mathcal{J}(v(0))| < \frac{d}{2} \text{ for all } t \in [0, \min [C, T]].$$

Here V is the phase space of the system, and G is the domain of the mapping ω: $G \to \mathbf{R}^s$.

9.2. PROPOSITION. *Suppose that the system* $(X, \mathcal{J}, \omega)$ *has the following properties*:

A. *It satisfies the condition for forbidden motions with the parameters* N, ν, b, *and* T (*see Definition* 5.1.D).

B. G *is an open set, and* ω: $G \to \mathbf{R}^s$ *is continuous and satisfies the extended condition for the non-overlapping of resonances with the parameters* N, $\nu_0, \nu_1, \ldots, \nu_{s-1}$, b, $d_0, \ldots, d_{s-1}$.

C. *These parameters satisfy the inequalities*

$$N \leqslant N_1, \tag{9.1}$$

$$\nu \leqslant \min [\nu_i,\ i = 0, 1, \ldots, s-1], \tag{9.2}$$

$$b \leqslant b_1. \tag{9.3}$$

Then $(X, \mathcal{J}, \omega)$ *is d-stable during the time T, where d is any number satisfying the inequalities*

$$d > 4b, \tag{9.4}$$

$$d \geqslant 4(d_0 + d_1 + \ldots + d_{s-1}). \tag{9.5}$$

Everywhere in the proof of Proposition 9.2, by blocks and discs we mean blocks of a partition of G and discs on G.

For the proof of Proposition 9.2 we need the concept of a trap for the point I; this is a special form of a neighbourhood of I.

9.3. Traps. We fix a mapping ω and consider a partition of G (relative to ω) into blocks, with parameters N, $\nu_1, \ldots, \nu_s$.

Let I be any point of G, and B_1 any of the blocks containing I whose multiplicity is the lowest of those of the blocks containing I (in general, there can be several such blocks). Let $b > 0$ be any fixed number.

9.3.A. DEFINITION. A disc D_1 of thickness b of the rigging of a block B_1 at I is called a *disc of the first rank* for I. A disc D_2 of the same thickness that is a disc of the first rank for some point of the lateral surface of D_1 is called a disc of the *second rank* for I. Similarly, a disc of thickness b of the first rank that lies on the lateral surface of a disc of the $(j-1)$th rank for I is called a disc of the *jth rank* for I.

9.3.B. DEFINITION. We call the union of all discs of all ranks for $I \in G$ a *trap in G for I with the parameters* $N, \nu_1, \ldots, \nu_s, b$, and denote it by $\mathscr{L}(I)$.

9.3.C. LEMMA. *Suppose that a continuous mapping* ω: $G \to \mathbf{R}^s$ *satisfies the extended condition for the non-overlapping of resonances with the parameters* $N, \nu_0, \nu_1, \ldots, \nu_{s-1}, b, d_0, \ldots, d_{s-1}$. *We denote* $d_0 + d_1 + \ldots + d_{s-1}$ *by* d'. *Then, for every* $I \in G - 4d'$, *the diameter of the trap* $\mathscr{L}(I)$ *in G with the parameters* $N, \nu_1, \ldots, \nu_{s-1}, \nu_s = \nu_{s-1}, b$ *is not greater than* $2d' - d_0$.

PROOF. By 8.5.A and 8.4.A, the number of discs of the first rank for any point in the interior of $G - 2d'$ is 1. Let D_1 be a disc of the first rank for $I \in G - 4d'$. Then there are two possibilities: $r = 0$ and $r = 1$, where r is the multiplicity of D_1. In the first case $\mathscr{L}(I) = D_1$ since, by definition, the disc of zero multiplicity does not have a lateral surface. Hence and from 8.5.B,

$$\operatorname{diam} \mathscr{L}(I) = \operatorname{diam} D_1 \leqslant d_0 \leqslant 2d' - d_0.$$

Let $r > 0$. It follows from 8.5.A and 8.4.B that the multiplicity of every disc of the second rank for I is less than r. We prove the following assertion by induction on $j = 1, 2, \ldots, r + 1$, using the same condition and proposition, and the fact that an 0-fold disc does not have a lateral surface. The multiplicity of every disc of rank j for I is not greater than $r - j + 1$, and there are no discs of rank greater than $r + 1$. By using condition 8.5.B, we obtain

$$\operatorname{diam} \mathscr{L}(I) \leqslant d_r + 2d_{r-1} + 2d_{r-2} + \ldots + 2d_0 = 2d' - d_0.$$

This proves Lemma 9.3.C.

9.4. PROOF OF PROPOSITION 9.2. Let $\mathscr{L}(I)$ be a trap for $I \in G - d$ with the parameters $N_1, \nu_1, \ldots, \nu_{s-1}, \nu_s = \nu_{s-1}, b_1$. Then, by Lemma 9.3.C and (9.5),

$$\operatorname{diam} \mathscr{L}(I) \leqslant \frac{d}{2} - d_0. \tag{9.6}$$

Hence, to prove Proposition 9.2 it is sufficient to establish the following result.

9.4.A. ASSERTION. *For every* $I \in G - d$, *and for every solution* $v(t)$

such that $\mathcal{J}(v(0)) = I$ *and* $v(t) \in V - d/2$ *for all* $t \in [0, C]$, *where* $C > 0$, *we have*

$$\mathcal{J}(v(t)) \in \mathcal{L}(I) \quad \text{for all} \quad t \in [0, \min [C, T]].$$

It follows from (9.6) that $\mathcal{L}(I) \subset G - \frac{d}{2} - d_0$; therefore, all the discs forming the trap $\mathcal{L}(I)$ also lie in $G - \frac{d}{2} - d_0$. From the definition of a trap we see that to prove 9.4.A it is sufficient to establish the following result.

9.4.B. ASSERTION. *Let D be an arbitrary disc of the rigging of some block B at some point I_0. Suppose that $\bar{D}$ is contained in the interior of the set $G - d/2$. We consider any solution $v(t)$ such that*

$$\mathcal{J}(v(0)) = I_0 \ \text{and} \ \mathcal{J}(v(t)) \in \bar{D}, \quad v(t) \in V - \frac{d}{2} \quad \text{for all} \quad t \in [0, C_1],$$

where $C_1 > 0$. Then either

$$\mathcal{J}(v(t)) \in D \quad \text{for all} \quad t \in [0, \min [C_1, T]],$$

that is, for these t the trajectory $\mathcal{J}(v(t))$ is contained in the interior of D and does not touch its boundary, or we can find a $\tilde{t} \in (0, \min [C_1, T])$ such that $\mathcal{J}(v(\tilde{t}))$ belongs to the lateral surface of D.

Let us prove 9.4.B. Let Z_{N_1} be a resonance corresponding to B. It follows from condition 8.5.A, Proposition 8.4.C, and (9.5) that all the ν_r-resonance vectors k, $|k| \leqslant N_1$, for $\bar{D}$ belong to Z_{N_1}, where r is the multiplicity of Z_{N_1}. It follows from (9.1) and (9.2) that all the ν-resonance vectors k, $|k| \leqslant N$, for $\bar{D}$ belong to Z_{N_1}.

Now we use the fact that our system satisfies the condition for forbidden motions with the parameters N, ν, b and T (see Definition 5.1.D). For U we take $\bar{D}$. Then it follows from this definition and from (9.3) and (9.4) that $\mathcal{J}(v(t))$ during the time $\min [C_1, T]$ cannot leave D across its base (see Definitions 7.2 and 8.1). Hence, it can only leave across the lateral surface, as required. This proves Proposition 9.2.

9.5. Proof of the main theorem (Theorem 4.4). We consider the frequency system $(X, \mathcal{J}, \omega)$ obtained from a system with the Hamiltonian H (see Definition 5.2.A). By Propositions 9.2, 5.2.B, and 8.6, this system is d-stable during the time T, provided that d and T satisfy the following condition:

9.5.A. CONDITION. We can find positive numbers

$$(9.7) \qquad \varkappa, N, \nu, b, N_1, \nu_0, \ldots, \nu_{s-1}, b_1, d_0, \ldots, d_{s-1}, \varepsilon_1, \ldots, \varepsilon_{s-1},$$

where

$$(9.8) \qquad \varkappa < 1,$$

which, together with d and T, satisfy (9.1)–(9.5), (10.1)–(10.4), (5.2)–(5.4), and (8.1)–(8.9), where N and b in (8.1)–(8.9) are replaced by N_1 and b_1, respectively.

The technical Lemma 10.7 asserts that there is a constant $M_0 > 0$ such that for any positive $M < M_0$ the system 9.5.A of equalities and inequalities in the variables (9.7) has a solution such that $d(M)$ and $T(M)$ are given by (4.4) and (4.5), respectively. We take M_0 in Theorem 4.4 to be the constant whose existence is asserted by Lemma 10.7. Then we find that the frequency system in question is d-stable during the time T, where d and T are given by (4.4) and (4.5).

Now we assume that the assertion of Theorem 4.4 is not true. Then we can find a solution $v(t) = I(t), p(t), \varphi(t), q(t)$ of the system such that

$$I(0) = \mathcal{I}(v(0)) \in G - d, \quad p(t), q(t) \in D - d \quad \text{for all} \quad t \in [0, C],$$

where $C > 0$ is an arbitrary number and $|I(t_1) - I(0)| = d/2$ for some $t_1 \in [0, \min[C, T]]$. Hence, we can find a $t_2 \in [0, t_1]$ such that

$$|I(t_2) - I(0)| = d/2 \quad \text{and} \quad |I(t) - I(0)| \leqslant d/2 \quad \text{for all} \quad t \in [0, t_1].$$

Consequently, $v(t) \in V - d/2$ for $t \in [0, t_1] = [0, \min[t_1, T]]$, where V is given by (5.1). But the existence of such a solution contradicts the d-stability of the system during the time T. This proves Theorem 4.4.

§10. Statement of the lemma on the elimination of non-resonance harmonics, and of the technical lemmas used in the proof of the main theorem

As above, let $\mathbf{Z}^s$ be the lattice of s-dimensional vectors k with integer components.

10.1.A. DEFINITION. By the *order* of a vector $k = k_1, \ldots, k_s \in \mathbf{Z}^s$ we mean $|k| = \sum_{i=1}^{s} |k_i|$.

Let U be an arbitrary subset of either the complex space $\mathbf{C}^s$ or the real space $\mathbf{R}^s$. Let H_0 be a fixed function that is differentiable on U (that is, in some neighbourhood of U).

10.1.B. DEFINITION. A vector $k \in \mathbf{Z}^s$ is called a *ν-resonance* vector for U if for at least one point $I \in U$

$$|\langle k, \omega(I)\rangle| < \nu,$$

where $\langle\, ,\rangle$ is the scalar product and $\omega(I) = \operatorname{grad} H_0|_I$.

10.2. Description of H. In Lemma 10.3 below, H is a function of the form

$$H = H_0(I) + H_1(I, p, \varphi, q)$$

that is 2π-periodic in $\varphi_1, \ldots, \varphi_s = \varphi$, where I is an s-dimensional, and p and q are $(n - s)$-dimensional vectors, $n \geqslant s \geqslant 1$. We also assume that the following conditions hold (which differ slightly from those in §4.3).

10.2.A. H is analytic on the complex set

$$\Delta F\colon\ I,\ p,\ q \in A,\ \ |\operatorname{Im}\varphi| \leqslant \rho,$$

where A is a closed set and $\rho > 0$.

10.2.B. If we put $A_I = \{I \mid \exists\, p$ and q such that $I, p, q \in A\}$, then

$$\sup_{I \in A_I} \left\| \frac{\partial^2 H_0(I)}{\partial I^2} \right\| \leqslant m$$

for some $m < \infty$ (see (4.2)).

10.3. LEMMA (ON THE ELIMINATION OF NON-RESONANCE HARMONICS). *Let H be defined on ΔF, where H and ΔF are described above. Let M, N, and ν be positive numbers with the following three properties*:

A. $$\sup_{\Delta F} |\operatorname{grad} H_1| \leqslant M.$$

B. *Let λ denote the linear hull of all the $\frac{\nu}{2}$-resonance vectors k of order at most N, $|k| \leqslant N$, for A_I. Then* $\dim \lambda < s$.

C. *There is a $\varkappa \in (0, 1)$ such that for M, N, and ν*

$$N > N_0, \tag{10.1}$$

where

$$N_0 = N_0(s, \varkappa) = \\ = \left\{ \frac{27(1-\varkappa)(s+1)}{\varkappa \log 2} \log\left[\frac{3(1-\varkappa)(s+1)}{2\varkappa} \left(\frac{9}{\log 2}\right)^{\frac{1}{1-\varkappa}} (16L_1)^{\frac{\varkappa}{(1-\varkappa)(s+1)}} \right] \right\}^{\frac{1}{\varkappa}},$$

$$32 m L_2 \left(\frac{9}{\log 2}\right)^{2s+2} \frac{M N^{2(s+1)(1-\varkappa)}}{\nu^2} < 1, \tag{10.2}$$

$$\frac{4n\sqrt{n}}{sm} \left(\frac{\log 2}{27}\right)^{s} \frac{\nu}{N^{s(1-\varkappa)}} < 1, \tag{10.3}$$

and

$$L_1 = 6e\left(\frac{6s}{e}\right) \qquad L_2 = 4\pi s\sqrt{s}\, 3^{2s}.$$

We put

$$b = \rho L_3 \sqrt{\frac{M N^{(s+2)(1-\varkappa)}}{\nu}}, \tag{10.4}$$

where

$$L_3=\frac{8}{3}\sqrt{[2\pi n\sqrt{(sn)}3^s]}\left(\frac{9}{\log 2}\right)^{(s+2)/2}.$$

Then there is an analytic canonical diffeomorphism $B\colon J, P, \psi, Q\mapsto I, p, \varphi, q$ *with the following properties:*

D. *B maps a set contained in* ΔF *onto the set* ΔF_1:

$$\Delta F_1 = \{I,\ p,\ q\in A-b,\ |\operatorname{Im}\varphi|\leqslant\rho_1\},$$

where

$$\rho>\rho_1>0.$$

E. *The following estimate of the difference of B and the identity mapping* $\mathscr{E}$ *in the metric of C holds:*

$$\|B-\mathscr{E}\|<\frac{b}{15},$$

where

$$\|B-\mathscr{E}\|=\sup_{B^{-1}(\Delta F_1)}|B(J,\ P,\ \psi,\ Q)-J,\ P,\ \psi,\ Q|.$$

In terms of J, P, ψ, Q, *the Hamiltonian H has the form*

$$H'=\bar{H}(J,\ P,\ \psi,\ Q)+R(J,\ P,\ \psi,\ Q),$$

where $\bar{H}$ *and R have the following properties:*

F. *The Fourier series of* $\bar{H}$ *contains only those harmonics whose number k belongs to* $\lambda\cap\mathbf{Z}^s$, *that is,* $\bar{H}$ *has the form*

$$\bar{H}=\sum_{k\in\lambda\cap\mathbf{Z}^s} h_k(J,\ P,\ Q)e^{ik\psi}.$$

G. *For real* ψ, *that is, at all points* $J, P, \psi, Q\in B^{-1}(\Delta F_1)\cap\{\operatorname{Im}\psi=0\}$, *we have the following estimate for the derivatives* R_{ψ_j} *of R with respect to* ψ_j:

$$|R_{\psi_j}|<8\pi\ s^{1/2}M\exp(-\rho N^{1-\varkappa})\qquad (j=1,\ \ldots,\ s).$$

10.4. REMARK. In fact, it follows from the proof of Lemma 10.3 that $\bar{H}$ has the form

$$\bar{H}=H_0(J)+\bar{H}_1(J,\ P,\ Q)+\mathscr{H}(J,\ P,\ \psi,\ Q),$$

where $\bar{H}_1=\dfrac{1}{(2\pi)^s}\oint H_1(J,\ P,\ \psi,\ Q)$, and $|\mathscr{H}|$ and $|\operatorname{grad}\mathscr{H}|$ are of order M, more precisely, $|\mathscr{H}|<4\pi M\sqrt{s}$ and $|\operatorname{grad}\mathscr{H}|<\frac{8\pi M}{\rho}\sqrt{s}$. This refinement is not used in the proof of Theorem 4.4.

We turn to the technical lemmas.

10.5. LEMMA (ON A PROPERTY OF THE INTEGER LATTICE). *Suppose that for some vector* $x=x_1,\ \ldots,\ x_s$ *from* $\mathbf{E}^s$, *for a* $\nu>0$, *and* *for r linearly independent vectors k^j from $\mathbf{Z}^s$ with $|k^j| \leqslant N$ $(j = 1, \ldots, r)$ we have*

$$|\langle k^j, x\rangle| < \nu \qquad (j = 1, \ldots, r),$$

where $\langle\ ,\ \rangle$ is the scalar product. We denote by λ the r-dimensional linear subspace of $\mathbf{E}^s$ spanned by the k^j. Then the length of the orthogonal projection of x on λ is less than $rN^{r-1}\nu$, that is,

$$|\operatorname{pr}_\lambda x| < rN^{r-1}\nu.$$

10.6. Steep functions and almost plane curves. LEMMA. *Let H_0 be generalized steep in some domain $G \subset \mathbf{E}^s$ with coefficients g, m, $C_1, \ldots, C_{s-1}$, $\delta_1, \ldots, \delta_{s-1}$, and indices $\alpha_1, \ldots, \alpha_{s-1}$ (see Definition 7.3). Let λ be an arbitrary affine subspace of $\mathbf{E}^s$,* $\dim \lambda = r$, *where $r = 1, \ldots, s-1$. Let γ be a continuous curve with the following properties:*

A. *γ lies in a b-neighbourhood of the plane λ in $\mathbf{E}^s$: $\gamma \subset \lambda + b$, where $b > 0$.*

B. *γ joins two points I_0 and I_1 and is contained in the closed ball $\overline{U(I_0, d)}$ of radius d and with centre at I_0, and I_1 belongs to the boundary of this ball.*

C. *$\overline{U(I_0, d)}$ is contained in G.*

D. *d and b satisfy the inequalities:*

$$d < \min[g/3m, \delta_r], \tag{10.5}$$

$$\varepsilon < g/4, \tag{10.6}$$

$$b < \min[d/4, \varepsilon/4m], \tag{10.7}$$

where

$$\varepsilon = \frac{C_r}{5}\left(\frac{d}{2}\right)^{\alpha_r}.$$

Then we can find a point $\tilde{I}$ on γ such that

$$|\operatorname{grad}(H_0|_{\tilde{\lambda}})|_{\tilde{I}}| > \varepsilon,$$

where $\tilde{\lambda}$ is the affine subspace of $\mathbf{E}^s$ that is obtained by a translation from λ and passes through $\tilde{I}$.

10.7. LEMMA. *Let $s \geqslant 2$ be any integer. Let g, m, $C_1, \ldots, C_{s-1}$, $\delta_1, \ldots, \delta_{s-1}$, and ρ be arbitrary positive numbers, let $\alpha_1, \ldots, \alpha_{s-1}$ be greater than or equal to 1, and $n \geqslant s$. We call all these numbers and also functions of them, constants.*

We use the single word "relations" for equalities and inequalities. We consider relations involving the constants, a number $M > 0$, and the positive numbers

$$(10.8) \qquad \varkappa,\ N,\ N_1,\ \nu,\ \nu_0,\ \nu_1,\ \ldots,\ \nu_{s-1},\ b,\ b_1,\ d_0,\ \ldots,\ d_{s-1},\ \varepsilon_1,\ \ldots,\ \varepsilon_{s-1},\ d,\ T,$$

in which $r = 1, \ldots, s-1$:

$$(9.8) \qquad \varkappa < 1,$$

$$(9.1) \qquad N \leqslant N_1,$$

$$(8.5) \qquad \nu_{s-1} \leqslant g/2sN_1^{s-1},$$

$$(8.6) \qquad \nu_r \leqslant \varepsilon_r/rN_1^{r-1},$$

$$(8.9) \qquad \varepsilon_r = \frac{C_r}{5}\left(\frac{d_r}{2}\right)^{\alpha_r},$$

$$(8.7) \qquad d_{r-1} \leqslant \frac{1}{2mN_1}\nu_r,$$

$$(9.3) \qquad b \leqslant b_1,$$

$$(5.2) \qquad 4mbN \leqslant \nu,$$

$$(10.4) \qquad b = A_1 \sqrt{\left(\frac{MN^{(s+2)(1-\varkappa)}}{\nu}\right)},$$

where the constant A_1, *and the constants* A_2 *and* A_3 *occurring below, are given by* (10.9);

$$(10.1) \qquad N_0 < N,$$

where

$$N_0 = \left\{\frac{27(1-\varkappa)(s+1)}{\varkappa \log 2}\log\left[\frac{3(1-\varkappa)(s+1)}{2\varkappa}\left(\frac{9}{\log 2}\right)^{1/(1-\varkappa)} \cdot A_2^{\frac{\varkappa}{(1-\varkappa)(s+1)}}\right]\right\}^{1/\varkappa},$$

$$(10.2) \qquad A_3\frac{MN^{2(s+1)(1-\varkappa)}}{\nu^2} < 1,$$

$$(10.3) \qquad \frac{4n\sqrt{n}}{sm}\left(\frac{\log 2}{27}\right)^s \frac{\nu}{N^{s(1-\varkappa)}} < 1,$$

$$(5.3) \qquad b < \rho/2,$$

$$(5.4) \qquad T \leqslant \frac{\rho}{16\pi s} b \frac{1}{M}\exp(\rho N^{1-\varkappa}),$$

$$(9.5) \qquad 4(d_0 + d_1 + \ldots + d_{s-1}) \leqslant d,$$

$$(8.1) \qquad d_r < \min[g/3m,\ \delta_r],$$

$$(8.8) \qquad \nu_{r-1} \leqslant \nu_r/2,$$

$$(9.2) \qquad \nu \leqslant \min[\nu_0,\ \nu_1,\ \ldots,\ \nu_{s-1}],$$

$$(8.2) \qquad \varepsilon_r < g/2,$$

$$(8.4) \qquad b_1 < d_0,$$

$$(8.3) \qquad b_1 < \min[d_r/4,\ \varepsilon_r/4m],$$

$$(9.4) \qquad 4b < d.$$

Here

$$(10.9)\quad A_1=\frac{8}{3}\sqrt{[2\pi n\sqrt{(sn)}3^s]}\left(\frac{9}{\log 2}\right)^{(s+2)/2}\cdot\rho,\quad A_2=96e\left(\frac{6s}{e}\right)^s,$$

$$A_3=\frac{128}{9}\quad s^{3/2}\left(\frac{27}{\log 2}\right)^{2s+2}\cdot m.$$

Then there is a constant $M_0>0$ (the value of M_0 will be given in the proof of the lemma) with the following property: For every $M\in(0, M_0)$ we can find values of the quantities (10.8) *such that, together with M, they satisfy the above relations. Here,*

$$d=M^{\frac{3}{(12\zeta+3s+14)\alpha_{s-1}}},\qquad T=\frac{1}{M}\exp\left[\left(\frac{1}{M}\right)^{\frac{2}{12\zeta+3s+14}}\right],$$

where the constant ζ is given by (4.1).

§11. Remarks on the proof of the main theorem

In this section, as in §3, we discuss systems with a Hamiltonian of the form

$$(11.1)\qquad H=H_0(I)+\varepsilon H_1(I,\ \varphi).$$

11.1. The power of a harmonic, and the dependence on the order of its number. Let $\Sigma\,\varepsilon h_k(I)e^{ik\varphi}$ be the Fourier series of the perturbation εH_1, and let $I(t)$, $\varphi(t)$ be an arbitrary solution of the system (11.1). By the power of the harmonic $\varepsilon h_k(I)e^{ik\varphi}$ we mean the speed of the displacement of $I(t)$ stipulated by the influence of this harmonic. It follows from Hamilton's equations that this influence is proportional to the amplitude $\varepsilon h_k(I)$, and is roughly $\varepsilon\,|\,k\,|\,h_k(I)$. The amplitude of a harmonic decreases rapidly as the order $|\,k\,|$ of its number increases; for example, if H_1 is analytic, then we assume henceforth that the rate of decrease is exponential: $|\,h_k|<Me^{-|k|\rho}$, where $M>0$ and $\rho>0$ do not depend on k. Thus, the power of a harmonic decreases exponentially as $|\,k\,|$ increases linearly.

11.2. Dependence of the width of a resonance zone on the order of the number of a harmonic. The kth harmonic induces only small oscillations of $I(t)$ almost everywhere. This harmonic significantly affects the displacement of $I(t)$ only in a narrow resonance zone, a small neighbourhood of the surface $\langle k,\ \omega(I)\rangle=0$.

As a rule, the thickness of this zone is also bounded by a quantity that decreases exponentially as the order $|\,k\,|$ of the number of the harmonic increases linearly. To all appearances, this estimate holds for systems with any unperturbed Hamiltonian H_0, apart from those infinitely degenerate in the sense of §1.13.A. If H_0 is quasiconvex, then the thickness of the zone is close to the square root of the amplitude of the harmonic.

We note that this dependence of the thickness of a zone on the order $|\,k\,|$ of a zone was not taken into account in proving the main theorem. (The thickness of a zone is connected with its resonance width ν, and ν

was chosen to be independent of $|k|$.) A calculation of this dependence would possibly allow us to improve the value of the constant a, which determines the estimate of the time of stability obtained in this theorem (see (4.5) or (1.5)).

11.3. Dependence of the diameter of a trap on the multiplicity of a resonance and on minimal order $|k|$ of the vectors forming the resonance. In the proof of the main theorem, the diameters of the traps for the solutions with initial conditions $I(0)$, $\varphi(0)$ such that $I(0)$ belongs to an r-fold block are bounded by the quantities d_r, for which $d_{r+1} \gg d_r$ for $\varepsilon \ll 1$. This reflects the increase in the diameters of traps as the multiplicity of the resonance corresponding to this block increases.

The size of a trap depends on the minimal order of the vectors forming this resonance; the greater the order, the smaller the diameter of a trap, and moreover, the slower the drift into the trap (see §§11.1 and 11.2).

11.4. Overlapping of resonances. Arnol'd diffusion. One of the important parameters in the proof of the main theorem is the highest order N of the resonances in question. By means of the zones corresponding to these resonances we construct blocks in each of which fast drift can only occur along the planes λ obtained by a translation from the linear hull of those resonance vectors for this block whose order is not greater than N. All resonance vectors for the block that are not parallel to the planes λ are of order greater than N. The harmonics whose numbers are these vectors cause a drift across λ into the intersections of the block with the zones corresponding to these harmonics. When N is large, the speed of this drift is small, since the amplitudes of the harmonics causing the drift are small; this speed is estimated by a quantity close to the total power of these harmonics $(\sim \varepsilon \exp(-N))$.

When N is not too large, the point $I(t)$, which in each block moves close to the plane λ, is in a trap that is small together with $1/N$. By increasing N we obtain a better estimate of the speed of drift across λ, and this allows us to improve the estimate of the time $I(t)$ spends in the trap. For this reason we have to take N as large as possible, but the following situation shows that we cannot increase N unboundedly. As N increases, the network of resonance zones becomes more dense, and beginning from some value, the system of isolated traps loses its precise outline and becomes a more ramified gallery of transitions permeating the whole of the projection G of the phase space on the space of action variables. For any two points from G, the projections $I(t)$ of the solutions of the system, while moving along these transitions, can a priori go from any arbitrarily small neighbourhood of one point into such a neighbourhood of the other. All this points to the possibility of diffusion. It seems that the rate of diffusion is greatest in the zones of influence of the 1-fold resonances of small orders (Fig. 4).

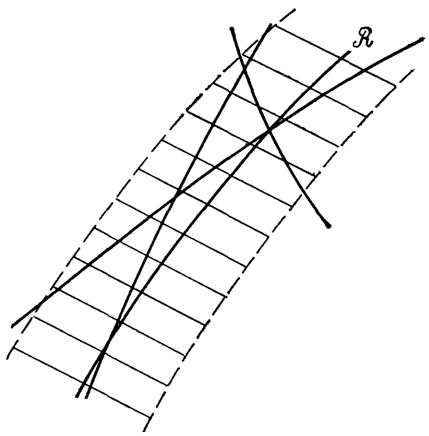

Fig. 4. The solid curves represent the intersections of 1-fold resonance surfaces with a section of the level surface of $H_0 = H_0(I_1, I_2, I_3)$. The order of the vector defining the surface $\mathcal{R}$ is small, and the orders of the "irreducible" vectors defining the other resonance surfaces intersecting the section, are large. The segments of parallel lines are the intersections with the resonance zone corresponding to $\mathcal{R}$, of the straight lines λ of fast drift. (This zone is outlined by a dotted line.) Every such segment intersects at least one resonance surface other than $\mathcal{R}$. Outside very narrow zones lying close to surfaces other than $\mathcal{R}$, the point $I(t)$ drifts only along the straight lines λ, and inside these zones it can also drift across these lines. Hence, it is possible, a priori, that if $I(t)$ receives a "lateral impulse" when it passes a resonance surface other than $\mathcal{R}$, then it jumps from one line λ to another. When $I(t)$ moves in this way close to $\mathcal{R}$, then it leaves its initial position.

The mechanism of diffusion has not yet been completely explained, and a strict proof of the existence of unstable solutions for a system of "general position" (see §2.1.A) is very difficult. For details of the mechanism of diffusion, see [11], [14], and [29].

11.5. Factors restricting the movement of $I(t)$. In proving the main theorem we only took into account one factor, the existence of almost integrals. In other words, we only considered the existence of those planes λ across which the speed of average motion $J(t)$ of the points $I(t)$ is very small; here $I(t)$, $\varphi(t)$ is a solution of (11.1), and J, ψ are the variables constructed in Lemma 10.3 (see §3.2.C). But there is another obvious restriction. By Remark 10.4, $J(t)$, $\psi(t)$ is a solution of a system with a Hamiltonian of the form $H_0(J) + \varepsilon\mathcal{H}_1(J, \psi)$, which differs only slightly from $H_0(J)$; here $\mathcal{H}_1$ is a certain function. Hence, while $I(t)$ is in the block in question, $J(t)$ moves close to the level surface of H_0.

By taking account of this restriction on the movement of $I(t)$ we could possibly improve the value of the constant a, which determines the estimate (4.5). It is curious that when we take account of this factor, we can obtain, in the following situation, a significantly better estimate of the time of stability than in the main theorem. Namely, for the solutions of systems with a quasiconvex unperturbed Hamiltonian H_0, with initial conditions $I(0)$ lying in the resonance zones of maximum multiplicity (that is, $s - 1$, where s is the number of frequencies) and of small order. By means of Lemma 10.3 it is easy to obtain for these solutions an estimate of the form (4.5), in which as s increases a decreases like $1/s$ and not like $1/s^2$, as in the main theorem. It is unlikely that by using equivalent arguments we could obtain such a good estimate for all solutions.

This fact is of interest in connection with the stability of the resonance relations that is observed in practice and is produced by non-Hamiltonian perturbations (see §2.1.D). If in a resonance zone of maximum multiplicity and small order the diffusion produced only by Hamiltonian perturbations is slower than in the whole space, then the transition of a real system into such a zone gives it greater stability with respect to all the action variables.

11.6. Behaviour of solutions in a resonance domain. Suitable coordinates for investigating resonance solutions. So far we have been interested mainly in the behaviour of the projections $I(t)$ of the solutions $I(t)$, $\varphi(t)$ of the system (11.1). How do the solutions behave? We assume that H_0 is quasiconvex and consider any r-fold block B, where $r = 1, \dots, s - 1$. It turns out that in the domain $B \times T^s$ of the phase space consisting of points I, φ such that $I \in B$, the system (11.1) is in a certain sense close to the system describing the motion of a material point along an r-dimensional torus in a field of force with potential energy of order ε.

To see this, we construct in $B \times T^s$ certain coordinates J, ψ, which may also be useful for a more detailed investigation of solutions in this domain. To do this we first transform to the coordinates constructed in Lemma 10.3 (here we denote them by J', ψ'). Then we make a linear substitution of the form $J = KJ'$, $\psi = (K^T)^{-1}\psi'$, where K is a unimodular matrix with integer elements, and $(K^T)^{-1}$ is the inverse of the transpose of K; a substitution of this form is canonical. Let $k', \dots, k^r$ be the vectors defining B. We choose K so that in the coordinates J, the planes λ of fast drift (that is, planes obtained by a translation from the linear hull of $k', \dots, k^r$) are given by $J_{r+1} = \text{const}, \dots, J_s = \text{const}$. This matrix always exists (see, for example, [28], Ch. 7, §1).

Note that the resonance surface $\mathscr{R}$ corresponding to B, which in the coordinates J' has the form

$$\mathscr{R} = \left\{ J' \,\middle|\, \left\langle k^i, \frac{\partial H_0(J')}{\partial J'} \right\rangle = 0,\ i = 1, \dots, r \right\},$$

has a simpler form in the coordinates J:

$$(11.2)\quad \mathscr{R} = \left\{ J \,\middle|\, \frac{\partial H_0(K^{-1}J)}{\partial J_i} = 0,\ i = 1, \dots, r \right\} \qquad \text{(see §3.3.D).}$$

We express the Hamiltonian of the system in terms of J', ψ' (see (3.2)), discard the remainder term $R(J', \psi')$, and make the change of variables

J', $\psi' \to J$, ψ. Then we obtain a function $\mathscr{H}$ independent of $\psi_{r+1}, \ldots, \psi_s$. It follows from the proof of Lemma 10.3 (see Remark 10.4) that $\mathscr{H}$ actually has the form

$$\mathscr{H} = H_0(K^{-1}J) + \varepsilon\mathscr{H}_1(J_1, \ldots, J_s, \psi_1, \ldots, \psi_r),$$

where $\mathscr{H}_1$ is a certain function.

Since $\mathscr{H}$ does not depend on $\psi_{r+1}, \ldots, \psi_s$, the variables $J_{r+1}, \ldots, J_s$ are first integrals of the system with the Hamiltonian $\mathscr{H}$, which is essentially a system with r degrees of freedom. (The functions $J_{r+1}, \ldots, J_s$ are crucial for the existence of the almost integrals of the original system (11.1).) Let λ be an arbitrary plane of fast drift that intersects B, and let $c_{r+1}, \ldots, c_s$ be constants such that

$$\lambda = \{J \mid J_i = c_i \ (i = r+1, \ldots, s)\}.$$

Let $c_1, \ldots, c_r, c_{r+1}, \ldots, c_s$ be the coordinates of a point of intersection of λ and $\mathscr{R}$. Then the values of the vector of the variables $J_1, \ldots, J_r$ at the points of intersection of λ and B lie in a neighbourhood of the constant vector $c_1, \ldots, c_r$. The diameter of this neighbourhood has the same order as the thickness of the zone close to $\mathscr{R}$ that we used to construct the block B, which is part of it. We assume that this thickness is small together with ε.

We fix the values of the arguments $J_{r+1}, \ldots, J_s$ of $\mathscr{H}$ by putting $J_i = c_i$, and approximate the resulting function in $\lambda \cap B$ more simply as follows. It follows from (11.2) that with these values of $J_{r+1}, \ldots, J_s$ the expansion of $H_0(K^{-1}J)$ in terms of $J_1, \ldots, J_r$ at the point $c_1, \ldots, c_r$, after neglecting terms of the third and higher orders, has the form

$a_0 + \sum_{i,j=1}^{r} a_{ij}(J_i - c_i)(J_j - c_j)$, where a_0 and a_{ij} are constants. We approximate the perturbation $\varepsilon\mathscr{H}_1$ by the function

$$\varepsilon\mathscr{H}_1(c_1, \ldots, c_s, \psi_1, \ldots, \psi_r),$$

which we denote by $\varepsilon U(\psi_1, \ldots, \psi_r)$. As a result we find that for fixed values of $J_{r+1}, \ldots, J_s$, $\mathscr{H}$ is close to the function

$$a_0 + \sum_{i,j=1}^{r} a_{ij}(J_i - c_i)(J_j - c_j) + \varepsilon U(\psi_1, \ldots, \psi_r), \tag{11.3}$$

where U is 2π-periodic with respect to its arguments. Since H_0 is quasi-convex, the symmetric matrix (a_{ij}) formed from the constants a_{ij}, is of fixed sign.

The system with the Hamiltonian (11.3) describes the motion of a material point over an r-dimensional torus with a constant metric, under a field of force with a potential energy εU. Here the ψ_i are the coordinates of a point of the torus, and the $J_i - c_i$ are the components of its momentum ($i = 1, \ldots, r$). Zero momentum corresponds to the point in the space of action variables that lies on the resonance surface $\mathcal{R}$.

§12. Application of the main theorem to the many-body problem

12.1. A planetary system. We consider a system of $s + 1$ points attracted to one another by Newton's law. We assume that the mass of one of them (the sun) is much greater than the mass of any of the others (the planets). We say that this system is *perturbed* relative to the following system: the planets are attracted towards a fixed centre by Newton's law, and there are no mutual interactions. The motion of the planets in the unperturbed system is the limit of their motions in the perturbed one, as their masses tend to zero. Note that the motion of the planets in the unperturbed system does not depend on their masses.

We assume that the masses of the sun and of the centre of attraction are each 1, and that the ratios of the masses m_i of the planets are fixed, that is, $m_i = \mu\varkappa_i$ ($i = 1, \ldots, s$), where the $\varkappa_i > 0$ are constants and μ is a small parameter.

The unperturbed system has been thoroughly studied, since it is integrable by quadrature. We assume that the energy of any planet is insufficient for it to go to infinity, and that the moment of momentum of every planet relative to the centre of attraction is non-zero. Then each planet moves in a closed elliptic orbit with one focus at the centre of attraction. The important parameters of these solutions are the lengths a_i of the semi-major axes of the orbits and their eccentricities e_i ($i = 1, \ldots, s$), and the angles b_{ij} ($i, j = 1, \ldots, s$) between the planes of the orbits.

Points in the phase space of the unperturbed system are characterized by the positions and speeds of the planets relative to the centre of attraction, and for the perturbed system, by the positions and speeds of the planets relative to the centre of mass of all the bodies in the system. This allows us in an obvious way to identify the configuration and phase spaces of both systems; in particular, we assume that the centre of attraction in the unperturbed system is the same point as the centre of mass in the perturbed one; we denote it by O.

Let V_e be the domain in phase space that corresponds to the elliptic orbits of all the planets in the unperturbed system. We regard a_i, e_i, and b_{ij} as functions on V_e; they are first integrals of the unperturbed system, and on solutions of the perturbed system they vary with time. If the distances between the bodies of the system are not too small, then the speed of displacement is of order μ. If the distance between at least two bodies is small, then this speed is large, and tends to infinity as the distance tends to zero.

Theorem 12.3 below asserts that V_e contains subdomains with the property that the values of the functions a_i on every solution with an initial condition belonging to any of these subdomains hardly change during a time interval that is exponentially large in comparison with $1/\mu$. Moreover, during the whole of this time, collisions or even close approaches of the bodies of the system are impossible. These subdomains are called domains of planetary motions, since they correspond to motions similar to those of the bodies of our solar system. We now define these subdomains.

12.2. **Domains of planetary motions.** Let G_i be the moment of momentum vector of the ith planet relative to O, G_H the moment of momentum vector of the whole unperturbed system relative to O, $G_H = \sum_{i=1}^{s} G_i$, and $N = \frac{1}{\mu} G$. It is not difficult to see that in V_e we have $|G_i| = m_i\sqrt{(a_i(1-e_i^2))}$. Hence, the length of N in V_e depends only on a_i, e_i, b_{ij}, and $\varkappa_i$ but not on μ.

Let $\alpha = \alpha_1, \ldots, \alpha_s$ and $\beta = \beta_1, \ldots, \beta_s$ be any two vectors, and γ a number. We denote by $B(\alpha, \beta, \gamma)$ the subset of V_e defined by:

$$\left\{\begin{array}{l} \alpha_i \leqslant a_i(v) \leqslant \beta_i \qquad (i=1, \ldots, s), \\ |N(v)| \geqslant \gamma, \end{array}\right. \tag{12.1}$$

where the $v \in V_e$ are points of phase space. This subset does not depend on μ. It is invariant for the unperturbed system. It is also clear that the conditions defining $B(\alpha, \beta, \gamma)$ are conditions on a_i, e_i, and b_{ij}.

Let α and β be arbitrary fixed vectors such that

$$0 < \alpha_1 \leqslant \beta_1 < \alpha_2 \leqslant \beta_2 < \ldots < \alpha_s \leqslant \beta_s. \tag{12.2}$$

Let us find the form of $B(\alpha, \beta, \gamma)$ for various values of γ. It is empty if γ is sufficiently large. It is not difficult to see that there is a largest value of γ for which $B(\alpha, \beta, \gamma)$ is non-empty; we denote it by $\gamma_m(\alpha, \beta)$. (To the points of $B(\alpha, \beta, \gamma_m(\alpha, \beta))$ correspond motions of the planets in circular orbits of radius β_i lying in the same plane.) Another critical value of γ is the largest value for which the boundary (in the whole of the phase space) of $B(\alpha, \beta, \gamma)$ contains points corresponding to collisions of the planets, either with one another or with the centre of attraction O. We denote this value by $\gamma_0(\alpha, \beta)$. Clearly, $0 < \gamma_0(\alpha, \beta) < \gamma_m(\alpha, \beta)$.

12.2.A. DEFINITION. A set $B(\alpha, \beta, \gamma)$ such that (12.1) holds and $\gamma_0(\alpha, \beta) < \gamma < \gamma_m(\alpha, \beta)$ is called a *domain of planetary motions.*

12.3. THEOREM. *Let B be any domain of planetary motions. Then there are positive constants C_1, C_2, C_3, and μ, depending only on the constants $\varkappa = \varkappa_1, \ldots, \varkappa_s$ and on B, with the following property.*

Let μ be an arbitrary point of $(0, \mu_0)$. Let $v(t)$ be any solution of the perturbed system with this value of μ, such that $v(0) \in B$. We put

$$T = \frac{1}{M} \exp \frac{1}{M^a}, \tag{12.3}$$

where

$$(12.4)\qquad M=C_1\mu,\quad a=\frac{2}{6s^2-3s+14}.$$

Then the bodies in the system cannot collide during the time interval $[0, T]$. *Moreover, for all* $t \in [0, T]$,

$$(12.5)\qquad \begin{gathered} |a_i(v(t))-a_i(v(0))|<C_2\mu^b \ (i=1,\ldots,s),\ \textit{where}\ b=\frac{3}{2}a,\\ |N(v(t))-N(v(0))|<C_3\mu, \end{gathered}$$

that is, the lengths a_i *of the semi-major axes of the orbits and the sum* N *of the moments of momentum hardly change.*

The rest of §12 is devoted to the proof of Theorem 12.3.

12.4. Poincaré variables. In this subsection we prove that in certain variables, the so-called Poincaré variables, the equations of motion are Hamilton's equations with a Hamiltonian of the same form as in Theorem 4.4.

12.3.A. LEMMA. *Every domain of planetary motions is compact.*

PROOF. Every domain B of planetary motions is bounded. On the other hand, B is contained in the interior of V_e, for otherwise the boundary of B would contain points corresponding to the collision of at least one of the planets with the centre O. Hence, B is closed not only in V_e (see (12.1)) but in the whole phase space, which is complete. This proves the lemma.

In the Euclidean space $\mathbf{E}^3$ containing the bodies of the system we introduce an arbitrary Cartesian coordinate system with origin at O. Let $x_i = x_{i1}, x_{i2}, x_{i3}$ be the coordinates of the ith planet, and $x_0 = x_{01}, x_{02}, x_{03}$ those of the sun. Then the kinetic energy T and potential energy U of the perturbed system have the form

$$(12.6)\qquad T=\frac{1}{2}\sum_{i=0}^{s} m_i\frac{\dot{x}_i^2}{2},\quad U=-\sum_{0\le i<j\le s}\frac{m_im_j}{|x_i-x_j|}.$$

Since the centre of mass of the system is at 0 and we have taken the mass m_0 of the sun to be 1,

$$(12.7)\qquad x_0=-\sum_{i=1}^{s} m_ix_i,\quad \dot{x}_0=-\sum_{i=1}^{s} m_i\dot{x}_i.$$

By using these equations we express the Lagrangian $L = T - U$ in terms of $x, \dot{x} = x_1, \ldots, x_s, \dot{x}_1, \ldots, \dot{x}_s$. Let $P_i = P_{i1}, P_{i2}, P_{i3}$ be the generalized momenta corresponding to $x_i = x_{i1}, x_{i2}, x_{i3}$:

$$(12.8)\qquad P_i=\frac{\partial L}{\partial \dot{x}_i}=m_i\dot{x}_i+m_i\sum_{j=1}^{s} m_j\dot{x}_j\qquad (i=1,\ldots,s).$$

We express the total energy $E = T + U$ of the system in terms of $P, x = P_1, \ldots, P_s, x_1, \ldots, x_s$, and represent $E(P, x)$ in the form

(12.9) $$E=\frac{1}{2}\sum_{i=1}^{s}\frac{P_i^2}{m_i}-\sum_{i=1}^{s}\frac{m_i}{|x_i|}+\mu^2R_\mu(P, x).$$

Let B' be any domain of planetary motions. By using the definition of this domain, the fact that B' is compact, and (12.6)–(12.9), it is not difficult to prove the following lemma.

12.4.B. LEMMA. *We can complete the definition of R, where $R(\mu, x, \dot{x}) = R_\mu(P(\mu, \dot{x}), x)$, for such values of μ, x, and $\dot{x}$ that it is analytic in the direct product of a neighbourhood of $\mu = 0$ on the complex line of the parameter μ, with a neighbourhood of B' in the complex space of x and $\dot{x}$.*

Now we define the Poincaré variables Y, where $Y = \Lambda, \xi, p, l, \eta, q$. Each of the vectors Λ, ξ, p, l, η, and q is s-dimensional, for example, $\Lambda = \Lambda_1, \ldots, \Lambda_s$. First we define the Poincaré variables for the unperturbed system. For every $i = 1, \ldots, s$, $\Lambda_i, \xi_i, p_i, l_i, \eta_i, q_i = Y_i$ are functions of x_i, $\dot{x}_i$, and $\varkappa_i$ that are independent of μ and i:

(12.10) $$Y_i = \Phi(\varkappa_i, x_i, \dot{x}_i).$$

These functions are described in [9], 140 and 131 (English); we do not repeat this description, but in what follows refer to it several times.

The Poincaré variables for the perturbed system are denoted by Y_μ, and are defined by the following condition: the functions defining the dependence of these variables on the momenta and coordinates of the planets for both systems (perturbed and unperturbed) are the same for the same values of μ. In the unperturbed system the momenta have the form $m_i\dot{x}_i$. Hence and from (12.10) we find that the variables $Y_\mu = Y_{\mu 1}, \ldots, Y_{\mu s}$ have the form

(12.11) $$Y_{\mu i}=\Phi\left(\varkappa_i, x_i, \frac{P_i}{\mu\varkappa_i}\right) \qquad (i=1, \ldots, s).$$

The relation (12.8) allows us to express them in terms of x and $\dot{x}$:

(12.12) $$Y_{\mu i}=\Phi\Big(\varkappa_i, x_i, \dot{x}_i+\mu\sum_{j=1}^{s}\varkappa_j\dot{x}_j\Big) \qquad (i=1, \ldots, s).$$

Note that when $\mu = 0$, the formulae (12.12) define the Poincaré variables for the unperturbed system (see (12.10)).

Even for $\mu = 0$ the Poincaré variables are not defined on the whole of V_e (the domain in the phase space corresponding to the elliptic orbits of the planets in the unperturbed system). Below we consider certain subsets of V_e on which these coordinates are defined.

Let ν be an arbitrary non-zero vector in $\mathbf{E}^3$, and $\delta > 0$. We denote by $K(\nu, \delta)$ the subset of the phase space in which to every point there correspond those positions and speeds of the planets such that the angle between the moment of momentum G_i of the ith planet and the vector ν, is not greater than $\pi - \delta$ for all $i = 1, \dots, s$:

$$(12.13) \qquad K(\nu, \delta) = \{v \mid G_i(v) \neq 0, \ \widehat{G_i(v), \ \nu} \leqslant \pi - \delta \ \ (i = 1, \dots, s)\}.$$

Let e_1, e_2, and e_3 be basis vectors for our Cartesian coordinate system in $\mathbf{E}^3$. Let B' be an arbitrary domain of planetary motions. By definition, $B' \subset V_e$. Hence and from the description of the Poincaré coordinates (see [9]) it follows that when $\mu = 0$ these coordinates are defined and analytic on $B' \cap K(e_3, \delta)$ for every $\delta > 0$. Moreover, the following lemma holds.

12.4.C. LEMMA. *For every $\delta > 0$ there is a neighbourhood $U_{x\dot{x}}$ of $B' \cap K(e_3, \delta)$ in the complex space of variables x, $\dot{x}$, with the following property: We can find a neighbourhood U_μ of $\mu = 0$ on the complex line of the parameter μ such that the Poincaré variables $Y = Y_1, \dots, Y_s$ as functions of μ, x and $\dot{x}$ (see (12.12)), are analytic in $U_\mu \times U_{x\dot{x}}$. In addition, $Y_\mu(x, \dot{x}) = Y(\mu, x, \dot{x})$ are coordinates on $U_{x\dot{x}}$ for every $\mu \in U_\mu$.*

The proof of the lemma follows from the form of the dependence of the Poincaré variables on μ (see (12.12)) and from the fact that $B' \cap K(e_3, \delta)$ is compact. The latter is a consequence of the compactness of B' (Lemma 12.4.A), the inequalities $\min_{B'} |\frac{1}{\mu} G_i| > 0$ $(i = 1, \dots, s)$ for every domain of planetary motions, and the form of the sets K (see (12.13)). Here $\frac{1}{\mu} G_i$ is the moment of momentum of the ith planet.

It follows from the description of the Poincaré variables that $\Lambda_{\mu i} = x_i\sqrt{a_i}$ for $\mu = 0$. The next lemma follows from this and Lemma 12.4.C.

12.4.D. LEMMA. *For every $\delta > 0$ there are numbers $C_4 > 0$ and $\mu_1 > 0$ such that for all $\mu \in (0, \mu_1)$,*

$$\max_{v \in B' \cap K(e_3, \delta)} |\Lambda_{\mu i}(v) - \varkappa_i \sqrt{a_i(v)}| \leqslant C_4 \mu \qquad (i = 1, \dots, s),$$

where

$$\Lambda_{\mu i}(v) = \Lambda_{\mu i}(x(v), \dot{x}(v)).$$

By using (12.11) we express P, x in terms of the Poincaré variables $Y = Y_\mu$. We denote $H = H_\mu$ the function $H_\mu(Y_\mu) = \frac{1}{\mu} E_\mu(P(Y_\mu), x(Y_\mu))$, and by $\mathcal{R}_\mu$ the function $\mathcal{R}_\mu(Y_\mu) = R_\mu(P(Y_\mu), x(Y_\mu))$ (see 12.9)). The next lemma follows from the description of the Poincaré variables (see [9]) and the form (12.9) of the energy function $E = E_\mu(P, x)$.

12.4.E. LEMMA. *In the domain where the Poincaré variables $Y = Y_\mu$ are defined, H has the form*

$$H(Y) = -\sum_{i=1}^{s} \frac{\varkappa_i^3}{2\Lambda_i^2} + \mu\mathcal{R}_\mu(Y).$$

Here $\mathcal{R}_\mu(Y)$ is 2π-periodic in $l = l_1, \ldots, l_s$. The equations of motion in the variables Y have the form of Hamilton's equations with the Hamiltonian $H(Y)$.

The last assertion follows from the fact that, although the change of variables $P, x \to Y_\mu$ is not canonical, in the variables $Y = Y_\mu$ the 2-form $dP \wedge dx = \sum_{i,j=1}^{s,3} dP_{ij} \wedge dx_{ij}$ is

$$\mu \sum_{i=1}^{s} (d\Lambda_i \wedge dl_i + d\xi_i \wedge d\eta_i + dp_i \wedge dq_i).$$

Lemma 12.4.E shows that $H = H(Y)$ has the same form as in the main theorem, with Λ and l playing the role of the action variables and the angle.

We put

$$H_0(\Lambda) = -\sum_{i=1}^{s} \frac{\varkappa_i^3}{2\Lambda_i^2}.$$

12.4.F. LEMMA. *The functions giving the dependence of the Poincaré variables Y on μ, x, and $\dot{x}$ do not depend on the choice of the coordinate system in $\mathbf{E}^3$ with origin O. For any fixed B' and δ, the form of $B' \cap K(e_3, \delta)$ in the coordinates x, $\dot{x}$ is also independent of the choice of this system. Moreover, H_0 and $\mathcal{R}_\mu$ do not depend on this choice.*

The first assertion follows from the definition of the Poincaré variables (see [9]) and from (12.12). The second follows from the definitions of B' and K. The third follows from the first and the fact that the form of all three terms on the right-hand side of (12.9) is independent of the choice of the coordinate system in $\mathbf{E}^3$ with origin O.

12.5. An application of the main theorem. In the main theorem the domain of the Hamiltonian has the form of a certain direct product. Here we consider subsets of the phase space of a special form; for each of these we can construct Poincaré variables whose range has the form of this direct product, and whose domain contains this subset. We prove that the conclusions of the main theorem hold in these subsets.

The next lemma is easily proved by using Definition 12.2.A.

12.5.A. LEMMA. *Let B be an arbitrary domain of planetary motions. Then there is another domain of planetary motions B' whose interior $\overset{0}{B}{}'$ contains B: $\overset{0}{B}{}' \supset B$.*

Let $\alpha = \alpha_1, \ldots, \alpha_s$ be an s-dimensional, and $\nu \neq 0$ a 3-dimensional vector. Let V denote the set

$$(12.14)\qquad W(\alpha, \nu) = \{v \in V_e \mid a_i(v) = \alpha_i \quad (i = 1, \ldots, s), N(v) = \nu\}.$$

Note that if W is contained in some domain of planetary motions, then it is compact (since this domain is compact (see Lemma 12.4.A)).

By the distance in the phase space we mean that defined by the Euclidean metric induced by the Euclidean metric in $\mathbf{E}^3$: the distance between two points is the square root of the squares of the differences of the coordinates x_i and speeds $\dot{x}_i$ of the planets. By $X + \delta$, where X is a subset of the phase space, we mean a δ-neighbourhood of X.

Let B and B' be arbitrary domains of planetary motions such that $\overset{0}{B}{}' \supset B$.

12.5.B. LEMMA. *There are positive constants δ_1 and δ_2 such that for every set $W \subset B$,*

$$W + \delta_1 \subset B' \cap K(\nu, \delta_2),$$

where ν is the vector such that $W = W(\alpha, \nu)$.

The lemma follows from the definitions of a domain of planetary motions and of the sets W and K (see (12.14)), and from the compactness of B and B'.

Let $W = W(\alpha, \nu)$ be an arbitrary set of the form (12.14). We define in $\mathbf{E}^3$ a Cartesian coordinate system with origin O by taking e_3 to be the vector $\frac{1}{|\nu|}\nu$ and choosing the other two basis vectors arbitrarily. Just as in §12.4 we introduce the Poincaré variables corresponding to this system in $\mathbf{E}^3$ and denote them by $Y_{\mu W}$. Suppose that W is contained in some domain of planetary motions. It follows from Lemmas 12.4.C, 12.5.B, and 12.5.A that the $Y_{\mu W}$ are defined in some neighbourhood of W for all μ in some neighbourhood of $\mu = 0$.

The Y_{0W} denote the variables $Y_{\mu W}$ for $\mu = 0$. The next lemma follows from the description of the Poincaré variables in [9] (we need to use, in particular, the relation $\Lambda_{0i} = x_i\sqrt{a_i}$ $(i = 1, \ldots, s)$ in §12.4).

12.5.C. LEMMA. *Let W be an arbitrary set of the form* (12.14), *contained in some domain of planetary motions. Then in the variables $Y_{0W} = \Lambda, \xi, p, l, \eta, q$, this set has the form*

$$W = \{v \mid \Lambda(v) = \text{const}, \quad (\xi, p, \eta, q)(v) \in \widetilde{D}_{\xi p \eta q}\},$$

where $\widetilde{D}_{\xi p \eta q}$ is a certain subset in the $4s$-dimensional space of the variables ξ, p, η, q.

The next lemma allows us to apply the main theorem to a neighbourhood of W.

12.5.D. LEMMA. *Let B be an arbitrary domain of planetary motions. Then there are positive constants C_1, C_5, μ_2, δ_3, δ_4, and ρ such that for every set W of the form* (12.14) *contained in B we can find open sets D_Λ*

and $D_{\xi p \eta q}$ *in the s-dimensional and* 4*s-dimensional spaces of variables* Λ *and* ξ, p, η, q, *respectively, with the following properties*:

a) *The closure of* D_Λ *is compact and does not intersect the coordinate hyperplanes* $\Lambda_i = 0$ $(i = 1, \ldots, s)$.

b) *Let F be a complex set of the same form as in the main theorem*:

$$F = \{Y = \Lambda,\ \xi,\ p,\ l,\ \eta,\ q \mid \operatorname{Re}\Lambda \in D_\Lambda,\quad \operatorname{Re}\xi,\ p,\ \eta,\ q \in D_{\xi p\eta q},\ \ |\operatorname{Im} Y| \leqslant \rho\}.$$

Then H_0 *and* $\mathcal{R}_\mu$ *(defined before Lemmas* 12.4.E *and* 12.4.F*) are analytic on this set, and*

$$\sup_{Y\in F}\left\|\frac{\partial^2 H_0(\Lambda)}{\partial\Lambda^2}\right\| \leqslant C_5, \qquad \sup_{Y\in F,\ \mu\in(0,\ \mu_2)} |\operatorname{grad}\mathcal{R}_\mu(Y)| \leqslant C_1.$$

c) *For every* $\mu \in (0, \mu_2)$ *the set*

$$\{v \mid \Lambda(v) \in D_\Lambda - \delta_3,\quad (\xi,\ p,\ \eta,\ q)(v) \in D_{\xi p\eta q} - \delta_3\},$$

where Λ, ξ, p, l, η, $q = Y_{\mu W}$, *contains* $W + \delta_4$. *Here* $D_\Lambda - \delta_3$ *is the set of points of the s-dimensional Euclidean space of variables* $\Lambda = \Lambda_1, \ldots, \Lambda_s$ *that are contained in* D_Λ *together with their* δ_3*-neighbourhoods. The set* $D_{\xi p\eta q} - \delta_3$ *is defined similarly.*

PROOF. Let B' be a domain of planetary motions such that $\overset{0}{B'} \supset B$. Let W be an arbitrary set of the form (12.14) such that $W \subset B$. It follows from Lemmas 12.5.B and 12.4.F that there are constants $\delta_2 > 0$ and $\delta > 0$, depending only on B, such that the δ-neighbourhood of W in the complex space of variables x, $\dot{x}$ is contained in the neighbourhood of $B' \cap K(e_3, \delta_2)$ whose existence is proved in Lemma 12.4.C. It follows from Lemmas 12.4.C and 12.4.F that there is a complex neighbourhood U'_μ of $\mu = 0$ depending on B, such that the Poincaré variables $Y_{\mu W}$ are defined in a complex $\delta/2$-neighbourhood of W for all $\mu \in U'_\mu$. The metrics in this $\delta/2$-neighbourhood induced by the Euclidean metric in the space of variables Y by means of the mapping $Y_{\mu W}^{-1}$: $Y \to x$, $\dot{x}$, are equivalent for all $\mu \in U'_\mu$ and $W \in B$.

It follows from Lemmas 12.5.C and 12.4.F that there are a constant $\delta_4 = \delta_4(B) > 0$ and open sets D_Λ and $D_{\xi p\eta q}$ such that for $\mu = 0$ the set

$$\{v \mid \Lambda(v) \in D_\Lambda,\quad (\xi,\ p,\ \eta,\ q)(v) \in D_{\xi p\eta q}\},$$

where Λ, ξ, p, l, η, $q = Y_{\mu W}$, is contained in $W + \frac{\delta}{4}$ and contains $W + 2\delta_4$. Hence assertion c) follows. The second estimate in b) follows from Lemmas 12.4.B and 12.4.F. The first estimate and assertion a) follow from Lemma 12.4.D and the definition of a domain of planetary motions. This proves

Lemma 12.5.D.

The function $H_0 = H_0(\Lambda)$ is quasiconvex outside the coordinate hyperplanes $\Lambda_i = 0$ $(i = 1, \ldots, s)$ (see §1.11 for the definition of a quasiconvex function). Hence, any compact set lying outside the coordinate hyperplanes has a neighbourhood in which H_0 is steep with every steepness index equal to 1 (see §1.11). The next lemma follows from this, Lemmas 12.4.E, 12.5.D and 12.4.D, and from Theorem 4.4.

12.5.E. LEMMA. *Let B be an arbitrary domain of planetary motions. Then there are positive constants C_1, C_2, μ_3, and δ_4 with the following property: Let W be an arbitrary set of the form* (12.14) *such that $W \subset B$; μ any point of* $(0, \mu_3)$; *$v(t)$, $0 \leqslant t \leqslant \tau$, any solution of the perturbed system with this value of μ whose trajectory lies entirely in $W + \delta_4$; $\tau > 0$ arbitrary. Then*

$$| a_i(v(t)) - a_i(v(0)) | < C_2\mu^b \qquad (i = 1, \ldots, s)$$

for all $t \in [0, \min[\tau, T]]$, where b and $T = T(C, \mu)$ are given by (12.3), (12.4), *and* (12.5).

12.6. Stability of the lengths of the semi-major axes, and of the sum of the moments of momentum of the planets during an exponentially large interval of time.

12.6.A. LEMMA. *Let B be a domain of planetary motions. Then there are positive constants δ_5 and C_3 with the following property: Let $\mu > 0$ be an arbitrary number, let $v(t)$, $0 \leqslant t < \tau$, any solution of the perturbed system with this value of μ whose trajectory lies entirely in $B + \delta_3$; $\tau > 0$ can also take the value ∞. Then on this solution the difference between the sum N of the moments of momentum of the planets at the time t and its initial value is bounded by a quantity of order μ:*

$$\sup_{t \in [0, \tau)} | N(v(t)) - N(v(0)) | < C_3\mu.$$

PROOF. Let B' be another domain of planetary motions such that $\overset{0}{B'} \supset B$. Since B and B' are compact, we have $\rho(\partial B', B) > 0$. We take this distance for δ_5; then, by hypothesis, the trajectories of the solutions $v(t)$ in question lie in B'.

Let G be the moment of momentum vector of the perturbed system relative to O, $G = \sum_{i=0}^{s} G_i$, where G_i for $i > 0$ is the moment of momentum of the ith planet, and G_0 that of the sun. B' is compact, and so it follows from (12.7) that we can find a constant $C_3 > 0$ such that

$$\max_{v \in B'} | G_0(v) | < \frac{C_3}{2}\mu^2. \tag{12.15}$$

We recall that the moment of momentum G_H of the unperturbed system is connected with the G_i and N by $G_H = \mu N = \sum_{i=1}^{s} G_i$. From (12.15) we obtain

$$\max_{v \in B^*} |G - G_H| < \frac{C_3}{2} \mu^2.$$

But G is a first integral of the perturbed system. Hence we obtain the assertion of the lemma.

Let $W = W(\alpha, \nu)$ be a set of the form (12.14), B a domain of planetary motions, and $O_{\mu W B}$ the neighbourhood of W given by

$$(12.16) \quad O_{\mu W B} = \{v\colon\ |a_i(v) - \alpha_i| < C_2 \mu^b \ (i = 1, \ldots, s),\ |N(v) - \nu| < C_3 \mu\},$$

where b, $C_2 = C_2(B)$, and $C_3 = C_3(B)$ are as in Lemmas 12.5.E and 12.6.A.

12.6.B. LEMMA. *Let B be a domain of planetary motions. Then for every $\delta > 0$ we can find a constant $\mu_4 = \mu_4(\delta) > 0$ such that $O_{\mu W B} \subset W + \delta$ for every $\mu \in (0, \mu_4)$ and $W \subset B$.*

12.6.C. PROOF OF THEOREM 12.3. This follows from Lemmas 12.5.E, 12.6.A, and 12.6.B.

Let B be a domain of planetary motions. We take for C_1, C_2, and C_3 the constant whose existence is proved in Lemmas 12.5.E and 12.6.A. We take $\mu_0 = \min\,[\mu_3, \mu_4(\delta)]$, where $\delta = \frac{1}{2} \min\,[\delta_4, \delta_5]$; here μ_3, $\mu_4(\delta)$, δ_4, and δ_5 are the quantities whose existence is proved in Lemmas 12.5.E, 12.6.A, and 12.6.B.

We prove Theorem 12.3 by contradiction. Assume that it is not true. Then there exists a solution $v(t)$ of the perturbed system with μ equal to some $\bar{\mu}$, $\bar{\mu} \in (0, \mu_0)$, such that $v(0) \in B$, and with the following property: Let $\bar{W}$ be a set of the form (12.14) such that $\bar{W} = W(\bar{\alpha}, \bar{\nu})$, where $\bar{\alpha} = a(v(0))$ and $\bar{\nu} = N(v(0))$. Then we can find a $\tau \in [0, T]$ such that $v(t) \in O_{\bar{\mu}\bar{W}B}$ for all $t \in [0, \tau)$ and $v(t) \in \partial O_{\bar{\mu}\bar{W}B}$, where $\partial O_{\bar{\mu}\bar{W}B}$ is the boundary of the set $O_{\bar{\mu}\bar{W}B}$ defined by (12.15). But this contradicts Lemmas 12.5.E and 12.6.A, and so Theorem 12.3 is proved.

References

[1] N. N. Nekhoroshev, Behaviour of nearly integrable Hamiltonian systems, Funktsional. Analiz i Prilozhen. 5:4 (1971), 82–83.
= Functional Anal. Appl. 5 (1971), 338–339. MR **45** #3881.

[2] H. Poincaré, Les méthodes nouvelles de le mécanique céleste. 3 vols., Paris 1892, 1893, 1899.
Translation: *Izbrannye trudy*, Nauka, Moscow 1971–1974.

[3] G. Birkhoff, Dynamical systems, Amer. Math. Soc. Coll. Publ. **9**, Amer. Math. Soc., New York 1927.
Translation: *Dinamicheskie sistemy*, Gostekhizdat, Moscow–Leningrad 1941.

[4] C. L. Siegel, Über die Existenz einer Normalform analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung, Math. Ann. **128** (1952), 144–170. MR **16**–704. MR **16**–704.
= Matematika **5**:2 (1961), 129–156.

[5] C. L. Siegel, Vorlesungen über Himelsmechanik, Springer–Verlag, Berlin–Göttingen–Heidelberg 1956. MR **18**–178.
Translation: *Lektsii po nebesnoi mekhanike*, Izdat. Inostr. Lit., Moscow 1959.

[6] A. N. Kolmogorov, On the conservation of conditionally periodic motions for a small change in the Hamiltonian, Dokl. Akad. Nauk SSSR **98** (1954), 527–530. MR **16**–924.

[7] A. N. Kolmogorov, The general theory of dynamical systems and classical mechanics, *Mezhdunarodnyi Matematicheskii Kongress v Amsterdame,* (Internat. Congress Mathematicians, Amsterdam), Fizmatgiz, Moscow 1961, pp. 187–208.

[8] V. I. Arnol'd, Proof of a theorem of A.N. Kolmogorov on the invariance of conditionally periodic motions under small perturbations of the Hamiltonian, Uspekhi Mat. Nauk **18**:5 (1963), 13–40. MR **29** # 328.
= Russian Math. Surveys **18**:5 (1963), 9–36.

[9] V. I. Arnol'd, Small denominators and problems of stability of motion in classical and celestial mechanics, Uspekhi Mat. Nauk **18**:6 (1963), 91–192. MR **30** # 943.
= Russian Math. Surveys **18**:6 (1963), 85–191.

[10] V.I. Arnold and A. Avez, Ergodic problems of classical mechanics, Benjamin, New York–Amsterdam 1968. MR **38** # 1233.

[11] V. I. Arnol'd, Instability of dynamical systems with many degrees of freedom, Dokl. Akad. Nauk SSSR **156** (1964), 9–12. MR **29** # 329.
= Soviet Math. Dokl. **5** (1964), 581–585.

[12] J. K. Moser, On invariant curves of area-preserving mappings of an annulus, Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. II **1962**, 1–20. MR **26** # 5255.
= Matematika **6**:5 (1962), 51–67.

[13] J. K. Moser, A rapidly convergent iteration method and non-linear partial differential equations, I, II, Ann. Scuola Norm. Sup. Pisa (3), **20** (1961), 265–315; 499–535. MR **33** # 7667; **34** # 6280.
= Uspekhi Mat. Nauk **23**:4 (1968), 179–238.

[14] G. M. Zaslavskii and B. V. Cherikov, Stochastic instability of non-linear oscillations, Uspekhi Fiz. Nauk **105**:1 (1971), 3–40.
= Soviet Physics Upskehi **14** (1971/2), 549–568.

[15] V. I. Arnol'd, A stability problem and ergodic propetties of classical dynamical systems, Proc. Internat. Congress Mathematicians (Moscow 1966), "Mir", Moscow 1968, 387–392. MR **39** # 574.
= Amer. Math. Soc. Transl. (2) **70** (1968), 5–11.

[16] J. Glimm, Formal stability of Hamiltonian systems, Comm. Pure Appl. Math. **17** (1964), 509–526. MR **32** # 7876.

[17] N. N. Nekhoroshev, Stable lower estimates for smooth mappings and for gradients of smooth functions, Mat. Sb. **90** (1973), 432–478.
= Math. USSR–Sb. **19** (1973), 425–467.

[18] J. K. Moser, On the elimination of the irrationality condition and Birkhoff's concept of complete stability, Bol. Soc. Mat. Mexicana (2) **5** (1960), 167–175. MR **24** # A3375.

[19] M. M. Khapaev, Generalization of Lyapunov's second method and the investigation of some resonance problems, Dokl. Akad. Nauk SSSR **193** (1970), 46–49. MR **44** # 4300.
= Soviet Math. Dokl. **11** (1970), 868–872.

[20] V. V. Beletskii, *Dvizhenie sputnika otnositel'no tsentra mass v gravitatsionnom pole* (Motion of a satellite relative to its centre of mass in a gravitational field), Izv. Moskov. Gos. Univ., Moscow 1975.

[21] J. K. Moser, Lectures on Hamiltonian systems, Mem. Amer. Math. Soc. **81**, Amer. Math. Soc., Providence, R.I., 1968. MR **37** #6060.
Translation: *Lektsii o gamil'tonovykh sistemakh,* Mir, Moscow 1973.

[22] J. K. Moser, Stabilitätsverhalten kanonischer Differentialgleichungssysteme, Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. IIa **1955**, 87–120. MR **18**–40.

[23] J. K. Moser, New aspects in the theory of stability of Hamiltonian systems, Comm. Pure Appl. Math. **11** (1958), 81–114. MR **20** #3354.

[24] G. Contopoulos, On the existence of a third integral of motion, Astronomical J. **68** (1963), 1–14.

[25] M. Hénon and C. Heiles, The applicability of the third integral of motion; some numerical experiments, Astronomical J. **69** (1964), 73–79.

[26] A. D. Bryuno, Formal stability of Hamiltonian systems, Mat. Zametki **1** (1967), 325–330. MR **35** #4527.
= Math. Notes **1** (1967), 216–219.

[27] V. I. Arnol'd, *Matematicheskie metody klassicheskoi mekhaniki* (Mathematical methods of classical mechanics), Nauka, Moscow 1974.

[28] N. Bourbaki, Eléments de mathématique. Partie I. Les structures fondamentales de l'analyse. Livre III, Topologie générale, Chaps. III–VIII, Actualities Sci. Ind. No. 916 and No. 1029, Hermann, Paris 1942 and 1947. MR **5**–102; **9**–261.
Translation: *Obshchaya topologiya. Chisla i svyazannye s nimi gruppy i prostranstva,* Nauka, Moscow 1969. MR **41** #984.

[29] B. V. Cherikov, Investigation of the theory of a non-linear resonance and the stochastic property, Dissertation, Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, 1969.

[30] V. I. Arnol'd, Small denominators, I, Mapping the circle onto itself, Izv. Akad. Nauk SSSR Ser. Mat. **25** (1961), 21–86. MR **25** #4113.

[31] H. Rüssman, Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes (Kleine Nenner. I), Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. IIa **1970**, 67–105. MR **42** #8037.

[32] H. Rüssman, Kleine Nenner II. Bermerkungen zur Newtonschen Methode, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II **1972**, 1–10. MR **46** #8407.

[33] E. Zehnder, Generalized implicit function theorems with applications to some small divisor problems. I, Comm. Pure Appl. Math. **28** (1975), 91–140. MR **52** # 1764.

[34] E. Zehnder, Generalized implicit function theorems with applications to some small divisor problems. II, Comm. Pure Appl. Math. **29** (1976), 49–111.

[35] A. I. Neishtadt, Passage through a resonance in a two frequency problem, Dokl. Akad. Nauk SSSR **221** (1975), 301–304.
= Soviet Physics Dokl. **20** (1975), 189–191.

[36] A. I. Neishtadt, Certain resonance problems in non-linear systems, Dissertation, Moscow State Univ., Moscow 1976.

[37] V. I. Arnol'd, *Nekotorye zadachi teorii differentsial'nykh uravnenii. Tezisy lektsii* (Lectures notes: Some problems in the theory of differential equations), Moscow 1974.

Received by the Editors 4 February 1977

Translated by L. W. Longdon