ICARUS 8, 203–215 (1968)

The Resonant Structure of the Solar System

The Law of Planetary Distances

A. M. MOLCHANOV

Communicated by B. J. Levin

Received December 16, 1966

It has been shown that the relations existing in the series of planetary distances appear as a result of more simple and exact relations in the series of their frequencies. The structure of the solar system is determined by a table of resonance relations. Analogous tables of integers determine the constitution of the systems of planetary satellites. It will be our aim to explain theoretically that the appearance of resonant structure in a given oscillating system is inevitable. A small value of coefficients (one, two, or three) of the resonant relation is more probable if the evolution occurs under the action of dissipative perturbations of mutually comparable size. If the dissipative factors are small in comparison with the conservative influences (of the same type as the attraction of the planets) then the system may conceivably remain in "distant" resonance.

I. Introduction

In celestial mechanics, conservative (Hamiltonian) forces prove to be a fundamental object of study. They are characterized by stationary modes of the same type as uniform equilibrium. In the potential field, arbitrary values of orbital parameters are possible.

Systems in which dissipative terms are dominant are characterized by quite different properties. They have isolated stable states—limiting cycles, which may be reduced, with growing of dissipation, to the stable position of equilibrium. However, the division of terms into dissipative and conservative categories is an idealization useful only for the description of extreme limiting cases. For the solar system in its present aspect, there is no doubt that the potential field is dominant. However, the system has existed for such a long time, 5×10^9 years, that even very small dissipative terms have had sufficient time to exert an influence. As will be shown below, small dissipative terms, especially in resonance zones, have a tendency to create stable configurations. Since resonance zones are determined by sets of integers, a discrete number of possible stationary states occurs. The solar system then, finds itself in one of these states.

II. Internal Resonance of Nonlinear Systems

The concept of resonance usually arises during the study of an external periodic impulse or an oscillating system. This method of consideration assumes that a small parameter exists in the problem which permits one to neglect the reverse influence of the oscillating system on the outside medium. In this simplified form the phenomenon of resonance has been extensively studied, and the results are well known. In systems where there is no distinct disparity of the components which would permit separation of the "medium" from the "system," the phenomenon of resonance is considerably more interesting and complex.

In what follows we shall limit ourselves to systems consisting of weakly bound components. The equations of motion of such a system are of the form

$$dx_i/dt = A_i(x_i) + \epsilon B_i(x_1, \ldots x_l; \epsilon); \qquad i = 1, 2, \ldots l. \quad (1)$$

Similar notation may also be used to describe the planetary system. The components in this case prove to be pairs of vectors—the radius-vector of the planet and the velocity vector (in the heliocentric coordinate system). The principal term in each equation describes the mutual influence of the planet and central body. The perturbation terms contain the mutual influence of the planets upon one another as well as other incidental influences; and depend, generally speaking, on all components $(x_1, \ldots x_l)$. The nondimensional parameter ϵ characterizes the smallness of the perturbations. Since the fundamental contributions to the perturbations arise from the general interaction of the planets, the size of ϵ is of the same magnitude as the ratio of the masses of all the planets to the mass of the central body. In particular, for the solar system $\epsilon \sim 10^{-3}$.

It is more convenient to study the perturbed system by changing to new variables—phases ϕ_i and first integrals I_k of the unperturbed system. In the planetary system the variables ϕ_1 are mean anomalies (one for each planet), and the variables I_k are the orbital elements (five to each planet). In terms of these new variables,

$$\begin{aligned} d\phi/dt &= \omega(I) + \epsilon\omega(I, \phi; \epsilon) \\ dI/dt &= \epsilon\mathfrak{F}(I, \phi; \epsilon) \end{aligned} \qquad (2)$$

are smooth functions of the slow variables $I = \{I_k\}$ and are periodic (with period 2π) in each of the rapid variables $\phi = \{\phi_i\}$. The variables ϕ_i are well defined in the torus $\tau = \mathfrak{M}[\phi_i; 0 \leqslant \phi_i < 2\pi]$.

Unperturbed motion in the system (2) is very easily obtained—the slow variables remain constant

$$I = \text{const.} \qquad (3)$$

and the rapid variables increase in direct proportion to the time (more accurately, they run through the torus τ many times)

$$\phi = \omega t + \phi_0. \qquad (4)$$

However, at various points of the space I the character of the motion of the phase variables differs. Generally speaking, the trajectory (4) is everywhere well-defined on the torus τ. However, for the points I satisfying the resonance relation

$$(\mathbf{n}, \boldsymbol{\omega}) \equiv n_1\omega_1(I) + \cdots + n_l\omega_l(I) = 0, \qquad (5)$$

the torus τ becomes divided into tori of smaller dimensions, defined by the equation

$$\psi \equiv n_1\phi_1 + \cdots + n_l\phi_l = \text{const.} \qquad (6)$$

As a result, instead of changing—as do the phase variables—the resonance phase ψ acts the part of a first integral and remains constant. However, it is impossible to confound the resonance phase with a first integral because the resonance phase is constant only at points of the resonance surface (5). At a given point, the resonance phase ψ, and the other phases as well, increase linearly as

$$\psi = \nu t + \psi_0, \qquad (7)$$

where $\nu = \nu(I)$, the corresponding resonance frequency

$$\nu(I) = n_1\omega_1(I) + \cdots + n_l\omega_l(I). \qquad (8)$$

Consequently, the full phase space is the direct product of the torus τ and the space I. The space containing the slow variables I proves to be a great number of equal stationary modes of the unperturbed system. To each point I there corresponds a motion of the rapid variables on the τ torus determined, in agreement with the formula (4) by the frequency vector $\boldsymbol{\omega}$

$$\boldsymbol{\omega} \equiv \boldsymbol{\omega}(I). \qquad (9)$$

At the points of the space I situated on resonance surfaces (5), there appears a supplementary "incidental" first integral—the resonance phase

$$\psi = \text{const.} \qquad (10)$$

Each resonance surface is characterized by the integer vector of resonance

$$\mathbf{n} \equiv (n_1, \ldots n_l). \qquad (11)$$

Thus there is a finite manifold of resonance planes, and they constitute a manifold of measure zero. Surfaces (on a unit of lesser dimensions) obtained by the intersection of two resonance surfaces consist of points of double resonance.

Double resonances are distributed on a given resonance surface analogously to the manner in which resonance surfaces are distributed in the space of slow variables I, forming a manifold of measure zero. Triple resonances form a manifold of measure zero on the surfaces of double resonance, and so on.

III. The Resonance of the Solar System

It may seem that the "and so on" which concludes the foregoing paragraph is simply a mathematical way of speech, an attempt to achieve formal completeness. From the point of view of the theory of measure (Birkhoff, 1933) the presence of even one resonance is surprising; and double or triple resonances are extremely improbable. However, nature does not limit itself to these very convincing reasons but seemingly is drawn to the opposite extremes. Analysis (Molchanov, 1966) of the observed data (see Whipple, 1948) brings us to the conclusion that the solar system possesses the maximum possible number of resonance relations—eight for nine planets. We shall postpone for now the question, "How does this occur?" (Schroedinger, 1948) and we shall formulate some results of this important experimental fact.

The resonance relations can be regarded as a family of equations which are linear in the frequencies. The number of independent resonances cannot, therefore, be equal to the number of frequencies, since a homogeneous system has only a null solution. If the number of resonances is smaller by one, then all the frequencies can be expressed in terms of one. This one then proves to be the only dimensional parameter and determines the scale of the system.

As a result, systems with maximum resonance are uniquely determined (with comparable accuracy) by a set of whole numbers—a table of resonance vectors. The resonance relations for the solar system (obtained by the analysis of a table of frequencies) appear as follows:

$$
\begin{aligned}
\omega_1 - \omega_2 - 2\omega_3 - \omega_4 &= 0,\\
\omega_2 - 3\omega_4 - \omega_6 &= 0,\\
\omega_3 - 2\omega_4 + \omega_5 - \omega_6 + \omega_7 &= 0,\\
\omega_4 - 6\omega_5 - 2\omega_7 &= 0,\\
2\omega_5 - 5\omega_6 &= 0,\\
\omega_5 - 7\omega_7 &= 0,\\
\omega_7 - 2\omega_8 &= 0,\\
\omega_7 - 3\omega_9 &= 0.
\end{aligned}
\tag{12}
$$

Choosing Jupiter's frequency as the unit of measurement, we find the frequencies[1] of the other planets. The frequencies and the resonance vectors,[2] the coefficient matrix of the system (12), are given in Table I.

The results obtained cannot be explained

TABLE I
Frequencies and Resonance Vectors of the Solar System

Planet	ω_{obs}	ω_{theor}	$\Delta\omega/\omega$									
Mercury	49.22	49.20	0.0004	1	−1	−2	−1	0	0	0	0	0
Venus	19.29	19.26	0.0015	0	1	0	−3	0	−1	0	0	0
Earth	11.862	11.828	0.0031	0	0	1	−2	1	−1	1	0	0
Mars	6.306	6.287	0.0031	0	0	0	1	−6	0	−2	0	0
Jupiter	1.000	1.000	0.0000	0	0	0	0	2	−5	0	0	0
Saturn	0.4027	0.400	0.0068	0	0	0	0	1	0	−7	0	0
Uranus	0.14119	0.14286	−0.0118	0	0	0	0	0	0	1	−2	0
Neptune	0.07197	0.07143	0.0075	0	0	0	0	0	0	1	0	−3
Pluto	0.04750	0.04762	−0.0025	0	0	0	0	0	1	0	−5	1

[1] Which are, of course, rational and rounded off in the table to four significant figures.

[2] The last vector proves not to be a resonance vector and is chosen so that the square matrix obtained should have a determinant equal to unity. The meaning of this procedure will be explained in Section III.

by chance. The resonance is very simple.[3] Most of the positions of the table are occupied by zero, ones and twos, and there are too many of them to allow us to believe that this has occurred by a chance. The explanation must be causal; it would appear inescapable that the resonance relations are the result of evolution. However, it would be hazardous, on the basis of one example, to draw the more important conclusion that maximum resonance is a structural principle.

Fortunately, in the solar system there are at least three other subsystems on which it is possible to test the predicted strength of the principle of maximum resonance—these are the systems of Jupiter's, Saturn's, and Uranus' satellites. Their frequencies and resonance vectors[2] are given in Table II, directly confirming the truth of the principle.

TABLE II

FREQUENCIES AND RESONANCE VECTORS OF SATELLITE SYSTEMS

Satellite	ω_{obs}	ω_{theor}	$\Delta\omega/\omega$								
			Saturn's Satellites								
Mimas	16.918	16.800	0.0070	−1	0	2	0	0	0	0	0
Enceladus	11.639	11.600	0.0035	0	−1	0	2	0	0	0	0
Phoebe	8.448	8.400	0.0057	0	0	−1	0	2	1	0	2
Diana	5.826	5.800	0.0045	0	0	0	−1	2	−1	0	−1
Rhea	3.530	3.500	0.0086	0	0	0	0	−1	2	2	0
Titan	1	1	0.0000	0	0	0	0	0	−3	4	0
Hyperion	0.7494	0.7500	0.0008	0	0	0	0	0	−1	0	5
Japetus	0.2010	0.2000	0.0050	0	0	0	0	0	0	−1	4
			Jupiter's Satellites								
Io	4.044	4.000	0.0110	1	−2	0	0				
Europa	2.015	2.000	0.0075	0	1	−2	0				
Ganymede	1.000	1.000	0.0000	0	0	−3	7				
Callisto	0.4288	0.4285	0.0008	0	0	−1	2				
			Uranus' Satellites								
Miranda	6.529	6.545	−0.0025	−1	1	1	1	0			
Ariel	3.454	3.454	−0.0000	0	−1	1	2	−1			
Umbriel	2.100	2.091	0.0043	0	0	−2	1	5			
Titania	1.000	1.000	0.0000	0	0	1	−4	3			
Oberon	0,6466	0.6364	0.0160	0	0	1	−2	0			

The law of maximum resonance, which now may be accepted at least as an heuristic principle, determines a discrete set of states in which an oscillating system can find itself. From this stems the necessity of changing the constitution of the question concerning the law of planetary distances (Schmidt, 1950) since simple integral value relations occur not for the distances, but for the frequencies, through which the distances are uniquely determined. Besides this there arises the problem of theoretically explaining the law of the maximum resonances itself.

IV. THE STRUCTURE OF THE RESONANCE ZONE ($t \sim \epsilon^{-1/2}$) ESSENTIAL (PRINCIPAL) RESONANCE

It is not possible to understand the preference shown to resonance surfaces by examining only the unperturbed system. In this system all the points of the space are quite equal—they are all points of uniform

[3] Of 160 positions in the tables of resonances, 0 occupies 98 places and 1 occupies 33 positions. The remaining numbers occur remarkably less often. Two occupies 16 positions 3 only five positions. Numbers greater than 3 occupy only eight vacancies—5 occurs three times, 7 and 4 each arise twice and 6 is found in the last vacant place.

equilibrium. A more accurate model of the full system has been obtained (Molchanov, 1963) in studying a new motion (with velocity of the order of ϵ)[4] in the space, defined

$$\begin{aligned} d\phi/dt &= \omega(I), \\ dI/dt &= \epsilon G(I, \epsilon), \end{aligned} \tag{13}$$

where

$$\begin{aligned} G(I, \epsilon) &= \langle \mathfrak{F}(I, \phi, \epsilon) \rangle_{\text{av}} \\ &= \frac{1}{2\pi} \int_{\tau} \mathfrak{F}(I, \phi, \epsilon) d\phi. \end{aligned} \tag{14}$$

The system (13) already, in principle, have preferred points or regimes, in which sooner or later a given trajectory will be drawn, independently of the initial conditions. But these separated states, even if they exist, are entirely determined by the function $G(I, \epsilon)$ and do not depend at all on the frequencies $\omega(I)$. Their chance coincidence with the resonance surfaces is "miraculous" but in the end not less so than the "miracle" of resonance in the unperturbed system. From this follows the rather discouraging result that the averaged system, just as an unperturbed system, cannot explain resonance modes.

Luckily the drift approximation possesses a property which would be a drawback in all situations except the one under investigation. This approach lends itself to the majority of trajectories but fails in the vicinity of resonance surfaces. The reason is easy to understand. The system is obtained by averaging over a rapid variable. But one of them, a resonance phase, stops being rapid especially close to the resonance surface, and necessitates a more detailed investigation.

We shall pass over to the investigation of the given resonance surface (5). We introduce new phase variables

$$\psi = A\phi; \qquad \phi = B\psi, \tag{15}$$

so that the resonance phase becomes one of the new variables, and the right-hand side of the system remains, as before, periodic (with period 2π).

Thus it is sufficient that the matrix A be integral and its determinant should equal unity

$$\det A = 1. \tag{16}$$

It is known (Molchanov, 1966) that a given set of S independent resonances can be brought to the canonical form

$$\omega_1 = 0, \ . \ . \ . \ , \qquad \omega_S = 0$$

by the substitution of variables such as (15). In the four sets of resonance vectors analyzed above, expansion into unimodular matrices A is assuredly allowed. Each of these matrices transforms the corresponding system to the canonical form (see footnote 2).

Let us now choose the convenient slow variables. As one of them, let us take the resonance frequency ν. Such a choice is allowed, provided the frequencies are functionally independent. This condition can be expressed in the form

$$\operatorname{rank}\left(\left(\frac{d\omega}{dI}\right)\right) \equiv \operatorname{rank}\left(\left(\frac{\partial \omega_i}{\partial I_k}\right)\right) = l. \tag{17}$$

For the equations describing the motion of the planetary systems, this condition is fulfilled.

The remaining slow variables we shall denote by the letter K, and we shall choose them so that the average of their real parts is zero. Here two cases may occur. If $G \equiv 0$, then the variables K may be chosen as convenient, only if together with ν they give a full system of coordinates in the space I.

If $G \not\equiv 0$, then the variables K must be stationary along drift trajectories. Such a choice is possible everywhere besides the points where drift trajectories touch the resonance surfaces. At all other points the following system is obtained:

$$\begin{aligned} d\phi/dt &= \omega(\nu, K) + \epsilon\omega(\nu, K, \phi, \psi, \epsilon), \\ d\psi/dt &= \nu + \epsilon m(\nu, K, \phi, \psi, \epsilon), \\ d\nu/dt &= \epsilon f(\nu, K, \phi, \psi, \epsilon), \\ dK/dt &= \epsilon\mathfrak{F}(\nu, K, \phi, \psi, \epsilon). \end{aligned} \tag{18}$$

In this system ψ and ν are scalar quantities, while K and ϕ are vectors, each having one component less than in the initial system.

[4] Multiple-frequency analog of the "drift-approximation," well known in physics. Particularly it leads to Laplace's method in celestial mechanics and to the method of Krylov-Bogolyubov in the theory of oscillation (Bogolyubov, 1955).

The right-hand sides are periodic over ψ and each of the ϕ's, and the function $\mathfrak{F}$ has a mean over the phases equal to zero.

Our task consists of studying the neighborhood of the coordinate surface $\nu = 0$, into which the resonance surface under investigation has been transformed in the new variables. However, not all the variables in the system (18) are yet sufficiently well chosen. The structure of these equations would indicate that ψ is a rapid variable and ν a slow one. This is what actually occurs, as a matter of fact, at large values of ν, but we are interested in the region of small ν's. Let us perform then yet another change of variables

$$\nu = \epsilon^{\alpha}\mu \tag{19}$$

to study small values of ν. Judging from the substitution in the system we may surmise that the index α may be taken equal to one-half, rendering

$$\nu = \mu\epsilon^{1/2}, \tag{20}$$

and then the equations take the form

$$\begin{aligned} d\phi/dt &= \omega(\mu\epsilon^{1/2}, K) + \epsilon\omega(\mu\epsilon^{1/2}, K, \phi, \psi, \epsilon), \\ d\psi/dt &= \mu\epsilon^{1/2} + \epsilon m(\mu\epsilon^{1/2}, K, \phi, \psi, \epsilon), \\ d\mu/dt &= \epsilon^{1/2} f(\mu\epsilon^{1/2}, K, \phi, \psi, \epsilon), \\ dK/dt &= \epsilon\mathfrak{F}(\mu\epsilon^{1/2}, K, \phi, \psi, \epsilon). \end{aligned} \tag{21}$$

The structure alone of this system makes clear the main features of what takes place in the neighborhood of the resonance surface. The rapid variable ψ certainly ceases to be rapid, but neither does it become a real slow one, since it varies with the velocity $\epsilon^{1/2}$. The slow variable ν engenders a new variable μ, changing with the same speed as ψ. This variable differs from ν by the characteristic scale factor $\epsilon^{1/2}$, which determines the extent of the resonance zone. It follows that there exist three types of variables—rapid phases, slow integrals, and intermediate "semirapid" resonance variables.

This division according to the scale of speed of variation permits full investigation of the structure of the resonance zone. In order to study the system (21) let us once again refer to the drift approach, but subject only the "intrinsically" rapid phases ϕ to averaging, leaving the resonant phase ψ as a slow variable. The right parts of the resulting system are related by the factor $\epsilon^{1/2}$. The introduction of the semi-slow time

$$\tau = \epsilon^{1/2}t \tag{22}$$

converts the system[5] into the form

$$\begin{aligned} d\psi/d\tau &= \mu + \epsilon^{1/2}m(\mu\epsilon^{1/2}, K, \psi, \epsilon), \\ d\mu/d\tau &= f(\mu\epsilon^{1/2}, K, \psi, \epsilon), \\ dK/d\tau &= \epsilon^{1/2}\mathfrak{F}(\mu\epsilon^{1/2}, K, \psi, \epsilon). \end{aligned} \tag{23}$$

Setting ϵ equal to zero (what else is a small parameter good for?) we obtain the system

$$\begin{aligned} d\psi/d\tau &= \mu, \\ d\mu/d\tau &= f(K, \psi), \\ dK/d\tau &= 0, \end{aligned} \tag{24}$$

which may be regarded as unperturbed in relation to the system (23). It follows immediately then from (24) that

$$K = \text{const.}, \tag{25}$$

which means much in the solution of the two first equations for μ and ψ, wherein K enters as a parameter.

It is of great interest that these two equations coincide with Hamilton's equations of one-dimensional motion, the phase ψ playing the role of coordinate, and μ the role of impulse. The field is given by the function $f(K, \psi)$ which also determines the structure of the resonance zone. The problem of one-dimensional motion has been studied in detail, and it is possible to derive the following results of significance for what follows. If the function $f(K, \psi)$ does not vanish anywhere, then the variable μ increases (or decreases) monotonically along the given trajectory and all integral curves cut the resonance surface, as in Fig. 1. If on the other hand, the function $f(K, \psi)$ vanishes, then in the vicinity of the resonance surface a vortex region occurs, consisting of closed trajectories. As a matter of fact, the function $f(K, \psi)$ is periodic in ψ and for fixed K has an even number of zeros. Every one of the zeros on the axis $\mu = 0$ is a stationary point of the

[5] In the context, the letters f, $\mathfrak{F}$, and m are used to denote general functional relations and may have various interpretations in various formulas.

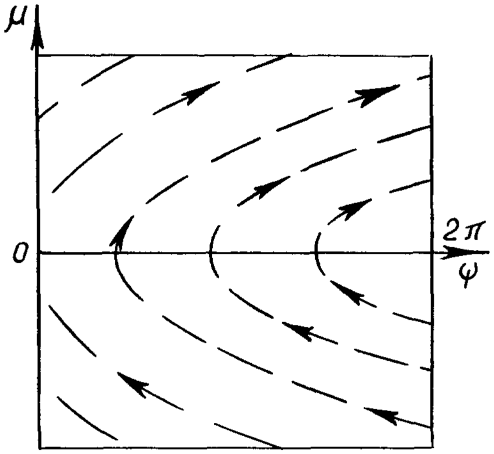

FIG. 1. Weak resonance.

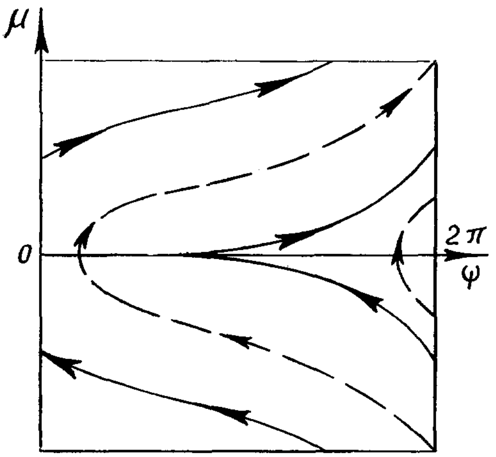

FIG. 2. Limits of strong resonance.

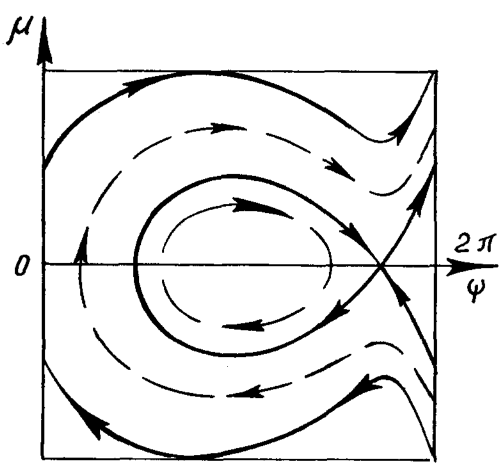

FIG. 3 Strong resonance: dissipative case.

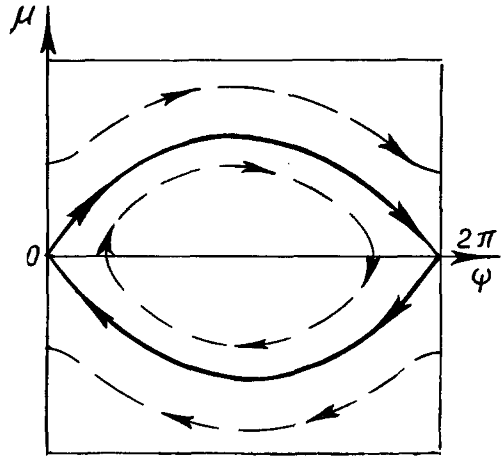

FIG. 4. Strong resonance: conservative case.

system (24); while exactly half of them, stationary points of the same type as the center, are surrounded by closed trajectories. Each such region is bounded by a loop-shaped boundary, beginning and ending in a stationary saddle point, in which case the two other lobes of the boundary either also close into a loop or continue to infinity, one of them receding and the other approaching. In the simplest case the resulting figure is reminiscent of a figure eight with one loop broken. The point at which it intersects itself is a stationary saddle point. The center is found within the complete boundary loop and the broken loop yields two boundaries—approaching and receding.

All the remaining integral curves cut the resonance surface, deviating little from the integral curves of the drift approximation. In this case, in the neighborhood of some resonance surfaces, a vertical zone may arise with extent of the order of $\epsilon^{1/2}$. Let us evaluate the fraction of the phase space occupied by the points of the eightfold (as in the solar system) resonance. This fraction as to order of magnitude is equal to $(\epsilon^{1/2})^8 = \epsilon^4$. Adopting the value of ϵ characteristic of the solar system as 1.34×10^{-3}, we obtain[6]

[6] Direct multiplication of the quantities $\Delta\omega/\omega$ for all the planets (except. of course, Jupiter!) yields a value smaller than 10^{-20}, which unequivocally corroborates the validity of the evaluation (26) and even points to the fact that the system is situated well inside the resonance zone.

evaluation for phase volume

$$\triangle V/V \sim 3 \times 10^{-12}. \qquad (26)$$

The resonance zone and the supplements to it comprise a fragmentation of phase space which is invariant relative to the system (24). Each trajectory (generally speaking) lies in its entirety in one of these two regions. For this reason in particular, even a more accurate version of the drift approximation is not capable of rationally explaining the resonance structure. Were every one of the 10^{11} stars of our galaxy to have a planetary system, it follows from the evaluation (26) that even then, in the absence of dissipation, the solar system would remain unique in its properties.

This statement may be generalized. A given model (in particular, the theory of the perturbations of Hamiltonian systems), for which there exists a mentioned division, appears too crude for clarification of the resonance structure. It is noteworthy that particularly mathematicians expanding the theory of perturbations of Hamiltonian systems (see, for example, Arnold, 1963) clearly understand, it seems, the abstractness of such a model.

In this way, a more precise version of the drift approximation enables one to understand why the system cannot deviate from a resonance surface. It remains a mystery, however, how it got in resonance zone. It seems that this is in general impossible to understand without taking into account nonconservative factors.

V. The Thin Structure of Resonant Zones ($t \sim \epsilon^{-1}$) Drawing into Resonance

The foregoing investigations make clear the special role of resonance zones. In the models treated resonance configurations form an insignificantly small, closed, privileged group, into which entry is forbidden to outsiders. The fact that the solar system belongs to this select group indicates that an explanation of the resonance structure in terms of evolution is inescapable. The refusal to attempt such an explanation would be equivalent to a return to the Newtonian concept (Levin, 1964) of the uniqueness of the solar system.

However, the system (24), an analysis of which brings one to such troublesome conclusions, is obtained by the averaging and transition to the limit from a full system (18)—that is, in the final analysis, by ignoring certain small terms. It follows that the hope of investigating the breach within the isolation of resonance zones can only be based upon the legitimacy of our neglect of the terms. Let us define, therefore, the structure of the field of integral curves, regarding the more important neglected terms.

Let us look at the model system (23) and abandon terms of the order of $\epsilon^{1/2}$: we obtain

$$\begin{aligned} d\psi/d\tau &= \mu &&+ \epsilon^{1/2} m(K,\psi), \\ d\mu/d\tau &= f(K,\psi) &&+ \epsilon^{1/2}\mu f_1(K,\psi), \\ dK/d\tau &= &&+ \epsilon^{1/2}\mathfrak{F}(K,\psi). \end{aligned} \qquad (27)$$

In order to understand the qualitative picture a special case is useful, $m = 0$, $f_1 =$ const., and $f(K, \psi) \equiv f(\psi)$. Such a system of the form

$$\begin{aligned} d\psi/d\tau &= \mu \\ d\mu/d\tau &= f(\psi) - \delta\mu \end{aligned} \qquad (28)$$

can be reduced to one equation

$$\frac{d^2\psi}{d\tau^2} + \delta\frac{d\mu}{d\tau} - f(\psi) = 0, \qquad (29)$$

which, as is known, describes motion with friction in a field of force $f(\psi)$.

It follows that the neglect of the deleted terms brings into existence a new aspect. The resonance zone loses its isolated character and begins to take place in the general movement. Even the second loop of the boundary, which in system (24) began and ended in the saddle point, now breaks. One of the ends remains at the saddle point, and the other winds around the second stationary point, the former center of the system (24). If the approaching end remains in the saddle, as in diagram 6, then between the approaching boundaries there appears a thin flow of phase volume $\epsilon^{1/2}$, running into the resonance

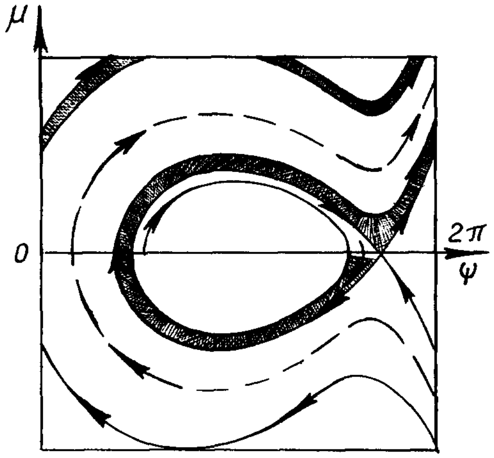

FIG. 5. Instability.

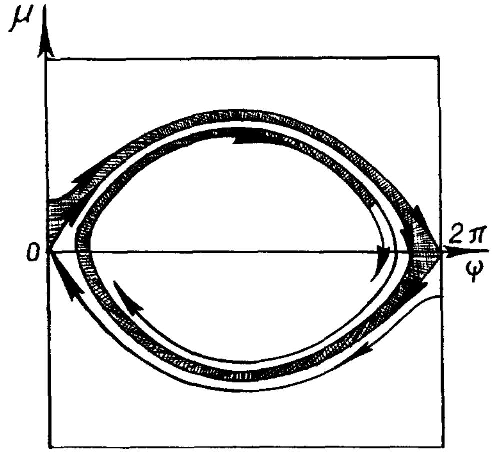

FIG. 8. Full stability.

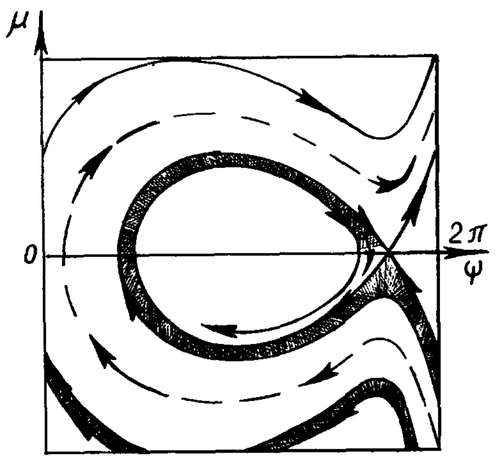

FIG. 6. Stability.

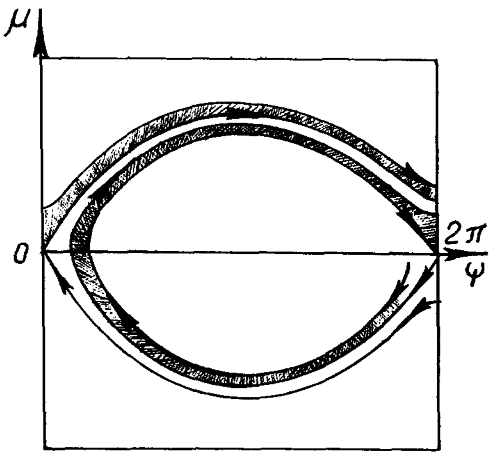

FIG. 9. Instability with flow.

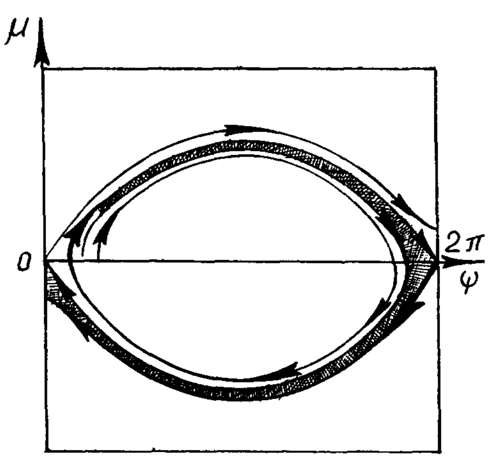

FIG. 7. Full instability.

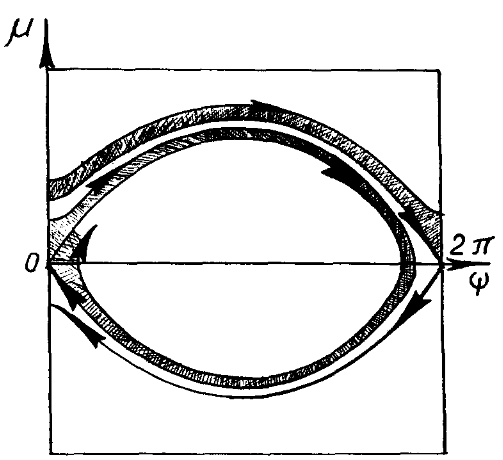

FIG. 10. Stability with flow.

zone. In the opposite case the receding boundary is separated and the flow runs out of the resonance zone.

With the help of the drift approximation it would be possible to carry out a full quantitative investigation of the system (27). However, the formulas obtained are rather cumbersome and, in the general case, not very informative. We shall therefore limit ourselves to the remark that the drift approximation works well only within the loop of the boundary—on the closed trajectories of the system (24)—and that the investigation itself is more usefully carried out in passing over to first integrals and the phase variable of systems (24); similarly in the general case it is useful to pass from system (1) to system (2).

In what follows a qualitative picture of the motion in the system (27) is important, arising as it does from the discontinuity of the change of variables K. It can be produced by moving slowly (with speed $\epsilon^{1/2}$) the figure 4 perpendicularly to the plane of the figure and as slowly as possible deforming the diagram of the integral curves. It is very important, that the movement along the K axis should take place with different (even though small) speed on different ovals. During the movement down the axis K, stability may change to instability, the loop of the boundary may shrink to a point, and the zone of fundamental resonance may disappear. This whole quite peculiar picture becomes yet more complex if the number of variables is greater than unity.

Comparison of the results of the Sections IV and V is instructive. In the main term we obtained Hamilton's system and the strict isolation of the resonance zone. The following approximation shows itself to be in general nonconservative, and the emigration (or immigration) is unexpectedly large—the width of the flow is of the order of $\epsilon^{1/2}$, and not ϵ.

Evidently this phenomenon is a general law. Hamilton's systems determine the main terms of the motion, but over short times. Nonconservative connections are negligibly small for short periods, but they are responsible for determining the evolution of the system.

VI. The Evolutionary Formation of a Resonance Structure

Above we examined separately three fundamental elements indispensable for the evolutionary explanation of resonance structure—zones of fundamental resonance, the drift approximation, and the drawing inward of the resonance zone.

Let us now follow through the full picture of evolutionary process, bringing into existence the resonance structures independently of the initial conditions. At the initial moment[7] let us choose at random the point I and follow through the fate of all possible systems, differing by the initial values of the phase ϕ. The points depicting these systems fill the torus τ in phase space, over which a rapid motion exists, superimposed on a slow drift over the space I.

After some time the drift trajectory cuts the resonance surface. Part of our torus (very small,[8] as we saw) turns out to be drawn into the resonance surface. The main part of the torus will be drawn along further by the drift motion and will drive down on the following surface, where the detachment of an alternate position takes place. For this reason sooner or later the given trajectory becomes pulled into the resonance zone. If at the initial moment the systems uniformly disintegrate over all phase space, then as the result of evolution they will congregate in volumetrically trivially small resonance zones. A picture of evolutionary maturity arises, directly opposed to the ergodic concepts, characteristic for the initial stages of evolution.

However, the conclusiveness of the representations arrived at depends in a decisive way upon the amount of time necessary for the creation of a resonance structure from the initial amorphous configuration. Let us turn next to the evaluation of the time of evolution I. First of all we notice that Hamilton's perturbation does not usually lead to

[7] Which we choose at will during, let us say, the first million of the five billion years which contemporary science alotts to the present situation.

[8] See in what follows, the text that follows formula (33).

the creation of a drift transverse to the resonance surface. For this reason it is necessary to discount two factors—Hamilton's perturbations of magnitude ϵ calling for the formation of resonance zones and dissipative factors of magnitude δ, leading to the transverse drift and the draining into the resonance zones. Motion within these zones is carried out, as we saw, along closed trajectories with period of the order of $\epsilon^{-1/2}$. Dissipation turns these curves into spirals, decreasing the amplitude to the order of magnitude $\delta\epsilon^{-1/2}$ at each revolution. The phase volume of the flow running in will be of the same order

$$\triangle V/V \sim \delta/\epsilon^{1/2}. \tag{30}$$

Strictly speaking the resonance surfaces are everywhere concisely laid out. However, our interest lies only in the fundamental resonance having small coefficients (see footnote 1). It follows then that the distances between resonance surfaces are of the order of unity. This means that the "mean free time" between surfaces (or the time of return during multiple crossing of the same surface) is evaluated as the reciprocal of the drift velocity

$$T_d \sim \delta^{-1}. \tag{31}$$

Increasing the quantity of intersections $V/\triangle V$ to the mean free time, indispensable to full "fragmentations" of the space, we obtain an evaluation of the time of evolution

$$T \sim \delta^{-2}\epsilon^{1/2}. \tag{32}$$

In a single frequency case the calculations may be carried out more accurately and the estimate obtained is remarkably better

$$T \sim T_d \sim \delta^{-1}. \tag{33}$$

The cause of such a fundamental divergence lies in the fact that where the dissipative perturbation is superimposed on the resonance zone which obeys Hamiltonian relations, the dimension of the inflowing stream turns out to be finite, and not infinitely small. However, in the general case the fact is not yet proven and in what follows it is suitable to satisfy ourselves with a rough evaluation (32).

The careful reader will have certainly noticed already that everything that has been derived uses little of the specific properties of the solar system, that even the Newtonian potential has not once been brought to mind. In justification, the author can only fall back on the words of Eddington (1933). The principal laws of gases hold, not because a gas is made "that way," but because it is made "just anyhow." The analysis carried out leads to the conclusion that, without discounting the dissipative factors, it is not possible to explain the resonance structure of the solar system. But in the presence of weak dissipative factors, a given oscillating system in the end takes on a resonance structure. It follows that the resonance structure is, in the main part, a developing category—it is the result and mark of evolutionary maturity of the system.

VII. Possible Factors in the Evolution of the Solar System

Since at the present time dissipative terms are without doubt much smaller than Newtonian interactions, and are generally not considered, they have been considerably less studied. It is possible, all the same, to make some quantitative statements directly on the basis of formula (32).

Of decisive importance to the argument is the "age" of the solar system, which we shall take as being five billion years old, that is (in measurement of the characteristic time of the system—the period of Jupiter)

$$T = 4 \times 10^8. \tag{34}$$

We find the magnitude of δ from the condition that basically the evolution was completed in the first one hundred million years.

Inserting in (32) $\epsilon = 1.34 \times 10^{-3}$ and $T = 10^7$ we obtain $\delta = 6 \times 10^{-5}$. Let us accept, further, that δ decreases by a factor of 1½ or 2 in the time equal to the time of evolution. The following hypothetical picture emerges from the results.

During the first hundred million years the dissipative factor (twenty times smaller than the mutual perturbations of the planets or smaller by four orders than the main gravitational field of the Sun) safeguards the construction of a resonance structure from a randomly chosen initial configuration. In the following forty periods (or the same length) this factor diminishes one-million-fold, becoming completely imperceptible.

It is evident that the calculations we have made are quite arbitrary, and demonstrate more a deferential delight in the age of the solar system and the properties of exponents than the real magnitude. However, they demonstrate also that it is quite possible with quite modest assumptions concerning the size of the dissipative terms to deduce that evolution of the resonance system is inescapable.

The question of the reality of the evolutionary terms is very interesting. We shall demonstrate two possible contentions. The first of them—purely friction arising in the motion of already almost finished planets in the remnants of a gas-dust cloud. This example illustrates an important characteristic—a term which calls for evolution unmistakably weakening and "dying out" in the process of evolution, which permits the system to remain in its complexly organized configuration. It is understood that systems exist in which the dissipative factor does not weaken, but these belong rather to the group of "has beens" and they interest us but little at the moment.

Another process may play an important role—the process of ejection of matter. This is, so to speak, the "cometary" variant as opposed to, or rather complementary to, the "asteroidal" variant. The cometary variant is methodically more interesting. It shows that evolutionary problems may be set up in the framework of purely Hamiltonian[9] equations, in which case dissipation must be understood as a dispersion, an escape of particles from the system.

Evaluations (see Levin, 1960) of the ejection of mass point out the great importance of the role played by this process in the beginning stage of evolution of a gas-dust cloud.

The formation of "repeated" vibrating structures could take place even under the influence of other dissipative factors. The decisive role played by tidal friction in the evolution of the Earth–Moon system is not doubted by anyone. As a result the Moon turns one side always towards us, clearly demonstrating the properties of 1:1 resonance. Such resonances are characteristic of systems in evolution whose dissipative factors are large enough that they do not allow the system to "build up" in "shallow" resonances. In the case where conservative perturbations dominate, resonances with large coefficients may be produced. In this way, 1:7 resonance of the frequencies of Uranus and Jupiter brings to mind that these may be the "youngest" of the resonances of the solar system, arising on the background of already decaying dissipation. It is possible to think that the trio of far planets with good 1:2:3 resonances came into resonance with the remainder of the solar system quite late.

During consideration of this work B. J. Levin suggested the hypothesis that the plane resonance structure in its rough form already was in the process of accumulating the planets, and in the process of further evolution a comparatively thin "branch" arose. It is most probable that the evolution of the solar system did in fact progress in this way.

It is worthwhile, however, to underline that the conclusion about the inevitability of resonance does not depend on the mechanism of formations of an oscillating system, and is based only upon its sufficiently long duration. Thus even if the sun were to capture, not necessarily at one time, already formed planets, five billion years would suffice, it seems, to form a resonance structure, possibly at the cost of ejecting some of the material.

[9] Sometimes one encounters the opinion that the Lionville theorem precludes the possibility of evolution in Hamiltonian systems. The example of the Hamiltonian function, $H = pq$, shows that evolution is possible. The prejudice is seemingly founded on the custom of thinking that the level surfaces of a Hamiltonian function are closed.

Acknowledgments

The author is grateful for helpful discussion to the members of the seminars of G. H. Duboshin and B. Yu. Levin. The author is also indebted to M. V. Keldysh who brought to his attention the work of

N. G. Chetajev, where for the first time, really, the idea of the resonance nature of the quantization conditions has been expressed.

References

Arnold, V. I. (1963). "Problems of the Motions of Artificial Bodies," p. 7.

Birkhoff, G. D. (1941). "Dynamical Systems."

Bogolyubov, N. N., and Mitropolsky, U. A. (1955). "Asymptotical Methods in the Theory of Nonlinear Oscillations."

Chetayev, N. G. (1958). *Prikl. Math. Mechanika* **22,** 487–489.

Eddington, A. S. (1933). "The Mathematical Theory of Relativity." Cambridge Univ. Press, London.

Levin, B. Yu. (1960). *Voprosy Kosmogonii* **7,** 55–58.

Levin, B. Yu. (1964). "Origin of the Earth and Planets," 4th ed. Moscow.

Molchanov, A. M. (1963). "Symposium on Dynamics of Satellites," p. 40. Springer Verlag, Berlin.

Molchanov, A. M. (1966). *Dokl. Akad. Nauk SSSR* **168,** 1966.

Schmidt, O. U. (1950). Four Lectures on the Theory of the Origin of the Earth.

Schroedinger, E. (1948). "Statistical Thermodynamics," p. 63.

Whipple, F. L. (1948). "The Earth, Moon, and Planets." Harvard Univ. Press, Cambridge, Massachusetts.