STABILITY THEORY IN CELESTIAL MECHANICS

J. MOSER

Courant Institute of Mathematical Sciences, New York University, New York, N.Y. 10012, U.S.A.

Abstract. This expository lecture surveys recent progress of the stability theory in Celestial Mechanics with emphasis on the analytical problems. In particular, the old question of convergence of perturbation series are discussed and positive results obtained, in the light of the work by Kolmogorov Arnold and Moser. For the three body problem, classes of quasi-periodic solutions and doubly asymptotic (or homoclinic) orbits are discussed.

1. Introduction

Over the years celestial mechanics has presented scientists with a variety of difficult problems challenging to astronomers, numerical analysts and mathematicians. One of the most puzzling questions is related to the stability of the solar system and the long time behavior of the solutions of the *n*-body problem which has been attacked again and again over the last hundred years. The analytical difficulties of these problems has to do with the presence of the notorious 'small divisors'. In the last decade definite progress has been made in this direction and the main goal of this lecture is to report on some of these advances and their significance for celestial mechanics. Speaking as a mathematician the emphasis will naturally be on the theoretical aspects and much remains to be done in the applications of these results to the realistic problems.

We want to point out that the recent work has led to a new concept of stability. It is the conventional view that one can delineate some 'blobs' or open regions in phase space in which the solutions remain bounded or stable, while outside of such regions the solutions escape or behave unboundedly. A sharp mathematical formulation and related results were developed by Liapounov. His work refers to dissipative systems, however, and is not applicable to the conservative systems of celestial mechanics. The recent mathematical work in this area has shown that for Hamiltonian systems this crude picture has to be replaced by another model: One finds complicated Cantor sets, which we may compare with a sponge, in which the solutions are stable and bounded for all time while the solutions lying in the many fine holes of the sponge may gradually seep out and become unstable. The filament of these holes is connected and give rise to a slow diffusion while the majority of the solutions belong to the solid part of the sponge consisting of stable solutions.

From a physical point of view this model is obviously hard to accept, but one cannot escape these conclusions if one idealizes the problem mathematically and studies the motion for *all* time and not only for a reasonable finite time interval. In fact, the idealization goes further: We are not talking about the motion of the planets under realistic forces but of rigorous solutions of the *n*-body problem taking into account only Newton's force law and referring to mass points with some smallness

restrictions on the masses. However, all these restrictions are less severe since the theory to be discussed allows for small additional forces as long as they are conservative and time independent, as well as for spherical homogeneous masses.

The complicated structure of the stability region is due to resonance phenomena which occur between the various frequencies $\omega_1, \omega_2, \ldots, \omega_n$. We speak of resonance, whenever one has a relation $\sum_{k=1}^{n} j_k \omega_k = 0$ with integer $j_1, j_2, \ldots, j_n$ which do not all vanish. While for practical purposes only resonances of small order matter, i.e. when $0 < |j_1| + |j_2| + \cdots + |j_n|$ is reasonably small, for the mathematical problem of long time behavior *all* these resonances affect the motion over long time. But even the definition of frequency requires the knowledge of the solution over all time, and a frequency is really defined by a limit for $t \to \infty$. A solution, oscillating for a very long time, but finally escaping, can at most approximately, but never rigorously, be assigned frequencies. On the other hand the above definition of resonances requires the exact knowledge of frequencies and we find ourselves in a vicious circle.

The way out of this dilemma is to give up the *initial value problem* which asks for the long time behavior of a solution for given initial data and instead to ask for solutions which are almost periodic with prescribed frequencies. Geometrically this requires that the solutions lie on a torus in phase space and gives rise to a *boundary value problem*. Historically, the expansion methods described by H. Poincaré are precisely based on such a boundary value problem, the boundary condition expressing the periodicity of the unknown function.

Therefore we begin with the classical approach to approximate the orbits by perturbation series in terms of trigonometric expansions. Poincaré's work put these techniques on a firm foundation, but it came as a disappointment to many that his investigations indicated that these series expansions were divergent. In fact, this was the starting point of 'asymptotic expansions' which have proved so useful in fluid dynamics and other fields of mathematical physics. It is perhaps the most remarkable recent discovery that some of these classical series expansions in celestial mechanics are actually convergent and give rise to a rigorous description of solutions of the n-body problem valid for all time. The convergence proof is based on new techniques in functional analysis which are due to Kolmogorov, Arnold and the author (Moser, 1973*). We will describe more precisely the type of expansions whose convergence can be established and explain why this result is not in contradiction to Poincaré's assertion about divergence.

2. Convergence of Perturbation Series

A. VARIABLE FREQUENCY EXPANSION

The main idea of the classical expansion techniques is *not* to solve in a straightforward way the initial value problem of integrating the equations with prescribed initial data. Instead we seek such initial conditions for which the solutions are pe-

* We refer to this booklet for further information.

riodic or quasi-periodic, i.e. remain on a torus in phase space. This leads to a boundary value problem – mathematically speaking – which has very different features from the initial value problem. From the practical point of view this amounts to the construction of a reference orbit.

Consider a Hamiltonian,

$$H(x, y, \varepsilon) = \sum_{\nu=0}^{\infty} \varepsilon^{\nu} H^{(\nu)}(x, y),$$

where $H^0(x, y) = F(x)$ independent of y, $H(x, y, \varepsilon)$ of period 2π in $y_1 \dots y_m$, and real analytic in x, y and ε, for $x \in D$, all real y and $|\varepsilon| < \varepsilon_0$. Roughly, the goal of Lindstedt's method is to find

$$S(\xi, y, \varepsilon) = \sum_{\nu=0}^{\infty} \varepsilon^{\nu} S^{(\nu)}(\xi, y),$$

such that

$$H(S_y, y, \varepsilon) = \Phi(\xi, \varepsilon)$$

is independent of η and where S_ξ and S_y must be real analytic and of period 2π in $y_1 \dots y_m$ and

$$\det S_{\xi y} \neq 0.$$

A solution to this problem would mean that the canonical transformation $(x, y) \to (\xi, \eta)$,

$$\eta = S_\xi, \qquad x = S_y,$$

leads to a Hamiltonian independent of η, showing that the system is integrable for all small $|\varepsilon|$. It is well known that this is not possible, in general, and we require only that

$$H(S_y, y, \varepsilon) = \Phi(\xi, \varepsilon) + O(|\xi|^2); \qquad \Phi(\xi, \varepsilon) = \Phi_0(\varepsilon) + \sum_{k=1}^{n} \Phi_k \xi_k, \tag{1}$$

so that the new Hamiltonian, say $\tilde{\Phi}(\xi, \eta, \varepsilon)$, satisfies

$$\tilde{\Phi}_\xi = \Phi_\xi, \qquad \tilde{\Phi}_\eta = 0 \quad \text{for} \quad \xi = 0,$$

and, thus

$$\xi = 0, \qquad \eta = \Phi_\xi(0, \varepsilon) t + \eta(0)$$

gives rise to particular solutions of the system. These solutions are called quasi-periodic, with frequencies

$$\omega_k(\varepsilon) = \Phi_{\xi_k}(0, \varepsilon), \quad (k = 1, \dots, n),$$

which are, in general, dependent on ε.

We describe three types of standard expansion methods: (i) a variable frequency expansion, (ii) a fixed frequency expansion, and (iii) a fixed frequency ratio expansion.

For (i) assume that x^0 can be chosen so that the quantities,

$$\omega_k = F_{x_k}(x^0),$$

are rationally independent, and set

$$
\begin{aligned}
S^{(0)}(\xi, y) &= (x^0 + \xi, y), \\
S^{(\nu)}(\xi, y), &\quad \text{period } 2\pi \text{ in } y \text{ and linear in } \xi \text{ for } \nu \geqslant 1.
\end{aligned}
\tag{2}
$$

Inserting the series of S into Equation (1) we get by comparison of coefficients of ε^ν,

$$(F_x(x^0 + \xi), S_y^{(\nu)}) = \Phi^{(\nu)} + g(\xi, y) + O(|\xi|^2),$$

where $g(\xi, y)$ is a known function, determined by $S^{(0)}, \ldots, S^{(\nu-1)}$, $\Phi^{(0)}, \ldots, \Phi^{(\nu-1)}$. Separating the coefficients of ξ_k by setting

$$
g = g_0(y) + \sum_{k=1}^{n} g_k(y)\,\xi_k + O(|\xi|^2), \qquad S^{(\nu)} = S_0^{(\nu)}(y) + \sum_{k=1}^{n} S_k^{(\nu)}(y)\,\xi_k,
$$

we get

$$
\begin{aligned}
(\omega, S_{0y}^{(\nu)}) &= \Phi_0^{(\nu)} + g_0(y), \\
(\omega, S_{ky}^{(\nu)}) + (F_{x_k x}(x^0),\ S_{0y}^{(\nu)}) &= \Phi_k^{(\nu)} + g_k(y).
\end{aligned}
\tag{3}
$$

It is well known how to solve such equations for the functions $S_k^{(\nu)}$ and constants $\Phi_k^{(\nu)}$ ($k = 0, 1, \ldots, n$) with the help of Fourier series, provided the frequencies $\omega = (\omega_1, \ldots, \omega_n)$ are not only rationally independent but also satisfy a condition of the type,

$$|(j, \omega)| \geqslant c|j|^{-\mu},\tag{4}$$

for all integer vectors $j = (j_1, \ldots, j_m) \neq 0$ with some positive constants c and μ. We have to check that the mean value of the right-hand side vanishes, which is achieved by choice of $\Phi_0^{(\nu)}$ and $\Phi_k^{(\nu)}$, thus determining $\Phi(\xi, \varepsilon) = \Phi_0 + \sum_k \Phi_k \xi_k$ in (1). The resulting expansion approximates a motion with frequencies,

$$\omega_k(\varepsilon) = \Phi_k(\varepsilon), \quad (k = 1, \ldots, n),$$

which generally depends on ε. Therefore we refer to it as variable frequency expansion. Incidentally, S is determined up to an additive function $\sigma(\xi, \varepsilon)$ independent of y, and S can be normalized uniquely by the requirement that

$$\int_{T^n} (S - S^0)\,dy = 0.$$

This type of expansion* was described in Poincaré's *Méthodes nouvelles de la mécanique céleste*, Vol. 2, and attributed to Lindstedt.

B. FIXED FREQUENCY EXPANSION

The above expansion has the shortcoming that the irrationality conditions (4) get destroyed for $\varepsilon \neq 0$. But it is easy to modify the above procedure so as to have the frequencies ω_k independent of ε; in fact, such a procedure was also described in Poincaré's *Méthodes nouvelles de la mécanique céleste*, Vol. 2. One simply replaces S by

$$S + \sum_{k=1}^{n} \alpha_k(\varepsilon) \, y_k = S + (\alpha, y); \qquad \alpha(0) = 0.$$

Then the above procedure is the same except that $S^{(\nu)}$ has to be replaced by $S^{(\nu)} + (\alpha^{(\nu)}, y)$, so that Equations (3) take the form

$$(\omega, S_{0y}^{(\nu)}) + (\omega, \alpha^{(\nu)}) = \Phi_0^{(\nu)} + g_0(y),$$
$$(\omega, S_{ky}^{(\nu)}) + (F_{x_k x}(x^0), \; S_{0y}^{(\nu)} + \alpha^{(\nu)}) = g_k(y),$$

(5)

where we set $\Phi_k^{(\nu)} = 0$ to assure that the frequencies are independent of ε. To make the mean values equal on both sides we use the constants $\alpha^{(\nu)} = (\alpha_1^{(\nu)}, \ldots, \alpha_n^{(\nu)})$ in the last n equations, assuming that

$$\det(F_{xx}(x^0)) \neq 0,$$

and $\Phi_0^{(\nu)}$ in the first equation. Then the above equation can again be solved uniquely if we require

$$\int_{T^n} S^{(\nu)} \, dy = 0, \quad \text{for} \quad \nu = 1, 2, \ldots,$$

(6)

yielding the fixed frequency expansion.

C. FIXED FREQUENCY RATIO EXPANSION

Actually, in order to preserve the irrationality conditions for small ε it suffices to keep $\omega_k(\varepsilon) = \lambda(\varepsilon) \, \omega_k(0)$ proportional to themselves, so that the frequency ratio is independent of ε. We use the extra freedom to make

$$\Phi(0, \varepsilon) = F(x^0),$$

independent of ε, ensuring that the resulting solutions lie on a fixed energy surface $H(x, y, \varepsilon) = F(x^0)$. To obtain such an expansion we set

$$\Phi_0^{(\nu)} = 0; \qquad \Phi_k^{(\nu)} = \lambda^{(\nu)} \omega_k, \quad \text{for} \quad \nu = 1, 2, \ldots.$$

The resulting equations for $S^{(\nu)}$, $\alpha^{(\nu)}$ and $\lambda^{(\nu)}$ are, instead of (5), of the form

$$(\omega, S_{0y}^{(\nu)}) + (\omega, \alpha^{(\nu)}) = g_0,$$
$$(\omega, S_{ky}^{(\nu)}) + (F_{x_k x}(x^0), \; S_{0y}^{(\nu)} + \alpha^{(\nu)}) = \lambda^{(\nu)} \omega_k + g_k.$$

To balance the $n + 1$ mean values of these equations we use the $n + 1$ constants

$\alpha_1^{(\nu)}, \ldots, \alpha_n^{(\nu)}$ and $\lambda^{(\nu)}$, and need the assumption that

$$\det \begin{pmatrix} 0 & \omega \\ \omega^T & F_{xx}(x^0) \end{pmatrix} \neq 0.$$

Under this assumption again $S^{(\nu)}$, $\alpha^{(\nu)}$ and $\lambda^{(\nu)}$ are uniquely determined if normalized by (6).

These expansion procedures have been known since the last century. Also the solvability of Equations (3) or (5) at each step has been known for a long time under an irrationality condition (4) showing that all coefficients of ε^ν are well defined by convergent Fourier series. However, the convergence question of such series for small positive $|\varepsilon|$ is a much more difficult problem as it relates to a nonlinear system of differential equations. The results of the work by Kolmogorov, Arnold and the author imply, in fact, that the fixed frequency expansion and the fixed frequency ratio expansions both converge for $|\varepsilon| < \varepsilon_0$ where ε_0 is a positive number depending on the given Hamiltonian and the constants c and μ in (4), but not on ω itself. Thus these expansions do not only yield an approximation for solutions useful for long time intervals but the existence of solutions valid for all real time which can be used as reference orbits. On the other hand the variable frequency expansion can, in general, not be expected to converge because of the violation of the nonresonance conditions (4) for arbitrary small $|\varepsilon|$. Thus the crucial feature for the convergence is that the non-resonance conditions are maintained under perturbation. At first such conditions can be imposed only on the formal expansion; the convergence of the resulting series ensures that these formal expansions actually belong to existing solutions. Incidentally, the convergence proof does not succeed with the conventional majorant method but a new rapidly convergent iteration process. A posteriori, one can identify the resulting solutions with those obtained by formal expansion (Moser, 1967). Presently the relevant existence theorems are being derived from an abstract implicit function theorem in Banach spaces (Zehnder, 1974).

3. Applications

The above results, and their extensions, have a great many applications in celestial mechanics. We list a few:

(a) The plane restricted three-body problem. It is possible to give bounds for the eccentricity valid for all time, i.e. given $\delta > 0$ one can establish for the eccentricity ε,

$$\varepsilon(0) - \delta \leqslant \varepsilon(t) \leqslant \varepsilon(0) + \delta \quad \text{for all} \quad t,$$

if the mass ratio μ is small enough. The periodic solutions of the first kind can be shown to be stable for small μ if resonances of order $\leqslant 3$ are avoided. The stability of the equilibria L_4 and L_5 can be established, provided one has linear stability and excludes 3 more exceptional values for μ.

More generally, one finds stability criteria for equilibria and periodic solutions for systems of two degrees of freedom.

(b) For systems of higher degrees of freedom such stability criteria have not been found except in the trivial case where the Hamiltonian is positive or negative definite. However, another weaker concept of 'stability in measure' can be verified: An equilibrium solution is called stable in measure if for any $\varepsilon > 0$ any neighborhood (in phase space) contains an open set V, containing the equilibrium, and a measurable set $S \subset V$ invariant under the flow such that

$$m(V - S) < \varepsilon m(V).$$

Here m denotes the Lebesgue or Liouville measure in phase space. In other words, most solutions starting in V will remain in V for all time, while no assertion is made about the exceptional set $V - S$. This is decidedly a weaker concept of stability than the usually accepted one due to Liapounov where it is required that $S = V$. But it has the advantage that one can give a stability criterion in terms of inequalities on finitely many coefficients of the Hamiltonian. Moreover, one can choose the set S in such a way that all solutions in S are given by quasi-periodic solutions obtained by the series expansions described above.

Of course, there are a great number of different stability concepts but it is a basic requirement of stability that the entire orbit changes little if the initial data are changed by a small amount. In addition, it is of importance that the orbit changes little under small perturbation of the system of the differential equation, at least if one restricts oneself to the class of Hamiltonian systems of differential equations. This latter requirement reflects simply that the differential equations, or some parameters in them, are not known with absolute precision but only to a limited degree of accuracy. To put this into more mathematical terms, we should like to aim at a stability concept in which one can predict the orbit within narrow bounds for initial data in an open set V in phase space and all Hamiltonian systems in an open set $\mathscr{W}$ of Hamiltonian functions in an appropriate topology. The latter condition would be satisfied, for example, if our desired stability criterion imposes only finitely many inequalities on finitely many derivatives of the Hamiltonian and is independent of the 'tail' of the expansion.

To summarize the result presently known: There are such stability criteria for Hamiltonian systems of $n \leqslant 2$ degrees of freedom, but for $n \geqslant 3$ degrees of freedom it is at present not possible to give a stability criterion of the above type except in the trivial case of an equilibrium with definite Hamiltonian. But one does have such criteria for stability in measure.

(c) The results are applicable to the three-body problem and yield Cantor sets of positive measure containing quasi-periodic motions, provided two masses are much smaller than the third mass. The solution constructed are close to circular orbits with zero inclination and a large ratio of the major axes. Thus the solutions in the set S are bounded for all time and never experience collisions. Since S has positive measure one concludes that the flow of the three-body problem is not ergodic. On the other hand the complement of S is a connected set and solutions arbitrarily close to one in S may become unbounded. This result is due to Arnold (1963) who also considered generalizations to the n-body problem for $n > 3$.

4. Random Motions, open Problems

The weakened type of stability concept discussed above brings up the question about the behavior of the solution in the exceptional sets which are frequently called regions of instability. Although little is known about these solutions there are results showing the presence of orbits of very erratic behavior. In fact, their long time behavior is topologically related to that of the shift of doubly infinite sequences. Such orbits for the restricted three-body problem were discussed by Sitnikov (1960) and Alekseev (1970). This topic is discussed at length by Moser (1973) and we forego a further discussion here. Instead we list a few problems showing how little is known about the n-body problem, in spite of the progress stated.

(1) Although the existence theorems mentioned above give a finite range for the perturbation, say, $|\varepsilon| < \varepsilon_0$, in which quasi-periodic solutions can be found this range seems to be much too pessimistic. Numerical calculations (especially by Hénon and his collaborators) indicate that a realistic range is several magnitudes larger. So far this phenomenon is not understood. In particular, what is the dependence of ε_0 on n, the number of degrees of freedom?

(2) Can one find stability criteria for generic Hamiltonian systems of $n \geqslant 3$ degrees of freedom in the sense of Liapounov, i.e. without exceptional regions? For example, is there a criterion for stability of an equilibrium solution in terms of finitely many coefficients of the Taylor expansion of the Hamiltonian? At present the expectation is that this question has a negative answer which would justify the use of the weaker concept of stability in measure.

(3) Are there solutions of the n-body problem which become unbounded as t approaches a finite time t^*? Sundman's study showed that this is impossible for $n \leqslant 3$ but it has not been excluded that such a solution exists for $n \geqslant 4$.

References

Alekseev, V. M.: 1970, *Actes Congrès Int. Math.* **2**, 893–907, Gauthier-Villars, Paris.
Arnold, V. I.: 1963, *Sup. Math. Nauk* **18**, No. 6 (114), 91–192.
Moser, J.: 1967, *Math. Ann.* **169**, 136–76.
Moser, J.: 1973, *Stable and Random Motions in Dynamical Systems*, Princeton University Press, Annals of Mathematics Studies 77.
Sitnikov, K.: 1960, *Dokl. Akad. Nauk U.S.S.R.* **133**, No. 2, 303–06.
Zehnder, E.: 1974, *Bull. Am. Math. Soc.* **80**, No. 1, 174–179.

DISCUSSION

D. Saari: I would like to comment on your last problem. As you know, if noncollision singularities do exist in the n-body problem, then n is greater than 3.

    J. Moser: Yes, indeed.

J. Palous: Since an infinite number of rational numbers exist we may conclude that an infinite number of resonant periodic orbits exist. Does this mean the existence of ergodic orbits?

J. Moser: Yes. There exist resonance regions for each rational number in some interval, and for each of these regions the complicated motion modelled by the sequence shift will, in general, take place. Unfortunately, the set of these orbits form a set of measure zero. Whether there exist generic systems having a region of positive measure of ergodicity and also quasi-periodic motion is not known.

L. Perek: With a student of mine, we got interested in studying in some detail the boundary of ergodic orbits in a problem of stellar dynamics. Instead of finding a simple boundary we found a very complicated outline of many lobes.

J. Moser: Thank you.