2

The n-Body Problem

2.1 The Two-Body Problem

2.1.1 Orbits

Suppose that two points $(\mathbf{r}_1, m_1)$ and $(\mathbf{r}_2, m_2)$ interact with each other with potential energy $U(|\mathbf{r}_1 - \mathbf{r}_2|)$, so that the equations of motion have the form

$$m_1\ddot{\mathbf{r}}_1 = -\frac{\partial U}{\partial \mathbf{r}_1}, \qquad m_2\ddot{\mathbf{r}}_2 = -\frac{\partial U}{\partial \mathbf{r}_2}.$$

Proposition 2.1. *The relative position vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ in the two-body problem varies in the same way as for the motion of a point of mass $m = m_1m_2/(m_1 + m_2)$ in the central force field with potential $U(|\mathbf{r}|)$.*

If

$$\boldsymbol{\xi} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2}$$

is the centre of mass of the points m_1 and m_2, then obviously

$$\mathbf{r}_1 = \boldsymbol{\xi} + \frac{m_2}{m_1 + m_2}\mathbf{r}, \qquad \mathbf{r}_2 = \boldsymbol{\xi} - \frac{m_1}{m_1 + m_2}\mathbf{r}.$$

It follows from these formulae that in a barycentric frame of reference the trajectories of the material points are similar planar curves (with similarity ratio m_2/m_1). Thus, the problem reduces to studying the single equation

$$m\ddot{\mathbf{r}} = -\frac{\partial U}{\partial \mathbf{r}}, \qquad \mathbf{r} \in \mathbb{R}^3.$$

Let x, y be Cartesian coordinates in the plane of the orbit. Then $K_z = m(x\dot{y} - y\dot{x}) = \text{const}$. In polar coordinates $x = r\,\cos\varphi,\ y = r\,\sin\varphi$ we clearly have $K_z = mr^2\dot{\varphi}$. Consequently, $r^2\dot{\varphi} = c = \text{const}$. If $c = 0$, then $\varphi = \text{const}$ (the point moves along a straight line). We assume that $c \neq 0$. Then φ is a monotonic function of t, and therefore locally there exists the inverse function $t = t(\varphi)$. As the point m moves, its position vector sweeps out some curvilinear sector of area

$$S(t) = \frac{1}{2}\int\limits_{\varphi(0)}^{\varphi(t)} r^2\, d\varphi = \frac{1}{2}\int\limits_0^t r^2\dot{\varphi}\, dt = \frac{ct}{2}.$$

Thus, $\dot{S} = c/2 = \text{const}$ (the "sector" velocity is constant). This fact is usually referred to as the *area integral* or *Kepler's second law*, and the constant c is called the *area constant.*

Proposition 2.2 (Newton). *For a fixed value of the area constant c we have*

$$m\ddot{r} = -\frac{\partial U_c}{\partial r}, \qquad \text{where } U_c = U + \frac{mc^2}{2r^2} \qquad (r > 0). \tag{2.1}$$

This equation describes the motion of a point of mass m along the straight line $\mathbb{R} = \{r\}$ under the action of the conservative force with potential U_c. We can integrate this equation by quadratures using the energy integral

$$\frac{m\dot{r}^2}{2} + U_c = h.$$

The function U_c is called the *effective* (or *amended*, or *reduced*) *potential.*

Using the energy and area integrals we can find the equation of orbits without solving (2.1). Indeed, since $\dot{r} = \sqrt{2(h - U_c)/m}$ and $r^2\dot{\varphi} = c$, we have

$$\frac{dr}{d\varphi} = \frac{dr}{dt}\,\frac{dt}{d\varphi} = \frac{r^2}{c}\sqrt{\frac{2(h - U_c)}{m}}.$$

Integrating this equation we obtain

$$\varphi = \int \frac{c\, dr}{r^2\sqrt{\frac{2(h-U_c)}{m}}}.$$

In calculations of orbits it is sometimes useful to bear in mind the following proposition.

Proposition 2.3 (Clairaut). *Let $\rho = 1/r$ and let $\rho = \rho(\varphi)$ be the equation of the orbit. Then*

$$m\frac{d^2\rho}{d\varphi^2} = -\frac{1}{c^2}\,\frac{d}{d\rho}\,U_c\left(\frac{1}{\rho}\right).$$

For fixed values of h and c the orbit is contained in the region

$$B_{c,\,h} = \left\{(r, \varphi) \in \mathbb{R}^2 :\ U + \frac{mc^2}{2r^2} \leqslant h\right\},$$

which is the union of several annuli. Suppose that h is a regular value of the effective potential U_c and suppose that the region $B_{c,h}$ is the annulus $0 < r_1 \leqslant r \leqslant r_2 < \infty$. We claim that in this case $r(t)$ is a periodic function of time, and

$$\min r(t) = r_1, \qquad \max r(t) = r_2.$$

For the proof we set

$$u = \frac{\pi}{\tau} \int_{r_1}^{r} \frac{dx}{\sqrt{\frac{2}{m}(h - U_c(x))}}, \qquad \tau = \int_{r_1}^{r_2} \frac{dx}{\sqrt{\frac{2}{m}(h - U_c(x))}}.$$

It is obvious that $r(u)$ is a periodic function of u with period 2π and that $\dot{u} = \pi/\tau = \text{const}$. The period of the function $r(\cdot)$ is clearly equal to 2τ.

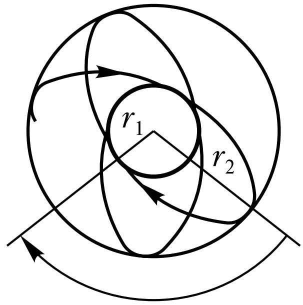

Fig. 2.1. Orbit in a central field

The angle φ changes monotonically (of course, if $c \neq 0$). The points on the orbit that are least distant from the centre are called *pericentres*, and the most distant, *apocentres.* The orbit is symmetric with respect to the straight lines passing through the point $r = 0$ and the pericentres (apocentres). The angle Φ between the directions to the adjacent apocentres (pericentres) is called the *apsidal angle.* The orbit is invariant under the rotation by the angle Φ. If the apsidal angle

$$\Phi = 2 \int_{r_1}^{r_2} \frac{c\, dr}{r^2 \sqrt{\frac{2}{m}(h - U_c)}}$$

is commensurable with π, then the orbit is closed. Otherwise it fills the annulus $B_{c,h}$ everywhere densely. If $r_2 = \infty$, then the orbit is unbounded.

The motion of the point along a circle $r = r_0$ is called a *relative equilibrium.* It is obvious that such a motion is uniform and the values of r_0 coincide with the critical points of the effective potential U_c. If the function U_c has a local minimum at a point $r = r_0$, then the corresponding circular motion is orbitally stable.

Theorem 2.1 (Bertrand). *Suppose that for some $c \neq 0$ there is a stable relative equilibrium and the potential U_c is analytic for $r > 0$. If every orbit sufficiently close to a circular one is closed, then up to an additive constant U is either γr^2 or $-\gamma/r$ (where $\gamma > 0$).*

In the first case the system is a harmonic oscillator; the orbits are ellipses centred at the point $r = 0$. The second case corresponds to the gravitational attraction. The problem of the motion of a point in the force field with potential $U = -\gamma/r$ is usually called *Kepler's problem.*

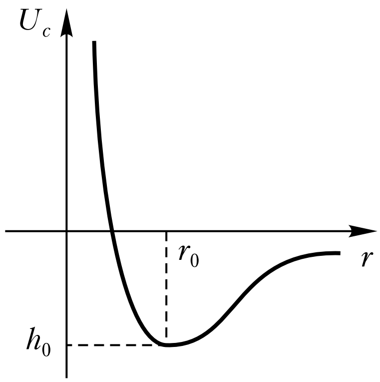

Fig. 2.2. Effective potential of Kepler's problem

The effective potential of Kepler's problem is

$$U_c = \frac{c^2}{2r^2} - \frac{\gamma}{r}.$$

According to *Clairaut's equation* in Proposition 2.3,

$$\frac{d^2\rho}{d\varphi^2} = -\rho + \frac{\gamma}{c^2}.$$

This linear non-homogeneous equation can be easily solved:

$$\rho = A\cos(\varphi - \varphi_0) + \frac{\gamma}{c^2} = \frac{1}{p}(1 + e\cos(\varphi - \varphi_0)), \tag{2.2}$$

where e and φ_0 are some constants and $p = c^2/\gamma > 0$. Hence,

$$r = \frac{p}{1 + e\cos(\varphi - \varphi_0)}$$

and therefore the orbits of Kepler's problem are conic sections with a focus at the centre of attraction (*Kepler's first law*).

Another proof of this law (based on the amazing duality between the orbits of Newtonian gravitation and Hooke's ellipses in the theory of small oscillations) is given below.

For fixed $c \neq 0$ there exists a unique relative equilibrium $r_0 = c^2/\gamma$. Its energy $h_0 = -\gamma^2/2c^2$ is minimal. Using the simple formula

$$v^2 = \dot{r}^2 + r^2\dot{\varphi}^2 = c^2\big(\rho^2 + \rho'^2\big), \qquad \rho' = \frac{d\rho}{d\varphi},$$

we can represent the energy integral in the form

$$\frac{c^2}{2}\big(\rho'^2 + \rho^2\big) - \gamma\rho = h.$$

Substituting into this formula the orbit's equation (2.2) we obtain the expression for the eccentricity $e = \sqrt{1 + 2c^2h/\gamma^2}$. Since $h \geqslant h_0 = -\gamma^2/2c^2$, the eccentricity takes only real values.

If $h = h_0$, then $e = 0$ and the orbit is circular. If $h_0 < h < 0$, then $0 < e < 1$; in this case the orbit is an ellipse. If $h = 0$, then $e = 1$ and the orbit is a parabola. For $h > 0$ we have $e > 1$; in this case the point moves along one of the branches of a hyperbola.

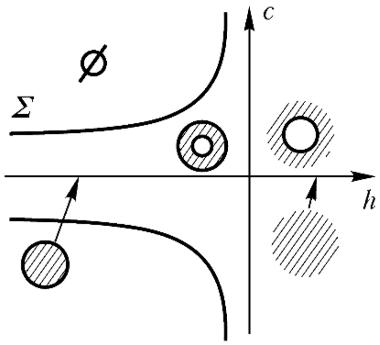

Fig. 2.3.

Fig. 2.3 depicts the *bifurcation set* Σ in the plane of the parameters c, h. The set Σ consists of the curve $h = -\gamma^2/2c^2$ and the two coordinate axes $c = 0$ and $h = 0$. The regions of possible motion $B_{c,\,h}$ (shaded areas in the figure) change the topological type at the points of Σ.

In the case of harmonic oscillator the period of revolution in an orbit is independent of the initial state. This is not the case in Kepler's problem. For elliptic motions "*Kepler's third law*" holds: $a^3/T^2 = \gamma/4\pi^2 = \text{const}$, where a is the major semiaxis of the ellipse and T is the period of revolution. Since

$$a = \frac{p}{1 - e^2} = \frac{\gamma}{2|h|},$$

the period depends only on the energy constant.

We shall now regard the Euclidean plane where the motion takes place as the plane of complex variable $z = x + iy$.

Proposition 2.4 (Bohlin). *The conformal map $w = z^2$ transforms the trajectories of a Hooke (linear) oscillator (ellipses with centre at zero) into Keplerian ellipses (with a focus at zero).*

◁ Zhukovskij's function $z = \xi + 1/\xi$ transforms the circles $|\xi| = c$ into arbitrary ellipses ($x = (c+1/c)\cos\varphi,\ \ y = (c-1/c)\sin\varphi$) with centre at zero. But $w = (1+1/\xi)^2 = \xi^2 + 1/\xi^2 + 2$ for such an ellipse; hence the map $\xi^2 \mapsto w$ is also Zhukovskij's function, but with an additional summand 2. It is easy to calculate that the distance from the centre to a focus of such an ellipse is equal to 2 for any c, so that adding 2 shifts the centre to a focus, as required. (The semiaxes $c + 1/c = a,\ \ c - 1/c = b$ give the square of the distance from the centre to a focus equal to $a^2 - b^2 = 4$.) ▷

This transformation of oscillatory orbits into Keplerian orbits is a special case of the following amazing fact.

Theorem 2.2 (Foure). *A conformal map $w \mapsto W(z)$ transforms the orbits of motion in the field with potential energy $U(z) = |dw/dz|^2$ (for the total energy constant h) into the orbits of motion in the field with potential energy $V(w) = -|dz/dw|^2$ (for the total energy constant $-1/h$).*

◁ The easiest way to prove this theorem is to compare the Lagrangians of the corresponding Maupertuis variational principles; see Ch. 4. (Incidentally, this comparison shows that the result remains valid also for the quantum-mechanical Schrödinger equation, where too there are "dual" variational principles.) According to Maupertuis' principle for natural systems (see § 4.1) a trajectory on the plane of complex variable z is a stationary curve for the length functional in the Jacobi metric, that is, in the Riemannian metric with length element

$$|ds| = \sqrt{2(h - U(z))}\,|dz|.$$

Passing to the plane of variable w we can write down the same length functional as the length in the metric with element

$$|ds| = \frac{|dw|}{\sqrt{U}}\sqrt{2(h-U)} = |dw|\sqrt{2\left(\frac{h}{U} - 1\right)} = \sqrt{h}\sqrt{2(h' - V(w))}\,|dw|,$$

where $h' = -1/h$ and $V(w) = -1/U(z)$. Up to the constant factor $\sqrt{h}$, we have obtained the metric for the potential energy V and the kinetic energy $|dw|^2/2$. Therefore our conformal map transforms the trajectories of motion with potential energy $U(z)$ into the trajectories of motion with potential energy $V(w)$, as required. ▷

Example 2.1. The conformal map $w = z^\alpha$ transforms the orbits of motion in a planar central field with a homogeneous force of degree a into the orbits of motion in a planar central field with a homogeneous force of the dual degree b, where $(a+3)(b+3) = 4$. For example, Hooke's force (linear oscillator) corresponds to $a = 1$, and Newton's gravitational force corresponds to $b = -2$, so that these forces are dual.

The exponent α is a linear function of the degree: $\alpha = (a + 3)/2$. But the theorem can also be applied to $w = e^z$ (or $w = \ln z$). △

2.1.2 Anomalies

To solve Kepler's problem completely it remains to determine the law of motion along the already known orbits. We choose the coordinate axes x and y along the major axes of the conic section representing the orbit. The equation of the orbit can be represented in the following parametric form:

$$\begin{array}{llll} x = a(\cos u - e), & y = a\sqrt{1-e^2}\,\sin u & (0 \leqslant e < 1) & \text{if } h < 0; \\ x = a(\cosh u - e), & y = a\sqrt{e^2-1}\,\sinh u & (e > 1) & \text{if } h > 0; \\ x = \frac{1}{2}(p - u^2), & y = \sqrt{p}\,u & & \text{if } h = 0. \end{array} \tag{2.3}$$

In astronomy the auxiliary variable u is called the *eccentric anomaly*, and the angle φ between the direction to the pericentre of the orbit (x-axis) and the position vector of the point, the *true anomaly.*

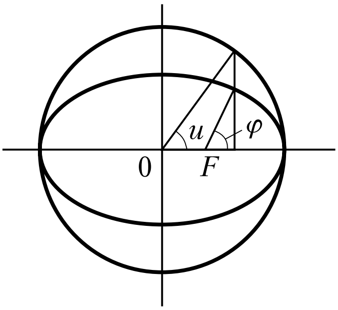

Fig. 2.4.

We have the following formulae:

$$\tan\frac{\varphi}{2} = \begin{cases} \sqrt{\dfrac{1+e}{1-e}}\,\tan\dfrac{u}{2} & \text{if } h < 0; \\[2ex] \sqrt{\dfrac{e+1}{e-1}}\,\tanh\dfrac{u}{2} & \text{if } h > 0; \\[2ex] \dfrac{u}{\sqrt{p}} & \text{if } h = 0. \end{cases}$$

Substituting formulae (2.3) into the area integral $x\dot{y} - y\dot{x} = c$ and integrating we obtain the following relations between time and the eccentric anomaly:

$$u - e\sin u = n(t - t_0), \qquad n = \frac{\sqrt{\gamma}}{p^{3/2}} \qquad \text{if } h < 0;$$

$$u - e\sinh u = n(t - t_0), \qquad n = -\frac{\sqrt{\gamma}}{p^{3/2}} \qquad \text{if } h > 0;$$

$$u + \frac{u^3}{3p} = n(t - t_0), \qquad n = \frac{2\sqrt{\gamma}}{p} \qquad \text{if } h = 0.$$

Here t_0 is the time when the point passes the pericentre. These equations (at least the first one) are called *Kepler's equations.* The linear function $\zeta = n(t - t_0)$ is usually called the *mean anomaly.*

Thus, in the elliptic case of Kepler's problem we have to solve the transcendental Kepler's equation

$$u - e\sin u = \zeta.$$

It is clear that for $0 \leqslant e < 1$ this equation has an analytic solution $u(e, \zeta)$, and the difference $u(e, \zeta) - \zeta$ is periodic in the mean anomaly ζ with period 2π. There is a choice of two ways of representing the function $u(e, \zeta)$ in a form convenient for calculations:

1) one can expand the difference $u - \zeta$ for fixed values of e in the Fourier series in ζ with coefficients depending on e;
2) one can try to represent $u(e, \zeta)$ as a series in powers of the eccentricity e with coefficients depending on ζ.

In the first case we have

$$u = \zeta + 2\sum_{m=1}^{\infty} \frac{J_m(me)}{m} \sin m\zeta, \tag{2.4}$$

where

$$J_m(z) = \frac{1}{2\pi}\int_0^{2\pi} \cos\,(mx - z\sin x)\,dx = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{m+2k}}{k!(m+k)!} \qquad (m = 0, 1, \dots)$$

is the *Bessel function* of order m. "These ... functions ... have been used extensively, precisely in this connection (which is that of Bessel), and more than half a century prior to Bessel, by Lagrange and others."[1]

[1] See Wintner [52].

The proof of formula (2.4) is based on the simple calculation

$$\begin{aligned}\frac{du}{d\zeta} &= \frac{1}{1-e\cos u}\\ &= \frac{1}{2\pi}\int_0^{2\pi}\frac{d\zeta}{1-e\cos u} + \sum_{m=1}^{\infty}\frac{\cos m\zeta}{\pi}\int_0^{2\pi}\frac{\cos m\zeta\, d\zeta}{1-e\cos u}\\ &= \frac{1}{2\pi}\int_0^{2\pi} du + \sum_{m=1}^{\infty}\frac{\cos m\zeta}{\pi}\int_0^{2\pi}\cos\left[m(u-e\sin u)\right] du\\ &= 1 + 2\sum_{m=1}^{\infty} J_m(me)\cos m\zeta.\end{aligned}$$

It remains to integrate this formula with respect to ζ.

Under the second approach we have the expansion

$$u(e,\zeta) = \sum_{m=0}^{\infty} c_m(\zeta)\frac{e^m}{m!}, \tag{2.5}$$

where

$$c_m(\zeta) = \left.\frac{\partial^m u(e,\zeta)}{\partial e^m}\right|_{e=0}.$$

Using the well-known Lagrange formula for the local inversion of holomorphic functions[2] ([603], § 7.32) we obtain the following formulae for the coefficients of this series:

$$c_0(\zeta) = \zeta; \qquad c_m(\zeta) = \frac{d^{m-1}}{d\zeta^{m-1}}\sin^m\zeta, \qquad m \geqslant 1.$$

The functions $c_m(\zeta)$ are trigonometric polynomials in the mean anomaly ζ. One can obtain the expansion (2.4) by rearranging the terms of the series (2.5). This is how Lagrange arrived at formula (2.4).

By the implicit function theorem (and in view of the periodicity of the function $u(e,\zeta)-\zeta$) the series (2.5) converges on the entire real axis $\zeta \in R$ for small e. A detailed analysis of the expansion (2.4) shows that Lagrange's series converges for $e \leqslant 0.6627434\ldots$.[3]

2.1.3 Collisions and Regularization

Above we were assuming that the area constant c is non-zero. Now suppose that $c = 0$. The motion of the point will be rectilinear and we can assume that

---

[2] Obtained by Lagrange precisely in connection with solving Kepler's equation.

[3] "In fact, a principal impetus for Cauchy's discoveries in complex function theory was his desire to find a satisfactory treatment for Lagrange's series" (Wintner [52]).

it takes place along the x-axis. If at some instant the velocity $\dot{x}$ is directed to the centre of attraction, then $x(t) \to 0$ and $\dot{x}(t) \to \infty$ as t approaches some t_0. Thus, the two bodies will collide at time $t = t_0$. It is clear that for $c = 0$ the function $x(t)$, $t \in \mathbb{R}$, necessarily has a singularity of this kind.

We now show that the eccentric anomaly u is a *regularizing* variable that resolves the singularity of the analytic function $x(t)$. If $c = 0$, then $e = 1$ in the elliptic and hyperbolic cases, and $p = 0$ in the parabolic case. Consequently, formulae (2.3) take the form

$$x = a(\cos u - 1), \qquad x = a(\cosh u - 1), \qquad x = -\frac{u^2}{2}. \tag{2.6}$$

In accordance with these formulae, for $h < 0$ the collisions take place at $u = 2\pi k,\ k \in \mathbb{Z}$; and for $h \geqslant 0$, only at $u = 0$. In the elliptic case it is also sufficient to consider the case $u = 0$.

We assume for simplicity that $t_0 = 0$. It is easy to obtain from Kepler's equations (for $e = 1$) that

$$t = u^3 f(u)$$

in a neighbourhood of the point $u = 0$, where f is an analytic function in a neighbourhood of zero such that $f(0) \neq 0$. From (2.6) we obtain a similar representation

$$x = u^2 g(u)$$

with an analytic function g such that $g(0) \neq 0$. Eliminating the eccentric anomaly u from these formulae we obtain Puiseux's expansion

$$x(t) = \left(\sqrt[3]{t}\,\right)^2 \sum_{n=0}^{\infty} c_n \left(\sqrt[3]{t}\,\right)^n.$$

The coefficients c_n with odd indices are obviously equal to zero, and $c_0 \neq 0$. Consequently, $x(t)$ is an even function of time, that is, the moving point is reflected from the centre of attraction after the collision. If x and t are regarded as complex variables, then $t = 0$ is an algebraic branching point of the analytic function $x(t)$. The three sheets of its Riemann surface meet at the collision point $t = 0$, and the real values of $x(t)$ for $t > 0$ and $t < 0$ lie only on one of the sheets. Consequently, the function $x(t)$ admits a unique real continuation.[4]

In conclusion we mention that regularization of the two-body problem in the general elliptic case (where $h < 0$) can be achieved by the transformation of coordinates $z = x + iy \mapsto w$ and time $t \mapsto \tau$ given by the formulae

$$z = w^2, \qquad t' = \frac{dt}{d\tau} = 4|w^2| = 4|z|. \tag{2.7}$$

This transformation takes the motions in Kepler's problem with constant energy $h < 0$ to the motions of the harmonic oscillator $w'' + 8|h|w = 0$ on the

[4] Regularization of collisions in the two-body problem goes back to Euler.

energy level

$$\frac{|w'|^2}{2} = 4\gamma + 4h|w^2| \tag{2.8}$$

(cf. Proposition 2.4).

The regularizing variable τ depends linearly on the eccentric anomaly u. Indeed, since

$$|z| = r = a(1 - e\cos u) \qquad \text{and} \qquad nt = u - e\sin u,$$

we have

$$\frac{du}{dt} = \frac{n}{1 - e\cos u} = \frac{na}{r},$$

whence $u = 4na\tau$.

2.1.4 Geometry of Kepler's Problem

Moser observed that by using an appropriate change of the time variable one can transform the phase flow of Kepler's problem into the geodesic flow on a surface of constant curvature. We shall follow [488] in the exposition of this result.

Lemma 2.1. *Let $x(t)$ be a solution of a Hamiltonian system with Hamiltonian $H(x)$ situated on the level $H = 0$. We change the time variable $t \mapsto \tau$ along the trajectories by the formula $d\tau/dt = G^{-1}(x) \neq 0$. Then the function $x(\tau) = x(t(\tau))$ is a solution of the Hamiltonian system (in the same symplectic structure) with the Hamiltonian $\widetilde{H} = HG$. If $G = 2(H + \alpha)$, then one can take $\widetilde{H} = (H + \alpha)^2$.*

We write down the Hamiltonian of Kepler's problem in the notation of § 2.1.3: $H = |p|^2/2 - \gamma/|z|$, where $p = \dot{z}$. We change the time variable $\dot{\tau} = |z|^{-1}$ on the manifold $H = h$ (cf. (2.7)). By Lemma 2.1 this corresponds to passing to the Hamiltonian function $|z|(H - h) = |z|\big(|p|^2 - 2h\big)/2 - \gamma$. We perform another change of the time variable $\tau \mapsto {}'\tau$, $d('\tau)/d\tau = \big(2(|z|(H - h) + \gamma)\big)^{-1}$ on the same level $H = h$. In the end we obtain a Hamiltonian system with the Hamiltonian function

$$\widetilde{H} = |z|^2 \frac{\big(|p|^2 - 2h\big)^2}{4}.$$

Finally we perform the Legendre transformation regarding p as a coordinate, and z as the momentum. As a result we obtain a natural system with the Lagrangian

$$L = \frac{|p'|^2}{(2h - |p|^2)^2}. \tag{2.9}$$

This function defines a Riemannian metric of constant Gaussian curvature (positive for $h < 0$, and negative for $h > 0$). In the case $h < 0$ the geodesics of the metric (2.9) (defined for all $p \in \mathbb{R}^2$) are the images of the great circles of the sphere under the stereographic projection, and in the case $h > 0$ (in which the metric is defined in the disc $|p|^2 < 2h$) the geodesics are the straight lines of the Lobachevskij plane (in Poincaré's model).

Remark 2.1 (A. B. Givental'). Let the plane (x, y) be the configuration plane of Kepler's problem with Lagrangian $L = (\dot{x}^2 + \dot{y}^2)/2 + 1/\sqrt{x^2 + y^2}$. In the space (x, y, z) we consider the right circular cone $z^2 = (x^2 + y^2)$ and the family of inscribed paraboloids of revolution $z = (x^2 + y^2)/4\alpha + \alpha$, where α is a parameter. By "projection" we shall mean the projection of the space (x, y, z) onto the plane (x, y) parallel to the z-axis. One can show that

1) the trajectories of Kepler's problem are the projections of the planar sections of the cone (in particular, the vertex of the cone is a focus of the projections of its planar sections),
2) the trajectories with the same value of the total energy are the projections of the sections of the cone by the planes tangent to one and the same paraboloid,
3) the trajectories with the same value of the angular momentum are the projections of the sections of the cone by the planes passing through one and the same point of the z-axis.

2.2 Collisions and Regularization

2.2.1 Necessary Condition for Stability

We now turn to the general *n-body problem* dealing with n material points $(m_1, \mathbf{r}_1), \dots, (m_n, \mathbf{r}_n)$ attracted to each other according to the law of universal gravitation. The kinetic energy is

$$T = \frac{1}{2} \sum m_i \dot{\mathbf{r}}_i^2$$

and the force function

$$V = \sum_{j<k} \frac{m_j m_k}{r_{jk}}, \qquad r_{jk} = |\mathbf{r}_j - \mathbf{r}_k|,$$

is always positive. We introduce an inertial frame of reference with origin at the centre of mass, and let the $\mathbf{r}_i$ be the position vectors of the points in the new frame. The equations of the n-body problem have the form of Lagrange's equations with the Lagrangian $L = T + V$.

We say that a motion $\mathbf{r}_s(t)$ $(1 \leqslant s \leqslant n)$ is *stable* if the following two conditions hold:

a) $r_{ij}(t) \neq 0$ for all values of t and all $i \neq j$ (there are no collisions);

b) $|r_{ij}(t)| \leqslant c$, where $c = \text{const}$.

Theorem 2.3 (Jacobi). *If a motion is stable, then the total energy $h = T - V$ is negative.*

◁ We apply Lagrange's formula

$$\ddot{I} = 2V + 4h, \tag{2.10}$$

where $I = \sum m_i r_i^2$ is the polar moment of inertia. If $h \geqslant 0$, then the function $I(t)$, $t \in \mathbb{R}$, is convex and therefore cannot be simultaneously bounded below and above. To complete the proof it remains to use Lagrange's identity:

$$I \sum m_i = \sum_{j<k} m_j m_k r_{jk}^2 + \Big(\sum m_i \mathbf{r}_i\Big)^2.$$ ▷

Under the additional assumption that the mutual distances be bounded below ($|r_{ij}(t)| \geqslant c > 0$) it follows from the energy integral and Lagrange's formula (2.10) that along a stable motion the mean values

$$\lim_{s\to\infty} \frac{1}{s} \int_0^s V(t)\,dt, \qquad \lim_{s\to\infty} \frac{1}{s} \int_0^s 2T(t)\,dt$$

exist and are equal to $-2h > 0$.

The necessary condition for stability $h < 0$ is not sufficient if $n > 2$.

2.2.2 Simultaneous Collisions

If the position vectors $\mathbf{r}_i(t)$ of all points have one and the same limit $\mathbf{r}_0$ as $t \to t_0$, then we say that a simultaneous collision takes place at time t_0. The point $\mathbf{r}_0$ clearly must coincide with the centre of mass, that is, $\mathbf{r}_0 = \mathbf{0}$. A simultaneous collision occurs if and only if the polar moment of inertia $I(t)$ tends to zero as $t \to t_0$.

Theorem 2.4. *If $I(t) \to 0$ as $t \to t_0$, then the constant vector of angular momentum is equal to zero:*

$$K = \sum m_i(\mathbf{r}_i \times \dot{\mathbf{r}}_i) = \mathbf{0}.$$

For $n = 3$ this theorem was already known to Weierstrass.

◁ Since $V(t) \to +\infty$ as $t \to t_0$, by the equation $\ddot{I} = 2V + 4h$ we have $\ddot{I}(t) > 0$ for the values of time close to t_0. Consequently, $I(t)$ is monotonically decreasing before the collision.

We use the inequality $K^2 \leqslant 2IT$ (see § 1.1), which is equivalent to the inequality

$$\ddot{I} \geqslant \frac{K^2}{I} + 2h$$

by Lagrange's formula. We multiply this inequality by the positive number $-2\dot{I}$ and integrate it on the interval (t_1, t) for $t < t_0$:

$$\dot{I}^2(t_1) - \dot{I}^2(t) \geqslant 2K^2 \ln \frac{I(t_1)}{I(t)} + 4h(I(t_1) - I(t)).$$

All the more we have the inequality

$$2K^2 \ln \frac{I(t_1)}{I(t)} \leqslant \dot{I}^2(t_1) + 4|h|I(t_1).$$

This implies the existence of a positive lower bound for $I(t)$ on the interval (t_1, t_0) if $K^2 \neq 0$. ▷

2.2.3 Binary Collisions

We say that a binary collision happens at time t_0 if the distance between two points, say, m_1 and m_n, tends to zero as $t \to t_0$, while the mutual distances between the other points are bounded below by some positive quantity for the values of t close to t_0. For such values of t the influence of the points $m_2, \dots, m_{n-1}$ on the motion of m_1 and m_n is clearly negligible by comparison with the interaction of m_1 and m_n. Therefore it is natural to expect that at times t close to t_0 the behaviour of the vector $\mathbf{r}_{1n}(t) = \mathbf{r}_1(t) - \mathbf{r}_n(t)$ is approximately the same as in the problem of collision of two bodies (see § 2.1). In the two-body problem a locally uniformizing variable was the true anomaly $u(t)$, which is proportional to the integral of the inverse of the distance between the points. Therefore in the case of a binary collision it is natural to try to regularize the solution by the variable

$$u(t) = \int_{t_0}^{t} \frac{ds}{|\mathbf{r}_{1n}(s)|}. \tag{2.11}$$

One can show that this consideration indeed achieves the goal: the functions $\mathbf{r}_k(u)$ are regular near the point $u = 0$ (corresponding to the binary collision) and in addition, $t(u) - t_0 = u^3 p(u)$, where $p(\cdot)$ is a function holomorphic near $u = 0$ and such that $p(0) \neq 0$. Thus, in the case of a binary collision, just like in the two-body problem, the coordinates of the points $\mathbf{r}_k$ are holomorphic functions of the variable $\sqrt[3]{t - t_0}$ and therefore admit a unique real analytic continuation for $t > t_0$. One can show that the functions $\mathbf{r}_2(t), \dots, \mathbf{r}_{n-1}(t)$ are even holomorphic in a neighbourhood of the point t_0.

To make the uniformizing variable $u(t)$ suitable for any pair of points and any instant of a binary collision one should replace (2.11) by the formula

$$u(t) = \int_0^t V(s)\, ds = \int_0^t \sum_{j<k} \frac{m_j m_k}{|\mathbf{r}_{jk}(s)|}\, ds.$$

If the polar angular momentum is non-zero, then binary collisions are the only possible singularities in the three-body problem. As shown by Sundman, the functions $\mathbf{r}_k(u)$ $(1 \leqslant k \leqslant 3)$ are holomorphic in some strip $|\operatorname{Im} u| < \delta$ of the complex plane $u \in \mathbb{C}$ containing the real axis. We now map this strip conformally onto the unit disc $|\omega| < 1$ by the transformation

$$\omega = \frac{e^{\pi u/2\delta} - 1}{e^{\pi u/2\delta} + 1},$$

which takes the real axis $-\infty < u < +\infty$ to the segment $-1 < \omega < 1$. As a result the coordinates of the points $\mathbf{r}_k$ become holomorphic functions in the disc $|\omega| < 1$ and can be represented as converging power series in the new variable ω. These series represent the motion of the three bodies for all values of time $t \in (-\infty, +\infty)$.[5]

This result is due to Sundman (1913); he followed the earlier work of Poincaré and Weierstrass, who obtained expansions of the solutions of the n-body problem in converging power series in the auxiliary variable ω in the absence of collisions. As for the possibility of collisions, they are infinitely rare in the three-body problem. Using the theorem on simultaneous collisions and the regularization of binary collisions one can show that in the twelve-dimensional state space of the three-body problem (for a fixed position of the centre of mass) the collision trajectories lie on certain singular analytic surfaces of dimension 10. Their measure is, of course, equal to zero. However, it is not known whether these singular surfaces can fill everywhere densely entire domains in the state space.

In conclusion we give as an illustration the results of numerical calculations in the "Pythagorean" variant of the three-body problem where the bodies with masses 3, 4, 5 are initially at rest in the (x, y)-plane at the points with coordinates $(1, 3)$, $(-2, -1)$, $(1, -1)$. The centre of mass of this system is at the origin.

The calculations of the Pythagorean three-body problem were started by Burrau back in 1913 and were continued in modern times by Szebehely using computers. In Fig. 2.5–2.7 one can see close encounters of the points, their binary collisions, and the dispersal of the triple system. Fig. 2.8 shows a "final" motion: the point of mass $m = 5$ is moving away along a straight line from the "double star" formed by the points $m = 3$ and $m = 4$, which periodically collide with each other. It is interesting that no triple collisions occur, although the angular momentum is equal to zero in this case.

[5] The power series in ω are absolutely useless for practical computations because of their extremely slow rate of convergence.

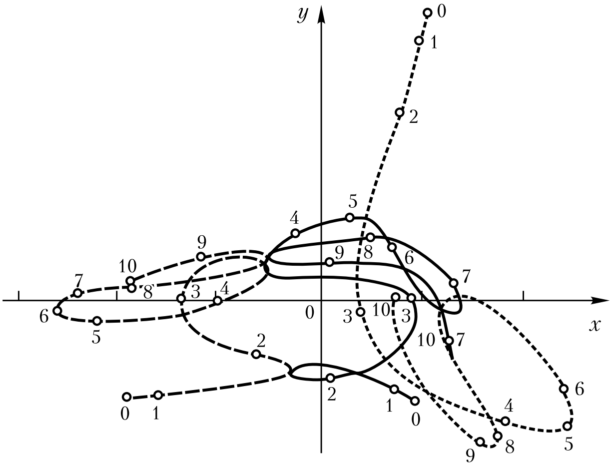

Fig. 2.5. Motion of gravitating masses in the Pythagorean three-body problem in the time interval from $t = 0$ to $t = 10$

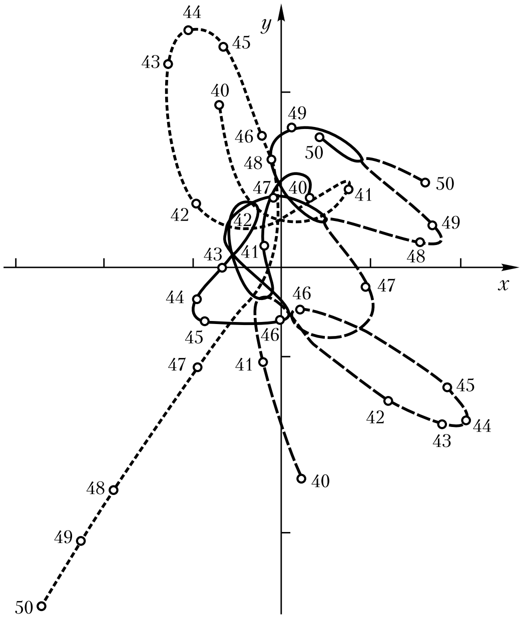

Fig. 2.6. Form of the orbits in the Pythagorean three-body problem in the time interval from $t = 40$ to $t = 50$

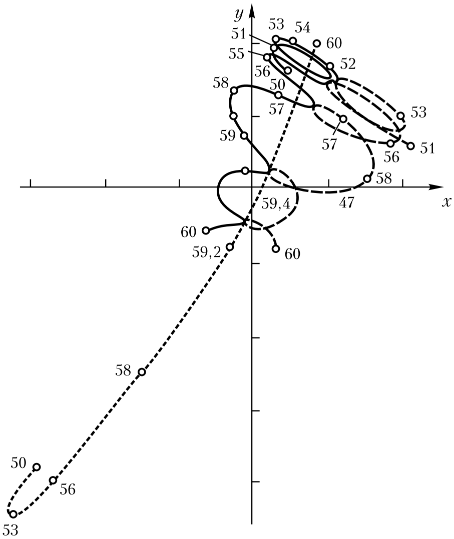

Fig. 2.7. Evolution of the orbits of the Pythagorean three-body problem in the time interval from $t = 50$ to $t = 60$

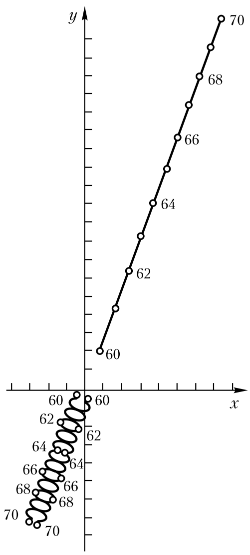

Fig. 2.8. Formation of a double star in the Pythagorean three-body problem (from $t = 60$ to $t = 70$)

2.2.4 Singularities of Solutions of the n-Body Problem

In the case of multiple collisions, when simultaneous collisions of $k \geqslant 3$ points occur, the singular points of the coordinates $\mathbf{r}_s$ of gravitating points as functions of time have a much more complicated structure from the analytic viewpoint. Generally speaking, these singularities are not algebraic; moreover, the functions $\mathbf{r}_s(t)$ $(1 \leqslant s \leqslant n)$ have no real analytic continuation after the instant of collision.

This can be seen even in examples of simultaneous collision in the three-body problem. It turns out that for arbitrary values of the masses m_1, m_2, m_3 there exist solutions of the form

$$\mathbf{r}_i(t) = t^{2/3} \sum_{m=0}^{\infty} \mathbf{a}_{im} t^{\alpha m}. \tag{2.12}$$

The positive number α is a non-constant algebraic function of the masses m_1, m_2, m_3, and the coefficients $\mathbf{a}_{i1}$ are not all equal to zero. At time $t = 0$ a triple collision took place. In a typical case where α is irrational, the series (2.12) has an isolated logarithmic singular point at $t = 0$. In particular, this solution, which is real for $t > 0$, has infinitely many different analytic branches for $t < 0$, but all these branches turn out to be complex.

Solution (2.12) was found by Block (1909) and Chazy (1918) using the following method. For any values of the masses the equations of the three-body problem admit a "homographic" solution such that the triangle formed by the bodies always remains similar to itself. This solution is analytically represented by the formula

$$\mathbf{r}_i(t) = \mathbf{a}_{i0} t^{2/3} \qquad (1 \leqslant i \leqslant 3). \tag{2.13}$$

Among the characteristic roots of the variational equation for this solution there is a negative number $(-\alpha)$. According to the well-known results of Lyapunov and Poincaré the equations of motion have a solution (2.12) that is asymptotic to solution (2.13). We remark that the method of Block and Chazy had already been applied by Lyapunov (1894) for proving that the solutions of the equations of rotation of a heavy rigid body with a fixed point are not single-valued as functions of complex time.

Consider a particular solution $\mathbf{r}_k(t)$ of the three-body problem. Suppose that at the initial time t_0 we have $r_{jk} \neq 0$ for all $j \neq k$. We trace this solution for $t > t_0$. There are three possibilities:

(a) there are no collisions for any $t > t_0$; then this motion proceeds without singularities up to $t = +\infty$;

(b) at some instant $t_1 > t_0$ a collision occurs that admits an analytic continuation;

(c) at some instant a collision occurs that does not admit an analytic continuation.

Suppose that case (b) takes place. Then for $t > t_1$ again one of the variants (a)–(c) is possible. Continuing this process we can either arrive after finitely many steps at one of the cases (a), (c), or have infinitely many continuable collisions occurring at times $t_1, t_2, \dots, t_k, \dots$. One can show that for $n = 3$ in the latter case we have

$$\lim_{k\to\infty} t_k = +\infty.$$

However, in the n-body problem for $n \geqslant 4$ a fundamentally different type of singularities is possible. Even in the four-body problem on a straight line there exist motions such that infinitely many binary collisions occur over a finite time interval $[0, t_*)$. Moreover, in the end, as $t \to t_*$ three of the bodies move away to infinity: one in one direction, and two others in the opposite direction, as in the Pythagorean three-body problem. But unlike the case of three bodies, the colliding bodies approach each other arbitrarily closely, which is what gives the energy for going to infinity over a finite time. The fourth body oscillates between the bodies going to infinity in opposite directions. When the oscillating body approaches closely the cluster of two bodies, an almost triple collision occurs. The existence of such a motion was proved by Mather using McGehee's regularization of simultaneous collisions in the three-body problem (see [419]).

In the spatial five-body problem there are collision-free singularities: over a finite time the bodies move away to infinity without ever having collisions [531, 607]. The existence of collision-free singularities was also proved for the planar $3N$-body problem for sufficiently large N; see [255].

2.3 Particular Solutions

Only a few exact solutions have been found in the n-body problem. For the case of bodies of different masses practically all of these solutions had already been known to Euler and Lagrange.

2.3.1 Central Configurations

We say that n material points $(m_i, \mathbf{r}_i)$ form a *central configuration* in a barycentric frame of reference if

$$\frac{\partial V}{\partial \mathbf{r}_i} = \sigma \frac{\partial I}{\partial \mathbf{r}_i}, \qquad 1 \leqslant i \leqslant n, \tag{2.14}$$

where

$$V = \sum_{k<j} \frac{m_k m_j}{r_{kj}}$$

is the potential of gravitational interaction, $I = \sum m_i r_i^2$ is the polar moment of inertia, and σ is a scalar function independent of the index i. It follows from Euler's formula for homogeneous functions that $\sigma = -V/2I$. Thus, formula (2.14) can also be written in the form

$$I\frac{\partial V}{\partial \mathbf{r}_i} = -\frac{1}{2}V\frac{\partial I}{\partial \mathbf{r}_i}.$$

Consequently, central configurations correspond to the critical points of the function IV^2. Since this function is homogeneous, the set $(m_i, \mathbf{r}_i)$ is a central configuration simultaneously with the set $(m_i, \alpha\mathbf{r}_i)$ for any $\alpha \neq 0$. We shall not distinguish between such configurations.

Finding all central configurations for any number of points n is a complicated algebraic problem, which is still unsolved. Leaving aside the trivial case $n = 2$, we list the known results in this area.

For $n = 3$ the only non-collinear central configuration is the equilateral triangle (Lagrange). For $n = 4$ the only non-coplanar configuration is the regular tetrahedron.

If the masses of all bodies are equal, then for $n = 4$ the only planar central configurations are those in which the bodies are situated either on one straight line, or in the vertices of a square, or in the vertices and in the centre of an equilateral triangle, or in the vertices and on the symmetry axis of an isosceles triangle [62, 63].

The collinear central configurations are described by the following *Moulton's theorem* [52]: corresponding to every numbering of the point masses there is a unique central configuration in which the points are situated on one straight line in the given order. Thus, there exist exactly $n!/2$ different collinear central configurations. For $n = 3$ there are exactly three such configurations; they were discovered by Euler.

There is a conjecture that for a given n and given masses the number of central configurations is finite [166] and, moreover, is bounded by a constant independent of the masses [52]. These problems are open also for the planar case [559]. (A planar central configuration is a relative equilibrium configuration of the n-body problem; see § 2.3.3.) For $n = 4$ in the planar case the conjecture of [166, 52] was proved in [274]; the case $n = 5$ remains open.

The concept of a central configuration is useful in the analysis of simultaneous collisions: it turns out that the configuration of gravitating points at the instant of a simultaneous collision is central (in the asymptotic sense). It follows from (2.14) that if initially the points formed a central configuration and were at rest, then their configuration clearly does not change up to the instant of a simultaneous collision.

2.3.2 Homographic Solutions

We say that a solution of the n-body problem is *homographic* if in a barycentric reference frame the configurations formed by the bodies remain similar to each other at all times. If in addition the configuration is not rotating, then such a solution is said to be *homothetic.* The solutions mentioned at the end of the preceding subsection may serve as an example. If the configuration remains congruent to itself, then the solution is called a *relative equilibrium.*

It is easy to show that

a) a homographic solution is homothetic if and only if the polar angular momentum is equal to zero;
b) a homographic solution is a relative equilibrium if and only if it is planar and its configuration rotates with constant angular velocity.

The proof of the following facts is more difficult:

c) if a homographic solution is non-coplanar, then it is homothetic;
d) if a homographic solution is coplanar, then it is planar.

In particular, every homographic solution is either planar or homothetic. In the three-body problem all the homographic solutions have the property that in a barycentric reference frame the three bodies lie in an invariable plane containing the centre of mass (Lagrange).

Proposition 2.5. *If a solution is homographic, then the bodies form a central configuration at all times.*

This proposition provides a method for constructing homographic solutions. We give as an example the well-known *Lagrange's theorem* (1772).

Theorem 2.5. *For arbitrary values of the masses, the three-body problem admits an exact solution such that*

1) *the plane containing these points is invariant in a barycentric reference frame,*
2) *the resultant of the two Newtonian gravitational forces applied to each of the three material points passes through their common centre of mass,*
3) *the triangle formed by the three bodies is equilateral,*
4) *the trajectories of the three bodies are conic sections similar to each other with a focus at the common centre of mass.*

In the special case of equal masses the conic sections are congruent and differ from one another by a rotation through 120°. This remark can be generalized: the problem of n points of equal masses has a solution in which each body is describing a conic section with a focus at the centre of mass, the trajectories are congruent and differ from one another by a rotation through $2\pi/n$.

2.3.3 Effective Potential and Relative Equilibria

Proposition 2.6. *The configurations of relative equilibria with polar angular momentum K coincide with the critical points of the function*

$$U_K = U + \frac{K^2}{2I}, \qquad \text{where } U = -V.$$

The function U_K is called the *effective* (or *amended*, or *reduced*) *potential.* We used it § 1.1 for describing the regions of possible motion in the planar n-body problem, and in § 2.1 for finding the trajectories of two bodies.

◁ Suppose that the configuration of a relative equilibrium is rotating around the centre of mass with constant angular velocity ω. Then, clearly, $K = I\omega$. We pass to a reference frame with coordinates u, v rotating with the angular velocity ω; in this frame the configuration of the relative equilibrium is stationary. In the new frame the Lagrangian function is

$$L = T + V = \frac{1}{2}\sum m_i\bigl(\dot u_i^2 + \dot v_i^2\bigr) + \omega \sum m_i(u_i\dot v_i - \dot u_i v_i) + V_\omega,$$

where $V_\omega = V + I\omega^2/2$. The equations of motion are

$$m_i\ddot u_i = 2m_i\omega\dot v_i + \frac{\partial V_\omega}{\partial u_i}, \qquad m_i\ddot v_i = -2m_i\omega\dot u_i + \frac{\partial V_\omega}{\partial v_i}. \tag{2.15}$$

One can easily derive Proposition 2.6 from these equations using the following observation: the functions U_K and V_ω have the same critical points, since $K = I\omega$ at these points. ▷

2.3.4 Periodic Solutions in the Case of Bodies of Equal Masses

If all the n bodies have the same mass, then one can seek periodic solutions in which all the bodies move along the same trajectory lagging one behind another by equal time intervals. The law of motion of the jth body ($j = 1, \dots, n$) is sought in the form

$$\mathbf{r}_j(t) = \mathbf{r}\bigl(t - (j-1)T/n\bigr), \tag{2.16}$$

where $\mathbf{r}(\cdot)$ is a periodic function with period T. Such solutions are called *simple choreographies* (this term was suggested by Simó). The function $\mathbf{r}(\cdot)$ can be determined from the condition that a periodic solution is an extremal of the action functional (see § 1.2.3).

For the three-body problem ($n = 3$) a solution of the form (2.16) was first found numerically [431]. Then an analytic proof of its existence was given [170]. In this solution the three bodies describe one and the same planar curve having the shape of 8 with equal loops. Over the period T each body passes twice the self-intersection point of the trajectory, and at these instants all the three bodies are situated on one straight line and form a collinear central configuration (cf. § 2.3.1; one of the bodies bisects the segment between the other two bodies). This periodic solution is stable in the linear approximation. Furthermore, the nonlinear terms in the expansion of the Hamiltonian of the problem about this periodic solution are such that KAM theory (§ 6.3) guarantees "stability with respect to the measure of initial data": a small neighbourhood of this periodic motion is foliated, up to a remainder of small relative measure, into the invariant tori, on which the motion is conditionally periodic [551, 553].

In the case $n > 3$ solutions of the form (2.16) have so far been found only numerically; all these solutions proved to be unstable. The existence of such solutions is at present established analytically for the interaction potential $U = -\gamma/r_{ij}^a$, $a \geqslant 2$ (the Newtonian potential corresponds to $a = 1$) [551, 552].

The variational approach was also used in the search for periodic solutions in which the orbits of all bodies are congruent curves permuted by a symmetry. Such a periodic solution was found in the four-body problem [171]. Over the period of the motion the bodies form a central configuration four times: twice they are situated in the vertices of a square, and twice in the vertices of a tetrahedron.

2.4 Final Motions in the Three-Body Problem

2.4.1 Classification of the Final Motions According to Chazy

Here, dealing with the three-body problem, we shall denote by $\mathbf{r}_k$ the vector from the point mass m_i to the point mass m_j for $i \neq k$, $j \neq k$, $i < j$.

Theorem 2.6 (Chazy, 1922). *Every solution of the three-body problem* $\mathbf{r}_k(t)$ $(k = 1, 2, 3)$ *belongs to one of the following seven classes:*

1°. H (*hyperbolic motions*): $|\mathbf{r}_k| \to \infty$, $|\dot{\mathbf{r}}_k| \to c_k > 0$ *as* $t \to +\infty$;

2°. HP_k (*hyperbolic-parabolic*): $|\mathbf{r}_i| \to \infty$, $|\dot{\mathbf{r}}_k| \to 0$, $|\dot{\mathbf{r}}_i| \to c_i > 0$ $(i \neq k)$;

3°. HE_k (*hyperbolic-elliptic*): $|\mathbf{r}_i| \to \infty$, $|\dot{\mathbf{r}}_i| \to c_i > 0$ $(i \neq k)$, $\sup\limits_{t \geqslant t_0} |\mathbf{r}_k| < \infty$;

4°. PE_k (*parabolic-elliptic*): $|\mathbf{r}_i| \to \infty$, $|\dot{\mathbf{r}}_i| \to 0$ $(i \neq k)$, $\sup\limits_{t \geqslant t_0} |\mathbf{r}_k| < \infty$;

5°. P (*parabolic*): $|\mathbf{r}_i| \to \infty$, $|\dot{\mathbf{r}}_i| \to 0$;

6°. B (*bounded*): $\sup\limits_{t \geqslant t_0} |\mathbf{r}_k| < \infty$;

7°. OS (*oscillating*): $\overline{\lim\limits_{t \to +\infty}} \sup\limits_{k} |\mathbf{r}_k| = \infty$, $\underline{\lim\limits_{t \to +\infty}} \sup\limits_{k} |\mathbf{r}_k| < \infty$.

Examples of motions of the first six types were known to Chazy. The existence of oscillating motions was proved by Sitnikov in 1959.

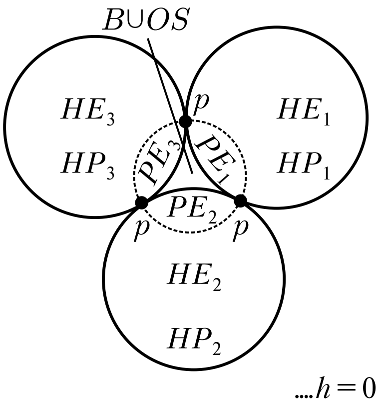

Fig. 2.9.

It is natural to associate with the seven types of final motions listed above the subsets of the twelve-dimensional phase space of the three-body problem M^{12} with a fixed position of the centre of mass: these subsets are composed entirely of the phase trajectories corresponding to the motions of a given type. The qualitative picture of the partition of M^{12} into the classes of final motions is represented by Fig. 2.9. The sets H and HP_k are entirely contained in the domain where the constant of total energy h is positive; P lies on the hypersurface $h = 0$; the sets B, PE_k, OS, in the domain $h < 0$; and motions in the class HE_k are possible for any value of h. It is known that H and HE_k are open in M^{12}, HP_k consists of analytic manifolds of codimension 1, and P consists of three connected manifolds of codimension 2 (represented by the three points in Fig. 2.9) and one manifold of codimension 3 (which is not shown in Fig. 2.9). The topology of the other classes has not been studied sufficiently.

2.4.2 Symmetry of the Past and Future

By Chazy's theorem one can introduce seven analogous final classes of motions when t tends not to $+\infty$, but to $-\infty$. To distinguish the classes in the cases $t \to \pm\infty$ we shall use the superscripts $(+)$ and $(-)$: H^+, HE_3^-, and so on. In one of Chazy's papers (1929) a false assertion was stated that in the three-body problem the two final types, for $t \to \infty$ and $t \to -\infty$, of the same solution coincide. The misconception of the "symmetry" of the past and future had been holding ground for a fairly long time, despite the numerical counterexample constructed by Bekker (1920), which asserted the possibility of "exchange": $HE_1^- \cap HE_2^+ \neq \varnothing$. Bekker's example had been "explained" by errors in numerical integration. In 1947 Shmidt produced an example of "capture" in the three-body problem: $H^- \cap HE^+ \neq \varnothing$. This example, which was also constructed by a numerical calculation, was given by Shmidt in support of his well-known cosmogony hypothesis. A rigorous proof of the possibility of capture was found by Sitnikov in 1953.

The current state of the problem of final motions in the three-body problem is concisely presented in Tables 2.1 and 2.2, which we borrowed from Alekseev's paper [3]. Each cell corresponds to one of the logically possible combinations of the final types in the past and future. The Lebesgue measure of the corresponding sets in M^{12} is indicated (where it is known).

Table 2.1.

<table>
<tr><td colspan="2" rowspan="2">$h > 0$</td><td colspan="2">$t \to +\infty$</td></tr>
<tr><td>H^+</td><td>HE_i^+</td></tr>
<tr><td rowspan="3">t
$\downarrow$
$-\infty$</td><td>H^-</td><td>Lagrange, 1772
(isolated examples);
Chazy, 1922
Measure > 0</td><td>PARTIAL CAPTURE
Measure > 0
Shmidt (numerical example), 1947;
Sitnikov (qualitative methods), 1953</td></tr>
<tr><td rowspan="2">HE_j^-</td><td rowspan="2">COMPLETE DISPERSAL
Measure > 0</td><td>$i = j$ Measure > 0
Birkhoff, 1927</td></tr>
<tr><td>$i \neq j$ EXCHANGE, Measure > 0
Bekker (numerical examples), 1920;
Alekseev (qualitative methods), 1956</td></tr>
</table>

Table 2.2.

<table>
<tr><td colspan="2" rowspan="2">$h < 0$</td><td colspan="3">$t \to +\infty$</td></tr>
<tr><td>HE_i^+</td><td>B^+</td><td>OS^+</td></tr>
<tr><td rowspan="4">t
$\downarrow$
$-\infty$</td><td rowspan="2">HE_j^-</td><td>$i = j$ Measure > 0
Birkhoff, 1927</td><td rowspan="2">COMPLETE
CAPTURE
{ Measure $= 0$
Chazy, 1929 and
Merman, 1954;
Littlewood, 1952;
Alekseev, 1968,
$\neq \varnothing$</td><td rowspan="2">{ Measure $= 0$
Chazy, 1929 and
Merman, 1954;
Alekseev, 1968,
$\neq \varnothing$</td></tr>
<tr><td>EXCHANGE
$i \neq j$ Measure > 0
Bekker, 1920
(numerical examples);
Alekseev, 1956
(qualitative methods)</td></tr>
<tr><td>B^-</td><td>PARTIAL
DISPERSAL
$\neq \varnothing$
Measure $= 0$</td><td>Euler, 1772
Lagrange, 1772,
Poincaré, 1892
(isolated examples);
Measure > 0
Arnold, 1963</td><td>Littlewood, 1952
Measure $= 0$
Alekseev, 1968,
$\neq \varnothing$</td></tr>
<tr><td>OS^-</td><td>$\neq \varnothing$
Measure $= 0$</td><td>$\neq \varnothing$
Measure $= 0$</td><td>Sitnikov, 1959,
$\neq \varnothing$
Measure $= ?$</td></tr>
</table>

2.5 Restricted Three-Body Problem

2.5.1 Equations of Motion. The Jacobi Integral

Suppose that the Sun S and Jupiter J are revolving around the common centre of mass in circular orbits (see Fig. 2.10). We choose the units of length, time, and mass so that the magnitude of the angular velocity of the rotation, the sum of masses of S and J, and the gravitational constant are equal to one. It is easy to show that then the distance between S and J is also equal to one.

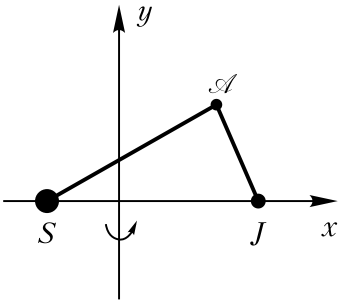

Fig. 2.10. Restricted three-body problem

Consider the motion of an asteroid $\mathscr{A}$ in the plane of the orbits of S and J. We assume that the mass of the asteroid is much smaller than the masses of the Sun and Jupiter and neglect the influence of the asteroid on the motion of the large bodies.

It is convenient to pass to a moving frame of reference rotating with unit angular velocity around the centre of mass of S and J; the bodies S and J are at rest in this frame. In the moving frame we introduce Cartesian coordinates x, y so that the points S and J are situated invariably on the x-axis and their centre of mass coincides with the origin. The equations of motion of the asteroid take the following form (see (2.15)):

$$\ddot{x} = 2\dot{y} + \frac{\partial V}{\partial x}, \quad \ddot{y} = -2\dot{x} + \frac{\partial V}{\partial y}; \quad V = \frac{x^2 + y^2}{2} + \frac{1-\mu}{\rho_1} + \frac{\mu}{\rho_2}, \tag{2.17}$$

where μ is Jupiter's mass and ρ_1, ρ_2 are the distances from the asteroid $\mathscr{A}$ to S and J. Since the coordinates of S and J are $(-\mu, 0)$ and $(1-\mu, 0)$, we have

$$\rho_1^2 = (x+\mu)^2 + y^2, \qquad \rho_2^2 = (x - 1 + \mu)^2 + y^2.$$

Equations (2.17) have the integral

$$\frac{\dot{x}^2 + \dot{y}^2}{2} - V(x, y) = h,$$

called the *Jacobi integral*, which expresses the conservation of energy in the relative motion of the asteroid.

For a fixed value of h the motion of the asteroid takes place in the domain

$$\{(x, y) \in \mathbb{R}^2 : \quad V(x, y) + h \geqslant 0\},$$

which is called a *Hill region.*

2.5.2 Relative Equilibria and Hill Regions

The form of Hill regions depends on the positions of the critical points of the function $V(x, y)$. Corresponding to each critical point (x_0, y_0) there is an "equilibrium" solution $x(t) \equiv x_0, \;\; y(t) \equiv y_0$, which can naturally be called a relative equilibrium. We claim that for every value of $\mu \in (0, 1)$ there are exactly five such points.

We calculate

$$V'_y = yf, \qquad f = 1 - \frac{1-\mu}{\rho_1^3} - \frac{\mu}{\rho_2^3},$$
$$V'_x = xf - \mu(1-\mu)\left(\frac{1}{\rho_1^3} - \frac{1}{\rho_2^3}\right)$$

and solve the system of algebraic equations $V'_x = V'_y = 0$. First suppose that $y \neq 0$. Then $f = 0$ and therefore, $\rho_1 = \rho_2 = \rho$. From the equation $f = 0$ we obtain that $\rho = 1$. Thus, in this case the points S, J, and $\mathscr{A}$ are in the vertices of an equilateral triangle. There are exactly two such relative equilibria, which are called *triangular* (or *equilateral*) *libration points.* They should be viewed as a special case of Lagrange's solutions of the general "unrestricted" three-body problem (see § 2.3). Lagrange himself regarded these solutions as a "pure curiosity" and considered them to be useless for astronomy. But in 1907 an asteroid was discovered, named Achilles, which moves practically along Jupiter's orbit being always ahead of it by 60°. Near Achilles there are 9 more asteroids (the "Greeks"), and on the other side there were discovered five asteroids (the "Trojans"), which also form an equilateral triangle with the Sun and Jupiter.

Now consider the relative equilibria on the x-axis. They are the critical points of the function

$$g(x) = \frac{x^2}{2} + \frac{1-\mu}{|x+\mu|} + \frac{\mu}{|x-1+\mu|}.$$

Since $g(x) > 0$ and $g(x) \to +\infty$ as $x \to \pm\infty, \;\; x \to -\mu$, or $x \to 1-\mu$, there exist three local minima of the function g in the intervals $(-\infty, -\mu)$, $(-\mu, 1-\mu), \;\; (1-\mu, +\infty)$, into which the points S and J divide the x-axis. In view of the inequality $g''(x) > 0$ these points are the only critical points of the function g. These *collinear libration points* were found by Euler.

One can show that the collinear libration points (we denote them by L_1, L_2, L_3)[6] are of hyperbolic type, and the triangular libration points (L_4 and L_5) are points of non-degenerate minimum of the function V. Fig. 2.11 depicts the transformation of the Hill regions as the Jacobi constant h changes from $-\infty$ to $+\infty$, under the assumption that Jupiter's mass is smaller than the Sun's mass (the complement of the Hill region is shaded).

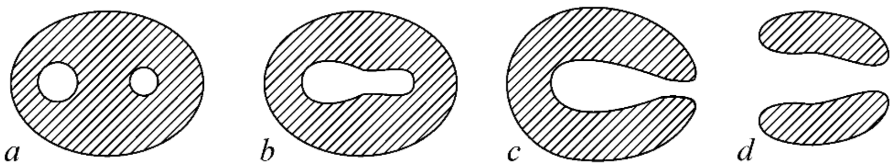

Fig. 2.11.

If h is greater than the negative number

$$-\frac{1}{2}(3-\mu+\mu^2),$$

then the Hill region coincides with the entire plane $\mathbb{R}^2 = \{x, y\}$. For $\mu = 1/2$ the Hill regions are symmetric not only with respect to the x-axis, but also with respect to the y-axis.

The collinear libration points are always unstable: among the roots of the characteristic equation of the variational equations there are two real roots of different signs; two other roots are purely imaginary complex conjugates.[7] For the triangular libration points the roots of the characteristic equation are purely imaginary and are distinct only when

$$27\mu(1-\mu) < 1. \tag{2.18}$$

Under this condition the triangular relative equilibrium points are stable in the first approximation. The problem of their Lyapunov stability proved to be much more difficult; we postpone the discussion of this problem until Ch. 6. In conclusion we remark that condition (2.18) is known to be certainly satisfied for the real system Sun–Jupiter.

2.5.3 Hill's Problem

Let us choose the origin of the rotating frame of reference at the point where the body of mass μ is situated. Then the coordinates x, y of the third body

[6] Here L_1 is between the Sun and Jupiter, L_2 beyond Jupiter, and L_3 beyond the Sun.

[7] Two more purely imaginary complex conjugates are added to these roots in the spatial restricted three-body problem, where motions of the asteroid across the plane of the orbits of the Sun and Jupiter are also considered.

of small mass must be changed to $x - (1 - \mu)$, y. Renaming these variables again by x, y we see that the equations of motion have the same form (2.10), only the potential should be replaced by the function

$$V = (1-\mu)x+\frac{1}{2}(x^2+y^2)+(1-\mu)(1+2x+x^2+y^2)^{-1/2}+\mu(x^2+y^2)^{-1/2}. \quad (2.19)$$

We now make another simplification of the problem, which was introduced by Hill and is taken from astronomy. Let the body of mass $1-\mu$ again denote the Sun, μ the Earth, and suppose that the third body of negligible mass – the Moon – moves near the point $(0, 0)$, where the Earth is invariably situated. We neglect in (2.17) all the terms of order at least two in x, y. This is equivalent to discarding in (2.19) the terms of order at least three in x, y. With the required accuracy, V is replaced by the function

$$V = \frac{\mu}{2}(x^2 + y^2) + \frac{3}{2}(1 - \mu)x^2 + \mu(x^2 + y^2)^{-1/2}.$$

Since the mass of the Earth μ is much smaller than the mass of the Sun $1-\mu$, we can neglect the first summand in this formula.

It is convenient to change the units of length and mass by making the substitutions

$$x \to \alpha x, \qquad y \to \alpha y, \qquad \mu \to \beta\mu, \qquad 1-\mu \to \beta(1-\mu),$$

where

$$\alpha = \left(\frac{\mu}{1-\mu}\right)^{1/3}, \qquad \beta = (1-\mu)^{-1}.$$

After this transformation the equations of motion of the Moon take the form

$$\ddot{x} - 2\dot{y} = \frac{\partial V}{\partial x}, \qquad \ddot{y} + 2\dot{x} = \frac{\partial V}{\partial y}; \qquad V = \frac{3}{2}x^2 + (x^2 + y^2)^{-1/2}. \quad (2.20)$$

These equations have a first integral – the Jacobi integral

$$\frac{\dot{x}^2 + \dot{y}^2}{2} - V(x, y) = h.$$

It is easy to see that on passing from the restricted three-body problem to its limiting variant called *Hill's problem* the two triangular and one collinear libration points disappear. Indeed, the system of equations $V'_x = V'_y = 0$ has only two solutions $(x, y) = (\pm 3^{-1/3}, 0)$. The Hill regions

$$\{V(x, y) + h \geqslant 0\}$$

are symmetric with respect to the x- and y-axes for all values of h. If $h \geqslant 0$, then the Hill region coincides with the entire plane. For $h < 0$ the boundary has asymptotes parallel to the y-axis: $x = \pm\left(-\frac{2}{3}h\right)^{1/2}$. The form of the Hill regions depends on whether the constant $(-h)$ is greater than, equal to, or less than the unique critical value of the function V, which is equal to $\frac{3}{2}\,3^{1/3}$. These three cases are shown in Fig. 2.12 (the Hill regions are shaded). Only case (a) is of interest for astronomical applications and, moreover, only the domain around the origin.

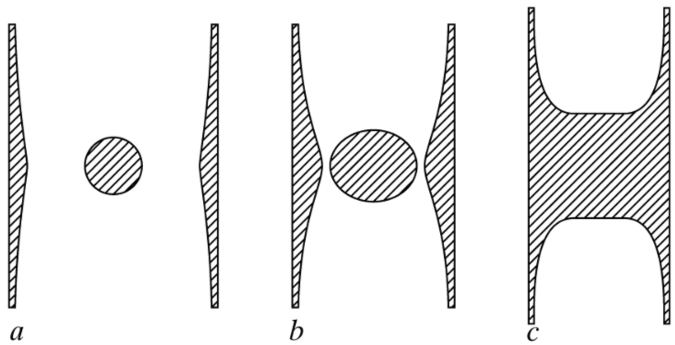

Fig. 2.12.

We now consider the questions related to regularization of Hill's problem. For that we pass to the new parabolic coordinates by the formulae $x = \xi^2 - \eta^2$, $y = 2\xi\eta$ and change the time variable $t \mapsto \tau$ along the trajectories:

$$\frac{dt}{d\tau} = 4(\xi^2 + \eta^2).$$

Denoting differentiation with respect to τ by prime we write down the equations of motion in the new variables:

$$\xi'' - 8(\xi^2 + \eta^2)\eta' = \widehat{V}'_\xi, \qquad \eta'' + 8(\xi^2 + \eta^2)\xi' = \widehat{V}'_\eta,$$

where

$$\widehat{V} = 4 + 4(\xi^2 + \eta^2)h + 6(\xi^2 - \eta^2)(\xi^2 + \eta^2).$$

The energy integral takes the form

$$\frac{\xi'^2 + \eta'^2}{2} - \widehat{V}(\xi, \eta, h) = 0.$$

This regularization of Hill's problem suggested by Birkhoff allows one to easily investigate the analytic singularities of solutions corresponding to collisions of the Moon with the Earth. Suppose that a collision occurs at time $t = 0$ and let $\tau(0) = 0$. Then obviously,

$$\xi = \left(\sqrt{8}\sin\alpha\right)\tau + \cdots, \qquad \eta = \left(\sqrt{8}\cos\alpha\right)\tau + \cdots; \qquad t = \frac{32}{3}\tau^3 + \cdots,$$

where α is an integration constant. Thus, the new time τ is a uniformizing variable and, as in the case of binary collisions in the general three-body problem, the solution $\xi(t)$, $\eta(t)$ admits a unique real analytic continuation after the collision.

As already mentioned, only the motions that take place near the point $\xi = \eta = 0$ are of interest for astronomy. For large negative values of h it is convenient to pass to the new variables

$$\varphi = 2\xi\left[-2h - 3(\xi^2 - \eta^2)^2\right]^{1/2}, \qquad \psi = 2\eta\left[-2h - 3(\xi^2 - \eta^2)^2\right]^{1/2}.$$

After this change of variables the energy integral takes quite a simple form

$$\xi'^2 + \eta'^2 + \varphi^2 + \psi^2 = 8.$$

This is the equation of a three-dimensional sphere in the four-dimensional phase space of the variables $\xi', \eta', \varphi, \psi$. Since points (ξ, η) and $(-\xi, -\eta)$ correspond to the same point in the (x, y)-plane, the Moon's states $(\xi', \eta', \varphi, \psi)$ and $(-\xi', -\eta', -\varphi, -\psi)$ should be identified. As a result we have obtained that for large negative h the connected component of the three-dimensional energy level that we are interested in is diffeomorphic to the three-dimensional projective space. This remark is of course valid for all $h < -\frac{3}{2}\sqrt[3]{3}$.

In conclusion we discuss periodic solutions of Hill's problem, which have important astronomical applications. The question is about the periodic solutions $x(t)$, $y(t)$ close to the Earth (the point $x = y = 0$) with a small period ϑ whose orbits are symmetric with respect to the x- and y-axes. More precisely, the symmetry conditions are defined by the equalities

$$x(-t) = x(t) = -x\left(t + \frac{\vartheta}{2}\right), \qquad y(-t) = -y(t) = y\left(t + \frac{\vartheta}{2}\right).$$

Consequently, these solutions should be sought in the form of the trigonometric series

$$x(t) = \sum_{n=-\infty}^{\infty} a_n(m) \cos{(2n+1)}\frac{t}{m}, \qquad y(t) = \sum_{n=-\infty}^{\infty} a_n(m) \sin{(2n+1)}\frac{t}{m},$$

where

$$m = \frac{\vartheta}{2\pi}.$$

Substituting these series into the equations of motion (2.20) we obtain an infinite nonlinear system of algebraic equations with respect to infinitely many unknown coefficients. Hill (1878) showed that this system has a unique solution, at least for small values of m (see [46, 52]). The value $m_0 = 0.08084\ldots$ for the real Moon lies in this admissible interval. The convergence of Hill's series was proved by Lyapunov in 1895.

One can show that the following asymptotic expansions hold for the coefficients $a_k(m)$:

$$a_0 = m^{2/3}\left(1 - \frac{2}{3}m + \frac{7}{18}m^2 - \cdots\right),$$

$$\frac{a_1}{a_0} = \tfrac{3}{16}m^2 + \tfrac{1}{2}m^3 + \cdots , \qquad \frac{a_{-1}}{a_0} = -\frac{19}{16}m^2 - \frac{5}{3}m^3 - \cdots ,$$

$$\frac{a_2}{a_0} = \tfrac{25}{256}m^4 + \cdots , \qquad \frac{a_{-2}}{a_0} = 0 \cdot m^4 + \cdots , \qquad \ldots .$$

This shows that for small values of m the main contribution to Hill's periodic solutions is given by the terms

$$x_0(t) = m^{2/3}\cos\frac{t}{m}, \qquad y_0(t) = m^{2/3}\sin\frac{t}{m},$$

which represent the law of motion of the Moon around the Earth without taking into account the influence of the Sun. The presence of the coefficient $m^{2/3}$ is a consequence of Kepler's third law. For small values of the parameter m, the Sun, perturbing the system Earth–Moon, does not destroy the periodic circular motions of the two-body problem, but merely slightly deforms them.

2.6 Ergodic Theorems of Celestial Mechanics

2.6.1 Stability in the Sense of Poisson

Let (M, S, μ) be a complete space with measure; here S is the σ-algebra of subsets of M, and μ a countably additive measure on S. Let g be a measure-preserving automorphism of the set M. We call the set

$$\Gamma_p = \bigcup_{n\in\mathbb{Z}} g^n(p), \qquad g^0(p) = p$$

the trajectory of a point $p \in M$, and

$$\Gamma_p^+ = \bigcup_{n\geqslant 0} g^n(p)$$

its positive semitrajectory.

Poincaré's Recurrence Theorem. *Suppose that $\mu(M) < \infty$. Then for any measurable set $V \in S$ of positive measure there exists a set $W \subset V$ such that $\mu(W) = \mu(V)$ and for every $p \in W$ the intersection $\Gamma_p^+ \cap W$ consists of infinitely many points.*

Following Poincaré we apply this result to the restricted three-body problem. In the notation of the preceding section the equations of motion of the asteroid have the form of Lagrange's equations with the Lagrangian

$$L = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + x\dot{y} - y\dot{x} + V, \qquad V = \frac{x^2 + y^2}{2} + \frac{1-\mu}{\rho_1} + \frac{\mu}{\rho_2}.$$

These equations can be represented in the Hamiltonian form with the Hamiltonian

$$H = \frac{1}{2}\big(X^2 + Y^2\big) + yX - xY - G, \qquad G = \frac{\mu}{\rho_1} + \frac{1-\mu}{\rho_2},$$

where $X = \dot{x} - y,\ \ Y = \dot{y} + x$ are the canonical momenta conjugate to the coordinates x, y. By Liouville's theorem the phase flow of this system, which we denote by $\{g^t\}$, preserves the ordinary Lebesgue measure in $\mathbb{R}^4 = \{X, Y, x, y\}$.

Consider the set of all points of the phase space for which the inequality $c_1 < -H < c_2$ holds, where c_1 and c_2 are sufficiently large positive constants. As we saw in § 2.5, under this assumption a point (x, y) belongs to one of the three connected subregions of the Hill region $\{V \geqslant c_1\}$. We choose one of the two domains containing the Sun or Jupiter. The corresponding connected domain in the phase space is clearly invariant under the action of g^t. From this domain we delete the collision trajectories, whose union has zero measure. We denote the remaining set by M and claim that M has finite measure. Indeed, the coordinates (x, y) of points in M belong to a bounded subset of the plane $\mathbb{R}^2$. The admissible momenta X, Y satisfy the inequalities

$$2(V - c_2) < (Y + x)^2 + (X - y)^2 < 2(V - c_1),$$

which follow from the Jacobi integral. In the plane $\mathbb{R}^2$ with Cartesian coordinates X, Y these inequalities define a circular annulus, whose area is at most $2\pi(c_2 - c_1)$. These remarks imply that $\mu(M)$ is finite. Therefore we can apply Poincaré's recurrence theorem: for almost every $p \in M$ the semitrajectory $\{g^t(p),\ t \geqslant 0\}$ intersects any neighbourhood of the point p for arbitrarily large values of t. Poincaré called such motions *stable in the sense of Poisson.*

We give a quantitative version of Poincaré's recurrence theorem, which was established in [140, 443] (see also [548]) for the case where M is an n-dimensional smooth manifold.

Theorem 2.7. *Suppose that a positive function $\psi(t)$ is arbitrarily slowly increasing to $+\infty$ as $t \to +\infty$, and $\psi(t)/t^{1/n}$ is monotonically decreasing to zero as $t \to +\infty$. Then for almost every $x \in M$ there exists a sequence $t_\nu \to +\infty$ such that*

$$\rho(g^{t_\nu} x, x) < t_\nu^{-1/n} \psi(t_\nu).$$

Here ρ is some distance on M. In [443] an example is given of a volume-preserving translation g on the n-dimensional torus $\mathbb{T}^n$ such that

$$\rho(g^t x, x) > C t^{-1/n}, \qquad C = \text{const}$$

for all $t \in \mathbb{N}$ and $x \in \mathbb{T}^n$.

2.6.2 Probability of Capture

Again let V be a measurable set of positive measure. For $n \in \mathbb{N}$ we denote by V^n the set of points in V such that $g^k(p) \in V$ for all $0 \leqslant k \leqslant n$. Obviously, $V^{n_1} \supset V^{n_2}$ if $n_1 < n_2$. The set

$$B = \bigcap_{n \geqslant 0} V^n$$

is measurable and $\mu(B) < \infty$. If $p \in B$, then of course $\Gamma_p^+ \in V$ for all $n \geqslant 0$.

Let $B^n = g^n(B)$. All the sets B^n are measurable, and again $B^{n_1} \supset B^{n_2}$ if $n_1 < n_2$. The set

$$D = \bigcap_{n \geqslant 0} B^n \subset B$$

is also measurable. If $p \in D$, then clearly $\Gamma_p \in V$.

Proposition 2.7. $\mu(B \setminus D) = 0$.

For this assertion not to be vacuous, one has to show first that $\mu(B) > 0$. But in concrete problems the proof of this fact may turn out to be a considerable difficulty. Proposition 2.7, which goes back to Schwarzschild, is of course valid also in the case where the time n is continuous.

For example, suppose that the system Sun–Jupiter has "captured" from the surrounding space the asteroids (the "Greeks" and "Trojans") into a neighbourhood of the triangular libration points. Proposition 2.7 immediately tells us that the probability of this event is zero. Thus, the phenomena of "capture" in celestial mechanics should be considered only in mathematical models that take into account dissipation of energy.

The following argument of Littlewood is a more interesting application. Consider the n-body problem with the centre of mass at rest. The motion of the points is described by a Hamiltonian system; the Hamiltonian function H is regular in the domain where the mutual distances are $r_{kl} > 0$. For arbitrary $c > 1$ we consider the open set $A(c)$ of points of the phase space where

$$c^{-1} < r_{kl} < c \qquad (1 \leqslant k < l \leqslant n), \qquad -c < H < c.$$

Since $A(c)$ is bounded, we have $\mu(A(c)) < \infty$. Consequently, by Proposition 2.7 the set $B(c)$ of points remaining in $A(c)$ for $t \geqslant 0$ is larger merely by a set of measure zero than the set $D(c)$ of points that are in $A(c)$ for all $t \in \mathbb{R}$.

If $c_1 < c_2$, then clearly $A(c_1) \subset A(c_2)$, $B(c_1) \subset B(c_2)$, and $D(c_1) \subset D(c_2)$. Hence the corresponding assertion is also valid for the sets

$$A = \bigcup_{c>1} A(c), \qquad B = \bigcup_{c>1} B(c), \qquad D = \bigcup_{c>1} D(c).$$

For points $p \in B$ the mutual distances r_{kl} for all $t \geqslant 0$ remain bounded above and below by some positive constants depending on p. For points $p \in D$ this property holds for all values of t. Almost all points of D belong to B.

For example, suppose that a planet system was stable "in the past". If it captures a new body, say, a speck of dust arriving from infinity, then the resulting system of bodies will no longer have the stability property: with probability one, either a collision will occur or one of the bodies will again move away to infinity. Moreover, it is not necessarily the speck of dust that will leave the Solar System; it may be Jupiter or even the Sun that may be ejected.

2.7 Dynamics in Spaces of Constant Curvature

2.7.1 Generalized Bertrand Problem

The potential of gravitational interaction has two fundamental properties. On the one hand, this is a harmonic function in three-dimensional Euclidean space (which depends only on the distance and satisfies Laplace's equation). On the other hand, only this potential (and the potential of an elastic spring) generates central force fields for which all the bounded orbits are closed (Bertrand's theorem). It turns out that these properties can be extended to the more general case of three-dimensional spaces of constant curvature (the three-dimensional sphere $\mathbb{S}^3$ and the Lobachevskij space $\mathbb{L}^3$).

For definiteness we consider the case of a three-dimensional sphere. Suppose that a material particle m of unit mass moves in a force field with potential V depending only on the distance between this particle and a fixed point $M \in \mathbb{S}^3$. This problem is an analogue of the classical problem of motion in a central field. Let θ be the length (measured in radians) of the arc of a great circle connecting the points m and M. Then V is a function depending only on the angle θ. Laplace's equation must be replaced by the Laplace–Beltrami equation:

$$\Delta V = \sin^{-2}\theta \frac{\partial}{\partial \theta}\left(\sin^{2}\theta \frac{\partial V}{\partial \theta}\right) = 0.$$

This equation can be easily solved:

$$V = -\gamma \frac{\cos\theta}{\sin\theta} + \alpha; \qquad \alpha, \gamma = \text{const.} \tag{2.21}$$

The additive constant α is inessential. For definiteness we consider the case $\gamma > 0$. The parameter γ plays the role of the gravitational constant. Apparently, the potential (2.21) was for the first time considered by Schrödinger for the purposes of quantum mechanics [536]. In addition to the attracting centre M this force field has a repelling centre M' at the antipodal point (when $\theta = \pi$). If we regard this force field as a stationary velocity field of a fluid on $\mathbb{S}^3$, then the flux of the fluid across the boundary of any closed domain not containing the points M or M' is equal to zero. These singular points M and M' can be interpreted as a sink and a source.

In the general case, where V is an arbitrary function of θ, the trajectories of the point m lie on the two-dimensional spheres $\mathbb{S}^2$ containing the points M and M'. This simple fact is an analogue of Corollary 1.3 (in § 1.1), which relates to motion in Euclidean space.

It is also natural to consider the generalized Bertrand problem: among all potentials $V(\theta)$ determine those in whose field almost all orbits of the point m on a two-dimensional sphere are closed. This problem (from various viewpoints and in various generality) was solved in [177, 281, 351, 557]. The solution of the generalized Bertrand problem (as in the classical case) is given by the two potentials

$$V_1 = -\gamma \cot \theta, \qquad V_2 = \frac{k}{2} \tan^2 \theta; \qquad k, \gamma = \text{const} > 0.$$

The first is an analogue of the Newtonian potential and the second is an analogue of Hooke's potential (with k being the "elasticity coefficient"). As shown in [351], the generalized Bohlin transformation (see § 2.1.3) takes the trajectories of the particle in the field with potential V_1 to the trajectories of the particle in the field with potential V_2.

Since the orbits are closed in these two problems, by Gordon's theorem [263] the periods T of revolution in the orbits depend only on the energy h. We now give explicit formulae for the function $T(h)$ obtained in [343].

It is well known that in Euclidean space the period of oscillations of a weight on an elastic spring is independent of the energy. This is no longer true for the sphere:

$$T = \frac{2\pi}{\sqrt{k + 2h}}.$$

For the potential of Newtonian type the dependence of the period on the energy is given by the formula

$$T = \frac{\pi}{\sqrt{\gamma}} \frac{\sqrt{\frac{h}{\gamma} + \sqrt{\frac{h^2}{\gamma^2} + 1}}}{\sqrt{\frac{h^2}{\gamma^2} + 1}}. \tag{2.22}$$

The case of the Lobachevskij space can be considered in similar fashion.

2.7.2 Kepler's Laws [177, 343]

First law. *The orbits of a particle are quadrics on $\mathbb{S}^3$ with a focus at the attracting centre M.*

A quadric is the intersection line of the sphere with a cone of the second order whose vertex coincides with the centre of the sphere. Spherical quadrics have many properties typical of conic sections on Euclidean plane (see, for example, [105]). In particular, one can speak about their foci F_1 and F_2: any ray of light outgoing from F_1 on reflection from the quadric necessarily passes through the point F_2 (rays of light are, of course, great circles on $\mathbb{S}^2$).

It was shown in [351] that the orbits of the generalized Hooke problem (the motion of a point in the field with potential $k(\tan^2\theta)/2$) are also quadrics whose centres coincide with the attracting centre M.

At each instant there is a unique arc of a great circle connecting the centre M and the material point m (the "position vector" of the point m). Unfortunately one cannot claim that the area on $\mathbb{S}^2$ swept out by this arc is uniformly increasing with time. To improve this situation we introduce an imaginary point m' by replacing the spherical coordinates θ, φ of the point m by $2\theta, \varphi$. Clearly the point m' is at double the distance from the attracting centre M.

Second law. *The arc of a great circle connecting M and m' sweeps out equal areas on the sphere in equal intervals of time.*

This law is of course valid for the motion in any central field on a surface of constant curvature.

Let F_1 and F_2 be the foci of a quadric. There is a unique great circle of the sphere $\mathbb{S}^2$ passing through these points. The quadric divides this circle into two parts; the length of each of these two arcs may be called the major axis of the quadric. Their sum is of course equal to 2π.

Third law. *The period of revolution in an orbit depends only on its major axis.*

The main point of the proof is in verifying the equality

$$\tan a = -\frac{\gamma}{h}, \tag{2.23}$$

where a is the length of the major axis. Then it remains to use the formula for the period (2.22). Note that relation (2.23) does not depend on which of the two major axes of the quadric is chosen.

In [343] an analogue of Kepler's equation was obtained connecting the position of the body in an orbit and the time of motion. The "eccentric" and "mean" anomalies were introduced based on appropriate spheroconical coordinates on $\mathbb{S}^2$ and elliptic functions.

2.7.3 Celestial Mechanics in Spaces of Constant Curvature

Having the formula for the interaction potential of Newtonian type (2.21), we can define the potential energy of n gravitating points with masses $m_1, \dots, m_n$:

$$V = -\sum_{i<j} \gamma m_i m_j \cot\theta_{ij}, \tag{2.24}$$

where θ_{ij} is the distance between the points m_i and m_j on the three-dimensional sphere. Formula (2.24) allows one to write down the equation of motion of n gravitating points on $\mathbb{S}^3$.

This problem has many common features with the classical n-body problem in Euclidean space. However, there are also essential differences. First, the two-body problem on $\mathbb{S}^3$ proves to be non-integrable: there are not sufficiently many first integrals for its solution and its orbits look quite complicated (see [137]). Here the main difficulty is related to the fact that the Galileo–Newton law of inertia does not hold: the centre of mass of gravitating points no longer moves along an arc of a great circle.

Furthermore, as in the classical case, binary collisions admit regularization. However, the question whether the generalized Sundman theorem is valid for the three-body problem in spaces of constant curvature remains open. This question essentially reduces to the problem of elimination of triple collisions. Recall that in the ordinary three-body problem the absence of simultaneous collisions is guaranteed by a non-zero constant value of the angular momentum of the system of n points with respect to their centre of mass (Theorem 2.3).

Of interest is the problem of finding partial solutions for n gravitating bodies in spaces of constant curvature (similar to the classical solutions of Euler and Lagrange). Results in this direction can be found in the book [137]. The restricted three-body problem was studied in this book: relative equilibria were found and the Hill regions were constructed.

2.7.4 Potential Theory in Spaces of Constant Curvature

As established by Newton, a homogeneous sphere in three-dimensional Euclidean space does not attract interior points, and the exterior points are attracted as if by a single material point located at the centre of the sphere whose mass is equal to the mass of the sphere. Newton's theorem on the sphere immediately implies that a homogeneous ball attracts points in the exterior in the same way as if its mass was concentrated at the centre, while the attraction force on interior points depends linearly on the distance to the centre (by Hooke's law).

It is also known that the level surfaces of the gravitational potential of a homogeneous rod is a confocal family of ellipsoids of revolution whose foci coincide with the ends of the rod. This result was generalized by Ivory.

Consider an infinitesimally thin homogeneous layer between two similar concentric ellipsoids with common centre and the same directions of the axes, which is called an elliptic layer. It turns out that the gravitational potential inside the elliptic layer is constant (*Newton's theorem* generalizing the theorem on the gravitation of a sphere) and the level surfaces of the potential in the exterior are ellipsoids confocal to the layer (*Ivory's theorem*). The proofs can be found, for example, in [83].

It is easier to think of an elliptic layer as an ellipsoid with, generally speaking, non-constant *homeoid* density

$$\frac{d\sigma}{|\nabla f|},$$

where $d\sigma$ is the area element of the ellipsoid and ∇f is the gradient of the quadratic form defining the surface of the ellipsoid. The homeoid density of a sphere and of a segment is obviously constant.

It turns out that the basic theorems of the theory of Newtonian potential in $\mathbb{E}^3$ can be carried over (with certain reservations) to the case of a space of constant curvature [349]. For definiteness we consider the case of a sphere, which is a space with non-trivial topology. A gravitational potential of Newtonian type is defined by (2.24).

Consider on $\mathbb{S}^3$ a two-dimensional sphere $\mathbb{S}^2$ with a homogeneous mass distribution. Let θ be the latitude on $\mathbb{S}^3$ measured from the centre of $\mathbb{S}^2$, so that $\mathbb{S}^2 = \{\theta = \theta_0\}$, $0 < \theta_0 \leqslant \pi/2$. One should bear in mind that on $\mathbb{S}^3$ there is another sphere $\mathbb{S}^2_- = \{\theta = \pi - \theta_0\}$ congruent to $\mathbb{S}^2$, whose points produce a repulsive action. The three-dimensional sphere $\mathbb{S}^3$ is divided by the two-dimensional spheres $\mathbb{S}^2$ and $\mathbb{S}^2_-$ into three connected domains.

The following analogue of Newton's theorem holds: the sphere $\mathbb{S}^2$ does not attract points lying "inside" $\mathbb{S}^2$ and $\mathbb{S}^2_-$, while "exterior" points are attracted in exactly the same way as if the sphere was replaced by a single point at the centre of $\mathbb{S}^2$, with mass equal to the mass of the whole sphere. This immediately implies the theorem on the gravitation of a homogeneous ball bounded by the sphere $\theta = \theta_0$ $(\theta_0 < \pi/2)$: exterior points $(\theta_0 \leqslant \theta \leqslant \pi - \theta_0)$ are attracted in the same way as if the mass of the ball was concentrated at the centre. The potential inside the ball of unit density is given by the formula

$$\frac{\pi\gamma(2\theta - \sin 2\theta)\cos\theta}{\sin\theta},$$

which is different from Hooke's potential $(k/2)\tan^2\theta$. Only for small θ we obtain the potential of elastic interaction

$$\frac{4\gamma\pi\theta^2}{3} + o(\theta^2).$$

In the case of Euclidean space the problem of the gravitation of a segment is essentially a planar one: in any plane containing the gravitating segment the level lines of the potential form a family of ellipses with foci at the ends of the segment. The situation is similar for a space of constant curvature. In the case of $\mathbb{S}^3$ the role of a plane is played by a two-dimensional sphere of unit radius.

Thus, on the two-dimensional sphere

$$x^2 + y^2 + z^2 = 1 \tag{2.25}$$

we consider a segment – an arc of a great circle with end points $F_1 = (\alpha, \beta, 0)$ and $F_2 = (\alpha, -\beta, 0)$. Of course, $\alpha^2 + \beta^2 = 1$. To make the arc uniquely determined we assume that it contains the point with coordinates $(1, 0, 0)$. This arc admits the parametrization

$$x = \sin\varphi, \qquad y = \cos\varphi, \qquad z = 0; \qquad \frac{\pi}{2} - \varphi_* \leqslant \varphi \leqslant \frac{\pi}{2} + \varphi_*,$$

where $\cos\varphi_* = \alpha$, $\sin\varphi_* = \beta$.

At a point with coordinates x, y, z the value of the potential (up to a constant factor) is equal to

$$V = \int_{\pi/2-\varphi_*}^{\pi/2+\varphi_*} \frac{\cos\widetilde{\theta}}{\sin\widetilde{\theta}}\, d\varphi,$$

where $\cos\widetilde{\theta} = x\sin\varphi + y\cos\varphi$. As an analogue of a confocal family of ellipses we have the family of ovals which is the result of the intersection of the cones

$$\frac{c^2x^2}{c^2-\alpha^2} + \frac{c^2y^2}{c^2+\beta^2} + z^2 = 0 \tag{2.26}$$

and the sphere (2.25); here c is a parameter. As $c \to 0$, the ovals converge to the original segment. As already mentioned, by analogy with the Euclidean case these ovals may be called spherical conics with foci at the points F_1 and F_2.

One can verify that the level lines of the potential V created by an arc of a great circle on $\mathbb{S}^2$ is a family of spherical conics with foci at the ends of the arc.

These observations admit a generalization. Let A be a symmetric operator in Euclidean space $\mathbb{R}^4$, and I the identity operator. The operator $(A-\lambda I)^{-1}$ (the resolvent of A, where λ is a spectral parameter) is also a symmetric operator, which defines the pencil of quadratic forms

$$f(x) = \bigl((A-\lambda I)^{-1}x, x\bigr).$$

Equating these forms to zero we obtain a family of cones, which intersect the three-dimensional sphere

$$g(x) = (x, x) = 1$$

in two-dimensional surfaces. These surfaces may be called spherical confocal quadrics.

Example 2.2. Dividing equation (2.26) by c^2 we obtain equations of the form $f(x) = 0$, where $A = \operatorname{diag}\,(-\alpha^2, \beta^2, 0)$, $\lambda = c^2$. △

On the quadrics one can define the homeoid density

$$\frac{d\sigma}{W_2},$$

where $d\sigma$ is the area element of the quadric as a surface in $\mathbb{R}^4$ and W_2 is the Euclidean area of the parallelogram constructed on the gradients of the functions f and g as vectors.

Theorem 2.8. *Let k be a quadric on $\mathbb{S}^3$ with homeoid mass density, and k_- the antipodal quadric. The potential created by k is constant in the two ball-shaped domains on $\mathbb{S}^3$ bounded by the quadrics k and k_-, and the level surfaces of this potential in the complementary domain form a family of quadrics confocal with k.*

This is an analogue of the classical Newton–Ivory theorem.

Ivory's theorem admits a generalization to quadratic forms of other signatures (with ellipsoids replaced by the corresponding hyperboloids in $\mathbb{R}^n$); see [76, 593]. The simplest case is a one-sheet hyperboloid in $\mathbb{R}^3$.

The gravitational potentials are replaced by differential forms, whose degrees are determined by the signature. For a one-sheet hyperboloid in $\mathbb{R}^3$ the result is 2-forms that are harmonic outside the hyperboloid and whose kernels are directed along the parallels of the elliptic coordinate system in the multiply connected component of the complement of the hyperboloid, and along the meridians, in the simply connected one. These forms can also be described as flows of an incompressible fluid along the fields of the kernels of the forms.

The corresponding magnetic fields are given by the Biot–Savart integrals over currents (generalizing the homeoid density) flowing along the meridians of the surface in the first case, and along the parallels, in the second (the field created by the current in the second component is zero).