Chapter 7
Orbits

7.1 Fundamental Physics

After ascent, we are now in outer space. How does a spacecraft move under the influence of the gravitational forces of the Sun, planets, and moons? This is the question we will deal with in this chapter, and we are pursuing general answers to it. Let us face reality from the start: The details of motions are usually very complicated and can be determined sufficiently accurately only numerically on a computer. This is exactly how real missions are planned. But the goal for us is not numerical accuracy, but to understand the basic behavior of a spacecraft. To achieve this, it suffices to study some crucial cases. The easiest and by far the most important case is the mutual motion of two point-like (a.k.a. ideal) bodies in the gravitational field of each other, the so-called *(ideal) two-body problem* (2BP), which we study in this chapter, such as the Moon in the gravitational field of the Earth. More complicated cases can often be traced back to the two-body problem by minor simplifications.

Before we derive the corresponding equation of motion, solve it, and thus describe the motion of orbiting bodies, we want to gain insight into the basic principles of gravitation and show that even Newton's laws are the outcome of these.

7.1.1 Gravitational Potential

The existence of forces seems to be so self-evident that we deem them to be the foundation of nature. But appearances can be deceptive, and also Newton succumbed to this in the late seventeenth century. It is not forces that are fundamental, but so-called potentials that cause such forces. This was shown by Laplace one century later. The gravitational potential U is a property of space induced by the

U. Walter, *Astronautics*, https://doi.org/10.1007/978-3-319-74373-8_7

mass of a body and its surrounding. Like a force, you would not see it by itself. Only if you insert a test mass into this space does the potential act on it and generates an attractive force.

Poisson's Equation

The basic mutual interaction between masses and space, in which they are embedded, is described by the famous Einstein field equations of his theory of general relativity

$$G_{ik} = \frac{8\pi G}{c^4} T_{ik} \qquad \textbf{Einstein field equations}$$

> **Remark** *The Einstein field equations are the components of a tensor equation, a system of 10 coupled partial differential equations of second order in the coordinates to determine G_{ik} from the given T_{ik}. Here the cosmological constant, which recently turned out to be significant on a cosmological scale, has been neglected. You do not really have to understand this equation and the meaning of its terms. We start our considerations with Einstein's equations to show that the origin of Newton's gravitational field is the theory of general relativity.*

G_{ik} is the so-called Einstein tensor. It describes the basic structure of space, its spacetime curvature; T_{ik} is the so-called stress-energy tensor that describes the energy and the inertial moment distribution of matter or fields in space. It corresponds to the classic energy and mass density ρ; G is the *gravitational constant*, and c is the velocity of light. The Einstein field equations tell us that matter and energy of the universe on one hand, and the curvature of space on the other, determine each other. To put it in a different way: Masses tell space how to curve, and space tells the masses how to move. In contradiction to classical Newtonian mechanics, space without any ingredient (Newton's absolute space) cannot exist.

If the curvature of space is weak and the planetary motions are far below relativistic speeds, and if the pressure in the state equation of the local matter/energy distribution is small, it is possible to show that the Einstein field equations turn into the classic potential equation called *Poisson's equation* (a.k.a. *Newtonian field equation*):

$$\nabla^2 U(\boldsymbol{r}) = 4\pi G \cdot \rho(\boldsymbol{r}) \qquad \textbf{Poisson's equation} \tag{7.1.1}$$

where

$$\begin{aligned} \nabla^2 &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \\ &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) + \frac{1}{r^2\sin^2\vartheta}\frac{\partial^2}{\partial\varphi^2} \end{aligned}$$

is the so-called *Laplace operator*, expressed here in cartesian and spherical coordinates.

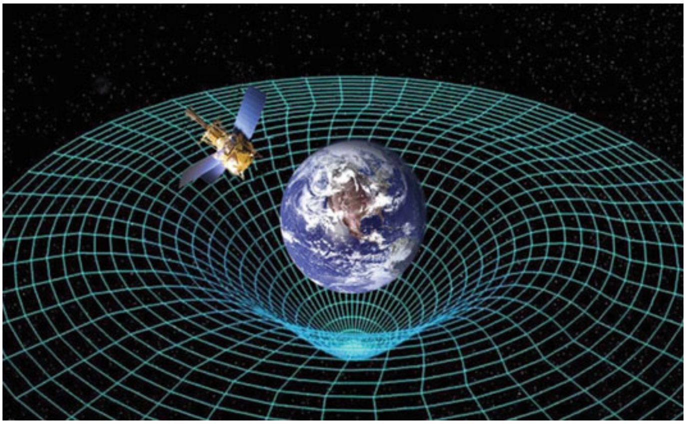

Fig. 7.1 According to the theory of general relativity mass, such as Earth, curves space, and the curvature of space is equivalent to the gravitational potential of this mass that attracts other bodies such as an orbiting satellite. Actually, space is curved (distorted) towards Earth in all three dimensions, which is hard to depict. *Credit* NASA

Equation (7.1.1) is a differential equation of second order. Reading it from right to left, it states that a given mass density $\rho(\boldsymbol{r})$ generates a gravitational potential $U(\boldsymbol{r}) \equiv U_G(\boldsymbol{r})$, which reflects the curvature of space. So, the potential can be conceived as the mass-induced curvature of space (see Fig. 7.1). As we will see in Sect. 7.1.2, the gradient of space curvature, or of the gravitational potential, respectively, in turn acts as a force on other bodies. Mathematically, the gradient of a potential is a force field. It is exactly this force field that we commonly interpret as the cause of gravitational attraction. Thus, the key statement of general relativity reads: Mass-induced gravitation and curvature of space around that mass are the same, they are just different in appearance.

Gravitational Potential of an Isotropic Mass

We now want to determine from Poisson's equation the gravitational potential of a celestial body that exhibits an isotropic mass distribution, $\rho(\boldsymbol{r}) = \rho(r)$. That is, the body is spherical with given radius R. (Gravitational anisotropies of a celestial body will be treated in Sects. 7.7.1 and 12.3.) The isotropic mass sphere of the body under consideration has total mass

$$M = 4\pi \int_0^R \rho(r) r^2 dr$$

and is generally centered at the origin O of our reference system (Fig. 7.2). This is why M is usually called the *central body*. The vector $\boldsymbol{r}$ is the radial vector, a.k.a. *position vector*, from O to any position outside the mass sphere.

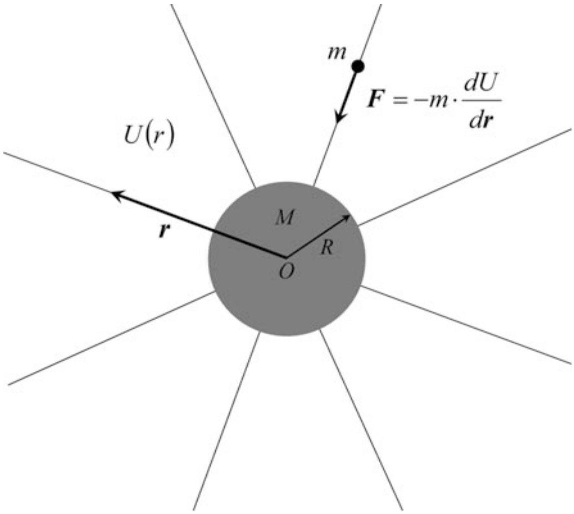

Fig. 7.2 The gravitational potential of a central body with radius *R* and mass *M* and in its surroundings with the resulting force on a test mass *m*. The potential is isotropic in all 3 dimensions

To determine the gravitational potential of such a body, we make use of the fact that the potential must also be isotropic. By adapting spherical coordinates r, ϑ, φ we hence have $\partial U(\boldsymbol{r})/\partial\vartheta = \partial U(\boldsymbol{r})/\partial\varphi = 0$ and $U(\boldsymbol{r}) = U(r)$. Thus, Eq. (7.1.1) reduces to

$$\frac{d}{dr}\left(r^2 \frac{dU}{dr}\right) = 4\pi G \rho r^2 \tag{7.1.2}$$

We now perform the radial integral over the sphere, which yields

$$r^2 \frac{dU}{dr} = 4\pi G \int_0^R \rho(r') r'^2 dr' = GM \quad @\ r \geq R$$

This is a telling interim result. It shows that the potential outside the body does not depend on the specifics of the radial density distribution, but just on the total mass of the central body. Note that this implies that the potential not only is independent of the actual radius of the sphere, but that a central body, such as the Sun, does not need to have a well-defined radius at all. We thus arrive at the famous result, already established by Newton, that

> Isotropic bodies of different size (including a point mass), but same total mass, generate the same gravitational potential outside their body.

An indefinite integration of the above equation finally provides the wanted gravitational potential

$$U(r) = -\frac{GM}{r} + U_0 \quad @\ r \geq R$$

We are free to choose the integration constant, $U_0 = U(r \to \infty)$, which is the zero reference. For instance, the potential energy of a body near Earth's surface is usually measured in terms of its altitude $h = r - R$ above Earth's surface at radius R, i.e. $U(R) = 0 \to U_0 = GM/R$, and hence $U(h) = -GM/(R+h) + U_0 \approx GMh$. In astrophysics, though, one always chooses $U_0 = 0$. So, for $r \to \infty$ the potential energy is defined to be zero, that is

$$U(r) = -\frac{\mu}{r} \quad @\ r \geq R \quad \textbf{gravitational potential} \tag{7.1.3}$$

with

$$\mu := GM \qquad \textbf{standard gravitational parameter}$$

Remark *The standard gravitational parameters of the Sun and Earth are called the* heliocentric *and* geocentric gravitational constants, *respectively. The square root of the heliocentric gravitational constant as determined from Kepler's third law* (see Eq. (7.4.12)) with $T = 365.256363$ *days and* $a = 1\ AU = 1.49597870 \times 10^{11}$ *m is called the* Gaussian gravitational constant *and deviates marginally from the heliocentric gravitational constant* (see Sect. 7.1.5).

Potential Energy

A body with small mass m placed into this gravitational potential, by definition, acquires the potential energy

$$E_{pot} := mU(r) = -\frac{\mu m}{r} \tag{7.1.4}$$

Motivated by this relation, one could also consider the gravitational potential as potential energy per mass.

Note *Here, M and m characterize the gravitational property of the masses, in contrast to their inertial property, which they also bear and to which we come in a moment.*

7.1.2 Gravitational Force

Because the potential energy varies from point to point in space and since a body tries to minimize its potential energy, a test mass m with zero initial velocity will move along the steepest descent, the gradient of the potential energy. We therefore interpret the gradient of the potential energy as a force, the gravitational force, which mathematically is described as

$$\boldsymbol{F}(\boldsymbol{r}) = -\frac{d}{d\boldsymbol{r}}E_{pot} = -m\frac{d}{d\boldsymbol{r}}U(r) \tag{7.1.5}$$

where

$$\frac{d}{d\boldsymbol{r}} \equiv grad \equiv \nabla := \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$

The negative sign occurs because the gravitational force $\boldsymbol{F} \equiv \boldsymbol{F}_G$ points in the direction of decreasing energy E_{pot}. As an illustrative example for the calculation of a gradient, let us calculate the gradient of the radial distance

$$\begin{aligned} r &= \sqrt{x^2+y^2+z^2} \\ \frac{dr}{d\boldsymbol{r}} &= \left(\frac{\partial r}{\partial x}, \frac{\partial r}{\partial y}, \frac{\partial r}{\partial z}\right) = \left(\frac{1}{2r}2x,\ \frac{1}{2r}2y,\ \frac{1}{2r}2z\right) = \frac{\boldsymbol{r}}{r} = \hat{\boldsymbol{r}} \end{aligned} \tag{7.1.6}$$

where $\hat{\boldsymbol{r}}$ is the unit vector in the direction of $\boldsymbol{r}$. Applying $U(r)$ from Eq. (7.1.3) to (7.1.5) and because of Eq. (7.1.6), we get for the gravitational force at any point in space (which we can also interpret as a gravitational force field)

$$\boldsymbol{F}(\boldsymbol{r}) = \mu m\frac{d}{d\boldsymbol{r}}\frac{1}{r} = -\frac{\mu m}{r^2}\frac{dr}{d\boldsymbol{r}}$$

yielding with Eq. (7.1.6) *Newton's law of gravitation* for the gravitational force

$$\boxed{\boldsymbol{F}(\boldsymbol{r}) = -\frac{\mu m}{r^2}\hat{\boldsymbol{r}}} \qquad \textbf{gravitational force} \tag{7.1.7}$$

It states that the gravitational force decreases with the square of the distance from the mass at the origin. We define

$$\boldsymbol{F}(\boldsymbol{r}) =: m\boldsymbol{g}(\boldsymbol{r}) = -mg(r)\hat{\boldsymbol{r}}$$

and call $\boldsymbol{g}(\boldsymbol{r})$ the **gravitational field**, which is formally the gravitational force field per unit test mass.

Central Force Properties

Any force that like the gravitational force exhibits the factorial property $\boldsymbol{F}(\boldsymbol{r}) = -F(r)\hat{\boldsymbol{r}}$ and hence

$$\boldsymbol{F}(\boldsymbol{r}) = -cf(r)\hat{\boldsymbol{r}} \quad \textbf{central force}$$

where c is the force-specific coupling constant, is called a *central force*. If $c > 0$, i.e. if it is attractive, it points to the origin. The vector field formed by the central forces at any point in space is called a **central force field**. Gravitational force and Coulomb force are two familiar examples with $f(r) \propto 1/r^2$.

Let us assume $\boldsymbol{F}$ to be an arbitrary central force with continuous function $f(r)$, and $U(r)$ is its antiderivative (primitive integral or potential function), i.e. $f(r) = dU/dr$. We then have

$$\boldsymbol{F}(\boldsymbol{r}) = -cf(r)\hat{\boldsymbol{r}} = -c\frac{dU}{dr}\hat{\boldsymbol{r}} = -c\frac{dU}{dr}\frac{dr}{d\boldsymbol{r}} = -c\frac{dU}{d\boldsymbol{r}}$$

So, we have shown that

Any central force can be written as the gradient of a central potential function.

We now show that the work done by a central force field on a body that moves between two points depends only on these points and not on the path followed. To prove this, we express the work W done by the central force field as an integral over an arbitrary path between two points A and B

$$W = \int_A^B \boldsymbol{F} \cdot d\boldsymbol{r} = -c\int_A^B \frac{dU}{d\boldsymbol{r}} \cdot d\boldsymbol{r} = -c\int_{U_A}^{U_B} dU = E_{pot,A} - E_{pot,B}$$

So, the work done depends only on the potential energy of the field at the terminal points.

Forces that obey the above definition are called **conservative forces**. Therefore, any central force is a conservative force (but not necessarily vice versa; see the following box).

Conservative Force

According to the above, one possible definition of a conservative force is:

> *The work done by a conservative force field on a body that moves between two points depends only on these points and not on the path followed.*

Note that from this it follows that:

> *The work done by a conservative force field on a body that moves on a round trip is zero.*

This is an alternative and more common definition of a conservative force. In addition, because the total energy, equaling the kinetic energy plus potential energy, must be conserved (see Sect. 7.1.3) on a round trip, we can characterize a conservative force also by:

After a round trip in a conservative force field a body regains its initial kinetic energy.

In turn, *a* ***non-conservative force*** *(a.k.a. dissipative force) can be characterized as one in which a body after a round trip has lost kinetic energy.*

Observe that the above equations for the work done also work backward if for total derivative $c \cdot dU = \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r}$ holds, i.e., if U is a function of the body position only. Then the force is $\boldsymbol{F}(\boldsymbol{r}) \equiv -c \cdot dU/d\boldsymbol{r}$ and also depends solely on position. Therefore, all solely position-dependent forces are conservative forces.

If a force is velocity-dependent, time-dependent, or dependent on any other variable, it is usually not conservative. A textbook example for a non-conservative force is friction. Here $\boldsymbol{F}(\boldsymbol{v}) = -k\boldsymbol{v}$, with $k > 0$, and hence $dW/dt = \boldsymbol{F} \cdot d\boldsymbol{r}/dt = -kv^2 < 0$, the system constantly dissipates energy.

A special case is the electromagnetic field. Here $U = \phi - \boldsymbol{v}\boldsymbol{A}$, where $\phi(\boldsymbol{r}, t)$ and $\boldsymbol{A}(\boldsymbol{r}, t)$ are scalar and vector potentials, respectively. It can be shown that for the electromagnetic Lorentz force $\boldsymbol{F} = -q(\partial U/\partial \boldsymbol{r} - d\boldsymbol{A}/dt) = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$ holds, with q the electric charge (coupling constant) of the affected particle, and $\boldsymbol{E} = -\nabla\phi - \partial\boldsymbol{A}/\partial t$ the electric field and $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ the magnetic field. So, even though $U(\boldsymbol{r}, \boldsymbol{v}, t)$, we obtain $\boldsymbol{F} = -q \cdot \partial U/\partial \boldsymbol{r}$ if $\boldsymbol{A}$ and hence $\boldsymbol{B}$ (and $\boldsymbol{E}$, which both go hand in hand) does not vary in time. In this case the electromagnetic field is also a conservative force field. If $\boldsymbol{B}$ (and $\boldsymbol{E}$) varies in time, the electromagnetic field is not conservative.

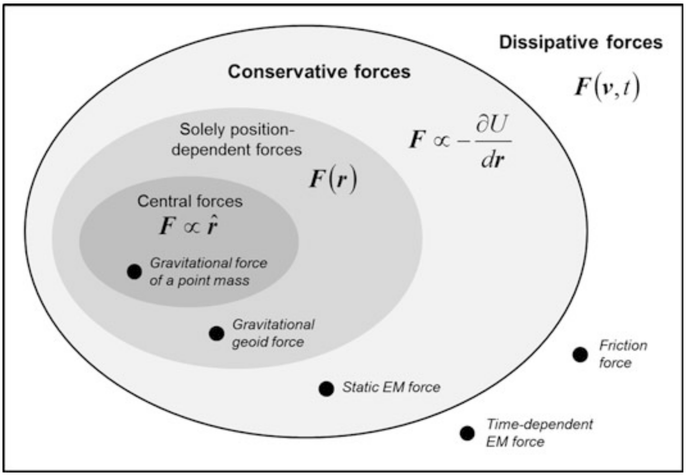

Fig. 7.3 Conservative versus dissipative forces and the different types of conservative forces with some typical examples (black dots)

We summarize all the above in Fig. 7.3 and by the following relations

$$\boldsymbol{F}(\boldsymbol{r}) = -cf(r)\hat{\boldsymbol{r}} \text{ (central force)}$$
$$\Downarrow$$
$$\boldsymbol{F}(\boldsymbol{r}) \text{ and } U(\boldsymbol{r}) \text{ are solely position-dependent}$$
$$\Downarrow$$
$$\boxed{\boldsymbol{F} = -c \cdot \partial U/\partial \boldsymbol{r} \quad \Leftrightarrow \quad \text{conservative force}}$$

A good example of a conservative force field that is even solely position-dependent, but yet not a central force, is Earth's true gravitational field (actually, the non-spherical part of it) as examined in Sect. 12.2.2.

In the above considerations we have assumed that the force in question is the only one that acts on the body. If more than one force act, the effects are separately attributable to each such force. For example, apart from the conservative gravitational force, we often have to consider external non-conservative forces such as atmospheric drag, solar radiation, or eddy fields, the last of which arise from time-dependent external magnetic fields. They change the kinetic energy of the body and hence are also called *dissipative forces*. As we will see from Sect. 7.2.1, external forces also imply that the angular momentum is no longer preserved because the system is no longer a closed system.

Earth's Gravitational Field

In Earth's environment $\boldsymbol{g}(\boldsymbol{r}) \equiv \boldsymbol{g}_\oplus(\boldsymbol{r})$ is Earth's gravitational field. Its absolute value is

$$g(r) = \frac{GM_\oplus}{r^2} = g_0 \frac{R_\oplus^2}{r^2} \qquad \textbf{Earth's mean gravitational acceleration}$$

Note *This definition contrasts with the one in physical geodesy, where g is* Earth's gravity*, which is the distinct acceleration at any point on Earth's surface. Earth's gravity g results jointly from its gravitational force plus the centrifugal force due to the rotation of the Earth and therefore is dependent on the local altitude and the local geographical latitude. Both forces in turn are derived from the (effective) geopotential, which comprises Earth's gravitational potential and Earth's centrifugal potential* (see Sect. 7.2.4). *The difference in these two definitions is reflected by the use of the discriminative words "gravitation" versus "gravity", and "Earth's gravitational potential" versus "geopotential".*

In particular, and for practical purposes, in the following we define

$$g_0 := \frac{GM_\oplus}{R_\oplus^2} = 9.7982876 \ \text{m s}^{-2} \qquad \textbf{Earth's mean gravitational acceleration at its surface}$$

where the numerical value follows from $\mu_\oplus = GM_\oplus = g_0 R_\oplus^2 = 3.98600442 \times 10^{14}\,\mathrm{m^3\,s^{-2}}$ (see Appendix A) and $R_\oplus = 6378.1363$ km, which is the equatorial scale factor of the Earth gravitational model EGM96 (see Sect. 12.2.2) equaling Earth's mean equatorial radius at any practical rate.

Note *Because Earth's mass is not perfectly homogeneous and spherically symmetrical $U_\oplus(r, \beta, \lambda)$ varies slightly with geocentric latitude β and geographic longitude λ* (see Eq. 12.2.3). *Therefore the gravitational acceleration at Earth's surface locally deviates from the mean value g_0, and in Earth's orbits $g(r, \beta, \gamma)$ locally deviates from Earth's mean gravitational acceleration $GM_\oplus / r^2$.*

While the gravitational attraction of masses hinge on this gravitational property their motion in a gravitational field is additionally determined by the inertial properties of the body. We will now show that Newton's laws, which are closely related to the inertia of a body, are also based on very fundamental properties of our universe.

7.1.3 Conservation Laws

In the literature it is common to assume Newton's laws and Newton's equation of motion as given and to apply them to gravitation and to derive the conservation of angular momentum and energy from them. This might be correct on mathematical grounds, but it does not mean that the conservation laws result from Newton's laws. It rather shows that the conservation laws also hold for motion under a gravitational force. Could that mean that they would not be valid in other cases? The conservation laws actually are very first principles in nature: In systems not affected by non-conservative interactions conservation laws are always valid. This property stems from very basic features of our universe, namely that time t and space $x,\ y,\ z$ are homogeneous, and the direction φ in space is isotropic.

Remark *According to Einstein's equations, space and time are homogeneous and isotropic because the masses are distributed evenly in the universe on a cosmic scale. All masses in the universe have to be considered here, because only in their entirety do they determine the gross spatial structure of the cosmos.*

The so-called *Noether's theorem* (Emmy Noether, 1918) of physics tells us that these basic features result in the following conservation laws.

Law of Conservation of Energy

Homogeneity of time, that is, the invariance of the physical action integral against continuous time shifts $t \to t + \delta t$, results in the conservation of energy

$$\boxed{\sum_{\text{all energy forms } j \text{ of all bodies } i} E_{ji} = const} \tag{7.1.8}$$

Law of Conservation of Linear Momentum

Homogeneity of space, that is, the invariance of the action integral against continuous spatial shifts $\boldsymbol{r} \rightarrow \boldsymbol{r} + \delta \boldsymbol{r}$, results in the conservation of linear momentum

$$\boxed{\sum_{all\ bodies\ i} \boldsymbol{p}_i = const} \tag{7.1.9}$$

Law of Conservation of Angular Momentum

Isotropy of the direction in space, that is, the invariance of the action integral against continuous spatial rotations $\varphi \rightarrow \varphi + \delta\varphi$, results in the conservation of angular momentum

$$\boxed{\sum_{all\ bodies\ i} \boldsymbol{L}_i = const} \tag{7.1.10}$$

Remark *You do not need to understand why symmetries correspond to conservation laws. Here is just a short summary. The pair of variables (energy, time), (linear momentum, location), and (angular momentum, rotation angle) are so-called "canonically conjugated parameters", generally written as (p_i, q_i), for every particle i. If one takes the difference between kinetic and potential energies for all particles under consideration, which is called the Lagrangian L, then from the energy minimization principle Euler's equation is: $d(\partial L/\partial \dot{q}_i)/dt - \partial L/\partial q_i = 0$ with $p_i \equiv \partial L/\partial \dot{q}_i$. The invariance of the universe and hence of its Lagrangian L with regard to the shifts $q_i \rightarrow q_i + \delta q_i$ implies $d\left(\sum_i \partial L/\partial q_i\right)/dt = 0$, which in turn implies $\sum_i \partial L/\partial \dot{q}_i = \sum_i p_i = const$. These are the said conservation laws.*

7.1.4 Newton's Laws of Motion

We are now set to derive the equation of motion in a gravitational field. First, it is important to note that Eq. (7.1.5) generally describes the relation between any type of energy and the force derived from it. So when taking the gradient of the energy conservation Eq. (7.1.8) and employing Eq. (7.1.5) we get for our test mass m (i = 1)

$$0 = \sum_j -\frac{dE_j}{d\boldsymbol{r}} = \sum_j \boldsymbol{F}_j \quad \textbf{Newton's third law} \tag{7.1.11}$$

The running index j indicates all relevant energies and consequent forces. This equation states that the sum of all forces that a mass is subject to vanishes. This is a generalization of Newton's third law that Newton established for only one acting force causing a reacting force: action equals reaction. The energies relevant to our point mass are: potential energy in the gravitational field, E_{pot}, and kinetic energy, E_{kin}; there are possibly also other energies from electric, magnetic, or chemical

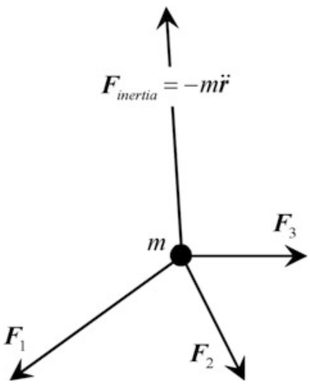

Fig. 7.4 According to Newton's second law, a mass *m* moves such that the inertial force caused by this accelerated motion counterbalances all other external forces acting on the mass

potentials that we will, however, neglect for our further considerations. The gravitational force derived from the potential energy is already described in Eq. (7.1.7). What is still missing though is the force derived as a gradient from the kinetic energy

$$\frac{dE_{kin}}{d\boldsymbol{r}} = \frac{1}{2} m \frac{d\boldsymbol{v}^2}{d\boldsymbol{r}} = m\boldsymbol{v}\frac{d\boldsymbol{v}}{d\boldsymbol{r}} = m\frac{d\boldsymbol{r}}{dt}\frac{d\boldsymbol{v}}{d\boldsymbol{r}} = m\frac{d\boldsymbol{v}}{dt} = m\ddot{\boldsymbol{r}}$$

According to Eq. (7.1.11) this is the so-called *inertial force*

$$\boldsymbol{F}_{inert} = -\frac{dE_{kin}}{d\boldsymbol{r}} = -m\ddot{\boldsymbol{r}} \qquad \textbf{inertial force}$$

Observe that the inertial force is antiparallel to the direction of acceleration: When you are push-starting a car, you accelerate it forward, but its inertial force pushes backwards against your palms. This derivation shows that

The inertial force is the force field of the kinetic energy.

Note *Here, m now characterizes the inertial property of the mass.*

Because external forces are a given, while a body can acquaint any state of motion, we see that a proper acceleration and hence an adjustable inertial force is the elegant means by which nature constantly complies with Newton's third law (see Fig. 7.4). Inserting this result into Eq. (7.1.11), one gets the Newton's well-known second law

$$\boxed{m\ddot{\boldsymbol{r}} = \sum_j \boldsymbol{F}_j} \qquad \textbf{Newton's second law} \qquad (7.1.12)$$

where the summation is over all external forces.

Remark *To be precise, Newton's second law states that* $\boldsymbol{F} = d\boldsymbol{p}/dt$*. But since* $\boldsymbol{p} = m\boldsymbol{v}$*, this together with* Eq. (7.1.11) *is equivalent to Eq.* (7.1.12).

If the external forces vanish, Eq. (7.1.12) reduces to $\ddot{\boldsymbol{r}} = 0$ with the solution

$$\boldsymbol{r} = \boldsymbol{v}_0 t + \boldsymbol{r}_0 \qquad \textbf{Newton's first law}$$

where $\boldsymbol{v}_0$ and $\boldsymbol{r}_0$ are the initial values of our mass m. This equation states that

> *Every body persists in a state of rest or of uniform motion in a straight line unless it is compelled to change that state by forces impressed on it.*

These are the words of Newton, by which he described his first law.
In conclusion, we have shown that

> Classical Newtonian physics, in particular the equation of motion in a gravitational field, is an outcome of the theory of general relativity by taking into account the homogeneity and isotropy properties of space and time in our universe.

Significance of Newton's Second Law
At first glance, Newton's second law (Eq. (7.1.12)) seems inconspicuous. Yet, it is the general starting point for the description of the classical motion of a body. It therefore deserves a second glance. It tells us two things. First, it ensures that whatever the external forces acting on a body may be, the inertial force $\boldsymbol{F}_{inertial} = -m\ddot{\boldsymbol{r}}$ will be such that it cancels them all out. Inertia therefore is nature's wild card to achieve this. The second point is that inertial force is available only at the expense of accelerated motion of the body. So,

> Accelerated motion, and with it inertial force, is the natural response of a body to external forces. Its trajectory $\boldsymbol{r}(t)$ will generally be determined by solving the differential equation of motion as given by Newton's second law.

This is why Newton's second law is so valuable for classical physics: Any determination of the motion of a body starts out at Eq. (7.1.12).

However, there is even more to Eq. (7.1.12). It gives us an immediate answer to the question why there is weightlessness all over the place in space: If the forces on a body are canceled out by inertial behavior, there will be no residual gravitational force causing weight on a scale. It tells us that this happens whenever a body is free to move. Therefore, when jumping from a tower into a pool of water, we experience the same weightlessness during the free fall as an astronaut does during his uniform free fall around a planet, called circular orbit. In the latter case, the circular motion causes a constant inertial centrifugal force counterbalancing the gravitational force. But why is an astronaut even weightless when he is flying in a spacecraft outbound on a straight line from Earth, say to the Moon, as happened nearly so during the Apollo flights? This seems a much trickier question, but the answer is found again in Newton's second law: The gravitational force pulls the astronaut back and hence decelerates his outward motion, which in turn entails a breaking force exactly in the same manner as when you slam on the breaks in your car. This deceleration force, which now points tangentially along the trajectory, counterbalances the gravitational force in the same way as the perpendicular centrifugal force in a circular orbit.

Equation of Motion

Assuming an ideal two-body problem with an inertial point mass M centered at O and by applying Newton's second law, we finally get for the motion of a body m under the gravitational force of the central mass M, as described in Eq. (7.1.7), Newton's gravitational equation of motion:

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{r^3}\boldsymbol{r} = -\frac{\mu}{r^2}\hat{\boldsymbol{r}} \tag{7.1.13}$$

with

$$\mu = GM$$

Note *Observe that the body's mass m no longer appears in this equation! The trajectory of a body is thus independent of its mass.*

Remark *In order that the masses cancel out in* Eq. (7.1.13), *we have to assume that gravitational mass and inertial mass are identical. Newton's theory is not able to explain why the gravitational and inertial mass of a body should be identical. They could just as well be different. Only the theory of general relativity provides us with a seamless explanation: acceleration forces (inertial forces) and gravitational forces are two sides of the same coin, the curvature of space. So, a body must react on acceleration and gravitation in exactly the same way: inertial force = weight force. Let us illustrate this with an example due to Einstein: If you would be standing in an elevator at an unknown place in outer space, you could not tell whether your weight is due to external gravitation or due to an acceleration of the elevator.*

7.1.5 General Two-Body Problem

The assumption that the central body M is fixed at O and the body m moves within its potential—which implies that the body m is negligibly small with respect to the central body M, $m \ll$ M—is a constraint that can easily be eliminated. Let us have a look at two bodies with unrestricted mass m_1 and m_2, moving around each other under the influence of their mutual gravitational potential. Now that we have two bodies on an equal footing, there is no exceptional point for the origin O of our reference system. We can place it wherever we want. Let $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ be the position vectors from an arbitrary O to m_1 and m_2, and $\boldsymbol{r} := \boldsymbol{r}_2 - \boldsymbol{r}_1$ the connecting vector

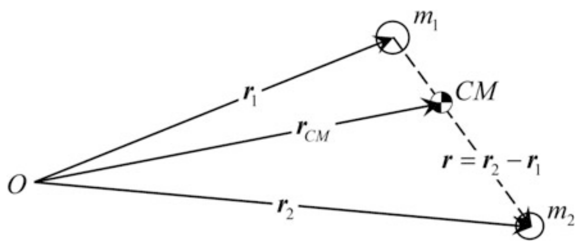

Fig. 7.5 Relevant vectors in the general two-body system

(see Fig. 7.5). According to Eq. (7.1.13), the vectorial equation of motion for each of the two bodies then is

$$\begin{aligned} m_2\ddot{\boldsymbol{r}}_2 &= -\frac{Gm_1m_2}{r^3}(\boldsymbol{r}_2 - \boldsymbol{r}_1) = -\frac{Gm_1m_2}{r^3}\boldsymbol{r} \\ m_1\ddot{\boldsymbol{r}}_1 &= -\frac{Gm_1m_2}{r^3}(\boldsymbol{r}_1 - \boldsymbol{r}_2) = +\frac{Gm_1m_2}{r^3}\boldsymbol{r} \end{aligned} \tag{7.1.14}$$

It is possible to trace bak these equation to that of the ideal two-body problem. To do so, one cancels m_1 from the first equation and m_2 from the other and then subtracts both equations from each other. This yields

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{r^3}\boldsymbol{r} \tag{7.1.15}$$

with

$$\mu := G(m_1 + m_2), \quad \boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1 \tag{7.1.16}$$

This is Newton's gravitational EoM for the connecting vector $\boldsymbol{r}$ between m_1 and m_2 describing the joint synchronous motion of two masses about each other.

Motion of the Center of Mass

The vector $\boldsymbol{r}_{CM}$ to the center of mass (CM, a.k.a. *barycenter*) per definition is the mass-weighted average of the position vectors to both bodies:

$$\boldsymbol{r}_{CM} := \frac{m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2}{m_1 + m_2} \tag{7.1.17}$$

Because of Eq. (7.1.14) $m_1\ddot{\boldsymbol{r}}_1 + m_2\ddot{\boldsymbol{r}}_2 = 0$. This implies $\ddot{\boldsymbol{r}}_{CM} = 0$ and hence

$$\boldsymbol{r}_{CM} = \boldsymbol{v}_0 t + \boldsymbol{r}_0 \tag{7.1.18}$$

with the initial conditions $\boldsymbol{v}_0, \boldsymbol{r}_0$. So, with no external forces acting, the CM moves along a straight line in space. This is Newton's third law applied to the CM. As this happens without any acceleration, the CM is an inertial system. According to Eq. (7.1.15) both bodies move synchronously around their common CM.

Motion in the CM System

Physically, one rather would like to describe the motion of each mass in an inertial reference system. Having found with the CM a natural point in space that exhibits inertial properties it lends itself to place the origin O into it: $\boldsymbol{r}_{CM} = 0$. Then $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the relative vectors with regard to the CM, and Eq. (7.1.17) results in

$$\boldsymbol{r}_{CM} = \frac{m_1 \boldsymbol{r}_1 + m_2 \boldsymbol{r}_2}{m_1 + m_2} = 0$$

From this, it follows that

$$m_1 \boldsymbol{r}_1 = -m_2 \boldsymbol{r}_2 \quad \text{and} \quad m_1 r_1 = m_2 r_2$$

In addition, we have for the absolute value of the connecting vector $r = r_1 + r_2$. With this and from Eqs. (7.1.14) we get after some simple modifications

$$\ddot{\boldsymbol{r}}_1 = -\frac{\mu_1}{r_1^3}\boldsymbol{r}_1, \quad \ddot{\boldsymbol{r}}_2 = -\frac{\mu_2}{r_2^3}\boldsymbol{r}_2$$

where

$$\mu_1 = \frac{Gm_2}{(1 + m_1/m_2)^2}, \quad \mu_2 = \frac{Gm_1}{(1 + m_2/m_1)^2}$$

These are the equations of motion of each of the two masses in the CM reference system.

For all following relevant considerations and hence for the remainder of this book let m be the mass under consideration and M be the other mass. Then the equation of motion for m with position vector $\boldsymbol{r}$ in the CM system reads

$$\boxed{\ddot{\boldsymbol{r}} = -\frac{\mu}{r^2}\hat{\boldsymbol{r}} = -\frac{\mu}{r^3}\boldsymbol{r}} \qquad \textbf{Newton's gravitational equation of motion} \tag{7.1.19}$$

with

$$\boxed{\mu = \frac{GM}{(1 + m/M)^2}}$$

Small Mass Approximation

To account for a small mass m moving about a large mass M (a.k.a. central body, e.g., small moon orbiting a planet), we linearly approximate Eq. (7.1.19) in m/M, which yields

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{r^2}\hat{\boldsymbol{r}} \tag{7.1.20}$$

with

$$\mu := G(M - 2m) \quad @ \; m \ll M$$

This is the Newtonian equation of motion relevant for all planets in the solar system. The factor μ differs from the one in Eq. (7.1.15) by $3mG$, which, in the case of the Moon circling the Earth, amounts to a non-negligible 3.7%. The μ values for all planets are given in Appendix A.

Of course for $m \to 0$ (e.g., spacecraft orbiting a planet) Eq. (7.1.20) crosses over into our primordial Eq. (7.1.13).

7.2 General Principles of Motion

Having derived the equation of motion, the next step would be to solve it in order to determine the precise motion of a body in a gravitational field. However, before we do that we will first study the general characteristics of a body's motion.

7.2.1 Vector Derivatives

We assume the most general situation, namely that a body moves in an arbitrary way (including rotation about a reference point) in space and attach an arbitrary co-moving reference system to this body, which we denote by $\left(\boldsymbol{u}_x, \boldsymbol{u}_y, \boldsymbol{u}_z\right)$. Examples are the $(\boldsymbol{u}_t, \boldsymbol{u}_n, \boldsymbol{u}_h)$-system in Fig. 6.5, the $(\boldsymbol{u}_r, \boldsymbol{u}_\theta, \boldsymbol{u}_h)$-system in Fig. 7.6,

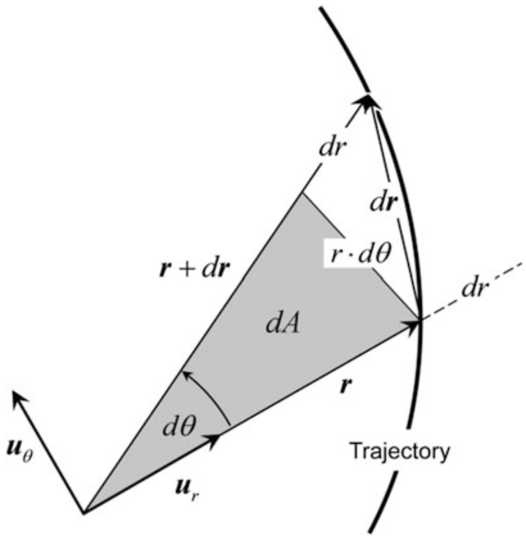

Fig. 7.6 Decomposition of the differential position vector d$\boldsymbol{r}$ in the co-moving reference system

or the $(\boldsymbol{u}_x, \boldsymbol{u}_y, \boldsymbol{u}_z)$-system in Fig. 8.20. Any vector $\boldsymbol{a}$ at a given instance t can therefore be expressed in terms of this reference frame

$$\boldsymbol{a}(t) = a_x \boldsymbol{u}_x + a_y \boldsymbol{u}_y + a_z \boldsymbol{u}_z$$

If we denote the unit vector of $\boldsymbol{a}$ by $\boldsymbol{u}_a(t)$, $\boldsymbol{a}(t)$ can also be written as

$$\boldsymbol{a}(t) = a(t) \cdot \boldsymbol{u}_a(t)$$

To derive the components of any differential vector, first have a look at Fig. 7.6 where the differential position vector

$$d\boldsymbol{r} = dr \cdot \boldsymbol{u}_r + d\boldsymbol{\theta} \times \boldsymbol{r} = dr \cdot \boldsymbol{u}_r + r \cdot d\theta \cdot \boldsymbol{u}_\theta$$

is pictured. From this we recognize that $dr \cdot \boldsymbol{u}_r$ is the radial, while $d\boldsymbol{\theta} \times \boldsymbol{r}$ is its lateral component of $d\boldsymbol{r}$. If we denote the angular velocity of $\boldsymbol{a}(t)$ as $\boldsymbol{\omega}(t) := d\boldsymbol{\theta}/dt$, its lateral unit motion therefore is

$$\dot{\boldsymbol{u}}_a = \boldsymbol{\omega} \times \boldsymbol{u}_a$$

We now aim to express the derivatives of this arbitrary $\boldsymbol{a}$ in the co-moving reference frame at the given instant t. By differentiating $\boldsymbol{a} = a \cdot \boldsymbol{u}_a$ twice, we obtain

$$\dot{\boldsymbol{a}} = \dot{a} \cdot \boldsymbol{u}_a + a\dot{\boldsymbol{u}}_a = \dot{\boldsymbol{a}}_r + \boldsymbol{\omega} \times \boldsymbol{a} \tag{7.2.1}$$

$$\ddot{\boldsymbol{a}} = \ddot{\boldsymbol{a}}_r + \underbrace{2\boldsymbol{\omega} \times \dot{\boldsymbol{a}}_r}_{\text{Coriolis force}} + \underbrace{\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{a})}_{\text{centrifugal force}} + \dot{\boldsymbol{\omega}} \times \boldsymbol{a} \tag{7.2.2}$$

where we have defined

$$\dot{\boldsymbol{a}}_r := \dot{a} \cdot \boldsymbol{u}_a = \dot{a}_x \boldsymbol{u}_x + \dot{a}_y \boldsymbol{u}_y + \dot{a}_z \boldsymbol{u}_z$$
$$\ddot{\boldsymbol{a}}_r := \ddot{a} \cdot \boldsymbol{u}_a = \ddot{a}_x \boldsymbol{u}_x + \ddot{a}_y \boldsymbol{u}_y + \ddot{a}_z \boldsymbol{u}_z$$

The annotated terms in the acceleration vector correspond to the said well-known forces in physics.

7.2.2 *Motion in a Central Force Field*

Let us consider the real-world case that a body m moves in a gravitational field, or more generally in a central force field, with position vector $\boldsymbol{r}$ and velocity vector $\boldsymbol{v}$. Under these conditions, its motion is determined by the equation of motion, in particular Eq. (7.1.19) for motion in a gravitational field. Before solving this differential vector equation we first want to derive some general properties of the motion.

Conservation of Angular Momentum

We have seen in Sect. 7.1.3 that, if the body m and the central body is a closed system, i.e. if the body m is not subject to any external interaction, its angular momentum L is a conserved quantity that is given by physics as

$$\boxed{\boldsymbol{h} := \frac{\boldsymbol{L}}{m} = \boldsymbol{r} \times \boldsymbol{v} = const} \quad \textbf{mass-specific angular momentum} \text{ (invariant)} \tag{7.2.3}$$

where h is the *mass-specific angular momentum*. We will denote it *angular momentum* for short. Vector representations of $\boldsymbol{h}$ in various reference systems are found in Sect. 13.1.5.

To verify that for this motion the angular momentum is indeed conserved, we take the time-derivative of $\boldsymbol{L}$

$$\dot{\boldsymbol{L}} = m\boldsymbol{v} \times \boldsymbol{v} + \boldsymbol{r} \times m\ddot{\boldsymbol{r}} = \boldsymbol{r} \times \boldsymbol{F}$$

where the latter follows from Newton's second law, Eq. (7.1.12). We now see that the conservation of angular momentum, i.e. $\dot{\boldsymbol{L}} = 0$, hinges on the fact that the gravitational force is a central force $\boldsymbol{F} = -m \cdot g(r) \cdot \hat{\boldsymbol{r}}$ (see Sect. 7.1.2). Note that $\boldsymbol{r} \times \boldsymbol{F}$ is the torque that the force exerts on a moving body. A central force therefore is also characterized in that it affects a body without any torque.

The Orbital Plane

Let us assume that the body m has the initial velocity $\boldsymbol{v}_0$ at the initial position $\boldsymbol{r}_0$. $\boldsymbol{r}_0$ and $\boldsymbol{v}_0$ span a plane. Because of Eq. (7.2.3), the initial angular momentum $\boldsymbol{h}_0$ is vertical on $\boldsymbol{r}_0$ and $\boldsymbol{v}_0$, and also at later times $\boldsymbol{h} \cdot \boldsymbol{r} = \boldsymbol{h} \cdot \boldsymbol{v} = 0$ holds. So, r as well as v is always vertical to h. In other words, because $\boldsymbol{h} = const$, the body m always maintains its motion in the plane, that was spanned by the initial $\boldsymbol{r}_0, \boldsymbol{v}_0$.

> **Note** *Strictly speaking, the motion in a plane with* $\boldsymbol{r}, \boldsymbol{v} \perp \boldsymbol{h}$ *is valid only for* $h \neq 0$. *For* $h = 0$, *the motion is on a line* (see Sect. 7.5).

Therefore, the plane spanned by $\boldsymbol{r}, \boldsymbol{v}$ does not change with time. As will be shown later, the pair $(\boldsymbol{r}, \boldsymbol{v})$ at any point in the orbit also determines the shape of an orbit. Hence, $(\boldsymbol{r}, \boldsymbol{v})$ determines both the orientation and the shape, i.e., the full state, of an orbit, which is why it is rightly called state vector. We conclude that

The motion of a body m always takes place in a constant plane, the orbital plane, through the center of mass common with M, perpendicular to the angular momentum $\boldsymbol{h}$, and spanned by $\boldsymbol{r}$ and $\boldsymbol{v}$.

As conservation of the angular momentum is a very general property, independent of the details of gravitational force or its potential, it is even true for spaces with dimension other than three. We will come back to this peculiarity in Sect. 7.6.

General Radial and Lateral Motion in a Plane
To obtain the general properties of the planar motion we apply the results of Sect. 7.2.1 to For convenience we choose the co-moving cylindrical reference system $(\boldsymbol{u}_r, \boldsymbol{u}_\theta, \boldsymbol{u}_h)$ with the instantaneous radial unit vector $\boldsymbol{u}_r = \hat{\boldsymbol{r}}$, $\boldsymbol{u}_\theta$ perpendicular to it in the motion plane, and $\boldsymbol{u}_r = \hat{\boldsymbol{h}}$ as the basis vectors (see Fig. 7.6). In this system any other vector in the plane is described as $\boldsymbol{a} = (a, \theta, 0)$. We also define the angular velocity

$$\omega := \dot{\theta} \qquad \textbf{angular velocity} \tag{7.2.4}$$

It is now easy to show that for $\boldsymbol{a} \equiv \boldsymbol{r} = r \cdot \boldsymbol{u}_r$ we get with $\boldsymbol{\omega} = \omega \cdot \boldsymbol{u}_h$ from Eqs. (7.2.1) and (7.2.2)

$$\dot{\boldsymbol{r}} = \boldsymbol{v} = \dot{r} \cdot \boldsymbol{u}_r + r\omega \cdot \boldsymbol{u}_\theta \tag{7.2.5}$$

$$\ddot{\boldsymbol{r}} = \left(\ddot{r} - \omega^2 r\right) \cdot \boldsymbol{u}_r + (2\omega\dot{r} + \dot{\omega}r) \cdot \boldsymbol{u}_\theta \tag{7.2.6}$$

So, while $v_r = \dot{r}$ is the radial component $v_\theta = \omega r$ is the lateral component of the velocity (see Fig. 7.5 for the differential analysis).

Note *$\dot{r}$ is only the radial part of the velocity vector $\dot{\boldsymbol{r}}$, i.e. $\dot{r} = \hat{\boldsymbol{r}}\dot{\boldsymbol{r}}$, and not its value, $\dot{r} \neq |\dot{\boldsymbol{r}}|$. To avoid confusion, we will therefore from now on always write $\boldsymbol{v}$ rather than $\dot{\boldsymbol{r}}$* (cf. Note in Sect. 6.3).

With Eq. (7.2.5) we obtain for the angular momentum $\boldsymbol{h} = \boldsymbol{r} \times \dot{\boldsymbol{r}} = r \cdot r\omega \cdot \boldsymbol{u}_h$, and hence for its absolute value

$$\boxed{h = \omega\, r^2 = \dot{\theta} r^2 = const} \tag{7.2.7}$$

Note *It is the angular momentum h and its conservation in time that makes a body to orbit* (ω!) *steadily around a central mass and prevents the masses in our universe to instantly collapse* (cf. Sect. 7.5 for trajectories with $h = 0$).

This equation is of notable significance. It states that the further a body departs on its orbit from the origin the less its angular velocity becomes, and vice versa.
From Eq. (7.2.5) finally follows with Eq. (7.2.7)

$$v^2 = \dot{r}^2 + \omega^2 r^2 = \dot{r}^2 + \frac{h^2}{r^2} \tag{7.2.8}$$

Kepler's Second Law
The conservation of angular momentum has an immediate and important geometrical implication. The infinitesimal area dA that the position vector r sweeps by advancing through $d\theta$ is determined according to Fig. 7.6, Eqs. (7.2.7) and (7.2.4) by

$$dA = \frac{1}{2} r \left(r \cdot d\theta \right) = \frac{1}{2} r^2 \cdot d\theta = \frac{1}{2} r^2 \omega \cdot dt = \frac{h}{2} dt$$

That is, $h = 2\dot{A} = const$. So if we consider a constant time interval Δt, we find

$$\Delta A = \int \frac{h}{2} dt = \frac{h}{2} \int dt = \frac{1}{2} h \cdot \Delta t = const \qquad \textbf{Kepler's second law} \qquad (7.2.9)$$

> The angular momentum can be interpreted as a constant areal velocity of the body: The area that the position vector sweeps in equal time intervals is constant.

Kepler's second law is valid not only for bound orbits (ellipses, circles) in a gravitational field, as Kepler postulated, but for any motion of a body in any central force field. This is because the above derivation rest solely on the conservation of angular momentum, which implies a central force as we saw in Sect. 7.2.1.

Equation of Radial and Lateral Motion under a Gravitational Force

From Eq. (7.2.6) we get with Newton's gravitational equation of motion (7.1.19), $\ddot{\boldsymbol{r}} = -\mu/r^2 \cdot \boldsymbol{u}_r$ for the radial and lateral component separately

$$\ddot{r} - \omega^2 r = -\frac{\mu}{r^2}$$
$$2\omega\dot{r} + \dot{\omega} r = 0$$

By multiplying the second equation by r and by direct integration we recover $r^2\omega = h$, which is Eq. (7.2.7). Inserting this result into the first equation, one obtains

$$\ddot{r} = -\frac{\mu}{r^2} + \frac{h^2}{r^3} \qquad \textbf{Leibniz's equation} \qquad (7.2.10)$$

This is an equation of radial motion (cf. Eqs. (7.1.19) and (7.6.3)), which, derived by Leibniz, historically is of high relevance. Multiplication by $\dot{r}$ and direct integration delivers the equation of radial energy

$$\frac{1}{2}\dot{r}^2 = \frac{\mu}{r} - \frac{h^2}{2r^2} + \varepsilon$$

As we will see later in Eq. (7.3.19), the energy integration constant can be written as $\varepsilon = -\mu/2a$ where a is geometrically the semi-major axis of the orbit (i.e., $a > 0$ for ellipses and $a < 0$ for hyperbolas). This explicitly proves that the gravitational field as a conservative force conserves both the angular momentum and the total energy of the moving body. In summary we have

$$\dot{r}^2 = \frac{2\mu}{r} - \frac{h^2}{r^2} - \frac{\mu}{a} \quad \textbf{gravitational equation of radial motion} \tag{7.2.11}$$

$$\dot{\theta} = \frac{h}{r^2} \quad \textbf{equation of lateral motion} \tag{7.2.12}$$

which are two scalar differential equations that describe the motion of the body in polar coordinates in the orbital plane. As we will see later from Eq. (7.3.7), $h^2 = \mu a(1 - e^2)$ and the second differential equation can be decoupled from the first via Eq. (7.3.5). In principle one could solve both equations of motion to derive a body's motion in the plane. However, as it turns out, rather than $r(t),\ \theta(t)$ they deliver the inverse solutions $t(r),\ t(\theta)$, which are useless for practical purposes. We therefore have to retreat to Kepler's solution by introducing an easy to treat auxiliary variable $E(t)$ (Kepler transformation, see Sect. 7.4.3), from which $r(E),\ \theta(E)$ follows.

What is the condition that Eqs. (7.2.11) and (7.2.12) are valid? The steadiness of the orbital plane in an inertial reference frame (and thus that the orbit can be described by a polar coordinate system in this plane) and $h = const$ is equivalent to the conservation of angular momentum. This in turn hinges on the fact, as shown in Sect. 7.2.1, that the potential U_G causes a central force field $\boldsymbol{F}_G = -m \cdot g(r) \cdot \hat{\boldsymbol{r}}$. However, this central force field condition generally does not hold, for instance for Earth's non-spherically symmetric gravitational potential (see Sect. 12.3.1, General Considerations) or for external perturbations on Earth orbits such as lunisolar perturbations (see Sect. 12.7.1). All non-central forces are able to tilt an orbital plane, and h may even vary over one orbital revolution, staying constant only on average over one orbital period. For this reason only Newton's second law (7.1.12) serves as a general differential equation to describe the motion of a body under the influence of any kind of forces.

7.2.3 Vis-Viva Equation

For the motion of a body in a gravitational field, the general law of conservation of energy Eq. (7.1.8) reduces to

$$\varepsilon_{kin} + \varepsilon_{pot} =: \varepsilon = const \tag{7.2.13}$$

where $\varepsilon_{pot} := E_{pot}/m = U(r)$ is the *specific potential energy*, $\varepsilon_{kin} = E_{kin}/m = \frac{1}{2}v^2$ is the specific *kinetic energy* and $\varepsilon := E_{tot}/m$ is the so-called *specific orbital energy* (a. k.a. *specific mechanical energy*). This results in the following important equations:

$$\frac{1}{2}v^2 = \frac{\mu}{r} + \varepsilon \tag{7.2.14}$$

$$\boxed{v^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right)} \qquad \textbf{vis-viva equation} \tag{7.2.15}$$

with $a := -\mu/2\varepsilon$. As we will see later in Eq. (7.3.19), a is geometrically the semi-major axis of the orbit (i.e., $a > 0$ for ellipses and $a < 0$ for hyperbolas). Observe that the gravitational equation of radial motion (7.2.11) can be also obtained by inserting Eq. (7.2.8) into Eq. (7.2.15).

Bear in mind that the vis-viva equation is crucial for any orbital motion and will be used frequently throughout this book, because at any orbit position it directly relates the orbit velocity and orbit radius. This is why it bears this very historic name.

Historic Remark
The naming *vis-viva equation*, earlier also called *equatio elegantissima*, originates with the German mathematician Leibniz (1646–1716). He coined the term "vis viva" (Latin, meaning "living force") for mv^2, which he considered as the real measure of force, to separate it from Newton's gravitational force, which people called just "vis". Later when "vis viva" was recognized as an energy, "vis" was also called by the Leibnizians "vis morte" (a.k.a. "vis mortua", meaning "dead force") to distinguish the concept "energy" from "force". You can imagine that this caused a serious fight between Newtonians and Leibnizians on whether their forces were dead or alive. Actually, this all only shows that at that time people did not really understand the difference between force, linear momentum, and energy and the relation between each other: $\boldsymbol{F} = d\boldsymbol{p}/dt = -dE/d\boldsymbol{r}$.

We now show explicitly that the energy of a body moving in a conservative field and hence in any gravitational field is indeed conserved and therefore the vis-viva equation with $a = -\mu/2\varepsilon = const$ holds. We first generalize Eq. (7.2.14) for an arbitrary potential U

$$\varepsilon = \frac{1}{2}v^2 + U$$

From this we obtain with $dv^2/dt = d\dot{\boldsymbol{r}}^2/dt = 2\dot{\boldsymbol{r}}\ddot{\boldsymbol{r}}$

$$\frac{d\varepsilon}{dt} = \dot{\boldsymbol{r}}\ddot{\boldsymbol{r}} + \frac{dU}{dt}$$

Since for the potential of a conservative field (see Sect. 7.1.2) $dU/dt = dU/d\boldsymbol{r} \cdot \dot{\boldsymbol{r}} = -\boldsymbol{f}(\boldsymbol{r})\dot{\boldsymbol{r}}$ holds, we find with Newton's second law (7.1.12), $\ddot{\boldsymbol{r}} = \boldsymbol{f}(\boldsymbol{r})$,

$$\frac{d\varepsilon}{dt} = \dot{\boldsymbol{r}}\ddot{\boldsymbol{r}} - \dot{\boldsymbol{r}}\ddot{\boldsymbol{r}} = 0$$

and therefore $\varepsilon = const$.

7.2.4 Effective Radial Motion

Rotational Potential

On pure mathematical grounds the expression $h^2/2r^2$ in Eq. (7.2.11) like the preceding term μ/r could be considered a potential: a radial angular momentum potential, the so-called *rotational potential,* a.k.a. *centrifugal potential*

$$U_\omega(r) := \frac{h^2}{2r^2} = \frac{1}{2}\omega^2 r^2 \tag{7.2.16}$$

The latter follows from Eq. (7.2.7). This also physically makes sense, because according to Eqs. (7.1.5), (7.2.7), and (7.1.6) the corresponding centrifugal force would be

$$\boldsymbol{F}_\omega = m\ddot{\boldsymbol{r}}_\omega = -m\frac{d}{d\boldsymbol{r}}\left(\frac{h^2}{2r^2}\right) = m\frac{h^2}{r^3}\frac{dr}{d\boldsymbol{r}} = m\frac{h^2}{r^3}\frac{\boldsymbol{r}}{r} = m\omega^2\boldsymbol{r} \tag{7.2.17}$$

This is the well-known formula for the centrifugal force in physics. It pushes the orbiting body toward the outside (positive sign). For example, for a circular orbit $r = const$, or $\dot{r} = 0$, it follows from Eq. (7.2.11) that $\mu/r = \mu/a = -h^2/2r^2$, meaning that the centrifugal force compensates the gravitational force at any point of the orbit. Generalizing Eq. (7.1.4), the corresponding rotational energy would be

$$E_\omega = mU_\omega(r) = \frac{1}{2}m\omega^2 r^2 \tag{7.2.18}$$

which is also a quite familiar expression in physics. From this equation the physical meaning of the rotational potential becomes clear: The constancy of angular momentum $h = \omega\, r^2 = const$ implies that the lateral velocity $v_\theta = \omega\, r = h/r$ increases with decreasing orbit radius. lateral kinetic energy $E_\omega = mU_\omega = \frac{1}{2}mv_\theta^2 = \frac{1}{2}m\omega^2 r^2$ increases correspondingly. The rotational potential $U_\omega(r) = \frac{1}{2}\omega^2 r^2 = h^2/2r^2$ is just a virtual potential of this effect. In summary, the closer a body comes to the central mass the more energy is required for its revolving motion. At its closest point of approach all available kinetic energy is rotational energy.

Effective Potential

It is convenient to lump together the gravitational and the rotational potential into a pseudopotential the so-called

$$U_{eff}(r) := U_G + U_\omega = -\frac{\mu}{r} + \frac{h^2}{2r^2} \qquad \textbf{effective potential} \tag{7.2.19}$$

With this it follows from Eq. (7.2.11) that

$$\frac{1}{2}\dot{r}^2 + U_{eff}(r) = -\frac{\mu}{2a} \tag{7.2.20}$$

The term $\dot{r}^2/2$ describes the specific radial kinetic energy. In Fig. 7.7 the effective potential for an orbiting body with a given specific angular momentum h is plotted. We recognize that if the body acquires a negative total specific energy, $\varepsilon = -\mu/2a < 0$, its motion is limited to a radial interval bounded by an inner point A and an outer point B. For $\varepsilon > 0$ the orbit is unbounded at the outer end. We determine the outer and inner bound radii from Eq. (7.2.20) and Eq. (7.3.20) with $\dot{r} = 0$ as

$$r_{A,B} = a\left(1 \pm \sqrt{1 - \frac{h^2}{\mu a}}\right) = a(1 \pm e)$$

which agrees with the result of Eq. (7.4.6) in Sect. 7.4.2. Note that the lower bound exists due to the rotational potential, i.e., due to fact that the body has to maintain its

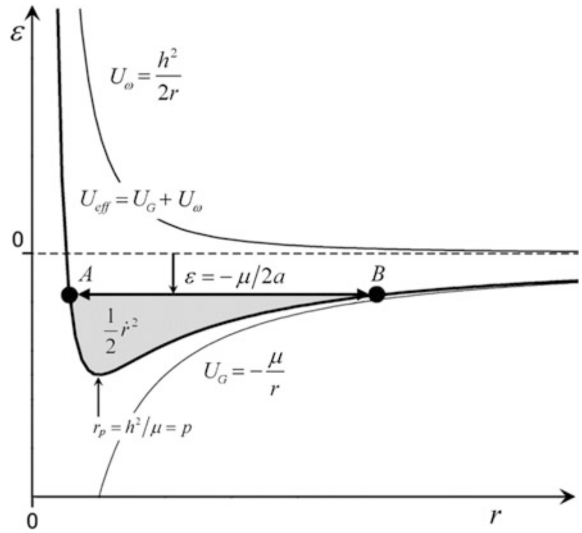

Fig. 7.7 The shape of the effective potential U_{eff} limits the motion of an orbiting body with a given h and specific total energy $\varepsilon < 0$ to the radial end points A and B. The area marked gray is the contribution of the specific radial kinetic energy

angular momentum. At the end points the radial velocity vanishes, while at the radial position $r_p = h^2 / \mu = p$, where the effective potential a minimum value, it attains the maximum value $\dot{r}_{\max} = e/h = e\sqrt{\mu/p}$, as derived from Eq. (7.2.11). As we will see later in Eq. (7.3.6), this particular point is called *semi-latus rectum*. The detailed behavior of the different types of orbits will be discussed in Sect. 7.4.

We close this section by summarizing the motion principles brought about by the laws of conservation.

For an object in a gravitational field:

- The law of conservation of angular momentum restricts its motion to a plane and imposes a simple relation between its angular velocity and radial distance, $\omega(r) = h/r^2$, which gives rise to Kepler's second law.
- The law of conservation of energy requires a one-to-one relation between its velocity and radial distance as specified by the vis-viva Eq. (7.2.15).
- The law of conservation of angular momentum jointly with the law of conservation of energy impose limits on its range of motion as given by $r_{A,B} = a\left(1 \pm \sqrt{1 - h^2/\mu a}\right) = a(1 \pm e)$ for $e < 1$.

7.3 Motion in a Gravitational Field

7.3.1 Orbit Equation

So far by applying the equation of motion and general conservation laws we were able to determine general features of the motion without knowing the explicit solution. To obtain further details of the orbit, we have to solve the equation of motion (7.1.19). However, it is in general not possible to find the desired $\boldsymbol{r}(t)$ because when explicitly written for each of the coordinate components we have three equations of motion that are coupled via the $1/r^3 = 1/(x^2 + y^2 + z^2)^{3/2}$ term. Yet, it is possible to find a general expression for the orbital shape the so-called *orbit equation*. To find that we apply an elegant method that Pierre-Simon Laplace (1749–1827) is credited for. However, it dates back to an article of Jakob Hermann (1668–1733, a pupil of Bernoulli and friend of Leibniz) in 1710 in the journal *Giornale de Letterati d'Italia*, Vol. 2, pp. 447–467, which in fact provides the first solution of Newton's gravitational equation of motion at all (see Volk (1976)).

In preparation to this method, we note that according to Eq. (7.2.5)

$$\boldsymbol{r} \cdot \boldsymbol{v} = \boldsymbol{r} \cdot (\dot{r} \cdot \boldsymbol{u}_r + r\omega \cdot \boldsymbol{u}_\theta) = r\dot{r}$$

which we will make use in Hermann's smart approach: For $h \neq 0$ (see Note following Eq. (7.3.7)) take the cross product of $\boldsymbol{h}$ with Eq. (7.1.19), which yields

$$\begin{aligned}\boldsymbol{h}\times\ddot{\boldsymbol{r}} &= -\frac{\mu}{r^3}(\boldsymbol{h}\times\boldsymbol{r}) = -\frac{\mu}{r^3}[(\boldsymbol{r}\times\boldsymbol{v})\times\boldsymbol{r}] = -\frac{\mu}{r^3}[\boldsymbol{v}(\boldsymbol{r}\cdot\boldsymbol{r}) - \boldsymbol{r}(\boldsymbol{r}\cdot\boldsymbol{v})] \\ &= -\frac{\mu}{r^3}\left(r^2\cdot\boldsymbol{v} - r\dot{r}\cdot\boldsymbol{r}\right) = -\mu\left(\frac{1}{r}\boldsymbol{v} - \frac{\dot{r}}{r^2}\boldsymbol{r}\right) = -\mu\frac{d}{dt}\left(\frac{\boldsymbol{r}}{r}\right) = -\mu\frac{d\hat{\boldsymbol{r}}}{dt}\end{aligned} \tag{7.3.1}$$

with $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ the unit vector in $\boldsymbol{r}$ direction. This approach is smart because owing to $\boldsymbol{h} = const$, this equation can be integrated directly to give

$$\boldsymbol{h}\times\boldsymbol{v} = -\mu\hat{\boldsymbol{r}} - \boldsymbol{A} = -\mu(\hat{\boldsymbol{r}} + \boldsymbol{e}) \tag{7.3.2}$$

with $\boldsymbol{A} = \mu\boldsymbol{e}$ the integration constant, determined by the initial conditions. Apart from $\boldsymbol{h}$, ε, and $\boldsymbol{e}$ (or $\boldsymbol{A}$, respectively) are also invariants of the system. $\boldsymbol{e}$ is called the *eccentricity vector* and $\boldsymbol{A}$ is called the *Laplace–Runge–Lenz vector* (a.k.a. *Runge-Lenz vector* or *Laplace vector*). From Eq. (7.3.2), we get

$$\boldsymbol{e} = \frac{1}{\mu}\boldsymbol{v}\times\boldsymbol{h} - \hat{\boldsymbol{r}} = \left(\frac{1}{r} - \frac{1}{a}\right)\boldsymbol{r} - \frac{1}{\mu}(\boldsymbol{r}\boldsymbol{v})\boldsymbol{v} \quad \textbf{eccentricity vector} \text{ (invariant)} \tag{7.3.3}$$

where the latter follows from Eqs. (7.2.3) and (7.2.15). Because $\boldsymbol{v}\times\boldsymbol{h}$ and $\hat{\boldsymbol{r}}$ lie in the motion plane, so must $\boldsymbol{e}$ (and also $\boldsymbol{A}$), that is $\boldsymbol{h}\cdot\boldsymbol{e} = 0$. Vector representations of $\boldsymbol{e}$ in various reference systems are given in Sect. 13.1.5.

To directly derive the equation for the trajectory of the orbit, we multiply Eq. (7.3.2) with r and with $\boldsymbol{r}(\boldsymbol{h}\times\boldsymbol{v}) = -\boldsymbol{h}(\boldsymbol{r}\times\boldsymbol{v}) = -\boldsymbol{h}^2$ get the orbit equation in vectorial form (i.e., independent of a reference system):

$$r + \boldsymbol{e}\cdot\boldsymbol{r} = \frac{h^2}{\mu} =: p \tag{7.3.4}$$

Introducing polar coordinates (r, θ) with $\cos\theta := \hat{\boldsymbol{e}}\hat{\boldsymbol{r}}$, which are most suited and common for orbital trajectories, we obtain

$$\boxed{r = \frac{p}{1+e\cdot\cos\theta}} \quad @\ h \neq 0 \quad \textbf{orbit equation} \tag{7.3.5}$$

with

$$p := \frac{h^2}{\mu} =: a\left(1 - e^2\right) > 0 \qquad \textbf{semi-latus rectum} \tag{7.3.6}$$

$$h^2 = \mu a\left(1 - e^2\right) \tag{7.3.7}$$

Note *The orbit equation is valid only as long as* $h \neq 0$. *If the body m takes on* $h = 0$, *then according to* Eq. (7.2.7) $\omega = 0$ *and therefore* $\theta = const$. *Thus, the body falls toward the central body M or moves away from it radially on a straight line* (see Sect. 7.5 *for details and compare Note following* Eq. (7.2.7)).

Historic Remark
The fight between Newtonians and Leibnizians as indicated by the Historic Remark in Sect. 7.2.3 was even more profound. Newtons derivation of trajectories under gravitational action was based on his three laws as outlined in his *Principia*. Having derived a vector of the gravitational forces he derived a vectorial equation of motion Eq. (7.1.19), which fully defines the two-dimensional motion in a plane. Leibniz, on the other hand, based his physics solely on his vis viva, mv^2, from which he derived the gravitational equation of radial motion Eq. (7.2.11) and from this by differentiation Leibniz's Eq. (7.2.10). This, however, is just one equation, from which the full description of two-dimensional motion cannot be derived. This weakness of his physics made him resort to his *best of all possible worlds* (most efficient orbit) *principle*, which today is known as the *principle of least action*. Only later in 1833 did William Hamilton show that from the principle of least action the laws of motion indeed follow. Newtonians, not aware of this equivalence, blamed Leibniz for this metaphysics and therefore believed that Newton's physics was superior. If Leibniz would only have known about conservation of angular momentum, he would have derived the equation of lateral motion Eq. (7.2.12). It is now quite easy to show (see Problem 7.4) that the solution of Leibniz's equation with the equation of lateral motion also delivers the orbit Eq. (7.3.5). Therefore, Newton's and Leibniz's physics stand on equal footing.

Analysis of Trajectories
A geometrical analysis of the polar Eq. (7.3.5) shows that it describes four types of trajectories (see Fig. 7.8): circle ($e = 0$), ellipse ($0 < e < 1,\ 0 < a < \infty$), parabola ($e = 1,\ a = \infty$), and hyperbola ($e > 1,\ -\infty < a < 0$). These trajectories turn out to be conic sections, a.k.a. Keplerian orbits, which will be discussed in detail in Sect. 7.4. Common to all Keplerian orbits is the symmetry line, the so-called *line of apsides*. The geometric interpretation of the elements a, e, p is (see Fig. 7.8b):

- a as defined in Eq. (7.3.6) is the semi-major axis of a Keplerian orbit and as such is a direct measure of the orbital energy ε (see Eq. (7.3.19)).
- The eccentricity e determines the type of orbit and its shape.
- The semi-latus rectum p (a.k.a. *orbital parameter*) is a direct measure of the orbital angular momentum via $p = h^2/\mu$ and as such is an important parameter

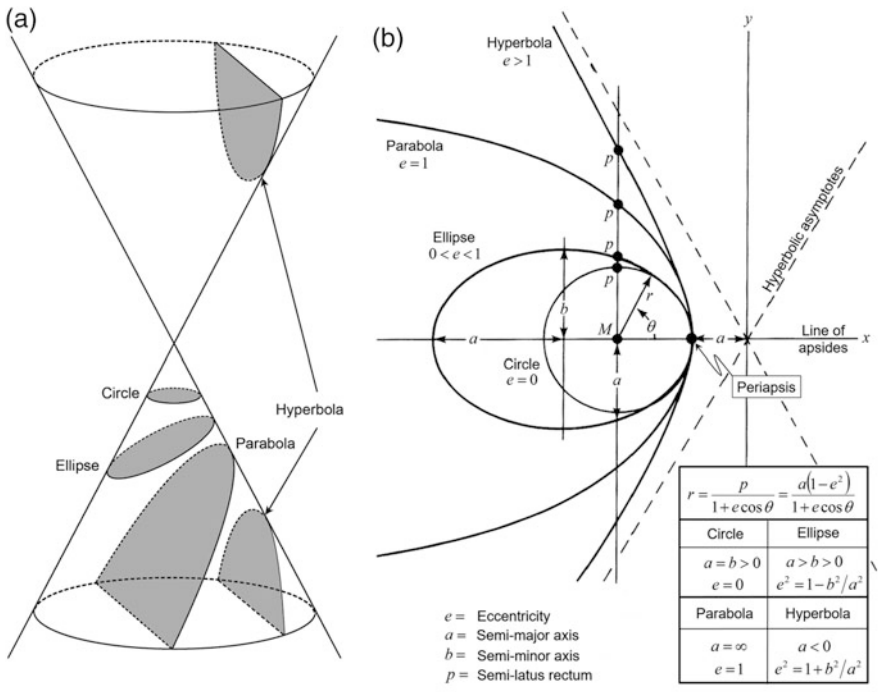

Fig. 7.8 (**a**) Keplerian orbits are conic sections. (**b**) Geometrical presentation of the parameters a, e, b, p for the four conic sections

to characterize any Keplerian orbit. Because $r(\theta = 90^\circ) = p$, it geometrically is the distance from the focal point to the intersection of the line normal to the line of apsides with the trajectory. At this intersection point the body also achieves its maximum radial velocity (see Fig. 7.7). For any Keplerian orbit, p is also the orbit's curvature radius at periapsis (Exercise).

Having achieved geometrical interpretations of a and e, we see that the vis-viva Eq. (7.2.15)

$$v^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right)$$

and Eq. (7.2.7) with (7.3.7)

$$h = \omega\, r^2 = \sqrt{\mu a(1 - e^2)} = \sqrt{\mu p}$$

are important, simple, and useful equations that link physical state properties, namely h, r, v, ω, to geometric properties of the orbit, a, e, p.

The Periapsis

Common to all Keplerian orbits is that they have a point of closest approach to the focal point. This point of closest approach is called the *periapsis*. The time of the passage of the body through the periapsis is defined as t_p. Sometimes t_p is called the (astronomical) **epoch**. This epoch is to be distinguished from the notion "standard epoch J2000" (= January 1, 2000, 11:58:55.816 UTC; see Sect. 13.2). So, the general meaning of "epoch" is "a reference point in time".

Significance of the Eccentricity Vector

We read from Eq. (7.3.5) that r becomes minimal when $\cos\theta = \hat{\boldsymbol{e}}\hat{\boldsymbol{r}} = 1$, i.e., if $\boldsymbol{r}$ points along $\boldsymbol{e}$. Thus, the eccentricity vector $\boldsymbol{e}$ points from the focal point (central body) to the periapsis (see Fig. 7.10). Its absolute value is the orbit eccentricity e, which describes the elongation of the orbit. Because both the focal point and the periapsis lie on the line of apsides, the eccentricity vector defines the line of apsides.

Because $\hat{\boldsymbol{e}}\hat{\boldsymbol{r}} = \cos\theta$, θ measures the angle between the radial vector to the periapsis and to the current position of the body and therefore $\theta(t_p) = 0$. Hence, we can determine the exact position on the orbit with θ. This important parameter $\boldsymbol{\theta}$ is called the **true anomaly** (a.k.a. *orbit angle*). So, the orbit equation provides the orbital radius r as a function of the true anomaly θ (Figs. 7.8 and 7.10).

Remark *The weird term "anomaly" for θ and later in this book for the angles M, E, F, and G dates back to the Ptolemaic astronomical system. In ancient times, any angle that could not be traced back to a true circular motion, appeared to be wrong or "anomalous".*

In summary, we figured out that

> The eccentricity vector e is the base vector relative to which the position on the orbit, the true anomaly θ, is measured: $\hat{\boldsymbol{e}}\hat{\boldsymbol{r}} = \cos\theta$. It points from the orbit's focal point (central body) to the smallest approach distance, the periapsis (see Figs. 7.8b and 7.10). Its absolute value describes the elongation of the orbit.

7.3.2 Position on the Orbit

How did we succeed in solving the apparently difficult vectorial equation of motion so swiftly? We made use of our previous knowledge that the momentum is a constant of motion and of its relation to $\boldsymbol{r}$ and $\boldsymbol{v}$, namely $\boldsymbol{h} = \boldsymbol{r} \times \boldsymbol{v}$. Therefore, we just had to integrate only once to find, besides $\boldsymbol{h}$, the second integral of motion $\boldsymbol{e}$. However, note that with Eq. (7.3.5), we actually did not achieve our goal to find $\boldsymbol{r}(t)$, we rather only found $r(\theta)$. So we still have to derive $\theta(t)$, a problem we will address now.

Because the true anomaly $\theta(t)$ evolves with time, there must exist a differential equation describing this evolution. We have found this already, it is the equation of lateral motion (7.2.12). In conjunction with the orbit Eq. (7.3.5) and $h = \sqrt{p\mu}$ it reads

$$\dot{\theta} = \frac{h}{r^2} = \sqrt{\frac{\mu}{p^3}}(1 + e \cdot \cos\theta)^2 \quad \textbf{equation of lateral motion} \tag{7.3.8}$$

Separating the variables and integrating on the right-hand side with respect to time results in

$$\int_0^\theta \frac{d\theta'}{(1 + e \cdot \cos\theta')^2} = \sqrt{\frac{\mu}{p^3}}(t - t_p) \tag{7.3.9}$$

Keplerian Problem

Equation (7.3.9) is to be read as follows: Given t_p, e, $p = a(1 - e^2)$. To find the true anomaly at any time t, the integral on the left hand side needs to be solved. Except for a circular and a parabolic orbit, this is too complicated to do by regular means. Even if we were able to solve, the solutions are very complicated (see Eqs. (7.4.17) and (7.4.33)). The crucial problem, however, is that these solutions display the time dependency not explicitly, that is, $\theta = \theta(t)$, but only implicitly, that is, $t = t(\theta)$. One could, though, for any given point in time solve $t = t(\theta)$ for θ numerically. But this is quite an effort. In the face of this problem, Kepler at the beginning of the seventeenth century proposed a method, which shifts the problem analytically to a simpler one, which can be solved with less effort, though still numerically. We will follow Kepler's elegant method in the next chapter. In fact, as we will see later, his method provides an easy way to solve Eq. (7.3.9) analytically also for elliptic and hyperbolic orbits.

The so-called "Keplerian problem" is historically the problem of finding the orbit position at a given time, if it was known at an earlier time. The background is the astronomical problem, even today, to find a celestial body back if it was observed at earlier times and to determine its orbital elements. Since the orbital path is constant in time, Kepler's problem lies in the difficulty of determining $\theta = \theta(t)$.

In Sect. 7.4 we will examine the specific properties of each type of orbit and the solutions to the Keplerian problem, separately.

Mean Lateral Motion

The lateral motion $\omega = \dot{\theta}(t)$ may be quite uneven. We can facilitate further calculations a lot by introducing an artificial constant lateral motion, the so-called *mean motion n* (a.k.a. *orbital frequency*). We define it as the average lateral motion $\langle\omega\rangle_T$, which of course is one orbit 2π over the orbital period T. Therefore,

$$n := \langle\omega\rangle_T = \frac{2\pi}{T} = \begin{cases} \sqrt{\mu/a^3} & \text{@ elliptic orbits} \\ \sqrt{-\mu/a^3} & \text{@ hyperbolic orbits} \end{cases} \quad \textbf{mean motion} \tag{7.3.10}$$

where the latter follows from Eqs. (7.4.12) and (7.4.31). Observe that while $\omega = d\theta/dt$ is the (instantaneous) angular velocity the mean motion n is formally a frequency. Only for circular orbits $n \equiv \omega$. Though the definition of n as based on a closed periodic orbit, it can be extended also to hyperbolic orbits by definition $n = \sqrt{\mu/(-a^3)}$ (see Sect. 7.4.3).

This idealized motion gives rise to the further definition to the *mean anomaly*

$$\boxed{M := n \cdot \left(t - t_p\right)} \qquad \textbf{mean anomaly} \tag{7.3.11}$$

Because both n and M can be applied to elliptic and hyperbolic orbits they are highly useful. The particular usefullness of M is due to the fact that it serves two different purposes in one parameter. First, by definition M is an angle that by intention advances linearly with time. Just as for the true anomaly θ the mean anomaly M is measured relative to the periapsis. But, in contrast to θ, M is not cyclically limited to the interval [0°, 360°]. As an unambiguous angle M perfectly serves as an orbital element (see Sect. 7.3.5). Second, because M is strictly linear with time and in addition is dimensionless, it is a perfect substitution for time and hence a orbit position sequencer.

> Because the mean anomaly M is an unambiguous angle, even for multiple revolutions, and in addition advances linearly in time, M is an ideal substitute for the true anomaly θ (cf. Sect.7.3.5) *and* the time variable t. Therefore, it is used in celestial mechanics as a standard orbital element. In particular, because $dM \propto dt$ the mean anomaly M is regularly used as the integrating variable for time averages (see, e.g., Sect.12.3).

For later purposes we derive from Eqs. (7.3.8) and (7.3.9) the following useful relation

$$\frac{dM}{d\theta} = \frac{\left(1 - e^2\right)^{3/2}}{\left(1 + e \cdot \cos\theta\right)^2} = \left(\frac{r}{a}\right)^2 \frac{1}{\sqrt{1 - e^2}} \tag{7.3.12}$$

7.3.3 Orbital Velocity

We now want to determine the orbital velocity $\boldsymbol{v}$ at any point on the orbit (see Fig. 7.9). With the identity $\hat{\boldsymbol{h}} \times \left(\boldsymbol{v} \times \hat{\boldsymbol{h}}\right) = \boldsymbol{v}$ we derive from Eq. (7.3.2)

$$\boxed{\boldsymbol{v} = \frac{\mu}{h} \hat{\boldsymbol{h}} \times (\boldsymbol{e} + \hat{\boldsymbol{r}})} \qquad \textbf{orbital velocity} \tag{7.3.13}$$

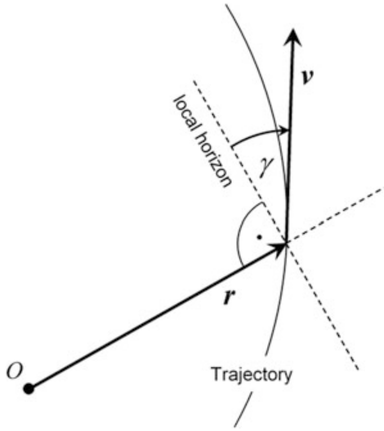

Fig. 7.9 Flight path angle γ measured relative to the local horizon in the co-moving reference system

For its absolute value we find

$$v = \frac{\mu}{h}\sqrt{\left[\hat{\boldsymbol{h}} \times (\boldsymbol{e} + \hat{\boldsymbol{r}})\right]\left[\hat{\boldsymbol{h}} \times (\boldsymbol{e} + \hat{\boldsymbol{r}})\right]}$$
$$= \frac{\mu}{h}\sqrt{\hat{\boldsymbol{h}}^2 (\boldsymbol{e} + \hat{\boldsymbol{r}})^2 - \left[\hat{\boldsymbol{h}}(\boldsymbol{e} + \hat{\boldsymbol{r}})\right]^2}$$

Because $\boldsymbol{h}$ is orthogonal on both $\boldsymbol{e}$ and $\boldsymbol{r}$, the second term in the radicand vanishes and we have with $\boldsymbol{e}\hat{\boldsymbol{r}} = e \cos\theta$

$$v = \frac{\mu}{h}|\boldsymbol{e} + \hat{\boldsymbol{r}}| = \frac{\mu}{h}\sqrt{1 + 2e\cos\theta + e^2} = \sqrt{\mu\left(\frac{2}{r} - \frac{1}{a}\right)} \tag{7.3.14}$$

where the latter follows from Eq. (7.2.15). For the radial part of the velocity we find

$$\dot{r} = \boldsymbol{v}\hat{\boldsymbol{r}} = \frac{\mu}{h}\left[\hat{\boldsymbol{h}} \times (\boldsymbol{e} + \hat{\boldsymbol{r}})\right]\hat{\boldsymbol{r}} = \frac{\mu}{h}\hat{\boldsymbol{h}}[(\boldsymbol{e} + \hat{\boldsymbol{r}}) \times \hat{\boldsymbol{r}}] = \frac{\mu}{h}\hat{\boldsymbol{h}}(\boldsymbol{e} \times \hat{\boldsymbol{r}})$$

and because $\boldsymbol{e} \times \hat{\boldsymbol{r}} = e \sin\theta \cdot \hat{\boldsymbol{h}}$ we finally have

$$\dot{r} = \frac{e\mu}{h}\sin\theta \tag{7.3.15a}$$

Vector representations of $\boldsymbol{v}$ in various reference frames are given in Sect. 13.1.5.

Flight Path Angle
We now define the much used flight path angle γ. This is the angle that the velocity vector $\boldsymbol{v}$ makes with the *local horizon*, the vertical on the radial vector. According to Fig. 6.9 we have

$$\dot{r} = \boldsymbol{v}\,\hat{\boldsymbol{r}} = v\,\sin\gamma \tag{7.3.15b}$$

$$r\dot{\theta} = \boldsymbol{v}(\boldsymbol{h}\times\hat{\boldsymbol{r}}) = v\,\cos\gamma \tag{7.3.15c}$$

and therefore with Eqs. (7.3.15a), (7.3.14) and (7.2.8)

$$\begin{aligned}
\sin\gamma &= \frac{e\,\sin\theta}{\sqrt{1+2e\,\cos\theta+e^2}} = \sqrt{1-\frac{\mu p}{r^2v^2}} = \sqrt{1-\frac{p}{r(2-r/a)}}\\
\cos\gamma &= \frac{1+e\,\cos\theta}{\sqrt{1+2e\,\cos\theta+e^2}} = \frac{\sqrt{\mu p}}{r\,v} = \sqrt{\frac{p}{r(2-r/a)}}
\end{aligned} \tag{7.3.16}$$

where we have also made use of relation $r\omega = \sqrt{\mu p}/r = h/r$.

We thus can express the orbital velocity in the orbit plane also as (cf. Eq. (13.1.10))

$$\boldsymbol{v} = \frac{\mu}{h}\begin{pmatrix} e\,\sin\theta \\ 1+e\,\cos\theta \end{pmatrix}_{RSW} = v\begin{pmatrix} \sin\gamma \\ \cos\gamma \end{pmatrix}_{RSW} \tag{7.3.17}$$

7.3.4 Orbital Energy

We intuitively know that the orbital energy must somehow depend on its shape or size or on both. To derive the relationship, we square Eq. (7.3.2) on both sides

$$\mu^2(\boldsymbol{e}+\hat{\boldsymbol{r}})^2 = \mu^2\left(e^2+2\boldsymbol{e}\hat{\boldsymbol{r}}+1\right) = (\boldsymbol{h}\times\boldsymbol{v})^2 = h^2\cdot v^2$$

The latter holds because of $\boldsymbol{h}\perp\boldsymbol{v}$. From Eq. (7.3.4) it follows that $\boldsymbol{e}\hat{\boldsymbol{r}} = h^2/\mu r - 1$. Equation (7.2.14) states $v^2 = 2\mu/r + 2\varepsilon$. This applied to the above equation leads to

$$2h^2\varepsilon = \mu^2\left(e^2-1\right) \tag{7.3.18}$$

From this we derive:

> The specific orbital energy ε is negative for ellipses with $e<1$, zero for parabolas with $e=1$, or positive for hyperbolas with $e>1$.

With Eqs. (7.3.7) and (7.3.18) can be transformed into the simple expression for the specific orbital energy

$$\boxed{\varepsilon = -\frac{\mu}{2a}} \tag{7.3.19}$$

Equation (7.3.19) proves that the parameter a (semi-major axis) defined in Eq. (7.3.6) is identical to the one in the vis-viva Eqs. (7.2.11) and (7.2.15), and thus the geometrical interpretation of the semi-major axis formerly announced is correct.

Equation (7.3.19) not only is remarkably simple, but is also remarkable because as the vis-viva Eq. (7.2.15), it directly relates an important physical quantity, here the "total orbital energy", to the geometrical size of the orbit, the semi-major axis. Thereby,

The orbital energy can be directly read from the orbital size: The larger the orbit, the larger (negative sign!) the orbital energy.

This interrelation can be easily viewed from Fig. 7.7.

Does also the angular momentum h directly relate to a geometrical property of the orbit? Absolutely, Eq. (7.3.6) is the one. It relates h to the semi-latus rectum p: $h^2 = \mu p$. However, p is not frequently used to characterize an orbit, except for a parabola. Rather a and e are used more often because they are orbital elements (see Sect. 7.3.5). In fact, and according to Eq. (7.3.18), the eccentricity can be employed to gauge the angular momentum at a given orbital energy: If $e = 1$ (radial orbit, see Sect. 7.5), then the angular momentum vanishes, and for $e = 0$ (circular orbit) the angular momentum becomes maximum, $h = \mu\sqrt{a}$. At a given orbital energy, the circular orbit is hence that bounded orbit with the biggest angular momentum.

Virial Theorem

A body on its trajectory continuously changes position and velocity, and along with this its potential and kinetic energy. Since the orbital energy is constant, kinetic energy is transformed into potential energy and the other way round, i.e., $\dot{\varepsilon}_{kin} \propto -\dot{\varepsilon}_{pot}$. Only on a circular orbit, ε_{kin} and ε_{pot} remain constant according to (see Eq. (7.4.5))

$$\varepsilon_{pot} = -2\varepsilon_{kin} = const \quad @ \text{ circular orbit}$$

However, from Sect. 11.1.2 it follows that, if $\langle\varepsilon_{kin}\rangle$ and $\langle\varepsilon_{pot}\rangle$ are the kinetic and potential energies time-averaged over an orbital period, then the following holds

$$\langle\varepsilon_{pot}\rangle = -2\langle\varepsilon_{kin}\rangle = const \quad \textbf{virial theorem} \tag{7.3.20}$$

7.3.5 Orbital Elements

An orbital element is any quantity that specifies an orbit under consideration. We have already derived orbital elements that remain constant even under orbital motion: angular momentum $\boldsymbol{h}$, eccentricity vector $\boldsymbol{e}$, and orbital energy ε. Constant orbital elements are also called *integrals of motion* (a.k.a. *invariants of motion*) and, evidently, they are of particular interest to characterize an orbit. Although the vectors $\boldsymbol{h}$, $\boldsymbol{e}$ are of high analytical relevance, we need to embed them into an inertial frame of reference, in which they constitute descriptive values.

Invariant Angular Elements

The classical way of doing this is using the geocentric equatorial coordinate system *IJK* (see Sect. 13.1.4), in which the Cartesian coordinate system $(\boldsymbol{u}_I, \ \boldsymbol{u}_J, \ \boldsymbol{u}_K)$ has the following orientation (see Fig. 13.2): The unit vector $\boldsymbol{u}_I$ points to the vernal point, $\boldsymbol{u}_K$ to Earth's north pole, and $\boldsymbol{u}_J = \boldsymbol{u}_K \times \boldsymbol{u}_I$. With this we establish the unit vector

$$\boldsymbol{u}_n = \frac{\boldsymbol{u}_K \times \hat{\boldsymbol{h}}}{\left|\boldsymbol{u}_K \times \hat{\boldsymbol{h}}\right|} \quad \text{unit vector to the ascending node}$$

These unit vectors determine the following angular orbital elements (see Fig. 7.10)

$$i = \arccos\left(\boldsymbol{u}_K \hat{\boldsymbol{h}}\right) \qquad 0 \le i < 180° \tag{7.3.21}$$

$$\Omega = \begin{cases} \arccos(\boldsymbol{u}_I \boldsymbol{u}_n) & @ \ \boldsymbol{u}_J \boldsymbol{u}_n \ge 0 \\ 2\pi - \arccos(\boldsymbol{u}_I \boldsymbol{u}_n) & @ \ \boldsymbol{u}_J \boldsymbol{u}_n < 0 \end{cases} \qquad 0 \le \Omega < 360° \tag{7.3.22}$$

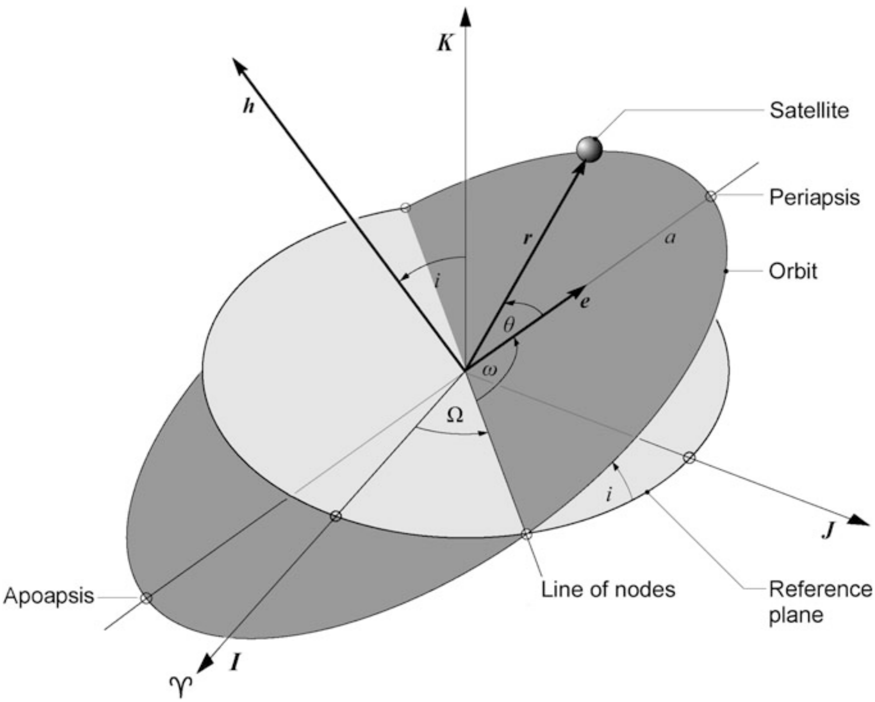

Fig. 7.10 The graphical representation of orbital elements of an elliptic satellite orbit in a geocentric equatorial coordinate system *IJK* (see Sect. 13.1.3), where the plane spanned by *I* and *J* with the *I*-axis oriented toward the vernal point ♈ is the reference plane. The orbital plane and reference plane intersect in the so-called line of nodes. The line of nodes in turn intersects the orbit in the so-called ascending node (where the satellite moves toward the upper side with respect to ***h*** of the reference plane) and in the descending node (vice versa)

$$\omega = \begin{cases} \arccos(\boldsymbol{u}_n \hat{\boldsymbol{e}}) & @ \ \boldsymbol{u}_K \hat{\boldsymbol{e}} \geq 0 \\ 2\pi - \arccos(\boldsymbol{u}_n \hat{\boldsymbol{e}}) & @ \ \boldsymbol{u}_K \hat{\boldsymbol{e}} < 0 \end{cases} \quad 0 \leq \omega < 360^\circ \tag{7.3.23}$$

The following classification of orbits according to their inclination is common:

$i = 0^\circ$	*Equatorial Orbit*. Prominent examples are the geostationary orbits.
$0^\circ < i < 90^\circ$	*Prograde Orbit*. This orbit is used by most satellites in Low Earth Orbits.
$i \approx 90^\circ$	*Polar Orbit*. Orbits used for sun-synchronous Earth observation satellites.
$i > 90^\circ$	*Retrograde Orbit*. Very few satellites have been put into this type of orbit.

Invariant Orbit Vectors

In terms of these angular elements, the eccentricity vector can be represented in the *IJK* system as follows (see also Sect. 13.1.5)

$$\boldsymbol{e} = e \begin{pmatrix} \cos\Omega \cos\omega - \sin\Omega \sin\omega \cos i \\ \sin\Omega \cos\omega + \cos\Omega \sin\omega \cos i \\ \sin\omega \sin i \end{pmatrix}_{IJK} \quad \textbf{eccentricity vector} \tag{7.3.24}$$

Equivalently, the angular momentum unity vector $\hat{\boldsymbol{h}}$, also known as the *three-dimensional inclination vector* (a.k.a. *orbit pole vector*), given as (see Eq. (13.1.5)) as

$$\boldsymbol{I} \equiv \hat{\boldsymbol{h}} = \begin{pmatrix} \sin\Omega \sin i \\ -\cos\Omega \sin i \\ \cos i \end{pmatrix}_{IJK} \quad \textbf{three-dimensional inclination vector} \tag{7.3.25}$$

For instance, in Fig. 12.28 this 3D-inclination vector depicts the evolution of i, Ω under lunisolar perturbations. Its projection onto the *IJ* reference plane is known as

$$\boldsymbol{i} = \sin i \begin{pmatrix} \sin\Omega \\ -\cos\Omega \end{pmatrix}_{IJ} \quad \text{(two-dimensional)} \ \textbf{inclination vector} \tag{7.3.26}$$

Note that for $i \to 0$ conveniently

$$\boldsymbol{i} = i \begin{pmatrix} \sin\Omega \\ -\cos\Omega \end{pmatrix}_{IJ} \quad @ \ i \to 0$$

We now rotate the inclination vector in the IJ reference plane by 90° counter-clockwise into the direction of the ascending node and thus get the *ascending node vector*

$$\boldsymbol{n} = \sin i \begin{pmatrix} \cos\Omega \\ \sin\Omega \end{pmatrix}_{IJ} = \sin i \begin{pmatrix} \cos\Omega \\ \sin\Omega \end{pmatrix}_{EQ} \quad \textbf{ascending node vector} \qquad (7.3.27)$$

All the above invariant quantities are orbital elements. This gives rise to the question of whether there is a minimum number of orbital elements that uniquely specify an orbit and how many of them are constant.

State Vector

We investigate the above question by going back to the basics of orbital motion. Depending on the initial conditions, orbital motion takes place in 6-dimensional phase space $\boldsymbol{r}$, $\boldsymbol{v}$, a space in which all possible states of a system are represented, with each possible state corresponding to one unique point in the phase space. Once an initial state $(\boldsymbol{r}_0, \boldsymbol{v}_0)$ is given, the two first-order differential vector equations

$$\dot{\boldsymbol{r}} = \boldsymbol{v}, \quad \dot{\boldsymbol{v}} = -\frac{\mu}{r^3}\boldsymbol{r}$$

derived from Newton's gravitation equation of motion (7.1.19) determine the future motion in phase space. Obviously, phase space has 6 dimensions and therefore the 6 elements of an initial state vector are the minimum number of orbital elements plus the independent parameter t. This is the significance of the state vector $(\boldsymbol{r}, \boldsymbol{v})$ as introduced in Sect. 7.2.2. It also has the advantage that it is the direct outcome of ranging and range rate measurements as part of orbit determination procedures (see Sect. 14.2.1). On the other hand it has the major drawback that $\boldsymbol{r}$ and $\boldsymbol{v}$ continuously change with time.

Number of Constant Orbital Elements

So, what is the maximum number of constant orbital elements? We have seen in Sect. 7.1.3 that linear momentum, angular momentum, and total energy always need to be conserved. Because the motion of a 2-body system in a central force field, such as the gravitational field and as shown in Sect. 7.2.2, always takes place in just 2 dimensions, we expect that any orbital motion exhibits $2 + 2 + 1 = 5$ constant (integrals) of motion and hence 5 constant orbital elements. This leaves one degree of freedom to the orbital motion, which is the 1-dimensional line of orbit. We therefore expect one additional orbital element to determine the moving position of the body on its orbit. Hence, in total there must be 5 constant orbital elements and 1 time-dependent orbital element. Obviously $\boldsymbol{h}$, $\boldsymbol{e}$ and $\theta(t)$ lend themselves as such orbital elements. Although they seem to make up $3 + 3 + 1 = 7$ elements, one has to consider the constraint $\boldsymbol{h} \cdot \boldsymbol{e} = 0$, and therefore the set $(\boldsymbol{h}, \boldsymbol{e}, \theta(t))$ actually represents the set of 6 independent elements and hence a minimum complete set of orbital elements.

We have seen that we may choose many different sets of 5 constant orbital elements, and any set is equivalent to any other. Next we to explore some common sets of orbital elements.

Common Sets of Orbital Elements
The most vivid orbital elements are the angular orbital elements i, Ω, ω (see Fig. 7.10) as given in Eq. (7.3.21), (7.3.22), and (7.3.23). In addition and from Eq. (7.3.7) we have from $\boldsymbol{h}$, $\boldsymbol{e}$

$$e = |\boldsymbol{e}|$$
$$a = \frac{\boldsymbol{h}^2}{\mu(1 - e^2)}$$

We thus obtain the highly important and so-called six *classical orbital elements* (see Fig. 7.10)

Classical Orbital Elements $\left(a, e, i, \Omega, \omega \,/\, t - t_p \leftrightarrow M\right)$

Two elements, so-called *metric elements*, describe the dimension of the orbit:

- ***a*** **Semi-major axis**, $-\infty < a < \infty$.
  It specifies the size of the orbit. Hyperbolic orbits have negative values.
- ***e*** **Eccentricity**, $0 \le e < \infty$.
  It specifies the shape and jointly with a the type of the orbit.

The next three elements are called *angular elements*. Two elements describe the orientation of the orbital plane:

- ***i*** **Inclination**, $0 \le i < 180°$
  Angle between the angular momentum vector h and the z-direction of the reference frame = angle between the corresponding planes perpendicular to these vectors.
- **Ω** **Right ascension of ascending node** (abbreviated: RAAN), $0 \le \Omega < 360°$
  Longitude of the ascending node = angle between the line from the origin O of the reference frame to the vernal point and from O to the ascending node.

One element *determines* the orientation of the orbit in the orbital plane:

- ***ω*** **Argument of periapsis**, $0 \le \omega < 360°$
  Angle in the orbital plane between line of nodes and the periapsis measured in the direction of the motion.

The sixth parameter is the orbital state sequencer, it determines the location on the orbit. The following two sequencers can be used alternatively:

- $\boldsymbol{t - t_p}$ **Time after periapsis passage**
  $\boldsymbol{t_p}$ is the time, when the satellite was at periapsis (a.k.a. *epoch*).

or

M **Mean anomaly**, $M > 0$
$M := n \cdot (t - t_p)$, dimensionless (see Sect. 7.3.2).

Although M formally is an angle (however not having an geometrical interpretation) with dimension [rad], it commonly is treated as dimensionless and hence in modern orbital mechanics is a more convenient variant of time. Both sequencers have the disadvantage that their relationship to r and θ (see Sect. 7.4.2) is transcendental.

There is an alternative and more traditional set of orbital elements, called *Keplerian elements*

Keplerian Elements $(a, e, i, \Omega, \omega \ / \ \theta \leftrightarrow E,\ F, G)$
The Keplerian elements share the metric and angular classical orbital elements. The following two sequencers can be used alternatively:

θ **True anomaly** (a.k.a. *orbit angle*), $0 \leq \theta < 360°$
Angle between the direction to the periapsis and the current position vector $\boldsymbol{r}$.

or

E **Eccentric anomaly** for elliptic orbits (see Sect. 7.4.2), $0 \leq E < 2\pi$
F **Hyperbolic anomaly** for hyperbolic orbits (see Sect. 7.4.3), $0 \leq F < \infty$
G **Universal anomaly** for all orbits (see Sect. 7.4.6), $0 \leq G < \infty$

The advantage of this set is that the determination of $r(\theta)$, or equivalently $r(E, F, G)$, is algebraic and hence straightforward. The disadvantage is that θ is not a convenient orbital state sequencer, but time t. The inclusion of θ, which Leonhard Euler called "angle of elements", into the set of elements dates back to Johann Bernoulli.

Sometimes the compound angle

$$u = \omega + \theta \qquad \textbf{argument of latitude}$$

i.e., the position of the orbit measured relative to the line of nodes, is a more valuable parameter for orbit analysis. Note that, though $\boldsymbol{h},\ \boldsymbol{e},\ \theta(t)$ usually are defined in the geocentric equatorial coordinate system, of course any inertial reference system will do and only the angular elements will change by a different choice of reference system.

Degenerate Orbits
There exist two particular cases where some of the angular elements become undefined:

1. ω and θ (or M) will become undefined for $e = 0$.
2. Ω and ω will become undefined for $i = 0$.

Both happens, for example, for satellites in geostationary orbit. These degenerate cases, however, pose no problem.

In the first case, $e = 0$, where ω becomes irrelevant the argument of latitude $u = \theta$ is used as a substitute to determine the orbital position, i.e. the true anomaly is measured relative the line of nodes.

In the second case, the compound angles

$\varpi = \Omega + \omega$	**longitude of periapsis**	@ $i = 0$
$L = \Omega + \omega + \theta = \varpi + \theta = \Omega + u$	**orbital longitude**	@ $i, e = 0$
$l = \Omega + \omega + M = \varpi + M$	**mean longitude**	@ $i, e = 0$

measured relative to the vernal point ϒ are used as substitutes for ω and θ or M (Ω is irrelevant for $i = 0$). The line of nodes and the vernal point then become the respective common reference directions in the orbital plane. Note that Ω is measured in the reference plane, while ω and θ are measured in the orbital plane. So, for $i > 0$, the compound angles ϖ, L, and l would be "doglegged" and no longer be true angles, but nevertheless they can be defined as mathematical quantities and treated like that. These doglegged compound angles are frequently used in the literature, but will not be used in this book.

Equinoctial Elements and Coordinate System

One problem of orbit degeneracy with angular elements arises because Ω and ω are measured relative to the ascending line of nodes, which vanishes for $i = 0$. This problem can be removed by rotating the *IJ* reference plane along the line of nodes by the inclination angle into the orbital plane and renaming the axes $K \to W$, $I \to E$, $J \to Q$. This reference frame is called *equinoctial coordinate system EQW* (see Fig. 7.11). By construction, for $i \to 0$ it crosses over into the *IJK*

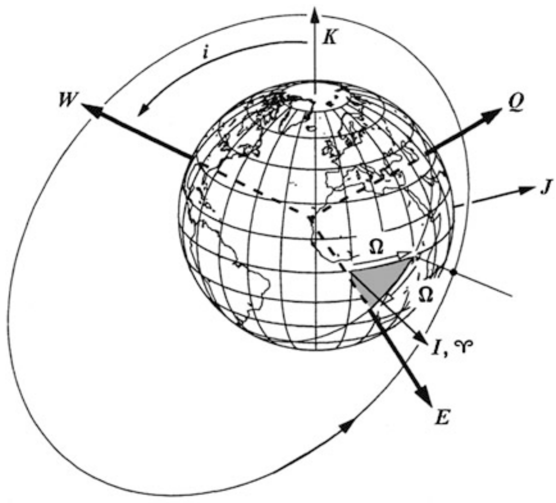

Fig. 7.11 The equinoctial coordinate system EQW. This system is formed from the IJK system by rotating IJK by the inclination angle about the line of nodes. *Credit* Vallado (2007)

coordinate system. The other orbit degeneracy problem, namely that ω and θ (or M) will become undefined for $e = 0$, can be removed by making the eccentricity vector part of the orbital elements. Both strategies are adopted to form a nonsingular set of orbital elements, the so-called **standard equinoctial orbital elements**, which read as follows:

(Standard) Equinoctial Elements (a, h, k, p, q, l)

a Semi-major axis, $-\infty < a < \infty$

Two components of the eccentricity vector:

h $= e\ \sin(\omega + \Omega)$
k $= e\ \cos(\omega + \Omega)$

Two components of the rescaled ascending node vector:

p $= \tan\frac{i}{2}\sin\Omega$
q $= \tan\frac{i}{2}\cos\Omega$

The sixth parameter again determines the location of the periapsis in time:

l $= \Omega + \omega + M(t)$, mean longitude, $0 \le l < 360°$

Note that the RAAN angle Ω now lies in the orbital plane and therefore the mean longitude $l = \Omega + \omega + M$ is no longer doglegged, but a true angle in the orbital plane. Since the equinoctial elements exhibit no singularities they are frequently employed in perturbational equations for equatorial orbits. For more details of equinoctial elements see Battin (1987), and for coordinate transformations to and from the equinoctial coordinate system see Vallado (2007).

In terms of the invariant orbit vectors we can identify

$$\boldsymbol{e} = \begin{pmatrix} h \\ k \end{pmatrix}_{EQ} = e\begin{pmatrix} \sin(\omega + \Omega) \\ \cos(\omega + \Omega) \end{pmatrix}_{EQ} \tag{7.3.28}$$

and the vector (q, p) actually is a rescaled ascending node vector

$$(q, p) = \frac{1}{1 + \cos i}\boldsymbol{n} = \tan\frac{i}{2} \cdot \begin{pmatrix} \cos\Omega \\ \sin\Omega \end{pmatrix}_{EQ}$$

The reason for scaling by $1/(1 + \cos i)$ lies in the better adaptation to Lagrange's planetary equations where the equinoctial elements are frequently used for GEO orbits.

The introduction of the singular free parameters (h, k) and (q, p) is most probably due to Joseph-Louis Lagrange. Today, there are some variants of the standard equinoctial orbital elements, most importantly the so-called **modified equinoctial orbital elements** (Walker et al. 1985) with the only modification $a \to p = a(1 - e^2)$. But also the replacements $\tan\frac{i}{2} \to \tan i$, $\sin i$, or $\sin\frac{i}{2}$ are quite common.

Conversion: State Vector → Orbital Elements

A frequent problem encountered in astrodynamics, in particular with orbit determination, is the conversion of the state vector $(\boldsymbol{r}, \boldsymbol{v})$ into (osculating) Keplerian elements and vice versa. In the following, we will have a look at this problem.

First, we want to convert a state vector $(\boldsymbol{r}, \boldsymbol{v})$ to Keplerian elements. To start with, we compute

$$\boldsymbol{h} = \boldsymbol{r} \times \boldsymbol{v}$$
$$\hat{\boldsymbol{h}} = \boldsymbol{h}/h, \quad \hat{\boldsymbol{r}} = \boldsymbol{r}/r, \quad \hat{\boldsymbol{v}} = \boldsymbol{v}/v$$

The semi-major axis then is determined from the vis-viva Eq. (7.2.14) as

$$a = \frac{\mu r}{2\mu - rv^2} \tag{7.3.29}$$

Observe that the semi-major axis and thus the orbital period $T = 2\pi\sqrt{a^3/\mu}$ can be determined merely from the absolute values of the position vector and orbital velocity. For the eccentricity we have from Eq. (7.3.3) and with Eq. (7.3.6)

$$e = |\boldsymbol{e}| = \sqrt{1 - \frac{h^2}{\mu a}}$$
$$\hat{\boldsymbol{e}} = \boldsymbol{e}/e$$

$$\theta = \begin{cases} \arccos(\hat{\boldsymbol{r}}\hat{\boldsymbol{e}}) @ \quad \hat{\boldsymbol{r}}\hat{\boldsymbol{v}} \geq 0 \\ 2\pi - \arccos(\hat{\boldsymbol{r}}\hat{\boldsymbol{e}}) @ \quad \hat{\boldsymbol{r}}\hat{\boldsymbol{v}} < 0 \end{cases} \tag{7.3.30}$$

The angular orbital elements i, Ω, ω are given in the geocentric equatorial reference system IJK by Eqs. (7.3.21), (7.3.22), and (7.3.23). To determine the mean anomaly M, we derive from Eq. (7.4.13a)

$$\begin{array}{ll} E = \arccos[(1 - r/a)/e] & @ \quad |\hat{\boldsymbol{r}} \times \hat{\boldsymbol{e}}| > 0 \\ E = 2\pi - \arccos[(1 - r/a)/e] & @ \quad |\hat{\boldsymbol{r}} \times \hat{\boldsymbol{e}}| < 0 \end{array} \tag{7.3.31}$$

Hence from the Kepler's Eq. (7.4.15) we derive

$$M = E - e \sin E$$

and from Eq. (7.4.14d) we have alternatively

$$\theta = 2 \arctan\left(\sqrt{\frac{1+e}{1-e}} \tan\frac{E}{2}\right)$$

Conversion: Orbital Elements → State Vector
Let the orbital elements $a, e, i, \omega, \Omega, \theta$ or $a, e, i, \omega, \Omega, M$ be given. The orbit radius then is determined as

$$r = \frac{a(1-e^2)}{1+e\cos\theta} = a(1-e\cos E)$$

where E is the solution of $M = E - e\sin E$ (see Newton's method in Sect. 7.4.2). As shown in Sect. 13.1.5, the radial vector and the orbital velocity vector in the *IJK* system (see Sect. 13.1.3) are then determined as

$$\boldsymbol{r} = r\begin{pmatrix} \cos\Omega\cos u - \sin\Omega\sin u\cos i \\ \sin\Omega\cos u + \cos\Omega\sin u\cos i \\ \sin u\sin i \end{pmatrix}_{IJK} \tag{7.3.32}$$

$$\boldsymbol{v} = \sqrt{\frac{\mu}{a(1-e^2)}}\begin{pmatrix} -\cos\Omega(\sin u + e\sin\omega) - \sin\Omega(\cos u + e\cos\omega)\cos i \\ -\sin\Omega(\sin u + e\sin\omega) + \cos\Omega(\cos u + e\cos\omega)\cos i \\ (\cos u + e\cos\omega)\sin i \end{pmatrix}_{IJK} \tag{7.3.33}$$

with $u = \omega + \theta$ the argument of latitude. For equatorial orbits Ω is undefined and can be set to zero. The same holds for ω in circular orbits, and hence $u = \theta$.

7.4 Keplerian Orbits

In this chapter we study the detailed properties of the different Keplerian orbits (a.k.a. *conic orbits*) and derive and present their basic results. This includes, in particular, the analytic solutions to $t = t(\theta)$ and the numeric algorithms to calculate $\theta = \theta(t)$ for elliptic and hyperbolic orbits.

7.4.1 Circular Orbit

The most common orbits of spacecraft around celestial bodies are circular or near-circular in shape, since they take on the highest minimum orbital altitude (implies least atmospheric drag) at a given orbital energy (cf. Fig. 7.7) and provide steady orbit conditions. According to Eq. (7.3.18) they also feature maximum angular momentum at a given orbital energy. For $e = 0$, Eq. (7.3.9) can be solved immediately and according to Eq. (7.3.5) we get

$$\begin{aligned} r &= a = p = const \\ \theta &= \sqrt{\frac{\mu}{a^3}}\cdot\left(t - t_p\right) = n\left(t - t_p\right) \end{aligned} \tag{7.4.1}$$

which is a circular orbit with radius a and with a negative specific orbital energy according to Eq. (7.3.19). From Eq. (7.4.1) we find $2\pi = T\sqrt{\mu/a^3}$ for a full revolution, and hence for the orbital period T (cf. Eq. (7.4.12))

$$\boxed{T = 2\pi\sqrt{\frac{a^3}{\mu}}} \qquad \textbf{orbital period} \tag{7.4.2}$$

To determine the orbital velocity, we use vis-viva Eq. (7.2.15) from which with $r = a$ we directly derive

$$\boxed{v = \sqrt{\frac{\mu}{r}} = \sqrt{\frac{\mu}{a}} = const} \tag{7.4.3}$$

The orbital velocity decreases with the root of the orbital radius; for example, the velocity of a body that circles the Earth in the theoretically lowest orbit possible, $r = R_\oplus$, is with $\mu_\oplus = g_0 R_\oplus^2$ (see Eq. (7.1.20))

$$v_\triangleright = \sqrt{g_0 R_\oplus} = 7.905 \text{ km/s} \qquad \textbf{first cosmic velocity} \tag{7.4.4}$$

This is the highest possible orbital circular velocity around the Earth, as according to Eq. (7.4.3), the circular velocity decreases with an increasing orbital altitude r. In a typical LEO (such as the ISS) of 400 km, it is only 7.67 km s^{-1}. Though speed decreases for higher orbits you still need more energy to reach higher orbits. This is because from Eq. (7.3.19) $\varepsilon = -\mu/2a = -\mu/2r$, i.e.,

$$2\varepsilon = \varepsilon_{pot} = -2\varepsilon_{kin} = const \tag{7.4.5}$$

The latter follows because $\varepsilon_{kin} = v^2/2 = \mu/2a$. In words, this means that the orbital energy of a circular orbit is negative and its absolute value equals that of the kinetic energy but, most important, that (cf. Eq. (7.3.20))

At any point in the circular orbit, the absolute value of the potential energy is twice that of the kinetic energy.

Note *Here and in general the potential energy is not the positive energy measured relative to the surface of the celestial body being circled, but the negative energy relative to infinity* (cf. Eq. (7.1.3) *and* Fig. 7.7).

Therefore, the energy for lifting a body into a higher circular orbit is determined as follows:

$$\Delta\varepsilon = \Delta\varepsilon_{kin} + \Delta\varepsilon_{pot} = -\frac{1}{2}\Delta\varepsilon_{pot} + \Delta\varepsilon_{pot} = \frac{1}{2}\Delta\varepsilon_{pot} > 0$$

Orbit lifting reduces the kinetic energy by a given amount, but it increases potential energy by double that amount. Therefore, although orbit lifting increases the orbital energy the orbital velocity decreases.

On the other hand, this leads to the paradoxical situation (a.k.a. *satellite orbit paradox*) that a S/C in a circular LEO, decelerated due to drag in effect gains velocity because it spirals down to lower altitudes.

7.4.2 Elliptic Orbit

For $0 < e < 1$, the orbit is an ellipse, which is the most general bounded orbit in a gravitational field (**Kepler's first law**). According to Eq. (7.3.19), its orbital energy is negative. According to Eq. (7.3.5) and with Eq. (7.3.6), there is a minimum and a maximum distance to the central body at the focal point (see Fig. 7.12):

$$r_{per} = a(1 - e) \qquad \textbf{periapsis} \tag{7.4.6}$$

$$r_{apo} = a(1 + e) \qquad \textbf{apoapsis} \tag{7.4.7}$$

The general terms periapsis and apoapsis are also called *pericenter* and *apocenter*. Depending on the central body, the specific terms are peri-/apogee (Earth), peri-/aphelion (Sun), peri-/aposelene (Moon), etc. The line through the periapsis and apoapsis is called line of apsides (a.k.a. *apse line*). From Eqs. (7.4.6) and (7.4.7), the semi-major axis *a,* the eccentricity *e*, and the semi-minor axis *b* are derived as

$$a = \frac{1}{2}\left(r_{apo} + r_{per}\right) \tag{7.4.8}$$

$$e = \frac{r_{apo} - r_{per}}{r_{apo} + r_{per}} \tag{7.4.9}$$

$$b = a\sqrt{1 - e^2} = h\sqrt{\frac{a}{\mu}} = \langle r \rangle_\theta$$

Geometrically, the eccentricity is the distance Δ from the center of the ellipse to its focal point in units of a (see Fig. 7.10): $\Delta/a = \left(a - r_{per}\right)/a = ae/a \ = e.$

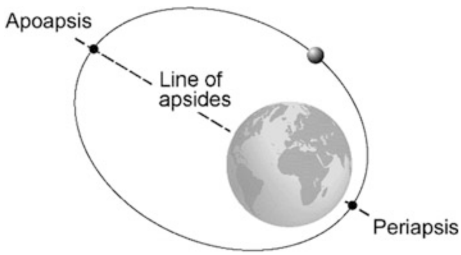

Fig. 7.12 Apoapsis, periapsis, and line of apsides of an elliptic orbit

According to Appendix A.1.2 the semi-minor axis b corresponds to the mean orbit radius $\langle r \rangle_\theta$ averaged over θ.

The velocities at the periapsis and apoapsis follow from Eq. (7.3.14)

$$v_{per} = \frac{\mu}{h}(1+e) = \sqrt{\frac{\mu}{a}}\sqrt{\frac{1+e}{1-e}} \tag{7.4.10}$$

$$v_{apo} = \frac{\mu}{h}(1-e) = \sqrt{\frac{\mu}{a}}\sqrt{\frac{1-e}{1+e}} \tag{7.4.11}$$

According to Kepler's second law, Eq. (7.2.9), the area ΔA swept by the orbit radius during the time Δt is given by $\Delta A = \Delta t \cdot h/2$. If one integrates over a full revolution, the swept area ΔA is the area of the ellipse, πab, and Δt is the orbital period T, and thus $T = 2\pi ab/h$. Because of Eq. (7.4.9), $b = h\sqrt{a/\mu}$, we get

$$\boxed{T = 2\pi\sqrt{\frac{a^3}{\mu}}} \qquad \textbf{orbital period (Kepler's third law)} \tag{7.4.12}$$

Remark 1 *Actually, Kepler stated in his third law the relation of two orbit periods around the same central body: $T_1^2/T_2^2 = a_1^3/a_2^3$. This follows from the above equation.*

Remark 2 *It is quite remarkable that the orbital period, just as the specific orbital energy $\varepsilon = -\mu/2a$, does not depend on eccentricity, but only on the semi-major axis.*

Remark 3 *Kepler's third law can be used to determine the mass M of a celestial body: By precise determination of the orbital period and semi-major axis of an object around the celestial body the standard gravitational parameter $\mu = GM$ and hence mass M can be determined. For instance, by measuring the orbits of the stars around the black hole at the center of our Milky Way, its mass was thus determined.*

Given the orbital period we find for the mean motion (see Eq. (7.3.10))

$$n = \frac{2\pi}{T} = \sqrt{\frac{\mu}{a^3}} \tag{7.4.13}$$

Kepler Transformation

We now seek to tackle the Keplerian problem, as already discussed in Sect. 7.3.2, and solve Eq. (7.3.9). To do so, we apply Kepler's method, which transfers the problem to a new angle parameter, the so-called **(elliptic) eccentric anomaly *E*** (see Fig. 7.13).

Remark *You may be worried that a Roman letter rather than a Greek symbol symbolizes an angle. The answer to such "anomalies" is as always: for historic reason. Introduced by Kepler and getting used to it for centuries nobody dares to switch to modern standards.*

The transformation is performed geometrically by drawing a great circle around the ellipse with radius a and projecting the position vector onto the horizontal and the vertical axis. From analysis in Fig. 7.13 the segments $r\cos\theta = x = a\cos E - ae$ and $r\sin\theta = y = b\sin E = a\sqrt{1-e^2}\sin E$ we get

$$r\cos\theta = a\cos E - ae = a(\cos E - e)$$
$$r\sin\theta = a\sqrt{1-e^2}\sin E$$

Squaring and applying the equations to each other results in

$$\boxed{r = a(1 - e\cos E)} \qquad \textbf{orbit equation}$$

$$\begin{aligned} \cos\theta &= \frac{\cos E - e}{1 - e\cos E}, & \cos E &= \frac{e + \cos\theta}{1 + e\cos\theta} \\ \sin\theta &= \frac{\sqrt{1-e^2}\sin E}{1 - e\cos E}, & \sin E &= \frac{\sqrt{1-e^2}\sin\theta}{1 + e\cos\theta} \\ \tan\frac{\theta}{2} &= \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2} \end{aligned} \tag{7.4.14}$$

The first equation is nothing else than the orbit equation (7.3.5), with the orbit angle substituted by the eccentric anomaly. It is exactly this simplification of the orbit equation that will make it possible to perform the final required integration, which we will see later.

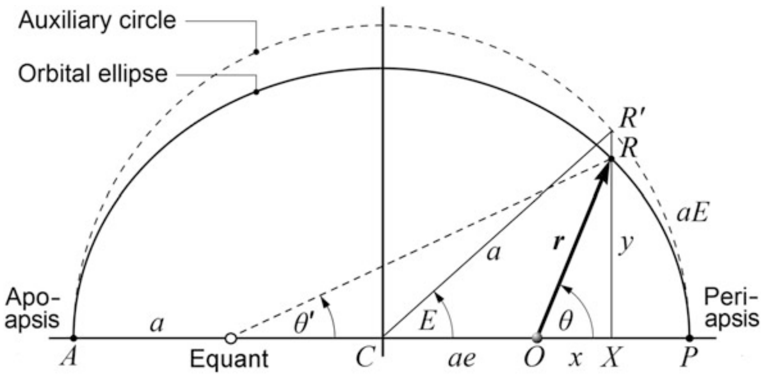

Fig. 7.13 Geometric interpretation of the eccentric anomaly

Note *The mean advancing orbit angle is not E as one would expect intuitively by geometric reasoning from* Fig. 7.13 *but* $M = n(t - t_p)$. *Because* $M = E(t) - e\ \sin E(t)$(see Eq. (7.4.15)), *M does not have a general simple geometric interpretation. Only if* $e \ll 1$, *M can be shown to be close to the orbit angle* θ' *as measured from the empty focal point* (see Fig. 7.13 *and Historic Remarks below*).

Kepler's Equation

With E as a practical orbit angle, we will now determine $E(t)$ to finally derive via Eq. (7.4.14b) the wanted $\theta(t)$. For this we determine with Eqs. (7.3.12) and (7.4.14)

$$\frac{dM}{dE} = \frac{dM}{d\theta}\frac{d\theta}{dE} = \frac{\left(1 - e^2\right)^{3/2}}{\left(1 + e\ \cos\theta\right)^2}\frac{1 + e\ \cos\theta}{\sqrt{1 - e^2}} = \frac{1 - e^2}{1 + e\ \cos\theta} = \frac{r}{a} = 1 - e\ \cos E$$

This equation can readily be integrated to deliver

$$\boxed{E(t) - e\ \sin E(t) = M = n\cdot\left(t - t_p\right)} \quad \textbf{Kepler's equation} \qquad (7.4.15)$$

with t_p is the time of passage through periapsis, where $E(t_p) = \theta(t_p) = 0$. This equation is written the other way round as in the literature to demonstrate that usually the time t or M is given (right-hand side) and $E(t)$ (left-hand side) needs to be determined from this, which obviously is not easy to achieve, because the relation is transcendental. Because according to Eq. (7.4.14b) E is directly linked to θ, we have in principle achieved our goal of determining the body on its orbit as a function of time.

While the orbit equation (7.3.5) determines the shape of an elliptic orbit, Kepler's equation determines the body's position in the orbit at a given time. Therefore both are key equations celestial mechanics and therefore enframed.

Customizing Kepler's Equation

Kepler's equation is of limited practical use, because it is tied to the time of passage through periapsis t_p, which usually is not known. Rather, if at any given time t_0 the state vector $\boldsymbol{r}_0$, $\boldsymbol{v}_0$ (or a set of orbital elements) is known, then from Eqs. (7.2.15), (7.3.3), and (7.4.14a) it follows that

$$a = \frac{\mu}{2\mu/r_0 - v_0^2}$$

$$\mu\boldsymbol{e} = \mu\left(\frac{1}{r_0} - \frac{1}{a}\right)\boldsymbol{r}_0 - (\boldsymbol{r}_0\boldsymbol{v}_0)\boldsymbol{v}_0 \ \rightarrow\ e = \sqrt{\boldsymbol{ee}}$$

$$E_0 = \arccos\left[\frac{1}{e}\left(1 - \frac{r_0}{a}\right)\right]$$

We now extend the above integration over the equation of lateral motion to the intervall $[t_0, t]$

$$\sqrt{\frac{\mu}{a^3}}\int_{t_0}^{t} dt' = \int_{E_0}^{E} (1 - e\cdot\cos E')\cdot dE' = [(1 - e\cdot\sin E)]_{E_0}^{E}$$

which yields

$$E(t) - e\ \sin E(t) = \sqrt{\frac{\mu}{a^3}}(t - t_0) + E_0 - e\ \sin E_0 \tag{7.4.15b}$$

For a given t_0, a, e, E_0 this Kepler's equation can be solved for $E(t)$, which the state vector propagation method in Sect. 7.4.7 makes use of.

If in addition we define $\Delta t = t - t_0$, $\Delta E = E - E_0$, then Kepler's equation can be rewritten (without proof) as

$$\frac{r_0}{a}\sin\Delta E + \frac{r_0 v_0}{\sqrt{\mu a}}(1 - \cos\Delta E) + (\Delta E - \sin\Delta E) = \sqrt{\frac{\mu}{a^3}}\cdot\Delta t \tag{7.4.15c}$$

This a remarkably simple (no eccentricity!) dimensionless equation, on which the widely used *universal variable formulation* described in Sect. 7.4.8 is based on.

Regularization

Kepler's transformation serves two purposes. The first and most obvious is to transform Kepler's problem of integration (see Eq. (7.3.9)) to the simpler problem to finding the root of Kepler's Eq. (7.4.15). The second and often missed benefit is that it regularizes the description of motion. Let us see what regularization means.

As long as the body moves on an almost circular orbit, the motion is obviously quite regular. But if the eccentricity becomes large, there are times when the orbiting body comes close to the central body where it moves very fast and other times where it is far from the central body moving very slowly—the motion is irregular. To demonstrate this mathematically, we set up the time and angular variations of orbit radius. Equation (7.2.11) and differentiating the orbit Eq. (7.3.5) yields

$$\frac{dr}{dt} = \sqrt{\frac{2\mu}{r} - \frac{h^2}{r^2} - \frac{\mu}{a}}$$
$$\frac{dr}{d\theta} = \frac{e}{p}r^2\sin\theta$$

So, if the eccentricity is high the orbit radius r may change by orders of magnitude and so does the time and angular variations. From a numerical point of view this is very undesirable, because the time steps Δt and angular steps $\Delta\theta$ would need to be adjusted correspondingly. If the steps are nevertheless fixed, then the precision of the determination of the orbit position varies between close approaches and remote motion.

Now let's see what Kepler's transformation does for this problem. From (7.4.14a) we derive

$$\frac{dr}{dE} = ea\,\sin\theta$$

Now the variation of the orbit radius with angle E is quite regular. This is not surprising, because if we position ourselves at the center of the orbit ellipse, radial changes look indeed much more regular than from the position at one of the focal points.

Kepler's transformation does not help to fully regularize angular velocities, though, because both

$$\frac{d\theta}{dt} = \frac{h}{r^2}$$
$$\frac{dE}{dt} = \sqrt{\frac{\mu}{a}}\frac{1}{r}$$

exhibit singularities, although for the eccentric angle velocity, dE/dt, it is less pronounced. This remaining irregular variation is embraced in Kepler's equation and it should be kept in mind that it causes numerical problems if $e \gg 0$. Only the universal solution in Sect. 7.4.6 will also regularize this problem.

So, overall we derive the following rule

For numerical purposes the use of the eccentric anomaly E over the orbit angle θ as a variable is to be prefered. For high eccentricity orbits E is superior, for very high eccentricity orbits, mandatory.

Later in Sect. 7.4.6 we will see that the so-called universal anomaly G is the best choice to regularize any orbit with $h > 0$, in particular, high eccentricity orbits, and even orbits with $e \to 1$. Analytically, of course, it does not make any difference, except for $e = 1$.

Historic Remarks

Historically, Fig. 7.13 is of major relevance. Beginning with Apollonius (3rd century BC) the Greek astronomers assumed the so-called eccentric model, in which the planets, in particular the Moon and the Sun, move in circles (see auxiliary circles in Fig. 7.13) about Earth, located somewhat off the center of the circle at position O. However, starting with Ptolemy they evaluated all orbital parameters in terms of the so-called *equant* (a.k.a. *punctum aequans*) which is the empty focal point. The equant has the remarkable property that for the true anomaly θ' (angle between the apsidal line to apogee and the radius vector from the equant to the revolving body) holds (see Problem 7.7)

$$\theta' = M - \frac{1}{4}e^2 \sin(2M) + O(e^3)$$

while for the common true anomaly θ (see Eq. (7.4.19))

$$\theta = M + 2e \sin M + \frac{5}{4}e^2 \sin(2M) + O(e^3).$$

Therefore, as seen from the equant, the angular motions of all planets in our solar system with their small eccentricities look uniform. This fact, which the Greeks knew by observational data and geometrical considerations, together with Plato's doctrine that planets move on circular orbits, made the ancient Greek astronomers, and initially also Kepler, believe in the eccentric model and the epicyclic model, which Ptolemy showed to be equivalent to each other for uniform motion.

Kepler as a geodesist was a mathematician with geometric reasoning. He did not like calculus, a modern science in those days. So, and owing to Copernicus, Kepler applied the eccentric model to the planets (incl. Earth) circling the Sun located in one focal point. But, by studying the precise observational data taken by Tycho Brahe of Mars' orbit, Kepler realized that its orbit must be an ellipse rather than a circle. That the position on the ellipse R at a given time was not on the radius vector CR' to the corresponding position on the auxiliary circle, as he assumed initially, but on the side XR' opposite to E (which is an important ingredient of Kepler's transformation), came to him in a flash of insight by adjusting the precise Mars locations to his elliptical model. With this situation in mind it is obvious, how Kepler was able to think up such an ingeniously simple transformation to solve the Keplerian problem: Drawing circles around the ellipse's center arose quite naturally.

It is interesting to see how Kepler was able to solve the Keplerian problem and arrive at his famous solution Eq. (7.4.15) although he could not apply infinitesimal calculus, as we did, because Newton and Leibniz developed this tool only about one century later. He rather studied the equivalent motion of the body on the auxiliary circle, in particular the area the corresponding position vector OR' sweeps from perigee in a given time interval. He knew from his second law Eq. (7.2.9) that this is proportional to the time interval. From Fig. 7.13 one recognizes that this area equals the circle sector $CR'P$ with area $\frac{1}{2}a^2 E$ minus the triangle $CR'O$, with area $\frac{1}{2}a^2 e \sin E$. So he arrived at $E - e \sin E \propto t$. The proportionality factor he called mean motion. But, owing to the significance of the equant and in line with Greek geometers he defined his eccentric anomaly relative to the apogee, i.e., $E' = 180° - E$. So in his original work the swept area is the sum of the said two subareas and hence $E' + e \sin E' \propto t$.

It was only Leonhard Euler at the beginning of the 18th century, who provided a rigorous analytic derivation of the eccentric anomaly and its relation to the true anomaly. Thus he also regularized the problem and derived Kepler's equation.

Solving Kepler's Equation

Kepler's equation has still the drawback that it cannot be solved analytically for E at a given t. This is achieved only numerically. A common way is Newton's method (a.k.a. Newton-Raphson method). For this one defines the function

$$f(E) = E - e \sin E - M$$

which transforms the problem to finding the root of $f(E)$. Newton's method states that you quickly get it in quadratic convergence (that means very fast if you are close to the solution) by the iteration

$$E_{i+1} = E_i - \frac{f(E_i)}{f'(E_i)} = E_i - \frac{E_i - e \sin E_i - M}{1 - e \cos E_i} \tag{7.4.16}$$

with

$$0 \le M = n \cdot \left(t - t_p\right) \le \pi$$

and

$$E_0 = M + e^2\left[(6M)^{1/3} - M\right] \qquad @ \quad 0 \le M < 0.25$$
$$E_0 = M + \frac{e \sin M}{1 - \sin(M + e) + \sin M} \qquad @ \quad 0.25 \le M \le \pi$$

the empirically optimal initial E values for $0 \le e < 1$ (see Esmaelzadeh, Ghadiri (2014)). Observe that Newton's method becomes unstable if the denominator in Eq. (7.4.16) vanishes, i.e. for $e \cos M \to 1$ and therefore care has to be taken for $e > 0.95$ and $M < 0.1$.

We thus have finally reached our goal of determining the orbital position at any time:

Calculation scheme for orbit propagation

For an orbit with mean motion $n = \sqrt{\mu/a^3}$, eccentricity e, and a given time span $t - t_p$ from the last periapsis passage t_p

1. Calculate $M(t)$ from Eq. (7.3.11)
2. If $\pi < M \le 2\pi$ then reduce it to the interval $[0, \pi]$ by $2\pi - M \to M$
3. Calculate $E(t)$ from Newton's method Eq. (7.4.16)
4. Apply this to Eqs. (7.4.14a) and (7.4.14b)

whereby one gets $r = r(t)$ and $\theta = \theta(t)$.

Analytical Solution

Kepler's Eq. (7.4.15) provides a means to find analytical solutions to $r(t)$ and $\theta(t)$. Since from Eqs. (7.4.14a) and (7.4.14c) $\cos E = (r - a)/(ae)$ and $\sin E = (\sqrt{1 - e^2} \cdot \sin\theta)/(1 + e \cos\theta)$, we immediately get from Eq. (7.4.15) the analytical solution to the Keplerian problem

$$M = n \cdot \left(t - t_p\right) = \begin{cases} \arcsin\lambda - e\lambda & @ -1 \le \lambda \le 0 \\ \pi - \arcsin\lambda - e\lambda & @\ 0 \le \lambda \le 1 \end{cases}$$

with

$$\lambda = \sin E = \frac{\sqrt{1 - e^2}\sin\theta}{1 + e\cos\theta} = \operatorname{sgn}(\sin E)\sqrt{1 - \left(\frac{r - a}{ea}\right)^2}$$

or

$$M = n \cdot \left(t - t_p\right) = \frac{\pi}{2} + \arcsin\rho - e\sqrt{1 - \rho^2} \quad @ -1 \le \rho \le 1 \tag{7.4.17}$$

with

$$\rho = \cos E = \frac{\cos\theta + e}{1 + e\cos\theta} = \frac{r - a}{ea}$$

$-1 \le \rho \le 1$ is equivalent to $0 \le \theta \le \pi$ or $r_{per} \le r \le r_{apo}$. It should be mentioned that Eq. (7.4.17) can be derived directly by integration of Eq. (7.2.11) (see Problem 7.8).

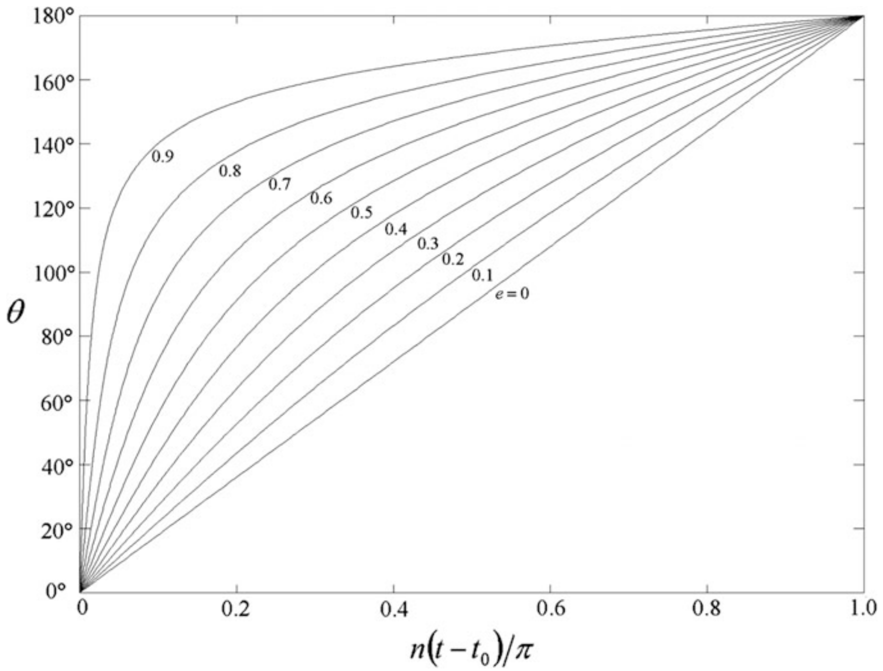

Fig. 7.14 A graphical solution to the Keplerian problem

We recall from Sect. 7.3.2 that this analytical solution is of no practical use. It suffers from the fact that the solution is only implicit. Though, like Kepler's equation, it could also be solved with the Newton method with regard to θ or r, it would be too much effort to calculate the function $f(E)$ and even more so its derivative $f'(E)$, even if done numerically. So when its comes to have a fast algorithm, e.g., an orbit propagator, that every millisecond calculates the exact orbit position of a spacecraft, Kepler and Newton are invincible. If a high accuracy of the result is not decisive, then a graphical depiction of Eq. (7.4.17) such as that given in Fig. 7.14 might be a favorable solution to the Keplerian problem. In addition, it provides a good overview.

Series Expansion of E(M)—Orbit Propagator

Because Newton's method Eq. (7.4.16) converges quadratically, we make use of this to determine a series approximation of $E(t)$ for e. It is easy to show (exercise) that for a more convenient iteration the Newton iteration Eq. (7.4.16) can be rewritten as

$$K_{i+1} = e\,\frac{\sin(M + K_i) - K_i \cos(M + K_i)}{1 - e\,\cos(M + K_i)}$$

where

$$K := E - M$$

Because this iteration equation is too complicated, we freeze in the slope of the Newton iteration at the starting value $E = M$ (Convince yourself by a depiction of Newton's method, that near the solution, the exact local slope is insignificant for the quadratic convergence.)

$$f' = 1 - e\,\cos E_i \approx 1 - e\,\cos M$$

This delivers the reduced Newton equation

$$E_{i+1} = E_i - \frac{f(E_i)}{f'(E_i)} = E_i - \frac{E_i - e\,\sin E_i - M}{1 - e\,\cos M}$$

i.e.

$$K_{i+1} = \frac{e}{1 - e\,\cos M}[\sin(M + K_i) - K_i \cos M]$$

or

$$K_{i+1} = s\,\cos K_i + c(\sin K_i - K_i)$$

with

$$s := \frac{e\,\sin M}{1 - e\,\cos M}, \quad c := \frac{e\,\cos M}{1 - e\,\cos M}$$

With this abridged iteration equation and the starting value $K_0 = 0$ we perform three iterations. This delivers

$$K_3 = s\left[1 - \frac{1}{2}s^2\left(1 + \frac{1}{3}c\right) + \frac{1}{24}s^4\left(13 + \frac{51}{5}c\right) + O\left(s^6\right)\right]$$

We therefore obtain with $M = n \cdot \left(t - t_p\right)$ the following result

$$E(t) = M + s\left[1 - \frac{1}{2}s^2\left(1 + \frac{1}{3}c\right) + \frac{1}{24}s^4\left(13 + \frac{51}{5}c\right) + O\left(s^6,\ s^4c^2\right)\right] \quad @\, e < 1 \tag{7.4.18}$$

The convergence of this series is so fast that we have the following small errors

$E(t) \approx M + s$	$\Delta E < 0.03° @\ e < 0.1$		
$E(t) \approx M + s - \frac{1}{2}s^3$	$\Delta E < 0.01° @\ e < 0.2;$	$\Delta E < 0.07°$	$@\ e < 0.3$
Eq.(7.4.18)	$\Delta E < 0.00062° @\ e < 0.2;$	$\Delta E < 0.013°$	$@\ e < 0.3$

We therefore have the following simple algorithm, a so-called orbit propagator, for elliptic orbits

Orbit Propagator $r(t),\ \theta(t)$, with $\Delta E < 0.00062°$ for $e < 0.2$
Given $a, e, t_0, r_0;\ t > t_0$

(a) Derive E_0 from $r_0 = a(1 - e\cos E_0)$
(b) Definitions

$$M := \sqrt{\frac{\mu}{a^3}}(t - t_0) + E_0 - e\ \sin E_0$$

$$s := \frac{e\ \sin M}{1 - e\ \cos M}, \quad c := \frac{e\ \cos M}{1 - e\ \cos M}$$

(c) Determine

$$E(t) = M + s\left[1 - \frac{1}{2}s^2\left(1 + \frac{1}{3}c\right) + \frac{1}{24}s^4\left(13 + \frac{51}{5}c\right)\right]$$

$$\rightarrow \quad r(t) = a(1 - e\cos E)$$

$$\rightarrow \quad \theta(t) = 2\arctan\left(\sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}\right)$$

Function Series Expansions

Other series expansions are provided in the literature (see e.g. Murray and Dermott (1999)). For a derivation see Problem 7.7) for small eccentricities, $e < 0.6627434$. They read

$$\begin{aligned} E &= M + e\sin M + \frac{1}{2}e^2\sin(2M) - \frac{1}{8}e^3[\sin M - 3\sin(3M)] + O\left(e^4\right) \\ \frac{r}{a} &= 1 - e\cos M + e^2\sin^2 M + \frac{3}{2}e^3\cos M\sin^2 M + O\left(e^4\right) \\ \theta &= M + 2e\sin M + \frac{5}{4}e^2\sin(2M) - \frac{1}{4}e^3\left[\sin M - \frac{13}{3}\sin(3M)\right] + O\left(e^4\right) \end{aligned} \tag{7.4.19}$$

Note that these series expansions exhibit only linear convergence and therefore Eq. (7.4.19a) converges much more slowly than Eq. (7.4.18).

Summary

In hindsight the achievement of Kepler's transformation tells us a remarkable story: Time t might be a good world-configuration state sequencer and hence a good state sequencer in physics overall, but it is not a good gravitationally bounded state sequencer. For two reasons: Firstly, it is a linear sequencer while bounded gravitational motion is periodic. Secondly, geometric elliptic motion is highly irregular in terms of time. The true anomaly θ remedies only the first problem, the Kepler transformation remedies both. With it the radial motion becomes remarkably simple. According to the results of Sect. 7.3.4 and Eq. (7.4.14a) the gravitational motion is basically determined by its mean energy state in the gravitational potential manifested in the semi-mayor axis a. The radial trajectory in terms of E then is simply a sinusoidal swing around this average radius a with amplitude ae, which equals the distance from the center of the ellipse to a focal point. This is why Eq. (7.4.14a) is so important and of practical relevance. The only hassle of this new view, namely the relation between E and t, is encapsuled in Kepler's equation (7.4.15). In Sect. 7.4.5 we will make use of this facilitated approach to find state vector solutions $\boldsymbol{r}(E)$, $\boldsymbol{v}(E)$ for any gravitational motion.

As we will see later, E is a good sequencer, but not a perfect one. With the universal anomaly $G = E\big/\sqrt{1-e^2}$ we will find in Sect. 7.4.6 the perfect sequencer with the additional benefit that it works equally well for elliptic, parabolic and hyperbolic trajectories.

7.4.3 Hyperbolic Orbit

The most general unbounded trajectory about a celestial body is a flyby hyperbola with $e > 1$. Note that for a hyperbola a is negatively defined, $a < 0$ (see Eq. (7.3.6)). Because of Eq. (7.3.19), the orbital energy is positive, $\varepsilon > 0$.

According to Eqs. (7.3.5) and (7.3.14) a hyperbola has its closest approach to the focal point at the periapsis with

$$r_{per} = \frac{h^2}{\mu(e+1)} = -a(e-1) \tag{7.4.20}$$

$$v_{per} = \frac{\mu}{h}(e+1) = \sqrt{-\frac{\mu}{a}}\sqrt{\frac{e+1}{e-1}} \tag{7.4.21}$$

Therefore its focal point lies at the distance $|a| + r_{per} = e|a| = -ea$ from the origin. A hyperbola does not possess an apoapsis. It reaches infinity at an asymptote, the angle of which is determined from Eq. (7.3.5) for $r \to \infty$ as

$$\cos\theta_\infty = -\frac{1}{e} \tag{7.4.22}$$

The so-called *impact parameter* Δ (a.k.a. *aiming radius*), which is the distance between the focal point and the asymptote measured normal to the asymptote, is found from Fig. 7.15 to be

$$\Delta = -\left(r_{per} + a\right)\sin\beta = -ae\ \sin\theta_\infty$$

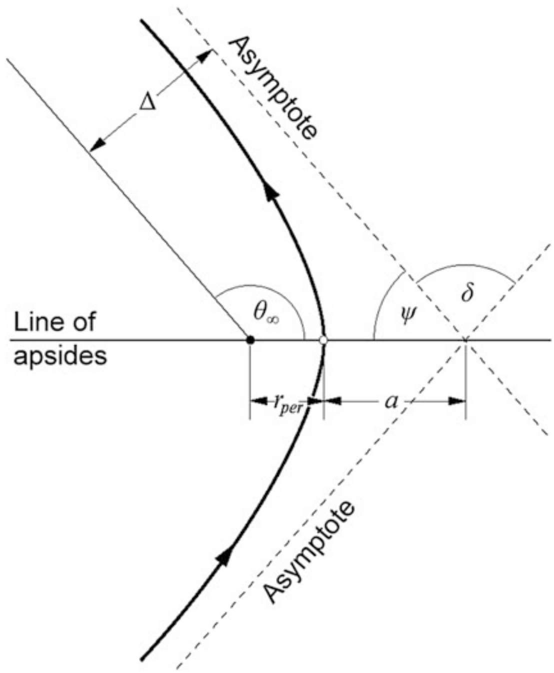

Fig. 7.15 The characteristic parameters of a hyperbola

Here we have chosen a negative sign because Δ should be a positive value, but a is negative. With Eqs. (7.4.20) and (7.4.22) one gets

$$\Delta = -a\sqrt{e^2 - 1} = r_{per}\sqrt{\frac{e+1}{e-1}} \qquad \textbf{impact parameter} \tag{7.4.23}$$

According to Eq. (7.3.14), the body at infinity has the so-called *hyperbolic excess velocity*

$$v_\infty = \frac{\mu}{h}\sqrt{e^2 - 1} = \sqrt{-\frac{\mu}{a}} \qquad \textbf{hyperbolic excess velocity} \tag{7.4.24}$$

For interplanetary flight, the parameter

$$C_3 := v_\infty^2 = -\frac{\mu}{a} \qquad \textbf{characteristic energy} \tag{7.4.25}$$

is often used. To attain a given hyperbolic excess velocity, for instance, for an interplanetary transfer, a speed of $v > v_{\triangleright\triangleright}$ is required at the departure orbit, where $v_{\triangleright\triangleright}$ is the escape velocity (see Sect. 7.4.2). It is determined from Eqs. (7.2.14) and (7.4.34) as

$$v^2 = \frac{2\mu}{r} + v_\infty^2 = v_{\triangleright\triangleright}^2 + v_\infty^2 \tag{7.4.26}$$

The related boost to reach this departure velocity is called *Oberth maneuver*. In fact, this equation is nothing other than the energy conservation equation, where only the kinetic energy shows up, because the potential energy vanishes at infinity.

Asymptotic Motion

When approaching $\theta \to \theta_\infty$ the body recedes asymptotically. In this limit it is easy to show from the orbit Eq. (7.3.5) that the following holds for $r(\theta)$

$$r(\theta) = \frac{p}{\sqrt{e^2-1}(\theta_\infty - \theta) + \frac{1}{2}(\theta_\infty - \theta)^2} \quad @\ \theta \to \theta_\infty \tag{7.4.27}$$

To determine $r(t)$ in the asymptotic limit, we recall the radial velocity as given in Eq. (7.2.11), which for $r \to \infty$ in first order perturbation calculation reads

$$\dot{r}^2 = -\frac{\mu}{a} \cdot \left(1 - \frac{2a}{r}\right) \quad @\ r \to \infty$$

With $\sqrt{1-2a/r} \approx 1 - a/r$ this yields the differential equation for asymptotic radial recession (+) or approach (−)

$$\begin{aligned} \frac{dr}{\sqrt{1-2a/r}} &\approx \frac{r \cdot dr}{r-a} = \left(1 - \frac{-a}{r-a}\right) dr = \left(1 - \frac{1}{1-r/a}\right) dr = \pm\sqrt{-\frac{\mu}{a}} \cdot dt \\ &= \pm v_\infty \cdot dt \end{aligned}$$

Solving with the initial condition $r(t=0) = r_0$ we derive

$$\left[r + a\ \ln\left(1 - \frac{r}{a}\right)\right]_{r_0}^{r} = \pm v_\infty (t - t_0)$$

hence

$$r + a\ \ln\left(1 - \frac{r}{a}\right) = \pm v_\infty (t - t_0) + r_0 + a\ \ln\left(1 - \frac{r_0}{a}\right) =: s$$

For $r \to \infty$ we have in 1st order approximation $r = s$, and in 2nd order approximation

$$\begin{aligned} r(t) &= s - a \cdot \ln\left(1 - \tfrac{s}{a}\right) \\ s &= \pm v_\infty (t - t_0) + r_0 + a \cdot \ln\left(1 - \tfrac{r_0}{a}\right) \end{aligned} \quad \begin{aligned} &+ @\ r \to \infty \\ &- @\infty \to r \end{aligned} \tag{7.4.28}$$

with initial condition $r(t_0) = r_0$. We recall that $a < 0$.

Given this result we are able to easily derive $\theta(t \to \infty)$. According to Eq. (7.2.7) we have

$$\int_{\theta}^{\theta_\infty} d\theta' = \int_t^\infty \frac{h}{r^2(t')} dt'$$

We insert the result Eq. (7.4.28), approximate for $t, s \to \infty$, and make use of

$$h\frac{dt}{ds} = \frac{h}{v_\infty} = \frac{p}{\sqrt{e^2-1}} = -a\sqrt{e^2-1}$$

Thus

$$\theta_\infty - \theta = \int_t^\infty \frac{h}{[s - a \cdot \ln(-s'/a)]^2} dt' \approx -a\sqrt{e^2-1} \int_s^\infty \frac{1 + 2a/s' \cdot \ln(-s'/a)}{s'^2} ds'$$

$$\int_s^\infty \frac{1 + 2a/s' \cdot \ln(-s'/a)}{s'^2} ds' = \int_s^\infty \left[\frac{1}{s'^2} - \frac{2}{a^2}\frac{\ln(-s'/a)}{(-s'/a)^3}\right] ds' = \frac{1}{s} + \frac{2}{a}\left(-\frac{1}{2}\frac{\ln x}{x^2}\right)\Bigg|_{-s/a}^\infty$$

$$= \frac{1}{s} + \frac{1}{a}\frac{\ln(-s/a)}{(-s/a)^2} + O\left(\frac{1}{s^3}\right)$$

Hence

$$\begin{aligned} \theta(t) &= \theta_\infty + \sqrt{e^2-1} \cdot \frac{a}{s}\left[1 + \frac{a}{s}\ln\left(-\frac{s}{a}\right) + O\left(\frac{1}{s^3}\right)\right] \\ s(t) &= v_\infty(t - t_0) + r_0 + a \cdot \ln\left(1 - \frac{r_0}{a}\right) \qquad @\ t \to \infty \\ \theta_\infty &= \arccos\left(-\frac{1}{e}\right) \end{aligned} \tag{7.4.28b}$$

again with initial condition $r(t_0) = r_0$ and $a < 0$.

Kepler's Equation and Solutions

One can show that the essential Eqs. (7.4.14) and (7.4.15) of the ellipse can be expressed in a similar way for a hyperbola, namely that there is also a hyperbolic eccentric anomaly F for the hyperbola. We can shorten its derivation by the statement that it has the following relation to the elliptic eccentric anomaly E: $E = iF$ with $i = \sqrt{-1}$ (for a proof see Sect. 7.4.6). If one applies this relation to Eqs. (7.4.14), one directly obtains

$$\boxed{r = a(1 - e\cosh F)} \qquad \textbf{orbit equation}$$

$$\begin{aligned} \cos\theta &= \frac{e - \cosh F}{e \cosh F - 1}, & \cosh F &= \frac{e + \cos\theta}{1 + e\cos\theta} \\ \sin\theta &= \frac{\sqrt{e^2 - 1}\cdot\sinh F}{e\cosh F - 1}, & \sinh F &= \frac{\sqrt{e^2-1}\cdot\sin\theta}{1 + e\cos\theta} \\ \tan\frac{\theta}{2} &= \sqrt{\frac{e+1}{e-1}}\tanh\frac{F}{2} \end{aligned} \tag{7.4.29}$$

and

$$M = n\cdot\left(t - t_p\right) = e\ \sinh F(t) - F(t) \tag{7.4.30}$$

with

$$n = \sqrt{-\frac{\mu}{a^3}} \tag{7.4.31}$$

Here, t_p the time of passage through the periapsis, and M the mean anomaly (cf. Eq. (7.3.11)).

To determine the orbit position as a function of time just as with the ellipse, one has to solve Eq. (7.4.30) with Newton's method

$$F_{i+1} = F_i - \frac{e\ \sinh F_i - F_i - M}{1 - e\ \cosh F_i} \tag{7.4.32}$$

with $F_0 = e\ \sinh\ M - M$ the initial value. If the result is inserted into Eqs. (7.4.29a) and (7.4.29b), one obtains the parameterized orbit $r = r(F(t))$ and $\theta = \theta(F(t))$.

Analytical Solution

In analogy to the elliptic case, we find from Eq. (7.4.30) with $\sinh F = \sqrt{e^2-1}\cdot\sin\theta\Big/(1 + e\cos\theta)$ from Eq. (7.4.29c) an implicit solution for r and θ

$$\begin{aligned} M = n\cdot\left(t - t_p\right) &= e\lambda - \operatorname{arsinh}\lambda \\ &= e\lambda - \ln\frac{\sqrt{e+1} + \sqrt{e-1}\cdot\tan(\theta/2)}{\sqrt{e+1} - \sqrt{e-1}\cdot\tan(\theta/2)} \end{aligned}$$

with

$$\lambda = \sinh F = \frac{\sqrt{e^2-1}\cdot\sin\theta}{1 + e\cos\theta} = \sqrt{\left(\frac{r-a}{ea}\right)^2 - 1}$$

or

$$M = n \cdot \left(t - t_p\right) = e\sqrt{\rho^2 - 1} - \operatorname{arcosh}\rho \\ = e\sqrt{\rho^2 - 1} - \ln\left(\rho + \sqrt{\rho^2 - 1}\right) \tag{7.4.33}$$

with

$$\rho = \cosh F = \frac{e + \cos\theta}{1 + e\cos\theta} = \frac{r - a}{-ea} = \frac{r + |a|}{e|a|}$$

Here again, this analytical solution is of no practical use, because it suffers from the fact that the solution is only implicit.

7.4.4 Parabolic Orbit

For $e = 1$, we get a parabolic orbit. According to Eq. (7.3.18), its orbital energy is $\varepsilon = 0$ and according to Eq. (7.3.7), its semi-major axis is $a = \infty$. However, because the semi-latus rectum $p = a(1 - e^2)$ is still a finite nonzero number, it is used as the sole orbit element to describe the shape of a parabolic orbit—p is the abeam semi-width of the parabola (see Fig. 7.8). Note that the semi-latus rectum, in general, is an excellent orbit element that steadily transforms between the transitions from circle to ellipse to parabola to hyperbola (see Fig. 7.8).

According to the vis-viva Eq. (7.2.15) and to the orbit equation the velocity on a parabolic orbit with $a = \infty$ is

$$v_{esc} = \sqrt{\frac{2\mu}{r}} = 2\sqrt{\frac{\mu}{p}}\cos\frac{\theta}{2} \tag{7.4.34}$$

The parabolic orbit is a limiting orbit where the body is able to just escape the central body reaching infinity with zero velocity, $v(r = \infty) = 0$. Equation (7.4.34), therefore, determines the so-called **escape velocity**, v_{esc}, which is the velocity a body at any position r from the barycenter requires to achieve parabolic escape.

Example

The velocity of a body at the surface of Earth needed to escape Earth's gravitation—the so-called *second cosmic velocity*—is with $\mu_\oplus = g_0 R_\oplus^2$ (see Eq. (7.1.20))

$$v_{\triangleright\triangleright} = \sqrt{2 g_0 R_\oplus} = 11.180\ \mathrm{km\,s^{-1}} \quad \textbf{second cosmic velocity} \tag{7.3.35}$$

With orbit equation (7.3.5) and Eq. (7.4.34) we have at periapsis, i.e. the closest approach to the barycenter

$$r_{per} = \frac{p}{2}$$
$$v_{per} = 2\sqrt{\frac{\mu}{p}}$$

Observe that the escape velocity at a given orbital radius enables a spacecraft to go to infinity independent of the direction in which the final velocity vector points. In Sect. 8.4.1 we will see how a transfer from any Keplerian orbit to a parabolic escape orbit is performed.

To determine the explicit orbit equations $r(t)$, $\theta(t)$ for a parabolic orbit, we recognize that for $e = 1$ and with the substitution $x := \theta/2$, one is able to directly integrate Eq. (7.3.9)

$$\begin{aligned}\sqrt{\frac{\mu}{p^3}}(t - t_p) &= \int_0^\theta \frac{d\theta'}{(1+\cos\theta')^2} = \int_0^\theta \frac{d\theta'}{(2\cos^2\theta'/2)^2} = \frac{1}{2}\int_0^{\theta/2} \frac{dx}{\cos^4 x} \\ &= \frac{1}{2}\int_0^{\theta/2}\left(\frac{1}{\cos^2 x} + \frac{\tan^2 x}{\cos^2 x}\right)dx = \frac{1}{2}\left(\tan x + \frac{1}{3}\tan^3 x\right)\Bigg|_0^{\frac{\theta}{2}} = \frac{1}{2}\left(\tan\frac{\theta}{2} + \frac{1}{3}\tan^3\frac{\theta}{2}\right)\end{aligned}$$

Thus, the position equation can be provided analytically

$$\begin{aligned}\sqrt{\frac{\mu}{p^3}}(t - t_p) &= \frac{1}{2}\tan\frac{\theta}{2} + \frac{1}{6}\tan^3\frac{\theta}{2} \qquad \textbf{Barker's equation} \\ &= \frac{1}{2}G + \frac{1}{3}G^3\end{aligned} \tag{7.4.36}$$

where we have defined the universal anomaly (cf. Sect. 7.4.6)

$$G := \tan\frac{\theta}{2} = \sqrt{2\frac{r}{p} - 1}$$

Considering Barker's equation as a cubic function for $\tan(\theta/2)$ we find the roots by applying Cardano's method and Descartes' rule of signs. From this, it can be shown (exercise, Problem 7.5) that there is only one real solution to this equation for θ and r, namely,

$$\begin{aligned}\tan\frac{\theta}{2} &= \left(\sqrt{q^2+1}+q\right)^{1/3} - \left(\sqrt{q^2+1}-q\right)^{1/3} = 2\sinh\left(\frac{1}{3}\operatorname{arsinh} q\right) \\ 2\frac{r}{p} &= \left(\sqrt{q^2+1}+q\right)^{2/3} + \left(\sqrt{q^2+1}-q\right)^{2/3} - 1 = 2\cosh\left(\frac{2}{3}\operatorname{arsinh}|q|\right) - 1\end{aligned} \tag{7.4.37}$$

with

$$q = 3\sqrt{\frac{\mu}{p^3}}(t - t_p)$$

where the later expressions follow from Chebyshev polynomials, which are particularly useful for small time intervals $t - t_p$. So, also for the parabolic case, we were able to solve the Keplerian problem analytically.

Asymptotic Motion

What is the mathematical description of the asymptotic motion, i.e. for $t \to \infty$? With the definition

$$\lambda^3(t) := \frac{1}{2q} = \frac{1}{6\sqrt{\mu/p^3}(t - t_p)} \to 0$$

we find for Eq. (7.4.37) the power series expansions

$$\begin{aligned} \tan\tfrac{\theta}{2} &= \tfrac{1}{\lambda} - \lambda + \tfrac{1}{3}\lambda^5 + O(\lambda^7) = (2q)^{1/3} - (2q)^{-1/3} + \cdots \\ 2\tfrac{r}{p} &= \tfrac{1}{\lambda} - 1 + \lambda + \tfrac{2}{3}\lambda^5 + O(\lambda^7) = (2q)^{1/3} - 1 + (2q)^{-1/3} + \cdots \end{aligned} \quad @\ \lambda(t) \to 0 \tag{7.4.38}$$

Because for $\varphi \to \pi/2$ we have the power series expansion

$$\tan\left(\frac{\pi}{2} - \varphi\right) = \frac{1}{\varphi} - \frac{1}{3}\varphi - \frac{1}{45}\varphi^3 - \frac{2}{945}\varphi^5 \quad @\ \varphi \to 0$$

we can remove the divergence of $\tan\theta(\lambda)$ for $\theta \to \pi$ by finding that φ, which satisfies the above $\tan\theta/2$ expansion in λ. We do so by making an ansatz of a power series approximation $\varphi = a\lambda + b\lambda^3 + c\lambda^5 + d\lambda^7$, calculating $1/\varphi - \frac{1}{3}\varphi - \frac{1}{45}\varphi^3 - \frac{2}{945}\varphi^5$ in terms of λ, and equate it to the $\tan\theta/2(\lambda)$ power series expansion. We thus get the result

$$\theta(t) = \pi - 2\varphi = \pi - 2\left(\lambda + \frac{2}{3}\lambda^3 - \frac{1}{45}\lambda^5 - \frac{76}{189}\lambda^7 + \ldots\right) \quad @\ \lambda(t) \to 0 \tag{7.4.39}$$

Near-Parabolic Orbits

The parabolic case is exceptional in that it is not transcendental like the elliptic and hyperbolic cases, where we achieved the $r(t), \theta(t)$ solutions only numerically by solving Kepler's equation. We would like to convey the parabolic non-transcendental nature to both elliptic and hyperbolic orbits close to the parabolic orbit, i.e. for $e \approx 1$. We therefore define a closeness parameter $\varepsilon = 1 - e$. We therewith expand the integral in Eq. (7.3.9) as

$$\int_0^\theta \frac{d\theta'}{[1+(1-\varepsilon)\cos\theta']^2} \approx \int_0^\theta \frac{d\theta'}{(1+\cos\theta')^2} + 2\varepsilon \int_0^\theta \frac{\cos\theta'}{(1+\cos\theta')^3} d\theta' + 3\varepsilon^2 \int_0^\theta \frac{\cos^2\theta'}{(1+\cos\theta')^4} d\theta'$$

The solutions to the integrals are (Exercise. Use a symbolic integrator and express results in terms of $\cos^2(\theta/2)$, finally transform into $1/\cos^2(\theta/2) = 1 + \tan^2(\theta/2)$)

$$\int_0^\theta \frac{1}{(1+\cos\theta')^2} d\theta' = \frac{1}{2}\tan\frac{\theta}{2}\left(1+\frac{1}{3}\tan^2\frac{\theta}{2}\right)$$

$$\int_0^\theta \frac{\cos\theta'}{(1+\cos\theta')^3} d\theta' = \frac{1}{4}\tan\frac{\theta}{2}\left(1-\frac{1}{5}\tan^4\frac{\theta}{2}\right)$$

$$\int_0^\theta \frac{\cos^2\theta'}{(1+\cos\theta')^4} d\theta' = \frac{1}{8}\tan\frac{\theta}{2}\left(1-\frac{1}{3}\tan^2\frac{\theta}{2}-\frac{1}{5}\tan^4\frac{\theta}{2}+\frac{1}{7}\tan^6\frac{\theta}{2}\right)$$

With the above definition $G := \tan\theta/2$ Kepler's equation therefore reads

$$\begin{aligned}\frac{q}{3} &:= \sqrt{\frac{\mu}{p^3}}(t-t_p)\\ &\approx \frac{1}{2}G\left(1+\frac{1}{3}G^2\right) + \frac{2\varepsilon}{4}G\left(1-\frac{1}{5}G^4\right) + \frac{3\varepsilon^2}{8}G\left(1-\frac{1}{3}G^2-\frac{1}{5}G^4+\frac{1}{7}G^6\right)\end{aligned} \tag{7.4.40}$$

Because ε is small, we solve the equation for G by one approximation iteration. In the first step we assume $\varepsilon = 0$. The solution is that of the parabolic orbit Eq. (7.4.38). We insert this approximate result into the residual. By defining

$$\frac{s}{3} := \frac{q}{3} - (1-e)\frac{1}{2}G\left(1-\frac{1}{5}G^4\right) - (1-e)^2\frac{3}{8}G\left(1-\frac{1}{3}G^2-\frac{1}{5}G^4+\frac{1}{7}G^6\right)$$

where now $G(q(t))$, we obtain the new Kepler's equation

$$\frac{s(t)}{3} = \frac{1}{2}G\left(1+\frac{1}{3}G^2\right)$$

The solution is again given by Eq. (7.4.38). We therefore have the following solution for near-parabolic orbits

Given $t > t_p$. Determine

$$G = 2\,\sinh\left\{\frac{1}{3}\operatorname{arsinh}\left[3\sqrt{\frac{\mu}{p^3}}\left(t - t_p\right)\right]\right\}$$

$$s := \sqrt{\frac{\mu}{p^3}}(t - t_0) - (1 - e)\frac{3}{2}G\left(1 - \frac{1}{5}G^4\right)$$
$$- (1 - e)^2\frac{9}{8}G\left(1 - \frac{1}{3}G^2 - \frac{1}{5}G^4 + \frac{1}{7}G^6\right) + O\left((1 - e)^3\right)$$

Then

$$\tan\frac{\theta}{2} = 2\sinh\left(\frac{1}{3}\operatorname{arsinh} s\right)$$
$$\frac{r}{p} = \cosh\left(\frac{2}{3}\operatorname{arsinh}|s|\right) - \frac{1}{2} \qquad (7.4.41)$$

7.4.5 ε-Based Transformation

Having identified Kepler's equations as the root of the transcendentality of Kepler's problem, we are now seeking to eliminate transcendentality from Newton's gravitational equation of motion (7.1.19) and thus find for a given initial state $\boldsymbol{r}(t_0)$, $\boldsymbol{v}(t_0)$ the solution $\boldsymbol{r}(t)$, $\boldsymbol{v}(t)$. If $t > t_0$ this problem is called state (vector) propagation. Let us reconsider Kepler's equations. They read

$$n\left(t - t_p\right) = E - e\ \sin E \qquad @\ \text{elliptic orbits}$$
$$n\left(t - t_p\right) = e\ \sinh F - F \qquad @\ \text{hyperbolic orbits}$$

By differentiating we obtain with Eq. (7.4.14a) and Eq. (7.4.29a)

$$n\frac{dt}{dE} = n\frac{dt}{dF} = \frac{r}{|a|}$$

From this follows

$$dE = dF = \sqrt{\frac{\mu}{|a|}}\frac{1}{r}\cdot dt = \sqrt{2|\varepsilon|}\frac{1}{r}\cdot dt \qquad \textbf{ε-based transformation} \qquad (7.4.42)$$

The latter expression, which follows from Eq. (7.3.19), is indicative of the fact that the transformation is only suitable as long as the specific orbital energy is <u>not</u> $\varepsilon \approx 0$. This is why we call it energy-based (ε-based) transformation. Comparison with sub-section *Regularization* in Sect. 7.4.2 and with Sect. 7.4.3 reveals that the ε-based transformation is a generalized Kepler transformation, which regularizes the orbit equation. The expression $dE = \sqrt{\mu/a}/r \cdot dt$ is the (regular) Kepler transformation, which applies to elliptic orbits only.

That the transformation is just bound to the requirement of a finite orbital energy implies two things. On one hand, and unfortunately, it implies that it is not applicable for parabolic or near-parabolic orbits having $\varepsilon \approx 0$. But, according to Eq. (7.3.18), $2\varepsilon \cdot h^2 = \mu^2(e^2 - 1)$. So on the other hand, as long as $\varepsilon \neq 0$, it does apply for orbits exhibiting a vanshing angular momentum, i.e. $h \approx 0$, in case of which $e = 1$. These are the so-called radial trajectories. We will make use of Kepler's equation with $e = 1$ for radial trajectories in Sects. 7.5.1 and 7.5.2. That the Kepler transformation applies for radial motion is a surprising result, because then $\theta = const$, in case of which the native geometrical concepts of mean, eccentric, and hyperbolic anomaly (i.e., orbit angles) are meaningless.

Algebraic Form of Newton's Gravitational Equation of Motion

To obtain an regularized algebraic form of Newton's gravitational equation of motion (7.1.19) we rewrite it as

$$\frac{d^2\boldsymbol{r}}{dt^2} + \frac{\mu}{r^3}\boldsymbol{r} = 0$$

We eliminate transcendentality by applying the ε-based transformation to

$$\frac{d^2\boldsymbol{r}}{dt^2} = \frac{d}{dt}\frac{d\boldsymbol{r}}{dt} = \frac{\mu}{|a|}\frac{1}{r}\frac{d}{dE}\left(\frac{1}{r}\frac{d\boldsymbol{r}}{dE}\right) = \frac{\mu}{|a|}\frac{1}{r}\left(\frac{1}{r}\boldsymbol{r}'' - \frac{r'}{r^2}\boldsymbol{r}'\right)$$

where

$$(\cdot)' := \frac{d}{dE}$$

Thus transformed, Newton's gravitational EoM reads

$$\boldsymbol{r}'' - \frac{r'}{r}\boldsymbol{r}' + |a|\hat{\boldsymbol{r}} = 0$$

This equation is still not solvable analytically. But the unit radial vector $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ hints at the eccentricity vector, which reads from Eq. (7.3.3) and with $\boldsymbol{r} \cdot \boldsymbol{v} = \boldsymbol{r} \cdot \boldsymbol{v}_r = r\dot{r}$

$$\mu\boldsymbol{e} = \mu\left(\frac{1}{r} - \frac{1}{a}\right)\boldsymbol{r} - (\boldsymbol{r}\boldsymbol{v})\boldsymbol{v} = \mu\hat{\boldsymbol{r}} - \frac{\mu}{a}\boldsymbol{r} - r\dot{r} \cdot \dot{\boldsymbol{r}}$$

We apply Kepler's transformation to this expression and get

$$\frac{r'}{r} \cdot \boldsymbol{r}' = |a|\hat{\boldsymbol{r}} - \mathrm{sgn}(a)(\boldsymbol{r} + a\boldsymbol{e})$$

This result inserted into the above EoM delivers

$$\boldsymbol{r}''(E, F) = -\text{sgn}(a) \cdot (\boldsymbol{r} + a\boldsymbol{e}) \quad \text{Newton's gravitational EoM} \tag{7.4.43}$$

where sgn(x) is the sign function. This is an algebraic form of Newton's gravitational equation of motion, which is beautifully simple and analytically solvable. To solve it we make a final substitution

$$\boldsymbol{u} := \boldsymbol{r} + a\boldsymbol{e} \rightarrow \boldsymbol{u}'' = \boldsymbol{r}''$$

which delivers

$$\boldsymbol{u}'' = -\text{sgn}(a) \cdot \boldsymbol{u}$$

Solution for Elliptic Orbits

For elliptic orbits with $a > 0$ the EoM delivers the harmonic oscillator $\boldsymbol{u}'' = -\boldsymbol{u}$ having the general solution

$$\boldsymbol{u}(E) = \boldsymbol{r}(E) + a\boldsymbol{e} = \boldsymbol{\alpha} \sin E + \boldsymbol{\beta} \cos E$$
$$\boldsymbol{u}'(E) = \boldsymbol{r}'(E) = \sqrt{\frac{a}{\mu}} r \cdot \boldsymbol{v}(E) = \boldsymbol{\alpha} \cos E - \boldsymbol{\beta} \sin E$$

subject to the initial condition

$$\boldsymbol{r}_0 + a\boldsymbol{e} = \boldsymbol{\alpha} \sin E_0 + \boldsymbol{\beta} \cos E_0$$
$$\sqrt{\frac{a}{\mu}} r_0 \boldsymbol{v}_0 = \boldsymbol{\alpha} \cos E_0 - \boldsymbol{\beta} \sin E_0$$

Solving for the unknow constants $\boldsymbol{\alpha}, \boldsymbol{\beta}$ delivers

$$\boldsymbol{\alpha} = (\boldsymbol{r}_0 + a\boldsymbol{e}) \sin E_0 + \cos E_0 \sqrt{\frac{a}{\mu}} r_0 \boldsymbol{v}_0$$
$$\boldsymbol{\beta} = (\boldsymbol{r}_0 + a\boldsymbol{e}) \cos E_0 - \sin E_0 \sqrt{\frac{a}{\mu}} r_0 \boldsymbol{v}_0$$

From this follows for an ellipse

$$\boldsymbol{r}(E) + a\boldsymbol{e} = (\boldsymbol{r}_0 + a\boldsymbol{e}) \cos(E - E_0) + \sqrt{\frac{a}{\mu}} r_0 \boldsymbol{v}_0 \sin(E - E_0) \tag{7.4.44}$$

Differentiation delivers

$$\boldsymbol{r}' = -(\boldsymbol{r}_0 + a\boldsymbol{e}) \sin(E - E_0) + \sqrt{\frac{a}{\mu}} r_0 \boldsymbol{v}_0 \cos(E - E_0)$$

Applying Eq. (7.4.42) yields

$$\boldsymbol{v}\cdot\sqrt{\frac{a}{\mu}}\cdot r = -(\boldsymbol{r}_0 + a\boldsymbol{e})\sin(E-E_0) + \sqrt{\frac{a}{\mu}} r_0\boldsymbol{v}_0\cos(E-E_0)$$

hence

$$r\cdot\boldsymbol{v}(E) = -\sqrt{\frac{\mu}{a}}(\boldsymbol{r}_0 + a\boldsymbol{e})\sin(E-E_0) + r_0\boldsymbol{v}_0\cos(E-E_0) \tag{7.4.45}$$

Solution for Hyperbolic Orbits

For hyperbolic orbits with $a<0$ the substitution delivers $\boldsymbol{u}'' = \boldsymbol{u}$ with the general solution

$$\boldsymbol{u}(F) = \boldsymbol{r}(F) + a\boldsymbol{e} = \boldsymbol{\alpha} e^F + \boldsymbol{\beta} e^{-F}$$

$$\boldsymbol{u}'(F) = \boldsymbol{r}'(F) = \sqrt{\frac{|a|}{\mu}} r\cdot\boldsymbol{v}(F) = \boldsymbol{\alpha} e^F - \boldsymbol{\beta} e^{-F}$$

By the same token as above for elliptic orbits, it is easy to show that for a hyperbola

$$\boldsymbol{r}(F) + a\boldsymbol{e} = (\boldsymbol{r}_0 + a\boldsymbol{e})\cosh(F-F_0) + \sqrt{\frac{|a|}{\mu}} r_0\boldsymbol{v}_0\sinh(F-F_0) \tag{7.4.46}$$

$$r\cdot\boldsymbol{v}(F) = \sqrt{\frac{\mu}{|a|}}(\boldsymbol{r}_0 + a\boldsymbol{e})\sinh(F-F_0) + r_0\boldsymbol{v}_0\cosh(F-F_0) \tag{7.4.47}$$

RSW Coordinate System $\boldsymbol{e},\ \boldsymbol{h}\times\boldsymbol{e}$

For highly elliptic, parbolic, or hyperbolic orbits and at large distances from periapsis the vectors $\boldsymbol{r}_0, \boldsymbol{v}_0$ exhibit the property $\angle(\boldsymbol{r}_0, \boldsymbol{v}_0) \to 0$ and hence do not establish a good basis for the propagated state vector $\boldsymbol{r}, \boldsymbol{v}$. A better choice then is an orthogonal RSW coordinate system $\boldsymbol{e},\ \boldsymbol{h}\times\boldsymbol{e}, \boldsymbol{h}$ that is not derived from $\boldsymbol{r}_0, \boldsymbol{v}_0$ and hence well established. Thus we get the corresponding results by projecting $\boldsymbol{r}, \boldsymbol{v}$ on $\boldsymbol{e},\ \boldsymbol{h}\times\boldsymbol{e}$:

For elliptic orbits

$$\begin{aligned}
\boldsymbol{r} + a\boldsymbol{e} &= \left[(\boldsymbol{r}_0\boldsymbol{e} + a)\cdot\cos(E-E_0) + \boldsymbol{v}_0\boldsymbol{e}\cdot r_0\sqrt{a/\mu}\cdot\sin(E-E_0)\right]\cdot\boldsymbol{e} \\
&\quad + \left[\boldsymbol{r}_0(\boldsymbol{h}\times\boldsymbol{e})\cdot\cos(E-E_0) + \boldsymbol{v}_0(\boldsymbol{h}\times\boldsymbol{e})\cdot r_0\sqrt{a/\mu}\cdot\sin(E-E_0)\right]\cdot\boldsymbol{h}\times\boldsymbol{e} \\
&= \cos\theta\cdot\boldsymbol{e} + \sin\theta\cdot\boldsymbol{h}\times\boldsymbol{e} \\
&= \frac{\cos E - e}{1 - e\cos E}\cdot\boldsymbol{e} + \frac{\sqrt{1-e^2}\sin E}{1 - e\cos E}\cdot\boldsymbol{h}\times\boldsymbol{e}
\end{aligned} \tag{7.4.48}$$

and

$$\begin{aligned} r \cdot \boldsymbol{v} &= \left[-(\boldsymbol{r}_0\boldsymbol{e}+a)\sqrt{\mu/a}\cdot\ \sin(E-E_0)+\boldsymbol{v}_0\boldsymbol{e}\cdot r_0\cos(E-E_0)\right]\cdot\boldsymbol{e} \\ &\quad+\left[-\boldsymbol{r}_0(\boldsymbol{h}\times\boldsymbol{e})\sqrt{\mu/a}\cdot\ \sin(E-E_0)+\boldsymbol{v}_0(\boldsymbol{h}\times\boldsymbol{e})\cdot r_0\cos(E-E_0)\right]\cdot\boldsymbol{h}\times\boldsymbol{e} \\ \boldsymbol{v} &= \frac{h}{r^2}(-\sin\theta\cdot\boldsymbol{e}+\cos\theta\cdot\boldsymbol{h}\times\boldsymbol{e}) \\ &= \frac{h}{r^2}\left(-\frac{\sqrt{1-e^2}\sin E}{1-e\cos E}\cdot\boldsymbol{e}+\frac{\cos E-e}{1-e\cos E}\cdot\boldsymbol{h}\times\boldsymbol{e}\right) \end{aligned} \tag{7.4.49}$$

and for hyperbolic orbits

$$\begin{aligned} \boldsymbol{r}+a\boldsymbol{e} &= \left[(\boldsymbol{r}_0\boldsymbol{e}+a)\cosh(F-F_0)+\boldsymbol{v}_0\boldsymbol{e}\cdot r_0\sqrt{|a|/\mu}\cdot\ \sinh(F-F_0)\right]\cdot\boldsymbol{e} \\ &\quad+\left[\boldsymbol{r}_0(\boldsymbol{h}\times\boldsymbol{e})\cdot\ \cosh(F-F_0)+\boldsymbol{v}_0(\boldsymbol{h}\times\boldsymbol{e})\cdot r_0\sqrt{|a|/\mu}\cdot\ \sinh(F-F_0)\right]\cdot\boldsymbol{h}\times\boldsymbol{e} \\ &= \cos\theta\cdot\boldsymbol{e}+\sin\theta\cdot\boldsymbol{h}\times\boldsymbol{e} \\ &= \frac{e-\cosh F}{e\cosh F-1}\cdot\boldsymbol{e}+\frac{\sqrt{e^2-1}\cdot\ \sinh F}{e\cosh F-1}\cdot\boldsymbol{h}\times\boldsymbol{e} \end{aligned} \tag{7.4.50}$$

and

$$\begin{aligned} r\cdot\boldsymbol{v} &= \left[(\boldsymbol{r}_0\boldsymbol{e}+a)\sqrt{\mu/|a|}\cdot\ \sinh(F-F_0)+\boldsymbol{v}_0\boldsymbol{e}\cdot r_0\cosh(F-F_0)\right]\cdot\boldsymbol{e} \\ &\quad+\left[\boldsymbol{r}_0(\boldsymbol{h}\times\boldsymbol{e})\sqrt{\mu/|a|}\cdot\ \sinh(F-F_0)+\boldsymbol{v}_0(\boldsymbol{h}\times\boldsymbol{e})\cdot r_0\cosh(F-F_0)\right]\cdot\boldsymbol{h}\times\boldsymbol{e} \\ \boldsymbol{v} &= \frac{h}{r^2}(-\sin\theta\cdot\boldsymbol{e}+\cos\theta\cdot\boldsymbol{h}\times\boldsymbol{e}) \\ &= \frac{h}{r^2}\left(-\frac{\sqrt{e^2-1}\cdot\ \sinh F}{e\cosh F-1}\cdot\boldsymbol{e}+\frac{e-\cosh F}{e\cosh F-1}\cdot\boldsymbol{h}\times\boldsymbol{e}\right) \end{aligned} \tag{7.4.51}$$

7.4.6 h-Based Transformation

The energy-based transformation suffers from the fact that it does not include the parabolic orbit and becomes numerically unstable for near-parabolic orbits, i.e. for $\varepsilon \approx 0$. We therefore seek a transformation that stably covers these cases.

Guided by the Eq. (7.3.18), $2\varepsilon \cdot h^2 = \mu^2(e^2 - 1)$, which claims that $\varepsilon \approx 0 \leftrightarrow e \approx 1$ is sensible if $h = \sqrt{\mu p} \neq 0$, we define a new transformation with variable G as

$$dG := \sqrt{\frac{\mu}{p}}\frac{1}{r}\cdot dt = \frac{\mu}{h}\frac{1}{r}\cdot dt \quad \textbf{\textit{h}-based transformation} \tag{7.4.52}$$

Because it requires $h \neq 0$ and in line with Eq. (7.4.42) we call it *angular-momentum-based (h-based) transformation*. Comparison with Eq. (7.4.42) reveals that $dE = dF = dG\sqrt{|1-e^2|}$. Therefore we obtain for elliptic und hyperbolic orbits the identities

$$\boxed{G = \frac{E}{\sqrt{1-e^2}} = \frac{F}{\sqrt{e^2-1}}} \quad \textbf{universal anomaly} \tag{7.4.53}$$

Observe that G is an angle, by the same token as E, F are, and therefore it formally is an anomaly (see Remark at the end of Sect. 7.3.1). This is why we call it *universal anomaly*. However, otherwise it lacks a geometrical interpretation—at least I do not know of one, to date. Observe that from this result, it follows

$$E = iF$$

which we made use of by a handwaving argument in Sect. 7.4.3 to derive Kepler's equation for hyperbolic orbits from elliptic orbits. The above h-based transformation retains the regularization property of the ε-based transformation and in addition passes smoothly between elliptic, parabolic, and hyperbolic orbits because it remains numerically stable for $|a| \to \infty$.

From Eq. (7.4.43) we find with $dE = dF = dG\sqrt{|1-e^2|}$ the corresponding Newton's gravitational equation of motion

$$\boldsymbol{r}''(G) = -\left(1-e^2\right)(\boldsymbol{r} + a\boldsymbol{e}) \quad \text{Newton's gravitational EoM} \tag{7.4.54}$$

This equation is also a beautifully simple algebraic form of Newton's gravitational equation of motion, which yields the same results as those in Sect. 7.4.5 for

$$\begin{aligned} E &\to G\sqrt{1-e^2} \quad &@ \text{ elliptic orbits} \\ F &\to G\sqrt{e^2-1} \quad &@ \text{ hyperbolic orbits} \end{aligned}$$

In order to derive a Kepler equation $t(G)$ applying for any $0 \leq e < \infty$, we integrate the above h-based transformation equation and find with $G(t_p) = 0$, where t_p is the epoch, i.e. the time at periapsis,

$$\sqrt{\frac{\mu}{p}}\left(t - t_p\right) = \int_0^G r(G') \cdot dG' \quad @ \; G\left(t_p\right) = 0 \tag{7.4.55}$$

To solve this equation we need to find $r(G)$. The road to $r(G)$ is to start with Leibniz's radial differential equation $\ddot{r}(t)$, transform it into $r''(G)$, and solve it to obtain $r(G)$. According to Eq. (7.2.10) Leibniz's equation reads

$$\ddot{r} = -\frac{\mu}{r^2} + \frac{\mu p}{r^3}$$

Applying the h-based transformation yields

$$\ddot{r} = -\frac{\mu}{p}\frac{r''}{r^2} - \frac{\dot{r}^2}{r}$$

From Eq. (7.2.11) we have

$$\dot{r}^2 = \frac{2\mu}{r} - \frac{h^2}{r^2} - \frac{\mu}{a}$$

By inserting these two equations into the above Leibniz's equation we find its G-form to be

$$r''(G) = -\left(1-e^2\right)r + p = -\left(1-e^2\right)(r-a) \quad \text{Leibniz's equation} \tag{7.4.56}$$

We solve this Leibniz's equation by substitution $u := r - a$, which delivers $u'' = r''$ and hence

$$u'' = -\left(1-e^2\right)u$$

The general solution of this equation satisfying the condition $r(G=0) = a(1-e)$ and $r'(G=0) = 0$ at the periapsis reads

$$r - a = -ae\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}G^{2n}\left(1-e^2\right)^n = -ae\cos\left(G\sqrt{1-e^2}\right) \tag{7.4.57}$$

From this follows

$$\frac{r}{p} = \frac{1}{1-e^2}\left[1 - e\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}G^{2n}\left(1-e^2\right)^n\right]$$

or

$$\boxed{\begin{aligned}\frac{r}{p} &= \frac{1}{1+e} + e\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+2)!}G^{2k+2}\left(1-e^2\right)^k\\ &= \frac{1}{1+e} + eG^2c_2\left(G^2\left(1-e^2\right)\right)\end{aligned}} \quad @\ 0 \le e < \infty \tag{7.4.58}$$

With this we derive the universal Kepler equation by carrying out the integral in Eq. (7.4.55)

$$\boxed{\begin{aligned}\sqrt{\frac{\mu}{p^3}}\left(t-t_p\right) &= \frac{G}{1+e} + e\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+3)!}G^{2k+3}\left(1-e^2\right)^k\\ &= \frac{G}{1+e} + eG^3c_3\left(G^2\left(1-e^2\right)\right)\end{aligned}} \quad \begin{aligned}&\textbf{universal Kepler equation}\\ &@\ 0 \le e < \infty,\ G\left(t_p\right) = 0\end{aligned}$$

$$\tag{7.4.59}$$

Here $c_2(x)$, $c_3(x)$ are the so-called *Stumpff functions* of second and third order (see box in Sect. 7.4.8). Observe that in the above two enframed equations the singularity at $e = 1$ has been removed and therefore they apply steadily to all Keplerian orbits $0 \le e < \infty$.

We finally provide the expression for $\theta(G)$. From Eq. (7.3.5) we have

$$e \cdot \cos\theta(G) = \frac{p}{r} - 1 = \frac{\cos G\sqrt{1-e^2} - e}{1 - e\cos G\sqrt{1-e^2}} \quad @\ 0 \le e < \infty \tag{7.4.60}$$

or alternatively with Eq. (7.4.14c)

$$\begin{aligned}\tan\frac{\theta(G)}{2} &= \sqrt{\tfrac{1+e}{1-e}}\tan\left(\tfrac{1}{2}G\sqrt{1-e^2}\right)\\ &= \frac{1+e}{2}G\sum_{n=0}^{\infty}(-1)^n\cdot\frac{4\left(2^{2n+2}-1\right)}{(2n+2)!}B_{2n+2}G^{2n}\left(1-e^2\right)^n \quad @\ 0 \le e < \infty\\ &= \frac{1+e}{2}G\left[1 + \tfrac{1}{12}G^2\left(1-e^2\right) + \tfrac{1}{120}G^4\left(1-e^2\right)^2 + \ldots\right]\end{aligned} \tag{7.4.61}$$

where

$$\begin{aligned} B_{2n} &= 0, \frac{1}{6}, -\frac{1}{30}, \frac{1}{42}, -\frac{1}{30}, \frac{5}{66} \qquad n = 0,1,2,3,4,5\\ B_{2n} &= (-1)^{n+1}\frac{2(2n)!}{(2\pi)^{2n}}\left(1 + \frac{1}{2^{2n}} + \frac{1}{3^{2n}}\right) \quad @\ n \ge 5,\ \frac{\Delta B_{2n}}{B_{2n}} < 10^{-6}\end{aligned}$$

are the Bernoulli numbers. Observe that in the above power series expansion the singularity at $e = 1$ is removed, as well.

Elliptic and Hyperbolic Orbits

From Eq. (7.4.57) follows

$$\begin{aligned} r(G) &= a\left[1 - e\cos\left(G\sqrt{1-e^2}\right)\right] \quad @\ e < 1\\ r(G) &= a\left[1 - e\cosh\left(G\sqrt{e^2-1}\right)\right] \quad @\ e > 1\end{aligned}$$

and from the universal Kepler Eq. (7.4.59)

$$\begin{aligned} &\sqrt{\tfrac{\mu}{p^3}}\left(t - t_p\right) = \tfrac{1}{1-e^2}\left[G - e\tfrac{1}{\sqrt{1-e^2}}\sin\left(G\sqrt{1-e^2}\right)\right]\\ \to\ &\sqrt{\tfrac{\mu}{a^3}}\left(t - t_p\right) = G\sqrt{1-e^2} - e\sin\left(G\sqrt{1-e^2}\right) \quad @\ e < 1\end{aligned}$$

and

$$\begin{aligned} &\sqrt{\tfrac{\mu}{p^3}}\left(t - t_p\right) = \tfrac{1}{e^2-1}\left[e\tfrac{1}{\sqrt{e^2-1}}\sinh\left(G\sqrt{e^2-1}\right) - G\right]\\ \to\ &\sqrt{\tfrac{\mu}{a^3}}\left(t - t_p\right) = e\sinh\left(G\sqrt{e^2-1}\right) - G\sqrt{e^2-1} \quad @\ e > 1\end{aligned}$$

Comparison of this result with Eqs. (7.4.14a), (7.4.29a) and Eqs. (7.4.29a), (7.4.30) corroborates the transformation rule $G\sqrt{1-e^2} \rightarrow E$, $G\sqrt{e^2-1} \rightarrow F$, and $G \rightarrow \theta$ for $e = 0$.

The general solution procedure to find for a given $t > t_p$ the state vector $\boldsymbol{r}(t)$,$v(t)$ for elliptic and hyperbolic orbits is therefore:

Solve Eq. (7.4.59) by Newton's method.

$\rightarrow E(t) = G(t)\sqrt{1-e^2}$ @ **elliptic orbits**

$r(t) = a(1 - e\cos E) \rightarrow \cos\theta(t) = \frac{1}{e}\left(\frac{p}{r} - 1\right)$

and state vector $\boldsymbol{r}(t)$,$v(t)$ from Eq. (7.4.44), Eq. (7.4.45)

or

$\rightarrow F(t) = G(t)\sqrt{e^2-1}$ @ **hyperbolic orbits**

$r(t) = a(1 - e\cosh F) \rightarrow \cos\theta(t) = \frac{1}{e}\left(\frac{p}{r} - 1\right)$

and state vector $\boldsymbol{r}(t)$,$\boldsymbol{v}(t)$ from Eqs. (7.4.46), (7.4.47).

Parabolic Orbits

For parabolic orbits with $e = 1$ we derive from Eq. (7.4.59)

$$\frac{q}{3} := \sqrt{\frac{\mu}{p^3}}(t - t_p) = \frac{1}{2}G + \frac{1}{6}G^3 \quad @\ G(t_p) = 0 \tag{7.4.62}$$

From Eq. (7.4.37) follows the solution

$$G(t) = \left(\sqrt{q^2+1} + q\right)^{1/3} - \left(\sqrt{q^2+1} - q\right)^{1/3} = 2\sinh\left(\frac{1}{3}\operatorname{arsinh} q\right)$$

Equation (7.4.58) yields

$$r(t) = \frac{1}{2}p\left(1 + G^2\right) \tag{7.4.63}$$

hence

$$\cos\theta(t) = \frac{p}{r} - 1 = \frac{1 - G^2}{1 + G^2}$$

or

$$\boxed{G = \tan\frac{\theta}{2}} \tag{7.4.64}$$

This agrees with the result derived from Eq. (7.4.61) for $e = 1$. The state vector then is determined by

$$\begin{aligned} \boldsymbol{r}(t) &= \cos\theta \cdot \boldsymbol{e} + \sin\theta(\boldsymbol{h}\times\boldsymbol{e}) \\ \boldsymbol{v}(t) &= \frac{h}{r^2}[-\sin\theta\cdot\boldsymbol{e} + \cos\theta(\boldsymbol{h}\times\boldsymbol{e})] \end{aligned} \tag{7.4.65}$$

Note that with the identity $G \equiv \tan\theta/2$ we recover all the results already derived in Sect. 7.4.4 for the parabolic orbit.

Near-Parabolic Orbits

For near-parabolic orbits with $e \approx 1$ Eq. (7.4.59) delivers

$$\sqrt{\frac{\mu}{p^3}}(t - t_p) = \frac{G}{1+e} + \frac{e}{6}G^3 - (1-e^2)\frac{e}{120}G^5 \tag{7.4.66}$$

$\rightarrow$ $G(t)$ by Newton's method

From Eq. (7.4.58) we find

$$r(t) = \frac{1}{2}p(G^2+1) - \frac{1}{4}(1-e^2)p\left(\frac{1}{6}G^2+1\right)G^2 \tag{7.4.67}$$

$$\rightarrow \qquad \cos\theta(t) = \frac{1}{e}\left(\frac{p}{r} - 1\right)$$

and state vector as given by Eq. (7.4.65).

For $e = 1$ these results obviously pass smoothly into the parabolic case above. For later purposes we solve Eq. (7.4.67) iteratively for G and find

$$\frac{1}{2}p(G^2+1) = r + (1-e^2)\frac{p}{24}\left(\frac{2r}{p}+5\right)\left(\frac{2r}{p}-1\right)$$

From this follows

$$G = \sqrt{\left(\frac{2r}{p}-1\right)\left[1 + \frac{1-e^2}{12}\left(\frac{2r}{p}+5\right)\right]} \tag{7.4.68}$$

Universal Optimal Solver

Based on the above results we have the following universal solver for state vector propagation, which is optimal in the sense of regularization.

Let $\boldsymbol{r}_0, \boldsymbol{v}_0$ be the state vector given at time t_0.

If the type of orbit and the metric elements of the orbit are not known, determine them by

$$\begin{array}{ll} \boldsymbol{h} = \boldsymbol{r}_0 \times \boldsymbol{v}_0 & \to p = \boldsymbol{h}^2/\mu \\ \boldsymbol{e} = \frac{1}{\mu}\boldsymbol{v}_0 \times \boldsymbol{h} - \frac{1}{r_0}\boldsymbol{r}_0 & \to e = \sqrt{\boldsymbol{ee}} \\ \dfrac{1}{a} = \dfrac{2}{r_0} - \dfrac{\boldsymbol{v}_0^2}{\mu} & \end{array}$$

Derive G_0 from the following relevant equation

$$\begin{array}{ll} e\cos\left(G_0\sqrt{1-e^2}\right) = 1 - r_0/a & @\ e<1,\ \text{elliptic orbit} \\ e\cosh\left(G_0\sqrt{e^2-1}\right) = 1 + r_0/|a| & @\ e>1,\ \text{hyperbolic orbit} \\ G_0 = \sqrt{\left(\frac{2r_0}{p}-1\right)\left[1+\frac{1-e^2}{12}\left(\frac{2r_0}{p}+5\right)\right]} & @\ e\approx 1,\ \text{near-parabolic orbit} \end{array}$$

Let $\quad R(G) = \frac{G}{1+e} + e\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+3)!}G^{2k+3}(1-e^2)^k$

$\to \quad R'(G) = \frac{1}{1+e} + e\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+2)!}G^{2k+2}(1-e^2)^k = \frac{r}{p}$

Then G_0, t_0 solves

$$\sqrt{\frac{\mu}{p^3}}\left(t_0 - t_p\right) = R(G_0)$$

With this we eliminate the unknown t_p and derive the customized transcendental equation

$$\sqrt{\frac{\mu}{p^3}}(t - t_0) + R(G_0) = R(G)$$

Therefore any instance of the transcendental equation above is solved for G by the Newton iteration

$$G_{i+1} = G_i - \frac{R(G_i) - R(G_0) - \sqrt{\mu/p^3}(t-t_0)}{R'(G_i)}, \quad i = 1, \ldots$$

Optimal initial values for elliptic orbits with $M = \sqrt{\mu/a^3}(t-t_0) + (1-e^2)^{3/2}R(G_0)$ are

$$\begin{array}{ll} G_1 = \dfrac{M + e^2\left[(6M)^{1/3} - M\right]}{\sqrt{1-e^2}} & @\ e<1,\ 0\le M<0.25 \\ G_1 = \dfrac{1}{\sqrt{1-e^2}}\left[M + \dfrac{e\sin M}{1-\sin(M+e)+\sin M}\right] & @\ e<1,\ 0.25\le M\le\pi \end{array}$$

and for parabolic or near-parabolic orbits

$$G_1 = \sqrt{\left(\frac{2r_0}{p} - 1\right)\left[1 + \frac{1-e^2}{12}\left(\frac{2r_0}{p} + 5\right)\right]} \quad @\ e \approx 1$$

The denominator of the iteration fraction, $R'(G)$, delivers directly the universal result

$$r(t) = p \cdot R'(G) \to e\ \cos\theta(t) = \frac{p}{r} - 1$$

and

$$\begin{aligned} E &= G\sqrt{1-e^2} \quad &@\ \text{elliptic orbits} \\ F &= G\sqrt{e^2-1} \quad &@\ \text{hyperbolic orbits} \end{aligned}$$

This inserted into the relevant solutions $\boldsymbol{r}(E)$, $\boldsymbol{v}(E)$ or $\boldsymbol{r}(F)$, $\boldsymbol{v}(F)$ in Sect. 7.4.5, or Eq. (7.4.65) for near-parabolic orbits, respectively, delivers the propagated state vector.

7.4.7 Conventional State Vector Propagation

A more classical approach for state vector propagation is encapsulated in the following algorithm

Algorithm for Universal State Vector Propagation $\boldsymbol{r}(t_0), \boldsymbol{v}(t_0) \to \boldsymbol{r}(t), \boldsymbol{v}(t)$

1. Let $\boldsymbol{r}_0, \boldsymbol{v}_0$ be given at time t_0.
2. If the type of orbit and the metric elements of the orbit are not known, determine them by

$$\begin{aligned} \boldsymbol{h} &= \boldsymbol{r}_0 \times \boldsymbol{v}_0 & &\to & p &= \boldsymbol{h}^2/\mu \\ \boldsymbol{e} &= \frac{1}{\mu}\boldsymbol{v}_0 \times \boldsymbol{h} - \frac{1}{r_0}\boldsymbol{r}_0 & &\to & e &= \sqrt{\boldsymbol{ee}} \\ \frac{1}{a} &= \frac{2}{r_0} - \frac{\boldsymbol{v}_0^2}{\mu} \end{aligned}$$

3. If $e < 1$, then <u>elliptic orbit</u>

   a. Determine E_0 from $e\cos E_0 = 1 - r_0/a$
   b. For $t > t_0$ solve $E(t) - e\ \sin E(t) = \sqrt{\mu/a^3} \cdot (t - t_0) + E_0 - \sin E_0$ for $E(t)$, for instance with Newton's method (see Eq. (7.4.16)).
   c. If $e \ll 1$, determine $\boldsymbol{r}(t)$ from Eq. (7.4.44) and $\boldsymbol{v}(t)$ with $r = \sqrt{\boldsymbol{rr}}$ from Eq. (7.4.45).
   If $e \gg 0$ and $\angle(\boldsymbol{r}_0, \boldsymbol{v}_0) \to 0$ determine $\boldsymbol{r}(t)$ from Eq. (7.4.48) and $\boldsymbol{v}(t)$ with $r = \sqrt{\boldsymbol{rr}}$ from Eq. (7.4.49).

4. If $e > 1$, then hyperbolic orbit
   a. Determine F_0 from $e \cosh F_0 = 1 + r_0/|a|$
   b. For $t > t_0$ solve $e \sinh F(t) - F(t) = \sqrt{\mu/|a|^3} \cdot (t - t_0) + e \sinh F_0 - F_0$ for $F(t)$, for instance with Newton's method (see Eq. (7.4.32)).
   c. Determine $\boldsymbol{r}(t)$ from Eq. (7.4.46) and $\boldsymbol{v}(t)$ with $r = \sqrt{\boldsymbol{rr}}$ from Eq. (7.4.47). If $\angle(\boldsymbol{r}_0, \boldsymbol{v}_0) \to 0$ determine $\boldsymbol{r}(t)$ from Eq. (7.4.50) and $\boldsymbol{v}(t)$ with $r = \sqrt{\boldsymbol{rr}}$ from Eq. (7.4.51).

5. If $e \approx 1$, then parabolic orbit or near-parabolic orbits
   a. Initialization:

$$G_0 = \sqrt{\left(\frac{2r_0}{p} - 1\right)\left[1 + \frac{1-e^2}{12}\left(\frac{2r_0}{p} + 5\right)\right]}$$

$$R(G) := \frac{G}{1+e} + \frac{e}{6}G^3 - \left(1 - e^2\right)\frac{e}{120}G^5$$

   b. For $t > t_0$ use initial value

$$G_1 = 2\sinh\left[\frac{1}{3}\operatorname{arsinh}\left(3\sqrt{\frac{\mu}{p^3}}(t - t_0) + 3R(G_0)\right)\right]$$

   and determine $G(t)$ from the Newton iteration

$$G_{i+1} = G_i - \frac{R(G_i) - R(G_0) - \sqrt{\mu/p^3}(t - t_0)}{R'(G_i)}, \quad i = 1, \ldots$$

   c. Determine

$$r(t) = \frac{1}{2}p\left(G^2 + 1\right) - \frac{1}{4}(1 - e^2)p\left(\frac{1}{6}G^2 + 1\right)G^2$$

$$\cos\theta(t) = \frac{1}{e}\left(\frac{p}{r} - 1\right), \sin\theta(t) = \sqrt{1 - \frac{1}{e^2}\left(\frac{p}{r} - 1\right)^2}$$

   d. Final state vector:

$$\boldsymbol{r}(t) = \cos\theta \cdot \boldsymbol{e} + \sin\theta \cdot (\boldsymbol{h} \times \boldsymbol{e})$$

$$\boldsymbol{v}(t) = \frac{h}{r^2}[-\sin\theta \cdot \boldsymbol{e} + \cos\theta \cdot (\boldsymbol{h} \times \boldsymbol{e})]$$

7.4.8 Universal Variable Formulation

There exist other propagation methods: for instance, the *Lagrange–Gibbs f and g solution method* (see Schaub and Junkins (2003) or Chobotov (2002)), but most importantly the so-called *universal variable formulation (a.k.a. universal approach)*. In the following we give only the basic approach without providing the full algorithm (cf. Danby 2003).

The starting point of the universal variable formulation is the Sundman transformation

$$ds = \frac{1}{r} \cdot dt \quad \textbf{Sundman transformation} \tag{7.4.69}$$

Here s is the so-called *universal variable* of dimension [time]/[length] and therefore it is formally not an anomaly, that is not an angle (see Remark at the end of Sect. 7.3.1). Obviously, the Sundman transformation is essentially the same as the h-based transformation Eq. (7.4.52). Due to its different normalization it delivers a slightly different transformed gravitational equation of motion

$$\boldsymbol{r}'' = -\alpha^2 \boldsymbol{r} - \mu \boldsymbol{e}$$

with

$$\alpha^2 := \frac{\mu}{a} = \frac{2\mu}{r_0} - \boldsymbol{v}_0^2$$

Rather than turning it into an harmonic oscillation EoM by the transformation $\boldsymbol{u} := \boldsymbol{r} + a\boldsymbol{e}$ as done in Sect. 7.4.5, it is differentiated to deliver

$$\boldsymbol{r}''' = -\alpha^2 \boldsymbol{r}' \tag{7.4.70}$$

It can be shown (see Danby 2003, Sect. 6.9) that with initial conditions $\boldsymbol{r}_0, \boldsymbol{v}_0, t_0$ the solution to this vector differential equation reads

$$\begin{aligned} \boldsymbol{r}(s) &= f \cdot \boldsymbol{r}_0 + g \cdot \boldsymbol{v}_0 \\ \boldsymbol{v}(s) &= \dot{f} \cdot \boldsymbol{r}_0 + \dot{g} \cdot \boldsymbol{v}_0 \end{aligned} \tag{7.4.71}$$

with

$$\begin{aligned} f &= 1 - (\mu/r_0) s^2 c_2\left(\alpha^2 s^2\right) \\ \dot{f} &= -[\mu/(r r_0)] s c_1\left(\alpha^2 s^2\right) \\ g &= t - t_0 - \mu s^3 c_3\left(\alpha^2 s^2\right) \\ \dot{g} &= 1 - (\mu/r) s^2 c_2\left(\alpha^2 s^2\right) \end{aligned} \tag{7.4.72}$$

Here the functions $c_n(\alpha^2 s^2)$ are the so-called Stumpff functions as expounded in the box.

Stumpff Functions

The Stumpff functions $c_n(x)$ are defined as

$$c_n(x) := \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{(2k+n)!}, \quad n \geq 0$$

with properties

$$\begin{aligned}
c_n(0) &= \frac{1}{n!} \\
c_n(x) &= \frac{1}{n!} - x \cdot c_{n+2}(x) \\
c_0(4x) &= 2c_0^2(x) = 2[1 - xc_2(x)]^2 \\
c_1(4x) &= c_0(x)c_1(x) = [1 - xc_2(x)][1 - xc_3(x)] \\
c_2(4x) &= \tfrac{1}{2}c_2^2(x) \\
c_3(4x) &= \tfrac{1}{4}c_2(x) + \tfrac{1}{4}c_0(x)c_3(x) = \tfrac{1}{4}c_2(x) + \tfrac{1}{4}[1 - xc_2(x)]c_3(x)
\end{aligned}$$

The latter four equations are useful for two things: First, for reducing an angle x by factors of 4 to a value smaller than a given amount (less than half an orbit). Second, only the power series for c_2 and c_3 need to be used. Explicitely, the first four Stumpff functions read

$$c_0(x) := \begin{cases} \cos\sqrt{x} & x > 0 \\ \cosh\sqrt{-x} & x < 0 \\ 1 & x = 0 \end{cases}$$

$$c_1(x) := \begin{cases} \sin\sqrt{x}/\sqrt{x} & x > 0 \\ \sinh\sqrt{-x}/\sqrt{-x} & x < 0 \\ 1 & x = 0 \end{cases}$$

$$c_2(x) := \begin{cases} (1 - \cos\sqrt{x})/x & x < 0 \\ (\cosh\sqrt{-x} - 1)/(-x) & x < 0 \\ 1/2 & x = 0 \end{cases}$$

$$c_3(x) := \begin{cases} (\sqrt{x} - \sin\sqrt{x})\big/x^{3/2} & x > 0 \\ (\sinh\sqrt{-x} - \sqrt{-x})\Big/(-x)^{3/2} & x < 0 \\ 1/6 & x = 0 \end{cases}$$

Since $x = \alpha^2 s^2$, the upper of the three expressions holds for elliptic orbits with $\alpha^2 > 0$, the middle for hyperbolic orbits with $\alpha^2 < 0$, and the lower for parabolic orbits with $\alpha = 0$.

Given the Stumpff functions we find

$$\alpha \cdot sc_1 = \begin{cases} \sin(\alpha s) \\ \sinh(\alpha s) \\ \alpha s \end{cases}$$

$$\alpha^2 s^2 c_2 = \begin{cases} 1 - \cos(\alpha s) \\ 1 - \cosh(\alpha s) \\ \alpha^2 s^2 / 2 \end{cases}$$

$$\alpha^3 s^3 c_3 = \begin{cases} \alpha s - \sin(\alpha s) \\ \alpha s - \sinh(\alpha s) \\ \alpha^3 s^3 / 6 \end{cases}$$

where again the upper of the three expressions holds for elliptic orbits, the middle for hyperbolic orbits, and the lower for parabolic orbits.

The initial conditions read $\boldsymbol{r}_0, \boldsymbol{v}_0, t_0$. So, to apply Eq. (7.4.72) we still have to determine $s(t)$ for a given time $t > t_0$. For this we have to rewrite Kepler's equation. It is easily verified that any of the above mentioned basic solution functions $f(\alpha^2 s^2)$ exhibit the key property

$$f'''(\alpha^2 s^2) = -\alpha^2 \cdot f'(\alpha^2 s^2)$$

A general solution of Eq. (7.4.70) therefore is

$$\boldsymbol{r}(s) = \boldsymbol{r}_0 + \boldsymbol{r}'_0 s c_1(\alpha^2 s^2) + \boldsymbol{r}''_0 s^2 c_2(\alpha^2 s^2)$$

With this Kepler's equation can be derived by integration of Eq. (7.4.69) as

$$t - t_0 = \int_0^s r(\bar{s}) \cdot d\bar{s} = r_0 s c_1(\alpha^2 s^2) + r_0 \dot{r}_0 s^2 c_2(\alpha^2 s^2) + \mu s^3 c_3(\alpha^2 s^2) \qquad (7.4.73)$$

where the constants $r_0, \dot{r}_0, \alpha$ need to be determined from the initial conditions $\boldsymbol{r}_0, \boldsymbol{v}_0, t_0$. This form of Kepler's equation is equivalent to Eq. (7.4.15c), as can easily be verified, and is therefore convenient, because it does not make use of the time of passage through periapsis in the original Kepler's Eq. (7.4.15), which usually is not known, but of the known state vector at a given time t_0. The root of this equation delivers the variable $s(t)$, which has the same property as the eccentric anomaly for the Kepler transformation.

This universal variable formulation is widely discussed in the literature and used in practice. Battin (1964, 1987), Chobotov (2002, Sect. 4.3), and Danby (2003, Sect. 6.9) provide a good overview, and Chobotov (2002) presents in his Sect. 4.5 a practicable algorithm for this propagator; there are many examples and a MATLAB code for this propagator in Curtis (2005, Sect. 3.7).

7.5 Radial Trajectories

For a vanishing angular momentum, $h = 0$, it is inadmissible to take the cross product of $\boldsymbol{h}$ with Eq. (7.1.19) as done in Eq. (7.3.1). Therefore the general orbit Eq. (7.3.5) is not a valid solution in this case. According to Eq. (7.2.7) $h = 0$ happens whenever the transverse velocity vanishes, $v_\theta = 0$ (of course, as $h = r \cdot v_\theta = const$, v_θ is either always zero or never zero)—for instance, if one positions a body at an arbitrary distance from the central body with an initial velocity $v = 0$. To derive the equation of motion we resort to the equations of radial motion (7.2.11) where we set $h = 0$. This results in

$$\begin{aligned} \dot{r}^2 &= \mu\left(\frac{2}{r} - \frac{1}{a}\right) \\ \dot{\theta} &= 0 \end{aligned} \tag{7.5.1}$$

For the lateral motion we hence find the solution

$$\theta = const$$

That is, the body moves on a straight line toward the central body (or away from it) until it crashes into the central body. This is a radial trajectory, a.k.a. *rectilinear orbit*, however, descripively they do not orbit the central mass. Observe that if we convey the definition Eq. (7.3.6) $p := h^2/\mu =: a(1 - e^2)$ to radial trajectories, we may consider them as degenerate elliptic, parabolic or hyperbolic orbits with $e = 1,\ a \neq 0$. Accordingly, we will classify them in the following as radial elliptic, radial parabolic, and radial hyperbolic trajectories. Owing to this degeneration, e and p are no longer discriminatory orbital elements, but a is. We therefore will adapt a as a discriminative orbital parameter.

The equation of radial motion for radial trajectories is just the square root of the above radial differential equation

$$\frac{dr}{dt} = \pm\sqrt{\mu}\sqrt{\frac{2}{r} - \frac{1}{a}} \quad \textbf{equation of radial motion} \tag{7.5.2}$$

The parameter a is the only characteristic orbital element of the radial trajectory shape and signifies its total specific orbital energy ε according to the general Eq. (7.3.19)

$$\varepsilon = -\frac{\mu}{2a}$$

The two signs in Eq. (7.5.2) indicate two different modes of motion of the body: At the plus sign the body moves outbound (increasing dr with increasing dt) while at negative sign it moves inbound. The sign of course depends on the initial conditions r_0, v_0, t_0. If the body initially moves outward, $\dot{r}_0 = v_0 > 0$, then $\dot{r} > 0$ also for $t > 0$ until $v = 0$, and vice versa.

Also the orbital element a of the radial trajectory is determined by the initial conditions. We find from Eq. (7.5.1)

$$\frac{v_0^2}{\mu} = \frac{2}{r_0} - \frac{1}{a}$$

and hence

$$a = \frac{\mu r_0}{2\mu - r_0 v_0^2} \tag{7.5.3}$$

We see from this equation that a may be finite positive or negative, or infinite, depending on the initial conditions. This gives rise to three different types of radial trajectories characterized by a

1. Radial elliptic trajectory: $a > 0, \quad 0 < r \leq 2a$
2. Radial parabolic trajectory: $a = \infty, \quad 0 < r < \infty$
3. Radial hyperbolic trajectory: $a < 0, \quad 0 < r < \infty$

For each type of trajectory, the body may move inward or outward, so we have to discern in total six different cases depending on the initial velocity v_0, which we are now going to investigate.

7.5.1 Radial Elliptic Trajectory

We first assume $a > 0$ and $0 < r \leq 2a$.

> **Note** The radial ellipse is an ellipse with $e = 1$. For this we have from Eqs. (7.4.6) *and* (7.4.7) $r_{per} = 0$ *and* $r_{apo} = 2a$: *The focal point coincides with the periapsis at* $r = 0$ *and the empty focus with the apoapsis at* $r = 2a$. *This is the reason for the seemingly odd condition* $r \leq 2a$.

Inward Motion: $v_0 < 0$
From Eq. (7.5.2) the equation of radial motion is

$$\frac{dr}{dt} = -\sqrt{\mu}\sqrt{\frac{2}{r} - \frac{1}{a}} \tag{7.5.4}$$

Separating the variables yields

$$-\frac{dr/a}{\sqrt{2a/r - 1}} = \sqrt{\frac{\mu}{a^3}} \cdot dt$$

Substituting $x = r/a \rightarrow dx = dr/a$ and integrating leads to

$$(t - t_0)\sqrt{\frac{\mu}{a^3}} = -\int_{r_0/a}^{r/a} \frac{dx}{\sqrt{2/x - 1}} = \left[-x\sqrt{\frac{2}{x} - 1} + 2\arcsin\sqrt{\frac{x}{2}}\right]_{r/a}^{r_0/a}$$

With

$$\frac{r_0}{a}\sqrt{\frac{2a}{r_0} - 1} = \frac{r_0 v_0}{\sqrt{\mu a}}$$

we finally get

$$(t - t_0)\sqrt{\frac{\mu}{a^3}} = \frac{r}{a}\sqrt{\frac{2a}{r} - 1} - \frac{r_0 v_0}{\sqrt{\mu a}} - 2\arcsin\sqrt{\frac{r}{2a}} + 2\arcsin\sqrt{\frac{r_0}{2a}} \tag{7.5.5}$$

This is the trajectory equation of the body moving inward. Because this is an implicit solution, $t = t(r)$, it needs to be solved for r, for instance with Newton's method to find $r = r(t)$ explicitly.

It can be shown (see Problem 7.9) that for $a \rightarrow \infty$, $v_0 = 0$, and $t - t_0 \ll \sqrt{\mu/r_0^3}$, the explicit solution to this equation is a near-radial parabolic orbit (see Sect. 7.5.3) with orbit equation

$$r^{3/2}(t) = r_0^{3/2} - \sqrt{\frac{9\mu}{2}} \cdot (t - t_0)\left[1 - \frac{r_0}{4a}\right] \quad @\ a \rightarrow \infty,\ v_0 = 0,\ t - t_0 \ll \sqrt{\mu/r_0^3}$$

Outward Motion: $v_0 > 0$
From Eq. (7.5.2) we get

$$\frac{dr}{dt} = \sqrt{\mu}\sqrt{\frac{2}{r} - \frac{1}{a}} \tag{7.5.6}$$

With just the opposite sign, we derive from the above

$$(t - t_0)\sqrt{\frac{\mu}{a^3}} = -\frac{r}{a}\sqrt{\frac{2a}{r} - 1} + \frac{r_0 v_0}{\sqrt{\mu a}} + 2\arcsin\sqrt{\frac{r}{2a}} - 2\arcsin\sqrt{\frac{r_0}{2a}} \tag{7.5.7}$$

This is the implicit trajectory equation of the body moving outward. Again it needs to be solved by Newton's method for $r = r(t)$.

Numerical Determination
As with all Keplerian equations, Eq. (7.5.7) suffers from the fact that it is implicit for $r(t)$ and this in quite an intricate way. So, for numerical solutions we try to

resort to simpler solutions. In 7.4.5 we have seen that the ε-based transformation is applicable for $h \to 0$. We therefore can determine the trajectory and its velocity as a function of time by applying the corresponding Kepler's Eq. (7.4.15b) with $e = 1$ (see also Herrick (1971)).

$$\begin{array}{ll} M = \sqrt{\frac{\mu}{a^3}}(t - t_0), \quad E_0 = \arccos\left(1 - \frac{r_0}{a}\right) & \\ E - \sin E = M + E_0 - \sin E_0 \quad \rightarrow \quad E(t) & @\ r_0 v_0^2 < 2\mu \\ r = a(1 - \cos E), \quad v = \frac{\sqrt{\mu a}}{r}\sin E & \end{array} \tag{7.5.8}$$

7.5.2 *Radial Hyperbolic Trajectory*

We now assume $a < 0$ and $0 < r < \infty$.

Inward Motion: $v_0 < 0$
For $a < 0$ and from Eq. (7.5.2) we get

$$\frac{dr}{dt} = -\sqrt{\mu}\sqrt{\frac{2}{r} + \frac{1}{|a|}} \tag{7.5.9}$$

By the same token as in Sect. 7.5.1, we obtain

$$(t - t_0)\sqrt{\frac{\mu}{|a|^3}} = -\int_{r_0/|a|}^{r/|a|} \frac{dx}{\sqrt{2/x + 1}} = \left[x\sqrt{\frac{2}{x} + 1} - 2\ \text{arcsinh}\sqrt{\frac{x}{2}}\right]_{r/|a|}^{r_0/|a|}$$

and finally for $a < 0$

$$(t - t_0)\sqrt{-\frac{\mu}{a^3}} = \frac{r}{a}\sqrt{1 - \frac{2a}{r}} + 2\ \text{arcsinh}\sqrt{-\frac{r}{2a}} + \frac{r_0 v_0}{\sqrt{-\mu a}} - 2\ \text{arcsinh}\sqrt{-\frac{r_0}{2a}} \tag{7.5.10}$$

A Newton iteration yields $r = r(t)$.

Outward Motion: $v_0 > 0$
From $a < 0$ and

$$\frac{dr}{dt} = \sqrt{\mu}\sqrt{\frac{2}{r} + \frac{1}{|a|}} \tag{7.5.11}$$

we find in analog to the above

$$(t-t_0)\sqrt{-\frac{\mu}{a^3}} = -\frac{r}{a}\sqrt{1-\frac{2a}{r}} - 2\,\mathrm{arcsinh}\sqrt{-\frac{r}{2a}} - \frac{r_0 v_0}{\sqrt{-\mu a}} + 2\,\mathrm{arcsinh}\sqrt{-\frac{r_0}{2a}} \tag{7.5.12}$$

A Newton iteration yields $r = r(t)$.

Numerical Determination

By the same token as for radial elliptic trajectories, radial hyperbolic trajectories are numerically more easily determined from a customized Eq. (7.4.30) by the procedure

$$\begin{aligned} &M = \sqrt{-\tfrac{\mu}{a^3}}\,(t-t_0), F_0 = \mathrm{arcosh}\left(1-\tfrac{r_0}{a}\right) \\ &\sinh F - F = M + \sinh F_0 - F_0 \quad \rightarrow \quad F(t) \\ &r = a(1-\cosh F), v = \tfrac{\sqrt{-\mu a}}{r}\sinh F \end{aligned} \qquad @\ r_0 v_0^2 > 2\mu \tag{7.5.13}$$

7.5.3 Radial Parabolic Trajectory

Comets that arrive from the border of our solar system (from the so-called Oort cloud) exhibit $e \approx 1$, $a \approx \infty$. This is a parabolic orbit. In addition, if they exhibit $h \approx 0$, we have in this special case $a = \infty$.

Inward Motion: $v_0 < 0$

If $a = \infty$, Eq. (7.5.2) reads for inward motion

$$\frac{dr}{dt} = -\sqrt{\frac{2\mu}{r}} \tag{7.5.14}$$

Separating the variables and integration leads to

$$(t-t_0)\sqrt{2\mu} = -\int_{r_0}^{r} \sqrt{r} = \frac{2}{3}\left(r_0^{3/2} - r^{3/2}\right)$$

where t counts the time from a given initial position $r_0 = r(t_0)$ (for instance, the point of first comet sighting) to a given later trajectory position r. Hence, we obtain

$$r^{3/2}(t) = r_0^{3/2} - \sqrt{\frac{9\mu}{2}}\cdot(t-t_0) \tag{7.5.15}$$

Outward Motion: $v_0 > 0$
The equation of radial motion in this case is

$$\frac{dr}{dt} = \sqrt{\frac{2\mu}{r}} \tag{7.5.16}$$

and by the same token as above we get

$$r^{3/2}(t) = r_0^{3/2} + \sqrt{\frac{9\mu}{2}} \cdot (t - t_0) \tag{7.5.17}$$

Both radial parabolic solutions are explicit in t.

Note that the radial parabolic trajectory is the only solution to Newton's gravitational EoM, which is not regularizable by an ε-based or h-based transformation (see Sect. 7.4.5 and 7.4.6) because in this case $\varepsilon = 0$ and $h = 0$ simultaneously. Fortunatelly, this case does not need regularization because it is not transcendental and therefore exhibits an explicit solution $r(t)$.

7.5.4 Free Fall

Inbound trajectories bear the special situation where the body is placed at a certain distance from the origin and after free fall collides with the surface of the center body or its origin after time t_{col}. To determine t_{col}, let R be the radius of the center body and $t_0 = 0$. Then we get from Eqs. (7.5.5), (7.5.10), and (7.5.15)

$$t_{col} = \begin{cases} \sqrt{\frac{a^3}{\mu}}\left(\frac{R}{a}\sqrt{\frac{2a}{R}-1} - 2\arcsin\sqrt{\frac{R}{2a}} - \frac{r_0 v_0}{\sqrt{\mu a}} + 2\arcsin\sqrt{\frac{r_0}{2a}}\right) & @\ 0 < a < \infty \\ \sqrt{\frac{2}{9\mu}}\left(r_0^{3/2} - R^{3/2}\right) & @\ a = \infty \\ \sqrt{-\frac{a^3}{\mu}}\left(\frac{R}{a}\sqrt{1-\frac{2a}{R}} + 2\,\mathrm{arcsinh}\sqrt{-\frac{R}{2a}} + \frac{r_0 v_0}{\sqrt{-\mu a}} - 2\,\mathrm{arcsinh}\sqrt{-\frac{r_0}{2a}}\right) & @\ a < 0 \end{cases} \tag{7.5.18}$$

For the time to collision with the origin we set $R = 0$ and get

$$t_{col} = \begin{cases} \sqrt{\frac{a^3}{\mu}}\left(-\frac{r_0 v_0}{\sqrt{\mu a}} + 2\arcsin\sqrt{\frac{r_0}{2a}}\right) & @\ 0 < a < \infty \\ \sqrt{\frac{2}{9\mu}}\, r_0^{3/2} & @\ a = \infty \\ \sqrt{-\frac{a^3}{\mu}}\left(\frac{r_0 v_0}{\sqrt{-\mu a}} - 2\,\mathrm{arcsinh}\sqrt{-\frac{r_0}{2a}}\right) & @\ a < 0 \end{cases} \tag{7.5.19}$$

Radial Elliptic Trajectory

If $v_0 = 0$ then $a = r_0/2$ and we get for the radial elliptic trajectory

$$t_{col} = \sqrt{\frac{r_0^3}{8\mu}} \left(\frac{2R}{r_0} \sqrt{\frac{r_0}{R} - 1} + 2 \arccos \sqrt{\frac{R}{r_0}} \right) \quad @ \; v_0 = 0 \tag{7.5.20}$$

If even $R = 0$, then

$$t_{col} = \frac{\pi}{\sqrt{8\mu}} r_0^{3/2} \quad @ \; v_0 = 0, \;\; R \equiv r_{col} = 0 \tag{7.5.21}$$

Note that this last result could also be derived if applying the orbiting time of an elliptic orbit (Eq. (7.4.12)) $T = 2\pi\sqrt{a^3/\mu}$ to this case with $a = r_0/2$. This is possible because it is independent of the orbit is eccentricity. The collision time then is half an orbit revolution.

Example

Let us assume that the Earth's motion would abruptly be stopped. When would it crash into the center of the Sun (provided that the total masses of the Sun were combined in the center)? When would it crash onto the surface of the Sun?

From the mean Earth to Sun distance we get $2a = r_0 = 149.6 \times 10^6$ km. Since $\mu_\otimes = 1.327 \times 10^{11}\ \mathrm{km^3\,s^{-2}}$, the time to the center of the Sun according to Eq. (7.5.21) is

$$t_{col} = \pi \sqrt{\frac{(149.6 \times 10^6)^3}{8 \cdot 1.327 \times 10^{11}}}\, s = 64.57 \text{ days}$$

The Sun has a radius of $r = 0.696 \times 10^6$ km. According to Eq. (7.5.20) the body crashes onto the surface of the Sun after

$$t_{col} = \frac{64.57}{\pi} (0.1361 + 2 \cdot 1.5025) \text{ days} = 64.56 \text{ days}$$

According to Eq. (7.5.4), it would have an impact velocity of 41.7 km $\mathrm{s^{-1}}$.

7.5.5 Bounded Vertical Motion

A body moving vertically upward with initial velocity $v_0 > 0$ will either leave the gravitational potential for $a < 0$ or come to a halt somewhere and reverse its path for $a > 0$. In the following, we will study this later case.

First the body will move upward on a path with orbital element

$$a = \frac{\mu r_0}{2\mu - r_0 v_0^2} > 0$$

Gravitation will bring the body to a halt at $r_{up} = 2a$. Therefore, the altitude that the body attains is

$$h_{up} = r_{up} - r_0 = \frac{r_0^2 v_0^2}{2\mu - r_0 v_0^2} \tag{7.5.22}$$

If we assume $t_0 = 0$, we get from Eq. (7.5.7) for the time the body moves up to the apex point

$$t_{up} = \sqrt{\frac{a^3}{\mu}}\left(\pi + \frac{r_0 v_0}{\sqrt{\mu a}} - 2\arcsin\sqrt{\frac{r_0}{2a}}\right) = \sqrt{\frac{a^3}{\mu}}\left(\frac{r_0 v_0}{\sqrt{\mu a}} + 2\arccos\sqrt{\frac{r_0}{2a}}\right)$$

At the apex point, the body has velocity $v = 0$. From there on, it will fall backward. According to Eq. (7.5.5), the time it will be at position $r < r_{up}$ is

$$t_{down} = \sqrt{\frac{a^3}{\mu}}\left(\frac{r}{a}\sqrt{\frac{2a}{r} - 1} + 2\arccos\sqrt{\frac{r}{2a}}\right)$$

Note that this "down trajectory" has the same orbital element a as the "up trajectory". So in total the flight time is

$$t_{tot} = \sqrt{\frac{a^3}{\mu}}\left(\frac{r_0 v_0}{\sqrt{\mu a}} + 2\arccos\sqrt{\frac{r_0}{2a}} + \frac{r}{a}\sqrt{\frac{2a}{r} - 1} + 2\arccos\sqrt{\frac{r}{2a}}\right) \tag{7.5.23}$$

Vertical Shot up from a Planetary Surface

We now assume a shot up from the surface of a planet with radius R without an atmosphere, i.e., without drag. In this case $r_0 = r = R$. With

$$a = \frac{\mu R}{2\mu - R v_0^2} > 0$$

Equation (7.5.23) therefore reduces to

$$t_{tot} = 2\sqrt{\frac{a^3}{\mu}}\left(\frac{R v_0}{\sqrt{\mu a}} + 2\arcsin\sqrt{\frac{R v_0^2}{2\mu}}\right) \tag{7.5.24}$$

We finally want a power series approximation of the flight time of a body being shot up from the planet's surface with low kinetic energy, $s_0 := R v_0^2 / \mu \ll 2$. Then

$$a = \frac{R}{2}\left(\frac{1}{1 - s_0/2}\right) = \frac{R}{2}\left[1 + \frac{1}{2}s_0 + \frac{1}{4}s_0^2 + \frac{1}{8}s_0^3 + O\left(s_0^4\right)\right]$$

$$\sqrt{\frac{a^3}{\mu}} = \sqrt{\frac{R^3}{8\mu}}\left[1 + \frac{3}{4}s_0 + \frac{15}{32}s_0^2 + \frac{35}{128}s_0^3 + O\left(s_0^4\right)\right]$$

$$\frac{Rv_0}{\sqrt{\mu a}} = \sqrt{2s_0}\left[1 - \frac{1}{4}s_0 - \frac{1}{32}s_0^2 - \frac{1}{256}s_0^3 + O\left(s_0^4\right)\right]$$

$$2\arcsin\sqrt{\frac{s_0}{2}} = \sqrt{2s_0}\left[1 + \frac{1}{12}s_0 + \frac{3}{160}s_0^2 + \frac{15}{2688}s_0^3 + O\left(s_0^4\right)\right]$$

We then find from Eq. (7.5.24) after some calculation

$$t_{tot} = 2\sqrt{\frac{s_0 R^3}{\mu}}\left[1 + \frac{2}{3}s_0 + \frac{2}{5}s_0^2 + \frac{8}{35}s_0^3 + O\left(s_0^4\right)\right]$$

Because $g_0 = \mu/R^2$ is the planet's mean gravitational acceleration at its surface, we finally get

$$t_{tot} = \frac{2v_0}{g_0}\left[1 + \frac{2}{3}s_0 + \frac{2}{5}s_0^2 + \frac{8}{35}s_0^3 + O\left(s_0^4\right)\right] \tag{7.5.25}$$

with

$$s_0 = \frac{Rv_0^2}{\mu} = \frac{v_0^2}{g_0 R}.$$

Since $v_{\triangleright} = \sqrt{g_0 R}$ is the so-called first cosmic velocity of a planet, Eq. (7.5.25) holds for the condition $(v_0/v_{\triangleright})^8 \ll 1$. This result is in line with the classical throw upward where $t_{tot} = 2v_0/g_0$ for $s_0 \to 0$. The power series expansion (7.5.25) diverges for velocities larger than the second cosmic velocity $v \geq v_{\triangleright\triangleright} = \sqrt{2g_0 R}$, i.e., $s_0 \geq 2$.

Example

Let us assume that on the Moon ($R = 1737.4$ km, $\mu = 4.9028 \times 10^3\ \text{km}^3\text{s}^{-2}$) a bullet is shot straight up with initial speed $v_0 = 1\ \text{km s}^{-1}$. What is the altitude and return time it travels?

With $s_0 = Rv_0^2/\mu = 0.35437$ and $a = R/(2 - s_0) = 1055.76$ km, we have according to Eq. (7.5.22) and (7.5.24) $h_{up} = R/(2/s_0 - 1) = 374.13$ km and $t_{tot} = 26.705$ min. Classically, i.e., at a constant gravitational force, we would derive $h_{up} = v_0^2 R^2/2\mu = 307.84$ km and $t_{tot} = 2v_0 R^2/\mu = 20.523$ min.

7.6 Life in Other Universes?

Mathematically, and also physically, it is quite possible that other universes with other spatial dimensions in principle might exist. String theory for example states that at the time of the Big Bang our universe started out with nine spatial dimensions. According to current beliefs (Brandenberger and Vafa 1989), initially all these nine dimensions were curled up on the Planck scale (10^{-35} m), that is, there were no macroscopic dimensions as what we have today. So-called strings that make up our elementary particles were "living" on these curled-up spaces. Very shortly after the Big Bang, an antistring crashed into one of these rolled up strings and, according to the belief, they eliminated each other and generated an uncurled space dimension: The first macroscopic dimension was born. In one dimension the probability that a string and an antistring meet is still very high. A string and an antistring annihilated anew, leading to a second macroscopic dimension. The question is whether two dimensions offer enough space, so that the strings no longer meet each other. Obviously not: the third dimension was born. Will someday another string and antistring meet again in our three dimensions to open up a fourth dimension? Nobody knows. However, we do know that coincidences play an important role in quantum mechanics. It could have been quite possible that not only two, but even four or five macroscopic dimensions could have formed, especially when the universe was still very small. Could we live in such a universe? Life in our universe, apart from many other factors, decisively depends on whether we have stable planetary orbits around a central star. So if we want to have an answer to the question of whether life would be possible in universes with other dimensions, first of all we would have to find out whether there would be stable planetary orbits. This is exactly what we will figure out now.

7.6.1 Equation of Motion in n Dimensions

First of all, one has to consider that according to Noether's theorem the conservation laws (see Eqs. (7.1.8)–(7.1.10)), especially the law of conservation of angular momentum, are independent of the dimension of the space. They are determined only by the homogeneity and isotropy of spacetime, and not by its dimensionality. The conservation laws are thus valid in all homogeneous n dimensional universes. Also, the law of conservation of energy is valid:

$$E_{kin} + E_{pot} = m\varepsilon = const$$

and the two expressions (see Eq. (7.1.4))

$$E_{kin} = \frac{1}{2}mv^2$$
$$E_{pot} = mU(r)$$

are independent of dimension. The angular momentum is defined as the cross product of position vector and velocity vector. Vectors are one-dimensional entities. That angular momentum is conserved means that these two vectors in a given n dimensional space open up a hyperspace. Two non-collinear vectors open up a plane. So for $n \geq 2$, the gravitationally determined motion in general is restricted to a plane spanned by the initial vectors. (If by coincidence the initial vectors were collinear, which is inevitably in the case $n = 1$, the motion for $n \geq 1$ would be one dimensional.) Hence, independent of n, the general gravitational motion of a body in an n-dimensional space is always in a plane, and we therefore will measure it by means of polar coordinates (r, θ). As the motion takes place in a plane, the considerations leading to Eq. (7.2.8)

$$v^2 = \dot{r}^2 + \frac{h^2}{r^2}$$

still remain correct. The only thing that changes with space dimensions is the gravitational potential. We find it by solving the corresponding Poisson's Eq. (7.1.1). Its general solution in $n \geq 3$ dimensions is (exercise, Problem 7.1)

$$U(r) = -\frac{\mu}{r^{n-2}} \tag{7.6.1}$$

whereby μ carries the unit $[\mu]$ = [m^n/s^2]. By applying all the above expressions to the energy conservation equation, we get the vis-viva equation for n dimensions:

$$\dot{r}^2 = \frac{2\mu}{r^{n-2}} - \frac{h^2}{r^2} + 2\varepsilon \qquad \textbf{vis-viva equation in } \boldsymbol{n} \textbf{ dimensions} \tag{7.6.2}$$

We seek for the equation of motion, and differentiate Eq. (7.6.2) to get

$$\ddot{r} = -\frac{(n-2)\mu}{r^{n-1}} + \frac{h^2}{r^3} \tag{7.6.3}$$

A big advantage of this differential equation is the fact that it is not vectorial, but scalar. A disadvantage is the sum of two terms on the right-hand side of the equation, as they both contain r. With this, we are no longer able to find a simple analytical solution for the differential equation by just separating the variables. So we have to look for other approaches. An important feature to solve differential equations in a smart way is to make a solution ansatz or a substitution, that comprises as much advance information as possible. Apparently, we have expressions of the form $1/r$. We assume that the solution is of the same form, and thus we change to the new radial variable: $\rho := 1/r$. (In the literature, this substitution is known as the *Burdet transformation*.) The second piece of previous knowledge we have is the conservation of the angular momentum. Therefore, we select the fixed angular momentum as one coordinate axis z, and for the other coordinates we use the rotating system of polar coordinates $(\rho, \ \theta)$. All in all, we now have a system of cylindrical coordinates $(\rho, \ \theta, \ z)$, which will later prove to be naturally adapted to this problem. We also know that the motion is periodic, whereas the time variable is linear. So it is a good idea to change for $h > 0$ from the time variable t to the orbit angle variable θ. Substituting $\rho = 1/r$ or $r = 1/\rho$, respectively, results with Eq. (7.2.7) in

$$\dot{r} = -\frac{1}{\rho^2}\frac{d\rho}{dt} = -\frac{1}{\rho^2}\frac{d\rho}{d\theta}\frac{d\theta}{dt} = -\frac{1}{\rho^2}\rho'\omega = -\frac{1}{\rho^2}\rho'\frac{h}{r^2} = -h\rho'$$

and

$$\ddot{r} = -h\frac{d\rho'}{dt} = -h\rho''\frac{h}{r^2} = -h^2\rho^2\rho''$$

With this we obtain for Eq. (7.6.3)

$$\rho'' + \rho = \frac{(n-2)\mu}{h^2}\rho^{n-3}$$

This equation of motion indeed looks easier. But a relation between the second variable θ and ρ is still missing. We get the missing equation from Eq. (7.3.8).

$$\dot{\theta} = \frac{h}{r^2} = h\rho^2$$

This concludes our search for differential equations of motion in n dimensions.

$$\begin{aligned} &\rho'' + \rho = \tfrac{(n-2)\mu}{h^2}\rho^{n-3} \\ &\dot{\theta} = h\rho^2 \\ &\ddot{u}_i = 0 \quad i = 3, \ldots, n \end{aligned} \qquad \textbf{Newton's gravitational EoM in } \boldsymbol{n} \textbf{ dimensions} \qquad (7.6.4)$$

where u_i are the remaining radial vector components that do not lie in the motion plane.

3-dim Universe

To verify the equation of motion in n dimensions we test for our well-known three dimensions

$$\rho'' + \rho = \frac{\mu}{h^2}$$

To solve it, we rewrite it to

$$\rho'' = -\left(\rho - \frac{\mu}{h^2}\right)$$

With the substitution $\lambda := \rho - \mu/h^2$ we get $\lambda'' = \rho''$, which results in the new simple differential equation $\lambda'' = -\lambda$. It has with the general solution $\lambda = \lambda_0 \cos(\theta - \theta_0)$. By resubstitution we obtain

$$\rho = \frac{\mu}{h^2} + \rho_0 \cos(\theta - \theta_0)$$

The two integration constants ρ_0 and θ_0 are determined by the specific initial conditions. Resubstituting $\rho = 1/r$ results in the well-known orbit equation (see Eq. (7.3.5))

$$r = \frac{p}{1 + e \cdot \cos\theta}$$

with $$p := h^2/\mu,\ e := p\rho_0.$$

7.6.2 4-Dimensional Universe

For $n = 4$ dimensions, the first equation of Eq. (7.6.4) reads

$$\rho'' = -\rho\left(1 - \frac{2\mu}{h^2}\right) = \pm k^2 \rho$$

with

$$k := \sqrt{\left|1 - \tfrac{2\mu}{h^2}\right|}.$$

We have to distinguish three cases:

1. Case $1 - \frac{2\mu}{h^2} > 0$

In this case $\rho'' = -k^2\rho$ and the solution is $\rho = \rho_0 \sin(k\theta + \varphi)$ or

$$r = \frac{r_0}{\sin(k\theta)} \tag{7.6.5}$$

where we chose $\varphi = 0°$ as an initial condition at $\theta = \pi/(2k)$. That is for $\theta = 0$ a planet is at infinity and approaches the scene. With $r = r_0$, it attains its smallest distance to the star, and then recedes into infinity. In total, this process can be considered as a flyby at a star as shown in Fig. 7.16.

2. Case $1 - \frac{2\mu}{h^2} < 0$

Here $\rho'' = k^2\rho$ and the solution is $\rho = \rho_0 e^{k\theta + \varphi}$ or

$$r = r_0 e^{-k\theta} \tag{7.6.6}$$

where we again chose an initial $\varphi = 0°$ at $\theta = \pi/2k$. In other words, the planet spirals exponentially toward the star until it crashes into it (see Fig. 7.16).

Note: *The other possible mathematical solution $r = r_0 e^{k\theta}$ is unphysical, as it would imply a repelling gravitational force.*

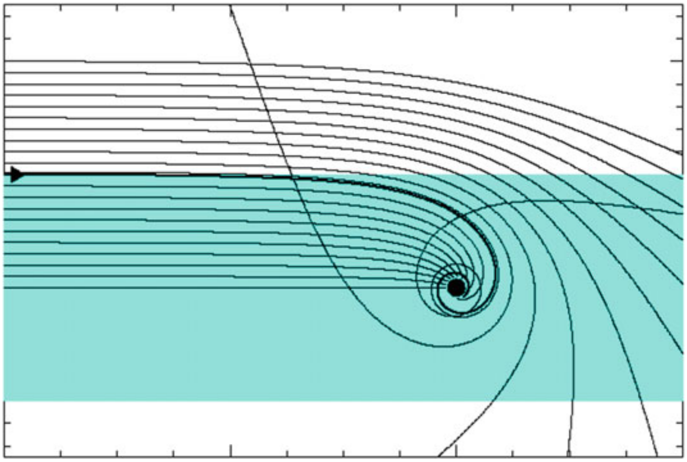

Fig. 7.16 Numerical simulation of the two-body problem in a four-dimensional space. Lightweight bodies with the same momentum, but different impact parameters, approach the central body from the left side. They escape into infinity again or crash into the center, depending on whether their impact parameter is in the shaded area or not. There are no stable orbits. Credit: Tegmark (1997)

3. Case $1 - \frac{2\mu}{h^2} = 0$

   Here, $\rho'' = 0$ and thus $\rho = a\theta + b$ or

$$r = \frac{1}{a\theta} \tag{7.6.7}$$

   Here, $b = 0$ was selected, i.e., $r(\theta = 0) = \infty$. On this borderline orbit the planet spirals toward the star, inversely proportional to θ, until it crashes onto its surface (bold trajectory in Fig. 7.16).

The orbits of three cases are shown in Fig. 7.16 taken from the literature, which were obtained by numerical simulation.

Conclusion: In a four-dimensional universe, none of the three possible orbits are stable or bounded, and therefore no planetary systems can exist and hence no life would be possible.

7.6.3 Universes with ≥ 5 Dimensions

In universes with $n \geq 5$ dimensions, the inverse radial acceleration is

$$\rho'' = \frac{(n-2)\mu}{h^2}\rho^{n-3} - \rho = \rho\left[\frac{(n-2)\mu}{h^2}\rho^{n-4} - 1\right] \tag{7.6.8}$$

Let us start our examination with a planet at $r = \infty$, i.e., $\rho = 0$, approaching the star. As long as the distance is large enough, so that

$$r^{n-4} > \frac{(n-2)\mu}{h^2}$$

or

$$\rho^{n-4} < \frac{h^2}{(n-2)\mu}$$

is valid, the expression in brackets of Eq. (7.6.8) is smaller than zero, and therefore $\rho'' < 0$. Let us assume that $\rho'' = -a = const < 0$ for a short period. Then we get the solution

$$\rho = \rho_0 + b\theta - \frac{a}{2}\theta^2$$

Even if the inverse radial velocity b were slightly positive in the beginning, ρ will further decrease after a certain advance of the orbit angle. If ρ decreases, ρ'' will become even more negative, and we get a runaway effect with the limiting value $\rho \to 0$, implying $r \to \infty$. This means that our body is gravitationally not bound to the star. At a certain position, it will approach the star to a minimum distance, depending on the incidence angle and velocity, without falling below the critical radius

$$\rho_{crit} = \frac{1}{r_{crit}} = \left[\frac{h^2}{\mu(n-2)}\right]^{\frac{1}{n-4}}$$

On its further track, it recedes and disappears in the depths of the universe. In total, its track will be more or less deflected. Qualitatively, this case corresponds to the first case in four dimensions.

If the body falls below the critical inverse radius due to its initial conditions, then the term in brackets in Eq. (7.6.8) becomes positive and thus $\rho'' > 0$. Let us assume that $\rho'' = a = const > 0$ for a moment. Then we get the solution

$$\rho = \rho_0 + b\theta + a\theta^2$$

As the body approaches from outside, its inverse radial velocity is $b > 0$, which means that in total ρ increases even faster (i.e., r decreases more). Then the term in brackets in Eq. (7.6.8) attains even larger positive values. Therefore, $\rho'' > 0$ increases further more, and ρ increases even faster. So we get an opposite runaway effect with the limit value $\rho \to \infty$, implying $r \to 0$: The planet approaches the star faster and faster, until it crashes into the star.

Conclusion: Also, in universes with dimensions $n \geq 5$, no stable planetary systems can exist, and thus life is not possible.

How would a $n \geq 4$-dimensional universe then evolve? Probably after its birth and after a very short period, the so-called epoch of radiation, when the masses form, these masses would immediately clot together to form black holes, and they would never form any gravitationally coupled stellar system or even galaxies. The black holes would merge into even bigger holes within a short period because of their large critical radius, and finally there would only be a huge black hole left that would absorb all the radiation and matter of the universe: That would be the quick end of a universe, hardly had it begun to exist.

Remark *Historically, in 1917 Ehrenfest already showed qualitatively, and in 1963 Büchel showed by general energy considerations, that for $n \geq 4$ there is no possibility that stable planetary orbits can exist. They are either deflected by the central body, or crash into it within a very short period of time.*

7.6.4 Universes with ≤ 2 Dimensions

In 1984, Deser and Jackiw and independently from them, Gott and Alpert applied the theory of general relativity to $n \leq 2$ spatial dimensions and found that the space surrounding a point mass would not have a curvature (the Riemann tensor and with it Einstein's curvature tensor would vanish). This means that other particles would not experience any gravitational pull. So, in $n \leq 2$-dimensional spaces there is no gravitational attraction at all, let alone an answer to the question of stable orbits. Classic astrodynamics erroneously has a different point of view. From Poisson's Eq. (7.1.1), it follows that in two dimensions a gravitational potential $U(r) = -\mu \ln r$ with force $F \propto -\mu/r$ exists (exercise, Problem 7.1). Since $U(r)$ diverges for $r \to \infty$, this already shows us that this solution is quite far from reality. The inconsistency between the theory of relativity and Newton's physics can be explained by the fact that for $n \leq 2$ in the theory of general relativity a correspondence to classical physics no longer exists.

Conclusion: Since in universes with $n \leq 2$ dimensions a gravitational force is not existent, planets also cannot exist, let alone life.

So, with our three spatial dimensions, we live on an island of stability, and we can only assume and hope that it is not as coincidental as string theory currently suggests.

7.7 Stellar Orbits

7.7.1 Motion in General Gravitational Potentials

In this section we consider the most general case, the motion of a body in a gravitational potential $U(\boldsymbol{r})$ that is generated by an arbitrary mass distribution $\rho(\boldsymbol{r})$.

To determine $U(\boldsymbol{r})$ from $\rho(\boldsymbol{r})$ we recall from Eq. (7.1.3) that the potential of a point mass M at origin (location $\boldsymbol{0}$) is given by

$$U(r) = -\frac{GM}{r} = -G\frac{M}{|\boldsymbol{r} - \boldsymbol{0}|}$$

Therefore, the potential of many masses M_i ($i = 1, \ldots, n$) at positions $\boldsymbol{r}_i$ is

$$U(\boldsymbol{r}) = -G\sum_i^n \frac{M_i}{|\boldsymbol{r} - \boldsymbol{r}_i|}$$

which entails that the potential may no longer be isotropic. If we assume a celestial body with a continuous mass distribution described by the density distribution function $\rho(\boldsymbol{r})$, then we have to carry out the transition $M_i \to \rho(\boldsymbol{r}) \cdot dV = \rho(\boldsymbol{r}) \cdot d^3\boldsymbol{r}$, whereby the sum becomes a volume integral

$$U(\boldsymbol{r}) = -G\iiint_V \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}d^3\boldsymbol{r}' \tag{7.7.1}$$

From Eq. (7.1.5) and Newton's second law Eq. (7.1.12) the motion within this potential is determined by the following equation of motion

$$\ddot{\boldsymbol{r}} = -\frac{d}{d\boldsymbol{r}}U(\boldsymbol{r}) \tag{7.7.2}$$

In view of the application to galaxies having a center of a point symmetric mass distribution, i.e. a location where $U(-\boldsymbol{r}) = U(\boldsymbol{r})$, we place the origin of our inertial reference frame (see Sect. 13.1.1) at this center. Thus

$$\frac{dU}{d\boldsymbol{r}}(\boldsymbol{r} = 0) =: \left.\frac{dU}{d\boldsymbol{r}}\right|_0 = 0$$

By employing cylindrical coordinates $\boldsymbol{u}_r, \boldsymbol{u}_\theta, \boldsymbol{u}_z$ and with Eq. (7.2.6) we can decompose the vector equation of motion into

$$\ddot{\boldsymbol{r}} = \begin{pmatrix} \ddot{r} - \omega^2 r \\ 2\omega\dot{r} + \dot{\omega}r \\ \ddot{z} \end{pmatrix}_{r\theta z} = -\frac{dU}{d\boldsymbol{r}} = -\left(\frac{\partial U}{\partial r}, \frac{1}{r}\frac{\partial U}{\partial \theta}, \frac{\partial U}{\partial z}\right)^T_{r\theta z} \tag{7.7.3}$$

The above choice of the origin at the center thus yields

$$\left.\frac{\partial U}{\partial r}\right|_0 = 0, \left.\frac{\partial U}{\partial z}\right|_0 = 0$$

Orbits in Axisymmetric Potentials

Our first restrictive, yet good, assumption (so-called *Wentzel-Kramers-Brillouin approximation*) is that, though billions of stars are moving in a galaxy, they on average generate a gravitational potential that is constant in time. In addition, we assume that the potential is laterally symmetric (axisymmetric), i.e. $\partial U/\partial\theta = 0$, everywhere. From Eq. (7.7.3) we then have $2\omega\dot{r} + \dot{\omega}r = 0$. We multiply this equation by r and find

$$\frac{d}{dt}\left(\omega r^2\right) = \frac{dh}{dt} = 0$$

This implies that in an axisymmetric potential the angular momentum is conserved

$$h = \omega r^2 = const$$

In respect to the radial motion of the body we employ from Sect. 7.2.4 the concept of the effective potential, which is defined as

$$U_{eff}(\boldsymbol{r}) := U(\boldsymbol{r}) + \frac{h^2}{2r^2} = U(\boldsymbol{r}) + \frac{1}{2}\omega^2 r^2 \quad \textbf{effective potential} \tag{7.7.4}$$

For an axisymmetric potential with $h = const$ we have

$$\frac{\partial U_{eff}}{\partial r} = \frac{\partial U}{\partial r} - \frac{h^2}{r^3} = \frac{\partial U}{\partial r} - \omega^2 r$$

Therefore we find from Eq. (7.7.3) two scalar equations of motion

$$\begin{aligned} \ddot{r} &= -\frac{\partial U_{eff}}{\partial r} \\ \ddot{z} &= -\frac{\partial U_{eff}}{\partial z} = -\frac{\partial U}{\partial z} \end{aligned} \quad \textbf{equations of motion} \tag{7.7.5}$$

Circular Guiding Orbits

What is the condition for stable orbits in an axisymmetric potential? For this to happen the effective potential must have a minimum, both in r and z, i.e.

$$0 = \frac{\partial U_{eff}}{\partial r} = \frac{\partial U}{\partial r} - \omega^2 r \tag{7.7.6}$$

$$0 = \frac{\partial U_{eff}}{\partial z} = \frac{\partial U}{\partial z} \tag{7.7.7}$$

We have seen that $\partial U/\partial z = 0$ is fullfilled anywhere in the plane vertical to the symmetry (a.k.a. *equatorial plane*). At the radial minimum we have $\dot{r} = 0$ and hence a stable circular orbit. Observe that every stable orbit has its individual angular momentum $h = \omega r^2$ and hence there may be infinitely many of them. We denote the radius of a given circular orbit as R and its orbital frequency as $\Omega := \omega(R)$. Thus, we have stable circular orbits in the equatorial plane, which we will call guiding orbits, if the following condition holds

$$\left.\frac{\partial U}{\partial r}\right|_R = R\Omega^2 \quad @ \text{ guiding orbits} \tag{7.7.8a}$$

Since U is a given, this equation in fact is a conditional equation for the angular frequency of a guiding orbit with orbital radius R

$$\Omega = \sqrt{\frac{1}{R}\left.\frac{\partial U}{\partial r}\right|_R} = \frac{h}{R^2} \tag{7.7.8b}$$

Stable Near-Circular Orbits

We want to know whether there also exist stable non-circular orbits in the vicinity of a circular guiding orbit and what are the conditions for them. For a stable near-to-guiding orbit the effective potential

$$U_{eff}(\boldsymbol{r}) = U(\boldsymbol{r}) + \frac{1}{2}\omega^2 r^2$$

needs to have a positive curvature at $r = R$, i.e.

$$\left.\frac{\partial^2 U_{eff}}{\partial r^2}\right|_R = \left.\frac{\partial^2 U}{\partial r^2}\right|_R - \left.\frac{\partial}{\partial r}\frac{h^2}{r^3}\right|_R = \left.\frac{\partial^2 U}{\partial r^2}\right|_R + 3\frac{h^2}{R^4} = \left.\frac{\partial^2 U}{\partial r^2}\right|_R + 3\Omega^2 > 0$$

Owing to Eq. (7.7.8a), we thus find the stability condition at the circular guiding orbit

$$\left.\left(\frac{\partial^2 U}{\partial r^2} + \frac{3}{r}\frac{\partial U}{\partial r}\right)\right|_R > 0 \tag{7.7.9}$$

Since according to Eq. (7.7.6) the local minimum condition $r\omega^2 = \partial U/\partial r$ holds, we have

$$\frac{\partial^2 U}{\partial r^2} = \omega^2 + 2r\omega\frac{\partial \omega}{\partial r}$$

Inserting this into the above stability conditional equation we get the alternative stability condition

$$\left.\left(\frac{r}{\omega}\frac{\partial\omega}{\partial r}\right)\right|_R = \left.\frac{\partial \ln\omega}{\partial \ln r}\right|_R > -2 \tag{7.7.10}$$

Let us assume that at the radial distance $r = R$ of the guiding orbit the potential quite generally behaves as $U \propto r^p$ with $p \in \mathbb{R}$. Then

$$\left.\left(\frac{\partial^2 U}{\partial r^2} - \frac{p-1}{r}\frac{\partial U}{\partial r}\right)\right|_R > 0$$

Inserting this result into Eq. (7.7.9) we find $p > -2$. Because also $U = c \cdot \ln r$ fulfills Eq. (7.7.9), we get for the gravitational potential about the guiding orbit and for attractive forces

$$U = \left\{\begin{array}{ll} cr^p & p > 0 \\ c \cdot \ln r & \\ -c/r^p & -2 < p < 0 \end{array}\right\} \quad @\ r = R \tag{7.7.11}$$

It can be shown (see also below) that all galaxies obey the conditional equation (7.7.9) and Eq. (7.7.10) and hence Eq. (7.7.11). Therefore, as long as stars do not enter the SOI of a neighboring star, which means that they do suffer mutual gravitational interaction, they move on stable near-circular orbits in any galaxy.

7.7.2 *Stellar Motion in General Galaxies*

We now study stable near-circular orbits in general galaxies and we presume that also some form of epicyclic motion exists. To study that motion in detail, we need to evaluate the equations of motion of the excursions

$$x = r - R$$

and of z from the equilibrium point $(R, 0)$ of the guiding orbit, which according to Eq. (7.7.5) reads

$$\ddot{z} = -\frac{\partial U_{eff}}{\partial z}, \ddot{x} = -\frac{\partial U_{eff}}{\partial x}$$

In view of the small excursions we expand U_{eff} into a Taylor series about the equilibrium point $(R, 0)$

$$U_{eff} = U_{eff}(R,0) + \frac{1}{2}\left(\frac{\partial^2 U_{eff}}{\partial r^2}\right)\bigg|_{R,0} x^2 + \frac{1}{2}\left(\frac{\partial^2 U_{eff}}{\partial z^2}\right)\bigg|_{R,0} z^2 + O(xz^2) \qquad (7.7.12)$$

and neglect terms of $O(xz^2)$ or higher. Note that the term of $O(xz)$ vanishes, because $U_{eff}(-z) = U_{eff}(z)$. From this we find the equations of a radial and vertical oscillating motion

$$\ddot{x} = -\frac{\partial U_{eff}}{\partial r} = -x \cdot \frac{\partial^2 U_{eff}}{\partial r^2}\bigg|_{R,0} =: -\kappa^2 x \qquad \textbf{radial epicycle} \qquad (7.7.13)$$

$$\ddot{z} = -\frac{\partial U_{eff}}{\partial z} = -z \cdot \frac{\partial^2 U_{eff}}{\partial z^2}\bigg|_{R,0} =: -\nu^2 z \qquad \textbf{vertical oscillation} \qquad (7.7.14)$$

Relation of Epicycle to Guiding Orbit Oscillation

We now are interested in how κ relates to Ω of the guiding orbit, about which we have the epicyclic oscillations. From Eqs. (7.7.4) and (7.7.8a) we derive

$$\kappa^2 = \frac{\partial^2 U_{eff}}{\partial r^2}\bigg|_R = \frac{\partial}{\partial r}\left(\frac{\partial U}{\partial r}\right)\bigg|_R + \frac{\partial^2}{\partial r^2}\left(\frac{h^2}{2r^2}\right)\bigg|_R = \frac{\partial}{\partial r}\left(r\omega^2\right)\bigg|_R + 3\omega^2\big|_R$$
$$= \Omega^2 + 2R\Omega\frac{\partial\omega}{\partial r}\bigg|_R + 3\Omega^2 = 4\Omega^2\left(1 + \frac{R}{2\Omega}\frac{\partial\omega}{\partial r}\bigg|_R\right)$$

Hence

$$\kappa^2 = 4\Omega^2\left(1 + \frac{1}{2}\frac{\partial\ \ln\omega}{\partial\ \ln r}\bigg|_R\right) \qquad (7.7.15)$$

Owing to the positive definite value $\kappa^2 > 0$ we again (see Eq. (7.7.10)) derive the condition for a stable near-circular orbit

$$\frac{\partial\ \ln\omega}{\partial\ \ln r}\bigg|_R > -2.$$

There seem to be two limiting cases. On one hand we have a point mass at the galactic center, which causes $U = -\mu/r$ and thus via the equilibrium condition $r\omega^2 = \partial U/\partial r$ delivers $\omega^2 = \mu/r^3$. Hence $\partial\ \ln\omega/\partial\ \ln r = -3/2$, and hence $\kappa = \Omega$. On the other hand $h = \omega r^2 = const$ holds, implying $d\omega/dr < 0$ and hence $d\ \ln\omega/d\ \ln r < 0$. So, for any mass distribution we have the limits

$$\Omega \le \kappa \le 2\Omega \qquad (7.7.16)$$

Epicyclic Motion

In the above we saw that the guiding orbit at $r = R$ and orbital frequency Ω is superimposed by the radial excursions with EoM $\ddot{x} = -\kappa^2 x$ and hence with path

$$x = x_0 \sin(\kappa t + \varphi)$$

The angular epicylic motion can be derived from the law of the conservation of angular momentum $\omega r^2 = \dot{\theta} r^2 = h = const$ as

$$\dot{\theta} = \frac{h}{R^2} = \frac{\Omega R^2}{(R+x)^2} \approx \Omega\left(1 - \frac{2x}{R} + \ldots\right) = \Omega + \frac{2\Omega}{R} x_0 \sin(\kappa t + \varphi)$$

Direct integration yields

$$\theta(t) = \theta_0 + \Omega \cdot t + \frac{2\Omega}{R\kappa} x_0 \cos(\kappa t + \varphi) \tag{7.7.17}$$

Since $r = R + x$, the radial variation of the stellar orbit about its guiding orbit is given as

$$r = R + x_0 \sin(\kappa t + \varphi) = R\left[1 + \frac{x_0}{R} \sin\left(\frac{\kappa}{\Omega}\theta + \phi\right)\right] \tag{7.7.18}$$

where we have transformed from the time domain into the angular domain employing the mean motion $\theta(t) = \theta_0 + \Omega \cdot t$ from Eq. (7.7.17). Equation (7.7.18) together with Eq. (7.7.17) describes the epicyclic motion.

Orbital Phase Shift

If we compare Eq. (7.7.18) with the general solution for a Keplerian orbit

$$r = \frac{p}{1 + e\cos\theta} \approx R(1 + e\cos\theta) \quad @\ e \to 0$$

we see that the eccentricity $e = x_0/R$ is given by the relative amplitude of the oscillation, and κ/Ω is the commensurability between epicycle and guiding motion. If we define the angular orbital period θ_P as

$$\theta_P \frac{\kappa}{\Omega} := 2\pi \quad \text{or} \quad \theta_P = 2\pi \frac{\Omega}{\kappa}$$

the orbit phase shift, i.e. the periapsis angle after one epicycle with respect to a full guiding orbit cycle, is

$$\Delta\theta = \theta_P - 2\pi = 2\pi\left(\frac{\Omega}{\kappa} - 1\right) \quad \textbf{orbital phase shift} \tag{7.7.19}$$

Owing to the above restriction $\Omega \le \kappa \le 2\Omega$, we have $-180^\circ \le \Delta\theta \le 0$. If $\kappa/\Omega = n/m$ the epicycle is commensurable, i.e., the stellar path closes after n epicycles and

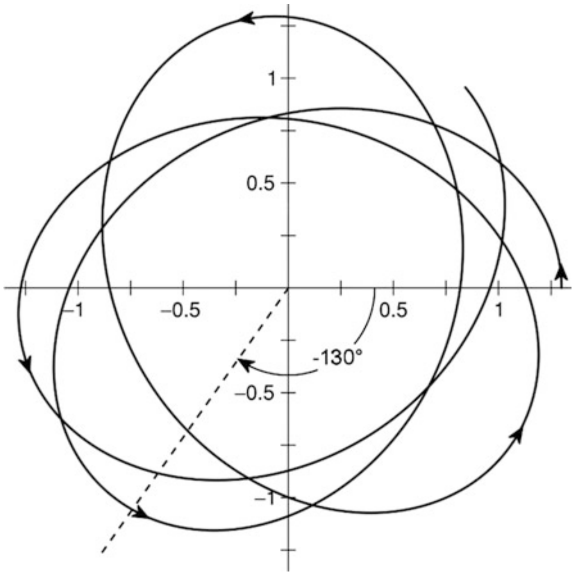

Fig. 7.17 An epicyclic motion with $\kappa = 1.36\,\Omega$ and according orbital phase shift $\Delta\theta = -130°$

m guiding orbit cycles. Usually stellar epicycles are not commensurable. Figure 7.17 depicts the epicyclic motion with $\kappa = 1.56 \cdot \Omega$ and hence $\Delta\theta = -130°$.

Example

For our Sun in the Milky Way we have

$$\kappa = 36.7 \text{ km s}^{-1}\text{kpc}^{-1}$$
$$\Omega = 27.2 \text{ km s}^{-1}\text{kpc}^{-1} = 8.81 \times 10^{-16} \text{ s}^{-1}$$

So

$$\kappa = (1.35 \pm 0.05)\Omega$$

The Sun therefore makes 1.35 oscillations in and out for every circuit of the galaxy. It therefore takes the Sun $T = 2\pi/\Omega = 226$ Million years to circle the galactic center once. The Sun's orbital phase shift after one epicycle is

$$\Delta\theta = 360° \left(\frac{\Omega}{\kappa} - 1 \right) = -93°$$

7.7.3 Stellar Orbits in Globular Cluster Galaxies

Suppose we have a globular cluster galaxy that is a spherical distribution of stars with radius R and mass $M = \frac{4}{3}\pi\rho R^3$. Owing to its billions of stars, it is fair to

Fig. 7.18 The great globular cluster M13 in the constellation Hercules. Credit Marco Burali, Tiziano Capecchi, Marco Mancini/Osservatorio MTM

assume that it is isotropic and homogeneous with constant mass density $\rho = const$ (see Fig. 7.18). We therefore can apply the isotropic form of Poisson's Eq. (7.1.2)

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)U = 4\pi G\rho \quad @ \ \rho = const$$

This can be solved by two succeeding direct integrations

$$\begin{aligned} r^2\frac{\partial U}{\partial r} &= \frac{4}{3}\pi G\rho r^3 \\ U(r) &= \frac{2}{3}\pi G\rho r^2 \end{aligned} \tag{7.7.20}$$

with the choice (see Sect. 7.3.4) $U(r \to 0) = 0$.

Since this potential satisfies the stability condition (7.7.11), stable periodic orbits must exist. To determine their orbit equation we apply Eqs. (7.7.2) and (7.1.6) to find the equation of motion

$$\ddot{\boldsymbol{r}} = -\frac{d}{d\boldsymbol{r}}U(\boldsymbol{r}) = -\frac{4}{3}\pi G\rho\boldsymbol{r} = -\frac{GM}{R^3}\boldsymbol{r}$$

With the substitution $\kappa = \sqrt{4\pi G\rho/3}$ this vector equation can be decomposed into

$$\begin{aligned} \ddot{x} &= -\kappa^2 x \\ \ddot{y} &= -\kappa^2 y \end{aligned}$$

The general solution of each these well-known differential equations is the harmonic oscillator

$$u(t) = u_0 + a\ \cos(\kappa t)\cos\varphi + b\ \sin(\kappa t)\sin\varphi$$

This is the equation of an ellipse, where u_0 marks the center and φ is the angle between the coordinate axis and the mayor axis of the ellipse. If we locate the center at the origin and place the x-axis along the major axis ($\varphi = 0$) and the y-axis along the minor axis ($\varphi = 90°$), we obtain

$$\begin{aligned} x &= a\ \cos(\kappa t) \\ y &= b\ \sin(\kappa t) \end{aligned}$$

This solution is an ellipse

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

with orbital frequency $\kappa = \sqrt{4\pi G\rho/3}$, and hence period $T = 2\pi/\kappa$, and centered (not focused) at the center of the galaxy $(0, 0)$. This motion can be considered as an epicyclic motion about a guiding orbit with $\kappa = 2\Omega$, i.e. the epicycle period is in 2:1 correspondence with the guidance orbital period, and orbital phase shift $\Delta\theta = -180°$ (see Sect. 7.7.2).

7.7.4 Stellar Motion in Disk-Shaped Galaxies

Most galaxies such as our Milky Way, however, are thin rotating disks (see Fig. 7.19) with, say, radius a and thickness $h \ll a$. Assuming again a constant mass density $\rho = const$, and assuming no vertical movements of stars within the disk, the gravitational potential in the galactic plane can be shown (Danby 2003, Chap. 5, problem 8) to be

$$U = -2\pi G\rho h a\left[1 - \frac{1}{4}\left(\frac{r}{a}\right)^2 - \frac{3}{4^3}\left(\frac{r}{a}\right)^4 - \frac{5}{4^4}\left(\frac{r}{a}\right)^6 - \frac{7\cdot5\cdot3}{4^7}\left(\frac{r}{a}\right)^8 - \mathrm{K}\right] \quad @\ r < a$$

$$\frac{dU}{dr} = \pi G\rho h\left[\frac{r}{a} + \frac{3}{8}\left(\frac{r}{a}\right)^3 + \frac{15}{64}\left(\frac{r}{a}\right)^5 + \frac{105}{1\,024}\left(\frac{r}{a}\right)^7 + \cdots\right]$$

$$\frac{d^2U}{dr^2} = \pi G\rho\frac{h}{a}\left[1 + \frac{9}{8}\left(\frac{r}{a}\right)^2 + \frac{75}{64}\left(\frac{r}{a}\right)^4 + \frac{735}{1\,024}\left(\frac{r}{a}\right)^6 + \cdots\right]$$

Fig. 7.19 The unbarred spiral galaxy NGC 4565 in the constellation Coma Berenices. *Credit* Adam Block/Kitt Peak National Observatory

and therefore the force acting on a star with mass m_* is

$$\boldsymbol{F} = -m_* \frac{dU}{d\boldsymbol{r}} = -\pi G m_* \rho h \frac{\boldsymbol{r}}{a}\left[1 + \frac{3}{8}\left(\frac{r}{a}\right)^2 + \frac{15}{64}\left(\frac{r}{a}\right)^4 + \frac{105}{1\,024}\left(\frac{r}{a}\right)^6 + \cdots\right]$$

Because $dU/dr > 0$ and $d^2U/dr^2 > 0$ at any radial distance from the center, the stability condition Eq. (7.7.9)

$$\frac{\partial^2 U}{\partial r^2} + \frac{3}{r}\frac{\partial U}{\partial r} > 0$$

holds and therefore there exist stable near-circular orbits orbits at any radial distance.

The radial value of the gravitational force function is shown in Fig. 7.20. It can be seen that toward the outer regions of the Milky Way, the absolute value of the gravitational force gradually increases over the force $\boldsymbol{F}(\boldsymbol{r}) = -G\rho m_* \boldsymbol{r}$ as in globular cluster galaxies. We therefore expect, in general, also ellipses as stellar orbits centered on the center of the Milky Way. However, when a star on its orbit moves outward, it is subject to an excessive gravitational force, which can be interpreted as an increasing standard gravitational parameter μ. At a given radial distance, this implies an excessive orbital velocity $v + \Delta v = \sqrt{(\mu + \Delta\mu)/r}$, which in turn leads to a prograde rotation of the line of apsides. Therefore and as shown in Sect. 7.7.3, such ellipses in a flattened galaxy do not close themselves, but get offset each time a star completes one revolution. The Sun, for example, takes about 230 million years

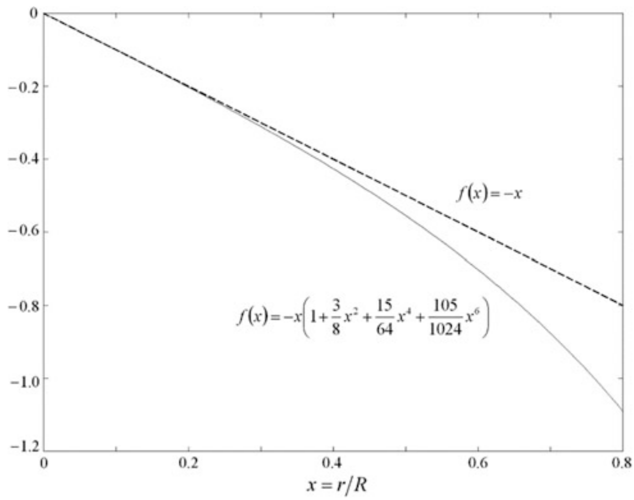

Fig. 7.20 The force function within a flattened galaxy at position $x = r/a$

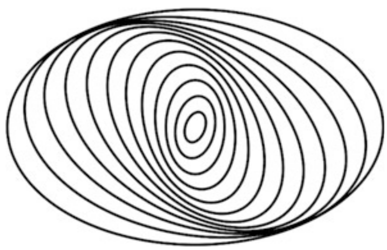

Fig. 7.21 Spiral waves in a flattened galaxy arise when the elliptic star orbits move in unison but are slightly skewed compared to its neighbors. The density of stars is highest where the ellipses crowd together

to go around its elliptic orbit. In that time the orbit gets offset by 105°. In spiral galaxies the elliptic orbits of all stars, though at different mean distances from the center, rotate in lockstep at constant offsets, which causes the spiral arms (see Fig. 7.21). The spiral arms bring about an inhomogeneous galactic density, which via a gravitational instability causes the ellipses to rotate in lockstep.

7.8 Problems

Problem 7.1 *Solutions of Poisson's Equation*
Show that Poisson's equation (7.1.1) in n-dimensional space outside an n-dimensional isotrophic mass distribution has the following solutions for $U(r)$ $(\mu = GM)$:

$$U(r) \propto \begin{cases} -\frac{\mu}{r^{n-2}} & @\ n \geq 3 \\ -\mu \ln r & @\ n = 2 \\ \mu \cdot r & @\ n = 1 \end{cases}$$

Consider in particular the case at $r = 0$.

Problem 7.2 *Eccentricity from Eccentricity Vector*
Derive from the eccentricity vector $\boldsymbol{e} = (\dot{\boldsymbol{r}} \times \boldsymbol{h})/\mu - \hat{\boldsymbol{r}}$ directly that for the absolute value of the eccentricity holds $1 - e^2 = h^2/\mu a$.

Problem 7.3 *Virial Theorem of a Two-Body System*
(a) Show that in a two-body system for each body with mass m, which orbits the common barycenter thus having the inertial moment $I = mr^2$,

$$\ddot{I} = 4E_{kin} + 2E_{pot}$$

generally holds.
(b) Then prove the virial theorem Eq. (7.3.20), $2\langle E_{kin}\rangle + \langle E_{pot}\rangle = 0$, for a bounded orbit.

Problem 7.4 *Orbit Equation—Fast Track*
(a) Starting out from $r^2 = \boldsymbol{r} \cdot \boldsymbol{r}$ show with Eq. (7.2.5) that $\ddot{r} = h^2/r^3 - \mu/r^2$
(b) Apply the Burdet transformation $\rho := 1/r$ (see Sect. 7.6.1), solve the equivalent equation of motion in ρ, and show that the orbit equation follows

$$r = \frac{p}{1 + e\,\cos\theta}$$

with $p := h^2/\mu$ and $e := p\rho_0$, where $\rho_0 = 1/r_0$ is the initial orbital radius.

Problem 7.5 *Solutions to Barker's Equation*
Show with Cardano's method and Descartes' rule of signs that the unique real solutions to Barker's equation from Sect. 7.4.4 are

$$\tan\frac{\theta}{2} = \left(\sqrt{q^2+1}+q\right)^{1/3} - \left(\sqrt{q^2+1}-q\right)^{1/3}$$

and

$$2\frac{r}{p} = \left(\sqrt{q^2+1}+q\right)^{2/3} + \left(\sqrt{q^2+1}-q\right)^{2/3} - 1$$

with

$$q = 3\sqrt{\frac{\mu}{p^3}}(t - t_0)$$

Problem 7.6 *Generalized Solution of Kepler's Problem*
Prove that by the application of the transformation Eqs. (7.4.14) quite generally the following relation holds

$$\int \frac{d\theta}{(1+e\cdot\cos\theta)^n} = \frac{1}{\sqrt{(1-e^2)^{2n-1}}}\int (1-e\cdot\cos E)^{n-1}dE \quad @\ e<1, n\geq 1$$

Problem 7.7 *Series Expansions*
(a) Prove the series expansions (cf. Danby (2003), Sect. 6.14)

$$E = M + e\sin M + \frac{e^2}{2}\sin(2M) - \frac{e^3}{8}[\sin M - 3\sin(3M)] + O(e^4)$$
$$\frac{r}{a} = 1 - e\cos M + e^2\sin^2 M + \frac{3}{2}e^3\cos M\sin^2 M + O(e^4)$$

for an elliptic orbit by applying the Banach fixed point theorem to $E = M + e\sin E = f(E)$ under the constraint that f is Lipschitz continuous for $e<0.6627434\ldots$. Then apply the result to $r/a = 1 - e\cos E$.

Remark *This solution procedure may sound like elementary mathematics. In fact the solution algorithm, called contraction mapping, is only a generalization of Newton's method. Just the verification that it works is elementary mathematics.*

Note *Contraction mapping is a very convenient method to solve implicit functional relations if the function is Lipschitz continuous. Practically, Lipschitz continuity is not checked beforehand, but contraction mapping is just applied and only then observed whether the series converges. In fact, we made use of the contraction mapping without saying when deriving* Eq. (12.5.15).

(b) Prove that for an elliptic orbit

$$\theta = M + 2e\sin M + \frac{5}{4}e^2\sin(2M) + O(e^3)$$

Hint: Apply contraction mapping to the integral equation (7.3.9)

$$\int_0^\theta \frac{dx}{(1+e\cdot\cos x)^2} = \frac{\mu^2}{h^3}(t-t_p) = \frac{M}{(1-e^2)^{3/2}}$$

Note that alternatively holds

$$\cos\theta(t) = \cos M(t) + e(\cos 2M - 1) + O\left(e^2\right)$$

(c) By the same token prove that for the true anomaly θ' (angle between the apsidal line to apogee and the radius vector from the empty focal point to the revolving body) holds

$$\theta' = M - \frac{1}{4}e^2\sin(2M) + O\left(e^3\right)$$

or alternatively

$$\cos\theta'(t) = \cos M(t) + \frac{1}{8}e^2(\cos M - \cos 3M) + O\left(e^2\right)$$

Problem 7.8 *Radial Position from the Equation of Radial Motion*
Prove that both, Eq. (7.4.17) for the radial position in time of an elliptic orbit and Eq. (7.4.33) for that of a hyperbolic orbit, and Barker's equation for a parabolic orbit, follows from Eq. (7.2.11) by direct integration.

Hint: Show first that (7.2.11) can be rewritten as

$$\begin{array}{lll} r\cdot\dot{r} = na^2e\sqrt{1-\rho^2}, & \rho = \frac{r-a}{ae} & @\text{ ellipse} \\ r\cdot\dot{r} = na^2e\sqrt{\rho^2-1}, & \rho = \frac{r-a}{-ae} & @\text{ hyperbola} \\ r\cdot\dot{r} = \sqrt{\mu p}\sqrt{2\rho-1}, & \rho = \frac{r}{p} & @\text{ parabola} \end{array}$$

Problem 7.9 *Near-Radial Parabolic Orbits*
(a) Show that from the orbit equation for a radial elliptic orbit with initial conditions r_0, v_0, t_0

$$(t-t_0)\sqrt{\frac{\mu}{a^3}} = \frac{r}{a}\sqrt{\frac{2a}{r}-1} - \frac{r_0}{a}\sqrt{\frac{2a}{r_0}-1} - 2\arcsin\sqrt{\frac{r}{2a}} + 2\arcsin\sqrt{\frac{r_0}{2a}}$$

for $a \to \infty$ and $v_0 = 0$ follows the orbit equation for the radial parabolic orbit

$$r^{3/2}(t) = r_0^{3/2} - \sqrt{\frac{9\mu}{2}}\cdot(t-t_0)$$

(b) Show that for a radial elliptic orbit, which is nearly radial parabolic, $a \gg r_0$ and $v_0 = 0$, one gets

$$r^{3/2}\left[1 + \frac{3}{20}\frac{r}{a}\right] \approx r_0^{3/2}\left[1 + \frac{3}{20}\frac{r_0}{a}\right] - \sqrt{\frac{9\mu}{2}}(t-t_0)$$

Because for $t - t_0 \ll \sqrt{\mu/r_0^3}$ from the radial parabolic orbit equation follows

$$\frac{r}{a} \approx \frac{r_0}{a}\left[1 - (t - t_0)\sqrt{\frac{2\mu}{r_0^3}}\ \right]$$

show that

$$\begin{aligned} r^{3/2}(t) &\approx r_0^{3/2} - \sqrt{\frac{9\mu}{2}}(t - t_0)\left[1 - \frac{r_0}{4a}\left(1 - \frac{12}{10}(t - t_0)\sqrt{\frac{\mu}{2r_0^3}}\right)\right] \\ &\approx r_0^{3/2} - \sqrt{\frac{9\mu}{2}}(t - t_0)\left[1 - \frac{r_0}{4a}\right] \end{aligned}$$