Lecture notes of SPACE MECHANICS

based on the notes for the SPACECRAFT ORBITAL DYNAMICS AND CONTROL course of the University of Bologna written by Dr. Elisa Maria Alessi and Dr. Dario Modenini

Academic Year 2018–2019
Lecturers: Prof. Giacomo Tommei and Dr. Stefano Marò

Contents

The Two-Body Problem 1

The *Two–Body Problem* is the easiest model that can be adopted for describing the motion of two bodies subject to the mutual gravitational attraction. Its assumptions and derivations apply both for natural and artificial bodies, for instance, to study the motion of a planet around the Sun or a spacecraft around a planet. This dynamical model represents the only case in astrodynamics which admits a general solution, as there exist at least sixth first integrals. Moreover, though it does not comprehend all the forces playing a role, it still provides a good initial approximation for more precise analysis, because the other effects are some orders of magnitude lower on condition to be far enough to any other massive bodies.

Here, we will consider the *Kepler's Problem* or *Restricted Two–Body Problem*, whose main hypothesis is that one of the two bodies has a mass so small that it does not affect the motion of the other body, which therefore is at rest. An example of this situation is represented by a satellite ($m \approx 10^3$ kg) orbiting around the Earth ($m \approx 10^{24}$ kg), which is assumed spherical with a radially symmetrical internal distribution of mass.

1.1 Equation of Motion

Let us consider two isolated point masses m and M, respectively, such that $m \ll M$. The only force acting on the system is the mutual gravitational attraction between the two bodies: m moves due to the gravitational force exerted by M, which is at rest.

To study the problem, let us consider an inertial reference system $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$, whose origin is fixed where M is located (see Fig. 1.1). The position and the velocity of the mass m in this framework are $\mathbf{r}$ and $\mathbf{v}$. By applying the Newton's Second Law of Motion, namely,

$$\mathbf{F} = m_I \frac{d^2\mathbf{r}}{dt^2}, \tag{1.1}$$

to the gravitational attraction due to M we obtain

$$m_I \frac{d^2\mathbf{r}}{dt^2} = -\frac{GMm_G}{r^2}\hat{\mathbf{r}}, \tag{1.2}$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$ is the radial unit vector, $G = 6.67259 \times 10^{-1}$ $\mathrm{m^3 kg^{-1} s^{-2}}$ is the Gravitational Constant and m_I and m_G are the inertial and gravitational mass associated with the second body, respectively. Following the Equivalence Principle according to

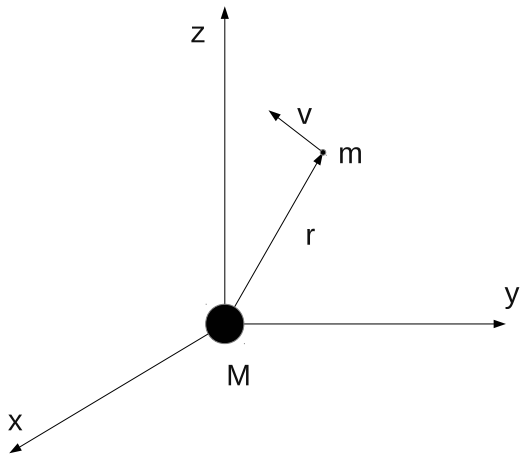


Figure 1.1: The Restricted Two–Body Problem in an inertial reference system centered at the body with the most significant mass.

$m_I = m_G$, we have

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\hat{\mathbf{r}}, \tag{1.3}$$

where $\mu = GM$ is the so-called *gravitational parameter*. In the case of the Earth $\mu_\oplus \approx 398604.3\ \mathrm{km}^3/\mathrm{s}^2$.

We notice that to assume $m_I = m_G = m$ implies that the acceleration acting on a body moving in a gravitational field does not depend on its mass: this is what Galileo showed with the well-known experiment of the falling bodies from the leaning tower in Pisa. Moreover, from Eq. (1.3) we see that the acceleration is always aligned with the radial direction, that is, the gravitational field is a central field. The gravitational attraction exerted by the Earth on a body on the terrestrial surface is

$$g = \frac{\mu}{R_\oplus^2} \approx 9.8\ \mathrm{m/s}^2, \tag{1.4}$$

being $R_\oplus \approx 6378$ km the equatorial radius of the Earth.

1.2 First Integrals

The equation of motion (1.3) is a non linear differential system of the second order. It follows that, to obtain a general solution and thus be able to describe the motion of the second body, we need 6 boundary conditions. If these are given at the initial time t_0, we deal with a Cauchy problem which admits a unique solution. In particular, we can

define the Cauchy problem as

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\hat{\mathbf{r}}, \qquad \mathbf{r}(t_0) = \mathbf{r}_0, \qquad \mathbf{v}(t_0) = \mathbf{v}_0. \tag{1.5}$$

If instead we have some boundary conditions at time t_0 and some at time t_1, we speak of Lambert problem or two-point boundary value problem. For instance, we can have

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\hat{\mathbf{r}}, \qquad \mathbf{r}(t_0) = \mathbf{r}_0, \qquad \mathbf{r}(t_1) = \mathbf{r}_1. \tag{1.6}$$

In this case, a general solution may not exist nor be unique. To show this, let us consider the linear differential equation

$$m\dot{x} = -kx,$$

which describes a harmonic motion with frequency $\omega = \sqrt{k/m}$. The boundary conditions are required to define the phase and the amplitude of the motion. If they are of the kind

$$x(t_0) = x_0, \qquad x(t_0 + T) = x_1,$$

where $T = 2\pi/\omega$ is the period of the motion, then the problem results to be either impossible or undetermined. The first case occurs if $x_1 \neq x_0$, because after one period the variable must take the same value. If instead $x_1 = x_0$ we deal with two equivalent conditions.

Let us now go back to the Kepler's problem written in (1.5) and let us look for 6 first integrals in order to solve it.

1.2.1 The Specific Angular Momentum

If we cross multiply the differential equation (1.3) by $\mathbf{r}$, we get

$$\mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\mathbf{r} \times \hat{\mathbf{r}} = 0, \tag{1.7}$$

because $\mathbf{r}$ and $\hat{\mathbf{r}}$ are aligned on the same direction. It follows that also the acceleration vector $\frac{d^2\mathbf{r}}{dt^2}$ is directed along the radius vector $\mathbf{r}$, that is, we deal with a central field.

Now, let us sum on the left-hand side of (1.7) the null term $\frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt}$, namely,

$$\mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} + \frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} = \frac{d}{dt}\left(\mathbf{r} \times \frac{d\mathbf{r}}{dt}\right) = 0,$$

which integrated gives

$$\mathbf{r} \times \frac{d\mathbf{r}}{dt} = \mathbf{h}, \tag{1.8}$$

where $\mathbf{h}$ is a constant (in modulus and direction) vector, called *specific angular momentum*, that is, the angular momentum per unit of mass. As a consequence, the Keplerian motion is a planar problem in the sense that the motion takes plane on the invariant plane defined by $\mathbf{r} \cdot \mathbf{h} = 0$. This plane contains the central body and is perpendicular to the vector $\mathbf{h}$. Clearly, the radius and the velocity vectors are always orthogonal to the direction of $\mathbf{h}$.

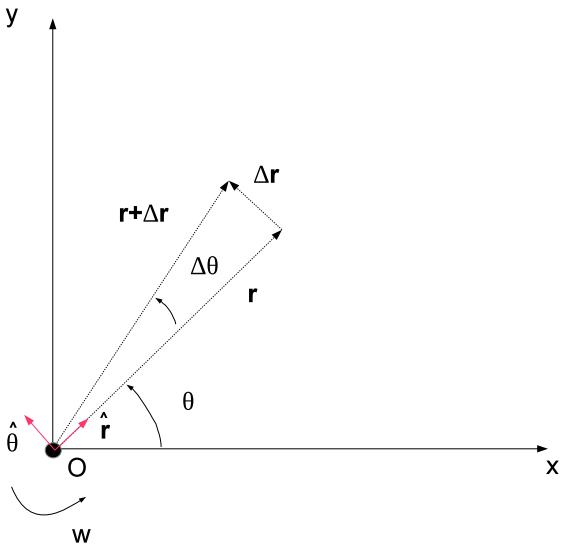


Figure 1.2: The planar reference system and the polar coordinates corresponding to the Two–Body Problem.

Polar Coordinates

Let us consider the polar reference system displayed in Fig. 1.2, where the position of the small body is defined by the two coordinates (r, θ). This is a non inertial reference frame, defined on the plane of the Keplerian motion in such a way that the pole coincides with the position of the central body and the polar axis with one of the semi-axes of the inertial reference system introduced before (see Fig. 1.1). The unit vectors of the system are one, say $\hat{\mathbf{r}}$, directed along $\mathbf{r}$ and the other, say $\hat{\boldsymbol{\theta}}$, perpendicular to $\hat{\mathbf{r}}$ and aligned along the direction of increasing θ. They can be expressed as

$$\hat{\mathbf{r}} = \cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}}, \qquad \hat{\boldsymbol{\theta}} = -\sin\theta\hat{\mathbf{i}} + \cos\theta\hat{\mathbf{j}}. \tag{1.9}$$

Since they are constant in modulus and the polar reference system rotates with respect to the inertial one with angular velocity $\mathbf{w}$ directed along $\mathbf{h}$ but with modulus $w = \dot{\theta}$ not necessarily constant, the corresponding time derivatives are

$$\frac{d\hat{\mathbf{r}}}{dt} = \mathbf{w} \times \hat{\mathbf{r}} = \dot{\theta}\hat{\boldsymbol{\theta}}, \tag{1.10}$$

and

$$\frac{d\hat{\boldsymbol{\theta}}}{dt} = \mathbf{w} \times \hat{\boldsymbol{\theta}} = -\dot{\theta}\hat{\mathbf{r}}. \tag{1.11}$$

The radius and velocity vectors in the inertial plane are

$$\mathbf{r} = r\cos\theta\hat{\mathbf{i}} + r\sin\theta\hat{\mathbf{j}}, \qquad \frac{d\mathbf{r}}{dt} \equiv \dot{\mathbf{r}} = \dot{r}(\cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}}) + r\dot{\theta}(-\sin\theta\hat{\mathbf{i}} + \cos\theta\hat{\mathbf{j}}), \quad (1.12)$$

which can be written in the polar frame as

$$\mathbf{r} = r\hat{\mathbf{r}}, \qquad \dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}. \quad (1.13)$$

We notice that the velocity vector is decomposed in the *radial* and *transversal* components, namely,

$$v_r = \dot{r}, \qquad v_\theta = r\dot{\theta}. \quad (1.14)$$

The modulus of the velocity vector is

$$v = \sqrt{v_r^2 + v_\theta^2}. \quad (1.15)$$

Now, by means of Eqs. (1.8) and (1.13) we get to

$$\mathbf{h} = r\hat{\mathbf{r}} \times \left(\dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}}\right) = r^2\dot{\theta}\left(\hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}}\right),$$

that is,

$$h = r^2\dot{\theta}. \quad (1.16)$$

The acceleration vector is computed by deriving the velocity vector and reads

$$\frac{d^2\mathbf{r}}{dt^2} \equiv \ddot{\mathbf{r}} = \ddot{r}\hat{\mathbf{r}} + \dot{r}\frac{d\hat{\mathbf{r}}}{dt} + (\dot{r}\dot{\theta} + \dot{r}\ddot{\theta})\hat{\boldsymbol{\theta}} + r\dot{\theta}\frac{d\hat{\boldsymbol{\theta}}}{dt} = (\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}}. \quad (1.17)$$

The radial and transversal components in this case are

$$a_r = \ddot{r} - r\dot{\theta}^2, \qquad a_\theta = 2\dot{r}\dot{\theta} + r\ddot{\theta}. \quad (1.18)$$

As we have seen before, the Kepler's problem is a central problem and thus, using Eq. (1.18), the equation of motion (1.3) in polar coordinates becomes

$$\begin{cases} \ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2}, \\ 2\dot{r}\dot{\theta} + r\ddot{\theta} = 0. \end{cases} \quad (1.19)$$

1.2.2 The Eccentricity Vector

Now let us cross multiply Eq. (1.3) by $\mathbf{h}$ and sum the null term $\dot{\mathbf{r}} \times \dot{\mathbf{h}}$ on the left-hand side:

$$\ddot{\mathbf{r}} \times \mathbf{h} + \dot{\mathbf{r}} \times \dot{\mathbf{h}} = -\frac{\mu}{r^2}\hat{\mathbf{r}} \times \mathbf{h},$$

that is, using (1.10) and (1.16),

$$\frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) = \frac{\mu h}{r^2}\hat{\boldsymbol{\theta}} = \mu\dot{\theta}\hat{\boldsymbol{\theta}} = \mu\frac{d\hat{\mathbf{r}}}{dt} = \mu\frac{d}{dt}\left(\frac{\mathbf{r}}{r}\right).$$

By integration of this expression, we obtain

$$\dot{\mathbf{r}} \times \mathbf{h} - \mu \frac{\mathbf{r}}{r} = \mu \mathbf{e}, \tag{1.20}$$

where $\mathbf{e}$ is a constant non dimensional vector, called *Laplace vector* or *Runge–Lenz vector* or *eccentricity vector*.

Having now two first integrals, $\mathbf{h}$ and $\mathbf{e}$, both three-dimensional vectors, we may think to be able to solve the Kepler's problem. Actually, $\mathbf{h}$ and $\mathbf{e}$ are not independent, because they are orthogonal and in particular satisfy

$$\mathbf{h} \cdot \mathbf{e} = 0. \tag{1.21}$$

Therefore they provide only 5 independent integration parameters.

We notice that the angular momentum $\mathbf{h}$ defines the orbital plane where $\mathbf{e}$ lies. We will see that $\mathbf{e}$ defines instead the geometry of the orbit in the plane, in particular its shape (circular, elliptic, ..), direction (through the line of apsides) and the pericenter direction.

1.2.3 The Specific Mechanical Energy

Let us dot multiply Eq. (1.3) by $\dot{\mathbf{r}}$:

$$\ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} = -\frac{\mu}{r^2} \hat{\mathbf{r}} \cdot \dot{\mathbf{r}},$$

that is,

$$\frac{d}{dt}\left(\frac{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}}{2}\right) + \frac{\mu}{r^3}\frac{d}{dt}\left(\frac{\mathbf{r} \cdot \mathbf{r}}{2}\right) = 0.$$

By integrating this expression, using the properties of the scalar product and recalling that $||\dot{\mathbf{r}}|| = v$, we get to

$$\frac{v^2}{2} - \frac{\mu}{r} = \mathcal{E}, \tag{1.22}$$

where $\mathcal{E}$ is another constant of the motion and it represents the specific mechanical energy of the moving body, that is, the energy per unit of mass. Looking to Eq. (1.22), it is clear that the first term is related to the kinetic energy, the second to the potential one.

1.3 The Polar Equation for the Orbit

Let us dot multiply the eccentricity vector by $\mathbf{r}$:

$$\mathbf{r} \cdot \mathbf{e} = \mathbf{r} \cdot \left(\dot{\mathbf{r}} \times \frac{\mathbf{h}}{\mu}\right) - \mathbf{r} \cdot \frac{\mathbf{r}}{r},$$

and recall that in general for the scalar triple product we have $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$. By using the properties of the scalar product, we find

$$\begin{aligned} \mathbf{r} \cdot \mathbf{e} + \mathbf{r} \cdot \frac{\mathbf{r}}{r} &= (\mathbf{r} \times \dot{\mathbf{r}}) \cdot \frac{\mathbf{h}}{\mu}, \\ \implies \quad re\cos\theta + r &= \mathbf{h} \cdot \frac{\mathbf{h}}{\mu}, \\ \implies \quad r(e\cos\theta + 1) &= \frac{h^2}{\mu}, \end{aligned} \tag{1.23}$$

where e is the *eccentricity* and θ is the angle going from the eccentricity vector to the radius vector and it is called *true anomaly*.

From Eq. (1.23), we obtain the *polar equation of the orbit*, namely,

$$r = \frac{p}{1 + e\cos\theta}, \tag{1.24}$$

where $p = h^2/\mu$ is a geometrical constant of the orbit, called *parameter* or *semi-latus rectum*. We notice that the true anomaly is one of the polar coordinates introduced before and it is related to the time.

Eq. (1.24) represents the equation of a conic section, which means that an object moving under the hypotheses of the Two–Body Problem can only orbit on a curve of this kind.

A conic section is, by definition, the locus of a point which moves in a such a way that the ratio of its distance from a given point, called *focus*, and its distance from a given line, called *directrix*, is a positive constant. This constant is, in particular, the eccentricity of the conic. To prove that, let us refer to Fig. 1.3 and call for the moment the constant $K = r/d = r_1/d_1$. We have

$$\begin{aligned} & r_1 + d_1 = r\cos\theta + d, \\ \implies \quad & d_1\left(\frac{r_1}{d_1} + 1\right) = d\left(\frac{r\cos\theta}{d} + 1\right), \\ \implies \quad & d_1(K+1) = d(K\cos\theta + 1), \\ \implies \quad & \frac{d}{d_1} = \frac{1+K}{1+K\cos\theta}, \\ \implies \quad & \frac{r}{r_1} = \frac{1+K}{1+K\cos\theta}, \\ \implies \quad & r = \frac{r_1(1+K)}{1+K\cos\theta}. \end{aligned}$$

Because of the way we have defined r_1, the associated true anomaly θ_1 is 0 and thus, using (1.24), the expression just derived becomes

$$r = \left(\frac{p}{1+e\cos\theta_1}\right)\left(\frac{(1+K)}{1+K\cos\theta}\right) = \left(\frac{p}{1+e}\right)\left(\frac{(1+K)}{1+K\cos\theta}\right).$$

From (1.24) we have

$$r = \frac{p}{1+e\cos\theta} = \left(\frac{p}{1+e}\right)\left(\frac{(1+K)}{1+K\cos\theta}\right),$$

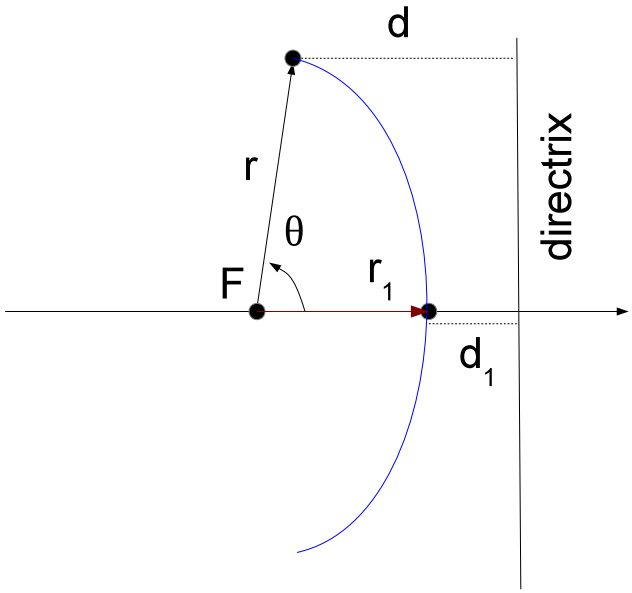


Figure 1.3: The definition of a conic section.

which is satisfied if and only if $K = e$.

According to the value assumed by e, we can distinguish among different kinds of conic section. We deal with

- a circle if $e = 0$;
- an ellipse if $0 < e < 1$;
- a parabola if $e = 1$;
- an hyperbola if $e > 1$.

As we will see, in the first two cases the specific energy of the orbiting body is negative, which means that it cannot escape from the gravitational field of the central body. In the second case, the specific energy is zero, and in the last one is positive.

Historically, there exists another definition for a conic section, that is, as the curve of intersection of a plane with a right circular cone (see Fig. 1.4). We have:

- an ellipse if the intersection takes place on one half of the cone;
- a circle if the intersection takes place on one half of the cone and the plane is parallel to the base of the cone;

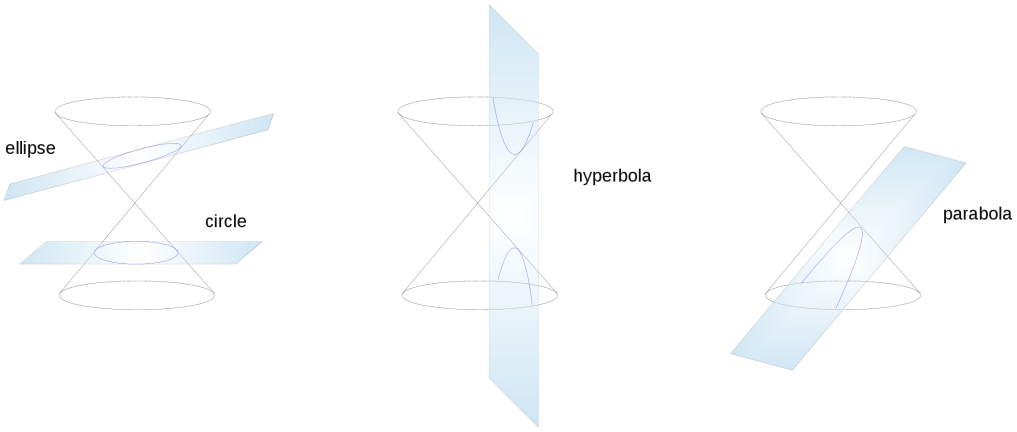


Figure 1.4: The definition of a conic section as the curve of intersection of a plane with a right circular cone.

- a parabola if the intersection takes place on one half of the cone and the plane is parallel to a line in the surface of the cone;
- a hyperbola if the plane cuts both halves of the cone.

There is also the possibility to deal with degenerate conics, if the intersection is a point and one or two straight lines.
Each type of conic section has two foci, say F and F^*. The first one F has a physical meaning, in the sense that it is where the attracting body is located. The second one F^* has just a geometrical meaning and in Sec. 6.2 we will see how to identify it, in the case of the ellipse. For the parabola, it is at an infinite distance with respect to F, while in the case of the hyperbola we have two branches, each of them corresponding to one different focus.

1.3.1 Elliptical Orbit

In Cartesian coordinates the equation of a generic ellipse with respect to the origin reads

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{1.25}$$

where a and b are the *semi-major axis* and *semi-minor axis*, respectively. In general, the length of the chord passing through the two foci is the major axis $2a$ of a conic section. The line passing through the foci is called *line of the apsides* and the points of intersection between this line and the ellipse are said *pericenter* and *apocenter*. The

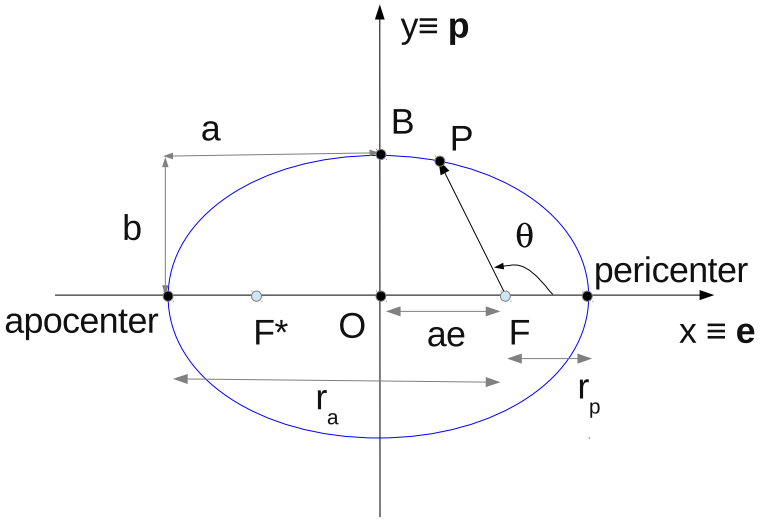


Figure 1.5: Semi-major axis, semi-minor axis and foci of an ellipse.

former is the closest point of the orbit to the focus F, the latter the farthest. In Tab. 1.1, we specify these distances by means of (1.24) and Fig. 1.5. They can be used to compute the value of the eccentricity, namely,

$$p = r_p(1+e) = r_a(1-e), \qquad \Longrightarrow \qquad e(r_p + r_a) = r_a - r_p,$$

that is,

$$e = \frac{r_a - r_p}{r_p + r_a}. \tag{1.26}$$

From Fig. 1.5, we have that the semi-major axis is given by

$$a = \frac{r_p + r_a}{2}. \tag{1.27}$$

From Tab. 1.1 and Eqs. (1.26)–(1.27) it follows

$$r_a = a(1+e), \qquad r_p = a(1-e). \tag{1.28}$$

Concerning the semi-focal distance $f = OF = OF^*$, we have

$$OF = a - r_p = \frac{r_a + r_p}{2} - r_p = \frac{r_a - r_p}{2} = \left(\frac{r_a - r_p}{r_a + r_p}\right)\left(\frac{r_a + r_p}{2}\right) = ea,$$

that is,

$$f = ea. \tag{1.29}$$

Definition	θ	r
Pericenter Radius r_p	0	$\frac{p}{1+e}$
Apocenter Radius r_a	π	$\frac{p}{1-e}$
Semi-latus Rectum	$\pi/2$	p

Table 1.1: Absolute value of the radius vector and true anomaly associated with the pericenter, apocenter and semi-latus rectum for an ellipse.

To compute the semi-minor axis b we recall that an ellipse can also be defined as the locus of a point such that the sum of the distances between the point and the foci is constant, i.e.,

$$PF + PF^* = constant, \qquad \Longrightarrow r_p + (r_p + 2f) = 2r_p + 2f = 2a(1-e) + 2ae = 2a.$$

This means that any point Q lying on the ellipse fulfills the condition

$$QF + QF^* = 2a. \tag{1.30}$$

In particular, if we consider the point B such that $b = OB$, then we also have (see Fig. 1.5) that $BF = BF^*$ and thus $BF = BF^* = a$ because of (1.30). Thus it turns out that

$$b = OB = \sqrt{BF^2 - OF^2} = \sqrt{a^2 - a^2e^2} = a\sqrt{1-e^2}, \tag{1.31}$$

which in modulus is valid for any type of conic section, except the parabola. From Tab. 1.1 and Eq. (1.28) we can derive the semi-latus rectum p as

$$p = a(1-e^2), \tag{1.32}$$

which is a valid expression for any type of conic section, except the parabola.

1.4 The Laplace Reference System

Let us define the *Laplace reference system* $\{\mathbf{e}, \mathbf{p}, \mathbf{h}\}$ in the following way:

- the $x-$axis corresponds to the direction defined by the eccentricity vector;
- the $y-$axis is obtained by rotating the $x-$axis counterclockwise by an angle equal to $\pi/2$ (see Fig. 1.5);
- the $z-$axis is directed along the angular momentum vector $\mathbf{h}$.

If we rewrite Eq. (1.20) as

$$\mathbf{e} + \hat{\mathbf{r}} = -\frac{1}{\mu}\left(\mathbf{h} \times \dot{\mathbf{r}}\right),$$

and cross multiply it by $\mathbf{h}$, namely,

$$\mathbf{h} \times (\mathbf{e} + \hat{\mathbf{r}}) = -\frac{1}{\mu}\mathbf{h} \times (\mathbf{h} \times \dot{\mathbf{r}}),$$

using the properties of the triple cross product $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}$ and the fact that $\mathbf{h}$ and $\dot{\mathbf{r}}$ are orthogonal, we obtain

$$\mathbf{h} \times (\mathbf{e} + \hat{\mathbf{r}}) = -\frac{1}{\mu}\left[(\mathbf{h} \cdot \dot{\mathbf{r}})\,\mathbf{h} - h^2\dot{\mathbf{r}}\right] = \frac{h^2}{\mu}\dot{\mathbf{r}},$$

that is,

$$\dot{\mathbf{r}} = \frac{\mu}{h}\hat{\mathbf{h}} \times (\mathbf{e} + \hat{\mathbf{r}}). \tag{1.33}$$

Since

$$\hat{\mathbf{h}} \times \hat{\mathbf{e}} = \hat{\mathbf{p}}, \qquad \hat{\mathbf{h}} \times \hat{\mathbf{r}} = \hat{\boldsymbol{\theta}}, \tag{1.34}$$

we get to

$$\dot{\mathbf{r}} = \frac{\mu}{h}\left(e\hat{\mathbf{p}} + \hat{\boldsymbol{\theta}}\right). \tag{1.35}$$

The above relationship tells us that the velocity of a small body moving on an ellipse can be decomposed in two components, both constant in modulus, namely,

- a vector of modulus $\mu e/h$ along the orthogonal direction corresponding to the semi-major axis and towards increasing true anomaly;
- a vector of modulus μ/h along the orthogonal direction to the radius vector in the orbital plane.

In the particular cases of the pericenter and the apocenter, the two units vectors $\hat{\mathbf{p}}$ and $\hat{\boldsymbol{\theta}}$ are along the same direction and there does not exist a radial component of the velocity. Therefore,

$$v_p = \frac{\mu}{h}(1+e) = \sqrt{\frac{\mu}{a}\left(\frac{1+e}{1-e}\right)}, \qquad v_a = \frac{\mu}{h}(1-e)\sqrt{\frac{\mu}{a}\left(\frac{1-e}{1+e}\right)}, \tag{1.36}$$

being the speed at the pericenter the maximum possible, the one at the apocenter the minimum.

1.5 Specific Energy and Semi-major Axis

We can use the just derived expressions for the radius and speed at pericenter and apocenter to obtain a different expression for the energy associated with the orbit. This can be computed at any point, as it is a constant of motion. Let us choose the pericenter for which Eq. (1.22) reads

$$\begin{aligned}
\mathcal{E} &= \frac{v_p^2}{2} - \frac{\mu}{r_p} = \frac{\mu^2}{2h^2}(1+e)^2 - \frac{\mu}{a(1-e)} = \frac{\mu^2}{2p\mu}(1+e)^2 - \frac{\mu}{a(1-e)} = \\
&= \frac{\mu}{2a(1-e^2)}(1+e)^2 - \frac{\mu}{a(1-e)} = \frac{\mu}{2a(1-e)}(1+e) - \frac{2\mu}{2a(1-e)} = \\
&= -\frac{\mu}{2a}.
\end{aligned} \tag{1.37}$$

In other words, the energy depends only on the semi-major axis of the orbit.

In an analogous way, we can find the relationship between the energy and the eccentricity of the orbit, namely,

$$\begin{aligned}\mathcal{E} &= \frac{v_p^2}{2} - \frac{\mu}{r_p} = \frac{\mu^2}{2h^2}(1+e)^2 - \frac{\mu}{a(1-e)} = \frac{\mu^2}{2h^2}(1+e)^2 - \frac{\mu}{p}(1+e) = \\ &= \frac{\mu^2}{2h^2}(1+e)^2 - \frac{\mu^2}{h^2}(1+e) = \frac{\mu^2}{2h^2}\left(1+e^2+2e-2-2e\right) = \\ &= -\frac{\mu^2}{2h^2}(1-e^2), \end{aligned} \tag{1.38}$$

that is,

$$e = \sqrt{1 + \frac{2\mathcal{E}h^2}{\mu^2}}. \tag{1.39}$$

Because of that, the energy integral does not provide any further information with respect to the other first integrals found, namely the angular momentum $\mathbf{h}$ and the eccentricity vector $\mathbf{e}$.

1.6 The Other Conic Sections

Let us take advantage of (1.37) to understand how to discriminate among the different conic sections, that is, when the satellite will move on one or another. On an elliptic orbit, the semi-major axis is positive and therefore the energy is negative. For the parabola, the semi-major axis takes value infinity by definition and, as a consequence, the energy is zero: in some sense we can think to an ellipse whose semi-major axis tends to increase indefinitely. In the hyperbolic case, the semi-major axis is negative and assumes only a geometrical meaning, not a physical one, and thus the corresponding energy is positive. From a physical point of view, this distinction can be made by thinking whether it is necessary to provide energy to the spacecraft to make it escape from the gravitational sink of the Earth. In the case of more attracting bodies, the various gravitational sinks sum up and the energy required by the spacecraft to move from an orbit of semi-major axis a_1 to another of semi-major axis a_2 depends on all the gravitational fields it experiences.

1.6.1 Circular Orbit

The circular orbit is a special case of the elliptical one, in the sense that $e = 0$ and the velocity vector is always perpendicular to the radius vector, i.e., we do not have a radial component. Also, the semi-major axis a is equal to r, which is constant in time, and the specific mechanical energy is

$$\frac{v_{circ}^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2r}, \tag{1.40}$$

that is, the modulus of the velocity is constant as well

$$v_{circ} = \sqrt{\frac{2\mu}{r} - \frac{\mu}{r}} = \sqrt{\frac{\mu}{r}}, \tag{1.41}$$

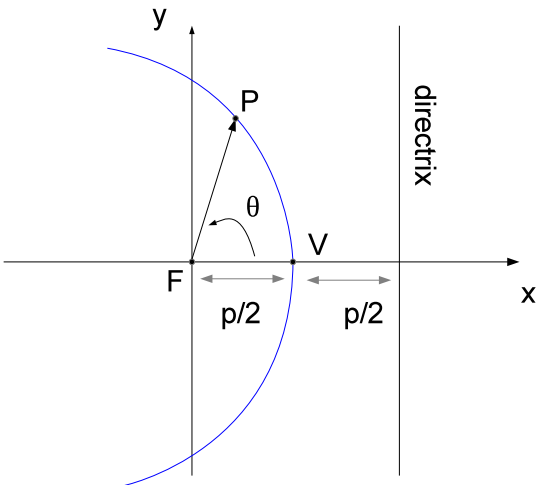


Figure 1.6: The parabola.

and it is called *first cosmic velocity*.

The larger the distance to the central body, the lower the velocity required to move on a circular orbit. The maximum value takes place (ideally) on zero-altitude orbit and for the case of the Earth is

$$v^*_{circ} = \sqrt{\frac{\mu_\oplus}{R_\oplus}} \approx 7.9 \frac{km}{s}. \tag{1.42}$$

1.6.2 Parabolic Orbit

The parabolic orbit is characterized by $e = 1$ and $\mathcal{E} = 0$. This means that the small body approaches the central one from infinity to go then toward infinity as well. The semi-major is by definition equal to ∞, the focus coincides with the origin and the parameter p is the distance between the focus and the directrix (see Fig. 1.6). We also have

$$r_p \equiv r(\theta = 0) = \frac{p}{2}. \tag{1.43}$$

In Cartesian coordinates the equation of the parabola is

$$y^2 = 2p\left(\frac{1}{2}p - x\right). \tag{1.44}$$

Again exploiting the energy equation (1.37), we can compute the velocity on a parabolic orbit for a given radius vector, namely,

$$v_{par} = \sqrt{\frac{2\mu}{r} - \cancelto{0}{\frac{\mu}{a}}} = \sqrt{\frac{2\mu}{r}} = \sqrt{2} v_{circ}, \tag{1.45}$$

where v_{circ} is the velocity on a circular orbit of radius r.

If the velocity is evaluated at the pericenter of the parabola, we speak about *second cosmic velocity* or *escape velocity*, because it represents the minimum velocity required by a small body to move off from the central gravitational field. When the effect of the gravitational field considered becomes negligible with respect to other forces, we say that the body *has approached infinity*: the escape velocity allows to get there. For satellites orbiting around the Earth, this distance corresponds to about 10^6 km and the escape velocity is maximum on the surface of the Earth and its value is

$$v^*_{par} = \sqrt{2} v^*_{circ} \approx 11.2 \frac{km}{s}. \tag{1.46}$$

At infinity the residual velocity on the parabola is zero. Indeed, let us consider the case $\mathbf{h} = 0$. This configuration occurs when $\dot{\mathbf{r}}$ is parallel to $\mathbf{r}$: these are the straight line orbits already introduced at the beginning of the section by means of Fig. 1.4. Let us assume to be on the surface of the Earth and to launch a body along the radial direction. According to the velocity given to the body, this will move on a different path. If the velocity is lower than a certain value, then it will reach a maximum altitude and then fall back to the Earth. This is the case of the *rectilinear ellipse*. There exists a well-defined value of the initial velocity such that the maximum altitude is attained at 'infinity': in this case the body will not fall back to the Earth, but stay there at rest. This is the *rectilinear parabola*. If, instead, the velocity is larger than the threshold, then the body will approach infinity with a non-zero velocity and its orbit will resemble the hyperbolic one.

1.6.3 Hyperbolic Orbit

For the hyperbolic we have $e > 1$ and $\mathcal{E} > 0$. From the energy equation (1.37) we get $a < 0$. Let us set $a = -a'$ where $a' = |a|$. We will see that the negative semi-major axis does not have a physical meaning, but its modulus has a well-defined geometrical interpretation. Since $a < 0$ and $e > 1$, the parameter p is positive as we expect, because it is related to the angular momentum $h = \sqrt{\mu p}$. Concerning the semi-minor axis b, since $1 - e^2 < 0$, we have

$$b = ja\sqrt{e^2 - 1} = -ja'\sqrt{e^2 - 1} = -jb', \tag{1.47}$$

where j is the imaginary unit and $b' = a'\sqrt{e^2 - 1} > 0$.

The Cartesian equation of the hyperbola, in a reference system with origin at the center of symmetry, is

$$\frac{x^2}{a'^2} - \frac{y^2}{b'^2} = 1. \tag{1.48}$$

The hyperbola is composed by two different branches. Once we fixed the main focus F, the small body will be able to move only on the closest branch to F, because the gravitational field is attractive and not repulsive (see Fig. 1.7). The pericenter distance is again

$$r_p = \frac{p}{1 + e} = a'(e - 1), \tag{1.49}$$

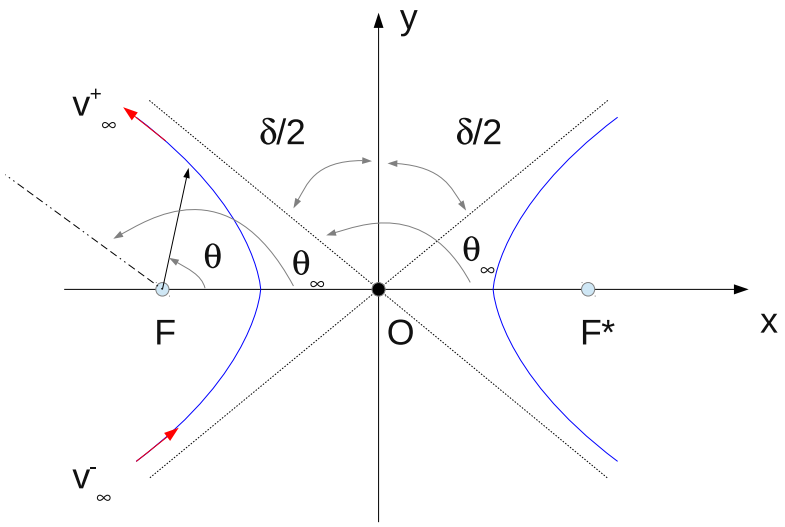


Figure 1.7: The hyperbola.

and the corresponding velocity is derived from the energy equation, namely,

$$v = \sqrt{\frac{2\mu}{r} + \frac{\mu}{a'}} \qquad \Longrightarrow \qquad v_p = \sqrt{\frac{\mu}{a'}\frac{e+1}{e-1}}, \tag{1.50}$$

which is larger than the escape velocity computed at the same altitude.

In the hyperbolic case, the residual velocity at infinity is called *third cosmic velocity* or *hyperbolic excess velocity* and is

$$v_\infty = \lim_{r \longrightarrow \infty} v = \sqrt{\frac{\mu}{a'}}. \tag{1.51}$$

Its direction is defined by the asymptote and corresponds to the true anomaly θ_∞, as shown in Fig. 1.7. In other words, at infinity we have

$$r \longrightarrow \infty \qquad \Longrightarrow \qquad 1 + e\cos\theta_\infty \longrightarrow 0,$$

that is,

$$\cos\theta_\infty = -\frac{1}{e}. \tag{1.52}$$

Moreover, the angle between the two asymptotes δ is called *deflection angle* and we have (see Fig. 1.7)

$$-\frac{1}{e} = \cos\theta_\infty = \cos\left(\frac{\pi}{2} + \frac{\delta}{2}\right) = -\sin\frac{\delta}{2},$$

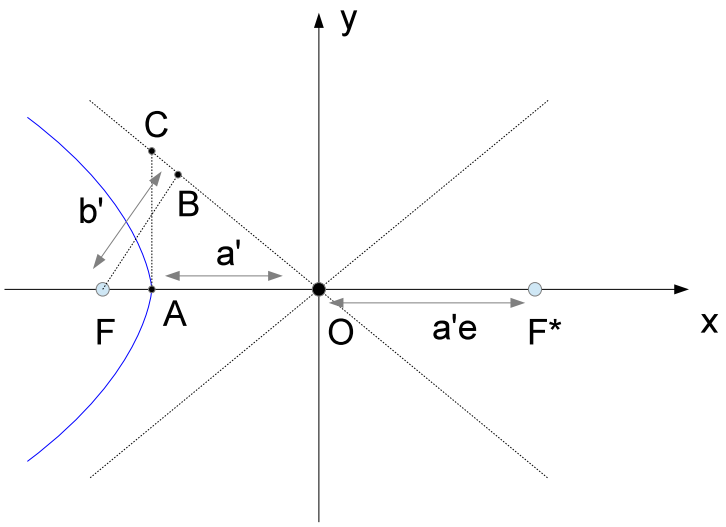


Figure 1.8: The geometrical meaning of the parameter b' for the hyperbola.

that is,

$$\sin\frac{\delta}{2} = \frac{1}{e}. \tag{1.53}$$

Let us refer to Fig. 1.8. The semi-focal distance in the hyperbolic case is

$$f = a'e. \tag{1.54}$$

The geometrical meaning of a' is the distance OA from the pericenter to the point of the intersection of the asymptotes. Indeed, we have

$$f = OF = OA + AF = OA + r_p = OA + a'(e-1) = OA + a'e - a' = ae \quad \Longrightarrow \quad OA = a'.$$

In the same figure we have drawn the segment AC perpendicular to the $x-$axis, i.e., the line joining the two foci, and the one FB perpendicular to one asymptote. The two triangles OAC and OBF are equivalent, because the both have a right angle, they share the angle in O and the two sides OA and OB are equal. Therefore,

$$CA = FB = \sqrt{OF^2 - OB^2} = \sqrt{f^2 - OA^2} = \sqrt{a'^2e^2 - a'^2} = a'\sqrt{e^2-1} = b'.$$

This is also said *impact parameter*.

Analogous to the case of the ellipse, the hyperbola can also be defined as the locus of a point such that the difference of the distances between the point and the foci is constant, i.e.,

$$QF - QF^* = 2a'.$$

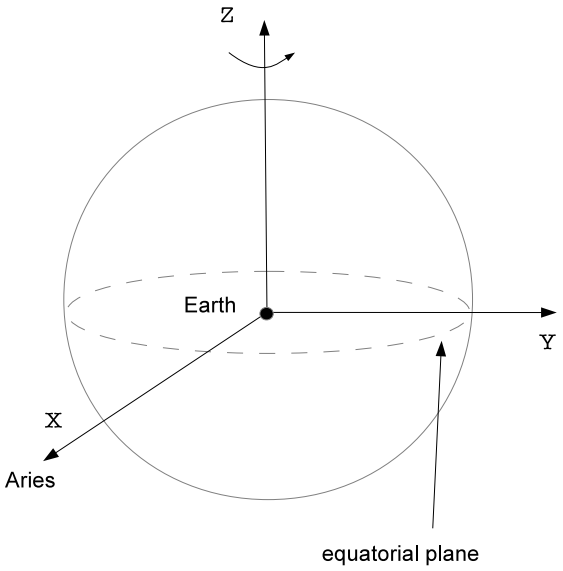


Figure 1.9: The pseudo-inertial reference system for a satellite orbiting around the Earth.

1.7 Orbital Elements

So far we have shown that, if we know the shape of the orbit, the motion of the small body can be described easily with respect to one of the reference systems defined by its orbit. We have proved that the orbital plane is constant, now we have to situate it with respect to a physical framework. This can be done by referring the $\{\mathbf{e}, \mathbf{p}, \mathbf{h}\}$ directions to a *pseudo-inertial* $\{\hat{\mathbf{I}}, \hat{\mathbf{J}}, \hat{\mathbf{K}}\}$ frame.

Let us consider the case of a satellite orbiting around the Earth. What follows will hold in general, for instance also for a planet moving around the Sun, in that case the reference plane will be the ecliptic one (see also Sec. **??**). The pseudo-inertial reference system can be set in this way (see Fig. 1.9):

- the origin is at the center of the Earth;
- the $X-$axis is placed in the Earth's equatorial plane and points toward the First Point of Aries Υ;
- the $Z-$axis is aligned along the Earth' spin axis and thus it is orthogonal to the $X-$axis;
- the $Y-$axis completes the right-handed coordinate system.

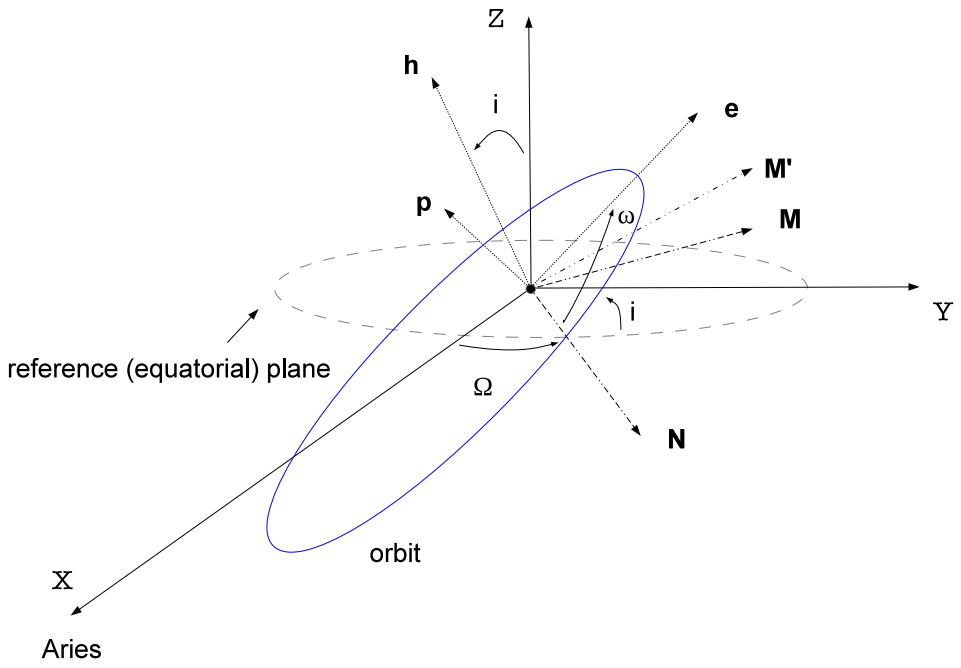


Figure 1.10: The rotations to define the local orbital reference system with respect to the pseudo-inertial one.

The reference system is said 'pseudo-inertial' and not inertial, because the origin orbits around the Sun.

We need to perform 3 rotations and thus 3 parameters are required to move from the pseudo-inertial reference system to the one defined by $\{\mathbf{e}, \mathbf{p}, \mathbf{h}\}$. Let us start from the $\{\hat{\mathbf{I}}, \hat{\mathbf{J}}, \hat{\mathbf{K}}\}$ coordinate system and use the classical Euler angles, in particular those called $3-1-3$ in astrodynamics. The transformation required consists in (see Fig. 1.12):

1. a rotation around the $Z-$axis of an angle Ω, called *longitude of the ascending node*, in such a way that the first axis coincides with the *line of nodes* $\mathbf{N}$ and the reference axes are $\{\mathbf{N}, \mathbf{M}, \hat{\mathbf{K}}\}$;

2. a rotation around $\mathbf{N}$ of an angle i, called *inclination* of the orbital plane, in such a way that the third axis is aligned along $\mathbf{h}$ and the reference axes are $\{\mathbf{N}, \mathbf{M}', \mathbf{h}\}$;

3. a rotation around $\mathbf{h}$ of an angle ω, called *argument of pericenter*, and we get to $\{\mathbf{e}, \mathbf{p}, \mathbf{h}\}$.

We notice that the first two rotations locate the orbital plane, while the third one helps

to define the position of the orbit in the plane.

The line of nodes is the line of intersection between the orbital plane and the reference plane (in our case the Earth's equatorial one) and clearly goes always through the center of the main body, i.e., the Earth in our case. The angle $\Omega \in [0°, 360°]$ is said longitude of the ascending node because, for $90° > i > 0°$, the satellite passes through it by 'going upward'. Opposite to the ascending node on the same line, we have the descending node. The argument of pericenter $\omega \in [0°, 360°]$ tells us how far the pericenter is from the line of nodes, while the inclination $i \in [0°, 180°]$ quantifies how much the orbital plane is tilted with respect to the reference one. If $90° > i > 0°$, then the small body moves counterclockwise and the orbit is said *prograde*, if $180° > i > 90°$, then we speak of *retrograde* orbit and the motion occurs clockwise. *Polar* orbits are such that $i = 90°$. If instead $i = 0°$, then the line of nodes cannot be defined nor Ω and the orbit, for an object moving around the Earth, is equatorial. On the other hand, if $e = 0$, that is, we deal with a circular orbit, then ω is not defined. To overcome these singularities, we can introduce new angular variables, in particular the *longitude of the pericenter* as $\tilde{\omega} = \omega + \Omega$. There exist other possibilities, for instance, the equinoctial orbital elements, which have the advantage to be not singular.

So far we have faced the Two–Body Problem from a geometrical point of view, demonstrating that the motion of the small body takes place on a conic section. If we focus on the specific case of the ellipse, its size and shape can be derived from the semi-major axis a and eccentricity e, while its orientation in space from the Euler angles Ω, i, ω. These 5 parameters are also known as *Lagrangian Parameters*. We notice that it is equivalent to define the ellipse by means of these independent parameters or by means of $\mathbf{h}$ and $\mathbf{e}$, which we proved (see Sec. 1.2) provide 5 independent parameters as well. As a matter of fact (see also Sec 1.11),

- from $\mathbf{h}$ it is possible to compute i and Ω and from $||\mathbf{h}||$ we obtain $p = h^2/\mu$ and then $a = p/(1 - e^2)$, where $e = ||\mathbf{e}||$;
- from $\mathbf{e}$ the pericenter direction is obtained using the dot and the cross products with the direction of the line of nodes.

1.7.1 Time of Pericenter Passage

Let us look for the sixth parameter, which allows to link the position and velocity on the orbit to a given instant of time. Indeed both r and v depends on t through θ. From (1.16) and (1.24) we have

$$\frac{d\theta}{dt} = h\frac{(1 + e\cos\theta)^2}{p^2},$$

which integrated gives

$$t - t_0 = \frac{p^2}{h}\int_{\theta_0}^{\theta} \frac{1}{(1 + e\cos\theta)^2} d\theta. \tag{1.55}$$

Let us take as reference position the pericenter, that is, $\theta_0 = 0$. The time t_0 is then said *time of pericenter passage* and it is denoted by τ_0. If we know this parameter, then we can solve the problem and associate a value of time to each value of true anomaly.

1.7.2 Circular Case

If the orbit is circular, i.e., $e = 0$, then

$$t - \tau_0 = \frac{a^2}{h}\int_{\theta_0}^{\theta} d\theta = \frac{a^2}{\sqrt{a\mu}}(\theta - \theta_0) = \sqrt{\frac{a^3}{\mu}}(\theta - \theta_0)\,. \tag{1.56}$$

In a period T, we sweep an angle equal to 2π and thus

$$T = 2\pi\sqrt{\frac{a^3}{\mu}}. \tag{1.57}$$

1.7.3 Parabolic Case

Let us consider the case of the parabolic orbit is parabolic, i.e., $e = 1$, and rewrite Eq. (1.55)

$$t - \tau_0 = \frac{p^2}{\sqrt{p\mu}}\int_0^{\theta} \frac{1}{(1+\cos\theta)^2} d\theta. \tag{1.58}$$

By using the trigonometric identity $1 + \cos\theta = 2\cos^2\frac{\theta}{2}$, the integral above results in

$$\begin{aligned}
\int_{\theta_0}^{\theta} \frac{1}{(1+\cos\theta)^2} d\theta &= \int_{\theta_0}^{\theta} \frac{1}{4\cos^4\frac{\theta}{2}} d\theta = \frac{1}{2}\int_{\theta_0}^{\theta} \frac{1}{\cos^2\frac{\theta}{2} 2\cos^2\frac{\theta}{2}} d\theta = \\
&= \frac{1}{2}\int_{\theta_0}^{\theta} \frac{\cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}}{\cos^2\frac{\theta}{2}} \frac{1}{2\cos^2\frac{\theta}{2}} d\theta = \\
&= \frac{1}{2}\int_{\theta_0}^{\theta} \left(1 + \tan^2\frac{\theta}{2}\right) \frac{1}{2\cos^2\frac{\theta}{2}} d\theta = \\
&= \frac{1}{2}\int_{\theta_0}^{\theta} \frac{1}{2}\left(1 + \tan^2\frac{\theta}{2}\right)^2 d\theta = \\
&= \frac{1}{2}\int_{\theta_0}^{\theta} \left(1 + \tan^2\frac{\theta}{2}\right) d\left(\tan\frac{\theta}{2}\right),
\end{aligned}$$

and in this way we find the *Barker's Equation*

$$t - \tau_0 = \frac{1}{2}\sqrt{\frac{p^3}{\mu}}\left(\tan\frac{\theta}{2} + \frac{1}{3}\tan^3\frac{\theta}{2}\right). \tag{1.59}$$

Given θ is easy to solve for t, but the inverse procedure requires to solve a cubic equation.

1.8 Elliptic Case: Eccentric Anomaly

For the elliptic orbit, i.e., $0 < e < 1$, we need to define a new reference system and a new geometrical definition for this conic section. We have already seen the equation in

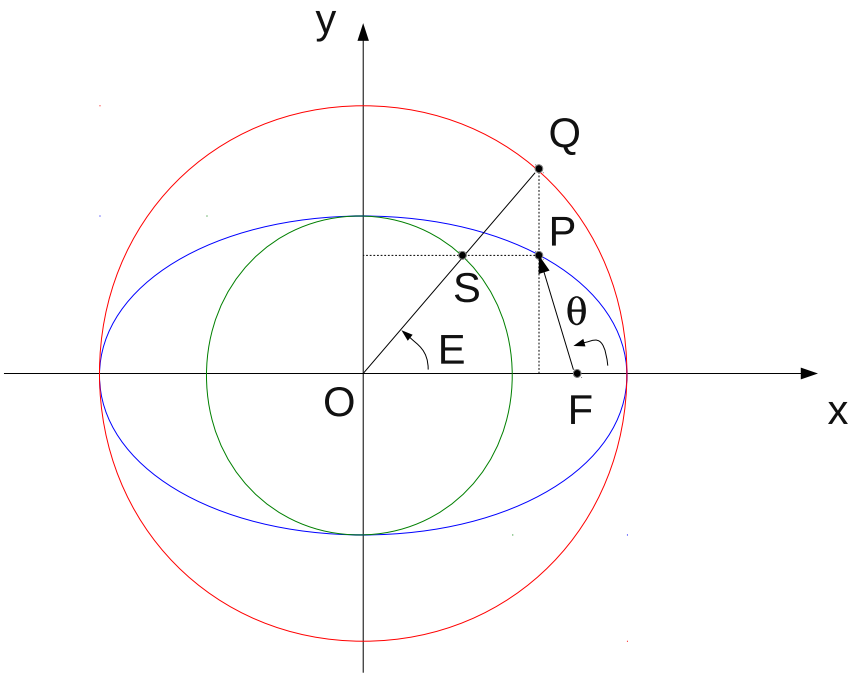


Figure 1.11: The definition of eccentric anomaly.

Cartesian coordinates (1.25) and in polar ones (1.24), now we will derive the parametric form. Let us consider Fig. 1.11, where we have drawn the circumscribed circle to the ellipse of radius a and the inscribed one of radius b, both with origin at the center of the ellipse. The line parallel to the semi-minor axis and passing through a given point P on the ellipse intersects the outer circle in Q, while the line parallel to the semi-major axis and passing through P intersects the inner circle in S. The origin and the points Q and S belong to the same line, which can be defined by the central angle E, which is called *eccentric anomaly*. Using this angle, the coordinates of any point P on the ellipse with respect to the origin are

$$x = a\cos E, \qquad\qquad y = b\sin E. \tag{1.60}$$

We notice that by adding up the squares of the two coordinates we find Eq. (1.25).
Let us derive r as a function of E instead of θ. From the same figure and Eq. (1.24), the $x-$coordinate is

$$x = a\cos E = OF + r\cos\theta = ae + r\cos\theta = ae + \frac{p-r}{e}.$$

If we multiply by e, then

$$ae\cos E = ae^2 + p - r = ae^2 + a(1-e^2) - r,$$

that is,

$$r = a\,(1 - e\cos E)\,, \tag{1.61}$$

which tells us that r depends linearly on $\cos E$, but is inversely proportional to $\cos\theta$.

1.8.1 True Anomaly and Eccentric Anomaly

Let us look for the relationship linking the eccentric anomaly to the true anomaly. We have already seen that

$$x = a\cos E = ae + r\cos\theta, \qquad y = b\sin E = r\sin\theta, \tag{1.62}$$

from which it follows, using the polar equation of the orbit (1.24),

$$\begin{aligned}\cos E &= e + \frac{1-e^2}{1+e\cos\theta}\cos\theta = \frac{e(1+e\cos\theta)+(1-e^2)\cos\theta}{1+e\cos\theta} = \frac{\cancel{e+e^2\cos\theta}+\cos\theta\cancel{-e^2\cos\theta}}{1+e\cos\theta},\\ \sin E &= \frac{a(1-e^2)}{a\sqrt{1-e^2}}\frac{\sin\theta}{1+e\cos\theta},\end{aligned}$$

that is,

$$\cos E = \frac{e+\cos\theta}{1+e\cos\theta}, \qquad \sin E = \frac{\sqrt{1-e^2}\sin\theta.}{1+e\cos\theta}. \tag{1.63}$$

Moreover, the first expression in (1.62) can be written using (1.61) as

$$\begin{aligned} r + r\cos\theta = r + a\cos E - ae \quad &\Longrightarrow \quad r(1+\cos\theta) = a\,(1-e\cos E) + a\cos E - ae,\\ \Longrightarrow \quad & 2r\cos^2\frac{\theta}{2} = a - ae\cos E + a\cos E - ae,\\ \Longrightarrow \quad & 2r\cos^2\frac{\theta}{2} = a(1-e)\cos E + a(1-e) = a(1-e)\,(1+\cos E)\,,\\ \Longrightarrow \quad & 2r\cos^2\frac{\theta}{2} = 2a(1-e)\cos^2\frac{E}{2}, \end{aligned} \tag{1.64}$$

because in general $1+\cos\alpha = 2\cos^2\frac{\alpha}{2}$. The second expression in (1.62), using $\sin\alpha = 2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}$, becomes instead

$$2r\sin\frac{\theta}{2}\cos\frac{\theta}{2} = 2b\sin\frac{E}{2}\cos\frac{E}{2}. \tag{1.65}$$

From the ratio between (1.65) and (1.64) we get

$$\tan\frac{\theta}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}. \tag{1.66}$$

It turns out that we always have $\theta > E$ and that $\frac{\theta}{2}$ and $\frac{E}{2}$ belong to the same quadrant. For sake of completeness, we give also the expressions for the radial and transversal components of the velocity both in terms of true and eccentric anomaly, namely,

$$\begin{cases} v_r = \sqrt{\frac{\mu}{a(1-e^2)}}e\sin\theta = \sqrt{\frac{\mu}{a}}\frac{e\sin E}{1-e\cos E}, \\ v_\theta = \sqrt{\frac{\mu}{a(1-e^2)}}\,(1+e\cos\theta) = \sqrt{\frac{\mu(1-e^2)}{a}}\frac{1}{1-e\cos E}. \end{cases} \tag{1.67}$$

1.8.2 Kepler's Equation

Let us see how the eccentric anomaly varies with time. If we derive with respect to time Eq. (1.61) we find

$$\dot{r} = ae\dot{E}\sin E. \tag{1.68}$$

Now let us project the velocity (1.35) onto the radial direction, namely,

$$\dot{r} = \dot{\mathbf{r}}\cdot\hat{\mathbf{r}} = \frac{\mu}{h}\left(e\hat{\mathbf{p}}\cdot\hat{\mathbf{r}} + \hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{r}}\right).$$

Since $\hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{r}} = 0$, we have (see also (1.67))

$$\dot{r} = \frac{\mu}{h}e\cos\left(\frac{\pi}{2}-\theta\right) = \frac{\mu}{h}e\sin\theta. \tag{1.69}$$

From (1.60), we get then

$$\sin\theta = \frac{b}{r}\sin E \qquad \Longrightarrow \qquad \dot{r} = \frac{\mu e b}{hr}\sin E,$$

which combined with (1.68) gives

$$\dot{E} = \frac{\mu b}{ahr}, \tag{1.70}$$

which can be written also either as

$$\dot{E} = \frac{h}{b}\frac{1}{r}, \tag{1.71}$$

or

$$\dot{E} = \sqrt{\frac{\mu}{a}}\frac{1}{r}. \tag{1.72}$$

It is interesting to notice that $\dot{E}$ is proportional to $1/r$, while $\dot{\theta}$ to $1/r^2$ (see Eq. (1.16)). By plugging (1.61) into (1.72) we obtain

$$dt = \sqrt{\frac{a^3}{\mu}}\left(1 - e\cos E\right)dE, \tag{1.73}$$

which can be integrated to get to

$$t - \tau_0 = \sqrt{\frac{a^3}{\mu}}\left(E - e\sin E\right), \tag{1.74}$$

where τ_0 is the value of time corresponding to $E = 0$, that is, the pericenter as before. Let us define the *mean anomaly* as

$$M = \sqrt{\frac{\mu}{a^3}}t, \tag{1.75}$$

in such a way that (1.74) can be written as

$$M - M_0 = E - e\sin E, \tag{1.76}$$

which is said *Kepler's equation*.

1.8.3 How to Solve the Kepler's Equation

Given e and M_0 at some initial time t_0, if we know E then it is easy to compute M at some time t, but the inverse procedure requires to solve a transcendental equation. This can be done by means of iterative methods. Here, we will present two of them, but there exist others, for instance by taking advantage of the Bessel's functions.

Successive Substitutions

The first method we can consider is established on the assumption that we know a value for the eccentric anomaly, say E_0 corresponding to $t = t_0$, close enough to the one we aim at computing, say E at t.

The procedures consists in neglecting the term containing the trigonometric function at the beginning and then to evaluate it at the eccentric anomaly E computed at the previous iteration, that is,

$$\begin{aligned} E_0 &= M - M_0 \\ E_1 &= M - M_0 + e\sin E_0 \\ E_{i+1} &= M - M_0 + e\sin E_i. \end{aligned}$$

The longer we iterate the smaller the correction to be applied to E. We can stop when

$$|E_{i+1} - E_i| < \epsilon, \tag{1.77}$$

where ϵ is a given tolerance. The convergence is usually reached after 30-40 iterations, so not quickly.

Newton's Method

In this case, we look at the problem as at the computation of the zeros of the function

$$f(E) = E - e\sin E - (M - M_0)\,, \tag{1.78}$$

and we exploit the fact that this function can be derived with respect to E. In particular, a Taylor series expansion of the first order gives

$$f(E) = f(E_0) + f'(E_0)(E - E_0) = 0, \tag{1.79}$$

which can be generalized to the $i-$step as

$$f'(E_i)(E_{i+1}-E_i) = -f(E_i) \quad \Longrightarrow \quad E_{i+1} = E_i - \frac{f(E_i)}{f'(E_i)} = E_i - \frac{E_i - e\sin E_i - (M - M_0)}{1 - e\cos E_i}.$$

The procedure ends whenever a condition of the type (1.77) is met. With the Newton's method the convergence is usually much faster.

1.8.4 Orbital Period and Mean Motion

The orbital period of an elliptic orbit can be derived by evaluating (1.74) at $E = 2\pi$, namely,

$$T = 2\pi\sqrt{\frac{a^3}{\mu}}, \tag{1.80}$$

which is the same expression found for the circular case, Eq. (1.57). We notice that the period does not depend on the eccentricity of the orbit, but just on the semi-major axis. We call *mean motion* the mean angular velocity of the small body along the orbit, namely,

$$n = \frac{2\pi}{T} = \sqrt{\frac{\mu}{a^3}}, \tag{1.81}$$

that can be used to write the Kepler's equation (1.76) as

$$E - e\sin E = n\,(t - \tau_0)\,. \tag{1.82}$$

Also,

$$\dot{M} = n = \sqrt{\frac{\mu}{a^3}} = \sqrt{\frac{\mu p}{a^3 p}} = \frac{h}{a}\frac{1}{\sqrt{ap}} = \frac{h}{a}\frac{1}{\sqrt{a^2(1-e^2)}} = \frac{h}{ab},$$

which is a constant. The mean anomaly is therefore the most regular of the anomalies considered so far, though it does not have any geometrical meaning.

1.9 Hyperbolic Case

The eccentric anomaly in the hyperbolic case is

$$\dot{E} = \frac{h}{r(-jb')} = j\frac{h}{rb'} = j\dot{E}', \tag{1.83}$$

where we have defined $\dot{E}'$ as the real number such that

$$\dot{E}' = \frac{h}{rb'}. \tag{1.84}$$

As a consequence, the parametric form for the coordinates with respect to the origin becomes

$$x = a\cos E = -a'\cos(jE'), \qquad y = b\sin E = -jb'\sin(jE'). \tag{1.85}$$

Since in general for a complex number α it holds

$$\sin\alpha = \frac{e^{j\alpha} - e^{-j\alpha}}{2j}, \qquad \cos\alpha = \frac{e^{j\alpha} + e^{-j\alpha}}{2},$$

and for a real number β

$$\sinh\beta = \frac{e^{\beta} - e^{-\beta}}{2}, \qquad \cosh\beta = \frac{e^{\beta} + e^{-\beta}}{2},$$

we have

$$x = -a'\frac{e^{-E'} - e^{E'}}{2} = -a'\cosh E', \qquad y = -jb'\frac{e^{-E'} - e^{E'}}{2j} = b'\sinh E'. \tag{1.86}$$

Hence, the parametric form for the radius vector as a function of E' is

$$r = a'(e\cosh E' - 1). \tag{1.87}$$

We notice that

$$\frac{x^2}{a'^2} - \frac{y^2}{b'^2} = \cosh^2 E' - \sinh^2 E' = 1,$$

as stated by (1.48).

Concerning the mean anomaly, we have

$$\dot{M} = \frac{h}{ba} = \frac{h}{(-a')(-jb')} = -j\frac{h}{a'b'} = -j\dot{M}', \tag{1.88}$$

where

$$\dot{M}' = \frac{h}{a'b'}. \tag{1.89}$$

The Kepler's equation (1.76) is now

$$e\sinh E' - E' = M' - M'_0, \tag{1.90}$$

while the relationship between true anomaly and the new defined eccentric anomaly reads

$$\tan\frac{\theta}{2} = \sqrt{\frac{e+1}{e-1}}\tanh\frac{E'}{2}. \tag{1.91}$$

Finally, the analogous expressions of the radial and transversal components of the velocity (1.67) are

$$v_r = \frac{na'e\sinh E'}{e\cosh E' - 1}, \qquad v_\theta = \frac{na'\sqrt{e^2-1}}{e\cosh E' - 1}, \tag{1.92}$$

where the mean motion n is given by definition as

$$n = \sqrt{\frac{\mu}{a'^3}}, \tag{1.93}$$

though it does not have a physical meaning.

1.10 From Position and Velocity to Orbital Elements

Let us assume to the position and velocity vector, say $(\mathbf{r}_0, \mathbf{v}_0)$, of a satellite at time t_0 orbiting around the Earth. In order to obtain the corresponding orbital elements the steps to follow are:

- compute the angular momentum

$$\mathbf{h} = \mathbf{h}_0 = \mathbf{r}_0 \times \mathbf{v}_0,$$

from which the inclination i of the orbit can be computed as

$$\cos i = \frac{h_z}{||\mathbf{h}||}, \tag{1.94}$$

and the longitude of the ascending node Ω as

$$\tan \Omega = \left(-\frac{h_x}{h_y}\right). \tag{1.95}$$

Indeed, from (1.8) and Fig. 1.12 we have

$$\begin{cases} h_x = y\dot{z} - z\dot{y} = h \sin i \sin \Omega, \\ h_y = z\dot{x} - x\dot{z} = -h \sin i \cos \Omega, \\ h_z = x\dot{y} - y\dot{x} = h \cos i; \end{cases} \tag{1.96}$$

- compute the eccentricity vector

$$\mathbf{e} = \mathbf{e}_0 = -\hat{\mathbf{r}}_0 - \frac{1}{\mu}\mathbf{h}_0 \times \mathbf{v}_0,$$

whose modulus is the eccentricity e;

- the semi-major axis a can be derived from

$$p = a\left(1 - e^2\right) = \frac{h^2}{\mu} \qquad \Longrightarrow \qquad a = \frac{h^2}{\mu\left(1 - e^2\right)};$$

- once computed Ω, the direction of the line of nodes is given by

$$\mathbf{N} = \begin{pmatrix} \cos \Omega \\ \sin \Omega \\ 0 \end{pmatrix}, \tag{1.97}$$

and the argument of pericenter ω can be computed from

$$\cos \omega = \mathbf{N} \cdot \hat{\mathbf{e}}, \qquad \sin \omega = \hat{\mathbf{h}} \cdot (\mathbf{N} \times \hat{\mathbf{e}}); \tag{1.98}$$

- if the time t_0 corresponds to the time of pericenter, then we already have the sixth parameter. Otherwise, we compute the true anomaly by means of

$$\cos \theta_0 = \hat{\mathbf{e}} \cdot \hat{\mathbf{r}}_0,$$

and consequently the eccentric anomaly by means of Eq. (1.66) and the time of pericenter passage from Eq. (1.74). In other words, the sixth parameter to define the orbit can be equivalently the time of pericenter passage, the true anomaly or the eccentric anomaly.

1.11 From Orbital Elements to Position and Velocity

Once solved for the orbital elements, to know the position and velocity vectors corresponding to a time $t \neq t_0$ we have two options:

1. to vary the true anomaly θ;
2. to vary the time t.

In the first case,

- the corresponding eccentric anomaly is given by Eq. (1.66);
- the corresponding time by Eq. (1.74);
- the modulus of the corresponding radius vector by either (1.24) or (1.61);
- since we know θ, $\hat{\mathbf{r}}$ and $\hat{\theta}$ can be computed by means of (1.9);
- therefore both $\mathbf{r}(t)$ and $\mathbf{v}(t)$ are known from (1.13).

In the second case,

- the corresponding eccentric anomaly by solving iteratively the Kepler's equation (1.76);
- the corresponding true anomaly is found by means of Eq. (1.66);
- the final steps are the same as in the first case.

Clearly, the same procedure hold whenever we know sixth orbital elements and we aim at computing the state vector at a given time t or true anomaly θ.
Moreover, following Sec. 1.7 to transform the coordinates given in the polar frame to the pseudo-inertial reference system we need first to move to the Laplace reference system by rotating around the axis defined by $\mathbf{h}$ of an angle θ and then perform the three rotations corresponding to the angles (ω, i, Ω). The latter is the inverse transformation of the one described in Sec. 1.7. In this way, the coordinates in the pseudo-inertial reference systems are given by

$$\begin{cases} x = r\left[\cos\Omega\cos\left(\omega+\theta\right) - \sin\Omega\sin\left(\omega+\theta\right)\cos i\right], \\ y = r\left[\sin\Omega\cos\left(\omega+\theta\right) + \cos\Omega\sin\left(\omega+\theta\right)\cos i\right], \\ z = r\sin\left(\omega+\theta\right)\sin i, \end{cases} \tag{1.99}$$

and

$$\begin{cases} \dot{x} = -\frac{na}{\sqrt{1-e^2}}\left\{\left[\sin\left(\omega+\theta\right) + e\sin\omega\right]\cos\Omega + \left[\cos\left(\omega+\theta\right) + e\cos\omega\right]\sin\Omega\cos i\right\}, \\ \dot{y} = -\frac{na}{\sqrt{1-e^2}}\left\{\left[\sin\left(\omega+\theta\right) + e\sin\omega\right]\sin\Omega - \left[\cos\left(\omega+\theta\right) + e\cos\omega\right]\cos\Omega\cos i\right\}, \\ \dot{z} = \frac{na}{\sqrt{1-e^2}}\left[\cos\left(\omega+\theta\right) + e\cos\omega\right]\sin i. \end{cases} \tag{1.100}$$

As a final note, we remark that throughout the chapter, we have solved the Kepler's problem by looking for the associated first integrals. This is the so-called *method of parameters*, which allows to compute easily an analytical solution of the problem. However, there also exists the possibility to integrate numerically the equation of motion (1.3) in order to get the time evolution of position and velocity of the satellite very accurately. This is called *method of coordinates*.

1.12 The Inverse Problem

So far we have faced the so-called direct problem: from the physical laws of dynamics and gravitation we have solved for the motion of a small body in a gravitational field. This is a deductive approach. However, from a historical point of view the procedure was opposite: Johannes Kepler enunciated three laws to describe the motion of the planets around the Sun, deriving these laws from the observations. Afterward, thanks to the experiments carried out by Galileo Galilei, Isaac Newton provided a general understanding of the problem of the mutual attraction of two bodies by means of the Universal Law of Gravitation.

In this section, we deal with this inductive approach: from the Kepler's Laws and the Newtons's Second Law of Motion we will deduce the Universal Law of Gravitation. This is a very actual approach, in the sense that often both for space mission and more general problems the desired motion (or effect) is known and the corresponding forces (or causes) to this to occur have to be derived.

Our hypotheses are the Newtons's Second Law of Motion $\mathbf{F} = m\mathbf{a}$ and the three *Kepler's Laws*, namely,

1. *Law of Ellipses*: the planets move on ellipses, being the center of the Sun located at one focus of the ellipse;
2. *Law of Equal Areas*: a line going from the center of the Sun to the planet sweeps out equal areas in equal intervals of time (see Fig. 1.12);
3. *Law of Harmonies*: the ratio between the square of the period of revolution and the cube of the semi-major axis of the ellipse is a constant.

We notice that the second law implies, in particular, that the speed at the pericenter is the maximum one, the one at the apocenter the minimum one.

Let us translate the Kepler's Laws in mathematical jargon:

1. the position of the planet with respect to the Sun is given by the polar equation of the ellipse Eq. (1.24)
$$r = \frac{p}{1 + e\cos\theta};$$
2. the area of the elliptic infinitesimal sector is
$$dA = \frac{1}{2} r r d\theta, \tag{1.101}$$

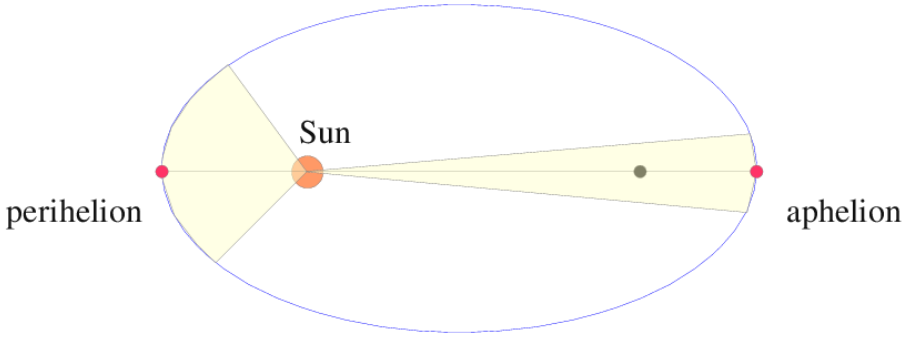


Figure 1.12: The Law of Areas.

and thus the areolar velocity is

$$\dot{A} = \frac{1}{2} r^2 \dot{\theta} = constant. \tag{1.102}$$

In the same way we obtained Eq. (1.16), we find

$$h = r^2 \dot{\theta} = 2\dot{A}, \tag{1.103}$$

and thus

$$\mathbf{h} = constant.$$

Indeed, the angular momentum vector is by definition

$$\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}} = \mathbf{r} \times \left(\dot{r}\hat{\mathbf{r}} + \mathbf{w} \times \mathbf{r} \right),$$

where $\mathbf{w} = \dot{\theta}\hat{\mathbf{k}}$ is angular velocity of the polar reference system with respect to the inertial one (see Fig. 1.2).

Using the properties of the triple cross product, we have

$$\mathbf{h} = \mathbf{r} \times (\mathbf{w} \times \mathbf{r}) = r^2 \mathbf{w} - \cancelto{0}{(\mathbf{r} \cdot \mathbf{w})\mathbf{r}} = r^2 \dot{\theta} \hat{\mathbf{k}}.$$

3.

$$\frac{T^2}{a^3} = constant.$$

Since the angular momentum vector is constant, the motion is planar and

$$\frac{d}{dt}\mathbf{h} = \frac{d}{dt}\left(\mathbf{r}\times\dot{\mathbf{r}}\right) = \dot{\mathbf{r}}\times\dot{\mathbf{r}} + \mathbf{r} + \ddot{\mathbf{r}} = 0 \quad \Longrightarrow \quad \mathbf{r}\times\ddot{\mathbf{r}} = 0.$$

For this cross product to be always null, the acceleration vector $\ddot{\mathbf{r}}$ must be always parallel to $\mathbf{r}$, which means that we deal with a central field. We can write

$$\ddot{\mathbf{r}} = V\hat{\mathbf{r}}, \tag{1.104}$$

where V is a scalar unknown function. Let us define now the eccentricity vector as the vector of magnitude equal to the eccentricity of the ellipse and direction the one going from the center of the Sun to the pericenter of the orbit. In this way Eq. (1.16) becomes

$$r = \frac{p}{1+\mathbf{e}\cdot\hat{\mathbf{r}}} \quad \Longrightarrow \quad p = r + \mathbf{e}\cdot\mathbf{r}.$$

By taking the first and second derivative of p with respect to time, we obtain

$$\begin{aligned} 0 = \dot{p} &= \dot{r} + \mathbf{e}\cdot\dot{\mathbf{r}}, \\ 0 = \ddot{p} &= \ddot{r} + \mathbf{e}\cdot\ddot{\mathbf{r}} = \ddot{r} + V\mathbf{e}\cdot\hat{\mathbf{r}}. \end{aligned} \tag{1.105}$$

Since

$$\dot{r} \;=\; \frac{d}{dt}r = \frac{pe\sin\theta}{\left(1+e\cos\theta\right)^2}\dot{\theta} = \frac{pe\sin\theta}{\left(1+e\cos\theta\right)^2}2\frac{\dot{A}}{r^2} = 2\dot{A}\frac{e\sin\theta}{p},$$

we have

$$\ddot{r} \;=\; \frac{d}{dt}\dot{r} = 2\dot{A}\frac{e\cos\theta}{p}\dot{\theta} = 2\dot{A}\frac{e\cos\theta}{p}2\frac{\dot{A}}{r^2} = 4\dot{A}^2\frac{e\cos\theta\left(1+e\cos\theta\right)^2}{p^3}.$$

The area of the ellipse is given by $A=\pi ab$ and is swept in one orbital period, thus we can write

$$\dot{A} = \frac{\pi ab}{T}, \tag{1.106}$$

and

$$\begin{aligned} \ddot{r} \;&=\; 4\frac{\pi^2a^2b^2}{T^2}\frac{e\cos\theta\left(1+e\cos\theta\right)^2}{p^3} = 4\frac{\pi^2a^2a^2(1-e^2)}{T^2}\frac{e\cos\theta\left(1+e\cos\theta\right)^2}{a^3(1-e^2)^3} = \\ &=\; 4\frac{\pi^2a}{T^2}\frac{e\cos\theta\left(1+e\cos\theta\right)^2}{(1-e^2)^2}. \end{aligned}$$

In this way, Eq. (1.105) reads

$$Ve\cos\theta = -\ddot{r} = -4\frac{\pi^2a}{T^2}\frac{e\cos\theta\left(1+e\cos\theta\right)^2}{(1-e^2)^2},$$

that is,

$$\begin{aligned} V &= -4\frac{\pi^2 a}{T^2}\frac{(1+e\cos\theta)^2}{(1-e^2)^2} = -4\frac{\pi^2 a}{T^2}\frac{a^2(1-e^2)^2}{r^2(1-e^2)^2} = \\ &= -4\frac{\pi^2 a^3}{T^2 r^2}. \end{aligned} \tag{1.107}$$

If we define the constant μ as

$$\mu = 4\frac{\pi^2 a^3}{T^2}, \tag{1.108}$$

and plug V into (1.104) we arrive to the Universal Law of Gravitation, namely

$$\ddot{\mathbf{r}} = -\frac{\mu}{r^2}\hat{\mathbf{r}}. \tag{1.109}$$

We notice that (1.108) represents the Kepler's third law and can be also written as

$$\mu = n^2 a^3. \tag{1.110}$$

1.13 The f and g series

The position vector $\mathbf{r}(t)$ at a given time $t > t_0$ can be expressed as a Taylor expansion in the neighborhood of t_0 in this way:

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \dot{\mathbf{r}}(t_0)(t-t_0) + \ddot{\mathbf{r}}(t_0)\frac{(t-t_0)^2}{2!} + \cdots = \sum_{n=0}^{\infty}\frac{(t-t_0)^n}{n!}\frac{d^n\mathbf{r}}{dt^n}|_{t=t_0}. \tag{1.111}$$

If the motion is Keplerian, then $\mathbf{r}(t_0)$, $\dot{\mathbf{r}}(t_0)$, $\mathbf{r}(t)$ and $\dot{\mathbf{r}}(t)$ lie on the same plane and thus we can find two functions $f(t)$ and $g(t)$ such that

$$\mathbf{r}(t) = f(t)\mathbf{r}(t_0) + g(t)\dot{\mathbf{r}}(t_0). \tag{1.112}$$

Moreover, $\mathbf{r}(t), \dot{\mathbf{r}}(t)$ represents a basis for every vector belonging to their plane and thus also the $n-$derivative of $\mathbf{r}(t)$ can be written as a linear combination of $\mathbf{r}(t), \dot{\mathbf{r}}(t)$, namely,

$$\frac{d^n\mathbf{r}(t)}{dt^n} = f_n(t)\mathbf{r}(t) + g_n(t)\dot{\mathbf{r}}(t), \tag{1.113}$$

where f_n, g_n are two scalar functions depending on the time.

It can be demonstrated that f_n, g_n can be expressed as a function of three scalar quantities, called *Lagrange's Fundamental Invariants*:

$$\epsilon := \frac{\mu}{r^3}, \qquad \lambda := \frac{\mathbf{r}\cdot\dot{\mathbf{r}}}{r^2}, \qquad \psi = \frac{\dot{\mathbf{r}}\cdot\dot{\mathbf{r}}}{r^2},$$

having the dimension of $[t]^{-2}$, $[t]^{-1}$ and $[t]^{-2}$, respectively.

They are said *invariant* because they do not depend on the reference system (they are scalar quantities); *fundamental* because their derivatives can be expressed as a function

of the invariants themselves and thus they form a closed set with respect to the operation of derivation.

Example:

$$\dot{\epsilon} = -3\epsilon\lambda, \qquad \dot{\lambda} = \psi - \epsilon - 2\lambda^2, \qquad \dot{\psi} = -2\lambda(\epsilon + \psi).$$

Let us notice that

$$\begin{aligned}
\frac{d^1\mathbf{r}}{dt^1} &= \dot{\mathbf{r}}, \\
\frac{d^2\mathbf{r}}{dt^2} &= -\frac{\mu}{r^3}\mathbf{r} = -\epsilon\mathbf{r}, \\
\frac{d^3\mathbf{r}}{dt^3} &= 3\epsilon\lambda\mathbf{r} - \epsilon\dot{\mathbf{r}}, \\
\frac{d^4\mathbf{r}}{dt^4} &= (-15\epsilon\lambda^2 + 3\epsilon\psi - 2\epsilon^2)\mathbf{r} + 6\epsilon\lambda\dot{\mathbf{r}}.
\end{aligned}$$

If we apply (1.113) to the first two expressions, we get

$$\begin{aligned}
\frac{d^1\mathbf{r}}{dt^1} &= f_1\mathbf{r} + g_1\dot{\mathbf{r}} = \dot{\mathbf{r}}, \\
\frac{d^2\mathbf{r}}{dt^2} &= f_2\mathbf{r} + g_2\dot{\mathbf{r}} = -\epsilon\mathbf{r}.
\end{aligned}$$

that is,

$$f_1 = 0, \qquad g_1 = 1, \qquad f_2 = -\epsilon, \qquad g_2 = 0.$$

In an analogous way, we have

$$f_0 = 1, \qquad g_0 = 0, \qquad f_3 = 3\epsilon\lambda, \qquad g_3 = -\epsilon, \qquad f_4 = -15\epsilon\lambda^2 + 3\epsilon\psi - 2\epsilon^2, \qquad g_4 = 6\epsilon\lambda.$$

This means that (1.113) is fulfilled for

$$\begin{aligned}
f_n(t) &= f_n(\epsilon(t), \lambda(t), \psi(t)), \\
g_n(t) &= g_n(\epsilon(t), \lambda(t), \psi(t)).
\end{aligned}$$

and that

$$\frac{d^n\mathbf{r}}{dt^n}|_{t=t_0} = f_n(t_0)\mathbf{r}(t_0) + g_n(t_0)\dot{\mathbf{r}}(t_0).$$

Hence, the Taylor series expansion (1.111) becomes

$$\begin{aligned}
\mathbf{r}(t) &= \sum_{n=0}^{\infty} \frac{(t-t_0)^n}{n!} \frac{d^n\mathbf{r}}{dt^n}|_{t=t_0} \\
&= \sum_{n=0}^{\infty} \frac{(t-t_0)^n}{n!} f_n(t_0)\mathbf{r}(t_0) + \sum_{n=0}^{\infty} \frac{(t-t_0)^n}{n!} g_n(t_0)\dot{\mathbf{r}}(t_0) \\
&= f(t)\mathbf{r}(t_0) + g(t)\dot{\mathbf{r}}(t_0),
\end{aligned}$$

being $f(t)$ and $g(t)$ the so-called *Lagrange coefficients*:

$$\begin{aligned} f(t) &= \sum_{n=0}^{\infty} \frac{(t-t_0)^n}{n!} f_n(t_0), \\ g(t) &= \sum_{n=0}^{\infty} \frac{(t-t_0)^n}{n!} g_n(t_0). \end{aligned} \tag{1.114}$$

In the same way, the velocity vector can be approximated as

$$\dot{\mathbf{r}}(t) = \dot{f}(t)\mathbf{r}(t_0) + \dot{g}(t)\dot{\mathbf{r}}(t_0). \tag{1.115}$$

From a practical point of view, (1.112) and (1.115) can be used if $t - t_0$ is small enough, typically $\mu(t-t_0)^2/r_0^3 \leq 0.01$.

Notice that $f, g, \dot{f}, \dot{g}$[1] are not linear independent: if we know three of them, then we can compute the other one. As a matter of fact,

$$\begin{aligned} \mathbf{r} \times \dot{\mathbf{r}} &= (f\mathbf{r}_0 + g\dot{\mathbf{r}}_0) \times (\dot{f}\mathbf{r}_0 + \dot{g}\dot{\mathbf{r}}_0) \\ &= f\dot{f}(\mathbf{r}_0 \times \mathbf{r}_0) + f\dot{g}(\mathbf{r}_0 \times \mathbf{r}_0) + g\dot{f}(\dot{\mathbf{r}}_0 \times \mathbf{r}_0) + g\dot{g}(\dot{\mathbf{r}}_0 \times \dot{\mathbf{r}}_0) \\ &= f\dot{g}(\mathbf{r}_0 \times \dot{\mathbf{r}}_0) - g\dot{f}(\mathbf{r}_0 \times \dot{\mathbf{r}}_0), \end{aligned}$$

that is,

$$f\dot{g} - g\dot{f} = 1, \tag{1.116}$$

because $\mathbf{r}_0 \times \dot{\mathbf{r}}_0 = \mathbf{r} \times \dot{\mathbf{r}} = \mathbf{h}$.

Moreover, from

$$\begin{aligned} \mathbf{r} \times \dot{\mathbf{r}}_0 &= f\mathbf{r}_0 \times \dot{\mathbf{r}}_0 = f\mathbf{h}, \\ \mathbf{r} \times \mathbf{r}_0 &= -g\mathbf{r}_0 \times \dot{\mathbf{r}}_0 = -g\mathbf{h}, \end{aligned}$$

it follows

$$f = 1 - \frac{a}{r_0}\left[1 - \cos\left(E - E_0\right)\right], \tag{1.117}$$

$$g = t - t_0 - \frac{1}{n}\left[E - E_0 - \sin\left(E - E_0\right)\right], \tag{1.118}$$

and

$$f = 1 - \frac{r}{p}\left[1 - \cos\left(\theta - \theta_0\right)\right], \tag{1.119}$$

$$g = \frac{rr_0}{\sqrt{\mu p}} \sin\left(\theta - \theta_0\right). \tag{1.120}$$

[1] From now on we simplify the notation, omitting the dependence from t. The 0 subscript refers to the dependence to t_0.

To prove, for instance, (1.119), let us recall Eq. (1.12), namely,

$$\mathbf{r} = r\cos\theta\hat{\mathbf{i}} + r\sin\theta\hat{\mathbf{j}}, \qquad \frac{d\mathbf{r}}{dt} \equiv \dot{\mathbf{r}} = \dot{r}(\cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}}) + r\dot{\theta}(-\sin\theta\hat{\mathbf{i}} + \cos\theta\hat{\mathbf{j}}),$$

and Eq. (1.67), namely,

$$\dot{r} = \sqrt{\frac{\mu}{p}}e\sin\theta, \qquad r\dot{\theta} = \sqrt{\frac{\mu}{p}}\left(1 + e\cos\theta\right),$$

which give

$$\begin{aligned}
\dot{\mathbf{r}} &= \sqrt{\frac{\mu}{p}}e\sin\theta(\cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}}) + \sqrt{\frac{\mu}{p}}\left(1 + e\cos\theta\right)(-\sin\theta\hat{\mathbf{i}} + \cos\theta\hat{\mathbf{j}}) \\
&= \left[\sqrt{\frac{\mu}{p}}e\sin\theta\cos\theta - \sqrt{\frac{\mu}{p}}\left(1 + e\cos\theta\right)\sin\theta\right]\hat{\mathbf{i}} + \left[\sqrt{\frac{\mu}{p}}e\sin^2\theta + \sqrt{\frac{\mu}{p}}\left(1 + e\cos\theta\right)\cos\theta\right]\hat{\mathbf{j}} \\
&= -\sqrt{\frac{\mu}{p}}\sin\theta\hat{\mathbf{i}} + \sqrt{\frac{\mu}{p}}\left[e\sin^2\theta + \cos\theta + e\cos^2\theta\right]\hat{\mathbf{j}} \\
&= -\sqrt{\frac{\mu}{p}}\sin\theta\hat{\mathbf{i}} + \sqrt{\frac{\mu}{p}}\left(e + \cos\theta\right)\hat{\mathbf{j}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbf{r}\times\dot{\mathbf{r}}_0 &= \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ 0 \end{pmatrix} \times \begin{pmatrix} -\sqrt{\frac{\mu}{p}}\sin\theta_0 \\ \sqrt{\frac{\mu}{p}}\left(e + \cos\theta_0\right) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ r\cos\theta\sqrt{\frac{\mu}{p}}\left(e + \cos\theta_0\right) + r\sin\theta\sqrt{\frac{\mu}{p}}\sin\theta_0 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ r\sqrt{\frac{\mu}{p}}\left[e\cos\theta + \cos\theta\cos\theta_0 + \sin\theta\sin\theta_0\right] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ r\sqrt{\frac{\mu}{p}}\left[e\cos\theta + \cos\left(\theta - \theta_0\right)\right] \end{pmatrix},
\end{aligned}$$

and thus

$$\begin{aligned}
f &= \frac{r\sqrt{\frac{\mu}{p}}\left[e\cos\theta + \cos\left(\theta - \theta_0\right)\right]}{\sqrt{\mu p}} = \frac{r}{p}\left[e\cos\theta + 1 - 1 + \cos\left(\theta - \theta_0\right)\right] \\
&= 1 - \frac{r}{p}\left[1 - \cos\left(\theta - \theta_0\right)\right].
\end{aligned}$$

1.14 Orbits at the Earth

To conclude the chapter we describe how the orbits for satellites around the Earth are usually distinguished . According to their altitude, eccentricity and inclination, they can serve to different purposes. Some of them are used especially to build constellations to ensure a constant global coverage.

1.14.1 Low Earth Orbits

Low Earth Orbits (LEO) are characterized by an altitude between about 200 km and 2000 km. They define a spherical region around the Earth, which is the most crowded one, mainly because higher altitudes require more powerful launchers. Various values of inclination with respect to the Earth equatorial plane are used, depending on the aim of the mission but also on the latitude of the launch site. We will see that they are strong affected by the perturbations due to the geopotential (see Sec. 3.1) and the atmospheric drag (see Sec. 3.2), the latter being responsible of a rapid decay towards the Earth. To mention some examples, the International Space Station (ISS) is orbiting on an almost circular LEO at an average altitude of 400 km and inclination of 51.6°. The Hubble Space Telescope is orbiting on a nearly circular orbit at an average altitude of 570 km and inclination of 28.5°. Earth observation satellites are also placed in LEO. They provide high resolution data to monitor, for instance, oceans, climate and geological features and for military and security services. Constellations of satellites in LEO are commonly used for communication purposes, e.g. the Iridium system. Equatorial LEO satellites are considered to observe equatorial regions, but have the drawback to be subject to challenging thermal and radiation conditions, which strongly affect the design of the payload. A special class of LEO is represented by the so-called *sun-synchronous orbits*, which ensure a constant orientation with respect to the Sun and thus illumination conditions. This property can be required, for instance, for the solar panels or to maintain a well-defined ground lighting. In Sec. 3.1.4 we will see how they are designed. The GOCE mission, whose aim was to measure the Earth's gravity field and modeling the geoid with an extremely high accuracy and spatial resolution, had a almost circular sun-synchronous orbit, with inclination of about 96.7°.

1.14.2 Geostationary Orbits

Geostationary Earth Orbits (GEO) are defined in such a way that the satellite appears always at a given fixed position in the sky to a ground observer, that is, it does not move relative to the Earth. To achieve this condition, it is required that its orbital period equals the Earth's rotation one and, as a consequence, GEO are characterized by zero inclination, zero eccentricity and orbital altitude equal to 35786 km. The main drawback in the choice of such orbits is that more energy must be spent both to reach this altitude and to zero the inclination (none of the existing launch sites is fully equatorial).

The *solar day* is the interval of time occurring between two subsequent noons and it consists of 24 hours. The rotation period of the Earth around its spin axis with respect to an inertial reference system, the so-called *sidereal day*, is shorter than one solar day, because in the time required to complete one rotation around its axis, the Earth has also moved around the Sun.

Let us refer to Fig. 1.13. Since in one solar year, i.e., 365.25 solar days, the Earth sweeps 360° along its orbit, then in one solar day it travels an angular distance equal to

$$\alpha = \frac{360^\circ}{365.25} = 0.986^\circ. \tag{1.121}$$

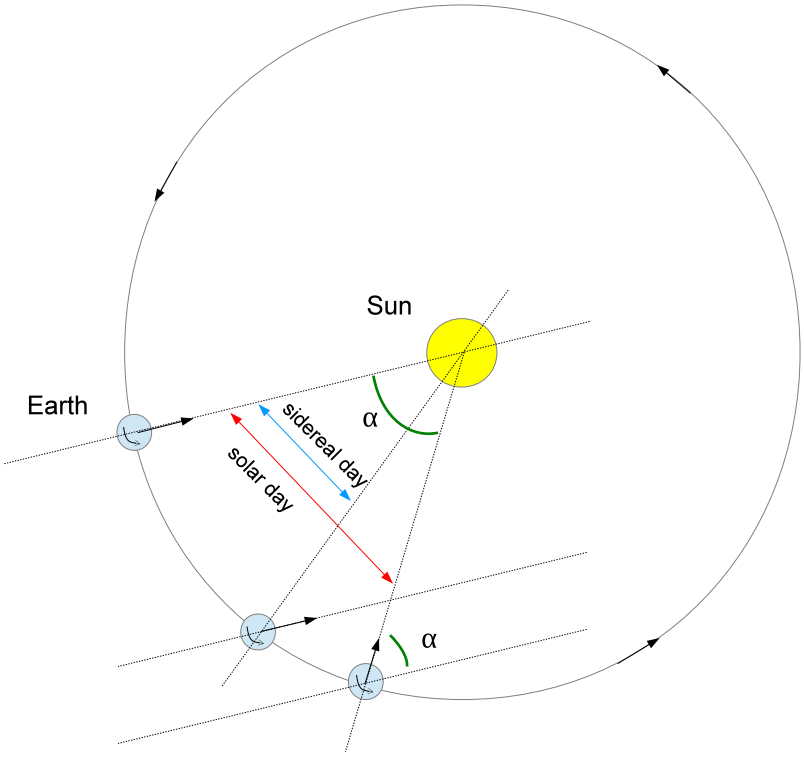


Figure 1.13: The solar and the sidereal day (not in scale).

This is the angle that must balanced by a further rotation of the Earth to get to the next noon. In other words,

$$\frac{T_{sol}}{360^\circ + \alpha} = \frac{T_{sid}}{360^\circ}, \tag{1.122}$$

where $T_{sol} = 24$ h and here 360° refers to the angle swept by the Earth in its rotation. Therefore,

$$T_{sid} = 23 \text{ h } 56'4'' = 86164 \text{ s}. \tag{1.123}$$

A satellite having period equal to one sidereal day is called *geosynchronous*. From the Kepler's third law, we derive the semi-major axis corresponding to this period, namely,

$$T_{GEO} \equiv T_{sid} = 2\pi\sqrt{\frac{a_{GEO}^3}{\mu_\oplus}} \qquad \Longrightarrow \qquad a_{GEO} = 42164 \text{ km}. \tag{1.124}$$

GEO are thus geosynchronous orbits, but they must be also circular. Otherwise, the orbital velocity will be not constant, and this fact would make it moves in longitude with respect to the ground station. The inclination must be zero to maintain the pointing in latitude.

These orbits are mainly used for telecommunication purposes, but also for weather satellites when the concern is not to obtain high resolution data, but to analyze global phenomena. Due to the high altitude, a GEO satellite can observe a large region on Earth and, as a matter of fact, 3 GEO satellites displaced by 120° in mean anomaly are enough to provide a global coverage (excluding the polar regions).

To transfer spacecraft from LEO to GEO they are usually employed the so-called *Geostationary Transfer Orbits* (GTO), which are elliptical orbits characterized by a perigee within the LEO region and an apogee near or above GEO. Their inclination is determined by the latitude of the launch site.

1.14.3 Medium Earth Orbits

Medium Earth Orbits (MEO) generally include all the orbits lying in altitude between LEO and GEO. The corresponding region around the Earth has gained interest recently because of the Global Navigation Satellite Systems (GNSS), which are spacecraft constellations able to provide autonomous geo-spatial positioning (and relative velocity), plus accurate timing, with global coverage. At the moment the American NAVSTAR GPS and the Russian GLONASS systems are operational, while the European Galileo system and the Chinese Beidou (Compass) one are in the building phase. They are placed in circular, inclined orbits ($\approx 56°$ for all of them, except Glonass for which $i \approx 65°$) with an altitude between approximately 19000 and 24000 km. They exploit 3 or 6 planes (only for GPS), i.e., the satellites are equally displaced in Ω.

Orbital Perturbations 2

2.1 Introduction

In the real world the orbits of the satellites, as well as the planetary ones, are not exactly Keplerian conic sections. Indeed, the hypothesis of a perfectly symmetric gravitational field, i.e. equivalent to the one due to a pointless mass, is never verified. Moreover, we can never avoid the presence of other external forces affecting the motion of the body. However, as we mentioned in the previous chapter, in many situations the Keplerian solution can be considered a good initial solution for two main reasons:

- the Kepler's problem is the only dynamical model that can be solved in closed form;
- for Earth's satellites, any other effect is some orders of magnitude lower than the Earth's monopole.

So far we have seen that, under the Keplerian assumptions, the equation of motion of a massless particle is (1.3), namely,

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\hat{\mathbf{r}},$$

and that the corresponding motion can be described by means of 6 independent parameters which are constant in time.

In general, when there exist additional forces acting on the spacecraft, the equation of motion can be written as

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\hat{\mathbf{r}} + \mathbf{f}, \tag{2.1}$$

where $\mathbf{f}$ includes all the dynamical contributions, which can be distinguished from the gravitational attraction exerted by the pointless body of mass M. We notice that $\mathbf{f}$ is, more properly, an acceleration and not a force. It is usually referred as a *perturbation* and does not necessarily represent a problem to the mission designer. Indeed these accelerations, and the deviations from the pure Keplerian motion they generate, can help in the design of the trajectory.

In the perturbative case, the orbital elements are no longer constant, but vary with time. Let us assume to know the orbital elements that at a given value of time t^* define the Keplerian orbit which, as a first approximation, describes the motion of the particle.

This orbit is said *osculating orbit* at t^*, because at this epoch position and velocity on it correspond to position and velocity on the true orbit. If we are able to compute the function that at t^* provides the future evolution of the osculating orbital elements, then we can compute a new unique set of orbital elements, which define the new orbit at the next epoch.

To solve the perturbed problem there exist various methods, which are usually divided in two main classes:

- *special perturbation methods*, which integrate numerically a given problem with given initial conditions and therefore aim at obtaining specific solutions holding only in well-defined cases;
- *general perturbation methods*, aiming to provide approximate analytical solutions, no matter on the initial conditions.

Moreover, as we will see in Chap. 3, the orbital perturbations can be classified in

- *secular*: the orbital element affected by the perturbation varies linearly with time (or according to some power of time);
- *long-period*: the effects due to the perturbation repeat with a period which is at least 1 order of magnitude larger than the orbital period;
- *short-period*: the effects due to the perturbation repeat after few or less orbital periods.

2.2 Special Perturbation Theory

Our aim is to compute a solution for

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^2}\hat{\mathbf{r}} + \mathbf{f}, \qquad \mathbf{r}(t_0) = \mathbf{r}_0, \qquad \mathbf{v}(t_0) = \mathbf{v}_0. \tag{2.2}$$

This is a Cauchy problem, which admits a unique solution.

Let us start with the special perturbation techniques, which include:

- *the Cowell's method*: the equation of motion is integrated numerically without any preliminary manipulation. The drawback is that the main gravitational term is at least 10^3 times bigger than the perturbations and this, together with the unavoidable errors due to the numerical approximation, makes the perturbative effects almost disappear.
- *The Encke's method*: it consists in integrating numerically only the perturbative terms in the equation of motion and thus terms of the different order of magnitude are treated separately. The whole motion is then built by superimpose all the solutions. The reference conic section is updated once the deviation from the Keplerian motion overcomes a given threshold.
- *The Herrick's method*: the approach is the same as in the Encke's method, but now the Keplerian conic section is updated at each time step.

2.2.1 Cowell's Approach

The Cowell's method consists in integrating numerically the whole equation of motion (2.2), but turns out to be efficient only if the perturbative acceleration $\mathbf{f}$ has almost the same order of magnitude as the inverse-square central gravitational one. Otherwise, it requires relatively small integration steps, which may affect the computational time and the growth of errors associated with roundoff. In any case, once chosen a given numerical integration method, the differential system to be considered is

$$\begin{cases} \dot{x} = v_x, \\ \dot{y} = v_y, \\ \dot{z} = v_z, \\ \dot{v_x} = -\frac{\mu}{r^3}x + f_x, \\ \dot{v_y} = -\frac{\mu}{r^3}y + f_y, \\ \dot{v_z} = -\frac{\mu}{r^3}x + f_z, \end{cases} \tag{2.3}$$

where $r = \sqrt{x^2 + y^2 + z^2}$ and the perturbative terms are usually given in some analytical form (see Chap. 3 for some examples).

2.2.2 Encke's Approach

With the Encke's method, only the differential equation corresponding to the perturbative term in (2.2) is integrated numerically. In this way, a given accuracy can be obtained also with relatively large integration steps. The main idea is to start from a reference solution at a given time, the osculating orbit, that would result in absence of any perturbing accelerations. At the time chosen, the true solution and the osculating one coincide. If the perturbations are some orders of magnitude lower than the central gravitational acceleration, then over short intervals of time the actual position and velocity will differ from the ones corresponding to the osculating orbit by a small amount. Whenever the true orbit deviates from the osculating one, then a *rectification* is applied, that is, a new initial time and corresponding reference solution are considered.

Let us assume to know the position and velocity on the orbit at time t^* and associate them to the osculating orbit at t^*, namely,

$$\mathbf{r}(t^*) = \mathbf{r}_{osc}(t^*), \qquad \mathbf{v}(t^*) = \mathbf{v}_{osc}(t^*).$$

At $t = t^* + \Delta t$, we can write

$$\mathbf{r}(t) = \mathbf{r}_{osc}(t) + \boldsymbol{\delta}(t), \qquad \mathbf{v}(t) = \mathbf{v}_{osc}(t) + \boldsymbol{\nu}(t), \tag{2.4}$$

where $\boldsymbol{\delta}(t)$ and $\boldsymbol{\nu}(t)$ are the 'small' differences in position and velocity between the actual and the osculating orbits. In particular, $\boldsymbol{\delta}(t)$ satisfies the differential equation

$$\frac{d^2\boldsymbol{\delta}}{dt^2} = \frac{d^2\mathbf{r}}{dt^2} - \frac{d^2\mathbf{r}_{osc}}{dt^2} = -\frac{\mu}{r^3}\mathbf{r} + \mathbf{f} + \frac{\mu}{r_{osc}^3}\mathbf{r}_{osc} = -\frac{\mu}{r^3}\mathbf{r} + \mathbf{f} + \frac{\mu}{r_{osc}^3}\left(\mathbf{r} - \boldsymbol{\delta}\right),$$

that is,

$$\frac{d^2\boldsymbol{\delta}}{dt^2} + \frac{\mu}{r_{osc}^3}\boldsymbol{\delta} = \frac{\mu}{r_{osc}^3}\left(1 - \frac{r_{osc}^3}{r^3}\right)\mathbf{r} + \mathbf{f}, \tag{2.5}$$

with initial conditions

$$\boldsymbol{\delta}(t^*) = 0, \qquad \dot{\boldsymbol{\delta}}(t^*) \equiv \boldsymbol{\nu}(t^*) = 0.$$

The main problem from a numerical point of view arises due to the term

$$1 - \frac{r_{osc}^3}{r^3},$$

which represents the difference of two almost equal numbers. One way to overcome this is to define

$$q = 1 - \frac{r^2}{r_{osc}^2},$$

in such a way that

$$\frac{r_{osc}^3}{r^3} = (1-q)^{-\frac{3}{2}}. \tag{2.6}$$

Written in this way, we can apply the series expansion

$$1 - \frac{r_{osc}^3}{r^3} = 1 - (1-q)^{-\frac{3}{2}} = 3q - \frac{3\cdot 5}{2!}q^2 + \frac{3\cdot 5\cdot 7}{3!}q^3 - .. = qf(q), \tag{2.7}$$

where

$$f(q) = 3\left(1 - \frac{5}{2}q + \frac{5\cdot 7}{3!}q^2 - ..\right).$$

Therefore, the equation to be integrated is

$$\frac{d^2\boldsymbol{\delta}}{dt^2} = \frac{\mu}{r_{osc}^3}\left[qf(q)\left(\mathbf{r}_{osc} + \boldsymbol{\delta}\right) - \boldsymbol{\delta}\right] + \mathbf{f}. \tag{2.8}$$

To summarize, from a practical point of view the procedure to follow is

1. compute a reference orbit, i.e. $\mathbf{r}_{osc}(t^*)$ and $\mathbf{v}_{osc}(t^*)$;
2. in the first step $\boldsymbol{\delta}(t^*) = 0$ and $\boldsymbol{\nu}(t^*) = 0$, that is, $\mathbf{r}(t^*) = \mathbf{r}_{osc}(t^*)$, $\mathbf{v}(t^*) = \mathbf{v}_{osc}(t^*)$ and thus $q(t^*) = 0$;
3. integrate Eq. (2.8) up to $t = t^* + \Delta t$ starting from these initial conditions;
4. in this way you find $\boldsymbol{\delta}(t^* + \Delta t)$ and $\dot{\boldsymbol{\delta}}(t^* + \Delta t)$ and thus $\mathbf{r}(t^* + \Delta t) = \mathbf{r}_{osc}(t^* + \Delta t) + \boldsymbol{\delta}(t^* + \Delta t)$ and $\mathbf{v}(t^* + \Delta t) = \mathbf{v}_{osc}(t^* + \Delta t) + \boldsymbol{\nu}(t^* + \Delta t)$, where $\mathbf{r}_{osc}(t^* + \Delta t)$ and $\mathbf{v}_{osc}(t^* + \Delta t)$ are computed from the osculating orbital elements;
5. compute $q(t^* + \Delta t)$ and the corresponding $f(q)$;
6. start over from step 3. now using as initial conditions the ones just computed and $t^* = t^* + \Delta t$.

The Encke's method allows large integration step, because the time variation of $\boldsymbol{\delta}$ is expected to be much slower than the one corresponding to $\mathbf{r}$. This fact suggests one possible criterion to decide when to apply the rectification of the orbit. This is, when $\frac{\mu}{r_{osc}^3}\left[qf(q)\left(\mathbf{r}_{osc} + \boldsymbol{\delta}\right) - \boldsymbol{\delta}\right]$ becomes much smaller than $\mathbf{f}$. A different rule can be to change the osculating orbit when $\frac{\delta}{r_{osc}} > 0.01$ (Bate, Mueller & White, 1971). In both cases, any growth in roundoff and truncation errors due to the numerical integration is bounded.

2.3 General Perturbation Theory

General perturbation methods include:

- *the method of the variation of parameters*: the motion can be described using the same 6 parameters, that are constant in the Keplerian case but evolve in time in the perturbed one. This is an effective method for many reasons: the parameters vary less rapidly than the Cartesian coordinates; their variation provide us with an immediate insight on the gap between the Keplerian motion and the perturbed one and finally the orbital parameters have a physical significance. There exist two formulations associated with this methodology, the first leads to write the so-called *Gauss planetary equations*, which relate the components of the perturbation $\mathbf{f}$ to the variations in the orbital elements. This is also called *the geometrical method* and will be faced in details in what follows. The second derivation consists in the so-called *Lagrange planetary equations* and can be used only if the perturbation is conservative.
- *The method of the perturbation of the coordinates*: the equation of motion is split in two terms, one containing the Keplerian contribution, the other all the perturbations. The latter is then integrated by means of a series expansion up to the first meaningful orders.

2.3.1 Gauss Planetary Equations

Let us start by understanding if the angular momentum and eccentricity vectors, defined by Eq. (1.8) and (1.20) respectively, are still constant. To this end, let us cross multiply Eq. (2.2) by $\mathbf{r}$, namely,

$$\mathbf{r} \times \ddot{\mathbf{r}} = -\frac{\mu}{r^2}(\mathbf{r} \times \hat{\mathbf{r}}) + \mathbf{r} \times \mathbf{f}. \tag{2.9}$$

Since the first term can be written as

$$\dot{\mathbf{r}} \times \dot{\mathbf{r}} + \mathbf{r} \times \ddot{\mathbf{r}} = \frac{d}{dt}(\mathbf{r} \times \dot{\mathbf{r}}) = \frac{d\mathbf{h}}{dt},$$

and $\mathbf{r} \times \hat{\mathbf{r}} = 0$, we have

$$\dot{\mathbf{h}} \equiv \frac{d\mathbf{h}}{dt} = \mathbf{r} \times \mathbf{f}. \tag{2.10}$$

The angular momentum vector varies in time, and, as a consequence, so the orbital plane does. This implies, in particular, that the polar reference system $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{h}}\}$ rotates with respect to the inertial one with angular velocity $\mathbf{w}$ which is now not always directed along $\mathbf{h}$, i.e. $\mathbf{w}$ is not always perpendicular to the orbital plane.

In the polar reference system system, the velocity vector can be expressed as

$$\dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + \mathbf{w} \times \mathbf{r}. \tag{2.11}$$

By cross multiplication by $\mathbf{r}$, we get

$$\mathbf{r} \times \dot{\mathbf{r}} = \dot{r}\cancelto{0}{(\mathbf{r} \times \hat{\mathbf{r}})} + \mathbf{r} \times (\mathbf{w} \times \mathbf{r}) = r^2\mathbf{w} - (\mathbf{r} \cdot \mathbf{w})\,\mathbf{r},$$

where $\mathbf{r}\cdot\mathbf{w}$ is not null, because $\mathbf{w}$ is not normal to the orbital plane. In other words, recalling Eq. (1.8),

$$\mathbf{h} = r^2\mathbf{w} - (\mathbf{r}\cdot\mathbf{w})\,\mathbf{r}, \tag{2.12}$$

or

$$\mathbf{w} = \frac{\mathbf{h}}{r^2} + \frac{\mathbf{r}\cdot\mathbf{w}}{r^2}\mathbf{r}, \tag{2.13}$$

that is,

$$\mathbf{w} = \frac{\mathbf{h}}{r^2} + \gamma\hat{\mathbf{r}}, \tag{2.14}$$

where we have defined

$$\gamma = \frac{\mathbf{r}\cdot\mathbf{w}}{r} \equiv \mathbf{w}\cdot\hat{\mathbf{r}} \equiv w_r, \tag{2.15}$$

which is an unknown. The vector $\mathbf{w}$ can thus be decomposed in a component along $\hat{\mathbf{h}}$, one along $\hat{\mathbf{r}}$ and none along $\hat{\boldsymbol{\theta}}$.

Let us cross multiply Eq. (2.2) by $\mathbf{h}$, namely,

$$\mathbf{h}\times\ddot{\mathbf{r}} = -\frac{\mu}{r^2}(\mathbf{h}\times\hat{\mathbf{r}}) + \mathbf{h}\times\mathbf{f}. \tag{2.16}$$

The term on the left-hand side can be written as

$$\mathbf{h}\times\ddot{\mathbf{r}} = \frac{d}{dt}(\mathbf{h}\times\dot{\mathbf{r}}) - \dot{\mathbf{h}}\times\dot{\mathbf{r}},$$

while the first term on the right-hand side as (since $\hat{\mathbf{r}}\times\hat{\mathbf{r}} = 0$)

$$-\frac{\mu}{r^2}(\mathbf{h}\times\hat{\mathbf{r}}) = -\mu\left(\frac{\mathbf{h}}{r^2} + \gamma\hat{\mathbf{r}}\right)\times\hat{\mathbf{r}} = -\mu\mathbf{w}\times\hat{\mathbf{r}} = -\mu\dot{\hat{\mathbf{r}}},$$

and thus Eq. (2.16) can be rewritten as

$$\frac{d}{dt}\left(\frac{1}{\mu}(\mathbf{h}\times\dot{\mathbf{r}}) + \hat{\mathbf{r}}\right) = \frac{1}{\mu}\left(\dot{\mathbf{h}}\times\dot{\mathbf{r}}\right) + \frac{1}{\mu}(\mathbf{h}\times\mathbf{f}), \tag{2.17}$$

that is, recalling Eq. (1.20),

$$\frac{d\mathbf{e}}{dt} = -\frac{1}{\mu}\left(\dot{\mathbf{h}}\times\dot{\mathbf{r}}\right) - \frac{1}{\mu}(\mathbf{h}\times\mathbf{f}). \tag{2.18}$$

We notice that since the orbital elements $(a, e, i, \Omega, \omega)$ depend on the angular momentum and eccentricity vectors, Eqs. (2.10) and (2.18) state the time evolution of the orbital elements depends on the magnitude of the perturbation. This is another advantage of the method of the variation of parameters. We also point out that $\mathbf{h}$ and $\mathbf{e}$ are not constant, but they are still orthogonal, like in the Keplerian case. This means that also now a sixth parameter is required to solve the time problem.

Let us consider the specific mechanical energy, namely,

$$\mathcal{E} = \frac{1}{2}v^2 - \frac{\mu}{r} = T + \Pi = -\frac{\mu}{2a}, \tag{2.19}$$

where we have specified the kinetic term T and the potential one Π.

The work-energy theorem states that the variation in kinetic energy is the work L done by all the external forces acting on the particle, that is,

$$dT = dL, \tag{2.20}$$

where the work done can be split in two terms, one accounting for the conservative forces $\mathbf{f}_c$, the other for dissipative ones $\mathbf{f}_d$, in this way

$$dL = dL_c + dL_d. \tag{2.21}$$

The work done by the conservative contributions is the opposite of the variation of potential energy, that is,

$$dL_c = -d\Pi, \tag{2.22}$$

while the one corresponding to dissipative forces is given by

$$dL_d = \mathbf{f}_d \cdot d\mathbf{s}, \tag{2.23}$$

where $d\mathbf{s}$ is a line element. So the total variation in kinetic energy can be expressed as

$$dT = -d\Pi + \mathbf{f}_d \cdot d\mathbf{s}, \tag{2.24}$$

and the specific mechanical energy change as

$$d\mathcal{E} = d(T + \Pi) = \mathbf{f}_d \cdot d\mathbf{s}. \tag{2.25}$$

The corresponding variation in time is

$$\dot{\mathcal{E}} \equiv \frac{d\mathcal{E}}{dt} = \mathbf{f} \cdot \mathbf{v} = f_d^t v, \tag{2.26}$$

where f_d^t is the tangential component of $\mathbf{f}_d$, that is, the one along the velocity vector.

Eq. (2.26) states that the energy of a perturbed orbit can vary only when there exists a dissipative force acting along the tangential direction of the orbit. If the energy changes, so the semi-major axis (and the orbital period) does, namely,

$$\dot{\mathcal{E}} = \frac{d}{dt}\left(-\frac{\mu}{2a}\right) = \frac{\mu}{2a^2}\frac{da}{dt} = f_d^t v \qquad \Longrightarrow \qquad \frac{da}{dt} = \frac{2a^2}{\mu} f_d^t v. \tag{2.27}$$

In other words, if only conservative forces are acting on the spacecraft or the dissipative ones are not directed along the tangent to the orbit at the given point, then the semi-major axis and the energy remain constant.

Let us derive now the *Gauss planetary equations*, which give the time variation of the orbital parameters starting from the assumption that the perturbation $\mathbf{f}$ can be decomposed in the polar reference system $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{h}}\}$, namely,

$$\mathbf{f} = f^r \hat{\mathbf{r}} + f^\theta \hat{\boldsymbol{\theta}} + f^h \hat{\mathbf{h}}. \tag{2.28}$$

The equation we have just found for the semi-major axis (2.27) considered instead the $\{\hat{\mathbf{t}}, \hat{\mathbf{n}}, \hat{\mathbf{h}}\}$ reference system (see Fig. 2.1), defined as

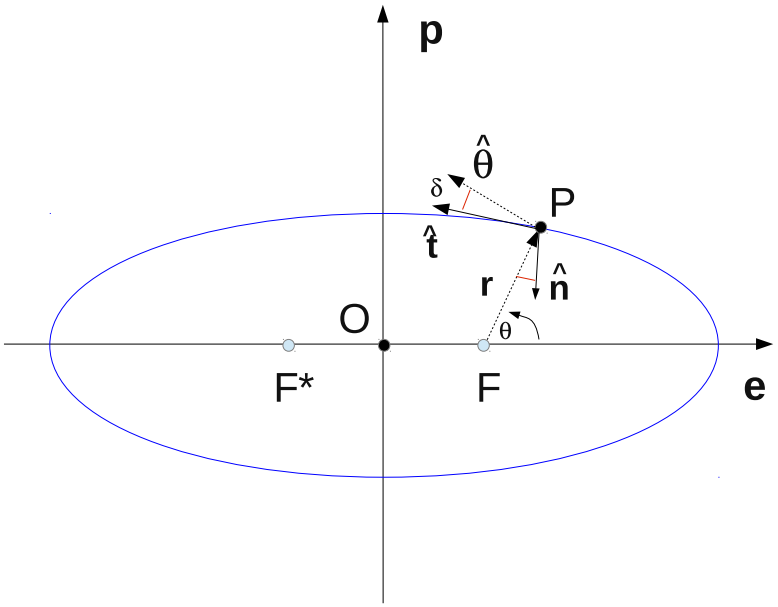


Figure 2.1: The $\{\hat{\mathbf{t}}, \hat{\mathbf{n}}, \hat{\mathbf{h}}\}$ reference system. The out-of-plane direction is simply directed along $\mathbf{h}$ and thus perpendicular to the orbital plane displayed.

- the first axis is along the tangential direction $\hat{\mathbf{t}}$ to the orbit at a given point P;
- the second axis is the normal direction $\hat{\mathbf{n}}$ to $\hat{\mathbf{t}}$ in the orbital plane (positive towards the interior of the ellipse);
- the out-of-plane axis is the normal direction $\hat{\mathbf{h}}$ to the orbital plane.

By introducing the angle δ as the angle between the transversal direction $\hat{\boldsymbol{\theta}}$ and the tangential one $\hat{\mathbf{t}}$ (see Fig. 2.1), Eq. (2.27) becomes

$$\frac{da}{dt} = \frac{2a^2}{\mu} v \left(-f_d^r \sin\delta + f_d^\theta \cos\delta \right). \tag{2.29}$$

Since

$$\sin\delta = -\frac{1}{v}\frac{\mu}{h} e \sin\theta, \qquad \cos\delta = \frac{1}{v}\frac{\mu}{h}\left(e\cos\theta + 1\right), \tag{2.30}$$

we have

$$\frac{da}{dt} = \frac{2a^2}{\mu} v \frac{\mu}{hv}\left[f_d^r e \sin\theta + f_d^\theta (1 + e\cos\theta) \right] = \frac{2a^2}{h}\left[f_d^r e \sin\theta + f_d^\theta (1 + e\cos\theta) \right]. \tag{2.31}$$

We notice again that the out-of-plane component of the perturbation does not have any effect on the semi-major axis.

Let us derive the equations relative to the other orbital parameter, by projecting Eq. (2.10) onto the polar reference system $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{h}}\}$. We have

$$\begin{aligned} \dot{\mathbf{h}} = \dot{h}\hat{\mathbf{h}} + \mathbf{w} \times \mathbf{h} &= \mathbf{r} \times \mathbf{f}, \\ \dot{h}\hat{\mathbf{h}} + \left(\frac{\mathbf{h}}{r^2} + \gamma\hat{\mathbf{r}}\right) \times \mathbf{h} &= \begin{vmatrix} \hat{\mathbf{r}} & \hat{\boldsymbol{\theta}} & \hat{\mathbf{h}} \\ r & 0 & 0 \\ f^r & f^\theta & f^h \end{vmatrix}, \\ \dot{h}\hat{\mathbf{h}} - \gamma h\hat{\boldsymbol{\theta}} &= -rf^h\hat{\boldsymbol{\theta}} + rf^\theta\hat{\mathbf{h}}, \end{aligned} \tag{2.32}$$

that gives

$$\dot{h} = rf^\theta, \qquad \gamma = \frac{rf^h}{h}. \tag{2.33}$$

In this way, we have an explicit expression for γ, which depends on f^h. This is the term that makes the difference between $\mathbf{w}$ in the perturbative case and the Keplerian one. In other words, $\mathbf{w}$ varies according to the time variation in i, Ω, which is responsible of the time variation of the orbital plane.

If we recall the findings of Sec. 1.7, the vector $\mathbf{w}$ can decomposed in terms of the Euler angles as

$$\mathbf{w} = \dot{\Omega}\hat{\mathbf{K}} + \dot{i}\hat{\mathbf{N}} + \dot{\tilde{\theta}}\hat{\mathbf{h}}, \tag{2.34}$$

where $\tilde{\theta} = \omega + \theta$ is the angle between the line of nodes and the radius vector $\mathbf{r}$. We notice that the $\{\hat{\mathbf{K}}, \hat{\mathbf{N}}, \hat{\mathbf{h}}\}$ reference system is not orthogonal. In polar components, we have

$$w_r = \gamma = \hat{\mathbf{r}} \cdot \left(\dot{\Omega}\hat{\mathbf{K}} + \dot{i}\hat{\mathbf{N}} + \dot{\tilde{\theta}}\hat{\mathbf{h}}\right), \tag{2.35}$$

and since

$$\begin{aligned} \hat{\mathbf{r}} \cdot \hat{\mathbf{K}} &= \sin\tilde{\theta}\sin i, \\ \hat{\mathbf{r}} \cdot \hat{\mathbf{N}} &= \cos\tilde{\theta}, \end{aligned}$$

it turns out that

$$\gamma = \dot{\Omega}\sin\tilde{\theta}\sin i + \dot{i}\cos\tilde{\theta}. \tag{2.36}$$

Moreover,

$$w_\theta = 0 = \hat{\boldsymbol{\theta}} \cdot \left(\dot{\Omega}\hat{\mathbf{K}} + \dot{i}\hat{\mathbf{N}} + \dot{\tilde{\theta}}\hat{\mathbf{h}}\right), \tag{2.37}$$

and since

$$\begin{aligned} \hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{K}} &= \cos\tilde{\theta}\sin i, \\ \hat{\boldsymbol{\theta}} \cdot \hat{\mathbf{N}} &= \cos\left(\frac{\pi}{2} + \tilde{\theta}\right) = -\sin\tilde{\theta}, \end{aligned}$$

we get to

$$0 = \dot{\Omega}\cos\tilde{\theta}\sin i - \dot{i}\sin\tilde{\theta}. \tag{2.38}$$

The last component reads

$$w_h = \frac{h}{r^2} = \hat{\mathbf{h}} \cdot \left(\dot{\Omega} \hat{\mathbf{K}} + \dot{i} \hat{\mathbf{N}} + \dot{\tilde{\theta}} \hat{\mathbf{h}} \right), \tag{2.39}$$

and using

$$\hat{\mathbf{h}} \cdot \hat{\mathbf{K}} = \cos i,$$

we obtain

$$\frac{h}{r^2} = \dot{\Omega} \cos i + \dot{\tilde{\theta}}. \tag{2.40}$$

Let us multiply Eq. (2.36) by $\sin \tilde{\theta}$ and Eq. (2.38) by $\cos \tilde{\theta}$ and sum them up. Then, let us make the difference between Eq. (2.36) multiplied by $\cos \tilde{\theta}$ and Eq. (2.38) by $\sin \tilde{\theta}$. We obtain

$$\dot{\Omega} \sin i = \gamma \sin \tilde{\theta}, \tag{2.41}$$

$$\dot{i} = \gamma \cos \tilde{\theta}, \tag{2.42}$$

$$\dot{\tilde{\theta}} = \frac{h}{r^2} - \dot{\Omega} \cos i. \tag{2.43}$$

Using (2.33), the first two become

$$\dot{\Omega} \sin i = \frac{r}{h} f^h \sin \tilde{\theta}, \tag{2.44}$$

$$\dot{i} = \frac{r}{h} f^h \cos \tilde{\theta}, \tag{2.45}$$

that is, the time variation on i, Ω depends only on the out-of-plane component of the perturbation.

Let us consider now the time variation of the eccentricity vector, namely,

$$\frac{d\mathbf{e}}{dt} = \dot{e} \hat{\mathbf{e}} + \mathbf{w}_e \times \mathbf{e}, \tag{2.46}$$

where

$$\mathbf{w}_e = \mathbf{w} - \dot{\theta} \hat{\mathbf{h}} \tag{2.47}$$

is the angular velocity of $\hat{\mathbf{e}}$ with respect to the inertial reference system. By means of (2.14), we obtain

$$\mathbf{w}_e = \left(\frac{h}{r^2} - \dot{\theta} \right) \hat{\mathbf{h}} + \gamma \hat{\mathbf{r}} = \left(\frac{h}{r^2} - \dot{\tilde{\theta}} + \dot{\tilde{\theta}} - \dot{\theta} \right) \hat{\mathbf{h}} + \gamma \hat{\mathbf{r}}, \tag{2.48}$$

and using (2.40),

$$\mathbf{w}_e = \left(\dot{\Omega} \cos i + \dot{\omega} \right) \hat{\mathbf{h}} + \gamma \hat{\mathbf{r}}, \tag{2.49}$$

where $\dot{\omega}$ is the time derivative of the argument of pericenter ω. In this way, (2.46) becomes

$$\frac{d\mathbf{e}}{dt} = \dot{e} \hat{\mathbf{e}} + \left(\dot{\Omega} \cos i + \dot{\omega} \right) \hat{\mathbf{h}} \times \mathbf{e} + \gamma \hat{\mathbf{r}} \times \mathbf{e}. \tag{2.50}$$

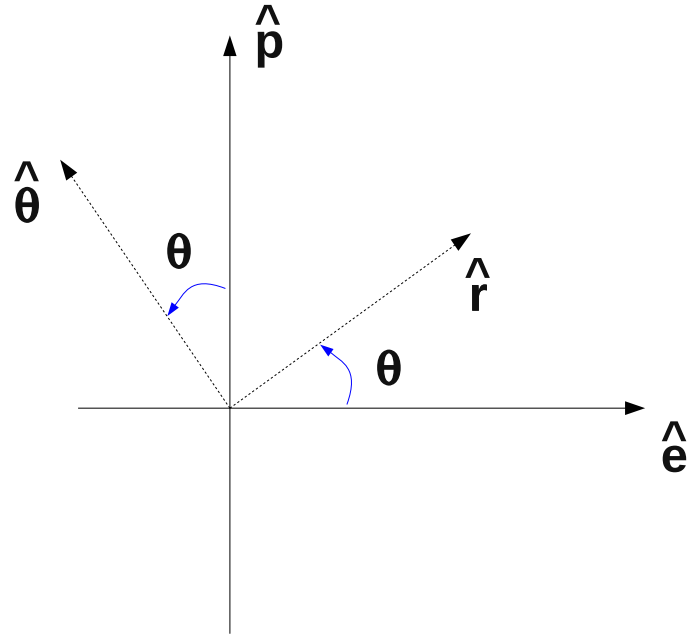


Figure 2.2: The unit vectors $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}$ with respect to the unit vectors $\hat{\mathbf{e}}, \hat{\mathbf{p}}$.

Since (see Sec. 1.4)

$$\hat{\mathbf{h}} \times \mathbf{e} = e\hat{\mathbf{p}}, \qquad \hat{\mathbf{r}} \times \mathbf{e} = -e \sin\theta \hat{\mathbf{h}}, \tag{2.51}$$

we get to

$$\frac{d\mathbf{e}}{dt} = \dot{e}\hat{\mathbf{e}} + \left(\dot{\Omega}\cos i + \dot{\omega}\right)\hat{\mathbf{p}} - \gamma e \sin\theta \hat{\mathbf{h}}. \tag{2.52}$$

Let us cross multiply Eq. (2.10) with Eq. (1.35), which holds also in the perturbative case,

$$\begin{aligned}
\dot{\mathbf{h}} \times \dot{\mathbf{r}} &= \frac{\mu}{h}(\mathbf{r} \times \mathbf{f}) \times \left(e\hat{\mathbf{p}} + \hat{\boldsymbol{\theta}}\right) \\
&= -\frac{\mu}{h}\left[e\hat{\mathbf{p}} \times (\mathbf{r} \times \mathbf{f}) + \hat{\boldsymbol{\theta}} \times (\mathbf{r} \times \mathbf{f})\right] \\
&= -\frac{\mu}{h}\left[e\hat{\mathbf{p}} \times \left(-rf^h\hat{\boldsymbol{\theta}} + rf^\theta\hat{\mathbf{h}}\right) + \hat{\boldsymbol{\theta}} \times \left(-rf^h\hat{\boldsymbol{\theta}} + rf^\theta\hat{\mathbf{h}}\right)\right] \\
&= -\frac{\mu}{h}\left(erf^h \sin\theta\hat{\mathbf{h}} + erf^\theta\hat{\mathbf{e}} + rf^\theta\hat{\mathbf{r}}\right).
\end{aligned} \tag{2.53}$$

Moreover,

$$-\frac{1}{\mu}(\mathbf{h} \times \mathbf{f}) = -\frac{1}{\mu}\left(-hf^\theta\hat{\mathbf{r}} + hf^r\hat{\boldsymbol{\theta}}\right). \tag{2.54}$$

Therefore, Eq. (2.52) can be written as

$$\dot{e}\hat{\mathbf{e}} + \left(\dot{\Omega}\cos i + \dot{\omega}\right)\hat{\mathbf{p}} - \gamma e \sin\theta \hat{\mathbf{h}} = \frac{1}{h}\left(erf^h \sin\theta \hat{\mathbf{h}} + erf^\theta \hat{\mathbf{e}} + rf^\theta \hat{\mathbf{r}}\right) - \frac{\mu}{h}\left(f^r \hat{\boldsymbol{\theta}} - f^\theta \hat{\mathbf{r}}\right). \tag{2.55}$$

By considering that (see Fig. 2.2)

$$\hat{\mathbf{r}} = \cos\theta\hat{\mathbf{e}} + \sin\theta\hat{\mathbf{p}}, \qquad \hat{\boldsymbol{\theta}} = -\sin\theta\hat{\mathbf{e}} + \cos\theta\hat{\mathbf{p}}, \tag{2.56}$$

and equating components, we obtain

$$\gamma e \sin\theta = \frac{e}{h} r f^h \sin\theta, \tag{2.57}$$

and

$$\begin{aligned} \dot{e} &= \frac{e}{h} r f^\theta + \frac{r}{h} f^\theta \cos\theta + \frac{h}{\mu} f^r \sin\theta + \frac{h}{\mu} f^\theta \cos\theta \\ &= \frac{r}{h}\left[\left(\cos\theta + e + \frac{h^2}{\mu r}\cos\theta\right) f^\theta + \frac{h^2}{\mu r} f^r \sin\theta\right]. \end{aligned} \tag{2.58}$$

Since

$$r(t) = \frac{\frac{h(t)}{\mu^2}}{1 + e(t)\cos\theta}, \tag{2.59}$$

Eq. (2.58) turns out to be

$$\begin{aligned} \dot{e} &= \frac{r}{h}\{[\cos\theta + e + (1 + e\cos\theta)\cos\theta]\, f^\theta + (1 + e\cos\theta)\, f^r \sin\theta\} \\ &= \frac{r}{h}\left[\left(2\cos\theta + e + e\cos^2\theta\right) f^\theta + (1 + e\cos\theta)\, f^r \sin\theta\right]. \end{aligned} \tag{2.60}$$

Along the $\hat{\mathbf{p}}$ direction we have instead

$$\begin{aligned} e\left(\dot{\Omega}\cos i + \dot{\omega}\right) &= \frac{r}{h} f^\theta \sin\theta - \frac{h}{\mu} f^r \cos\theta + \frac{h}{\mu} f^\theta \sin\theta \\ &= \frac{r}{h}\left[\left(\sin\theta + \frac{h^2}{\mu r}\sin\theta\right) f^\theta - \frac{h^2}{\mu r} f^r \cos\theta\right] \\ &= \frac{r}{h}\{[\sin\theta + (1 + e\cos\theta)\sin\theta]\, f^\theta - (1 + e\cos\theta)\, f^r \cos\theta\} \\ &= \frac{r}{h}\left[\sin\theta\,(2 + e\cos\theta)\, f^\theta - \cos\theta\,(1 + e\cos\theta)\, f^r\right], \end{aligned} \tag{2.61}$$

which seems to tell us that the time variation in the argument of pericenter $\dot{\omega}$ is related to $\dot{\Omega}$ and thus on f^h. This cannot occur, because the orbital elements defining the shape of the orbit, namely a, e, ω reasonably vary only due to the components of the perturbation acting in the plane f^r, f^θ, while the orbital elements defining the orientation of the orbital plane, i.e. i, Ω, vary due to the out-of-place component of $\mathbf{f}$. Let us assume that the perturbation is only along the out-of-plane direction. Then, the orbital plane will change, in particular, the line of nodes will move. The argument of pericenter is defined

as the angle between this line and the eccentricity vector $\mathbf{e}$, which is constant since we are supposing $f^r = f^\theta = 0$. The $e\dot{\Omega}\cos i$ term in Eq. (2.61) does not refer thus to an absolute displacement of the pericenter, but to a change in the reference system.

Let us summarize the Gauss Planetary Equations derived so far in a more compact form

$$\begin{aligned}
\dot{a} &= \frac{2}{n\sqrt{1-e^2}}\left(e\sin\theta f^r + \frac{p}{r}f^\theta\right), \\
\dot{e} &= \frac{\sqrt{1-e^2}}{na}\left[\sin\theta f^r + (\cos\theta + \cos E)\, f^\theta\right], \\
\dot{i} &= \frac{r}{na^2\sqrt{1-e^2}}\cos(\theta+\omega) f^h, \\
\dot{\Omega} &= \frac{r}{na^2\sqrt{1-e^2}}\frac{\sin(\theta+\omega)}{\sin i} f^h, \\
\dot{\omega} &= \frac{\sqrt{1-e^2}}{nae}\left[-\cos\theta f^r + \left(1+\frac{1}{1+e\cos\theta}\right)\sin\theta f^\theta - \dot{\Omega}\cos i\right].
\end{aligned} \tag{2.62}$$

In the original formulation, instead of $\dot{a}$, we find

$$\dot{p} = \frac{2p}{h} r f^\theta. \tag{2.63}$$

Concerning the time parameter, it can be proved that its variation is given by

$$\dot{M} = n - \frac{2r}{\sqrt{\mu a}} f^r - \sqrt{1-e^2}\left(\dot{\omega} + \dot{\Omega}\cos i\right), \tag{2.64}$$

or, equivalently,

$$\dot{M} = n + \frac{1-e^2}{nae}\left[\left(-\frac{2e}{1+e\cos\theta} + \cos\theta\right) f^r - \left(1+\frac{1}{1+e\cos\theta}\right)\sin i f^\theta\right], \tag{2.65}$$

We notice that some of Eqs. (2.62) are not determined for $i = 0$ ($\dot{\Omega}$) and $e = 0$ ($\dot{\omega}$). This is due in particular to the choice of the Euler angles and to the fact that the argument of pericenter is no longer defined for circular orbits. As we mentioned in the previous chapter, in those cases it is more convenient to adopt a different set of orbital elements.

2.3.2 Lagrange Planetary Equations

It is possible to prove that if the perturbation is conservative, that is, it can be written as the gradient of a potential function, namely,

$$\mathbf{f} = \frac{\partial \mathcal{R}}{\partial \mathbf{r}}, \tag{2.66}$$

where $\mathcal{R}$ is the so-called *perturbing function*, then we can apply the *Lagrange planetary equations* to compute the corresponding variation in orbital elements. They are

$$\begin{aligned}
\dot{a} &= \frac{2}{na}\frac{\partial \mathcal{R}}{\partial M}, \\
\dot{e} &= \frac{1-e^2}{na^2 e}\frac{\partial \mathcal{R}}{\partial M} - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial \mathcal{R}}{\partial \omega}, \\
\dot{i} &= -\frac{1}{na^2\sqrt{1-e^2}\sin i}\left(\frac{\partial \mathcal{R}}{\partial \Omega} - \cos i \frac{\partial \mathcal{R}}{\partial \omega}\right), \\
\dot{\Omega} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial \mathcal{R}}{\partial i}, \\
\dot{\omega} &= \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial \mathcal{R}}{\partial e} - \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial \mathcal{R}}{\partial i}, \\
\dot{M} &= n - \frac{2}{na}\frac{\partial \mathcal{R}}{\partial a} - \frac{1-e^2}{na^2 e}\frac{\partial \mathcal{R}}{\partial e}.
\end{aligned} \tag{2.67}$$

We summarize the Gauss and Lagrange planetary equations in Tabs. 2.1-2.2.

	f^r	f^θ	f^h
Δa	$\surd$	$\surd$	
Δe	$\surd$	$\surd$	
Δi			$\surd$
$\Delta\Omega$			$\surd$
$\Delta\omega$	$\surd$	$\surd$	$\surd$
ΔM	$\surd$	$\surd$	$\surd$

Table 2.1: The table shows when a given orbital element will change, depending on the components of the perturbation $\mathbf{f}$ in the polar reference system $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{h}}\}$.

	$\mathcal{R}(a)$	$\mathcal{R}(e)$	$\mathcal{R}(i)$	$\mathcal{R}(\Omega)$	$\mathcal{R}(\omega)$	$\mathcal{R}(M)$
Δa						$\surd$
Δe					$\surd$	$\surd$
Δi				$\surd$	$\surd$	
$\Delta\Omega$			$\surd$			
$\Delta\omega$		$\surd$	$\surd$			
ΔM	$\surd$	$\surd$				

Table 2.2: The table shows when a given orbital element will change, depending on the perturbing function $\mathcal{R}$.

Major Orbital Perturbations on Earth Satellites

3

In this chapter we present the major orbital perturbations acting on satellites around the Earth. Each contribution will be first modeled, and then the associated effect studied by means of either the Gauss or the Lagrange planetary equations, depending whether the perturbation is conservative or not. In particular, we will deal with:

- the fact that the Earth is not a perfect sphere with a radially symmetrical internal distribution of mass;
- the atmospheric drag;
- the solar radiation pressure;
- the presence of other massive bodies, in particular Moon and Sun.

In Fig. 3.1 the order of magnitude of the corresponding (excluded drag) acceleration is given as a function of the distance from the center of the Earth (Valk, Lemaître & Anselmo, 2008). In a nutshell, the main perturbations acting on satellites in LEO are the J_2 term and the atmospheric drag; in MEO J_2 and luni-solar perturbations; in GEO J_{22} and luni-solar perturbations. The solar radiation pressure must be taken into account in MEO and GEO for high enough values of area-to-mass ratio. There exist other effects that can play a role on the orbit of a spacecraft around the Earth, for instance, tidal friction, albedo, relativistic effects, but their order of magnitude is less significant.

The effects due to a given orbital perturbation will be distinguished in

- *secular*, if they do not depend on the mean (or true) anomaly nor in the argument of perigee;
- *long-period*, if they depend on the argument of perigee, but not on the mean anomaly;
- *short-period*, if they depend on the mean anomaly.

For a conservative perturbation, the semi-major axis does not experience any secular or long-period change.

Finally, we will also describe the station-keeping strategies, to be applied to play against the long-period variations induced by J_{22}, atmospheric drag and luni-solar perturbations.

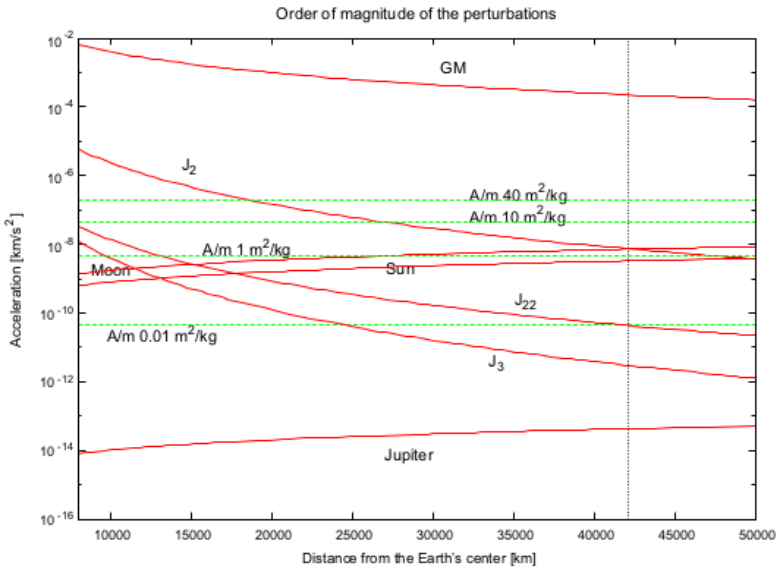


Figure 3.1: Order of magnitude of the major perturbing accelerations acting on satellites around the Earth as a function of the distance to the Earth's center (Valk, Lemaître & Anselmo, 2008). The effect due to the atmospheric drag is not shown. The one corresponding to the solar radiation pressure depends on the area-to-mass A/m ratio.

3.1 Earth Gravitational Potential

Let us consider a satellite subject to the gravitational potential of the Earth, and let $\mathbf{r}$ be its geocentric radius vector. In an inertial reference system whose origin is located at the center of mass of the Earth, the equation of motion of the satellite can be written as

$$\frac{d^2\mathbf{r}}{dt^2} = \frac{\partial \mathcal{U}(\mathbf{r})}{\partial \mathbf{r}}, \tag{3.1}$$

where $\mathcal{U}(\mathbf{r})$ is the gravitational potential associated with the Earth. In Chap. 1, we made the assumption that the central body, the Earth in this case, is a point mass, but in real world situations this is never true. However, that hypothesis holds if the central body is perfectly spherical and the total mass is distributed uniformly in its interior.

We recall that the gravitational potential $\mathcal{U}(\mathbf{r})$ must satisfy the Laplace's equation in any region exterior to the attracting mass, namely,

$$\nabla^2 \mathcal{U} = 0. \tag{3.2}$$

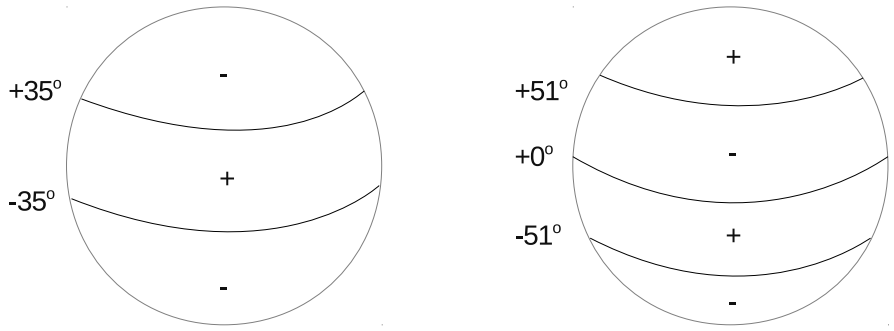


Figure 3.2: Zonal harmonics of degree 2 (left) and 3 (right).

A function $\mathcal{U}(\mathbf{r})$ fulfilling the above equation is said *harmonic function*. By considering the Earth as extended attracting body, the general solution of (3.2) can be expressed as

$$\mathcal{U}(\mathbf{r}) = \frac{\mu}{r}\left\{1 + \sum_{l=2}^{\infty}\sum_{m=0}^{l} P_{lm}\left(\sin\phi\right)\left(\frac{r_{\oplus}}{r}\right)^{l}\left[C_{lm}\cos m\lambda + S_{lm}\sin m\lambda\right]\right\}, \tag{3.3}$$

where

- $r_{\oplus}$ the equatorial radius of the Earth (or any massive body);
- ϕ the geocentric latitude;
- λ the longitude;
- $P_{lm}\left(\sin\phi\right)$ the associated Legendre polynomials of degree l and order m;
- C_{lm}, S_{lm} coefficients determined experimentally.

Moreover, μ corresponds to $\mu_{\oplus}$, but the treatment of the problems is general. We notice that in (3.3) μ/r is the fundamental harmonic or monopole which depends only on r and corresponds to the potential exerted by a perfect sphere. Since in (3.3) we use spherical coordinates we speak of *spherical harmonics*: they are periodic functions in the unit sphere.

If the planet is characterized by axial symmetry, then also the gravitational field is symmetric with respect to the polar axis and $\mathcal{U}(\mathbf{r})$ is independent from the longitude. This is $m = 0$ and

$$\mathcal{U}(\mathbf{r}) = \frac{\mu}{r}\left\{1 - \sum_{l=2}^{\infty} P_{l}\left(\sin\phi\right)\left(\frac{r_{\oplus}}{r}\right)^{l} J_{l}\right\}, \tag{3.4}$$

where $P_l(\sin\phi) \equiv P_{l0}(\sin\phi)$ is said *zonal harmonic* and $J_l = -C_l$. In (3.3) or in (3.4) J_0 is the fundamental harmonic, while J_1 is missing because we set the origin of the reference system at the center of mass of the Earth.

These harmonics are called zonal because for any l there exist l latitude circles along to $P_l(\sin\phi) = 0$ and thus $(l+1)$ regions where the function is alternately positive, i.e., increasing, or negative, i.e., decreasing. For instance, for $l = 2$ we have

$$P_2(\sin\phi) = \frac{1}{2}\left(3\sin^2\phi - 1\right) = 0 \qquad \Longrightarrow \qquad \phi \approx \pm 35^\circ, \tag{3.5}$$

and for $l = 3$

$$P_3(\sin\phi) = \frac{1}{2}\sin\phi\left(5\sin^2\phi - 3\right) = 0 \qquad \Longrightarrow \qquad \phi = 0 \text{ or } \phi \approx \pm 51^\circ. \tag{3.6}$$

In Fig. 3.2 we show the corresponding regions on a planet. Concerning the J_l coefficients, in Tab. 3.1 we give the first values in the case of the Earth.

Zonal Coefficient	**Approximated Value**
J_2	1.08×10^{-3}
J_3	-2.53×10^{-6}
J_4	-1.61×10^{-6}
J_5	-2.27×10^{-7}
J_6	5.40×10^{-7}

Table 3.1: Approximated values of the first zonal coefficients in the case of the Earth.

If $m \neq 0$ and $m = l$, then we deal with the so-called *sectorial harmonics*, namely the functions

$$P_{ll}(\sin\phi) \begin{matrix}\sin(l\lambda) \\ \cos(l\lambda)\end{matrix}, \tag{3.7}$$

which do not depend on the latitude. Indeed, since

$$P_{ll}(\sin\phi) = \frac{(2l)!}{2^l l!}(\cos^2\phi)^{\frac{l}{2}},$$

the only values of ϕ such that a sectorial harmonic vanishes correspond to the poles, that is, when $\phi = \pm\pi/2$. On the other hand, $\sin(l\lambda)$ and $\cos(l\lambda)$ cancel out for $2l$ different values of λ, that is along $2l$ meridians and the sphere is split in $2l$ orange slices, see Fig. 3.3 on the left.

Otherwise, if $m \neq 0$ and $m \neq l$ we have the *tesseral harmonics*, that is, the functions

$$P_{lm}(\sin\phi) \begin{matrix}\sin(m\lambda) \\ \cos(m\lambda)\end{matrix}. \tag{3.8}$$

In this case we have $(l - m)$ values of longitude along each meridian such that the harmonic is zero and $2m$ values of latitude along each parallel and therefore the sphere is split in $2m(l - m + 1)$ tiles, see Fig. 3.3 on the right.

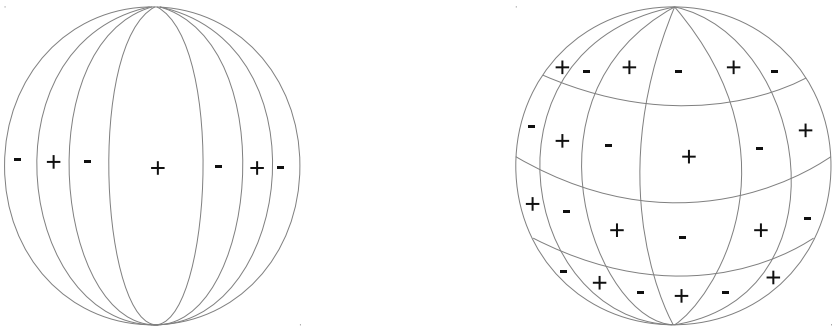


Figure 3.3: Sectorial harmonics of degree and order 7 (left) and tesseral harmonics of degree 9 and order 6 (right).

In what follows, we analyze the consequences on the orbit of a satellite around the Earth due to terms associated with J_2, J_3 and J_{22}. To this end, we remark that an expression for the gravitational potential $\mathcal{U}(\mathbf{r})$, equivalent to (3.3), is

$$\mathcal{U}(\mathbf{r}) = \frac{\mu}{r}\left\{1 - \sum_{l=2}^{\infty} J_l P_l\left(\sin\phi\right)\left(\frac{r_\oplus}{r}\right)^l + \sum_{l=2}^{\infty}\sum_{m=1}^{l} J_{lm} P_{lm}\left(\sin\phi\right)\left(\frac{r_\oplus}{r}\right)^l \cos\left(m\lambda - m\lambda_{lm}\right)\right\}, \tag{3.9}$$

where $J_{lm} = \sqrt{S_{lm}^2 + C_{lm}^2}$ and

$$\lambda_{lm} = \frac{1}{m}\tan^{-1}\left(\frac{S_{lm}}{C_{lm}}\right) \tag{3.10}$$

are the equilibrium longitudes for J_{lm}, whose meaning will be specified later.

Finally, we notice that the perturbation due to the mass distribution of the Earth is conservative: it depends only on position, not on velocity and it can be described by means of a potential function.

3.1.1 Effects due to J_2

The term of the potential associated with J_2 is responsible of the oblateness of the Earth. In particular, it is known that the equatorial diameter of the Earth is longer than the polar diameter of about 20 km. We see here the secular effects on the orbital elements due to that.

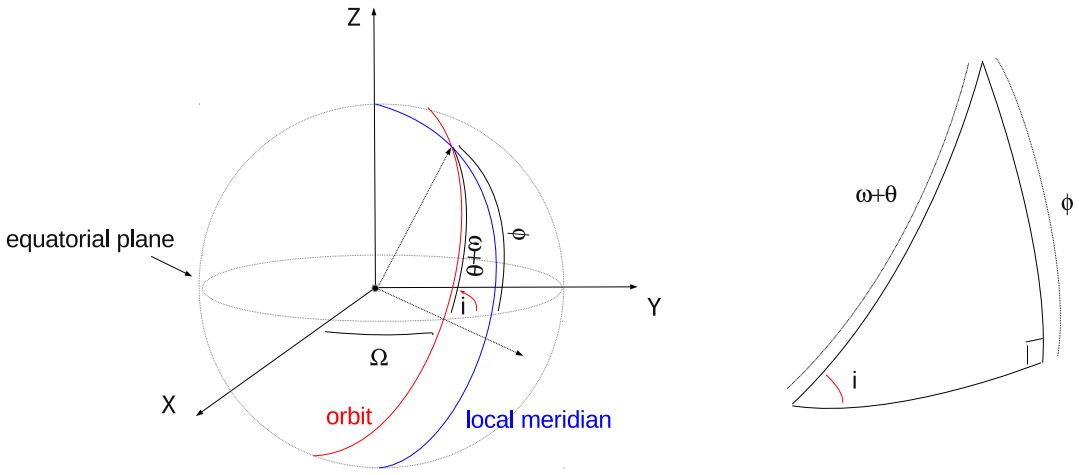


Figure 3.4: The relationship between geocentric latitude and the angular orbital elements of the orbit.

By considering only the term of degree 2, Eq. (3.4) reads

$$\begin{aligned} \mathcal{U}(\mathbf{r}) &= \frac{\mu}{r}\left[1 - P_2\left(\sin\phi\right)\left(\frac{r_\oplus}{r}\right)^2 J_2\right] \\ &= \frac{\mu}{r}\left[1 - \frac{1}{2}\left(3\sin^2\phi - 1\right)\left(\frac{r_\oplus}{r}\right)^2 J_2\right], \end{aligned} \tag{3.11}$$

that is, the perturbing function due to J_2 is

$$\mathcal{R} = -\frac{\mu}{2} J_2 r_\oplus^2 \frac{(1 + e\cos\theta)^3}{p^3}\left(3\sin^2 i \sin^2(\omega + \theta) - 1\right). \tag{3.12}$$

Indeed, if we refer to Fig. 3.4 and apply the rules of spherical trigonometry we get

$$\frac{\sin\phi}{\sin i} = \frac{\sin\left(\omega + \theta\right)}{\sin\left(\pi/2\right)}. \tag{3.13}$$

Since $\mathcal{R}$ is a periodic function in θ of period 2π, it can be expanded as a Fourier series

$$\mathcal{R} = \bar{\mathcal{R}} + \sum_{n=1}^{\infty} a_n \cos n\theta + b_n \sin n\theta, \tag{3.14}$$

where $\bar{\mathcal{R}}$ is the mean value of the perturbing function over one orbit, namely,

$$\bar{\mathcal{R}} = \frac{1}{2\pi}\int_0^{2\pi} \mathcal{R} dM. \tag{3.15}$$

From the definition of mean anomaly, recall Eq. (1.75), we have

$$dM = ndt,$$

and from the angular momentum Eq. (1.16)

$$dt = \frac{r^2}{h}d\theta.$$

This is,

$$dM = n\frac{r^2}{h}d\theta, \tag{3.16}$$

and thus, using also Eqs. (1.23) and (1.81),

$$\begin{aligned}
\bar{\mathcal{R}} &= \frac{1}{2\pi}\int_0^{2\pi}\mathcal{R}n\frac{r^2}{h}d\theta \\
&= -\frac{\mu r_{\oplus}^2}{4\pi}J_2\int_0^{2\pi}\frac{(1+e\cos\theta)^3}{p^3}\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]n\frac{r^2}{h}d\theta \\
&= -\frac{\mu r_{\oplus}^2}{4\pi}J_2\int_0^{2\pi}\frac{1+e\cos\theta}{p}\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]\sqrt{\frac{\mu}{a^3}}\frac{1}{\sqrt{p\mu}}d\theta \\
&= -\frac{J_2 r_{\oplus}^2}{4\pi}\frac{\mu}{a^3(1-e^2)^{3/2}}\int_0^{2\pi}(1+e\cos\theta)\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]d\theta. \qquad (3.17)
\end{aligned}$$

By solving the integral, we get to

$$\bar{\mathcal{R}} = \frac{J_2 r_{\oplus}^2}{4}\frac{\mu}{a^3(1-e^2)^{3/2}}\left(2-3\sin^2 i\right). \tag{3.18}$$

Indeed,

$$\begin{aligned}
\int_0^{2\pi}(1+e\cos\theta)\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]d\theta &= \int_0^{2\pi}\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]d\theta + \\
&\quad \int_0^{2\pi}e\cos\theta\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]d\theta,
\end{aligned}$$

with

$$\begin{aligned}
\int_0^{2\pi}\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]d\theta &= 3\sin^2 i\frac{1}{2}\left[\omega+\theta-\sin(\omega+\theta)\cos(\omega+\theta)\right]|_0^{2\pi}-2\pi \\
&= 3\sin^2 i\frac{1}{2}\left[\omega+2\pi-\omega-\sin\omega\cos\omega+\sin\omega\sin\omega\right]-2\pi \\
&= 3\pi\sin^2 i-2\pi,
\end{aligned}$$

and

$$\begin{aligned}
\int_0^{2\pi}e\cos\theta\left[3\sin^2 i\sin^2(\omega+\theta)-1\right]d\theta &= \cancelto{0}{\left[3e\sin^2 i\sin\theta\sin^2(\omega+\theta)\right]|_0^{2\pi}} - \\
&\quad \cancelto{0}{\int_0^{2\pi}3e\sin^2 i2\sin(\omega+\theta)\cos(\omega+\theta)\sin\theta d\theta}.
\end{aligned}$$

We notice that (3.18) depends only the orbital elements a, e, i and thus represents the secular perturbation due to J_2. Also, the Lagrange planetary equations (2.67) applied to $\bar{\mathcal{R}}$ provide

$$\frac{da}{dt} = \frac{de}{dt} = \frac{di}{dt} = 0, \tag{3.19}$$

while for the longitude of the ascending node Ω the variation in time is

$$\begin{aligned}
\frac{d\Omega}{dt} &= \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial \bar{\mathcal{R}}}{\partial i} \\
&= -\frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{J_2 r_\oplus^2}{4}\frac{\mu}{a^3(1-e^2)^{3/2}} 6\sin i\cos i \\
&= -\frac{3}{2}\frac{J_2 r_\oplus^2 n}{p^2}\cos i,
\end{aligned} \tag{3.20}$$

for the argument of perigee ω it is

$$\begin{aligned}
\frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial \bar{\mathcal{R}}}{\partial e} - \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial \bar{\mathcal{R}}}{\partial i} \\
&= \frac{\sqrt{1-e^2}}{na^2 e}\frac{J_2 r_\oplus^2}{4}\frac{\mu}{a^3}\left(2-3\sin^2 i\right)\frac{3e}{(1-e^2)^{5/2}} + \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{J_2 r_\oplus^2}{4}\frac{\mu}{a^3(1-e^2)^{3/2}} 6\sin i\cos i \\
&= \frac{J_2 r_\oplus^2 n}{4a^2(1-e^2)^2}\left(6+9\cos^2 i - 9 + 6\cos^2 i\right) \\
&= \frac{J_2 r_\oplus^2 n}{4p^2}\left(-3+15\cos^2 i\right) \\
&= \frac{3}{4}\frac{J_2 r_\oplus^2 n}{p^2}\left(5\cos^2 i - 1\right),
\end{aligned} \tag{3.21}$$

and for the mean anomaly M it is

$$\begin{aligned}
\frac{dM}{dt} &= n - \frac{2}{na}\frac{\partial \bar{\mathcal{R}}}{\partial a} - \frac{1-e^2}{na^2 e}\frac{\partial \bar{\mathcal{R}}}{\partial e} \\
&= n + \frac{2}{na}3\frac{J_2 r_\oplus^2}{4}\frac{\mu}{a^4(1-e^2)^{3/2}} - \frac{1-e^2}{na^2 e}\frac{J_2 r_\oplus^2}{4}\frac{\mu}{a^3}\left(2-3\sin^2 i\right)\frac{3e}{(1-e^2)^{5/2}} \\
&= n + \frac{3n}{a^2}\frac{J_2 r_\oplus^2}{2(1-e^2)^2}(1-e^2)^{1/2} - \frac{n(1-e^2)^{1/2}}{a^2}\frac{J_2 r_\oplus^2}{4}\left(2-3\sin^2 i\right)\frac{3}{(1-e^2)^2} \\
&= n + \frac{3}{2}\frac{J_2 r_\oplus^2 n}{p^2}\left(1-\frac{3}{2}\sin^2 i\right)\sqrt{1-e^2}.
\end{aligned} \tag{3.22}$$

We notice that if $i \in (0, 90^\circ)$, then $\dot{\Omega} < 0$. This is, for prograde orbits the line of nodes regresses (see Fig. 3.5), for retrograde orbits it precesses. We recall that for an observer looking to the North pole from above the prograde motion is counterclockwise, the retrograde one instead is clockwise.

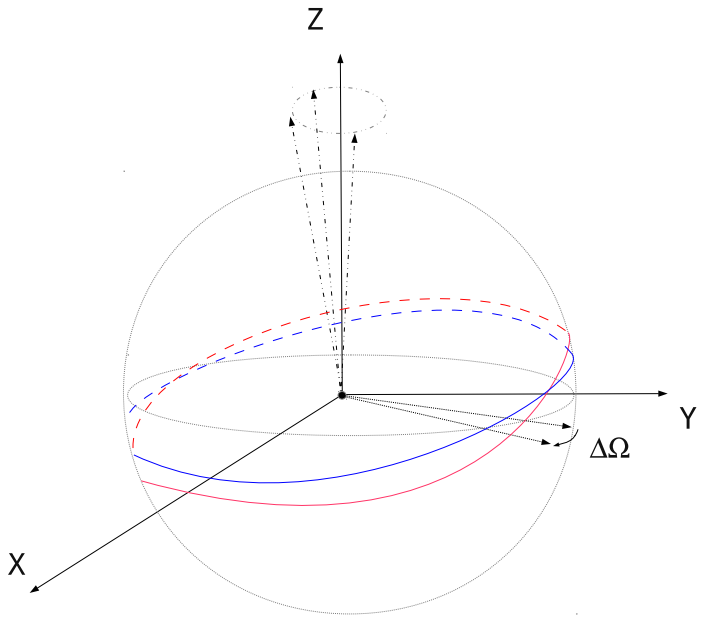


Figure 3.5: Due to J_2 the longitude of ascending node of prograde orbits moves westward. The orbital plane tends to turn toward the equator and the angular momentum vector rotates about the polar axis because of that.

On the other hand, the line of apsides does not move because of the perturbation due to J_2 if

$$5\cos^2 i - 1 = 0,$$

that is, for the *critical values of inclination*

$$i_1 \approx 63.43^\circ, \qquad i_2 \approx 116.57^\circ,$$

whose values hold for any orbit and massive body, i.e., not only for the Earth, that is, they are independent from the other orbital elements and the value of J_2. If $i < i_1$ then the line of apsides advances; if $i > i_1$ it regresses.

3.1.2 Effects due to J_3

The spherical harmonic corresponding to J_3 is responsible of the 'pear-shape' of the Earth and it provides a long-term periodic perturbation on the orbit of the satellite.

The perturbing function associated with the long-period effects due to J_3 is

$$\mathcal{R} = \frac{3}{2}\mu J_3 \frac{r_\oplus^3}{a^4} \frac{e}{(1-e^2)^{\frac{5}{2}}} \sin i \left(1 - \frac{5}{4}\sin^2 i\right) \sin\omega, \tag{3.23}$$

which does not depend on the true anomaly. The just written expression is derived from Eq. (3.3), using the identity

$$\sin^3(\theta+\omega) = \frac{3}{4}\sin(\theta+\omega) - \frac{1}{4}\sin(3\theta+3\omega),$$

and taking the mean value with respect to the mean anomaly (Roy, 2005). The perturbing function (3.23) depends on the argument of perigee ω and thus now $\dot{e} \neq 0$ and $\dot{i} \neq 0$. In particular, we have

$$\frac{\partial \mathcal{R}}{\partial \omega} = \frac{3}{2}\mu J_3 \frac{r_\oplus^3}{a^4} \frac{e}{(1-e^2)^{\frac{5}{2}}} \sin i \left(1 - \frac{5}{4}\sin^2 i\right) \cos\omega, \tag{3.24}$$

and therefore, by applying the Lagrange planetary equations, the variation in eccentricity due to J_3 is

$$\begin{aligned} \frac{de}{dt} &= -\frac{3}{2}\frac{\sqrt{1-e^2}}{na^2 e}\mu J_3 \frac{r_\oplus^3}{a^4} \frac{e}{(1-e^2)^{\frac{5}{2}}} \sin i \left(1 - \frac{5}{4}\sin^2 i\right) \cos\omega \\ &= -\frac{3}{2} n J_3 \frac{r_\oplus^3}{a^3(1-e^2)^2} \sin i \left(1 - \frac{5}{4}\sin^2 i\right) \cos\omega, \end{aligned} \tag{3.25}$$

and the one in inclination is

$$\begin{aligned} \frac{di}{dt} &= \frac{3}{2}\frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\mu J_3 \frac{r_\oplus^3}{a^4} \frac{e}{(1-e^2)^{\frac{5}{2}}} \sin i \left(1 - \frac{5}{4}\sin^2 i\right) \cos\omega \\ &= \frac{3}{2} e J_3 \frac{r_\oplus^3}{a^3(1-e^2)^3} \cos i \cos\omega \left(1 - \frac{5}{4}\sin^2 i\right). \end{aligned} \tag{3.26}$$

In order to get the long-term period effect, we integrate the above expressions with respect to ω, namely,

$$\begin{aligned} \Delta i &= \int_0^{\omega^*} \frac{di}{d\omega} d\omega = \int_0^{2\pi} \frac{di}{dt}\frac{dt}{d\omega} d\omega = \\ &= \int_0^{\omega^*} \left[\frac{3}{2} e J_3 \frac{r_\oplus^3}{a^3(1-e^2)^3} \cos i \cos\omega \left(1 - \frac{5}{4}\sin^2 i\right)\right] \left[3\frac{J_2 r_\oplus^2 n}{p^2}\left(1 - \frac{5}{4}\sin^2 i\right)\right]^{-1} d\omega \\ &= \int_0^{\omega^*} \frac{1}{2}\frac{J_3}{J_2}\frac{r_\oplus}{a}\frac{e}{1-e^2} \cos i \cos\omega d\omega \\ &= \frac{1}{2}\frac{J_3}{J_2}\frac{r_\oplus}{a}\frac{e}{1-e^2} \cos i \sin\omega^*. \end{aligned} \tag{3.27}$$

In an analogous way, for the eccentricity we find

$$\Delta e = -\frac{1}{2}\frac{J_3}{J_2}\frac{r_\oplus}{a}\frac{e}{1-e^2}\sin i \sin \omega^*. \tag{3.28}$$

We notice that in (3.27) we exploit the fact that the argument of perigee varies because of J_2, according to Eq. (3.21). This is justified by the order of magnitude of J_2 which is about 2.5 times larger than the one associated with J_3 (see Tab. 3.1).

Moreover, Eqs. (3.20) and (3.27) tell us that for polar orbits secular and long-period effects in Ω and i do not take place. For equatorial orbits instead the eccentricity is constant.

3.1.3 Effects due to J_{22}

The term $J_{22} = \sqrt{C_{22}^2 + S_{22}^2} \approx 1.8 \times 10^{-6}$ is associated with the fact that the equatorial section of the Earth is not a circle, but an ellipse. We consider the consequences of that on a geostationary orbit (see Sec. 1.14.2), for which the effect is more pronounced. We recall that such orbit is characterized by

- the orbital period is equal to the period of rotation of the Earth around its spin axis;
- the orbit is circular, i.e., $e = 0$;
- the orbit lies on the equatorial plane, i.e., $i = 0$.

It follows that $r = a = 42164$ km and $v = \sqrt{\mu/r} \approx 3.075$ km/s. From the point of view of an observer on the Earth's surface, a geostationary satellite appears fixed in the sky, that is, the ground track is a point on the equator with latitude $\phi = 0$ and given longitude λ. The effect due to the ellipticity of the Earth's equator is to make the satellite moving with respect to this fixed position, as seen from the Earth.

Let us consider a reference system which rotates together with the Earth, see Fig. 3.6. If the equatorial section was circular, in any point of the orbit there would exist only a central component of acceleration, say a_r. Since it is elliptic, there also exists a transversal component, say a_t, directed towards the major axis of the ellipse corresponding to the equatorial section. On the points labelled as $1, 2, 3, 4$ in Fig. 3.6 $a_t = 0$: these are equilibrium points; $2, 4$ are in particular unstable, while $1, 3$ stable. Though the ellipticity tends to move the satellite toward the major axis of the Earth's equator, what happens is that the satellite moves toward the minor axis. We have thus a longitudinal drift, which depends on the position of the satellite with respect to the stable equilibrium points.

In order to explain this paradox, let us recall Eq. (1.81), namely,

$$\mu = n^2 a^3,$$

and derive it with respect to time, that is,

$$0 = 2n\frac{dn}{dt}a^3 + 3a^2n^2\frac{da}{dt} \qquad \Longrightarrow \qquad \frac{dn}{dt} = -\frac{3}{2}\frac{n}{a}\frac{da}{dt}. \tag{3.29}$$

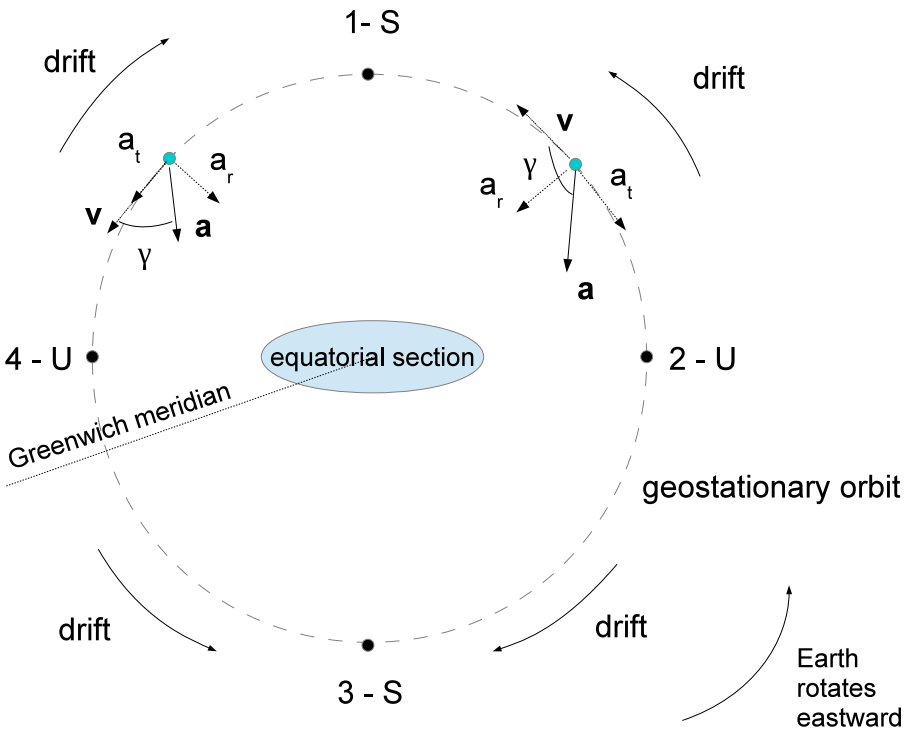


Figure 3.6: The acceleration due to the ellipticity of the Earth's equator on a geostationary orbit. The points labeled as U are the unstable equilibrium points, the ones labeled as S the stable ones.

Also, from Eq. (1.37), namely,

$$\mathcal{E} = -\frac{\mu}{2a},$$

we get to

$$\frac{d\mathcal{E}}{dt} = \frac{\mu}{2a^2}\frac{da}{dt}. \tag{3.30}$$

Since it also holds (see Eq. (2.26)) that

$$\frac{d\mathcal{E}}{dt} = \mathbf{a}\cdot\mathbf{v}, \tag{3.31}$$

where $\mathbf{v}$ is the velocity on the orbit and $\mathbf{a} = (a_r, a_t)$ is the acceleration due to J_{22}, we have

$$\frac{da}{dt} = \frac{2a^2}{\mu}\mathbf{a}\cdot\mathbf{v}. \tag{3.32}$$

This tells us that

- if $\mathbf{a}\cdot\mathbf{v}>0$, that is, the two vectors form an angle $\gamma<90^\circ$, then $\dot{a}>0$, the satellite moves on a higher orbit and the mean motion decreases $\dot{n}<0$, i.e., the orbital period increases. The satellite slows down and goes toward West.
- If $\mathbf{a}\cdot\mathbf{v}<0$, that is, $\gamma>90^\circ$, then $\dot{a}<0$ and $\dot{n}>0$. The satellite speeds up and moves toward East, like if it preceded the Earth.

Let us see specifically how the longitude corresponding to the satellite ground track behaves. The perturbing function due to J_{22} is

$$\mathcal{R}=\frac{\mu}{r}J_{22}\left(\frac{r_\oplus}{r}\right)^2 P_{22}(\sin\phi)\cos\left(2\lambda-2\lambda_{22}\right), \tag{3.33}$$

where λ_{22} is the longitude corresponding to the equilibrium positions mentioned above, that is,

$$\lambda_{22}=\frac{1}{2}\tan^{-1}\left(\frac{S_{22}}{C_{22}}\right)=\begin{cases}75.09^\circ \text{ East: stable (point 3 in Fig. 3.6)}\\ 165.09^\circ \text{ East: unstable (point 2 in Fig. 3.6)}\\ 104.91^\circ \text{ West: stable (point 1 in Fig. 3.6)}\\ 14.91^\circ \text{ West: unstable (point 4 in Fig. 3.6)}\end{cases} \tag{3.34}$$

Since $e=0$ and $i=0$, the Lagrange planetary equations applied to the semi-major axis read

$$\frac{da}{dt}=\frac{2}{na}\frac{\partial\mathcal{R}}{\partial M}=\frac{2}{na}\frac{\partial\mathcal{R}}{\partial\lambda}, \tag{3.35}$$

or, equivalently,

$$\begin{aligned}\frac{d^2\lambda}{dt^2} &= \frac{dn}{dt}=-\frac{3}{2}\frac{n}{a}\frac{da}{dt}=-\frac{3}{2}\frac{n}{a}\frac{2}{na}\frac{\partial\mathcal{R}}{\partial\lambda}\\ &= \frac{3}{a^2}\frac{\mu}{r}J_{22}\left(\frac{r_\oplus}{r}\right)^2 P_{22}(\sin\phi)2\sin\left(2\lambda-2\lambda_{22}\right).\end{aligned} \tag{3.36}$$

The Legendre polynomial $P_{22}(\sin\phi)$ for $\phi=0$ is equal to

$$P_{22}(\sin\phi)=3(1-\sin^2\phi)=3, \tag{3.37}$$

and this is a further confirmation that the effect is more relevant for equatorial orbits. Geostationary orbits have in addition the property to be in resonance with the Earth's rotation period.

Because of (3.37), Eq. (3.36) becomes

$$\frac{d^2\lambda}{dt^2}=18J_{22}\left(\frac{r_\oplus}{r}\right)^2 n^2\sin\left(2\lambda-2\lambda_{22}\right), \tag{3.38}$$

which means that the satellite oscillates around the equilibrium positions associated with λ_{22}. This is indeed the equation of the pendulum, for which we can depict a phase portrait like in Fig. 3.7. Let us multiply Eq. (3.38) by $\dot{\lambda}$, namely,

$$\dot{\lambda}\ddot{\lambda}-18\dot{\lambda}J_{22}\left(\frac{r_\oplus}{r}\right)^2 n^2\sin\left(2\lambda-2\lambda_{22}\right)=0,$$

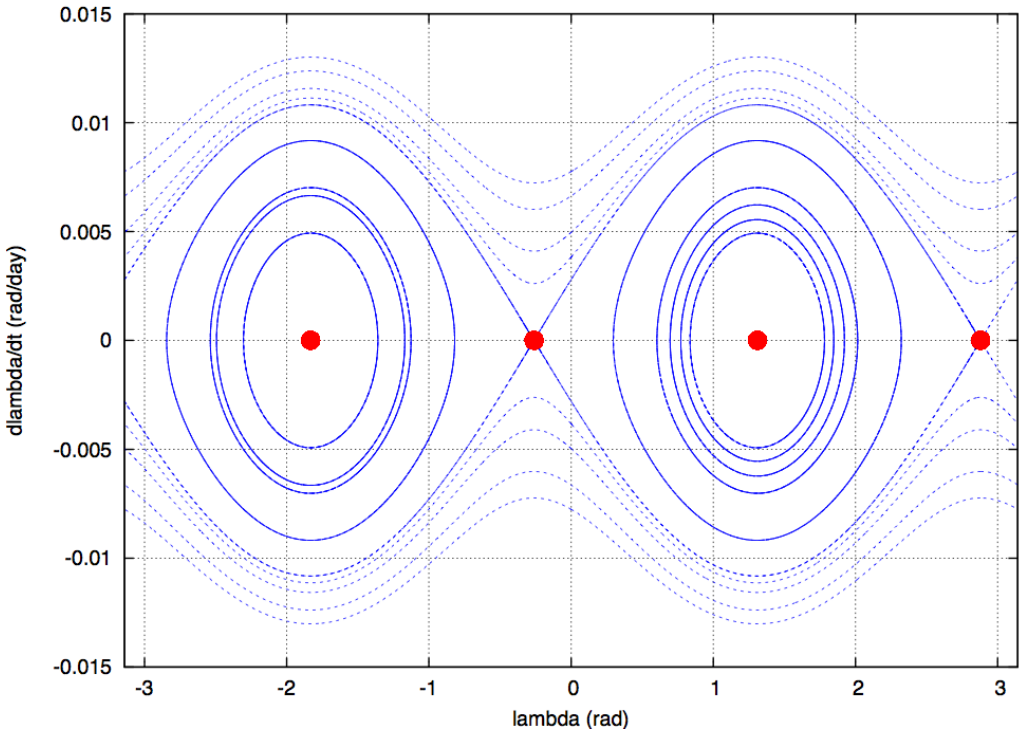


Figure 3.7: Phase portrait corresponding to the drift in longitude λ for geostationary orbits due to J_{22}. The red points are the equilibrium positions λ_{22}, the motion is clockwise.

that is,

$$\frac{d}{dt}\left(\frac{1}{2}\dot{\lambda}^2\right) + \frac{18}{2}J_{22}\left(\frac{r_\oplus}{r}\right)^2 n^2 \frac{d\cos\left(2\lambda - 2\lambda_{22}\right)}{dt} = 0.$$

By integration in time, we obtain

$$\dot{\lambda}^2 + 18J_{22}\left(\frac{r_\oplus}{r}\right)^2 n^2 \cos\left(2\lambda - 2\lambda_{22}\right) = \dot{\lambda}^2 + 18J_{22}\left(\frac{r_\oplus}{r}\right)^2 n^2 \left[1 - 2\sin^2\left(\lambda - \lambda_{22}\right)\right] = constant,$$

that is,

$$\dot{\lambda}^2 - 36J_{22}\left(\frac{r_\oplus}{r}\right)^2 n^2 \sin^2\left(\lambda - \lambda_{22}\right) = constant. \tag{3.39}$$

As a function of the value of this constant, which depends on the initial conditions considered, we find a librational or a rotational motion.

3.1.4 Applications: Sun-Synchronous Orbits and Molniya Orbits

A *sun-synchronous orbit* is an orbit such that it passes above the same point on the Earth at the same local solar hour. This always ensures equal lighting conditions and it is useful especially for meteorological and atmospheric studies and remote sensing. The orbital plane maintains a fixed orientation with respect to the Sun–Earth direction all

over the year. In order to satisfy this constraint, the secular variation of Ω must be equal to the mean velocity of the Sun in its apparent motion at the Earth (see Sec. 1.14.2), namely,

$$\frac{d\Omega}{dt} = \frac{360^\circ}{365.25} \approx \frac{0.986^\circ}{day}, \tag{3.40}$$

which is a positive quantity and thus the sun-synchronous orbit must be retrograde, the line of nodes moves toward East and chases the Sun. Typically sun-synchronous orbits are quasi-circular with $i \approx 98.6^\circ$ and an altitude of about 800 km. We notice that Eq. (3.20) implies that, assuming $e = 0$, the inclination is determined by the altitude required.

A *Molniya orbit* is a semi-synchronous orbit, i.e., with an orbital period of 12 hours, with high eccentricity $e = 0.73$, critical inclination $i = i_1$ and semi-major axis $a = 26562$ km. Because of the critical inclination, the argument of perigee ω remains constant, typically $\omega = 270^\circ$ and the satellite spends most of the time at the apogee, where it can be observed from the ground stations on the North hemisphere. As a matter of fact, if $\omega = 270^\circ$ then at the apogee the sub-satellite point corresponds to a latitude equal to i_1 North. These orbits are named after Russian communication satellites, which indeed were supposed to be observed from Russia.

Other orbits which take advantage of the critical values of inclination are the *Tundra* orbits, characterized by an orbital period of 24 hours.

3.2 Atmospheric Drag

Now let us consider the consequences due to the atmospheric drag. This is a non-conservative perturbation, directed mostly on the opposite direction of the velocity of the satellite computed with respect to the atmospheric flow. If there also exists a component perpendicular to such velocity, then we speak of *atmospheric lift*. The atmospheric drag has significant effects up to an altitude of about 800–1000 km, because the Earth's atmospheric density decreases exponentially with the altitude. For an elliptic orbit, it must be taken into account at the perigee. The two main difficulties in analyzing this kind of perturbation arise because it is a challenging task to obtain an accurate modeling of the shape of the satellite and its orientation with respect to the surrounding atmosphere and also an accurate modeling of the atmospheric density, which varies as a function of altitude and time.

3.2.1 Modeling

The force due to the interaction between the satellite and the atmosphere is usually decomposed in two directions, the first along the relative velocity satellite-atmosphere, say $\hat{\mathbf{v}}_a$, the other perpendicular to $\hat{\mathbf{v}}_a$, say $\hat{\mathbf{v}}_{a_\perp}$. Let us assume that the atmosphere rotates together with the Earth with an angular velocity $w_\oplus = 4.178 \times 10^{-3}$ deg/s. Then

$$\mathbf{v}_a = \mathbf{v} - \mathbf{w}_\oplus \times \mathbf{r}, \tag{3.41}$$

where $\mathbf{w}_\oplus = w_\oplus \hat{\mathbf{k}}$, and $\mathbf{r}, \mathbf{v}$ are the radius and velocity vector of the satellite in its orbit around the Earth. For satellites on orbits low enough to consider not negligible the effect of the atmosphere, we have

$$||\mathbf{v}|| >> ||\mathbf{w}_\oplus \times \mathbf{r}|| \qquad \Longrightarrow \qquad \mathbf{v}_a \approx \mathbf{v},$$

and the perturbing force can be written as

$$\mathbf{F} = -D\hat{\mathbf{v}} + L\hat{\mathbf{v}}_\perp, \tag{3.42}$$

where D is the *drag* and L is the *lift*.

Let us neglect the term due to the lift and write the drag as

$$D = \frac{1}{2}\rho A v^2 C_D, \tag{3.43}$$

where ρ is the atmospheric density, A is the area of the satellite along the direction of motion, v the modulus of the orbital velocity and C_D the drag coefficient, typically between 1.5 and 3. The perturbing acceleration, as the one considered in Chap. 2, is thus a negative tangential force per unit of mass, namely,

$$\mathbf{f} = \frac{\mathbf{F}}{m} = -\frac{1}{2}\rho \frac{A}{m} v^2 C_D \hat{\mathbf{v}}, \tag{3.44}$$

and the greatest consequences are found for high area-to-mass ratios A/m. Since $\mathbf{f}$ is directed along $\hat{\mathbf{v}}$, it does not have out-of-plane components, that is, in the $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{h}}\}$ reference system it can be decomposed as

$$\mathbf{f} = f^r \hat{\mathbf{r}} + f^\theta \hat{\boldsymbol{\theta}} + \cancel{f^h \hat{\mathbf{h}}}.$$

3.2.2 Effects

By applying the Gauss planetary equations (2.62) to the atmospheric perturbation, it turns out that

$$\frac{di}{dt} = \frac{d\Omega}{dt} = 0. \tag{3.45}$$

In order to find the variation in time on the other orbital elements, let us recall Eq. (1.15), namely,

$$v = \sqrt{v_r^2 + v_\theta^2} = \sqrt{(\dot{r})^2 + (r\dot{\theta})^2},$$

and Eq. (1.67), namely,

$$v_r = \sqrt{\frac{\mu}{a(1-e^2)}} e \sin\theta, \qquad v_\theta = \sqrt{\frac{\mu}{a(1-e^2)}} \left(1 + e\cos\theta\right),$$

from which it follows

$$v = \sqrt{\frac{\mu}{a(1-e^2)}} \sqrt{e^2 + 1 + 2e\cos\theta} = \frac{r\dot{\theta}}{1 + e\cos\theta} \sqrt{e^2 + 1 + 2e\cos\theta}. \tag{3.46}$$

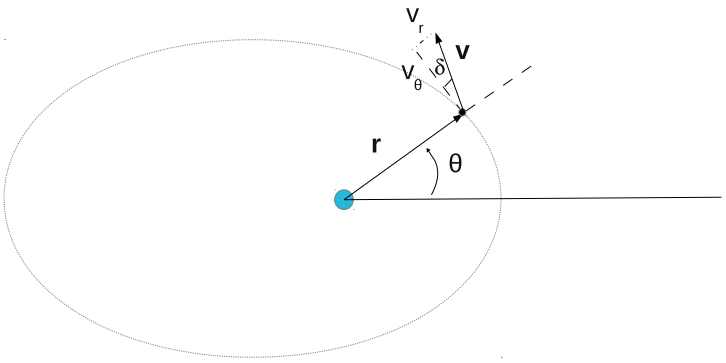


Figure 3.8: The flight path angle δ.

Moreover, the flight path angle δ introduced in Sec. 2.3.1 between the tangential and the transversal direction (see Fig. 3.8), is such that

$$\sin\delta = \frac{\dot{r}}{v} = \frac{e\sin\theta}{\sqrt{e^2+1+2e\cos\theta}}, \qquad \cos\delta = \frac{r\dot{\theta}}{v} = \frac{1+e\cos\theta}{\sqrt{e^2+1+2e\cos\theta}}. \tag{3.47}$$

In this way the components of the perturbation can be written as

$$f^r = -\frac{D}{m}\sin\delta, \qquad f^\theta = -\frac{D}{m}\cos\delta, \tag{3.48}$$

and by applying the Gauss planetary equations (2.62) we get the variation in time of the semi-major axis as

$$\begin{aligned}\frac{da}{dt} &= \frac{2}{n\sqrt{1-e^2}}\left(e\sin\theta f^r + \frac{p}{r}f^\theta\right)\\ &= -\frac{2D}{mn\sqrt{1-e^2}}\left(\frac{e^2\sin^2\theta+(1+e\cos\theta)^2}{\sqrt{e^2+1+2e\cos\theta}}\right)\\ &= -\frac{2D}{mn\sqrt{1-e^2}}\sqrt{e^2+1+2e\cos\theta}.\end{aligned} \tag{3.49}$$

The eccentricity variation is instead, using also Eq. (1.63),

$$\begin{aligned}\frac{de}{dt} &= \frac{\sqrt{1-e^2}}{na}\left[\sin\theta f^r + (\cos\theta+\cos E)\, f^\theta\right]\\ &= -\frac{D}{m\sqrt{1+e^2+2e\cos\theta}}\frac{\sqrt{1-e^2}}{na}\left[e\sin^2\theta + (1+e\cos\theta)\left(\cos\theta+\frac{e+\cos\theta}{1+e\cos\theta}\right)\right]\\ &= -\frac{D}{m\sqrt{1+e^2+2e\cos\theta}}\frac{\sqrt{1-e^2}}{na}\left(e\sin^2\theta+\cos\theta+e\cos^2\theta+e+\cos\theta\right)\\ &= -\frac{D}{m\sqrt{1+e^2+2e\cos\theta}}\frac{\sqrt{1-e^2}}{na}\left(2e+2\cos\theta\right),\end{aligned} \tag{3.50}$$

while the one in the argument of perigee is

$$\begin{aligned}\frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{nae}\left[-\cos\theta f^r + \left(1+\frac{1}{1+e\cos\theta}\right)\sin\theta f^\theta\right] \\ &= -\frac{D}{m\sqrt{1+e^2+2e\cos\theta}}\frac{\sqrt{1-e^2}}{nae}\left[-e\cos\theta\sin\theta + \left(1+\frac{1}{1+e\cos\theta}\right)\sin\theta\,(1+e\cos\theta)\right] \\ &= -\frac{2D\sin\theta}{m\sqrt{1+e^2+2e\cos\theta}}\frac{\sqrt{1-e^2}}{nae}. \end{aligned} \tag{3.51}$$

The phenomenon taking place due to the atmospheric drag is a circularization of the orbit: when almost zero eccentricity is reached, the trajectory resembles a spiral, which oscillates a little in eccentricity.

In order to compute the corresponding variations over one orbital period and the secular effects, we have to integrate

$$\begin{aligned}\Delta a &= \int_0^{2\pi}\frac{da}{dE}dE = \int_0^{2\pi}\frac{da}{dt}\frac{dt}{d\theta}\frac{d\theta}{dE}dE, \\ \Delta e &= \int_0^{2\pi}\frac{de}{dE}dE = \int_0^{2\pi}\frac{de}{dt}\frac{dt}{d\theta}\frac{d\theta}{dE}dE, \\ \Delta\omega &= \int_0^{2\pi}\frac{d\omega}{dE}dE = \int_0^{2\pi}\frac{d\omega}{dt}\frac{dt}{d\theta}\frac{d\theta}{dE}dE. \end{aligned} \tag{3.52}$$

Because of (1.16) and (1.23), we have

$$\begin{aligned}\frac{dt}{d\theta} &= \left(\frac{d\theta}{dt}\right)^{-1} = \frac{r^2}{h} = \frac{r^2}{na^{\frac{3}{2}}r^{1/2}(1+e\cos\theta)^{1/2}} \\ &= \frac{a^2(1-e^2)^2}{na^{\frac{3}{2}}a^{\frac{1}{2}}(1-e^2)^{\frac{1}{2}}(1+e\cos\theta)^2} \\ &= \frac{(1-e^2)^{\frac{3}{2}}}{n(1+e\cos\theta)^2}. \end{aligned} \tag{3.53}$$

Therefore,

$$\begin{aligned}\frac{da}{d\theta} &= -\frac{2D}{mn\sqrt{1-e^2}}\sqrt{e^2+1+2e\cos\theta}\frac{(1-e^2)^{\frac{3}{2}}}{n(1+e\cos\theta)^2} \\ &= -\rho\frac{A}{m}v^2C_D\frac{(1-e^2)}{n^2(1+e\cos\theta)^2}\sqrt{e^2+1+2e\cos\theta} \\ &= -\rho\frac{A}{m}C_D\frac{\mu}{n^2a(1-e^2)}(e^2+1+2e\cos\theta)\frac{(1-e^2)}{(1+e\cos\theta)^2}\sqrt{e^2+1+2e\cos\theta} \\ &= -\rho\frac{A}{m}C_Da^2\frac{(e^2+1+2e\cos\theta)^{\frac{3}{2}}}{(1+e\cos\theta)^2}. \end{aligned} \tag{3.54}$$

In an analogous way, it can be proven that

$$\frac{de}{d\theta} = -\frac{A}{m}\rho C_Da(1-e^2)\frac{(e^2+1+2e\cos\theta)^{\frac{1}{2}}}{(1+e\cos\theta)^2}(e+\cos\theta), \tag{3.55}$$

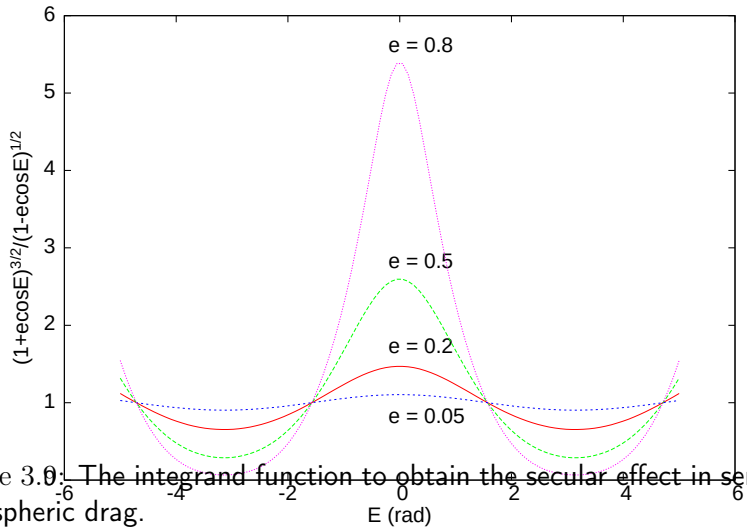


Figure 3.6: The integrand function to obtain the secular effect in semi-major axis due to the atmospheric drag.

and

$$\frac{d\omega}{d\theta} = -\frac{A}{m}\rho C_D a \frac{1-e^2}{e}\frac{(e^2+1+2e\cos\theta)^{\frac{1}{2}}}{(1+e\cos\theta)^2}\sin\theta. \tag{3.56}$$

Eq. (3.56) contains a $\sin\theta$ term and thus the integral over one orbit will be zero. This means that the atmospheric drag does not give rise to secular effects in ω, but only periodic ones.

From Eq. (1.63) it can be proven that

$$\frac{d\theta}{dE} = \frac{\sqrt{1-e^2}}{1-e\cos E}, \tag{3.57}$$

and also

$$\cos\theta = \frac{\cos E - e}{1-e\cos E}. \tag{3.58}$$

in such a way that we obtain

$$\begin{aligned}
\frac{da}{dt}\frac{dt}{d\theta}\frac{d\theta}{dE} &= -\rho\frac{A}{m}C_D a^2\frac{(e^2+1+2e\cos\theta)^{\frac{3}{2}}}{(1+e\cos\theta)^2}\frac{\sqrt{1-e^2}}{1-e\cos E}\\
&= -\rho\frac{A}{m}C_D a^2\frac{(e^2+1+2e\frac{\cos E-e}{1-e\cos E})^{\frac{3}{2}}}{(1+e\frac{\cos E-e}{1-e\cos E})^2}\frac{\sqrt{1-e^2}}{1-e\cos E}\\
&= -\rho\frac{A}{m}C_D a^2\frac{(e^2(1-e\cos E)+1-e\cos E+2e\cos E-2e^2)^{\frac{3}{2}}}{(1-e\cos E+e\cos E-e^2)^2(1-e\cos E)^{\frac{1}{2}}}\sqrt{1-e^2}\\
&= -\rho\frac{A}{m}C_D a^2\frac{(-e^2-e^3\cos E+1+e\cos E)^{\frac{3}{2}}}{(1-e^2)^2}\frac{\sqrt{1-e^2}}{(1-e\cos E)^{\frac{1}{2}}}\\
&= -\rho\frac{A}{m}C_D a^2\frac{(1-e^2)^{\frac{3}{2}}(1+e\cos E)^{\frac{3}{2}}}{(1-e^2)^2}\frac{\sqrt{1-e^2}}{(1-e\cos E)^{\frac{1}{2}}}\\
&= -\rho\frac{A}{m}C_D a^2\frac{(1+e\cos E)^{\frac{3}{2}}}{(1-e\cos E)^{\frac{1}{2}}}. \qquad (3.59)
\end{aligned}$$

Analogously, for the eccentricity we have

$$\frac{de}{dE} = -\rho\frac{A}{m}C_D a(1-e^2)\cos E\left(\frac{1+e\cos E}{1-e\cos E}\right)^{\frac{1}{2}}. \qquad (3.60)$$

The corresponding integrals (3.52) can be solved numerically once given a model for ρ. We notice that we always have $\Delta a < 0$, because (see Fig. 3.9)

$$\frac{(1+e\cos E)^{\frac{3}{2}}}{(1-e\cos E)^{\frac{1}{2}}} > 0.$$

Moreover, $\Delta e < 0$ because the function $\cos E\left(\frac{1+e\cos E}{1-e\cos E}\right)^{\frac{1}{2}}$ behaves in the way depicted in Fig. 3.9 and it can be proved that the area underneath the $x-$axis is smaller than the one above it.

The main effect due to the atmospheric drag is to shrink the orbit and as long as the satellite approaches the Earth, its orbital velocity increases. This is sometimes referred as the *drag paradox*. Let us consider a circular orbit and the expressions (3.29) and (3.30) derived from the energy equation, namely,

$$\Delta n = -\frac{3}{2}\frac{n}{a}\Delta a = -\frac{3}{na^2}\Delta\mathcal{E},$$

that is,

$$\Delta T = 3\frac{a}{\mu}T\Delta\mathcal{E}, \qquad (3.61)$$

where T is the orbital period. Moreover, for a circular orbit the speed is

$$v = \sqrt{\frac{\mu}{a}},$$

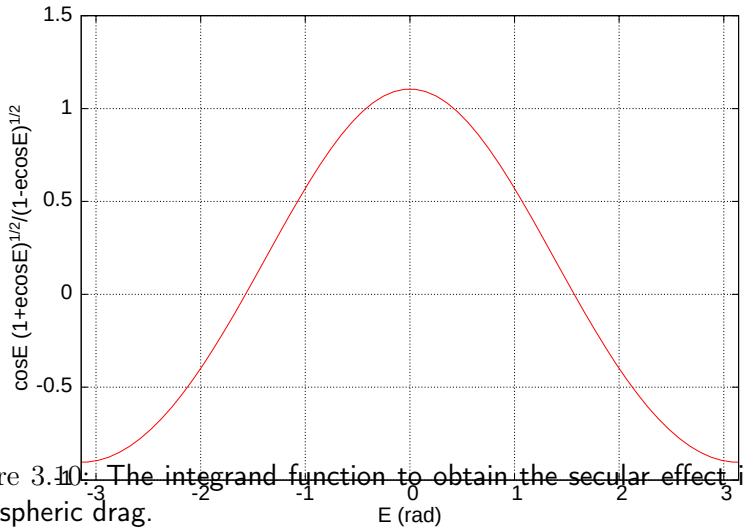


Figure 3.10: The integrand function to obtain the secular effect in eccentricity due to the atmospheric drag.

and thus

$$\sqrt{a}\Delta v + \frac{1}{2}\frac{v}{\sqrt{a}}\Delta a = 0 \qquad \Longrightarrow \qquad \Delta v = -\sqrt{\frac{a}{\mu}}\Delta\mathcal{E}. \tag{3.62}$$

In other words, as a general rule, if the force is such that the semi-major axis a decreases then

- the specific energy decreases;
- the orbital period decreases, i.e., the mean motion increases;
- the circular velocity increases.

3.2.3 Atmosphere Density Models

An accurate analysis of the perturbations induced by the drag on the orbit depends inevitably on the atmospheric density model adopted. As a matter of fact, ρ is determined empirically on the basis of the altitude and the time. In this process the solar flux variability is crucial and at least the following factors must be considered:

- solar activity, which is periodic in 11 years;
- day-night cycle, ρ being maximum at noon and minimum at midnight;
- geomagnetic storms;
- ultraviolet wavelength radiation, with a period of 27 days.

Since the '50s different models have been developed, we recall in particular the Jacchia -Roberts and the NRLMSIS-00 ones. In a first approximation, we can consider the static isotherm exponential one, namely,

$$\rho = \rho_0 \exp\left(-\frac{h - h_0}{H}\right), \tag{3.63}$$

where

- h is the altitude of the s/c;
- h_0 a reference altitude;
- ρ_0 the density corresponding to h_0;
- H a given scale factor, which represents the altitude at which the density is ρ_0/e.

This model is quite reliable if h does not differ too much from h_0, in order to ensure the isotherm assumption. In Mengali & Quarta (2006) and Vallado (2007), for instance, you can find the values corresponding to h_0, ρ_0, H.

3.2.4 Orbital Decay and Satellite's Lifetime

For what seen so far, LEO objects (see Sec 1.14.1) experience an orbital decay due to the atmospheric drag. If corrections maneuvers are not applied, then in a sufficient amount of time they can can reentry to the Earth. This mechanism is exploited as natural removal of space debris in LEO.

We can define the satellite lifetime as the interval of time going from the launch to the atmospheric reentry. It is simply

$$\mathcal{L} = NT = 2\pi N \left[\frac{1}{\mu}\left(\frac{h_p + r_\oplus}{1 - e}\right)^3\right]^{\frac{1}{2}}, \tag{3.64}$$

where h_p is the perigee altitude and N is the estimated number of revolutions. In particular, N depends on ρ and the ballistic coefficient $m/(AC_D)$, which measures how much the satellite is sensitive to the drag effect. Low values of $m/(AC_D)$ result in high orbital elements variations.

For a spacecraft in circular orbit at an altitude of 250 km amd $A/m \approx 0.01$ m^2/kg, we have $\mathcal{L} \approx 20$ days and the station-keeping $\Delta v-$budget required is about 570 m/s per year. If the altitude is 400 km the lifetime increases of about 200 days and $\Delta v \approx 24$ m/s per year, at 800 km the drag takes hundreds of years to remove the satellite from the orbit.

3.3 Solar Radiation Pressure

The solar radiation pressure is a non-conservative perturbation, due to the electromagnetic radiation coming from the Sun. In this case, the difficulties consist in modeling

the solar cycle and its variation and in knowing accurately the area of the spacecraft exposed to the solar radiation, along with its shape, orientation and optical properties. All these satellite's parameters are not constant in time, let us think, for instance, to the orbital attitude or to the fact that the material of the surface degrades and thus the reflectivity coefficients change.

In what follows, we consider only the direct radiation from the Sun and not the one reflected by the Earth to the satellite, this is the albedo. Also, the Sun is assumed as a point mass source and the solar rays as perfectly parallel.

3.3.1 Modeling

The solar radiation, that is, a photon beam, propagates in the vacuum with velocity equal to the speed of light $c = 299792458$ m/s and at striking the surface of the satellite it exerts a pressure on it.

Generally speaking, the pressure is proportional to the energy flux Φ, which is the energy E passing through a given area A per unit of time, namely,

$$\Phi = \frac{\Delta E}{A\Delta t}. \tag{3.65}$$

Indeed, the pressure due to a force F on a surface A is

$$P = \frac{F}{A},$$

where the force can be written in terms of a linear momentum change in this way

$$F = \frac{\Delta p}{\Delta t}.$$

Now, a single photon of energy E carries an impulse

$$p = \frac{E}{c},$$

and a photon beam hitting a satellite's surface changes its impulse of an amount

$$\Delta p = \frac{\Delta E}{c} = \frac{\Phi A \Delta t}{c}. \tag{3.66}$$

It follows that the perturbing force due to the solar radiation pressure is in modulus equal to

$$F = \frac{\Phi A}{c}, \tag{3.67}$$

and the pressure is

$$P = \frac{\Phi}{c}. \tag{3.68}$$

At 1 AU = 1.49597870691×10^8 km, the mean value of P is

$$\bar{P} = 4.56 \times 10^{-6} \quad \text{Pascal.}$$

Since the orbit of the Earth is not circular the solar radiation pressure P varies during the year in this way

$$P = \bar{P}\left(\frac{a_\odot}{||\mathbf{r} - \mathbf{r}_\odot||}\right)^2 \approx \bar{P}\left(\frac{a_\odot}{r_\odot}\right)^2, \tag{3.69}$$

where $a_\odot = 1$ AU is the mean distance between the Sun and the Earth, $\mathbf{r}$ is the geocentric radius vector of the satellite and $\mathbf{r}_\odot$ is the geocentric position of the Sun.

Not all the photons are absorbed by the surface: some are absorbed, some are reflected specularly, some are reflected diffusely according to the corresponding *reflectivity* coefficients ρ_a, ρ_s, ρ_d, such that

$$\rho_a + \rho_s + \rho_d = 1. \tag{3.70}$$

Let us look to Fig. 3.11, where

- A is the area of the satellite exposed to the Sun;
- $\hat{\mathbf{n}}$ is the unit direction normal to A;
- $\hat{\mathbf{s}}$ is the unit direction between A and the Sun, directed toward the Sun;
- ϕ is the *solar incident angle* between $\hat{\mathbf{n}}$ and $\hat{\mathbf{s}}$.

The accelerations caused by three portions of the photon beam are, respectively,

$$\begin{aligned} \mathbf{f}_a &= -P\frac{A}{m}\rho_a \cos\phi \hat{\mathbf{s}}, \\ \mathbf{f}_s &= -2P\frac{A}{m}\rho_s \cos^2\phi \hat{\mathbf{n}}, \\ \mathbf{f}_d &= -P\frac{A}{m}\rho_d \cos\phi \left(\hat{\mathbf{s}} + \frac{2}{3}\hat{\mathbf{n}}\right), \end{aligned} \tag{3.71}$$

and thus

$$\mathbf{f} = -P\frac{A}{m}\cos\phi\left[(\rho_a + \rho_d)\,\hat{\mathbf{s}} + \left(2\rho_s \cos\phi + \frac{2}{3}\rho_d\right)\hat{\mathbf{n}}\right]. \tag{3.72}$$

In a first approximation, we can consider $\hat{\mathbf{s}} \parallel \hat{\mathbf{n}}$, this is the so-called *cannonball model*, in such a way that

$$\mathbf{f} = -\frac{A}{m}P\left(\rho_a + 2\rho_s + \frac{5}{3}\rho_d\right)\frac{\mathbf{r}_\odot}{||\mathbf{r}_\odot||}. \tag{3.73}$$

Some authors denote $\rho_a + 2\rho_s + \frac{5}{3}\rho_d$ as $1 + \beta$ or $C_R \in [0, 2]$. In particular β is the *index of reflection* and we have

- $\beta = 0$ for a perfectly absorbing surface;
- $\beta = 1$ for a perfectly reflecting surface;
- $\beta = -1$ for a perfectly transmitting surface, i.e., transparent.

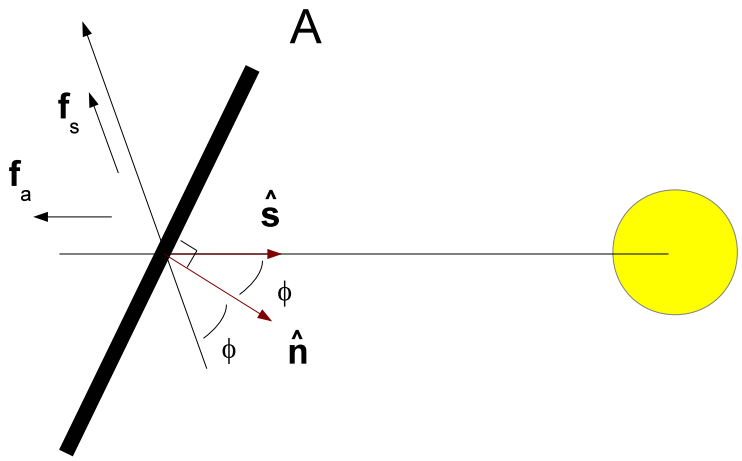


Figure 3.11: The geometry associated with the solar radiation pressure perturbation. A is the area of the satellite exposed to the Sun, $\hat{\mathbf{n}}$ is the unit direction normal to A, $\hat{\mathbf{s}}$ is the unit direction between A and the Sun directed toward the Sun and ϕ is the solar incident angle. $\mathbf{f}_a$ and $\mathbf{f}_s$ are the perturbing accelerations due the photons, which are absorbed and reflected specularly, respectively.

3.3.2 Effects

The perturbation due to the solar radiation pressure in modulus varies linearly with the area-to-mass ratio A/m and in a first approximation we can consider it independent from the altitude of the orbit, see Eq. (3.69) and Fig. 3.1. Compared to other effects,

- for low values of A/m, up 0.01 m^2/kg, it is some orders of magnitude less significant than the perturbation due to J_2;
- for $A/m \approx 1.63$ m^2/kg the two effects are comparable for a GEO s/c;
- for $A/m > 10 - 20$ m^2/kg the solar radiation pressure becomes the dominant perturbation for GEO satellites.

Since we deal with a non-conservative perturbation, we have to apply the Gauss planetary equations to (3.73). In the $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{h}}\}$ reference system, the unit components of the

perturbing acceleration can be written as

$$\begin{aligned}
f^r &= \cos^2\frac{i}{2}\cos^2\frac{\epsilon}{2}\cos\left(\lambda_\odot - u - \Omega\right) - \sin^2\frac{i}{2}\sin^2\frac{\epsilon}{2}\cos\left(\lambda_\odot - u - \Omega\right) - \\
&\quad -\frac{1}{2}\sin i\sin\epsilon\left[\cos\left(\lambda_\odot - u\right) - \cos\left(-\lambda_\odot - u\right)\right] - \\
&\quad -\sin^2\frac{i}{2}\cos^2\frac{\epsilon}{2}\cos\left(-\lambda_\odot - u - \Omega\right) - \cos^2\frac{i}{2}\sin^2\frac{\epsilon}{2}\cos\left(-\lambda_\odot - u - \Omega\right), \\
f^\theta &= \cos^2\frac{i}{2}\cos^2\frac{\epsilon}{2}\sin\left(\lambda_\odot - u - \Omega\right) - \sin^2\frac{i}{2}\sin^2\frac{\epsilon}{2}\sin\left(\lambda_\odot - u - \Omega\right) - \\
&\quad -\frac{1}{2}\sin i\sin\epsilon\left[\sin\left(\lambda_\odot - u\right) - \sin\left(-\lambda_\odot - u\right)\right] - \\
&\quad -\sin^2\frac{i}{2}\cos^2\frac{\epsilon}{2}\sin\left(-\lambda_\odot - u - \Omega\right) - \cos^2\frac{i}{2}\sin^2\frac{\epsilon}{2}\sin\left(-\lambda_\odot - u - \Omega\right), \\
f^h &= \sin i\cos^2\frac{\epsilon}{2}\sin\left(\lambda_\odot - \Omega\right) - \sin i\sin^2\frac{\epsilon}{2}\sin\left(\lambda_\odot + \Omega\right) - \cos i\sin\epsilon\sin\lambda_\odot,
\end{aligned}$$

where ϵ is the obliquity of the ecliptic (see Sec. **??**), $\lambda_\odot$ the ecliptic longitude of the Sun and $u = \omega + \theta$ is the *argument of latitude* ($\tilde{\theta}$ in Sec. 2.3.1). By considering these expressions, it turns out that the solar radiation pressure causes periodic variations in all orbital elements, especially long-period ones in e, ω, which can be written as

$$\begin{aligned}
\frac{de}{dt} &= -\frac{3}{2}C_R P\frac{A}{m}\frac{\sqrt{1-e^2}}{na}\left[C_1\sin A_1 + C_2\sin A_2 + C_3\left(\sin A_3 + \sin A_4\right) + C_4\sin A_5 + C_5\sin A_6\right], \\
\frac{d\omega}{dt} &= \frac{3}{2}C_R P\frac{A}{m}\frac{\sqrt{1-e^2}}{nae}\left[C_1\cos A_1 + C_2\cos A_2 + C_3\left(\cos A_3 + \cos A_4\right) + C_4\cos A_5 + C_5\cos A_6\right],
\end{aligned} \tag{3.74}$$

where

$$\begin{aligned}
C_1 &= \cos^2\frac{i}{2}\cos^2\frac{\epsilon}{2}, \\
C_2 &= \sin^2\frac{i}{2}\sin^2\frac{\epsilon}{2}, \\
C_3 &= \frac{1}{2}\sin i\sin\epsilon, \\
C_4 &= -\sin^2\frac{i}{2}\cos^2\frac{\epsilon}{2}, \\
C_5 &= -\cos^2\frac{i}{2}\sin^2\frac{\epsilon}{2},
\end{aligned}$$

and

$$\begin{aligned} A_1 &= \lambda_\odot - \omega - \Omega, \\ A_2 &= \lambda_\odot - \omega + \Omega, \\ A_3 &= \lambda_\odot - \omega, \\ A_4 &= \lambda_\odot + \omega, \\ A_5 &= \lambda_\odot + \omega - \Omega, \\ A_6 &= \lambda_\odot + \omega + \Omega. \end{aligned}$$

We notice that for GEO satellites, C_2, C_3, C_4 vanish.

3.4 Third-Body Perturbation

Let us consider the perturbation due to the presence of a third body, either Sun or Moon for satellites around the Earth. It becomes significant as long as the altitude increases and thus it strongly affects the apogee of elliptic orbits and cannot be neglected to predict the dynamics in MEO and GEO. The main issue in this case regards the position of Sun and Moon, especially the latter, with respect to the equatorial plane.

3.4.1 Modeling

Let us consider a system composed by three bodies, m_1, m_2 and m_3. In an inertial reference system $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ the equation of motion for each mass m_l with $l = 1, 2, 3$ can be written as

$$m_l \frac{d^2\mathbf{r}_l}{dt^2} = G \sum_{q=1, l\neq q}^{3} \frac{m_l m_q}{r_{lq}^3} \left(\mathbf{r}_q - \mathbf{r}_l\right), \tag{3.75}$$

where $\mathbf{r}_l = x_l\hat{\mathbf{i}} + y_l\hat{\mathbf{j}} + z_l\hat{\mathbf{k}}$ is the radius vector corresponding to m_l and $\mathbf{r}_{lq} = \mathbf{r}_q - \mathbf{r}_l$, see Fig. 3.12. We notice that (3.75) can be generalized to n masses, not just 3. In that case, we speak of *the n-Body Problem*.

Let us assume that m_1 is the Earth, m_2 the perturbing body and m_3 the satellite. Eq. (3.75) gives

$$\frac{d^2\mathbf{r}_1}{dt^2} = G\frac{m_3}{r_{13}^3}\left(\mathbf{r}_3 - \mathbf{r}_1\right) + G\frac{m_2}{r_{12}^2}\left(\mathbf{r}_2 - \mathbf{r}_1\right), \tag{3.76}$$

and

$$\frac{d^2\mathbf{r}_3}{dt^2} = G\frac{m_1}{r_{31}^3}\left(\mathbf{r}_1 - \mathbf{r}_3\right) + G\frac{m_2}{r_{32}^3}\left(\mathbf{r}_2 - \mathbf{r}_3\right). \tag{3.77}$$

Now, let us translate the origin of the reference system at the Earth and define (see Fig. 3.12)

$$\mathbf{r} := \mathbf{r}_3 - \mathbf{r}_1 \equiv \mathbf{r}_{13}, \qquad \mathbf{r}_p := \mathbf{r}_2 - \mathbf{r}_1 \equiv \mathbf{r}_{12}, \qquad \boldsymbol{\rho} := \mathbf{r}_2 - \mathbf{r}_3 \equiv \mathbf{r}_{32}.$$

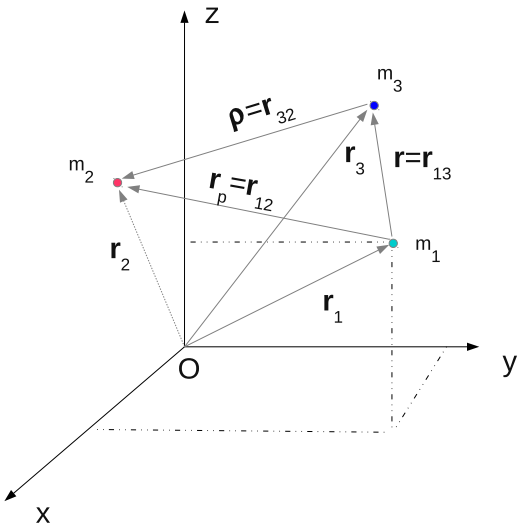


Figure 3.12: Relative position of three bodies in an inertial reference system.

Then, the equation of motion for the satellite relative to the Earth is

$$\begin{aligned}\frac{d^2\mathbf{r}}{dt^2} &= \frac{d^2\mathbf{r}_3}{dt^2} - \frac{d^2\mathbf{r}_1}{dt^2} = G\frac{m_1}{r_{31}^3}(\mathbf{r}_1 - \mathbf{r}_3) + G\frac{m_2}{r_{32}^3}(\mathbf{r}_2 - \mathbf{r}_3) - G\frac{m_3}{r_{13}^3}(\mathbf{r}_3 - \mathbf{r}_1) - G\frac{m_2}{r_{12}^3}(\mathbf{r}_2 - \mathbf{r}_1) \\ &= -G(m_1 + m_3)\frac{\mathbf{r}}{r^3} + Gm_2\left(\frac{\boldsymbol{\rho}}{\rho^3} - \frac{\mathbf{r}_p}{r_p^3}\right). \end{aligned} \tag{3.78}$$

We notice that the first term on the right-hand side is the classical Keplerian one, the other represents the perturbation on the satellite due to the mass m_2. This perturbation consists in particular in one direct term $\boldsymbol{\rho}/\rho^3$, i.e., m_2 attracts gravitationally directly m_3, and in one indirect term $\mathbf{r}_p/r_p^3$, i.e., m_2 also attracts the Earth and thus move it. We can write

$$Gm_2\left(\frac{\boldsymbol{\rho}}{\rho^3} - \frac{\mathbf{r}_p}{r_p^3}\right) = \frac{\partial}{\partial \mathbf{r}}\left[Gm_2\left(\frac{1}{\rho} - \frac{\mathbf{r}\cdot\mathbf{r}_p}{r_p^3}\right)\right], \tag{3.79}$$

that is, the perturbing function for the third-body effect is

$$\mathcal{R} = Gm_2\left(\frac{1}{\rho} - \frac{\mathbf{r}\cdot\mathbf{r}_p}{r_p^3}\right), \tag{3.80}$$

which is called *disturbing function* associated with the disturbing mass m_2. It follows that the third-body perturbation is conservative.

If $r << r_p$, then the disturbing function can be written as

$$\mathcal{R} = \frac{Gm_2}{r_p} \sum_{k=2}^{\infty} P_k(\cos S) \left(\frac{r}{r_p}\right)^k \tag{3.81}$$

where S is the angle between r and r_p and P_k is the Legendre polynomial of degree k.

3.4.2 Effects

The third-body perturbation gives rise to secular variations in Ω, ω, which can be expressed, by means of (3.81) and for a third body moving on a circular orbit, as

$$\frac{d\Omega}{dt} = -\frac{3Gm_2\left(2+e^2\right)\left(2-3\sin^2 i_p\right)}{16r_p^3 n\sqrt{1-e^2}}\cos i, \tag{3.82}$$

$$\frac{d\omega}{dt} = \frac{3Gm_2\left(2-3\sin^2 i_p\right)}{16r_p^3 n\sqrt{1-e^2}}\left(e^2+4-5\sin i\right), \tag{3.83}$$

where i_p is the orbital inclination of the third body with respect to the equatorial plane. The variation in the longitude of the ascending node is a precession about the pole of the ecliptic if the third body m_2 is the Sun, about an axis which is normal to the lunar orbital plane if the third body m_2 is the Moon. The effect reminds what we have seen is caused by J_2 and analogous considerations can be drawn depending on whether the orbit is prograde or not. Actually, the effects due to both Sun and Moon are to be considered at high altitudes: the Sun is very massive, while the Moon is very close to the Earth. The resulting precession due to Sun, Moon and J_2 will be about a mean pole between the Earth's and the ecliptic ones.

Moreover, we have long-period perturbations in e, i, Ω, ω. The long-period variation in inclination is especially important for GEO satellites: it can be proved that, by considering at the same time the perturbation due to Sun, Moon and J_2 and eliminating monthly and yearly variations, we have

$$\frac{di}{dt} = \sum_{j=1}^{2} \frac{3}{8}\frac{n_j^2}{n} k_j \left[\cos i \sin\left(2i_j\right)\sin\left(\Omega-\Omega_j\right) + \sin i \sin^2 i_j \sin\left(2\Omega-2\Omega_j\right)\right], \tag{3.84}$$

$$\begin{aligned}\frac{d\Omega}{dt} = &\sum_{j=1}^{2} \frac{3}{16}\frac{n_j^2}{n\sin i} k_j \left[\sin\left(2i\right)\left(1-3\cos^2 i_j\right) + 2\cos\left(2i\right)\sin\left(2i_j\right)\cos\left(\Omega-\Omega_j\right)\right] + \\ &+\sum_{j=1}^{2} \frac{3}{16}\frac{n_j^2}{n\sin i} k_j \left[\sin\left(2i\right)\sin^2 i_j \cos\left(2\Omega-2\Omega_j\right)\right] - \frac{3}{2}J_2\left(\frac{r_\oplus}{a}\right)^2 n\cos i,\end{aligned} \tag{3.85}$$

where the subscript j indicates Sun and Moon, $k_1 = 1$ for the Sun, $k_2 = 1/82.3$ for the Moon. We notice that (3.85) reflects a coupling between the variation in i and the one in Ω.

Orbital Transfers 4

4.1 Introduction

The motion of a satellite can be controlled by means of a propulsion system, which helps either to change the orbital parameters to achieve some specific goals or to correct any deviations to the nominal orbit due to the perturbations acting on the body. There exist different options that can be applied, in particular we distinguish between impulsive maneuvers and low thrust strategies. The formulae we will give in Sec. 4.2 are valid for any artificial object embedded with a propulsion system and explain how the distinction can be made. In the whole chapter, we assume that the only natural force acting on the spacecraft is given by one central body, that is, the Keplerian assumptions are considered.

Concerning the impulsive case, we recall that two orbits sharing one focus can intersect at most in two points and that to move from one to the other by means of a single burn is possible only when there exists at least one intersection. Otherwise, if the two orbits do not intersect, then at least two impulsive maneuvers are required.

Here we are not going to describe all the possible transfers between two nominal orbits and, in the second part of the chapter, we will introduce the basis of low thrust transfers.

4.2 Tsiolkowsky's Equation

Let us consider a rocket of mass m_0 at some initial time t_0 moving with velocity $\mathbf{v}_0$ with respect to an inertial reference system in the vacuum in gravity-free space. The ejection of some propellant by the propulsion system makes the rocket to move. Let $\mathbf{v}_e$ be the velocity of exhaustion of the propellant mass with respect to the rocket. Its modulus $v_e = ||\mathbf{v}_e||$ is called *effective exhaust velocity* or *equivalent exhaust velocity*. The mass of the rocket changes in time according to

$$m(t) = m_e + m_p(t), \tag{4.1}$$

where m_e is the empty mass which remains constant and $m_p(t)$ is the mass corresponding to the propellant.

At time t, the rocket has mass $m(t)$ and it is moving with a velocity $\mathbf{v}(t)$. At $t + dt$, its mass is $m(t) - dm$ and its velocity $\mathbf{v}(t) + d\mathbf{v}$, while the expelled propellant has mass dm and velocity $\mathbf{v}(t) - \mathbf{v}_e$. The law of conservation of the linear momentum reads

$$m(t)\mathbf{v}(t) = (m(t) - dm)(\mathbf{v}(t) + d\mathbf{v}) + dm(\mathbf{v}(t) - \mathbf{v}_e),$$

that is,

$$0 = m(t)d\mathbf{v} - dmd\mathbf{v} - \mathbf{v}_e dm.$$

By neglecting the second order term $dmd\mathbf{v}$, we obtain

$$m(t)d\mathbf{v} = \mathbf{v}_e dm, \tag{4.2}$$

which can be integrated to give

$$v_f - v_0 = v_e \ln\Big(\frac{m_0}{m_f}\Big), \tag{4.3}$$

where we have considered that $\mathbf{v}_e$ is constant and its direction is opposite to that of $\mathbf{v}_f$ and $\mathbf{v}_0$.

Moreover, the magnitude of the thrust provided by the propulsion system is[1]

$$T = -v_e \frac{dm}{dt}. \tag{4.4}$$

In the propulsion design, the *specific impulse* is defined as the ratio between the magnitude of the thrust and the product between the magnitude of the acceleration of gravity at sea level g_0 and the propellant mass flow rate, that is,

$$I_{sp} := \frac{v_e}{g_0}. \tag{4.5}$$

This is a characteristic of the type of propellant used and has the dimensions of time. In Tab. 4.1 we give the values corresponding to some propulsion systems. From (4.3) and (4.5), we get

$$v_f - v_0 = g_0 I_{sp} \ln\Big(\frac{m_0}{m_f}\Big), \tag{4.6}$$

which is said *rocket equation* or *Tsiolkowsky's equation*. This indicates how much mass must be consumed to get to a nominal speed v_f starting from v_0, namely,

$$\frac{m_0 - m_f}{m_0} = 1 - \exp\Big(-\frac{v_f - v_0}{g_0 I_{sp}}\Big). \tag{4.7}$$

4.2.1 Gravity Losses

Now, let us assume that the rocket is ascending against a gravity force with a flight path angle γ.

[1] This is true if we consider negligible the difference between the pressure at the exit from the nozzle p_e and the free stream pressure p_0. Otherwise,

$$T = -v_e \frac{dm}{dt} + (p_e - p_0)A,$$

where A is the exit area of the nozzle.

Type	$I_{sp}(s)$	$T(N)$
N_2	60	0.1–50
H_2	250	0.1–50
Mono-propellant Liquid	140–235	$0.1–12\times10^6$
Bi-propellant Liquid	320–460	$0.1–12\times10^6$
Solid	260–300	$0.1–12\times10^6$
Hybrid	290–350	$0.1–12\times10^6$
Solid Core Nuclear	800–1100	up to 12×10^6
Liquid Core Nuclear	3000	up to 12×10^6
Gas Core Nuclear	6000	up to 12×10^6
Thermoelectric	500–1000	10^{-4}–20
Electromagnetic	1000–7000	10^{-4}–20
Electrostatic	2000–10000	10^{-4}–20

Table 4.1: Specific impulse and thrust level corresponding to some propulsion systems (Mengali & Quarta, 2006).

Eq. (4.2) turns out to be

$$m_0 dv = -v_e dm - mg\sin\gamma dt, \tag{4.8}$$

where we take into account the sense of the direction of the acceleration of gravity $\mathbf{g}$. By integrating (4.8), we obtain

$$v - v_0 = v_e \ln\Big(\frac{m_0}{m}\Big) - \int_{t_0}^{t} g(h)\sin\gamma dt, \tag{4.9}$$

where we consider g to vary with the altitude h. Let us analyze the just written equation. First of all, to increase v we need either to increase v_e or m_0/m. The term $\int_0^t g(h)\sin\gamma dt$ is said *gravity loss* and results to be zero in the ideal case where the impulsive burn takes place instantaneously. Also, it should be clear that the lower the altitude we thrust the larger the effort. Gravity losses are also significant for high values of γ.

In real world situations, speaking in general terms for any artificial orbital body, the thrust is given for a finite amount of time. If this interval is negligible with respect to the whole duration of the mission, the change in velocity can be considered impulsive and in particular, for this approximation to be true it is necessary to estimate the ratio T/W_s where W_s is the local weight of the object, namely,

$$W_s := m\frac{\mu}{r^2}. \tag{4.10}$$

We can distinguish between

- *high thrust*: $T/W_s \geq 0.5$, that is, the thrust is dominant and we can consider the impulsive approximation;

- *low thrust*: $T/W_s \leq 10^{-5}$, that is, the thrust acts continuously and the body moves on a spiral.

One possible way to study the second case is using the perturbations theory approach. In the intermediate cases, the force due to the thrust and the one due to the gravity are both important and it is advisable to exploit numerical tools starting from semi-analytical models as first approximation.

Finally, for sake of completeness, for a launcher in Eq. (4.9) it is required to add one term, which accounts for the perturbation due to the atmospheric drag. Following Sec. 3.2, we have

$$v - v_0 = v_e \ln\Big(\frac{m_0}{m}\Big) - \int_{t_0}^{t} g(h)\sin\gamma dt - \int_{t_0}^{t} \frac{D}{m} dt, \tag{4.11}$$

being $D = 1/2\rho A v^2 C_D$.

4.3 Impulsive Coplanar Transfers

Let us consider two nominal coplanar circular orbits of radius r_i and r_f, respectively, with $r_i < r_f$ at a given central body of mass parameter μ. We are going to describe the two classical options that can be applied to move the spacecraft from one to the other. In both cases, the maneuvers are assumed as completely impulsive, the flight path angle is zero and therefore there is not any gravity loss.

4.3.1 Hohmann Transfer

The Hohmann transfer is the minimum cost two-impulses transfer between coplanar circular orbits. A first maneuver Δv_1 is applied along the direction of motion of the spacecraft to transfer it onto an elliptic orbit of semi-major axis

$$a = \frac{r_i + r_f}{2}. \tag{4.12}$$

This intermediate ellipse is such that the pericenter radius is equal to r_i and the apocenter one to r_f. After half of one period of the ellipse, namely,

$$\text{time of flight} = \pi\sqrt{\frac{a^3}{\mu}}, \tag{4.13}$$

a second maneuver Δv_2 is applied again along the tangent direction of motion to inject the spacecraft into the final orbit. See Fig. 4.1.

From the energy equation (1.37), the magnitude of the velocity on the ellipse at the pericenter is

$$v_{peri} = \sqrt{\mu\Big(\frac{2}{r_i} - \frac{1}{a}\Big)} = \sqrt{2\mu\Big(\frac{1}{r_i} - \frac{1}{r_i + r_f}\Big)},$$

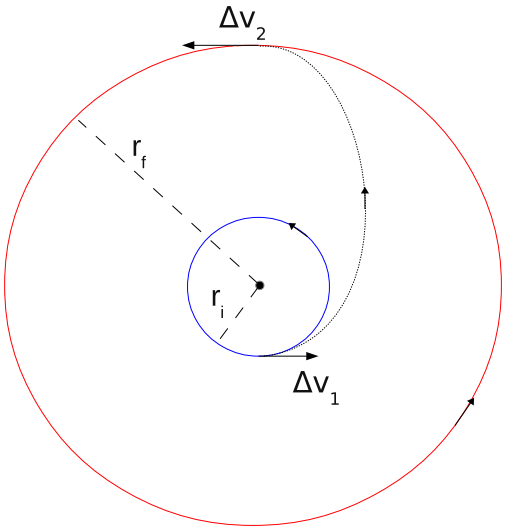


Figure 4.1: The Hohmann transfer.

the one at the apocenter is

$$v_{apo} = \sqrt{\mu\Big(\frac{2}{r_f} - \frac{1}{a}\Big)} = \sqrt{2\mu\Big(\frac{1}{r_f} - \frac{1}{r_i + r_f}\Big)},$$

and the ones on the initial and final circular orbits

$$v_{circ_i} = \sqrt{\frac{\mu}{r_i}}, \qquad\qquad v_{circ_f} = \sqrt{\frac{\mu}{r_f}}.$$

The two maneuvers are thus

$$\Delta v_1 = v_{peri} - v_{circ_i}, \tag{4.14}$$

and

$$\Delta v_2 = v_{circ_f} - v_{apo}. \tag{4.15}$$

The total cost of the transfer is

$$\Delta v_{Hoh} = \Delta v_1 + \Delta v_2.$$

In the case where the aim is to move from the outer to the inner orbit the two maneuvers are applied in the direction opposite to the motion of the spacecraft.

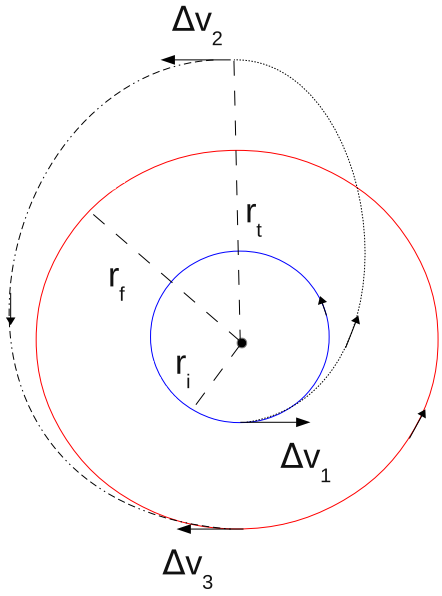


Figure 4.2: The bi-elliptic transfer.

Notice that

$$\begin{aligned}\frac{\Delta v_{Hoh}}{v_{circ_i}} &= \frac{\sqrt{2\mu\left(\frac{1}{r_i}-\frac{1}{r_i+r_f}\right)}}{\sqrt{\frac{\mu}{r_i}}} - 1 + \sqrt{\frac{\mu r_i}{\mu r_f}} - \frac{\sqrt{2\mu\left(\frac{1}{r_f}-\frac{1}{r_i+r_f}\right)}}{\sqrt{\frac{\mu}{r_i}}} = \\ &= \sqrt{2-\frac{2r_i}{r_i+r_f}} - 1 + \sqrt{\frac{r_i}{r_f}} - \sqrt{\frac{2r_i}{r_f}-\frac{2r_i}{r_i+r_f}} = \\ &= \sqrt{\frac{2R}{1+R}}\left(1-\frac{1}{R}\right) - 1 + \sqrt{\frac{1}{R}}, \end{aligned} \tag{4.16}$$

where $R = r_f/r_i$. By differentiating the expression just written in terms of R, it is possible to prove that the function $\Delta v/v_{circ_i}$ has a maximum at $R_{max} \approx 15.58$.

4.3.2 Bi-elliptic Transfer

In the case of the bi-elliptic transfer, three maneuvers are applied and two intermediate ellipses are considered. The first impulse along the direction of motion aims at moving the spacecraft from the inner circular orbit to an elliptic orbit whose pericenter coincides with r_i, while the apocenter is equal to $r_t > r_f$, where r_t is a parameter chosen by the

orbit designer. After a time

$$t_1 = \pi\sqrt{\frac{a_t^3}{\mu}}, \qquad \text{with} \qquad a_t = \frac{r_i + r_t}{2}, \tag{4.17}$$

a second impulse is given tangentially to transfer to an ellipse of semi-major axis

$$a_f = \frac{r_t + r_f}{2}, \tag{4.18}$$

being the apocenter equal to r_t and the pericenter equal to r_f. After a time

$$t_2 = \pi\sqrt{\frac{a_f^3}{\mu}}, \tag{4.19}$$

the last maneuver is applied to insert into the final orbit. This time the maneuver is in the direction opposite to the motion of the spacecraft. See Fig. 4.2. The total time of flight is thus

$$\text{time of flight} = t_1 + t_2,$$

and the total cost

$$\Delta v_{be} = \Delta v_1 + \Delta v_2 + \Delta v_3,$$

where

$$\begin{aligned}
\Delta v_1 &= \sqrt{2\mu\Big(\frac{1}{r_i} - \frac{1}{r_i + r_t}\Big)} - v_{circ_i}, \\
\Delta v_2 &= \sqrt{2\mu\Big(\frac{1}{r_t} - \frac{1}{r_f + r_t}\Big)} - \sqrt{2\mu\Big(\frac{1}{r_t} - \frac{1}{r_i + r_t}\Big)}, \\
\Delta v_3 &= -v_{circ_f} + \sqrt{2\mu\Big(\frac{1}{r_f} - \frac{1}{r_f + r_t}\Big)}.
\end{aligned} \tag{4.20}$$

Let us introduce the parameter $R^* = r_t/r_i$ and refers each of the three maneuvers to the initial orbital velocity of the spacecraft. We get

$$\begin{aligned}
\frac{\Delta v_1}{v_{circ_i}} &= \sqrt{\frac{2R^*}{1 + R^*}} - 1, \\
\frac{\Delta v_2}{v_{circ_i}} &= \sqrt{2\Big(\frac{1}{R^*} - \frac{1}{R + R^*}\Big)} - \sqrt{2\Big(\frac{1}{R^*} - \frac{1}{1 + R^*}\Big)}, \\
\frac{\Delta v_3}{v_{circ_i}} &= -\sqrt{\frac{1}{R}} + \sqrt{2\Big(\frac{1}{R} - \frac{1}{R + R^*}\Big)},
\end{aligned} \tag{4.21}$$

that is,

$$\frac{\Delta v_{be}}{v_{circ_i}} = \sqrt{\frac{2R^*}{1 + R^*}} - 1 - \sqrt{2\Big(\frac{1}{R^*} - \frac{1}{1 + R^*}\Big)} + \sqrt{2\Big(\frac{1}{R^*} - \frac{1}{R + R^*}\Big)} + \sqrt{2\Big(\frac{1}{R} - \frac{1}{R + R^*}\Big)} - \sqrt{\frac{1}{R}}. \tag{4.22}$$

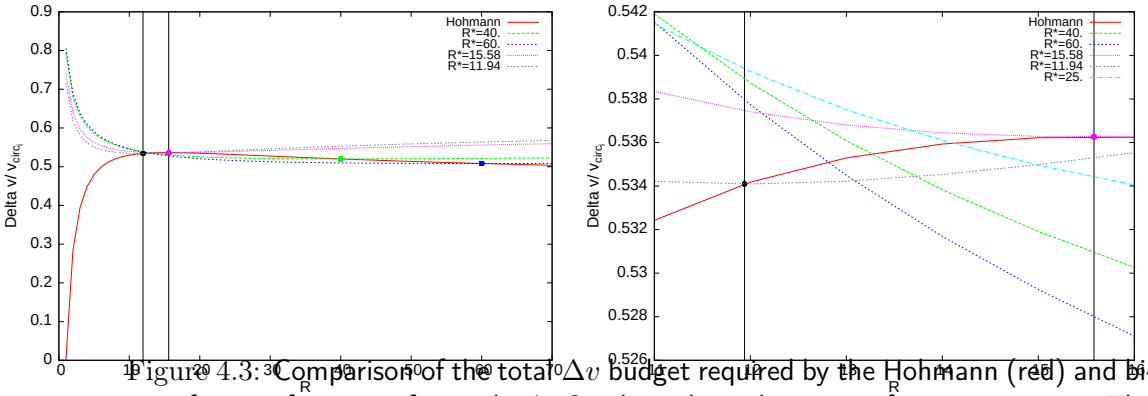


Figure 4.3: Comparison of the total Δv budget required by the Hohmann (red) and bi-elliptic transfer as a function of R and R^*. On the right a closer view for $11 < R < 16$. The circles in color report when $R^* = R$.

If $R^* \to R$ then clearly we fall back into the Hohmann case. Instead for $R^* \to \infty$, the transfer is said *bi-parabolic* and the total cost is

$$\Delta v_\infty = (\sqrt{2} - 1)\Big(1 + \sqrt{\frac{1}{R}}\Big) v_{circ_i}. \tag{4.23}$$

Notice that this is the ideal situation when $\Delta v_{be}/v_{circ_i}$ reaches its minimum as a function of R^* and can be thus considered for comparing the two strategies described. Indeed, the fact that the third maneuver is applied on the direction opposite to the orbital motion makes Δv_3 a wasted energy. As a consequence, the larger the semi-major axis of the two intermediate orbits the more efficient the bi-elliptic transfer.

4.3.3 Comparison between Hohmann and Bi-elliptic Transfers

Let us look to Fig. 4.3, where we have plotted the total Δv budget relative to the initial circular orbital velocity for the Hohmann and the bi-elliptic transfer as a function of R and R^*, see (4.16) and (4.22). To analyze this plot, let us first notice that the Hohmann transfer is as much expensive as the bi-parabolic one in terms of impulsive burns when $\Delta v_{Hoh} = \Delta v_\infty$, that occurs for $\tilde{R} \approx 11.94$. In other words, for any $R > \tilde{R}$ the bi-parabolic transfer is more convenient than the Hohmann one (to the detriment of the time of flight), while for $R < \tilde{R}$ the Hohmann case is preferable to the bi-elliptic (not just bi-parabolic) one.

Moreover, we can see that a curve corresponding to the bi-elliptic transfer for a given value of R^* intersects the curve corresponding to the Hohmann strategy twice. The second time, in particular, takes place when $R^* = R$. From that point on $R^* < R$ and thus it is more convenient to apply the Hohmann transfer. On the other hand, from Sec. 4.3.1 we know that at $R_{max} \approx 15.58$ the cost associated with the Hohmann case attains a maximum and thus for any $R^* > R > R_{max}$ the bi-elliptic transfer is less expensive.

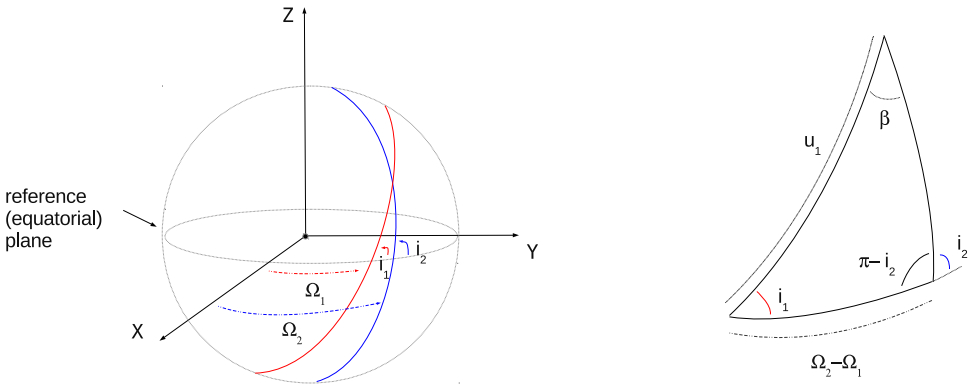


Figure 4.4: The geometry associated to a plane change maneuver.

Finally, the first intersection between the two kinds of curve takes place for $\tilde{R} \leq R < R_{max}$, that is, the choice between the two designs must be done according to R^*. It is possible to prove that it is better to adopt the bi-elliptic strategy if $R^* > R_0^*$, the Hohmann one otherwise. The value R_0^* is defined as

$$R_0^* = \frac{A + \sqrt{B}}{2}, \tag{4.24}$$

with

$$\begin{aligned} A &= \frac{1}{R}\left[\frac{(M-3N)^2+1}{(M+N)^2-1}\right] + \frac{1}{R^2}\left[\frac{1}{1-(M+N)^2}\right], \\ B &= A^2 + \frac{4}{R[(M+N)^2-1]}, \\ M &= \frac{P^2 - 1/R}{2P}, \\ N &= \frac{1}{2P}, \\ P &= \frac{R-1}{R}\sqrt{\frac{R}{1+R}} + \sqrt{\frac{2}{R}}. \end{aligned}$$

4.4 Impulsive Non-coplanar Transfers

Let us consider two circular orbits of given radius and different inclination, say i_1 and i_2 respectively. To move from one to the other, it is required to apply one maneuver at one of the nodes (ascending or descending).

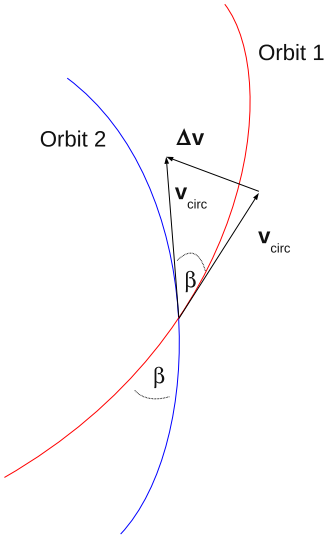


Figure 4.5: Maneuver required to change the inclination of a circular orbit.

Indeed, let us refer to Fig. 4.4, where we have depicted two generic orbits, i.e., not necessarily circular. At the intersection point the maneuver is performed to accomplish the transfer. The change in velocity causes a plane change equal to the angle β. Using the spherical law of cosines, we have

$$\cos(\pi - i_2) = -\cos i_2 = -\cos i_1 \cos\beta + \sin i_1 \sin\beta \cos u_1, \tag{4.25}$$

where $u_1 = \theta_1 + \omega_1$ is the argument of latitude ($\tilde{\theta}$ in Sec. 2.3.1), being θ_1 the true anomaly on the first orbit at the time of the maneuver. Also,

$$\begin{aligned} \cos\beta &= -\cos i_1 \cos(\pi - i_2) + \sin i_1 \sin(\pi - i_2)\cos(\Omega_2 - \Omega_1) = \\ &= \cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos(\Omega_2 - \Omega_1). \end{aligned} \tag{4.26}$$

In other words, if the intersection of the two orbits, that is, the time of the maneuver, occurs on the reference plane, then $\Omega_2 = \Omega_1$ and thus the final orbit will differ from the initial one only in inclination.

So, if our aim to maintain semi-major axis and argument of ascending node of a nominal circular orbit and vary its inclination, then it means that the modulus of the velocity on the final orbit will also not change and that the maneuver must be applied at one of the nodes. Its modulus must be equal to (see Fig. 4.5)

$$\Delta v = 2 v_{circ} \sin\frac{\beta}{2}. \tag{4.27}$$

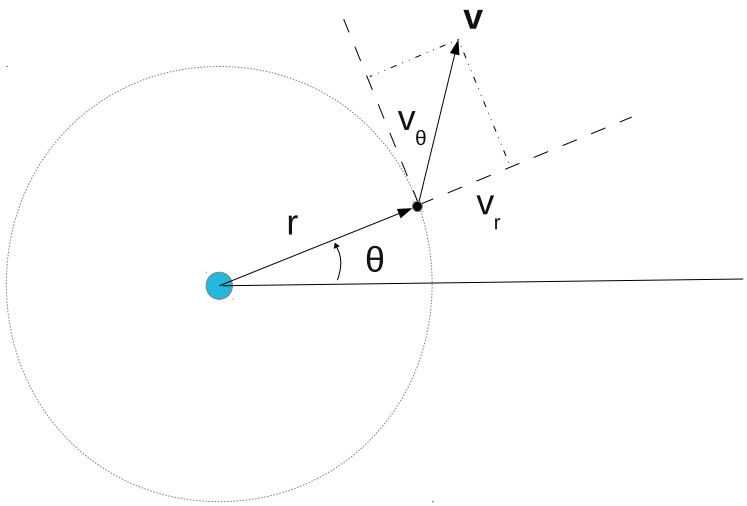


Figure 4.6: Polar coordinates.

4.5 Coplanar Continuous Thrust Transfers

Low thrust devices can be exploited to change size, shape and plane of a nominal orbit, as well as impulsive burns do. The classic example considers to use an electric engine to raise the altitude of a spacecraft at the Earth, but more complex missions have flown under low thrust configurations. We mention Deep Space 1 to the comet 19P/Borrelly at the end of the '90s, SMART-1 to the Moon in 2003 and the future BepiColombo mission to Mercury.

Low thrust trajectories are explained here by making simplified assumptions on the magnitude and direction of the corresponding acceleration. This approach gives us some analytical insights on the problem. In more realistic situations, numerical methods and optimization techniques are needed.

Let us rewrite the equations of motion of a spacecraft moving because of a thrust and the gravitational acceleration due to a nominal central body. Following Sec. 4.2, we have[2]

$$\begin{cases} \dot{\mathbf{r}} = \mathbf{v}, \\ \dot{\mathbf{v}} = \mu \frac{\mathbf{r}}{r^3} + \frac{\mathbf{T}}{m}, \\ \dot{m} = -\frac{T}{g_0 I_{sp}}. \end{cases} \tag{4.28}$$

[2]Notice that the third equation can be adopted only when $m(t)$ does not present a discontinuity, which occurs, for instance, when some modules are detached from the main body of the object.

Under the hypothesis of constant low thrust, we have

$$\frac{\partial T}{\partial t} = \frac{\partial T}{\partial r} = 0.$$

Let us also assume that

- at time $t = t_0$ the spacecraft is moving on a circular orbit of radius r_0;
- the thrust direction always belongs to the orbital plane and thus this will not change;
- the acceleration due to the thrust is constant in modulus, namely $a_T = T/m$. This means, in particular, that we neglect the mass variation due the propellant consumption.

In an inertial reference system with origin at the central body (see Fig. 4.6), the equations of motion read (recall Eqs. (1.14) and (1.19))

$$\begin{cases} \dot{r} = v_r, \\ r\dot{\theta} = v_\theta, \\ \dot{v_r} = \ddot{r} - r\dot{\theta}^2 =: a_r, \\ \dot{v_\theta} = 2\dot{r}\dot{\theta} + r\ddot{\theta} =: a_\theta. \end{cases} \tag{4.29}$$

4.5.1 Constant Tangential Acceleration

Let us consider the case when the only acceleration on the radial direction is the one due to the gravity and the thrust due to the propulsion system takes place on the tangential direction, this is,

$$a_r = -\frac{\mu}{r^2}, \qquad a_\theta = a_T.$$

From Eqs. (4.29), we have

$$\begin{cases} \ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2}, \\ 2r\dot{r}\dot{\theta} + r^2\ddot{\theta} = ra_T. \end{cases} \tag{4.30}$$

Let us assume that for any t the orbit of the spacecraft can be approximated by a circular orbit, this is, a_T is low enough that $\ddot{r} \approx 0$. With this hypothesis, the centrifugal term is compensated by the gravitational acceleration at each instant, namely,

$$r\dot{\theta}^2 \approx \frac{\mu}{r^2}, \qquad \Longrightarrow \qquad \dot{\theta} \approx \sqrt{\frac{\mu}{r^3}}. \tag{4.31}$$

Notice that, because of this assumption, we are allowed to consider as coincident the transversal and the tangential directions.

From the last equation in (4.30), we get

$$2r\dot{r}\dot{\theta} + r^2\ddot{\theta} = \frac{d(r^2\dot{\theta})}{dt} = \frac{d}{dt}\sqrt{\mu r} = \frac{1}{2}\dot{r}\sqrt{\frac{\mu}{r}} = ra_T, \qquad \Longrightarrow \qquad \frac{1}{2}\sqrt{\frac{\mu}{r^3}}dr = a_T dt.$$

By integration, we then obtain

$$-\sqrt{\frac{\mu}{r}}+\sqrt{\frac{\mu}{r_0}}=a_T(t-t_0),$$

that is,

$$r=\frac{r_0v_0^2}{[v_0-a_T(t-t_0)]^2}, \tag{4.32}$$

where $v_0=\sqrt{\mu/r_0}$ is the modulus of the velocity on the circular orbit at t_0.

Now, we can ask ourselves when the spacecraft will manage to escape from a close orbit around the central orbit, this is, when the orbital energy varies from $\mathcal{E}<0$ to $\mathcal{E}=0$ thanks to the low thrust. Because of (4.31), we have

$$\frac{v^2}{2}=\frac{1}{2}(v_r^2+v_\theta^2)=\frac{1}{2}(\dot r^2+r^2\dot\theta^2)=\frac{1}{2}\Big(\dot r^2+\frac{\mu}{r}\Big),$$

and therefore

$$\mathcal{E}=\frac{v^2}{2}-\frac{\mu}{r}=\frac{\dot r^2}{2}-\frac{\mu}{2r}=0 \qquad\Longleftrightarrow\qquad \dot r=\sqrt{\frac{\mu}{r}}.$$

Moreover, from (4.31)

$$\ddot\theta=-\frac{3}{2}\dot r\sqrt{\frac{\mu}{r^5}},$$

and from the last equation in (4.30) we get

$$-\frac{3}{2}r\dot r\sqrt{\frac{\mu}{r^5}}+2\dot r\sqrt{\frac{\mu}{r^3}}=\frac{1}{2}\dot r\sqrt{\frac{\mu}{r^3}}=a_T \qquad\Longrightarrow\qquad \frac{1}{2}\sqrt{\frac{\mu}{r}}\sqrt{\frac{\mu}{r^3}}=\frac{1}{2}\frac{\mu}{r^2}=a_T.$$

Since $\mu=v_0^2r_0$, the radius of the escape orbit as a function of the initial radius and velocity and of the thrust acceleration is

$$r_e=\frac{r_0v_0}{\sqrt{2r_0a_T}}. \tag{4.33}$$

By means of (4.32) we know the time which is required to achieve the escape condition, namely,

$$r_e=\frac{r_0v_0^2}{[v_0-a_T(t_e-t_0)]^2}=\frac{r_0v_0}{\sqrt{2r_0a_T}}, \qquad\Longrightarrow\qquad a_T(t_e-t_0)=v_0-\sqrt{v_0}(2r_0a_T)^{1/4},$$

this is,

$$t_e=t_0+\frac{v_0}{a_T}\Big(1-\frac{(2r_0a_T)^{1/4}}{\sqrt{v_0}}\Big). \tag{4.34}$$

4.5.2 Constant Radial Acceleration

Let us consider now the other case, when the thrust acceleration is given along the radial direction. From Eqs. (4.29), we have now

$$\begin{cases} \ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2} + a_T, \\ 2r\dot{r}\dot{\theta} + r^2\ddot{\theta} = \frac{d(r^2\dot{\theta})}{dt} = 0. \end{cases} \tag{4.35}$$

The last equation tells us that the angular momentum $h = r^2\dot{\theta}$ is an integral of motion. Thus

$$r^2\dot{\theta} = r_0^2\dot{\theta}_0 = r_0 v_0 = \sqrt{\mu r_0},$$

since at t_0 we assumed the orbit to be circular. The first equation in (4.35) results in

$$\ddot{r} - r\dot{\theta}^2 + \frac{\mu}{r^2} = \ddot{r} - \frac{\mu r_0}{r^3} + \frac{\mu}{r^2} = a_T, \qquad \Longrightarrow \qquad a_T = \ddot{r} + \frac{\mu}{r^3}(r - r_0).$$

By multiplication for $\dot{r}$ we get

$$\dot{r}a_T = \dot{r}\ddot{r} + \dot{r}\frac{\mu}{r^3}(r - r_0),$$

which integrated (recall that a_T is assumed constant) gives

$$ra_T + constant = \frac{1}{2}\dot{r}^2 - \frac{\mu}{r} + \frac{1}{2}\mu\frac{r_0}{r^2}.$$

Since at t_0 we have $\dot{r}_0 = 0$,

$$constant = -r_0 a_T - \frac{\mu}{2r_0}, \tag{4.36}$$

and hence

$$ra_T - r_0 a_T - \frac{\mu}{2r_0} = \frac{1}{2}\dot{r}^2 - \frac{\mu}{r} + \frac{1}{2}\mu\frac{r_0}{r^2}, \qquad \Longrightarrow \qquad \dot{r}^2 = 2a_T(r - r_0) + \mu\frac{-r_0^2 + 2rr_0 - r^2}{r_0 r^2}.$$

The radial velocity as a function of the radius is thus

$$\dot{r}^2 = (r - r_0)\Big(2a_T - \mu\frac{r - r_0}{r_0 r^2}\Big). \tag{4.37}$$

In order to find the radius corresponding to the escape condition, let us proceed as in the previous section. The kinetic contribution is

$$\frac{v^2}{2} = \frac{1}{2}(\dot{r}^2 + r^2\dot{\theta}^2) = \frac{1}{2}\Big[(r - r_0)\Big(2a_T - \mu\frac{r - r_0}{r_0 r^2}\Big) + \frac{\mu r_0}{r^2}\Big],$$

and the condition $\mathcal{E} = 0$ is fulfilled for r_e such that

$$\frac{1}{2}\Big[(r_e - r_0)\Big(2a_T - \mu\frac{r_e - r_0}{r_0 r_e^2}\Big) + \frac{\mu r_0}{r_e^2}\Big] - \frac{\mu}{r_e} = 0.$$

With some manipulation, we obtain

$$a_T(r_e - r_0) = \mu \frac{r_e^2 + \cancel{r_0^2} - \cancel{2r_e r_0} - \cancel{r_0^2} + \cancel{2r_e r_0}}{2r_e r_0^2} = \mu \frac{\cancel{r_e^2}}{2r_0 \cancel{r_e^2}} = \frac{\mu}{2r_0}.$$

It follows that the escape radius is

$$r_e = r_0 + \frac{\mu}{2a_T r_0}. \tag{4.38}$$

If we use this expression in (4.37), we obtain

$$\begin{aligned} \dot{r}^2 &= (r - r_0)\Big(\frac{\mu}{r_0(r_e - r_0)} - \mu \frac{r - r_0}{r_0 r^2}\Big) = \mu \frac{r - r_0}{r_0 r^2 (r_e - r_0)}\Big(r^2 - (r - r_0)(r_e - r_0)\Big) \\ &= \frac{2a_T}{r^2}(r - r_0)\Big(r^2 - (r - r_0)(r_e - r_0)\Big). \end{aligned} \tag{4.39}$$

The above expression tells us that there exists a configuration for which the radial velocity is zero (apart from the trivial case of the initial circular orbit). This may happen before the escape condition is accomplished, if the thrust acceleration is not high enough. Indeed,

$$r^2 - (r - r_0)(r_e - r_0) = 0 \qquad \Longleftrightarrow \qquad r = \frac{(r_e - r_0) \pm \sqrt{(r_e - r_0)(r_e - 5r_0)}}{2}.$$

It exists r such that the escape condition is not achieved if the argument of the square root is positive, that is, if and only if

$$r_e > 5r_0.$$

Since we don't want this to happen, using (4.38), it is required that

$$a_T > \frac{\mu}{8r_0^2}, \tag{4.40}$$

in order to escape from the initial circular orbit with a constant thrust on the radial direction. In this case, the expression for the time of escape t_e is more complicated than the tangential case and we do not consider it here.

Interplanetary Trajectories 5

5.1 Patched Conic Approach

In the design of the trajectory required by an interplanetary mission, the first approximation considered is usually the Two–Body Problem. This dynamical model provides a good description of the motion of the probe, by assuming that one massive body at a time is responsible of the behavior of the particle. Indeed, in the space between two given planets, the dominant acceleration is the one due to the gravitational attraction of the Sun, the other perturbations can be neglected and the orbit can be considered a heliocentric Keplerian ellipse. In approaching a planet, the Keplerian orbit is instead a planetocentric hyperbola and the interaction between the probe and the massive body is assumed to be instantaneous, that is, the corresponding interval of time can be considered much smaller than the whole duration of the mission. Due to the planetary encounter, the heliocentric velocity of the spacecraft changes, both in modulus and direction, according to the relative planet-s/c velocity and to the distance of closest approach. During the hyperbolic approach the Sun affects the motion of the planet almost as much as it affects the motion of the probe and this is why it can be 'turned off'.

In this way, the whole trajectory is decomposed in arcs of conic sections, which are patched in a suitable manner. This is why the method is called *patched conic approach*. It ensures to analyze the problem easily and with a sufficient precision, at least for the preliminary phase of the orbit design. Starting from an initial approximation of this kind, the accurate computation of the interplanetary trajectory requires, in a second step, the numerical integration of the equations of motion which account for all the perturbations.

As example, we will consider a hypothetic mission from the Earth to Mars. The probe is located initially on a parking orbit around the Earth (e.g., a LEO) and must be transferred onto a parking orbit around Mars. From a macroscopic perspective, the spacecraft leaves the heliocentric orbit of the Earth and moves onto a different heliocentric orbit in order to reach the heliocentric orbit of Mars. During the transfer, the main attractor is the Sun and the study can be carried on by assuming a heliocentric Keplerian motion. In the proximity of the Earth (Mars), the probe is assumed to be subject only to the gravitational acceleration of the planet. In the following, we will consider that all the orbits lie on the same orbital plane.

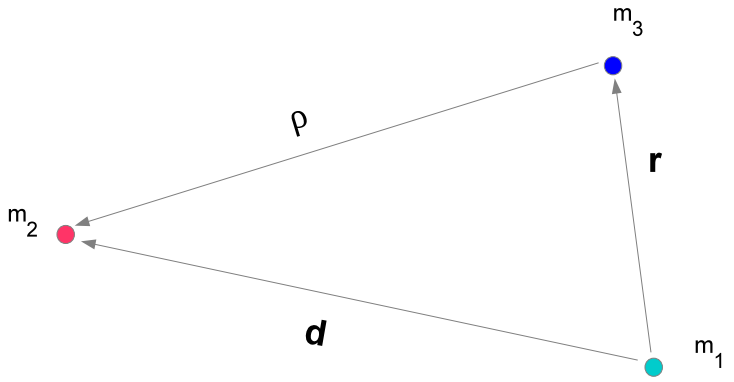


Figure 5.1: Relative position of three bodies in an inertial reference system.

5.2 Sphere of Influence

The most critical issue in the patched conic approximation is to define the point of transition between the realms of the two massive bodies (Sun and planet), say P_1 and P_2 of mass m_1 and m_2, respectively. To this end, we introduce the concept of *sphere of influence*, which, according to Laplace, is the locus of points, measured with respect to P_1 and P_2, where the ratios between the main gravitational acceleration and the perturbing one are equal.

Let us recall the third-body perturbation considered in Sec. 3.4 and refer to Fig. 5.1. If m_2 is the perturbing body, then the equation of motion of the probe with respect m_1 is

$$\frac{d^2\mathbf{r}}{dt^2} = -G(m_1+m_3)\frac{\mathbf{r}}{r^3} + Gm_2\left(\frac{\boldsymbol{\rho}}{\rho^3} - \frac{\mathbf{d}}{d^3}\right) \equiv \mathbf{f}_1^K + \mathbf{f}_1^p, \tag{5.1}$$

where m_3 the mass of the s/c, $\mathbf{r}$ is the radius vector of the s/c with respect to m_1, $\mathbf{d}$ the radius vector of m_2 with respect to m_1, $\boldsymbol{\rho}$ the radius vector of m_2 with respect to m_3, and the subscripts K and p indicate the Keplerian and the perturbing accelerations, respectively. Instead, if the perturbing body is assumed to be m_1, then the equation of motion of the probe with respect to m_2 is

$$\frac{d^2\boldsymbol{\rho}}{dt^2} = -G(m_2+m_3)\frac{\boldsymbol{\rho}}{\rho^3} + Gm_1\left(\frac{\mathbf{r}}{r^3} - \frac{\mathbf{d}}{d^3}\right) \equiv \mathbf{f}_2^K + \mathbf{f}_2^p. \tag{5.2}$$

In each case, the ratio between the modulus of the perturbing acceleration and the Keplerian one can be interpreted as the error introduced when the former term is neglected.

In this sense, the sphere of influence is the locus of points where the error occurred by neglecting the gravitational acceleration of one body or the other is the same. We have

$$\begin{aligned}\frac{f_1^p}{f_1^K} &= \frac{Gm_2\left|\frac{\boldsymbol{\rho}}{\rho^3}-\frac{\mathbf{d}}{d^3}\right|}{G(m_1+m_3)\left|\frac{\mathbf{r}}{r^3}\right|} \\ &= \frac{m_2}{m_1+m_3}r^2\left|\frac{\boldsymbol{\rho}}{\rho^3}-\frac{\mathbf{d}}{d^3}\right|, \end{aligned} \tag{5.3}$$

and

$$\begin{aligned}\frac{f_2^p}{f_2^K} &= \frac{Gm_1\left|\frac{\mathbf{r}}{r^3}-\frac{\mathbf{d}}{d^3}\right|}{G(m_2+m_3)\left|\frac{\boldsymbol{\rho}}{\rho^3}\right|} \\ &= \frac{m_1}{m_2+m_3}\rho^2\left|\frac{\mathbf{r}}{r^3}-\frac{\mathbf{d}}{d^3}\right|, \end{aligned} \tag{5.4}$$

and we look for the distance r with respect to P_1 such that

$$\frac{f_1^p}{f_1^K}=\frac{f_2^p}{f_2^K}.$$

If we consider negligible m_3 with respect to m_1 and m_2 and $\rho \approx d$[1] in (5.3), then (see Fig. 5.1)

$$\begin{aligned}\frac{f_1^p}{f_1^K} &\approx \frac{m_2}{m_1}r^2\left|\frac{\boldsymbol{\rho}-\mathbf{d}}{\rho^3}\right| \\ &\approx \frac{m_2}{m_1}r^2\left|\frac{-\mathbf{r}}{\rho^3}\right| \\ &\approx \frac{m_2}{m_1}\frac{r^3}{\rho^3}. \end{aligned} \tag{5.5}$$

Instead, by assuming $d >> r$, Eq. (5.4) can be approximated as

$$\begin{aligned}\frac{f_2^p}{f_2^K} &\approx \frac{m_1}{m_2}\rho^2\left|\frac{\mathbf{r}}{r^3}\right| \\ &\approx \frac{m_1}{m_2}\frac{\rho^2}{r^2}. \end{aligned} \tag{5.6}$$

Therefore, by equating (5.5) and (5.6), we obtain

$$\frac{m_2}{m_1}\frac{r^3}{\rho^3}=\frac{m_1}{m_2}\frac{\rho^2}{r^2},$$

that is,

$$\left(\frac{\rho}{r}\right)^5=\left(\frac{m_2}{m_1}\right)^2. \tag{5.7}$$

[1]This approximation is reasonable for all the planets with respect to the Sun.

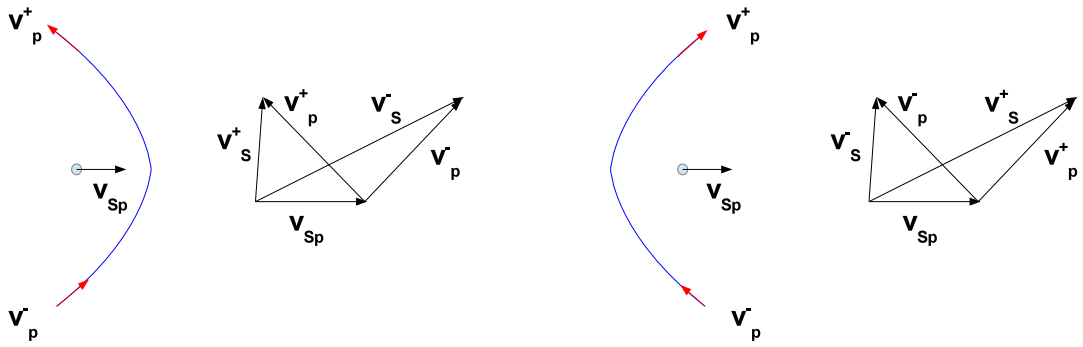


Figure 5.2: The gravity assist. Left: the case when the encounter takes place in front of the planet, or on its leading side. Right: the case when the encounter takes place behind the planet, or on its trailing side.

For instance, let us consider the sphere of influence of the Earth with respect to the Sun. We have $m_1 \approx 5.97 \times 10^{24}$ kg, $m_2 \approx 1.99 \times 10^{30}$ kg, $\rho \approx d \approx 1.49597870691 \times 10^8$ km and thus

$$r = \rho \left(\frac{m_1}{m_2} \right)^{2/5} \approx 924216 \text{ km}.$$

Furthermore, the sphere of influence of the Moon, taking the Earth as perturbing body, has radius

$$r = \rho \left(\frac{m_1}{m_2} \right)^{2/5} \approx 66171 \text{ km},$$

because $m_1 \approx 7.34 \times 10^{22}$ kg, $m_2 \approx 5.97 \times 10^{24}$ kg, $\rho \approx d \approx 384400$ km.

5.3 Gravity Assist

When the probe enters into the sphere of influence of a given planet, it can take advantage of the corresponding gravitational attraction in order to increase or decrease its heliocentric orbital velocity. This concept is called *gravity assist*, or *fly-by*, or *swing-by* or *slingshot effect*. Since the angular momentum of the Solar System (assumed as a $n-$Body Problem) is constant (see Sec. **??**), the effect can be seen as an exchange of energy and angular momentum between the s/c and the planet. Because of that, the semi-major axis, eccentricity and inclination of the heliocentric orbit of the s/c may change. At approaching the planet, the relative s/c-planet velocity does not change in modulus, but in direction, because we deal with a hyperbolic encounter. As a consequence, the absolute velocity of the probe with respect to the Sun changes in modulus and direction. The first missions to exploit this phenomenon were Pioneer 10 and Mariner 10, which were pulled, respectively, outward by the gravity of Jupiter in 1973 and toward Mercury by

the gravity of Venus in 1974. Later on, gravity assists became a common procedure in the design of interplanetary trajectories.

Let us denote as

- $\mathbf{v}_S$ the heliocentric velocity of the probe;
- $\mathbf{v}_p$ the planetocentric velocity of the probe;
- $\mathbf{v}_{Sp}$ the heliocentric velocity of the planet.

We have

$$\mathbf{v}_S = \mathbf{v}_p + \mathbf{v}_{Sp}. \tag{5.8}$$

When the s/c approaches the planet, two situations can occur, namely,

1. the s/c loses kinetic energy in its heliocentric motion (Fig. 5.2, left), that is,

$$||\mathbf{v}_S^+|| < ||\mathbf{v}_S^-||;$$

2. the s/c gains kinetic energy in its heliocentric motion (Fig. 5.2, right), that is,

$$||\mathbf{v}_S^+|| > ||\mathbf{v}_S^-||,$$

where the subscripts $-$ and $+$ indicate the entrance and the exit into the sphere of influence, respectively.

Let us assume that the radius of the sphere of influence, say r_∞, is infinitesimal with respect to the radius of the orbit (assumed circular) of the corresponding planet, $r_\oplus$ for the Earth and $r_{\sigma'}$ for Mars, and that the radius of the planet, say r_p, is infinitesimal with respect to the radius of its sphere of influence. For an observer on the planet the s/c is seen to arrive on a hyperbolic orbit, i.e., with semi-major axis $a < 0$ and eccentricity $e > 1$. We recall the concepts introduced in Sec. 1.6.3 for this kind of conic section. The planetocentric arrival and departure velocities at the sphere of influence, called *hyperbolic excess velocities*, $\mathbf{v}_\infty^+$ and $\mathbf{v}_\infty^-$, are directed along the asymptotes of the hyperbola and are such that

$$||\mathbf{v}_\infty^+|| = ||\mathbf{v}_\infty^-|| = \sqrt{\frac{\mu_p}{a'}},$$

where $a' = |a|$ and μ_p is the mass parameter corresponding to the given planet. Indeed, from the energy's equation (1.37)

$$\frac{v_\infty^2}{2} - \frac{\mu_p}{r} = -\frac{\mu_p}{2a} \qquad \Longrightarrow \qquad v_\infty = \lim_{r\to\infty} \sqrt{\frac{2\mu_p}{r} - \frac{\mu_p}{a}} = \sqrt{\frac{\mu_p}{a'}}.$$

Let us denote with β the angle between the direction of the asymptotes and the apsis line of the hyperbola (see Fig. 5.3). Since we assumed r_∞ to be much larger than the radius of the planet, then

$$\cos\theta_\infty = \lim_{r\to\infty} \frac{1}{e}\left(\frac{p}{r} - 1\right) = -\frac{1}{e}. \tag{5.9}$$

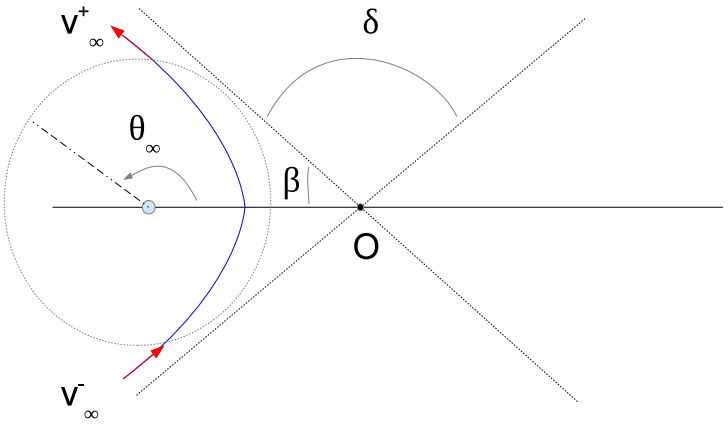


Figure 5.3: The hyperbolic excess velocity, the corresponding true anomaly and the deflection angle.

Moreover, since

$$\theta_\infty = \pi - \beta,$$

we have

$$\cos\beta = -\cos\theta_\infty = \frac{1}{e}. \tag{5.10}$$

Due to the hyperbolic encounter, the hyperbolic excess velocity is turned by an angle δ, the *deflection angle*, such that

$$2\beta + \delta = \pi,$$

that is,

$$\frac{1}{e} = \cos\beta = \cos\left(\frac{\pi}{2} - \frac{\delta}{2}\right) = \sin\frac{\delta}{2}. \tag{5.11}$$

If we consider negligible the interval of time spent inside the sphere of influence, then the interaction between the probe and the planet can be seen as an impulsive maneuver. For the law of cosines (see Fig. 5.4), we have

$$\begin{aligned} \Delta v &= \sqrt{v_\infty^2 + v_\infty^2 - 2v_\infty^2\cos\delta} \\ &= \sqrt{2v_\infty^2\,(1-\cos\delta)} \\ &= 2v_\infty \sin\frac{\delta}{2} \\ &= 2\frac{v_\infty}{e}, \end{aligned} \tag{5.12}$$

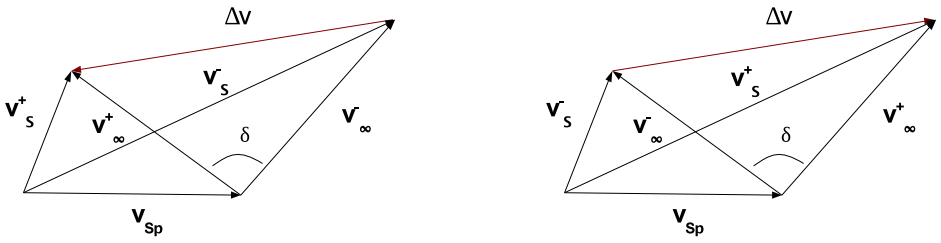


Figure 5.4: The impulse given by the hyperbolic encounter.

which states that the impulse is maximum when the eccentricity is minimum, that is, when $e = 1$. In this case, $\mathbf{v}_\infty$ is aligned along $\mathbf{v}_{Sp}$ and $\Delta v = 2v_\infty$. However, we must ensure that the probe does not collide with the planet, that is, that the pericenter of the hyperbola is higher than the radius of the planet, namely,

$$r_{peri} = a'(e-1) > r_{planet}, \tag{5.13}$$

or

$$\frac{\mu_p}{v_\infty^2}(e-1) > r_p. \tag{5.14}$$

This implies

$$e > 1 + r_p \frac{v_\infty^2}{\mu_p},$$

and thus

$$e_{min} = 1 + r_p \frac{v_\infty^2}{\mu_p}. \tag{5.15}$$

We notice that, from a practical point of view, this eccentricity is not attainable, but it provides the limit for the maximum impulse achievable for a given value of v_∞. Indeed, from (5.12) it follows

$$\Delta v_{max} = 2\frac{v_\infty}{1 + r_p \frac{v_\infty^2}{\mu_p}}. \tag{5.16}$$

Furthermore, we can look for the value of v_∞ such that Δv_{max} is maximum. To this end, let us compute

$$\begin{aligned} \frac{\partial \Delta v_{max}}{\partial v_\infty} &= \frac{2\left(1 + r_p \frac{v_\infty^2}{\mu_p}\right) - 2v_\infty 2v_\infty \frac{r_p}{\mu_p}}{\left(1 + r_p \frac{v_\infty^2}{\mu_p}\right)^2} \\ &= \frac{2 - 2r_p \frac{v_\infty^2}{\mu_p}}{\left(1 + r_p \frac{v_\infty^2}{\mu_p}\right)^2} = 0, \end{aligned}$$

that is,

$$v_\infty = \sqrt{\frac{\mu_p}{r_p}}. \tag{5.17}$$

This value corresponds to a maximum because the second derivative reads

$$\begin{aligned}
\frac{\partial^2 \Delta v_{max}}{\partial v_\infty^2} &= \frac{-4r_p\frac{v_\infty}{\mu_p}\left(1+r_p\frac{v_\infty^2}{\mu_p}\right)^2 - 2\left(1-r_p\frac{v_\infty^2}{\mu_p}\right)2\left(1+r_p\frac{v_\infty^2}{\mu_p}\right)2r_p\frac{v_\infty}{\mu_p}}{\left(1+r_p\frac{v_\infty^2}{\mu_p}\right)^4} \\
&= \frac{-4r_p\frac{v_\infty}{\mu_p}\left(1+r_p\frac{v_\infty^2}{\mu_p}\right)\left(1+r_p\frac{v_\infty^2}{\mu_p}+1-r_p\frac{v_\infty^2}{\mu_p}\right)}{\left(1+r_p\frac{v_\infty^2}{\mu_p}\right)^4} \\
&= \frac{-8r_p\frac{v_\infty}{\mu_p}}{\left(1+r_p\frac{v_\infty^2}{\mu_p}\right)^3} < 0.
\end{aligned}$$

5.4 From the Earth to Mars

Let us assume the heliocentric orbits of the Earth and Mars to be circular of radius $r_\oplus$ and $r_\male$, respectively. The aim of the transfer we are going to design is to move the s/c from a parking circular orbit around the Earth of radius r_i to a parking circular orbit around Mars of radius r_f. We decompose the trajectory in three parts, namely,

1. the planetocentric phase of escape from the gravitational field of the Earth;
2. the heliocentric transfer;
3. the planetocentric phase of capture by the gravitational field of Mars.

5.4.1 Heliocentric Transfer

Let us start from phase 2., because it defines the heliocentric velocity of the s/c at the boundary of the Earth's sphere of influence, say $\mathbf{v}_d$, and the one at the boundary of the Mars' sphere of influence, say $\mathbf{v}_a$. Let us assume, for the moment, that these spheres of influence are points, which coincide with the corresponding planet. We perform a Hohmann transfer (see Sec. 4.3.1), starting at the time when the angular distance between the two planets satisfies the requirement described in Sec. **??** for the rendezvous problem. The transfer orbit has semi-major axis equal to

$$a = \frac{r_\oplus + r_\male}{2},$$

and from the energy's equation (1.37) we compute the modulus of the departure velocity v_d, needed by the elliptic transfer. We have

$$\frac{v_d^2}{2} - \frac{\mu_\odot}{r_\oplus} = -\frac{\mu_\odot}{2a} = -\frac{\mu_\odot}{r_\oplus + r_\male},$$

that is,

$$v_d = \sqrt{2\mu_\odot \frac{r_{\male}}{r_\oplus \left(r_\oplus + r_{\male}\right)}}. \tag{5.18}$$

In the same way, we find for the arrival velocity

$$\frac{v_a^2}{2} - \frac{\mu_\odot}{r_{\male}} = -\frac{\mu_\odot}{r_\oplus + r_{\male}},$$

that is,

$$v_a = \sqrt{2\mu_\odot \frac{r_\oplus}{r_{\male} \left(r_\oplus + r_{\male}\right)}}. \tag{5.19}$$

5.4.2 Escape and Capture

Since we perform a Hohmann transfer and for that we assume that the sphere of influence of the Earth coincides with the planet, at the exit from the Earth's sphere of influence the heliocentric velocity of the s/c is orthogonal to the line joining the Earth and the Sun and lies on the same direction of the heliocentric velocity of the Earth, say $\mathbf{v}_\oplus$. Therefore, the velocity of the s/c with respect to the Earth at the exit from the sphere of influence is

$$v_\infty^+ = \sqrt{\frac{2\mu_\odot}{r_\oplus + r_{\male}} \left(\frac{r_{\male}}{r_\oplus}\right)} - v_\oplus, \tag{5.20}$$

where

$$v_\oplus = \sqrt{\frac{\mu_\odot}{r_\oplus}}.$$

We notice that since we are considering a transfer to an outer planet, it starts from behind the Earth (the energy increases). In the opposite case, i.e., when we design a transfer to an inner planet, v_∞^+ above has opposite sign and we depart in front of the Earth (the energy reduces).

We assumed that the s/c is initially located on a circular parking orbit of radius r_i at the Earth and thus velocity

$$v_c = \sqrt{\frac{\mu_\oplus}{r_i}}.$$

We have to apply a maneuver, say Δv_i, to move the probe from such parking orbit to the escape hyperbolic orbit with v_∞ given by (5.20). The minimum Δv_i required to this end is the tangential maneuver corresponding to the insertion into the perigee of the hyperbola. From the conservation of energy, at the perigee the velocity on the hyperbola is given by

$$\frac{v_i^2}{2} - \frac{\mu_\oplus}{r_i} = \frac{(v_\infty^+)^2}{2} \qquad \Longrightarrow \qquad v_i = \sqrt{(v_\infty^+)^2 + 2\frac{\mu_\oplus}{r_i}},$$

and thus the maneuver to apply is

$$\Delta v_i = v_i - v_c = \sqrt{(v_\infty^+)^2 + 2\frac{\mu_\oplus}{r_i}} - \sqrt{\frac{\mu_\oplus}{r_i}}. \tag{5.21}$$

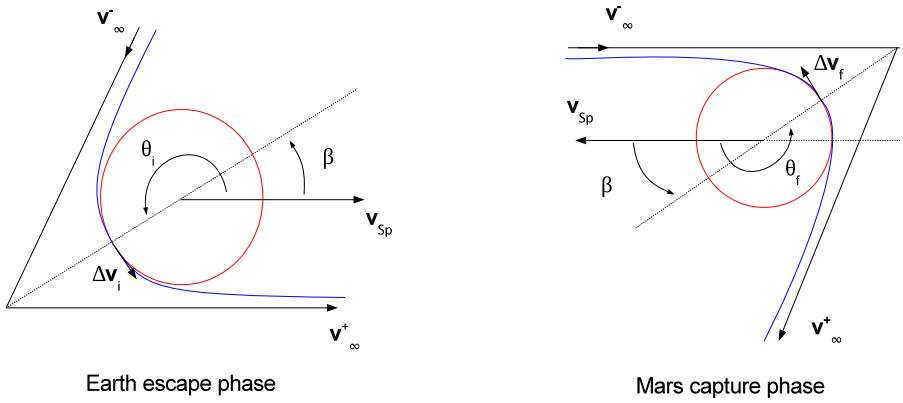


Figure 5.5: The escape and capture phases.

The semi-major axis of the hyperbola can be derived from v_∞^+ as

$$a = -\frac{\mu_\oplus}{(v_\infty^+)^2}, \tag{5.22}$$

and its eccentricity as

$$e = -\frac{r_i}{a} + 1 = \frac{r_i}{\mu_\oplus}(v_\infty^+)^2 + 1, \tag{5.23}$$

from which we can compute the angle β as

$$\beta = \cos^{-1}\left(\frac{1}{e}\right).$$

This angle is important because it tells us the phase on the circular orbit, at which the maneuver must be applied. For the escape stage, this is

$$\theta_i = \pi + \beta.$$

See Fig. 5.5 (left).

Concerning the capture at Mars, in an analogous way we find that the hyperbolic excess velocity at the entrance of the sphere of influence of Mars is

$$v_\infty^- = v_{\mathord{\male}} - \sqrt{\frac{2\mu_\odot}{r_\oplus + r_{\mathord{\male}}}\left(\frac{r_\oplus}{r_{\mathord{\male}}}\right)}, \tag{5.24}$$

where

$$v_{\mathord{\male}} = \sqrt{\frac{\mu_\odot}{r_{\mathord{\male}}}}.$$

To insert into the circular orbit of radius r_f at Mars, the maneuver Δv_f is

$$\Delta v_f = -\sqrt{\frac{\mu_{\mathord{\male}}}{r_f}} + \sqrt{(v_\infty^-)^2 + 2\frac{\mu_{\mathord{\male}}}{r_f}}. \tag{5.25}$$

In this case, the energy reduces and thus the encounter occurs in front of the planet.
The eccentricity of the hyperbola with pericenter r_f and v_∞^- as above is

$$e = \frac{r_f (v_\infty^-)^2}{\mu_{\sigma'}} + 1, \tag{5.26}$$

and the phase of insertion into the circular orbit, computed counterclockwise with respect to the direction given by the velocity of Mars, is

$$\theta_f = \pi + \beta = \pi + \cos^{-1}\left(\frac{1}{e}\right).$$

See Fig. 5.5 (right).
Finally, we remark that if the heliocentric transfer does not consist in a Hohmann transfer, but in a Lambert one, then in computing (5.20) and (5.24) we must take into account the angle between the heliocentric departure/arrival velocity of the s/c and the heliocentric velocity of the planet and thus use Eq. (**??**).

The Lambert's Theorem 6

6.1 Statement of the Problem

The problem of Lambert consists in the computation of the orbit which allows to transfer from a given position P_1 to another given position P_2 in a fixed time interval $\Delta t = t_2 - t_1$. This is a general issue, not applicable only to the trajectory design, but also to the preliminary orbit determination. The solution of the Lambert's problem, for instance, represents a tool we can exploit for space rendez-vous (see Chap. **??**).[1] The first geometrical insight of the problem is due to Gauss: the asteroid Ceres was discovered by Giuseppe Piazzi on January 1, 1801 in its way toward the Sun and Gauss used these data to determine its orbit and predict new observations one year later. This is why some authors refer to the practical solution of the Lambert's problem as to the Gauss's problem. In this chapter, we will see in particular how the tools developed in Sec. **??** are related to the Lambert's problem.

[1] We remark that the Lambert's problem is not the rendez-vous problem.

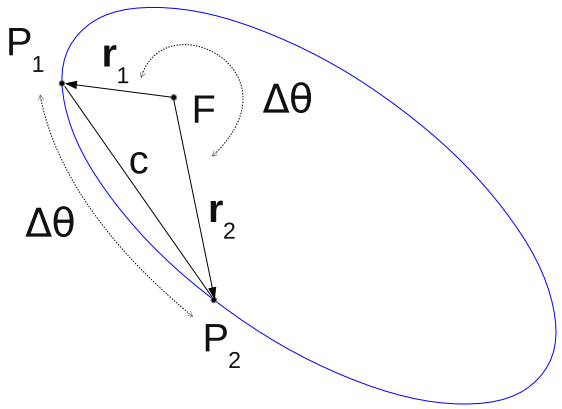


Figure 6.1: The geometry of the Lambert's problem.

In a fixed time of flight, there exist only two orbits that connect two nominal positions. Otherwise, if the duration of the flight is not defined a priori, then we have an infinite number of possibilities. Let us look to Fig. 6.1. The two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, going from the central body (e.g. Earth) to P_1 and P_2 respectively, determine the orbital plane associated with the desired transfer. To well pose the problem, we also need to define the direction of the transfer, that is, $\Delta\theta$. If the two vectors are opposite and collinear, then $\Delta\theta = \pi$ and the orbital plane is not determined. This means that there does not exist an unique solution. If, instead, $\Delta\theta = k2\pi$ with k integer and $\Delta t = 0$, then the solution is unique and it is a degenerate conic section.

6.1.1 The Lambert's Theorem

The Lambert's theorem states that the orbital transfer time depends on:

- the semi-major axis a of the conic section where the two points are located;
- the sum of the distances between the focus of the conic section and the points, i.e. $r_1 + r_2$;
- the length of the chord joining the two points, namely c.

In mathematical jargon:

$$\sqrt{\mu}\Delta t = f(a, r_1 + r_2, c), \tag{6.1}$$

where μ is the mass parameter corresponding to the nominal central body.

6.1.2 Proof

We prove the theorem assuming that there exists an elliptic orbit, of eccentricity e and semi-major axis a, joining the two points P_1 and P_2. Let E_1 and E_2 be the eccentric anomalies associated with P_1 and P_2, respectively. Following the Kepler's equation (see Eq. (1.76)),

$$\sqrt{\frac{\mu}{a^3}}\Delta t = E_2 - e\sin E_2 - E_1 + e\sin E_1 = E_2 - E_1 - e(\sin E_2 - \sin E_1). \tag{6.2}$$

Let us define

$$E_+ := \frac{E_2 + E_1}{2}, \qquad E_- := \frac{E_2 - E_1}{2},$$

in such a way that

$$\begin{aligned}
E_1 &= E_+ - E_-, \\
E_2 &= E_+ + E_-, \\
E_2 - E_1 &= 2E_-, \\
\sin E_2 - \sin E_1 &= \sin(E_+ + E_-) - \sin(E_+ - E_-) = 2\sin E_- \cos E_+, \\
\cos E_2 - \cos E_1 &= = -2\sin E_- \sin E_+, \\
\cos E_2 + \cos E_1 &= = 2\cos E_- \cos E_+,
\end{aligned}$$

and hence

$$\sqrt{\mu}\Delta t = 2\sqrt{a^3}(E_- - e\sin E_-\cos E_+). \tag{6.3}$$

Since we assumed the orbit to be elliptic and thus $e < 1$, we can introduce a new variable ψ such that

$$\cos\psi = e\cos E_+,$$

and (6.3) results in

$$\sqrt{\mu}\Delta t = 2\sqrt{a^3}(E_- - \sin E_-\cos\psi). \tag{6.4}$$

Now, from Eqs. (1.61)-(1.62), we have

$$\begin{aligned} r &= a(1 - e\cos E), \\ x &= a\cos E; \\ y &= a\sqrt{1-e^2}\sin E, \end{aligned}$$

and thus

$$r_1 + r_2 = a(2 - e(\cos E_2 + \cos E_1)) = 2a(1 - \cos E_-\cos\psi), \tag{6.5}$$

and

$$\begin{aligned} c^2 &= (x_2 - x_1)^2 + (y_2 - y_1)^2 \\ &= a^2(\cos E_2 - \cos E_1)^2 + a^2(1-e^2)(\sin E_2 - \sin E_1)^2 \\ &= 4a^2\sin^2 E_-(\sin^2 E_+ + \cos^2 E_+ - e^2\cos^2 E_+) \\ &= 4a^2\sin^2 E_-\sin^2\psi. \end{aligned} \tag{6.6}$$

From (6.5) and (6.6), if we define $\alpha = \psi + E_-$ and $\beta = \psi - E_-$ it turns out that

$$r_1 + r_2 + c = 2a(1 - \cos E_-\cos\psi + \sin E_-\sin\psi) = 2a(1-\cos\alpha) = 4a\sin^2\frac{\alpha}{2}, \tag{6.7}$$

and

$$r_1 + r_2 - c = 2a(1 - \cos E_-\cos\psi - \sin E_-\sin\psi) = 2a(1-\cos\beta) = 4a\sin^2\frac{\beta}{2}. \tag{6.8}$$

Since

$$\begin{aligned} E_- &= (\alpha - \beta)/2, \\ 2\sin E_-\cos\psi &= \sin(\psi + E_-) - \sin(\psi - E_-) = \sin\alpha - \sin\beta, \end{aligned}$$

Eq. (6.4) becomes

$$\sqrt{\mu}\Delta t = \sqrt{a^3}(\alpha - \beta - \sin\alpha + \sin\beta), \tag{6.9}$$

which proves the theorem. As a matter of fact, from (6.7) and (6.8), we have

$$\sin\frac{\alpha}{2} = \sqrt{\frac{r_1 + r_2 + c}{4a}}, \qquad \sin\frac{\beta}{2} = \sqrt{\frac{r_1 + r_2 - c}{4a}},$$

or, if s denotes the semi-perimeter of the triangle FP_1P_2, namely, $s = (r_1 + r_2 + c)/2$,

$$\sin\frac{\alpha}{2} = \sqrt{\frac{s}{2a}}, \qquad \sin\frac{\beta}{2} = \sqrt{\frac{s-c}{2a}}. \tag{6.10}$$

For the hyperbolic case, it can be proved that (6.9) reads

$$\sqrt{\mu}\Delta t = \sqrt{-a^3}(-\alpha + \beta + \sinh\alpha - \sinh\beta), \tag{6.11}$$

where

$$\sinh\frac{\alpha}{2} = \sqrt{-\frac{s}{2a}}, \qquad \sinh\frac{\beta}{2} = \sqrt{-\frac{s-c}{2a}}. \tag{6.12}$$

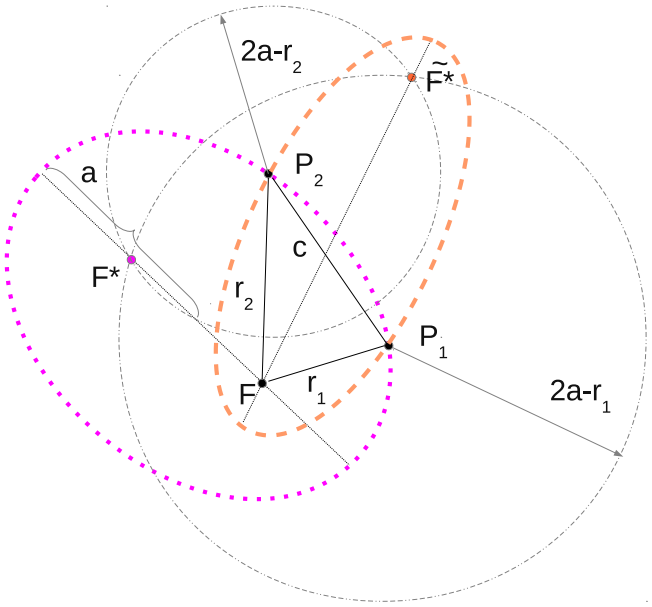


Figure 6.2: The two ellipses of semi-major axis $a > a_m$ joining P_1 and P_2.

6.2 Locus of the Vacant Foci

Since the angles α and β are defined through the sine of a square root, we must solve the ambiguity related to the quadrant they belong to. To this end, let us first understand where the second focus F^* of the ellipse, having semi-major axis a and going through P_1 and P_2, can be located. By definition, we have

$$P_1F + P_1F^* = 2a, \qquad P_2F + P_2F^* = 2a, \tag{6.13}$$

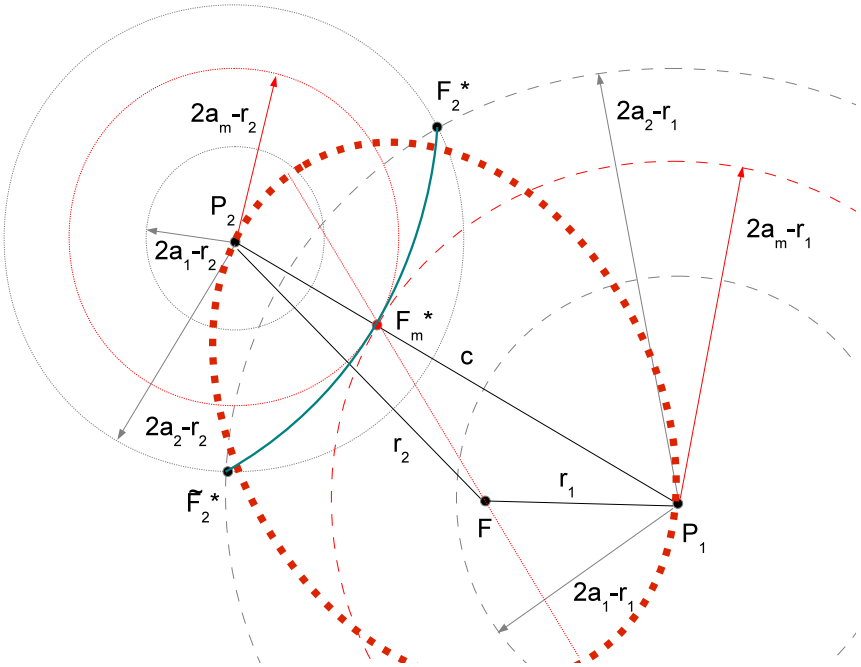


Figure 6.3: The ellipse of minimum energy of semi-major axis a_m and the geometry of the circles with radius $2a - r_1$ and $2a - r_2$ for different values of a. In green, the locus of the vacant foci.

this is,

$$P_1F^* = 2a - P_1F = 2a - r_1, \qquad P_2F^* = 2a - P_2F = 2a - r_2. \tag{6.14}$$

In other words, the second focus is at the point of intersection of two circles of radius $2a - r_1$ and $2a - r_2$ centered at P_1 and P_2, respectively (see Fig. 6.2). According to the value of the semi-major axis, it may happen that there do not exist two intersecting circles and therefore an ellipse joining the two points. The minimum value for the semi-major axis required to find the ellipse we are looking for, namely a_m, is the one associated with the *ellipse of minimum energy*. In this case, the two circles of radius $2a_m - r_1$ and $2a_m - r_2$ are tangent in one point F^*, which corresponds to the second focus and belongs to the chord c (see Fig. 6.3). The semi-major axis a_m is defined as

$$a_m = \frac{r_1 + r_2 + c}{4} = \frac{s}{2}, \tag{6.15}$$

and

$$P_1F^*_m = s - r_1, \qquad P_2F^*_m = s - r_2.$$

Indeed, from (6.13)

$$4a = P_1F + P_1F^* + P_2F + P_2F^* = r_1 + P_1F^* + r_2 + P_2F^*,$$

and the minimum value $P_1F^* + P_2F^*$ can take is c. The transfer time, according to (6.9), (6.10) and (6.15), on the ellipse of minimum energy is

$$\Delta t_m = \sqrt{\frac{s^3}{8\mu}}(\pi - \beta_m + \sin\beta_m). \tag{6.16}$$

For any $a > a_m$ the corresponding circles intersect in two points, F^* and $\tilde{F}^*$ equidistant from and on opposite sides with respect to the chord (see Figs. 6.2 and 6.3). Therefore there exist two conjugate ellipses with the same semi-major axis a satisfying our constraints and 4 elliptic arcs connecting P_1 and P_2, two such that $\Delta\theta < \pi$ and two such that $\Delta\theta > \pi$. Each one of these arcs is characterized by a different time of flight, because the two ellipses have different eccentricity, namely,

$$e = \frac{FF^*}{2a}, \qquad \tilde{e} = \frac{F\tilde{F}^*}{2a}. \tag{6.17}$$

Moreover, from (6.14)

$$P_2F^* - P_1F^* = -(r_2 - r_1),$$

which can be recognized as the equation of the hyperbola having semi-major axis

$$a_h = \frac{r_2 - r_1}{2}, \tag{6.18}$$

and foci P_1 and P_2. We can conclude that the locus of the vacant foci is a hyperbola of semi-major axis a_h and eccentricity

$$e_h = \frac{c}{r_2 - r_1}. \tag{6.19}$$

It can be also proved (Battin, 1999) that e_h is the reciprocal of the eccentricity, $e_F = 1/e_h$, of the fundamental ellipse, which is defined as the ellipse of minimum eccentricity containing P_1 and P_2. Such ellipse has the semi-major axis parallel to the chord, and equal to

$$a_F = \frac{r_1 + r_2}{2}. \tag{6.20}$$

6.3 Geometrical Interpretation of α and β

From Secs. 6.1.2 and 6.2, we know that for two fixed positions P_1 and P_2 and semi-major axis a, the shape of the ellipse connecting them can be changed by moving its foci, on condition that we do not modify $r_1 + r_2$ and the transfer time. In particular, since

$$P_1F + P_2F = r_1 + r_2,$$

the focus F can move on a ellipse of semi-major axis $(r_1 + r_2)/2$ of foci P_1 and P_2. Analogously, the focus F^* can move on an ellipse of semi-major axis $4a - r_1 - r_2$ confocal with the elliptic locus of F (see Fig. 6.4). Because of the way we defined such transformations, α and β are invariants.

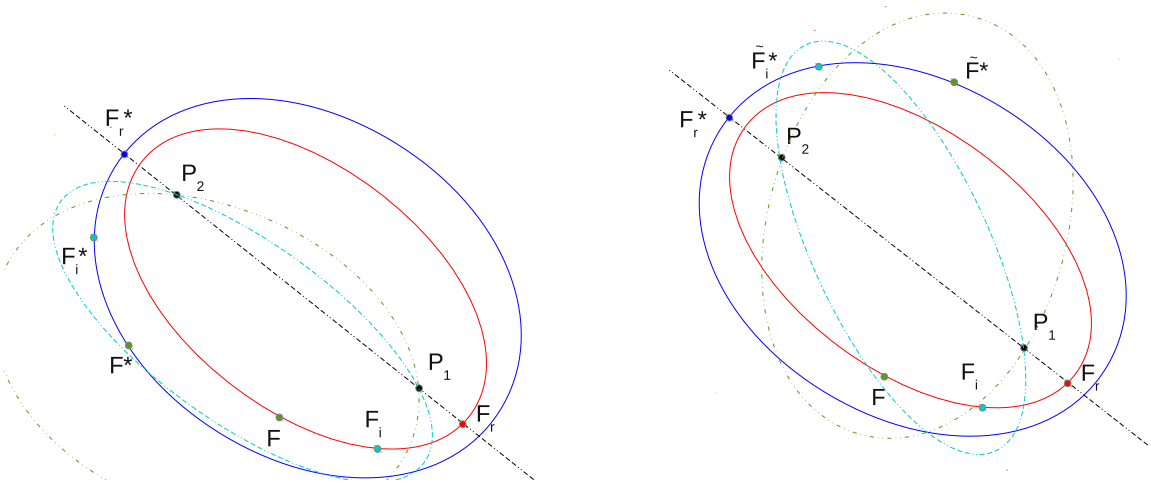


Figure 6.4: We show where (red and blue curves) the foci can move with the constraint of fixed $a, r_1 + r_2, c$. Left: the case of the vacant focus on the lower branch of the hyperbolic locus. Right: the one on the upper branch.

The limit case takes place when the two foci are altered up to lie on the line going from P_1 to P_2. The orbit linking P_1 and P_2 becomes a *rectilinear ellipse*, that is, an ellipse of eccentricity $e = 1$ as flat as possible (see also Sec. 1.6.2). In this case the chord c coincides with the elliptic arc going from P_1 to P_2 and we have

$$P_2F_r - P_1F_r = r_2 - r_1 = c,$$

where F_r is one of the foci of the rectilinear ellipse. Also, Eq. (6.2) with $e = 1$ results in

$$\sqrt{\mu}\Delta t = \sqrt{a^3}(E_2 - E_1 - \sin E_2 + \sin E_1), \tag{6.21}$$

and since for $e = 1$ we have

$$\begin{aligned} r_1 &= a(1 - \cos E_1) = 2a \sin^2\left(\frac{E_1}{2}\right), \\ r_2 &= a(1 - \cos E_2) = 2a \sin^2\left(\frac{E_2}{2}\right), \\ s &= \frac{r_1 + r_2 + c}{2} = \frac{r_1 + r_1 + c + c}{2} = r_1 + c, \\ s &= \frac{r_1 + r_2 + c}{2} = \frac{r_2 - c + r_2 + c}{2} = r_2, \end{aligned}$$

then

$$r_1 = s - c = 2a \sin^2\left(\frac{E_1}{2}\right), \qquad r_2 = s = 2a \sin^2\left(\frac{E_2}{2}\right). \tag{6.22}$$

By comparing (6.9) with (6.21) and (6.10) with (6.22), it is clear that α and β can be seen as the eccentric anomalies in the case of the rectilinear ellipse.

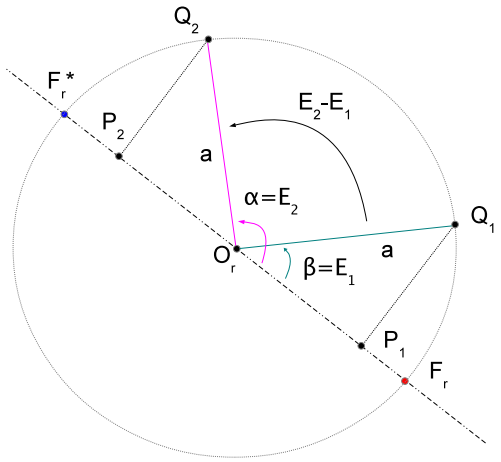


Figure 6.5: Geometrical interpretation of α and β in relation to the rectilinear ellipse.

Let us refer to Fig. 6.5 and recall the construction of Sec. 1.8. We draw a circle of radius a and origin at the center of the rectilinear ellipse, namely O_r, and two segments, say P_1Q_1 and P_2Q_2 perpendicular to the chord P_1P_2 and going, respectively, from P_1 and P_2 to the circle. The angles α and β are the angles between the chord and the segments going from O_r to Q_1 and Q_2, respectively.
Since $P_2F_r = r_2 = s$ and $P_2F_r + P_2F_r^* = 2a$, we have $P_2F_r^* = 2a - s$. Moreover, since the radius of the circle is a, we have $P_2O_r = a - 2a + s = s - a$. If we construct the circle centered in O_r starting from P_2 without using F_r and F_r^*, then

- if $0 \leq P_2O_r \leq c$, then O_r belongs to the chord;
- if $P_2O_r < 0$ then O_r lies exteriorly to P_2;
- if $P_2O_r > c$, then O_r is located exteriorly P_1.

Let us consider the following cases.

1. If $a = a_m$, then P_2 coincides with F_r^* (see Eq. (6.15)) and, as already noticed, $\alpha = \pi$ (see Eq. 6.16).
2. If $P_2O_r = \frac{c}{2}$, that is, the center of the rectilinear ellipse is the midpoint of the chord, then

$$s - a = r_1 + c - a = \frac{c}{2}, \qquad \Longrightarrow \qquad a = r_1 + \frac{c}{2} = \frac{2r_1 + r_2 - r_1}{2} = \frac{r_1 + r_2}{2},$$

which corresponds the semi-major axis associated with the fundamental ellipse (see (6.20)).

To conclude, since the known quantities of the problem are r_1, r_2, c, $\Delta\theta$ and Δt, case 1. tells us that the quadrant of the angle α is determined according to whether Δt is larger than Δt_m, while case 2. suggests that the quadrant associated with β has to be chosen according to the position of the vacant focus, which determines the eccentricity of the orbit as seen before. The difference between the case of the vacant focus along the lower branch of the hyperbolic locus and the one along the upper branch is in particular dictated by the angle $\Delta\theta$. The interpretation just given derives from Prussing (1979).

6.4 How to Solve the Lambert's Problem

Now, let us now assume that at the epoch t_1 a given spacecraft orbiting at the Earth is situated at P_1 and that it must transfer to P_2 in Δt units of time. For what explained in Secs. 6.2 and 6.3, if the orbit connecting the two given points in the given time of flight is an ellipse, then the angles α and β must follow the rule:

$$\begin{aligned}
\beta &= \begin{cases} \beta_0, & \text{if} \quad 0 \leq \Delta\theta < \pi, \\ -\beta_0, & \text{if} \quad \pi \leq \Delta\theta < 2\pi. \end{cases} \\
\alpha &= \begin{cases} \alpha_0, & \text{if} \quad \Delta t \leq \Delta t_m, \\ 2\pi - \alpha_0, & \text{if} \quad \Delta t > \Delta t_m, \end{cases}
\end{aligned}$$

where α_0 and β_0 are the principle values of the inverse sine functions used to solve (6.10). The corresponding cases are showed in Fig. 6.6. Actually, it can be proved that the condition on β holds also in the parabolic and hyperbolic cases (Prussing, 1979; Battin, 1999).

But still, we do not know what kind of conic section fulfills our constraints. To this end, we must solve Eq. (6.9) in terms of the semi-major axis a. This can be done only by means of iterative methods, e.g. the Newton's method introduced in Sec. 1.8.3, because the equation contains transcendental terms. Let us consider the parabolic case, letting a tend to infinity in (6.9) and (6.10) and assuming $0 \leq \Delta\theta < \pi$, namely,

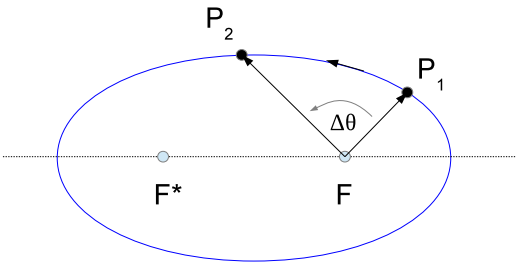
$$\lim_{a\to\infty} \sin\frac{\alpha}{2} = \lim_{a\to\infty} \sqrt{\frac{s}{2a}} = 0 \qquad \Longrightarrow \qquad \frac{\alpha}{2} \approx \sin\frac{\alpha}{2} = \sqrt{\frac{s}{2a}} \qquad \Longrightarrow \qquad \alpha \approx 2\sqrt{\frac{s}{2a}},$$

$$\lim_{a\to\infty} \sin\frac{\beta}{2} = \lim_{a\to\infty} \sqrt{\frac{s-c}{2a}} = 0 \qquad \Longrightarrow \qquad \frac{\beta}{2} \approx \sin\frac{\beta}{2} = \sqrt{\frac{s-c}{2a}} \quad \Longrightarrow \quad \beta \approx 2\sqrt{\frac{s-c}{2a}}.$$

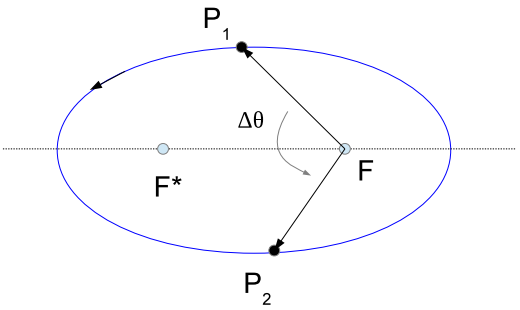
Also,

$$\alpha - \sin\alpha \approx \alpha - \left(\alpha - \frac{\alpha^3}{3!}\right) = \frac{\alpha^3}{6} \approx \frac{4s^{3/2}}{3(2a)^{3/2}},$$

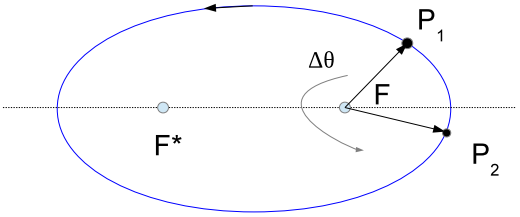
$$\beta - \sin\beta \approx \frac{\beta^3}{6} \approx \frac{4(s-c)^{3/2}}{3(2a)^{3/2}},$$



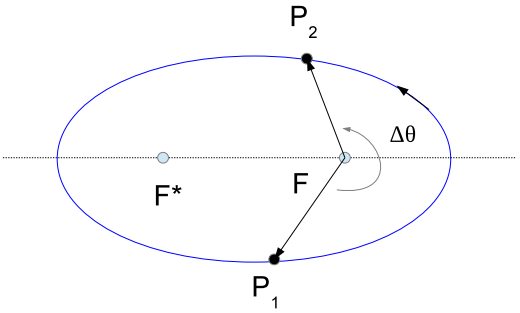
$\Delta t < \Delta t_m$, $\Delta\theta < \pi$
$\alpha = \alpha_0$, $\beta = \beta_0$



$\Delta t > \Delta t_m$, $\Delta\theta < \pi$
$\alpha = 2\pi - \alpha_0$, $\beta = \beta_0$



$\Delta t > \Delta t_m$, $\Delta\theta > \pi$
$\alpha = 2\pi - \alpha_0$, $\beta = -\beta_0$



$\Delta t < \Delta t_m$, $\Delta\theta > \pi$
$\alpha = \alpha_0$, $\beta = -\beta_0$

Figure 6.6: The different cases that can occur in the Lambert's problem if the connecting orbit is an ellipse.

In this way the transcendental functions cancel out and we obtain

$$\Delta t_p = \frac{\sqrt{2}}{3\sqrt{\mu}}\Big(s^{3/2} - (s-c)^{3/2}\Big). \tag{6.23}$$

If, instead, $\pi \leq \Delta\theta < 2\pi$, then $\beta < 0$ and

$$\Delta t_p = \frac{\sqrt{2}}{3\sqrt{\mu}}\Big(s^{3/2} + (s-c)^{3/2}\Big). \tag{6.24}$$

The two equations (6.23) and (6.24) can be summarized in

$$\begin{aligned} \Delta t_p &= \frac{\sqrt{2}}{3\sqrt{\mu}}\Big(s^{3/2} - \mathrm{sign}(\Delta\theta)(s-c)^{3/2}\Big) = \\ &= \frac{1}{6\sqrt{\mu}}\left[(r_1+r_2+c)^{3/2} - \mathrm{sign}(\Delta\theta)\,(r_1+r_2-c)^{3/2}\right], \end{aligned} \tag{6.25}$$

which tells us that in the parabolic case, the time of flight does not depend on the semi-major axis. It can be proved (Battin, 1999) that Eq. (6.25) corresponds to the Barker's equation (1.59) introduced in Sec. 1.7.3.

One of the possible procedures to solve the Lambert's equation is thus the following:

1. the quadrant of β is determined by $\Delta\theta$;
2. compute Δt_p;
3. if $\Delta t < \Delta t_p$, then the conic section we are looking for is a hyperbola;
4. if $\Delta t_p < \Delta t < \Delta t_m$, then the arc connecting P_1 to P_2 is short-way elliptic, then $\alpha = \alpha_0$;
5. if $\Delta t > \Delta t_m$, then the arc is long-way elliptic, then $\alpha = 2\pi - \alpha_0$;
6. with these input data find the zero of either

$$g(a) = \sqrt{a^3}(\alpha - \beta - \sin\alpha + \sin\beta) - \sqrt{\mu}\Delta t,$$

if the orbit is an ellipse, or

$$g_h(a) = \sqrt{-a^3}(-\alpha + \beta + \sinh\alpha - \sinh\beta) - \sqrt{\mu}\Delta t,$$

if the orbit is a hyperbola. In the latter case, $\alpha = \alpha_0$ always.

In the case of elliptic transfers, once computed a, the eccentricity e of the conic section can be recovered from (1.117) and (1.119). In particular,

$$1 - \frac{r_2}{p}(1 - \cos\Delta\theta) = 1 - \frac{a}{r_1}(1 - \cos\Delta E),$$

where $\Delta E \equiv 2E_{-} = \alpha - \beta$, while $\Delta\theta$, r_1 and r_2 are known. If follows

$$p = \frac{r_1 r_2}{a} \frac{1 - \cos\Delta\theta}{1 - \cos\Delta E},$$

and then

$$e = \sqrt{1 - \frac{p}{a}}.$$

At this point, it should be clear that the geometrical derivation of Gauss in terms of the area of the triangle and the sector defined by r_1 and r_2 (see Secs. **??** and **??**) was a first attempt to find a solution to the Lambert's problem.

Bibliography

Bate, R. R., Mueller, D. D., White, J. E., 'Fundamental of Astrodynamics', Dover Publications, Inc., New York, 1971.

Battin, R. H., 'An Introduction to the Mathematics and Methods of Astrodynamics', Revised Edition, AIAA Education Series, American Institute of Aeronautics and Astronautics, Inc., Reston, 1999.

Chao, C., 'Applied Orbit Perturbation and Maintenance', American Institute of Aeronautics and Astronautics/Aerospace Press, Reston, Virginia/El Segundo, California, 2005.

Chobotov, V. A., 'Orbital Mechanics', Third Edition, AIAA Education Series, American Institute of Aeronautics and Astronautics, Inc., Reston, 2002.

Gómez, G., Llibre, J., Martínez, R., Simó, C., 'Dynamics and Mission Design Near Libration Point Orbits – Volume 1: Fundamentals: The Case of Collinear Libration Points', World Scientific, Singapore, 2000a.

Gómez, G., Jorba, À., Masdemont, J., Simó, C., 'Dynamics and Mission Design Near Libration Point Orbits – Volume 3: Advanced Methods for Collinear Points', World Scientific, Singapore, 2000b.

Mengali, G., Quarta, A. A., 'Fondamenti di Meccanica del Volo Spaziale', Pisa University Press, Pisa, 2006.

Milani, A., Gronchi, G. F., 'Theory of Orbit Determination', Cambridge University Press, 2010.

Montebruck, O., Gill, E., 'Satellite Orbits. Model, Methods, and Applications', Springer, New York, 2001.

Prussing, J. E., 'Geometrical Interpretation of the Angles α and β in Lambert's Problem', J. Guidance and Control 2(5), 442-443, 1979.

Roy, A. E., 'Orbital Motion', Institute of Physics Publishing, Bristol, 2005.

Szebehely, V., 'Theory of orbits', Fourth Edition, Academic Press, New York, 1967.

Thornton, C. L., Border, J. S., 'Radiometric Tracking Techniques for Deep Space Navigation', First Edition, 2003.

Valk, S., Lemaître, A., Anselmo, L., 'Analytical and semi-analytical investigations of geosynchronous space debris with high area-to-mass ratios', Adv. Space Res. 41, 1077-1090, 2008.

Vallado, D. A., 'Fundamentals of Astrodynamics and Applications', Third Edition, Springer, New York, 2007.