Space Vehicles and Orbital Dynamics

Lecture Notes

Version 2020.2

Aerospace Engineering Department (UC3M)

Foreword

These lecture notes are only a summary of derivations, formulas, and brief descriptions for the course *Space Vehicles and Orbital Dynamics*. They are currently a work in progress, and as such they may include multiple typos, important omissions, and errors. We require all your help and feedback to improve them and make them more useful for future students. Please contact us if you see any areas of improvement: mario.merino@uc3m.es.

The notes do not represent the full contents of the course, and they must be complemented with proper books. Please refer to the bibliography of the course (in the course information webpage) to consult the recommended texts. In particular, the books [Curtis, 2013], [Junkins and Schaub, 2009], [Tewari, 2007] and [Vallado, 2001] are essential references.

Links to interesting topics outside of the scope of the course are given in footnotes marked with the keyword "*Rabbithole*." It is fun to follow one of this links, but one can easily lose track of time while researching these topics. Beware!

Above all, enjoy while you learn the beautiful subject of orbital dynamics and space vehicles.

Mario Merino,
Leganés, January 2020

Meaning of mathematical symbols

The same nomenclature as in the second year course "Applied Mechanics to Aerospace Engineering," is used in this course. All scalars are written in normal typeface, e.g. x, μ. Vectors, on the contrary, will always be presented in bold face: $\boldsymbol{v}$, $\boldsymbol{F}$. Tensors will be denoted with a double bar, like $\overline{\overline{\mathcal{U}}}$. The following standard nomenclature is used throughout the notes and problems:

$B_0 : \{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$: A vector basis, defined by the ordered set of three vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$.

$S_0 : \{O; B_0\}$: A reference frame, with origin at point O and vector basis B_0. By abuse of notation, a reference frame is often also denoted by $Oxyz$, where O is the origin, Ox Oy, Oz are the coordinate axis in the directions of the three vectors of the basis B_0, and Oxy, Oxz, Oyz are the coordinated planes.

$C : \{q_1, q_2, q_3\}$: A system of (generalized) coordinates.

$\dot{x}, \ddot{x}$: Dots over a scalar variable denote differentiation in time: $\dot{x} = \mathrm{d}x/\mathrm{d}t$, $\ddot{x} = \mathrm{d}^2x/\mathrm{d}t^2$.

$\mathrm{d}\boldsymbol{A}/\mathrm{d}t|_0$: Time derivative of some vector $\boldsymbol{A}$, taken in the reference frame S_0

∇ : Nabla operator. The vector gradient of a scalar function f is denoted ∇f. The divergence and curl (or rotor) of a vector $\boldsymbol{A}$ are denoted $\nabla \cdot \boldsymbol{A}$ and $\nabla \times \boldsymbol{A}$, respectively.

$\boldsymbol{r}_0^P$: Position vector of a point P as seen from reference frame S_0: $\boldsymbol{r}_0^P = \boldsymbol{OP}$. Dimensions: $[L]$.

$\boldsymbol{v}_0^P$: Velocity vector of a point P as seen from reference frame S_0: $\boldsymbol{v}_0^P = \mathrm{d}\boldsymbol{r}_0^P/\mathrm{d}t|_0$. Dimensions: $[L/T]$.

$\boldsymbol{a}_0^P$: Acceleration vector of a point P as seen from reference frame S_0: $\boldsymbol{v}_0^P = \mathrm{d}\boldsymbol{v}_0^P/\mathrm{d}t|_0$. Dimensions: $[L/T^2]$.

$\boldsymbol{\omega}_{10}$: Angular velocity vector of reference frame or rigid body 1 with respect to reference frame or rigid body 0. Dimensions: $[1/T]$.

$\boldsymbol{\alpha}_{10}$: Angular acceleration vector of reference frame or rigid body 1 with respect to reference frame or rigid body 0. Dimensions: $[1/T^2]$.

$\boldsymbol{p}_0$: Linear momentum as seen from reference frame S_0. Superindices like P or b are used to indicate the particle or body to which the linear momentum refers to. Dimensions: $[ML/T]$.

$\boldsymbol{H}_{A,0}$: Angular momentum about a point A as seen from reference frame S_0. Superindices like P or b are used to indicate the particle or body to which the angular momentum refers to. Dimensions: $[ML^2/T]$.

$\boldsymbol{h}_{A,0}$: Specific angular momentum per unit masss about a point A as seen from reference frame S_0. Superindices like P or b are used to indicate the particle or body to which the angular momentum refers to. Dimensions: $[L^2/T]$.

$\boldsymbol{F}$: Force vector. Subindices are used to label different forces. If necessary, a superindex can be used to denote the point of application of the force. Dimensions: $[ML/T^2]$.

$\boldsymbol{M}_A$: Torque or moment of force about a point A. Additional subindices are used to label the moment of different forces. Dimensions: $[ML^2/T^2]$.

$W_0, \dot{W}_0$: Work and power exerted by a force or torque as seen from reference frame S_0. In the case of work, note that the initial and final states of the motion, and in general the trajectory followed between them, must be specified. Additional subindices are used to label the work and power of different forces or torques. Dimensions: $[ML^2/T^2]$, $[ML^2/T^3]$.

V_0 : Potential energy associated to a conservative force as seen from reference frame S_0. Subindices are also used to label different forces as required. Dimensions: $[ML^2/T^2]$.

T_0, E_0 : Kinetic and mechanical energy as seen from reference frame S_0. Superindices like P or b are used to indicate the particle or body to which the variable refers to. Dimensions: $[ML^2/T^2]$.

ξ_0 : Specific mechanical energy per unit mass as seen from reference frame S_0. Superindices like P or b are used to indicate the particle or body to which the variable refers to. Dimensions: $[L^2/T^2]$.

$\mathcal{D}(\boldsymbol{r}_0^P)$: Mass distribution function of a material system. It can be discrete, i.e., composed of point particles, or continuous, i.e., given as a linear, area, or volume density.

$\mathcal{M}^{\mathcal{D}}_{O,n}$: n-th moment of a mass distribution function $\mathcal{D}(\boldsymbol{r}_0^P)$ about a point O. This is a n-th order tensor, i.e., $\mathcal{M}^{\mathcal{D}}_{O,0}$ is a scalar, $\mathcal{M}^{\mathcal{D}}_{O,1}$ is a vector, and $\mathcal{M}^{\mathcal{D}}_{O,2}$ is a second-order tensor. Dimensions: $[ML^n]$.

$m^{\mathcal{D}}$: Total mass of a mass distribution function $\mathcal{D}(\boldsymbol{r}_0^P)$. Dimensions: $[M]$.

$I^{\mathcal{D}}_O, I^{\mathcal{D}}_{Ox}, I^{\mathcal{D}}_{Oxy}$: Moment of inertia of a mass distribution function $\mathcal{D}(\boldsymbol{r}_0^P)$ with respect to a point O, a line Ox, and a plane Oxy, respectively. Dimensions: $[ML^2]$.

$P^{\mathcal{D}}_{Oxy,Oxz}$: Product of inertia of a mass distribution function $\mathcal{D}(\boldsymbol{r}_0^P)$ with respect to the orthogonal planes Oxy and Oxz. Dimensions: $[ML^2]$.

$\overline{\overline{\mathcal{I}}}^{\mathcal{D}}_O$: Tensor of inertia of a mass distribution function $\mathcal{D}(\boldsymbol{r}_0^P)$ about a point O Dimensions: $[ML^2]$.

$\overline{\overline{\mathcal{U}}}$: Unit diagonal tensor.

When it is necessary to explicitly state that a point A belongs to a rigid body or reference frame b (i.e., to indicate that the points moves jointly with the body or reference frame, so that the relative velocity of the point in a body-fixed reference frame is zero), this will be indicated by stating the rigid body or reference frame in parenthesis, e.g. $A(b)$.

Subindices are used with matrix or component notation of vectors and tensors to denote the basis used. For example, a vector $\boldsymbol{A}$ can be expressed in a basis $B_0 : \{\boldsymbol{i}_0, \boldsymbol{j}_0, \boldsymbol{k}_0\}$ as:

$$\boldsymbol{A} = x\boldsymbol{i}_0 + y\boldsymbol{j}_0 + z\boldsymbol{k}_0 = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_0,$$

and tensor $\boldsymbol{A} \otimes \boldsymbol{A}$ as:

$$\begin{aligned} \boldsymbol{A} \otimes \boldsymbol{A} = {} & xx\boldsymbol{i}_0 \otimes \boldsymbol{i}_0 + xy\boldsymbol{i}_0 \otimes \boldsymbol{j}_0 + xz\boldsymbol{i}_0 \otimes \boldsymbol{k}_0 \\ & + yx\boldsymbol{j}_0 \otimes \boldsymbol{i}_0 + yy\boldsymbol{j}_0 \otimes \boldsymbol{j}_0 + yz\boldsymbol{j}_0 \otimes \boldsymbol{k}_0 \\ & + zx\boldsymbol{k}_0 \otimes \boldsymbol{i}_0 + zy\boldsymbol{k}_0 \otimes \boldsymbol{j}_0 + zz\boldsymbol{k}_0 \otimes \boldsymbol{k}_0 \\ = {} & \begin{bmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{bmatrix}_{00}. \end{aligned}$$

Contents

1 Two-body problem

The two-body problem is an essential problem in orbital mechanics. It is not only an academic exercise: most computations and definitions used in modern astrodynamics and mission analysis are based on the two-body problem in one way or another, or on variants of it.

Problem statement: Analyze the motion of two bodies due solely to their mutual gravitational attraction.

Assumptions: The two particles are modeled as point particles and assumed to be alone in the universe. Other effects (the two bodies not being point particles, the presence of other bodies, etc) will be added to the problem as *perturbations* in a later chapter.

1.1 Mathematical model

Consider an inertial frame $S_0 : \{O; B_0\}$, where O is an origin point and $B_0 : \{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ is a right-handed orthonormal vector basis. We call the two point particles P_1 and P_2. Their position vectors are:

$$\boldsymbol{r}_0^{P_1} = x_1\boldsymbol{i} + y_1\boldsymbol{j} + z_1\boldsymbol{k}, \tag{1.1}$$

$$\boldsymbol{r}_0^{P_2} = x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}. \tag{1.2}$$

We define the difference of these position vector $\boldsymbol{r}$ (no indices) and the corresponding unit vector $\boldsymbol{u}_r$ as:

$$\boldsymbol{r} = \boldsymbol{r}_0^{P_2} - \boldsymbol{r}_0^{P_1}, \tag{1.3}$$

$$\boldsymbol{u}_r = \boldsymbol{r}/r. \tag{1.4}$$

We will also use:

$$\boldsymbol{v} = \left.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right|_0; \quad \boldsymbol{a} = \left.\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right|_0. \tag{1.5}$$

The gravitational pull that P_2 exerts on P_1 is:

$$\boldsymbol{F}_{2,1} = G\frac{m^{P_1}m^{P_2}}{r^2}\boldsymbol{u}_r. \tag{1.6}$$

By the third law of Newton, the gravitational pull of P_1 on P_2 is $\boldsymbol{F}_{1,2} = -\boldsymbol{F}_{2,1}$.

The equations of motion of the particle are:

$$\boxed{m^{P_1}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_1}}{\mathrm{d}t^2}\right|_0 = +G\frac{m^{P_1}m^{P_2}}{r^2}\boldsymbol{u}_r} \tag{1.7}$$

$$\boxed{m^{P_2}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_2}}{\mathrm{d}t^2}\right|_0 = -G\frac{m^{P_1}m^{P_2}}{r^2}\boldsymbol{u}_r} \tag{1.8}$$

1.1.1 Motion of the center of mass

We define the position of the center of mass G of the system as

$$\boxed{\boldsymbol{r}_0^G = \frac{m^{P_1}\boldsymbol{r}_0^{P_1} + m^{P_2}\boldsymbol{r}_0^{P_2}}{m},} \tag{1.9}$$

where $m = m^{P_1} + m^{P_2}$ is the total mass of the system. Adding equations (1.7) and (1.8) together we see that

$$m^{P_1}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_1}}{\mathrm{d}t^2}\right|_0 + m^{P_2}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_2}}{\mathrm{d}t^2}\right|_0 = m\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^G}{\mathrm{d}t^2}\right|_0 = \boldsymbol{0}, \tag{1.10}$$

i.e., G is not accelerated. Thus, the reference frame $S_G : \{G; B_0\}$ is also an inertial reference frame.

1.1.2 Two-body-problem equation

Simplifying the masses in equations (1.7) and (1.8) and then subtracting yields:

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_2}}{\mathrm{d}t^2}\right|_0 - \left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_1}}{\mathrm{d}t^2}\right|_0 = \left.\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right|_0 = -G\frac{m^{P_1}}{r^2}\boldsymbol{u}_r - G\frac{m^{P_2}}{r^2}\boldsymbol{u}_r. \tag{1.11}$$

This can be interpreted as the equation of motion of P_2 (where the mass of P_2 has been simplified), as seen from the non-rotating but non-inertial reference frame $S_1 : \{P_1; B_0\}$. Note that the second term is actually the inertial force in this reference frame, due to the acceleration of the origin. Dropping the subindex 0 that denotes differentiation in the only defined basis B_0 we may write:

$$\boxed{\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} = -\frac{\mu}{r^3}\boldsymbol{r},} \tag{1.12}$$

where $\mu = Gm$ is called the *gravitational parameter* of the system. When if $m^{P_1} \gg m^{P_2}$, one typically uses $\mu = Gm^{P_1}$ as an approximation. In this case, μ is usually written with the subindex of the primary body: Sun ☉, Mercury ☿, Venus ♀, Earth ⊕, Moon ☾, Mars ♂, Jupiter ♃, Saturn ♄, Uranus ♅, Neptune ♆, etc. Observe that this approximation is equivalent to dropping the small non-inertial term in equation (1.11).

1.1.3 Conservation of specific mechanical energy

The force in equation (1.12) is conservative, thus it defines a *conservative problem*. We define the potential energy per unit mass associated to it as $-\mu/r$, so that $-\mu\boldsymbol{r}/r^3 = -\nabla(-\mu/r)$. Observe that with this definition, the potential energy goes to zero far away from P_1. We define the mass-specific mechanical energy of P_2 in S_1 as

$$\boxed{\xi = \frac{v^2}{2} - \frac{\mu}{r}.} \tag{1.13}$$

Dot-multiplying each side of equation (1.12) by $\boldsymbol{v}$,

$$\boldsymbol{v}\cdot\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v^2}{2}\right),$$
$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\cdot\left(-\frac{\mu}{r^3}\boldsymbol{r}\right) = -\frac{\mu}{r^2}\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mu}{r}\right).$$

Thus we may write

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v^2}{2} - \frac{\mu}{r}\right) = \frac{\mathrm{d}\xi}{\mathrm{d}t} = 0 \Rightarrow \xi = \text{const}, \tag{1.14}$$

i.e., ξ is conserved, as expected in a conservative problem. Observe that with the definition of the potential energy used, the specific mechanical energy ξ can be negative.

1.1.4 Conservation of specific angular momentum

Note that the equation of motion (1.12) defines a *central force problem*. We define the mass-specific angular momentum vector of P_2 about P_1 in S_1 as

$$\boxed{\boldsymbol{h} = \boldsymbol{r}\times\boldsymbol{v}.} \tag{1.15}$$

Cross-multiplying each side of equation (1.12) by $\boldsymbol{r}$,

$$\boldsymbol{r}\times\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}\times\boldsymbol{v}) = \frac{\mathrm{d}\boldsymbol{h}}{\mathrm{d}t},$$
$$\boldsymbol{r}\times\left(-\frac{\mu}{r^3}\boldsymbol{r}\right) = \boldsymbol{0}.$$

Thus we may write

$$\frac{\mathrm{d}\boldsymbol{h}}{\mathrm{d}t} = \mathbf{0} \Rightarrow \boldsymbol{h} = \mathbf{const}, \tag{1.16}$$

i.e., $\boldsymbol{h}$ is conserved, as expected for a central force problem. Since $\boldsymbol{r}$ and $\boldsymbol{v}$ are always perpendicular to $\boldsymbol{h}$, we conclude that the motion is *planar*. Taking polar coordinates in this plane we can write the magnitude of $\boldsymbol{h}$ as:

$$\boxed{h = rv_\theta = r^2\dot{\theta}.} \tag{1.17}$$

1.1.5 Eccentricity vector

We define the eccentricity vector as

$$\boxed{\boldsymbol{e} = \frac{\boldsymbol{v}\times\boldsymbol{h}}{\mu} - \frac{\boldsymbol{r}}{r} = \left(v^2 - \frac{\mu}{r}\right)\frac{\boldsymbol{r}}{\mu} - \frac{\boldsymbol{r}\cdot\boldsymbol{v}}{\mu}\boldsymbol{v}.} \tag{1.18}$$

Cross-multiplying each side of equation (1.12) by $\boldsymbol{h}$,

$$\begin{aligned}
\boldsymbol{h}\times\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(-\boldsymbol{v}\times\boldsymbol{h}\right), \\
\boldsymbol{h}\times\left(-\frac{\mu}{r^3}\boldsymbol{r}\right) = \frac{\mu}{r^3}\boldsymbol{r}\times(\boldsymbol{r}\times\boldsymbol{v}) &= \frac{\mu}{r^3}\left[\left(\boldsymbol{r}\cdot\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right)\boldsymbol{r} - r^2\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right] = \mu\left(\frac{\boldsymbol{r}}{r^2}\frac{\mathrm{d}r}{\mathrm{d}t} - \frac{1}{r}\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(-\mu\frac{\boldsymbol{r}}{r}\right).
\end{aligned}$$

Thus we may write

$$\frac{\mathrm{d}\boldsymbol{e}}{\mathrm{d}t} = \mathbf{0} \Rightarrow \boldsymbol{e} = \mathbf{const}, \tag{1.19}$$

i.e., $\boldsymbol{e}$ is conserved. This vector always points to the periapsis of the orbit, and its magnitude e is the eccentricity of the orbit.

1.2 Perifocal reference frame

So far, we had not defined the directions of the vectors of B_0. Since the motion has two constant vectors $\boldsymbol{h}$ and $\boldsymbol{e}$, without any loss of validity on the former, we choose $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ as follows:

$$\boldsymbol{i} = \boldsymbol{e}/e, \tag{1.20}$$
$$\boldsymbol{j} = \boldsymbol{k}\times\boldsymbol{i}, \tag{1.21}$$
$$\boldsymbol{k} = \boldsymbol{h}/h. \tag{1.22}$$

When $\boldsymbol{e} = \mathbf{0}$ (circular orbits), any direction in the orbital plane is chosen for vector $\boldsymbol{i}$. With this definition of B_0, we call the reference frame $S_1 : \{P_1; B_0\}$ the *perifocal reference frame*.

We will denote the standard Cartesian coordinates in this reference frame as (x, y, z). Note that the motion takes place in the plane P_1xy. The perifocal reference frame is sometimes denoted by the letters PQW, and the unit vectors along the coordinated axes called $\boldsymbol{p}$, $\boldsymbol{q}$ and $\boldsymbol{w}$.

1.3 Trajectory equation

Dot-multiplying definition (1.18) by $\boldsymbol{r}$, and introducing the *true anomaly* θ as the polar angle between $\boldsymbol{e}$ and $\boldsymbol{r}$:

$$er\cos\theta = \frac{\boldsymbol{r}\cdot(\boldsymbol{v}\times\boldsymbol{h})}{\mu} - r = \frac{\boldsymbol{h}\cdot\boldsymbol{h}}{\mu} - r = \frac{h^2}{\mu} - r. \tag{1.23}$$

Solving for r,

$$\boxed{r = \frac{h^2/\mu}{1 + e\cos\theta}} \tag{1.24}$$

This equation gives the trajectory $r = r(\theta)$ of P_2 in S_1 in polar coordinates (r, θ) in the perifocal reference frame. Equation (1.24) is the expression of a *conic section* of orbit parameter $p = h^2/\mu$ and eccentricity e, with one of the foci at P_1, and the closest vertex in the direction of P_1x. There are three types of non-degenerate conic sections: ellipses, parabolas, and hyperbolas. Two-body problem orbits are also called *Keplerian orbits*.

The trajectory equation can be alternatively obtained by writing the $\boldsymbol{u}_r$-component of equation (1.12) in arbitrary polar coordinates (r, θ) in the plane of motion,

$$\ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2}. \tag{1.25}$$

Noting that $h = r^2\dot{\theta}$ and using the transformation

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \dot{\theta}\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{h}{r^2}\frac{\mathrm{d}r}{\mathrm{d}\theta} = -h\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1}{r}\right), \tag{1.26}$$

$$\frac{\mathrm{d}^2r}{\mathrm{d}^2t} = -\frac{h^2}{r^2}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\frac{1}{r}\right), \tag{1.27}$$

we can write:

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\frac{1}{r}\right) + \left(\frac{1}{r}\right) = \frac{\mu}{h^2} \tag{1.28}$$

This is the equation of a *forced harmonic oscillator* where the dependent variable is $(1/r)$ and the independent variable is θ, whose complete solution reads

$$\frac{1}{r} = \frac{\mu}{h^2}\left[A\cos(\theta + \theta_0) + 1\right], \tag{1.29}$$

with A and θ_0 constants of integration, which resolve to e and 0 respectively after identifying the trajectory with a conic section of eccentricity e with one of the foci at $r = 0$, and defining the origin of θ in the direction of the closest vertex (pericenter).

The parametric equations of the trajectory in Cartesian coordinates in the perifocal reference frame, noting that $x = r\cos\theta$ and $y = r\sin\theta$, read simply:

$$x = \frac{h^2\cos\theta/\mu}{1 + e\cos\theta}; \qquad y = \frac{h^2\sin\theta/\mu}{1 + e\cos\theta}. \tag{1.30}$$

1.4 Lagrange coefficients

The equation of motion (1.12) is a second order differential equation on $\boldsymbol{r}$. Given the initial position $\boldsymbol{r}_0$ and velocity $\boldsymbol{v}_0$, the trajectory is fully determined. Since motion is planar, $\boldsymbol{r}$ and $\boldsymbol{v}$ at any later time will be a linear combination of $\boldsymbol{r}_0$ and $\boldsymbol{v}_0$, and we can write

$$\boldsymbol{r}(t) = f(t)\boldsymbol{r}_0 + g(t)\boldsymbol{v}_0, \tag{1.31}$$

$$\boldsymbol{v}(t) = \dot{f}(t)\boldsymbol{r}_0 + \dot{g}(t)\boldsymbol{v}_0. \tag{1.32}$$

The unknown functions $f(t)$ and $g(t)$ are termed *Lagrange coefficients*. Taking $t = 0$ it is clear that these functions satisfy:

$$f(0) = 1; \quad \dot{f}(0) = 0; \quad g(0) = 0; \quad \dot{g}(0) = 1; \tag{1.33}$$

The conservation of angular momentum implies an identity of these functions:

$$f(t)\dot{g}(t) - g(t)\dot{f}(t) = 1 \tag{1.34}$$

Expressions of $f(t)$ and $g(t)$ in terms of θ can easily be found using the results already obtained for the two-body problem. These functions are useful in a number of applications, in particular the formulation and solution of *Lambert's problem*.

The functions $f(t)$ and $g(t)$ can be expanded into an infinite series of powers in time, and used to solve the time law $r = r(t)$, $\theta = \theta(t)$ in small time intervals about the initial condition. However, this series converges very slowly. Thus, this approach is seldom used, and Kepler's equation (given further below) is used instead.

1.5 Conic sections

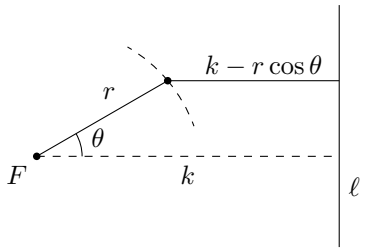
Conic sections are curves that can be defined in multiple forms and appear in many fields of physics and mathematics:

1. As the intersection between a double-cone and a plane.

2. Given a point F (focus) and a line ℓ (directrix), call the distance between them k (focal parameter). A conic section is the locus of points for which the ratio e of the distance $k - r\cos\theta$ to ℓ and the radius r from F is constant:

$$e = \frac{k - r\cos\theta}{r} \tag{1.35}$$

(in the hyperbolic case, r is allowed to be negative in the unoccupied branch).



Calling $p = ek$ (semi-latus rectum), we find the polar expression of a conic section as seen from one of its foci, where e is the eccentricity:

$$r = \frac{p}{1 + e\cos\theta}. \tag{1.36}$$

Comparing with equation (1.24), $p = h^2/\mu$.

3. Any quadratic form in two variables x, y gives a conic section in Cartesian coordinates:

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0. \tag{1.37}$$

A quadratic form can also be represented in matrix form as[1]:

$$\left[\begin{array}{cc|c} x & y & 1 \end{array}\right] \cdot \left[\begin{array}{cc|c} A & B & D \\ B & C & E \\ \hline D & E & F \end{array}\right] \cdot \left[\begin{array}{c} x \\ y \\ \hline 1 \end{array}\right] = 0. \tag{1.39}$$

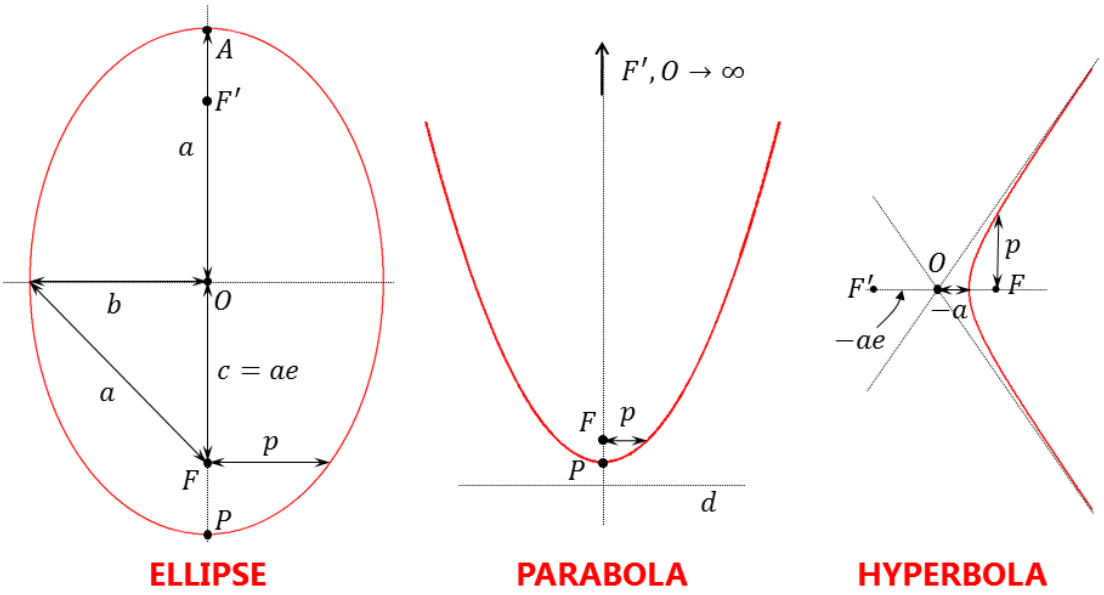
The matrix of non-degenerate conics have a non-zero determinant.

Non-degenerate conics are classified into ellipses, parabolas, and hyperbolas. The ellipse is the only closed non-degenerate conic section, and the circle is a particular case of ellipse. An example of the three non-degenerate conic sections can be observed below:

[1] *Rabbithole*: As a variant of the matrix notation, using homogeneous coordinates X, Y, W it is possible to study the intersection of a conic section with infinity (e.g. the asymptotes of an hyperbola) as regular points, so that all conic types can be shown to be the same projective object, with zero (ellipse), one (parabola) or two (hyperbola) intersections with the infinity line:

$$\left[\begin{array}{ccc} X & Y & W \end{array}\right] \cdot \left[\begin{array}{ccc} A & B & D \\ B & C & E \\ D & E & F \end{array}\right] \cdot \left[\begin{array}{c} X \\ Y \\ W \end{array}\right] = 0; \quad \text{with: } x = \frac{X}{W}, y = \frac{Y}{W}. \tag{1.38}$$

In this formulation, the intersections with the infinity line are those coordinate sets for which $W = 0$.



Conic sections have a center, two foci, two vertices or apsides, and two directrices. There is a major axis and a minor axis that intersect perpendicularly at the center of the conic section. The vertices of the conic section are its intersections with the major axis. Several parameters are defined for conic sections:

Semi-major axis a**:** half of the major axis. It is negative for hyperbolas.

Semi-minor axis b**:** half of the minor axis.

Semi-interfocal distance c**:** half distance between the two foci

Semi-latus rectum p**:** also known as orbit parameter, it is the distance from one of the foci to the conic in the direction perpendicular to the major axis. It is finite for all non-degenerate conic sections.

Eccentricity e**:** parameter that defines the shape and type of the conic section, defined as $e = c/a$.

Only two parameters are needed to define the size and shape of the conic section, e.g. a and e. For non-degenerate conics, the type of conic section is reflected on the value of its parameters:

	a	e
(circle)	$a > 0$	$e = 0$
ellipse	$a > 0$	$0 \leqslant e < 1$
parabola	$a = \infty$	$e = 1$
hyperbola	$a < 0$	$e > 1$

The two vertices of the conic section are denoted periapsis (or pericenter) and the apoapsis (or apocenter), i.e., the points of the conic section nearest and farthest away from the occupied focus P_1:

$$\boxed{r_p = a(1-e); \quad r_a = a(1+e); \quad 2a = r_p + r_a; \quad e = \frac{r_a - r_p}{r_a + r_p}.} \tag{1.40}$$

In the case of parabolic orbits, the expression above for the position of the periapsis is indeterminate. Introducing the definition of the orbit parameter is easy to find $r_p = p/2$ in this case.

The tangent to the conic section at the periapsis and apoapsis is perpendicular to the major axis. For this reason, $\boldsymbol{r}$ and $\boldsymbol{v}$ are perpendicular to each other at those points.

From equation (1.36) its is evident that when $\theta = \pi/2$ the value of the radius is $r = p$.

Important relations of conic section parameters:

$$\boxed{a^2 = b^2 + c^2; \quad b = a\sqrt{1-e^2}; \quad c = ea \quad p = a(1-e^2) = b^2/a.}$$

1.6 Trajectory types

As seen above, three types of orbits exist in the two-body problem, which are non-degenerate conic sections: *ellipses*, *parabolas*, and *hyperbolas*.

1.6.1 Elliptic orbits

Elliptic trajectories are the only closed orbits in the two-body problem. In Cartesian coordinates in the perifocal reference frame, the elliptic orbit can be expressed as

$$\frac{(x+ea)^2}{a^2}+\frac{y^2}{b^2}=1. \tag{1.41}$$

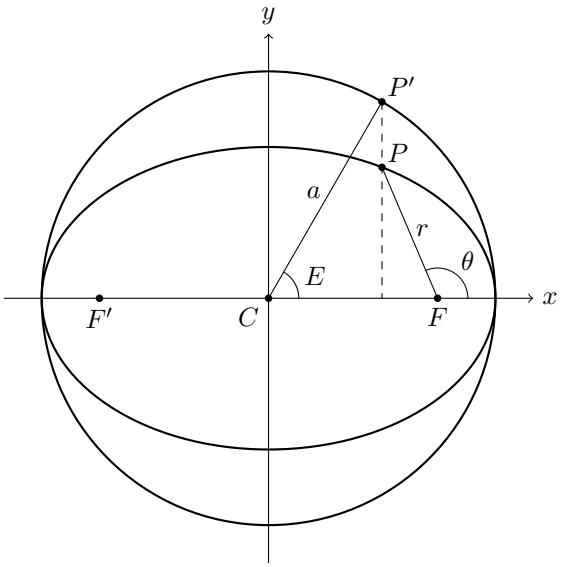
The mechanical energy for an elliptic orbit is $\xi < 0$. From the vis-viva equation (1.65), it is evident that the velocity at any point of the elliptic orbit satisfies $v^2 = v_e^2 - \mu/a$, where v_e is the local escape velocity (remember that the value of v_e depends on r as given further below).

The intersections of the ellipse with the minor axis can be found by finding the maximum (and minimum) of y in equation (1.30). These points are found at $\theta = \arccos(-e)$, $r = a$.

A number of properties of an ellipse stem from its relation with what is called the (major) *auxiliary circle*, a circle with the same center C as the ellipse with radius a:

> *For every point P in the ellipse, there is a corresponding point P' in the auxiliary circle. The y and y' coordinates of these two points are related by:*
>
> $$\boxed{\frac{y'}{y}=\frac{a}{b}.} \tag{1.42}$$

Since the area of the auxiliary circle is πa^2 and the y coordinate of every point in and on the ellipse is scaled by the factor b/a, it is easy to show that the area of the ellipse is $A_e = \pi ab$.



For every point P on the ellipse, it is useful to define the *eccentric anomaly* E as the angle from the center C of the ellipse to the corresponding point P' on the auxiliary circle. Point P on the ellipse can be defined with its x, y coordinates in the perifocal reference frame:

$$x = a(\cos E - e); \quad y = b \sin E. \tag{1.43}$$

The eccentric and the true anomalies are related by:

$$\sin E = \frac{\sqrt{1-e^2}\sin\theta}{1+e\cos\theta}; \qquad \cos E = \frac{e+\cos\theta}{1+e\cos\theta}; \tag{1.44}$$

$$\sin\theta = \frac{\sqrt{1-e^2}\sin E}{1-e\cos E}; \qquad \cos\theta = \frac{\cos E - e}{1-e\cos E}. \tag{1.45}$$

In the new variable, the radius from F can be written as

$$\boxed{r = a\,(1 - e\cos E)\,.} \tag{1.46}$$

The *orbital period* (sidereal period) τ of an elliptic orbit can be computed from the conservation of angular momentum. Noting that the area $\mathrm{d}A$ swept by $\boldsymbol{r}$ in a differential time $\mathrm{d}t$ can be expressed as the area of a triangle, $\mathrm{d}A = r^2\mathrm{d}\theta/2$, we can write:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{1}{2}r^2\dot{\theta} = \frac{1}{2}h = \text{const} \tag{1.47}$$

Integrating in a full orbit, and noting that the area of the ellipse is $A_e = \pi ab$,

$$\boxed{\tau = 2\pi\sqrt{\frac{a^3}{\mu}}.} \tag{1.48}$$

Associated to this we define the *mean elliptic angular rate* n_e as

$$\boxed{n_e = \sqrt{\frac{\mu}{a^3}}.} \tag{1.49}$$

1.6.2 Parabolic trajectories

Parabolic trajectories are the lowest-energy orbits that reach $r = \infty$. With $e = 1$ they must be considered a divisory case between ellipses ($e < 1$) and hyperbolas ($e > 1$). In Cartesian coordinates in the perifocal reference frame, the parabolic trajectory can be expressed as

$$x = \frac{p}{2} - \frac{y^2}{2p}. \tag{1.50}$$

The periapsis is located on $r = p/2$, the midpoint between the focus F and the directrix ℓ of the parabola.

The mechanical energy of a parabolic trajectory is $\xi = 0$. The magnitude of the velocity satisfies $v = v_e$ at all points (remember that the value of v_e depends on r as given below). The flight path angle in a parabolic trajectory satisfies $\gamma = \theta/2$ at all times.

While it is dissimilar to the elliptic and the hyperbolic cases, where an auxiliary anomaly can be defined with trigonometric or hyperbolic functions respectively, it is still useful to define a *parabolic anomaly*, defined simply as $P = \tan(\theta/2)$.

1.6.3 Hyperbolic trajectories

Hyperbolic orbits are extensively used in interplanetary flight. In Cartesian coordinates in the perifocal reference frame, the hyperbolic trajectory can be expressed as

$$\frac{(x + ea)^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{1.51}$$

The mechanical energy of an hyperbolic trajectory is $\xi > 0$. The velocity in an hyperbolic trajectory satisfies $v^2 = v_e^2 + v_h^2$ at all points (remember that the value of v_e depends on r as given below).

The true anomaly of the asymptotes θ_∞ is found by equating to zero the denominator of equation (1.24). This is related to the angle β of the asymptotes with the line of apses (i.e., the major axis of the hyperbola):

$$\cos\theta_\infty = -\frac{1}{e}; \quad \beta = \pi - \theta_\infty. \tag{1.52}$$

The *turning angle* δ is the angle between the two asymptotes, and for a spacecraft arriving traveling along the hyperbolic branch represents the change in direction at infinity, far away from the central body, after flying by it:

$$\delta = 2\theta_\infty - \pi = \pi - 2\beta \tag{1.53}$$

The occupied branch of the hyperbola lies in $-\theta_\infty < \theta < \theta_\infty$. The rest of the angular domain of θ is the vacant branch of the hyperbola for which $r < 0$. This second branch is only relevant in repulsive force problems, which is not the case for gravity.

The distance between the asymptote and a parallel line that passes by the focus F is called the *impact parameter* or *aiming radius* B, which coincides with $|b|$ is the magnitude of the (imaginary) semi-minor axis of the hyperbola,

$$B = -ea\sin\beta = -ea\sin\theta_\infty = -a\sqrt{e^2-1} = |b|. \tag{1.54}$$

Just as with the ellipse, a point P on the hyperbola can be defined with its coordinates x, y in the perifocal reference frame as:

$$x = |a|(e - \cosh H); \quad y = |b|\sinh H. \tag{1.55}$$

where H is the so-called *hyperbolic anomaly*. Remember that $\cosh^2 H - \sinh^2 H = 1$. The relation of H with the true anomaly θ is given by

$$\boxed{\sinh H = \frac{\sqrt{e^2-1}\sin\theta}{1+e\cos\theta}; \qquad \cosh H = \frac{e+\cos\theta}{1+e\cos\theta};} \tag{1.56}$$

$$\boxed{\sin\theta = -\frac{\sqrt{e^2-1}\sinh H}{1-e\cosh H}; \qquad \cos\theta = \frac{\cosh H - e}{1-e\cosh H}.} \tag{1.57}$$

1.7 Kepler's laws of orbital motion

Enunciated by Johannes Kepler (1571–1630) as empirical laws, based on the observations of Tycho Brahe:

1. The orbit of a planet is an ellipse, with the Sun at one of the two foci (more generally, orbits in the two-body problem can also be parabolas and hyperbolas).
2. A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.
3. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

These laws are easily demonstrated from the mathematical model introduced above, which is based on Newton's laws of motion and Newton's law of gravitation.

1.8 Velocity components and flight path angle

We can decompose the velocity vector into its radial and polar components along the unit vectors $\boldsymbol{u}_r$ and $\boldsymbol{u}_\theta$, $v_r = \dot{r}$ and $v_\theta = r\dot{\theta}$.

Taking derivatives in equation (1.24) we obtain an expression for the radial rate:

$$\boxed{v_r = \dot{r} = \frac{r\dot{\theta}e\sin\theta}{1+e\cos\theta} = \frac{\mu}{h}e\sin\theta.} \tag{1.58}$$

And, from equation (1.17), we find:

$$\boxed{v_\theta = r\dot{\theta} = \frac{h}{r} = \frac{\mu}{h}(1+e\cos\theta).} \tag{1.59}$$

The flight path angle γ is the angle of the velocity vector with the local horizontal (i.e., with the direction of $\boldsymbol{u}_\theta$):

$$\boxed{\tan\gamma = \frac{e\sin\theta}{1+e\cos\theta}.} \tag{1.60}$$

After transformation, the components of the velocity in Cartesian coordinates read:

$$v_x = \dot{x} = -\frac{\mu}{h}\sin\theta \qquad\qquad v_y = \dot{y} = \frac{\mu}{h}(\cos\theta + e) \tag{1.61}$$

1.8.1 Whittaker's theorem

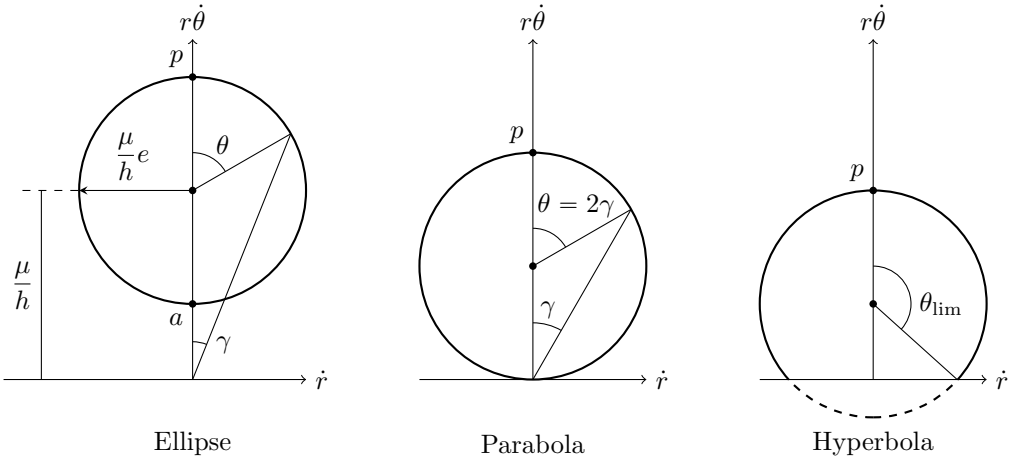
Inspecting equations (1.58) and (1.59) it is clear that we can write $\boldsymbol{v}$ as the sum of two components, one of constant magnitude in the variable direction of $\boldsymbol{u}_\theta$, and one of constant magnitude and constant direction along $\cos\theta\boldsymbol{u}_\theta + \sin\theta\boldsymbol{u}_r = \boldsymbol{j}$:

$$\boldsymbol{v} = \frac{\mu}{h}\boldsymbol{u}_\theta + \frac{\mu e}{h}\boldsymbol{j} \tag{1.62}$$

At the periapsis the two components are parallel and add together, while at the apoapsis they partially cancel out. The existence this decomposition of the velocity vector is known as *Whittaker's theorem*.

1.8.2 Velocity diagrams

The *hodograph* is the representation of the velocity of a point particle as a function of time *in velocity space*. The shape of the hodograph of a two-body problem in polar coordinates is a displaced circle (or part of a circle, in the hyperbolic case), as can be inferred from equations (1.58) and (1.59). The center of the circle is at a distance μ/h from the origin along the v_θ axis, and the radius of the circle is $\mu e/h$. Below is a schematic plot of the hodograph for the elliptic, parabolic, and hyperbolic cases:



Ellipse Parabola Hyperbola

Many observations can be made from these plots:

1. The pericenter is the point of maximum velocity, the apocenter (relevant only in the elliptic case) is the point of minimum velocity.

2. The figure is symmetric with respect to the v_θ axis. This means that the velocity in the part of the orbit from pericenter to apocenter (or infinity, in the parabolic and hyperbolic case), which has $\dot{r} > 0$, is symmetric with the part of the orbit from apocenter (or infinity, in the parabolic and hyperbolic case) to pericenter, which has $\dot{r} < 0$.

3. In the parabolic case the flight path angle is $\gamma = \theta/2$.

4. In the parabolic case as P_2 approaches infinity ($\theta \to \pi$), its velocity goes to zero. This is not the case in the hyperbolic case as P_2 goes to infinity ($\theta \to \theta_{\lim}$), where the velocity remains finite.

1.9 Vis-viva equation and fundamental velocities

Particularizing equation (1.13) at the periapsis $r = r_p = a(1-e)$, noting that the velocity vectors there are perpendicular to $\boldsymbol{r}$ so that we can write $h = r_p v_p$, and rearranging the equations, we obtain:

$$\xi(1-e) = \frac{h^2}{2a^2(1-e)} - \frac{\mu}{a}, \tag{1.63}$$

Using $p = h^2/\mu = a(1-e^2)$,

$$\xi = -\frac{\mu}{2a} \tag{1.64}$$

Knowing the value of the specific mechanical energy, we write the version of equation (1.13) that is known as *vis-viva equation*:

$$\boxed{\frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}.} \tag{1.65}$$

Observe that the mechanical energy ξ is negative for all elliptical orbits, zero for parabolic orbits, and positive for hyperbolic orbits. Observe also that if a spacecraft is at a radius r with velocity magnitude v, then the semi-major axis a, and thus its mechanical energy and period, are fully determined, regardless of the direction of the velocity vector (naturally, though, the angular momentum vector $\boldsymbol{h}$ and the eccentricity e of the orbit do depend on the direction of the velocity vector).

We recapitulate two useful relations between conic section parameters and the two-body problem parameters:

$$\boxed{p = \frac{h^2}{\mu}; \quad a = -\frac{\mu}{2\xi}.} \tag{1.66}$$

Equation (1.65) is extremely useful to find several important velocities:

Circular velocity: A circular orbit has $r = a$. At any given radius r, the velocity that an object must have to be in a circular orbit is

$$\boxed{v_c = \sqrt{\frac{\mu}{r}},} \tag{1.67}$$

in a direction perpendicular to $\boldsymbol{r}$.

Escape velocity: The minimum energy that allows P_2 to reach $r \to \infty$ and fully escape from the attraction of the gravitational well of P_1 is[2]:

$$\boxed{v_e = \sqrt{\frac{2\mu}{r}}.} \tag{1.68}$$

If P_2 has $v = v_e$ (regardless of the direction of the velocity vector), it is in a parabolic trajectory, which is the lowest-energy orbit that is not closed (and goes to infinity).

Velocity at periapsis: When $r = r_p = a(1-e)$, the velocity is perpendicular to $\boldsymbol{r}$ and has magnitude:

$$\boxed{v_p = \sqrt{\frac{\mu}{a}\left(\frac{1+e}{1-e}\right)}.} \tag{1.69}$$

Velocity at apoapsis: Only meaningful for elliptical orbits. When $r = r_a = a(1+e)$, the velocity is perpendicular to $\boldsymbol{r}$ and has magnitude:

$$\boxed{v_a = \sqrt{\frac{\mu}{a}\left(\frac{1-e}{1+e}\right)}.} \tag{1.70}$$

Remember that $\boldsymbol{r} \perp \boldsymbol{v}$ at periapsis and apoapsis (i.e., $\gamma = 0$ there), and that $r_a v_a = r_p v_p = h$.

[2] *Rabbithole*: For a very massive body, the escape velocity can equal the speed of light c. We define the Schwarzschild radius r_s as the radius at which $v_e = c$. No particles (even photons) can escape from a body whose radius reaches this critical condition. The Schwarzschild radius is used to define the event horizon of black holes.

Excess hyperbolic velocity: When in a hyperbolic orbit, P_2 will go to $r \to \infty$ with a non-zero velocity. This is known as the excess hyperbolic velocity:

$$\boxed{v_h = \sqrt{-\frac{\mu}{a}}.} \tag{1.71}$$

The square of v_h is denoted c_3 and termed *characteristic energy* of the hyperbolic trajectory, $c_3 = v_h^2$.

1.10 Common reference frames

The plane of motion of the Earth about the Sun is called the *ecliptic*. The plane of the equator of the Earth is the *equatorial plane*.

We can define an ecliptic Earth-centered inertial reference frame (ECI-ecliptic) as:

1. The origin O of the reference frame is at Earth's center of mass
2. The Oxy plane is the plane of the ecliptic
3. The Oz axis points in the direction of the angular momentum vector of Earth's orbit
4. The Ox axis points in the direction of the vernal equinox (a.k.a. first point Aries, ♈), i.e., the direction of the Sun as seen from the Earth on the first day of spring.

Likewise, we can define an equatorial Earth-centered inertial reference frame (ECI-equatorial) as:

1. The origin O of the reference frame is at Earth's center of mass
2. The Oxy plane is the equatorial plane
3. The Oz axis points in the direction of Earth's North pole
4. The Ox axis points in the direction of the vernal equinox (a.k.a. first point Aries, ♈), i.e., the direction of the Sun as seen from the Earth on the first day of spring.

The Ox axis as defined above is the same for both reference frames, and coincides with the intersection line between the ecliptic and the equatorial planes.

Apart from these two inertial reference frames, we can define others by translating the origin. For example, we can define a Sun-centered inertial reference frame (ecliptic or equatorial).

To be precise, none of these reference frames can be strictly called *inertial* (a more correct term is *pseudo-inertial*), as their origins are accelerated by other bodies in the universe. To keep the definition of these reference frames simple and useful, the inertial accelerations are typically treated as perturbations. Moreover, it should be noted that these definitions are not rigorous: the plane of the ecliptic and the equator, as well as the direction of the vernal equinox and Earth's North pole, evolve slowly in time due to many effects. The most common solution to this problem is to fix a date (or *epoch*) for the definition of the reference frames. Currently, the epoch 12:00 TT January 1 of year 2000 is used (J2000.0)[3]

An Earth-centered Earth-fixed reference frame (ECEF) can be defined as follows:

1. The origin O of the reference frame is at Earth's center of mass
2. The Oxy plane is the equatorial plane
3. The Oz axis points in the direction of Earth's North pole
4. The Ox axis passes through Greenwich's meridian

[3] *Rabbithole*: Noteworthy, the definition of time itself is problematic: relativistic effects cause clocks at different states of motion to tick differently. This gives rise to various time references that are used in astrodynamics problems.

The angular velocity of the ECEF reference frame with respect to ECI-equatorial is $7.2921 \cdot 10^{-5}$ rad/s, and the rotation period is one sidereal day, 23.93447 h. This is the time it takes Earth to complete one full revolution, and is slightly less than one solar day, which is the time it takes the Earth to repeat its orientation with respect to the Sun, including the effect of its orbital motion around it.

Clearly, transforming position, velocity and acceleration vectors from one reference frame to another requires using the well-known connection formulas: see the 2nd year Mechanics lecture notes. Expressing a vector in a different vector basis requires transforming the basis (e.g. by means of a rotation matrix, or by expressing each unit vector of one basis in terms of the other).

A table of position, velocity, orbital elements, or other means that allow to identify where an object is in space and how it is moving in a given reference frame is termed *ephemeris*[4].

1.10.1 Celestial sphere

The celestial sphere associated to an inertial reference frame is the projection of all objects in space into an imaginary sphere centered at the origin of an inertial reference frame. All points of the sphere are identified by the two angles in spherical coordinates, and usually ignoring the radius r (a.k.a. range).

In the case of ecliptic inertial reference frames, the two angles are called α and δ:

Right ascension: α is the azimuthal angle, defined with respect to the direction of the Ox axis (vernal equinox).

Declination: δ is the latitude angle, defined with respect to the direction of the Oxy plane (ecliptic plane).

1.11 Classical orbital elements

The state vector of an object in space can be defined as the 6-dimensional vector that results from the concatenation of its position vector $\boldsymbol{r}$ and its velocity vector $\boldsymbol{v}$ in a given reference frame:

$$\boldsymbol{X} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]. \tag{1.72}$$

While the state vector contains all the information on the motion of the object, it is not the preferred representation in many circumstances, as the type of orbit, its orientation in space, etc. are not obvious from $\boldsymbol{X}$.

The six *classical orbital elements* (COE) provide the same information while describing the orbit at the same time. The first three elements actually are the three Euler angles that define the orientation of the perifocal reference frame in space. Unless otherwise noted, an ECI-equatorial reference frame is used as the starting point for the Euler angles of geocentric orbits:

Right ascension of the ascending node (RAAN): Ω is the azimuthal angle from the Ox axis to the *line of nodes* (the intersection line between the Oxy plane and the orbital plane), in the direction where the body crosses the Oxy plane from below to above (i.e., the *ascending node*). It ranges from 0 to 2π.

Inclination: the orbit inclination i is the angle between the Oxy plane and the orbital plane. It ranges from 0 to π.

Argument of periapsis: ω is the angle from the line of nodes to the periapsis of the orbit. It ranges from 0 to 2π.

The next two give the size and shape of the orbit:

Semi-major axis: a gives the size of the orbit, and also its energy and its period. The orbital parameter p is sometimes used instead of a, especially for near-parabolic orbits ($e \simeq 1$).

Eccentricity: e describes the shape and type of trajectory. It goes without saying, e is a nondimensinal parameter without units.

[4] A common repository to obtain ephemerides of celestial bodies is the NASA's JPL Horizons database.

Finally, the last one identifies the current position of the body in the orbit:

True anomaly: the angle θ between the periapsis and the current position of the body. It ranges from 0 to 2π. In some references, the Greek letter ν is used to denote the true anomaly. The time since passage by the periapsis, or any other alternative anomaly (e.g. eccentric anomaly E) can also be used.

Converting to and from COE and state vector is a central aspect of many calculations.

1. To go from COE to state vector, first compute $\boldsymbol{r}$ and $\boldsymbol{v}$ in the perifocal reference frame vector basis. Then, transform the matrix to the original reference frame in which the state vector should be provided (e.g. ECI-equatorial), by undoing the three rotations associated to the Euler angles.

2. To go from state vector to COE, first compute $\boldsymbol{h}$, $\boldsymbol{n} = \boldsymbol{k} \times \boldsymbol{h}$ (the ascending node vector), and $\boldsymbol{e}$. Using these vectors, the value of ξ and h, and the angles between $\boldsymbol{i}$ and $\boldsymbol{n}$, between $\boldsymbol{n}$ and $\boldsymbol{e}$, and between $\boldsymbol{e}$ and $\boldsymbol{r}$, it is possible to compute the COE. Beware of correctly resolving the quadrants for all angles.

1.11.1 Singular cases

The classical orbital elements have several singularities that must be dealt with:

1. In prograde ($i = 0$) and retrograde ($i = \pi$) equatorial orbits, there is no line of nodes. Thus Ω and ω are not well defined. Instead, the *true longitude* (or right ascension) *of periapsis* is used: $\tilde{\omega}_{\text{true}} = \Omega + \omega$. This is the usual singularity that exists for every set of Euler angles.

2. Circular orbits ($e = 0$) have no periapsis. Thus, ω and θ are not well defined. Instead, the *argument of latitude* is used: $u = \omega + \theta$.

3. Lastly, equatorial circular orbits have problems with Ω, ω and θ. They are replaced by the *true longitude* (or right ascension) *of the body*: $\lambda_{\text{true}} = \Omega + \omega + \theta$.

Other alternatives to the COE exist which are more suited to study particular orbit types, or possess advantages for numerical computing. Examples include the so-called equinoctial elements.

1.12 Kepler's equation

So far the problem of relating time t with position (r, θ) has not been addressed. Analytically, starting from equation (1.17), we can find this relation by integration:

$$h\mathrm{d}t = r^2\mathrm{d}\theta; \tag{1.73}$$

$$\int_{t_p}^{t} \frac{\mu^2}{h^3}\mathrm{d}t = \int_0^{\theta} \frac{\mathrm{d}\theta}{(1 + e\cos\theta)^2}, \tag{1.74}$$

with t_p the value of time at periapsis passage, where $\theta = 0$. This integral can be found in mathematical handbooks, and is treated separately for $e < 1$, $e = 1$, and $e > 1$. The results are cumbersome, and in order to obtain workable expressions, one typically resorts to changing the variable θ into the auxiliary anomalies (E, P, H) for each case. There is also a universal formulation that can be used for the three conic sections, but which involves advanced functions.

Geometrically, constructions have been used since the time of Kepler to relate the auxiliary anomalies to time. Kepler's equations are nonlinear, so when the time is known and we solve for the auxiliary anomaly, an iterative numerical method (such as Newton-Raphson method) is typically used. Note that when the anomaly is known and we solve for time, computation is straightforward and no iterative numerical method is needed.

1.12.1 Elliptic case

When $e < 1$, equation (1.74) can be integrated after changing the integration variable into the eccentric anomaly E,

$$\boxed{M_e = E - e\sin E,} \tag{1.75}$$

where $M_e = n_e(t - t_0)$ is the *mean anomaly in the elliptic case*, with $n_e = \sqrt{\mu/a^3}$ the corresponding *mean angular rate*, and t_0 is the time of passage by the periapsis.

From periapsis to apoapsis, $M_e \leqslant E \leqslant \theta$. From apoapsis to periapsis , $M_e \geqslant E \geqslant \theta$. At peri- and apoapsis, $M_e = E = \theta$. Note that for $e = 0$ (circular case), the three anomalies coincide at all times.

1.12.2 Parabolic case

When $e = 1$, equation (1.74) can be integrated into:

$$\boxed{M_p = \frac{1}{2}\tan\frac{\theta}{2} + \frac{1}{6}\tan^3\frac{\theta}{2},} \tag{1.76}$$

where $M_p = n_p(t - t_0)$ is the *mean anomaly in the parabolic case*, with $n_p = \sqrt{\mu/p^3}$ the corresponding *mean angular rate*, and t_0 is the time of passage by the periapsis.

1.12.3 Hyperbolic case

When $e > 1$, equation (1.74) can be integrated after changing the integration variable into the hyperbolic anomaly H,

$$\boxed{M_h = e\sinh H - H,} \tag{1.77}$$

where $M_h = n_h(t - t_0)$ is the *mean anomaly in the hyperbolic case*, with $n_h = \sqrt{-\mu/a^3}$ the corresponding *mean angular rate*, and t_0 is the time of passage by the periapsis.

Observe that the elliptic and hyperbolic cases are the same, if we make $H = iE$, $n_h = in_e$.

1.12.4 Universal formulation

Kepler's equation can be also cast in terms of a *universal anomaly* χ, which has units of $[L^{1/2}]$:

$$\sqrt{\mu}(t - t_0) = e\chi^3 \mathrm{S}\left(\frac{\chi^2}{a}\right) + a(1 - e)\chi. \tag{1.78}$$

In this expression $\mathrm{S}(z)$ belongs to the functions of the Strumpff class:

$$\mathrm{S}(z) = \begin{cases} \dfrac{\sqrt{z} - \sin\sqrt{z}}{\sqrt{z}^3} & (z > 0) \\ \dfrac{1}{6} & (z = 0) \\ \dfrac{\sinh\sqrt{-z} - \sqrt{-z}}{\sqrt{-z}^3} & (z < 0) \end{cases} \tag{1.79}$$

and it can be shown that equation (1.78) reduces to the elliptic, parabolic, and hyperbolic cases for the corresponding sign of $z = \chi^2/a$. In fact, the relation of χ with the elliptic, parabolic and hyperbolic anomalies is given by:

$$\chi = \begin{cases} \sqrt{a}E & \text{(ellipse)} \\ \sqrt{p}P & \text{(parabola)} \\ \sqrt{-a}H & \text{(hyperbola)} \end{cases} \tag{1.80}$$

2 Orbital Maneuvering

Spacecraft with propulsion systems can change their trajectory. Moreover, some spacecraft must counteract external perturbations to remain in the mission orbit. In this chapter, we model rocket propulsion in the equation of motion of a spacecraft. The fundamental orbital maneuvers are then discussed.

2.1 Thrust Model

The thrust force $\boldsymbol{F}$ provided by a rocket can be modeled as

$$\boldsymbol{F} = F\boldsymbol{u}_F, \quad \text{with: } F = \dot{m}c, \tag{2.1}$$

where $\boldsymbol{u}_F$ is the unit vector of the thrust force, $\dot{m}$ is the *mass flow rate* of propellant being ejected from the rocket; and c is the *effective exhaust velocity*, which includes any pressure contributions. For historical reasons, instead of c it is common to use the *specific impulse*, I_{sp}. This is equal to the exhaust velocity divided by the gravity acceleration constant at sea level, $g_0 = 9.81 \simeq 10$:

$$I_{sp} = \frac{c}{g_0} \tag{2.2}$$

While c has units of velocity, I_{sp} has units of time, and it is always given in seconds.

The two-body problem equation of motion of the spacecraft is now extended to include the acceleration due to thrust, Γ,

$$\Gamma(t) = \frac{F}{m^P} = \frac{\dot{m}g_0 I_{sp}}{m^P}, \tag{2.3}$$

obtaining as a result the following differential equation:

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t} = -\frac{\mu}{r^3}\boldsymbol{r} + \Gamma(t)\boldsymbol{u}_F(t). \tag{2.4}$$

The mass of the spacecraft changes according to the ejection of propellant,

$$\frac{\mathrm{d}m^P}{\mathrm{d}t} = -\dot{m} = -\frac{F}{g_0 I_{sp}} = -\frac{m^P\Gamma}{g_0 I_{sp}} \tag{2.5}$$

The mass of the spacecraft, therefore, is no longer constant and becomes an additional state variable, whose evolution must be computed along with its position and velocity. Hence, the state vector of the spacecraft must include this new parameter, $(m^P, \boldsymbol{r}, \boldsymbol{v})$.

The higher I_{sp} is, the lower the total propellant mass consumption for a given space maneuver. This is evident if we integrate equation (2.4) for a spacecraft far away from any other body, flying and firing its rockets along the same direction:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{g_0 I_{sp}}{m^P}\frac{\mathrm{d}m^P}{\mathrm{d}t} \Rightarrow \Delta v = g_0 I_{sp} \ln\left(\frac{m_0^P}{m_f^P}\right), \tag{2.6}$$

where m_0^P and m_f^P stand for the initial and final mass of the spacecraft. This is the famous Tsiolkovsky's rocket equation, which relates the propellant mass consumption with the total change in velocity Δv (read *delta vee*) of the maneuver and the specific impulse of the rockets. Note that, in general, $\Delta\boldsymbol{v}$ has to be treated as a vector, unless the initial velocity and the firing direction are equal.

The main relevant performance figures of space propulsion system are, thus

1. The maximum thrust force, and its *throttle range*
2. The specific impulse, and how constant it is in the throttle range
3. The rocket lifetime (i.e., how long can it operate before it is irreversibly damaged)
4. Rocket inert mass, volume...

Current rocket technologies can be broadly categorised into two large groups: **chemical propulsion**, with high thrust and relatively low specific impulse (up to about $I_{sp} = 500$ s), and **electric propulsion**, with low thrust and high specific impulse (up to about $I_{sp} = 5000$ s, which corresponds to about 50 km/s effective exhaust velocity). Space maneuvers with chemical and electric propulsion are very different and require different analysis techniques and planning techniques. The typical high thrust maneuver corresponds, for example, to a launcher injection in orbit and chemical rockets, such as hydrazine rocket systems. These maneuvers will be modeled as **impulsive maneuvers**. On the other hand, the typical electric propulsion maneuver is done for station keeping in GEO or for slow orbit raising. These are known as **low-thrust maneuvers**.

The analysis of impulsive high thrust maneuvers is discrete because maneuvers bring the spacecraft from one two-body problem conic to another, which intersect at the maneuver point. This model is simple because it does not required the integration of equations (2.4) and (2.5). It suffices to "patch" together several conic sections, which are characterized analytically. The analysis of low thrust maneuvers is clearly more complex, and its full solutino requires the integration of equations (2.4) and (2.5).

2.2 Propagation and optimal control Problems

The dynamical model of a spacecraft that performs maneuvers is the set of ODEs introduced above, equations (2.4) and (2.5). Depending on whether the thrust acceleration $\Gamma(t)$ is known (or prescribed) a priori or not, we consider two different types of maneuver problems:

1. When the thrust is known, i.e. when $\Gamma(t)$ is established *a priori*, the objective is usually to predict with the required accuracy the future position and velocity of the spacecraft. For instance, assuming that the rocket is providing the maximum thrust tangential to the trajectory during 20 seconds, determine the final position and velocity, and the resulting orbit after the maneuver. This is known as **Propagation Problem** and can be formulated as:

   > *Given* $\Gamma(t)$, $\boldsymbol{u}(t)$ *and initial conditions obtain the time evolution of the system using* (2.4) *and* (2.5).

2. On the other hand, when the thrust law $\Gamma(t)$ is not known, the problem is usually to determine the control law that fulfils certain constraints and minimizes a certain objective. For instance, to compute the maneuver that brings the spacecraft to a 600 km altitude circular orbit from a 500 km altitude circular orbit with minimum propellant consumption. This is known as **Optimal Control Problem** and can be formulated as:

   > *Given boundary conditions and a cost function* J *obtain* $\Gamma(t)$, $\boldsymbol{u}(t)$ *using* (2.4) *and* (2.5).

   The usual cost/objective functions (the function to be minimized) are mass and time. Mass is a critical parameter in any space mission, and its value imposes requirements to all the spacecraft subsystems. In turn, time is crucial in manned missions, in which the time spent in orbit is relevant for human beings. There can be other objective functions, but mass and/or time are (almost) always included.

   The mathematical way of defining the objective function may vary. For example, the objective function J may be defined in terms of a *characteristic velocity*:

   $$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{\Gamma m}{c} \Rightarrow c \log\left(\frac{m_0}{m_f}\right) = \int_{t_0}^{t_f} \Gamma \mathrm{d}t = J,$$

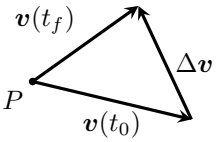
   which is equivalent to the accumulated Δv due to the operation of the rockets throughout the duration of the mission. Observe that minimizing this J is the same as minimizing the propellant mass consumption.

2.3 Impulsive Maneuvers

High thrust maneuvers (i.e., done with chemical rockets) with short burn times much shorter than the orbital period are assumed to be *impulsive* or *instantaneous*: there is an instantaneous change in velocity without any change in the position of the spacecraft. That is,

$$\boldsymbol{r}(t_f) - \boldsymbol{r}(t_0) = \boldsymbol{0}, \qquad \boldsymbol{v}(t_f) - \boldsymbol{v}(t_0) = \Delta\boldsymbol{v}.$$

This approximation simplifies the computation of maneuvers substantially: The trajectory is the result of patching together the two-body problem solution *before* and *after* the impulsive burn, introducing a discontinuous change of the velocity vector at that instant. This means a change in the classical orbital elements of the spacecraft, too. Between successive rocket burns, the spacecraft flies without firing its rockets. This is called *coasting*.



2.3.1 Hohmann Transfer

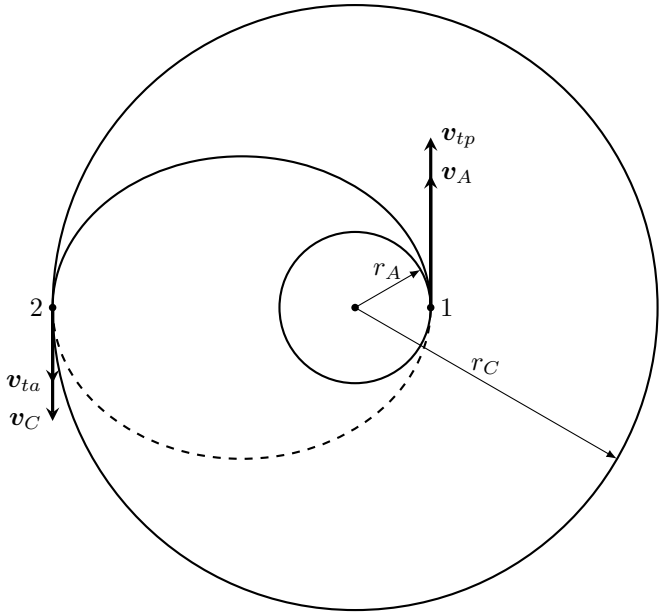
A German engineer, Walter Hohmann (1880-1945) published a book entitled "Die Erreichbarkeit der Himmelskörper" (*The Attainability of Heavenly Bodies*) in 1925, where he demonstrated which is the minimum energy transfer between two circular orbits, using only two impulses. This type of transfer is quite relevant because the orbits of the planets around the Sun are almost circular, and, therefore, Hohmann was able to compute the minimum Δv associated to a direct interplanetary mission to visit another planet in the Solar System using impulsive maneuvers. The problem faced by Hohmann is similar to the problem of a satellite that is orbiting the Earth and is wanted to go from a circular orbit of radius r_A to another of radius r_C.

The transfer is done with an elliptic arc that is tangent to the two circular orbits. The points of tangency are the peri- and apocenter of the transfer ellipse, and the two firings are done in the tangential direction, parallel to the velocity vector at those points.

Data of the problem:

1. Initial circular orbit of radius r_A.
2. Final circular orbit of radius r_C, coplanar with the initial one.

We will assume $r_C > r_A$ (the other possibility swaps the peri- and apocenter of the transfer orbit, and requires a few sign changes in the equations below).



The transfer ellipse has radius at pericenter $r_p = r_A$, radius at apocenter $r_p = r_C$, and semi-major axis $a_t = (r_A + r_C)/2$.

Since the firings are parallel to the velocity, we can work with scalars rather than vectors, and the required Δv at each firing can be computed using only the vis-viva equation. The first firing makes the rocket leave the initial orbit and enter the tranfer ellipse (at pericenter); the second firing circularizes the orbit at the desired altitude (at apocenter):

$$\Delta v_1 = v_{tp} - v_A = \sqrt{\frac{2\mu}{r_A} - \frac{2\mu}{r_A + r_C}} - \sqrt{\frac{\mu}{r_A}} \tag{2.7}$$

$$\Delta v_2 = v_C - v_{ta} = \sqrt{\frac{\mu}{r_C}} - \sqrt{\frac{2\mu}{r_C} - \frac{2\mu}{r_A + r_C}} \tag{2.8}$$

The total Δv of the transfer is the sum of Δv_1 and Δv_2.

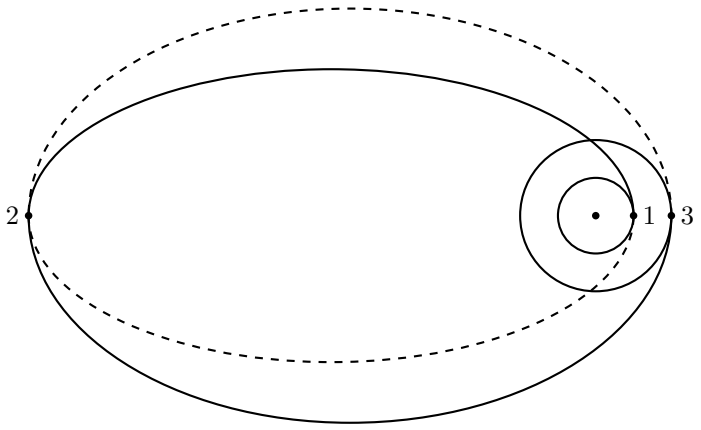
2.3.2 Bi-elliptic Hohmann Transfer

The Hohmann transfer is the optimal one between circular coplanar orbits when we restrict ourselves to the analysis of two-burns maneuvers. Nevertheless, other strategies that drop this constraint may result in more efficient maneuvers in terms of Δv. In this section, we analyse one of this strategies: the bi-elliptic Hohmann transfer (BHT). It is a transfer trajectory between two co-planar circular orbits, as in the case of the "pure" Hohmann transfer. Nonetheless, the BHT is a three-impulse maneuver. Therefore, there is an initial circular orbit A (with radius r_A), a final circular orbit C (with radius r_C) and an intermediate point B at a distance r_B:

$$\Delta v_1 = \sqrt{\frac{2\mu}{r_A} - \frac{2\mu}{r_A + r_B}} - \sqrt{\frac{\mu}{r_A}} \tag{2.9}$$

$$\Delta v_2 = \sqrt{\frac{2\mu}{r_B} - \frac{2\mu}{r_B + r_C}} - \sqrt{\frac{2\mu}{r_B} - \frac{2\mu}{r_A + r_B}} \tag{2.10}$$

$$\Delta v_3 = \sqrt{\frac{\mu}{r_C}} - \sqrt{\frac{2\mu}{r_C} - \frac{2\mu}{r_B + r_C}} \tag{2.11}$$



The case of interest is when intermediate point B has $r_B \gg r_A, r_C$. The rationale behind this strategy lies in the fact that, when we consider the needed Δv to come back from a very distant position (i.e., a near parabolic trajectory), we get:

$$r_B \to \infty \Rightarrow \Delta v_B \to 0 \tag{2.12}$$

That is, as $r_B \to \infty$, the cost of the maneuver at B is null. The question that arises is, then, whether or not the BHT is more efficient than a "pure" Hohmann transfer. In other words, which of the two transfers is more efficient in terms of propellant mass consumption[5]. The following non-dimensional

[5] In terms of time, BTH is always worse than a pure Hohmann transfer.

variables are defined:

$$\alpha = \frac{r_C}{r_A}; \quad \beta = \frac{r_B}{r_A}; \quad \Delta\tilde{v}_{\mathrm{HT}} = \frac{\Delta v_{\mathrm{HT}}}{v_A}; \quad \Delta\tilde{v}_{\mathrm{BHT}} = \frac{\Delta v_{\mathrm{BHT}}}{v_A} \tag{2.13}$$

In these variables we find:

$$\Delta\tilde{v}_{\mathrm{HT}} = \frac{1}{\sqrt{\alpha}} - \frac{\sqrt{2}(1-\alpha)}{\sqrt{\alpha}\sqrt{1+\alpha}} - 1 \tag{2.14}$$

$$\Delta\tilde{v}_{\mathrm{BHT}} = \sqrt{\frac{2(\alpha+\beta)}{\alpha\beta}} - \frac{1+\sqrt{\alpha}}{\sqrt{\alpha}} - \frac{\sqrt{2}}{\sqrt{\beta}\sqrt{1+\beta}}(1-\beta) \tag{2.15}$$

Both functions depend on α, but only the latter depends on β. We arrive to the following conclusions:

1. The pure Hohmann transfer is cheaper if $\alpha \leqslant 11.94$.
2. The Bi-elliptic transfer is cheaper if $\alpha \geqslant 15.58$.
3. For $11.94 < \alpha < 15.58$, results depend on β.

2.3.3 Phasing Maneuver

When we do not want to change the orbit of a spacecraft, but move it forward or backward a given phase angle or anomaly $\Delta\theta$, a phasing maneuver is carried out. One example is the rendezvous maneuver between two space vehicles, when both are in the same orbit: the object at the final position that we want to reach is the "target", and the spacecraft that will maneuver to arrive there is the "interceptor" or "chaser". It is assumed that the target does not perform any maneuver, and merely moves in two-body problem motion. Another example is a station keeping maneuver, in which a spacecraft must change its location within its orbit, e.g. to correct the East-West deviations of GEO satellites. In this case, we can consider the desired final position that we want to attain as a "virtual target satellite" that we want to rendezvous with.

In all cases, we must account for the motion of the target with time, not only that of the chaser. Here we consider only the case where the target and the chaser are on a given circular orbit of radius a_c. We denote the true anomaly of the target and chaser as $\theta_t(t)$ and $\theta_c(t)$ respectively. We define the difference in phase as

$$\Delta\theta(t) = \theta_t(t) - \theta_c(t) \tag{2.16}$$

Calling $n_c = \sqrt{\mu/a_c^3}$ the mean angular rate at the circular orbit, we can write $\theta_t(t) = \theta_t(0) + n_c t$. Observe that, if the chaser does not perform any maneuvers, $\theta_c(t) = \theta_c(0) + n_c t$ as well, and $\Delta\theta$ remains constant in time.

There are many ways to accomplish a phasing maneuver. One possibility is to drop the spacecraft into a coplanar elliptic waiting orbit with different semi-major axis and thus with a different period. After a predefined amount of time, which must equal an integer number of periods of the waiting ellipse, the spacecraft is maneuvered back into the initial orbit, but its phase relative to the target will have changed. If the waiting orbit and the waiting time are chosen correctly, then the new $\Delta\theta$ will be zero.

The waiting time is:

$$t_w = 2\pi N\sqrt{\frac{a_w^3}{\mu}} \tag{2.17}$$

The achieved $\Delta\theta$ relative to the target, when the chaser is inserted back into the circular orbit at the same angular position whence it left, is:

$$\Delta\theta = n_c t_w = 2\pi N\sqrt{\frac{a_w^3}{a_c^3}} + 2\pi k \tag{2.18}$$

In this equation there are two free parameters: the integer N, and the real number a_w. Moreover, since we are dealing with angles, we have added a factor $2\pi k$, with k integer, to account for the periodicity of $\Delta\theta$. Therefore, multiple solutions exist to the phasing problem. Clearly, the smaller the change of semimajor axis $(a_w - a_c)$, the smaller the cost of the maneuver; on the other hand, the total time for the maneuver may decrease if $(a_w - a_c)$ is increased.

Another possibility is to perform a Hohmann transfer to a circular phasing orbit of radius a_p with two impulses, then come back to the initial orbit with the reverse Hohmann transfer (another two impulses). Two major differences exist in this case: we do not need to reenter the initial orbit at the same angular position as before, and the waiting time can be selected with more freedom. Calling $t_t = \pi\sqrt{a_t^3/\mu}$ the time on each of the two elliptic arcs in the Hohmann transfers (with $a_t = (a_c + a_p)/2$ the semimajor axis of the transfer ellipse), and t_p the phasing time in the phasing orbit, the total waiting time in this case is:

$$t_w = 2t_t + t_p \tag{2.19}$$

The achieved $\Delta\theta$ relative to the target is:

$$\Delta\theta = 2n_c t_t + (n_c - n_p)t_p + 2\pi k \tag{2.20}$$

This choice brings more freedom and more parameters to optimize the maneuver. However, the number of required thruster firings doubles.

2.3.4 Plane Change Maneuver

So far we have considered coplanar maneuvers. When the orbit inclination or the RAAN need to be changed, it is necessary to perform a plane change maneuver. We must retain the full vector character of $\Delta\boldsymbol{v}$ in this case.

The two orbits, initial and final, must be defined a priori and they must intersect on at least one point, location where the single rocket firing of the maneuver will take place. If the velocity at that point for the initial and final orbit is $\boldsymbol{v}_i$ and $\boldsymbol{v}_f$ respectively, then the required $\Delta\boldsymbol{v}$ is

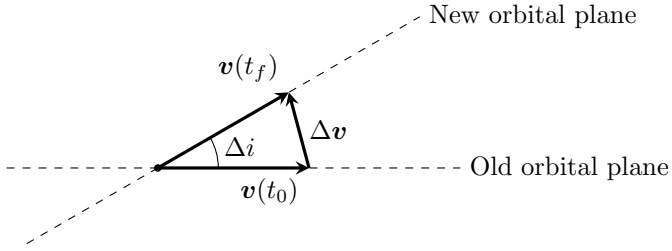
$$\Delta\boldsymbol{v} = \boldsymbol{v}_f - \boldsymbol{v}_i. \tag{2.21}$$

In the particular case where the two orbits are circular with equal radius a and different inclination i_i, i_f, the maneuver would be carried out at the ascending or descending node. In this case, the velocity on each orbit is constant and equal to $v = \sqrt{\mu/a}$. The maneuver is easily understood looking from the line of intersection of the two orbital planes, which in this case is the mutual line of nodes. We want to rotate the velocity vector at this point, while keeping its magnitude constant Applying trigonometry, the magnitude of $\Delta\boldsymbol{v}$ is:

$$\Delta v = 2v \sin\left(\frac{\Delta i}{2}\right) \simeq v\Delta i. \tag{2.22}$$

The approximation is only valid for small angles Δi.

Plane change maneuvers are quite expensive in terms of Δv far more than the coplanar maneuvers seen so far, and therefore consume substantial propellant mass. For a given plane rotation, the cost of the maneuver decreases the smaller the velocity at the initial and final orbits, hence, it is always desirable to make the plane changes at higher radius r from the central body (i.e., in the case of elliptic orbits, at apocenter). For low Earth orbits, it might be cheaper to raise the orbit first, perform the plane change maneuver, then reduce the orbit radius.



2.4 Low-thrust Maneuvers

In non-impulsive maneuvers, a small thrust force is applied continuously for long periods of time, comparable with the orbital period. Therefore, the assumptions of the impulsive model are not met and this model is not usable. The equations of motion of the spacecraft, (2.4) and (2.5), must be integrated to solve the low-thrust maneuver problem. There are very few circumstances where simplifications can be made. A particular case is the one discussed in the following subsection.

2.4.1 Low-thrust orbit raising

Assume that a spacecraft is maneuvered from a circular orbit with radius r_A to a higher circular, coplanar orbit of radius r_C, using a small tangential thrust acceleration $\Gamma(t)$ (small compared with the circular velocity divided by the orbital period, $\Gamma(t) \ll v/\tau$). The spacecraft will spiral slowly and perform the low-thrust transfer over many orbital periods of the initial orbit. At any point of the transfer, the spacecraft is in a near-circular orbit about the central body.

Since the thrust is applied tangentially along the direction of the velocity, we can compute the change in specific mechanical energy as

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \Gamma v \tag{2.23}$$

Noting that $\xi = -\mu/(2a)$, and approximating v as the local circular velocity v_c at the current radius a, we obtain the following differential equation on the semi-major axis,

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{\mu}{2a^2}\frac{\mathrm{d}a}{\mathrm{d}t} = \Gamma\sqrt{\frac{\mu}{a}} \quad \rightarrow \quad \sqrt{\frac{\mu}{4a^3}}\mathrm{d}a = \Gamma \mathrm{d}t. \tag{2.24}$$

Noting that the total Δv is the integral of the right-hand side,

$$\sqrt{\frac{\mu}{r_A}} - \sqrt{\frac{\mu}{r_C}} = v_A - v_C = \Delta v \tag{2.25}$$

This result shows that, as long as the assumptions of this approximation hold, the Δv for an electric, low-thrust orbit raising equals the difference of circular velocities of the initial and final orbits.

2.5 Optimal Control

TBW

3 Orbital perturbations

So far, we have considered the ideal two-body problem, where the only force acting on our spacecraft is the central force from the central body. In real life, many phenomena exert additional forces on the spacecraft, modifying its trajectory from the Keplerian conic sections. As long as these additional forces are small, they can be considered *perturbations* of the two-body problem. In general, inclusion of any perturbation means that we must integrate the full equation of motion of the spacecraft, and this typically needs to be approached numerically. Alternatively, we can try to derive approximate equations for the evolution of the classical orbital elements of the spacecraft, which are no longer constant values but vary slowly in time.

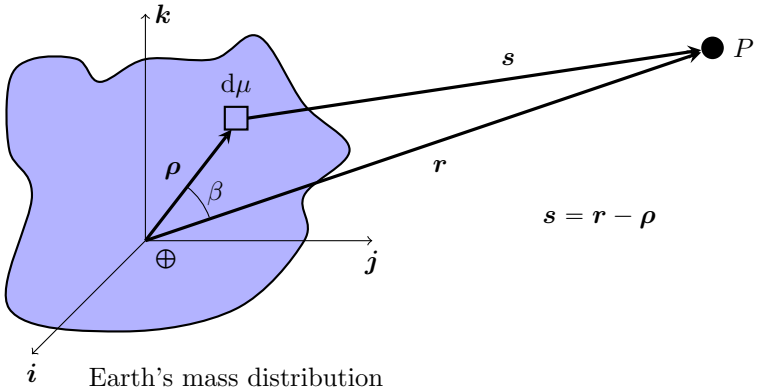
In this Chapter we overview the main perturbations on an Earth satellite, their relative importance as a function of height, how to model them, and their major effects on the trajectory. We will distinguish between *periodic* effects (those that have zero net drift over sufficiently long periods of time) and *secular* effects (those that involve a continuous drift in the orbital parameters of the spacecraft. Perturbations are also important for the evolution of the attitude (i.e. the orientation) of the spacecraft. Perturbation torques will be subject of another Chapter.

3.1 Non-spherical Earth gravity field

We have modeled the *main acceleration* $\boldsymbol{a}_m$ of the spacecraft as that due to the gravitational attraction of the Earth as

$$\boldsymbol{a}_m = -\frac{\mu}{r^3}\boldsymbol{r} = \nabla\left(\frac{\mu}{r}\right). \tag{3.1}$$

This potential would be exact if the Earth were a point mass or otherwise its mass distribution were spherical-symmetric. In reality, the Earth is not spherical but oblate, and its mass distribution lacks any symmetries. The difference between the actual acceleration and the main two-body problem acceleration given above is the *perturbing acceleration* due to the non-sphericity of the Earth.



Earth's mass distribution

Integrating over Earth's mass distribution with a mass differential $\mathrm{d}\mu = G\mathrm{d}m$, and calling s the distance between $\mathrm{d}\mu$ and the spacecraft P, the exact potential is:

$$\int \frac{\mathrm{d}\mu}{s} = \int \frac{\mathrm{d}\mu}{\sqrt{r^2+\rho^2-2r\rho\cos\beta}}. \tag{3.2}$$

With the assumtion $r > \rho$, the denominator can be series-expanded as"

$$\begin{aligned}
\frac{1}{\sqrt{r^2+\rho^2-2r\rho\cos\beta}} &= \frac{1}{r}\left[1+\frac{\rho}{r}\cos\beta+\frac{1}{2}\left(\frac{\rho}{r}\right)^2(3\cos^2\beta-1)+\frac{1}{2}\left(\frac{\rho}{r}\right)^3\cos\beta(5\cos^2\beta-3)+\ldots\right] \\
&= \frac{1}{r}\left[P_0(\cos\beta)+\frac{\rho}{r}P_1(\cos\beta)+\left(\frac{\rho}{r}\right)^2P_2(\cos\beta)+\left(\frac{\rho}{r}\right)^3P_3(\cos(\beta)+\ldots\right] \\
&= \frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{\rho}{r}\right)^n P_n(\cos\beta),
\end{aligned}$$

where $P_n(x)$ is the Legendre polynomial of degree n. Legendre polynomials are well-known polynomials with many useful properties. The first polynomials of the set are:

$$P_0(x) = 1; \qquad P_1(x) = x; \qquad P_2(x) = \frac{1}{2}(3x^2 - 1); \qquad P_3(x) = \frac{1}{2}(5x^3 - 3x).$$

Thus, the gravitational potential can be expressed as

$$\int \frac{\mathrm{d}\mu}{r} \sum_{n=0}^{\infty} \left(\frac{\rho}{r}\right)^n P_n(\cos\beta) \tag{3.3}$$

We observe that the zeroth term of the series expansion of the potential yields simply

$$\frac{\mu}{r}, \tag{3.4}$$

i.e., the potential of the ideal two-body problem with a point mass as the central body.

We consider next the case of an axisymmetric planet. It can be shown that the first-order term of the series vanishes in this case, and the series expansion can be written as:

$$\frac{\mu}{r} + \int \frac{\mathrm{d}\mu}{r} \sum_{n=2}^{\infty} \left(\frac{\rho}{r}\right)^n P_n(\cos\beta) \tag{3.5}$$

After some manipulation (see Tewari's book page 49 and following), and exploting the symmetry of the body, it is possible to write this result as

$$\frac{\mu}{r}\left[1 - \sum_{n=2}^{\infty} \left(\frac{R_\oplus}{r}\right)^n J_n P_n(\cos\phi)\right], \tag{3.6}$$

where $R_\oplus$ is the equatorial radius of the Earth, introduced here for normalization purposes only, ϕ is the colatitude angle taken from the North pole, and J_n are Jeffery's constants, which depend on the mass distribution of the planet. In the case of the Earth, the most important of these constants is J_2, which has the value

$$J_2 = 0.00108263. \tag{3.7}$$

TBC

3.2 Third-body perturbations

Other bodies in the universe, in particular the Sun and the Moon (lunisolar effects), exert an additional gravitational pull on our spacecraft, perturbing its trajectory. The position of these "third bodies" is usually known, either from ephemerides tables or from the propagation of their position in time. However, the geometric problem is quite complex as it involves the three-dimensional position of the bodies with respect to the orbital plane of the spacecraft and the ecliptic plane of the Earth.

Consider an inertial reference frame S_0. Consider also a non-rotating reference frame with origin on Earth's center, S_1, and a parallel one with origin on Moon's center, S_2. The position of our spacecraft P satisfies:

$$\boldsymbol{r}_2^P = \boldsymbol{r}_1^P - \boldsymbol{r}_1^{\leftmoon}. \tag{3.8}$$

Remember that, while S_1 is an ECI reference frame, this frame is not strictly inertial: in fact, it suffers the acceleration of the origin due to Moon's attraction of the Earth,

$$\boldsymbol{a}_0^{\oplus} = \frac{\mu_{\leftmoon}}{(r_1^{\leftmoon})^3}\boldsymbol{r}_1^{\leftmoon} \tag{3.9}$$

This must be taken into account as an inertial acceleration to be subtracted when expressing the spacecraft acceleration in S_1,

$$\boldsymbol{a}_1^P = -\mu_\oplus \frac{\boldsymbol{r}_1^P}{(r_1^P)^3} - \mu_{\leftmoon}\left(\frac{\boldsymbol{r}_2^P}{(r_2^P)^3} + \frac{\boldsymbol{r}_1^{\leftmoon}}{(r_1^{\leftmoon})^3}\right). \tag{3.10}$$

Using equation (3.8) and the cosine theorem it is easy to write:

$$r_2^P = r_1^{\leftmoon}\sqrt{1 - 2\varepsilon\cos\beta + \varepsilon^2}, \tag{3.11}$$

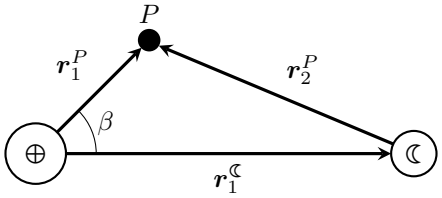
where $\varepsilon = r_1^P / r_1^{\leftmoon}$ and β is the angle between $\boldsymbol{r}_1^{\leftmoon}$ and $\boldsymbol{r}_1^P$ as defined in the figure below. When $\varepsilon \ll 1$, we can take the first-order approximation

$$r_2^P \simeq r_1^{\leftmoon}(1 - \varepsilon\cos\beta). \tag{3.12}$$

In this case, the expression of $\boldsymbol{a}_1^P$, keeping only first-order terms in ε, simplifies to

$$\boldsymbol{a}_1^P = -\left(\frac{\mu_\oplus}{(r_1^P)^3} + \frac{\mu_{\leftmoon}}{(r_1^{\leftmoon})^3}\right)\boldsymbol{r}_1^P + 3\varepsilon\frac{\mu_{\leftmoon}}{(r_1^{\leftmoon})^3}\cos\beta\boldsymbol{r}_1^{\leftmoon}. \tag{3.13}$$

The effect of the Moon therefore includes a small correction to the central acceleration term along $\boldsymbol{r}_1^P$, plus a *tidal* acceleration term proportional to $\cos\beta$. The net effect of this tidal term is to push the spacecraft "away" from the Earth when the spacecraft is in the Moon-Earth line. This term is maximal when the spacecraft is located directly between the Earth and Moon or directly behind the Earth. This is the term responsible for the ocean tides. The Sun plays an analogous effect.



The third-body perturbations become more relevant the farther away from the Earth, until eventually the study of the spacecraft motion as a two-body problem with the Earth as the central body stops being useful, and it is preferable to study the motion as a two-body problem with the Sun as the central body (including, then, the Earth as a perturbing third-body object).

The full expression of the gravitational acceleration in the presence of the Moon as a third-body can be obtained using Lagrange polynomials P_n:

$$\boldsymbol{a}_1^P = -\frac{\mu_\oplus}{(r_1^P)^3}\boldsymbol{r}_1^P + \frac{\mu_{\leftmoon}}{(r_1^{\leftmoon})^2}\sum_{n=1}^{\infty}\varepsilon^n\left[P'_{n+1}(\cos\beta)\frac{\boldsymbol{r}_1^{\leftmoon}}{r_1^{\leftmoon}} - P'_n(\cos\beta)\frac{\boldsymbol{r}_1^P}{r_1^P}\right]. \tag{3.14}$$

See Tewari's book for the derivation of this result.

The third-body perturbations may bring in periodic effects and secular effects on the trajectory of the spacecraft. Among the secular ones, it is common that the Sun and the Moon alter the right-ascension of the ascending node (i.e., the position of the line of nodes) and the argument of the pericenter.

3.3 Atmospheric drag

While very tenuous, the thin atmosphere at low Earth orbit altitudes exerts a continuous drag force on our satellite. The effect becomes the dominant perturbation at low altitudes, and in fact grows dramatically as the orbital height decreases below 300 km. As the drag force opposes the velocity vector, the net effect is to decrease the mechanical energy of the spacecraft, and thus its semi-major axis a.

Moreover, as the effect is larger near the pericenter than near the apocenter, drag tends to lower the apocenter first, circularizing the orbit and thus decreasing its eccentricity. This effect is exploited in aerocapture/aerobraking maneuvers in interplanetary probes, which use the target planet atmosphere to slow down and reach a quasi-circular orbit while minimizing the propellant consumption.

Modeling the drag force accurately is complex for several reasons. Firstly, the atmospheric density at high altitudes varies substantially with time. The major cause for this variation is solar illumination and solar activity: during the more intense years of solar activity the Earth's atmosphere warms up

and the density at higher altitudes increases. Secondly, the interaction of the thin atmosphere with the surface of the satellite occurs at very high Knudsen numbers, so the gas cannot be modeled accurately as a continuum and requires us to consider the free molecular flow interactions. Thirdly, the drag force depends on the geometry of the spacecraft and its current orientation or *attitude*. Thus, the trajectory problem becomes coupled with the attitude problem.

A crude model for the drag acceleration is to use a *drag coefficient* C_D,

$$\boldsymbol{a}_D = -\frac{1}{2}\frac{C_D S}{m^P}\rho V \boldsymbol{V}, \tag{3.15}$$

where $\boldsymbol{V}$ denotes the velocity of the spacecraft relative to the air, and S is the characteristic surface area of the spacecraft. If we neglect the velocity of the atmosphere in the ECI reference frame, $\boldsymbol{V} \simeq \boldsymbol{v}$, and introduce the *ballistic coefficient*,

$$B_c = \frac{m^P}{C_D S}, \tag{3.16}$$

then we can write

$$\boldsymbol{a}_D = -\frac{1}{2B_c}\rho v \boldsymbol{v}. \tag{3.17}$$

At high altitudes, the atmosphere is nearly-isothermal, and thus its density follows a near-exponential law:

$$\rho = \rho_0 \exp\left(-\frac{\zeta - \zeta_0}{H}\right), \tag{3.18}$$

where ρ_0 is the density at the reference height ζ_0, and H is a scale height.

The (negative) work per unit time and mass exerted on the spacecraft equals the rate of change of its mechanical energy, $\xi = -\mu/(2a)$,

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{\mu}{2a^2}\frac{\mathrm{d}a}{\mathrm{d}t} = \boldsymbol{a}_D \cdot \boldsymbol{v} = -\frac{\rho}{2B_c}v^3 \tag{3.19}$$

For the particular case of a circular orbit with $v = v_c = \sqrt{\mu/a}$, we can consider an analogous model to that of a low-thrust orbit raising (in section 2.4.1), except that in this case the force is a decelerating force.

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{\rho}{B_c}\sqrt{\mu a}. \tag{3.20}$$

Introducing the exponential atmosphere model, and noting that $\mathrm{d}a = \mathrm{d}\zeta$,

$$-\frac{B_c \mathrm{d}\zeta}{\rho_0\sqrt{\mu a}} \exp\left(\frac{\zeta - \zeta_0}{H}\right) = \mathrm{d}t. \tag{3.21}$$

If we integrate this equation between two instants of time, and approximate a as constant in this interval, we obtain the following analytical approximation to the time it takes our satellite to decrease its orbital height from ζ_1 to a lower ζ_2 due to drag:

$$\Delta t = \frac{H B_c}{\sqrt{\mu a}}\left(\frac{1}{\zeta_1} - \frac{1}{\zeta_2}\right). \tag{3.22}$$

The spacecraft will spiral down until the atmosphere is so thick that it will either burn up or precipitate to Earth's surface, at which point this model stops being valid. Clearly, the higher the ballistic coefficient, the longer it will take for the spacecraft to reenter on Earth's atmosphere.

3.4 Solar radiation pressure

The photons in sunlight not only carry energy but also momentum, according to the expressions

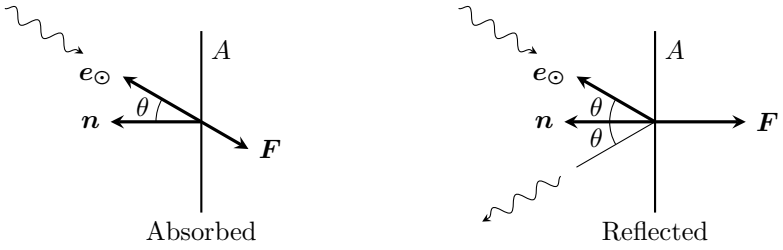
$$E = h\nu; \qquad p = \frac{h\nu}{c}, \tag{3.23}$$

where h is Planck's constant, ν is the photon frequency, and c is the speed of light.

At one astronomical unit (i.e., the mean Earth-Sun distance), the average flux of power per unit area is solar radiation is $S_\odot = 1367$ W/m^2. This means the average flux of momentum per unit time and unit area is $P_\odot = 4.560 \cdot 10^{-6}$ Nm^2. This is a tiny amount of pressure, but it can create a non-negligible perturbation force (and torque) on the spacecraft, especially if it features large surfaces like e.g. solar arrays.

Let us introduce the unit sun vector $\boldsymbol{e}_\odot$, which points from our spacecraft to the Sun. For a flat, specular surface of area A and external normal unit vector $\boldsymbol{n}$ that forms an angle θ with $\boldsymbol{e}_\odot$ and reflects a fraction R of the received light symmetrically with respect to its normal, the net force due to solar radiation pressure is:

$$\boldsymbol{F} = -P_\odot A \cos\theta[(1-R)\boldsymbol{e}_\odot + 2R\cos\theta\boldsymbol{n}] \tag{3.24}$$



The net effect of solar radiation pressure is hard to assess, but it may involve periodic as well as secular effects on the trajectory of the spacecraft.

3.5 General perturbations

TBW

4 Initial orbit determination

Orbit determination is a set of techniques that allow us to infer the orbital elements of a spacecraft from a number of observations. In real life, the available observations typically include azimuth and elevation measurements from a ground station, and the rates at which these parameters change with time. It is also common to have range and range-rate measurements available. In all cases, it should be clear in which reference frame we are obtaining the measurements, be it in ECEF, ECI, or a topocentric (station-centered) Earth-fixed frame.

In general, multiple measurements are taken over time, each with a characteristic error. The goal of orbit determination is to find the "best" approximation to the true orbital parameters of the spacecraft with this information. This is done with mathematical filters of different types (e.g. Kalman filter).

Initial orbit determination, on the other hand, uses only a *minimal set* of parameters to compute the orbital elements. Since there are 6 degrees of freedom to be determined (6 COEs, or the 6 components in the $\boldsymbol{r}$ and $\boldsymbol{v}$ vectors), a set of 6 independent pieces of information are necessary to perform the initial orbit determination. The schemes used to do this initial determination are quick and are often used to provide a first guess to the algorithms of advanced orbit determination.

In a way, many of the problems that we have solved in the first chapters of this course can be categorized within the initial orbit determination class. The algorithms that we will learn in this chapter are also used in mission analysis, to determine the best orbit to go from one position to another. This is the case, in particular, of Lambert's problem. Lambert's problem is specially important as its solution applies not only to initial orbit determination, but also to problems of intercept, rendezvous, and targeting.

4.1 From a single, full observation

Consider a ground station capable of determining the range ρ, azimuth β and elevation ε, as well as the rates $\dot{\rho}$, $\dot{\beta}$, and $\dot{\varepsilon}$. With these 6 pieces of information, it is possible to determine $\boldsymbol{r}$ and $\boldsymbol{v}$ or the COEs directly, without any special procedure other than simple vector calculus.

We introduce the ECI (S_0), ECEF (S_1), and topocentric (S_2) reference frames. The latter is centered at the ground station, and the unit vectors $\boldsymbol{i}_2, \boldsymbol{j}_2, \boldsymbol{k}_2$ of its vector basis B_2 points South, East, and to the Zenith respectively. This is why this reference frame is also known as SEZ. The ground station S position in ECEF, $\boldsymbol{r}_1^S$, is known in terms of its longitude λ, geodetic latitude ϕ, and radius r_S (Earth radius plus station altitude),

$$\boldsymbol{r}_1^S = r_S \cos\phi \cos\lambda \boldsymbol{i}_1 + r_S \cos\phi \sin\lambda \boldsymbol{j}_1 + r_S \sin\phi \boldsymbol{k}_1. \tag{4.1}$$

The observation allows to locate the satellite P in S_2 in terms of its range, the azimuth with respect to the North pole (measured toward the East) and its elevation from the local horizontal:

$$\boldsymbol{r}_2^P = -\rho \cos\varepsilon \cos\beta \boldsymbol{i}_2 + \rho \cos\varepsilon \sin\beta \boldsymbol{j}_2 + \rho \sin\varepsilon \boldsymbol{k}_2. \tag{4.2}$$

Likewise, the relative velocity in S_2 is known in terms of the rates:

$$\begin{aligned} \boldsymbol{v}_2^P = {} & [-\dot{\rho}\cos\varepsilon\cos\beta + \rho\dot{\varepsilon}\sin\varepsilon\cos\beta + \rho\dot{\beta}\cos\varepsilon\sin\beta]\boldsymbol{i}_2 \\ & + [\dot{\rho}\cos\varepsilon\sin\beta - \rho\dot{\varepsilon}\sin\varepsilon\sin\beta + \rho\dot{\beta}\cos\varepsilon\cos\beta]\boldsymbol{j}_2 \\ & + [\dot{\rho}\sin\varepsilon + \rho\dot{\varepsilon}\sin\varepsilon]\boldsymbol{k}_2. \end{aligned} \tag{4.3}$$

To obtain the position and velocity of P in ECI (S_0), we use the formulas that relate relative with absolute quantities. We need to express all vectors in the same basis before operating (ideally, in the basis B_0 of S_0):

$$\boldsymbol{r}_0^P = \boldsymbol{r}_1^P = \boldsymbol{r}_1^S + \boldsymbol{r}_2^P, \tag{4.4}$$

$$\boldsymbol{v}_0^P = \boldsymbol{v}_2^P + \boldsymbol{\omega}_{10} \times (\boldsymbol{r}_1^S + \boldsymbol{r}_2^P), \tag{4.5}$$

where $\boldsymbol{\omega}_{10}$ is the (sidereal) angular velocity of the Earth. Once $\boldsymbol{r}_0^P$ and $\boldsymbol{v}_0^P$ are known, the standard procedure to determine the COEs can be applied.

4.2 General approach

We often do not know all aspects of a single observation, and more than one observation (at different times) or station needs to be used for the initial orbit determination. In general, however, it is possible to find from the observations a unit vector $\boldsymbol{w}$ perpendicular to the orbital plane that goes in the same direction as the angular momentum, $\boldsymbol{h}$.

1. If we know a position vector and a velocity vector in ECI, the cross product gives $\boldsymbol{h}$ directly

2. If we know two or more position vectors, their cross product can be used to compute $\boldsymbol{w}$. The positive direction of this vector can be chosen so that it is parallel (instead of antiparallel) with $\boldsymbol{h}$, i.e., so that the satellite orbital motion goes counterclockwise with respect to $\boldsymbol{w}$.

Once we know $\boldsymbol{w}$, we know the orbital plane, and we can easily compute Ω and i, the first two Euler angles of the COE.

Fully determining the orbit amounts then to finding the location of the pericenter on this orbital plane with ω, and two parameters of the conic section such as a and e. Finally, the value of the true anomaly at one observation point, θ, can be computed if needed. These parameters can be found using e.g. the trajectory equation, particularized at each observation point. For example, for three observations:

$$r_1 = \frac{p}{1 + e\cos\theta_1}; \qquad r_2 = \frac{p}{1 + e\cos(\theta_1 + \Delta\theta_{12})}; \qquad r_3 = \frac{p}{1 + e\cos(\theta_1 + \Delta\theta_{13})}. \tag{4.6}$$

This is a nonlinear system of equations with three unknowns: p, e, and θ_1. Once θ_1 is known, we know the direction of the eccentricity vector $\boldsymbol{e}$, and with it we can find the angle ω.

The equations to use in each case may vary, depending on the information that is available for us to make the initial orbit determination. While this approach is valid, there are better solution schemes available that we can use depending on the type of data that we have. The following sections provide two important examples for this, but are not the only ones[6]. The following relations are useful in many of these developments:

1. Remember the eccentricity vector

$$\boldsymbol{e} = \frac{\boldsymbol{v} \times \boldsymbol{h}}{\mu} - \frac{\boldsymbol{r}}{r}. \tag{4.7}$$

   If we dot-multiply by $\boldsymbol{r}$ the result is

$$\boldsymbol{e} \cdot \boldsymbol{r} = p - r \tag{4.8}$$

2. Whenever we have more than two observations, the vectors should all be coplanar in ECI. Hence, there should be a linear combination of them that yields zero:

$$a_1\boldsymbol{r}_1 + a_2\boldsymbol{r}_2 + a_3\boldsymbol{r}_3 + \ldots = \boldsymbol{0} \tag{4.9}$$

Observe that we only need 6 scalar pieces of information: any additional data is redundant, and in the worst case, incompatible. The incompatibility of the data can be due to measurement errors, and also due to perturbations that change the orbit slightly between successive observations. Whenever more data is available than needed, we speak of general orbit determination (not initial orbit determination), and we use techniques such as least squares to solve the overdetermined problem, or filters to update the orbit determination as new data becomes available.

[6] *Rabbithole*: This is actually a field of astrodynamics full of history. The interested student is directed to chapter 7 in Vallado's book [Vallado, 2001] for an overview of Laplace's and Gauss's techniques for angles-only observations. Gauss's technique, in particular, uses the f and g functions introduced in chapter 1.

4.3 Gibbs' Problem

Consider that we have reconstructed three position vectors of the satellite in ECI $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ at different observation times t_1, t_2, t_3. Gibbs' method for preliminary orbit determination employs vector calculus and is best suited to cases where $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ are sufficiently apart from each other, i.e., when the angle between them is not small.

Input data: three sequential position vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$. Coplanar, but not collinear.

We begin by writing the central measurement $\boldsymbol{r}_2$ as a linear combination of the other two:

$$\boldsymbol{r}_2 = c_1\boldsymbol{r}_1 + c_3\boldsymbol{r}_3. \tag{4.10}$$

Dotting this equation with $\boldsymbol{e}$, using equation (4.8), we obtain:

$$p - r_2 = c_1(p - r_1) + c_3(p - r_3) \tag{4.11}$$

From here, knowing c_1 and c_3 we can solve for p. However, there is an elegant way to solve for p that avoids computing c_1 and c_3 from equation (4.10). If we take equation (4.11) and multiply it by $\boldsymbol{r}_3 \times \boldsymbol{r}_1$,

$$0 = c_1\boldsymbol{r}_3 \times \boldsymbol{r}_1(p - r_1) - \boldsymbol{r}_3 \times \boldsymbol{r}_1(p - r_2) + c_3\boldsymbol{r}_3 \times \boldsymbol{r}_1(p - r_3). \tag{4.12}$$

Crossing equation (4.10) with $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$ yields:

$$\boldsymbol{r}_2 \times \boldsymbol{r}_1 = c_3\boldsymbol{r}_3 \times \boldsymbol{r}_1, \qquad \boldsymbol{r}_2 \times \boldsymbol{r}_3 = -c_1\boldsymbol{r}_3 \times \boldsymbol{r}_1, \tag{4.13}$$

Eliminating c_1 and c_3 above yields

$$0 = \boldsymbol{r}_2 \times \boldsymbol{r}_3(p - r_1) + \boldsymbol{r}_3 \times \boldsymbol{r}_1(p - r_2) + \boldsymbol{r}_1 \times \boldsymbol{r}_2(p - r_3). \tag{4.14}$$

Rearranging:

$$\boldsymbol{N} = p\boldsymbol{D}, \tag{4.15}$$

where

$$\boldsymbol{N} = r_1(\boldsymbol{r}_2 \times \boldsymbol{r}_3) + r_2(\boldsymbol{r}_3 \times \boldsymbol{r}_1) + r_3(\boldsymbol{r}_1 \times \boldsymbol{r}_2). \tag{4.16}$$

$$\boldsymbol{D} = \boldsymbol{r}_2 \times \boldsymbol{r}_3 + \boldsymbol{r}_3 \times \boldsymbol{r}_1 + \boldsymbol{r}_1 \times \boldsymbol{r}_2; \tag{4.17}$$

Observe that $\boldsymbol{N}$ and $\boldsymbol{D}$ are parallel and point in the same direction. The value of p is found simply by taking the magnitude of previous equation, $p = N/D$.

Moreover, this direction is perpendicular to the orbital plane, so we can easily compute a unit vector in this direction $\boldsymbol{w} = \boldsymbol{D}/D$. To complete a perifocal vector basis, we need also the unit vectors

$$\boldsymbol{p} = \boldsymbol{e}/e, \qquad \boldsymbol{q} = \boldsymbol{w} \times \boldsymbol{p}, \qquad \boldsymbol{w} = \boldsymbol{D}/D. \tag{4.18}$$

To find $\boldsymbol{q}$, we compute the vector $\boldsymbol{S}$ with the following cross product, making use of the relation $(\boldsymbol{A} \times \boldsymbol{B}) \times C = (\boldsymbol{C} \cdot \boldsymbol{A})\boldsymbol{B} - (\boldsymbol{C} \cdot \boldsymbol{B})\boldsymbol{A}$:

$$De\boldsymbol{q} = \boldsymbol{D} \times \boldsymbol{e} = (r_2 - r_3)\boldsymbol{r}_1 + (r_3 - r_1)\boldsymbol{r}_2 + (r_1 - r_2)\boldsymbol{r}_3 = \boldsymbol{S} \tag{4.19}$$

From here, the eccentricity can be found as $e = S/D$. Provided that $\boldsymbol{S}$ is not zero, then the pericenter direction is given by

$$\boldsymbol{p} = \boldsymbol{q} \times \boldsymbol{w} = \frac{1}{SD}\boldsymbol{S} \times \boldsymbol{D}. \tag{4.20}$$

To be rigorous, it is important to note that the direction of $\boldsymbol{D}$ and $\boldsymbol{S}$ depend on the true anomaly difference between $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, between $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$, and between $\boldsymbol{r}_3$ and $\boldsymbol{r}_1$. If these angles are all $< \pi$, then the direction of $\boldsymbol{D}$ coincides with that of $\boldsymbol{h}$, otherwise it is the opposite. While not indispensable, this possible change of sign should be taken into account to define $\boldsymbol{w}$ and $\boldsymbol{q}$ as in chapter 1.

If we want to compute the velocity vector for one or more of the measurements, we can compute it directly with the eccentricity vector (equation (4.7)), taking the cross product with $\boldsymbol{h}$:

$$\frac{1}{\mu}\boldsymbol{v} \times \boldsymbol{h} = \frac{\boldsymbol{r}}{r} + \boldsymbol{e} \quad \Rightarrow \quad \frac{h^2}{\mu}\boldsymbol{v} = \frac{\boldsymbol{h} \times \boldsymbol{r}}{r} + \boldsymbol{h} \times \boldsymbol{e} \quad \Rightarrow \quad \boldsymbol{v} = \frac{1}{r}\sqrt{\frac{\mu}{ND}}\boldsymbol{D} \times \boldsymbol{r} + \sqrt{\frac{\mu}{ND}}\boldsymbol{S}. \tag{4.21}$$

where $\boldsymbol{h} \times (\boldsymbol{h} \times \boldsymbol{v}) = h^2\boldsymbol{v}$ has been used. Note that the sign of $\boldsymbol{v}$ in this equation is affected by the aforementioned considerations on the sign of $\boldsymbol{D}$ and $\boldsymbol{S}$.

4.4 Lambert's problem

It is often the case that we must determine the orbit from two position vectors and the time of flight between them. This is the two-boundary value problem in Keplerian motion, and is often referred to as Lambert's problem. Many algorithms have been proposed to solve it (and new continue to be developed). The Lambert's problem is also useful in the mission analysis of rendezvous trajectories and interplanetary orbital transfers, where it is necessary not only to change orbit but also to arrive at destination at a particular time (i.e. when the target planet is there).

Input data: two position vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ and the time of flight between them Δt.

We begin with the obvious. If $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are part of the sought orbit, as long as they are not co-linear (more on this later) a vector perpendicular to the orbital plane is $\boldsymbol{k} = \boldsymbol{r}_1 \times \boldsymbol{r}_2 / |\boldsymbol{r}_1 \times \boldsymbol{r}_2|$.

The difference in true anomaly between the two positions, positive if the rotation is in the direction of $\boldsymbol{k}$, is

$$\Delta\theta = \arccos\left(\frac{\boldsymbol{r}_1 \cdot \boldsymbol{r}_2}{r_1 r_2}\right) + 2\pi m, \tag{4.22}$$

where the arccos function ranges between 0 and π. When $m = 0$, we have the so-called direct or **short arc solution** of the problem. When $m = -1$, we have the so-called **long arc solution**. Other values of m imply complete "waiting" revolutions in the direction of $\boldsymbol{k}$ ($m > 0$) or in the direction of $-\boldsymbol{k}$ ($m < -1$).

The key insight of the problem must be understood is that, given two position vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, there is a one-parameter family of conic sections that passes by them. Our sought solution is one of these conics, for which the flight time between the two positions is exactly Δt.

The solution algorithm described next is due to Avanzini [Avanzini, 2008]. We introduce the chord vector

$$\boldsymbol{c} = \boldsymbol{r}_2 - \boldsymbol{r}_1, \tag{4.23}$$

and the vector basis (note that this is different from $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{w}$ in general):

$$\boldsymbol{i} = \boldsymbol{c}/c; \qquad \boldsymbol{j} = \boldsymbol{k} \times \boldsymbol{i}; \qquad \boldsymbol{k}. \tag{4.24}$$

We decompose the eccentricity vector in two parts, one along the chord direction $\boldsymbol{i}$ and one along $\boldsymbol{j}$:

$$\boldsymbol{e} = e_F \boldsymbol{i} + e_T \boldsymbol{j}. \tag{4.25}$$

The projection of the eccentricity vector $\boldsymbol{e}$ along the chord vector is constant:

$$\boldsymbol{e} \cdot (\boldsymbol{r}_2 - \boldsymbol{r}_1) = \boldsymbol{e} \cdot \boldsymbol{c} = e_F c = -(r_2 - r_1). \tag{4.26}$$

Thus, one component of the eccentricity vector in this basis, e_F, is constant and known:

$$e_F = -\frac{r_2 - r_1}{c} \tag{4.27}$$

(note that this component can be positive or negative). The family of conics depends on the other free component of the eccentricity vector, e_T. The eccentricity of the orbit is $\sqrt{e_F^2 + e_T^2}$ and the minimum possible eccentricity is $e = |e_F| \leqslant 1$. This minimum eccentricity case corresponds to the so called *fundamental ellipse*. In this fundamental ellipse, the chord $\boldsymbol{c}$ is parallel to the major axis, and from the symmetry of this ellipse we find

$$a_F = \frac{r_1 + r_2}{2}; \qquad p_F = a_F(1 - e_F^2). \tag{4.28}$$

It can be shown that a general conic section passing by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ has[7]:

$$p = p_F - e_T \frac{r_1 r_2}{c} \sin(\Delta\theta) \tag{4.29}$$

[7] *Rabbithole*: The derivation is not trivial and involves geometry and vector calculus. A good derivation can be found on [Battin, 1999], section 6.3.

And the semimajor axis obeys $a = p/(1 - e_F^2 - e_T^2)$. This allows parametrizing the family of conic sections in terms of e_T. The main advantages of this parametrization are that (1) as e_T varies from $-\infty$ to $+\infty$ we run along all possible conics, and (2) the time of flight for the long or the short arc solutions are monotonic on e_T, which facilitates the search for the solution to the Lambert problem. Figure 4.1 shows this family of conics.

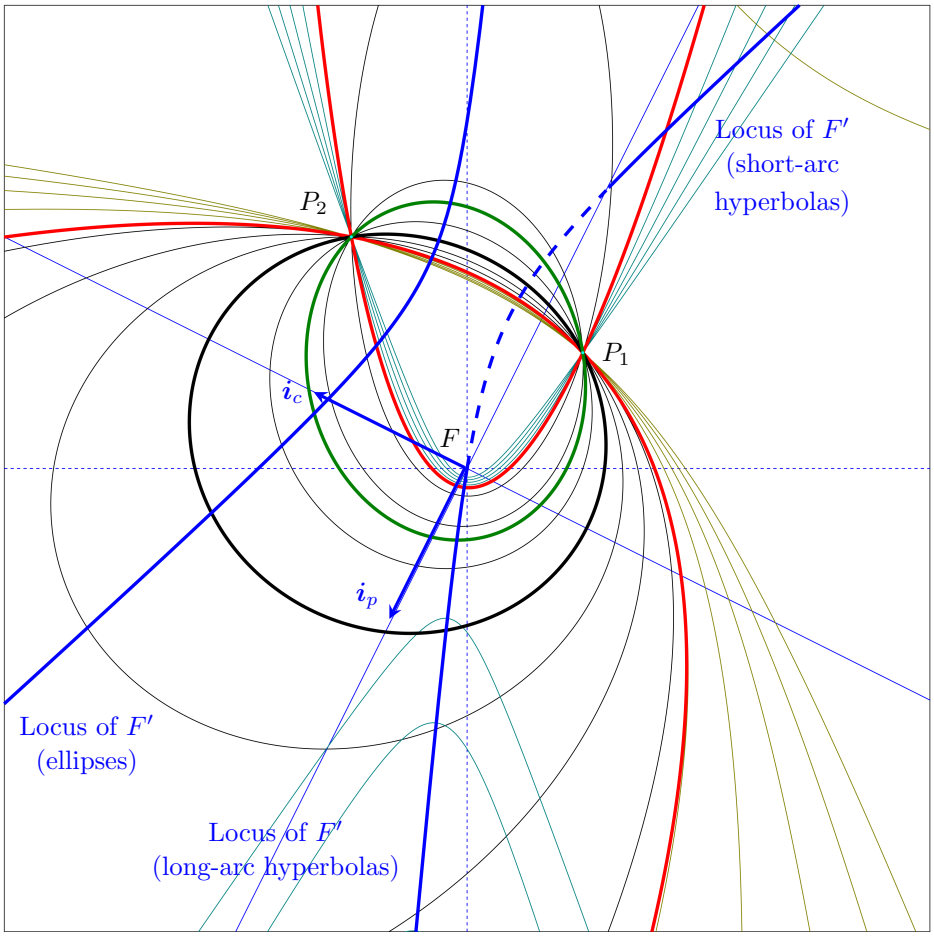


Figure 4.1: Family of conic sections in Lambert's problem. The attracting center (primary focus) is at the center of the figure. P_1 and P_2 are easily identifiable as the intersection points of all conic sections. Ellipses are drawn in black; the two possible parabolas in red; hyperbolas in light blue and light green. Note that ellipses are valid conics for short arc and long arc solutions, whereas parabolas and hyperbolas only provide one of the two. The fundamental ellipse is thick black. The minimum energy ellipse is thick green. The blue hyperbola denotes the locus of points for the secondary focus. The dashed part of this hyperbola corresponds to inadmissible hyperbolas (the arc between P_1 and P_2 occurs in the forbidden branch of the hyperbola, the repulsive one). The directions of $\boldsymbol{i}_c$ and $\boldsymbol{i}_p$ are shown.

The last step is to bring in the information on the time of flight. This is done with Kepler's equation for the elliptic, parabolic, or hyperbolic case, depending on the value of e. Note that the pericenter (and thus the origin for the angle θ) is in the direction of the $\boldsymbol{e}$ vector, which changes as e_T is varied. The valid range of e_T is not all the real line. Defining the limit values $e_{TP} = \sqrt{1 - e_F^2}$ and $e_{TD} = p_F c/(r_1 r_2 \sin\Delta\theta)$, and assuming $\sin\Delta\theta > 0$, we find the following:

1. For $e_T \in [-e_{TP}, +e_{TP}]$, the conic sections are ellipses. Both long arc and short arc solutions exist.

2. For $e_T = -e_{TP}$ we have a parabola that allows only a short arc solution.

3. Similarly, $e_T = +e_{TP}$ gives a parabola that allows only a long arc solution.
4. For $e_T < -e_{TP}$ we have hyperbolic short arcs.
5. For $e_T \in [+e_{TP}, e_{TD}]$ we have hyperbolic long arcs that pass very close to the attracting center F.
6. For $e_T = e_{TD}$ the solution degenerates with $p = 0$ and $F \equiv F'$ into a pair of lines that intersect at the origin.
7. For $e_T > e_{TD}$ we have inadmissible hyperbolic solutions, where the arc that joins P_1 and P_2 is the unoccupied branch of the hyperbola (i.e., the "repulsive" branch).

One must check that the selected arc does not hit the Earth. This is specially worrysome for the higher values of e_T (the last ones in the list above). Figure 4.2 shows the time of flight as a function of e_T, from $-e_{TP}$ to e_{TP} (elliptic cases). The long arc and the short arc solutions are both shown. The horizontal line represents the desired time-of-flight, for which there are two solutions. For higher times of flight, other solutions exist that include full revolutions in the same orbit. These solutions differ from the short arc and long arc in integer orbital periods, τ.

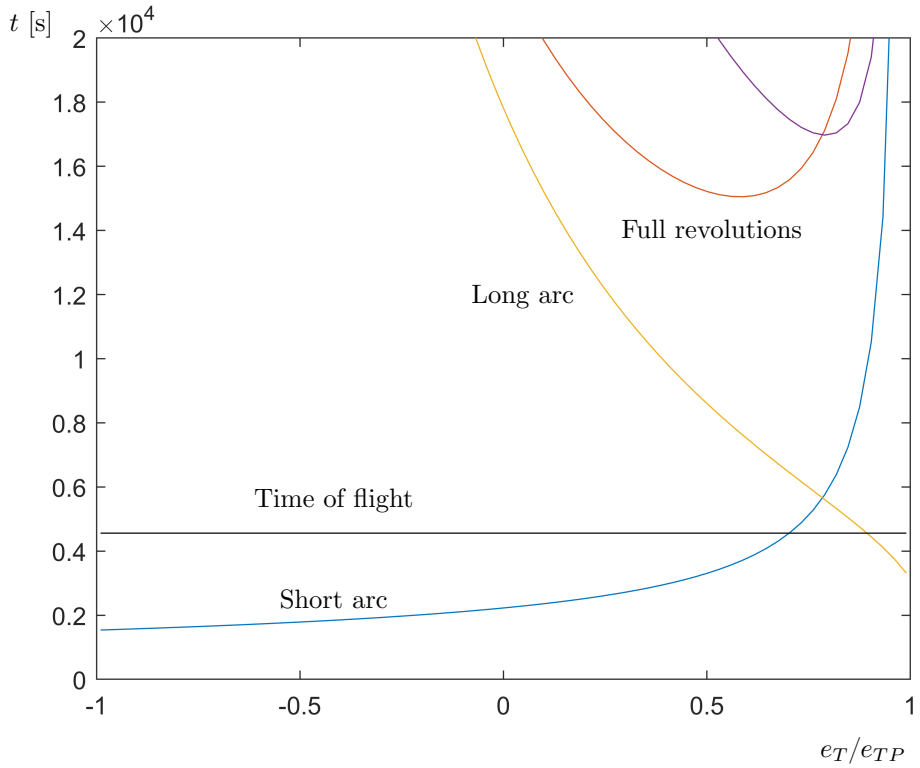


Figure 4.2: Time of flight for the short arc (blue) and long arc (yellow) solutions as a function of e_T, for the elliptical cases, in an example Lambert problem. For a given time of flight (e.g. the horizontal line), there are two solutions. The same conic sections, but with one additional full revolution, also appear as valid solutions at larger times of flight (red and purple for the short arc direction and the long arc direction, respectively).

4.4.1 Minimum energy transfer

Lambert also found the orbit of minimum semi-major axis a, which is the ellipse with minimum mechanical energy. Note that, in general:

1. This ellipse is not the fundamental ellipse (minimum eccentricity)

2. When used as a transfer orbit, this minimum energy ellipse does not imply minimum flight time between the two points P_1 and P_2, nor minimum Δv.

We begin with the realization that the sum of the distances from a point of an ellipse to the two foci, F and F', is constant and equal to $2a$, the length of the major axis. The focus F is the attracting center, so its location is known; the focus F', on the other hand, is unknown. Fixing the location of this second focus completely determines the conic section.

For a given value of a, F' lies on a circle of radius $2a - r_1$ from P_1, and on a circle of radius $2a - r_2$ from P_2. The intersection between these two circles defines the locus of points for F': this happens to be one branch of an hyperbola of eccentricity equal to $1/e_F$. The other branch is associated to hyperbolic solutions of the Lambert problem. There are three possibilities:

1. For large a, there are two possible intersections between these two circles.
2. For $a = a^*$, the two circles are tangent and there is only one possible solution for F'.
3. No intersections for $a < a^*$.

The value $a = a^*$ that makes the circles tangent is therefore the minimum semimajor axis, and corresponds to the minimum energy orbit. This minimum happens when F' is on the chord line that joins P_1 with P_2. Using the chord c and the semiperimeter of the FP_1P_2 triangle s:

$$c = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \Delta\theta}, \tag{4.30}$$

$$s = \frac{r_1 + r_2 + c}{2}, \tag{4.31}$$

we can write:

$$c = (2a^* - r_1) + (2a^* - r_2) \quad \Rightarrow \quad a^* = \frac{s}{2}. \tag{4.32}$$

Knowing that the distance between foci is $2ea$, it is possible to obtain the value of the eccentricity of this minimum energy ellipse.

5 Interplanetary flight

Interplanetary flight is an example of the N body problem: the effect of gravity of at least three celestial bodies (Earth, the Sun, and the destination planet) must be taken into account. Additionally, one or more intermediate fly-bys with gravitational assist maneuvers at auxiliary celestial bodies may be carried out to save propellant. Designing and optimizing the trajectory of an interplanetary mission is therefore a complex task.

Fortunately, the motion of the spacecraft is dominated by the gravity of just a single body for long parts of the journey. This enables a simplified analysis wherein the trajectory is divided into segments, each of them treated as a different two-body problem (potentially with perturbations due to the other bodies). This is known as the *patched conics method*, and is commonly used to get a first estimate on the cost and time of a potential trajectory and do quick assessments and trade-offs. In later stages of mission analysis more refined models (including the numerical integration of the N body problem) are used.

Problem statement: Analyze and design the trajectory of a spacecraft in the Solar System.

Assumptions: All bodies in the problem are treated as point particles; the scale of the heliocentric problem is assumed to be much larger than the planet problems. Most of the mission time happens in the heliocentric problem and the duration of the departure and arrival problems will be neglected. The motion of the spacecraft will be treated as a set of two-body problems with respect to the planets and the Sun.

Reference frames: We denote the Sun as $\odot$. Assuming the departure and target planets are Earth and Mars we denote them as $\oplus$ and $\mars$, respectively (we will change these as needed if the planets are others). The spacecraft will be denoted as P. We define:

- A non-rotating basis $B : \{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$, with $\boldsymbol{k}$ perpendicular to the ecliptic plane.
- The reference frames $S_0 : \{\odot; B\}$, $S_1 : \{\oplus; B\}$ and $S_2 : \{\mars; B\}$. In practice, S_0 can be considered inertial, and S_1 and S_2 can be considered pseudo-inertial when describing the motion of the spacecraft in the neighborhood of these planets.
- The polar unit vectors $\boldsymbol{u}_r$, $\boldsymbol{u}_\theta$ with respect to the Sun in the plane defined by $\boldsymbol{i}$, $\boldsymbol{j}$.

5.1 Sphere of influence

To formulate the patched conics method, we want to assess when it is preferable to describe the motion of spacecraft P as a two-body problem about the planet as the central body, with Sun's gravity as a third-body perturbation, and when it is preferable to describe the motion of P as a two-body problem about the Sun with the planet as a perturbation instead. Since the Sun is many times more massive than any of the planets in the solar system, we anticipate that the region where the former is preferred will be a small region in the neighborhood of the planet, and that the later will be preferred in the vast majority of the solar system.

The two-body problem equation of motion about a planet (e.g. planet Earth), considering the Sun as a third-body perturbation, is:

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}_1^P}{\mathrm{d}t^2}\right|_1 = -\mu_\oplus \frac{\boldsymbol{r}_1^P}{(r_1^P)^3} - \mu_\odot \left(\frac{\boldsymbol{r}_0^P}{(r_0^P)^3} - \frac{\boldsymbol{r}_0^\oplus}{(r_0^\oplus)^3}\right) \simeq -\mu_\oplus \frac{\boldsymbol{r}_1^P}{(r_1^P)^3} - \mu_\odot \frac{\boldsymbol{r}_1^P}{(r_0^\oplus)^3}. \tag{5.1}$$

Likewise, the two-body problem equation of motion about the Sun, considering the planet as a third-body perturbation, is:

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^P}{\mathrm{d}t^2}\right|_0 = -\mu_\odot \frac{\boldsymbol{r}_0^P}{(r_0^P)^3} - \mu_\oplus \left(\frac{\boldsymbol{r}_1^P}{(r_1^P)^3} - \frac{\boldsymbol{r}_1^\odot}{(r_1^\odot)^3}\right) \simeq -\mu_\odot \frac{\boldsymbol{r}_0^\oplus}{(r_0^\oplus)^3} - \mu_\oplus \frac{\boldsymbol{r}_1^P}{(r_1^P)^3}. \tag{5.2}$$

In both cases, the approximation of the right hand side has been carried out taking into account that $\boldsymbol{r}_0^P = \boldsymbol{r}_0^\oplus + \boldsymbol{r}_1^P$, accepting that $r_1^P \ll r_0^\oplus$, and keeping only zeroth-order terms in the expansion in $r_1^P/r_0^\oplus$. The ratio of the magnitude of the perturbing acceleration to the primary acceleration in each case is, respectively:

$$\frac{\text{Perturbation}}{\text{Primary}} \simeq \frac{\mu_\odot (r_1^P)^3}{\mu_\oplus (r_0^\oplus)^3} \qquad \text{and} \qquad \frac{\text{Perturbation}}{\text{Primary}} \simeq \frac{\mu_\oplus (r_0^\oplus)^2}{\mu_\odot (r_1^P)^2} \tag{5.3}$$

We define the *sphere of influence of the planet* (Earth in the example) as the sphere centered on the planet where the perturbation-to-primary acceleration ratios in the two-body problem about the planet and in the two-body problem about the Sun are roughly equal. The radius of this sphere is:

$$\boxed{R_{\mathrm{SOI},\oplus} = r_0^{\oplus}\left(\frac{\mu_\oplus}{\mu_\odot}\right)^{2/5},} \tag{5.4}$$

which increases the larger the mass of the planet, and the further away it is from the Sun. In the cases of interest, the sphere of influence is always very small compared to the scale of the solar system. The concept of sphere of influence helps us decide when to switch from a two-body problem formulation to another; it does not, however, provide an estimate of the error committed when the perturbation accelerations are neglected. To assess the error, one should evaluate the ratios in equation (5.3) directly.

In practice, the error committed is small for preliminary mission design. An exception to this are Earth-Moon missions (in this case, the Earth takes the role of the Sun, and the Moon the role of the planet): in the Earth Moon system, the sphere of influence of the Moon is not small in the scale of the full system problem, and the error committed by treating the motion as two-body problems can be large. For these reasons, cislunar problems are better treated with a *three-body problem* formulation.

5.2 Heliocentric problem

In the *patched conics method*, interplanetary flight from one planet to another is divided in three parts: a first one called departure from the home planet (here, $\oplus$); a second one where the spacecraft performs a heliocentric arc; and a third one consisting on the arrival at the target planet (here, σ). The patched conics method is a *two-scale method*: planetary problems occur in a scale that is negligible in the scale of the heliocentric problem, i.e., the spheres of influence of the planets are tiny in the scale of the solar system and treated as a point. In contrast, in the scale of the planetary problems, the spheres of influence of the planets are huge and treated as infinitely large. The heliocentric part of the trajectory is always studied first, as it sets boundary conditions to the other two parts. The mission duration is approximately equal to the duration of the heliocentric transfer, and the departure and arrival problem durations are usually neglected in first approximation.

5.2.1 Restricted patched conics method

In the *restricted patched conics method* it is further assumed that all planets move in coplanar ($i = 0$), circular ($e = 0$, $r = a$) orbits, and that the heliocentric arc is a Hohmann transfer, i.e., half an ellipse that is tangent to the orbits of the two planets around the Sun. These simplifications make the preliminary analysis of a mission even easier. Later on, we will relax these assumptions and consider the actual orbits of the planets and other possible arcs (using Lambert's algorithm to determine the transfer).

Heliocentric Hohmann transfer

We will denote with subindex t all variables related to the heliocentric transfer ellipse, which has pericenter radius $r_{tp} = a_\oplus$, apocenter radius $r_{ta} = a_\sigma$, semi-major axis $a_t = (a_\oplus + a_\sigma)/2$, and eccentriciy $e_t = (a_\sigma - a_\oplus)/(a_\sigma + a_\oplus)$. The velocity of the spacecraft P with respect to S_0 at peri- and apocenter are:

$$\boxed{\boldsymbol{v}_{0,tp}^P = \sqrt{\frac{2\mu_\odot}{a_\oplus + a_\sigma}\left(\frac{a_\sigma}{a_\oplus}\right)}\boldsymbol{u}_\theta; \qquad \boldsymbol{v}_{0,ta}^P = \sqrt{\frac{2\mu_\odot}{a_\oplus + a_\sigma}\left(\frac{a_\oplus}{a_\sigma}\right)}\boldsymbol{u}_\theta.} \tag{5.5}$$

This is the velocity with respect to reference frame S_0 (absolute velocity), with which the spacecraft must leave Earth's SOI and the velocity with which it approaches Mars' SOI.

In the restricted patched conics method, the absolute velocity of the planets around the Sun is simply the circular velocity:

$$\boldsymbol{v}_0^\oplus = \sqrt{\frac{\mu_\odot}{a_\oplus}}\boldsymbol{u}_\theta; \qquad \boldsymbol{v}_0^\sigma = \sqrt{\frac{\mu_\odot}{a_\sigma}}\boldsymbol{u}_\theta. \tag{5.6}$$

To obtain the relative velocity of spacecraft P in S_1 and S_2, i.e. with respect to the planets, we must subtract these planet velocities to the absolute spacecraft velocity in S_0 (equations (5.5)).

Rendezvous opportunities

The relative position of Earth and Mars when the mission begins is important: we must launch knowing that Mars will be at the apocenter of the transfer orbit when the spacececraft reaches there. This determines the *rendezvous opportunities*.

Calling $n_\oplus$ and n_σ the (constant) angular rates of Earth and Mars about the Sun, the true anomaly of each planet increases linearly as:

$$\theta_\oplus(t) = \theta_\oplus(t_0) + n_\oplus(t - t_0), \tag{5.7}$$
$$\theta_\sigma(t) = \theta_\sigma(t_0) + n_\sigma(t - t_0). \tag{5.8}$$

The *relative phase angle* $\phi \equiv \theta_\sigma(t) - \theta_\oplus(t)$ of Mars as seen from Earth is:

$$\boxed{\phi(t) = \phi(t_0) + \Delta n(t - t_0),} \tag{5.9}$$

where $\Delta n = n_\sigma - n_\oplus$ (negative for this choice of planets). The time it takes for the relative configuration to repeat is known as the *synodic period*,

$$\boxed{\tau_s = \frac{2\pi}{|\Delta n|}.} \tag{5.10}$$

Since $a_\sigma > a_\oplus$, then $n_\sigma < n_\oplus$, and Mars lags behind the Earth over time. The opposite would be true if we had chosen e.g. Venus instead of Mars, as it lies closer to the Sun than Earth.

The duration of the transfer is half a period of the elliptic orbit:

$$\Delta t_t = \pi\sqrt{\frac{(a_\oplus + a_\sigma)^3}{8\mu_\odot}}. \tag{5.11}$$

Let us call t_1 and t_2 the instants of departure from the origin planet, arrival at the target planet, and likewise t_3 and t_4 the instants of departure and arrival for the return trip. The necessary condition for rendezvous is that at time $t_2 = t_1 + \Delta t_t$ Mars be 180 deg ahead of the initial position of Earth at t_1:

$$\theta_\sigma(t_2) = \theta_\oplus(t_1) + \pi. \tag{5.12}$$

Therefore, the initial relative phase angle of Mars as seen from Earth must be:

$$\phi(t_1) = \pi - n_\sigma \Delta t_t. \tag{5.13}$$

Upon arrival at Mars, the relative phase angle is

$$\phi(t_2) = \pi - n_\oplus \Delta t_t. \tag{5.14}$$

Later, to begin the return trip, we must ensure that Earth at t_4 will be 180 deg ahead of Mars at t_3:

$$\phi(t_3) = n_\oplus \Delta t_t - \pi. \tag{5.15}$$

The minimum wait time between t_2 and t_3 is Δt_w, given by

$$\phi(t_3) - \phi(t_2) = \Delta n \Delta t_w \quad \Rightarrow \quad \Delta t_w = 2\frac{n_\oplus}{\Delta n}\Delta t_t + \tau_s k \tag{5.16}$$

with the smallest k that makes $t_w \geqslant 0$.

5.2.2 Complete patched conics method

The *complete patched conics method* solves the 3D problem, including non-coplanar and non-circular planetary orbits, using Lambert's algorithm. The heliocentric transfer need not be a Hohmann transfer in this case.

As inputs we have the positions of the departure and arrival planets as a function of time, the desired time of departure, and the desired time of flight. Solving the heliocentric Lambert's problem allows finding the value of the velocities of the spacecraft at the beginning and the end of the heliocentric transfer, which provides the equivalent information to that of equations (5.5) in the restricted patched conics method.

The position of the planets as a function of time can of course be obtained by direct integration (a.k.a. orbital propagation), but it is usually easier to resort to planetary *ephemeris* databases such as NASA's JPL Horizons database.

Lambert's solution is typically used to trade off different launch dates with their corresponding cost (e.g. the characteristic hyperbolic energy c_3, or the total Δv once the departure and arrival problems are integrated into the solution) and travel time. Given a departure time and a duration of flight, it is possible to compute where the target planet will be at arrival date and solve for the cost of that transfer. This information is usually represented in so-called *porkchop plots*, name motivated by their resemblance with the piece of meat.

5.3 Planetary departure problem

We focus now on the departure from the home planet. In this scale, the sphere of influence of the planet is very large, and it can be considered to lie at infinity as a first approximation.

5.3.1 Escape hyperbola

We assume the spacecraft is waiting in a circular parking orbit (subindex c) of radius r_c around the planet, from where it will be injected into a hyperbolic escape trajectory (subindex e). The pericenter of this hyperbolic trajectory must coincide with r_c,

$$r_{ep} = a_e(1 - e_e) = r_c \tag{5.17}$$

and the excess hyperbolic velocity at infinity $\boldsymbol{v}_{1,eh}^P$ must equal the desired relative velocity in the heliocentric problem. If we have solved first the heliocentric problem, then we know the necessary value of $\boldsymbol{v}_{1,eh}^P$ to enter the desired arc.

In the *restricted* patched conics method, the condition is given by equations (5.5) and (5.6):

$$\boldsymbol{v}_{1,eh}^P = \boldsymbol{v}_{0,tp}^P - \boldsymbol{v}_0^{\oplus} \tag{5.18}$$

(the aphelion velocity $\boldsymbol{v}_{0,ta}^P$ would be used instead of $\boldsymbol{v}_{0,tp}^P$ in the case that the target planet were closer to the Sun than Earth). In the *complete* patched conics method, $\boldsymbol{v}_{1,eh}^P$ is also given by the difference of the absolute departure velocity from the Lambert solution minus the instantaneous absolute velocity of the planet.

When we go to a planet farther away from the Sun than Earth, we want to abandon Earth in its direction of motion around the Sun, so that Earth's velocity and the hyperbolic excess velocity add up (in order to travel to a planet closer to the Sun, we would leave Earth in the opposite direction). The opposite is true when traveling to more inner planets.

In the particular case of the restricted patched conics problem, the solution is:

$$v_{1,eh}^P = \sqrt{-\frac{\mu_\oplus}{a_e}} = \sqrt{\frac{\mu_\odot}{a_\oplus}}\left(\sqrt{\frac{2a_\sigma}{a_\oplus + a_\sigma}} - 1\right). \tag{5.19}$$

With these two equations we can solve for a_e and e_e, determining the shape of the escape hyperbola:

$$\boxed{a_e = -\frac{\mu_\oplus}{(v_{1,eh}^P)^2}; \qquad e_e = 1 + \frac{r_c(v_{1,eh}^P)^2}{\mu_\oplus}.} \tag{5.20}$$

5.3.2 Escape impulse

The speed of P in the circular parking orbit is

$$v_{1,cc}^{P} = \sqrt{\frac{\mu_{\oplus}}{r_c}}. \tag{5.21}$$

The speed at the periapsis of the hyperbola is

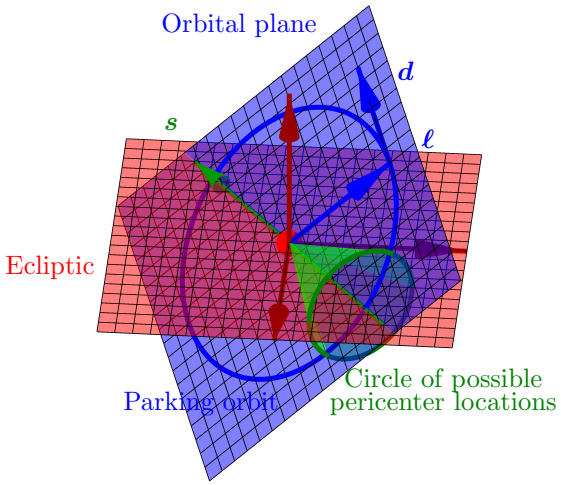
$$v_{1,ep}^{P} = \sqrt{\frac{\mu_{\oplus}}{a_e}\left(\frac{1+e_e}{1-e_e}\right)}. \tag{5.22}$$

Therefore, a tangential impulsive maneuver with $\Delta v = v_{1,ep}^{P} - v_{1,cc}^{P}$ is needed to start the interplanetary flight.

The resulting heliocentric transfer arc and the time of arrival are very sensitive to errors in the magnitude and direction of the escape impulse, Δv. Hence, one or several midcourse trajectory correction maneuvers (TCMs) are used to correct any deviations from the planned heliocentric arc. These maneuvers usually have a tiny Δv cost if executed long in advance of arrival at the target planet.

5.3.3 Launch into parking orbit

Let us call O the center of the Earth, C the center of the hyperbola, L the point on Earth surface from where the spacecraft is launched into the parking orbit, $\boldsymbol{s}$ a unit vector in the direction of $\boldsymbol{v}_{1,eh}^{P}$ (i.e., in the direction of the escape asymptote), $\boldsymbol{\ell}$ a unit vector that points from O to L, and $\boldsymbol{d}$ a unit vector in the direction of launch as seen from the ECI reference frame S_1. Note that the plane of the ecliptic is given by vectors $\{O; \boldsymbol{i}, \boldsymbol{j}\}$, and that the direction of vectors $\boldsymbol{\ell}$ and $\boldsymbol{d}$ depend on the local time of launch due to the rotation of Earth.



The angle β_e of the asymptote with respect to the major axis of the hyperbola must satisfy

$$\cos\beta_e = \frac{1}{e_e}. \tag{5.23}$$

Therefore, the major axis of the hyperbola must lie on the cone of half angle β_e and axis along $-\boldsymbol{s}$. This condition defines a set of possible impulse points at the parking orbit. The intersection of this cone with the sphere of radius r_c is a circle. The plane of the parking orbit, defined by vectors $\boldsymbol{\ell}$ and $\boldsymbol{d}$, must contain vector $\boldsymbol{s}$ so that the parking orbit passes by the center of this circle and the injection into the escape trajectory can be a tangential maneuver. This sets a condition on the direction of launch, $\boldsymbol{d}$, and/or the local time of launch.

A variant of this strategy is to launch directly into the escape hyperbola, without a parking orbit. In this case, the orbital plane defined by $\boldsymbol{\ell}$ and $\boldsymbol{d}$ still must contain $\boldsymbol{s}$.

5.4 Planetary arrival problem

Again, in the planetary problem scale, the sphere of influence is very large and can be considered to lie at infinity. The geometry of the arrival problem is identical to the escape problem; the only difference is that the spacecraft travels along the hyperbola branch in the opposite direction.

5.4.1 Arrival hyperbola

With respect to the target planet (Mars in our case), the spacecraft enters the sphere of influence in an arrival hyperbolic trajectory (subindex a). Similarly to the planetary departure problem, the spacecraft arrives with a hyperbolic excess velocity $\boldsymbol{v}^P_{2,ah}$ with respect to Mars.

In the *restricted* patched conics method, this velocity is given by

$$\boldsymbol{v}^P_{2,ah} = \boldsymbol{v}^P_{0,ta} - \boldsymbol{v}^♂_0 \tag{5.24}$$

(replacing $\boldsymbol{v}^P_{0,ta}$ with $\boldsymbol{v}^P_{0,tp}$ for more inner planets). At aphelion of the heliocentric Hohmann transfer, the spacecraft is slower than Mars, so the spacecraft enters the sphere of influence of the planet from ahead, with respect to the planet's motion around the Sun (for more inner planets, the spacecraft would catch the planet from behind). Then:

$$v^P_{2,ah} = \sqrt{-\frac{\mu_♂}{a_a}} = \sqrt{\frac{\mu_⊙}{a_♂}}\left(1 - \sqrt{\frac{2a_⊕}{a_⊕ + a_♂}}\right). \tag{5.25}$$

In the *complete* patched conics method, $\boldsymbol{v}^P_{2,ah}$ is given by the difference of the absolute arrival velocity in Lambert's solution minus the instantaneous absolute velocity of the target planet.

Midcourse trajectory correction maneuvers can select very precisely the entry point of the sphere of influence (i.e., infinity with respect to the planet) as discussed in section 5.4.2). We may want to insert the spacecraft into a target science orbit (subindex s) of pericenter radius r_{sp} around the target planet. In this case, we want to tailor the arrival hyperbola so that its pericenter radius equals r_{sp}:

$$r_{ap} = a_a(1 - e_a) = r_{sp} \tag{5.26}$$

or we may want to simply let it fly by Mars and leave again, with a velocity turn angle δ_a given by

$$\sin\left(\frac{\delta_a}{2}\right) = \frac{1}{e_a} \tag{5.27}$$

Condition (5.25) plus either (5.26) or (5.27) allow finding a_a and e_a, thus determining the arrival hyperbola:

$$\boxed{a_a = -\frac{\mu_♂}{(v^P_{1,ah})^2}; \qquad e_a = 1 + \frac{r_{sp}(v^P_{1,ah})^2}{\mu_♂}.} \tag{5.28}$$

To fully determine the orbital plane of the hyperbola and relate this with the relevant aiming parameters one must use B-plane targeting.

5.4.2 B-plane targeting

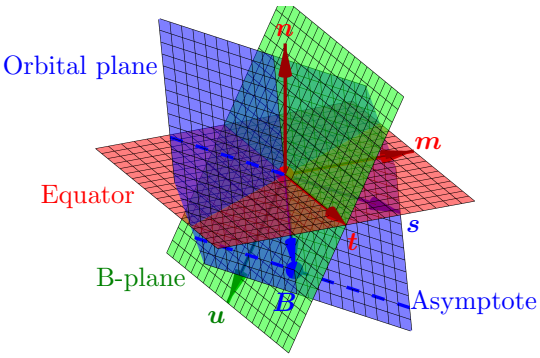
We have postponed the details of precisely aiming the spacecraft to enter Marsspace with the desired arrival hyperbolic orbit. During the heliocentric phase of interplanetary flight, small trajectory correction maneuvers (TCMs) are carried out. Typically, the sooner these maneuvers are done, the lower their Δv cost.

We call O the center of the target planet, and define the unit vectors $\boldsymbol{s}$ in the direction of the incoming spacecraft velocity with respect to the planet, $\boldsymbol{v}^P_{1,ah}$; $\boldsymbol{n}$ in the direction of the North pole of the target planet; $\boldsymbol{t} = \boldsymbol{s} \times \boldsymbol{n}/|\boldsymbol{s} \times \boldsymbol{n}|$, a vector that lies in the equatorial plane of the target planet and is perpendicular to $\boldsymbol{s}$; $\boldsymbol{m} = \boldsymbol{n} \times \boldsymbol{t}$, a vector in the equatorial plane perpendicular to $\boldsymbol{t}$; and $\boldsymbol{u} = \boldsymbol{s} \times \boldsymbol{t}$. Note that $\{\boldsymbol{t}, \boldsymbol{u}\}$ is a basis for the plane perpendicular to $\boldsymbol{s}$ that passes by O and intersects the equatorial plane of the planet along the direction of $\boldsymbol{t}$. We will call this plane the *B-plane*. Note also that the reference frame $S_* : \{O, B_*\}$ with $B_* = \{\boldsymbol{t}, \boldsymbol{m}, \boldsymbol{n}\}$ can be regarded as a planet-centered inertial equatorial reference frame.

We define the *impact parameter vector* $\boldsymbol{B}$ as the vector that goes from O to the intersection point of the incoming asymptote with the B-plane. Observe that $\boldsymbol{B}$ is perpendicular to the incoming asymptote (direction $\boldsymbol{t}$), so it is contained in the B-plane. We may write:

$$\boldsymbol{B} = B_t\boldsymbol{t} + B_u\boldsymbol{u} = B\cos\theta\boldsymbol{t} + B\sin\theta\boldsymbol{u}, \tag{5.29}$$

where B is the (scalar) impact parameter of the arrival hyperbola and θ is the angle in the B-plane from $\boldsymbol{t}$ to $\boldsymbol{B}$. Vectors $\boldsymbol{s}$ and $\boldsymbol{B}$ define the plane of the arrival hyperbolic orbit.



If our goal is to insert the spacecraft into a final science orbit, we want this plane to coincide with the science orbital plane to avoid expensive plane-change maneuvers, and the pericenter of the hyperbola to match the science orbit pericenter. The impact parameter is related to the arrival hyperbola pericenter radius through

$$\boxed{B = \sqrt{r_{ap}^2 + \frac{2r_{ap}\mu_{\mathcal{O}}}{(v_{2,ah}^P)^2}}.} \tag{5.30}$$

The inclination angle (with respect to the planet's equatorial plane) of the orbital plane is related to θ through:

$$\boxed{\cos i = (\boldsymbol{s}\cdot\boldsymbol{m})\cos\theta} \tag{5.31}$$

Observe that TCMs can change $\boldsymbol{B}$ easily, allowing us to select i and r_{ap}, but one can hardly change $\boldsymbol{t}$ or $v_{2,ah}^P$ without a large Δv. This means that it is hard to change the Ω and ω of the science orbit with TCMs. Note also that there is a maximum possible value of the inclination i that can be achieved this way.

5.4.3 Capture impulse

In case insertion into a science orbit is desired, the velocity at the pericenter of the arrival hyperbola is

$$v_{2,ap}^P = \sqrt{\frac{\mu_{\mathcal{O}}}{a_a}\left(\frac{1+e_a}{1-e_a}\right)}. \tag{5.32}$$

We must decelerate the spacecraft to the velocity at the pericenter of the science orbit, $v_{2,sp}^P$,

$$v_{2,sp}^P = \sqrt{\frac{\mu_{\mathcal{O}}}{a_s}\left(\frac{1+e_s}{1-e_s}\right)}. \tag{5.33}$$

This is a tangential maneuver, so the Δv may be computed with a scalar operation, i.e. $\Delta v = v_{2,ap}^P - v_{2,sp}^P$.

The total Δv budget for the interplanetary mission is then the sum of the escape and the capture impulses. As mentioned above, the TCMs typically have a very small Δv cost, and can therefore be ignored in first approximation (it is typically included in the mission margins, as there is some uncertainty to it). If the target planet has atmosphere, the capture impulse may be carried out by *aerobraking*, i.e., by using drag forces to decelerate the spacecraft, rather than rockets, saving Δv.

5.4.4 Planetary fly-by and gravity assist

If no capture impulse takes place, the spacecraft will fly by the planet and leave its sphere of influence again. In the process, the velocity of the spacecraft relative to the planet $\boldsymbol{v}_1^P$ will have been deflected a turn angle δ.

For an heliocentric observer in S_0, all this occurs in a tiny portion of space and in a short period of time, in which the velocity of the spacecraft seems to change abruptly (i.e. impulsively) from $\boldsymbol{v}_{0,\text{before}}^P$ to $\boldsymbol{v}_{0,\text{after}}^P$:

$$\boldsymbol{v}_{0,\text{before}}^P = \boldsymbol{v}_0^{\sigma} + \boldsymbol{v}_{2,\text{before}}^P, \tag{5.34}$$

$$\boldsymbol{v}_{0,\text{after}}^P = \boldsymbol{v}_0^{\sigma} + \boldsymbol{v}_{2,\text{after}}^P, \tag{5.35}$$

and thus it will experience a $\Delta\boldsymbol{v}$ without expending any propellant. This is known as a *gravity assist maneuver*. The energy for this change is "stolen" from the planet's orbital energy.

Using the $\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{w}\}$ vector basis of the perifocal reference frame for the arrival hyperbola in S_2, this $\Delta\boldsymbol{v}$ can be expressed as

$$\Delta\boldsymbol{v} = \boldsymbol{v}_{2,\text{after}}^P - \boldsymbol{v}_{2,\text{before}}^P = -2v_{2,ah}^P \sin\left(\frac{\delta_a}{2}\right)\boldsymbol{p} = -\frac{2v_{2,ah}^P}{1 + r_{sp}(v_{2,ah}^P)^2/\mu_{\sigma}}\boldsymbol{p}. \tag{5.36}$$

The $\Delta\boldsymbol{v}$ always lies along the apse line, in the $-\boldsymbol{p}$ direction. The spacecraft gains momentum (in S_0) when it performs a trailing side fly-by of the planet; it loses momentum when it performs a leading side fly-by of the planet.

Note that the direction of $\boldsymbol{p}$ can be described with the unit vectors introduced in section 5.4.2:

$$\boldsymbol{p} = \cos\beta_a \boldsymbol{s} + \sin\beta_a \frac{\boldsymbol{B}}{B}. \tag{5.37}$$

5.5 Pork-chop analysis

Pork-chop analysis is a fundamental optimization tool for interplanetary mission analysis. By solving the *complete* patched conics method with the associated Lambert problem for all possible combinations of departure and arrival dates, it is possible to identify low-cost launch and time of flight options and windows. The graphical representation of this analysis is done in the so-called pork-chop plots, which consist in a contour plot of the total Δv as a function of the departure date and the time of flight (or the arrival date). Alternatively, the characteristic energy c_3 of the departure and arrival hyperbolas is represented instead of Δv.

The steps to create a pork-chop diagram are as follows:

1. Create a sufficiently fine mesh for the range of departure times and durations of flight to be investigated.
2. For each point in this mesh:
   (a) Find out the initial position of the departure planet and the final position of the target planet. This can be done by integrating their equations of motion or by querying an ephemeris database.
   (b) Solve the short or long arc Lambert problem, depending on the value of $\Delta\theta$ between the departure and arrival positions. We can restrict our search to ellipse trajectories only, as high energy transfers are seldom of interest. Then, we would discard any parabolic or hyperbolic solutions.
   (c) Get the initial and final absolute velocities, compute the relative velocities to the planets, and solve the planetary sub-problems to find out the total Δv.
3. Plot the contours of Δv in the time and duration grid, and identify windows of interest.

When the $\Delta\theta$ to be traveled by the spacecraft becomes 180 deg, the Lambert problem is ill-posed as the normal plane is not well defined. In this case, we must choose the orbital plane to use, e.g. the plane of the ecliptic. An example porkchop diagram is shown in figure 5.1.

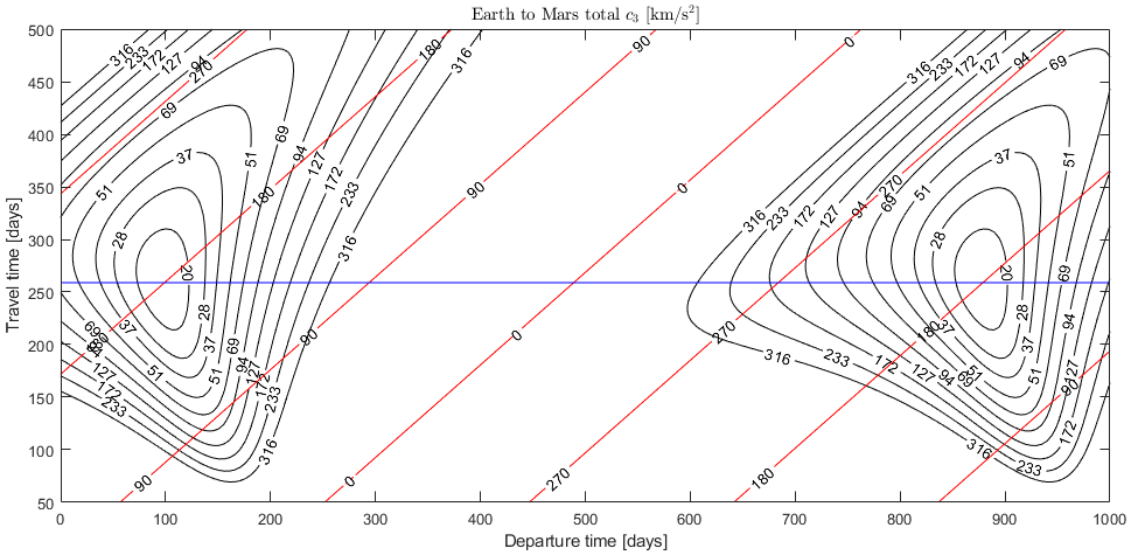


Figure 5.1: Example porkchop diagram for an Earth-Mars like flight. In this example, Earth and Mars have been assumed to have circular coplanar orbits. The black contour levels show $c_3 = (v^P_{1,eh})^2 + (v^P_{2,ah})^2$. Red lines are lines of constant $\theta_{\mathcal{O}}(t_2) - \theta_{\oplus}(t_1)$. The blue line shows the travel time for a Hohmann transfer, t_t. All points between 0 and 180 deg correspond to short arc transfers; from 180 to 360, long arc transfers. Only elliptic transfers have been considered. As it can be seen, the optimum in this case occurs for $180deg$ transfers with a flight duration t_t. The pattern repeats every 780 days, which corresponds to the synodic period of Earth and Mars.

6 Three-body problem

In comparison with the two-body problem, the additional presence of one more body in the three-body problem makes it analytically unsolvable (except in some trivial cases). Moreover, the problem exhibits chaos: small variations in the initial conditions can mean large deviations in the subsequent trajectory.

Early studies on the three-body problem were motivated by the motion of the Moon around the Earth including the effect of the Sun. Virtually all the fundamental work on the three-body problem is due to Lagrange. This problem is also useful to explain the existence of "trapped" asteroids in the orbit of Jupiter, known as Trojans and Greeks. More importantly to us, it allows to design cislunar space missions and missions to other binary systems (e.g. Sun and Jupiter, Jupiter and one of its Moons, etc) and exploit many interesting concepts such as Lagrange libration points and invariant manifolds.

The general three-body problem is far too ample, and except for some preliminary considerations, in this lecture we shall focus our attention on the *circular restricted three-body problem* (CR3BP), where two of the three masses are much larger than the third and perform known circular orbits about each other.

Problem statement: Analyze the motion of a three bodies in the due to their mutual gravitational attractions.

Assumptions: The three masses are modeled as point particles and assumed to be alone in the universe.

6.1 General mathematical model

The equations of motion of the three particles P_1, P_2 and P_3 in an arbitrary inertial reference frame S_0 are:

$$m^{P_1}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_1}}{\mathrm{d}t^2}\right|_0 = +G\frac{m^{P_1}m^{P_2}}{|\boldsymbol{r}_0^{P_2}-\boldsymbol{r}_0^{P_1}|^3}(\boldsymbol{r}_0^{P_2}-\boldsymbol{r}_0^{P_1}) + G\frac{m^{P_1}m^{P_3}}{|\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_1}|^3}(\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_1}), \tag{6.1}$$

$$m^{P_2}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_2}}{\mathrm{d}t^2}\right|_0 = -G\frac{m^{P_1}m^{P_2}}{|\boldsymbol{r}_0^{P_2}-\boldsymbol{r}_0^{P_1}|^3}(\boldsymbol{r}_0^{P_2}-\boldsymbol{r}_0^{P_1}) + G\frac{m^{P_2}m^{P_3}}{|\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_2}|^3}(\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_2}), \tag{6.2}$$

$$m^{P_3}\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_3}}{\mathrm{d}t^2}\right|_0 = -G\frac{m^{P_1}m^{P_3}}{|\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_1}|^3}(\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_1}) - G\frac{m^{P_2}m^{P_3}}{|\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_2}|^3}(\boldsymbol{r}_0^{P_3}-\boldsymbol{r}_0^{P_2}). \tag{6.3}$$

Adding the three equations, it is straightforward to see that the center of mass G of the system is a non-accelerated point,

$$\boxed{M\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{G}}{\mathrm{d}t^2}\right|_0 = 0,} \tag{6.4}$$

where $M = (m^{P_1} + m^{P_2} + m^{P_3})$ is the total mass of the system. It is also easy to demonstrate that:

1. The total mechanical energy of the system is constant (hint: dot-multiply each equation by the corresponding velocity vector and add them up).

2. The total angular momentum vector $\boldsymbol{H}_0$ of the system is constant (hint: cross-multiply each equation by the corresponding position vector and add them up).

Vector $\boldsymbol{H}_0$ and the center of mass define an invariant plane in the three-body problem. Observe, however, that the three-body problem is in general not planar (i.e., the three masses can move out of this plane).

Thus, we choose our inertial reference frame with origin at G, and a non-rotating right-handed orthornormal vector basis $B = \{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$, with $\boldsymbol{k}$ in the direciton of $\boldsymbol{H}_0$, i.e., $S_0 : \{G; B\}$.

6.2 Lagrange solutions

There is not much we can do analytically with the general three-body problem. The only known exceptions are the equilateral and collinear solutions of Lagrange, which are found by imposing that the system remains in the invariant plane perpendicular to $\boldsymbol{H}_0$ and maintains its relative shape in time. In other words, the three position vectors must rotate at the same rate $\dot{\theta}(t)$ to keep the angles between them constant, and their magnitudes must accept the expressions:

$$r_0^{P_1}(t) = r_0^{P_1}(0)f(t); \qquad r_0^{P_2}(t) = r_0^{P_2}(0)f(t); \qquad r_0^{P_3}(t) = r_0^{P_3}(0)f(t), \tag{6.5}$$

with $f(0) = 1$. Defining a polar vector basis for each particle $i = 1, 2, 3$, $\{\boldsymbol{u}_{ri}, \boldsymbol{u}_{\theta i}\}$ we can write:

$$\boldsymbol{r}_0^{P_i}(t)/r_0^{P_i}(0) = f(t)\boldsymbol{u}_{ri}; \tag{6.6}$$

$$\boldsymbol{v}_0^{P_i}(t)/r_0^{P_i}(0) = \dot{f}(t)\boldsymbol{u}_{ri} + f(t)\dot{\theta}(t)\boldsymbol{u}_{\theta i}; \tag{6.7}$$

Using these expressions, the total angular momentum of the system is

$$\boldsymbol{H}_0 = f^2(t)\dot{\theta}(t)\sum_{i=1}^{3} m^{P_i}[r_0^{P_i}(0)]^2\boldsymbol{k} = \text{const} \tag{6.8}$$

From where it is easy to see that:

1. The factor $f^2(t)\dot{\theta}(t)$ must be constant, and thus the angular rate must be $\dot{\theta}(t) \propto 1/f^2(t)$.
2. The angular momentum of each individual particle is constant, since

   $$\boldsymbol{H}_0^{P_i} = \boldsymbol{r}_0^{P_i} \times m^{P_i}\boldsymbol{v}^{P_i} = m^{P_i}[r_0^{P_i}(0)]^2 f^2(t)\dot{\theta}(t)\boldsymbol{k}. \tag{6.9}$$

   Hence,

   $$\mathrm{d}\boldsymbol{H}_0^{P_i}/\mathrm{d}t|_0 = \boldsymbol{r}_0^{P_i} \times \boldsymbol{F}_{P_i} = \boldsymbol{0}, \tag{6.10}$$

   so the resultant force on each particle passes through the center of mass of the system, $\boldsymbol{F}_{P_i} = F_{P_i}\boldsymbol{u}_{ri}$, and thus the motion of each particle is a **central force problem**.

Thus, cross-multiplying the equation of motion (6.1) by $\boldsymbol{r}_0^{P_1}$ results in

$$\boxed{\boldsymbol{r}_0^{P_i} \times \left(\frac{m^{P_2}}{|\boldsymbol{r}_0^{P_2} - \boldsymbol{r}_0^{P_1}|^3}\boldsymbol{r}_0^{P_2} + \frac{m^{P_3}}{|\boldsymbol{r}_0^{P_3} - \boldsymbol{r}_0^{P_1}|^3}\boldsymbol{r}_0^{P_3}\right) = \boldsymbol{0}.} \tag{6.11}$$

Similar expressions result for (6.2) and (6.3) when cross-multiplied by $\boldsymbol{r}_0^{P_2}$ and $\boldsymbol{r}_0^{P_3}$. Analysis of these expressions shows that there are only two types of configurations that satisfy our hypotheses:

1. Collinear configurations, in which $\boldsymbol{r}_0^{P_1}, \boldsymbol{r}_0^{P_2}$ and $\boldsymbol{r}_0^{P_3}$ are parallel, i.e., the three particles lie on a line that rotates with angular rate $\dot{\theta}(t)$. Then, $\boldsymbol{r}_0^{P_1} \times \boldsymbol{r}^{P_2} = 0$, etc. The possible positions for the particles on this line can be found from the condition that $\dot{\theta}(t)$ be the same for the three particles, which results in Lagrange's quintic equation. There are only three possible configurations of this type.
2. Equilateral configurations, in which $r_0^{P_1} = r_0^{P_2} = r_0^{P_3}$ and the vectors form 60 deg angles with each other. There are two configurations of this type.

In both cases, it is possible to write the equation of motion of each particle $i = 1, 2, 3$ after some manipulation as:

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}_0^{P_1}}{\mathrm{d}t^2}\right|_0 + \frac{k}{(r_0^{P_i})^3}\boldsymbol{r}_0^{P_i} = \boldsymbol{0}, \tag{6.12}$$

where k is some constant with units $[L^3/T^2]$. By analogy with the two-body problem, the possible trajectories of each particle in these configurations are conics with a focus at the center of mass of the system—ellipses, parabolas, hyperbolas.

6.3 Circular restricted three-body problem

To progress analytically beyond the Lagrange solutions for a more general case, it is necessary to take additional assumptions. The circular restricted three-body problem (CR3BP) assumes also that

1. The masses of the particles satisfy $m^{P_1} > m^{P_2} \gg m^{P_3}$. Thus, the motion of P_1 and P_2 (the *primaries*) is assumed to be unaffected by P_3.

2. The primaries orbit each other in circles.

We shall call $M^* = m^{P_1} + m^{P_2}$ the total mass of the system (i.e., of the primaries, ignoring the small m^{P_3}), and L the constant distance between P_1 and P_2.

6.3.1 Synodic reference frame

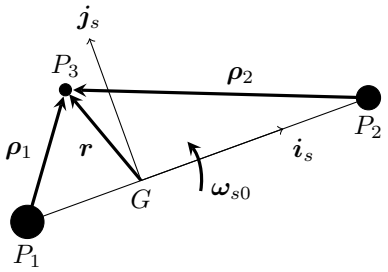
We define the non-inertial reference frame $S_s : \{G; B_s\}$, where $B_s = \{\boldsymbol{i}_s, \boldsymbol{j}_s, \boldsymbol{k}_s\}$ is a rotating vector basis with

- $\boldsymbol{k}_s \equiv \boldsymbol{k}$ in the direction of the angular momentum vector of the primaries.
- $\boldsymbol{i}_s$ in the direction of the vector that goes from P_1 to P_2, i.e., on the axis of the primaries.
- As a consequence of this, $\boldsymbol{j}_s = \boldsymbol{k}_s \times \boldsymbol{i}_s$ lies in the invariant plane of the system together with $\boldsymbol{i}_s$

In this reference frame the primaries are fixed. S_s rotates with respect to inertial space with the constant angular rate of the primaries,

$$\boxed{\boldsymbol{\omega}_{s0} = \sqrt{\frac{GM^*}{L^3}}\boldsymbol{k}_s.} \tag{6.13}$$

Observe that, since this reference frame is non-inertial, any analysis in this frame must account for the relevant inertial forces: centrifugal force and Coriolis force.



The equation of the center of mass, ignoring the small m^{P_3}, projected along the line of the primaries, shows that

$$m^{P_1} r_0^{P_1} = m^{P_2} r_0^{P_2}. \tag{6.14}$$

Since $r_0^{P_1} + r_0^{P_2} = L$, we can write

$$\boxed{r_0^{P_1} = \frac{m^{P_2}}{M^*}L; \qquad\qquad r_0^{P_2} = \frac{m^{P_1}}{M^*}L.} \tag{6.15}$$

6.3.2 Dimensionless equation of motion

To work with the CR3BP we normalize all magnitudes with M^*, L, $1/\omega_{s0}$. For simplicity, we shall also drop all unambiguous sub- and super-indices. With this convention we define:

- Vectors $\boldsymbol{r}$, $\boldsymbol{v}$, $\boldsymbol{a}$: the normalized position, velocity and acceleration vectors of P_3 relative to S_s. Likewise, the Cartesian coordinates in S_s will be denoted simply as x, y, z.
- Vector $\boldsymbol{\rho}_1$: the normalized vector from P_1 to P_3.
- Vector $\boldsymbol{\rho}_2$: the normalized vector from P_2 to P_3.

We also introduce the non-dimensional parameter

$$\boxed{\mu^* = \frac{m^{P_2}}{M^*},} \tag{6.16}$$

not to be confused with the gravitational parameter μ of the two-body problem.

The non-dimensional position of the primaries along the Gx axis is then $x = -\mu^*$ for P_1, and $x = 1 - \mu^*$ for P_2. Hence, the magnitudes of $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are given by:

$$\boxed{\rho_1^2 = (x + \mu^*)^2 + y^2 + z^2,} \tag{6.17}$$

$$\boxed{\rho_2^2 = (x + \mu^* - 1)^2 + y^2 + z^2.} \tag{6.18}$$

The nondimensional equation of motion for P_3 reads

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right|_s = -\frac{1-\mu^*}{\rho_1^3}\boldsymbol{\rho}_1 - \frac{\mu^*}{\rho_2^3}\boldsymbol{\rho}_2 - \boldsymbol{k}\times(\boldsymbol{k}\times\boldsymbol{r}) - 2\boldsymbol{k}\times\left.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right|_s. \tag{6.19}$$

Projected along the vectors of B_s yields:

$$\boxed{\ddot{x} = -\frac{1-\mu^*}{\rho_1^3}(x+\mu^*) - \frac{\mu^*}{\rho_2^3}(x+\mu^*-1) + x + 2\dot{y},} \tag{6.20}$$

$$\boxed{\ddot{y} = -\frac{1-\mu^*}{\rho_1^3}y - \frac{\mu^*}{\rho_2^3}y + y - 2\dot{x},} \tag{6.21}$$

$$\boxed{\ddot{z} = -\frac{1-\mu^*}{\rho_1^3}z - \frac{\mu^*}{\rho_2^3}z.} \tag{6.22}$$

Observe that the two gravitational forces and the centrifugal force admit a potential form, but the Coriolis force does not. For the first two, then:

$$\boxed{U = -\frac{1-\mu^*}{\rho_1} - \frac{\mu^*}{\rho_2} - \frac{1}{2}(x^2+y^2)} \tag{6.23}$$

It is then possible to write the equations of motion as

$$\ddot{x} - 2\dot{y} = -\frac{\partial U}{\partial x} \tag{6.24}$$

$$\ddot{y} + 2\dot{x} = -\frac{\partial U}{\partial y} \tag{6.25}$$

$$\ddot{z} = -\frac{\partial U}{\partial z} \tag{6.26}$$

6.3.3 Lagrange libration points

If we make $\dot{x}, \ddot{x}, \dot{y}, \ddot{y}, \dot{z}, \ddot{z} = 0$ in the equations above, we can find the equilibrium positions for P_3 in S_s.

These equilibrium positions coincide with critical points of the potential, i.e., $\nabla U = \mathbf{0}$. They are positions where the combined force of the two gravitational forces *and* the centrifugal force cancel out. All of them are located in the invariant plane Gxy. There is a total of five equilibrium positions in the CR3BP, which coincide with the positions of P_3 in the Lagrange solutions of section 6.2 when we apply the additional assumptions of the CR3BP, and they are called *Lagrange libration points*:

1. L_1, L_2, and L_3 are the collinear Lagrange points. They are labeled according to the energy needed to access them (see section 6.3.6), with L_1 located to the between P_1 and P_2, L_2 located to the further out than P_2, and L_3 located further out than P_1. Their exact position depends on μ^*.

2. L_4 and L_5 are the equilateral Lagrange points. They are usually labeled in the following order: L_4 is the point lying on $y > 0$, and L_5 is the point lying on $y < 0$. Their position vectors in S_s are given by:

$$\boldsymbol{r}^{L_4} = \left(\frac{1}{2} - \mu^*\right)\boldsymbol{i}_s + \frac{\sqrt{3}}{2}\boldsymbol{j}_s, \tag{6.27}$$

$$\boldsymbol{r}^{L_5} = \left(\frac{1}{2} - \mu^*\right)\boldsymbol{i}_s - \frac{\sqrt{3}}{2}\boldsymbol{j}_s. \tag{6.28}$$

6.3.4 Linearized motion about Lagrange libration points

To gain some insight on the motion of a particle in the neighborhood of the Lagrange libration points we will linearize the equations of motion about each point L_i for $i = 1, \ldots, 5$. Among other things, this will allow us to assess the (linear) stability of these equilibrium points, and in case of stability, compute the oscillation (a.k.a. libration) angular frequencies.

If the position of the Lagrange point L_i under analysis is given by $\boldsymbol{r}^{L_i} = x^{L_i}\boldsymbol{i}_s + y^{L_i}\boldsymbol{j}_s + z^{L_i}\boldsymbol{k}_s$ we will denote the position of P_3 *relative to* L_i as $\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{r}^{L_i} = x'\boldsymbol{i}_s + y'\boldsymbol{j}_s + z'\boldsymbol{k}_s$. In this new co-rotating reference frame $S_{L_i} : \{L_i, B_s\}$ the position of L_i is given by $\boldsymbol{r}' = \boldsymbol{0}$. The gradient ∇U near L_i is approximated by its Taylor's series expansion about L_i to first order in x', y', and z', i.e., $\nabla U(\boldsymbol{r}') = \nabla U(\boldsymbol{0}) + \nabla\nabla U(\boldsymbol{0}) \cdot \boldsymbol{r}'$, where $\nabla\nabla U(\boldsymbol{0})$ is the Hessian matrix of U at L_i. Since $\nabla U(\boldsymbol{0}) = \boldsymbol{0}$ (we are at a Lagrange point), the first term in the series expansion is zero:

$$\frac{\partial U}{\partial x}(\boldsymbol{r}') = \cancel{\frac{\partial U}{\partial x}(\boldsymbol{0})} + \frac{\partial^2 U}{\partial x^2}(\boldsymbol{0})x' + \frac{\partial^2 U}{\partial x \partial y}(\boldsymbol{0})y' + \frac{\partial^2 U}{\partial x \partial z}(\boldsymbol{0})z', \tag{6.29}$$

$$\frac{\partial U}{\partial y}(\boldsymbol{r}') = \cancel{\frac{\partial U}{\partial y}(\boldsymbol{0})} + \frac{\partial^2 U}{\partial x \partial y}(\boldsymbol{0})x' + \frac{\partial^2 U}{\partial y^2}(\boldsymbol{0})y' + \frac{\partial^2 U}{\partial y \partial z}(\boldsymbol{0})z', \tag{6.30}$$

$$\frac{\partial U}{\partial z}(\boldsymbol{r}') = \cancel{\frac{\partial U}{\partial z}(\boldsymbol{0})} + \frac{\partial^2 U}{\partial x \partial z}(\boldsymbol{0})x' + \frac{\partial^2 U}{\partial y \partial z}(\boldsymbol{0})y' + \frac{\partial^2 U}{\partial z^2}(\boldsymbol{0})z'. \tag{6.31}$$

The second derivatives of U at L_i ($\boldsymbol{r}' = \boldsymbol{0}$), which constitute the elements of matrix $\nabla\nabla U(\boldsymbol{0})$, are given by:

$$\frac{\partial^2 U}{\partial x^2}(\boldsymbol{0}) = -3\frac{1-\mu^*}{(\rho_1^{L_i})^5}(x^{L_i} + \mu^*)^2 - 3\frac{\mu^*}{(\rho_2^{L_i})^5}(x^{L_i} + \mu^* - 1)^2 + \frac{1-\mu^*}{(\rho_1^{L_i})^3} + \frac{\mu^*}{(\rho_2^{L_i})^3} - 1 \tag{6.32}$$

$$\frac{\partial^2 U}{\partial x \partial y}(\boldsymbol{0}) = -3\frac{1-\mu^*}{(\rho_1^{L_i})^5}(x^{L_i} + \mu^*)y^{L_i} - 3\frac{\mu^*}{(\rho_2^{L_i})^5}(x^{L_i} + \mu^* - 1)y^{L_i} \tag{6.33}$$

$$\frac{\partial^2 U}{\partial y^2}(\boldsymbol{0}) = -3\frac{1-\mu^*}{(\rho_1^{L_i})^5}(y^{L_i})^2 - 3\frac{\mu^*}{(\rho_2^{L_i})^5}(y^{L_i})^2 + \frac{1-\mu^*}{(\rho_1^{L_i})^3} + \frac{\mu^*}{(\rho_2^{L_i})^3} - 1 \tag{6.34}$$

$$\frac{\partial^2 U}{\partial x \partial z}(\boldsymbol{0}) = 0, \tag{6.35}$$

$$\frac{\partial^2 U}{\partial y \partial z}(\boldsymbol{0}) = 0, \tag{6.36}$$

$$\frac{\partial^2 U}{\partial z^2}(\boldsymbol{0}) = \frac{1-\mu^*}{(\rho_1^{L_i})^3} + \frac{\mu^*}{(\rho_2^{L_i})^3}. \tag{6.37}$$

Denoting $\partial^2 U/\partial x^2(\boldsymbol{0})$ as U_{xx0}, etc, for brevity, we can write the linearized equations of motion in matrix form as:

$$\ddot{x}' = -U_{xx0}x' - U_{xy0}y' + 2\dot{y}', \tag{6.38}$$

$$\ddot{y}' = -U_{xy0}x' - U_{yy0}y' - 2\dot{x}', \tag{6.39}$$

$$\ddot{z}' = -U_{zz0}z'. \tag{6.40}$$

The motion in the $x'y'$ plane is decoupled from the motion in the z' direction in the linearized equations. This allows solving each part separately.

Plane $x'y'$ problem

The $x'y'$ problem can be written in matrix form as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x' \\ y' \\ \dot{x}' \\ \dot{y}' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -U_{xx0} & -U_{xy0} & 0 & 2 \\ -U_{xy0} & -U_{yy0} & -2 & 0 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ \dot{x}' \\ \dot{y}' \end{bmatrix}. \tag{6.41}$$

This is a first-order differential linear system with constant coefficients. Assuming all eigenvalues of the matrix above have maximum geometric multiplicity[8], the general solution can be written as

$$\begin{bmatrix} x' \\ y' \\ \dot{x}' \\ \dot{y}' \end{bmatrix} = C_1\boldsymbol{b}_1\exp(\lambda_1 t) + C_2\boldsymbol{b}_2\exp(\lambda_2 t) + C_3\boldsymbol{b}_3\exp(\lambda_3 t) + C_4\boldsymbol{b}_4\exp(\lambda_4 t), \tag{6.42}$$

where C_1, C_2, C_3 and C_4 are constants to be determined from the initial conditions and $\boldsymbol{i}$, λ_i for $i = 1, 2, 3, 4$ are the eigenvectors and eigenvalues of the matrix.

The characteristic polynomial of the matrix can be used to compute the eigenvalues:

$$\boxed{\lambda^4 + (4 + U_{xx0} + U_{yy0})\lambda^2 + (U_{xx0}U_{yy0} - U_{xy0}^2) = 0} \tag{6.43}$$

The motion about L_i is:

1. Asymptotically stable if $\Re(\lambda_i) < 0$ for the four λ_i.
2. Linearly stable if $\Re(\lambda_i) \leqslant 0$ for the four λ_i, with at least one λ_i with $\Re(\lambda_i) = 0$, and those eigenvalues with real part equal to zero have equal geometric and algebraic multiplicities. Note that non-linear effects can destroy this weak marginal stability.
3. Unstable if at least one λ_i has $\Re(\lambda_i) > 0$, or if an eigenvalue with $\Re(\lambda_i) = 0$ has defective geometric multiplicity (degeneracy).

Equation (6.43) is a bi-quadratic equation with two solutions for $\zeta = \lambda^2$. Each solution ζ_1 and ζ_2 is associated to two eigenvalues λ_+ and λ_- with equal modulus and diametrically opposed in the complex plane. We can state the following:

- It is impossible[9] to have both λ_+ and λ_- in the semiplane $\Re(\lambda) < 0$. Hence, it is impossible for the system to be asymptotically stable. This makes physical sense, as asymptotic stability typically requires some form of damping, and our system is conservative in S_s.
- It is easy to see that if both solutions ζ_1, ζ_2 are real and negative, then the four eigenvalues λ fall in the imaginary axis. The system is linearly stable if and only if $\zeta_1, \zeta_2 < 0$ and all eigenvalues have maximum geometric multiplicity. This second condition is guaranteed if $\zeta_1 \neq \zeta_2$.
- Any other scenario results in at least one λ with positive real part, and thus the system is unstable.

For the **collinear Lagrange points** L_1, L_2, L_3, $y^{L_i} = 0$, $\rho_1^{L_i} = |x^{L_i} + \mu^*|$, and $\rho_2^{L_i} = |x^{L_i} + \mu^* - 1|$. It is easy to see that $U_{xx0} = -2K - 1$, $U_{xy0} = 0$, $U_{yy0} = K - 1$, where the value K is $K = (1-\mu^*)/(\rho_1^{L_i})^3 + \mu^*/(\rho_2^{L_i})^3$. Then, equation (6.43) becomes:

$$\zeta^2 + (2 - K)\zeta + (-2K - 1)(K - 1) = 0. \tag{6.44}$$

We can find that $K \geqslant 1$ for all three collinear lagrange points, and in all cases there is one root $\zeta_1 > 0$ and another root $\zeta_2 < 0$. Hence, the root $\zeta_1 > 0$ makes the collinear Lagrange points always unstable.

[8] Otherwise, solutions of the type $t^n\boldsymbol{b}\exp(\lambda t)$ would exist.
[9] Can you see why?

For the **equilateral Lagrange points** L_4, L_5, $x^{L_i} = 1/2 - \mu^*$, $y^{L_i} = \sqrt{3}/2$, and $\rho_1^{L_i} = \rho_2^{L_i} = 1$. It is easy to see that $U_{xx0} = -3/4$, $U_{xy0} = \pm 3\sqrt{3}(2\mu^* - 1)/4$, and $U_{yy0} = -9/4$. Then, equation (6.43) becomes:

$$\zeta^2 + \zeta + \frac{27}{4}\mu^*(1 - \mu^*) = 0. \tag{6.45}$$

The two values ζ_1, ζ_2 are real and negative if $\mu^* \leqslant 0.0385$. If this condition is met, the equilateral Lagrange points are stable. This is true in the majority of problems of interest (Sun-Earth system, Earth-Moon system, Sun-Jupiter system, etc)[10].

Out of plane z' problem

Regarding the problem in the z' direction, noting that $U_{zz0} > 0$ always, it is analogous to an undamped harmonic oscillator and its general solution is:

$$z' = C_5 \exp(-i\sqrt{U_{zz0}}t) + C_6 \exp(+i\sqrt{U_{zz0}}t) = C_5' \cos(\sqrt{U_{zz0}}) + C_6' \sin(\sqrt{U_{zz0}}), \tag{6.46}$$

where C_5, C_6 are two additional constants (and C_5', C_6' two dependent ones). The linearized motion in the out of plane z' direction is always stable and uncoupled from the motion in x', y'. The angular frequency of the z' oscillations is $\sqrt{U_{zz0}}$.

Lissajous orbits

When Lagrange points L_4 and L_5 are stable, there exist 3 superposed oscillatory motions with different angular frequencies. Ordering the eigenvalues of equation (6.42) such that $\lambda_1 = -\lambda_2 = is$ and $\lambda_3 = -\lambda_4 = il$ with $s \geqslant l$, we distinguish:

1. Short-period oscillation, with angular frequency s in the $x'y'$ plane.
2. Long-period oscillation, with angular frequency l in the $x'y'$ plane.
3. Out-of-plane oscillation, with angular frequency $\sqrt{U_{zz0}}$.

The orbits created by the superposition of the oscillatory movements are called *Lissajous orbits*.

Even at the unstable L_1, L_2, L_3 Lagrange points, it is possible to place a spacecraft in motion around them by selecting the initial conditions such that the constants C_i in equation (6.42) corresponding to unstable modes are equal to zero. The result is a Lissajous orbit with two angular frequencies:

1. Short-period oscillation, with angular frequency s in the $x'y'$ plane.
2. Out-of-plane oscillation, with angular frequency $\sqrt{U_{zz0}}$.

Naturally, any perturbation can feed the unstable modes, so station keeping maneuvers are needed to maintain this flight condition. In practice, the Δv budget associated to correcting the unstable component of the solution is small, and in reality many missions use Lagrange points L_1, L_2, L_3 (e.g. L_1 in the Sun-Earth system is a good location for Sun observation missions). Indeed, these points are energetically cheaper to access than L_4 and L_5, as discussed in 6.3.6, making them more appealing in many cases.

Moreover, even around the conditionally-stable L_4, L_5 points, non-linear effects and perturbations can affect the orbit of the spacecraft and make it unstable. Hence, station keeping maneuvers are needed in general for all Lagrange points.

[10] It is interesting to note that points L_1, L_2, L_3 are saddle point of the restriction of the potential U to the invariant plane xy, while points L_4, L_5 are maxima. Expectedly, L_1, L_2, L_3 are unstable equilibrium points. It is not so evident why L_4, L_5 are conditionally stable depending on the value of μ^*. Can you see how this is possible? Hint: there is a force of the equation of motion in S_s not described by U.

6.3.5 Jacobi's energy integral

If we dot-multiply the equation of motion (6.19) by $\boldsymbol{v}$, the velocity of P_3 in S_s, then we obtain

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right|_s \cdot \left.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right|_s + \nabla U \cdot \left.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right|_s = 0. \tag{6.47}$$

This expression is readily integrated to obtain the conservation of mechanical energy in the synodic reference frame, known as Jacobi's energy integral:

$$\boxed{\frac{v^2}{2} + U = E.} \tag{6.48}$$

Often, Jacobi's constant $C = -2E$ is used instead of E for historical reasons. Note that the mechanical energy of P_3 is not conserved in the inertial reference frame S_0, where the motion of the primaries prevents the definition of a time-independent potential.

The potential U is composed of three contributions—the gravitational potential due to P_1, the gravitational potential due to P_2, and the centrifugal force potential. The first two create negative cusps of the potential at the positions of the primaries. The third one drops parabolically away from the $x = y = 0$ axis. Once constant E is fixed by the initial conditions of the problem, the topology of this potential governs the possible motions of P_3 as seen from S_s.

6.3.6 Hill's surface

In equation (6.48), the condition $v^2 \geqslant 0$ defines the regions of existence of solution for the position of P_3. These regions are limited by the surface $U = E = \text{const}$, at which $v^2 = 0$. This surface is sometimes referred to as *Hill's surface*. We distinguish five different topologies of the Hill surface, separated by four critical values of E, E_1, E_2, E_3, E_4:

1. For $E < E_1$, there are three separate regions of existence: one blob around P_1, another blob around P_2 and the exterior space far from the primaries. For these energy values, it is impossible for P_3 to move from the neighborhood of P_1 to the neighborhood of P_2, or to escape the system. Likewise, for a particle very far away, it is impossible to get close to either of the primaries. As E increases, the inner blobs grow and the external constrain shrinks.

2. For $E = E_1$, the two inner blobs touch each other at Lagrange point L_1, between the primaries.

3. For $E_1 < E < E_2$, there are two regions of existence: an hourglass-shaped inner region, which covers the neighborhood of P_1 and P_2, and an external region far away from the primaries. For these energies P_3 can cross from P_1 to P_2 and vice-versa, but it cannot escape the system. Likewise, for a particle very far away, it is impossible to get close to either of the primaries. As the energy increases, the inner region continues to grow and the outer constraint to shrink.

4. For $E = E_2$, the inner and the outer regions touch at Lagrange point L_2, behind P_2.

5. For $E_2 < E < E_3$, the existence regions have merged into one. It is possible to enter and leave the system, and to travel to either primary, through the aperture that exists in Hill's surface around L_2. Increasing the energy makes the forbidden region limited by Hill's surface to recede.

6. For $E = E_3$, a new aperture is created between the inner region and exterior space at Lagrange point L_3, behind P_1.

7. For $E_3 < E < E_4$, similar to before, it is possible to travel to either primary and to escape the system, but now there are two doors between inner and outer space. Increasing the energy makes the forbidden region limited by Hill's surface to recede further.

8. For $E = E_4$, the forbidden parts of the invariant plane have shrunk to two points, the equilateral Lagrange points L_4 (located on $y > 0$) and L_5 (located on $y < 0$).

9. For $E > E_4$, it is possible to visit all points of the invariant plane. However, two forbidden regions still exist out of this plane, so not all three-dimensional motions are allowed. Further increasing the energy makes these forbidden regions recede to higher values of $|z|$.

6.3.7 Motion near Hill's surface

If P_3 reaches Hill's surface $U = E$, its velocity relative to the synodic reference frame S_s will become zero. However, in general, the resultant force on P_3 is not zero at this surface, so it will turn around and continue its motion.

Calling $t = t_0$ the instant when the particle touches Hill's surface at a point $\boldsymbol{r}^*$, we can use the Taylor series expansion of the position vector $\boldsymbol{r}$ and write:

$$\boldsymbol{r}(t) = \boldsymbol{r}^* + \cancel{\left.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right|_{s,t_0} t} + \frac{1}{2}\left.\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right|_{s,t_0} t^2 + \frac{1}{6}\left.\frac{\mathrm{d}^3\boldsymbol{r}}{\mathrm{d}t^3}\right|_{s,t_0} t^3 + O(t^4). \tag{6.49}$$

It is easy to see that the second derivative term is a vector normal to the surface that points toward the existence region:

$$\left.\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right|_{s,t_0} = -\nabla U(\boldsymbol{r}^*) = \boldsymbol{n} \tag{6.50}$$

Since $\boldsymbol{v} = \boldsymbol{0}$ at $\boldsymbol{r}^*$, taking derivatives in the equation of motion one can show that, at this point:

$$\dddot{x} = 2\ddot{y}; \qquad \dddot{y} = -2\ddot{x}; \qquad \dddot{z} = 0, \tag{6.51}$$

Thus, the third derivative term in the expansion is a tangent vector $\boldsymbol{s}$ to the surface, contained in the $\boldsymbol{i}_s\boldsymbol{j}_s$ plane. Therefore, the expansion can be written as:

$$\boldsymbol{r}(t) = \boldsymbol{r}^* + \frac{1}{2}\boldsymbol{n}t^2 + \frac{1}{6}\boldsymbol{s}t^3. \tag{6.52}$$

The trajectory of P_3 has a cusp point upon reaching Hill's surface, entering and leaving it parallel to the xy plane.

6.3.8 Non-linear periodic orbits

The CR3BP is rich with complex physics. One of its peculiarities is the existence of periodic orbits, not only in the linearized equations of motion about a Lagrange point, but also in the full, non-linear equations. These orbits can be categorized in families. Some of them are orbitally stable, while others are not. Current research is still finding new types of orbits and new properties of the CR3BP, which have direct applications in the design of interplanetary missions.

Examples of non-linear periodic orbits include the planar Lyapunov orbits and the non-planar vertical and halo orbits. The latter are of particular interest in, e.g., the Earth-Moon system, where L_2 halo orbits are a large, non-planar, oval-shaped orbits. When seen from the Earth, a spacecraft in such a halo orbit seems to move slowly around the Moon, and has continuous visibility with Earth. These orbits are proposed, for example, for a communications relay with lunar missions on the far side of the Moon.

7 Relative motion

The motion of a spacecraft with respect to another, near spacecraft is of great practical importance in astrodynamics. For example, formation-flying missions use two or more spacecraft that fly closely together to perform a variety of tasks. Rendezvous and docking, as well as spacecraft inspection, require closely orbiting a target spacecraft. Finally, we can study the perturbed motion of a satellite with respect to its nominal orbital position, as if this nominal position were a target satellite.

In all these cases, what matters is the relative motion of the chaser spacecraft with respect to the target spacecraft, or at least more so than their absolute motion of the two spacecraft with respect to the Earth. This relative motion is actually the *difference* between the two absolute motions.

In this chapter we will derive the equations of motion of a chaser satellite relative to an accelerated, rotating reference frame centered on the target satellite. Both the non-linear and the linearized equations of motion are presented. The latter enable much more insight into the problem of relative motion, albeit they incur in an error with respect to the full non-linear equations, which becomes more important the larger the distance of the chaser with respect to the target, and which accumulates over time.

7.1 Hill's reference frame

We define Hill's reference frame S_1 as follows

1. Origin on the target satellite A.
2. Ax axis pointing along the radial direction, outward: $\boldsymbol{i}_1 = \boldsymbol{r}_0^A / r_0^A$.
3. Ay axis pointing along the local horizontal, in the direction of motion of A: $\boldsymbol{j}_1 = \boldsymbol{k}_1 \times \boldsymbol{i}_1$.
4. Az axis pointing in the out of plane direction, along the angular momentum vector of A: $\boldsymbol{k}_1 = \boldsymbol{h}_0^A / h_0^A$.

This reference frame is accelerating and rotating, so it is clearly *non-inertial*. It is also referred to as the local-vertical local-horizontal (LVLH) reference frame.

The acceleration of the origin is given by the two-body-problem acceleration of A;

$$\boldsymbol{a}_0^A = -\frac{\mu}{(r_0^A)^3}\boldsymbol{r}_0^A. \tag{7.1}$$

The angular velocity vector can be written using the angular momentum vector:

$$\boldsymbol{\omega}_{10} = \dot{\theta}_A \boldsymbol{k}_1 = \frac{\boldsymbol{h}_0^A}{(r_0^A)^2}, \tag{7.2}$$

and, taking derivatives, the angular acceleration vector is:

$$\boldsymbol{\alpha}_{10} = -2\frac{\boldsymbol{h}_0^A}{(r_0^A)^3}\dot{r}_0^A = -2\frac{\boldsymbol{h}_0^A}{(r_0^A)^4}\boldsymbol{r}_0^A \cdot \boldsymbol{v}_0^A. \tag{7.3}$$

7.2 Relative kinematics

The kinematics of the chaser satellite B with respect to A are given by the well-known expressions of relative kinematics:

$$\boldsymbol{r}_0^B = \boldsymbol{r}_0^A + \boldsymbol{r}_1^B, \tag{7.4}$$

$$\boldsymbol{v}_0^B = \boldsymbol{v}_0^A + \boldsymbol{v}_1^B + \boldsymbol{\omega}_{10} \times \boldsymbol{r}_1^B, \tag{7.5}$$

$$\boldsymbol{a}_0^B = \boldsymbol{a}_0^A + \boldsymbol{a}_1^B + \boldsymbol{\alpha}_{10} \times \boldsymbol{r}_1^B + \boldsymbol{\omega}_{10} \times (\boldsymbol{\omega}_{10} \times \boldsymbol{r}_1^B) + 2\boldsymbol{\omega}_{10} \times \boldsymbol{v}_1^B. \tag{7.6}$$

These expressions, in particular the last one, are essential to write down the equation of motion of B in S_1 in the next section.

7.3 Relative dynamics

The absolute motion of B is given by

$$\boldsymbol{a}_0^B = -\frac{\mu}{(r_0^B)^3}\boldsymbol{r}_0^B. \tag{7.7}$$

Using the expression for the acceleration of B above, it is possible to write

$$\boldsymbol{a}_1^B = -\frac{\mu}{|\boldsymbol{r}_0^A + \boldsymbol{r}_1^B|^3}(\boldsymbol{r}_0^A + \boldsymbol{r}_1^B) + \frac{\mu}{(r_0^A)^3}\boldsymbol{r}_0^A - \boldsymbol{\alpha}_{10} \times \boldsymbol{r}_1^B - \boldsymbol{\omega}_{10} \times (\boldsymbol{\omega}_{10} \times \boldsymbol{r}_1^B) - 2\boldsymbol{\omega}_{10} \times \boldsymbol{v}_1^B \tag{7.8}$$

This is the full, non-linear equation of relative motion of B in S_1. At this point we could integrate this equation directly, and obtain the trajectory of B in this reference frame.

The next section linearizes the equations of motion to better understand the main terms that drive relative motion in a simplified way. Naturally, one must turn back to the full, non-linear equations in order to integrate with a precision the motion of the chaser spacecraft.

7.4 Clohessy-Wiltshire equations

For all cases of practical interest, the distance between A and B is smaller than the radius r_0^A. Hence, we can establish the ordering $r_1^B \ll r_0^A$ and expand the denominator in equation (7.8) to first order as:

$$|\boldsymbol{r}_0^A + \boldsymbol{r}_1^B|^{-3} = [(r_0^A)^2 + (r_1^B)^2 + 2\boldsymbol{r}_0^A \cdot \boldsymbol{r}_1^B]^{-3/2} \simeq (r_0^A)^{-3} - 3(r_0^A)^{-5}\boldsymbol{r}_0^A \cdot \boldsymbol{r}_1^B. \tag{7.9}$$

Using this approximation, the gravitational terms in equation (7.8) become:

$$-\frac{\mu}{|\boldsymbol{r}_0^A + \boldsymbol{r}_1^B|^3}(\boldsymbol{r}_0^A + \boldsymbol{r}_1^B) + \frac{\mu}{(r_0^A)^3}\boldsymbol{r}_0^A \simeq -\frac{\mu}{(r_0^A)^3}\left(\boldsymbol{r}_1^B - \frac{3}{(r_0^A)^2}\boldsymbol{r}_0^A \cdot \boldsymbol{r}_1^B\right). \tag{7.10}$$

Then, projecting equation (7.8) into the vector basis of S_1 we obtain:

$$\ddot{x} - 2\omega\dot{y} = +\frac{2\mu}{R^3}x + \omega^2 x + \alpha y, \tag{7.11}$$

$$\ddot{y} + 2\omega\dot{x} = -\frac{\mu}{R^3}y + \omega^2 y - \alpha x, \tag{7.12}$$

$$\ddot{z} = -\frac{\mu}{R^3}z, \tag{7.13}$$

where $R = r_0^A$ for short. These are the *linearized equations of motion* in the general case. In the particular case of a circular orbit, $\alpha = 0$, $R = \text{const}$, $\omega = \sqrt{\mu/R^3} = n = \text{const}$, and the equations are known as the *Clohessy-Wiltshire equations*:

$$\ddot{x} - 2n\dot{y} - 3n^2 x = 0, \tag{7.14}$$

$$\ddot{y} + 2n\dot{x} = 0, \tag{7.15}$$

$$\ddot{z} + n^2 z = 0, \tag{7.16}$$

As can be seen above, the linearized motion in z is decoupled from the motion in the x, y plane in the general case. The motion in z in the circular case is that of a harmonic oscillator with angular frequency n. Observe that the equations can be non-dimensionalized with any characteristic length of our choice and the characteristic time $1/n$ (this is however not done in the present notes).

In matrix form, the x, y problem can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 3n^2 & 0 & 0 & 2n \\ 0 & 0 & -2n & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} \tag{7.17}$$

And the z problem as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} z \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -n^2 & 0 \end{bmatrix} \cdot \begin{bmatrix} z \\ \dot{z} \end{bmatrix} \tag{7.18}$$

The general solution to the Clohessy-Wiltshire equations is[11]:

$$x = 2C_1 + C_2 \sin(nt) + C_3 \cos(nt), \tag{7.19}$$
$$y = -3C_1 nt + 2C_2 \cos(nt) - 2C_3 \sin(nt) + C_4, \tag{7.20}$$
$$z = C_5 \sin(nt) + C_6 \cos(nt), \tag{7.21}$$
$$\dot{x} = nC_2 \cos(nt) - nC_3 \sin(nt), \tag{7.22}$$
$$\dot{y} = -3nC_1 - 2nC_2 \sin(nt) - 2nC_3 \cos(nt), \tag{7.23}$$
$$\dot{z} = nC_5 \cos(nt) - nC_6 \sin(nt). \tag{7.24}$$

These expressions are linked to the inital conditions at $t = 0$ as follows:

$$C_1 = 2x_0 + \dot{y}_0/n; \quad C_2 = \dot{x}_0/n; \quad C_3 = -3x_0 - 2\dot{y}_0/n; \quad C_4 = y_0 - 2\dot{x}_0/n; \quad C_5 = \dot{z}_0/n; \quad C_6 = z_0.$$

7.4.1 Characteristic trajectories

Is easy to see that all coordinates and their derivatives have sinusoidal terms of angular frequency n, the orbital frequency of the reference orbit of satellite A.

Leaving aside the harmonic oscillator motion in the z direction, the more complex dynamics happen in the x, y plane, i.e. the orbital plane. By properly choosing the initial conditions, it is possible leave satellite B stationary directly ahead or directly behind with respect to satellite A, at a distance y_0. This requires that we excite only the integration constant $C_4 = y_0$. On the contrary, it is not possible to leave the satellite B stationary directly above or below A at a distance x_0: if we want to excite the integration constant C_1, we automatically introduce a new term in y. A satellite in this situation secularly lags behind (if $x_0 > 0$) or runs ahead of point A over time. The existence of this secular, linearly growing solutions is a major feature of the relative motion problem.

Exciting any of the other constants C_2, C_3, C_5, C_6 superposes different sinusoidal oscillations of angular frequency n to the basic motions described above.

We can get more insight into the problem by inspecting each type of motion from the absolute, ECI reference frame. By doing so, we can appreciate that the existence of a non-zero C_1 constant means that A and B have different semimajor axes, and hence, different orbital periods, so they will unavoidably separate from each other over time. Reciprocally, we can only have closed, periodic relative orbits if $C_1 = 0$ and the two periods coincide. If only C_4 is non-zero, the two satellites A and B share the same orbit, but have slightly different phase. Whenever C_2, C_3, C_5 and/or C_6 are non-zero, there are sinusoidal oscillations and the orbit of B is an ellipse; the maxima in x correspond to the apocenter, the minima to the pericenter.

Note that, while qualitatively identical, these characteristic trajectories differ from those of the full, non-linear problem. The error is small if the distance from A to B is small. However, the error may accumulate over time, and can become large in long duration missions such as formation flying missions.

7.4.2 Maneuvers

In relative motion problems it is common that we want to change the current trajectory of satellite B with respect to A into a new trajectory. This can be achieved with two ideal impulses: the first one sets the satellite B into an arc that transfers it to the new desired position in a prescribed amount of time. The second one corrects the velocity of B so it matches the required velocity in the new trajectory:

1. The state vector of the satellite initially is $\boldsymbol{r}(t_0)$, $\boldsymbol{v}(t_0^-)$.

2. Right after the first impulse, the position is still the same, $\boldsymbol{r}(t_0)$, but the velocity has changed to $\boldsymbol{v}(t_0^+)$. The first $\Delta\boldsymbol{v}$ is

$$\Delta\boldsymbol{v}_1 = \boldsymbol{v}(t_0^+) - \boldsymbol{v}(t_0^-). \tag{7.25}$$

[11] To obtain this solution, note that the eigenvalue $\lambda = 0$ of the matrix of the x, y problem has algebraic multiplicity equal to 2, but geometric multiplicity equal to 1. Hence, we must use Jordan chains of generalized eigenvalues to find the general solution.

3. After a fixed maneuver duration time $\Delta t = t_f - t_0$, the satellite has state vector $\boldsymbol{r}(t_f)$, $\boldsymbol{v}(t_f^-)$.

4. Finally, a second impulse is applied that does not change the position instantaneously, which remains equal to $\boldsymbol{r}(t_f)$, but changes the velocity to $\boldsymbol{v}(t_f^+)$:

$$\Delta \boldsymbol{v}_2 = \boldsymbol{v}(t_f^+) - \boldsymbol{v}(t_f^-). \tag{7.26}$$

Since we know the analytic solution to the linearized equations of motion, solving for Δv_1 and Δv_2 for a given maneuver time Δt is straightforward. The problem can be easily solved using the *state transition matrix* formulation of the solution:

$$\begin{bmatrix} \boldsymbol{r}(t) \\ \boldsymbol{v}(t) \end{bmatrix} = \Phi(t - t_0) \cdot \begin{bmatrix} \boldsymbol{r}(t_0) \\ \boldsymbol{v}(t_0) \end{bmatrix} \tag{7.27}$$

The state transition matrix $\Phi(t - t_0)$ applied to the state vector at time t_0 returns the state vector at time t. It is a 6x6 matrix composed of 4 blocks. Taking $t_0 = 0$ for compacity:

$$\Phi(t) = \left[\begin{array}{c|c} \Phi_{rr}(t) & \Phi_{rv}(t) \\ \hline \Phi_{vr}(t) & \Phi_{vv}(t) \end{array}\right] \tag{7.28}$$

so that

$$\boldsymbol{r}(t) = \Phi_{rr}(t) \cdot \boldsymbol{r}(0) + \Phi_{rv}(t) \cdot \boldsymbol{v}(0), \tag{7.29}$$
$$\boldsymbol{v}(t) = \Phi_{vr}(t) \cdot \boldsymbol{r}(0) + \Phi_{vv}(t) \cdot \boldsymbol{v}(0), \tag{7.30}$$

where:

$$\Phi_{rr}(t) = \begin{bmatrix} 4 - 3\cos(nt) & 0 & 0 \\ 6[\sin(nt) - nt] & 1 & 0 \\ 0 & 0 & \cos(nt) \end{bmatrix}; \tag{7.31}$$

$$\Phi_{rv}(t) = \begin{bmatrix} \sin(nt)/n & 2[1 - \cos(nt)]/n & 0 \\ 2[\cos(nt) - 1]/n & [4\sin(nt) - 3nt]/n & 0 \\ 0 & 0 & \sin(nt)/n \end{bmatrix}; \tag{7.32}$$

$$\Phi_{vr}(t) = \begin{bmatrix} 3n\sin(nt) & 0 & 0 \\ 6n[\cos(nt) - 1] & 0 & 0 \\ 0 & 0 & -n\sin(nt) \end{bmatrix}; \tag{7.33}$$

$$\Phi_{vv}(t) = \begin{bmatrix} \cos(nt) & 2\sin(nt) & 0 \\ 2\sin(nt) & 4\cos(nt) - 3 & 0 \\ 0 & 0 & \cos(nt) \end{bmatrix}. \tag{7.34}$$

This formulation is applied between times 0^+ and t_f^-, i.e., right after the first impulse and until right before the second impulse. In other words, we cover the coasting arc of the maneuver. Noting that $\boldsymbol{r}(0)$ and $\boldsymbol{r}(t_f)$ are known (inputs to the problem), the only unknowns are $\boldsymbol{v}(0^+)$ (velocity after first impulse) and $\boldsymbol{v}(f^-)$ (velocity before second impulse). The first can be obtained by inversion, and then the second can be obtained directly:

$$\boldsymbol{v}(0^+) = \Phi_{rv}^{-1}(t_f) \cdot [\boldsymbol{r}(t_f) - \Phi_{rr}(t_f) \cdot \boldsymbol{r}(0)], \tag{7.35}$$
$$\boldsymbol{v}(t_f^-) = \Phi_{vr}(t_f) \cdot \boldsymbol{r}(0) + \Phi_{vv}(t_f) \cdot \boldsymbol{v}(0^+). \tag{7.36}$$

Observe that $\Phi_{rr}(t_f)$ and $\Phi_{vv}(t_f)$ become the identity and $\Phi_{rv}(t_f)$ and $\Phi_{vr}(t_f)$ become the null matrix if $t_f = 0$.

In addition, $\Phi_{rv}(t_f)$ is singular for $t_f = k\pi/n$ (and also for other values of t_f). This means that the inverse in equation (7.35) is not defined. If we call $\boldsymbol{b} = \boldsymbol{r}(t_f) - \Phi_{rr}(t_f) \cdot \boldsymbol{r}(0)$, for these values of t_f, only those vectors $\boldsymbol{b}$ in the image of $\Phi_{rv}(t_f)$ yield a valid solution $\boldsymbol{v}(0^+)$. If a solution $\boldsymbol{v}^*(0^+)$ exists for equation (7.35), any other solution $\boldsymbol{v}(0^+) = \boldsymbol{v}^*(0^+) + \boldsymbol{c}$, where $\boldsymbol{c}$ is a vector of the kernel of $\Phi_{rv}(t_f)$, is also a solution. Typically, we want to choose the solution that minimizes $\Delta \boldsymbol{v}_1$.

8 Space Vehicles

This chapter firstly summarizes basic information on the space environment and the impact that it has on the design of space vehicles. Secondly, it continues with basic advice on the design and sizing of some of the fundamental subsystems of a spacecraft. The chapter concludes with an overview of the operation of one key type of space system: global navigation satellite systems.

8.1 Space environment

Isothermal atmosphere model Earth's atmosphere is composed of many different layers. Temperature varies in a non-trivial way, inverting its gradient several times. Pressure, however, decreases monotonically as we ascend in the atmosphere. Modeling the atmospheric pressure and density in the upper layers of the atmosphere is important to determine air drag on low-Earth-orbit satellites.

We can pose a very simple model for pressure if we assume that temperature is constant in the atmosphere $T = \text{const}$. This is only an approximation of course, and more advanced models exist that take temperature variation with altitude into account. We call this model the "isothermal atmosphere" model.

Consider a small cylindrical column of air of height $\mathrm{d}h$, and call A the area of its bottom and upper surfaces. If air density is ρ at this height, the weight of this column of air is simply $\rho g A \mathrm{d}h$, where g is the gravity acceleration.

Since this column is in equilibrium (i.e., it is not falling or rising), this weight must be compensated by the pressure difference $\mathrm{d}p$ between its bottom and upper surfaces. Then, we can write the following force balance equation:

$$Ap - A(p + \mathrm{d}p) = \rho g A \mathrm{d}h \Rightarrow \mathrm{d}p = -\rho g \mathrm{d}h \tag{8.1}$$

Using the ideal gas law $p = \rho RT/M$, where $R = 8.314$ J/(K·mol) is the universal gas constant and M is the gas molecular mass (0.029 kg/mol for air), we can substitute ρ and integrate this equation to obtain an exponential evolution of pressure with height:

$$\frac{\mathrm{d}p}{p} = -\frac{gM}{RT}\mathrm{d}h \Rightarrow p = p_0 \exp\left(\frac{-h}{h_0}\right) \tag{8.2}$$

where p_0 is the pressure at $h = 0$ and $h_0 = RT/(gM)$.

Solar radiation The Sun emits a tremendous amount of electromagnetic radiation at all frequencies and in all directions. At the Earth, we receive about

$$S_\oplus = 1366 \text{ W/m}^2 \tag{8.3}$$

of radiation. Naturally, the power per unit area (i.e. the *irradiance*) increases as we get closer to the Sun, and decreases as we get away from it.

Due to conservation of energy, the integrated irradiance $S(r)$ over a the surface of a sphere of radius r from the Sun is constant, independently of the chosen value of r:

$$4\pi r^2 S(r) = 4\pi r_\oplus S_\oplus \Rightarrow S(r) = S_\oplus \frac{r_\oplus^2}{r^2} \tag{8.4}$$

where $r_\oplus = 150$ million km (1 astronomical unit) is roughly the distance between the Earth and the Sun. This formula is useful to compute the solar irradiance anywhere in the solar system, if we know the distance to the Sun, r.

Solar irradiance affects how much electric power we can generate with solar arrays, and it defines how hot our spacecraft will get when illuminated. It also affects how much solar radiation pressure our spacecraft will feel.

8.2 Space systems

A space system consists of several segments working together, where at least one of them is spaceborne. The common segments of a space system are:

1. The space segment: this is the spacecraft (or the various spacecraft) that are in orbit
2. The ground segment: the control centers and ground stations used to track, monitor, and command the space segment
3. The launcher segment: the launch vehicles used to set the space segment in orbit.

Within the space segment, each spacecraft can be decomposed into the **payload** (i.e., the instruments, transponders, etc that the client wants to have in orbit) and the **spacecraft platform or bus** (consisting of all the subsystems needed to support the operation of the payload and keep the satellite alive).

There are several spacecraft platform subsystems:

1. Power subsystem: in charge of producing, accumulating, distributing electric power to the different loads.
2. Telecommunications: uplink (telecommand) and downlink (telemetry and mission data) require of a communication subsystem to work.
3. Attitude determination and control subsystem: this subsystem determines the current orientation (attitude) of the spacecraft using sensors, and then changes it to the desired orientation using actuators like momentum wheels.
4. Propulsion subsystem: needed if the satellite must perform propulsive maneuvers. Every now and then, small corrective maneuvers are required to maintain the same operational orbit, or to desaturate the ADCS.
5. Data handling: the on board computer must process the incoming commands from the ground segment, monitor and control all the other subsystems, produce the telemetry that will be sent to the ground, and store and sometimes process the mission data until it can be delivered to a ground station.
6. Structural subsystems: the structure is in charge of maintaining the geometrical distribution of the spacecraft, and enduring the stresses that occur during launch.
7. Thermal control subsystem: this subsystem must deal with the changing environmental conditions (sunlight/eclipse, different spacecraft orientations with respect to the Sun), and maintain the temperature of each component within the allowable ranges at all times.
8. Environmental control and life support: In case of manned missions, it is necessary to provide air, water, and food to the crew, keep temperature and humidity at reasonable levels, and to dispose or recycle the generated waste.

8.3 Spacecraft subsystem analysis and design

Telecommunications: link budget equation When trying to reach the ground from a spacecraft, the signal in our communication must be sufficiently focused and have enough power for the ground antenna to be able to receive it and tell the signal apart from the background electromagnetic noise.

If we have determined that the minimum energy per bit of information for a successful communication link is E_b, we can write down the link budget equation as follows:

$$E_b \leqslant \frac{P L_t L_r L_a L_s G_t G_r}{R} \tag{8.5}$$

where:

- P is the power, i.e. energy per unit time, spent by the spacecraft in sending the signal
- L_t are any efficiency losses in the transmitter. E.g., if only 80% of the input power eventually becomes the signal power, then $L_t = 0.8$.
- L_r contemplates any similar losses at the receiving end of the link.

- L_a are any losses that may exist due to the propagation of the signal through the atmosphere. The atmosphere may damp or absorb part of the signal power, meaning that only a fraction of the power makes it to the ground. The absorption depends on the wavelength of the signal, λ.
- L_s are the free space losses, which scale with the transmission distance d. Since the emitted power at the satellite expands outwards roughly spherically, the power per unit area decreases as $1/d^2$. L_s is formally defined as $L_s = \lambda^2/(4\pi d)^2$.
- G_t is the gain of the transmitting antenna, i.e., how well the antenna concentrates the radiated power in the desired direction, rather than emitting in all directions. It is defined as $G_t = 4\pi A_t/\lambda^2$, where A_t is the effective area of the antenna.
- G_r is the gain of the receiving antenna, i.e., how well the antenna selectively listens only to the desired direction, rather than all directions. It is defined as $G_r = 4\pi A_r/\lambda^2$, where A_r is the effective area of the antenna.
- R is the data rate of the transmission, i.e., how many bits per unit time are being sent.

The product PL_tG_t is known as the *effective isotropic radiated power* (EIRP).

Rotational dynamics: attitude determination and control The rotational dynamics of the spacecraft in orbit are governed by the angular momentum equation—the rotational analogous of the (linear) Newton's second law.

$$\frac{\mathrm{d}\boldsymbol{H}_G}{\mathrm{d}t} = \boldsymbol{M}_G \tag{8.6}$$

where $\boldsymbol{H}_G$ is the total angular momentum of the spacecraft about its center of mass G, and $\boldsymbol{M}_G = \sum(\boldsymbol{r} - \boldsymbol{r}_G) \times \boldsymbol{F}$ the moment of all forces about the center of mass.

While this equation may seem complicated, the important aspect to be understood in this course is that in the absence of external moments of forces (i.e., if $\boldsymbol{M}_G = 0$), the total angular momentum of the spacecraft is conserved. This means that we cannot change its state of rotation, unless the spacecraft is composed of two or more bodies and we transmit angular momentum from one body to another, or unless we use propulsion to change the total angular momentum.

For instance, imagine a cubic spacecraft floating in empty space, with a wheel in its interior. If the cube is not rotating initially, we can make it rotate about the axis of the wheel if we spin up or down the wheel. Observe that during the whole maneuver, the total angular momentum (sum of the angular momentum of the cube and the wheel) remain constant.

This mechanism is used in the ADCS to control the orientation and rotation state of the spacecraft in the presence of perturbations that create a moment of forces. However, when the angular momentum stored in the wheels is too large, the rotational speed of the wheels may reach its maximum, and we need to use the propulsion system to *desaturate* the wheels and start over.

Solar array and batteries sizing The most common electric power subsystem in a spacecraft is a combination of solar arrays that serve as a primary power source to produce power, plus a set of batteries as a secondary power source to store and manage power when the primary source is not available.

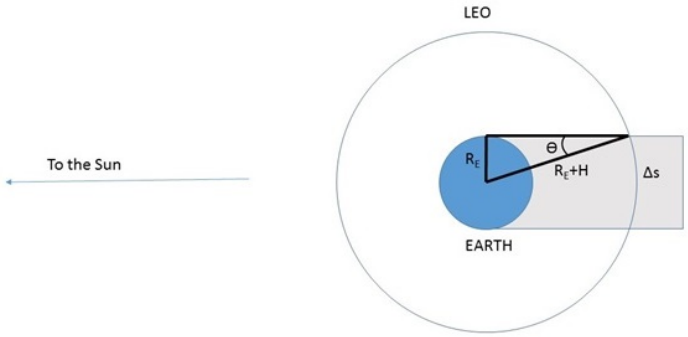
Solar arrays only convert a fraction of the incoming solar power into electric power: inefficiencies depend on the solar cell technology, the solar array geometry and temperature, the orientation of the solar arrays with respect to the Sun, and the age of the system (as solar cells degrade with time). All in all, we can write the produced electric power by the solar arrays (SA) as:

$$P_{SA} = \eta_{SA} A_{SA} S \tag{8.7}$$

where:

- η_{SA} is the overall solar array efficiency.
- A_{SA} is the area of the solar array.
- S is the solar irradiance at the spacecraft location, i.e. the power per unit area of the solar radiation. The SI units are W/m^2. For spacecraft orbiting the Earth, $S = S_\oplus \simeq 1366$ W/m^2. If the spacecraft travels closer to the Sun (e.g. to Venus or Mercury), this number increases; if it travel away from the Sun (e.g. to Mars or Jupiter), this number decreases.

The solar arrays must produce electricity to power the spacecraft systems during the sunlight period, and also to charge the batteries to sustain power consumption during the eclipse. An Earth satellite spends some time in the sunlight and some time in eclipse in each orbit. We can estimate the eclipse time in the case of a circular orbit assuming that the shadow of the Earth is cylindrical away from the Sun, and that the satellite track crosses fully this shade cylinder. The angle θ in the figure is $\theta = \arcsin[R_\oplus/(R_\oplus + H)]$, where $R_\oplus$ is the radius of the Earth and H the orbital altitude of the spacecraft. The fraction of the orbit that the spacecraft is in eclipse is then equal to $2\theta/(2\pi)$. For LEO satellites, this can be about 1/3 of the orbit.



If a satellite requires an electric power P continuously to operate, then the solar arrays must provide a power

$$P_{SA} = P\frac{T_{orb}}{T_{light}} \tag{8.8}$$

where $T_{orb} = 2\pi\sqrt{GM_\oplus/(R_\oplus + H)}$ is the orbital period around the Earth (G is the gravitational constant, and $M_\oplus$ is Earth's mass), i.e., the time it takes the satellite to complete one orbit, and T_{light} is the time spent in sunlight in one orbit.

In turn, the batteries must be sized so store and deliver every orbit the following energy:

$$E = PT_{eclipse} \tag{8.9}$$

where $T_{eclipse}$ is the time spent in eclipse each orbit. Batteries must not undergo full charge/discharge cycles often, as this will severely limit their durability. To extend the useful life of the batteries for the duration of the spacecraft mission, we must limit the *depth of discharge* (DOD) of the batteries to a given fraction that depends on the battery technology and the number of charge/discharge cycles expected in the mission (i.e., the total number of eclipses expected during the mission). Once the design DOD is known, the actual energy of the batteries can be sized as

$$E_{batt} = \frac{E}{DOD}. \tag{8.10}$$

Thermal control in space In space, the only heat transfer mechanism that exists is radiation. Also, within the spacecraft, conduction between the different, physically connected parts, can transmit heat.

The balance between all the heat that is generated in one component A by dissipation of electric power, plus any heat coming from the surrounding space environment and neighboring components into A, minus the heat that is conducted or radiated away from A, is related to the rate at which the temperature of component A changes in time:

$$C_A\frac{\mathrm{d}T_A}{\mathrm{d}t} = Q_{in,int} + Q_{in,env} - Q_{out}. \tag{8.11}$$

In this thermal energy balance equation, C_A is the thermal capacity of the component A, i.e. how much heat must be provided to this component to make its temperature increase by a given amount. The SI units of C_A are W/K.

Each component in the spacecraft has one such equation. The combination of all these coupled equations can be solved simultaneously in a computer, to find the evolution of the temperature of

each component under different operating conditions and environmental heat loads (eclipse, sunlight, etc).

If we consider an isolated component with no neighboring elements to conduct or radiate heat to and from, the term $Q_{in,env}$ represents the incoming radiation heat from space, and the term Q_{out} represents the outgoing radiation heat emitted from the object into space. Then, the external heat sources from the environment are the solar radiation, the planetary albedo (i.e. the reflection of solar radiation off the planet surface), and the planetary IR radiation (i.e. the heat emitted by the planet itself since it has a non-zero temperature). The background empty space can be considered to be at 4 K, and can be ignored in most situations. Of these loads, the Sun is the dominant one:

$$Q_{in,env} = Q_{in,Sun} + Q_{in,albedo} + Q_{in,planet} \tag{8.12}$$

$$Q_{in,Sun} = \alpha_{Sun} A_\alpha S \tag{8.13}$$

where:

- α_{Sun} is the solar absorptivity, i.e., what fraction of the incoming solar irradiance is absorbed by the component (the rest is reflected). It is a non-dimensional number that ranges from 0 (no absorption) to 1 (maximum absorption).
- A_α is the effective area of absorption of the component that is exposed to sunlight.
- S is the solar irradiance at the spacecraft location as explained before, i.e. the power per unit area of the solar radiation.

The heat emitted from tan isolated component is its radiation into space. Since the component is typically at a few hundreds of degrees Kelvin, its radiation is essentially in the infrared part of the spectrum, and the emitted radiation heat can be expressed as:

$$Q_{out} = \varepsilon_{IR} A_\varepsilon \sigma T_A^4 \tag{8.14}$$

where:

- ε_{IR} is the emissivity of the component in the infrared. It is a non-dimensional number that ranges from 0 (no emission) to 1 (maximum emission). The maximum emission is that of a so-called *black body* at the same temperature as the object.
- A_ε is the effective area for radiation of the component.
- σT_A^4 is the power radiated as heat of a perfect black body at a temperature T_A. The Stefan-Boltzmann constant is $\sigma = 5.67 \cdot 10^{-8}$ W/(m^{2}K^4).

Different materials have different optical properties α_{Sun} and ε_{IR}. By correctly selecting the materials of the surfaces of a spacecraft, we can exert some control over its temperature when exposed to the Sun and/or during eclipses (i.e. no Sun).

Rocket propulsion As the third law of Newton states, for every action there is an equal but opposite reaction elsewhere. By ejecting propellant, a rocket exploits this law to create a thrust force on the vehicle equal to the momentum expelled from the rocket per unit time:

$$F = \dot{m} c, \tag{8.15}$$

where $\dot{m} = -\mathrm{d}m/\mathrm{d}t$ is the mass flow rate at which propellant is being ejected, and c the exhaust velocity of the propellant leaving the rocket. The velocity c is also known as **specific impulse**, I_{sp}. Due to historical reasons, the I_{sp} is sometimes expressed in seconds instead of in velocity units, dividing c by $g_0 = 9.81$ m/s^2, the gravity acceleration on the ground:

$$I_{sp}^v = c \text{ (in velocity units)}; \quad I_{sp}\,[\mathrm{s}] = \frac{c\,[\mathrm{m/s}]}{g_0\,[\mathrm{m/s^2}]} \text{ (in seconds)}. \tag{8.16}$$

Tsiolkovsky's rocket equation The equation of motion of a rocket system of mass m and velocity v with thrust $F = \dot{m} c$ is:

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = \dot{m} c = -\frac{\mathrm{d}m}{\mathrm{d}t} c \tag{8.17}$$

integrating this equation between two instants of time t_1 and t_2, between which the rocket velocity changes from v_1 to v_2, and the rocket mass changes from m_1 to m_2, requires some knowledge of differential equations. The result is:

$$\frac{\mathrm{d}v}{c} = \frac{\mathrm{d}m}{m} \Rightarrow \frac{v_2 - v_1}{c} = \ln\left(\frac{m_1}{m_2}\right). \tag{8.18}$$

In this equation, "ln" is the natural logarithm. We define the quantity $\Delta v = v_2 - v_1$, pronounced *delta-vee*, as the velocity increase obtained thanks to the rocket, which has used a mass of propellant equal to $m_2 - m_1$. Each space mission has a well-defined "cost" in terms of the necessary Δv to accomplish it.

This equation can be inverted using the exponential function "exp" into

$$\frac{m_2}{m_1} = \exp\left(-\frac{\Delta v}{c}\right) \equiv \exp\left(-\frac{\Delta v}{I_{sp} g_0}\right). \tag{8.19}$$

This equation shows that unless the Δv of the mission that we want to carry out is smaller or comparable to the $I_{sp}g_0$ of our rocket technology, the final mass m_2 will be tiny compared to the initial mass m_1 of the rocket. Typically, sending a small satellite into space requires a large rocket launcher for this reason.

Rocket staging Tsiolkovsky's rocket equation includes in m_2 not only the payload m_{pay} of the rocket (e.g. the satellite we want to put into orbit), but also any inert mass in the system such as structural mass of the rocket itself, m_{strut}. Given the adverse scaling of the rocket size, anything that can reduce inert mass is desirable.

Staging consists in releasing structural mass as soon as it is no longer needed (e.g. depleted propellant tanks, used-up rockets, etc). This way, the remaining of the trip is done with a lower inert mass, improving the overall mass performance of the rocket system. In effect, this is the same as stacking several rockets on top of each other, and igniting them in series.

To solve staging problems, it is useful to write down the initial mass of the i-th rocket stage as

$$m_{0,i} = m_{fuel,i} + m_{struct,i} + m_{pay,i}, \tag{8.20}$$

and note that the payload of the i-th stage is actually the next stage, $(i + 1)$. The advantages of staging are great for systems with up to 3–4 stages, but beyond that, the reduction of the initial system mass is very limited.

Impulsive maneuvers Rocket maneuvers allow to change the trajectory (and thus the orbit) of a spacecraft. Chemical rockets provide large thrust levels for short periods of time; hence, we can simplify the analysis of chemical rocket maneuvers by imagining that the spacecraft changes instantaneously its velocity vector $\boldsymbol{v}$ when the maneuver takes place. This approximation is known as the *impulsive maneuver model*. Using vector operations, velocity changes from $\boldsymbol{v}_1$ to $\boldsymbol{v}_2 = \boldsymbol{v}_1 + \Delta\boldsymbol{v}$, while the position vector $\boldsymbol{r}$ remains unchanged during the instantaneous rocket firing.

The Hohmann transfer explained in the course is an example of application of the impulsive maneuver model.

Key parameters of electric propulsion While chemical propulsion is *energy limited*, as its performance depends on the amount of energy per unit mass stored in the chemical bonds of the propellant, electric propulsion is *power limited*: the performance depends on how much electric power is available on board.

As with chemical propulsion, there are two basic parameters that characterize an electric thruster:

1. Specific impulse I_{sp}: this is a measure of the velocity at which the propellant is ejected from the thruster. A higher specific impulse allows fulfilling a propulsive mission (i.e. providing a given ΔV to the spacecraft) at a much lower expense of propellant. Electric propulsion devices have I_{sp}^v in the order of 10–100 km/s (about 1000–10000 s if $I_{sp}$is expressed in seconds), whereas the best chemical rockets can only provide about 5 km/s (about 500 s).

2. Thrust F: the force the propulsion system can generate to accelerate the spacecraft. Contrary to chemical propulsion, where F can be very large, electric propulsion thrust levels are comparatively small: current thrusters provide < 1 N of force (typically, about 50–200 mN). Thrust is equal to the propellant mass flow rate $\dot{m}$ used in the thruster times the exhaust velocity, i.e., its specific impulse in velocity units:
$$F = \dot{m} I_{sp}^v. \tag{8.21}$$

Electric propulsion provides tiny thrust levels, but it does so very efficiently, using very little propellant. This enables space missions that would be too ambitious to be carried out only with chemical propulsion, and to lower the cost of existing missions. Note, however, that electric propulsion cannot fully replace chemical propulsion: large thrust levels are required to take off from the surface of a planet, and for some quick propulsive maneuvers.

Apart from I_{sp} and F, there is another important figure of merit of an electric thruster:

3. Thrust efficiency η_T: it is a measure of how the input power P is used for propulsion. The power contained in a jet of mass flow rate $\dot{m}$ and exhaust velocity I_{sp}^v is $\dot{m}(I_{sp}^v)^2/2$. The thrust efficiency is defined as the ratio of the jet power over the input power:
$$\eta_T = \frac{\dot{m}(I_{sp}^v)^2}{2P} = \frac{F I_{sp}^v}{2P} = \frac{F^2}{2\dot{m}P} \tag{8.22}$$

8.4 Global Navigation Satellite Systems

Operation principles of a GNSS position fix The basic ideas behind the operation of a GNSS can be summarized as follows:

1. The satellites of a GNSS constellation each have a very precise on board clock that is synchronized with all other clocks in the constellation.
2. The position vector of each satellite in the constellation is known (or can be made known) to any user of the system at all times.
3. Each satellite broadcasts a time-stamped, identifiable signal in all directions. Each signal encodes, among other pieces of information, the exact time at which the signal was broadcast.
4. Any number of users can receive these signals simultaneously: they are all passive (i.e. receiving) users.
5. A user receiving a satellite signal can compare the broadcast time with his/her own clock: the travel time for the signal $\Delta t = t_{RX} - t_{TX}$ is a direct measure of the distance (or range) d between the user and the emitting satellite, since the speed of light c is a physical constant: $d = c\Delta t$.
6. Assuming the instantaneous position of the satellites is known to the user, using 3 such distance measurements to 3 different satellites allows, in principle, to determine the 3 unknowns x, y, z of the position of the user.
7. However, for this to work, the user would need to have his/her clock perfectly synchronized with the constellation clocks. As this would be impractical, a 4th distance measurement with a 4th satellite is needed to determine the 4 unknowns of the problem: x, y, z and t, the time of the user.

Obtaining a position (and time) fix for a user requires therefore solving a system of 4 non-linear equations, one for each pseudo-range measurement:

$$c^2(t_{RX} - t_{TX,1})^2 = (x_{RX} - x_{TX,1})^2 + (y_{RX} - y_{TX,1})^2 + (z_{RX} - z_{TX,1})^2, \tag{8.23}$$
$$c^2(t_{RX} - t_{TX,2})^2 = (x_{RX} - x_{TX,2})^2 + (y_{RX} - y_{TX,2})^2 + (z_{RX} - z_{TX,2})^2, \tag{8.24}$$
$$c^2(t_{RX} - t_{TX,3})^2 = (x_{RX} - x_{TX,3})^2 + (y_{RX} - y_{TX,3})^2 + (z_{RX} - z_{TX,3})^2, \tag{8.25}$$
$$c^2(t_{RX} - t_{TX,4})^2 = (x_{RX} - x_{TX,4})^2 + (y_{RX} - y_{TX,4})^2 + (z_{RX} - z_{TX,4})^2, \tag{8.26}$$

where $x_{TX,i}, y_{TX,i}, z_{TX,i}$ is the position of the i-th transmitting satellite, $t_{TX,i}$ the time at which the signal from the i-th satellite was sent, and $x_{RX}, y_{RX}, z_{RX}, t_{RX}$ are the unknowns (position and time of the receiving user).

Operation principles of a GNSS velocity fix Once the procedure to determine the position is clear, it would be possible to compute the velocity of a user by observing how his/her position changes in time. In other words, by taking the derivative of the position vector:

$$\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} \simeq \frac{\boldsymbol{r}(t_2) - \boldsymbol{r}(t_1)}{t_2 - t_1}. \tag{8.27}$$

There is, however, another, more accurate way to determine the user's velocity: using the Doppler shift of the signal frequency. This is done according to the following basic ideas:

1. The exact frequency at which the satellite signals are emitted is known to the users of the system.
2. The velocity vector of each satellite in the constellation is known (or can be made known) to any user of the system at all times.
3. Whenever the distance between a satellite and a user changes in time (either because the satellite is moving or the user is moving), the perceived signal frequency by the user changes slightly: the frequency will increase if the two of them are coming closer together, and it will decrease if they are moving apart. This physical phenomenon is known as *Doppler shift*.
4. Any user listening to the satellite signals can determine the Doppler shift of any of them, $\Delta f = f_{RX} - f_{TX}$.
5. As the user knows the velocity of each satellite, combining 3 such measurements with 3 different satellites it is possible to determine the user's velocity (3 unknowns: v_x, v_y, v_z).
6. However, the cheap user's clock may have a drift with time. This may lead the user to wrong measurements of the Doppler shifts. Thus, a 4th measurement with a 4th satellite is needed to determine the drift rate of the user clock as an extra unknown.

Obtaining a velocity (and clock rate) fix for a user requires therefore solving the following system of 4 non-linear equations:

$$\dot{t}_{RX} f_{RX,1} - f_{TX,1} = \frac{f_{TX,1}}{c}(\boldsymbol{v}_{RX} - \boldsymbol{v}_{TX,1}) \cdot \frac{\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,1}}{|\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,1}|} \tag{8.28}$$

$$\dot{t}_{RX} f_{RX,2} - f_{TX,2} = \frac{f_{TX,2}}{c}(\boldsymbol{v}_{RX} - \boldsymbol{v}_{TX,2}) \cdot \frac{\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,2}}{|\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,2}|} \tag{8.29}$$

$$\dot{t}_{RX} f_{RX,3} - f_{TX,3} = \frac{f_{TX,3}}{c}(\boldsymbol{v}_{RX} - \boldsymbol{v}_{TX,3}) \cdot \frac{\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,3}}{|\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,3}|} \tag{8.30}$$

$$\dot{t}_{RX} f_{RX,4} - f_{TX,4} = \frac{f_{TX,4}}{c}(\boldsymbol{v}_{RX} - \boldsymbol{v}_{TX,4}) \cdot \frac{\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,4}}{|\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,4}|} \tag{8.31}$$

where $\boldsymbol{v}_{TX,i}$ is the velocity of the i-th transmitting satellite, $f_{TX,i}, f_{RX,i}$ the frequencies at which the signal from the i-th satellite was respectively broadcast and received, and $\boldsymbol{v}_{RX}, \dot{t}_{RX}$ are the unknowns (position and clock drift rate of the receiving user). Note that the vector $(\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,i})/|\boldsymbol{r}_{RX} - \boldsymbol{r}_{TX,i}|$ is a unit vector that points in the direction from the i-th satellite to the user.

9 Attitude determination and control

Most spacecraft can be accurately modeled as a rigid body. After reviewing the fundamentals of rigid body kinematics, inertia, and dynamics, this chapter explores torque-free motion for a spacecraft in orbit, motion under external torques, in particular the one due to gravity-gradient, and motion with internal momentum exchange devices (e.g. reaction wheels and control moment gyroscopes). We conclude with a brief overview of some key aspects of the attitude determination and control subsystem of a spacecraft. Quaternions, due to their advantages in describing rotational kinematics, are also briefly discussed.

9.1 Rotational kinematics

A rigid body is equivalent to a reference frame from the kinematics point of view. In the following we shall use a body-fixed reference frame $S_b : \{G; B_b\}$ with origin on its center of mass G, and a suitable body-fixed vector basis B_b.

The attitude of a spacecraft in an inertial reference frame S_0 can be given in terms of its 3-1-3 Euler angles[12]. These angles are called ψ, *precession*, θ, *nutation*, and ϕ, *spin*. This way, it is possible to go from the vector basis of the inertial reference frame to the body-fixed vector basis.

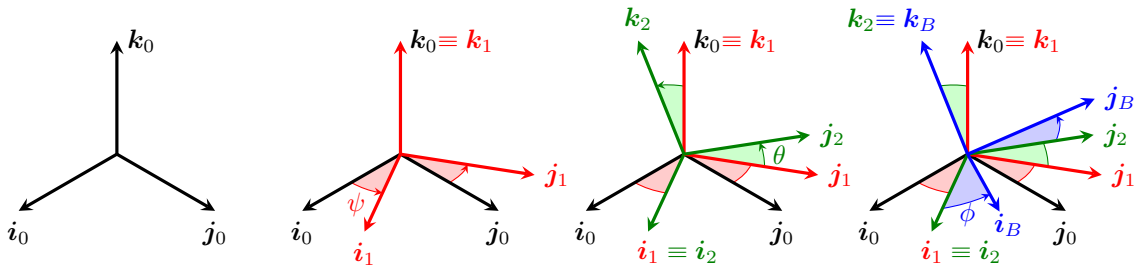


Figure 9.1: Precession, nutation, and spin.

The angular velocity vector $\boldsymbol{\omega}_{b0}$ takes the following form in terms of the derivatives of the Euler angles:

$$\boldsymbol{\omega}_{b0} = (\dot{\theta}\cos\psi + \dot{\phi}\sin\theta\sin\psi)\boldsymbol{i}_0 + (\dot{\theta}\sin\psi - \dot{\phi}\sin\theta\cos\psi)\boldsymbol{j}_0 + (\dot{\phi}\cos\theta + \dot{\psi})\boldsymbol{k}_0 \tag{9.1}$$

$$= (\dot{\theta}\cos\phi + \dot{\psi}\sin\theta\sin\phi)\boldsymbol{i}_b + (-\dot{\theta}\sin\phi + \dot{\psi}\sin\theta\cos\phi)\boldsymbol{j}_b + (\dot{\phi} + \dot{\psi}\cos\theta)\boldsymbol{k}_b. \tag{9.2}$$

The angular acceleration can be obtained by differentiation. Remember that $\boldsymbol{\omega}_{b0}$ satisfies:

$$\boldsymbol{\alpha}_{b0} = \left.\frac{\mathrm{d}\boldsymbol{\omega}_{b0}}{\mathrm{d}t}\right|_0 = \left.\frac{\mathrm{d}\boldsymbol{\omega}_{b0}}{\mathrm{d}t}\right|_b \tag{9.3}$$

because the $\boldsymbol{\omega}_{b0} \times \boldsymbol{\omega}_{b0}$ term vanishes when applying the Coriolis formula.

9.2 Inertia properties

The relevant inertia properties for the rotational dynamics of a rigid body b is the tensor of inertia. With respect to the center of mass G, the tensor of inertia is defined as:

$$\mathcal{I}_G^b = \int \left[(r_b^P)^2\overline{\overline{\mathcal{U}}} - \boldsymbol{r}_b^P \otimes \boldsymbol{r}_b^P\right] \mathrm{d}m = \begin{bmatrix} I_x & -P_{xy} & -P_{xz} \\ -P_{xy} & I_y & -P_{yz} \\ -P_{xz} & -P_{yz} & I_z \end{bmatrix}_1, \tag{9.4}$$

where P is a point that sweeps the whole body inside the integral. The matrix expression has the components of the tensor of inertia in a given basis B_1; these are called moments (e.g. I_x) and

[12] This is not the only way to parametrize the attitude of a rigid body. See, for instance, quaternions, described in §9.8.

products (e.g. P_{xy}) of inertia. The tensor of inertia is symmetric positive definite, and there is a basis B_* in which the tensor is diagonal: such basis shows the *principal directions of inertia*. The directions of the vectors of this basis B_* are the principal directions of inertia and correspond to the eigenvectors of $\mathcal{I}_G^b$. The moments of inertia in this basis are the principal moments of inertia, and correspond to the eigenvalues of $\mathcal{I}_G^b$, I_1, I_2 and I_3.

Depending on the value of these principal moments of inertia, we distinguish the following classes of bodies with regards to their inertia properties:

1. Bodies with spherical-symmetric inertia properties: $I_1 = I_2 = I_3$.
2. Bodies with axisymmetric inertia properties: $I_1 \neq I_2 = I_3$. We can subdivide these objects into oblate $I_1 > I_2 = I_3$ and prolate $I_1 < I_2 = I_3$.
3. Bodies with asymmetric inertia properties: $I_1 > I_2 > I_3$ (we customarily order the moments of inertia from largest to smallest).

Steiner's theorem allows computing the tensor of inertia about one point given the tensor of inertia about another. Tensors of inertia about the same point are additive. Naturally, all operations with tensor components must be carried out in the same basis.

9.3 Rotational Dynamics

The angular momentum of the spacecraft b about its center of mass G with respect to the reference frame S_0 is:

$$\boldsymbol{H}_{G0}^b = \mathcal{I}_G^b \cdot \boldsymbol{\omega}_{b0} \tag{9.5}$$

Euler equations describe the time evolution of $\boldsymbol{H}_{G0}^b$:

$$\left.\frac{\mathrm{d}\boldsymbol{H}_{G0}^b}{\mathrm{d}t}\right|_0 = \boldsymbol{M}_G, \tag{9.6}$$

where $\boldsymbol{M}_G$ is the resultant external torque about G acting on b. Since the expression of $\boldsymbol{H}_{G0}^b$ in terms of $\mathcal{I}_G^b$ is easier to work with in reference frame S_b, because the components of $\mathcal{I}_G^b$ are constant in the basis B_b, we typically write this equation in S_b as:

$$\left.\frac{\mathrm{d}\boldsymbol{H}_{G0}^b}{\mathrm{d}t}\right|_b = \boldsymbol{M}_G - \boldsymbol{\omega}_{b0} \times \boldsymbol{H}_{G0}^b \qquad \Rightarrow \qquad \mathcal{I}_G^b \cdot \boldsymbol{\alpha}_{b0} = \boldsymbol{M}_G - \boldsymbol{\omega}_{b0} \times (\mathcal{I}_G^b \cdot \boldsymbol{\omega}_{b0}) \tag{9.7}$$

Below and in following sections, we will assume that our body basis B_b is principal of inertia, so $\mathcal{I}_G^b$ is diagonal with moments of inertia I_x, I_y and I_z along $\boldsymbol{i}_b$, $\boldsymbol{j}_b$ and $\boldsymbol{k}_b$ respectively. Also, we will use the following nomenclature for the components of vector in the body-fixed vector basis B_b: $\boldsymbol{\omega}_{b0} = \omega_x \boldsymbol{i}_b + \omega_y \boldsymbol{j}_b + \omega_z \boldsymbol{k}_b$, $\boldsymbol{\alpha}_{b0} = \alpha_x \boldsymbol{i}_b + \alpha_y \boldsymbol{j}_b + \alpha_z \boldsymbol{k}_b$, and $\boldsymbol{M}_G = M_x \boldsymbol{i}_b + M_y \boldsymbol{j}_b + M_z \boldsymbol{k}_b$.

With these prescriptions, equation (9.7) projected along the vectors of this basis yields:

$$I_x \alpha_x = M_x + (I_y - I_z)\omega_y \omega_z, \; I_y \alpha_y = M_y + (I_z - I_x)\omega_x \omega_z, \tag{9.8}$$

$$I_z \alpha_z = M_z + (I_x - I_y)\omega_x \omega_y, \tag{9.9}$$

As it can be observed, the rotational dynamics about the three principal axes are coupled together, in the general case of an asymmetric body.

The kinetic energy of the rigid body b with respect to reference frame $S_0 : \{O; B_0\}$ is:

$$T_0^b = \frac{1}{2}(v_0^G)^2 + \frac{1}{2}\boldsymbol{\omega}_{b0} \cdot \mathcal{I}_G^b \cdot \boldsymbol{\omega}_{b0} = T_{0,T}^b + T_{0,R}^b \tag{9.10}$$

where we have identified the translational kinetic energy $T_{0,T}^b$ and the rotational kinetic energy $T_{0,R}^b$. Observe that $T_{0,R}^b$ is the kinetic energy in the auxiliary, non-rotating reference frame centered at G with the basis of S_0, i.e., $S_G : \{G; B_0\}$:

$$T_{0,R}^b \equiv T_G^b = \frac{1}{2}\boldsymbol{\omega}_{b0} \cdot \mathcal{I}_G^b \cdot \boldsymbol{\omega}_{b0}. \tag{9.11}$$

If we compute T_G^b in terms of the vector and tensor components in the principal vector basis, its expression is particularly simple:

$$T_G^b = \frac{1}{2}(I_x\omega_x^2 + I_y\omega_y^2 + I_z\omega_z^2). \tag{9.12}$$

9.4 Torque-free motion

In this and following sections we study the rotational motion of the spacecraft in the auxiliary reference frame S_G introduced above. Note that, in general, this reference frame is non-inertial, as the origin is accelerated. Luckily, the force of inertia due to the acceleration of the origin does not induce any torque of inertia, as can be easily checked[13]. Also, unless otherwise noted, we will assume that the body-fixed vector basis B_b is the principal basis of inertia described used in the previous section.

When $\boldsymbol{M}_G = \boldsymbol{0}$ we speak of torque-free motion. This is the problem of Poinsot. We can extract a great deal of information on the motion of a body in this condition without fully solving its equations of motion, by using conservation laws.

Since there are no torques acting on b, T_G^b is a conserved quantity of motion. We can compute the value T_G^b from the initial conditions:

$$\frac{I_x}{2}\omega_x^2 + \frac{I_y}{2}\omega_y^2 + \frac{I_z}{2}\omega_z^2 = T_G^b \tag{9.13}$$

This equation can be regarded as a geometrical constraint on $\boldsymbol{\omega}_{b0}$, whose tip must lie on the surface of a body-fixed ellipsoid of principal semi axes:

$$\sqrt{\frac{2T_G^b}{I_x}}, \quad \sqrt{\frac{2T_G^b}{I_y}}, \quad \text{and} \quad \sqrt{\frac{2T_G^b}{I_z}}. \tag{9.14}$$

This is known as *Poinsot's ellipsoid*, or the *inertia ellipsoid* of the rigid body.

The conservation of $\boldsymbol{H}_{G0}^b$ yields further insight. We can write:

$$\boldsymbol{H}_{G0}^b \cdot \omega_{b0} = 2T_G^b. \tag{9.15}$$

This equation shows that the projection of ω_{b0} (a variable vector) along $\boldsymbol{H}_{G0}^b$ (a constant vector in B_0) constant. Geometrically, this can also be interpreted as a constraint on $\boldsymbol{\omega}_{b0}$, whose tip must lie on an inertial plane perpendicular to $\boldsymbol{H}_{G0}^b$ and separated a distance $2T_G^b/H_{G0}^b$ from the origin.

These geometrical constraints must be understood in the vector space where vector $\boldsymbol{\omega}_{b0}$ lives. So far, we have seen that $\boldsymbol{\omega}_{b0}$ must lie simultaneously on a body-fixed ellipsoid and on an inertial plane. It is straightforward to see that these two figures are tangent to each other. If we find the gradient of the ellipsoid with respect to $\boldsymbol{\omega}_{b0}$,

$$\frac{\mathrm{d}T_G^b}{\mathrm{d}\boldsymbol{\omega}_{b0}} = \boldsymbol{H}_{G0}^b. \tag{9.16}$$

so its normal vector is in the direction of the normal vector of the plane.

We conclude that the rotational motion of the rigid body is equivalent in all aspects to the motion of the ellipsoid of inertia as it rolls without slipping over the $\boldsymbol{H}_{G0}^b$ plane. The path drawn by $\boldsymbol{\omega}_{b0}$ on the ellipsoid is called the *polhode*, and the path drawn on the plane is called the *herpolhode*.

Finally, to better understand the torque-free motion, we resort to a slightly different construction. Let $\boldsymbol{H}_{G0}^b = H_x\boldsymbol{i}_b + H_y\boldsymbol{j}_b + H_z\boldsymbol{k}_b$ be the expression of the angular momentum vector in the body-fixed reference frame. Observe that the components H_x, H_y and H_z are not constant. But, the magnitude of $\boldsymbol{H}_{G0}^b$ is. Hence, we can write:

$$(H_{G0}^b)^2 = H_x^2 + H_y^2 + H_z^2, \tag{9.17}$$

$$T_G^b = \frac{1}{2I_x}H_x^2 + \frac{1}{2I_y}H_y^2 + \frac{1}{2I_z}H_z^2. \tag{9.18}$$

[13] Exercise: demonstrate this statement by integrating the differential torque on a elemental mass of the rigid body.

These equation can be interpreted as geometric constraints of $\boldsymbol{H}_{G0}^b$ in its own vector space, namely, that the tip of vector $\boldsymbol{H}_{G0}^b$ must remain at the same time on the surface of a sphere of radius H_{G0}^b and on the surface of an ellipsoid (caution: this ellipsoid is different from the inertia ellipsoid discussed before). The intersection between the sphere and the ellipsoid gives the possible trajectories for the $\boldsymbol{H}_{G0}^b$ vector in the body-fixed reference frame. We make the following observations, assuming that $I_x > I_y > I_z$, the general case of a body with asymmetric inertia properties:

1. If $\boldsymbol{\omega}_{b0}$ is exactly oriented along a principal axis, it is parallel to $\boldsymbol{H}_{G0}^b$. This is an equilibrium rotation for $\boldsymbol{\omega}_{b0}$, i.e., its time variation is null. We say that these are pure rotations about the principal axes of inertia.

2. If $\boldsymbol{\omega}_{b0}$ is almost, but not exactly, oriented along the major or minor principal axes of inertia, i.e. those with maximal or minimal moment of inertia I_1 and I_3, respectively, then equation (9.17) and (9.18) intersect in a roughly circular path: $\boldsymbol{\omega}_{b0}$ precesses about the inertial direction of $\boldsymbol{H}_{G0}^b$, but it does not deviate too much from the equilibrium position that corresponds to that principal axis of inertia. We conclude that the pure rotation about the major and minor axes of inertia are stable.

3. If $\boldsymbol{\omega}_{b0}$ is almost, but not exactly, oriented along the intermediate (I_2) axis of inertia, then equation (9.17) and (9.18) intersect in roughly hyperbolic path with origin in the direction of the intermediate principal axis of inertia. Even a tiny initial deviation from the equilibrium position about this axis grows large over time. We conclude that the pure rotation about the intermediate axis of inertia is unstable, as any small perturbation that takes $\boldsymbol{\omega}_{b0}$ from its equilibrium orientation will initiate large excursions[14].

These conclusions are modified if any mechanical dissipation exists that damps rotational kinetic energy over time. The dissipation of kinetic rotational energy could be due, for example, to the sloshing of liquids in the propellant tanks of the spacecraft (which then is not exactly a rigid body). Note that, while T_G^b may vary over time due to these effects, H_{G0}^b can only change if there are external torques. The dissipation of T_G^b makes the ellipsoid of equation (9.18) to shrink over time, and the quasi-circular paths of the previous discussion are modified into spiralling paths. The new conclusion is that the minor axis of inertia becomes unstable, and only the major axis of inertia remains stable: this is the rotation that allows to minimize the kinetic energy T_G^b for a given $\boldsymbol{H}_{G0}^b$.

The Explorer-1 mission is an important example of how this phenomenon can make a space mission go wrong. This spacecraft was an elongated, pencil-like object, designed to rotate about its longitudinal axis (minimum inertia principal direction). As a consequence of kinetic energy dissipation by the slightly flexible structural components, the spacecraft started to precess and eventually to tumble, until it was rotating about its major axis of inertia.

9.5 External torques

External torques appear on the right hand side of equation (9.7), and change the total angular momentum of the spacecraft. These torques may be exerted by the spacecraft on purpose, using external-torque devices (see §9.7), or by external perturbations, such as:

1. Drag. If the spacecraft has an asymmetric forward surface, drag forces can exert a small torque.

2. Solar radiation pressure. Similarly, solar radiation can exert a torque.

3. Magnetic field. If the spacecraft has some magnetic dipole (either residual in the structure and components, or intentionally generated), then it will interact with the geomagnetic field and result in a torque.

4. Gravity gradient. The small variation of the strength of gravity with altitude means that the center of gravity does not coincide with the center of mass, resulting in a small torque.

Albeit small, external perturbation torques exert their influence over long periods of time, giving rise to oscillatory and secular variations of the angular momentum of the spacecraft.

[14] This gives rise to the curious, but perfectly analytical and well-understood, Dzhanibekov effect.

9.5.1 Gravity gradient

To better understand gravity gradient torque, consider the gravitational pull due to Earth on each mass differential of the rigid body, and the torque it generates about G. Expressing the position of a generic point of the rigid body as $\boldsymbol{r}_0^P = \boldsymbol{r}_0^G + \boldsymbol{r}_G^P$ we have:

$$\mathrm{d}F = -\frac{\mu}{|\boldsymbol{r}_0^G + \boldsymbol{r}_G^P|^3}(\boldsymbol{r}_0^G + \boldsymbol{r}_G^P)\mathrm{d}m \tag{9.19}$$

$$\mathrm{d}M_G = \boldsymbol{r}_G^P \times \mathrm{d}F = -\frac{\mu}{|\boldsymbol{r}_0^G + \boldsymbol{r}_G^P|^3}(\boldsymbol{r}_G^P \times \boldsymbol{r}_0^G)\mathrm{d}m. \tag{9.20}$$

Expanding the denominator and keeping only first order terms on $\boldsymbol{\rho}$, and calling $R = r_0^G$:

$$\mathrm{d}M_G \simeq -\frac{\mu}{R^3}\left(1 - 3\frac{\boldsymbol{r}_0^G \cdot \boldsymbol{r}_G^P}{R^2}\right)(\boldsymbol{r}_G^P \times \boldsymbol{r}_0^G)\mathrm{d}m. \tag{9.21}$$

When we integrate over the body mass distribution we find

$$\int \mathrm{d}M_G \simeq -\frac{\mu}{R^3}\int \boldsymbol{r}_G^P \mathrm{d}m \times \boldsymbol{r}_0^G + 3\frac{\mu}{R^5}\int \boldsymbol{r}_0^G \cdot \boldsymbol{r}_G^P(\boldsymbol{r}_G^P \times \boldsymbol{r}_0^G)\mathrm{d}m \tag{9.22}$$

The first term vanishes due to the definition of center of mass. Writing $\boldsymbol{r}_0^G$ in the body-fixed principal basis of inertia B_b as $\boldsymbol{r}_0^G = X\boldsymbol{i}_b + Y\boldsymbol{j}_b + Z\boldsymbol{k}_b$ the second term can be integrated to yield, in basis B_b:

$$M_x = 3\frac{\mu YZ}{R^5}(I_z - I_y), \qquad M_y = 3\frac{\mu XZ}{R^5}(I_x - I_z), \qquad M_z = 3\frac{\mu XY}{R^5}(I_y - I_x). \tag{9.23}$$

We conclude that the gravity gradient torque is directly related with the differences between principal moments of inertia, and that it scales as $1/R^3$. It is zero if one of the principal axes is aligned with the local vertical, which yield equilibrium attitudes. However, the only stable equilibrium attitude is with the minor axis of inertia is oriented along the local vertical.

9.6 Momentum exchange devices

Virtually all satellites are equipped with some form of momentum exchange devices: reaction wheels or control moment gyros are examples of this. In essence, these devices consist of a spinning disk inside the satellite, supported by a frame. A motor can spin up or spin down the wheel, effectively exchanging angular momentum between the wheel and the rest of the spacecraft. In more advanced devices, the direction of spin with respect to the spacecraft reference can also be altered with additional motors.

The study of a spacecraft with a momentum exchange device is always done in a similar way: we must treat the wheel w and the rest of the spacecraft b as separate rigid bodies. If we analyze the system composed by the spacecraft and the wheel, $b + w$, then the motor and frame torques between them are as internal torques to the system of analysis, and they do not appear in the right hand side of equation (9.7). Assuming that no external torques exist, the total angular momentum of the system is conserved:

$$\boldsymbol{H}_{G0}^{b+w} = \boldsymbol{H}_{G0}^{b} + \boldsymbol{H}_{G0}^{w}. \tag{9.24}$$

However, when we analyze separately the wheel w or the rest of the spacecraft b as different systems, then the torque of the motor and the reaction torque on the frame that supports the spinning wheel appear as external torques on the system of analysis and appear in the right hand side of equation (9.7). By themselves, the angular momentum of the rest of the spacecraft b, or the angular momentum of the wheel w, are not conserved and change in time as angular momentum is transferred from one body to the other.

TBC

9.7 Determination and control

The attitude determination and control subsystem (ADCS) of a spacecraft is responsible for identifying the current attitude and rotation state of the spacecraft, and for correcting and maintaining the desired attitude for the objectives of the mission.

To achieve its goals, the ADCS has sensors that allow it to determine the attitude of the spacecraft, and a set of actuators that affect the attitude of the spacecraft. Crucial to its operation, it also has filters and algorithms for the navigation, guidance, and control (GNC) of the attitude problem.

Different types of sensors exist. The following is just a list of examples:

1. Sun sensors allow determining the sun vector.

2. Horizon sensors watch the Earth IR radiation and allow determining the Earth vector.

3. Star trackers compare the visual of the sky with a starmap to determine the current orientation of the startracker.

4. Inertial measurement units with gyroscopes allow determining the angular rate of the spacecraft, and do not require any external references to do so. However, they tend to drift over time if they are not reseted by external observations.

5. Magnetometers resolve the direction of the magnetic field of the Earth, and can be used as a compass.

Likewise, there are various possible actuators. Here, we subdivide our actuators into internal momentum exchange devices, such as the wheels and control gyroscopes discussed in previous section, and external torque devices, such as the following:

1. Rockets. A set of small rockets appropriately placed constitute a reaction control system (RCS) which enables the spacecraft to experience an external torque.

2. Magnetotorquers. Essentially a electromagnet (magnetic dipole) that we can switch on and off. By interacting with the geomagnetic field it is possible to exert an external torque on the spacecraft.

3. Deployable panels. A panel asymmetrically deployed from the spacecraft can allow some solar radiation pressure or drag to exert an external torque on the spacecraft.

4. Ejectable masses. A one-shot mechanism. A spinning satellite, equipped with end masses attached with a cable to the spacecraft, cuts off the cables, releasing the masses to free space. The masses carry with them some angular momentum.

External torque devices are necessary, as internal momentum exchange devices have a maximum angular momentum that they can store (e.g. a maximum spin rate for the wheels) and tend to saturate. External torque devices then allow to desaturate them.

There are different strategies for attitude control that can be followed, depending on the space mission.

Firstly, it is possible to control the spacecraft in three axis. Wheels and external actuators are placed among the three body axes. For redundancy, a fourth wheel with an intermediate orientation to the other three is commonly mounted too. This is the most advanced option but also the most flexible.

Secondly, a spacecraft can be passively gravity-gradient stabilized. If we carefully design our satellite so that its tensor of inertia has (for example, by deploying booms along the local vertical)

Thirdly, one can impose a spin bias on the spacecraft, for example by having all the spacecraft, a part of it, or a large internal wheel, rotating continuously. This is a spin-stabilized spacecraft. The direction of spin acquires gyroscopic rigidity, and varies only slowly under the effect of perturbations. However, off-axis perturbations will induce some precession about the spin axis, which must still e controlled. If the RCS of the spacecraft is mounted on the rotating part, it is possible to use a lower number of rockets: the rockets need to be fired at the right times to exert an external torque in the desired inertial direction.

9.8 Introduction to quaternions

Euler angles, while intuitive to understand, have major drawbacks for computation. For one, they have singular, gimbal-lock configurations. To avoid Euler angle problems, another way to parametrize the attitude of a rigid body is to use quaternions.

The set of quaternions[15] $\mathbb{H}$ can be regarded as a multidimensional extension of complex numbers $\mathbb{C}$. A quaternion $(q) \in \mathbb{H}$ can be written as:

$$(q) = q_0 + q_1\mathrm{i} + q_2\mathrm{j} + q_3\mathrm{k}, \tag{9.25}$$

where $q_0, q_1, q_2, q_3 \in \mathbb{R}$, with $q_0 = \Re(\mathrm{q})$ often called the *scalar or real* part of the quaternion and $q_1\mathrm{i} + q_2\mathrm{j} + q_3\mathrm{k} = \Im(\mathrm{q})$ the *vector or imaginary* part. Abusing notation, the vector part of quaternion q will be identified with the vector $\boldsymbol{q} = q_1\boldsymbol{i} + q_2\boldsymbol{j} + q_3\boldsymbol{k}$ of $\mathbb{R}^3$. The special symbols i, j, k are extensions of the imaginary unit i in complex numbers, and satisfy:

$$\mathrm{ii} = -1; \qquad \mathrm{jj} = -1; \qquad \mathrm{kk} = -1; \tag{9.26}$$

$$\mathrm{ij} = -\mathrm{ji} = \mathrm{k}; \qquad \mathrm{jk} = -\mathrm{kj} = \mathrm{i}; \qquad \mathrm{ki} = -\mathrm{ik} = \mathrm{j}. \tag{9.27}$$

We can also write a quaternion by separating its scalar and vector parts as $(\mathrm{q}) = (q_0, \boldsymbol{q})$. The product of two quaternions (p) and (q) can be computed using the rules above, or more compactly, as:

$$(\mathrm{p})(\mathrm{q}) = (p_0 q_0 - \boldsymbol{p}\cdot\boldsymbol{q}, p_0\boldsymbol{q} + q_0\boldsymbol{p} + \boldsymbol{p}\times\boldsymbol{q}). \tag{9.28}$$

Observe that quaternion multiplication is non-conmutative.

The conjugate of the quaternion $(\mathrm{q}) = q_0 + q_1\mathrm{i} + q_2\mathrm{j} + q_3\mathrm{k} = (q_0, \boldsymbol{q})$ is defined as

$$(\mathrm{q}^*) = q_0 - q_1\mathrm{i} - q_2\mathrm{j} - q_3\mathrm{k} = (q_0, -\boldsymbol{q}). \tag{9.29}$$

The norm of the quaternion (q) is

$$|(\mathrm{q})| = \sqrt{(\mathrm{q}^*)(\mathrm{q})} = q_0^2 + q_1^2 + q_2^2 + q_3^2. \tag{9.30}$$

Finally, $(\mathrm{q}) = 1 = (1, \boldsymbol{0})$ is the identity quaternion, and the inverse of (q) is

$$(\mathrm{q}^{-1}) = \frac{(\mathrm{q}^*)}{(\mathrm{q}^*)(\mathrm{q})}. \tag{9.31}$$

Unit quaternions are those with $|(\mathrm{q})| = 1$ and represent rotations in $\mathbb{R}^3$, and are sometimes called *rotation quaternions*. The unit quaternion associated to a rotation of angle α about the unit vector $\boldsymbol{u} = u_x\boldsymbol{i} + u_y\boldsymbol{j} + u_z\boldsymbol{k}$ is defined as

$$(\mathrm{q}) = \left(\cos\frac{\alpha}{2}, \sin\frac{\alpha}{2}\boldsymbol{u}\right) \tag{9.32}$$

Observe that $(\mathrm{q}^{-1}) = (\mathrm{q}^*)$ for unit quaternions. It is easy to see that (q^{-1}) represents the inverse rotation to (q), and that (−q) is exactly the same rotation as (q).

To rotate a vector $\boldsymbol{v}$ of $\mathbb{R}^3$ using quaternions we define the quaternion $(\mathrm{v}) = (0, \boldsymbol{v})$ and then compute

$$\begin{aligned}
(\mathrm{v}') &= (\mathrm{q})(\mathrm{v})(\mathrm{q}^*) \\
&= (q_0, \boldsymbol{q})(\boldsymbol{v}\cdot\boldsymbol{q}, q_0\boldsymbol{v} - \boldsymbol{v}\times\boldsymbol{q}) \\
&= (0, q_0(q_0\boldsymbol{v} - \boldsymbol{v}\times\boldsymbol{q}) + (\boldsymbol{v}\cdot\boldsymbol{q})\boldsymbol{q} + \boldsymbol{q}\times(q_0\boldsymbol{v} - \boldsymbol{v}\times\boldsymbol{q})) \\
&= (0, (q_0^2 - \boldsymbol{q}\cdot\boldsymbol{q})\boldsymbol{v} + 2q_0\boldsymbol{q}\times\boldsymbol{v} + 2(\boldsymbol{v}\cdot\boldsymbol{q})\boldsymbol{q}) \\
&= (0, \underbrace{(\boldsymbol{v}\cdot\boldsymbol{u})\boldsymbol{u}}_{\boldsymbol{a}_1} + \underbrace{(\boldsymbol{v} - (\boldsymbol{v}\cdot\boldsymbol{u})\boldsymbol{u})\cos\alpha}_{\boldsymbol{a}_2} + \underbrace{\boldsymbol{u}\times\boldsymbol{v}\sin\alpha}_{\boldsymbol{a}_3}) = (0, \boldsymbol{v}')
\end{aligned} \tag{9.33}$$

[15] *Rabbithole*: Quaternions, introduced by Hamilton, form a four-dimensional associative normed division algebra over the real numbers. They have applications also in describing spin in quantum mechanics.

where $\boldsymbol{v}'$ is the rotated vector. It is interesting to understand the rotation in terms of the three vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ that have been identified in this last expression.

Consider now that we have two vector bases B_0 and B_1, and a vector $\boldsymbol{v} = v_{x0}\boldsymbol{i}_0 + v_{y0}\boldsymbol{j}_0 + v_{z0}\boldsymbol{k}_0 = v_{x1}\boldsymbol{i}_1 + v_{y1}\boldsymbol{j}_1 + v_{z1}\boldsymbol{k}_1$. We will use the notation (v_0) to indicate the vector part of the quaternion $v_{x0}\mathrm{i} + v_{y0}\mathrm{j} + v_{z0}\mathrm{k}$ constructed with the components of $\boldsymbol{v}$ in basis B_0. Likewise (v_1) will be $v_{x1}\mathrm{i} + v_{y1}\mathrm{j} + v_{z1}\mathrm{k}$. We will call $({}_0\mathrm{q}_1)$ the rotation quaternion that (v_1) to v_0 through expression (9.33) as

$$(\mathrm{v}_0) = ({}_0\mathrm{q}_1)(\mathrm{v}_1)({}_1\mathrm{q}_0) \tag{9.34}$$

where ${}_1\mathrm{q}_0 = {}_0\mathrm{q}_1{}^*$.

Quaternion rotations can be concatenated. As such,

$$({}_0\mathrm{q}_3) = ({}_0\mathrm{q}_1)({}_1\mathrm{q}_2)({}_2\mathrm{q}_3) \tag{9.35}$$

would be the quaternion that encodes the rotation from a basis B_3 to B_0. By writing the quaternions for the three simple rotations associated to the Euler angles, it is easy to find the expression of the rotation quaternion for a general rotation in terms of them.

We consider now a continuous rotation in time. If the components of $\boldsymbol{v}$ in B_1 are constant, then differentiating in equation (9.34) gives the time derivative of $\boldsymbol{v}$ in basis B_0 is

$$\begin{aligned}(\dot{\mathrm{v}}_0) &= ({}_0\dot{\mathrm{q}}_1)(\mathrm{v}_1)({}_1\mathrm{q}_0) + ({}_0\mathrm{q}_1)(\mathrm{v}_1)({}_1\dot{\mathrm{q}}_0) &(9.36)\\ &= ({}_0\dot{\mathrm{q}}_1)({}_1\mathrm{q}_0)(\mathrm{v}_0)({}_0\mathrm{q}_1)({}_1\mathrm{q}_0) + ({}_0\mathrm{q}_1)({}_1\mathrm{q}_0)(\mathrm{v}_0)({}_0\mathrm{q}_1)({}_1\dot{\mathrm{q}}_0) &(9.37)\\ &= ({}_0\dot{\mathrm{q}}_1)({}_1\mathrm{q}_0)(\mathrm{v}_0) + (\mathrm{v}_0)({}_0\mathrm{q}_1)({}_1\dot{\mathrm{q}}_0) &(9.38)\end{aligned}$$

with

$$({}_0\dot{\mathrm{q}}_1)({}_1\mathrm{q}_0) = (\dot{q}_0, \dot{\boldsymbol{q}})(q_0, -\boldsymbol{q}) = (q_0\dot{q}_0 + q_1\dot{q}_1 + q_2\dot{q}_2 + q_3\dot{q}_3, \dot{q}_0\boldsymbol{q} + q_0\dot{\boldsymbol{q}} + \dot{\boldsymbol{q}} \times \boldsymbol{q}) = (0, \boldsymbol{\omega}|_0/2), \tag{9.39}$$

where the scalar part of the resulting quaternion vanishes since $|\mathrm{q}| = 1 = \text{const}$, and we have labeled the vector part of the resulting quaternion as $\boldsymbol{\omega}|_0/2$. Similarly, we find:

$$({}_0\mathrm{q}_1)({}_1\dot{\mathrm{q}}_0) = (0, -\boldsymbol{\omega}|_0/2). \tag{9.40}$$

Bringing this back to expression (9.38) we have

$$\begin{aligned}(\dot{\mathrm{v}}_0) &= (0, \boldsymbol{\omega}|_0/2)(\mathrm{v}_0) + (\mathrm{v}_0)(0, -\boldsymbol{\omega}|_0/2)\\ \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}&\begin{bmatrix} v_{x0}\\ v_{y0}\\ v_{z0}\end{bmatrix} = \frac{1}{2}\begin{bmatrix}\omega_{x0}\\ \omega_{y0}\\ \omega_{z0}\end{bmatrix} \times \begin{bmatrix} v_{x0}\\ v_{y0}\\ v_{z0}\end{bmatrix} - \frac{1}{2}\begin{bmatrix} v_{x0}\\ v_{y0}\\ v_{z0}\end{bmatrix} \times \begin{bmatrix}\omega_{x0}\\ \omega_{y0}\\ \omega_{z0}\end{bmatrix} = \begin{bmatrix}\omega_{x0}\\ \omega_{y0}\\ \omega_{z0}\end{bmatrix} \times \begin{bmatrix} v_{x0}\\ v_{y0}\\ v_{z0}\end{bmatrix}\end{aligned}$$

from where we identify $\boldsymbol{\omega}|_0$ as the list of vector components of the angular velocity vector $\boldsymbol{\omega}_{10}$ in basis B_0. Introducing the angular velocity quaternion $(\omega|_0) = (0, \boldsymbol{\omega}|_0)$ we further find

$$(\omega|_0) = 2({}_0\dot{\mathrm{q}}_1)({}_1\mathrm{q}_0) \qquad \Rightarrow \qquad ({}_0\dot{\mathrm{q}}_1) = \frac{1}{2}(\omega|_0)({}_0\mathrm{q}_1) \tag{9.41}$$

These expressions allow to relate the components of $\boldsymbol{\omega}_{10}$ in basis B_0 with the quaternion $({}_0\mathrm{q}_1)$ and its derivative.

References

[Avanzini, 2008] Avanzini, G. (2008). A simple lambert algorithm. *Journal of guidance, control, and dynamics*, 31(6):1587–1594.

[Battin, 1999] Battin, R. H. (1999). *An introduction to the mathematics and methods of astrodynamics, revised edition*. American Institute of Aeronautics and Astronautics.

[Curtis, 2013] Curtis, H. D. (2013). *Orbital mechanics for engineering students*. Butterworth-Heinemann.

[Junkins and Schaub, 2009] Junkins, J. L. and Schaub, H. (2009). *Analytical mechanics of space systems*. American Institute of Aeronautics and Astronautics.

[Tewari, 2007] Tewari, A. (2007). *Atmospheric and space flight dynamics*. Springer.

[Vallado, 2001] Vallado, D. A. (2001). *Fundamentals of astrodynamics and applications*, volume 12. Springer Science & Business Media.