ORBIT DETERMINATION

Giovanni Federico Gronchi

Dipartimento di Matematica, Università di Pisa
Largo B. Pontecorvo 5, 56127 Pisa, Italy

Keywords: computation of orbits, two-body problem, algebraic methods

Contents

Summary

To determine the orbit of a solar system body means to compute its position and velocity at a certain time using the observations of the body, e.g. right ascension and declination if we use an optical telescope. This allows to compute ephemerides and predict the position of the body at different times.

This branch of Celestial Mechanics has attracted the interest of several scientists over the last centuries. However, the ongoing improvements of the observational technologies have set up new orbit determination problems in the recent years: this is partly due to the availability of different

observables (e.g. the range, with radar telescopes), but also to the huge amount of data that can be collected. For these reasons scientists have been induced to think about new algorithms to compute orbits.

In this chapter we present a review of some orbit determination methods, with particular care about the computation of preliminary orbits. Here we include both classical methods, due to Gauss and Laplace, and very recent ones, which are suitable for the sets of optical observations made with modern telescopes. Also the problem of alternative solutions is considered: we describe some results on the geometric characterization of the number of preliminary solutions. The last part of this chapter is devoted to the linkage of short arcs, that is an identification problem appearing with the very large amount of observations that can be made with modern instruments.

1 Introduction

The determination of the orbits of the solar system bodies is an important branch of Celestial Mechanics and has attracted the interest of several scientists over the last centuries. The main problem can be formulated as follows: given a set of observable quantities of a celestial body, made at different epochs (e.g. the angular positions of an asteroid on the celestial sphere), compute the position and velocity of the body at the average time of the observations, so that it is possible to predict the position of the body in the future.

The observations of a celestial body are affected by errors, e.g. due to the instruments, or to atmospheric effects. It is necessary to take into account the effect of these errors in an orbit determination procedure.

Here is a short (and incomplete) list of scientists who gave important contributions to this field: E. Halley, A. J. Lexell, J. L. Lagrange, A. M. Legendre, F. F. Tisserand, P. S. De Laplace, C. F. Gauss, O. Mossotti, H. Poincaré, C. W. L. Charlier, A. Leuschner.

A key event for the development of orbit determination methods was the discovery of Ceres, the first *main belt*[1] asteroid, by Giuseppe Piazzi (Observatory of Palermo, January 1, 1801). He could follow up Ceres in the sky for about 1 month, collecting about 20 observations. Then a problem was set up for the scientists of that epoch: to predict when and in which part of the sky Ceres could be observed again. Ceres was recovered one year later by H. W. Olbers and F. Von Zach, following the suggestions of C. F. Gauss, who among many other scientific interests, was attracted by astronomical problems and became the director of the Göttingen observatory in 1807. Gauss' method consists in two steps: compute a preliminary orbit (see Section 2.2), then apply an iterative method to obtain a solution of a least squares fit (see Section 3). Unfortunately, there can be more than one preliminary orbit: this problem is addressed in Section 4.

At the beginning of the XIX century an asteroid was typically observed only once per night; moreover the number of objects that could be observed was much smaller. The observations at the present days are different: we can detect many more asteroids and we compare images of the same field taken a few minutes apart to search for moving objects. In Figure 1 we show three images of the detection of an asteroid in September 2002.
Thus today there is also an identification problem, that is to join together sets of observations taken in different nights as belonging to the same observed object. The different cases occurring in the identification are described in Sections 5, 6.

There is a broad literature about orbit determination: here we restrict the exposition to the most famous classical methods and to some recent achievements concerning objects orbiting around

[1]The main belt asteroids (MBAs) are located between the orbits of Mars and Jupiter.

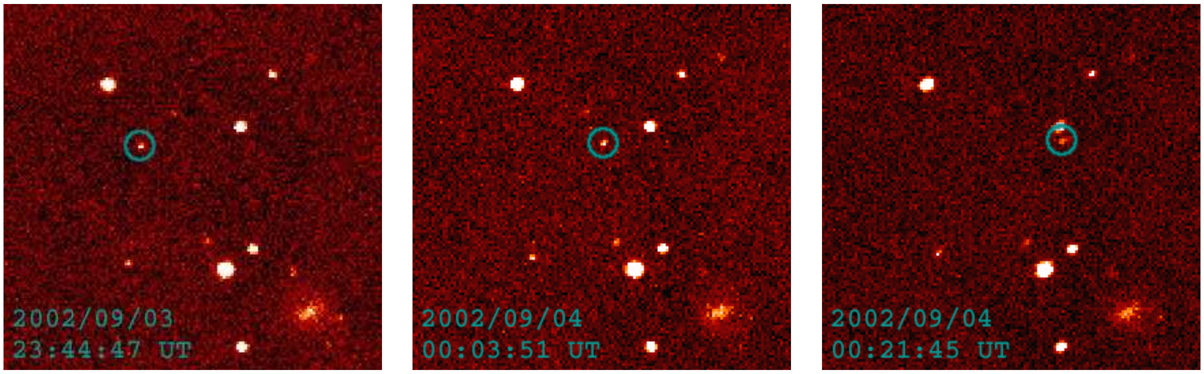

Figure 1: Three images showing the detection of an asteroid (encircled in the figures) during the night of September 3, 2002: the time interval between two consecutive images is ≈ 20 minutes. Courtesy of F. Bernardi.

the Sun (e.g. asteroids), observed with optical instruments.

2 Classical methods of preliminary orbit determination

We illustrate the two classical methods by Laplace and by Gauss to compute a preliminary orbit of a celestial body orbiting around the Sun and observed from the Earth.

2.1 Laplace's method

Assume we have the observations (α_i, δ_i) of a solar system body at times $t_i, i = 1 \ldots m$, $m \geq 3$; then we can interpolate for $\alpha, \delta, \dot{\alpha}, \dot{\delta}$ at a mean time $\bar{t}$, where the dots indicate the time derivative. To obtain an orbit we have to compute the radial distance ρ and the radial velocity $\dot{\rho}$ at the same time $\bar{t}$.

Let $\boldsymbol{\rho} = \rho\hat{\mathbf{e}}^\rho$ be the geocentric position vector of the observed body, with $\rho = \|\boldsymbol{\rho}\|$ and $\hat{\mathbf{e}}^\rho = (\cos\delta\cos\alpha, \cos\delta\sin\alpha, \sin\delta)$, where α, δ are the right ascension and declination. We denote by $\mathbf{q} = q\hat{\mathbf{q}}$ the heliocentric position of the center of the Earth, with $q = \|\mathbf{q}\|$, and by $\mathbf{r} = \mathbf{q} + \boldsymbol{\rho}$ the heliocentric position of the body.

We use the arc length s to parametrize the motion: s is related to the time t by

$$\frac{ds}{dt} = \sqrt{\dot{\alpha}^2\cos^2\delta + \dot{\delta}^2} \stackrel{def}{=} \eta \qquad \text{(proper motion)} .$$

We introduce the moving orthonormal basis

$$\hat{\mathbf{e}}^\rho, \quad \hat{\mathbf{e}}^v = \frac{d\hat{\mathbf{e}}^\rho}{ds}, \quad \hat{\mathbf{e}}^n = \hat{\mathbf{e}}^\rho \times \hat{\mathbf{e}}^v . \tag{1}$$

The relation

$$\frac{d\hat{\mathbf{e}}^v}{ds} = -\hat{\mathbf{e}}^\rho + \kappa\hat{\mathbf{e}}^n$$

defines the geodesic curvature κ. The second derivative of $\boldsymbol{\rho}$ with respect to t can be written as

$$\frac{d^2\boldsymbol{\rho}}{dt^2} = (\ddot{\rho} - \rho\eta^2)\hat{\mathbf{e}}^\rho + (\rho\dot{\eta} + 2\dot{\rho}\eta)\hat{\mathbf{e}}^v + (\rho\eta^2\kappa)\hat{\mathbf{e}}^n .$$

On the other hand, assuming the asteroid and the Earth move on Keplerian orbits, we have

$$\frac{d^2\boldsymbol{\rho}}{dt^2} = \frac{d^2}{dt^2}(\mathbf{r} - \mathbf{q}) = -\frac{\mu}{r^3}\mathbf{r} + \frac{\mu + \mu_\oplus}{q^3}\mathbf{q}\,,$$

with $r = \|\mathbf{r}\|$ and μ, $\mu_\oplus$ the masses of the Sun and of the Earth respectively.
Neglecting the mass of the Earth and projecting the equation of motion onto $\hat{\mathbf{e}}^n$ at time $\bar{t}$ we obtain the dynamical equation of Laplace's method

$$\mathcal{C}\frac{\rho}{q} = 1 - \frac{q^3}{r^3} \qquad \text{with} \qquad \mathcal{C} = \frac{\eta^2 \kappa q^3}{\mu(\hat{\mathbf{q}} \cdot \hat{\mathbf{e}}^n)}\,, \tag{2}$$

where $\rho, q, r, \eta, \hat{\mathbf{q}}, \hat{\mathbf{e}}^n, \mathcal{C}$ denote the values of these quantities at time $\bar{t}$.

In equation (2) ρ and r are unknown, while the other quantities can be computed by interpolation. Using (2) and the geometric equation

$$r^2 = q^2 + \rho^2 + 2q\rho\cos\epsilon\,, \tag{3}$$

where $\cos\epsilon = \mathbf{q} \cdot \boldsymbol{\rho}/(q\rho)$, we can write a polynomial equation of degree eight for r at time $\bar{t}$ by eliminating the geocentric distance:

$$\mathcal{C}^2 r^8 - q^2(\mathcal{C}^2 + 2\mathcal{C}\cos\epsilon + 1)r^6 + 2q^5(\mathcal{C}\cos\epsilon + 1)r^3 - q^8 = 0\ . \tag{4}$$

The occurrence of alternative solutions in equations (2), (3) is discussed in Section 4.1.
The projection of the equations of motion on $\hat{\mathbf{e}}^v$ gives

$$\rho\dot{\eta} + 2\dot{\rho}\eta = \mu(\mathbf{q} \cdot \hat{\mathbf{e}}^v)\left(\frac{1}{q^3} - \frac{1}{r^3}\right)\ . \tag{5}$$

We can use equation (5) to compute $\dot{\rho}$ from the values of r, ρ found by (4) and (2).

2.2 Gauss' Method

Assume we have three observations (α_i, δ_i), $i = 1, 2, 3$ of a solar system body at times t_i, with $t_1 < t_2 < t_3$. Let $\mathbf{r}_i, \boldsymbol{\rho}_i$ denote the heliocentric and topocentric position of the body respectively, and let $\mathbf{q}_i$ be the heliocentric position of the observer. Gauss' method uses the heliocentric positions

$$\mathbf{r}_i = \boldsymbol{\rho}_i + \mathbf{q}_i \quad i = 1, 2, 3\ . \tag{6}$$

We assume that $|t_i - t_j|, 1 \leq i, j \leq 3$, is much smaller than the period of the orbit and write $\mathcal{O}(\Delta t)$ for the order of magnitude of the time differences.

From the coplanarity condition we have

$$\lambda_1\mathbf{r}_1 - \mathbf{r}_2 + \lambda_3\mathbf{r}_3 = 0 \tag{7}$$

for $\lambda_1, \lambda_3 \in \mathbb{R}$. The vector product of both members of (7) with $\mathbf{r}_i, i = 1, 3$ and the fact that the vectors $\mathbf{r}_i \times \mathbf{r}_j, i < j$ have all the same orientation as $\mathbf{c} = \mathbf{r}_h \times \dot{\mathbf{r}}_h$, $h = 1, 2, 3$ implies

$$\lambda_1 = \frac{\mathbf{r}_2 \times \mathbf{r}_3 \cdot \mathbf{c}}{\mathbf{r}_1 \times \mathbf{r}_3 \cdot \mathbf{c}}\,, \qquad \lambda_3 = \frac{\mathbf{r}_1 \times \mathbf{r}_2 \cdot \mathbf{c}}{\mathbf{r}_1 \times \mathbf{r}_3 \cdot \mathbf{c}}\ .$$

Let $\boldsymbol{\rho}_i = \rho_i \hat{\mathbf{e}}^\rho_i$, $i = 1,2,3$. From the scalar product of $\hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3$ with both members of (7), using (6), we obtain

$$\rho_2[\hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3 \cdot \hat{\mathbf{e}}^\rho_2] = \hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3 \cdot [\lambda_1 \mathbf{q}_1 - \mathbf{q}_2 + \lambda_3 \mathbf{q}_3] \ . \tag{8}$$

The differences $\mathbf{r}_i - \mathbf{r}_2, i = 1,3$, are expanded in powers of $t_{ij} = t_i - t_j = \mathcal{O}(\Delta t)$ by the f, g series formalism; thus $\mathbf{r}_i = f_i \mathbf{r}_2 + g_i \dot{\mathbf{r}}_2$, with

$$f_i = 1 - \frac{\mu}{2}\frac{t_{i2}^2}{r_2^3} + \mathcal{O}(\Delta t^3)\,, \qquad g_i = t_{i2}\left(1 - \frac{\mu}{6}\frac{t_{i2}^2}{r_2^3}\right) + \mathcal{O}(\Delta t^4) \ . \tag{9}$$

Then $\mathbf{r}_i \times \mathbf{r}_2 = -g_i \mathbf{c}$, $\mathbf{r}_1 \times \mathbf{r}_3 = (f_1 g_3 - f_3 g_1)\mathbf{c}$ and

$$\lambda_1 = \frac{g_3}{f_1 g_3 - f_3 g_1}\,, \qquad \lambda_3 = \frac{-g_1}{f_1 g_3 - f_3 g_1}\,, \tag{10}$$

$$f_1 g_3 - f_3 g_1 = t_{31}\left(1 - \frac{\mu}{6}\frac{t_{31}^2}{r_2^3}\right) + \mathcal{O}(\Delta t^4) \ . \tag{11}$$

Using (9) and (11) in (10) we obtain

$$\lambda_1 = \frac{t_{32}}{t_{31}}\left[1 + \frac{\mu}{6 r_2^3}(t_{31}^2 - t_{32}^2)\right] + \mathcal{O}(\Delta t^3)\,, \tag{12}$$

$$\lambda_3 = \frac{t_{21}}{t_{31}}\left[1 + \frac{\mu}{6 r_2^3}(t_{31}^2 - t_{21}^2)\right] + \mathcal{O}(\Delta t^3) \ . \tag{13}$$

Let $V = \hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_2 \cdot \hat{\mathbf{e}}^\rho_3$. By substituting (12), (13) into (8), using relations $t_{31}^2 - t_{32}^2 = t_{21}(t_{31} + t_{32})$ and $t_{31}^2 - t_{21}^2 = t_{32}(t_{31} + t_{21})$, we can write

$$\begin{aligned} -V\rho_2 t_{31} &= \hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3 \cdot (t_{32}\mathbf{q}_1 - t_{31}\mathbf{q}_2 + t_{21}\mathbf{q}_3) + \\ &+ \hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3 \cdot \left[\frac{\mu}{6 r_2^3}[t_{32}t_{21}(t_{31}+t_{32})\mathbf{q}_1 + t_{32}t_{21}(t_{31}+t_{21})\mathbf{q}_3]\right] + \mathcal{O}(\Delta t^4) \ . \end{aligned} \tag{14}$$

If the $\mathcal{O}(\Delta t^4)$ terms are neglected, the coefficient of $1/r_2^3$ in (14) is

$$B(\mathbf{q}_1, \mathbf{q}_3) = \frac{\mu}{6} t_{32} t_{21} \hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3 \cdot [(t_{31} + t_{32})\mathbf{q}_1 + (t_{31} + t_{21})\mathbf{q}_3]. \tag{15}$$

Then multiply (14) by $q_2^3 / B(\mathbf{q}_1, \mathbf{q}_3)$ to obtain

$$-\frac{V\ \rho_2\ t_{31}}{B(\mathbf{q}_1, \mathbf{q}_3)}\ q_2^3 = \frac{q_2^3}{r_2^3} + \frac{A(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)}{B(\mathbf{q}_1, \mathbf{q}_3)}\,,$$

where

$$A(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = q_2^3\ \hat{\mathbf{e}}^\rho_1 \times \hat{\mathbf{e}}^\rho_3 \cdot [t_{32}\mathbf{q}_1 - t_{31}\mathbf{q}_2 + t_{21}\mathbf{q}_3] \ .$$

Setting

$$\mathcal{C} = \frac{V\ t_{31}\ q_2^4}{B(\mathbf{q}_1, \mathbf{q}_3)}\,, \qquad \gamma = -\frac{A(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)}{B(\mathbf{q}_1, \mathbf{q}_3)}\,; \tag{16}$$

we obtain the dynamical equation of Gauss' method:

$$\mathcal{C}\,\frac{\rho_2}{q_2} = \gamma - \frac{q_2^3}{r_2^3} \ . \tag{17}$$

After the possible values for r_2 have been found by (17) and by the geometric equation

$$r_2^2 = \rho_2^2 + q_2^2 + 2\rho_2 q_2 \cos\epsilon_2 \,, \tag{18}$$

then the velocity vector $\dot{\mathbf{r}}_2$ can be computed, e.g. from Gibbs' formulas.
The occurrence of alternative solutions of equations (17), (18) is discussed in Section 4.2.

We observe that in his original formulation Gauss used different quantities as unknowns, whose values could be improved by an iterative procedure (today called Gauss map).

3 Least squares orbits

We consider the differential equation

$$\frac{d\mathbf{y}}{dt} = (\mathbf{y}, \mathbf{t}, \boldsymbol{\mu}) \tag{19}$$

giving the state $\mathbf{y} \in \mathbb{R}^p$ of the system at time t (e.g. $p = 6$ if $\mathbf{y}$ is a vector of orbital elements). Here $\boldsymbol{\mu} \in \mathbb{R}^{p'}$ are some constants, called dynamical parameters.
The integral flow, solution of (19) for initial data $\mathbf{y}_0$ at time t_0, is denoted by $\Phi_{t_0}^t(\mathbf{y}_0, \boldsymbol{\mu})$.
We also introduce the *observation function*

$$\mathbf{R} = (R_1, \ldots, R_k)\,, \qquad R_j = R_j(\mathbf{y}, t, \boldsymbol{\nu})\,, \quad j = 1 \ldots k$$

depending on the state $\mathbf{y}$ of the system at time t, and on some constants $\boldsymbol{\nu} \in \mathbb{R}^{p''}$, called kinematical parameters. Moreover we define the *prediction function* $\tilde{\mathbf{r}}(t)$ as the composition of the integral flow with the observation function:

$$\tilde{\mathbf{r}}(t) = \mathbf{R}(\boldsymbol{\Phi}_{t_0}^t(\mathbf{y}_0, \boldsymbol{\mu}), t, \boldsymbol{\nu}) \ .$$

These functions gives a prediction for a specific observation at time t.

We can group the multidimensional data and predictions into two vectors[2], with components r_i, $r(t_i)$ and define the vector of the residuals

$$\boldsymbol{\xi} = (\xi_1 \ldots \xi_m)\,, \qquad \xi_i = r_i - r(t_i)\,, \qquad i = 1 \ldots m \ .$$

3.1 The least squares principle

We describe the least squares method, introduced by Gauss, whose first celebrated application was just to the orbit determination of the asteroid Ceres.[3]

The principle of least squares asserts that the solution of the orbit determination problem makes the target function

$$\mathcal{Q}(\boldsymbol{\xi}) = \frac{1}{m}\, \boldsymbol{\xi}^T\, \boldsymbol{\xi} \tag{20}$$

[2]For example, assume the available observations at time τ_j are the right ascension α_j and the declination δ_j, for $j = 1, \ldots, h$. Then

$$m = kh\,, \qquad k = 2\,, \qquad t_{2j-1} = t_{2j} = \tau_j\,, \qquad \left\{ \begin{array}{l} r_{2j-1} = \alpha_j \\ r_{2j} = \delta_j \end{array} \right. \,, \qquad \left\{ \begin{array}{l} r(t_{2j-1}) = \tilde{r}_1(\tau_j) \\ r(t_{2j}) = \tilde{r}_2(\tau_j) \end{array} \right. \,.$$

[3]There has been a dispute for the invention of the least squares method, that is so important and widely used in every field of the applied Sciences.

attain its minimal value. We observe that

$$\xi_i = \xi_i(\mathbf{y}_0, \boldsymbol{\mu}, \boldsymbol{\nu})$$

and select part of the components of $(\mathbf{y}_0, \boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathbb{R}^{p+p'+p''}$ to form the vector $\mathbf{x} \in \mathbb{R}^N$ of the *fit parameters*, i.e. the parameters to be determined by fitting them to the data. Let us define

$$Q(\mathbf{x}) = \mathcal{Q}(\boldsymbol{\xi}(\mathbf{x})) \ .$$

The remaining components of $(\mathbf{y}_0, \boldsymbol{\mu}, \boldsymbol{\nu})$ form the vector $\mathbf{k}$ of the *consider parameters*. An important requirement is that $m \geq N$. We introduce the $m \times N$ *design matrix*

$$B = \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}(\mathbf{x})$$

and search for the minimum of $Q(\mathbf{x})$ by looking for its stationary points:

$$\frac{\partial Q}{\partial \mathbf{x}} = \frac{2}{m}\, \boldsymbol{\xi}^T B = \mathbf{0} \ . \tag{21}$$

Equation (21) is generally nonlinear: we can use Newton's method to search for its solutions. The standard Newton's method involves the computation of the second derivatives of the target function:

$$\frac{\partial^2 Q}{\partial \mathbf{x}^2} = \frac{2}{m}\, (B^T B + \boldsymbol{\xi}^T H) \tag{22}$$

where

$$H = \frac{\partial^2 \boldsymbol{\xi}}{\partial \mathbf{x}^2}(\mathbf{x})$$

is a 3-index array of shape $m \times N \times N$. We set

$$C_{new} = B^T B + \boldsymbol{\xi}^T H \, ;$$

thus C_{new} is a $N \times N$ matrix, non-negative in the neighborhood of a local minimum[4]. Given the residuals $\boldsymbol{\xi}(\mathbf{x}_k)$ obtained from the value $\mathbf{x}_k$ of the fit parameters at iteration k, the linear approximation of $\frac{\partial Q}{\partial \mathbf{x}}$ in a neighborhood of $\mathbf{x}_k$, evaluated at the solution $\mathbf{x}^*$ of (21) gives

$$\frac{\partial Q}{\partial \mathbf{x}}(\mathbf{x}_k) + \frac{\partial^2 Q}{\partial \mathbf{x}^2}(\mathbf{x}_k)\ (\mathbf{x}^* - \mathbf{x}_k) = \mathbf{0} \, , \tag{23}$$

that is

$$C_{new}\ (\mathbf{x}^* - \mathbf{x}_k) = -B^T\, \boldsymbol{\xi} \ .$$

If $C_{new}(\mathbf{x}_k)$ is invertible then

$$\mathbf{x}_{k+1} = \mathbf{x}_k + C_{new}^{-1} D \qquad D = -B^T\, \boldsymbol{\xi} \, ,$$

where also $D = D(\mathbf{x}_k)$. The point $\mathbf{x}_{k+1}$ should be a better approximation to $\mathbf{x}^*$ than $\mathbf{x}_k$. In the case of orbit determination the convergence of Newton's method to solve the least squares fit is usually not guaranteed, depending on the choice of the first guess $\mathbf{x}_0$ selected to start the iterations, that is on the preliminary orbit.

[4]By $\boldsymbol{\xi}^T H$ we mean the matrix with components $\sum_i \xi_i\, \partial^2 \xi_i / \partial x_j \partial x_k$.

3.2 Differential corrections

A variant of Newton's method, known as *differential corrections*, is often used to minimize the target function $Q(\mathbf{x})$. At each iteration we have

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (B^T\, B)^{-1}\, B^T \boldsymbol{\xi}$$

where B is computed at $\mathbf{x}_k$, and $C = B^T\, B$ is called *normal matrix* and replaces the matrix C_{new}. In this way we are neglecting the term $\boldsymbol{\xi}^T\, H\, (\mathbf{x}^* - \mathbf{x}_k)$ in (23): this approximation works if the residuals are small enough.
One iteration of differential corrections is just the solution of a linear least squares problem

$$C\, (\mathbf{x}_{k+1} - \mathbf{x}_k) = -B^T\, \boldsymbol{\xi}'. \tag{24}$$

Equation (24) is called *normal equation* and this linear problem can be obtained by truncation of the target function:

$$Q(\mathbf{x}) \simeq Q(\mathbf{x}_k) + \frac{2}{m}\, \boldsymbol{\xi}^T\, B\, (\mathbf{x} - \mathbf{x}_k) + \frac{1}{m}\, (\mathbf{x} - \mathbf{x}_k)^T\, C\, (\mathbf{x} - \mathbf{x}_k)\ .$$

Let us denote by $\mathbf{x}_*$ the value of $\mathbf{x}$ at convergence. The inverse of the normal matrix

$$\Gamma = C^{-1} \tag{25}$$

is called *covariance matrix* and its value in $\mathbf{x}_*$ can be used to estimate the uncertainty of the solution of the differential correction algorithm. In fact the eigenvalues of Γ are proportional to the length of the axes of the *confidence ellipsoids*

$$\frac{1}{m}(\mathbf{x} - \mathbf{x}^*)^T C(\mathbf{x} - \mathbf{x}^*) \leq \sigma^2\,, \tag{26}$$

where σ is a real number that can be selected within a probabilistic interpretation of the observational errors.

4 Occurrence of alternative solutions

We describe Charlier's theory, concerning a geometric interpretation of the occurrence of alternative (or multiple) solutions in Laplace's method of preliminary orbit determination, that assumes geocentric observations.

In Section 4.2 we explain a generalization of this theory, allowing to take into account topocentric observations, that is observations made from the surface of the rotating Earth. This applies to Gauss' method, or to the extension of Laplace's method taking into account topocentric observations.

Both methods of preliminary orbit determination lead us to two algebraic equations, which differ only by the value of the coefficients $(\gamma, \mathcal{C}, \epsilon)$:

$$r^2 = q^2 + \rho^2 + 2q\rho\cos\epsilon \qquad \text{(geometric equation)} \tag{27}$$

$$\mathcal{C}\,\frac{\rho}{q} = \gamma - \frac{q^3}{r^3} \qquad \text{(dynamical equation)} \tag{28}$$

see (3), (2), (17). We introduce the intersection problem

$$\begin{cases} (q\gamma - \mathcal{C}\rho)r^3 - q^4 = 0 \\ r^2 - q^2 - \rho^2 - 2q\rho\cos\epsilon = 0 \\ r, \rho > 0 \end{cases}, \tag{29}$$

that is, given $(\gamma, \mathcal{C}, \epsilon) \in \mathbb{R}^2 \times [0, \pi]$ we search for pairs (r, ρ) of strictly positive real numbers, solutions of (28) and (27). We can eliminate the variable ρ from (29) and obtain the reduced problem of searching the values $r > 0$ that are roots of

$$\begin{cases} P(r) \stackrel{def}{=} \mathcal{C}^2 r^8 - q^2(\mathcal{C}^2 + 2\mathcal{C}\gamma\cos\epsilon + \gamma^2)r^6 + 2q^5(\mathcal{C}\cos\epsilon + \gamma)r^3 - q^8 \\ r > 0 \end{cases}. \tag{30}$$

Note that $P(r)$ has only four monomials, thus by Descartes' sign rule there are at most three positive roots of $P(r)$, counted with multiplicity. Note that, if $r = \bar{r}$ is a component of a solution of (29), from (28) we obtain a unique value $\bar{\rho}$ for the other component and, conversely, from a value $\bar{\rho}$ of ρ we obtain a unique $\bar{r}$. There are no more than three values of ρ that are components of the solutions of (29).

We define as *spurious solution* of (30) a positive root $\bar{r}$ of $P(r)$ that is not a component of a solution $(\bar{r}, \bar{\rho})$ of (29) for any $\bar{\rho} > 0$, that is it gives a non-positive ρ through the dynamical equation (28).

How many solutions are possible for the intersection problem? From each solution of (29) a full set of orbital elements can be determined, in fact the knowledge of the topocentric distance ρ allows to compute the corresponding value of $\dot{\rho}$. In case of alternative solutions all of them should be tested as first guess for the differential corrections.

4.1 Charlier's theory

Charlier's theory describes the occurrence of multiple solutions in the problem defined by equations (2), (3), with geocentric observations. Nevertheless, if in (27), (28) we interpret ρ and q as the geocentric distance of the observed body and the heliocentric distance of the center of the Earth, then equation (28) with $\gamma = 1$ corresponds to (2) and equation (27) corresponds to (3). Therefore we shall discuss Charlier's theory by studying the multiple solutions of (29) with $\gamma = 1$, and we shall see that in this case the solutions of (29) can be at most two.

Charlier realized that *'the condition for the appearance of another solution simply depends on the position of the observed body'*. We stress that this statement assumes that the two–body model for the orbit of the observed body is exact and neglects the observation and interpolation errors in the parameters $\mathcal{C}, \epsilon$. In particular we make the following assumption:

the parameters $\mathcal{C}, \epsilon$ are such that the corresponding intersection problem with $\gamma = 1$ admits at least one solution. (31)

In the real astronomical applications this assumption may not be fulfilled and the intersection problem may have no solution, due to the errors in the observations.

For each choice of $\mathcal{C}, \epsilon$ the polynomial $P(r)$ in (30) with $\gamma = 1$ has three changes of sign in the sequence of its coefficients, in fact the coefficient of r^3 is positive because from (28) and (27) we have

$$\mathcal{C}\cos\epsilon + 1 = \frac{1}{2\rho^2 r^3}\left[(r^3 - q^3)(r^2 - q^2) + \rho^2(r^3 + q^3)\right] > 0\,,$$

thus the positive roots of $P(r)$ can indeed be three.
Since $P(q)=0$, there is always the solution corresponding to the center of the Earth, in fact, from the dynamical equation, $r=q$ corresponds to $\rho=0$. This solution must be discarded for physical reasons. Using (31), Descartes' sign rule and the relations

$$P(0) = -q^8 < 0\,; \qquad \lim_{r\to+\infty} P(r) = +\infty\,,$$

we conclude that there are always three positive roots of $P(r)$, counted with multiplicity. By (31) at least one of the other two positive roots r_1, r_2 is not spurious: if either r_1 or r_2 is spurious the solution of (29) is unique, otherwise we have two non–spurious solutions.
To detect the cases with two solutions we write $P(r) = (r-q)P_1(r)$, with

$$P_1(r) = \mathcal{C}^2 r^6 (r+q) + (r^2+qr+q^2)\left[q^5 - (2\mathcal{C}\cos\epsilon + 1)q^2 r^3\right]\,,$$

so that

$$P_1(q) = 2q^7\mathcal{C}(\mathcal{C} - 3\cos\epsilon)\,.$$

From the relations

$$P_1(0) = q^7 > 0\,, \qquad \lim_{r\to+\infty} P_1(r) = +\infty$$

it follows that if $P_1(q) < 0$ then $r_1 < q < r_2$, while if $P_1(q) > 0$ then either $r_1, r_2 < q$ or $r_1, r_2 > q$. In the first case the dynamical equation gives us two values ρ_1, ρ_2 with $\rho_1\rho_2 < 0$, so that one root of $P_1(r)$ is spurious. In the second case both roots give rise to meaningful solutions of (29). If $P_1(q) = 0$ there is only one non–spurious root of $P(r)$.

We introduce two algebraic curves in geocentric polar coordinates (ρ, ψ), with $\psi = 0$ towards the opposition direction (corresponding to $\epsilon = 0$), by

$$\begin{array}{ll} \mathfrak{C}^{(1)}(\rho,\psi) = 0 & \qquad\text{(zero circle)}\,, \\ \mathfrak{C}^{(1)}(\rho,\psi) - 3\cos\psi = 0 & \qquad\text{(limiting curve)}\,, \end{array}$$

where

$$\mathfrak{C}^{(1)}(\rho,\psi) = \frac{q}{\rho}\left[1 - \frac{q^3}{r^3}\right]\,, \qquad r = \sqrt{\rho^2 + q^2 + 2q\rho\cos\psi}\,.$$

The limiting curve has a loop inside the zero circle and two unlimited branches with $r > q$. By the previous discussion the limiting curve and the zero circle divide the reference plane, containing the center of the Sun, the observer and the observed body at time $\bar{t}$, into four connected components (see Figure 2), separating regions with a different number of solutions of the orbit determination problem. Using heliocentric polar coordinates (r, ϕ), with $\rho^2 = r^2 + q^2 - 2qr\cos\phi$, the limiting curve is given by

$$4 - 3\frac{r}{q}\cos\phi = \frac{q^3}{r^3} \tag{32}$$

and, in heliocentric rectangular coordinates $(x, y) = (r\cos\phi, r\sin\phi)$, by

$$4 - 3\frac{x}{q} = \frac{q^3}{(x^2+y^2)^{3/2}}\,.$$

Figure 2 shows in particular that, if the celestial body has been observed close to the opposition direction, then the solution of Laplace's method of preliminary orbit determination is unique.

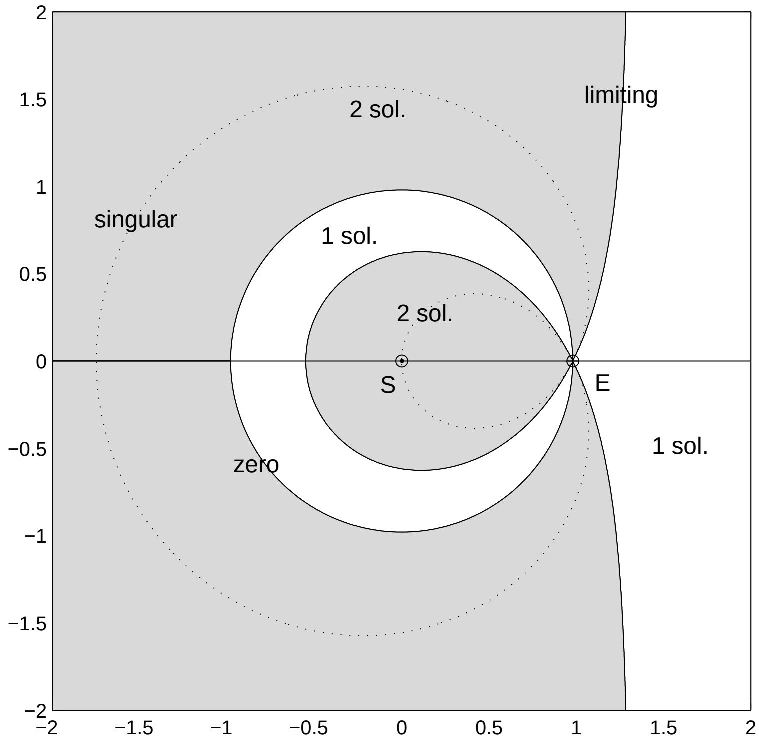

Figure 2: The limiting curve and the zero circle divide the reference plane into four connected regions, two with a unique solution of (29) and two with two solutions (shaded in this figure). The singular curve (dotted) divides the regions with two solutions into two parts, with one solution each. The Sun and the Earth are labeled with S and E respectively. We use heliocentric rectangular coordinates, and astronomical units (AU) for both axes. With kind permission from Springer.

In 1911 Charlier introduced the *singular curve*

$$4 - 3\frac{q}{r}\cos\phi = \frac{r^3}{q^3}\,, \tag{33}$$

corresponding to (32) by radial inversion, that divides the regions with two solutions into regions containing only one solution each (see Figure 2).

4.2 Generalization of Charlier's theory

We can generalize Charlier's theory of multiple solutions in preliminary orbit determination from three observations: this more general theory consider the problem (29) for $\gamma \in \mathbb{R}$.
The following assumption is introduced, that generalizes (31): the parameters $\gamma, \mathcal{C}, \epsilon$ are such that the corresponding intersection problem admits at least one solution.
In this case we can assert that for each given value of γ, the condition for the appearance of another solution simply depends on the position of the observed body.
The constant γ is a bifurcation parameter and qualitatively different results occurs depending on which of relations $\gamma \leq 0$, $0 < \gamma < 1$, $\gamma = 1$, $\gamma > 1$ holds.

Note that $r = q$ generically is not a root of $P(r)$, in fact

$$P(q) = q^8(1-\gamma)\left(2\mathcal{C}\cos\epsilon - (1-\gamma)\right)\,,$$

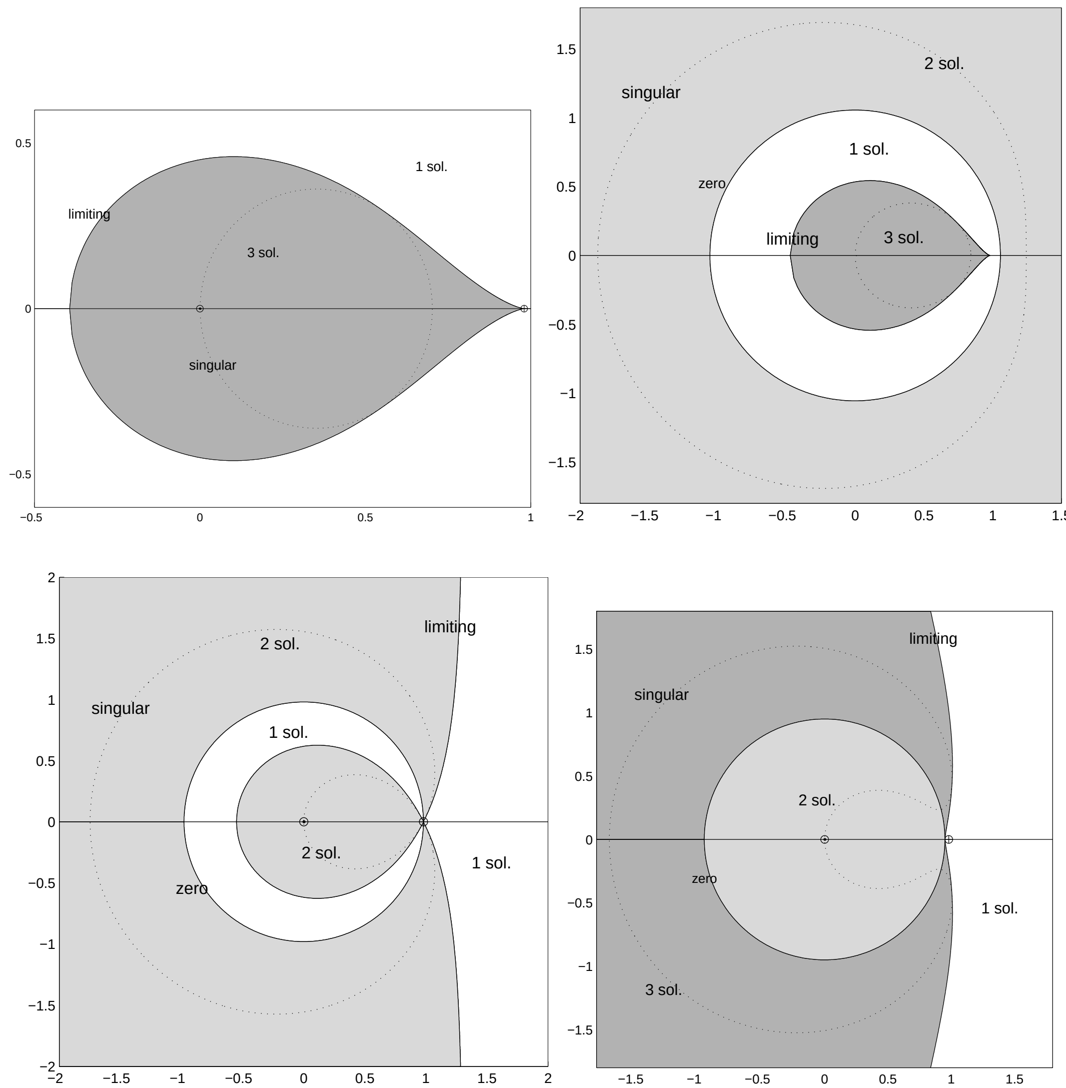

Figure 3: Summary of the results on multiple solutions of (29) for all the qualitatively different cases. The regions with a different number of solutions are enhanced with colours: we use *light gray* for two solutions, *dark gray* for three solutions. Top left: $\gamma = -0.5$. Top right: $\gamma = 0.8$. Bottom left: $\gamma = 1$ (Charlier's case). Bottom right: $\gamma = 1.1$. With kind permission from Springer.

thus we cannot follow the same steps of Section 4.1 to define the limiting curve. However, for each value of $\gamma \neq 1$ it is possible to perform a geometric construction of a curve delimiting regions with a different number of solutions. On the contrary, the definition of the zero circle and of the singular curve for a generic γ is immediate.

Figure 3 summarizes the results for all the qualitatively different cases: there are regions with a unique solution (white), with two solutions (light gray) and with three solutions (dark gray) of (29). On top–left of Figure 3 we show the results for $\gamma = -0.5$: there are only two regions, with either one or three solutions. On top–right we show the results for $\gamma = 0.8$: in the region outside the zero circle there are two solutions of (29) while the region inside is divided by the limiting curve into two parts, with either one or three solutions. On bottom–left we have Charlier's case ($\gamma = 1$), discussed in Section 4.1. On bottom–right we show the results for $\gamma = 1.1$: inside the zero circle there are two solutions, while the region outside can contain either one or three solutions.

Note that in each case the singular curve (dotted in the figure) separates the regions with multiple solutions into parts with only one solution each.

Thus the results on the multiple solutions are generically different from Charlier's: the solutions can be up to three and, if $\gamma > 1$, there are two solutions close to the opposition direction. However, in case of three solutions, one of them bifurcates from $r = q$ and, with ground–based observations, is usually close to that value, thus it is unlike that it corresponds to a good preliminary orbit.

5 New challenges with the modern surveys

The improvements of the observational technologies have produced new orbit determination problems, mostly due to the very large amount of data than can be collected. The main problem is to join together sets of observations, made in different nights, as belonging to the same observed object. This is called *identification problem* and is the subject of this and the next sections.

5.1 Very short arcs and attributables

The observations of a solar system body are grouped into a *very short arc* (VSA), also called *tracklet* in the astronomical literature.

The VSA is composed by $m \geq 3$ optical angular observations (α_i, δ_i) at different times t_i, $i = 1 \ldots m$ such that we can fit both angular coordinates as a function of time with a polynomial model of low degree. In most cases a degree 2 model is used, centered at the mean time $\bar{t} = \frac{1}{m}\sum_i t_i$:

$$\begin{aligned} \alpha(t) &= \alpha(\bar{t}) + \dot{\alpha}(\bar{t})\ (t - \bar{t}) + \frac{1}{2}\ddot{\alpha}(\bar{t})\ (t - \bar{t})^2 , \\ \delta(t) &= \delta(\bar{t}) + \dot{\delta}(\bar{t})\ (t - \bar{t}) + \frac{1}{2}\ddot{\delta}(\bar{t})\ (t - \bar{t})^2 . \end{aligned}$$

The vector $(\alpha, \dot{\alpha}, \ddot{\alpha}, \delta, \dot{\delta}, \ddot{\delta})$ is obtained as solution of a linear least squares problem, together with two 3×3 covariance matrices. If the second derivatives are poorly determined then we speak of a *too short arc* (TSA); in this case the data do not allow to compute a least squares orbit.

We shall call *attributable* a vector

$$A = (\alpha, \delta, \dot{\alpha}, \dot{\delta}) \in [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}^2 ,$$

representing the angular position and velocity of the body at the average time $\bar{t}$ of the observations. Given an attributable the radial distance ρ and the radial velocity $\dot{\rho}$ are completely undetermined.

5.2 Identification problems

We classify as follows the identification problems occurring with modern data:

1) **orbit identification**: join together two sets of observations related to two different orbits to form an orbit fitting all the data;

2) **attribution**: join together a TSA with the set of observations of an orbit to form an orbit fitting all the data;[5]

[5]The name attributable has been introduced just to indicate a set of data suitable for attribution to an already existing orbit.

3) **linkage**:[6] join together two TSAs of observations to form an orbit fitting all the data.

The linkage operation is the most difficult: in Section 6 we will discuss some algorithms to perform it. However we warn that an orbit produced by the linkage operation usually needs a confirmation, by attributing additional data, to be considered reliable.

5.3 Orbit identification

Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^6$ be two nominal orbits, solutions of least squares problems with normal and covariance matrices $C_1, C_2, \Gamma_1, \Gamma_2$. We can assume that $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^6$ are given at the same epoch, up to orbit and covariance propagation.

Assume the two separate sets of observations

$$(t_i, r_i)\ ,\ i = 1, m_1 \qquad (t_i, r_i)\ ,\ i = m_1 + 1, m_1 + m_2$$

have been used to determine $\mathbf{x}_1, \mathbf{x}_2$, with m_1 observations in the first arc and m_2 in the second arc. Moreover, denote by

$$\boldsymbol{\xi}_1 = (\xi_i), \quad i = 1, m_1 \qquad \boldsymbol{\xi}_2 = (\xi_i), \quad i = m_1 + 1, m_1 + m_2$$

the residuals with respect to the nominal solutions. We can compute the two separate target functions for $i = 1, 2$

$$Q_i(\mathbf{x}) = \frac{1}{m_i}\,\boldsymbol{\xi}_i \cdot \boldsymbol{\xi}_i = Q_i(\mathbf{x}_i) + \Delta Q_i(\mathbf{x}) = Q_i(\mathbf{x}_i) + \frac{1}{m_i}\,(\mathbf{x} - \mathbf{x}_i) \cdot C_i\,(\mathbf{x} - \mathbf{x}_i) + \ldots$$

where the dots represent the terms of degree three in $(\mathbf{x} - \mathbf{x}_i)$ and those of degree 2 containing the residuals. The joint target function Q is a linear combination Q_0 of the two separate minima $Q_1(\mathbf{x}_1), Q_2(\mathbf{x}_2)$ plus a penalty ΔQ measuring the increment of the target function resulting from the hypothesis that the two objects are the same:

$$\begin{aligned}
m\,Q(\mathbf{x}) &= \boldsymbol{\xi}_1 \cdot \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 \cdot \boldsymbol{\xi}_2 = m_1 Q_1(\mathbf{x}) + m_2 Q_2(\mathbf{x}) = m Q_0 + m \Delta Q(\mathbf{x})\,, \\
m\,Q_0 &= [m_1 Q_1(\mathbf{x}_1) + m_2 Q_2(\mathbf{x}_2)]\,, \\
m\,\Delta Q(\mathbf{x}) &= m_1 \Delta Q_1(\mathbf{x}) + m_2 \Delta Q_2(\mathbf{x}) = \\
&= (\mathbf{x} - \mathbf{x}_1) \cdot C_1\,(\mathbf{x} - \mathbf{x}_1) + (\mathbf{x} - \mathbf{x}_2) \cdot C_2\,(\mathbf{x} - \mathbf{x}_2) + \ldots
\end{aligned}$$

We can use the quadratic approximation for both ΔQ_i, and obtain an explicit formula for the solution of the identification problem. Neglecting the higher order terms we have

$$\begin{aligned}
&m\,\Delta Q(\mathbf{x}) \simeq (\mathbf{x} - \mathbf{x}_1) \cdot C_1\,(\mathbf{x} - \mathbf{x}_1) + (\mathbf{x} - \mathbf{x}_2) \cdot C_2\,(\mathbf{x} - \mathbf{x}_2) = \\
&= \mathbf{x} \cdot (C_1 + C_2)\,\mathbf{x} - 2\mathbf{x} \cdot (C_1\,\mathbf{x}_1 + C_2\,\mathbf{x}_2) + \mathbf{x}_1 \cdot C_1\,\mathbf{x}_1 + \mathbf{x}_2 \cdot C_2\,\mathbf{x}_2\ .
\end{aligned}$$

The minimum of ΔQ can be found by minimizing the non-homogeneous quadratic form of the formula above. If we denote this minimum by $\mathbf{x}_0$, then by expanding around $\mathbf{x}_0$ we have

$$m\,\Delta Q(\mathbf{x}) \simeq (\mathbf{x} - \mathbf{x}_0) \cdot C_0\,(\mathbf{x} - \mathbf{x}_0) + K$$

[6] In the context of space debris this operation is called *correlation*.

where

$$
\begin{aligned}
C_0 &= C_1 + C_2 \,, \\
C_0 \, \mathbf{x}_0 &= C_1 \, \mathbf{x}_1 + C_2 \, \mathbf{x}_2 \,, \\
K &= \mathbf{x}_1 \cdot C_1 \, \mathbf{x}_1 + \mathbf{x}_2 \cdot C_2 \, \mathbf{x}_2 - \mathbf{x}_0 \cdot C_0 \, \mathbf{x}_0 \ .
\end{aligned}
$$

If the matrix C_0 is positive definite, then we can solve for the new minimum point by the covariance matrix $\Gamma_0 = C_0^{-1}$:

$$\mathbf{x}_0 = \Gamma_0 \, (C_1 \, \mathbf{x}_1 + C_2 \, \mathbf{x}_2) \ . \tag{34}$$

The identification penalty K/m approximates the minimum of the penalty $\Delta Q(\mathbf{x})$, normalized by the number of observations m. In the linear approximation $K/m = \Delta Q(\mathbf{x}_0)$. We observe that K is translation invariant, that is the transformation

$$\mathbf{x}_0 \ \rightarrow \ \mathbf{x}_0 + \mathbf{v} \qquad \mathbf{x}_1 \rightarrow \mathbf{x}_1 + \mathbf{v} \qquad \mathbf{x}_2 \rightarrow \mathbf{x}_2 + \mathbf{v}$$

for an arbitrary vector $\mathbf{v}$ gives

$$K \ \rightarrow \ K + 2\mathbf{v} \cdot (C_1 \, \mathbf{x}_1 + C_2 \, \mathbf{x}_2 - C_0 \, \mathbf{x}_0) + \mathbf{v} \cdot (C_1 + C_2 - C_0)\mathbf{v} = K \ .$$

Therefore we can compute K with a translation by $-\mathbf{x}_1$, that is assuming $\mathbf{x}_1 \rightarrow \mathbf{0}$, $\mathbf{x}_2 \rightarrow \mathbf{x}_2 - \mathbf{x}_1 = \Delta\mathbf{x}$, and $\mathbf{x}_0 \rightarrow \Gamma_0 \, C_2 \, \Delta\mathbf{x}$:

$$K = \Delta\mathbf{x} \cdot C_2 \, \Delta\mathbf{x} - (\mathbf{x}_0 - \mathbf{x}_1) \cdot C_0 \, (\mathbf{x}_0 - \mathbf{x}_1) = \Delta\mathbf{x} \cdot (C_2 - C_2 \, \Gamma_0 \, C_2)\Delta\mathbf{x} \ . \tag{35}$$

Alternatively, we can compute K with a translation by $-\mathbf{x}_2$, that is with $\mathbf{x}_2 \rightarrow \mathbf{0}$, $\mathbf{x}_1 \rightarrow -\Delta\mathbf{x}$ and $\mathbf{x}_0 \rightarrow \Gamma_0 \, C_1 \, (-\Delta\mathbf{x})$:

$$K = \Delta\mathbf{x} \cdot C_1 \, \Delta\mathbf{x} - (\mathbf{x}_0 - \mathbf{x}_2) \cdot C_0 \, (\mathbf{x}_0 - \mathbf{x}_2) = \Delta\mathbf{x} \cdot (C_1 - C_1 \, \Gamma_0 \, C_1) \, \Delta\mathbf{x} \ .$$

Then, setting

$$C = C_2 - C_2 \, \Gamma_0 \, C_2 = C_1 - C_1 \, \Gamma_0 \, C_1 \,, \tag{36}$$

we can summarize the conclusions by the formula

$$Q(\mathbf{x}) \simeq Q_0 + \frac{1}{m} \, \Delta\mathbf{x} \cdot C \, \Delta\mathbf{x} + \frac{1}{m} \, (\mathbf{x} - \mathbf{x}_0) \cdot C_0 \, (\mathbf{x} - \mathbf{x}_0) \ . \tag{37}$$

Relation (37) allows to assess the uncertainty of the identified solution, by defining confidence ellipsoids with matrix C_0.

5.4 Attribution

We assume an orbit $\mathbf{x}_1$ has been fit to the first set of m_1 observations, at the mean epoch t_1, and the uncertainty is described by the covariance and normal matrices Γ_1, C_1. The second arc includes m_2 scalar observations: we assume they form a TSA and compute an attributable A, at the mean epoch t_2.

Prediction for an attributable

Let us consider a function G that maps an open set of the initial conditions space into the attributable 4-dimensional space, that is the vector of observables is

$$\mathbf{y}(\bar{t}) = (\alpha(\bar{t}), \delta(\bar{t}), \dot{\alpha}(\bar{t}), \dot{\delta}(\bar{t})) = G\left(\mathbf{x}(\bar{t})\right) \ .$$

Given initial conditions $\mathbf{x}$ at time t_0 with covariance Γ, the prediction function $F = G \circ \Phi_{t_0}^{t}$ is also 4-dimensional and its partial derivatives form the matrix DF of dimension 4×6. The covariance and normal matrix are the 4×4 matrices obtained from Γ by

$$\Gamma_{\mathbf{y}} = (DF)\, \Gamma\, (DF)^T \qquad C_{\mathbf{y}} = \Gamma_{\mathbf{y}}^{-1} \ .$$

The matrix $\Gamma_{\mathbf{y}}$ can be used to assess the uncertainty of all the components of the attributable; the normal matrix $C_{\mathbf{y}}$ can be used to define the metric used in the attribution algorithm.

Attribution penalty

Let $\mathbf{x}_1$ be the attributable, that is the 4-dimensional vector representing the set of observations to be attributed, and C_1 be the 4×4 normal matrix of the fit used to compute it. Let $\mathbf{x}_2$ be the predicted attributable, computed from the known least squares orbit, and Γ_2 be the covariance matrix of such 4-dimensional prediction, obtained by propagation of the covariance of the orbital elements. Then $C_2 = \Gamma_2^{-1}$ is the corresponding normal matrix. With this new interpretation for the symbols $\mathbf{x}_1, \mathbf{x}_2, C_1, C_2$, the algorithm for linear attribution uses the same formulae of Section 5.3 applied in the 4-dimensional attributable space:

$$\begin{aligned} C_0 &= C_1 + C_2\,, \qquad \Gamma_0 = C_0^{-1}\,, \\ \mathbf{x}_0 &= \Gamma_0\left[C_1\,\mathbf{x}_1 + C_2\,\mathbf{x}_2\right]\,, \\ K_4 &= (\mathbf{x}_2 - \mathbf{x}_1) \cdot \left[C_1 - C_1\,\Gamma_0\,C_1\right]\,(\mathbf{x}_2 - \mathbf{x}_1)\ . \end{aligned} \tag{38}$$

The attribution penalty K_4/m (m the number of scalar observations) is used to filter out the pairs orbit-attributable which cannot belong to the same object. For the pairs with K_4 below some control value, we select a preliminary orbit and perform the differential corrections.

6 Linkage

In this section we recall some methods used to deal with the linkage problem of two TSAs. The linkage is more difficult than the other identification problems because usually we cannot neglect nonlinear terms in the procedure, as we do for example in the orbit identification problem when we propagate the orbits with their covariance.

6.1 The admissible region method

Let A be an attributable at time $\bar{t}$ for a celestial body $\mathcal{A}$. As already mentioned, the information contained in A leaves completely unknown the topocentric distance ρ and the radial velocity $\dot{\rho}$. However, we can constrain the possible values of $\rho, \dot{\rho}$ by making some hypotheses on the physical and dynamical nature of the observed object.

We introduce the two-body energy of the heliocentric orbit of $\mathcal{A}$:

$$\mathcal{E}_\odot(\rho,\dot\rho) = \frac{1}{2}\|\dot{\mathbf{r}}(\rho,\dot\rho)\|^2 - k^2\frac{1}{\|\mathbf{r}(\rho)\|}\,, \tag{39}$$

where k is Gauss' constant. We consider the region excluding interstellar orbits, that is satisfying condition

$$\mathcal{E}_\odot(\rho,\dot\rho) \le 0\,. \tag{40}$$

The heliocentric position of $\mathcal{A}$ is

$$\mathbf{r} = \mathbf{q} + \rho\,\hat{\mathbf{e}}^\rho\,, \tag{41}$$

where $\hat{\mathbf{e}}^\rho$ is the unit vector in the observation direction and $\mathbf{q}$ the heliocentric position of the observer. Using as coordinates (ρ,α,δ), the heliocentric velocity $\dot{\mathbf{r}}$ of $\mathcal{A}$ is

$$\dot{\mathbf{r}} = \dot{\mathbf{q}} + \dot\rho\,\hat{\mathbf{e}}^\rho + \rho\,\cos\delta\dot\alpha\,\hat{\mathbf{e}}^\alpha + \rho\,\dot\delta\,\hat{\mathbf{e}}^\delta\,, \tag{42}$$

where

$$\hat{\mathbf{e}}^\alpha = \frac{1}{\cos\delta}\frac{\partial\hat{\mathbf{e}}^\rho}{\partial\alpha}\,, \qquad \hat{\mathbf{e}}^\delta = \frac{\partial\hat{\mathbf{e}}^\rho}{\partial\delta}$$

and $\dot{\mathbf{q}}$ is the heliocentric velocity of the observer. The vectors $\hat{\mathbf{e}}^\rho, \hat{\mathbf{e}}^\alpha, \hat{\mathbf{e}}^\delta$ form an orthonormal basis:

$$\hat{\mathbf{e}}^\rho\cdot\hat{\mathbf{e}}^\alpha = \hat{\mathbf{e}}^\rho\cdot\hat{\mathbf{e}}^\delta = \hat{\mathbf{e}}^\alpha\cdot\hat{\mathbf{e}}^\delta = 0\,, \quad \|\hat{\mathbf{e}}^\rho\| = \|\hat{\mathbf{e}}^\alpha\| = \|\hat{\mathbf{e}}^\delta\| = 1\,.$$

Thus the squared norms of the heliocentric position and velocity are

$$\|\mathbf{r}(\rho)\|^2 = \rho^2 + 2\rho\,\mathbf{q}\cdot\hat{\mathbf{e}}^\rho + \|\mathbf{q}\|^2\,, \tag{43}$$

$$\|\dot{\mathbf{r}}(\rho,\dot\rho)\|^2 \;=\; \dot\rho^2 + 2\dot\rho\dot{\mathbf{q}}\cdot\hat{\mathbf{e}}^\rho + \rho^2\big(\dot\alpha^2\cos^2\delta + \dot\delta^2\big) + 2\rho\big(\dot\alpha\cos\delta\dot{\mathbf{q}}\cdot\hat{\mathbf{e}}^\alpha + \dot\delta\dot{\mathbf{q}}\cdot\hat{\mathbf{e}}^\delta\big) + \|\dot{\mathbf{q}}\|^2\,. \tag{44}$$

We shall use the coefficients[7]

$$\begin{aligned} c_0 &= \|\mathbf{q}\|^2 & \qquad c_3 &= 2\dot\alpha\cos\delta\,\dot{\mathbf{q}}\cdot\hat{\mathbf{e}}^\alpha + 2\dot\delta\,\dot{\mathbf{q}}\cdot\hat{\mathbf{e}}^\delta \\ c_1 &= 2\dot{\mathbf{q}}\cdot\hat{\mathbf{e}}^\rho & c_4 &= \|\dot{\mathbf{q}}\|^2 \\ c_2 &= \dot\alpha^2\cos^2\delta + \dot\delta^2 = \eta^2 & c_5 &= 2\mathbf{q}\cdot\hat{\mathbf{e}}^\rho\,, \end{aligned} \tag{45}$$

and the polynomial expressions

$$\begin{aligned} 2\mathcal{T}_\odot(\rho,\dot\rho) &\stackrel{def}{=} \|\dot{\mathbf{r}}(\rho,\dot\rho)\|^2 = \dot\rho^2 + c_1\dot\rho + c_2\rho^2 + c_3\rho + c_4\,, \\ S(\rho) &\stackrel{def}{=} r^2 = \rho^2 + c_5\rho + c_0\,, \\ W(\rho) &\stackrel{def}{=} c_2\rho^2 + c_3\rho + c_4\,. \end{aligned} \tag{46}$$

By substituting the last expressions in (39), condition (40) reads

$$2\mathcal{E}_\odot(\rho,\dot\rho) = \dot\rho^2 + c_1\dot\rho + W(\rho) - 2k^2/\sqrt{S(\rho)} \le 0\,.$$

To have real solutions, the discriminant of $\mathcal{E}_\odot$ as a polynomial of degree 2 in $\dot\rho$ must be non-negative, i.e.

$$c_1^2/4 - W(\rho) + 2k^2/\sqrt{S(\rho)} \ge 0\,.$$

[7] To obtain more accurate results, the position $\mathbf{q}$ and the velocity $\dot{\mathbf{q}}$ at time $\bar t$ should be computed consistently with the interpolation used for $\hat{\mathbf{e}}^\rho$, according to a suggestion by Poincaré.

Let us set $\gamma = c_4 - c_1^2/4$ and define $P(\rho) = c_2\rho^2 + c_3\rho + \gamma$. Then condition (40) implies

$$2k^2/\sqrt{S(\rho)} \geq P(\rho) \ . \tag{47}$$

The polynomial $P(\rho)$ is non-negative for each ρ: it is the opposite of the discriminant of $\mathcal{T}_\odot(\rho, \dot\rho)$ as a polynomial in the variable $\dot\rho$. $\mathcal{T}_\odot$ is a kinetic energy and is non-negative, thus its discriminant is non-positive. Also $S(\rho)$ is non-negative, thus we can square both sides of (47) and obtain the polynomial inequality of degree 6

$$4k^4 \geq V(\rho) = P^2(\rho)S(\rho) = \sum_{i=0}^{6} A_i\,\rho^i \ , \tag{48}$$

with coefficients

$$\begin{aligned}
A_0 &= c_0\gamma^2\,, \quad A_1 = c_5\gamma^2 + 2c_0c_3\gamma\,, \quad A_2 = \gamma^2 + 2c_3c_5\gamma + c_0(c_3^2 + 2c_2\gamma)\,,\\
A_3 &= 2c_3\gamma + c_5(c_3^2 + 2c_2\gamma) + 2c_0c_2c_3\,,\\
A_4 &= c_3^2 + 2c_2\gamma + 2c_2c_3c_5 + c_0c_2^2\,, \quad A_5 = c_2(2c_3 + c_2c_5)\,, \quad A_6 = c_2^2\ .
\end{aligned}$$

The region defined by (40) has at most two connected components. In Figure 4 we plot the level curves of $\mathcal{E}_\odot$ for positive, zero and negative values, showing the qualitative change in the topology of these sets.

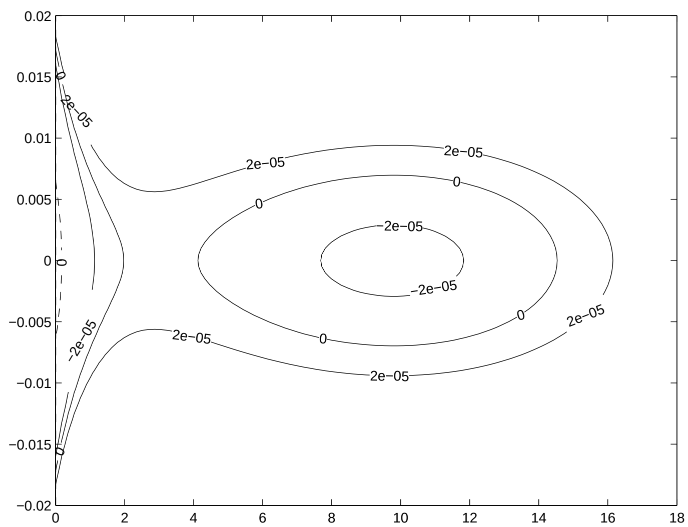

Figure 4: Three level curves of $\mathcal{E}_\odot$, including the zero level curve, and $\mathcal{E}_\oplus = 0$ (dashed curve) in the plane $(\rho, \dot\rho)$.

It is useful to introduce another constraint to exclude objects at an arbitrarily small distance from the observer. We can make different choices, for example

1) assign an inner boundary by requiring that $\mathcal{A}$ is not a satellite of the Earth, i.e. by imposing a condition on the geocentric energy $\mathcal{E}_\oplus(\rho, \dot\rho)$;

2) set a minimal distance by requiring that $\mathcal{A}$ is not too small. This is possible if photometric measurements are available.

Excluding satellites of the Earth

We describe the region defined by the condition $\mathcal{E}_\oplus(\rho,\dot\rho) \geq 0$. Assume for simplicity that the observations are geocentric: $\mathbf{q}_\oplus$ is the heliocentric position of the Earth center, $\boldsymbol{\rho} = \rho\hat{\mathbf{e}}^\rho$ is the geocentric position of the observed body, and $\mathbf{r} = \boldsymbol{\rho} + \mathbf{q}_\oplus$. The geocentric energy is

$$\mathcal{E}_\oplus(\rho,\dot\rho) = \frac{1}{2}\|\dot{\boldsymbol{\rho}}\|^2 - k^2\mu_\oplus\frac{1}{\rho} \geq 0 \ , \tag{49}$$

where $\mu_\oplus$ is the ratio between the mass of the Earth and the mass of the Sun. By using $\|\dot{\boldsymbol{\rho}}(\rho,\dot\rho)\|^2 = \dot\rho^2 + \rho^2\,\eta^2$, where $\eta = \sqrt{\dot\alpha^2\cos^2\delta + \dot\delta^2}$ is the proper motion, (49) becomes

$$\dot\rho^2 + \rho^2\,\eta^2 - 2k^2\mu_\oplus\frac{1}{\rho} \geq 0\,,$$

that is

$$\dot\rho^2 \geq G(\rho)\,, \qquad \text{with} \quad G(\rho) = \frac{2k^2\mu_\oplus}{\rho} - \eta^2\rho^2\ . \tag{50}$$

Note that $G(\rho) > 0$ for $0 < \rho < \rho_0 = \sqrt[3]{(2k^2\mu_\oplus)/\eta^2}$.

However, condition (49) is meaningful only inside the sphere of influence of the Earth, otherwise the dynamics of $\mathcal{A}$ is dominated by the Sun, not by the Earth. Thus we need to introduce the condition

$$\rho \geq R_{SI} = a_\oplus\ \sqrt[3]{\mu_\oplus/3}\,, \tag{51}$$

where R_{SI} is the radius of the sphere of influence, $a_\oplus$ is the semimajor axis of the Earth. To exclude the satellites of the Earth we have to assume that either (49) or (51) apply. If $\rho_0 \leq R_{SI}$ the region of the satellites to be excluded is defined simply by eq. (50); this occurs for

$$\rho_0^3 = 2k^2\mu_\oplus/\eta^2 \leq R_{SI}^3 = a_\oplus^3\ \mu_\oplus/3$$

thus, taking into account Kepler third law $a_\oplus^3\ n_\oplus^2 = k^2$, with $n_\oplus$ the mean motion of the Earth, we have $\rho_0 \leq R_{SI}$ if and only if $\eta \geq \sqrt{6}\ n_\oplus$. Otherwise, if $\rho_0 > R_{SI}$, the boundary of the region containing satellites of the Earth is formed by a segment of the straight line $\rho = R_{SI}$ and the two arcs of the $\dot\rho^2 = G(\rho)$ curve with $0 < \rho < R_{SI}$, as in Figure 5.

To understand the shape of the boundary of the Earth satellites region we need to find possible intersections between the curves $\mathcal{E}_\oplus = 0$ and $\mathcal{E}_\odot = 0$. However, if $\mathcal{E}_\oplus$ is computed in a geocentric approximation, these intersections are physically meaningful only if they occur for $R_\oplus < \rho < R_{SI}$, that is, during a close approach to the Earth, but above its physical surface. We can prove that for $R_\oplus \leq \rho \leq R_{SI}$ the condition $\mathcal{E}_\oplus(\rho,\dot\rho) \leq 0$ implies $\mathcal{E}_\odot(\rho,\dot\rho) \leq 0$.

This result shows that the region of solar system orbits excluding the satellites of the Earth does not have more connected components than the region satisfying condition (40) only. This happens only for particular values of the mass, radius and orbital elements of the Earth, and it is not a general property of whatever planet.

The tiny object boundary

An alternative method to assign a lower limit to the distance is to impose that the object is not very small and very close to the Earth. We assume that the size is controlled by setting a maximum value for the absolute magnitude H:

$$H(\rho) \leq H_{max}\ . \tag{52}$$

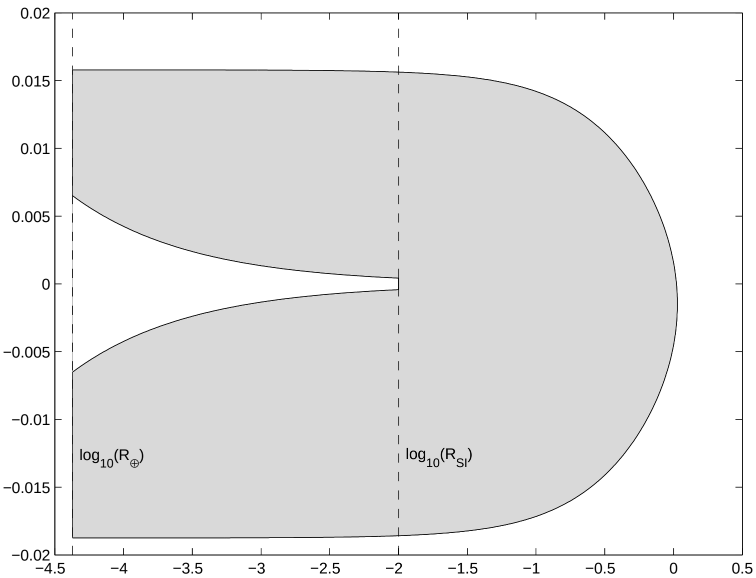

Figure 5: The qualitative features of the region of heliocentric orbits in the $(\log_{10}\rho, \dot\rho)$ plane: by combining conditions (40), (49), (51) and $\rho \geq R_\oplus$, we are left with the domain sketched. The $\log_{10}\rho$ scale is used to enhance the part of the region with small values of ρ.

If an average value h of the apparent magnitude is available, then H can be computed from h using the relation

$$H = h - 5\log_{10}\rho - x(\rho)\,, \tag{53}$$

where the correction $x(\rho)$ accounts for the distance from the Sun and the phase effect. For small values of ρ we can approximate $x(\rho)$ with a quantity x_0 independent of ρ. However, also for larger values of ρ this is an acceptable approximation. In this approximation, condition (52) becomes

$$\log_{10}\rho \geq \frac{h - H_{max} - x_0}{5} \stackrel{def}{=} \log_{10}\rho_H\,,$$

that is, given the apparent magnitude h, we have a minimum distance $\rho_H = \rho(H_{max})$ for the object to be of significant size. For example, using $H_{max} = 30$ we have $\rho \geq 0.01$ AU if $h = 20$, and $\rho \geq 0.001$ AU if $h = 15$. The region satisfying condition (52) is just a half plane $\rho \geq \rho_H$: we call *tiny object boundary* the straight line $\rho = \rho_H$.

Definition of admissible region

We can choose the inner boundary according to the type of objects whose orbit we want to determine, e.g. near-Earth asteroids, trans-Neptunian objects.

As an example, searching for objects in heliocentric orbit with significant size we can assume that $\rho(H_{max}) > R_{SI}$. Given an attributable A and a maximum value for the absolute magnitude H_{max}, we define as admissible region the set

$$\mathcal{D}(A) = \{(\rho, \dot\rho) : \rho \geq \rho_H,\ \mathcal{E}_\odot(\rho, \dot\rho) \leq 0\}\ . \tag{54}$$

The admissible region consists of at most two compact connected components. Its boundary has an outer part, given by arcs of the curve $\mathcal{E}_\odot(\rho, \dot\rho) = 0$, symmetric with respect to the line $\dot\rho = -c_1/2$. The boundary has also an inner part consisting, in the simplest case, of a segment of the line $\rho = \rho(H_{max})$. For smaller objects, with $\rho(H_{max}) < R_{SI}$, the inner boundary has a more complex shape, like the one shown in Figure 5.

Delaunay's triangulations

To sample the admissible region we start by sampling its boundary, selecting points that are as possible equispaced on the boundary. Then we construct a Delaunay triangulation of the region, that is a sampling with the properties described below.

Consider the polygonal domain $\tilde{\mathcal{D}}$ defined by connecting with edges the sample of boundary points of the admissible region $\mathcal{D}$. A triangulation of $\tilde{\mathcal{D}}$ is a pair (Π, τ), where $\Pi = \{P_1, \ldots, P_N\}$ is a set of points (the *nodes*) of the domain, and $\tau = \{T_1, \ldots, T_k\}$ is a set of triangles with vertexes in Π such that:

(i) $\bigcup_{i=1,k} T_i = \tilde{\mathcal{D}}$;

(ii) for each $i \neq j$ the set $T_i \bigcap T_j$ is either empty or a *vertex* or an *edge* of a triangle.

To each triangulation (Π, τ) we can associate the minimum angle, that is the minimum among the angles of all the triangles T_i. Among all possible triangulations of a convex domain the Delaunay triangulation is characterized by these properties:

(i) it maximizes the minimum angle;

(ii) it minimizes the maximum circumcircle;

(iii) for each triangle T_i, the interior part of its circumcircle does not contain any nodes of the triangulation.

These properties are all equivalent for convex domains.

If, in addition to the set of points Π, we give as input also some edges P_iP_j, for example the boundary edges of $\tilde{\mathcal{D}}$, we refer to the corresponding triangulation containing the prescribed edges as a constrained triangulation.

The domain $\tilde{\mathcal{D}}$ is in general not convex: in this case we need to give as input the edges along the boundary. Then there still exists a constrained triangulation such that (i), (ii) hold, called constrained Delaunay triangulation, but property (iii) is not guaranteed.

The definition of Delaunay's triangulation uses distances and angles, thus it depends on the metric selected for the space $(\rho, \dot{\rho})$. In particular we can select a strictly increasing function $f(\rho)$ and perform the triangulation of the admissible region with the metric $ds^2 = df(\rho)^2 + d\dot{\rho}^2$, i.e., we can work in the plane $(f(\rho), \dot{\rho})$ endowed with the Euclidean metric. If our purpose is to search for objects in a particular portion of the $(\rho, \dot{\rho})$ space, then we can use a metric selected *ad hoc.* For example, to enhance the region near the Earth we can use $f(\rho) = \log_{10}(\rho)$, as in Figure 5.

Recursive attribution

Each node of the admissible region corresponds to an orbit to which we can assign a degenerate covariance matrix. This orbit and its covariance can be propagated to the time of a second attributable in order to check the compatibility of both sets of observations. In this procedure we can take advantage of the concept of *Line of Variation* to perform constrained differential corrections.

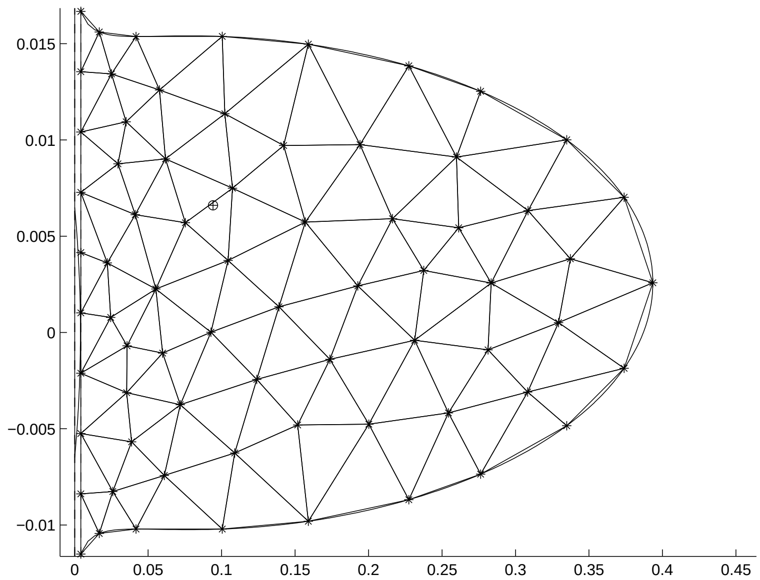

Figure 6: Triangulated admissible region in the plane $\rho, \dot\rho$ for the asteroid 2003 BH_{84}. The units are AU, AU/yr and the coordinates of the asteroid are marked with a circled cross.

6.2 Preliminary orbits with the two-body integrals

It is well known that Kepler's problem

$$\ddot{\mathbf{r}} = -\mu \frac{\mathbf{r}}{|\mathbf{r}|^3}\,, \qquad \mathbf{r} \in \mathbb{R}^3,\ \mu > 0$$

has the following integrals of motions:

$$\begin{aligned} &\mathbf{c} = \mathbf{r} \times \dot{\mathbf{r}} && \text{(angular momentum)}\,,\\ &\mathcal{E} = \frac{1}{2}|\dot{\mathbf{r}}|^2 - \frac{\mu}{|\mathbf{r}|} && \text{(energy)}\,,\\ &\mathbf{L} = \frac{1}{\mu}\dot{\mathbf{r}} \times \mathbf{c} - \frac{\mathbf{r}}{|\mathbf{r}|} && \text{(Laplace-Lenz vector)}\,. \end{aligned} \tag{55}$$

The integrals above give only 5 independent scalar conservation laws, in fact we have the relations

$$\mathbf{L} \cdot \mathbf{c} = 0\,, \qquad 2|\mathbf{c}|^2\mathcal{E} + \mu^2(1 - |\mathbf{L}|^2) = 0\,.$$

The linkage problem can be written using polynomial equations defined by some integrals of Kepler's problem. This approach together with the algebraic elimination of variables allows us to have a global control on the solutions.

The use of the integrals $\mathbf{c}, \mathcal{E}$ to write equations for the linkage problem was first suggested in 1977. The same equations have been reconsidered in 2010, and solved by means of algebraic methods.

In another recent work different equations are considered, writing a suitable projection of the Laplace-Lenz vector in place of the energy.

To set up the equations of the linkage problem we write the Keplerian integrals as function of $(\rho, \dot\rho)$. The angular momentum is

$$\mathbf{c}(\rho, \dot\rho) = \mathbf{r} \times \dot{\mathbf{r}} = \mathbf{D}\dot\rho + \mathbf{E}\rho^2 + \mathbf{F}\rho + \mathbf{G} \tag{56}$$

with

$$\mathbf{D} = \mathbf{q} \times \hat{\mathbf{e}}^\rho \,, \qquad \mathbf{E} = \dot\alpha \cos\delta \hat{\mathbf{e}}^\rho \times \hat{\mathbf{e}}^\alpha + \dot\delta \hat{\mathbf{e}}^\rho \times \hat{\mathbf{e}}^\delta \,,$$
$$\mathbf{F} = \dot\alpha \cos\delta \mathbf{q} \times \hat{\mathbf{e}}^\alpha + \dot\delta \mathbf{q} \times \hat{\mathbf{e}}^\delta + \hat{\mathbf{e}}^\rho \times \dot{\mathbf{q}} \,, \qquad \mathbf{G} = \mathbf{q} \times \dot{\mathbf{q}} \,,$$

where $\hat{\mathbf{e}}^\rho, \hat{\mathbf{e}}^\alpha, \hat{\mathbf{e}}^\delta$ are unit vectors depending only on α, δ, defined as in Section 6.1. Thus $\mathbf{D}, \mathbf{E}, \mathbf{F}, \mathbf{G}$ depend only on the attributable $\mathcal{A}$ and on the motion of the observer $\mathbf{q}, \dot{\mathbf{q}}$ at the mean time $\bar{t}$.

The dependence of the energy function

$$\mathcal{E} = \frac{1}{2}|\dot{\mathbf{r}}|^2 - \frac{\mu}{|\mathbf{r}|}$$

on $\rho, \dot\rho$ is described in Section 6.1.

The Laplace-Lenz vector is given by

$$\mu \mathbf{L}(\rho, \dot\rho) = \dot{\mathbf{r}} \times \mathbf{c} - \mu \frac{\mathbf{r}}{|\mathbf{r}|} = \left(|\dot{\mathbf{r}}|^2 - \frac{\mu}{|\mathbf{r}|}\right)\mathbf{r} - (\dot{\mathbf{r}} \cdot \mathbf{r})\dot{\mathbf{r}}$$

where

$$\begin{aligned} |\mathbf{r}| &= (\rho^2 + |\mathbf{q}|^2 + 2\rho \mathbf{q} \cdot \hat{\mathbf{e}}^\rho)^{1/2} \,, \\ |\dot{\mathbf{r}}|^2 &= \dot\rho^2 + (\dot\alpha^2 \cos^2\delta + \dot\delta^2)\rho^2 + 2\dot{\mathbf{q}} \cdot \hat{\mathbf{e}}^\rho \dot\rho + 2\dot{\mathbf{q}} \cdot (\dot\alpha \cos\delta \hat{\mathbf{e}}^\alpha + \dot\delta \hat{\mathbf{e}}^\delta)\rho + |\dot{\mathbf{q}}|^2 \,, \\ \dot{\mathbf{r}} \cdot \mathbf{r} &= \rho\dot\rho + \mathbf{q} \cdot \hat{\mathbf{e}}^\rho \dot\rho + (\dot{\mathbf{q}} \cdot \hat{\mathbf{e}}^\rho + \mathbf{q} \cdot \hat{\mathbf{e}}^\alpha \dot\alpha \cos\delta + \mathbf{q} \cdot \hat{\mathbf{e}}^\delta \dot\delta)\rho + \dot{\mathbf{q}} \cdot \mathbf{q} \,. \end{aligned}$$

First we describe the algorithm that employs the angular momentum and energy integrals. Given two attributables $\mathcal{A}_1, \mathcal{A}_2$ at times $\bar{t}_1, \bar{t}_2$, equating the angular momentum at the two times we obtain

$$\mathbf{D}_1 \dot\rho_1 - \mathbf{D}_2 \dot\rho_2 = \mathbf{J}(\rho_1, \rho_2) \tag{57}$$

where

$$\mathbf{J}(\rho_1, \rho_2) = \mathbf{E}_2 \rho_2^2 - \mathbf{E}_1 \rho_1^2 + \mathbf{F}_2 \rho_2 - \mathbf{F}_1 \rho_1 + \mathbf{G}_2 - \mathbf{G}_1 \,.$$

Hereafter we use the indexes 1, 2 to denote the quantities defined above at times $\bar{t}_1$, $\bar{t}_2$.

By scalar multiplication of (57) with $\mathbf{D}_1 \times \mathbf{D}_2$ we perform the elimination of the variables $\dot\rho_1, \dot\rho_2$. This yields

$$q(\rho_1, \rho_2) \stackrel{def}{=} \mathbf{D}_1 \times \mathbf{D}_2 \cdot \mathbf{J}(\rho_1, \rho_2) = 0 \,.$$

Now we use the conservation of the energy. By vector multiplication of (57) with $\mathbf{D}_1$ and $\mathbf{D}_2$, projecting on $\mathbf{D}_1 \times \mathbf{D}_2$, we obtain

$$\dot\rho_1(\rho_1, \rho_2) = \frac{(\mathbf{J} \times \mathbf{D}_2) \cdot (\mathbf{D}_1 \times \mathbf{D}_2)}{|\mathbf{D}_1 \times \mathbf{D}_2|^2} \,, \qquad \dot\rho_2(\rho_1, \rho_2) = \frac{(\mathbf{J} \times \mathbf{D}_1) \cdot (\mathbf{D}_1 \times \mathbf{D}_2)}{|\mathbf{D}_1 \times \mathbf{D}_2|^2} \,.$$

Substituting into $\mathcal{E}_1 = \mathcal{E}_2$, rearranging the terms and squaring twice we obtain the polynomial equation $p(\rho_1, \rho_2) = 0$, with total degree 24. We consider the semi-algebraic problem

$$p(\rho_1, \rho_2) = 0 \,, \qquad q(\rho_1, \rho_2) = 0 \,, \qquad \rho_1, \rho_2 > 0 \,. \tag{58}$$

We can write

$$p(\rho_1, \rho_2) = \sum_{j=0}^{20} a_j(\rho_2)\, \rho_1^j \,, \qquad q(\rho_1, \rho_2) = b_2\, \rho_1^2 + b_1\, \rho_1 + b_0(\rho_2)$$

for some coefficients a_i, b_j, depending only on ρ_2.

Now we eliminate the variable ρ_1. Consider the resultant $Res(\rho_2)$ of p, q with respect to ρ_1: it is a degree 48 polynomial defined as the determinant of the Sylvester matrix

$$\begin{pmatrix} a_{20} & 0 & b_2 & 0 & \dots & \dots & 0 \\ a_{19} & a_{20} & b_1 & b_2 & 0 & \dots & 0 \\ \vdots & \vdots & b_0 & b_1 & b_2 & \dots & \vdots \\ \vdots & \vdots & 0 & b_0 & b_1 & \dots & \vdots \\ a_0 & a_1 & \vdots & \vdots & \vdots & b_0 & b_1 \\ 0 & a_0 & 0 & 0 & 0 & 0 & b_0 \end{pmatrix} .$$

The roots of $Res(\rho_2)$ give us all the ρ_2 components of the solutions of (58). The computation of the solutions can be done as follows:

i) compute the positive roots ρ_2 of $Res(\rho_2)$;

ii) for each root find the corresponding values of ρ_1, $\dot\rho_1, \dot\rho_2$;

iii) discard the spurious solutions, obtained by squaring;

iv) compute the related Keplerian orbits at times $\tilde t_i = \bar t_i - \dfrac{\rho_i}{c}, i = 1, 2$, corrected by aberration, with c the speed of light.

One method to select among alternative solutions is to use compatibility conditions. The knowledge of the angular momentum vector and of the energy at a given time yields the values of

$$a, e, I, \Omega \ .$$

From $\mathbf{c}_1 = \mathbf{c}_2$, $\mathcal{E}_1 = \mathcal{E}_2$ we obtain the same values of a, e, I, Ω at times $\tilde t_1, \tilde t_2$, but we must check the compatibility conditions

$$\omega_1 = \omega_2 \,, \qquad \ell_1 = \ell_2 + n(\tilde t_1 - \tilde t_2) \,, \tag{59}$$

where n is the mean motion.

To take into account the errors in the observations we can consider the map

$$(\mathcal{A}_1, \mathcal{A}_2) \overset{\mathbf{\Psi}}{\mapsto} (\mathcal{A}_1, \rho_1, \dot\rho_1, \Delta_{1,2}) \ , \qquad \Delta_{1,2} = (\Delta\omega, \Delta\ell)$$

giving the orbit at time $\tilde t_1$, and the difference in ω, ℓ. Then we check whether $\Delta_{1,2} = \mathbf{0}$ is compatible with the observational errors by covariance propagation through the map $\mathbf{\Psi}$. This algorithm also allows to define covariance matrices for the preliminary orbits that we compute.

It is possible to reduce the algebraic degree of the linkage problem by writing different equations (i.e. using different integrals): we select a suitable component of the Laplace-Lenz vector in place of the energy. Given $\mathcal{A}_1, \mathcal{A}_2$ we equate $\mathbf{L}_1, \mathbf{L}_2$ projected along $\mathbf{v} = \hat{\mathbf{e}}_2^\rho \times \mathbf{q}_2$:

$$\mathbf{L}_1(\rho_1, \dot\rho_1) \cdot \mathbf{v} = \mathbf{L}_2(\rho_2, \dot\rho_2) \cdot \mathbf{v} \ . \tag{60}$$

We have

$$\Big(|\dot{\mathbf{r}}_1|^2 - \frac{\mu}{|\mathbf{r}_1|}\Big)(\mathbf{r}_1 \cdot \mathbf{v}) - (\dot{\mathbf{r}}_1 \cdot \mathbf{r}_1)(\dot{\mathbf{r}}_1 \cdot \mathbf{v}) = -(\dot{\mathbf{r}}_2 \cdot \mathbf{r}_2)(\dot{\mathbf{r}}_2 \cdot \mathbf{v}) \ .$$

Rearranging the terms and squaring we obtain

$$\tilde{p}(\rho_1, \rho_2) \stackrel{def}{=} \mu^2(\mathbf{r}_1 \cdot \mathbf{v})^2 - |\mathbf{r}_1|^2 \left\{ \left[|\dot{\mathbf{r}}_1|^2 \mathbf{r}_1 - (\dot{\mathbf{r}}_1 \cdot \mathbf{r}_1)\dot{\mathbf{r}}_1 + (\dot{\mathbf{r}}_2 \cdot \mathbf{r}_2)\dot{\mathbf{r}}_2 \right] \cdot \mathbf{v} \right\}^2 = 0 \; .$$

$\tilde{p}$ is a polynomial of total degree 10 in ρ_1, ρ_2, therefore the system

$$\tilde{p}(\rho_1, \rho_2) = 0 \, , \quad q(\rho_1, \rho_2) = 0 \qquad (\rho_1, \rho_2 > 0) \tag{61}$$

has degree 20.

Taking two sets of observations of asteroid (99942) Apophis, at mean epochs $\bar{t}_1 = 53175.59$, $\bar{t}_2 = 53357.45$ MJD, we show in Figure 7 the advantage of using equation (60) instead of the conservation of the energy $\mathcal{E}$.

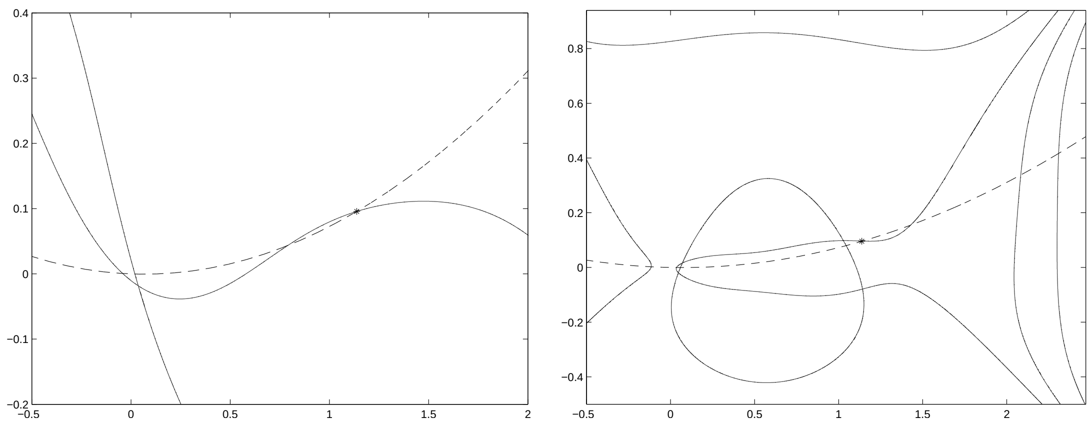

Figure 7: Comparison of the intersections of the algebraic curves computed for the linkage problem in the plane (ρ_1, ρ_2). Left: curves defined by $\tilde{p} = 0$ (solid) and $q = 0$ (dashed) in (61), using the integral $\mathbf{L} \cdot \mathbf{v}$. Right: curves defined by $p = 0$ (solid) and $q = 0$ (dashed) in (58), using $\mathcal{E}$. With kind permission from Springer.

We can choose among alternative solutions of the linkage problem also by means of the attribution algorithm. Let $\mathcal{E}_1$ be a set of orbital elements at time t_1, with covariance matrix Γ_1. Propagate orbit and covariance to the epoch $\bar{t}_2$ of $\mathcal{A}_2$, with covariance $\Gamma_{\mathcal{A}_2}$. Then extract a predicted attributable $\mathcal{A}_p$, at time $\bar{t}_2$, with covariance $\Gamma_{\mathcal{A}_p}$. We can compare $\mathcal{A}_p, \Gamma_{\mathcal{A}_p}$ with $\mathcal{A}_2, \Gamma_{\mathcal{A}_2}$ by defining an identification penalty χ_4, that gives the price to pay to assume that the observations of both attributables belong to the same celestial body.

Acknowledgments

The author is grateful to A. Milani and G. B. Valsecchi for several interesting discussions about different aspects of orbit determination.

Glossary

Admissible region: compact set of data in the plane $(\rho, \dot{\rho})$ of the radial distance and velocity of an observed asteroid. It is defined by imposing dynamical and physical constraints on the asteroid.
Asteroid survey: systematic scan of the sky, with a telescope, to produce an asteroid catalog.
Attributable: 4-dimensional vector giving the angular position and angular velocity of an asteroid on the celestial sphere at a certain time.

Differential corrections: Iterative algorithm that implements the least squares method.
Identification: to establish that two sets of asteroid observations belong to the same celestial body.
Line of Variation: one-dimensional set, in the orbital elements space, representing a simplified model for the confidence region.
Linkage: identification of two very short arcs of asteroid observations as belonging to the same celestial body.
Preliminary orbit: orbit to be used as starting guess for the differential corrections.
Very short arc: small set of asteroid observations, usually referred to the same observing night, that is used to define an attributable.

Biographical Sketch

Giovanni Federico Gronchi received his Ph.D. in Mathematics from the University of Pisa. At present he is Associate Researcher of Mathematical Physics at the Department of Mathematics, University of Pisa. His research is about Solar System body dynamics, perturbation theory, orbit determination, singularities and periodic orbits of the N-body problem.

References

[1] A. Celletti, G. Pinzari: 'Four Classical Methods for Determining Planetary Elliptic Elements: A Comparison', CMDA **93**, 1–52 (2005) [Description of some classical orbit determination methods, including the ones of Gauss and Laplace. Here the Gauss map is described in full details.]

[2] A. Celletti, G. Pinzari: 'Dependence on the observational time intervals and domain of convergence of orbital determination methods', CMDA **95**, 327–344 (2006) [This paper investigates how the variation of the time intervals between the observations affects the determination of the orbits using the classical methods. Sufficient conditions for the convergence of Gauss' method are also introduced.]

[3] C. V. L. Charlier: 'On Multiple Solutions in the Determination of Orbits from three Observation', MNRAS **71**, 120–124 (1910) [Geometric interpretation of the occurrence of alternative solutions in preliminary orbit determination from three observations. Here the observations are assumed to be geocentric.]

[4] C. V. L. Charlier: 'Second Note on Multiple Solutions in the Determination of Orbits from three Observation', MNRAS **71**, 454–459 (1911) [Additional study of alternative solutions and introduction of the singular curve.]

[5] G. Foderà Serio, A. Manara, P. Sicoli: 'Giuseppe Piazzi and the Discovery of Ceres', in Asteroids III, W.F. Bottke Jr. et al. eds., Arizona University Press (2003) [The history of the discovery of the asteroid Ceres, and of the determination of its orbit by Gauss.]

[6] C. F. Gauss: 'Theoria motus corporum coelestium in sectionibus conicis Solem ambientium' (1809), 'Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections', reprinted by Dover publications (1963) [Source reference for Gauss' method, introducing preliminary orbits, and their refinement by iterative methods. The least squares method is also presented here.]

[7] G. F. Gronchi: 'Multiple Solutions in Preliminary Orbit Determination from Three Observations', CMDA **103/4**, 301–326 (2009) [Generalization of Charlier's theory taking into account topocentric observations.]

[8] G. F. Gronchi, L. Dimare, A. Milani: 'Orbit determination with the two-body integrals', CMDA **107/3**, 299–318 (2010) [Introduction of algebraic methods to solve the linkage equations with the angular momentum and energy integrals.]

[9] G. F. Gronchi, D. Farnocchia, L. Dimare: 'Orbit determination with the two-body integrals. II', CMDA **110/3**, 257–270 (2011) [Algebraic methods applied to different linkage equations, with lower degree. A suitable component of Laplace-Lenz vector is used in place of the energy.]

[10] S. Herrick: 'Astrodynamics', Vol.1, Van Nostrand Reinhold, London (1971) [The derivation of Gibbs' formulae for the computation of the velocities can be found here.]

[11] P. S. Laplace: Mém. Acad. R. Sci. Paris, in *Laplace's collected works* **10**, 93–146 (1780) [Source reference for Laplace's method.]

[12] A. M. Legendre: 'Nouvelles Méthodes pour la déterminations des orbits des comètes' (1805) [This treatise is related to the dispute for the invention of the least squares method.]

[13] A. Milani: 'The Asteroid Identification Problem I: recovery of lost asteroids', Icarus **137**, 269–292 (1999) [Description of the orbit identification problems.]

[14] A. Milani, G. F. Gronchi, M. de' Michieli Vitturi and Z. Knežević: 'Orbit Determination with Very Short Arcs. I Admissible Regions', CMDA **90**, 59–87 (2004) [Here the admissible region is introduced, together with the detailed proofs of its properties.]

[15] A. Milani, A., G. F. Gronchi, Z. Knežević, M. E. Sansaturio and O. Arratia: 'Orbit Determination with Very Short Arcs. II Identifications', Icarus **79**, 350–374 (2005) [Introduction of the recursive attribution method, that employs the admissible region.]

[16] A. Milani, A. La Spina, M. E. Sansaturio and S. R. Chesley: 'The Asteroid Identification Problem III. Proposing identifications', Icarus **144**, 39–53 (2000) [Introduction of an orbit identification method.]

[17] A. Milani, S. R. Chesley, M. E. Sansaturio, G. Tommei, and G. B. Valsecchi: 'Nonlinear impact monitoring: line of variation searches for impactors', Icarus **173**, 362–384 (2005) [Introduction of the Line of Variation.]

[18] A. Milani, G. F. Gronchi, D. Farnocchia, Z. Knežević, R. Jedicke, L. Dennau, F. Pierfederici: 'Topocentric Orbit Determination: Algorithms for the Next Generation Surveys', Icarus **195**, 474–492 (2008) [Comparison of the classical methods, showing an extension of Laplace's method, that accounts for topocentric observations. There are also examples of occurrence of alternative solutions.]

[19] A. Milani, G. F. Gronchi: 'Theory of Orbit Determination', Cambridge University Press (2010) [Treatise about the modern methods of orbit determination. It describes in details most of the issues that we have presented.]

[20] O. F. Mossotti: 'Nuova analisi del problema di determinare le orbite dei corpi celesti', in *Mossotti, scritti*, Domus Galileiana, Pisa (1816-1818) [Articles with some classical orbit determination methods, different from the ones by Gauss and Laplace.]

[21] H. C. Plummer: 'An introductory treatise on Dynamical Astronomy', Cambridge University press, reprinted by Dover publications, New York (1960) [In this book there is a detailed review of Charlier's theory.]

[22] H. Poincaré: 'Sur la détermination des orbites par la méthode de Laplace', Bulletin astronomique **23**, 161–187 (1906) [It is about the use of Laplace's method, and introduces also a way to deal with topocentric observations.]

[23] S. Stigler: 'Gauss and the Invention of Least Squares', Ann. of Stat. **9/3** , 465–474 (1981) [The history of the dispute for the invention of the least squares method.]

[24] L. G. Taff, D. L. Hall: 'The use of angles and angular rates. I - Initial orbit determination', CMDA **16**, 481–488 (1977) [The equations for the linkage using the angular momentum and energy integrals are first introduced here.]

[25] F. Tisserand, J. Perchot: 'Lecons sur la détermination des orbites', Paris, Gauthier-Villars (1899) [Important treatise of orbit determination.]