Classical and modern orbit determination for asteroids

Giovanni F. Gronchi

Department of Mathematics, University of Pisa, Via Buonarroti 2, 56127 Pisa, Italy
email: gronchi@dm.unipi.it

Abstract. With the substantial improvements in observational techniques we have to deal with very big databases, consisting of a few positions of an object over a short time span; this is often not enough to compute a preliminary orbit with traditional tools. In this paper we first review a classical method by C.F. Gauss to compute a preliminary orbit for asteroids. This method, followed by a least squares fit to improve the orbit, still today gives successful results when we have at least three separate observations. Then we introduce the basics of a very recent orbit determination theory, that has been thought just to be used with modern sets of data. These data allow us in many cases to know the angular position and velocity of an asteroid at a given time, even though the radial distance and velocity \((r, \dot{r})\), needed to compute its full orbit, are unknown. The variables \((r, \dot{r})\) can be constrained to a compact set, that we call the admissible region (AR), whose definition requires that the body belongs to the Solar System, that it is not a satellite of the Earth, and that it is not a "shooting star" (i.e. very close and very small). We provide a mathematical description of the AR: its topological properties are surprisingly simple, in fact it turns out that the AR cannot have more than two connected components. A sampling of the AR can be performed by means of a Delaunay triangulation; a finite number of six-parameter sets of initial conditions are thus defined, with each node of the triangulation representing a possible orbit (a virtual asteroid).

1. Classical orbit determination: Gauss’s method

An important problem in orbit determination is the computation of the orbital elements of a celestial body using a given set of angular observations \((\alpha_i, \delta_i)\), that are, for example, the right ascension and the declination of the object on the celestial sphere.

This problem became particularly challenging when the Italian astronomer Giuseppe Piazzi discovered the first asteroid (Ceres) on 1 January 1801; he could follow this asteroid until 11 February 1801 writing for it 21 observations that covered only 3° in the sky (Foderà Serio et al. 2003).

At the beginning of the 19th century some methods for orbit determination had already been invented, but they did not allow the recovery of Ceres in the sky; the nature of the observations at disposal made the problem to be set as follows:

\[
\textit{determinare orbitam corporis coelestis, absque omni suppositione hypothetica, ex observationibus tempus haud magnum complectentibus neque adeo delectum, pro applicatione methodorum specialium, patientibus} \quad \text{† (C. F. Gauss 1809)}
\]

The two astronomers Franz Von Zach and Heinrich W. Olbers independently recovered

† ‘to determine the orbit of a celestial body, without making any hypothesis, from observations covering a space neither too large nor such as to allow the special methods to be applied’
Ceres (on 31 December 1801 and on 2 January 1802, respectively) following the indications of the German mathematician Carl Friedrich Gauss, who developed very important mathematical techniques just to deal with this problem (see Gauss 1809).

1.1. Gauss’s method for preliminary orbits

The orbital elements, defining the Keplerian orbit of an object, are six, so that we can try to determine completely an orbit using only three angular observations \( \{ (\alpha_i, \delta_i), \; i = 1, 2, 3 \} \).

Let us write three heliocentric positions \( \vec{r}_k \) of a celestial body as

\[
\vec{r}_k = \rho_k \vec{s}_k + \vec{R}_k; \quad k = 1, \ldots, 3, \tag{1.1}
\]

where \( \vec{R}_k = (R_{1,k}, R_{2,k}, R_{3,k}) \) is the (known) heliocentric position of the observer at time \( t_k \); \( \vec{s}_k = (s_{1,k}, s_{2,k}, s_{3,k}) \) is the (known) position of the body at time \( t_k \) on the unit sphere centered at the observer’s position,

\[
\vec{s}_k = (\cos \delta_k \cos \alpha_k, \cos \delta_k \sin \alpha_k, \sin \delta_k),
\]

and the three positive scalars \( \rho_k \) are unknown.

As a Keplerian orbit lies on a plane, the vectors \( \vec{r}_k \) must be linearly dependent: there exist \( c_1, c_2, c_3 \in \mathbb{R} \) such that

\[
c_1 \vec{r}_1 + c_2 \vec{r}_2 + c_3 \vec{r}_3 = 0. \tag{1.2}
\]

Set \( c_2 = -1 \), so that

\[
\vec{r}_2 = c_1 \vec{r}_1 + c_3 \vec{r}_3,
\]

and

\[
\begin{align*}
\vec{r}_1 \wedge \vec{r}_2 &= c_3 \vec{r}_1 \wedge \vec{r}_3, \\
\vec{r}_3 \wedge \vec{r}_2 &= c_1 \vec{r}_3 \wedge \vec{r}_1.
\end{align*}
\]

If \( \vec{r}_1 \wedge \vec{r}_3 \neq \vec{0} \) we can write the coefficients in terms of the \( f_k, g_k \) series (see Roy 1988) for \( k = 1, 3 \):

\[
c_1 = \frac{g_1}{f_1 g_3 - f_3 g_1}; \quad c_3 = \frac{-g_3}{f_1 g_3 - f_3 g_1},
\]

where \( f_k, g_k \) are such that

\[
\vec{r}_1 = f_1 \vec{r}_2 + g_1 \vec{r}_2; \quad \vec{r}_3 = f_3 \vec{r}_2 + g_3 \vec{r}_2.
\]

Taking \( t = t_2 \) for the origin of the time, the values \( t_1, t_3 \) are comparably small; thus we can approximate the \( f_k, g_k \) series up to the third order in \( t_k \):

\[
f_k \approx 1 - \frac{\mu t_k^2}{2r_2^2}; \quad g_k \approx t_k - \frac{\mu t_k^3}{6r_2^4},
\]

with \( \mu \) the total mass of the two bodies. In case of an asteroid (a small body) orbiting around the Sun, \( \mu \) can be taken as the mass of the Sun only. Note that the approximations of \( f_k, g_k \) depend only on \( r_2 \).

Substituting (1.1) in (1.2) we obtain

\[
c_1 \rho_1 \vec{s}_1 + c_2 \rho_2 \vec{s}_2 + c_3 \rho_3 \vec{s}_3 + c_1 \vec{R}_1 + c_2 \vec{R}_2 + c_3 \vec{R}_3 = 0, \tag{1.3}
\]

that gives a system of 3 equations in the 4 unknowns \( \rho_1, \rho_2, \rho_3, r_2 \). Note that we also have the relation

\[
r_2^2 = \rho_2^2 + 2 \rho_2 \vec{s}_2 \cdot \vec{R}_2 + R_2^2, \tag{1.4}
\]

that gives \( r_2 \) in terms of \( \rho_2 \).
Then we use the approximations
\[ c_1 = c_1(r_2) \approx \frac{t_3}{t_3 - t_1} \left[ 1 + \frac{\mu}{6t_2} ((t_3 - t_1)^2 - t_3^2) \right] = A_1 + \frac{B_1}{r_2}; \]
\[ c_3 = c_3(r_2) \approx -\frac{t_1}{t_3 - t_1} \left[ 1 + \frac{\mu}{6t_2} ((t_3 - t_1)^2 - t_1^2) \right] = A_3 + \frac{B_3}{r_2}. \]

**Remark:** the coefficients
\[ A_1 = \frac{t_3}{t_3 - t_1}; \quad B_1 = \frac{\mu t_3}{6(t_3 - t_1)} ((t_3 - t_1)^2 - t_3^2); \]
\[ A_3 = -\frac{t_1}{t_3 - t_1}; \quad B_3 = -\frac{\mu t_1}{6(t_3 - t_1)} ((t_3 - t_1)^2 - t_1^2); \]
are positive, and \( c_1, c_3 \) are functions of the variable \( r_2 \) only.

We can write equation (1.3) as follows
\[
\begin{pmatrix}
  s_{1,1} & s_{1,2} & s_{1,3} \\
  s_{2,1} & s_{2,2} & s_{2,3} \\
  s_{3,1} & s_{3,2} & s_{3,3}
\end{pmatrix}
\begin{pmatrix}
  c_{1,1} \\
  c_{2,1} \\
  c_{3,1}
\end{pmatrix}
= -
\begin{pmatrix}
  G_1 \\
  G_2 \\
  G_3
\end{pmatrix},
\]
where \( G_i = \sum_{k=1}^{3} R_{i,k} c_k \) and \( R_{i,k} \) is the \( i \)-th component of \( R_k \). Computing \( S^{-1} := (\sigma_{i,k})_{i,k} \), the inverse of the matrix \( S := (s_{i,k})_{i,k} \), we obtain
\[ c_k \rho_k = -\sum_{i=1}^{3} \sigma_{k,i} G_i; \quad k = 1, \ldots, 3 \]
and in particular
\[ \rho_2 = \sum_{i=1}^{3} \sigma_{2,i} G_i. \tag{1.5} \]

By substituting the approximated values of \( c_1, c_3 \) in (1.5) we obtain
\[ \rho_2 = \sum_{i=1}^{3} \sigma_{2,i} \sum_{k=1}^{3} R_{i,k} c_k = \tilde{A}_2 + \frac{\tilde{B}_2}{r_2^3}, \tag{1.6} \]
with
\[ \tilde{A}_2 = \sum_{i=1}^{3} \sigma_{2,i} \sum_{k=1}^{3} R_{i,k} A_k; \quad \tilde{B}_2 = \sum_{i=1}^{3} \sigma_{2,i} \sum_{k=1}^{3} R_{i,k} B_k, \]
where we have defined \( A_2 = 1, B_2 = 0 \).

Finally, substituting relation (1.6) into the equation (1.4), we obtain an 8th-degree equation for \( r_2 \):
\[ p(r_2) = r_2^8 - \left[ \tilde{A}_2^2 + 2 \tilde{A}_2 (\vec{s}_2, \vec{R}_2) + \|\vec{R}_2\|^2 \right] r_2^6 - 2 \tilde{B}_2 \left[ \tilde{A}_2 + \langle \vec{s}_2, \vec{R}_2 \rangle \right] r_2^4 - \tilde{B}_2^2 = 0, \tag{1.7} \]
whose positive solutions gives us preliminary orbits.

**Remarks on the Solutions of (1.7):**

(i) as \( p(r_2) \) has only 4 terms, by Cartesió’s sign rule we know that equation (1.7) cannot have more than 3 positive solutions;

(ii) the coefficient of the 6th-degree term is always non-positive because we have
\[ \left[ \tilde{A}_2^2 + 2 \tilde{A}_2 (\vec{s}_2, \vec{R}_2) + \|\vec{R}_2\|^2 \right] = \|\vec{A}_2 \vec{s}_2 + \vec{R}_2\|^2 \geq 0; \]
(iii) the scalar product $\langle \tilde{x}_2, \tilde{R}_2 \rangle$ is positive if the observations are done in the region of the sky from quadrature to opposition;

(iv) the only coefficient of $p(r_2)$ that may change sign is the coefficient of the 3rd-degree term (the only odd term) so that, apart from vanishing of the coefficients, the sequence of signs of the coefficients of $p(r_2)$ can be

$$[+, -, +, -] \text{ or } [+, -, -, -]$$

and, respectively, the sequence of signs of the coefficients of $q(r_2) := p(-r_2)$ can be

$$[+, -, -, -] \text{ or } [+,-,+,+].$$

From the previous remarks we notice that $p(r_2)$ has at most 4 roots, and at most 3 positive ones. In fact, the number of sign changes of either the coefficients of $p(r_2)$ or the coefficients of $q(r_2) := p(-r_2)$ is only one, so that Cartesio’s rule says that either on the positive axis $r_2 > 0$ or on the negative axis $r_2 < 0$ there is exactly one root.

In several cases we obtain only one preliminary orbit from the method explained before; this orbit can be improved when additional observations are available and Gauss himself introduced the least squares method just to deal with this problem: the improvement procedure is usually called differential corrections and the classical algorithm for orbit determination, often used still today, consists of the two steps:

(a) search for a preliminary orbit;

(b) apply the differential corrections to improve it.

**Hint:** When the arc of the observations is short, or when it does not present significant geodetic curvature, then Gauss’s method cannot be applied. In fact, the shortness of the arc corresponds to the smallness of the determinant of the matrix $S$ that is inverted in this algorithm, and the uncertainty in the observations may allow the situation shown in Fig. 1 in the case of short arcs.

**Figure 1.** Qualitative sketch in which the orbit given by Gauss’s method (preliminary orbit plus differential corrections) presents a curvature that is not reliable at all.

The observations of an asteroid/comet at Gauss’s epoch (19th century) were not more than one per night, while today we have to deal with sets of observations all referred to the same night, in particular the hypotheses used by Gauss are usually not fulfilled for the modern sets of data.
2. Modern orbit determination: too short arcs

We introduce some definitions and some mathematical tools that can be used in the orbit determination problem when the hypotheses of Gauss’s method are not fulfilled.

2.1. Preliminary definitions

A sequence of observations is a set of astrometric observations belonging to the same object:

\[ t_i, \alpha_i, \delta_i, h_i; \quad i = 1, \ldots, m; \quad m \geq 2, \]

where \( \alpha_i, \delta_i \) are angles (RA, DEC) and \( t_i \) are times, with \( t_i < t_{i+1} \). The quantities \( h_i \) are the (optional) values of the apparent magnitude.

A sequence of observations is called a very short arc if the observations are known to belong to the same object just because they can be fit together by some smooth curve, typically a low degree polynomial.

If either the preliminary orbit cannot be determined, or it can but the differential corrections procedure does not converge, we speak of a too short arc.

We recall the following fundamental concept, first introduced in (Milani et al. 2001):

**Definition:** we shall call attributable a vector

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

representing a sequence of observations at the average time \( t^* = \text{Mean}(t_i) \).

![Figure 2. The celestial sphere and the components of an attributable.](image)

The average apparent magnitude \( h^* \) is also part of the attributable if it is available, i.e. if at least one value \( h_i \) of the apparent magnitude is given.

Note that for a given attributable the (geocentric) radial position and velocity \( r, \dot{r} \) of the body are completely undetermined.

2.2. Attributables and orbit determination

If the angular arc length is large enough it may be possible to compute a preliminary orbit by classical methods, e.g. by Gauss’s method. Otherwise, an attributable can be used to perform the following operations:

(a) an attribution, that is a least squares fit identification of the object the attributable is referred to, with another object with a known orbit;

(b) a linkage, that is to join together two attributables, referring to two objects both without a full orbit, to form a set of observations allowing to compute an orbit;
(c) a recovery (resp. a precovery), that is to find in the sky (resp. in the archives) an object for which only an attributable is available; this operation may be difficult, but is possible for a short time after/before $t^*$. 

In the following we shall try to answer these basic questions:

(i) *Can we give some dynamical and physical constraints on $r, \dot{r}$ in order to extract additional orbital information from each single attributable?*

(ii) *Can we represent this information in a suitable way for applications?*

3. The admissible region

Given an attributable we do not have, from the observations, any information on the values of the topocentric distance $r$ and of the range rate $\dot{r}$; we shall make some assumptions on the nature of the object to constrain their values.

**Definition:** The admissible region (AR) is the subset of the $(r, \dot{r}) \in \mathbb{R}^+ \times \mathbb{R}$ defined using the following conditions:

(A) *The object is not a satellite of the Earth,* that is its geocentric energy

$$\mathcal{E}_\oplus(r, \dot{r}) \geq 0. \quad (3.1)$$

Note that this condition is meaningful only if

(B) *The object is inside the sphere of influence of the Earth:*

$$r \leq R_{SI}; \quad R_{SI} \approx 0.010044 \text{ AU},$$

that is we shall require that condition $[(A) \cup (B)^c]$ is fulfilled.

(C) *The object belongs to the Solar System,* that is its heliocentric energy

$$\mathcal{E}_\odot(r, \dot{r}) \leq 0. \quad (3.2)$$

(D) *The object is not a shooting star* (i.e. very small and very close). We can assume that the absolute magnitude

$$H \leq H_{max}. \quad (3.3)$$

![Figure 3. Qualitative sketch of an admissible region: it is defined by condition $[(A) \cup (B)^c] \cap (C) \cap (D)$.](image)

† We can easily modify this condition in order to obtain objects with semimajor axis less, for example, than a few hundreds of AU: the condition becomes $\mathcal{E}_\odot \leq -k^2/(2a_{max})$. 
We shall examine in more details the constraints given by the previous conditions:

**CONDITION [(A) ∪ (B)c]: NOT AN ARTIFICIAL SATIEL OF THE EARTH**

We can write condition (3.1) as

\[
\dot{r}^2 \geq G(r) = \frac{2k^2 \mu_\oplus}{r} - \eta^2 r^2,
\]

where \( \eta = \sqrt{\delta^2 + \bar{\alpha}^2 \cos^2 \delta} \) is the *proper motion*.

The inequality (3.4) gives a constraint to \((r, \dot{r})\) only in the region where

\[
r \leq R_{SI} = a_\oplus \left( \frac{\mu_\oplus}{3} \right)^{1/3}.
\]

Note that \(G(r)\) is positive for \(0 < r < r_0 = (2k^2 \mu_\oplus/\eta^2)^{1/3}\) and that \(r_0 \leq R_{SI}\) only for \(\eta \geq \sqrt{\delta n_\oplus}\), very fast moving objects.

**CONDITION (C): NOT AN INTERSTELLAR OBJECT**

Condition (3.2) is equivalent to

\[
V(r) = P^2(r) S(r) \leq 4k^4,
\]

where \(P(r)\) and \(S(r)\) are 2nd-degree polynomials, so that we have a 6th-degree polynomial inequality.

The admissible region, defined by \([ (A) \cup (B)c ] \cap (C) \cap (D)\), may have in principle a complicated structure due to possible intersections of the curves defining its boundary. Luckily, and surprisingly enough, the structure of the AR is rather simple and it is possible to have a precise understanding of its topological features, necessary to perform a suitable sampling of it.

**TOPOLOGY OF THE ADMISSIBLE REGION**

The following results have been proven in (Milani et al. 2004a):

**Theorem 3.1.** The equation \(V(r) = 4k^4\) has at most three positive roots. Thus the region \(\mathcal{E}_\oplus \leq 0, r > 0\) has either one or two connected components †.

**Theorem 3.2.** For \(R_\oplus < r < R_{SI}\) (\(R_\oplus = \text{radius of the Earth}\))

\[
\mathcal{E}_\oplus \leq 0 \Rightarrow \mathcal{E}_\odot \leq 0.
\]

**Corollary 1.** Intersections between the boundaries \(\mathcal{E}_\oplus = 0\) and \(\mathcal{E}_\odot = 0\) are excluded in the region with \(R_\oplus \leq r \leq R_{SI}\).

**CONDITION (D): NOT A SHOOTING STAR**

Suppose the average apparent magnitude \(h^*\) is available. Then the absolute magnitude \(H\) can be computed:

\[
H = h^* - 5 \log_{10} r - x(r).
\]

† HINT: Multiple roots are also constrained.
In the approximation with $x(r) = 0$, condition (3.3) is
\[ \log_{10} r \geq \frac{h^* - H_{\text{max}}}{5} = \log_{10} r_H, \]  

hence the object needs to be very bright to be very close, otherwise it is just a shooting star.

Note that condition (3.5) does not change the topological features of the AR obtained by the two previous theorems, as it says that $(r, \dot{r})$ lies in the half-plane $r \geq r_H$.

4. Sampling the admissible region: virtual asteroids

Given an admissible region, we sample it with virtual asteroids (VAs):

1. The boundary is densely sampled. This requires us to follow all the portions of the boundary, defined by
\[ E_\odot = 0; \quad r = R_{SI}; \quad r = r_H; \quad E_\odot = -\frac{k^2}{2\sigma_{\text{max}}}. \]

2. A first triangulation is formed by joining the VAs along the boundary, selecting the most regular among the possible ones with these nodes (Delaunay’s triangulation).

3. The triangulation is densified by adding points optimally selected to increase the regularity, until some target number of VAs is reached.

For each attributable, the nodes $N_i = (r_i, \dot{r}_i), i = 1, N$ are stored together with the incidence relations, that is the list of triples of nodes forming the triangles.

For each node $N_i$ there is a VA, i.e. a complete 6-elements orbit.

5. Examples

We present the case of the main belt asteroid 2003 BH$_{84}$: in Fig. 4 we draw the triangulated admissible region defined by the attributable obtained from the only observations of the day of its discovery, on 25 January 2003.

**Figure 4.** The AR for the asteroid 2003 BH$_{84}$: on the horizontal axis we plot $f(r) = 1 - e^{-r^2/(2s^2)}$, where $s = \max(r)$, in AU on the vertical axis we plot $\dot{r}$ in AU/day. The real position of the asteroid ($r \approx 1.9834, \dot{r} \approx 0.0066$) is marked with a plus.
Triangulated ephemerides for 2003 BH84

Once we have triangulated the AR, each of the nodes of the triangulation gives us a full orbit that can be propagated in the past or in the future. In Fig. 5 we draw the projections of the AR onto the planes \((\alpha, \delta), (\dot{\alpha}, \dot{\delta}), (\alpha, \dot{\alpha}), (\delta, \dot{\delta})\), after a propagation to 6 February 2003. We also plot with a circled asterisk the components of the attributable obtained from the only observations of 6 February, and we note that, while in the three planes \((\alpha, \delta), (\dot{\alpha}, \dot{\delta}), (\alpha, \dot{\alpha})\) these components lie inside the image of the triangulation, in the \((\delta, \dot{\delta})\) plane they fall outside it. Actually this feature can be explained only if we take into account the uncertainty of the orbits that we have propagated, as it is done in (Milani et al. 2004b).

Figure 5. Triangulated ephemerides for 2003 BH84. Top left: \((-\alpha, \delta)\) plane; top right: \((-\dot{\alpha}, \dot{\delta})\) plane; bottom left: \((-\alpha, -\dot{\alpha})\) plane; bottom right: \((\delta, \dot{\delta})\) plane. Angles are in degrees and angular velocities are in degrees per day.

6. Conclusions and future work

Given a set of observations of an asteroid, for which it is not possible to write a full orbit, we can often define an attributable. The latter, under natural assumptions about the object, can provide useful orbital information through its related admissible region. In fact, this region can be suitably sampled and for each point of the sampling we have a full orbit that can be propagated together with its uncertainty up to a certain time. We think that these mathematical tools can be successfully used in performing orbit determination with too short arcs and we are working on the creation of algorithms to do that.
Acknowledgements

I have reported work done in collaboration with A. Milani, M. de’ Michieli Vitturi and Z. Knežević. I’d like to thank here all of them. Thanks also to G. Tommei who helped me with suggestions and comments on this paper.

References


Gauss, C. F. 1809, Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections reprinted by Dover publications, 1963


Roy, A. E. 1988, Orbital Motion, Adam Hilger, Bristol

Discussion

KLEOMENIS TSIGANIS: If you compute the admissible region – so the object can be either main belt, or centaur or whatever – could you try to calculate a more probable admissible region by using the current distribution of asteroids? We know that it is much more likely to get an asteroid between 2 AU and 4 AU, so if you could weight your admissible region by the distribution of some statistical model, then you get a more restrictive zone.

GIOVANNI GRONCHI: Yes, you are completely right. We have already thought about this possibility and would like to do something like that. Our first thought is a simple one; for example, I was speaking about the inclination which is quite indicative. But if we have an asteroid population model, it would be better, because we have in principle a probability distribution in this plane.

MIKHAIL MAROV: Last March the astronomical community dealing with the minor bodies was very much disturbed with an unexpected appearance of one of the quite small asteroids – it looked like it was going to impact the Earth. Does your procedure allow for an accurate prediction of NEO bodies, in a quite limited time frame?

GIOVANNI GRONCHI: Yes. We are working on adding other constraints to the admissible region that give a lower boundary for the impact time. We hope in the future to have some methods to be able to discard the possibility of an NEO impact at once.

SIMON MITTON: Following the last question, I want to make a public-understanding-of-science point, and in doing so I’m not criticising the work of you and your colleagues. But I think that where predictions are based on very short arcs, and they appear to show near misses with the Earth in 100, 200 or 300 years time – and we have examples of this – it’s really not very helpful if this information is placed into the public domain in a very premature way which then a few days later has to be withdrawn on the basis that we have some more observations. It’s quite important if a very short arc observation shows up a near-Earth object possibly impacting the Earth, the professional community should not behave in a way that gets the general public too excited and makes it front
page news. Because it actually harms our profession if 10 days later all of this has to be withdrawn.

**GIOVANNI GRONCHI:** Yes, I agree completely. It’s difficult to explain this to the general public because of the concept of probability. Most of these possible collisional impacts, luckily for us, disappear in a few days because of additional observations. So, yes, it should be confined to a restricted community.