ORBIT DETERMINATION WITH VERY SHORT ARCS. I ADMISSIBLE REGIONS

ANDREA MILANI[1], GIOVANNI F. GRONCHI[1], MATTIA DE' MICHIELI VITTURI[1] and ZORAN KNEŽEVIĆ[2]

[1]*Department of Mathematics, University of Pisa, via Buonarroti 2, 56127 Pisa, Italy, e-mail: {milani; gronchi}@dm.unipi.it, demichie@mail.dm.unipi.it*

[2]*Astronomical Observatory, Volgina 7, 11160 Belgrade 74, Serbia and Montenegro, e-mail: zoran@aob.bg.ac.yu*

(Received: 24 November 2003; revised: 31 March 2004; accepted: 16 April 2004)

Abstract. Most asteroid discoveries consist of a few astrometric observations over a short time span, and in many cases the amount of information is too limited to compute a full orbit according to the least squares principle. We investigate whether such a Very Short Arc may nonetheless contain significant orbit information, with predictive value, e.g., allowing to compute useful ephemerides with a well defined uncertainty for some time in the future.

For short enough arcs, all the significant information is contained in an *attributable*, consisting of two angles and two angular velocities for a given time; an apparent magnitude is also often available. In this case, no information on the geocentric range r and range-rate $\dot{r}$ is available from the observations themselves. However, the values of $(r, \dot{r})$ are constrained to a compact subset, the *admissible region*, if we can assume that the discovered object belongs to the Solar System, is not a satellite of the Earth and is not a shooting star (very small and very close). We give a full algebraic description of the admissible region, including geometric properties like the presence of either one or two connected components.

The admissible region can be sampled by selecting a finite number of points in the $(r, \dot{r})$ plane, each corresponding to a full set of six initial conditions (given the four component attributable) for the asteroid orbit. Because the admissible region is a region in the plane, it can be described by a triangulation with the selected points as nodes. We show that triangulations with optimal properties, such as the *Delaunay triangulations*, can be generated by an effective algorithm; however, the optimal triangulation depends upon the choice of a metric in the $(r, \dot{r})$ plane.

Each node of the triangulation is a *Virtual Asteroid*, for which it is possible to propagate the orbit and predict ephemerides. Thus for each time there is an image triangulation on the celestial sphere, and it can be used in a way similar to the use of the nominal ephemerides (with their confidence regions) in the classical case of a full least square orbit.

Key words: asteroid recovery, Delaunay triangulation, ephemerides, orbit determination

1. Introduction

In the last few years there has been an enormous increase in the rate of asteroid discoveries. Most of this progress is due to the automated CCD

Celestial Mechanics and Dynamical Astronomy **90:** 59–87, 2004.

surveys, such as Spacewatch, LINEAR, LONEOS, Catalina, NEAT. The modes of operations of the automated surveys, although they may differ in some details, are essentially the same. A number N of digital images of the same area on the celestial sphere is taken within a short time span, typically within a single night.[1] Then the images are digitally blinked, that is a computer program is run on this set of frames to identify all changes among them. If an object is found to move along a straight line, with uniform velocity, in all N frames, then it should be the detection of a real moving object, provided the signal to noise ratio is large enough to make unlikely the presence of exactly aligned spurious signals. If the image is found in less than N frames it still can be a real object with marginal signal to noise, it could have been covered by a star image in some of the frames, but it could also be a spurious detection. Typically $3 \leqslant N \leqslant 5$, and 2 hours is the time span between the first and the last observation. Such a detection is reported to the Minor Planet Center (MPC) as a sequence of N observations; we shall call such a sequence a Very Short Arc.

This operation mode is optimal for detecting moving objects of asteroidal and cometary nature.[2] Unfortunately, it is not at all optimal for determining the orbit of the detected object: these arcs are, in most cases, *too short* for a full orbit determination. When this is the case, we call the set of observations a *Too Short Arc* (*TSA*). As it is well known from the theory of preliminary orbit determination (Gauss, 1809; Danby, 1989), when three observations are used to compute an orbit, the curvature of the arc appears as a divisor in the orbit solution of Gauss' method. The smaller is the curvature, the less accurate is the orbit; taking into account the observational errors, in most cases it turns out to be impossible to apply the usual computational algorithm, consisting of a preliminary orbit determination by means of Gauss' method followed by a least squares fit (differential corrections). When starting from a TSA, either Gauss' method fails, or the differential correction procedure does not converge. On the other hand, if the survey were to use longer intervals among the individual frames, the curvature of the observed arc would be significant, and this would enormously complicate the algorithms to detect from one frame to another the moving images of the same object.

For this reason the TSAs are not considered *discoveries*, but just *detections*; this does not indicate that the observed object is fictitious, but just that its nature cannot be determined. Indeed without an orbit it is not possible to discriminate among different classes of objects, it is not possible to predict

[1] This is why these short sequences of observations are called *One Night Stand* (*ONS*).

[2] For transneptunian objects, with a much smaller proper motion, longer intervals of time among the frames may be necessary to guarantee that the angular size of the observed arc corresponds to a large enough number of pixels.

ephemerides allowing for follow up and it is seldom possible to find an identification with a known object with a reliable orbit. This has created a complex tangle with discovery rights, accessibility of data, monopoly of processing of non-public data, disagreement on the significance and value of the ONS as the topics for hot and not always scientific discussions. We will not enter into these discussions at all, but we want to find a positive scientific solution to the problems created by the existence of large databases of TSAs.

Our research plan consists of several steps, of which only the first one is complete and is presented in this paper; the basic idea is the following. A TSA is recorded as a set of N observations, which means that a set of points on a straight line is what is actually detected, with deviations from alignment compatible with the random observational error. Thus from the TSA we can compute the straight line, either by linear regression or by other fitting procedure. Then a TSA is represented by an *attributable*,[3] consisting of a reference time (just the mean of the observing times), two average angular coordinates and two corresponding angular rates for the reference time. An attributable provides no information on the range (the radial distance) and range rate at the reference epoch.

Our goal is to prove that attributables, and therefore TSAs, contain useful information. They allow to extract information on the orbit of the object being detected, as discussed in this paper, Sections 2 and 3; in fact, the range and range rate are constrained if we assume that the object belongs to the solar system, but not to the Earth–Moon system. The *confidence region*, as defined in conventional orbit determination, is replaced by an *admissible region*. An example of such region is shown in Section 4. How the admissible regions can be efficiently sampled by Virtual Asteroids (VAs) is discussed in Section 5; the VAs are not just a set of isolated points, but have a two dimensional structure. The VAs allow to predict ephemerides in a generalized sense, as discussed in Section 6.

The procedure to compute attributables may provide curvature information, which can be used to decide which paradigm of orbit determination should be used. Attributables can be used in identifications, not only in the attribution case, but also to link together two TSAs with a preliminary orbit. They can be used to detect Virtual Impactors (VIs), that is low probability future collisions with our planet compatible with the TSA information. All this is, however, the subject of ongoing research and will be reported in future papers.

Please note that we are not defining rigorously what a TSA is; that is, we are not giving an upper bound on the number of observations and on the

[3] The name attributable was introduced by Milani et al. (2001) with the intention of using the same definition as a step for finding identifications with asteroids with known orbits; an identification of this kind is called an *attribution*.

length of the observed arc for a set of observations to be considered a TSA. The reason is that only experience can tell us if the methods we are now developing for one night arcs can be useful also for longer arcs; we suspect that many two-night arcs could be conveniently processed with the method we are developing for TSAs. Operationally, the definition is as follows: a TSA is a set of observations for which the conventional orbit determination process either fails, or does not provide useful information (i.e., the confidence region is too large for practical use in whatever prediction).

We need to comment on the relationship between our work and results already present in the literature. Virtanen et al. (2001) have introduced the method of statistical ranging in a similar context, but there are significant differences with respect to our approach. Instead of assuming the observation of two angles and two angular rates at the same time, they assume the observation of two angles at each one of two different epochs. Thus their space of unknowns consists of the ranges at the two epochs, instead of the range and range rate at the same epoch. Orbit determination can be performed all the same, by solving a Lambert problem (see Danby, 1989, Chapter 6). The disadvantage of their approach is that whenever two observations are selected among the available ones they are both affected by the random observational error. To take this error into account, they need to sample the probability space of the observational error with a Monte Carlo type method. An attributable is the result of a least squares fit to a line, thus part of the accidental random error is already removed whenever there are more than two observations. Moreover, Virtanen et al. sample at random two ranges space to identify the admissible region without exploiting any a priori geometric information on this region.

Tholen and Whiteley (submitted for publication) use a method in which the space of the unknowns is explored with a regular grid. Although this is more efficient than random sampling, it is less efficient than a sampling adapted to the shape and geometric properties of the admissible region. Anyway Virtanen et al. (2001) get to the main conclusion, on which we agree, that ephemerides prediction is often possible, with an accuracy compatible with, e.g., recovery planning, even when the conventional orbit determination is impossible. In conclusion, we owe to these authors important insights which have stimulated our research, but we are following a different approach.

2. The Admissible Region

We assume that at time t an asteroid $\mathcal{A}$ with heliocentric position P is observed from the Earth, which is at the same time in $P_{\oplus}$.

Let $(r, \epsilon, \theta) \in \mathbb{R}^+ \times [-\pi, \pi) \times (-\pi/2, \pi/2)$ be spherical coordinates for the geocentric position $P - P_\oplus$.

DEFINITION 1. We shall call *attributable* a vector $A = (\epsilon, \theta, \dot{\epsilon}, \dot{\theta}) \in [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}^2$, observed at a time t.

The reference system defining angles (ϵ, θ) can be selected as necessary. We almost always use an equatorial reference system (e.g., J2000), that is we use the right ascension α for ϵ and the declination δ for θ, but of course we could use an ecliptic system without changing the equations in this paper.

Usually (although not always) the attributable also contains an average apparent magnitude h if there is at least one measurement of the apparent magnitude available.

Note that the geocentric distance r (the range) and the range rate $\dot{r}$ are left undetermined by the attributable.

The purpose of this section is to find conditions on $r, \dot{r}$ under the hypothesis that the object $\mathcal{A}$ belongs to the solar system, but not to the Earth–Moon system. We use the following quantities:

Heliocentric two-body energy

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

where $k = 0.01720209895$ is Gauss' constant.

Geocentric two-body energy

$$\mathcal{E}_\oplus(r, \dot{r}) = \frac{1}{2}\|\dot{P} - \dot{P}_\oplus\|^2 - k^2 \mu_\oplus \frac{1}{\|P - P_\oplus\|}, \tag{2}$$

where $\mu_\oplus$ is the ratio between the mass of the Earth and the mass of the Sun.

Radius of the sphere of influence of the Earth

$$R_{\mathrm{SI}} = a_\oplus \left(\frac{\mu_\oplus}{3}\right)^{1/3} = 0.010044\,\mathrm{AU},$$

that is the distance from the Earth to the collinear Lagrangian point L_2, apart from terms of order $\mu_\oplus^{2/3}$. Here $a_\oplus$ is the semimajor axis of the orbit of the Earth.

Physical radius of the Earth

$$R_\oplus \simeq 4.2 \times 10^{-5}\,\mathrm{AU}$$

Note that we are using a system of units with 1 AU as unit of length and 1 ephemeris day as unit of time; we do not need to specify the unit of mass as $\mathcal{E}_\odot(r, \dot{r})$ and $\mathcal{E}_\oplus(r, \dot{r})$ are the energies per unit mass of the asteroid.

Given an attributable A, the following four conditions have obvious physical interpretation:

(A) $\mathcal{D}_1 = \{(r, \dot{r}) : \mathcal{E}_\oplus \geqslant 0\}$ ($\mathcal{A}$ is not a satellite of the Earth) ;
(B) $\mathcal{D}_2 = \{(r, \dot{r}) : r \geqslant R_{SI}\}$ (the orbit of $\mathcal{A}$ is not controlled by the Earth) ;
(C) $\mathcal{D}_3 = \{(r, \dot{r}) : \mathcal{E}_\odot \leqslant 0\}$ ($\mathcal{A}$ belongs to the Solar System) ;
(D) $\mathcal{D}_4 = \{(r, \dot{r}) : r \geqslant R_\oplus\}$ ($\mathcal{A}$ is outside the Earth).

DEFINITION 2. Given an attributable A, we define as *admissible region* the domain

$$\mathcal{D} = \{\mathcal{D}_1 \cup \mathcal{D}_2\} \cap \mathcal{D}_3 \cap \mathcal{D}_4.$$

Note that in setting the conditions (A)–(D) we have introduced the following assumptions:

1. The observer is assumed to be at the geocenter. This approximation could be removed by replacing $P_\oplus, \dot{P}_\oplus$ with the heliocentric position and velocity of the observer (see Section 3.3), but then condition (D) should be modified.
2. The orbits of asteroids passing close to the Earth are affected by both the attraction of the Sun and that of the Earth; taking into account a complete three-body model would be very complicated. Thus conditions (A) and (B) are approximate, and indeed there are objects in heliocentric orbit experiencing temporary capture as satellites of the Earth, with $\mathcal{E}_\oplus < 0$. However, this can happen only for very low relative velocities $\|\dot{P} - \dot{P}_\oplus\|$, and the objects found in these conditions are often artificial, such as the upper stages of interplanetary launch rockets (e.g. J002E3[4]).
3. When the object is much farther away from the Earth than the Moon, that is $r >> 60\ R_\oplus$, we should use for $\mu_\oplus$ the ratio between the mass of the Earth–Moon system and the mass of the Sun.
4. In computing the radius of the sphere of influence we are neglecting the eccentricity of the orbit of the Earth.

In spite of all these limitations, the conditions defining the admissible region are a good approximation, and to find analytical formulae based on definition 2 is a good starting point for further more accurate analysis.

2.1. EXCLUDING SATELLITES OF THE EARTH

We look for a simple analytical and geometric description of the region satisfying condition (A).

[4] http://planetary.org/html/news/articlearchive/headlines/2003/apollo12- debris.html.

Spherical coordinates. The heliocentric position of $\mathcal{A}$ is given by

$$P = P_\oplus + r\,\widehat{R},$$

where $\widehat{R}$ is the unit vector in the observation direction. Using spherical coordinates, the heliocentric velocity $\dot{P}$ of $\mathcal{A}$ is

$$\dot{P} = \dot{P}_\oplus + \dot{r}\,\widehat{R} + r\,\dot{\epsilon}\,\widehat{R}_\epsilon + r\,\dot{\theta}\,\widehat{R}_\theta,$$

where $\dot{P}_\oplus$ is the heliocentric velocity of the Earth,

$$\widehat{R}_\epsilon = \frac{\partial \widehat{R}}{\partial \epsilon}; \quad \widehat{R}_\theta = \frac{\partial \widehat{R}}{\partial \theta}.$$

Explicitly in coordinates

$$\widehat{R} = (\cos\epsilon\cos\theta, \sin\epsilon\cos\theta, \sin\theta);$$

$$\widehat{R}_\epsilon = (-\sin\epsilon\cos\theta, \cos\epsilon\cos\theta, 0);$$

$$\widehat{R}_\theta = (-\cos\epsilon\sin\theta, -\sin\epsilon\sin\theta, \cos\theta).$$

Furthermore we have

$$\langle \widehat{R}, \widehat{R}_\epsilon \rangle = \langle \widehat{R}, \widehat{R}_\theta \rangle = \langle \hat{r}_\epsilon, \widehat{R}_\theta \rangle = 0,$$

that is, the vectors $\widehat{R}$, $\widehat{R}_\epsilon$, $\widehat{R}_\theta$ define an orthogonal basis for $\mathbb{R}^3$. Note that $\|\widehat{R}\| = \|\widehat{R}_\theta\| = 1$ but $\|\widehat{R}_\epsilon\| = \cos\theta$, so that this basis is not orthonormal.

Geocentric energy. We shall use this formalism to compute the orbital energies. For the geocentric energy (Equation (2)) we have

$$\|P - P_\oplus\|^2 = r^2\,;$$

$$\|\dot{P} - \dot{P}_\oplus\|^2 = \dot{r}^2 + r^2\dot{\epsilon}^2\cos^2\theta + r^2\dot{\theta}^2 = \dot{r}^2 + r^2\,\eta^2,$$

where

$$\eta = \sqrt{\dot{\epsilon}^2\cos^2\theta + \dot{\theta}^2}$$

is the proper motion. Condition (A) becomes

$$2\mathcal{E}_\oplus(r, \dot{r}) = \dot{r}^2 + r^2\,\eta^2 - 2k^2\mu_\oplus\frac{1}{r} \geqslant 0,$$

that is

$$\dot{r}^2 \geqslant \frac{2k^2\mu_\oplus}{r} - \eta^2 r^2 := G(r),$$

where $G(r) > 0$ for

$$0 < r < r_0 = \sqrt[3]{\frac{2k^2\mu_\oplus}{\eta^2}}.$$

With regard to condition (B), if $r_0 \leqslant R_{\mathrm{SI}}$, the admissible region is defined by $\dot{r}^2 \geqslant G(r)$; this occurs for

$$r_0^3 = \frac{2k^2\mu_\oplus}{\eta^2} \leqslant R_{SI}^3 = a_\oplus^3 \frac{\mu_\oplus}{3}$$

and, taking into account Kepler's third law $a_\oplus^3 n_\oplus^2 = k^2$ ($n_\oplus$ is the mean motion of the Earth), we have

$$r_0 \leqslant R_{SI} \quad \Leftrightarrow \quad \eta \geqslant \sqrt{6}\, n_\oplus.$$

Otherwise, if $r_0 > R_{SI}$, the boundary of the region given by conditions (A), (B) is formed by a segment of the straight line $r = R_{SI}$ and two arcs of the $\dot{r}^2 = G(r)$ curve for $0 < r < R_{SI}$.

2.2. EXCLUDING INTERSTELLAR ORBITS

We look for the analytical and geometric description of the region satisfying condition (C), in particular we would like to know if it is a connected region. We will show it actually can have either one or two connected components.

Heliocentric energy. For the heliocentric energy (Equation (1)) we use the heliocentric position and velocity in spherical coordinates:

$$\|P\|^2 = r^2 + 2r\langle P_\oplus, \widehat{R}\rangle + \|P_\oplus\|^2;$$
$$\|\dot{P}\|^2 = \dot{r}^2 + 2\dot{r}\langle \dot{P}_\oplus, \widehat{R}\rangle + r^2\left(\dot{\epsilon}^2\cos^2\theta + \dot{\theta}^2\right)$$
$$+ 2r(\dot{\epsilon}\langle \dot{P}_\oplus, \widehat{R}_\epsilon\rangle + \dot{\theta}\langle \dot{P}\oplus, \widehat{R}_\theta\rangle) + \|\dot{P}_\oplus\|^2.$$

We introduce the notation:

$$\begin{aligned} &c_0 = \|P_\oplus\|^2; \qquad c_3 = \dot{\epsilon}\, c_{3,1} + \dot{\theta}\, c_{3,2};\\ &c_1 = 2\langle \dot{P}_\oplus, \widehat{R}\rangle; \quad c_4 = \|\dot{P}_\oplus\|^2;\\ &c_2 = \dot{\epsilon}^2\cos^2\theta + \dot{\theta}^2 = \eta^2; \quad c_5 = 2\langle P_\oplus, \widehat{R}\rangle, \end{aligned} \tag{3}$$

where

$$c_{3,1} = 2\langle \dot{P}_\oplus, \widehat{R}_\epsilon\rangle;$$
$$c_{3,2} = 2\langle \dot{P}_\oplus, \widehat{R}_\theta\rangle;$$

so that

$$\|P\|^2 = r^2 + c_5 r + c_0 := S(r);$$
$$\|\dot{P}\|^2 = 2\mathcal{T}_\odot(r, \dot{r}) = \dot{r}^2 + c_1\dot{r} + W(r);$$
$$W(r) := c_2 r^2 + c_3 r + c_4.$$

By substituting in Equation (1), condition (C) reads

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

Reality condition for the range rate. In order to have real solutions for $\dot{r}$, the discriminant of $\mathcal{E}_\odot$, regarded as a degree 2 polynomial in $\dot{r}$, must be non-negative:

$$\Delta_\odot(r) := \frac{c_1^2}{4} - W(r) + \frac{2k^2}{\sqrt{S(r)}} \geqslant 0.$$

Let us set $\gamma = c_4 - c_1^2/4$ (note that $\gamma \geqslant 0$), and define

$$P(r) := c_2 r^2 + c_3 r + \gamma;$$

then the energy condition (C) implies the following condition on r:

$$\frac{2k^2}{\sqrt{S(r)}} \geqslant P(r). \tag{4}$$

The degree 2 polynomial $P(r)$ is non-negative for each r: it is the opposite of the discriminant of the degree 2 polynomial $\mathcal{T}_\odot(r,\dot{r})$ (regarded as a function of $\dot{r}$). $\mathcal{T}_\odot$ is a kinetic energy and is non-negative, thus its discriminant is non-positive. Also $S(r)$ is non-negative, thus we can square the left and right hand side of (4) and obtain an inequality involving a polynomial of degree 6, namely

$$V(r) := P^2(r)S(r) = \sum_{i=0}^{6} A_i r^i, \tag{5}$$

with coefficients

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

After squaring, condition (4) becomes $V(r) \leqslant 4k^4$.

Connected components of $\mathcal{D}_3$. The main result of this section is the following

THEOREM 1. *The region $\mathcal{D}_3$, defined by condition* (*C*), *has at most two connected components.*

Proof. To prove the theorem we need the two following lemmas:

LEMMA 1 (Existence of solutions). *If either η or γ is positive,*[5] *then there are at least two solutions of*

[5] In the completely degenerate case $\eta = \gamma = 0$ there are solutions for $\dot{r}$ for all values of r. This is a very strange situation, with both $\hat{r}$ and the velocity of the asteroid parallel to the velocity of the Earth.

$$V(r) - 4k^4 = 0, \tag{6}$$

one positive and the other negative.

Proof. The value of the polynomial at the origin is

$$\begin{aligned} V(0) - 4k^4 &= c_0\, \gamma^2 - 4k^4 \leqslant c_0\, c_4^2 - 4k^4 \\ &= \|P_\oplus\|^2\, \|\dot{P}_\oplus\|^4 - 4k^4 \end{aligned}$$

which is always strictly negative because the heliocentric orbital energy of the Earth

$$\frac{1}{2}\|\dot{P}_\oplus\|^2 - \frac{k^2}{\|P_\oplus\|}$$

is strictly negative.[6] The leading coefficient $A_6 (= \eta^4)$ is non-negative. If it is not zero, we have $\lim_{r\to\pm\infty} V(r) = +\infty$ and we have at least one positive and one negative real root.

If A_6 vanishes, that is if $c_2 = 0$ (thus $\eta = \dot{\epsilon} = \dot{\theta} = c_3 = 0$) and $\gamma > 0$, then $V(r) - 4k^4$ is a degree 2 polynomial with two real roots, one positive and one negative. □

LEMMA 2. *The equation* $V'(r) = 0$ *cannot have more than three distinct solutions. If it has exactly three distinct real roots, then there cannot be any root with multiplicity* 2.

Proof. The derivative

$$V'(r) = P(r)[2P'(r)S(r) + P(r)S'(r)].$$

is a degree 5 polynomial.

Note that $P(r) \geqslant 0$ for all r, thus it has either no real roots or just one double real root.

In the first case, if $P(r) > 0$ then $V'(r)/P(r)$, being a degree 3 polynomial, has at most three real roots, and so does $V'(r)$.

In the second case, if $P(r)$ has a double root, then this is a root with multiplicity three of $V'(r)$, and there are at most two other roots.

If $V'(r)$ has three distinct real roots including a multiple one, then there is one root $\bar{r}$ of $P(r)$ which must be at least a triple root of $V'(r)$; the other two have to be simple. □

To conclude the proof of the theorem we observe that:

(i) by Rolle's theorem, between two roots of Equation (6) there must be a root of $V'(r)$;

[6] The topocentric correction would not be enough to make the heliocentric energy of the observer positive.

(ii) there cannot be an odd number of real roots of Equation (6) (counted with their multiplicity), as $V(r)$ is a *real* polynomial with even degree;
(iii) at least two real roots of Equation (6) have odd multiplicity.

Using Lemma 1, Lemma 2 and the remarks above we are left only with the following possibilities:

(i) four distinct and simple roots (see Figure 2);
(ii) three distinct roots, two simple and one with even multiplicity (the component at large r reduces to a point);
(iii) two distinct roots, one simple and the other one with odd multiplicity (the admissible region has only one component, see Figure 1). □

From the above arguments the theorem immediately follows.

In Figure 2 we show an example in which the region $\mathcal{D}_3$ has two connected components: we have used for the attributable the value $(\epsilon, \theta, \dot{\epsilon}, \dot{\theta}) = (0, 0, -0.09, 0.01)$, with ϵ, θ in degrees, $\dot{\epsilon}, \dot{\theta}$ in degrees per day. We have plotted also the level curves for small positive and small negative values of $\mathcal{E}_{\odot}$, showing the qualitative change.

2.3. CLOSE APPROACHES

To understand the global structure of the admissible region $\mathcal{D}$ we need to find possible intersections between the $\mathcal{E}_{\oplus} = 0$ curve and the $\mathcal{E}_{\odot} = 0$ curve. However, these intersections are physically meaningful only if they occur for $R_{\oplus} < r < R_{\mathrm{SI}}$, that is, during a close approach to the Earth, but above its physical surface. The following result indicates that these intersections occur

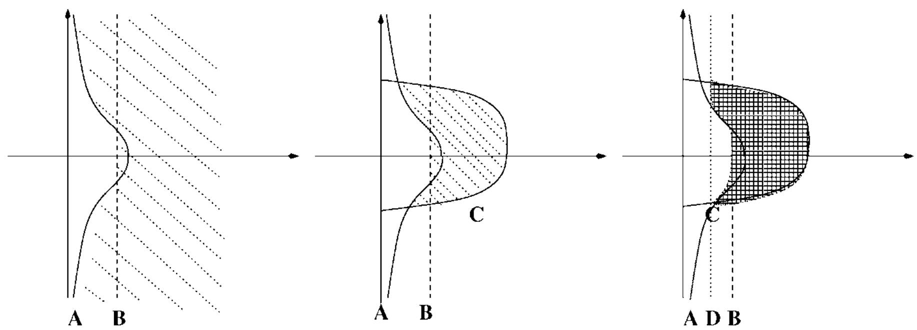

Figure 1. The qualitative features of the admissible region: if condition (A) or (B) are satisfied, we obtain the domain $\mathcal{D}_1 \cap \mathcal{D}_2$ drawn on the left. If we set also condition (C), we are left with the domain sketched in the middle plot. Adding condition (D), we end up with the admissible region $\mathcal{D}$, sketched in the plot on the right. We stress that this figure is only qualitative and that it refers to a case with only one connected component (see Section 2.2).

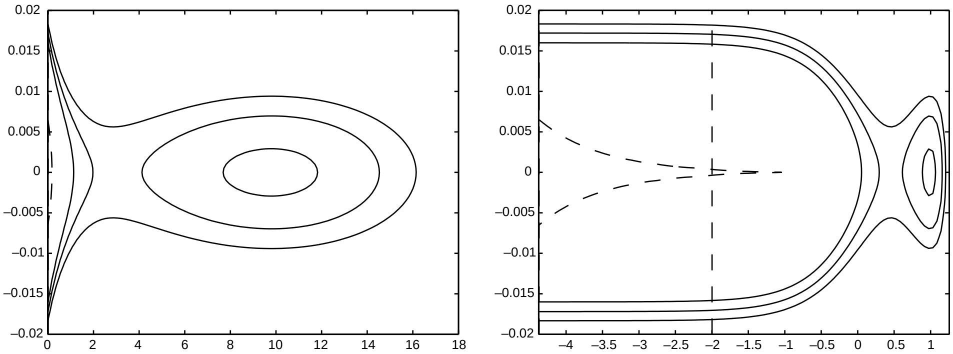

Figure 2. We show an example with two connected components. On the left we plot three level curves of $\mathcal{E}_\odot$, including the zero level curve, and $\mathcal{E}_\oplus = 0$ (dashed curve) in the plane $(r, \dot r)$; on the right we draw the same plot in the plane $(\log_{10}(r), \dot r)$. The dashed line denotes R_{SI}.

only where it does not matter; it also implies that the admissible region does not have more connected components than the region satisfying condition (C).

THEOREM 2. *For* $R_\oplus \leqslant r \leqslant R_{SI}$ *the condition* $\mathcal{E}_\oplus(r, \dot r) \leqslant 0$ *implies* $\mathcal{E}_\odot(r, \dot r) \leqslant 0$.

Proof. By the triangular inequality, to prove that $\mathcal{E}_\odot(r, \dot r) \leqslant 0$ it is enough to prove that

$$\left(\|\dot P - \dot P_\oplus\| + \|\dot P_\oplus\|\right)^2 \leqslant \frac{2k^2}{\|P - P_\oplus\| + \|P_\oplus\|}.$$

We observe that

$$\mathcal{E}_\oplus(r, \dot r) \leqslant 0 \quad \Leftrightarrow \quad \|\dot P - \dot P_\oplus\| \leqslant \sqrt{\frac{2k^2\mu_\oplus}{\|P - P_\oplus\|}}. \tag{7}$$

Using relation (7) we thus only have to prove that

$$\frac{2k^2\mu_\oplus}{r} + \|\dot P_\oplus\|^2 + 2\|\dot P_\oplus\|\sqrt{\frac{2k^2\mu_\oplus}{r}} \leqslant \frac{2k^2}{r + \|P_\oplus\|}$$

for $R_\oplus \leqslant r \leqslant R_{SI}$. This is equivalent to prove that in this interval the function

$$\begin{aligned} F(r) =& 2k^2\mu_\oplus(r + \|P_\oplus\|) + r(r + \|P_\oplus\|)\|\dot P_\oplus\|^2 \\ &+ 2\|\dot P_\oplus\|\sqrt{2k^2\mu_\oplus}\sqrt{r}(r + \|P_\oplus\|) - 2k^2 r \end{aligned}$$

is negative.

To describe the qualitative features of $F(r)$ we start by decomposing the derivative $F'(r)$ as follows:

$$F'(r) = \frac{g_1(r) + g_2(r)}{\sqrt{r}},$$

where

$$g_1(r) = \sqrt{r}[\mathcal{C} + 2\|\dot{P}_\oplus\|^2 r];$$

$$g_2(r) = \|\dot{P}_\oplus\|\sqrt{2k^2\mu_\oplus}[3r + \|P_\oplus\|];$$

with

$$\mathcal{C} := 2k^2\mu_\oplus + \|P_\oplus\|\|\dot{P}_\oplus\|^2 - 2k^2;$$

note that $\mathcal{C} \leqslant -2.8075 \times 10^{-4} < 0$.

The second derivative $g_1''(r)$ is positive for each $r > 0$ so that, for convexity reasons, the graphs of the functions g_1 and $-g_2$ can intersect at most twice for $r > 0$.

Hence $F'(r)$ has at most two zeros for $r > 0$. Taking also into account the following:

$$\lim_{r \to +\infty} F(r) = +\infty; \quad F(0) = 2k^2\mu_\oplus\|P_\oplus\| > 0; \quad \lim_{r \to 0^+} F'(r) = +\infty.$$

we conclude that $F(r)$ cannot have more than two zeros for $r > 0$.

Finally, by using the estimates $F(R_\oplus) \leqslant -2.49 \times 10^{-10} < 0$ and $F(R_{\mathrm{SI}}) \leqslant -2.6346 \times 10^{-6} < 0$, we conclude that $F(r) < 0$ for $R_\oplus \leqslant r \leqslant R_{\mathrm{SI}}$, and this completes the proof of the theorem. □

Note that Theorem 2 applies only for particular values of the mass, radius and orbital elements of the planet on which the observer is located. It is a physical property of the Earth, not a general property of whatever planet; it depends on the values of the parameters, which are detailed below. A larger planet, such as Jupiter, can have satellites whose velocity would be hyperbolic with respect to the Sun, if Jupiter was not controlling the orbit; the Earth can not have satellites with this behavior.

Numerical values used in the estimates: $\mu_\oplus = 1/328900.5614$; $R_\oplus = 4.24 \times 10^{-5}$; $k = 0.01720209895$; $a_\oplus = 1.0$; $e_\oplus = 0.0167$.

From the expressions

$$\|P_\oplus\| = a_\oplus(1 - e_\oplus \cos(u_\oplus));$$

$$\|\dot{P}_\oplus\|^2 = k^2\left[\frac{2}{r_\oplus} - \frac{1}{a_\oplus}\right];$$

it follows that

$$\max\|P_\oplus\| = 1.0167; \quad \max\|\dot{P}_\oplus\| = 0.0175.$$

We have obtained the upper bounds for $F(R_\oplus)$ and $F(R_{SI})$ by using the maximum values of the length of the heliocentric position and velocity of the Earth along an elliptic orbit with the osculating $e_\oplus$.

2.4. THE BOUNDARY OF THE ADMISSIBLE REGION

We can now give a complete description of the boundary of the admissible region $\mathcal{D}$. It consists of:

1. part of the algebraic curve $\mathcal{E}_\odot = 0$ for $r > 0$. If Equation (6) has three positive roots there is another component, consisting of a simple closed curve, at larger values of r: this includes the case when this curve reduces to a single point, if Equation (6) has a double positive root;
2. two segments of the straight line $r = R_\oplus$;
3. two portions of the curve $\dot{r}^2 = G(r)$ (corresponding to $\mathcal{E}_\oplus = 0$) and one segment of the straight line $r = R_{SI}$ if $R_{SI} < r_0$; if $R_{SI} \geqslant r_0$ the two portions of the $\dot{r}^2 = G(r)$ are joined at $r = r_0$.

Thus the admissible region consists of at most two connected components, and it is compact being the inside of a finite number of closed continuous curves.

2.5. A SIMPLIFIED CASE

Let us compute the admissible region in a simplified case, obtained assuming the Earth on a circular orbit: $e_\oplus = 0$. We also assume $a_\oplus = 1$ and $n_\oplus = k$, that is we are neglecting the terms of the order of $\mu_\oplus$ in the orbit of the Earth and all other planetary perturbations.

Let us consider, for instance, coordinates ϵ, θ such that θ is the ecliptic latitude of $\mathcal{A}$ and ϵ is the angle between the opposition and the projection of $\mathcal{A}$ onto the ecliptic plane. These coordinates are singular for $\theta = \pm\pi/2$, that is for observations at the ecliptic pole.

Within these approximations $P_\oplus = (1, 0, 0)$ and $\dot{P}_\oplus = (0, k, 0)$; thus

$$\begin{aligned} &c_0 = 1; \quad c_3 = 2k(\dot{\epsilon} \cos \epsilon \cos \theta - \dot{\theta} \sin \epsilon \sin \theta);\\ &c_1 = 2k \sin \epsilon \cos \theta; \quad c_4 = k^2; \\ &c_2 = \dot{\epsilon}^2 \cos^2 \theta + \dot{\theta}^2; \quad c_5 = 2 \cos \theta \cos \epsilon. \end{aligned} \tag{8}$$

Note that $c_3/2k$ is the time derivative of the y coordinate of $\widehat{R}$. The coefficients of the inequalities defining $\mathcal{D}_3$ are simpler than in the general case, but still too complicated to extract information on the number of connected components for all ϵ, θ; the computation becomes simple only for some special values of the angles.

Some special cases. For $\epsilon = \pm\pi/2$ (at quadrature) we have $c_5 = 0$ and $c_3 = \pm 2k\ \mathrm{d}(\cos\theta)/\mathrm{d}t$ and the coefficients of $V(r)$ are

$$A_6 = c_2^2 > 0; \quad A_2 = \gamma^2 + c_0(c_3^2 + 2c_2\gamma) > 0;$$
$$A_5 = 2c_2c_3; \quad A_1 = 2c_0c_3\gamma;$$
$$A_4 = c_3^2 + 2c_2\gamma + c_0c_2^2 > 0; \quad A_0 = c_0\gamma^2.$$
$$A_3 = 2c_3\gamma + 2c_0c_2c_3;$$

In this case $\mathrm{sgn}(A_5) = \mathrm{sgn}(A_3) = \mathrm{sgn}(A_1) = \mathrm{sgn}(c_3)$. For $c_3 \geqslant 0$ the number of variations of signs in the sequence of the coefficients of $V(r) - 4k^4$ is only one and we have only one positive root, that is only one connected component. Note that for $\theta = 0$, at the exact quadratures, we have $c_3 = 0$, and also in this case there is only one connected component.

For $\epsilon = 0$ (at opposition) $c_5 = 2\cos\theta > 0$ and $c_3 = 2k\dot{\epsilon}\cos\theta$. If $c_3 \geqslant 0$, that is $\dot{\epsilon} \geqslant 0$ (non-retrograde proper motion), again the coefficients $A_i, i = 1\ldots6$, are positive and there can be only one positive root of $V(r) = 4k^4$. However, unlike the case of the quadratures, at the exact opposition ($\epsilon = \theta = 0$) there could be two connected components, provided the motion is retrograde.

If the computation is performed with a different coordinate system $(r, \tilde{\epsilon}, \tilde{\theta})$ non-singular at the ecliptic poles, as an example such that

$$\widehat{R} = (\sin\tilde{\theta}, \cos\tilde{\epsilon}\cos\tilde{\theta}, \sin\tilde{\epsilon}\cos\tilde{\theta})$$

then in the plane $\tilde{\theta} = 0$ (orthogonal to the direction of the Sun) $c_5 = 0$ and $c_3 = -2k\dot{\tilde{\epsilon}}\sin\tilde{\epsilon} = -2k\mathrm{d}(\cos\tilde{\epsilon})/\mathrm{d}t$. By the same argument used above, for $c_3 \geqslant 0$ there can be only one connected component of the admissible region, but this is not guaranteed for $c_3 < 0$.

3. The Inner Boundary

The results of Section 2 provide a complete topological and analytical description of the admissible region. From the metric point of view, the results are not completely satisfactory, because the near edge is too close to the observer. Actually, if the topocentric correction is taken into account, the constraint (D): $r \geqslant R_\oplus$ has no meaning and the admissible region always extends down to an arbitrarily small distance from the observer. This is an unpleasant, but rather intuitive result: an object very close and heading directly toward the observer can have an arbitrary proper motion, including a very small one.

Thus we are led to consider other ways to constrain from below the distance from the observer. We are discussing in this section two conditions defining the inner boundary, which could be used as a replacement for condition (D).

3.1. SHOOTING STARS

An alternative condition giving a lower limit to the distance is that the object is not a shooting star (very small and very close). We can assume that the size is controlled by setting a maximum for the absolute magnitude H:

(E) $\mathcal{D}_5 = \{(r, \dot{r}) : H(r) \leqslant H_{\max}\}$.

If some value of the apparent magnitude is available, then the absolute magnitude H can be computed from h, the average of the measured apparent magnitudes:

$$H = h - 5\log_{10} r - x(r), \tag{9}$$

where the correction $x(r)$ accounts for the distance from the Sun and the phase effect (see the official IAU definition of absolute magnitude, Bowell et al., 1989). However, for small r (e.g., $r < 0.01$ a.u.) the correction $x(r)$ has a negligible dependence upon r because the distance from the Sun is $\simeq 1$ AU and the phase is close to the angle between the direction $\widehat{R}$ and the opposition direction. Thus we can approximate $x(r)$ with the quantity x_0 independent of r. Moreover, we are using r, the distance at the reference time t, for all the epochs of the observations including photometry; this is a fair approximation unless the relative change of distance during the time span of the observed arc is relevant, which can happen only for very small distances. In this approximation, condition (E) becomes

$$H_{\max} \geqslant H = h - 5\ \log_{10}\ r - x_0$$

or, equivalently

$$\log_{10} r \geqslant \frac{h - H_{\max} - x_0}{5} := \log_{10}\ r_{\mathrm{H}},$$

that is, given the apparent magnitude h, there is a minimum distance r_H for the object to be of significant size. If we use $H_{\max} = 30$ (a few meters diameter) then, for example, at opposition (where $x_0 = 0$):

$$h = 20 \quad \Rightarrow \quad r \geqslant 0.01\,\mathrm{AU}; \qquad h = 15 \quad \Rightarrow \quad r \geqslant 0.001\,\mathrm{AU}.$$

In any case, the absolute magnitude of the object is not a function of $\dot{r}$ and the region satisfying condition (E) is just a half plane $r \geqslant r_H$. Provided $r_H \geqslant R_\oplus$ (for $H_{\max} = 30$ this occurs for $h \geqslant 8.1$) it is possible to use the same arguments of Section 2.3 to show that the geometry of the admissible region does not become more complicated when condition (D) is replaced by condition (E). On the contrary it is quite possible that this geometry becomes simpler. If $h > 20$ the entire sphere of influence of the Earth is excluded by condition (E), thus conditions (B) is implied by (E) and condition (A) becomes irrelevant.

3.2. IMMEDIATE IMPACTORS AND JUST LAUNCHED BODIES

We may exclude from the admissible region $\mathcal{D}$ the values of $(r, \dot{r})$ such that an object with proper motion η cannot avoid collision with the Earth within a short time from the epoch of the attributable. In the same way we may exclude the values of $(r, \dot{r})$ leading to a contact with the Earth in the immediate past, that is we are not considering objects just launched from our planet. We assume that the motion is rectilinear, a valid approximation provided the time is short enough. We also neglect the topocentric correction (Figure 3).

This additional condition (F) can be expressed as

$$\frac{\eta r^2}{|\dot{r}|} \geqslant R_{\oplus}. \tag{10}$$

This condition may be included in a more restrictive definition of admissible region whenever we are not interested in a late warning system,

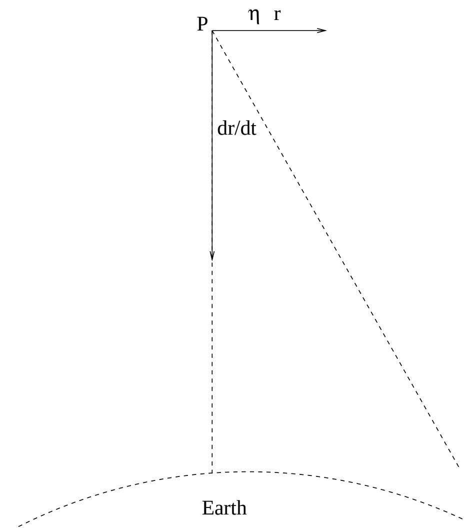

Figure 3. Immediate impact trajectory.

monitoring for possible impacts within a day or so. In other cases we may be interested also in the cases of orbits violating condition (F), at least for the immediate impact case: as an example, we may be interested in discovering fireballs before they enter the atmosphere. As it is clear by comparison with condition (E), these immediately impacting objects need to be very small, unless the detection is extremely bright. In practice condition (F) is, for most attributables, less important than condition (E), unless $H_{\max}$ is very large, that is, unless we are searching for very small meteoroids.

3.3. MODIFIED ADMISSIBLE REGION

In the following we shall use a definition of the admissible region modified as follows:

(1) condition (E) replaces condition (D);
(2) In the computation of condition (C), the topocentric correction is applied, that is $P_\oplus$ and $\dot{P}_\oplus$ actually indicate the heliocentric position and velocity of the observer;
(3) condition (C) is replaced by $\mathcal{E}_\odot \leqslant -k^2/(2a_{\max})$; this is done to exclude comets with a long period, e.g. with $a_{\max} = 100$ AU.

We are not using condition (F), and we are not applying the topocentric correction for the computation of condition (A); these additional modifications would not be effective in defining a more realistic admissible region while making its geometry much more complicated.

4. An Example

To provide an illustrative example, we have chosen the asteroid 2003 BH_{84}, discovered from the European Southern Observatory on 25 January 2003. The four observations on that night span only 1 hour and 40 minutes in time, and the proper motion was $\eta = 0.35°$ per day. We first computed the attributable by a linear fit[7] and then we computed the admissible region by applying the algorithms described in the previous sections. The maximum distance r compatible with the modified admissible region (that is, with a semimajor axis $a < 100$ AU) was 4.46 AU.

Nevertheless the information contained in the attributable, computed by using only the first night of observations, was enough to conclude that this

[7] A quadratic fit was also computed, but the curvature was found not to be significant with respect to the random observational errors.

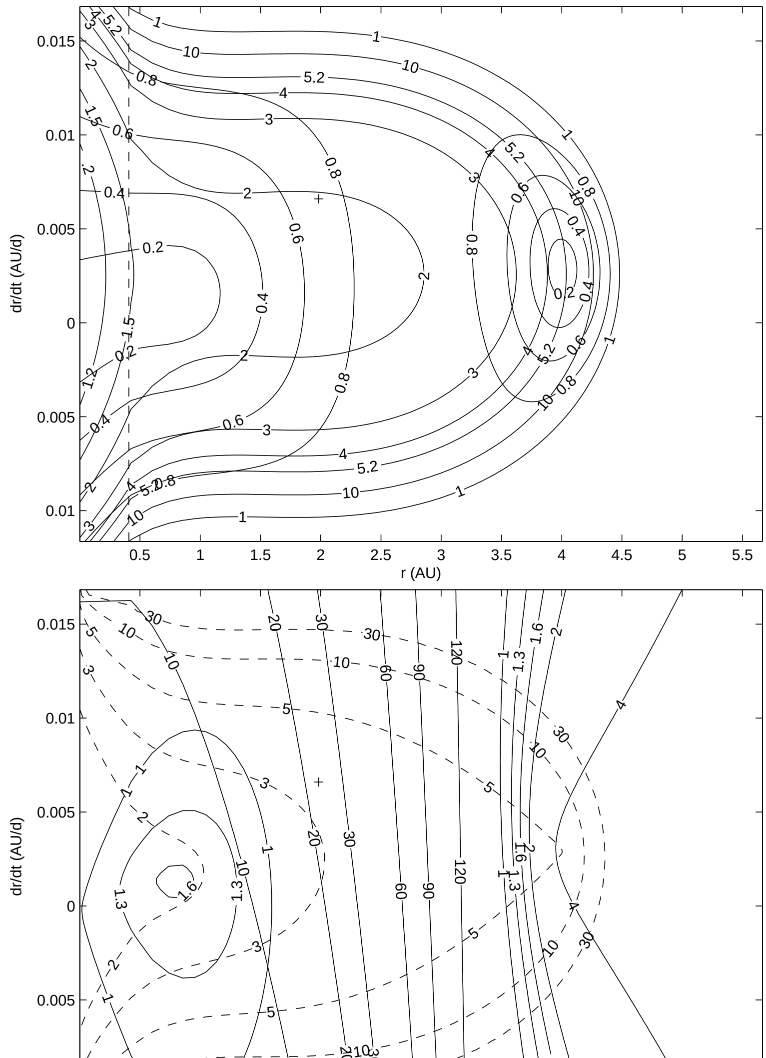

Figure 4. By using only the first night of observations of 2003 BH_{84} we have plotted in the $(r, \dot{r})$ plane: above, the level curves corresponding to $a = 1.2$, 1.5, 2, 3, 4, 5.2, 10 AU, $e = 0.2$, 0.4, 0.6, 0.8, 1 and the inner boundary corresponding to the magnitude limit $H_{\max} = 22$ (dashed vertical line); below, the level curves $q = 1$, 1.3, 1.6, 2, 4 AU (solid), $Q = 2$, 3, 5, 10, 30 AU (dashed) and $I = 10°$, 20°, 30°, 60°, 90°, 120° (solid).

object could be neither a main belt asteroid nor a Hungaria. For a fine grid of points in the $(r, \dot{r})$ plane we compute a set of orbital elements, uniquely determined by the values of $(\alpha, \delta, \dot{\alpha}, \dot{\delta}, r, \dot{r})$.[8] We then plot the contour lines of (a, e, I) on Figure 4, and find that a moderate value of e is possible for either a low value of a (e.g., an Athen type orbit) or for $I > 120$. By also inspecting the level curves of $q = a(1 - e)$, $Q = a(1 + e)$ and in the same figure we can conclude that the most likely interpretation of the first night of data is that the object is a Near Earth Asteroid (NEA), with $q < 1.3$, the other possibilities being a retrograde orbit and an object whose orbit is close to that of Mars. The crosses in the two plots mark the actual values of r, $\dot{r}$ as determined a posteriori, after the asteroid was recovered.

5. Sampling the Admissible Region

The admissible region is anyway an infinite set, thus we cannot proceed with computations (e.g., of ephemerides) for all points. We need to sample the admissible region with a finite, and not too large, subset of points. In order to sample we define an algorithm to triangulate the admissible region: the nodes of the triangulation will give us a sample, the edges joining them and the triangles provide additional structure. First we select a number of points on the boundary and produce an initial triangulation using these boundary points as nodes. Then we add nodes inside the admissible region and change the edges to achieve a triangulation with the optimal properties described below.

5.1. SAMPLING OF THE BOUNDARY

The boundary of the admissible region has an outer part, given by arcs of the curve $\mathcal{E}_\odot(r, \dot{r}) = 0$ (symmetric with respect to the line $\dot{r} = -c_1/2$); the curve $\mathcal{E}_\odot = -k^2/(2a_{\max})$ used as outer boundary of the modified admissible region is also symmetric. The boundary also has an inner part consisting of some combination of arcs of the curve $\mathcal{E}_\oplus(r, \dot{r}) = 0$ (symmetric with respect $\dot{r} = 0$) and of segments of the lines $r = r_{\mathrm{H}}, r = r_{\mathrm{SI}}$.

The symmetry with respect to the line $\dot{r} = -c_1/2$ allows us to perform the computations only for the lower region ($\dot{r} \leqslant -c_1/2$) of the exterior boundary, from $r = r_{\mathrm{H}}$ to the maximum value $r_{\max}$ such that the discriminant $\Delta_\odot(r)$ vanishes. Then we perform the sampling of the inner boundary, using the symmetry with respect to $\dot{r} = 0$ of the $\mathcal{E}_\oplus = 0$ curve.

[8] These quantities define a set of initial conditions for the asteroid orbit at epoch $t - \delta t$, where $\delta t = r/c$ is the light travel time.

The intersection points among the lines and the curves are always included in the boundary sampling, unless there are some too close: in this case we simplify the boundary by using a shortcut, at the price of including in the triangulation a small portion outside of the admissible region.

The admissible region may sometimes have two connected components when the discriminant $\Delta_{\odot}(r)$ has three positive roots. In this case we perform a separate sampling of the boundaries of the two components.

In addition we would like to select points that are *equispaced* on the boundary, that is, if the boundary is parameterized by the arc length s, then the distance of each couple of consecutive points corresponds to a fixed increment of s.

To avoid the computation of the arc length parameter we use the following idea. We choose a large number of points, equispaced in the abscissa and then we use an elimination rule to be iterated until we are left with a desired number of points. It can be shown (see the Appendix) that the remaining points are close to the ideal distribution, equispaced in arc length.

The number of points used to sample the boundary has to be limited for practical reasons, because, as will be shown in the following subsection, the final number of Virtual Asteroids is proportional to the number of boundary points, and the computational cost of whatever prediction depends upon the number of VAs.

5.2. OPTIMAL TRIANGULATION

Let us consider the domain $\widetilde{\mathcal{D}} \subset \mathcal{D}$, defined by connecting with edges the boundary points of the admissible region: we shall define a method to triangulate $\widetilde{\mathcal{D}}$. Let us start with some definitions.

A *triangulation* of the polygonal domain $\widetilde{\mathcal{D}}$ 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 vertices in Π.

We shall use T_i also for the convex hull of the points in T_i. With this notation we search for a triangulation with the following properties:

(i) $\bigcup_{i=1,k} T_i = \widetilde{\mathcal{D}}$;
(ii) for each $i \neq j$ the set $T_i \cap T_j$ is either the *empty set* 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, there is a well-known construction, called the *Delaunay triangulation* (see Bern and Eppstein, 1992), characterized by the following properties:

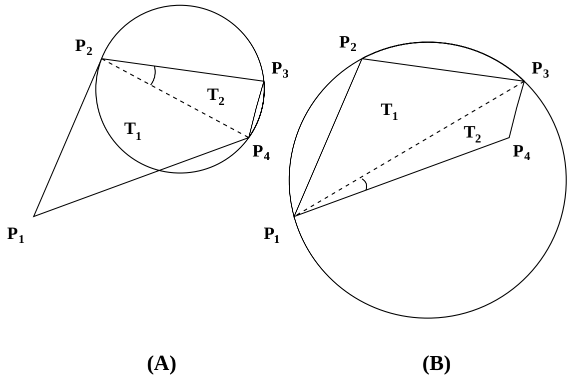

Figure 5. Possible triangulations of the quadrangle $P_1P_2P_3P_4$: the one in (A) is a Delaunay triangulation. We mark in both cases the minimum angle and we draw the circumcircles corresponding to triangle $P_2P_3P_4$ (left plot) and to triangle $P_1P_2P_3$ (right plot).

(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 (Risler, 1991).

The above properties are equivalent if the domain is convex.

If the domain is a convex quadrangle whose vertexes Π are not on the same circle, then there exist two possible triangulations (Π, τ_1), (Π, τ_2): by property (iii), only one of these is Delaunay's (see Figure 5). In this case the Delaunay triangulation can be obtained from the other one by an *edge-flipping* technique, which consists of substituting the diagonal P_1P_3 (*not-Delaunay edge*) of the quadrangle, corresponding to the common edge, with the diagonal P_2P_4 (*Delaunay edge*). Note that the edge-flipping also results in an increase of the minimum angle.

If, in addition to the set of points Π, we give as input also some edges P_iP_j, for example the boundary edges as we do for $\widetilde{\mathcal{D}}$, we refer to this input as *planar straight line graph*, and to the corresponding triangulation containing the prescribed edges as a *constrained triangulation*.

Note that the domain $\widetilde{\mathcal{D}}$ is in general not convex: in this case we can give as input the edges along the boundary. Then there still exists a constrained triangulation that maximizes the minimum angle (also minimizes the maximum circumcircles, that is with properties (i) and (ii)), called *constrained Delaunay triangulation* (Bern and Eppstein, 1992) but property (iii) is not guaranteed.

Figure 5 suggests how to transform any triangulation of $\widetilde{\mathcal{D}}$ into a constrained Delaunay: for each triangle T_i, we iterate a procedure over the adjacent triangles; if the common edge with an adjacent triangle is not a Delaunay, we apply the edge-flipping technique. Repeating this procedure until all edges of the triangulation are Delaunay's or edges of the boundary of $\widetilde{\mathcal{D}}$, at each step the minimum angle increases and at the end we obtain the triangulation that maximizes the minimum angle (Delaunay, 1934).

The procedure adopted to triangulate our domain uses as input the sampling of the boundary described in Section 5.1 and the polygon formed by these boundary points. The first phase is to generate a constrained Delaunay's triangulation (Π_0, τ_0) with these boundary points and boundary edges.

Once the initial triangulation is obtained, the second phase is to refine it by adding new points, internal to the domain, keeping at each insertion the Delaunay property. At each step we add a new point extending to the internal part of the domain the discrete density defined on the boundary points by the quantities[9]

$$\rho(P_j) = \min_{l \neq j} d(P_l - P_j),$$

where d is some distance (see Section 5.3). Let G_i be the barycenters of the triangles T_i; we define the corresponding densities

$$\tilde{\rho}(G_i) = \frac{1}{3} \sum_{m=1}^{3} \rho(P_{i_m})$$

($P_{i_m}, m = 1 \ldots 3$, belongs to the same triangle T_i) and we add as new point the barycenter $G_{\bar{k}}$ that maximizes the minimum distance (weighted with its density $\rho(G_{\bar{k}})$) from the nodes of the triangulation. Then we eliminate the corresponding triangle $T_{\bar{k}}$ and we add to τ the triangles obtained joining the edges of $T_{\bar{k}}$ with the new point (keeping at each triangle insertion the Delaunay's optimal property by means of the edge–flipping technique). We iterate this insertion procedure while

$$\max_{G_i} \left(\min_j \left\{ \frac{\mathrm{d}(G_i, P_j)}{\tilde{\rho}(G_i)} \right\} \right) > \sigma, \tag{11}$$

where σ is a fixed small parameter. In (de' Michieli Vitturi, 2004) it is shown that, if we denote with n_0 the number of points on the boundary of length $\mu(\partial\widetilde{\mathcal{D}})$, the following results holds:

THEOREM 3. *The algorithm converges and the final number of triangles is less than*

$$\frac{\mu(\partial\widetilde{\mathcal{D}})\, n_0}{\sqrt{3}\sigma}.$$

When no new points need to be added, either because some maximum number has been reached, or because the convergence criterion (11) has been reached, we begin the third phase of the procedure.

We apply to the triangulation obtained as above a *mesh improvement technique*, generalizing the *Laplacian smoothing* (see Winslow, 1964; Field, 1988). We

[9] ρ_{P_j} is indeed an approximation of the inverse of a density function.

move every internal point P_j of the triangulation to the center of mass (weighted with the density defined above) of the polygon formed by all its neighboring points (i.e. the ones connected to P_j by an edge), if it lies inside the polygon.

This technique improves the quality of the triangulation, but it can produce a triangulation that is not Delaunay's, so that we apply again the edge-flipping technique at the end of the smoothing algorithm.

The final result is a triangulation optimal from the point of view of property (i), that is avoiding as much as possible 'flattened' triangles.

5.3. SELECTION OF A METRIC

The definition of Delaunay triangulation uses distances and angles, thus it depends on the metric selected for the space $(r, \dot{r})$, in fact its own definition is based on computations of distances and angles. In particular we can select a strictly increasing function $f(r)$ and perform the triangulation of the admissible region with the metric

$$\mathrm{d}s^2 = \mathrm{d}f(r)^2 + \mathrm{d}\dot{r}^2,$$

in other words, we can work in the plane $(f(r), \dot{r})$ endowed with the Euclidean metric.

In our work we have selected an *adaptive metric*, defined by the function

$$f(r) = 1 - \exp\left(-\frac{r^2}{2\,s^2}\right), \tag{12}$$

Since $f'(r)$ is maximum at $r = s$, by choosing the parameter s we select which part of the admissible region should be more densely sampled. Currently we use $s = r_{\max}$, the largest root of the discriminant $\Delta_{\odot}(r)$, because with this choice we enhance the portion of the space $(r, \dot{r})$ close to $r = r_{\max}$. If our purpose is to search for objects in a particular portion of the $(r, \dot{r})$ space (e.g. NEAs or MBs or TNOs), then we can use a metric selected *ad hoc*, like using a smaller s to enhance the region near the Earth for NEA.

6. Triangulated Ephemerides

As first example of how to use the admissible region and its triangulation, we shall discuss the generation of ephemerides. If the orbit of an asteroid has been determined according to the least squares principle, the ephemerides at some time t_1 different from t are the predicted values of the angles α, δ with a confidence region on the celestial sphere. When the available observational data are not enough to compute a full least squares orbit (e.g., if there are only two observations, or if the observed arc is very short), the notion of ephemerides has to be redefined as the set of values for the angles α, δ at time t_1 which is compatible with the data.

If we can assume the object satisfies the conditions defining the admissible region, that is, if we can exclude interstellar orbits, satellites of the Earth and shooting stars, then there is a set of admissible values for α, δ at time t_1 which is, by continuity, a compact subset of the celestial sphere with at most two connected components. In practice, however, there is no known algorithm to compute explicitly this subset. The goal we can achieve is to sample this set by triangulation. Given a triangulation of the admissible region, computed from the attributable representing the observations available, we can compute the predicted observation at time t_1 for each node of the triangulation. Indeed, each node corresponds to a choice of the values of $(r, \dot{r})$ at the time t, and together with the four component attributable this provides a set of six initial conditions, that is a set of orbital elements at the epoch t (to be corrected for light travel time).

The problem is how dense the triangulation needs to be, that is how many orbits have to be propagated from time t to time t_1, to sample the ephemerides in a useful way, that is with distances among the sampling points comparable to the field of view of the telescope to be used. We have no rigorous answer to this question, but it is clear that a regular triangulation of the admissible region is a strategy to optimize this procedure, provided the selection of the metric (discussed in Section 5.3) is appropriate.

We conclude by providing an example. For the same object 2003 BH_{84} discussed in Section 4, we have computed the triangulation of the admissible region with the metric defined by Equation (12). The inner boundary of this example is defined using $H_{max} = 22$ as maximum value for the absolute magnitude. In Figure 6 (above) we are showing the admissible region and its triangulation; we also mark with $+$ the value of $(r, \dot{r})$ obtained in hindsight, that is resulting from the orbit determined by using also the recovery observations of 30 January and 6 February 2003.

Next, we have attempted to predict the recovery on 6 February by using the 25 January attributable only. The triangulated ephemerides are shown in Figure 6 (below), the actual recovery observation being marked with a circled asterisk. Note that the $H_{max} \leqslant 22$ condition is important to reduce the size of the recovery search area.

7. Conclusions and Future Work

When observations are available only over a very short time span, to the point that all the meaningful information can be summarized in an

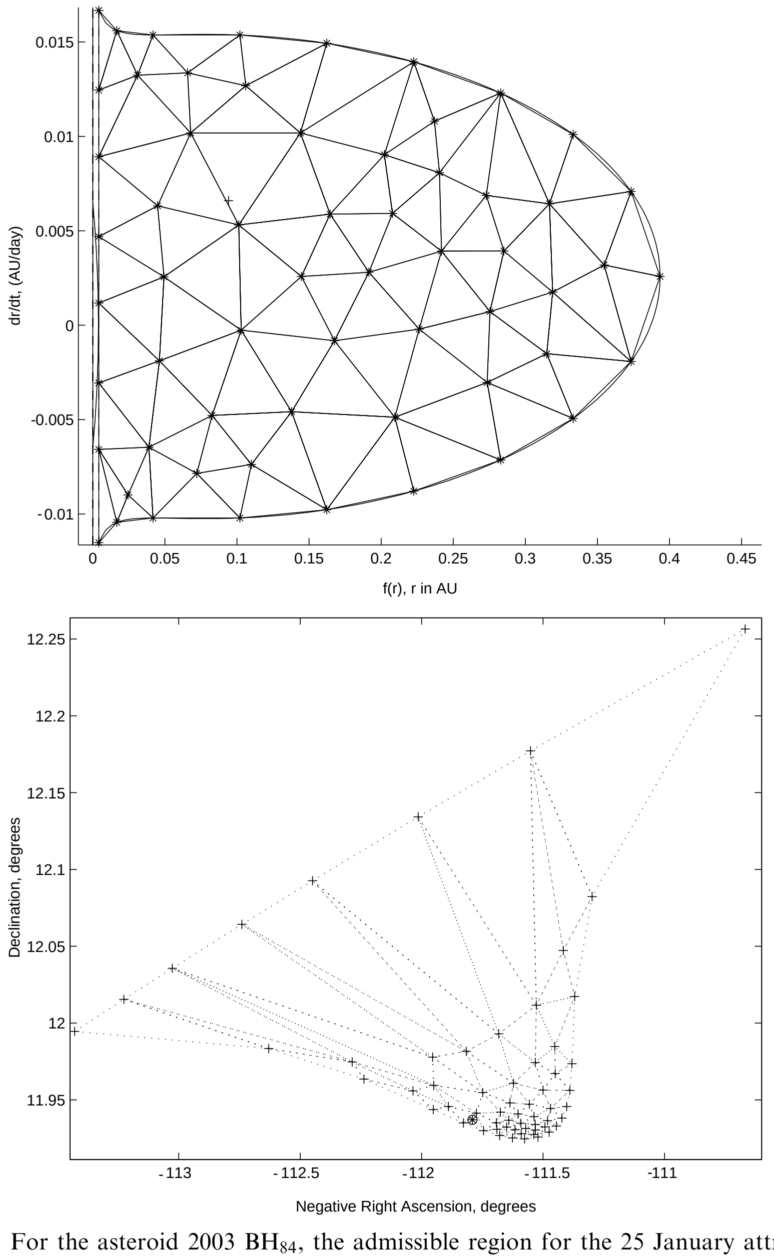

Figure 6. For the asteroid 2003 BH_{84}, the admissible region for the 25 January attributable and its triangulation (above); the actual value, as determined a posteriori, is marked with $+$. The image of the triangulation on the plane of the ephemerides for 6 February, with the actual observation marked with a circled asterisk (below).

attributable (with no significant curvature), it is not possible to perform an orbit determination in the usual sense.

We have introduced the concept of an attributable to summarize the available information, provided from the observations, in a four-dimensional vector. Given an attributable, a full orbit with six parameters cannot be computed, and the missing information can be represented by unknown range and range rate. We have shown that these two unknowns are however constrained to a compact subset, the admissible region, in the hypothesis that the object being observed has a solar system orbit.

Some orbital information is anyway available from the attributable: it can be exploited to assess the relevance of the discovery; as an example, it is possible to decide that a given attributable does not belong to a main belt asteroid, and this can be the case even when the proper motion is moderate.

The information contained in an attributable is enough to constrain the future observations: the admissible region can be mapped into the celestial sphere at some other time. To be able to approximate this map with finite computations we need to sample the admissible region in an efficient way, adjusted to its shape and size. This we have achieved by using an optimized triangulation of the admissible region. In this way, by propagating only the nodes of the triangulation, we can predict ephemerides, e.g., for the purpose of planning the recovery of the same object.

The next question is how this information can be combined with additional observations to provide an orbit. As an example, we can try to combine information from two attributables, that is eight scalar conditions, to compute a full 6-parameter orbit. This is discussed in a paper in preparation.

Acknowledgements

This research was supported by the *Italian Space Agency (ASI)*, the Spanish *Ministerio de Ciencia y Tecnología*, by the European funds *FEDER* through the grant AYA2001-1784 and by the *Ministry of Science, Technology and Development of Serbia* through the project no. 1238. We thank A. Lemaitre for suggesting many improvements.

Appendix A. Uniform sampling of a curve

Given n points on a rectifiable curve γ, with unitary length, we study the problem of selecting m among them $(m < n)$ such that the distance along the curve between two consecutive points is as close as possible to $1/(m-1)$. We search for an approximation of the solution of this problem avoiding to compute the *arc length*.

Without loss of generality we can assume that γ is the unit interval $[0,1] \subset \mathbb{R}$. Let $(P_k)_{k=1\ldots n}$ be a set of ordered points in $[0,1]$, with $P_1 = 0, P_n = 1$, and let $(Q_j)_{j=1,m}$ be the sequence of equispaced *ideal points*, with

$$Q_{j+1} - Q_j = \frac{1}{m-1} = h \text{ (ideal step).}$$

We define $d_k = P_k - P_{k-1}$ and $\delta_{k,j} = |Q_j - P_k|$; note that for each P_k there exists an ideal point Q_j such that $\delta_{k,j} \leqslant h/2$. We introduce an elimination rule in order to discard a point from the set $(P_k)_{k=1\ldots n}$:

Elimination rule: skip the point $P_{\bar{k}}$ such that $\bar{k}$ minimizes the function

$$f(k) = \frac{\min\{d_k, d_{k+1}\}}{1 + \min_{j=1\ldots m} \delta_{k,j}}, \quad k = 2 \ldots n-1. \tag{13}$$

We apply $(n-m)$ times the previous rule: note that at each step the values of d_k may change, due to the elimination of points in the set $(P_k)_{k=1\ldots n}$. We shall call $(\widehat{P}_j)_{j=1\ldots m}$ the subset of the points selected in $(P_k)_{k=1\ldots n}$. Now we prove that the above algorithm selects the ideal points if these are contained in the initial set $(P_k)_{k=1\ldots n}$.

PROPOSITION 1. *If the condition*

$$(Q_j)_{j=1\ldots m} \subseteq (P_k)_{k=1\ldots n} \tag{14}$$

holds, then $\widehat{P}_j = Q_j$ *for each* $j = 1\ldots m$.

Proof. If $f(\bar{k}) \leqslant f(k)$ for each k, we shall prove that for $n > m$

(i) $f(\bar{k}) < h$;
(ii) $P_{\bar{k}} \notin (Q_j)_{j=1\ldots m}$.

We observe that we cannot have $f(\bar{k}) > h$ because, for each k, $f(k) \leqslant \min\{d_k, d_{k+1}\}$ by definition, and $d_k \leqslant h$ by (14). We exclude also $f(\bar{k}) = h$ because $n > m$ implies that there are two points whose mutual distance is less than h, and the first property is proven.

If by contradiction $P_{\bar{k}} = Q_j$ for some index $j \in 1\ldots m$, then we show that there exists an index $k \neq \bar{k}$ in which the function f assumes a smaller value. Note that $f(\bar{k}) = \min\{d_{\bar{k}}, d_{\bar{k}+1}\}$ as the denominator in (13) reduces to 1. Assume that $d_{\bar{k}} = \min\{d_{\bar{k}}, d_{\bar{k}+1}\}$, then $\min\{d_{\bar{k}-1}, d_{\bar{k}}\} \leqslant d_{\bar{k}}$ and, by property (i), $d_{\bar{k}} < h$, so that $P_{\bar{k}-1}$ is different from each Q_j. This implies $\delta_{\bar{k}-1,j} > 0$ for each j, thus $f(\bar{k}-1) < f(\bar{k})$. The case $f(\bar{k}+1) = \min\{d_{\bar{k}}, d_{\bar{k}+1}\}$ is similar.

The skipped points are never ideal points, thus after $n-m$ steps the remaining points are the ideal points Q_j. $\square$

In the general case, when starting from points selected with a non optimal procedure, the remaining points will not be the ideal ones, exactly equispaced, but they will anyway be a better approximation.

References

Bern, M. and Eppstein, D.: 1992, 'Mesh Generation and Optimal Triangulation', In: D. -Z. Du and F. K. Hwang (eds), *Computing in Euclidean Geometry*, World Scientific, pp. 23–90.

Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J. and Harris, A. W.: 1989, 'Application of photometric models to asteroids', In: R. P. Binzel, T.Gehrels and M. S. Matthews (eds), *Asteroids II*, University of Arizona Press, Tucson pp. 524–556.

Danby, J. M. A.: 1989, *Fundam. Celestial Mech.*, Willmann-Bell, Richmond.

Delaunay, B.: 1934, 'Sur la sphere vide', *Izv. Akad. Nauk SSSR*, *Otdel. Mat. Estestv. Nauk* **7**, 793–800.

de' Michieli Vitturi, M.: 2004, 'Approximate gradient-based methods for optimum shape design in aerodynamic', Ph.D. Thesis, University of Pisa.

Field, D. A.: 1988, 'Laplacian smooting and Delaunay triangulations', *Communications Applied Methods* **4**, 709–712.

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

Milani, A., Sansaturio, M. E. and Chesley, S. R.: 2001, 'The Asteroid Identification Problem IV: Attributions', *Icarus* **151**, 150–159.

Risler, J. J.: 1991, 'Méthodes mathematiques pour le CAO, Collection Recherche en mathematiques appliquées', *RMA 18*, Masson.

Virtanen, J., Muinonen, K. and Bowell, E.: 2001, 'Statistical Ranging of Asteroid Orbits', *Icarus* **154**, 412–431.

Winslow, A. M.: 1964, 'An irregular triangle mesh generator' *Report UCXRL-7880*, National Technical Information Service, Springfield, VA.