

**FROM ASTROMETRY TO CELESTIAL MECHANICS:
ORBIT DETERMINATION WITH VERY SHORT ARCS**
(*Heinrich K. Eichhorn Memorial Lecture*)

ANDREA MILANI¹ and ZORAN KNEŽEVIĆ²

¹*Department of Mathematics, University of Pisa, via Buonarroti 2, 56127 Pisa, Italy,
e-mail: milani@dm.unipi.it*

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

(Received: 1 October 2004; revised: 15 February 2005; accepted: 7 March 2005)

Abstract. Contemporary surveys provide a huge number of detections of small solar system bodies, mostly asteroids. Typically, the reported astrometry is not enough to compute an orbit and/or perform an identification with an already discovered object. The classical methods for preliminary orbit determination fail in such cases: a new approach is necessary. When the observations are not enough to compute an orbit we represent the data with an attributable (two angles and their time derivatives). The undetermined variables range and range rate span an *admissible region* of solar system orbits, which can be sampled by a set of *Virtual Asteroids* (VAs) selected by an optimal triangulation. The attributable results from a fit and has an uncertainty represented by a covariance matrix, thus the predictions of future observations can be described by a quasi-product structure (admissible region times confidence ellipsoid), which can be approximated by a triangulation with each node surrounded by a confidence ellipsoid. The problem of identifying two independent short arcs of observations has been solved. For each VA in the admissible region of the first arc we consider prediction at the time of the second arc and the corresponding covariance matrix, and we compare them with the attributable of the second arc with its own covariance. By using the penalty (increase in the sum of squares, as in the algorithms for identification) we select the VAs which can fit together both arcs and compute a preliminary orbit. Even two attributables may not be enough to compute an orbit with a convergent differential corrections algorithm. The preliminary orbits are used as first guess for constrained differential corrections, providing solutions along the *Line Of Variations* (LOV) which can be used as second generation VAs to further predict the observations at the time of a third arc. In general the identification with a third arc will ensure a least squares orbit, with uncertainty described by the covariance matrix.

Key words: asteroid recovery, ephemerides, orbit determination

1. Introduction

The astrometric observations of a small body by themselves do not provide an orbit for the observed body, thus do not provide information on the

nature of the object (asteroid, comet, satellite, Transneptunian). The first complete mathematical method to convert astrometry into orbits had been established by Gauss (1809): he devised an algorithm to compute a preliminary orbit satisfying three given observations in different nights. When additional observations became available, Gauss proposed to correct the preliminary orbit by solving a least squares problem. This method is now called *differential corrections*, and this sequence, preliminary orbit followed by least squares,¹ is now the algorithm almost universally used and considered classic. At the time when the asteroids were detected as “intruders” not found in star charts, the observations were indeed typically only one per night, and the algorithms found by Gauss were the optimal solution of the orbit determination problem. The fact is, the circumstances of the observations of asteroids (and other small bodies) are now deeply changed: the historical discovery procedure should not be conditioning our way of thinking about orbit determination to be performed with modern data.

The number of asteroid observations has in recent years increased dramatically, mostly because of the automated surveys like LINEAR, LONEOS, Catalina, Spacewatch, NEAT. The procedures of operation of these surveys are basically the same, although they can differ in details. A number of images of the same area on the celestial sphere are taken over a short time span, typically within a single night. The images are then digitally blinked and all the changes from one to another logged. If an object moves from image to image, at a constant rate and along a straight line, this is probably the detection of a real body. The series of observations, usually consisting of 3–5 positions over a time span 1–2 hours, are reported as a sequence of individual observations of the same object (note that this initial identification is done by the observer); we shall refer to this sequence as a *very short arc*.

This method of work is optimal for the discovery of asteroids and comets, but it is not suited for the determination of their orbits. In Gauss’ method for preliminary orbit determination the curvature of the arc on the celestial sphere appears as divisor already in the first iteration. The smaller the curvature, the less accurate the resulting orbit: taking also into account the observational errors, often the standard algorithm fails to provide the solution. Either the preliminary orbit cannot be determined at all, or it can be computed, but the differential corrections do not converge. In such a case, we speak of a *Too Short Arc* (TSA), by which we mean too short for

¹In the modern orbit determination, the least squares solution is computed by solving numerically the perturbed N-body problem, while Gauss was using the analytical solution of the 2-body problem, but the spirit is not changed.

orbit determination. The reported sequence of observations can be considered the detection of a moving object, rather than a discovery.

This situation is unsatisfactory, given the wealth of information contained in these observations, which remains unused due to the failure of the orbit determination procedure. Without an orbit, on the other hand, we cannot determine what kind of body we are looking at, we can neither compute an ephemeris for the later follow up observations, nor identify the observed object with any other associated to a known orbit. We need to establish a new paradigm for the process leading from astrometry to orbit determination, working efficiently and reliably under the prevailing observing conditions of today.

Our goal is to develop the procedure which would allow to extract all the existing information from the TSA, and to combine it with some plausible assumptions about the nature of the motion of the detected body, in order to get preliminary orbits. Our research plan consists of several steps, of which three are completed (Milani et al., 2004, 2005a,b); these are described in the present paper. The basic idea is as follows: a TSA comprises a number of observed positions with deviations from alignment compatible with a random observational error. We can fit a straight line to the data and compute two average angular coordinates and their corresponding average angular rates, assigning the results to the reference epoch (simple mean of the observing times). We shall call such a set of data an *attributable*.

Note that an attributable does not provide any information on the geocentric distance (range) of the body and its radial velocity (range rate) at the reference time. However, the range and range rate are constrained if we assume that the body belongs to the solar system, but that it is not a satellite of the Earth. Hence we introduce a concept of *admissible region*, which in our algorithm replaces the conventional confidence region as defined in the classical orbit determination procedure. Such a region can be sampled by virtual asteroids, and we can compute a sort of generalized ephemerides which allow identification in the sense of attribution (Milani et al., 2001), linking of two TSA's and computation of preliminary orbit, detection of virtual impactors, etc.

Our method represents an extension and important upgrade of the existing methods developed by Virtanen et al. (2001), by Tholen and Whiteley (2003) and by Goldader and Alcock (2003). We introduced several significant improvements making our method more efficient and reliable, but we agree with these authors' main conclusion that ephemerides prediction is often possible, with an accuracy good enough e.g. for recovery planning, even when the orbit cannot be computed in the usual way.

2. Definition of 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, α, δ) be spherical coordinates for the geocentric position $P - P_{\oplus}$.

DEFINITION 1. *We shall call **attributable** a 4-dimensional vector $A = (\alpha, \delta, \dot{\alpha}, \dot{\delta})$, observed at a time t .*

Here t has to be interpreted as the mean of the observation times. The angles (α, δ) can be specified as necessary: usually the geocentric equatorial coordinates, right ascension and declination for the standard epoch J2000, are used. Also, with the data contained in the observations we optionally can have another component of the attributable – an average apparent magnitude h . Note that range and range rate (r, \dot{r}) are left completely undetermined by this definition.

The conditions to constrain (r, \dot{r}) make use of the following well-known quantities:

Heliocentric two-body energy

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

where $k = 0.01720209895$ is Gauss' constant;

Geocentric two-body energy

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

where μ_{\oplus} is the ratio (mass of the Earth)/(mass of the Sun);

Radius of the sphere of influence of the Earth

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

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

$$R_{\oplus} \simeq 4.2 \times 10^{-5} \text{ AU}.$$

The following four conditions make now obvious physical sense:

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

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

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

3. Borders of the Admissible Region

The multi-line border of the admissible region can be mathematically described in a rigorous way. The procedures and all the results are described in full detail in Milani et al. (2004).

The admissible region cannot have more than two connected components. More precisely, the degree six polynomial resulting from condition (C) cannot have more than three real positive roots: when there is only one such root, the admissible region has only one component, when there are three, it has two components (see Figures 1 and 2 in Milani et al., 2004).

The boundary of the admissible region consists of

1. part of the algebraic curve $\mathcal{E}_\odot = 0$. If the degree six polynomial has three positive roots there is another component, a simple closed curve, at larger values of r : this includes the case when this curve reduces to a single point, if there is 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 $r = R_{\text{SI}}$ if $R_{\text{SI}} < r_0$; if $R_{\text{SI}} \geq r_0$ the two portions of the $\dot{r}^2 = G(r)$ are joined at $r = r_0$.

Note that $G(r)$ derives from condition (A):

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

where η is the proper motion and $G(r) > 0$ for

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

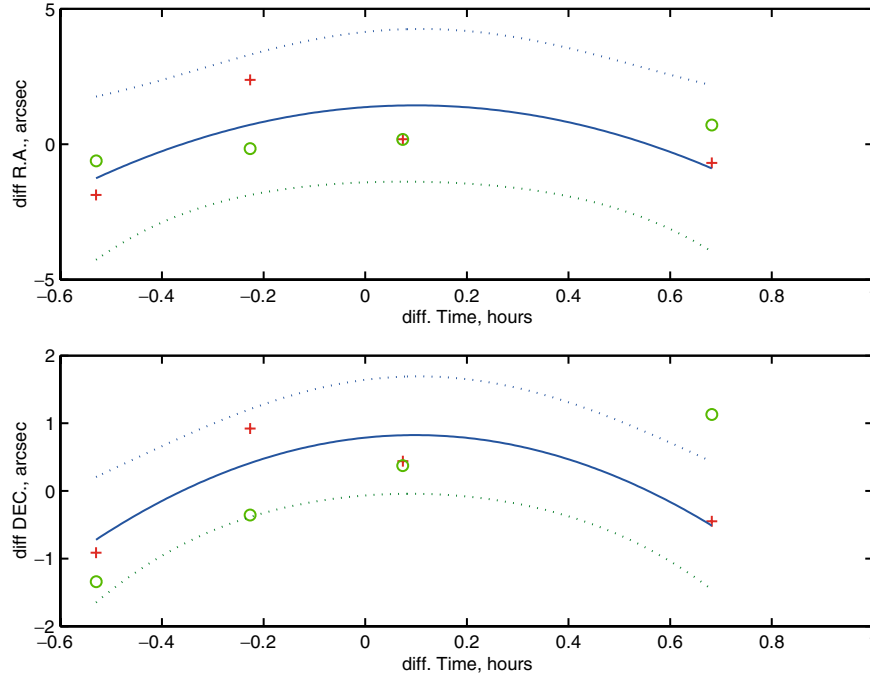


Figure 1. Residuals in right ascension (above) and declination (below) with respect to the fit of the observations of the asteroid 2004 SA₁ taken by LINEAR on 14 January 2004. The crosses represent the observations as reported, the continuous line is the best fitting parabola, the dotted lines the confidence boundaries for the parabolic fit, the circles are the observations “without astrometric error”, computed from a very well determined orbit known a posteriori.

This result provides full analytical and topological description of the admissible region. From the metric point of view, however, the definition of the region is not entirely satisfactory, since the inner boundary might be too close to the observer. Instead of condition (D), therefore, one can set an upper limit for the absolute magnitude of the body to exclude very small and very close objects of minor importance:

(E) the absolute magnitude H of the object is $\leq H_{\max}$.

The region defined by condition (E) is a half plane $r \geq r_H$ (the difference between the absolute and the observed apparent magnitude does not depend on the range rate). As an example, if we set $H_{\max} = 25$, corresponding to a body which most likely would not result in a very significant damage on the ground in case of impact, and the measured apparent magnitude is $\simeq 20$, the resulting $r_H \simeq 0.01$ AU.²

²For an accurate computation of r_H the phase effect has to be taken into account.

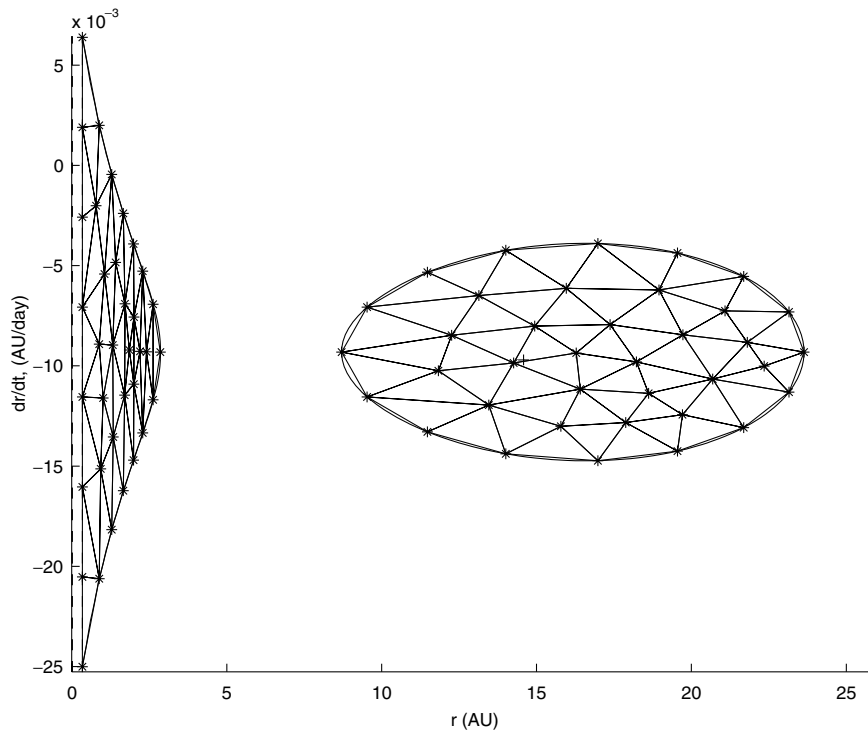


Figure 2. Triangulation of the modified admissible region for the precovery observations of 2000 EC₉₈. The cross marks the actual position, in the range, range rate plane, of the Centaur, as determined a posteriori with additional observations.

Sometimes it is also desirable to limit the upper boundary of the admissible region to exclude from consideration the long periodic comets with large orbital semimajor axes $a > a_{\max} = 100 \text{ AU}$; condition (C) in such a case becomes $(C_{100}) \mathcal{E}_{\odot} \leq -k^2/(2a_{\max})$. If we apply these alternative conditions (E) and (C_{100}) in place of (C) and (D), we speak of the *modified admissible region*.

4. Sampling the Admissible Region

Given a very short arc of observations and the corresponding attributable $A_0 = (\alpha, \delta, \dot{\alpha}, \dot{\delta})$ at the mean observation time t , we need to represent the uncertainty of the orbit, hence of the ephemerides, in a different way with respect to the traditional method of covariance matrices and confidence ellipsoids. This because a least square solution for the orbital ele-

ments X , with its normal matrix C and covariance matrix Γ , is in general not computable.

If the arc is very short, to the point that there is no significant information on the curvature of the path on the celestial sphere, the values of (r, \dot{r}) are not constrained by the observations. If, to the contrary, there is significant curvature, this allows to constrain both the range and the range rate (Danby, 1989, Ch. 6). Figure 1 shows a typical example of a set of four observations spanning only one hour, for which the curvature is not significant, that is less than the uncertainty resulting from the known performance of the station (Carpino et al., 2003.)

Even without curvature information, if we assume the object belongs to the solar system we can limit the uncertainty in the (r, \dot{r}) plane to the admissible region. The (modified) admissible region is a compact subset of the $r > 0$ half plane. However, this is still an infinite set, with a complex shape, thus there are infinite possible orbits and we need to find efficient ways to sample them with a finite number of points in the initial conditions space, the *Virtual Asteroids* (VAs) sharing the reality of the object, in the sense that the orbit through one of them is a good approximation of the orbit of the real object, but we do not know which one.

This requires to sample the admissible region \mathcal{D} with a number of points. The most natural and geometrically significant way to sample a 2-dimensional region is a *triangulation*, with nodes and sides joining them. Regions with simple boundaries (such as the $r > 0$ half plane) can be triangulated by equilateral triangles, but the admissible region has a complicated boundary. A *Delaunay triangulation* has a number of optimal properties, e.g., it is the triangulation with the largest minimum angle (among all the triangles).

There is an efficient algorithm to compute a Delaunay triangulation, starting from a finite sampling of the boundary. Thanks to the explicit analytic description of the admissible region we can sample the boundary with approximately uniform distances and compute a Delaunay triangulation with the given nodes on the boundary. Then the nodes are selected as points (r_i, \dot{r}_i) , $i = 1, N$ sampling the admissible region, with the sides and the triangles providing an additional geometric structure.³

As an example, from the precovery observations of the Centaur 2000 EC₉₈ taken on February 5, 2000 (one month before the official “discovery”) we have computed the attributable, the modified admissible region and its

³The properties defining the Delaunay triangulation are metric ones, thus the nodes selected depend upon the choice of a metric. Different metrics can be used to enhance the resolution in some portions of the admissible region, e.g., when the main goal is to search for either NEO, or main belt asteroids, or TNO.

Delaunay triangulation, shown in Figure 2. Note that in this case, as it is common for Centaurs and Transneptunians, the admissible region has two connected components and the real object is in the one farther from the Earth.

The set of VAs selected is then

$$X_i = (\alpha, \delta, \dot{\alpha}, \dot{\delta}, r_i, \dot{r}_i), \quad i = 1, N.$$

The set of six coordinates forming the vector $X_i = (\alpha, \delta, \dot{\alpha}, \dot{\delta}, r_i, \dot{r}_i)$ can be considered as a set of orbital elements.⁴ They can be converted into Cartesian position and velocity (topocentric, then heliocentric since the observer's position is well known), as well as into other types of elements, e.g., Keplerian.

To assess how representative are the set of orbital element vectors $\{X_i\}$ with respect to the full set of possible orbits we need to take into account that even the measured part of the X_i vectors, that is the 4-dimensional vector A_0 , has some uncertainty. The values of the angles and their rates are computed by least square fit to a set of observations, thus their uncertainty can be represented in the conventional way with a 4×4 covariance matrix Γ_A . If we assume the value of two variables (r_i, \dot{r}_i) , without uncertainty, then the uncertainty of the 6-vector X_i is represented by the *conditional covariance matrix*, a 6×6 matrix Γ_X

$$\Gamma_X = \begin{bmatrix} \Gamma_A & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}$$

with $\underline{0}$ suitable matrices with null coefficients. This matrix is obviously not positive-definite, but has the (r, \dot{r}) subspace as kernel (null space).

A geometric description of the confidence region for the orbits compatible with the given attributable with uncertainty (A_0, Γ_A) can be given as follows: for each point A of a confidence ellipsoid centered in A_0 (in the 4-dimensional attributable space) we need to compute the admissible region $\mathcal{D}(A)$. This gives as confidence region a subset of the 6-dimensional elements space with a positive 6-dimensional volume. However, if the attributable A_0 is well determined, the confidence region is flattened, with a two dimensional “base plate” $\{A_0\} \times \mathcal{D}(A_0)$. This set is not exactly a product (confidence ellipsoid \times admissible region), because the confidence region changes with the attributable; however, it is not very different from a product and the set of VAs $X_i, i = 1, N$ selected along $\{A_0\} \times \mathcal{D}(A_0)$ is quite representative.

⁴The epoch corresponding to the initial conditions is $t_j = t - r_j/c$, with c the speed of light, to take into account the light travel time.

5. Propagating to Another Epoch

Given a triangulation $B_i, i = 1, N$, each (X_i, Γ_X) is an orbital element set with uncertainty.

If these elements are converted into other coordinates Y , such as Keplerian elements, and propagated (nonlinearly) to some later epoch t_1 , the covariance matrix can be propagated (linearly) to a new one $\Gamma_Y(t_1)$. The fact that the covariance matrix Γ_X is not invertible does not matter, but of course $\Gamma_Y(t_1)$ will also have a 2-dimensional kernel. Then it is possible to compute the predicted attributable (for observation time t_1) with uncertainty (A^i, Γ_{A^i}) for each node. The ephemerides (with confidence ellipse) computed in the conventional case of a well determined orbit are thus replaced by the union of N confidence ellipsoids, one for each VA, in the 4-dimensional space of the attributables A' at the new epoch.

If another very short arc of observations is available at the second epoch, its attributable with uncertainty (A_1, Γ_{A_1}) can be compared with the predictions for each VA. As an example, for the same asteroid of Figure 2, we have computed the attributable A_1 corresponding to the official discovery observations of March 3–4, 2000: Figures 3 and 4 show the predictions A^i for the same attributable, computed by using only the data of February 5.

Both in angles and in proper motion the predictions can be good enough for pointing the telescope and to discriminate this object from others in the same fields, that is by using the triangulation technique the “discovery” could have been a targeted recovery. Note that the uncertainty arises from two mathematically very different sources: the span of the admissible region, and the confidence ellipsoids. In this example, if only the connected component farther from the Earth is considered the span of the predictions from the different VA selected in the admissible region is small, while the size of the ellipsoid is the main error term. On the contrary the other connected component would give predictions for the different VAs spread over a large part of the sky, none of them close to the recovery observations. This means that a targeted recovery after one month was possible only assuming that the object was on an orbit beyond Jupiter, an assumption which was by no means obvious given the available data.

6. Identification and Preliminary Orbits

When using the data from surveys, the correspondence between the very short arcs of observations and the physically distinct objects is not known. This is the problem of *asteroid identification*. Since the arcs are too short,

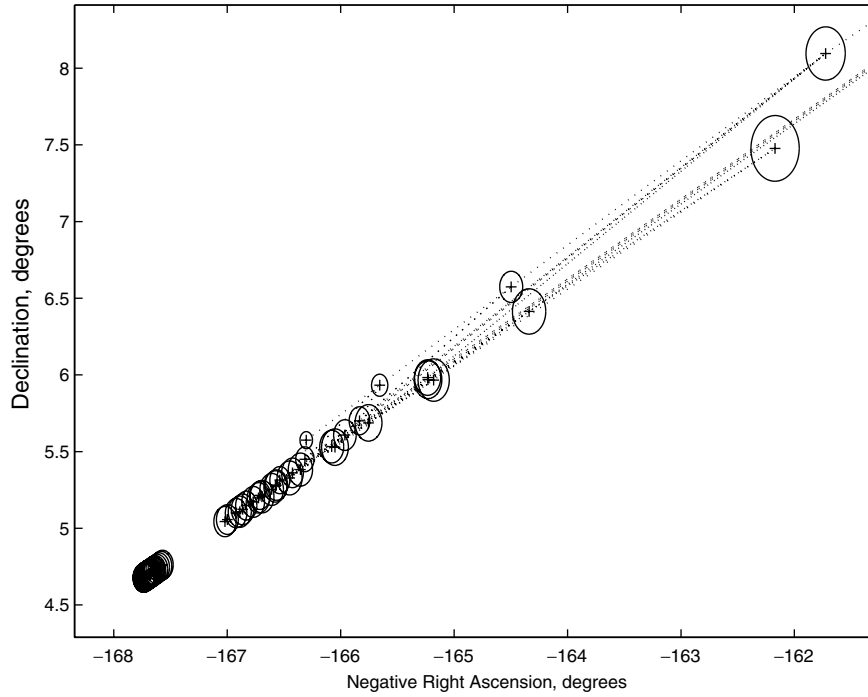


Figure 3. Prediction (propagated triangulation) for the observations of March 3–4, 2000 based only on the predisccovery observations of February 5, projected on the (α, δ) plane. The confidence ellipses are also drawn, but the uncertainty of the prediction is dominated by the unknown location on the (r, i) plane at the time of the predisccovery attributable.

more than one TSA needs to be used to constrain the orbit, to be able to obtain a least squares solution. However, how do we know whether two TSAs belong to the same object, having an orbit for neither one of the two?

An algorithm has been defined (Milani et al., 2001) to decide whether two arcs of observations are to be identified, provided that an orbit is available (for at least one of the two) with its uncertainty, represented by a covariance matrix: this case of identification is called an *attribution*, hence the term *attributable*. When both observed arcs are TSA, the only available orbits are the ones of the VA defined by the triangulation of the admissible region of the first attributable.⁵

The predicted attributable with uncertainty (A^i, Γ_{A^i}) is computed for each node and it is compared with the attributable computed from the

⁵Of course it is also possible to triangulate the admissible region of the second attributable: the algorithm is not invariant by exchanging the two observed arcs.

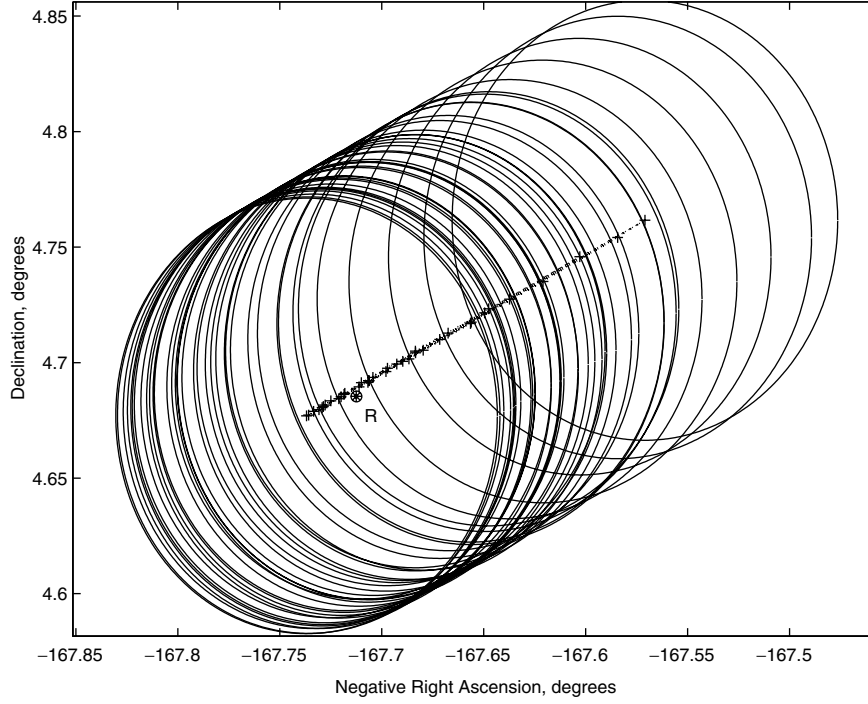


Figure 4. Detail of the previous figure, showing only the predictions from the nodes belonging to the second connected component of the admissible region, the one farther from the Earth. For this subset of VAs, the uncertainty is dominated by the uncertainty of the attributable, as expressed by its covariance matrix and represented here by the confidence ellipses. The recovery attributable as observed is labeled by the letter R.

observations of the second arc (A_1, Γ_{A_1}) . By means of the normal matrices

$$C_{A^i} = [\Gamma_{A^i}]^{-1}, \quad C_{A_1} = [\Gamma_{A_1}]^{-1}$$

the attribution penalty K_i is computed as

$$\begin{aligned} C_1^i &= C_{A^i} + C_{A_1}, \\ C &= C_{A^i} - C_{A^i} [C_1^i]^{-1} C_{A^i} = C_{A_1} - C_{A_1} [C_1^i]^{-1} C_{A_1}, \\ K_i &\simeq \frac{1}{2} [A^i - A_1]^T C [A^i - A_1]. \end{aligned}$$

If the attribution penalty K_i is low for some VA index i , then the two attributables A_0 and A_1 may belong to the same object. If this is the case, we need a preliminary orbit to be used as first guess to start the differential correction procedure. One possibility is to use X_i as the preliminary orbit, but there are ways to define a better one, taking into account that an optimal “compromise” attributable between the predicted A^i and the observed

A_1 is provided by the algorithms of Milani et al. (2001, 2005b, submitted):

$$A_1^i = [C_1^i]^{-1} (C_{A_1} A_1 + C_{A^i} A^i).$$

This algorithm has a geometrical interpretation in terms of intersections of the two families of confidence ellipsoids.

At this stage, the identification between the two attributables can only be hypothetical. The values of the attribution penalty to be considered acceptable cannot be very small, because the minimum of the values of K_i for $i = 1, N$ is not the same as the minimum for all possible choices of (r, \dot{r}) in $\mathcal{D}(A_0)$: the finite sampling implies an increase of the penalty with respect to the minimum possible. As a result, acceptable values may occur also for couples of attributables which either cannot be fit together, or would give large residuals.

The procedure outlined in this section has to be interpreted just as a step of a multi-stage filtering process to select the proposed identifications, to be confirmed by finding an orbit fitting both sets of observations according to the least square principle.

7. Constrained Differential Corrections

When only two very short arcs are available the classical differential correction procedure, starting from a rough preliminary orbit, may not converge. Even if it converges to a nominal least squares solution, the latter may be poorly determined and may a posteriori turn out to be very far from the true orbit. A good strategy is to seek for a *Line Of Variations* (LOV) orbit, by means of a *constrained differential corrections* algorithm (Milani et al., 2005a).

At a point X in the elements space (at an epoch near the observations) the orbit determination has a *weak direction* $V_1(X)$, corresponding to the long axis of the confidence ellipsoid computed with the normal matrix $C(X)$. The point X is on the LOV if the cost function Q restricted to the hyperplane $H(X)$ normal to the weak direction $V_1(X)$ has a local minimum. The differential corrections can be constrained to corrections lying on the hyperplane $H(X)$. If the iterations converge, then the limit point is on the LOV (Milani et al., 2005a). In other words, the constrained differential corrections are attempting to find a five parameter solution, with the assumed variable along the weak direction. It is often the case, when there are few data, that constrained differential corrections converge, but full differential corrections do not.

This procedure is very effective when combined with the computation of preliminary identification orbits described in the Section 6. For the same example of the previous Figures, namely the attributables formed with the pre-discovery data of February 5, 2000 and with the discovery data of March 3–4 for the Centaur 2000 EC₉₈, we have selected the 5 VAs with $\sqrt{K_i} \leq 6$, corresponding to the nodes 4, 15 and 25–27 of the triangulation, computed the preliminary orbits and used them to start the constrained differential corrections.

The LOV is represented by a straight line segment on the (r, \dot{r}) plane of the first attributable in Figure 5, with the 5 LOV solutions labeled with the node indexes. Note that 3 out of the 5 LOV solutions are outside the admissible region, and indeed they are hyperbolic; nevertheless, one of the elliptic solutions (number 4) is quite close to the “true” solution (as determined a posteriori with additional observations), which is marked with a crossed square.

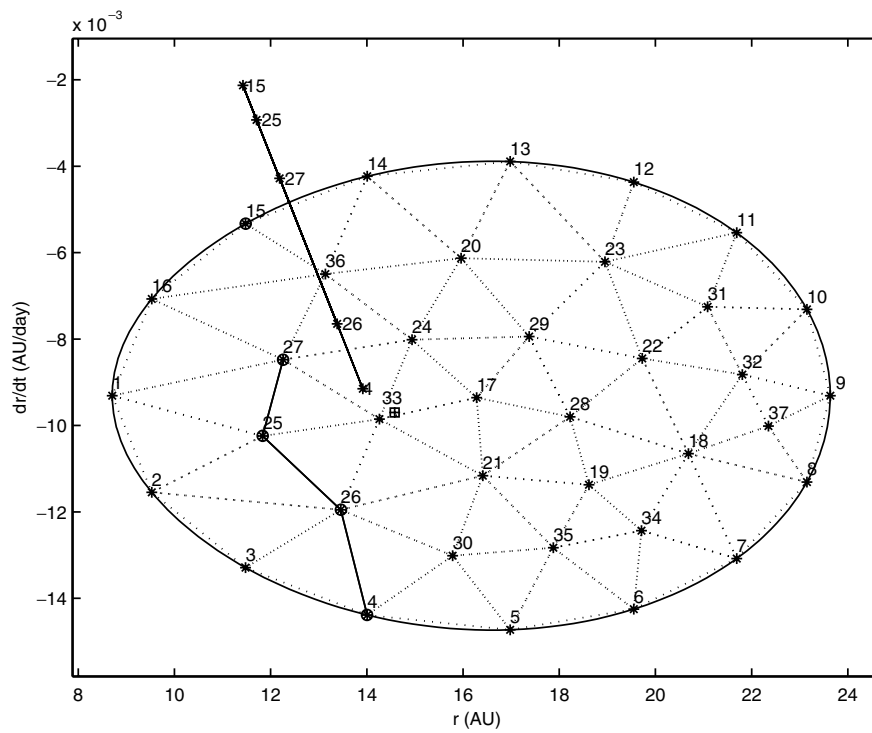


Figure 5. Identification of the discovery observations (March 3–4, 2000) with the preliminary orbits based upon the February 5 pre-discovery observations. The sides of the triangulation joining nodes with identification penalty ≤ 36 are marked with a full line, the others are dotted. The true solution is marked with a crossed square.

In this case the attempt to compute a nominal solution, even starting from a comparatively good guess (such as the LOV point number 4), fails. However, the two elliptic solutions (in particular number 4) are good enough to compute useful predictions, allowing identification with additional observations, as described in the next Section.

8. From Multiple Solutions to the Nominal Orbit

When a third very short arc is available, the observations can be compared with the predictions resulting from the LOV solutions. As an example, from LOV solution number 4 of Figure 5 we have computed the predictions for an observation of 2000 EC₉₈ on April 4, 2000. The confidence region is not small, although it is narrow.

The attribution of this observation to one of the LOV orbits computed as described in the Section 7 is quite obvious. Differential corrections provide an orbit fitting the three very short arcs with a normalized RMS of 0.35.

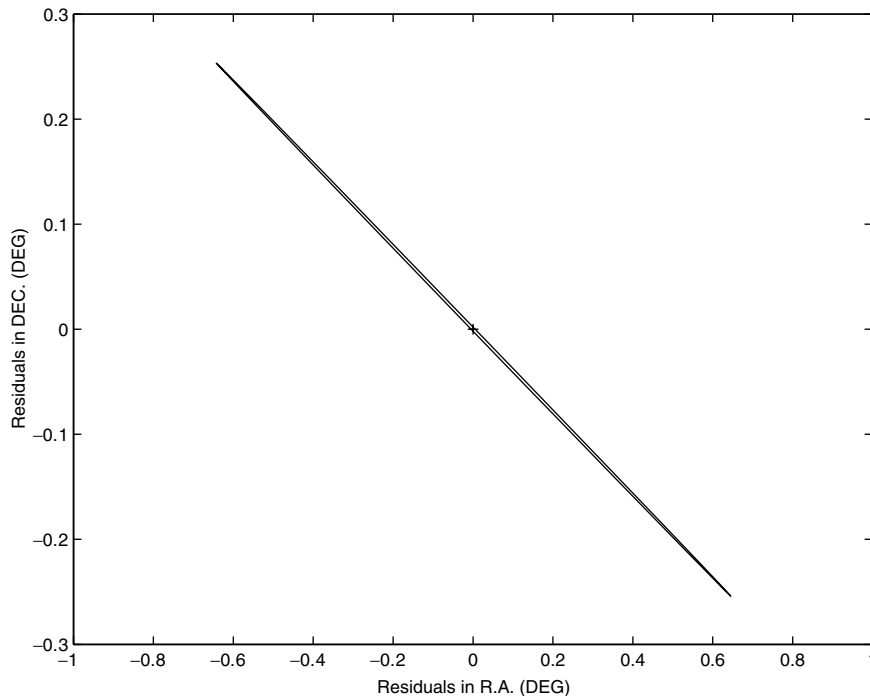


Figure 6. Confidence ellipse on the (α, δ) residuals plane for a recovery observation of 2000 EC₉₈ on 4 April 2000, computed by using one of the LOV solutions based upon the observations of February 5 and March 3–4. The cross is the recovery observation.

The orbit is still not tightly constrained by the observations, e.g., the semimajor axis is 11.3 ± 0.8 AU, the eccentricity 0.33 ± 0.15 , but the confidence ellipsoid is still a reasonably accurate description of the uncertainty region. As this example shows, the attribution of additional observations becomes easier and easier as more short arcs are available. After three short arcs have been fitted, the orbit determination procedure can follow the classical paradigm, with step by step improvements obtained by full differential corrections.

This could be the new paradigm, replacing the classical one (Gauss' preliminary orbit from three observations followed by full differential corrections). First, the attributables are computed by fitting the observations of the available very short arcs. Let us suppose three attributables belonging to the same object are available: the steps leading to the orbit determination are the following:

1. The admissible region of the first attributable is computed and sampled by a Delaunay triangulation, providing a set of VAs.
2. The predictions for the time of the second arc, computed from the VAs of the first, are compared with the second arc attributable.
3. For the VAs such that the attribution penalty (with the second attributable) is low a preliminary orbit is computed.
4. The above preliminary orbits are used as first guess in constrained differential corrections, providing, when there is convergence, LOV solutions fitting both very short arcs.
5. The LOV orbits are propagated and the predictions are compared with the attributable of the third arc.
6. For the LOV orbits such that the attribution penalty (with the third attributable) is low full differential corrections are used to check the attribution.
7. The orbit resulting from the fit to three very short arcs is used with its covariance, as in the classical procedure, for additional attributions when more observations are available.

In this way the orbit progresses from the stage in which there is an essentially 2-dimensional indetermination (steps 1–3), to a 1-dimensional indetermination (steps 4, 5), to a single nominal least squares solution (steps 6, 7).

9. Conclusions and Future Work

The new procedure, as described in the previous Section, has been rigorously defined, but to test it for reliability and efficiency in processing a large data set of astrometric observations is quite another matter.

Indeed the preceding discussion does not clearly indicate the main difficulty, namely, the false identifications. The problem to be solved is by no means to find an orbit fitting three very short arcs of observations, known to belong to the same object. If this was the case, by selecting one observation from each arc we could use Gauss' preliminary orbit method, then proceed with the classical paradigm. The problem is to compute $\simeq N$ orbits, given $\simeq N$ very short arcs in each one of three separate nights covering the same (large) region in the sky: for the current surveys, the number N is of the order of a few thousands, for the next generation surveys $N > 100,000$. It is computationally impossible to proceed, for each triple formed with one arc per each night, with the classical orbit determination procedure⁶. Thus we need to apply the most computationally intensive tests, such as the differential corrections, to a very small subset of the set of all triples.

The challenge is to define a sequence of filtering stages, with increasing computational load and screening a decreasing number of couples/triples of arcs. Of the steps listed in the previous Section, the ones with computational load linear in N , such as step 1, are not a problem. Step 2 is quadratic in N and requires a very efficient use of the orbit propagator, possibly a simplified one. Step 3 is quadratic, but simple; it should provide a number of couples candidate for identification by far smaller than N^2 , because step 4 is the most computationally intensive. Steps 5 and 6 are expensive, but they should be performed only on the confirmed couples. Step 7 needs to be performed only on the confirmed triples. All this would be inefficient if the number of spurious identifications resulting from steps 2–4 was large.

To test the algorithms of the new paradigm, and especially the global data flow and computational load, we need to use first simulated data. Real data are affected by too many errors and we can not know which identifications remain to be found. In a simulation, we have a “ground truth”, that is the catalog of objects used as input, thus we can exactly compute the level of completeness of a set of identifications obtained with a given procedure, and also the fraction of false identifications. By training our algorithms with the simulation test cases we can select the most efficient sequence of operations and the best values for the control parameters. After this, we have to test on real data.

So far the individual steps have been tested on both simulated and real data, but the full procedure has been tested only on comparatively small simulations. The next stages of our work will include testing on full scale

⁶For $\simeq 10^{15}$ preliminary orbits and a similar number of differential corrections, the CPU time with a current processor would be between 10^{12} and 10^{13} s.

simulations of the next generation surveys, and on the real data of the current surveys. Only after these tests are successfully concluded we will be able to claim we have established a new paradigm for the conversion of astrometric data into catalogs of orbits, and a new link between astrometry and celestial mechanics.

Acknowledgements

For this research we would like to acknowledge the essential contributions from O. Arratia, S.R. Chesley, M. de' Michieli Vitturi, G. Gronchi, M.E. Sansaturio, G. Tommei. We also thank R. Jedicke, G.B. Valsecchi and A. Boattini for giving important insight on the next generation surveys. We thank the organizers of the Humboldt Colloquium for the opportunity to present, as *Eichhorn Lecture*, our long term research program on asteroid orbit determination. Z.K. acknowledges support from the Ministry of Science and Environmental Protection of Serbia through project No 1238.

References

- Carpino, M., Milani, A. and Chesley, S. R.: 2003, 'Error statistics of asteroid optical astrometric observations', *Icarus*, **166**, 248–270.
- Danby, J. M. A.: 1989, *Fundamentals of Celestial Mechanics*, Willmann-Bell, Richmond.
- Gauss, C. F.: 1809, *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, reprinted by Dover publications, 1963.
- Goldader, J. D. and Alcock, C.: 2003, 'Constraining recovery observations for Trans-Neptunian objects with poorly known orbits', *Publ. Astron. Soc. Pacific*, **115**, 1330–1339.
- Milani, A., Sansaturio, M. E. and Chesley, S. R.: 2001, 'The asteroid identification problem IV: Attributions', *Icarus* **151**, 150–159.
- Milani, A., Gronchi, G. F., de' Michieli Vitturi, M. and Knežević, Z.: 2004, 'Orbit determination with Very Short Arcs. I Admissible Regions', *CMDA*, **90**, 59–87.
- Milani, A., Sansaturio, M. E., Tommei, G., Arratia, O. and Chesley, S. R.: 2005a, 'Multiple solutions for asteroid orbits: computational procedure and applications', *Astron. Astrophys.* **431**, 729–746.
- Milani, A., Gronchi, G. F., Knežević, Z., Sansaturio, M. E., and Arratia, O.: 2005b, 'Orbit determination with Very Short Arcs. II Identifications', *Icarus*, submitted.
- Tholen, D. and Whiteley, R. J.: 2003, 'Short Arc Orbit Computations', Preprint.
- Virtanen, J., Muinonen, K. and Bowell, E.: 2001, 'Statistical ranging of asteroid orbits', *Icarus*, **154**, 412–431.