Orbit determination with very short arcs. II Identifications

Andrea Milani [a], Giovanni F. Gronchi [a], Zoran Knežević [b]
Maria Eugenia Sansaturio [c], Oscar Arratia [c]

[a] *Dipartimento di Matematica, Università di Pisa, Via Buonarroti 2, 56127 Pisa, Italy*

[b] *Astronomical Observatory, Volgina 7, 11160 Belgrade 74, Serbia and Montenegro*

[c] *E.T.S. de Ingenieros Industriales, University of Valladolid Paseo del Cauce s/n 47011 Valladolid, Spain*

ABSTRACT

When the observational data available are not enough to compute a meaningful orbit for an asteroid/comet we can represent the data with an attributable (containing two angles and their time derivatives); then the undetermined variables range and range rate span an *admissible region* of solar system orbits, which can be represented by a set of Virtual Asteroids (VAs) selected by means of an optimal triangulation [Milani et al. 2004]. The attributable 4 coordinates are themselves the result of a fit and they have 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 computing a preliminary orbit starting from two short arcs of observations, represented by two attributables, can be solved by considering for each VA (selected in the admissible region of the first arc) the corresponding covariance matrix for the prediction at the time of the second arc, and by comparing it with the attributable of the second arc with its own covariance. By defining a suitable identification penalty we can select the VAs which allow to fit together both arcs at the attributable level, that is in the 8-dimensional

Email addresses: `milani@dm.unipi.it` (Andrea Milani), `gronchi@dm.unipi.it` (Giovanni F. Gronchi), `zoran@aob.bg.ac.yu` (Zoran Knežević), `genny@pisces.eis.uva.es` (Maria Eugenia Sansaturio), `oscarr@eis.uva.es` (Oscar Arratia).

Preprint submitted to ICARUS *Submitted 21 February 2005*

space. The compromise value of the attributable, compatible with both sets of observations, can be used to define a preliminary orbit.

Since even two attributables may not be enough to compute an orbit with a convergent differential corrections algorithm, the preliminary orbit obtained as above is used in a constrained differential correction, providing solutions along the *Line Of Variations* which can be used as second generation VAs to further predict the observations at the time of a third arc [Milani et al. 2005a]. In general the identification with a third arc will ensure a well determined orbit. Additional sets of observations can be attributed to the orbit obtained so far with well known methods [Milani et al. 2001].

After testing these algorithms in a few significant cases, we have used a large scale simulation to measure the completeness, the reliability and the efficiency of the overall procedure to build up orbits by accumulating identifications. Under the conditions expected for the next generation asteroid surveys, the methods developed in this and in the preceding papers are efficient enough to be used as primary identification methods, with very good results. One important property is that the completeness in finding the possible identifications is as good for comparatively rare orbits, such as the ones of Near Earth Objects, as for main belt orbits.

Key Words: Celestial Mechanics; Asteroids, Dynamics; Orbits

1 Introduction

Astrometric observations of asteroids/comets are reported by the observers as *Very Short Arcs*, that is sequences of observations closely spaced in time and assumed to be of the same physical object. When, as in most cases, the information contained in such a data set is not enough to compute a full, six parameters set of orbital elements, we refer to them as them *Too Short Arcs (TSAs).* In such a case the problem of orbit determination must begin with the task of *linkage*, that is identification of two TSAs belonging to the same physical object. Such a 2–identification is, in most cases, enough to allow for an orbit, although it will be, in most cases, of very poor accuracy. Next we need to find 3–identifications, that is to *attribute* another TSA to the 2-identification orbit, and so on.

This way of thinking of the orbit determination as a procedure inextricably connected to the identification problem is a significant change with respect to the classical paradigm, going back to [Gauss 1809]. The fact is, the procedure used by modern surveys to discover asteroids/comets (and other small bodies) is very different to the one of ancient times, thus the classical methods are a solution of a problem different from the one we are facing now [Milani and Knežević, 2005].

This paper continues a line of research meant to establish a new paradigm of *population orbit determination* suitable to handle the observational data of the current and next generation surveys. In [Milani et al. 2004], hereafter referred to as Paper I, we have found the following properties of the TSAs. First, the essential information contained in most TSAs can be summarized by an *attributable*, that is in the fit to the measurements of two angles and of their time derivatives at a reference time (corresponding to the mean of the actual observation times). Second, for each attributable we can define an *admissible region*, a subset of the half plane of the undetermined coordinates range and range–rate where the orbits of solar system objects can be found, thus excluding satellites of the Earth, heliocentric hyperbolic orbits and tiny shooting stars. Third, we have established an efficient algorithm to sample the admissible region by means of a *Delaunay triangulation*.

This paper is organized as follows. In Section 2 we describe the procedure actually used to compute from a TSA the corresponding attributable and its uncertainty; we also discuss the issues of quality control and whether the TSA contains information beside the one expressed by the attributable.

In Section 3 we define the *attributable orbit elements* with their uncertainties, a set of values defining the initial conditions of one orbit with the two angles and the two angular rates of the attributable plus the range and range rate (with respect to the observer). We demonstrate that it is possible to generalize the definition of the covariance matrix in such a way that it becomes applicable to an orbit determined by using one TSA and one node of the Delaunay triangulation. In this way we define a set of *Virtual Asteroids (VAs)* sampling the space of orbits compatible with the available observational data.

In Section 4 we show that, given a VA with a generalized covariance, it is possible to compute a prediction for future/past observations with a formal uncertainty like the one provided by the classical Gaussian algorithm. Then in Section 5 we define a criterion, based on an *identification penalty*, to assess the likelihood that another attributable, computed from an independently detected TSA, actually belongs to the same object. By scanning the triangulation, that is the swarm of VA associated with the first TSA, we select the VA for which the identification penalty is low enough, if any. We discuss different possibilities to compute a preliminary orbit for each of these which can fit 2 TSAs, that is 6 parameters constrained by 8 scalar observations. [1]

[1] In an old discipline such as Celestial Mechanics it is hard to find something never considered before. In [Celletti and Pinzari 2005] we have found a reference to an (unpublished) work done by Mossotti in 1816–1818, solving the problem of determining an orbit with 4 observations. We still have to fully understand the

In Section 6 we show how to apply a *constrained differential correction* iterative algorithm to find a set of orbits fitting all the observations belonging to two TSAs. A constrained solution is essentially a five parameter solution, with one additional parameter left at an arbitrary value. In this way we extract, from the 2–dimensional swarm of triangulation nodes, a 1–dimensional swarm of solutions. The procedure can be repeated to attribute to some of the second generation VAs a third TSA: in this case it is possible, in most cases, to compute a full 6–parameter vector of orbital elements according to the classical least squares principle. To further attribute other TSA to the 3–identification orbit we can use methods already established and well tested. Thus we have, in principle, defined a new procedure for orbit determination; for a synthetic presentation of the new paradigm, see [Milani and Knežević, 2005].

In Section 7 we have tested the new algorithms on a simulated next generation survey. The results are very encouraging, thus in Section 8 we conclude that our method is suitable to be used as primary orbit determination method, entirely replacing the classical paradigm for the processing of the data of the present and future surveys. We also outline some of the work needed to apply these methods to realistic full-scale simulations of future surveys and to real data, when available.

2 Attributables

We assume that a celestial body is at the heliocentric position P and is observed from the heliocentric position $P_\oplus$ on the Earth. Let $(r, \alpha, \delta) \in \mathbb{R}^+ \times [-\pi, \pi) \times (-\pi/2, \pi/2)$ be spherical coordinates for the topocentric position $P - P_\oplus$.

We shall call *attributable* a vector $A = (\alpha, \delta, \dot\alpha, \dot\delta) \in [-\pi, \pi) \times (-\pi/2, \pi/2) \times \mathbb{R}^2$, representing the angular position nd velocity of the body at a time $\bar{t}$. The geocentric distance r and its rate $\dot r$ (that is, range and range–rate) are left completely undetermined by the attributable.

The angular coordinates (α, δ) are defined by a reference system selected in an arbitrary way. In practice we use for α the right ascension and for δ the declination with respect to an equatorial reference system (J2000).

connections with our methods.

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

$$t_i, \alpha_i, \delta_i, h_i \quad ; \quad i = 1, m \quad ; \quad m \geq 2$$

where α_i, δ_i are angles (RA, DEC), t_i times with $t_i < t_{i+1}$. Moreover, h_i are (optional) apparent magnitudes.

Note that $m = 1$ is not really used in modern astrometry: how does the observer know a moving object has been detected? [2] If the detection is based on a trail, then both ends should be measured and reported with beginning and end of exposure as times, provided it is possible to decide the sense of motion (if not, there are two possible attributables).

A sequence of observations is a Very Short Arc if the observations are known to belong to the same object not because an orbit has already been fit to all of them, but just because they can be unambiguously fit by some smoothing curve, typically a low degree polynomial. In other words, the observations are joined together by the observer, not by the orbit computer. A Very Short Arc is a unique entity, it cannot be split and should be reported at once. *Then it should have a unique name.*

To fix the terminology, given a Very Short Arc the *arc length* $\Delta\phi$ is the maximum of the angle [3] between the unit vectors in the directions defined by the celestial coordinates (α_j, δ_j) and (α_k, δ_k) for $1 \leq j < k \leq m$.

The *arc time span* is $\Delta t = t_m - t_1$. The *central time* $\bar{t}$ is the average of the observation times. If the observations have equal weights, $\bar{t}$ is just the arithmetic mean; if there are unequal weights w_i, $\bar{t}$ should be computed by a weighed mean where weights w_i take into account the RMS of both α and δ, such as $w_i = 1/(RMS(\delta)\ RMS(\alpha)\ \cos(\delta))$. Normally the observations from the same station at the same date have the same weight, thus $\bar{t}$ is a simple arithmetic mean in most cases.

The question can be asked whether it is possible to form a Very Short Arc with observations from two different observatories. Indeed, if the topocentric correction is small, that is if the distance r is large, it may be possible to fit

[2] At the times of Piazzi, Olbers and Gauss, asteroids were detected by comparison of the observations with a star catalogue. Thus individual observations of an asteroid were indeed possible and required an amount of work such that multiple observations in the same night were rare.

[3] Because of loops due to the apparent retrograde motion, the arc length may not correspond to the angular distance between the first and the last observation.

observations from two or more observatories, taken within a short time span, to some smoothing curve. However, if the object is close, this fit will be poor and the available information will not be represented by the smoothed curve, because it would wipe out the parallax information.

2.2 Computation of Attributables

Given a Very Short Arc, the method to compute the corresponding attributable, if $m \geq 3$, is as follows.

The first step is the fit of δ_i to a quadratic function of $t_i - \bar{t}$; then the angle δ appearing in the attributable is the constant term, $\dot{\delta}$ is the linear coefficient. An acceleration $\ddot{\delta}$ is also estimated. This fit takes into account the weights assigned to each individual observation, although as noted above in most cases the weight is the same for all the observation of a single Very Short Arc. A full covariance matrix for the variables $\delta, \dot{\delta}, \ddot{\delta}$ is available from the fit; we shall denote this 3×3 matrix by Γ_δ. The square roots of the variances along the main diagonal of Γ_δ are the RMS uncertainties $\sigma(\delta), \sigma(\dot{\delta}), \sigma(\ddot{\delta})$.

The second step is to project onto the tangent plane to the point (α, δ) on the sphere, that is we compute the coordinates $\beta_i = \alpha_i \cos\delta$. The data β_i are fitted to a quadratic function of $t_i - \bar{t}$; then $\beta = \alpha\cos\delta$ is the constant term, $\dot{\beta} = \dot{\alpha}\cos\delta$ is the linear coefficient. The acceleration $\ddot{\beta} = \ddot{\alpha}\cos\delta$ is also estimated, and the covariance of the variables $\alpha, \dot{\alpha}, \ddot{\alpha}$ is represented by the 3×3 matrix Γ_α, with the squares of the uncertainties $\sigma(\alpha), \sigma(\dot{\alpha}), \sigma(\ddot{\alpha})$ along the main diagonal.

When there are only two observations, a linear fit must be used, and only $\sigma(\delta), \sigma(\dot{\delta}), \sigma(\alpha), \sigma(\dot{\alpha})$ are available; moreover, if the weights are equal the correlations $Corr(\alpha, \dot{\alpha}), Corr(\delta, \dot{\delta})$ are zero, that is the covariance matrices are diagonal.

2.3 Quality control

The observations belonging to the sequence are known to belong to the same object even before it is possible to perform an orbital fit. In practice, this means that the quadratic fit used to compute the attributables should have residuals small enough. This condition could be violated if the arc length is too large, to the point that there are significant terms cubic in time (and of even higher degree). In such a case, the attempt to compress the information contained in the Very Short Arc into a single attributable cannot succeed, and the computation of the attributable should be abandoned. On the other hand, if the arc length is large and the higher derivatives are significant the data should be good enough for a conventional orbit determination, e.g., either Gauss's method or Laplace's method for the preliminary orbit: in such cases new algorithms are not necessary.

If $m > 3$ it is possible that one of the observations does not fit: as a matter of principle, it could be discarded as an outlier and the quadratic fit repeated. However, such a rejection procedure would give low reliability results. If such a misfit occurs we are entitled to suspect that either the observations do not belong to the same object, or the overall data quality is very poor.

If $m = 2$ the attributable can be computed by using a straight line, the second derivatives are estimated at zero (and have infinite uncertainty). To the data of this class no quality control can be applied, but the attributable can be used if the observations are known to be good.

In conclusion, the output of the polynomial fits are the following: the attributable four coordinates $(\alpha, \delta, \dot{\alpha}, \dot{\delta})$; the central time $\bar{t}$; the estimated second derivatives $(\ddot{\alpha}, \ddot{\delta})$ and the 3×3 covariance matrices $\Gamma_\alpha, \Gamma_\delta$. Optionally, an estimate for the mean apparent magnitude is the average $\bar{h}$ of the measured apparent magnitudes h_i.

2.4 Covariance matrix

As a result of the least square fits to compute it, an attributable A has an uncertainty formally represented by a 4×4 covariance matrix Γ_A, obtained from the two covariance matrices Γ_α and Γ_δ as follows:

$$
\begin{aligned}
\Gamma_A &= [\gamma_{ik}]_{i,k=1,4} \\
\gamma_{1,1} &= \gamma_{\alpha,\alpha} \quad \gamma_{3,3} = \gamma_{\dot{\alpha},\dot{\alpha}} \\
\gamma_{1,3} &= \gamma_{3,1} = \gamma_{\alpha,\dot{\alpha}} \\
\gamma_{2,2} &= \gamma_{\delta,\delta} \quad \gamma_{4,4} = \gamma_{\dot{\delta},\dot{\delta}} \\
\gamma_{2,4} &= \gamma_{4,2} = \gamma_{\delta,\dot{\delta}}
\end{aligned}
$$

with all the other components zero [4].

The matrix Γ_A defined in this way is positive definite [5]. We are using the 2×2 sub-matrices of $\Gamma_\delta, \Gamma_\alpha$, that is the marginal uncertainty of the attributable

[4] We are assuming the error model for the astrometric observations does not include correlation between the measured α and δ, otherwise the matrix could be full. Such a correlation could be large if timing was a significant source of error.

[5] Provided the observation times are different; with multiple observations at the same time some degenerate cases can occur.

whatever the value of the accelerations $\ddot{\alpha}, \ddot{\delta}$. If there are only two observations with equal weight Γ_A is diagonal.

Of course the formal covariance matrix Γ_A has a probabilistic interpretation in terms of multivariate Gaussian probability distribution if (and only if) the error model for the astrometric measurements of α, δ is also Gaussian. Moreover, we have assumed in the above description of the method to compute the attributable that the observation errors are uncorrelated, and this is not the case in the most advanced error models [Carpino et al., 2003]. The algorithms described above could be suitably modified to take this into account.

2.5 Curvature

Even for a Very Short Arc, for which the classical orbit determination algorithm fails, it cannot be assumed a priori that all the information from the available astrometry can be compressed into the attributable. This hypothesis needs to be tested on the real data available, by measuring the curvature of the best fitting curve on the celestial sphere.

Curvature is in fact a vector, whose components can be computed from $\ddot{\alpha}, \ddot{\delta}$ (also dependent upon the components of the 4-vector A). The subject of curvature will be discussed in detail in the next paper in this series, we only anticipate here that there is a rigorous test to be applied to the data to decide whether there is significant curvature information. Of course, if there is curvature information, it should be used in the orbit determination. What follows in this paper applies only to the case in which the curvature information is either not significant (with respect to the assumed astrometric error model), or it is too poor to provide useful constraints on the variables $r, \dot{r}$ which are left undetermined by the attributable. In such a case the Very Short Arc is indeed a Too Short Arc, as discussed in Section 1.

3 Attributable orbital elements

Given a TSA, after computing the attributable (and assuming there is no significant curvature information) we are left with a totally undetermined point in the $(r, \dot{r})$ plane. Following Paper I, we can assume that this point belongs to a *modified admissible region* of solar system orbits, and we can sample this compact region by a finite Delaunay triangulation. Each node of this triangulation defines a Virtual Asteroid, that is a possible, but by no means determined, set of six quantities [6] :

$$X = [\alpha, \delta, \dot{\alpha}, \dot{\delta}, r, \dot{r}]$$

A set of six initial conditions uniquely determines the orbit of an asteroid, thus we can consider it as a set of orbital elements, belonging to a new type of coordinates (different from the classical Keplerian, equinoctial, cometary, Cartesian, etc., coordinates). We shall call such data a set of *attributable orbital elements.*

3.1 Distance dependent corrections

To describe a set of orbital elements we need, besides the initial conditions, an *epoch time* t_0 (compulsory) and an *absolute magnitude* H (optional, only if there are photometric measurements together with the astrometric ones). The values of these quantities are not, however, coincident with the observation time $\bar{t}$ and the apparent magnitude $\bar{h}$ computed together with the attributable, but require corrections dependent upon the value of the distance r, described below.

Aberration and Epoch

As it is well known, an observation at time $\bar{t}$ of an asteroid needs to be corrected for aberration [7] . This because the light spends a significant time $\delta t = r/c$, with c the speed of light, to reach the observer from the asteroid. That is, the asteroid is observed at time t for its position at the time $t - \delta t$. Thus the epoch time to be associated with the set of attributable orbital elements is

$$t_0 = \bar{t} - \delta t$$

For different VAs with the same attributable, that is for different triangulation nodes (or anyway for different selected points on the $(r, \dot{r})$ plane) the attributable orbital elements have different epoch times, that is $t_0 = t_0(\bar{t}, r)$.

The approximations used in this formula are: (1) second order aberration terms are neglected; (2) if the distance changes significantly during the arc

[6] Five of these are measured by real numbers, while α is an angle, defined *mod* 2π; this is important whenever we compute a difference of two such vectors, e.g., the angles $\alpha = \pi - \epsilon$ and $\alpha = -\pi + \epsilon$ are close for small ϵ.

[7] Both planetary and stellar aberration, that is the computation must be performed by using the topocentric position vector of the asteroid.

time span, that is if $\dot{r}\,\Delta t$ is of the order of r, the δt should be computed separately for each observation, thus $t_0 - \bar{t}$ would not be simply a function of the six coordinates of the orbital elements set. Approximation (1) can be a problem only for extremely accurate observations, of the kind possible with space-borne astrometry; approximation (2) could be a problem only for objects very close to the Earth, in most cases well below the *shooting star limit* (see Paper I, Section 3).

Absolute magnitude from the apparent one

The equation describing the apparent magnitude h as a function of the absolute magnitude H has the form:

$$h = H + Z(G, \phi, r, r_\odot)$$

with G the opposition effect coefficient (in principle, a physical property of the asteroid; in practice, for a recently discovered asteroid, it is assumed at the common value 0.15), ϕ the phase angle, $r_\odot$ the distance from the asteroid to the Sun [Bowell et al. 1989].

Now both ϕ and $r_\odot$ are functions of the coordinates of the attributable, in particular of r, but not of $\dot{r}$. For small r (as discussed in Paper I, Section 3.1) the values of both ϕ and $r_\odot$ are weakly dependent upon r, and can be approximated with the value ϕ^0 and $r_\odot^0$ for infinitesimal r (the order zero term in r in their Taylor series expansion). For large r such an approximation would be rough, and the values of $\phi(r), r_\odot(r)$ would need to be computed with full precision.

Thus, to obtain a reliable result in all cases, we need to compute H as follows:

$$H = \bar{h} - Z(0.15, \phi(r), r, r_\odot(r)) \ ,$$

and to use this as estimate of the absolute magnitude to be attached to the attributable orbital elements.

The main case in which this estimate may be rough is again near the shooting stars boundary: $\bar{h}$ being just an average of the observed apparent magnitudes, if $\dot{r}\,\Delta t$ is of the order of r the correction Z should be computed separately for each observation, thus $H - \bar{h}$ would not be simply a function of the orbital elements.

The problem we would like to solve in this subsection is how to represent the uncertainty of a set of attributable orbital elements, assuming that they have been obtained from a given attributable. This case is quite different from the customary one, in which the uncertainty of a set of orbital elements is described by a positive-definite 6×6 covariance matrix, computed in the differential corrections process, that is by a fit to ≥ 3 observations well separated in time and in direction. For a TSA this is not available.

Among the attributable orbital elements, the first four coordinates are the attributable A and are indeed computed by a well determined least squares fit, thus they have a positive-defined 4×4 covariance matrix Γ_A. The last two coordinates are the point B on the $(r, \dot{r})$ plane, to be selected in the admissible region. Thus the question arises about what can be used as descriptor of the uncertainty of the full set of variables $X = [A, B]$.

Conditional Covariance Matrix

To describe the uncertainty of the attributable orbital elements $X = [A, B]$ we need to translate into a mathematical formalism the intuitive statement that the attributable A is measured, the point $B = (r, \dot{r})$ is just conjectured.

The inverse of the covariance matrix Γ_A, which is implicitly used in the least squares fit to compute A, is the 4×4 normal matrix C_A. This matrix is the *conditional normal matrix*, appearing, in a probabilistic interpretation, in the Gaussian probability density for the variables A **assuming** that B has a given value, that is assuming the selected VA; it can be formally built with the design matrix, giving the partials of the observations (α_i, δ_i) with respect to the 4 coordinates of A. Thus also Γ_A is the *conditional covariance matrix*[8]. We can formally define the *conditional covariance matrix* for the 6-vector X as the 6×6 symmetric 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 B subspace as kernel (null space). The 2×2 sub-matrix in the lower right hand corner is Γ_B, the fact that it is zero expresses the fact that the values of B have been assumed at some exact value, no

[8] The conditional covariance matrix is the inverse of the conditional normal matrix.

uncertainty. The companion matrix

$$C_X = \begin{bmatrix} C_A & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}$$

is the normal matrix in 6-space, equally conditional. C_X and Γ_X are not inverse of each other, but *pseudo-inverse*, that is Γ_X is indeed the matrix providing the least squares differential correction for X when B is constrained to a fixed value.

A non positive-definite covariance matrix, such as Γ_X, can be used in very much the same way as a conventional covariance matrix, with some caution, e.g., in computing the uncertainty of predictions, such as future observations. As an example, the covariance Γ_X can be propagated and/or transformed to a covariance matrix in some other coordinate system, e.g., Cartesian coordinates Y (the heliocentric position is just $P_\oplus + r\,\hat{R}$, where $\hat{R}$ is the unit vector pointing in the observation direction, and similarly for the velocity). Then, given the Jacobian matrix $\partial Y/\partial X$

$$\Gamma_Y = \frac{\partial Y}{\partial X}\, \Gamma_X\, \frac{\partial Y}{\partial X}^T \tag{1}$$

is also not positive-definite, with a 2-dimensional null space, containing the radial direction in both position and velocity.

Notation problems

In the formulas of this Section we have used so far a rather standard notation; from now on we will face the following ambiguity. A normal matrix and a covariance matrix are functions of the values of the variables for which they are computed. Of course the matrices used as output of the differential correction process are the ones *at convergence*, e.g., if the vector A has to be determined, and the nominal least squares solution is A_0, the normal matrix C_A must be computed by using the design matrix (Jacobian matrix of the residuals with respect to the coordinates of A) computed in A_0: then the notation should stress this, that is, we must always use the notation

$$C_A\,|_{A=A_0} \quad ; \quad \Gamma_A\,|_{A=A_0}$$

or at least the abbreviated version C_{A_0}, Γ_{A_0}.

A similar problem occurs for partial derivatives: confusion is possible between the variable, with respect to which derivation is performed, and the value assumed by the corresponding argument; we shall use the short notation:

$$\left.\frac{\partial A'}{\partial A}\right|_{A_0} = \left.\frac{\partial A'}{\partial A}\right|_{A=A_0} .$$

Quasi-Product Structure

As discussed in Paper I, for each value A of the attributable (at the mean observation time $\bar{t}$) we can define a (modified) *Admissible Region* $\mathcal{D}(A)$ in the plane of $B = (r, \dot{r})$, such that for $B \in \mathcal{D}(A)$ the attributable orbital elements $X = [A, B]$ belong to a solar system significant body, that is: the osculating heliocentric two body orbit is elliptic, possibly with limited semimajor axis a, the geocentric orbit is hyperbolic (if it is inside the sphere of influence of the Earth) and the absolute magnitude H is below some limit H_{max} (we are excluding from consideration small shooting stars). The set $\mathcal{D}(A)$ is compact, in most cases connected (up to two connected components can occur), and its boundary can be explicitly computed.

If we cannot determine the value of B from the observations (no significant curvature information), we can nevertheless assume that, if the exact value of the attributable is A, the value of B is contained in $\mathcal{D}(A)$. The existence of an observable real body with B outside $\mathcal{D}(A)$ is not impossible, but is either very unlikely (observable hyperbolic comets are rare) or outside the scope of our investigation (artificial satellites of the Earth and shooting stars of course do exist, but we are not interested in them).

Thus the confidence region describing the uncertainty of the attributable orbital elements $X = [A, B]$, given the attributable computed from a given TSA, is defined by

$$Z_X(\sigma) = \Big\{[A, B] \,\Big|\, (A - A_0)^T \, C_{A_0} \, (A - A_0) \leq \sigma^2 \text{ and } B \in \mathcal{D}(A)\Big\} \qquad (2)$$

where $\sigma > 0$ is a parameter, A_0 is the nominal (least squares) value of the attributable 4 angular coordinates, and C_{A_0} is the corresponding normal matrix. This set is not a Cartesian product, although in many cases it can be approximated by the Cartesian product of a confidence ellipsoid in the A space times the admissible region computed with the nominal attributable A_0:

$$Z_X^0(\sigma) = \Big\{A \,\Big|\, (A - A_0)^T \, C_{A_0} \, (A - A_0) \leq \sigma^2\Big\} \times \mathcal{D}(A_0) = \qquad (3)$$

$$= \Big\{[A, B] \,\Big|\, (A - A_0)^T \, C_{A_0} \, (A - A_0) \leq \sigma^2 \text{ and } B \in \mathcal{D}(A_0)\Big\} \qquad (4)$$

Sampling the confidence region

The practical problem is how to sample the confidence region $Z_X(\sigma)$ with a finite number of VAs. Our approach is to use the VAs corresponding to the nodes of a Delaunay triangulation of the admissible region $\mathcal{D}(A_0)$. If the triangulation nodes are the points $\{B^i = (r_i, \dot{r}_i)\}_{i=1,k}$ in $\mathcal{D}(A_0)$, then the orbits of the VAs are defined by the attributable orbital elements

$$\{X^i = [A_0, B^i]\} \; i = 1, k$$

(with epoch times $t_0^i = \bar{t} - r_i/c$).

This sampling is adequate for some predictions if and only if

(1) the sampling of $\mathcal{D}(A_0)$ by the nodes $\{B^i\}$ is dense enough;
(2) the uncertainty in the A subspace is not too large, and anyway is appropriately accounted for by the covariance matrix Γ_{A_0};
(3) $\mathcal{D}(A)$ is not too different from $\mathcal{D}(A_0)$ for values of A far from the nominal, but still inside the confidence ellipsoid for A.

All the above are hypotheses to be verified in concrete cases. Some parameters, such as the number of points in the Delaunay triangulation, can be adjusted to meet the requirements of the condition 1. Condition 2 refers to the reliability of the astrometric measurement error model [Carpino et al., 2003], condition 3 remains to be investigated.

4 Predictions from an Attributable

We would like to discuss how to compute a prediction, starting from a set of VAs, that is from a set of attributable orbital elements with uncertainty:

$$X^i = [A_0, B^i], t_0^i, H \quad ; \quad \Gamma_{X^i}$$

obtained as described in the previous Section.

4.1 Propagation of Conditional Covariance

The process of prediction consists of two steps: the first is the *orbit propagation* Φ from X_0 at the epoch time t_0^i to the prediction time t_1; this gives a set of orbital elements with uncertainty

$$Y^i, \bar{t}_1, H \quad ; \quad \Gamma_{Y^i}$$

with the new covariance matrix Γ_{Y^i} given by the equation analogous to (1). As already mentioned, the elements Y^i can be in a different coordinate system, e.g., Cartesian coordinates. It follows again from formula (1) that Γ_{Y^i} has *rank* 4, that is, it is not positive-definite with a 2-dimensional null space and with four linearly independent rows.

Projection on the Attributable 4-space

The second step is the computation of the *observation function* $F : Y^i \mapsto A^i$ with A^i in some space of dimensionality lower than 6; in this paper we are interested in the case that this dimension is 4 and A^i is an attributable, predicted at the new epoch [9] t_1. The Jacobian matrix of partial derivatives of the prediction function F is

$$DF(Y^i) = \left.\frac{\partial A'}{\partial Y}\right|_{Y^i}$$

a 4×6 matrix. *Generically* [10] this Jacobian matrix will have rank 4.

A formula similar to (1) for covariance propagation holds also for mappings between spaces of different dimensions, provided the rank of the Jacobian matrix is maximum [Jazwinski, 1970]

$$\Gamma_{A^i} = \left.\frac{\partial A'}{\partial Y}\right|_{Y^i} \Gamma_{Y^i} \left[\left.\frac{\partial A'}{\partial Y}\right|_{Y^i}\right]^T .$$

By using the covariance propagation formula (1), taking into account the zeros of the covariance matrix Γ_{X^i}, this formula implies

$$\Gamma_{A^i} = \left.\frac{\partial A'}{\partial X}\right|_{X^i} \Gamma_{X^i} \left[\left.\frac{\partial A'}{\partial X}\right|_{X^i}\right]^T = \left.\frac{\partial A'}{\partial A}\right|_{X^i} \Gamma_{A_0} \left[\left.\frac{\partial A'}{\partial A}\right|_{X^i}\right]^T . \quad (5)$$

The question is then what is the rank of the 4×4 matrix Γ_{A^i}. This question cannot be answered with certainty in all cases, but the following statements can be rigorously proven.

For $t_1 \to \bar{t}$, A^i has A_0 as limit, the transformation between the two attributables approaches the identity, thus $\Gamma_{A^i} \to \Gamma_{A_0}$, the original 4×4 sub-matrix of Γ_X. Since Γ_{A_0} has rank 4 for $t_1 - t_0$ small enough the rank of Γ_{A^i} is 4. However, we do not know how small $t_1 - t_0$ has to be for this to be guaranteed.

[9] Of course the aberration correction needs to be applied again.
[10] The precise mathematical definition of a generic property is not simple; we can describe it by saying that this occurs with probability 1.

Generically the rows of $\partial A^i / \partial X^i$ are linearly independent, and they do not belong to the null space of Γ_{X^i}. Thus *generically* Γ_{A^i} has rank 4. However, a matrix can be of maximum rank and still be *numerically degenerate* if its conditioning number [11] is larger than the inverse of the machine accuracy. If this happens, the matrix has an inverse in exact arithmetic, but the computation of the inverse is numerically unstable and requires the utmost caution.

Thus we expect, in almost all cases, the matrix Γ_{A^i} to be invertible. We can think of Γ_{A^i} as the *marginal covariance matrix* associated to the subspace A' of a set of attributable orbital elements $X' = [A', B']$. Indeed, the uncertainty of the attributable A' is computed without making any assumption on the non-measured quantities $B' = (r', \dot{r}')$. By the rule dual to the one used for the conditional matrices [12] the *marginal normal matrix*

$$C_{A^i} = {\Gamma_{A^i}}^{-1}$$

generically exists, but it may be difficult to compute. If the inverse matrix

$$M = \left[\left. \frac{\partial A'}{\partial A} \right|_{X^i} \right]^{-1} \tag{6}$$

exists, then C_{A^i} can be computed by the formula derived from (5)

$$C_{A^i} = M^T \; C_{A_0} \; M \; . \tag{7}$$

Thus it is possible in most (maybe not all) cases, to define a *confidence ellipsoid* for the prediction A^i in 4-space of the attributables A' for time t_1:

$$Z_{A^i}(\sigma) = \left\{ A' \left| (A' - A^i)^T \; C_{A^i} \; (A' - A^i) \leq \sigma^2 \right. \right\}$$

where $A^i = F(\Phi(X^i))$ is the prediction (corresponding to the assumption B^i). This is in fact the *inside* of a 3-dimensional ellipsoid in the 4-dimensional space of the attributables, where the second attributable is predicted to be, within a confidence level [13] described by the parameter σ.

[11] For a positive-definite matrix, one definition of conditioning number is the ratio of the largest to the smallest eigenvalue.
[12] The marginal normal matrix is the inverse of the marginal covariance matrix.
[13] If the error model is purely Gaussian, it is reliable and linearity applies, we can use σ^2 as χ^2 to estimate the probability associated to a given confidence region.

We can draw the conclusions from the discussion in this Section and give a definition of the *confidence region for the prediction* A^i even in the case we are discussing, that is when the first Very Short Arc is a TSA, with all the significant information contained in the attributable.

The confidence region for the attributable orbital elements derived from the attributable A_0 is $Z_X(\sigma)$ defined by eq. (2); we assume it can be approximated by the product $Z^0_X(\sigma)$ defined by eq. (4).

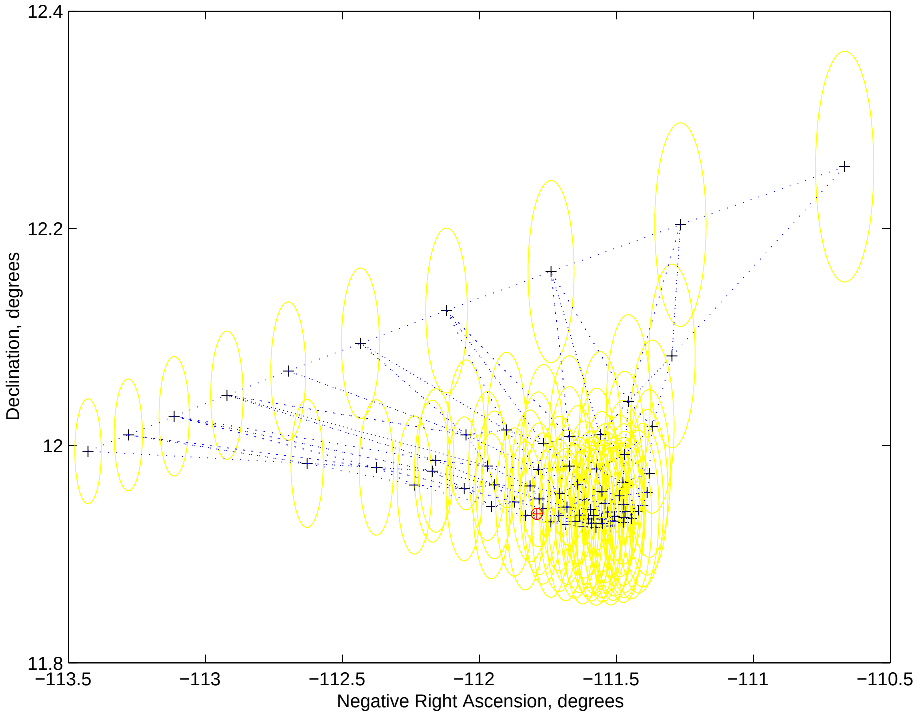

Fig. 1. For the asteroid 2003 BH_{84}, the observations 11 days after the discovery have been predicted in the *triangulated* form by using only the attributable computed with the observations of the discovery night. The ellipses indicate the projected uncertainty coming from the fit of the attributable. The $\oplus$ sign indicates the recovery attributable, computed with the actually observed data of the second night.

As a matter of principle, the image on the attributables space at time t_1 of the admissible region $\mathcal{D}(A_0)$ is a two dimensional manifold (compact, with boundary) $V = F\left(\Phi\left(\mathcal{D}(A_0)\right)\right)$. However, we have no way to explicitly compute this manifold as a function of $B = (r, \dot{r})$, because the map $X \longrightarrow A_1$ does not have an analytic expression (A_1 is the nominal attributable at second time). We can compute a triangulation of this manifold by using the image of the already computed triangulation $\{B^i\}, i = 1, k$ of $\mathcal{D}(A_0)$:

$$A^i = F(\Phi(X^i))$$

that is, the triangulation in the 4-dimensional observations space at t_1 has as nodes the predictions from the nodes of the triangulation in the X elements space (in turn defined by the nodes of the triangulation in the B space).

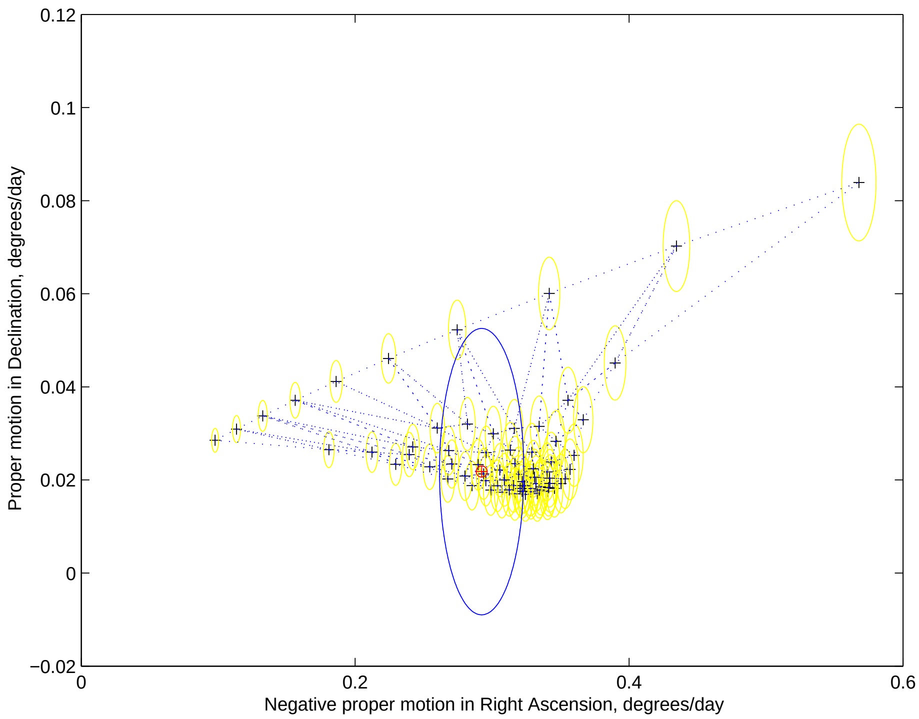

Fig. 2. The same triangulated ephemerides projected in the proper motion $(-\dot{\alpha}, \dot{\delta})$ plane. The largest ellipse indicates the projection of the uncertainty in the fit of the second attributable.

The idea of *triangulated ephemerides* was already discussed in Section 6 of Paper I. Here we are extending the same idea to a 4-dimensional predictions space, something more difficult to visualize, although some 2-dimensional projections can be used, especially on the planes $(-\alpha, -\dot{\alpha})$ and $(\delta, \dot{\delta})$, to have a good perception of the uncertainty of the attributable. Figures such as these can be used to assess the difficulty of a planned recovery.

In this paper we are going one step further, that is we associate with each node of the triangulated ephemerides its covariance. Geometrically, we have to think of each node surrounded by its own confidence ellipsoid; thus the projections, such as in Figures 1 and 2, would be surrounded by a confidence ellipse. This is an approximation to the *tubular neighborhood* $T(V)$ of the two-manifold V which would be obtained by the union of confidence ellipsoids centered in every point of V.

This tubular neighborhood, so difficult to be computed, plays an essential role in the problem of identification. Whenever we would like to obtain another attributable which could belong to the same object, we can scan either the sky with a telescope or an archive containing the astrometry already preprocessed.

In both cases, to decide if the objects are the same we need to assess not only how close is each observation to the predictions(s), but also if this discrepancy can be accounted for by the prediction uncertainty.

In the case of recovery, that is when planning what area in the sky has to be covered, the answer is simply that we need the covered area to include the projection on the celestial sphere of $T(V)$.

5 Linkage of two Attributables

The problem of asteroid identification has been classified into three main cases in [Milani, 1999]. The case we are going to discuss here is the one in which the available data consist only of a couple of TSAs, that is the significant information is contained in two attributables, A_0 at time $\bar{t}_0$ and A_1 at time $\bar{t}_1$. This implies that there is an orbit available neither for epoch $\bar{t}_0$ nor for epoch $\bar{t}_1$. However, by using the information from both A_0 and A_1 we can succeed in estimating an orbit; this kind of identification is called a *linkage*.

The attributables A_0 and A_1 have been computed from two sets of $m_0 \geq 2$ and $m_1 \geq 2$ observations. We are assuming that the attributables contain all the significant information, thus each of the two TSAs provides 4 equations in the 6 orbital elements. If the two data sets can be joined and proven to belong to the same object, we have at least 8 equations and the problem becomes overdetermined, thus a least squares solution can exist.

The problem would be quite simple, but for two main difficulties. First, the orbit propagation and prediction function are highly nonlinear, thus we can use the linearization of the problem, e.g., in the differential correction algorithm, only after determining a reasonably good first guess for the orbital elements. Second, the problem is of course not to analyze one couple of TSAs, one taken during the night of $\bar{t}_0$ and the other during the $\bar{t}_1$ night. In the modern surveys, many thousands of asteroid detections are reported every night of operation. In the near future, with the next generation surveys, we can expect this number to increase to somewhere between $100,000$ and 1 million per night.

Thus the problem can be formulated as follows: we need to find a way to decide which couples of attributables, one from the first night, another from the second, can belong to the same object, and we need to obtain some preliminary orbit for the identified object, roughly satisfying the observations from both nights [14]. Such preliminary orbits need to be *good enough* to be used as a

[14] This description is suitable for a main belt asteroid. For trans-neptunians, the attributables can be formed by a much longer arc, up to one month, and the reference to the two *nights* is not correct. For an object discovered near the Earth, the two

starting point for a differential correction procedure, which can converge and succeed in fitting in the least squares sense all the observations from the two TSAs. Last but not least, all this must be done with an algorithm of very limited computational complexity, to be repeated on millions (today) and trillions (tomorrow) of couples.

5.1 Identification penalty

The target functions of the separate fits for the attributables A_0 and A_1 are:

$$Q_0(A) = \frac{1}{4}(A - A_0) \cdot C_{A_0}\ (A - A_0) \tag{8}$$

$$Q_1(A') = \frac{1}{4}(A' - A_1) \cdot C_{A_1}\ (A' - A_1) \tag{9}$$

where C_{A_0} and C_{A_1} are the 4×4 normal matrices of the attributables, with central times $\bar{t}_0$ and $\bar{t}_1$, respectively [15] .

To test the hypothesis that the object is the same, we need to find a minimum for the joint target function, obtained from the weighed sum of squares of the discrepancies $A - A_0$ and $A' - A_1$

$$Q = \frac{1}{8}(4\ Q_0(A) + 4\ Q_1(A')) \tag{10}$$

under the assumption that there is a single orbit giving rise to the exact values A, A' at central times $\bar{t}_0$ and $\bar{t}_1$, respectively. To be able to speak of orbits, however, we have to assume the values of $(r(t_0), \dot{r}(t_0))$, that is, we need to select a VA

$$X^i = [A, B^i]$$

having for B component one of the triangulation nodes. Then we use the Jacobian matrix of the map $X^i \to A^i$ to constrain A in such a way that it can belong to the same modification of the VA giving A'

$$A' - A^i = \left.\frac{\partial A'}{\partial X}\right|_{X^i}\ (X - X^i) + \ldots$$

where the dots stand for higher order terms (this map is nonlinear).

We should not forget that $B = B^i$ is an assumption, not a measurement: we do not have an appropriate weight matrix to include a ΔB component in the

---

arcs may be separated by a few hours, thus they belong to the same night.

[15] There are no terms of degree higher than 2 because the two fits are linear.

differential correction. Thus we set $\Delta B = \underline{0}$ (not knowing any better) and the above equation becomes

$$A' - A^i = \left.\frac{\partial A'}{\partial A}\right|_{X^i} (A - A_0) + \dots \ .$$

Note that *generically* $\partial A'/\partial A$ is an invertible 4×4 matrix, thus we can use the inverse M defined in eq. (6) and write

$$A - A_0 = M\ (A' - A^i) + \dots \ .$$

and substitute into equations (9)

$$\begin{aligned} Q_0(A) &= \frac{1}{4}(A' - A^i) \cdot M^T\ C_{A_0}\ M\ (A' - A^i) + \dots = \\ &= \frac{1}{4}(A' - A^i) \cdot C_{A^i}\ (A' - A^i) + \dots \end{aligned}$$

where we have used the equation (7) for the propagation of the marginal normal matrix. Then we substitute in equation (10):

$$2\ Q = (A' - A_1) \cdot C_{A_1}\ (A' - A_1) + (A' - A^i) \cdot C_{A^i}\ (A' - A^i) + \dots$$

If the two attributables A and A' could be chosen independently, we could select $A = A_0$ and $A' = A^i$ and get a target function $Q = 0$. Thus the minimum value of Q we obtain under the assumption that the two are related (and that $B = B^i$) is in fact the *penalty*, measuring the increase in the target function which results from the identification. Neglecting the higher order terms, Q is the sum of two quadratic forms, generically positive-definite.

At this point the argument can proceed exactly as in [Milani et al. 2000], were we gave an explicit formula for the solution of the identification problem, the only difference being that we are working in a 4-dimensional space (rather than in a 6-dimensional space).

$$\begin{aligned} 2Q(A') &\simeq (A' - A_1) \cdot C_{A_1}\ (A' - A_1) + (A' - A^i) \cdot C_{A^i}\ (A' - A^i) = \\ &= A' \cdot (C_{A_1} + C_{A^i})\ A' - 2A' \cdot (C_{A_1}\ A_1 + C_{A^i}\ A^i) + \\ &\quad + A_1 \cdot C_{A_1}\ A_1 + A^i \cdot C_{A^i}\ A^i + \dots \end{aligned}$$

Neglecting the dots (terms of degree > 2), the minimum of the penalty Q can be found by minimizing the non-homogeneous quadratic form of the formula above. If the new joint minimum is A_1^i, then by expanding around A_1^i we have

$$2\ Q \simeq (A' - A_1^i) \cdot C_0^i\ (A' - A_1^i) + K^i$$

and by comparing the last two formulas we find:

$$\begin{aligned} C_0^i &= C_{A_1} + C_{A^i} \\ C_0^i\, A_1^i &= C_{A_1}\, A_1 + C_{A^i}\, A^i \\ K^i &= A_1 \cdot C_{A_1}\, A_1 + A^i \cdot C_{A^i}\, A^i - A_1^i \cdot C_0^i\, A_1^i \end{aligned}$$

If the matrix C_0^i, which is the sum of the two separate normal matrices C_{A_1} and C_{A^i}, is positive-definite, then it is invertible and we can solve for the new minimum point:

$$A_1^i = \left[C_0^i\right]^{-1} \left(C_{A_1}\, A_1 + C_{A^i}\, A^i\right) .$$

The computation of the *minimum identification penalty* $K^i = 2\ Q(A_1^i)$, as shown in [Milani et al. 2000], gives a simple expression of K^i as a quadratic form:

$$K^i = \Delta A^i \cdot C\, \Delta A^i \ .$$

where $\Delta A^i = A^i - A_1$ is the correction to be applied to the nominal attributable observed at time $\bar{t}_1$ and the matrix C is computed by one of the two alternative formulae

$$\begin{aligned} C &= C_{A^i} - C_{A^i}\ \left[C_0^i\right]^{-1}\ C_{A^i} = \\ &= C_{A_1} - C_{A_1}\ \left[C_0^i\right]^{-1}\ C_{A_1}\ . \end{aligned}$$

Note the above equations are true in exact arithmetic, but might be inaccurately fulfilled in a numerical computation if the matrix C_0^i is badly conditioned.

We can summarize the conclusions by the formula

$$2\, Q(A') \simeq \left[\Delta A^i\right]^T C\ \Delta A^i + (A' - A_1^i)^T\ C_0^i\ (A' - A_1^i)$$

which gives the minimum identification penalty $Q(A_1^i) = K^i/2$ and also allows one to assess the uncertainty of the identified solution for the attributable A', by defining a confidence ellipsoid with matrix C_0^i.

5.2 Scanning the Triangulation

It is important to realize that the identification penalty K^i, computed for a given node B^i of the triangulation of $\mathcal{D}(A_0)$, does not need at all to be small. First, we cannot know a priori whether the two asteroids observed at times $\bar{t}_0$ and $\bar{t}_1$ are indeed the same. Second, even if they are the same, the value of B^i

could be totally wrong with respect to the true values of the distance and its rate at time $\bar{t}_0$. In both cases the two attributables cannot fit, and this will be revealed by a large value of K^i.

As was shown in [Milani et al. 2000],[Milani et al. 2001], an identification procedure needs to be organized as a sequence of filters, each one selecting the couples candidate for identification with more and more strict conditions, and by using more and more computationally intensive algorithms.

The first filter we propose is based on the values of the penalty: given the attributable A_0 and the triangulation $\{B^i\}, i = 1, k$ of $\mathcal{D}(A_0)$, we scan the list of attributables of the second "night" $\bar{t}_1$. For each attributable A_1 we first compute the identification penalties $K^i, i = 1, k$. If they are all large, say $K^i > K_{max}$, then we discard the couple (A_0, A_1). If there are some nodes B^i, for some indexes $i \in I$, such that $K^i \leq K_{max}$, then we proceed to the next step, the computation of a preliminary orbit W_P^i for each $i \in I$. If a preliminary orbit is "reasonable", then we apply to it the following filtering stages, described later.

The value of the control K_{max} to be used is difficult to establish a priori, based only on an analytical theory. We cannot use the χ^2 tables for dimension 8, even in the assumption that the astrometric measurement error model is purely Gaussian and reliable (an already optimistic assumption). We are sampling the confidence region with a finite number of points B^i, thus we cannot assume that the minimum among the K^i is the absolute minimum we could get by trying all values of $B \in \mathcal{D}(A_0)$, that is

$$Min_{i=1,k} K^i \geq Min_{B \in \mathcal{D}(A_0)} K(B)$$

and we cannot compute analytically the safety margin to be left to take into account this difference. We conclude that the value of K_{max} to be used in large scale production of linkages can only be dictated by the analysis of the results of large scale tests.

As an example, we are considering the discovery and follow-up of the asteroid 2003 BH_{84} already used in [Milani et al. 2004], because it is an interesting case: a Near Earth Asteroid discovered very far from the Earth (almost 2 AU) during an experiment on the possibility of discovering very faint objects (with a 2.2 meter telescope) [Boattini et al., 2004].

Figure 3 shows the triangulation of the admissible region for the attributable computed by using only the 4 observations from the night of 25 January. Then we attempted the identification with another attributable, computed with the 3 observations of the night of 30 January, which are known a posteriori to belong to the same object. The sides of the triangles are color coded as follows. The nodes with $K^i \leq (0.6)^2$ are joined by continuous segments; the other sides

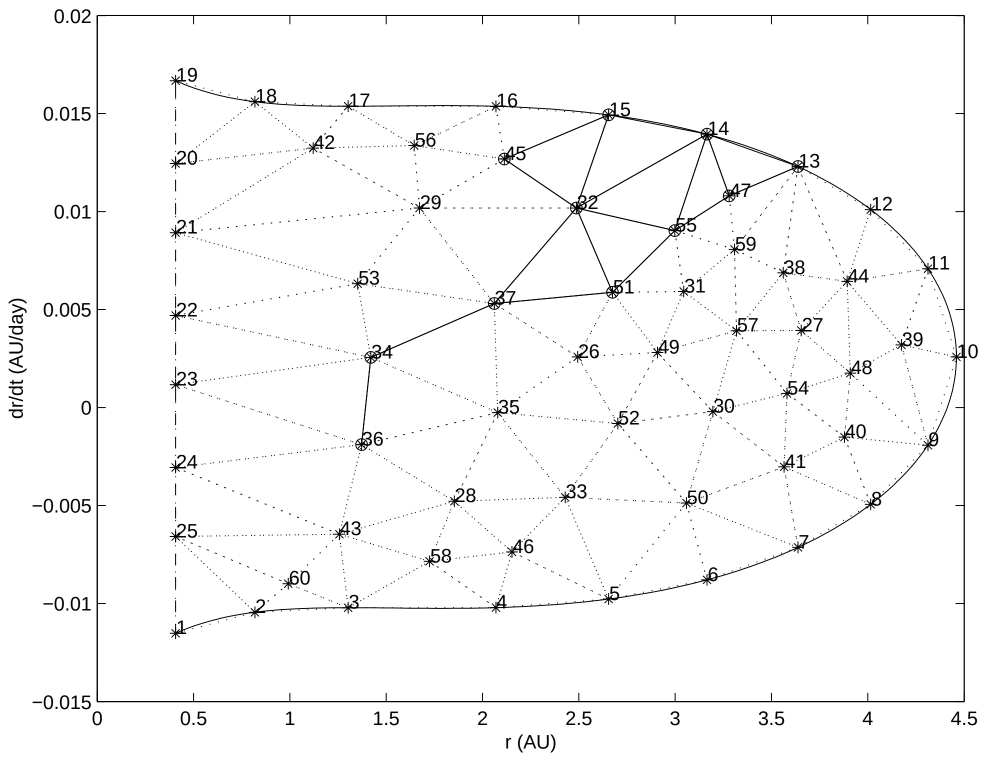

Fig. 3. The admissible region and its Delaunay triangulation for the attributable computed with the observations form the discovery night of $2003BH_{84}$. The continuous lines join the nodes with identification penalty $K^i < 0.6^2$.

of the triangles are dotted. Of course this example only shows that, when the identification is true, the values of some K^i (by no means all) can be small. The problem of the efficiency of this filter, that is the relative number of *false positives*, must be addressed by another, much larger test.

5.3 Selection of the Preliminary Orbits

The procedure described above provides us with a number of *best fitting corrected attributables* A_1^i, for $i \in I$, where I is a subset of $1 \leq i \leq k$. Each A_1^i comes with its penalty value K^i, which is not too large, that is, an orbit with B^i as distance and its rate at time t_0, and giving the attributable A_1^i as observation at time $\bar{t}_1$, can fit both A_0 and A_1 with not too large residuals; the fit is performed in the 8-dimensional space of the residuals of both attributables. We do not claim that this is a best fit, because we have used only the discrete set of points B^i rather than all the points $B \in \mathcal{D}(A_0)$.

To be able to start a differential correction process we need to compute a set of orbital elements to be used as first guess, by using a consistent set of six coordinates at the same epoch time. We have a number of options, the simpler ones being

(1) just use $X^i = [A_0, B^i]$, epoch time t_0^i.
(2) the attributable A_1^i and the value $B' = (r', \dot{r}')$ as computed for time $\bar{t}_1$ (from the orbit $X^i = [A_0, B^i]$ at $t_{0,i}$). The epoch is $\bar{t}_1 - r'/c$.
(3) the attributable back-propagated (linearly) to time $\bar{t}_0$, starting from A_1^i

$$A_0^i = A_0 + M\ (A_1^i - A^i)\ ,$$

where M is defined in eq. (6), and the value B^i of the node, epoch t_0^i.

Option 3 is the linear inverse image of the "best compromise" attributable A_1^i, which is at time $\bar{t}_1$, on the space of attributables at time $\bar{t}_0$. However, it can be shown that it is also the "best compromise" attributable in the space of attributables at time $\bar{t}_0$, in the linear approximation. That is, by converting the target function $Q_1(A')$ to a quadratic form in the space of A (by means of the linearized map), we could use the same algorithm, although with some additional complications, to find a minimum directly there. This minimum would coincide with A_0^i. This does not imply that option 3 is the same as option 2: even apart from the nonlinearity of the attributable transformation, the values of $(r, \dot{r})$ used in the two options correspond to different orbits.

The option 1 gives a preliminary orbit likely to be more inaccurate, thus we are using option 2. If the next step fails with the option 2 preliminary orbit, we try again with the option 3 preliminary orbit.

More complicated options may involve the use of additional propagations between the times $\bar{t}_0$ and $\bar{t}_1$, maybe even back and forth, in an iterative loop. According to the tests discussed in Section 7, such additional complications do not appear necessary.

6 Multiple Solutions from two Attributables

The next step should be to compute, starting from the preliminary orbits of the previous Section, least squares solutions. However, the observational data available are still very limited, amounting to only two TSAs (just enough to compute two attributables). This implies that the nominal orbit, according to the least squares principle, may not exist, may be impossible to find with the classical differential corrections procedure, and anyway will typically be very poorly determined. Indeed, the orbit determination procedure cannot be considered complete until at least a third attributable can be identified with the other two. Thus the least squares solutions at this stage are themselves only intermediate orbits, to be used to allow additional identifications.

A suitable algorithm to handle these cases, specially effective with “two nighters”, has been presented in [Milani et al. 2005a]. Here we will only recall the computational procedure, without details to be found in that paper.

6.1 Constrained solutions

Given a set of M observations, and some first guess value X for the orbital elements, we can compute the corresponding observation residuals Ξ (an $m = 2M$ dimensional vector) with their weight matrix W (an $m \times m$ symmetric matrix) and the normal matrix C at X:

$$B(X) = \frac{\partial \Xi}{\partial X}(X) \quad ; \quad C(X) = B(X)^T \, W \, B(X) \ .$$

Let $\lambda_j(X), j = 1, \ldots, 6$ be the eigenvalues of $C(X)$, with $\lambda_1(X)$ the smallest one; let $V_1(X)$ be an eigenvector with eigenvalue $\lambda_1(X)$, that is

$$C(X^*) \, V_1(X) = \lambda_1(X) \; V_1(X) \ ;$$

let $H(X)$ be the 5-dimensional hyperplane orthogonal to $V_1(X)$

$$H(X) = \{Y | (Y - X) \cdot V_1(X) = 0\} \ .$$

One step of *constrained differential correction* is the linearized correction to X to approach the minimum of the cost function

$$Q = \frac{1}{m} \, \Xi^T W \, \Xi$$

restricted to the hyperplane $H(X)$. After the correction $X' = X + \Delta X$, with $\Delta X \in H(X)$, has been applied, the observation residuals are recomputed and the new normal matrix $C(X')$ is computed. Then the new hyperplane $H(X')$ is used as new constraint and the correction is repeated, until convergence (the constrained correction becomes negligible) to the point $\overline{X}$.

The point $\overline{X}$ has the property of having the gradient of the cost function Q parallel to the eigenvector $V_1(\overline{X})$, that is it belongs to the *Line Of Variations (LOV)* of the least squares problem.

The LOV definition depends upon the coordinates used for the initial conditions X; for asteroids with few observations spanning a small arc on the celestial sphere, the use of Cartesian coordinates is recommended, although the attributable elements can also be used (with essentially equivalent results). For a detailed discussion of the dependence upon coordinates and metric see [Milani et al. 2005a].

In the previous Section we have shown how to compute a set of preliminary orbits X^i starting from a subset of the triangulation nodes, satisfying the condition of moderate identification penalty K^i. From each initial guess X^i we can start a constrained differential correction process, which will converge (in some cases) to a LOV point $\overline{X^i}$.

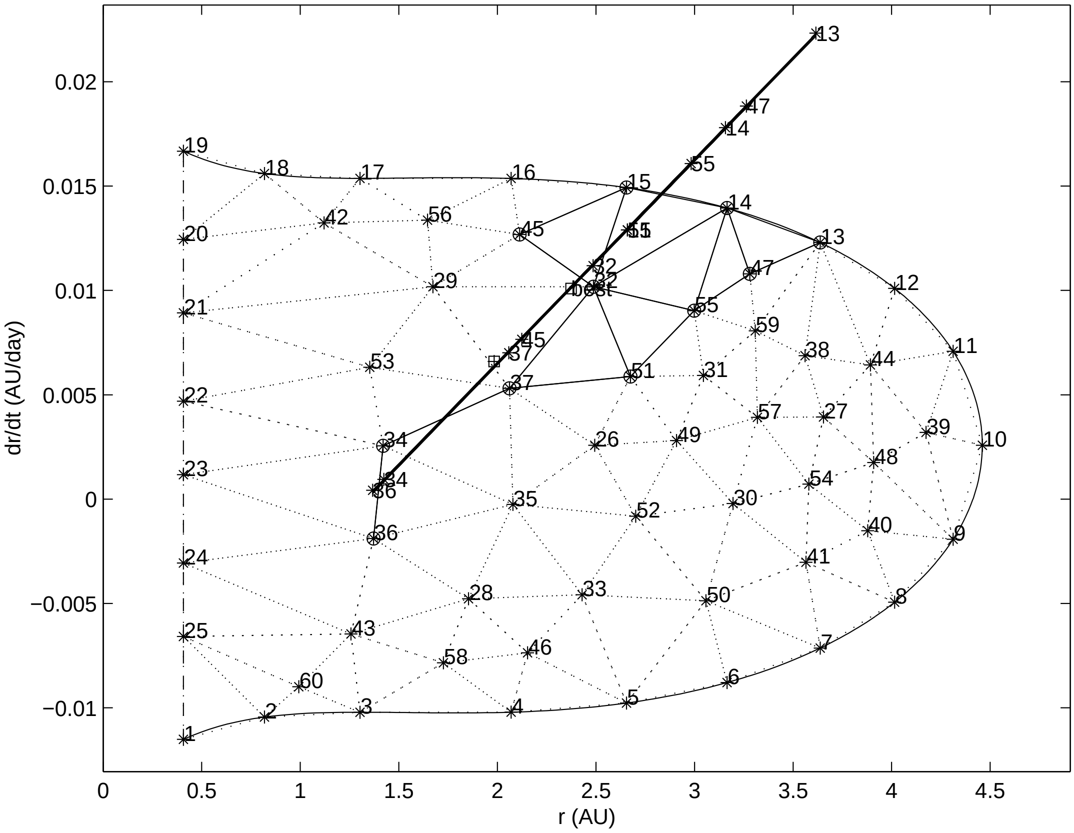

Fig. 4. Attributable from the discovery night of 2003 BH_{84} identified with the attributable of 5 days later. The continuous lines join the nodes with identification penalty $K^i < 0.6^2$, which have been used to compute preliminary orbits: from each of them we have started a constrained differential correction procedure and a LOV point has been determined if the solution converged.

We need to stress once more that this procedure depends upon the coordinate system used. As an example, in Figure 4 we show the case of the attributable computed from the discovery night of 2003 BH_{84} identified with the attributable formed with the observations of a night 5 days later. The diagonal line represents the LOV, on which the points $\overline{X^i}$ are marked with the triangulation index i. The nodes number 13, 14, 47 and 55 of the triangulation, which belong to the admissible region and therefore correspond to elliptic orbits, provide (with the procedure described above) LOV points well outside the Solar System boundary. That is, when the Cartesian coordinates used in the constrained differential corrections are converted to elements, the eccentricities are > 1. If the constrained differential corrections had been performed in elements singular for $e = 1$, such as Keplerian or equinoctial, the number of LOV points obtained would have been smaller.

In this case a nominal least squares solution exists and can obtained with unconstrained differential corrections starting from some of the LOV points; it is marked "best" in the Figure, and it is close to the LOV solution with index 32. However, the true solution (known a posteriori, that is by using also the data from a third night of observations) is marked by a crossed square sign, and is much closer to the LOV solution with index 37. This is a good example of the fact that the nominal solution, even if it exists, does not need to be an approximation of the true solution better than the other LOV solutions.

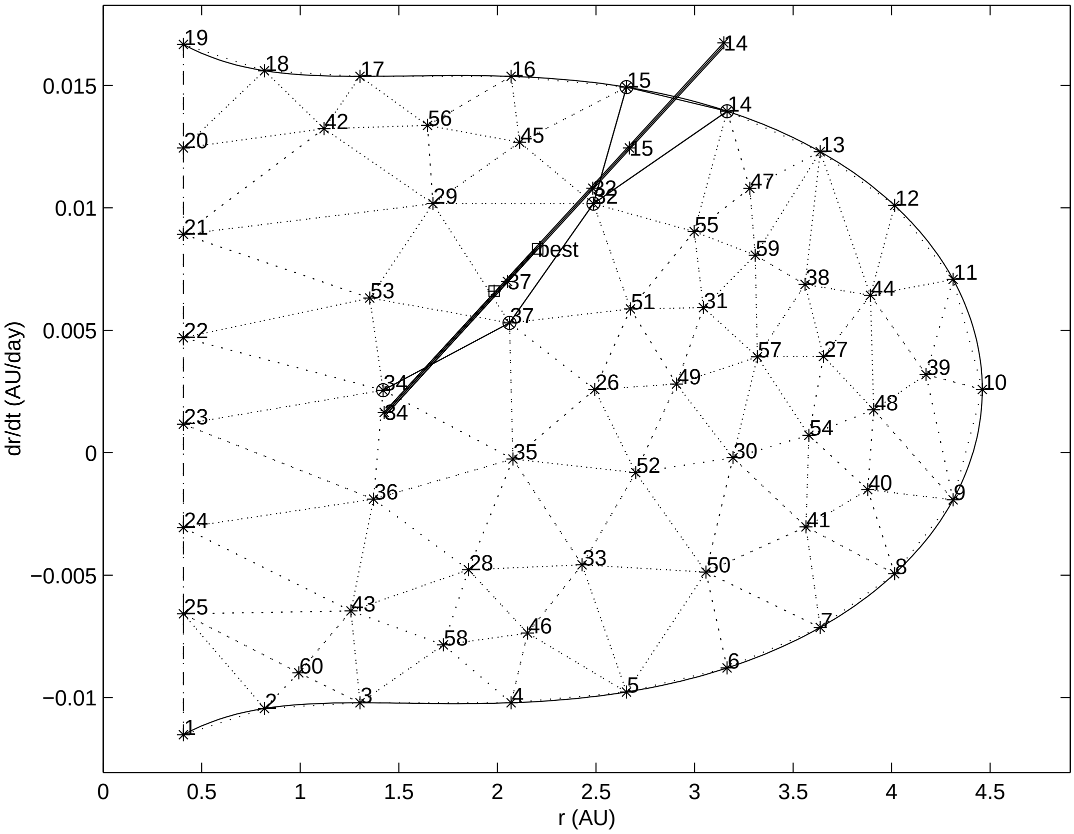

Fig. 5. Attributable from the discovery night of 2003 BH_{84} identified with the attributable formed with the observations performed 11 days later.

In Figure 5 we show, for the same asteroid, the identification of the discovery night attributable with the attributable based on the observations of a single night 11 days later. This case is slightly more difficult, and indeed the number of triangulation nodes with moderate identification penalty K^i is reduced to 5. All these 5 have provided, by convergent constrained differential corrections, LOV solutions; a nominal solution could also be computed. But the true solution is again closer to the LOV point obtained from the triangulation node number 37: indeed, the values of $(r, \dot{r})$ of the node B^{37} were the closest ones to the real values for the asteroid at the discovery time.

The above are of course artificial examples, obtained by splitting into Very Short Arcs the observations already known to belong to the same asteroid. However, we need to point out that 11 days is already a long interval for asteroid discovery surveys. Let us suppose that the data of the second night, 5 days after the discovery, were not available, or maybe were available but had not been identified with the discovery attributable. Then it would have been possible to recover the same object by conducting a scan of the region shown in Figure 1. It would also have been possible, if the data of 11 days after the discovery had been found by a survey without any knowledge that they belonged to the same object, to identify them. If this example is representative, a survey scanning large portions of the dark sky with repeat cycle as long as a week (and even longer) could successfully identify the Near Earth Asteroids detected. Of course this tentative conclusion needs to be confirmed by a statistically significant set of examples.

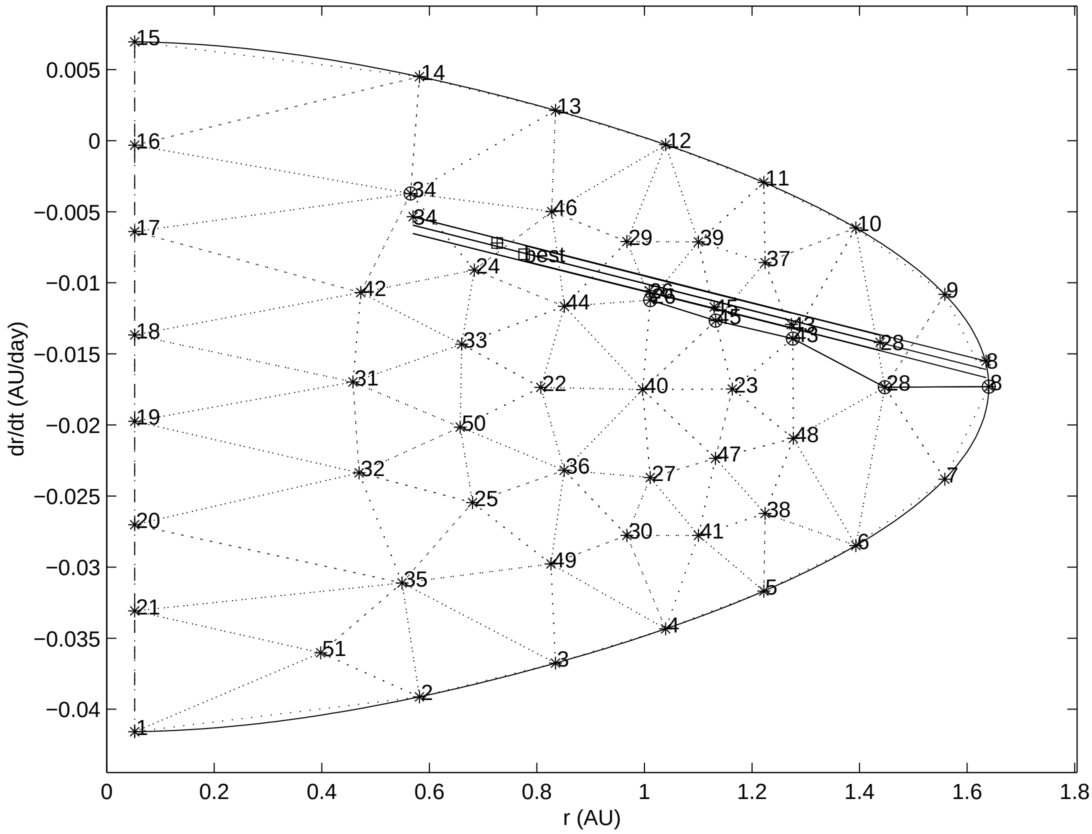

Fig. 6. Attributable from the November 25 precovery of 1998 XB identified with the attributable from the official discovery night, December 1. The continuous lines join the 5 nodes with identification penalty $K^i < 1.5^2$. The three parallel straight lines indicate a linear fit to the LOV and its uncertainty.

As a second example, we are using the asteroid 1998 XB. This object has been a challenge for orbit determination because it has been discovered at an elongation of 90°. This resulted in multiple solutions for Gauss' preliminary orbit, in multiple minima for the least squares fit [Milani et al. 2005a] and in very strong nonlinearity in all the orbit and observations predictions.

We use the attributable computed with the data of the first night of observations in the year 1998, the 25 November [16]. We have computed and triangulated the admissible region for the first night, and then tried to perform an identification with a second night of observations on 1 December. Figure 6 shows that a well defined LOV has been computed, with many constrained solutions; however, in this case the control value K_{max} for the penalty was 1.5^2. The nominal solution ("best") is very close to the true solution (crossed square).

The presence of three parallel lines indicating the LOV in Figure 6 can be explained as follows. Of course we do not know all the points on the LOV, but just the ones marked $(*)$ and labeled with the triangulation node index. The central line is obtained by linear regression, that is we assume that the projection of the LOV on the $(r, \dot{r})$ plane is a straight line. The two side lines indicate the uncertainty of the fit; that is, in this case there is a visible, although small, curvature. In the two previous figures there was no curvature visible to the human eye, and indeed the three lines were superimposed. We do not fully understand why the LOV is so well approximated by a straight line in $(r, \dot{r})$, although it is obvious, also from other experiments in orbit determination, that the attributable elements are superior, that is they suffer less from nonlinearity effects for short arcs.

In Figure 7 we are pushing the test much further. We are attempting to identify the attributable from the night of November 25 with another attributable, based only upon the data of December 26. That is, we assume that either no other observations were available, or none of them was identified with the same object, for a time span of one month. In this case we had to significantly increase the penalty control value K_{max} to 5^2. We were able to compute 5 LOV points; in this case a nominal solution was not found [17].

The third example is based on the discovery observations for the Centaur *(31824) Elatus*. The object was discovered on October 29, 1999 by the Catalina Survey, and the designation 1999 UG_5 was granted after follow up observations for 6 consecutive nights. Later a number of precovery observations were

[16] These observations were not credited as a discovery of 1998 XB, because they remained unidentified. The same object was rediscovered (from Beijing Observatory) on 1 December, then followed up by the same observatory on 2 December. Thus the designation and discovery were credited to Beijing Observatory.
[17] It may well exist, but be too far from the LOV solutions we have used as starting point for the unconstrained differential correction.

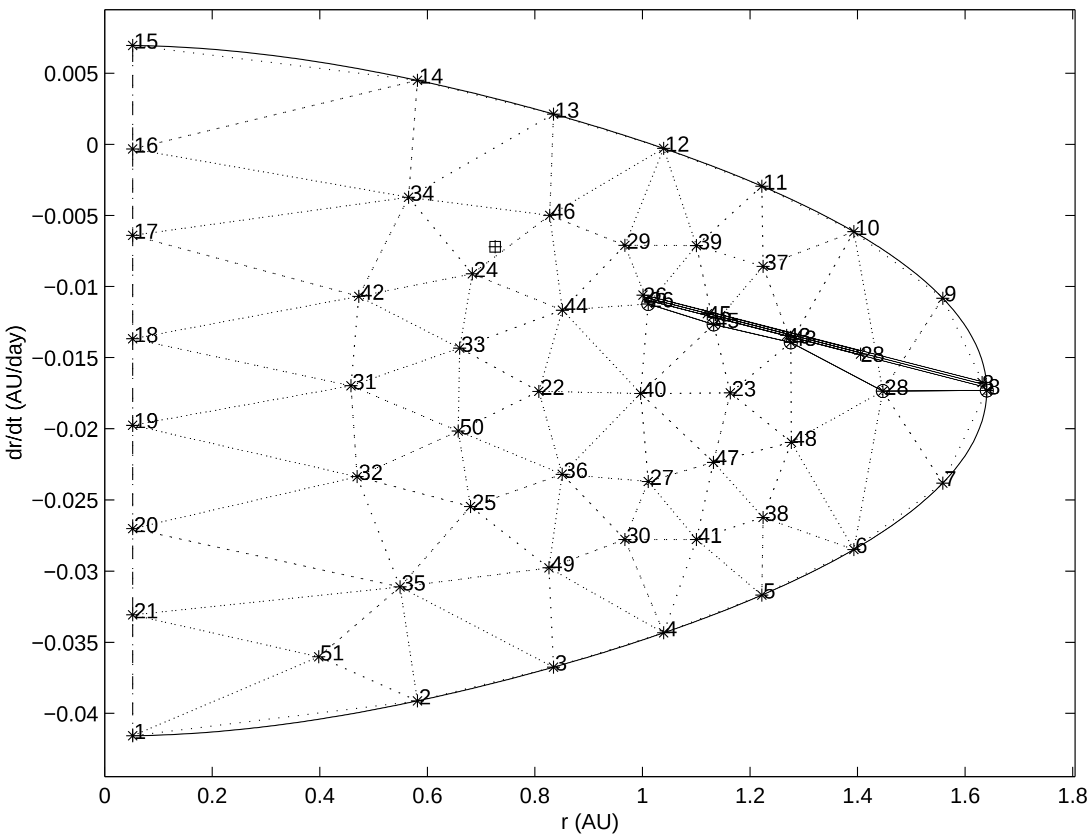

Fig. 7. Attributable from the November 25 precovery of 1998 XB identified with the attributable from the night of December 26. The continuous lines join the nodes with identification penalty $K^i < 5^2$. Note that a nominal solution has not been found, but still some portion of the LOV has been identified. The true solution (crossed square) is very close to the LOV, although not close to any of the LOV points computed.

discovered in the archives of One Night Stands (ONS), maintained, but at that time not yet published, by the Minor Planet Center (MPC). Later the asteroid was numbered and named.

We have selected the four discovery observations from Catalina, spanning just $\simeq 50$ minutes of time on October 29, 1999 and computed the attributable, the admissible region and its Delaunay triangulation, shown in Figure 8. The admissible region has in this case two connected components, one corresponding to Centaur type orbits, the other to much closer orbits.

Then we have selected one of the precoveries, the one by LONEOS on October 17, 1998, formed the corresponding attributable and computed the identification penalties. We had to use a rather high value $K_{max} = 5^2$, but we were able to obtain 4 LOV solutions, out of which one was hyperbolic (the one with index 7). The best solution could be found, by using Cartesian coordinates, but it also corresponds to a hyperbolic orbit. In this case, by combining the

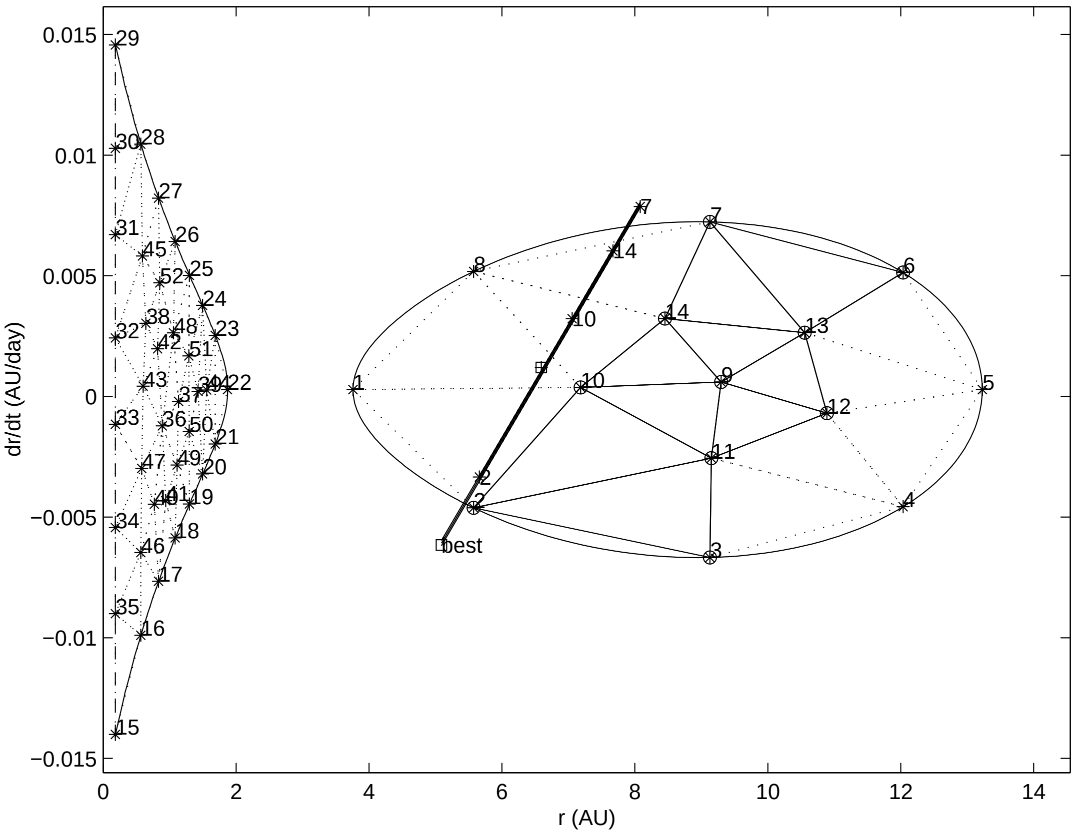

Fig. 8. (31824) 1999 UG_5 4 discovery observations; identification with ONS 1 year earlier. The continuous lines join the nodes with identification penalty $K^i < 5^2$. The nominal least squares solution, marked "best", is hyperbolic.

information from the observations with our a priori knowledge of the population densities, we can conclude that the initial conditions near the nominal solution are very unlikely! Note also that the values of the penalty are never small (in fact, they are huge) for all the nodes belonging to the connected component of the admissible region closer to the Earth. That is, the precovery data cannot have anything to do with 1999 UG_5 unless it is a Centaur.

The last two examples were still manufactured with the a posteriori knowledge that the observations do belong to the same object: as genuine identifications they would be paradoxically difficult. To detect a fast moving object and to ignore it for more than one month would be an example of observational malpractice. To scan the archives, going back one year, for a precovery of an object suspected to be a Centaur and observed for a single night, appears a forlorn hope. Nevertheless, we can find constrained solutions even in these extreme cases.

6.3 *Attributions of a third Very Short Arc*

The next step is to find a third attributable belonging to the same object of the the two already identified. The *attribution* of a set of observations, for which a set of orbital elements is not available, to another discovery for which there are enough observations to compute a nominal least squares orbit has been discussed in [Milani et al. 2001]. In our case the least squares orbit is replaced by multiple solutions along the LOV, as in [Milani et al. 2005a]. We do not repeat here the formulas, which result in the computation of an *attribution penalty* K_A^i for the third attributable identified with the orbit from the LOV point of index i obtained from the first two attributables. K_A^i has essentially the same interpretation as the K^i computed to identify the first two, that is, it is used to filter out the less likely identifications; if the value is below a control value K_A^{max} the triple identification is tested by differential corrections. If the full least squares fit with all the data succeeds, with low RMS of the residuals, the triple identification is confirmed and a comparatively well determined orbit is available.

As a first test, we use the same three "artificial" cases as in the previous Section.

For 2003 BH_{84}, after identifying the data of the discovery night with the ones taken 5 days later, we have tested the attribution of the data from the third night, 11 days after the discovery. We obtain a triple identification, with convergent full differential corrections and very small residuals, using as first guess the constrained orbits obtained from nodes 34, 36, 37, 45. Note that differential corrections are divergent when starting from the nominal orbit, and also from the constrained orbit from node 32, which is very close to the nominal one (see Figure 4). This is a comparatively easy case, in which the triple identification could have been obtained also by other well known methods; however, it is not a trivial case, since the identification cannot be obtained always, in particular cannot be obtained from the nominal solution. We have checked that by changing the order of the three attributables does not change the result, e.g, it is possible to identify the first night with the third, then with the second.

For 1998 XB, we have first identified the precovery data of November 25 with the ones of the official discovery night (December 1), as in Figure 6, then attempted the triple identification with the December 26 attributable. We have convergent full differential correction starting from the constrained solutions obtained from nodes 26, 34, 45 and also from the nominal solution.

For 1999 UG_5 we have first formed the identification of the discovery attributable with the 1998 precovery, as in Figure 8. Then we have attempted the triple identification with another precovery attributable of September 14, 1999. This case proved to be very difficult: with the option 2 preliminary orbits the identification was not found. It was found with the option 3 preliminary orbits and by using triangulation with more nodes. Thus this case can be handled only by a special effort, both in terms of the computational cost and of the human effort required to test different options. It is likely that extreme cases such as this one cannot be handled by routine, automatic processing but only as special efforts to follow up some particularly interesting discoveries.

The problem we cannot address with the "hindsight tests" like the three above is the number of spurious identifications which would be proposed by a massive use of our procedure, as well as the number of real identifications hidden in the data and not detected with these methods. This large scale test is the subject of the next Section.

7 Large scale test

After careful consideration, we have decided to use a large scale simulation to test the performance of our identification algorithms. Using real data, that is real TSA so far not identified, would have the following disadvantages:

(1) The non identified very short arcs are not entirely public. [18] Even if they were all available to us, the numbers would be representative of the state of the art in asteroid surveys rather than of the future needs.
(2) The identifications which can be found with well known algorithms have already resulted in removal of the corresponding very short arcs. Thus we could only measure marginal improvements, with respect to the identification procedures already applied to the data, rather than the overall performance of our new algorithms.
(3) The real asteroid astrometric data are of very uneven quality. This can be somewhat compensated by the use of a complex error model as the one described in [Carpino et al., 2003], but a fraction of the data is "bad", to be discarded, even by the standards of the observing site. Moreover, the file of non identified observations (the so-called One Night Stands file of the MPC) contains data which have been submitted to less rigorous quality control, and indeed some of them are contradictory, even bizarre.
(4) With real data, we have no way to assess the completeness of the obtained list of identifications. We can only say we have found identifications additional to the ones found by others before us, we have no way to guess

[18] This is due the usual discussion on open data policy versus proprietary rights. See resolution B.1 of the 25^{th} General Assembly of the IAU, Sidney 2003, IAU Information Bullettin 94, January 2004.

how many real identifications remain hidden in the data.

On the contrary, the use of a simulation has the corresponding advantages:

(1) The next generation surveys are expected to generate astrometric data at a rate two orders of magnitude larger than the current ones. Thus only a large simulation of a future survey can test the algorithms under conditions similar to the ones in which they will soon be used. This is especially important because of the presence of effects growing quadratically in the number of bodies observed (see Subsection 7.2).
(2) We would like to assess the performance of our algorithms in case they were used as the primary methods for identification and orbit determination for one or more surveys.
(3) In simulated data, we can assume that the error model (for both astrometry and photometry) is well known, and indeed the errors are added in by using random numbers with known probability densities. This assumption is of course optimistic: it allows us to separate the problem of quality control from the problem of identification, although in a real survey the two problems will have to be solved at once.
(4) The most important advantage of a simulation is that we know the "ground truth", that is, we have the list of objects which have been included in the simulation with their assumed orbits. The identification algorithms cannot use this *a priori "secret"* information, but after the list of identification has been generated we can compare it with the ground truth and find how many have been missed and how many are wrong.

It is our goal to apply our new algorithms to real data, but we want first to be convinced that they are reliable and computationally efficient, to the point that using them as the primary orbit determination method is justified.

Thus we have asked the team of one of the next generation surveys, Pan-STARRS [19], to provide us with a simulation of the moving objects data which could be obtained in one month of regular operation. Robert Jedicke of the Pan-STARRS project has suitably reformatted one of their simulations, including $\simeq 1,000,000$ objects observed for 4 separate nights, with a 4 days interval, and has made available both the set of simulated observations and the catalog of assumed objects with their orbits. Note that the purpose of the exercise is to assess the performance of our algorithms, not the performance of Pan-STARRS. The expected performance of a survey depends upon many assumptions, the most important being the detection model and the observations scheduling: the very preliminary simulation we are using was based upon a simplified detection model and did not even try to select an optimal scheduling.

[19] http://pan-starrs.ifa.hawaii.edu/public/

What matters at this stage is that the simulation has a number of observations of the right order of magnitude and they are obtained from a population model of Main Belt Asteroids (MBA) and Near Earth Asteroids (NEA). Trojans, Centaurs, Trans-neptunians and comets were not included, thus an efficient identification of objects belonging to these populations will require further optimization of the orbit determination procedure.

7.1 How to measure success

The properties of our identification and orbit determination procedure we want to measure are *completeness, reliability and efficiency.*

We assume the simulation consists of a finite number K of observing nights, in our case $K = 4$. We also assume that identifications are searched only for consecutive [20] observing nights; note the search can go both forward and backward in time. Let the simulated data set contain m_{tot} TSAs, belonging to n_{tot} objects, out of which $n(k)$ objects observed for exactly k consecutive nights (k–nighters), $n_+(k)$ objects observed for at least k consecutive nights (h–nighters with $h \geq k$), and $N_{id}(k)$ possible identifications joining the TSAs of k consecutive nights. If there are objects observed over $> k$ nights, that is $n_+(k) > n(k)$, then $N_{id}(k) > n_+(k)$ because each $(k+1)$–nighter can provide 2 identifications with k consecutive nights.

Having found $id(k)$ k–identifications (joining k TSAs in consecutive nights), by using the "ground truth" of the simulation we know we have found $id_T(k)$ true and $id_F(k)$ false identifications. For the identifications at the level k (for $k = 2, \ldots, K$) the ratio $compl(k) = id_T(k)/N_{id}(k)$ measures the completeness and the ratio $wrong(k) = id_F(k)/id(k)$ measures the (lack of) reliability.

For $k < K$, the difference $N_{id}(k) - n_+(k)$ indicates that there are objects for which several different k–identifications are possible. For these h–nighters, with $h > k$, we do not really need to find all possible k–identifications: it may be enough to find one of them to have the possibility to find a $(k+1)$–identification for that object, and so on up to h–identifications. Indeed, by achieving a large value of $compl(k)$ we may introduce duplications in finding h–identifications with $h > k$ and this redundancy may decrease the computational efficiency.

We would like to measure also the completeness of the procedure in terms of the number of simulated objects for which a more or less good orbit has been computed. Thus the overall completeness should be measured by the global number $\mathfrak{ID}(k,k)$ of true k–identifications of k-nighters, after removal

[20] This restriction can be removed later; see Section 8.2.

of all the duplications and contradictions contained in the merged list with $id(4) + id(3) + id(2)$ identifications. Reliability should be measured by the total number $\mathfrak{Wr}(k)$ of false k–identification we have not been able to remove by comparing with the others. Moreover, we need to take into account the number of true but incomplete identifications $\mathfrak{Id}(k, h)$, that is the number of k–nighters for which we have found only h–identifications, with $1 < h < k$. $\mathfrak{Id}(k, 1)$ is the number of total failures, that is the number of k–nighters for which we have found no identification at all.

To illustrate these definitions in the simplest example, let us suppose there are 6 Too Short Arcs in 4 observing nights: A and E in night 1, B and F in night 2, C in night 3 and D in night 4; let the ground truth list of objects include A=B=C=D and E=F. Then $n(4) = 1$ and $N_{id}(4) = 1$ (the only 4–identification possible is A=B=C=D), $n(3) = 0$, $n_+(3) = 1$ and $N_{id}(3) = 2$ (A=B=C and B=C=D), $n(2) = 1$, $n_+(2) = 2$ and $N_{id}(2) = 4$ (A=B, B=C, C=D, E=F). Note we indicate an identification by listing all the TSAs in order of time, with an equal sign as separator; however, the identifications may be found in a sequence not respecting the order of time, e.g., A=B=C can be found both from A=B going forward and from B=C going backward in time. This possible duplication may contribute to completeness but can also result in a decrease of efficiency.

Let us suppose the output of the identification procedure for the above example is A=B, F=C and E=F at level 2, A=B=C, E=F=C at level 3, A=B=C=D at level 4. Then the above defined metrics at level 2 are $compl(2) = 2/4$, $wrong(2) = 1/3$. At the next level, $compl(3) = 1/2$, $wrong(3) = 1/3$ and finally $compl(4) = 1$, $wrong(4) = 0$. Such results do not appear very good at level 2 and 3, but in fact the performance of the overall procedure has been perfect, both in completeness and in reliability.

To understand the last statement let us use this example to illustrate the final stage of the identification processing, the *normalization*. We first sort all the identifications found by "quality", that is, an identification comes before if it contains more nights[21]. In the example, we obtain

$$
\begin{array}{l}
A = B = C = D \\
A = B = C \\
E = F = C \\
A = B \\
F = C \\
E = F
\end{array}
$$

[21] In fact, we also sort the identifications with the same number of nights in ascending order of normalized RMS of residuals.

Then we scan this sorted list from the top to reduce it to a *normalized* list of identifications. The first one, A=B=C=D is kept in the normalized list. The second one is removed because it is *compatible* with the first. E=F=C is removed because it is *discordant* with A=B=C=D and inferior (less nights). A=B is removed because it is *compatible* with the first. F=C is removed because *discordant* with A=B=C=D and inferior. E=F is kept because it is *independent* from the first. The identifications left in the normalized list are only A=B=C=D, E=F, thus $\mathfrak{Id}(4,4) = N_{id}(4) = 1$ and $\mathfrak{Id}(2,2) = N_{id}(2) = 1$; $\mathfrak{Wr}(k) = 0$ for all k.

The normalization procedure is thus univocally defined by the binary relations among identifications: *compatible* (all the TSAs belonging to the first are among the TSAs of the second), *independent* (none of the TSAs belonging to the first are among the TSAs of the second) and *discordant* (neither compatible nor independent). Discordant identifications appear as contradictions, unless they are removed by an identification containing all the TSAs of both: e.g., A=B=C and B=C=D are discordant unless A=B=C=D is in the list.

The discordant identifications with the same number of nights are both removed from the normalized list at the end of the procedure. This choice is justified because a wrong identification results in "permanent" damage: once they are accepted (in the normalized list) the corresponding TSAs are removed from the list of TSAs to be identified, thus making impossible to find later the true identifications for them. Thus it is better to remove a true identification, which could be recovered later, rather than keeping a wrong one.

The question is when the removal of the TSAs belonging to identifications should be done. The outcome, that is the normalized identification list, does depend upon the order of the operations. Let us consider another example: we find the identifications C=D=E, A=B=C=D, E=F=G, in this order while executing the identification procedure. If normalization is applied by comparing each new identification with all the ones previously found, but the identified TSAs are not removed, the identification C=D=E would be removed because discordant and inferior to A=B=C=D; then E=F=G would not be removed, being independent. On the contrary, if the TSAs belonging to an identification were removed immediately after the identification has been found, the TSAs C, D, and E would not be used to look for other identifications after the first one and A=B=C=D would not be found. Thus, both by removing and by not removing the identified TSAs, the result would be different and less likely to be true with respect to the one obtained by a posteriori normalization. The only safe procedure is to complete the search for all possible identifications, then normalize the identifications list, then remove the TSAs belonging to normalized identifications. The results reported in the following show that this procedure is essential to achieve good completeness and reliability.

In conclusion, the completeness relative to k–nighters can be measured by $\mathfrak{Compl}(k) = \mathfrak{Id}(k,k)/n(k)$, the reliability by the fraction of lost objects $\mathfrak{Lost}(k) = \mathfrak{Id}(k,1)/n(k)$; the fraction remaining with an orbit based on less nights than the observed ones is $\mathfrak{Inc}(k,h) = \mathfrak{Id}(k,h)/n(k)$ for $1 < h < k$. Our results will be expressed by means of these ratios.

7.2 Strategies for optimization

To be able to test our algorithms on rather large simulations, and also to claim that they are efficient, we need to optimize our procedure. The relevance of this work is shown by the fact that, between our first attempt and the current version of the code implementing the algorithms of this paper, we have decreased the CPU times by a factor > 100. However, optimization is to a large extent based on many tricks, by themselves not worth reporting. To make our work reproducible, we only outline the three basic principles we have followed.

(1) **Remove the quadratic loops**: when the total number of TSAs is $m_{tot} \simeq 3.5 \times 10^6$ a nanosecond consumed for each couple results in almost 2 hours of CPU. Thus the loops on the couples need to be replaced by methods of computational complexity $O(m_{tot} \log(m_{tot}))$. There are well known algorithms such as heap sorting and binary search with this property [Knuth 1973].
(2) **Use filter stages of increasing computational cost**: for attribution, at each of the steps (2-, 3- and 4–identifications) we use in sequence three filtering stages [Milani et al. 2001]. Much care needs to be taken to avoid that the number of couples passing the first and second filtering stage has a significant quadratic component.
(3) **Use iterations separated by normalization**: currently we run the algorithms twice, the first time with a very small number of VAs (either 1 or 2) sampling the admissible region of each TSA, the second time with $\simeq 50$ VAs per TSA. Between the two we perform the normalization of the identifications list and remove the TSAs belonging to the normalized identifications. In this way the more intensive computations are applied only to a reduced subset with much less than m_{tot} TSAs.

The computational efficiency of the identification procedure could be described by a *CPU time model*, that is by measuring the computation time $c(m)$ (for a given hardware) as a function of the number m of TSAs. The algorithms include computations done for each TSAs, computations done for each couple of them, and binary searches in lists of TSAs. However, there is also a startup time for all the different programs involved. By using the CPU times of four tests with 1/10, 2/10, 3/10 and 10/10 of the simulated objects (with 2 iterations, see Section 7.3, and including also the programs used to compare with the ground truth), we have found a reasonable fit with the CPU time model

$$c(m) = 58 + 2.2 \times 10^{-4}\, m + 1.5 \times 10^{-10}\, m^2$$

in minutes [22] . That is, our code still contains a quadratic component with 18 nanoseconds consumed for each couple of TSAs.

7.3 Simulation Results

The dataset of the Pan-STARRS simulation contained $m_{obs} = 7,053,082$ observations, from which $m_{tot} = 3,525,714$ attributables have been computed [23] . The number of objects observed over k consecutive nights is given in Table 1.

Table 1
Simulated observations: number of objects observed over k nights.

k	4	3	2	1	total
$n(k)$	799909	58275	52118	46794	957096
for NEA only	1166	215	187	198	1766

The first iteration

The first iteration of the identification/orbit determination procedure consists of five steps:

(1) generation of Virtual Asteroids from each attributable;
(2) linkages, that is attribution of attributables to the VA generated in step 1, thus obtaining a 2–identification orbit;
(3) attribution of attributables to the 2–identification orbits of step 2, thus obtaining a 3–identification orbit.
(4) attribution of attributables to the 3–identification orbits of step 3, thus obtaining a 4–identification orbit.
(5) identification management, including normalization of the list of 3– and 4–identifications and removal of the corresponding attributables from the data set to be submitted to the second iteration.

[22] Xeon 3GHz, Linux Suse 9.1, Intel Fortran compiler 8.1.
[23] The discrepancy $m_{obs} - 2\, m_{tot}$ was due to quality control, in which simulated couples of observations which would result on the same pixel were removed.

This sequence of operations implements the scheme discussed in Section 8 of [Milani and Knežević, 2005]. The purpose of the first iteration is to remove as many attributables as possible from those to be identified, by solving the "easy" cases with a computationally cheap algorithm. Thus the average number of VAs per attributables needs to be close to 1.

We have used in Step 1 a method based on our theory of the admissible region: for each attributable we have computed, in the $(r, \dot{r})$ plane, the point with largest r among those resulting in a semimajor axis of 2.5 AU [24]. Only if the region with $a < 2.5$ AU had two connected components we have also added another VA located in the component farther from Earth [25].

The logic behind this choice is to force our VAs to be in the main belt, whenever the admissible region contains some main belt type orbits. This should give some advantage with respect to the classical Väisälä method [Väisälä and Oterma 1951], which assumes the asteroid is at perihelion when it is observed. Väisälä's assumption can be shown to be optimal if the object is indeed a main belt asteroid and it is observed near the opposition, our method should be less dependent upon the observing strategy. Anyway, the details of the method used to obtain VAs in the first iteration do not matter much, provided it achieves the goal of allowing the identification of most attributables. It is obvious that the assumptions made in selecting the VAs (both with Väisälä's and with our method) are creating a "computational bias" against identifying Near Earth Objects (NEO), but this does not matter in the first iteration.

Because in the first iteration computational efficiency has priority with respect to completeness (but not with respect to reliability) we have used in Steps 2, 3 and 4 comparatively low values for the controls, e.g., $K_{max} \leq 16^2$ in Step 2 and $\leq 20^2$ in Steps 3 and 4. In steps 2 and 3 we search for attributions only forward in time. An additional control to avoid false identifications is based on the comparison of the mean apparent magnitude $\bar{h}$ with the predicted one. To minimize the number of false and incomplete identifications (both would result in removal from the list of attributables, thus the second iteration could not recover the corresponding true and complete identifications) we have excluded the 2–identifications from Step 5.

The results of the first iteration are summarized in Table 2. Note that the level of completeness is already good for 2– and 3–identifications, it is extremely good for 4–identifications. The total CPU time of the first iteration was 37.2 hours, thus we have achieved the goal of making our algorithms ef-

---

[24] In the notation of Paper I, we have used for r the first positive root of the degree six equation associated with the level curve $E_\odot = -k^2/(2\, a_{max})$ for $a_{max} = 2.5$ AU, and $\dot{r} = -c_1/2$.

[25] Taking for r the average of the second and third positive root, and $\dot{r} = -c_1/2$.

Table 2
Results of the first iteration: for the symbols, see Section 7.1.

k	4	3	2
$compl(k)$	98.3 %	94.8 %	95.9 %
$wrong(k)$	0.0001%	1.4 %	4.1 %
$\mathfrak{Compl}(k)$	98.26 %	93.92 %	
for NEA only	45.5 %	33.0 %	
$\mathfrak{Inc}(k, k-1)$	0.02 %		
for NEA only	0.3 %		
$\mathfrak{Lost}(k)$	1.71 %	6.07 %	
for NEA only	54.2 %	67.0 %	
$\mathfrak{Wr}(k)$	0.0001 %	0.01 %	
for NEA only	0 %	0 %	

ficient enough to be used as primary orbit determination method. However, what we have been running in the first iteration is not really the algorithm described in this series of papers, because we have not yet exploited the triangulation of the admissible region.

However, the small fraction of $\mathfrak{Lost}(4) = 1.71\%$ includes 54.2% of the NEO! The situation is even worse for 3–nighters, with 67% of the NEO among the lost. This is the motivation for the use of two iterations: the fast method essentially does not work for NEO, but allows to reduce the data set, thus making feasible a much more computationally intensive search for identifications in the second iteration. After the first iteration, the number of attributables left after removal of the observations of the (normalized) 3– and 4–identifications was $213,606$, that is only 6.1% of the original data set.

The second iteration

The second iteration is performed on the reduced list of "leftover" attributables selected by Step 5 of the first iteration. There are also five steps corresponding to the ones of the first iteration, with the following specific differences:

(1) on average 50 VAs are generated for each attributable by using the De-

launay triangulation of the admissible region (see Paper I);
(2–4) the controls are much looser, e.g., $K_{max} = 100^2$;
(5) Identification management includes normalization of the list of 2–, 3– and 4–identifications.

In step 3 we look for attributions also going backward in time.

Table 3
Results of the second iteration.

k	4	3	2
$compl(k)$	99.8 %	96.3 %	96.9 %
$wrong(k)$	0.0001%	1.4 %	4.1 %
$\mathfrak{Compl}(k)$	99.84 %	98.27 %	93.69 %
for NEA only	99.5 %	99.5 %	96.3 %
$\mathfrak{Inc}(k, k-1)$	0.03 %	0.14 %	
for NEA only	0.3 %	0 %	
$\mathfrak{Inc}(k, k-2)$	0.04 %		
for NEA only	0 %		
$\mathfrak{Lost}(k)$	0.06 %	1.58 %	5.93 %
for NEA only	0.3 %	0.5 %	3.2 %
$\mathfrak{Wr}(k)$	0.0001 %	0.01 %	0.01 %
for NEA only	0 %	0 %	0 %

The results of the second iteration are summarized in Table 3. The fraction of objects lost among the 4–nighters becomes very small for both main belt and NEO. 3–nighters also have a very good level of completeness, even for 2–nighters the results are good. The fraction of wrong identifications is minute. The higher fraction of identifications for NEO is unexplained, but may be just the effect of small number statistics: indeed, there are only 10 lost NEO (6 being 2–nighters), plus 3 incomplete NEO identifications.

In fact the number of "failures" is so small that we might be tempted to look at them one by one and try to fix them: e.g., there are only 13 false identifications (none for NEO). However, to further reduce these numbers by ad hoc changes to the control parameters would give an illusory result. Additional tuning of the control parameters needs to be done on a much more realistic simulation and/or on real data.

The CPU time of the second iteration was 7.4 hours: more intensive computations were applied to much less numerous attributables to be identified, with an overall computational cost significantly less than that of the first iteration. This implies that, if in some future simulation the final results were not complete enough, we would have to further increase the computational intensity of the second iteration rather than of the first. Anyway the total CPU time of less than 2 days implies that the efficiency of the algorithm has exceeded the requirements for practical use as primary orbit determination method.

We are not claiming that it is impossible to find a method both fast and complete, to be used in a single iteration. E.g., it is clear that we could use a "smart triangulation" algorithm selecting the number of nodes in the triangulation depending upon some parameters of the attributable (the most obvious being the proper motion and the elongation), and even retrying with a larger number of nodes if no identification has been found: this would essentially be the same as our sequence of two iterations done at once. For the purpose of this paper, to keep the two iterations separate allows to better understand the main problem, which is the following.

For a computationally fast method we have to use few VAs per attributable: if we want the fraction of objects identified by this method to be large, we have to select these VAs in the portion of the admissible region where the most numerous population can be found, in practice to select main belt like VAs. In this way we selectively loose identifications for all the objects with unusual orbits, which are the most interesting! The second iteration needs to be totally unbiased, allowing for orbits known to be rare and even for ones never discovered before. This is obtained by using a triangulation of the entire admissible region.

8 Conclusions and future work

The goal of this research program was to solve the problem of orbit determination starting from a set of TSAs, each one by itself containing insufficient information for the classical orbit determination methods. The most difficult step was to obtain an orbit from a couple of TSAs obtained at different times. This required to define two algorithms, one for computation of a preliminary orbit, and another one to allow for differential corrections starting from the preliminary orbit, even for the cases in which the conventional algorithm (pseudo-Newton method) fails. A less challenging, but still not trivial, problem was to identify a TSA with one of more orbits computed from 2 other TSAs. As the number of TSAs already identified increases the problem becomes easier, until the algorithms already known [Milani et al. 2001] are perfectly adequate.

The difficulty of the task arises from the fact that we have to assume that the number of TSAs obtained from each night of observation is large, thus we need to use an algorithm computationally efficient, to scan all the possible couples of TSAs, and at the same time very accurate, to be able to discard false identifications when a rigorous least square fit leaves unacceptable residuals. Although the classical algorithms could in principle be used, they would not be efficient enough to cope with the data volume expected from the next generation surveys. For a discussion on this, see also [Milani and Knežević, 2005].

In one of our already published papers we had solved the problem of attributing a TSA to a LOV orbit [Milani et al. 2005a], and this method can be applied to the case in which we search for a TSA to be attributed to 2 TSAs already identified (Section 6.3). We had also solved in Paper I the problem of selecting Virtual Asteroids for an object about which we only have a single TSA. In [Milani and Knežević, 2005] we gave an outline of the overall orbit determination procedure, with several critical details to be filled in. With this paper we conclude the next main step.

8.1 Results

In this paper we have solved the problem of selecting a (small) subset of TSA couples for which an identification is worth trying, by using the *minimum identification penalty* (Section 5.1). For the selected couples, the algorithm also provides a preliminary orbit which could fit the two attributables of the TSAs with moderate residuals (Section 5.3).

From each preliminary orbit we can compute a constrained solution, along the LOV (line of weakness) of the least squares fit with the observation of both TSAs (Section 6.2).

From each LOV solution we can search for a third TSA to be attributed. This follows a similar scheme, with the minimum identification penalty acting as a filter control to select the possible triples of TSAs (Section 6.3). This procedure can be seen as one step in a recursive procedure to add more and more TSAs to an orbit becoming increasingly well determined.

To the above theoretical results we have added the results of a large scale simulation, with $\simeq 3.5 \times 10^6$ TSAs. With such a challenging test, of the same order of complexity of the orbit determination problems to be solved for the next generation surveys, we have been able to show that our algorithms can be implemented in a very efficient software. The computational resources required to use our algorithms as primary orbit determination method, applied to all observations, are modest (a couple of days of CPU with a standard workstation).

We have obtained a level of completeness above 99% for 4–nighters, both main belt and NEO, above 98% for 3–nighters and above 93% for 2–nighters. The leftover "One Night Stands" file after removal of the TSAs belonging to the identifications of the second iteration contains only 50,257 TSAs, out of which most belong to 1–nighters: only 3,493 (0.1% of the original dataset) belong to objects observed in more than 1 night, thus could have been identified. The number of wrong identifications is 13, negligibly small.

8.2 Open problems

One obvious problem is that it is not possible to find identifications with objects which have not been observed. To estimate the completeness of the simulated survey we should use the ratio between the total number of complete orbits in the final catalog and the total number of objects observed:

$$\mathfrak{Compl}_S = \sum_{k=2}^{K} n(k)\ \mathfrak{Compl}(k) / \sum_{k=1}^{K} n(k)$$

which turns out to be, for the simulation we have completed, 94.5%. This apparently satisfactory result depends upon the fact that, among the data of the observation simulation we started from, most observations belonged to 4–nighters (see Table 1). If there was a large proportion of 1–nighters, there is nothing our identification algorithm could do to avoid a much lower $\mathfrak{Compl}_S$. Moreover, if there was a large proportion of 2–nighters, we would be left with many very poor orbits, and also with a much larger number of false 2–identifications.

Unfortunately, with a more realistic detection model the proportion of objects successfully observed over all the nights would sharply decrease. Photon statistics, light curve effects, and different seeing contribute to give to the same asteroid a different signal to noise ratio on different nights, and even over the short time span between the two observations in the same night. Thus, for the objects with a nominal signal to noise ratio marginally above the detection threshold, the actual detection of a TSA becomes a random event.

In the simulation of this paper we have limited the search for identifications to consecutive observing nights, that is the propagation time span was limited to 4 days. This limitation could be removed, and the algorithms should perform well also over a propagation time span of 8, possibly also of 12 days, as the example of Figure 5, with a propagation time of 11 days, indicates. This would improve the situation in presence of a large fraction of objects observed, e.g., the first and the third night. However, the overall procedure would become somewhat more complicated.

An open problem is how to use the 2–identification orbits. One possibility is to identify them with k–identification orbits obtained in other months; if the survey has been operational over a time span of the order of the synodic periods of the main belt asteroids, most of the objects will have multiple orbits in the catalog and identification could be performed with standard algorithms. The problem of identifying even the constrained solutions of the 2–identifications orbits has already been solved [Milani et al. 2005a]. Another possibility is to propagate the 2-identification orbits to seek TSAs to be attributed in observing nights not too far in time, possibly even looking for detections with marginal signal to noise. With all this, we still think that the 2–nighters are not a real discovery, just a proposed discovery: their orbits are often so poorly determined that it is not even possible to decide if they are NEO, and some of them may even correspond to false identifications! In our opinion, an asteroid/comet survey should have as one of the design goals to minimize the number of 2–nighters, as well as the number of 1–nighters.

Last but not least, such a huge orbit determination would result in a large catalog of orbits, some of which would be compatible, within the uncertainties, with collisions with our planet. We have developed methods to solve this *impact monitoring* problem [Milani et al. 2005b]. However, the next generation surveys might generate such a large population of *Virtual Impactors* that the current algorithms might not be efficient enough. This problem is connected to the one of 2–nighters: they have such poor orbits, that a surprisingly large proportion of them would turn out to be compatible with impacts, although the probabilities of such impacts would be very small.

ACKNOWLEDGMENTS

We thank the Pan-STARRS project for their support, and in particular for having allowed the use of one of their preliminary simulations. This research has been funded by: the Italian *Ministero dell'Università e della Ricerca Scientifica e Tecnologica*, PRIN 2004 project "The Near Earth Objects as an opportunity to understand physical and dynamical properties of all the solar system small bodies", *Ministry of Science and Environmental Protection of Serbia* throught project 1238, the Spanish *Ministerio de Ciencia y Tecnología* and the European funds *FEDER* through the grant AYA2001-1784. We thank R. Jedicke, S. Chesley, G.Valsecchi, A. Boattini and A. Harris for illuminating discussions on the next generation surveys.

References

[Boattini et al., 2004] Boattini, A., et al., 2004, *Near Earth Asteroid search and follow-up beyond 22ⁿd magnitude*, *Astron. Astrophys.*, 418, 743–750.

[Bowell et al. 1989] Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J., Harris, A.W.: 1989, 'Application of photometric models to asteroids', in *Asteroids II*, R. P. Binzel, T.Gehrels and M.S. Mathews eds., pp. 524–556

[Carpino et al., 2003] Carpino, M., Milani, A., Chesley, S.R. 2003. *Error Statistics of Asteroid Optical Astrometric Observations*, Icarus, **166**, 248–270.

[Celletti and Pinzari 2005] Celletti, A., Pinzari, G.: 2005, *Four classical methods for determining planetary elliptic elements: a comparison*, preprint, `http://adams.dm.unipi.it/simca/preprints/2005/January05/celletti_11_01_2005.pdf`

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

[Jazwinski, 1970] Jazwinski, A. H. 1970. *Stochastic processes and filtering theory.* Academic Press.

[Knuth 1973] Knuth, D.E. 1973. *Sorting and Searching. The Art of Computer Programming, vol. 3*, Addison Wesley.

[Milani, 1999] Milani, A. 1999. *The Asteroid Identification Problem I: recovery of lost asteroids*, *Icarus*, **137**, 269–292.

[Milani and Knežević, 2005] Milani, A., Knežević, Z. 2005. *From Astrometry to Celestial Mechanics: Orbit Determination with Very Short Arcs*, CMDA, in press

[Milani et al. 2000] Milani, A., La Spina, A., Sansaturio, M. E., Chesley, S. R. 2000. *The Asteroid Identification Problem III. Proposing identifications*, *Icarus*, **144**, 39–53.

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

[Milani et al. 2004] Milani, A., Gronchi, G.F., de' Michieli Vitturi, M., Knežević, Z. 2004. *Orbit Determination with Very Short Arcs. I Admissible Regions*, CMDA, **90**, 59–87.

[Milani et al. 2005a] Milani, A., Sansaturio, M.E., Tommei, G., Arratia, O., Chesley, S.R.: 2005, *Multiple solutions for asteroid orbits: computational procedure and applications*, Astron. Astrophys., 431, 729–746.

[Milani et al. 2005b] Milani, A., Chesley, S.R., Sansaturio, M.E., Tommei, G., Valsecchi, G. 2005, *Nonlinear impact monitoring: Line Of Variation searches for impactors Icarus*, 173, 362–384.

[Väisälä and Oterma 1951] Väisälä, Y., Oterma, L. 1951, *Formulae and directions for computing the orbits of minor planets and comets Ann. Univ. Turkuensis, Ser. A*, t. X, No. 3, 1–32.