1981A&A...102...59N

Astron. Astrophy. 102, 59–64 (1981)

ASTRONOMY AND ASTROPHYSICS

A Simple Method of Orbit Determination

W. Neutsch

Astronomische Institute der Universität Bonn, Auf dem Hügel 71, D-5300 Bonn, Federal Republic of Germany

Received September 15, 1980; accepted May 5, 1981

Summary. In the following a simple and straightforward method of orbit determination is discussed. It may be used for all types of Kepler orbits (elliptic, parabolic or hyperbolic ones) and for any given number $N \geqq 3$ of observations all of which are treated symmetrically. This is advantageous if compared with the classical (unsymmetrical) methods with $N=3$ both for theoretical and practical reasons.

Furthermore the use of position and velocity at a fixed epoch as orbital elements makes it unnecessary to discuss some special cases, as e.g. small inclination or small excentricity, separately. Hence the formulae of the text are easily transformed into a computer code.

Key words: celestial mechanics – orbit determination

Introduction

This paper contains a short description of a very simple way to determine the elements of the (Keplerian) orbit of a celestial body around a given centre of attraction from $N \geqq 3$ observations. In contrast to the classical methods (Gauß, 1891; Klinkerfues, 1899; Bauschinger, 1928) which are designed for the use of logarithm tables ours is rather well-suited for electronic calculators. This is mainly so because we need not consider as special cases, say, orbits of small inclination to the ecliptic or nearly circular or parabolic orbits. In fact, the method works equally well for (minor) planets and for comets, as was shown by carrying out several numerical tests for objects with very different orbits.

1. Notation; Calculations of Position and Velocity

A certain disadvantage of the various known methods in celestial mechanics lies in the use of the classical orbital elements a, e, i, Ω, ω, T [with the usual meanings, see e.g. Stracke (1929), Bucerius (1966)], for some of them are not always continuous functions of position and velocity at a given moment. This makes a distinction of several special cases (small excentricity, small inclination,...) necessary. Mainly for this reason we suggest another possibility to describe the orbit of a celestial object (called the "planet") which moves under the attraction of a centre (the "sun") according to Newton's law

$$\ddot{\boldsymbol{r}} = -\frac{\mu}{|\boldsymbol{r}|^3}\boldsymbol{r} \tag{1}$$

where $\boldsymbol{r}$ is the position of the planet relative to the sun, a dot means a temporal derivative. μ is the product of Newton's constant of gravitation and the sum of the sun's and the planet's masses, which is assumed known. – From (1) we derive the well-known constancy of the vectors

$$\boldsymbol{C} = \boldsymbol{r} \times \dot{\boldsymbol{r}} \tag{2}$$

("angular momentum vector") and

$$\boldsymbol{A} = \dot{\boldsymbol{r}} \times \boldsymbol{C} - \frac{\mu}{|\boldsymbol{r}|}\boldsymbol{r} \tag{3}$$

("Pauli-Lenz-vector") which shows that both of them may be expressed by the classical elements. The representation is

$$\boldsymbol{C} = C \begin{pmatrix} \sin\Omega\sin i \\ -\cos\Omega\sin i \\ \cos i \end{pmatrix}; \quad C = |\boldsymbol{C}| = \sqrt{\mu p}; \quad p = a(1-e^2), \tag{4}$$

$$\boldsymbol{A} = A \begin{pmatrix} \cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i \\ \cos\omega\sin\Omega + \sin\omega\cos\Omega\cos i \\ \sin\omega\sin i \end{pmatrix}; \quad \mathrm{A} = |\boldsymbol{A}| = \mu e. \tag{5}$$

The directions of $\boldsymbol{C}$ and $\boldsymbol{A}$ determine the normal to the orbital plane and the perihelion direction respectively. From these formulae we can find all the classical elements, except the perihelion passage time T, from $\boldsymbol{A}$ and $\boldsymbol{C}$, i.e. from $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ at any time. That $\boldsymbol{C}$ and $\boldsymbol{A}$ are equivalent to only five elements is expressed by the fact that they are orthogonal to each other ($\boldsymbol{C}^{-}\boldsymbol{A}=0$). T is given by Kepler's equation and the fact that the true anomaly is the angle between $\boldsymbol{A}$ and $\boldsymbol{r}$. So the position and the velocity at any time contain the same information as the classical elements and we shall use them in the further developments as parameters of the orbit.

For the sake of simplicity and symmetry we introduce a more suitable notation:

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be the position and velocity vectors at time t, $\boldsymbol{a}^*$ and $\boldsymbol{b}^*$ those at (another) time t^*. Excluding the uninteresting case of a straight line orbit passing through the sun, we see that $\boldsymbol{a}$ and $\boldsymbol{b}$ (as well as $\boldsymbol{a}^*$ and $\boldsymbol{b}^*$) span the orbital plane and are linearly independent. So every vector in the orbital plane (especially $\boldsymbol{a}^*, \boldsymbol{b}^*$) is expressible in a unique way as a linear combination of $\boldsymbol{a}$ and $\boldsymbol{b}$. We write

$$\boldsymbol{a}^* = \alpha\boldsymbol{a} + \beta\boldsymbol{b}; \quad \boldsymbol{b}^* = \gamma\boldsymbol{a} + \delta\boldsymbol{b} \tag{6}$$

and, dually,

$$\boldsymbol{a} = \alpha^*\boldsymbol{a}^* + \beta^*\boldsymbol{b}^*; \quad \boldsymbol{b} = \gamma^*\boldsymbol{a}^* + \delta^*\boldsymbol{b}^* \tag{7}$$

It should be remarked that by some authors the coefficients $\alpha, \beta, \gamma, \delta$ are called f, g, $\dot{f}$, $\dot{g}$, respectively, but for reasons of symmetry and simplicity in notation we prefer the present nomenclature.

To determine $\boldsymbol{a}^*$ and $\boldsymbol{b}^*$ from $\boldsymbol{a}$ and $\boldsymbol{b}$ and vice versa, it is only necessary to know $\alpha, \beta, \gamma, \delta$; $\alpha^*, \beta^*, \gamma^*, \delta^*$ as functions of $\boldsymbol{a}, \boldsymbol{b}$ and t^*-t. By the constancy of $\boldsymbol{C}$, see (2), we observe that

$$\boldsymbol{C}=\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{a}^*\times\boldsymbol{b}^*=\boldsymbol{C}^* \tag{8}$$

and so

$$\alpha\delta-\beta\gamma=\alpha^*\delta^*-\beta^*\gamma^*=1. \tag{9}$$

The definition of the coefficients in (6) and (7) leads to

$$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\begin{pmatrix}\alpha^* & \beta^*\\ \gamma^* & \delta^*\end{pmatrix}=\begin{pmatrix}\alpha^* & \beta^*\\ \gamma^* & \delta^*\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \tag{10}$$

or, explicitly,

$$\alpha^*=\alpha^*(\boldsymbol{a},\boldsymbol{b};t^*-t)=\alpha(\boldsymbol{a}^*,\boldsymbol{b}^*;t-t^*)=\delta(\boldsymbol{a},\boldsymbol{b};t^*-t)=\delta, \tag{11}$$

$$\beta^*=\beta^*(\boldsymbol{a},\boldsymbol{b};t^*-t)=\beta(\boldsymbol{a}^*,\boldsymbol{b}^*;t-t^*)=-\beta(\boldsymbol{a},\boldsymbol{b};t^*-t)=-\beta, \tag{12}$$

$$\gamma^*=\gamma^*(\boldsymbol{a},\boldsymbol{b};t^*-t)=\gamma(\boldsymbol{a}^*,\boldsymbol{b}^*;t-t^*)=-\gamma(\boldsymbol{a},\boldsymbol{b};t^*-t)=-\gamma, \tag{13}$$

$$\delta^*=\delta^*(\boldsymbol{a},\boldsymbol{b};t^*-t)=\delta(\boldsymbol{a}^*,\boldsymbol{b}^*;t-t^*)=\alpha(\boldsymbol{a},\boldsymbol{b};t^*-t)=\alpha. \tag{14}$$

To get $\alpha, \beta, \gamma, \delta$ from $\boldsymbol{a}, \boldsymbol{b}$ and t^*-t, we first determine C and e, which are given by (4) and (5):

$$C=|\boldsymbol{C}|;\quad e=|\boldsymbol{A}|/\mu \tag{15}$$

and then the true anomaly v at the time t, i.e. the angle between $\boldsymbol{A}$ and $\boldsymbol{a}$. At the time t^* the planet has the anomaly v^* which is found with the help of the Kepler-Barker-equation:

$$t^*-t=\frac{C^3}{\mu^2}\int_v^{v^*}(1+e\cos x)^{-2}dx. \tag{16}$$

The integration can be carried out analytically, and then the solution of this implicit relation may be effected by standard methods, cf. e.g. Klinkerfues (1899) or Bauschinger (1928), so there is no need to discuss it more thoroughly.

With the knowledge of C, e, v and v^* it is easy to evaluate $\alpha, \beta, \gamma, \delta$: We derive from the definition (6) of $\alpha, \beta, \gamma, \delta$:

$$\frac{\partial\alpha}{\partial t^*}=\gamma;\quad \frac{\partial\beta}{\partial t^*}=\delta. \tag{17}$$

This and (14) allow us to calculate $\delta, \alpha=\delta^*$ and γ as soon as β is known. β is found most conveniently via

$$C|\boldsymbol{a}||\boldsymbol{a}^*|\sin(v^*-v)=\boldsymbol{C}(\boldsymbol{a}\times\boldsymbol{a}^*)=\beta\boldsymbol{C}(\boldsymbol{a}\times\boldsymbol{b})=\beta C^2. \tag{18}$$

Let

$$\Delta=v^*-v. \tag{19}$$

Then the result is

$$\alpha=1+\frac{\cos\Delta-1}{1+e\cos v^*}, \tag{20}$$

$$\beta=\frac{C^3}{\mu^2}\frac{\sin\Delta}{(1+e\cos v)(1+e\cos v^*)}, \tag{21}$$

$$\gamma=\frac{\mu^2}{C^3}(e\sin v-e\sin v^*-\sin\Delta), \tag{22}$$

$$\delta=1+\frac{\cos\Delta-1}{1+e\cos v}. \tag{23}$$

Although this is all we need for orbit determination (in fact only the formulae for α and β are necessary), some other results may be useful for taking into account perturbations (not considered here) or for improving a given orbit by a least-square-method.

If we change $\boldsymbol{a}$ and $\boldsymbol{b}$ into $\boldsymbol{a}+\boldsymbol{da}$ and $\boldsymbol{b}+\boldsymbol{db}$ respectively – where $\boldsymbol{da}$ and $\boldsymbol{db}$ are small – then position and velocity ($\boldsymbol{a}^*$ and $\boldsymbol{b}^*$) at time t^* will be altered by amounts of, say, $\boldsymbol{da}^*$ and $\boldsymbol{db}^*$. We have to calculate $\boldsymbol{da}^*$ and $\boldsymbol{db}^*$ as functions of $\boldsymbol{da}$ and $\boldsymbol{db}$ (with given t and t^*) and vice versa. This may be done in the following way:

By assumption (nonlinear orbit) $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{C}$ are linearly independent. Thus we may write down $\boldsymbol{da}$ and $\boldsymbol{db}$ as unique linear combinations, e.g.,

$$\boldsymbol{da}=i\boldsymbol{a}+j\boldsymbol{b}+k\boldsymbol{C};\quad \boldsymbol{db}=l\boldsymbol{a}+m\boldsymbol{b}+n\boldsymbol{C} \tag{24}$$

and analogously

$$\boldsymbol{da}^*=i^*\boldsymbol{a}^*+j^*\boldsymbol{b}^*+k^*\boldsymbol{C}^*;\quad \boldsymbol{db}^*=l^*\boldsymbol{a}^*+m^*\boldsymbol{b}^*+n^*\boldsymbol{C}^*. \tag{25}$$

One may easily prove that

$$i^*+m^*=i+m;\quad i+m=i^*+m^*, \tag{26}$$

$$\begin{aligned}&i^*+2m^*=\alpha(i+2m)-\beta(Sj+l),\\ &i+2m=\delta(i^*+2m^*)+\beta(S^*j^*+l^*),\end{aligned} \tag{27}$$

$$\begin{aligned}&S^*j^*+l^*=-\gamma(i+2m)+\delta(Sj+l);\\ &Sj+l=\gamma(i^*+2m^*)+\alpha(S^*j^*+l^*),\end{aligned} \tag{28}$$

$$\begin{aligned}&j^*=-3(t^*-t)(i+m)+j+J_\alpha(i+2m)-J_\beta(Sj+l);\\ &j=3(t^*-t)(i^*+m^*)+j^*+J_\alpha^*(i^*+2m^*)-J_\beta^*(S^*j^*+l^*),\end{aligned} \tag{29}$$

$$k^*=\alpha k+\beta n;\quad k=\delta k^*-\beta n^*, \tag{30}$$

$$n^*=\gamma k+\delta n;\quad n=-\gamma k^*+\alpha n^*, \tag{31}$$

where

$$J_\alpha=2\int_0^{t^*-t}\alpha(\boldsymbol{a},\boldsymbol{b};u)du;\quad J_\beta=2\int_0^{t^*-t}\beta(\boldsymbol{a},\boldsymbol{b};u)du, \tag{32}$$

$$J_\alpha^*=-\delta J_\alpha+\gamma J_\beta;\quad J_\beta^*=\beta J_\alpha-\alpha J_\beta, \tag{33}$$

$$S=\mu|\boldsymbol{a}|^{-3};\quad S^*=\mu|\boldsymbol{a}^*|^{-3}. \tag{34}$$

These equations contain the same information as the Jacobian

$$\frac{\partial(\boldsymbol{a}^*,\boldsymbol{b}^*)}{\partial(\boldsymbol{a},\boldsymbol{b})}. \tag{35}$$

It should be noted that only for evaluating (16) and (32) a distinction between elliptic, parabolic and hyperbolic orbits is necessary. All the other equations are independent of the type of the planet's motion. This is especially true for the orbit determination considered next.

2. Orbit Determination ($N \geqq 3$ Observations)

We are now prepared to solve our main problem, i.e. to determine the elements of the planet's (Keplerian) orbit from $N \geqq 3$ observations. The following data are provided by the i'th ($1 \leqq i \leqq N$) observation:

1. The time of the measurement t_i;

2. The position vector $\boldsymbol{E}_i=\begin{pmatrix}X_i\\ Y_i\\ Z_i\end{pmatrix}$ of the observer at that time relative to the sun;

3. The observed direction to the planet, given by a unit vector

$$e_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix};$$

and

4. The weight p_i of the observation.

Let $\boldsymbol{r}_i$ be the position which is observed at time t_i. The light from the planet has to travel along a distance

$$d_i = |\boldsymbol{r}_i - \boldsymbol{E}_i| \tag{1}$$

so $\boldsymbol{r}_i$ is the position of the planet at time $t_i - d_i/c$, where c = velocity of light ("planetary aberration").

We introduce a fixed epoch t_0 as the weighted mean of the observational times:

$$t_0 = \frac{\sum_{i=1}^{N} p_i t_i}{\sum_{i=1}^{N} p_i} \tag{2}$$

and call the planet's position and velocity at that time $\boldsymbol{a}$ and $\boldsymbol{b}$. By Eq. (6) of Sect. 1 we may write

$$\boldsymbol{r}_i = \alpha_i \boldsymbol{a} + \beta_i \boldsymbol{b} \tag{3}$$

with the coefficients given by

$$\alpha_i = \alpha(\boldsymbol{a}, \boldsymbol{b};\ t_i - t_0 - d_i/c), \tag{4}$$

$$\beta_i = \beta(\boldsymbol{a}, \boldsymbol{b};\ t_i - t_0 - d_i/c). \tag{5}$$

Because the planet in $\boldsymbol{r}_i$ is seen from $\boldsymbol{E}_i$ in the direction $\boldsymbol{e}_i$ at a distance d_i, we find with help of the fact that $\boldsymbol{e}_i$ is a unit vector:

$$\boldsymbol{e}_i^2 = \boldsymbol{e}_i^t \boldsymbol{e}_i = x_i^2 + y_i^2 + z_i^2 = 1 \tag{6}$$

a second expression for $\boldsymbol{r}_i$, namely

$$\boldsymbol{r}_i = \boldsymbol{E}_i + d_i \boldsymbol{e}_i. \tag{7}$$

We are not interested in $\boldsymbol{r}_i$, so we eliminate it from (3) and (7), getting

$$\alpha_i \boldsymbol{a} + \beta_i \boldsymbol{b} - d_i \boldsymbol{e} = \boldsymbol{E}_i \tag{8}$$

for all $i = 1, .., N$. With (2) we have to find the unknown quantities α_i, β_i, d_i; $\boldsymbol{a}$ and $\boldsymbol{b}$ from the nonlinear system (4), (5), (8).

The solution is by iteration. We begin with an estimation of α_i and β_i, then solve the *linear* system (8) for a_1, a_2, a_3; b_1, b_2, b_3; $d_1, \ldots, d_N$. By (4) and (5) we get a new approximation for α_i and β_i. We repeat this calculation until convergence occurs (normally only very few steps are needed). The linearity of (8) is most useful, because it allows to use any given number $N \geqq 3$ of observations and to pin down the normal equations immediately. We postpone a more thorough discussion until later on (Sect. 3 and 4).

The initial values of α_i and β_i may be derived by developing α_i and β_i into a power series of the time interval

$$\tau_i = t_i - t_0. \tag{9}$$

We find easily from the definitions

$$\alpha_i = \alpha(\boldsymbol{a}, \boldsymbol{b}; \tau_i - d_i/c) \approx \alpha(\boldsymbol{a}, \boldsymbol{b}; \tau_i) = 1 + O(\tau_i^2), \tag{10}$$

$$\beta_i = \beta(\boldsymbol{a}, \boldsymbol{b}; \tau_i - d_i/c) \approx \beta(\boldsymbol{a}, \boldsymbol{b}; \tau_i) = \tau_i + O(\tau_i^3). \tag{11}$$

So the relative error is only of order τ_i^2 (i.e. the square of the – usually small – heliocentric arc the planet moves along between t_0 and t_i), if we put

$$\alpha_i = 1; \quad \beta_i = t_i - t_0 \tag{12}$$

as the first approximation (this is the limiting case of linear unaccelerated motion). Of course, if we know something about the orbit in advance of the calculation, other choices could be better, but we shall not go into the details here.

The only step of our iteration scheme which should be discussed more thoroughly is the solution of (8). This is done in Sect. 3 (for the somewhat simpler case $N = 3$) and 4 (for general values of N).

3. The Special Case $N = 3$

The case $N = 3$ is of special interest because

1. 3 is the minimum number of observations needed to determine $\boldsymbol{a}$ and $\boldsymbol{b}$;
2. There are as many conditions as unknowns (9), so an exact solution is possible, while for $N \geqq 4$ we have to apply a least-square-method;
3. The calculations are simpler than in the general case.

We introduce some additional notation. Let $\boldsymbol{E}$ and $\boldsymbol{e}$ be the (3,3)-matrices

$$\boldsymbol{E} = (\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3) = \begin{pmatrix} X_1 & X_2 & X_3 \\ Y_1 & Y_2 & Y_3 \\ Z_1 & Z_2 & Z_3 \end{pmatrix};$$

$$\boldsymbol{e} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3) = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} \tag{1}$$

while $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ are the 3-vectors

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}; \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix};$$

$$\boldsymbol{\varphi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix} = \boldsymbol{\alpha} \times \boldsymbol{\beta} = \begin{pmatrix} \alpha_2\beta_3 - \beta_2\alpha_3 \\ \alpha_3\beta_1 - \beta_3\alpha_1 \\ \alpha_1\beta_2 - \beta_1\alpha_2 \end{pmatrix}. \tag{2}$$

We shall assume that $\varphi_1, \varphi_2, \varphi_3$ are all different from 0. Otherwise the planet would have moved between two of the three times t_i along a heliocentric arc which is an exact multiple of 180°, a case of little practical interest. We also exclude the possibility

$$\det \boldsymbol{e} = \boldsymbol{e}_1 (\boldsymbol{e}_2 \times \boldsymbol{e}_3) = 0 \tag{3}$$

which corresponds to coplanar $\boldsymbol{e}_i$ (all the directions lie on a great circle), for then there could be more than one solution of (2.8), cf. Klinkerfues (1899). Under these circumstances the classical methods of (e.g.) Gauß and Lagrange (see Bucerius, 1966) are not applicable either. – From these hypotheses it follows that the determinant of (2.8),

$$\mathrm{Det} = -\varphi_1 \varphi_2 \varphi_3 \det \boldsymbol{e} \tag{4}$$

does not disappear, so the system (2.8) admits a unique solution. Written down in full, (2.8) becomes

$$\left(\begin{array}{c|c|ccc} \alpha_1 \cdot \mathbf{1} & \beta_1 \cdot \mathbf{1} & -\boldsymbol{e}_1, & \mathbf{0}, & \mathbf{0} \\ \alpha_2 \cdot \mathbf{1} & \beta_2 \cdot \mathbf{1} & \mathbf{0}, & -\boldsymbol{e}_2, & \mathbf{0} \\ \alpha_3 \cdot \mathbf{1} & \beta_3 \cdot \mathbf{1} & \mathbf{0}, & \mathbf{0}, & -\boldsymbol{e}_3 \end{array}\right) \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{d} \end{pmatrix} = \begin{pmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \boldsymbol{E}_3 \end{pmatrix} \tag{5}$$

where $\mathbf{1}$ and $\mathbf{0}$ are the (3, 3) unit matrix and the 3-dimensional zero vector, respectively, while $\boldsymbol{d} = (d_1, d_2, d_3)^t$.

Multiplying from the left by

$$(\varphi_1\cdot\mathbf{1}\ \mid\ \varphi_2\cdot\mathbf{1}\ \mid\ \varphi_3\cdot\mathbf{1}) \tag{6}$$

we arrive at a system of 3 equations for d_1, d_2, d_3:

$$\begin{pmatrix} -\varphi_1 x_1 & -\varphi_2 x_2 & -\varphi_3 x_3 \\ -\varphi_1 y_1 & -\varphi_2 y_2 & -\varphi_3 y_3 \\ -\varphi_1 z_1 & -\varphi_2 z_2 & -\varphi_3 z_3 \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} \varphi_1 X_1 + \varphi_2 X_2 + \varphi_3 X_3 \\ \varphi_1 Y_1 + \varphi_2 Y_2 + \varphi_3 Y_3 \\ \varphi_1 Z_1 + \varphi_2 Z_2 + \varphi_3 Z_3 \end{pmatrix} = \boldsymbol{E}\cdot\boldsymbol{\varphi} \tag{7}$$

or, more explicitly,

$$\begin{pmatrix} \varphi_1 d_1 \\ \varphi_2 d_2 \\ \varphi_3 d_3 \end{pmatrix} = \boldsymbol{R}\cdot\boldsymbol{\varphi} \tag{8}$$

with the (3, 3)-matrix $\boldsymbol{R}$ given by

$$\boldsymbol{R} = -\,e^{-1}\cdot\boldsymbol{E} \tag{9}$$

which may be evaluated once and for all before the iteration, because it is independent of α_i and β_i (and φ_i).

We next insert the d_i from (8) into (5), which then splits into three subsystems ($k=1,2,3$):

$$\begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{pmatrix}\begin{pmatrix} a_k \\ b_k \end{pmatrix} = \begin{pmatrix} E_{1k} + d_1 e_{1k} \\ E_{2k} + d_2 e_{2k} \\ E_{3k} + d_3 e_{3k} \end{pmatrix}. \tag{10}$$

Multiplying from the left with the transposed matrix we find after a little bit of manipulation

$$\begin{pmatrix} A & C \\ C & B \end{pmatrix}\begin{pmatrix} a_k \\ b_k \end{pmatrix} = \begin{pmatrix} H_{k1}\alpha_1 + H_{k2}\alpha_2 + H_{k3}\alpha_3 \\ H_{k1}\beta_1 + H_{k2}\beta_2 + H_{k3}\beta_3 \end{pmatrix} \tag{11}$$

4. Solution for General $N \geqq 3$

In principle it would suffice to have a method for orbit determination using three observations, for one could select three measurements, calculate an orbit passing through them and improve it afterwards by a least-square-method.

This, however, is neither elegant nor the most effective way to proceed, for we do not know a priori which three observations give the best approximation. An unlucky choice leads to additional (and unnecessary) labour. So it is desirable to use all the observations according to their weights from the very beginning. Of course, for $N \geqq 4$ the system (2.8) only allows for a least-square-solution, because there are more Equations ($3N$) to be fulfilled than unknown numbers ($N+6$). –

To begin with, let us pin down (2.8) explicitly:

$$\left[\begin{array}{c|c|cccc} \alpha_1\cdot\mathbf{1} & \beta_1\cdot\mathbf{1} & -e_1, & 0, & \cdots, & 0 \\ \alpha_2\cdot\mathbf{1} & \beta_2\cdot\mathbf{1} & 0, & -e_2, & \cdots, & 0 \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot\ \cdots, & \cdot \\ \cdot & \cdot & & & & \\ \alpha_N\cdot\mathbf{1} & \beta_N\cdot\mathbf{1} & 0, & 0, & \cdots, & -e_N \end{array}\right]\left[\begin{array}{c} \boldsymbol{a} \\ \hline \boldsymbol{b} \\ \hline d_1 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{array}\right] = \left[\begin{array}{c} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{E}_N \end{array}\right]. \tag{1}$$

To get the set of normal equations we have to multiply from the left first with the weight matrix

$$\mathbf{Diag}(p_1, p_1, p_1;\ p_2, p_2, p_2;\ \cdots;\ p_N, p_N, p_N) \tag{2}$$

and then with the transposed to the $(3N, N+6)$-matrix in (1). The result is (summation always over i from 1 to N):

$$\left[\begin{array}{c|c|cccc} A\cdot\mathbf{1} & C\cdot\mathbf{1} & -p_1\alpha_1\boldsymbol{e}_1, & & \cdots, & -p_N\alpha_N\boldsymbol{e}_N \\ \hline C\cdot\mathbf{1} & B\cdot\mathbf{1} & -p_1\beta_1\boldsymbol{e}_1, & & \cdots, & -p_N\beta_N\boldsymbol{e}_N \\ \hline -p_1\alpha_1\boldsymbol{e}_1^t & -p_1\beta_1\boldsymbol{e}_1^t & p_1, & 0, & \cdots, & 0 \\ & & 0, & p_2, & \cdots, & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ -p_N\alpha_N\boldsymbol{e}_N^t & -p_N\beta_N\boldsymbol{e}_N^t & 0, & 0, & \cdots, & p_N \end{array}\right]\left[\begin{array}{c} \boldsymbol{a} \\ \hline \boldsymbol{b} \\ \hline d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{array}\right] = \left[\begin{array}{c} \sum p_i\alpha_i\boldsymbol{E}_i \\ \hline \sum p_i\beta_i\boldsymbol{E}_i \\ \hline -p_1\boldsymbol{E}_1^t\boldsymbol{e}_1 \\ -p_2\boldsymbol{E}_2^t\boldsymbol{e}_2 \\ \cdot \\ \cdot \\ \cdot \\ -p_N\boldsymbol{E}_N^t\boldsymbol{e}_N \end{array}\right], \tag{3}$$

with the abbreviations

$$A = \boldsymbol{\alpha}^t\boldsymbol{\alpha} = \boldsymbol{\alpha}^2 = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \tag{12}$$

$$B = \boldsymbol{\beta}^t\boldsymbol{\beta} = \boldsymbol{\beta}^2 = \beta_1^2 + \beta_2^2 + \beta_3^2 \tag{13}$$

$$C = \boldsymbol{\alpha}^t\boldsymbol{\beta} = \boldsymbol{\beta}^t\boldsymbol{\alpha} = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 \tag{14}$$

$$\boldsymbol{H} = \boldsymbol{E}^t + \boldsymbol{e}^t\cdot\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} = (H_{ij}) = (E_{ji} + d_j e_{ji}) \tag{15}$$

The determinant of (11) is

$$\det\begin{pmatrix} A & C \\ C & B \end{pmatrix} = AB - C^2 = \boldsymbol{\alpha}^2\boldsymbol{\beta}^2 - (\boldsymbol{\alpha}\cdot\boldsymbol{\beta})^2 = \boldsymbol{\varphi}^2$$

$$= \varphi_1^2 + \varphi_2^2 + \varphi_3^2 =: \varphi^2 > 0 \tag{16}$$

so the solution is unique, namely

$$\boldsymbol{a} = \varphi^{-2}\boldsymbol{H}\cdot(\boldsymbol{\beta}\times\boldsymbol{\varphi}) \qquad \boldsymbol{b} = \varphi^{-2}\boldsymbol{H}\cdot(\boldsymbol{\varphi}\times\boldsymbol{\alpha}) \tag{17),(18}$$

Thus we can find the values of $\boldsymbol{a}$, $\boldsymbol{b}$ and d_1, d_2, d_3 in (5) by the formulae (8), (17), and (18) in turn, where the auxiliary matrices $\boldsymbol{\varphi}$, $\boldsymbol{R}$, and $\boldsymbol{H}$ are defined in (2), (9), and (15), respectively. –

Note that only once (before the iteration) the (3,3)-matrix $\boldsymbol{e}$ has to be inverted while all the other operations are trivial.

where

$$A = \sum p_i\alpha_i^2, \tag{4}$$

$$B = \sum p_i\beta_i^2, \tag{5}$$

$$C = \sum p_i\alpha_i\beta_i. \tag{6}$$

If we introduce the (3,3)-matrices

$$\boldsymbol{F}_i = \boldsymbol{e}_i\boldsymbol{e}_i^t - \mathbf{1} = \begin{pmatrix} x_i^2 - 1 & x_i y_i & x_i z_i \\ x_i y_i & y_i^2 - 1 & y_i z_i \\ x_i z_i & y_i z_i & z_i^2 - 1 \end{pmatrix} \tag{7}$$

we derive from (3) the following six equations for the orbital elements $a_1, a_2, a_3;\ b_1, b_2, b_3$:

$$\left(\begin{array}{c|c} \sum p_i\alpha_i^2\boldsymbol{F}_i & \sum p_i\alpha_i\beta_i\boldsymbol{F}_i \\ \sum p_i\alpha_i\beta_i\boldsymbol{F}_i & \sum p_i\beta_i^2\boldsymbol{F}_i \end{array}\right)\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} = \begin{pmatrix} \sum p_i\alpha_i\boldsymbol{F}_i\boldsymbol{E}_i \\ \sum p_i\beta_i\boldsymbol{F}_i\boldsymbol{E}_i \end{pmatrix} = \begin{pmatrix} \sum p_i\alpha_i\boldsymbol{e}_i\times(\boldsymbol{E}_i\times\boldsymbol{e}_i) \\ \sum p_i\beta_i\boldsymbol{e}_i\times(\boldsymbol{E}_i\times\boldsymbol{e}_i) \end{pmatrix} \tag{8}$$

The matrix on the left hand side of (8) is a system of normal equations and therefore symmetric and (in general) positive definite and well-conditioned. It can be inverted by one of the standard procedures of Linear Algebra (e.g. that of Cholesky, cf. Isaacson/Keller (1966) or Jordan-Engeln/Reutter (1978)). We need not discuss this point further. Once $\boldsymbol{a}$ and $\boldsymbol{b}$ are known, it is quite trivial to calculate the d_i from (3). We find

$$d_i = \boldsymbol{e}_i \cdot (\alpha_i \boldsymbol{a} + \beta_i \boldsymbol{b} - \boldsymbol{E}_i) \tag{9}$$

which completes the solution of (1) or (3). ((9) might equally well be found from (2.8) directly). –

Of course, this method is applicable to the case $N=3$ too, but the calculations of the last section are simpler. The reason for this is that only three distances are to be found, therefore the d_i should be obtained first.

5. Remarks

At last we want to discuss the precision of the results of our method. Of course, the orbit which is determined in Sect. 3 passes exactly through the three observed positions. But (Sect. 4) if $N \geqq 4$ the problem is overdetermined and only a least-square-solution of (2.8) is possible. That means that the observed direction $\boldsymbol{e}_i$ and the direction to the planet derived from the elements $\boldsymbol{a}$ and $\boldsymbol{b}$ (given by $\boldsymbol{r}_i - \boldsymbol{E}_i$) may be different by an angle of, say, ε_i $(i=1, \ldots, N)$.

It can be shown that the calculations of Sect. 4 are equivalent to the solution of the minimum problem

$$\sum_{i=1}^{N} p_i d_i^2 \sin^2 \varepsilon_i = \min! \tag{1}$$

(i.e.: (1) leads to just the same system of normal equations).

The difference between the sine of ε_i and the angle itself is completely negligible, so (1) is very nearly the same as

$$\sum_{i=1}^{N} p_i d_i^2 \varepsilon_i^2 = \min! \tag{2}$$

The error ε_i of the i-th measurement is thus given a weight $p_i d_i^2$ instead of p_i. We may correct for the very slight error introduced by this simply if we replace p_i during the iteration by $p_i d_i^{-2}$.

Finally it is to be noted that – to the best of my knowledge – our method is the first one which allows to determine an orbit from N observations directly, without first choosing 3 measurements (cf. Sect. 4). So – as far as perturbations are vanishingly small – a further improvement of the orbit is not needed if we use all of the observations from the outset.

6. Conclusion

We construct a very simple procedure of orbit determination using $N \geqq 3$ observed directions. If compared with the classical methods it has some advantages:

1. It allows taking into account more than 3 observations with their weights and hence avoids the problem of selecting three measurements. If perturbations of the major planets are to be neglected, there is no need to improve the resulting orbit.

2. All the observations are treated in a symmetrical way. This is not true for most of the classical methods.

3. It does not make any difference between elliptic, parabolic and hyperbolic orbits. In fact, numerical tests show that it works equally well for (minor) planets and comets.

4. It is easily translated into a computer program.

On the other hand, the results of the calculation are the initial values of position and velocity from which the classical elements have to be found afterwards. But this never causes any trouble.

Because the coefficients α_i and β_i do not depend strongly on the elements $\boldsymbol{a}$ and $\boldsymbol{b}$, the iteration should normally converge quite rapidly. Several numerical examples (one is given in the appendix) of known orbits of minor planets and comets show that this is indeed the case, even if the heliocentric arc the planet moves along during the observation is rather large.

Acknowledgements. I am very grateful to Prof. Dr. Jörg Pfleiderer (Innsbruck) who carefully read a preliminary version of this paper and made a large number of comments which were of great value and helped to express some of the arguments more clearly.

References

Bauschinger, J.: 1928, Die Bahnbestimmung der Himmelskörper, Engelmann-Verlag, Leipzig
Bucerius, H.: 1966, Himmelsmechanik, Bibliograph. Inst. Mannheim
Gauß, C.F.: 1871, Theoria motus corporum coelestium in sectionibus conicis solem ambientium (Werke, Bd. 7), Perthes, Gotha
Isaacson, E., Keller, H.B.: 1966, Analysis of Numerical Methods, Wiley & Sons, New York London Sydney
Jordan-Engeln, G., Reutter, F.: 1976, Formelsammlung zur Numerischen Mathematik, Bibliograph. Inst. Mannheim
Klinkerfues, W.: 1899, Theoretische Astronomie, Vieweg & Sohn, Braunschweig
Stracke, G.: 1929, Bahnbestimmung der Planeten und Kometen, Springer-Verlag, Berlin

Appendix: A Numerical Example

As an application of our method we calculate the orbital elements of Ceres. We use the observations of Olbers, Harding and Bessel which are given in Gauß (1871, Chap. 159)

Obs.	Date (Paris Mean Time)	α (1806.0)	δ (1806.0)
1	1805, Sep. 5, $12^h19^m14^s$	95°59′23″.10	+22°21′27″.08
2	1806, Jan. 17, $10^h15^m\ 2^s$	101°18′40″.38	+30°21′24″.20
3	1806, May 23, $9^h33^m18^s$	121°56′ 8″.97	+28° 2′47″.04

For the vectors $\boldsymbol{e}$ and $\boldsymbol{E}$ we find

Obs.	x	y	z	X	Y	Z
1	−0.0964172	0.9951904	−0.0173129	0.9628573	−0.2958452	−0.0000001
2	−0.1692467	0.9773990	0.1266754	−0.4499487	0.8750783	−0.0000001
3	−0.4670685	0.8741417	0.1331285	−0.4760567	−0.8944019	0.0000008

(Lengths in a.u., unit of time: 1 day)

Iterating 7 times we get

It.	a_1	a_2	a_3	b_1	b_2	b_3
1	−0.6869997	2.4097041	0.1927119	−0.0102365	−0.0034614	0.0018869
2	−0.7019348	2.4963064	0.2041542	−0.0102740	−0.0036318	0.0017872
3	−0.6999414	2.4845916	0.2026191	−0.0102653	−0.0036135	0.0017966
4	−0.7001787	2.4859854	0.2028020	−0.0102663	−0.0036158	0.0017954
5	−0.7001498	2.4858160	0.2027797	−0.0102661	−0.0036155	0.0017955
6	−0.7001533	2.4858365	0.2027824	−0.0102661	−0.0036156	0.0017955
7	−0.7001529	2.4858340	0.2027821	−0.0102661	−0.0036155	0.0017955

These values lead to the classical elements:

semimajor axis:	2.7715064 a.u.
excentricity:	0.0823315
longitude of the perihelion:	65°36′39″
ascending node:	80°58′58″
inclination:	10°37′24″
mean anomaly (1806.0):	325°21′42″