orbit of the Sun. This project is greatly enhanced if some of the brighter stars are included in the display. Use data from Appendix G.

24. Project cometary orbits into the ecliptic, and include, optionally, orbits of the major planets.

25. Generalize ephemeris calculations to include range and range-rate. Include satellite-to-satellite observations.

26. We are not yet prepared to deal with the dynamics when more than two bodies are involved. But in many cases some idea of the motion can be generated by "patching together" parts of Keplerian orbits. Suppose that we have the Sun, with mass M, and a planet, with mass m, moving around the Sun in a circular orbit of radius r_1. Then (and this is discussed in section 11.18) the planet can be considered as surrounded by an "activity sphere" of radius $r_1(m/M)^{2/5}$. Now introduce a third body, a comet, say, with negligible mass. Assume that outside the activity sphere, this mass moves in a Keplerian orbit around the Sun, but inside the activity sphere it moves in a Keplerian orbit around the planet. Write a program for calculating orbits in this model. Iteration will be needed to find when the comet enters the activity sphere. For a start, anyway, restrict calculations to two dimensions. Look for cases where an encounter with the planet results in a significant change in the orbital energy of the comet. Interpreting the planet as Jupiter, look for situations in which a parabolic comet is "captured" by Jupiter (turned into a periodic comet with aphelion distance close to the radius of Jupiter's orbit) or expelled from the solar system. Also investigate how a spacecraft launched from the Earth, with aphelion distance r_1, can receive a boost in energy in order to reach, say, Saturn.

27. It has been suggested that the "star of Bethlehem" was Halley's comet, and that the Birth took place during the summer at about the time of the comet's appearance. Using the following elements (Ref. 71), discuss the hypothesis on the basis of geometrical reasoning. $T = -11$, Oct. 10.849, $q = 0.587\,20$, $e = 0.967\,37$, $\omega = 92°.544$, $\Omega = 35°.191$, $i = 163°.584$. (Note: T is given according to the Julian calendar.)

Chapter 7

The Determination of Orbits

7.1 Introduction

In Chapter Six we have seen how, once the elements of an orbit are known, the geocentric position on the celestial sphere can be calculated for any time. In this chapter we shall be concerned with the reverse situation, that of finding the elements of an orbit from observations. For convenience we shall refer to the observed body as a comet, but it could equally well be a minor planet or an interplanetary spacecraft; with slight modifications to the methods used, it could be an artificial satellite. The roughest glance at the process of ephemeris computation described in Section 6.16 will show that the work cannot practically be reversed, so that some new technique must be found. In fact no direct way is known for finding the elements of an orbit from observations, and it is necessary to proceed by approximations.

We divide the subject into two parts: finding a *preliminary*, approximate orbit from a minimum of observations, and, secondly, using many observations to improve the orbit. The second part is by far the more important; in fact the detailed analysis of many observations yields not only a *definitive* orbit, but also knowledge of the dynamical environment in which the motion takes place; i.e., the relevant forces, gravitational or otherwise. This is too big an area for the present text, and will be only briefly looked at in sections 7.4–7.6.

The classical problem of the determination of orbits is to find a preliminary orbit of a newly discovered comet or minor planet using data from three observations, each one comprising time, right ascension and declination. It has been argued forcibly that this is no longer an important problem; but it is one that still evokes a lot of interest, and it can be solved, elegantly and without drudgery, on the microcomputer. It is none the worse for having a long and sometimes romantic history. The potential solver faces some danger. Solution may be impossible if the observations are too close together in time, or if they are unfavorably placed geometrically relative to each other. The situation is complicated by the fact that the observations will contain errors. How do you locate the "position" of a fuzzy object with a little tail? All methods of solution

proceed by successive approximations. So in this work you will, with inaccurate data, be using mathematical steps that may not converge to find a solution that may not exist. With that understood, let us be optimists, relax and enjoy the subject. Otherwise, skip to section 7.4!

There are two traditional approaches to the subject, following the approaches of Laplace and Gauss. Laplace's method will be discussed, but without a program, although details of a program should be fairly clear. In section 7.3 two methods that have evolved from Gauss' approach will be described, with programing discussed.

An observer on another star would recognize the bodies in the solar system as moving in elliptic orbits about the Sun; but observations from the Earth are affected by the motion of the Earth. The observed geocentric path will obviously not be an ellipse, and Figure 7.1, showing part of the path of comet Arend-Roland (1956 h), demonstrates how complicated the observed path will become. The position of the Earth in the solar system at any time is, of course, accurately known. If we could observe the distance of the comet, then there would be no difficulty in calculating its position in the solar system; unfortunately only its direction can be observed, and the calculation of its distance is one of the processes of orbit determination. In astrodynamics more and different information may be available from observations. The processes of orbit determination can be modified (and simplified) to take advantage of this extra information.

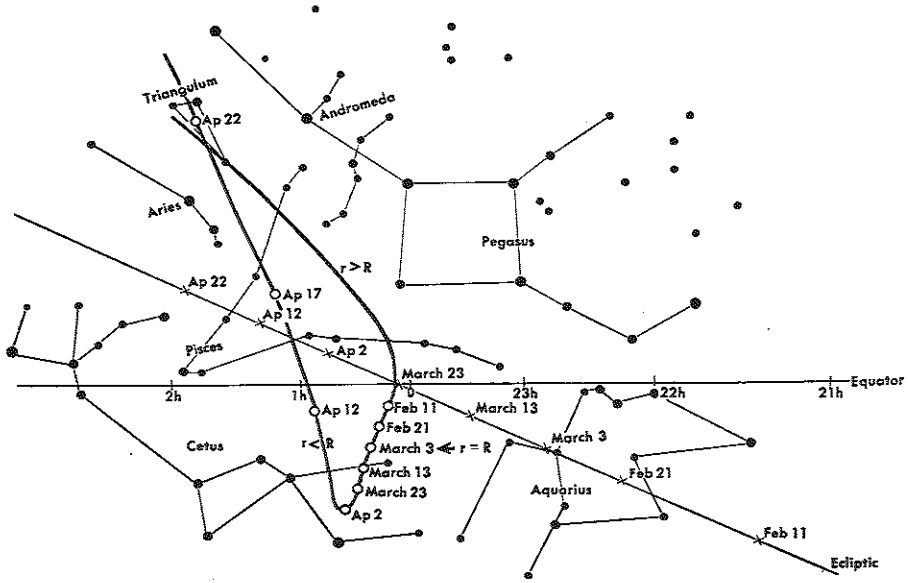

Figure 7.1 *The chart is a central projection, so that all great circles appear as straight lines.*

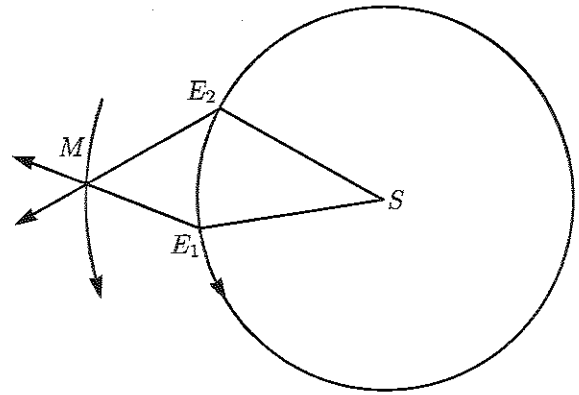

Figure 7.2

For interest we shall describe in principle the method used by Kepler to find the distance, and thence the orbit, of Mars. The sidereal period of Mars was accurately known, and Figure 7.2 shows the situation for two observations separated by one sidereal period, so that Mars has returned to the same position in the solar system. Since the sidereal period is 1.88 years, the Earth will have revolved through approximately 677°, so that the angle E_1SE_2 is known, as is the distance E_1E_2 (but only in terms of the astronomical unit). Observations furnish the angles E_1E_2M and E_2E_1M, so that the triangle E_1E_2M can be solved to find the lengths of the sides, and ultimately the distance MS.

The price of the simplicity of Kepler's method is that observations are needed over many revolutions, and this is a luxury that we cannot afford. The history of the discovery of Ceres will illustrate this. Ceres, the first of the minor planets to be discovered, was found by Piazzi in 1801, but only a few observations were possible before it approached conjunction and became too close to the Sun to be observed. Ceres is a faint object, and it was obviously important to predict when and where it could be observed again; this prediction could not be based on the leisurely study of several revolutions but had to depend on a small arc of one revolution. In this case the occasion was doubly historic because Gauss evolved a new method for orbit determination; the principles of this will be described later.

A single observation yields two angles and the time. Since six unknowns must be found before the orbit is determined, a minimum of three observations is necessary. An accurate orbit, the *definitive orbit*, is found from many more observations, but since three are enough, we shall be concerned here with the problem of determining the orbit from these.

It is instructive at this stage to consider the observed path geometrically. Throughout this chapter we shall use the notation of Figure 7.3, where S, E, and C apply to the Sun, the Earth, and the comet, respectively.

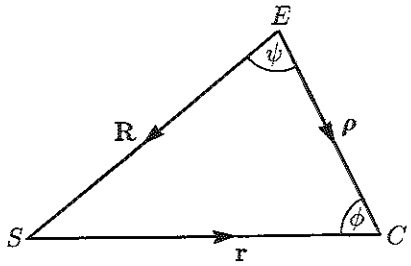

Figure 7.3

The observed quantity is the unit vector $\hat{\boldsymbol{\rho}}$, and it traces out a curve on the celestial sphere (which we assume to have unit radius). Then

$$\frac{d\hat{\boldsymbol{\rho}}}{ds} = \hat{\boldsymbol{\rho}}' = \hat{\mathbf{t}},$$

where s is the distance measured along the curve and $\hat{\mathbf{t}}$ is the unit vector tangent to the curve at $\hat{\boldsymbol{\rho}}$. (The prime will denote differentiation with respect to s in this section only.) Let

$$\hat{\mathbf{n}} = \hat{\boldsymbol{\rho}} \times \hat{\mathbf{t}}.$$

Since $\hat{\mathbf{t}}'$ is perpendicular to $\hat{\mathbf{t}}$, it can be resolved along $\hat{\boldsymbol{\rho}}$ and $\hat{\mathbf{n}}$, so that

$$\hat{\mathbf{t}}' = \lambda\hat{\boldsymbol{\rho}} + \kappa\hat{\mathbf{n}},$$

where, by differentiating $\hat{\boldsymbol{\rho}} \cdot \hat{\mathbf{t}} = 0$, we see that $\lambda = -1$. The component of $\hat{\mathbf{t}}'$ at right angles to the line of sight is $\kappa\hat{\mathbf{n}}$, and Figure 7.4 shows the situation where κ is positive; in this case the curve is concave toward the direction $\hat{\mathbf{n}}$.

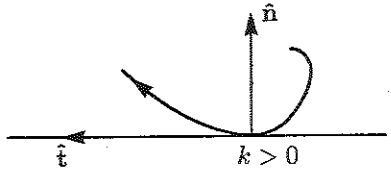

Figure 7.4

When κ is negative, the curve is convex toward $\hat{\mathbf{n}}$. Now

$$\mathbf{r} + \mathbf{R} = \boldsymbol{\rho} = \rho\hat{\boldsymbol{\rho}},$$

so

$$\begin{aligned}\dot{\mathbf{r}} + \dot{\mathbf{R}} &= \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\dot{\hat{\boldsymbol{\rho}}} \\ &= \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\frac{ds}{dt}\hat{\mathbf{t}}, \quad (dt \text{ refers to the time})\end{aligned}$$

and

$$\ddot{\mathbf{r}} + \ddot{\mathbf{R}} = \ddot{\rho}\hat{\boldsymbol{\rho}} + 2\dot{\rho}\frac{ds}{dt}\hat{\mathbf{t}} + \rho\frac{d^2s}{dt^2}\hat{\mathbf{t}} + \rho\left(\frac{ds}{dt}\right)^2(-\hat{\boldsymbol{\rho}} + \kappa\hat{\mathbf{n}}).$$

Multiply by $\cdot\,\hat{\mathbf{n}}$; then

$$(\ddot{\mathbf{r}} + \ddot{\mathbf{R}}) \cdot \hat{\mathbf{n}} = \rho\left(\frac{ds}{dt}\right)^2\kappa. \tag{7.1.1}$$

Now both the Earth and the comet move subject to the gravitational attraction of the Sun, and the mass of each can be neglected compared with the mass of the Sun. Hence

$$\ddot{\mathbf{r}} = -\mu\frac{\mathbf{r}}{r^3} \quad \text{and} \quad \ddot{\mathbf{R}} = -\mu\frac{\mathbf{R}}{R^3},$$

and so

$$\begin{aligned}\ddot{\mathbf{r}} + \ddot{\mathbf{R}} &= -\mu\left(\frac{\mathbf{r}}{r^3} + \frac{\mathbf{R}}{R^3}\right) \\ &= -\mu\left(\frac{\boldsymbol{\rho} - \mathbf{R}}{r^3} + \frac{\mathbf{R}}{R^3}\right).\end{aligned}$$

Multiply by $\cdot\,\hat{\mathbf{n}}$; then from (7.1.1) we find

$$\mu\left(\frac{1}{r^3} - \frac{1}{R^3}\right)\mathbf{R} \cdot \hat{\mathbf{n}} = \rho\left(\frac{ds}{dt}\right)^2\kappa. \tag{7.1.2}$$

Now suppose κ is positive; then $(r - R)$ and $\mathbf{R} \cdot \hat{\mathbf{n}}$ have opposite signs. If $r > R$, then the direction of the Sun makes an angle of more than 90° with $\hat{\mathbf{n}}$, and the curve is convex toward the Sun. If $r < R$, then the curve is concave toward the Sun. These results are not altered if κ is negative. Hence we have Lambert's theorem that the apparent path is convex toward the Sun when $r > R$ and concave toward the Sun when $r < R$. This is illustrated in Figure 7.1.

From the triangle SEC we have

$$r^2 = \rho^2 + R^2 - 2R\rho\cos\psi,$$

and, formally, this and (7.1.2) can be solved for the two unknowns r and ρ. But this will be impossible if $\kappa = 0$; then, the three observations will lie on a great circle, and the six quantities will not be independent, so more than three observations are needed. There will be trouble if κ is small. In fact the ability to determine an orbit from three observations depends upon the extent to which the observed arc departs from a great circle.

7.2 Laplace's Method

The basic formulas for Laplace's method are

$$\mathbf{r} + \mathbf{R} = \rho\hat{\boldsymbol{\rho}}, \tag{7.2.1}$$

$$\mathbf{r}' + \mathbf{R}' = \rho'\hat{\boldsymbol{\rho}} + \rho\hat{\boldsymbol{\rho}}', \tag{7.2.2}$$

$$-\frac{\boldsymbol{\rho}}{r^3} + \left(\frac{1}{r^3} - \frac{1}{R^3}\right)\mathbf{R} = \rho''\hat{\boldsymbol{\rho}} + 2\rho'\hat{\boldsymbol{\rho}}' + \rho\hat{\boldsymbol{\rho}}'', \tag{7.2.3}$$

and

$$r^2 = \rho^2 + R^2 - 2R\rho\cos\psi. \tag{7.2.4}$$

Here we have chosen the unit of time to make $k = 1$, and the unit of mass to be the mass of the Sun. If the modified time is referred to as τ, measured from some epoch t_0, then

$$\tau = k(t - t_0).$$

A prime will denote differentiation with respect to τ.

The observations furnish three values of $\hat{\boldsymbol{\rho}}$ for three values of τ. The four formulas given above represent exactly the geometrical and gravitational aspects of the motion. The initial approximation in Laplace's method involves finding values for $\hat{\boldsymbol{\rho}}'$ and $\hat{\boldsymbol{\rho}}''$ at some instant. Let this instant be t_0; then for small τ, $\hat{\boldsymbol{\rho}}$ can be expanded by a Taylor series as

$$\hat{\boldsymbol{\rho}} = (\hat{\boldsymbol{\rho}})_{\tau=0} + \tau(\hat{\boldsymbol{\rho}}')_0 + \tfrac{1}{2}\tau^2(\hat{\boldsymbol{\rho}}'')_0 + \cdots \tag{7.2.5}$$

If this series is truncated after the third term, then three observations are necessary to determine $(\hat{\boldsymbol{\rho}}')_0$ and $(\hat{\boldsymbol{\rho}}'')_0$. These values are not exact, so that the observed positions (and therefore the geometry of the problem) will not be represented accurately. It is best to choose t_0 as the arithmetic mean of the three times, so that the errors are minimized. (See problems 10 and 11 at the end of the section.) A simpler, but less satisfactory choice is to put t_0 equal to time of the central observation. If more than three observations are available, we can, of course, find better values for $(\hat{\boldsymbol{\rho}}')_0$ and $(\hat{\boldsymbol{\rho}}'')_0$.

Let us assume that $\hat{\boldsymbol{\rho}}'$ and $\hat{\boldsymbol{\rho}}''$ have been found. Taking $(\hat{\boldsymbol{\rho}} \times \hat{\boldsymbol{\rho}}') \cdot$ (7.2.3), we find

$$\left(\frac{1}{r^3} - \frac{1}{R^3}\right)[\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\rho}}', \mathbf{R}] = \rho[\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\rho}}', \hat{\boldsymbol{\rho}}''], \tag{7.2.6}$$

an equation which is similar to (7.1.2). Also from $(\hat{\boldsymbol{\rho}} \times \hat{\boldsymbol{\rho}}'') \cdot$ (7.2.3), we find

$$\left(\frac{1}{r^3} - \frac{1}{R^3}\right)[\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\rho}}'', \mathbf{R}] = -2\rho'[\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\rho}}', \hat{\boldsymbol{\rho}}'']. \tag{7.2.7}$$

Assuming that (7.2.4) and (7.2.6) can be solved for r and ρ, (7.2.7) can then be solved for ρ'. Then from (7.2.1) and (7.2.2), $\mathbf{r}$ and $\mathbf{r}'$ can be found, and the elements of the orbit can be calculated from these.

If these elements are then used to predict the positions at the times of observation the answers will not agree with the original observations. This is because we truncated the series (7.2.5). Some method must be found to improve the orbit, but before discussing this, we shall consider in more detail the solution of (7.2.4) and (7.2.6) from a theoretical angle; their practical solution is described in problem 4 at the end of this section.

By eliminating r, we can obtain an algebraic equation of the eighth degree in ρ, but rather than attempting to solve this, it is easier to make the following substitutions. From the triangle SEC we have (see Figure 7.3)

$$\frac{R}{\sin\phi} = \frac{r}{\sin\psi} = \frac{\rho}{\sin(\phi+\psi)}. \tag{7.2.8}$$

Write (7.2.6) in the form

$$\rho = A\left(\frac{1}{R^3} - \frac{1}{r^3}\right); \tag{7.2.9}$$

then, using (7.2.8), we find

$$R\sin\psi\cos\phi + \left(R\cos\psi - \frac{A}{R^3}\right)\sin\phi = -\frac{A}{R^3}\frac{\sin^4\phi}{\sin^3\psi}.$$

in order to simplify this, let

$$\left.\begin{aligned} N\sin m &= R\sin\psi, \\ N\cos m &= R\cos\psi - \frac{A}{R^3}, \\ M &= -\frac{NR^3}{A}\sin^3\psi, \end{aligned}\right] \tag{7.2.10}$$

where the sign N is chosen to make M positive. Substituting into (7.2.9) and simplifying, we find, eventually,

$$\sin^4\phi = M\sin(\phi + m). \tag{7.2.11}$$

The quantities M and m are known; assuming that a unique solution exists, (7.2.11) can be solved by successive approximations. (See problem 3 at the end of this section.)

It can easily be verified that the position of the observer satisfies the equations so that $\phi = \pi - \psi$ is a solution that must be rejected. Also, from the geometry of the triangle SEC,

$$\phi < \pi - \psi.$$

The solutions of (7.2.11) are the intersections of the curves

$$y_1 = \sin^4\phi$$

and

$$y_2 = M\sin(\phi + m).$$

Figure 7.5 shows these curves for m negative and M somewhat less than one. In this case there are three solutions, and in general we shall have either one or three real solutions for ϕ lying between 0 and π. But if there is only one

solution, it must be $\phi = \pi - \psi$; then no solution is left for the comet and the problem has no meaning.

Consider the case when A is positive. Then $r > R$, and since $\sin\psi$ is not negative, N must be negative; then m lies in the third or fourth quadrant. Figure 7.5 shows m in the fourth quadrant; as m decreases, the curve y_2 slides to the right and there will be a critical value of m, after which there is only one real solution (see problem 5). Since we require three roots, m certainly cannot be in the third quadrant. Similarly, if A is negative, it can be shown that m must certainly lie in the first quadrant.

Let the three roots be ϕ_1, ϕ_2, and ϕ_3, where $\phi_1 \le \phi_2 \le \phi_3$. If $\phi_1 = \pi - \psi_1$ the problem has no solution. If $\phi_2 = \pi - \psi$, there is a unique solution, ϕ_1; this can be found without difficulty by solving (7.2.11). If $\phi_3 = \pi - \psi$, there are two possible solutions, ϕ_1 and ϕ_2. It might be possible to judge between these, since one might require a solution for r which would be unreasonably large, but it may be necessary to use a fourth observation. In the latter case (7.2.11) might be formed for two choices of three dates from the four, having a common central date; the solution common to both would be used. (Or see problem 12 at the end of this section.)

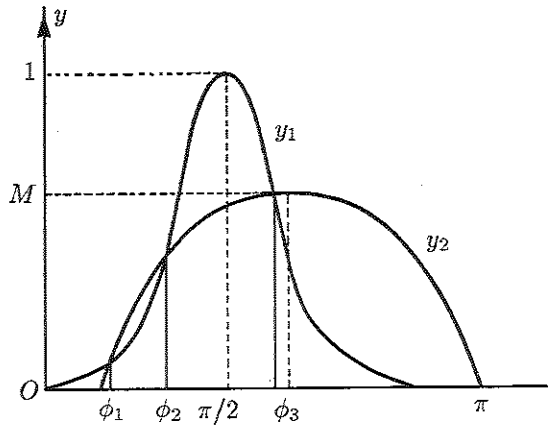

Figure 7.5

To find the condition for a unique solution, consider

$$F(\phi) \equiv \sin^4\phi - M\sin(\phi + m),$$

so that

$$\frac{\partial F}{\partial \phi} \equiv 4\sin^3\phi\cos\phi - M\cos(\phi + m).$$

Suppose A to be positive. If there are three solutions, we see from Figure 7.5 that

$$\frac{\partial F(\phi_2)}{\partial \phi} > 0.$$

The derivatives at the other two roots are negative, so that for a unique solution it is necessary and sufficient that

$$\phi_2 = \pi - \psi$$

and

$$-4\sin^3\psi\cos\psi + M\cos(\psi - m) > 0.$$

Using (7.2.10), this becomes

$$\frac{4MA\cos\psi}{NR^3} + \frac{M}{N}\left\{\cos\psi\left(R\cos\psi - \frac{A}{R^3}\right) + R\sin^2\psi\right\} > 0$$

or

$$\frac{MR}{N}\left(1 + \frac{3A\cos\psi}{R^4}\right) > 0.$$

Then, since N is negative, we have

$$1 + \frac{3A\cos\psi}{R^4} < 0. \tag{7.2.12}$$

It can be shown that the same condition holds if A is negative.

Consider the limiting case when

$$1 + \frac{3A\cos\psi}{R^4} = 0.$$

Eliminating $\cos\psi$ by means of (7.2.4), and using (7.2.9), we find

$$\rho^2 = r^2 + \frac{2}{3}\frac{R^5}{r^3} - \frac{5}{3}R^2. \tag{7.2.13}$$

This is the equation of a surface of revolution about the line SE. A section through SE is shown in Figure 7.6. The sign of the left-hand side of (7.2.12) changes on crossing the surface, and it can be verified that the inequality is satisfied in the shaded areas of the figure.

We shall next consider the problem of improving the approximate orbit. Now that we have approximate values for ρ, we can adjust for parallax and for planetary aberration. But the adjustments will not affect the discrepancies between the computed and observed positions for the first and third times of observation, and the removal of these is the main problem. The method given below is due to Leuschner. We assume t_0, the time of approximation using (7.2.5) to be the central, or second, time of observation.

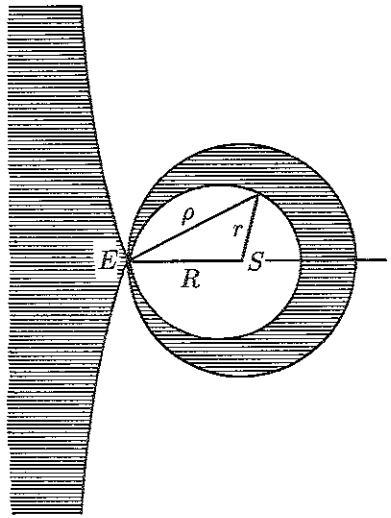

Figure 7.6

To find the predicted right ascension and declination for the two times of observation, we use the f and g functions,

$$\mathbf{r} = f\mathbf{r}_0 + g\mathbf{r}_0'$$

to find the values of $\mathbf{r}$. ($\mathbf{r}_0$ and $\mathbf{r}_0'$ apply to the second observation and are known.) Then we use the formulas (6.16.2) and (6.16.3) to find α and δ. Let the subscripts $_1$ and $_3$ apply to the first and third times of observation, and let $\Delta\alpha_1$ and $\Delta\delta_1$, etc., be the discrepancies between the observed and calculated values in the sense that they are corrections to be added to the calculated values. We then have four residuals, $\Delta\alpha_1$, $\Delta\delta_1$, $\Delta\alpha_3$, and $\Delta\delta_3$.

The errors in $\mathbf{r}$ stem from an incorrect value of ρ, while those in $\mathbf{r}'$ are due to a variety of causes; hence we want to find corrections $\Delta\rho$ and $\Delta\mathbf{r}'$ (four in all) in terms of the four residuals. $\mathbf{R}$ is, of course, not affected; so, from (7.2.1) we have, in general,

$$\Delta\mathbf{r} = \Delta\boldsymbol{\rho} = \Delta\rho\hat{\boldsymbol{\rho}} + \rho\Delta\hat{\boldsymbol{\rho}}. \tag{7.2.14}$$

Multiply by $\cdot\,\hat{\boldsymbol{\rho}}$; then

$$\Delta\rho = \hat{\boldsymbol{\rho}}\cdot\Delta\mathbf{r}$$

so

$$\rho\Delta\hat{\boldsymbol{\rho}} = \Delta\mathbf{r} - \hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}}\cdot\Delta\mathbf{r}). \tag{7.2.15}$$

$\Delta\mathbf{r}$ here is the correction to be applied to $\mathbf{r}$. If we assume that all corrections are of the first order of small quantities, and we neglect their squares and products,

we find, from the equation of motion for the comet,

$$\mathbf{r}'' + \Delta\mathbf{r}'' = -\frac{\mathbf{r}}{r^3} - \frac{\Delta\mathbf{r}}{r^3} + 3\mathbf{r}\frac{\Delta r}{r^4},$$

so that

$$\Delta\mathbf{r}'' = -\frac{\Delta\mathbf{r}}{r^3} + 3\mathbf{r}\frac{\Delta r}{r^4}.$$

But from (7.2.4),

$$r\Delta r = (\rho - R\cos\psi)\Delta\rho.$$

Now let $\Delta\mathbf{r}''$ be evaluated at the time of the second observation, when $\Delta\hat{\boldsymbol{\rho}} = \mathbf{0}$; so $\Delta\mathbf{r} = \hat{\boldsymbol{\rho}}\Delta\rho$. Then

$$\begin{aligned}\Delta\mathbf{r}'' &= \left\{-\frac{\hat{\boldsymbol{\rho}}}{r^3} + \frac{3\mathbf{r}}{r^5}(\rho - R\cos\psi)\right\}\Delta\rho \\ &= \mathbf{H}\Delta\rho, \quad \text{say.}\end{aligned} \tag{7.2.16}$$

Now

$$\Delta\mathbf{r}_1 = \Delta\mathbf{r} + \Delta\mathbf{r}'\tau_1 + \tfrac{1}{2}\Delta\mathbf{r}''\tau_1^2 + \cdots \tag{7.2.17}$$

and, ignoring terms of the order τ_1^3, we have

$$\Delta\mathbf{r}_1 = (\hat{\boldsymbol{\rho}} + \tfrac{1}{2}\mathbf{H}\tau_1^2)\Delta\rho + \tau_1\Delta\mathbf{r}'. \tag{7.2.18}$$

Substituting into (7.2.15), we have the three equations that are components of

$$\rho_1\Delta\hat{\boldsymbol{\rho}}_1 = \Delta\mathbf{r}_1 - \hat{\boldsymbol{\rho}}_1(\hat{\boldsymbol{\rho}}_1\cdot\Delta\mathbf{r}_1) \tag{7.2.19}$$

for the first time of observation, and another three for the third. The residuals $\Delta\hat{\boldsymbol{\rho}}_1$ and $\Delta\hat{\boldsymbol{\rho}}_3$ are known (if we let $\hat{\boldsymbol{\rho}} = (\lambda, \mu, \nu)$, they are $\Delta\lambda_1$, $\Delta\mu_1$, $\Delta\nu_1$, etc.), and it appears for a moment that we have six equations for the four unknowns $\Delta\rho$ and $\Delta\mathbf{r}'$. But the six equations are not independent, and it is better to work with the residuals $\Delta\alpha$ and $\Delta\delta$, where we have

$$\left.\begin{aligned}\cos^2\delta_1\Delta\alpha_1 &= \lambda_1\Delta\mu_1 - \mu_1\Delta\lambda_1, \\ \cos\delta_1\Delta\delta_1 &= \Delta\nu_1, \\ \cos^2\delta_3\Delta\alpha_3 &= \lambda_3\Delta\mu_3 - \mu_3\Delta\lambda_3, \\ \cos\delta_3\Delta\delta_3 &= \Delta\nu_3.\end{aligned}\right] \tag{7.2.20}$$

Substituting from (7.2.19) for the components $\Delta\lambda_1$, $\Delta\mu_1$, $\Delta\nu_1$, and doing the same for the third observation, we have four linear simultaneous equations for the unknowns $\Delta\rho$ and $\Delta\mathbf{r}'$, and these can be solved by the use of a program like the one listed in Appendix E.

If these residuals are applied as corrections to the original values of $\mathbf{r}$ and $\mathbf{r}'$ a better orbit will have been found, but it may still not be successful in predicting the observations for the first and third dates; however, the residuals should be less than they were in the first place. The entire process can now be

repeated and this is continued until the residuals become negligible. It should be noted that once the equations (7.2.20) have been set up, the coefficients of $\Delta\rho$ and $\Delta\mathbf{r}'$ will not vary through the successive approximations, so that later approximations can be accomplished relatively easily.

If extra observations are available, more residuals can be found; we are then able to set up more equations than there are unknowns, and these can be solved by the method of least squares. The resulting corrections to the orbit arise from data of more than three observations and are therefore more dependable than those found above. Observations over a longer arc can be used if, instead of the truncated series (7.2.17), the f and g functions are used, so that

$$\Delta\mathbf{r} = \mathbf{r}_0\Delta f + \mathbf{r}_0'\Delta g + f\Delta\mathbf{r}_0 + g\Delta\mathbf{r}_0'.$$

The resulting equations are more complicated than those described above; they are given by Herget (Ref. 24).

Problems

1. Show how the value of $\cos\psi$ can be calculated directly from the observations and from appropriate tables.

2. Discuss the problem of orbit determination, using (7.2.6) instead of (7.1.2). Show the relationships between the determinants of (7.2.6) and quantities such as κ or $\mathbf{R}\cdot\hat{\mathbf{n}}$.

3. Consider the solution (7.2.11). If an approximate value ϕ_0 is found, show that a correction $\Delta\phi_0$ can be calculated from

$$\Delta\phi_0 = -\frac{\sin^4\phi_0 - M\sin(\phi_0+m)}{4\sin^3\phi_0\cos\phi_0 - M\cos(\phi_0+m)}.$$

   Discuss the case when the denominator is small.

4. Show that the equations (7.2.4) and (7.2.6) can be solved by substituting trial values of ρ into (7.2.4) to find r and then substituting this value of r into (7.2.6) to find ρ. Use this method to solve

$$r^2 = 0.9734 + 1.1493\rho + \rho^2,$$
$$\rho = 2.703 - \frac{2.596}{r^3}.$$

   (The answer is $\rho = 2.631$.)

5. Show that the condition for (7.2.11) to have a double root is

$$4\sin^3\phi\cos\phi - M\cos(\phi+m) = 0.$$

   Hence show that m should satisfy the inequality

$$9 - 16\tan^2 m \geq 0,$$

   for there to be a possible solution to the problem, and find the possible limits of m and the corresponding limits of ϕ. Finally show that the maximum value of M which can result in (7.2.11) having three real roots is approximately 1.431.

6. Derive the inequality (7.2.12) when A is negative.

7. Show that the left-hand side of (7.2.12) changes sign on crossing a boundary in Figure 7.6, and that the solution for ρ is unique in the shaded region of the figure.

8. Find the octic equation for ρ_2 by eliminating r_2 between (7.2.4) and (7.2.6).

9. Consider the solution of (7.2.4) and (7.2.6) as follows: take, for convenience, $R = 1$, and (7.2.6) in the form

$$l\rho = 1 - \frac{1}{r^3}.$$

   Eliminate ρ, obtaining an octic in r, in which the coefficients of r^i for $i = 1, 2, 4, 5, 7$ are zero. Show that if the coefficient of r^8 is positive, then that of r^6 is negative and that of r^3 is positive. Hence, from the theory of equations, show that the octic has three positive, one negative, and four complex roots. Show that $r = 1$ is a real root, and that if the other two positive roots lie on either side of 1, then it is possible to distinguish between them in the solution. Show that the condition for this is that

$$l(l - 3\cos\psi) < 0.$$

10. Suppose we have three observations of λ, λ_1, λ_2, and λ_3, and that values λ_0, λ_0', and λ_0'', for time t_0 are found by assuming

$$\lambda = \lambda_0 + \lambda_0'\tau_0 + \tfrac{1}{2}\lambda_0''\tau_0^2.$$

   Better values, $\lambda_0 + \Delta\lambda_0$, etc., would be found if we could assume

$$\lambda = \lambda_0 + \lambda_0'\tau_0 + \tfrac{1}{2}\lambda_0''\tau_0^2 + \tfrac{1}{6}\lambda_0'''\tau_0^3 + \tfrac{1}{24}\lambda_0''''\tau_0^4.$$

   Find $\Delta\lambda_0$, $\Delta\lambda_0'$, and $\Delta\lambda_0''$ in terms of λ_0''' and λ_0'''', and discuss the errors involved in finding λ_0, λ_0', and λ_0'' from three observations (of which t_0 need not be a time of observation). Show that the error in λ_0'' is minimized if

$$\tau_1 + \tau_2 + \tau_3 = 0$$

   and find the corresponding value of t_0.

repeated and this is continued until the residuals become negligible. It should be noted that once the equations (7.2.20) have been set up, the coefficients of $\Delta\rho$ and $\Delta\mathbf{r}'$ will not vary through the successive approximations, so that later approximations can be accomplished relatively easily.

If extra observations are available, more residuals can be found; we are then able to set up more equations than there are unknowns, and these can be solved by the method of least squares. The resulting corrections to the orbit arise from data of more than three observations and are therefore more dependable than those found above. Observations over a longer arc can be used if, instead of the truncated series (7.2.17), the f and g functions are used, so that

$$\Delta\mathbf{r} = \mathbf{r}_0\Delta f + \mathbf{r}_0'\Delta g + f\Delta\mathbf{r}_0 + g\Delta\mathbf{r}_0'.$$

The resulting equations are more complicated than those described above; they are given by Herget (Ref. 24).

Problems

1. Show how the value of $\cos\psi$ can be calculated directly from the observations and from appropriate tables.

2. Discuss the problem of orbit determination, using (7.2.6) instead of (7.1.2). Show the relationships between the determinants of (7.2.6) and quantities such as κ or $\mathbf{R}\cdot\hat{\mathbf{n}}$.

3. Consider the solution (7.2.11). If an approximate value ϕ_0 is found, show that a correction $\Delta\phi_0$ can be calculated from

$$\Delta\phi_0 = -\frac{\sin^4\phi_0 - M\sin(\phi_0+m)}{4\sin^3\phi_0\cos\phi_0 - M\cos(\phi_0+m)}.$$

   Discuss the case when the denominator is small.

4. Show that the equations (7.2.4) and (7.2.6) can be solved by substituting trial values of ρ into (7.2.4) to find r and then substituting this value of r into (7.2.6) to find ρ. Use this method to solve

$$r^2 = 0.9734 + 1.1493\rho + \rho^2,$$
$$\rho = 2.703 - \frac{2.596}{r^3}.$$

   (The answer is $\rho = 2.631$.)

5. Show that the condition for (7.2.11) to have a double root is

$$4\sin^3\phi\cos\phi - M\cos(\phi+m) = 0.$$

   Hence show that m should satisfy the inequality

$$9 - 16\tan^2 m \geq 0,$$

   for there to be a possible solution to the problem, and find the possible limits of m and the corresponding limits of ϕ. Finally show that the maximum value of M which can result in (7.2.11) having three real roots is approximately 1.431.

6. Derive the inequality (7.2.12) when A is negative.

7. Show that the left-hand side of (7.2.12) changes sign on crossing a boundary in Figure 7.6, and that the solution for ρ is unique in the shaded region of the figure.

8. Find the octic equation for ρ_2 by eliminating r_2 between (7.2.4) and (7.2.6).

9. Consider the solution of (7.2.4) and (7.2.6) as follows: take, for convenience, $R = 1$, and (7.2.6) in the form

$$l\rho = 1 - \frac{1}{r^3}.$$

   Eliminate ρ, obtaining an octic in r, in which the coefficients of r^i for $i = 1, 2, 4, 5, 7$ are zero. Show that if the coefficient of r^8 is positive, then that of r^6 is negative and that of r^3 is positive. Hence, from the theory of equations, show that the octic has three positive, one negative, and four complex roots. Show that $r = 1$ is a real root, and that if the other two positive roots lie on either side of 1, then it is possible to distinguish between them in the solution. Show that the condition for this is that

$$l(l - 3\cos\psi) < 0.$$

10. Suppose we have three observations of λ, λ_1, λ_2, and λ_3, and that values λ_0, λ_0', and λ_0'', for time t_0 are found by assuming

$$\lambda = \lambda_0 + \lambda_0'\tau_0 + \tfrac{1}{2}\lambda_0''\tau_0^2.$$

   Better values, $\lambda_0 + \Delta\lambda_0$, etc., would be found if we could assume

$$\lambda = \lambda_0 + \lambda_0'\tau_0 + \tfrac{1}{2}\lambda_0''\tau_0^2 + \tfrac{1}{6}\lambda_0'''\tau_0^3 + \tfrac{1}{24}\lambda_0''''\tau_0^4.$$

   Find $\Delta\lambda_0$, $\Delta\lambda_0'$, and $\Delta\lambda_0''$ in terms of λ_0''' and λ_0'''', and discuss the errors involved in finding λ_0, λ_0', and λ_0'' from three observations (of which t_0 need not be a time of observation). Show that the error in λ_0'' is minimized if

$$\tau_1 + \tau_2 + \tau_3 = 0$$

   and find the corresponding value of t_0.

11. Suppose we have four observations of λ from which we want to find values of λ_0, λ_0' and λ_0''. Show that the error in λ_0'' is minimized if

$$\sum_{i<j=1}^{4} \tau_i \tau_j = 0.$$

Show that this leads to a quadratic in t_0 with real roots.

12. Given four observations, show that ρ can be found in the following way. From (7.2.6) and (7.2.7) we can find an expression of the form $\rho' = B\rho$; B depends on the derivatives of the direction cosines up to the second. Now, by differentiating this, we find $\rho'' = (B' + B^2)\rho$; B' involves third derivatives, but these can be found from four observations. Now find another relation between ρ and ρ'' (using 7.2.3)) of the form $\rho'' = C + D\rho$. Hence solve for ρ by eliminating ρ''.

13. Find approximations for the first and second derivatives of the direction cosines for the time of the second observation for the following set of observations:

Date				
Nov.	7.8205	+0.902 897	+0.060 606	+0.425 562
Nov.	26.7480	+0.922 476	+0.051 857	+0.382 554
Dec.	18.6262	+0.934 768	+0.080 227	+0.346 080.

(Answer: $\lambda' = +0.047\,391 \quad \mu' = +0.020\,560 \quad \nu' = -0.115\,775$
$\lambda'' = -0.078\,250 \quad \mu'' = +0.291\,339 \quad \nu'' = +0.100\,251.$)

14. If the apparent path of a body on the celestial sphere has a point of inflection, show that the tangent at that point passes through the Sun.

7.3 Gauss' Method

"Gauss' method" is a phrase applied to a class of methods originating with the ideas of Gauss, and further evolved by Gauss and many others. Two modern variants will be described here; they are based on a discussion by B. G. Marsden (Ref. 41) whom I would like to thank for his help and suggestions concerning the material for this section. After nearly two hundred years of polishing in the hands of experts, Gauss' method now forms one of the most beautiful structures in celestial mechanics.

The method starts with the geometrical condition that the vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ lie in a plane. Consequently, there exist scalars c_1 and c_3, independent of the coordinate system used, such that

$$\mathbf{r}_2 = c_1\mathbf{r}_1 + c_3\mathbf{r}_3. \tag{7.3.1}$$

To introduce the dynamics, let

$$\left.\begin{aligned} \mathbf{r}_1 &= f_1\mathbf{r}_2 + g_1\mathbf{v}_2, \\ \mathbf{r}_3 &= f_3\mathbf{r}_2 + g_3\mathbf{v}_2, \end{aligned}\right] \tag{7.3.2}$$

from which

$$\mathbf{r}_2 = c_1(f_1\mathbf{r}_2 + g_1\mathbf{v}_2) + c_3(f_3\mathbf{r}_2 + g_3\mathbf{v}_2). \tag{7.3.3}$$

If this is successively multiplied by $\times\mathbf{v}_2$ and $\mathbf{r}_2\times$, then there result

$$1 = c_1f_1 + c_3f_3 \quad \text{and} \quad 0 = c_1g_1 + c_3g_3,$$

from which

$$\left.\begin{aligned} c_1 &= \frac{g_3}{f_1g_3 - g_1f_3}, \\ \text{and} \\ c_3 &= -\frac{g_1}{f_1g_3 - g_1f_3}. \end{aligned}\right] \tag{7.3.4}$$

Alternative expressions for c_1 and c_3 involve Gauss' sector-triangle ratios. Let $[\mathbf{r}_i, \mathbf{r}_j]$ denote the area of the triangle formed by $\mathbf{r}_i$ and $\mathbf{r}_j$, and $(\mathbf{r}_i, \mathbf{r}_j)$ the orbital area swept out by the radius vector between the two vectors. Then

$$\left.\begin{aligned} c_1 &= \frac{|\mathbf{r}_2 \times \mathbf{r}_3|}{|\mathbf{r}_1 \times \mathbf{r}_3|} = \frac{[\mathbf{r}_2, \mathbf{r}_3]}{[\mathbf{r}_1, \mathbf{r}_3]} \\ \text{and} \\ c_3 &= \frac{|\mathbf{r}_1 \times \mathbf{r}_2|}{|\mathbf{r}_1 \times \mathbf{r}_3|} = \frac{[\mathbf{r}_1, \mathbf{r}_2]}{[\mathbf{r}_1, \mathbf{r}_3]}. \end{aligned}\right] \tag{7.3.5}$$

Let

$$y_1 = \frac{(\mathbf{r}_2, \mathbf{r}_3)}{[\mathbf{r}_2, \mathbf{r}_3]}, \quad y_2 = \frac{(\mathbf{r}_1, \mathbf{r}_3)}{[\mathbf{r}_1, \mathbf{r}_3]}, \quad y_3 = \frac{(\mathbf{r}_1, \mathbf{r}_2)}{[\mathbf{r}_1, \mathbf{r}_2]}. \tag{7.3.6}$$

Then

$$\left.\begin{aligned} c_1 &= \frac{(\mathbf{r}_2, \mathbf{r}_3)}{(\mathbf{r}_1, \mathbf{r}_3)} \frac{[\mathbf{r}_2, \mathbf{r}_3]}{(\mathbf{r}_2, \mathbf{r}_3)} \frac{(\mathbf{r}_1, \mathbf{r}_3)}{[\mathbf{r}_1, \mathbf{r}_3]} \\ &= \frac{(t_3 - t_2)}{(t_3 - t_1)} \frac{y_2}{y_1}. \\ \text{In the same way,} \\ c_3 &= \frac{(t_2 - t_1)}{(t_3 - t_1)} \frac{y_2}{y_3}. \end{aligned}\right] \tag{7.3.7}$$

In deriving these formulas we have used the law of areas.

Introducing the observer into (7.3.1), we have

$$c_1\rho_1\hat{\boldsymbol{\rho}}_1 - \rho_2\hat{\boldsymbol{\rho}}_2 + c_3\rho_3\hat{\boldsymbol{\rho}}_3 = c_1\mathbf{R}_1 - \mathbf{R}_2 + c_3\mathbf{R}_3. \tag{7.3.8}$$

Similarly, from equations (7.3.2),

$$\left.\begin{aligned} f_1\mathbf{r}_2 + g_1\mathbf{v}_2 &= \rho_1\hat{\boldsymbol{\rho}}_1 - \mathbf{R}_1, \\ \mathbf{r}_2 &= \rho_2\hat{\boldsymbol{\rho}}_2 - \mathbf{R}_2, \\ f_3\mathbf{r}_2 + g_3\mathbf{v}_2 &= \rho_3\hat{\boldsymbol{\rho}}_3 - \mathbf{R}_3. \end{aligned}\right\} \tag{7.3.9}$$

We shall see shortly that one way to determine the orbit is to solve the three scalar equations in (7.3.8) for the three geocentric distances using approximations for c_1 and c_3. These equations will be ill-conditioned; before being treated numerically, they will be changed to a triangular form using a beautiful *geometrical* transformation due to L. E. Cunningham. We introduce axes ξ, η and ζ, with the ξ-axis pointing toward the first observed position (i.e., parallel to $\hat{\boldsymbol{\rho}}_1$, so that $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\rho}}_1$), and so that the direction of the third observation, $\hat{\boldsymbol{\rho}}_3$, intersects the positive η-axis. See Figure 7.7.

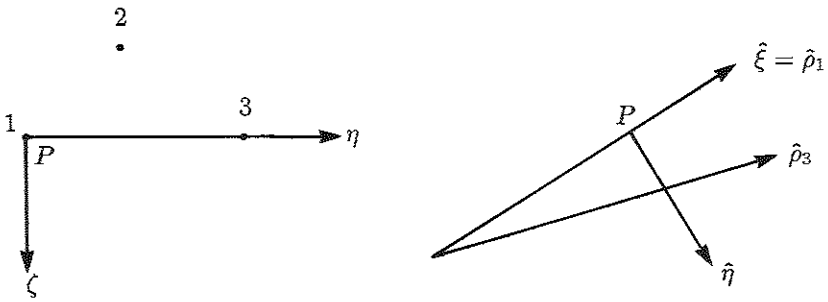

Figure 7.7

Then

$$\begin{aligned} \hat{\boldsymbol{\eta}} &= \frac{\hat{\boldsymbol{\rho}}_1 \times (\hat{\boldsymbol{\rho}}_3 \times \hat{\boldsymbol{\rho}}_1)}{|\hat{\boldsymbol{\rho}}_1 \times (\hat{\boldsymbol{\rho}}_3 \times \hat{\boldsymbol{\rho}}_1)|} \\ &= \frac{\hat{\boldsymbol{\rho}}_3 - \hat{\boldsymbol{\rho}}_1(\hat{\boldsymbol{\rho}}_3 \cdot \hat{\boldsymbol{\rho}}_1)}{\sqrt{1 - (\hat{\boldsymbol{\rho}}_3 \cdot \hat{\boldsymbol{\rho}}_1)^2}}, \end{aligned} \tag{7.3.10}$$

and

$$\hat{\boldsymbol{\zeta}} = \hat{\boldsymbol{\xi}} \times \hat{\boldsymbol{\eta}}. \tag{7.3.11}$$

Let us define a rotation matrix RM having as its *rows* the components of $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\zeta}}$. If some vector is resolved initially in the equatorial system (and

written as a column), then premultiplication by RM will give its components in the new system. All vectors in (7.3.8) and (7.3.9) must be resolved in this way. In particular, the components of the unit vectors in the directions of the observations become

$$\hat{\boldsymbol{\rho}}_1 = \begin{bmatrix} \lambda_1 \\ \mu_1 \\ \nu_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}}_2 = \begin{bmatrix} \lambda_2 \\ \mu_2 \\ \nu_2 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\rho}}_2 \cdot \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\rho}}_2 \cdot \hat{\boldsymbol{\eta}} \\ \hat{\boldsymbol{\rho}}_2 \cdot \hat{\boldsymbol{\zeta}} \end{bmatrix}, \quad \hat{\boldsymbol{\rho}}_3 = \begin{bmatrix} \lambda_3 \\ \mu_3 \\ \nu_3 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\rho}}_3 \cdot \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\rho}}_3 \cdot \hat{\boldsymbol{\eta}} \\ 0 \end{bmatrix}. \tag{7.3.12}$$

(In the program listing that follows, the equatorial components of these vectors form the rows of a matrix `RH`. The components just found are called `L(I)`, `M(I)` and `N(I)` for `I = 1,2,3`.)

The first treatment of these equations is characterized by Marsden as the "Gauss-Encke-Merton method" or simply by GEM, giving credit to the principal contributors. The operations start with approximations to c_1 and c_3. The most primitive approximation uses just the opening terms in the f and g series, (6.7.11). Let

$$\tau_1 = k(t_1 - t_2), \quad \tau_2 = k(t_3 - t_1), \quad \tau_3 = k(t_3 - t_2). \tag{7.3.13}$$

Then for the first approximation, put

$$c_1 = \frac{\tau_3}{\tau_3 - \tau_1}, \quad c_3 = -\frac{\tau_1}{\tau_3 - \tau_1}. \tag{7.3.14}$$

Gauss' method has been criticized (see Ref. 30) as being weak, or even invalid, because of the limited convergence of the f and g *series*, expanded in powers of the time. But these series are irrelevant for most purposes today. After an initial approximation, such as (7.3.14), needed to get things started, the rigorous expressions (7.3.7) for c_1 and c_3 will be used.

We address equation (7.3.8). From the ζ-component, using (7.3.12), we have

$$-\rho_2\nu_2 = c_1Z_1 - Z_2 + c_3Z_3.$$

So ρ_2 can be found in terms of c_1 and c_3 as

$$\rho_2 = \frac{-c_1Z_1 + Z_2 - c_3Z_3}{\nu_2}. \tag{7.3.15}$$

Note the division by ν_2. The magnitude of ν_2 is a measure of the departure of the observed arc from a great circle; if it is too small, then there is little hope of finding a useful orbit from the observations. From the η-component of (7.3.8) we find

$$\rho_3 = \frac{\rho_2\mu_2 + c_1Y_1 - Y_2 + c_3Y_3}{c_3\mu_3}. \tag{7.3.16}$$

Finally, from the ξ-component,

$$\rho_1 = \frac{\rho_2\lambda_2 - c_3\rho_3\lambda_3 + c_1X_1 - X_2 + c_3X_3}{c_1}. \tag{7.3.17}$$

We have simply solved (7.3.8), transformed into triangular form, by back substitution.

With these values for the geocentric distances, the heliocentric vectors $\mathbf{r}_i$ can be found from $\mathbf{r}_i = \boldsymbol{\rho}_i - \mathbf{R}_i$. The program must have a subroutine similar to the program in section 6.12 for calculating the sector triangle ratio, y. Then (7.3.6) and (7.3.7) are used to find new values for c_1 and c_2, and the cycle of calculations is repeated. The operation may end in disaster; it is not a bad idea to have the program exhibit the latest geocentric distances and to ask "do you want me to continue?" Then the operator might say "yes," or give it up, or insert geocentric distances of his own. Once the process is converging toward the correct solution, that convergence is slow but linear. Therefore it can be speeded up using Steffensen's method: three successive values of c_1 and c_3 are needed, and then a formula based on (6.6.17) is used to find the next pair of values. Be warned that the process may converge to a spurious solution with *negative* geocentric distances. It may then be necessary to start the iteration, not using (7.3.14), but with plausible values for the geocentric distances. Marsden warns that early iterations can be wild, with fluctuating signs for some geocentric distances, before finally settling down.

Once the process has been considered to have converged sufficiently, take values of $\mathbf{r}_i$ that go with the sector-triangle ratio most recently calculated. Then use this information to find the f and g functions relating the two times, and thence the velocity at one of the times. Finally, the elements are found, using the components of position and velocity.

The following listing contains the essential components of the GEM procedure using (7.3.14) to start the iterations.

```
10       REM THIS PROGRAM APPLIES THE GEM METHOD FROM THE TEXT.
20       REM THE LISTING DOES NOT INCLUDE THE CALCULATION OF THE
30       REM SECTOR-TRIANGLE RATIO (SEE SECTION 6.12) NOR THE
40       REM CALCULATION OF ELEMENTS, GIVEN COMPONENTS OF
50       REM POSITION AND VELOCITY. THE DECLARATIONS OF DOUBLE
60       REM PRECISION VARIABLES AND INTEGERS HAVE ALSO BEEN
70       REM OMITTED.
80       REM
90       REM YOU ARE ASSUMED TO HAVE ENTERED VALUES OF
100      REM PARAMETERS SUCH AS PI, GK, THE GAUSSIAN
110      REM GRAVITATIONAL CONSTANT, AND OBL, THE
120      REM OBLIQUITY OF THE ECLIPTIC.
130      REM
140      REM YOU ARE ASSUMED TO HAVE ENTERED THE TIME,
150      REM T(I), RIGHT ASCENSION, A(I), AND DECLINATION,
160      REM D(I), FOR I = 1 TO 3. CONVERT ANGLES TO RADIANS.
170      REM THE TIMES MIGHT BE CONVERTED TO JULIAN DATES.
180      REM
190      REM FIND DIRECTION COSINES IN EQUATORIAL COORDINATES.
200 FOR I = 1 TO 3
210    RH(I,1) = COS(D(I))*COS(A(I))
220    RH(I,2) = COS(D(I))*SIN(A(I))
230    RH(I,3) = SIN(D(I))
240 NEXT I
250     REM NEXT, SET UP THE MATRIX RM.
260 W1 = 0#
270 FOR J = 1 TO 3
280    RM(1,J) = RH(1,J)
290    W1 = W1 + RH(1,J)*RH(3,J)
300 NEXT J : REM LOOP FOR XI.
310 W2 = SQR(1# - W1*W1)
320 FOR J = 1 TO 3
330    RM(2,J) = (RH(3,J) - W1*RH(1,J))/W2
```



```
340 NEXT J : REM LOOP FOR ETA. SEE (7.3.10).
350 RM(3,1) = RM(1,2)*RM(2,3) - RM(1,3)*RM(2,2)
360 RM(3,2) = RM(1,3)*RM(2,1) - RM(1,1)*RM(2,3)
370 RM(3,3) = RM(1,1)*RM(2,2) - RM(1,2)*RM(2,1)
380     REM COMPONENTS OF ZETA. SEE (7.3.11).
390     REM NEXT, FIND THE DIRECTION COSINES LAMBDA,
400     REM MU AND NU GIVEN IN (7.3.12).
410 FOR I = 1 TO 3
420    L(I) = 0# : M(I) = 0# : N(I) = 0#
430 NEXT I
440 L(1) = 1#
450 FOR I = 1 TO 3
460    L(2) = L(2) + RM(1,I)*RH(2,I)
470    L(3) = L(3) + RM(1,I)*RH(3,I)
480    M(2) = M(2) + RM(2,I)*RH(2,I)
490    M(3) = M(3) + RM(2,I)*RH(3,I)
500    N(2) = N(2) + RM(3,I)*RH(2,I)
510 NEXT I
520     REM ENTER AND ROTATE SOLAR COORDINATES, PUTTING
530     REM INTO THE ARRAY SC(I,J), ROW BY ROW.
540 FOR I = 1 TO 3
550    REM ENTER OR CALCULATE SOLAR COORDINATES
560    REM FOR THE TIME T(I). THESE SHOULD BE
570    REM CORRECTED FOR THE POSITION OF THE OBSERVER,
580    REM USING (6.17.1) AND (6.17.2). CALL THESE
590    REM SUN(1), SUN(2), AND SUN(3).
600    FOR J = 1 TO 3
610       SC(I,J) = 0#
620       FOR K = 1 TO 3
630          SC(I,J) = SC(I,J) + RM(J,K)*SUN(K)
640       NEXT K
650    NEXT J
660 NEXT I
670 T1 = GK*(T(1) - T(2))
680 T3 = GK*(T(3) - T(2)) : REM (7.3.13).
690 C1 = T3/(T3 - T1)
700 C3 = - T1/(T3 - T1)
710     REM THESE ARE INITIAL APPROXIMATIONS, USING (7.3.14).
720     REM NEXT, FIND NEW GEOCENTRIC DISTANCES, AND THE SUM
730     REM OF THE ABSOLUTE VALUES OF (OLD - NEW) GEOCENTRIC
740     REM DISTANCES, DENOTED BELOW BY W2. IT MAY BE
750     REM NECESSARY TO INITIALIZE THESE TO ZERO AT THE
760     REM START OF THE PROGRAM.   PROVIDED THE GEOCENTRIC
770     REM DISTANCES ARE REASONABLE, THE TIMES SHOULD BE
780     REM MODIFIED USING (7.3.26).
790 W1 = (- C1*SC(1,3) + SC(2,3) - C3*SC(3,3))/N(2)
800 W2 = ABS(W1 - GD(2))
810 GD(2) = W1 : REM (7.3.15).
820 W1 = (GD(2)*M(2) + C1*SC(1,2) - SC(2,2)
       + C3*SC(3,2))/(C3*M(3))
830 W2 = W2 + ABS(W1 - GD(3))
840 GD(3) = W1 : REM(7.3.16).
850 W1 = (GD(2)*L(2) - C3*GD(3)*L(3) + C1*SC(1,1)
       - SC(2,1) + C3*SC(3,1))/C1
860 W2 = W2 + ABS(W1 - GD(1))
870 GD(1) = W1 : REM (7.3.17)
880 FOR I = 1 TO 3
890    PRINT GD(I)
900 NEXT I
910 IF W2 < .00001# THEN 1380
920     REM THE PROCESS HAS CONVERGED SUFFICIENTLY,
930     REM AND THE ELEMENTS ARE TO BE FOUND. THE CHOICE
940     REM OF THE "SMALL QUANTITY" IS YOURS.
950     REM
960     REM FIND THE HELIOCENTRIC COORDINATES.
970 FOR I = 1 TO 3
980    X(I,1) = GD(I)*L(I) - SC(I,1)
990    X(I,2) = GD(I)*M(I) - SC(I,2)
1000   X(I,3) = GD(I)*N(I) - SC(I,3)
1010 NEXT I
1020    REM NEXT FIND THE SECTOR-TRIANGLE RATIOS
1030    REM DEFINED IN (7.3.6).
1040 R1 = 0# : R2 = 0# : KAY = 0#
1050 FOR J = 1 TO 3
1060    R1 = R1 + X(2,J)*X(2,J)
```

```
1070    R2 = R2 + X(3,J)*X(3,J)
1080    KAY = KAY + X(2,J)*X(3,J)
1090 NEXT J
1100 R1 = SQR(R1) : R2 = SQR(R2)
1110 KAY = SQR(2*KAY + 2#*R1*R2)
1120 DT = T(3) - T(2)
1130 GOSUB 4000 : Y1 = YG1
1140 R1 = 0# : R2 = 0# : KAY = 0#
1150 FOR J = 1 TO 3
1160    R1 = R1 + X(1,J)*X(1,J)
1170    R2 = R2 + X(3,J)*X(3,J)
1180    KAY = KAY + X(1,J)*X(3,J)
1190 NEXT J
1200 R1 = SQR(R1) : R2 = SQR(R2)
1210 KAY = SQR(2#*KAY + 2#*R1*R2)
1220 DT = T(3) - T(1)
1230 GOSUB 4000 : Y2 = YG1
1240 R1 = 0# : R2 = 0# : KAY = 0#
1250 FOR J = 1 TO 3
1260    R1 = R1 + X(1,J)*X(1,J)
1270    R2 = R2 + X(2,J)*X(2,J)
1280    KAY = KAY + X(1,J)*X(2,J)
1290 NEXT J
1300 R1 = SQR(R1) : R2 = SQR(R2)
1310 KAY = SQR(2#*KAY + 2#*R1*R2)
1320 DT = T(2) - T(1)
1330 GOSUB 4000 : Y3 = YG1
1340 PRINT "RATIOS " ; Y1, Y2, Y3
1350 C1 = (Y2/Y1)*((T(3) - T(2))/(T(3) - T(1)))
1360 C3 = (Y2/Y3)*((T(2) - T(1))/(T(3) - T(1))) : REM (7.3.7)
1370 GOTO 790
1380    REM FIND HELIOCENTRIC COORDINATES IN ECLIPTIC
1390    REM COORDINATES FOR TIMES T(1) AND T(2).
1400 X = 0# : X2 = 0#
1410 FOR K = 1 TO 3
1420    X = X + RM(K,1)*X(1,K)
1430    X2 = X2 + RM(K,1)*X(2,K)
1440 NEXT K
1450 Y = 0# : Y2 = 0#
1460 FOR K = 1 TO 3
1470    Y = Y + RM(K,2)*X(1,K)
1480    Y2 = Y2 + RM(K,2)*X(2,K)
1490 NEXT K
1500 Z = 0# : Z2 = 0#
1510 FOR K = 1 TO 3
1520    Z = Z + RM(K,3)*X(1,K)
1530    Z2 = Z2 + RM(K,3)*X(2,K)
1540 NEXT K
1550 YT = Y*COS(OBL) + Z*SIN(OBL)
1560 Z =  - Y*SIN(OBL) + Z*COS(OBL) : Y = YT
1570 YT = Y2*COS(OBL) + Z2*SIN(OBL)
1580 Z2 =  - Y2*SIN(OBL) + Z2*COS(OBL) : Y2 = YT
1590 R = SQR(X*X + Y*Y + Z*Z)
1600 F = 1# - 2#*DT*DT*MU/(KAY*KAY*YG1*YG1*R)
1610 G = DT/YG1
1620 XV = (X2 - F*X)/G
1630 YV = (Y2 - F*Y)/G
1640 ZV = (Z2 - F*Z)/G
1650    REM X, Y, Z, XV, YV AND ZV ARE
1660    REM COMPONENTS OF POSITION AND VELOCITY AT TIME T(1).
1670    REM FINALLY, CALL A SUBROUTINE TO CALCULATE THE
1680    REM ELEMENTS.
1690 END
4000 REM SUBROUTINE FOR FINDING THE SECTOR-TRIANGLE RATIO.
```

The other method to be described is called the "Moulton–Väisälä–Cunningham" method by Marsden, or MVC for short. This uses equations (7.3.9) with components again resolved using the rotation matrix RM. In this reference system, let $\mathbf{r}_2$ and $\mathbf{v}_2$ be

$$\mathbf{r}_2 = (\xi_2, \eta_2, \zeta_2) \quad \text{and} \quad \mathbf{v}_2 = (v\xi_2, v\eta_2, v\zeta_2). \tag{7.3.18}$$

First, we shall resolve the equations in an order that will facilitate discussion:

$$\left.\begin{aligned} \xi_2 &= \rho_2\lambda_2 - X_2, \\ \eta_2 &= \rho_2\mu_2 - Y_2, \\ \zeta_2 &= \rho_2\nu_2 - Z_2, \end{aligned}\right] \tag{7.3.19}$$

$$\left.\begin{aligned} f_1\zeta_2 + g_1 v\zeta_2 &= -Z_1, \\ f_3\zeta_2 + g_3 v\zeta_2 &= -Z_3, \end{aligned}\right] \tag{7.3.20}$$

$$\left.\begin{aligned} f_1\eta_2 + g_1 v\eta_2 &= -Y_1, \\ f_3\eta_2 + g_3 v\eta_2 &= \rho_3\mu_3 - Y_3, \\ f_3\xi_2 + g_3 v\zeta_2 &= \rho_3\lambda_3 - X_3, \\ f_1\xi_2 + g_1 v\zeta_2 &= \rho_1 - X_1. \end{aligned}\right] \tag{7.3.21}$$

This method uses the f and g functions rather than the sector-triangle ratios. So (7.3.15) and (7.3.4) are combined to give

$$\rho_2 = \frac{1}{\nu_2}\left(\frac{-g_3 Z_1 + g_1 Z_3}{f_1 g_3 - g_1 f_3} + Z_2\right). \tag{7.3.22}$$

Finally, as initial approximations to the f and g functions,

$$\left.\begin{aligned} f_1 &= 1 - \tfrac{1}{2}\tau_1^2 r_2^{-3}, \quad g_1 = (\tau_1 - \tfrac{1}{6}\tau_1^3 r_2^{-3})/k, \\ f_3 &= 1 - \tfrac{1}{2}\tau_3^2 r_2^{-3}, \quad g_3 = (\tau_3 - \tfrac{1}{6}\tau_3^3 r_2^{-3})/k. \end{aligned}\right] \tag{7.3.23}$$

The basic quantity to be varied in the iteration is ρ_2, the geocentric distance at time t_2. So that method of solution starts with an estimate of ρ_2. This might be found from (7.3.15) as in the GEM method, or it might be an educated guess. From then on, there is a choice of routes that can be taken:

A. If the observations are close enough together in time for the approximations in (7.3.23) to be adequate, then the following steps might be taken

A1. Start with a value for ρ_2.

A2. Find $\mathbf{r}_2$ from (7.3.19).

A3. Find f_1, g_1, f_3, and g_3 for (7.3.23).

A4. Either

A4(a). Find ρ_2 from (7.3.22), return to A1, and interate; the process of iteration might be speeded up through the use of Steffensen's method.

or

A4(b). Find separate values for the quantity $v\zeta_2$ from the two equations in (7.3.20). The difference between these is a function of the value of ρ_2 used in A1:

$$\Delta(\rho_2) = \frac{-Z_1 - f_1\zeta_2}{g_1} + \frac{Z_3 + f_3\zeta_2}{g_3}. \tag{7.3.24}$$

The equation $\Delta(\rho_2) = 0$ can then be solved using the secant method.

A5. Once a satisfactory value for ρ_2 has been found, the successive application of the members of (7.3.21) will provide values for the components of $\mathbf{v}_2$.

An approximate orbit has now been found. This might be improved by the use of f and g functions, as described under B.

B. The feature of this approach is that after an initial approximation (B2), the f and g functions are used. These steps may or may not be used after those of A.

B1. Start with a value for ρ_2.

B2. Take steps A2 and A3.

B3. Take an average value of the discrepant values of $v\zeta_2$ found from (7.3.20)

$$v\zeta_2 = \left(\frac{-Z_1 - f_1\zeta_2}{g_1}\tau_3 + \frac{-Z_3 - f_3\zeta_2}{g_3}\tau_1\right)\frac{1}{\tau_1 + \tau_3}. \tag{7.3.25}$$

B4. Take the steps of A5.

B5. Using the f and g functions, find values for f_1, g_1, f_3, and g_3.

B6. Find ρ_2 from (7.3.22) and ζ_2 from the last equation in (7.3.19).

B7. Return to B3, and iterate. As before, if successive values for ρ_2 are progressing smoothly, then the process can be speeded up through the use of Steffensen's method. If an iteration is not going well, the operator can start again with a different initial guess for ρ_2. In this respect, the method is more versatile than GEM. Universal variables should be used for the computation of the f and g functions even if the final orbit is confidently expected to be elliptic. During an erratic iteration, hyperbolic orbits are likely to appear temporarily.

In either method, as soon as estimates are found for the geocentric distances, the times should be corrected for planetary aberration by subtracting

$$\delta t_{\text{days}} = 0.005\,768\rho, \tag{7.3.26}$$

ρ being measured in astronomical units. Right from the start the solar coordinates should be corrected for the position of the observer (the *topocentric* correction) using (6.17.1) and (6.17.2).

You will have seen that many different routes can be taken in solving this problem of orbit determination. There are, of course, many more that have not been mentioned here. For example, if an orbit is known to be nearly parabolic, the condition ($\alpha = 0$) can be imposed to stabilize the iterations. This sort of device introduces an element of art to the process. When you are learning these methods, enjoy them, but do not expect that the automation of just one method will solve every problem.

7.4 Herget's Method for a Preliminary Orbit Using More Than Three Observations.

An orbit determined from three observations might satisfy those observations precisely, but show an alarming discrepancy with a fourth observation. This is because any observation has an associated error. We can only begin to "smooth out" the effects of the errors when many observations are treated. Eventually, the "best" orbit to be found will not satisfy any one observation precisely, but the residuals will be distributed among the observations in a way that is satisfactory statistically.

Herget's method (Ref. 42) is a compromise; two observations (which are chosen by the operator) are satisfied precisely, and the residuals are distributed among the remainder according to the method of least squares. Consider a set of n observations, (α_i, δ_i) at times t_i, $i = 1, 2, \ldots, n$. These do *not* need to be any special order. For these times we have

$$\left.\begin{aligned}
t_1 &: \quad \mathbf{r}_1 = \rho_1\hat{\boldsymbol{\rho}}_1 - \mathbf{R}_1,\\
t_2 &: \quad \mathbf{r}_2 = \rho_2\hat{\boldsymbol{\rho}}_2 - \mathbf{R}_2,\\
&\vdots\\
t_i &: \quad \mathbf{r}_i = \rho_i\hat{\boldsymbol{\rho}}_i - \mathbf{R}_i,\\
&\vdots\\
t_{n-1} &: \mathbf{r}_{n-1} = \rho_{n-1}\hat{\boldsymbol{\rho}}_{n-1} - \mathbf{R}_{n-1},\\
t_n &: \quad \mathbf{r}_n = \rho_n\hat{\boldsymbol{\rho}}_n - \mathbf{R}_n.
\end{aligned}\right\} \tag{7.4.1}$$

The observations at times t_1 and t_n are to be satisfied precisely, and iterations are to be made by varying the quantities ρ_1 and ρ_n.

If ρ_1 and ρ_n are assigned values, then $\mathbf{r}_1$ and $\mathbf{r}_n$ can be found. Then the velocity at one of the times, say, t_1, can be found using one of the methods of sections 6.11-6.13, and the f and g functions used to find $\mathbf{r}_i = \mathbf{r}(t_i)$ at the remaining times. (Herget suggests using the sector-triangle ratios for all the times; but the course just outlined seems to be more direct, and uses programs that you already have.) The values calculated for $\mathbf{r}_i$ will lead to inconsistencies when substituted into (7.4.1). To quantify these, introduce the unit vectors

$$\left.\begin{aligned}
\mathbf{A}_i &= [-\sin\alpha_i, \cos\alpha_i, 0],\\
\mathbf{D}_i &= [-\sin\delta_i\cos\alpha_i, -\sin\delta_i\sin\alpha_i, \cos\delta_i],
\end{aligned}\right\} \quad i = 2, 3, \ldots, (n-1). \tag{7.4.2}$$

Note that for each i these form an orthogonal set with $\hat{\boldsymbol{\rho}}_i$. Define

$$\left.\begin{aligned}
P_i &= P_i(\rho_1, \rho_n) = (\mathbf{r}_i + \mathbf{R}_i)\cdot\mathbf{A}_i,\\
Q_i &= Q_i(\rho_1, \rho_n) = (\mathbf{r}_i + \mathbf{R}_i)\cdot\mathbf{D}_i,
\end{aligned}\right\} \quad i = 2, 3, \ldots, (n-1). \tag{7.4.3}$$

Were we dealing with precise observations and a correct orbit, the P_i and Q_i would all be zero. As it is, each will be a non-zero "residual". We need to

find values of ρ_1 and ρ_n so that these residuals are, firstly, as small as possible, and, secondly, distributed in a way that is statistically reasonable.

To start the discussion, assume for the moment that the observations are exact, so that ρ_1 and ρ_n are to be found to satisfy the equations

$$\left.\begin{aligned} P_i(\rho_1,\rho_n) &= 0,\\ Q_i(\rho_1,\rho_n) &= 0,\end{aligned}\right] \quad i=2,3,\ldots,(n-1). \tag{7.4.4}$$

We shall use Newton's method. Starting with approximate values ρ_1^a and ρ_n^a, we must find corrections $\Delta\rho_1$ and $\Delta\rho_n$ so that $\rho_1=\rho_1^a+\Delta\rho_1$ and $\rho_n=\rho_n^a+\Delta\rho_n$ reduce the P_i and Q_i. Ideally, we would like to have

$$\left.\begin{aligned} P_i(\rho_1^a+\Delta\rho_1,\ \rho_n^a+\Delta\rho_n) &= 0,\\ Q_i(\rho_1^a+\Delta\rho_1,\ \rho_n^a+\Delta\rho_n) &= 0,\end{aligned}\right] \quad i=2,3,\ldots,(n-1). \tag{7.4.5}$$

In Newton's method these are linearized, so that $\Delta\rho_1$ and $\Delta\rho_n$ are found from

$$\left.\begin{aligned} P_i(\rho_1^a,\rho_n^a)+\frac{\partial P_i}{\partial\rho_1}\Delta\rho_1+\frac{\partial P_i}{\partial\rho_n}\Delta\rho_n &= 0,\\ Q_i(\rho_1^a,\rho_n^a)+\frac{\partial Q_i}{\partial\rho_1}\Delta\rho_1+\frac{\partial Q_i}{\partial\rho_n}\Delta\rho_n &= 0,\end{aligned}\right] \quad i=2,\ldots,(n-1). \tag{7.4.6}$$

To deal with these equations, two questions must be answered: how do we find values for the partial derivatives, and how do we deal with $(2n-4)$ equations when there are only two unknowns?

The derivatives are found from the approximations:

$$\left.\begin{aligned} \frac{\partial P_i}{\partial\rho_1} &\simeq \frac{P_i(\rho_1+\Delta,\rho_n)-P_i(\rho_i-\Delta,\rho_n)}{2\Delta},\\ \frac{\partial P_i}{\partial\rho_n} &\simeq \frac{P_i(\rho_1,\rho_n+\Delta)-P_i(\rho_1,\rho_n-\Delta)}{2\Delta},\\ \frac{\partial Q_i}{\partial\rho_1} &\simeq \frac{Q_i(\rho_1+\Delta,\rho_n)-Q_i(\rho_1-\Delta,\rho_n)}{2\Delta},\\ \frac{\partial Q_i}{\partial\rho_n} &\simeq \frac{Q_i(\rho_1,\rho_n+\Delta)-Q_i(\rho_1,\rho_n-\Delta)}{2\Delta}\end{aligned}\right] \quad i=2,\ldots,(n-1). \tag{7.4.7}$$

The error in these approximations depends on Δ^3. Herget suggests $\Delta=0.1$, with lower values when the geocentric distances are small. If Δ is too small, then the round-off errors incurred in the subtraction of nearly equal quantities can become serious; but with double precision calculation, a value $\Delta=0.001$ should suffice for most purposes. But be careful! To use (7.4.7) a subroutine is needed that computes the P_i and Q_i with input quantities ρ_1 and ρ_n. This subroutine must be run five times, with inputs successively:

$$(\rho_1,\rho_n),\quad (\rho_1+\Delta,\rho_n),\quad (\rho_1-\Delta,\rho_n),\quad (\rho_1,\rho_n+\Delta),\quad (\rho_1,\rho_n-\Delta).$$

Next, write the equations (7.4.6) as

$$\left.\begin{aligned} b_1+a_{11}x_1+a_{12}x_2 &= 0,\\ b_2+a_{21}x_1+a_{22}x_2 &= 0,\\ \vdots\qquad\qquad &\quad\ \vdots\\ b_k+a_{k1}x_1+a_{k2}x_2 &= 0.\end{aligned}\right] \tag{7.4.8}$$

In particular, $x_1=\Delta\rho_1$, $x_2=\Delta\rho_n$ and $k=2n-4$. These are called "equations of condition": they cannot be solved. For any x_1 and x_2, each equation will have a residual, r_i, where

$$r_i(x_1,x_2)=b_i+a_{i1}x_1+a_{i2}x_2,\quad i=1,\ldots,k. \tag{7.4.9}$$

In the method of least squares, the "best" values for x_1 and x_2 are those that make the sum of the squares of the residuals a minimum. Let

$$\begin{aligned} L(x_1,x_2) &= \sum_{i=1}^{k} r_i^2\\ &= \sum_{i=1}^{k}(b_i+a_{i1}+a_{i2}x_2)^2. \end{aligned} \tag{7.4.10}$$

For this to be a minimum, it is necessary that the partial derivatives of L with respect to x_1 and x_2 be zero. Then

$$\sum_{i=1}^{k} a_{i1}(b_i+a_{i1}x_1+a_{i2}x_2)=0,$$

or

$$\left.\begin{aligned} \sum_{i=1}^{k} a_{i1}b_i+x_1\sum_{i=1}^{k}a_{i1}^2+x_2\sum_{i=1}^{k}a_{i1}a_{i2} &= 0,\\ \text{and similarly}\qquad\qquad &\\ \sum_{i=1}^{k} a_{i2}b_i+x_1\sum_{i=1}^{k}a_{i1}a_{i2}+x_2\sum_{i=1}^{k}a_{i2}^2 &= 0.\end{aligned}\right] \tag{7.4.11}$$

These are called the "normal equations" and they are solved for x_1 and x_2 — or for the corrections, $\Delta\rho_1$ and $\Delta\rho_n$.

The procedure can now be summarized. From a set of observations two are chosen to be represented exactly; if in doubt, choose the first and last. Geocentric distances for these times must be estimated. (These might have been inherited from an earlier calculation based on three observations, or from some plausible argument.) The basic subroutine referred to above is now run five times to find the P_i, Q_i and their derivatives. These are substituted into the

normal equations (7.4.11), which are solved for the correction for the geocentric distances. With these corrected values, the whole process is repeated, and repetitions continue until the changes from one iteration to the next can be neglected.

Exercises

1. Set up a program for implementing this method. To debug it, begin with a model where the observations are exact, but start with values of the geocentric distances slightly different from the correct ones. Successively start with poorer initial values to get some idea of the convergence of the method. Once the program has run well with correct observations, introduce observational errors, using a random number generator, controlling the rms value of the generator. (See Appendix F.) Experiment with different rms values. Herget comments that "when the observations are inconsistent, the successive iterations rapidly tend toward the solution $\rho_1 = \rho_n = 0$." Can you confirm this? Try including one very wrong observation.
2. Write equations (7.4.8) in the condensed form $\mathbf{b} + \mathbf{A}\mathbf{x} = \mathbf{0}$. Show that the normal equations can then be written as $\mathbf{A}^T\mathbf{b} + \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{0}$, so that the least squares solution is $\mathbf{x} = -(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$. Suppose now that $\mathbf{x}$ has m components instead of two. Show that the same notation gives the correct least squares solution.

7.5 The Differential Correction of Orbits

This subject merits an entire text; so the discussion that follows will be general and superficial. But I hope that it will contain enough detail to enable a reader to construct and run non-trivial programs.

Fundamental to the discussion is the assumption that we are dealing with some "model" which can be specified by giving numerical values to a set of parameters; through the use of observations, we want to derive improved values for these parameters. In dealing with orbital motion, the model might be that of the problem of two bodies, when the parameters might consist of a set of six Keplerian elements; but in more complicated cases the model might include masses of perturbing bodies, or parameters describing the effects of the non-sphericity of the Earth; or it might include a set of initial conditions, position and velocity, at a starting time. All of the parameters involved in describing the model will be included in a vector $\mathbf{X}$. Using the model, and a numerical set of parameters, we can predict numerical values for a set of observations: this set will make up the vector $\mathbf{Y}$. (The word "vector" is used here in the conventional sense of linear algebra to denote a matrix consisting of a single column.) For the model we assume that there is a subroutine that will calculate $\mathbf{Y}$, given $\mathbf{X}$. This can be described by a flowchart

Flowchart 7.1

or an equation.

$$\mathbf{Y}_c = \mathbf{Y}(\mathbf{X}). \tag{7.5.1}$$

The subscript c in (7.5.1) stands for *calculated*. The corresponding quantities will also be *observed*, and the observed values will be written as $\mathbf{Y}_o$. $\mathbf{Y}_o$ and $\mathbf{Y}_c$ will not be equal; this can be due to errors of observation, errors in the parameters and errors in the model. In the discussion that follows, we assume that the model is correct. We want to use the discrepancy between $\mathbf{Y}_o$ and $\mathbf{Y}_c$ to improve the value of $\mathbf{X}$.

We start the discussion by temporarily assuming that there are no errors of observation. Then the problem of finding $\mathbf{X}$, given $\mathbf{Y}_o$, is equivalent to trying to solve

$$\mathbf{Y}_o = \mathbf{Y}(\mathbf{X}) \tag{7.5.2}$$

for $\mathbf{X}$. The process of solution begins with an estimate for $\mathbf{X}$: call it $\mathbf{X}_0$. From this we calculate values for the observations

$$\mathbf{Y}_c = \mathbf{Y}(\mathbf{X}_0). \tag{7.5.3}$$

We then calculate the *residual*

$$\mathbf{y} = \mathbf{Y}_o - \mathbf{Y}_c. \tag{7.5.4}$$

With a correct model and no observational error, there will be a correct value of $\mathbf{X}$ where

$$\mathbf{X} = \mathbf{X}_0 + \mathbf{x}. \tag{7.5.5}$$

We are to use $\mathbf{y}$ to find an approximation for $\mathbf{x}$, under the assumption that the square and products of the members of the vector $\mathbf{x}$ can be neglected. Now (7.5.2) becomes

$$\mathbf{Y}_o = \mathbf{Y}(\mathbf{X}_0 + \mathbf{x}). \tag{7.5.6}$$

Let

$$\mathbf{J} = \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \tag{7.5.7}$$

be the Jacobian matrix having typical element $\partial Y_i/\partial X_j$, evaluated using the parameters $\mathbf{X}_0$. Then, neglecting $\mathbf{x}^2$, and higher powers, we have the approximation

$$\mathbf{Y}(\mathbf{X}_0 + \mathbf{x}) \simeq \mathbf{Y}(\mathbf{X}_0) + \mathbf{J}\mathbf{x}. \tag{7.5.8}$$

This uses the differential $\mathbf{Jx}$ to approximate the difference $\mathbf{Y}(\mathbf{X}_0+\mathbf{x})-\mathbf{Y}(\mathbf{X}_0)$; hence the phrase "differential correction." The method of correction is Newton's method. If (7.5.8) is taken to be precise, we then have the equation

$$\mathbf{Jx}=\mathbf{y}. \tag{7.5.9}$$

If the number of observations is equal to the number of unknowns (the method of "minimum data") then $\mathbf{J}$ will be square; if, in addition, $\mathbf{J}$ is invertible, then (7.5.9) can be solved for $\mathbf{x}$. The adjusted value $(\mathbf{X}_0+\mathbf{x})$ may need further improvement.

The requirement that $\mathbf{J}$ (if square) be invertible is equivalent to a requirement that the problem (given $\mathbf{Y}_o$, find $\mathbf{X}$) is "well defined." This is well expressed by the phrase that the observations must be able to "see" the components of $\mathbf{X}$, and, one can add, put them into focus. For example, the inclination of the orbital plane of a spectroscopic binary to the plane of the sky (perpendicular to the line of sight) cannot be found from observations of radial velocity; so that inclination must not be included among the parameters to be corrected. If the eccentricity of an orbit is small, then there will be numerical difficulties in trying to locate the position of pericenter. If too short an arc of an orbit is observed, it can be impossible to separate the members of $\mathbf{X}$ and put them into "focus." In each of these cases, there is a geometrical problem that will be revealed by the numerical difficulties encountered in solving (7.5.9).

In the discussion that follows it will be assumed that the numbers of components in $\mathbf{X}$ and $\mathbf{Y}$ are m and n, respectively. Normally, n is much larger than m. The "optics" through which the observations "see" the parameters to be corrected are contained in the matrix $\mathbf{J}$. There are several ways in which to find its components. If the model is that of Keplerian motion, explicit formulas are available. (See Ref. 35). They can be found as the solutions of special differential equations. The simplest way in which to approximate them is that used in the preceding section:

$$\begin{aligned}\frac{\partial Y_i}{\partial X_j}&=J_{ij}\\ &\simeq\frac{1}{2\delta_j}[Y_i(X_1,\ldots,X_{j-1},X_j+\delta_j,X_{j+1},\ldots,X_m)\\ &\quad -Y_i(X_1,\ldots,X_{j-1},X_j-\delta_j,X_{j+1},\ldots,X_m)].\end{aligned}\tag{7.5.10}$$

Suppose that (7.5.1) is invoked by a subroutine called `MODEL(X,Y)`. (The FORTRAN terminology will make the steps more explicit.) Let

$$\bar{\delta}_k=\begin{bmatrix}0\\0\\\vdots\\\delta_k\\\vdots\\0\end{bmatrix}$$

Let us successively `CALL MODEL(`$\mathbf{X}+\bar{\delta}_k$`,Y1)` and `CALL MODEL(`$\mathbf{X}-\bar{\delta}_k$`,Y2)`, and let $\mathbf{y}=\mathbf{Y1}-\mathbf{Y2}$. Using differentials, approximately

$$\begin{bmatrix}J_{11}&J_{12}&\cdots&J_{1k}&\cdots&J_{1m}\\J_{21}&J_{22}&\cdots&J_{2k}&\cdots&J_{2m}\\\vdots&&&&&\vdots\\\vdots&&&&&\vdots\\\vdots&&&&&\vdots\\J_{n1}&J_{n2}&\cdots&J_{nk}&\cdots&J_{nm}\end{bmatrix}\begin{bmatrix}0\\0\\\vdots\\2\delta_k\\\vdots\\0\end{bmatrix}\simeq\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}.$$

Therefore

$$\begin{bmatrix}J_{1k}\\J_{2k}\\\vdots\\J_{nk}\end{bmatrix}\simeq\begin{bmatrix}\frac{y_1}{2\delta_k}\\\frac{y_2}{2\delta_k}\\\vdots\\\frac{y_n}{2\delta_k}\end{bmatrix}\tag{7.5.11}$$

The following listing may help clarify the procedure. The component J_{ik} has been replaced by `JAC(I,K)`.

```
      DO 2 K = 1, N
       X(K) = X(K) + DEL(K)
       CALL MODEL(X,Y1)
       X(K) = X(K) - 2 * DEL(K)
       CALL MODEL(X,Y2)
       DO 1 I = 1, N
1       JAC(I,K) = (Y1(I) - Y2(I)) / (2 * DEL(K))
2      X(K) = X(K) + DEL(K)
```

A judicious choice of the size of the "tweak," `DEL(K)`, may not be easily made; it will depend, among other things, on the order of magnitude of the corresponding component, $X(K)$. If the $\mathbf{X}$ vector includes components of position and velocity, then these may have considerably disparate orders of magnitude. (It can be an advantage to use units internally in a calculation so that these orders of magnitude are similar.) If the order of magnitude of `X(K)` is `M(K)`, then, as a rule of thumb, I recommend trying `DEL(K) = M(K)*10`$^{-4}$; but experimentation may be needed. In particular, over long arcs, smaller `DEL` may be needed.

We can now return to the equations of condition (7.5.9). Since there will be more equations (n) than unknowns (m), it will be impossible to satisfy all of these equations, and they must be "solved" by some compromise, as were the equations (7.4.8) in the preceding section. Again, we use the method of least squares. Before outlining the procedure, let me point out again that we are using a linearized approximation to improve the estimate $\mathbf{X}_0$ of the parameters

of the model; additional improvement may be necessary to compensate for the linearization.

Let us assume that the only reason that the equations of condition cannot be satisfied is that there are unavoidable errors of observation in the $\mathbf{Y}_o$. (So we are assuming that the model itself is correct.) This means that an estimate of the vector $\mathbf{x}$ is equivalent to an estimate of the errors of observations, $\mathbf{v}$, where

$$\mathbf{v} = \mathbf{J}\mathbf{x} - \mathbf{y}. \tag{7.5.12}$$

If $\mathbf{x}$ were given its "correct" value, then (7.5.12) would give the actual errors of observation. But observational errors are samples of random variables, and, through (7.5.1) $\mathbf{x}$ becomes also associated with random variables; we can only *estimate* $\mathbf{x}$; but then we may be able to use statistical reasoning to evaluate the *quality* of the estimate.

Let $\mathbf{x}^*$ be the "best" estimate of $\mathbf{x}$ (subject to assumptions to be described shortly); then we have a "best" estimate of the observational errors, $\mathbf{v}^* = \mathbf{J}\mathbf{x}^* - \mathbf{y}$. We assume each observational error to be a sample value of a random variable with Gaussian probability distribution given by

$$\exp\left(-\tfrac{1}{2}v_i^2/\sigma_i^2\right). \tag{7.5.13}$$

Note that this implies that there is no *bias*, or *systematic error* in the observations.

The probability of the error v_i is given by $\exp\left(-\frac{1}{2}v_i^2/\sigma_i^2\right)$ and the joint probability of the errors $v_1, v_2, \ldots, v_n$, is

$$\exp\left(-\tfrac{1}{2}(v_1^2/\sigma_1^2 + v_2^2/\sigma_2^2 + \cdots + v_n^2/\sigma_n^2)\right).$$

Here we have assumed that there is no correlation between the errors of different observations; this may be too optimistic. The "best" estimate of $\mathbf{v}$ is that which makes the joint probability a maximum, or, equivalently, the sum of the squares

$$\frac{v_1^2}{\sigma_1^2} + \frac{v_2^2}{\sigma_2^2} + \cdots + \frac{v_n^2}{\sigma_n^2}$$

a minimum. The condition for this, found by elementary calculus, again leads to a set of normal equations.

Probably, the numbers σ_i will not be known; but an estimate may be possible of their relative values. Let

$$w_i = \frac{k}{\sigma_i}, \quad i = 1, 2, \ldots, n. \tag{7.5.14}$$

These are *relative weights*. Let the diagonal matrix $\mathbf{W}$ be defined by

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & \cdots\cdots & 0 \\ 0 & w_2 & 0 & \cdots\cdots & 0 \\ 0 & 0 & w_3 & & 0 \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots\cdots & w_n \end{bmatrix}. \tag{7.5.15}$$

Let

$$\tilde{\mathbf{J}} = \mathbf{W}\mathbf{J} \quad \text{and} \quad \tilde{\mathbf{y}} = \mathbf{W}\mathbf{y}. \tag{7.5.16}$$

The "weighted" equations of condition can be written as

$$\tilde{\mathbf{J}}\mathbf{x} = \tilde{\mathbf{y}} \tag{7.5.17}$$

and the normal equations as

$$\tilde{\mathbf{J}}^T\tilde{\mathbf{J}}\mathbf{x} = \tilde{\mathbf{J}}^T\tilde{\mathbf{y}}, \tag{7.5.18}$$

with their solution being the "best" estimate

$$\mathbf{x}^* = (\tilde{\mathbf{J}}^T\tilde{\mathbf{J}})^{-1}\tilde{\mathbf{J}}^T\tilde{\mathbf{y}}. \tag{7.5.19}$$

The quality of this solution will depend on the residuals that result when the solution is substituted into the equations of condition. Define

$$\sigma = \sqrt{\frac{1}{n-m}\sum_{i=1}^{n} w_i^2 v_i^2}. \tag{7.5.20}$$

This is the "rms error of unit weight". ("rms" stands for "root mean square.") Note the appearance of $(n - m)$ in the denominator. If n is less than m, no reasonable solution by these means is possible; if $n = m$, then we have the situation of "minimum data"; a solution may be possible, but no statistical information about that solution can be found. Let the ith diagonal element in $(\tilde{J}^T\tilde{J})^{-1}$ be s_i^2. Then the rms error associated with the estimate x_i^* is σs_i. Two other matrices are of importance. The *covariance matrix* of the estimate is

$$\mathbf{P} = \sigma^2(\tilde{\mathbf{J}}^T\tilde{\mathbf{J}})^{-1}. \tag{7.5.21}$$

If this matrix has typical element p_{ij}, then the *correlation matrix*, $\mathbf{C}$, has typical element $c_{ij} = p_{ij}/(p_{ii}p_{jj})^{1/2}$. The terms of the principal diagonal of $\mathbf{C}$ are all equal to one; if any off-diagonal term approaches one, then this *correlation coefficient* indicates a possible strong correlation (or lack of independence) between the errors of the corresponding observations.

The assumptions of independence, Gaussian distributions, and no bias (or systematic error) in these distributions can weaken the conclusions, which should, anyway, be treated with reserve.

We have used the language of "rms errors." It is common in astronomical applications to use "probable errors". There is nothing intrinsically "probable" about a probable error; it is a number such that, assuming a Gaussian distribution (7.5.13), the error is equally likely to be larger or smaller (in absolute magnitude). For a Gaussian distribution with rms error σ, the probable error is

$$pr \simeq 0.6745\sigma.$$

We finish with an additional FORTRAN listing designed to implement this procedure. The matrix A(I,J) is J^TJ, and B(I) is J^Ty.

```
C    FIND THE RESIDUALS, Y(I), I = 1,...,N, AND THE MATRIX
C    JAC(I,J), I = 1,...,N, J = 1,...,M. ALTERNATIVELY,
C    THESE MIGHT BE FOUND IN THE LOOP OVER INDIVIDUAL
C    OBSERVATIONS.
C    INITIALIZE A(J,K) AND B(J), J,K = 1,...,M, TO ZERO.
     DO 13 I = 1,N
C     APPLY WEIGHTS, W(I).
      DO 11 J = 1,M
   11  JAC(I,J) = W(I)*JAC(I,J)
      Y(I) = W(I)*Y(I)
C     AUGMENT THE MATRICES A AND B.
      DO 12 J = 1,M
       B(J) = B(J) + JAC(I,J)*Y(I)
       DO 12 K = 1,M
   12   A(J,K) = A(J,K) + JAC(I,J)*JAC(I,K)
   13 CONTINUE
C     SOLVE X = (A INVERSE) B. SAVE (A INVERSE).
C     NEXT FIND SIGMA FROM (7.5.20).
      SIGMA = 0
      DO 21 I = 1,N
       V = - Y(I)
       DO 20 J = 1,M
   20   V = V + JAC(I,J)*X(J)
   21  SIGMA = SIGMA + V*V
      SIGMA = DSQRT(SIGMA/DFLOAT(N-M))
C     CALCULATE RMS ERRORS OF THE ESTIMATE.
C     APPLY THE CORRECTIONS, X(J), TO THE INITIAL ESTIMATES
C     OF THE PARAMETERS. IF THE CORRECTIONS ARE CONSIDERED
C     TOO LARGE, THEN ANOTHER ITERATION WILL BE NEEDED.
```

7.5.1 Projects

These are designed so that you can, by stages, learn, program and debug the procedures. The "model" can be that of Keplerian motion. The parameters, X, can be Keplerian elements, or components of position and velocity at some given time. The observations might be geocentric right ascension and declination, or quantities such as range and range-rate.

1. For the first experiment with the Jacobian matrix $\mathbf{J}$, take $\mathbf{X}$ to consist of components of position and velocity at time t_0, and $\mathbf{Y}$ to consist of the same components at time t_1. $\mathbf{J}$ will then be a 6×6 matrix. In this special case (do not generalize) $\mathbf{J}$ is a type of matrix called symplectic. A consequence is that if $\mathbf{J}$ is subdivided into four 3×3 matrices as

$$\mathbf{J} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{W} & \mathbf{Y} \end{bmatrix}, \quad \text{then} \quad \mathbf{J}^{-1} = \begin{bmatrix} \mathbf{Y}^T & -\mathbf{V}^T \\ -\mathbf{W}^T & \mathbf{U}^T \end{bmatrix}.$$

That is, the inverse of $\mathbf{J}$ can be found by a rearrangement of its terms and some changes of sign. Find $\mathbf{J}$; rearrange to find $\mathbf{J}^{-1}$. Calculate the product of the two matrices and verify that one is truly the inverse of the other. Don't go any further until this is correct.

2. This is a continuation of the first project. You will already have computed $\mathbf{Y}$, given $\mathbf{X}$. Now introduce small changes into every component of $\mathbf{X}$, to give $\mathbf{X} + \mathbf{x}$, and find the resulting components of position and velocity at time t_1, and then $\mathbf{y}$, the difference between these numbers and the components of $\mathbf{Y}$. Verify that approximately $\mathbf{Jx} \simeq \mathbf{y}$.

3. Experiment with different "tweak" sizes for the effects on the components of $\mathbf{J}$. Consider cases where $t_1 - t_0$ spans at least one complete revolution.

4. If a tweaked orbit is sufficiently close to the reference, or untweaked orbit, then the linear approximation for the difference between the orbits will be adequate. But if a tweaked orbit strays too far from the reference orbit, the linear approximation is lost. One method of procedure is as follows. If the magnitude of the difference between the orbits exceeds a given amount, then all of the differences (in position and velocity) are reduced by a factor of ten. Then the tweaked orbit remains within the linear region around the reference orbit. When finding derivatives for times after this reduction, the quantities δ_j in (7.5.10) or DEL(K) in the listing must be reduced by a factor of ten. Experiment with this procedure, using orbits extending over several revolutions.

5. Consider the application of $\mathbf{J}$, as formulated above, to the two-point boundary value problem. We want to find the orbit that connects positions $\mathbf{r}_0$ at time t_0 and $\mathbf{r}_1$ at time t_1. We have an approximate value for the initial velocity $\mathbf{v}_0$, and this needs to be improved. $\mathbf{r}_0$ and this approximation for $\mathbf{v}_0$ constitute the vector $\mathbf{X}$. Using these as initial conditions, the value of $\mathbf{r}$ at $\mathbf{t}_1$ can be found; let the difference between this value and $\mathbf{r}_1$ be $\delta\mathbf{r}_1$. Show how $\mathbf{J}$ (more specifically, the components $\mathbf{U}$ and $\mathbf{V}$ from the first project) can be used to find a correction to the approximate velocity at t_0 in terms of $\delta\mathbf{r}_1$.

6. Let $\mathbf{X}$ contain six Keplerian elements and $\mathbf{Y}$ the observations at three different times. $\mathbf{J}$ will be a 6×6 matrix. Start this project with a set of elements $\mathbf{X}_p$ and calculate a set, $\mathbf{Y}_p$, of (correct) observations. Now let $\mathbf{X}$ consist of a set of elements changed slightly but deliberately from $\mathbf{X}_p$. Use this set to find calculated values of the observations, $\mathbf{Y}_c$. If $\mathbf{y} = \mathbf{Y}_p - \mathbf{Y}_c$, then corrections to take $\mathbf{X}$ to $\mathbf{X}_p$ should satisfy $\mathbf{Jx} = \mathbf{y}$. Verify that this is true. Find $\mathbf{x} = \mathbf{J}^{-1}\mathbf{y}$. If $\mathbf{X} + \mathbf{x}$ is close to $\mathbf{X}_p$ but not close enough, another iteration may be necessary. Show how this procedure might be used in orbit determination once an approximate solution is known.

7. Generalize the preceding problem. Let $\mathbf{Y}$ contain observations for more than three times. $\mathbf{J}$ no longer has an inverse, now $\mathbf{x} = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\mathbf{y}$. You will be using the program for setting up the normal equations. Verify that the solution is correct.

8. Again, we generalize the preceding problem. But now, instead of using the "perfect" observations, $\mathbf{Y}_p$, add random errors to each component to produce true observations, $\mathbf{Y}_o = \mathbf{Y}_p + \mathbf{e}$. The numbers in $\mathbf{e}$ can be generated by a subroutine that produces sample values of a Gaussian distribution. See Appendix F. For a start, use just one such distribution, so that all observations share the same rms error, and have, therefore, equal weights. Now your estimate of the corrections to $\mathbf{X}$ can include statistics. Also, the sum of the square of the residuals, v_i, should be such that

$$\sqrt{\frac{1}{n-m}\sum_{i=1}^{n} v_i^2}$$

should be close to the rms error that you inserted into the observations.

9. Continue to generalize. Different observations can be given errors with different rms values. Experiment with the use of different weights in the normal equations. Then introduce a greater variety of types of observation. They will have a greater variety of rms errors, and there will be more scope for weighting.

Note that these projects have all involved situations that have been artificially contrived; that is, we know the "true" answers. I suggest that they are worthwhile in order to gain experience and confidence.

7.6 Using a Previous Estimate: Recursive Methods

At the conclusion of the estimation described in the preceding section, we had the "best" estimate itself, and also the covariance matrix $\mathbf{P}$, of (7.5.21) that contained the statistics of that estimate. The covariance matrix of the errors of observation was

$$\mathbf{Q} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}. \tag{7.6.1}$$

Starting with equations of condition $\mathbf{Jx} = \mathbf{y}$ (7.5.9), the normal equations (7.5.18) would be $\mathbf{J}^T\mathbf{Q}^{-1}\mathbf{Jx} = \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y}$. This assumes that the σ_i are known a priori, and the assumption will be made in the discussion that follows.

Suppose that two mutually independent sets of observations, $\mathbf{Y}_1$ and $\mathbf{Y}_2$, are to be combined. If the sets have covariance matrices of errors $\mathbf{Q}_1$ and $\mathbf{Q}_2$, respectively, then the combined set of observations is contained in the vector

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \tag{7.6.2}$$

having the covariance matrix of errors

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix}. \tag{7.6.3}$$

(The zero matrices here are rectangular.) If the two sets were to be processed separately, then the associated equations of condition would be

$$\mathbf{J}_1\mathbf{x} = \mathbf{y}_1, \quad \text{and} \quad \mathbf{J}_2\mathbf{x} = \mathbf{y}_2, \tag{7.6.4}$$

and the corresponding normal equations would be

$$\mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{J}_1\mathbf{x} = \mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{y}_1, \quad \text{and} \quad \mathbf{J}_2^T\mathbf{Q}_2^{-1}\mathbf{J}_2\mathbf{x} = \mathbf{J}_2^T\mathbf{Q}_2^{-1}\mathbf{y}_2. \tag{7.6.5}$$

Combined, the equations of condition are

$$\begin{bmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \end{bmatrix}\mathbf{x} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \tag{7.6.6}$$

with normal equations

$$\begin{bmatrix} \mathbf{J}_1^T & \mathbf{J}_2^T \end{bmatrix}\begin{bmatrix} \mathbf{Q}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \end{bmatrix}\mathbf{x} = \begin{bmatrix} \mathbf{J}_1^T & \mathbf{J}_2^T \end{bmatrix}\begin{bmatrix} \mathbf{Q}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

or

$$(\mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{J}_1 + \mathbf{J}_2^T\mathbf{Q}_2^{-1}\mathbf{J}_2)\mathbf{x} = \mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{y}_1 + \mathbf{J}_2^T\mathbf{Q}_2^{-1}\mathbf{y}_2. \tag{7.6.7}$$

So, normal equations for the entire set can be derived by adding together the normal equations for the separate sets. But remember that we have assumed that there is no correlation between the two sets.

Next, suppose that the normal equations for the first set have been solved, with the solution

$$\begin{aligned} \mathbf{x}_1^* &= (\mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{J}_1)^{-1}\mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{y}_1 \\ &= \mathbf{P}_1\mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{y}_1, \end{aligned} \tag{7.6.8}$$

where $\mathbf{P}_1$ is the covariance matrix of errors of the estimate $\mathbf{x}_1^*$. Then the normal equations for the first set could be written as

$$\mathbf{P}_1^{-1}\mathbf{x}_1^* = \mathbf{J}_1^T\mathbf{Q}_1^{-1}\mathbf{y}_1$$

and consequently the normal equations for the entire set can be written as

$$(\mathbf{P}_1^{-1} + \mathbf{J}_2^T\mathbf{Q}_2^{-1}\mathbf{J}_2)\mathbf{x} = \mathbf{P}_1^{-1}\mathbf{x}_1^* + \mathbf{J}_2^T\mathbf{Q}_2^{-1}\mathbf{y}_2. \tag{7.6.9}$$

This leads to the following reinterpretation. We have a "best" estimate, $\mathbf{x}_1^*$, together with its statistics, $\mathbf{P}_1$. Then we have a further set of observations, $\mathbf{Y}_2$, and its statistics, $\mathbf{Q}_2$. (7.6.9) tells us how the previous estimate can be incorporated into the use of $\mathbf{Y}_2$ to improve the estimate further.

In some respects one can say that the first set of observations is not required since $\mathbf{x}_1^*$ and $\mathbf{P}_1$ are available. But the discarding of observations is never desirable (their statistics can no longer be tested). Also we are solving a linear problem here, and may eventually have to iterate in order to satisfy the fundamental non-linear generating problem.

Another way to derive this result is to treat the previous estimate as an "observation." We shall change the notation slightly. Suppose that we have an estimate, $\mathbf{x}_0$, together with its statistics, $\mathbf{P}_0$. Now we have a new set of observations, $\mathbf{Y}$, with residuals $\mathbf{y}$ and statistics $\mathbf{Q}$. Treating the initial elements as an observation leads to the equations of condition

$$\mathbf{J}\mathbf{x} = \mathbf{y} \quad \text{and} \quad \mathbf{x} = \mathbf{x}_0$$

or

$$\begin{bmatrix} \mathbf{J} \\ \mathbf{I}_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x}_0 \end{bmatrix}. \tag{7.6.10}$$

The covariance matrix of errors of all the "observations" is

$$\begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0 \end{bmatrix}$$

so that the normal equations are

$$\begin{bmatrix} \mathbf{J}^T & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{J} \\ \mathbf{I}_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{J}^T & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x}_0 \end{bmatrix}$$

or

$$(\mathbf{J}^T\mathbf{Q}^{-1}\mathbf{J} + \mathbf{P}_0^{-1})\mathbf{x} = (\mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y} + \mathbf{P}_0^{-1}\mathbf{x}_0), \tag{7.6.11}$$

which is the equivalent of (7.6.9). Its solution is

$$\mathbf{x}^* = (\mathbf{P}_0^{-1} + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{J})^{-1}(\mathbf{P}_0^{-1}\mathbf{x}_0 + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y}) \tag{7.6.12}$$

with covariance matrix of errors

$$\mathbf{P} = (\mathbf{P}_0^{-1} + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{J})^{-1}. \tag{7.6.13}$$

$\mathbf{P}$ can also be written in the alternative form

$$\mathbf{P} = \mathbf{P}_0 - \mathbf{P}_0\mathbf{J}^T(\mathbf{J}\mathbf{P}_0\mathbf{J}^T + \mathbf{Q})^{-1}\mathbf{J}\mathbf{P}_0. \tag{7.6.14}$$

To show this it is sufficient to show the truth of

$$(\mathbf{P}_0^{-1} + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{J})[\mathbf{P}_0 - \mathbf{P}_0\mathbf{J}^T(\mathbf{J}\mathbf{P}_0\mathbf{J}^T + \mathbf{Q})^{-1}\mathbf{J}\mathbf{P}_0] = \mathbf{I}_m. \tag{7.6.15}$$

So we can write (7.6.12) as

$$\mathbf{x}^* = \mathbf{P}(\mathbf{P}_0^{-1}\mathbf{x}_0 + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y}), \tag{7.6.16}$$

with $\mathbf{P}$ given by (7.6.14).

$\mathbf{X}$ has m components so in (7.6.14) we invert an m-by-m matrix. If $\mathbf{Y}$ has n components, remember that we are using them to improve a previous estimate, so there is no reason why n should not be less than m. So the use of (7.6.14) may look more complicated, but can lead to the inversion of a matrix of lower order. No truly algebraic problems of ill-conditioning are removed, but the arithmetic can be less hazardous.

Now consider the mathematics involved in "updating" the parameters after an estimate, and using the new values to predict the following observations. The whole process starts with an estimate, $\mathbf{X}_0$, which is used to predict the early observations, and as a result we find a correction $\mathbf{x}_0^*$. We then use this best estimate in $\mathbf{Y}_c(\mathbf{X}_0 + \mathbf{x}_0^*)$ to find residuals

$$\mathbf{y}^* = \mathbf{Y}_o - \mathbf{Y}_c(\mathbf{X}_0 + \mathbf{x}_0^*). \tag{7.6.17}$$

The improved parameters should also be used in the calculation of $\mathbf{J}$, to make the updating as efficient as possible. Linearizing (7.6.17),

$$\mathbf{y}^* = \mathbf{Y}_o - \mathbf{Y}_c(\mathbf{X}_0) - \mathbf{J}\mathbf{x}_0^*$$

or

$$\mathbf{y} = \mathbf{J}\mathbf{x}_0^* + \mathbf{y}^*. \tag{7.6.18}$$

Then using (7.6.16) and (7.6.13),

$$\begin{aligned} \mathbf{x}^* &= \mathbf{P}[\mathbf{P}_0^{-1}\mathbf{x}_0^* + \mathbf{J}^T\mathbf{Q}^{-1}(\mathbf{J}\mathbf{x}_0^* + \mathbf{y}^*)] \\ &= (\mathbf{P}_0^{-1} + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{J})^{-1}(\mathbf{P}_0^{-1}\mathbf{x}_0^* + \mathbf{J}^T\mathbf{Q}^{-1}\mathbf{J}\mathbf{x}_0^*) + \mathbf{P}\mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y}^* \\ &= \mathbf{x}_0^* + \mathbf{P}\mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y}^*. \end{aligned} \tag{7.6.19}$$

Using (7.6.14) this can be put into the alternative form

$$\begin{aligned} \mathbf{x}^* &= \mathbf{x}_0^* + [\mathbf{P}_0 - \mathbf{P}_0\mathbf{J}^T(\mathbf{J}\mathbf{P}_0\mathbf{J}^T + \mathbf{Q})^{-1}\mathbf{J}\mathbf{P}_0]\mathbf{J}^T\mathbf{Q}^{-1}\mathbf{y}^* \\ &= \mathbf{x}_0^* + \mathbf{P}_0\mathbf{J}^T[\mathbf{I}_n - (\mathbf{J}\mathbf{P}_0\mathbf{J}^T + \mathbf{Q})^{-1}(\mathbf{J}\mathbf{P}_0\mathbf{J}_T + \mathbf{Q} - \mathbf{Q})]\mathbf{Q}^{-1}\mathbf{y}^* \\ &= \mathbf{x}_0^* + \mathbf{P}_0\mathbf{J}^T(\mathbf{J}\mathbf{P}_0\mathbf{J}^T + \mathbf{Q})^{-1}\mathbf{y}^*. \end{aligned} \tag{7.6.20}$$

$\mathbf{P}$ is found from (7.6.14) as before. Remember again that $\mathbf{y}^*$ is found by using the current best estimate of the parameters.

The formulas just described have no particular merit in most problems of differential correction. But they can be useful if estimates are required in a

hurry while the observations are being made. To carry the point further, we shall introduce time explicitly into the model. Let us suppose that the set of elements to be determined consists of the coordinates in phase space (that is, components of position and velocity), $\mathbf{X}_0$ at time t_0, of the body whose motion is being observed. (It is quite simple to modify the formulas so that $\mathbf{X}_0$ might apply to osculating Keplerian elements at t_0.) Assume observations to be made at times t_1, t_2, ... As a result of observations made prior to time t_k, we have made an estimate $\mathbf{x}^*_{0,k-1}$ with statistics $\mathbf{P}_{0,k-1}$. Using this latest estimate, as we did in (7.6.17), we predict the observations $\mathbf{Y}_c$ for the time t_k, and we also observe $\mathbf{Y}_o$ for the same time, deriving the usual residual $\mathbf{y}_k$. We want to use $\mathbf{y}_k$ to improve the estimate of the conditions at time t_0: i.e., we want to find an estimate $\mathbf{x}^*_{0,k}$ to add to $\mathbf{X}_0$.

From (7.6.20), we than have

$$\mathbf{x}^*_{0,k} = \mathbf{x}^*_{0,k-1} + \mathbf{P}_{0,k-1}\mathbf{J}_k^T(\mathbf{J}_k\mathbf{P}_{0,k-1}\mathbf{J}_k^T + \mathbf{Q}_k)^{-1}\mathbf{y}_k. \tag{7.6.21}$$

Here $\mathbf{Q}_k$ is the covariance matrix of the errors of the observations $\mathbf{Y}_o$ made at time t_k. $\mathbf{J}_k$ is the Jacobian matrix relating small changes in $\mathbf{X}_0$ (or, better, $(\mathbf{X}_0 + \mathbf{x}^*_{k-1})$) to consequent changes in $\mathbf{Y}_c$ at time t_k. Then

$$\begin{aligned}\mathbf{J}_k &= \frac{\partial \mathbf{Y}_c(t_k)}{\partial \mathbf{X}(t_0)} \\ &= \frac{\partial \mathbf{Y}_c(t_k)}{\partial \mathbf{X}(t_k)}\frac{\partial \mathbf{X}(t_k)}{\partial \mathbf{X}(t_0)} \\ &= \mathbf{M}_k\boldsymbol{\Omega}(t_k, t_0).\end{aligned} \tag{7.6.22}$$

Since $\mathbf{Y}_c(t_k)$ is likely to be found from $\mathbf{X}(t_k)$ by elementary formulas, which are easily differentiated, $\mathbf{M}_k$ should be simple to compute.

$$\boldsymbol{\Omega}(t_k, t_0) = \frac{\partial \mathbf{X}(t_k)}{\partial \mathbf{X}(t_0)}$$

is the *matrizant* or *state transition matrix*. Its computation and some of its properties were considered in Project 1 of the preceding section. It will be further discussed in section 11.18. From (7.6.21) and (7.6.22),

$$\mathbf{x}^*_{0,k} = \mathbf{x}^*_{0,k-1} + \mathbf{P}_{0,k-1}\boldsymbol{\Omega}^T(t_k, t_0)\mathbf{M}_k^T\mathbf{N}(t_k, t_0)\mathbf{y}_k, \tag{7.6.23}$$

where

$$\mathbf{N}(t_k, t_0) = [\mathbf{M}_k\boldsymbol{\Omega}(t_k, t_0)\mathbf{P}_{0,k-1}\boldsymbol{\Omega}^T(t_k, t_0)\mathbf{M}_k^T + \mathbf{Q}_k]^{-1}. \tag{7.6.24}$$

The covariance matrix of errors of the estimate can be written as

$$\mathbf{P}_{0,k} = \mathbf{P}_{0,k-1} - \mathbf{P}_{0,k-1}\boldsymbol{\Omega}^T(t_k, t_0)\mathbf{M}_k^T\mathbf{N}(t_k, t_0)\mathbf{M}_k\boldsymbol{\Omega}(t_k, t_0)\mathbf{P}_{0,k-1}. \tag{7.6.25}$$

These formulas can appear very complicated. It is best if you rederive them, justifying each algebraic step. Once you are used to the method, the formulas are not hard to program, and the use of this recursive "filtering" can be very versatile. Methods of this kind are often referred to under the general label "Kalman filter"; they are widely used.

7.6.1 Exercises

1. Justify the identity (7.6.15).
2. Suppose that just one observation is processed, uncorrelated with the preceding observations, and having rms error σ. Show that $\mathbf{J}$ is just a row vector and that
$$\mathbf{P}^{-1} = \sigma^{-2}\mathbf{J}^T\mathbf{J} + \mathbf{P}_0^{-1}.$$
Deduce from this that, for algebraic reasons, the rms errors of the estimated parameters *must decrease* as more observations are processed. This is a weakness in the method. If there is an error in the model, the estimates can actually be getting worse, while the statistics say that the reverse is the case. Further, show that as the rms errors for the estimated quantities become smaller, the ability of later observations to influence the results is diminished. In some operations, over many observations, the P matrix is periodically increased in magnitude.
3. Show that the method might be applied if only the relative weights of the observational errors are initially known. Discuss the ultimate estimation of the true rms errors of observation.
4. Set up a program for the use of formulas (7.6.23) – (7.6.25). At the start, you will have an estimate of $\mathbf{X}$, but no statistics. One way to express this mathematically, is to let the initial $\mathbf{P}$ matrix be diagonal with very large components.
5. The method using equations (7.6.23) – (7.6.25) involves improving initial parameters at the initial time t_0. Suppose that the observations are for a mission having a motive (such as a rendezvous) at a final time t_f. Set up equations so that values of the parameters at t_f are estimated recursively.
6. Consider the following situation. Observations are made at times t_1, t_2, ... As a result of observations made prior to t_k, an estimate has been made of $\mathbf{X}$ (components of position and velocity) at time t_{k-1}; call this $\mathbf{X}^*_{k-1}$. We shall also have its statistics in the covariance matrix $\mathbf{P}^*_{k-1}$. Using this estimate, the value of $\mathbf{X}$ at time t_k will be found (call it $\mathbf{X}'_k$) and also the predicted observation $\mathbf{Y}_c(t_k)$. But $\mathbf{Y}$ will also be observed at time t_k, leading to a residual $\mathbf{y}_k$. We want to use $\mathbf{y}_k$ to find an estimate $\mathbf{X}^*_k$ and its statistics $\mathbf{P}^*_k$. Show that the covariance matrix for $\mathbf{X}'_k$ is
$$\mathbf{P}'_k = \boldsymbol{\Omega}(t_k, t_{k-1})\mathbf{P}^*_{k-1}\boldsymbol{\Omega}^T(t_k, t_{k-1})$$
and that if $\mathbf{X}^*_k = \mathbf{X}'_k + \mathbf{x}^*_k$,
$$\mathbf{x}^*_k = \mathbf{P}'_k\mathbf{M}_k^T(\mathbf{M}_k\mathbf{P}'_k\mathbf{M}_k^T + \mathbf{Q}_k)^{-1}\mathbf{y}_k,$$
and
$$\mathbf{P}^*_k = \mathbf{P}'_k - \mathbf{P}'_k\mathbf{M}_k^T(\mathbf{M}_k\mathbf{P}'_k\mathbf{M}_k^T + \mathbf{Q}_k)^{-1}\mathbf{M}_k\mathbf{P}'_k,$$
with
$$\mathbf{M}_k = \frac{\partial \mathbf{Y}_c(t_k)}{\partial \mathbf{X}(t_k)}.$$