1987CeMec..40..329M

A Cubic Approximation For Kepler's Equation

by Seppo Mikkola

Department of Physics, York University, 4700 Keele Street, Toronto,
Ontario, M3J 1P3, Canada
and
Turku University Observatory, Itäinen Pitkäkatu 1, 20520 Turku, Finland

(Received January 15, 1986; accepted May 19, 1987)

Abstract: We derive a new method to obtain an approximate solution for Kepler's equation. By means of an auxiliary variable it is possible to obtain a starting approximation correct to about three figures. A high order iteration formula then corrects the solution to high precision at once. The method can be used for all orbit types, including hyperbolic. To obtain this solution the trigonometric or hyperbolic functions must be evaluated only once.

1. Introduction

Despite the apparent simplicity of Kepler's equation, it has drawn much attention even during recent years. This interest is explained by the fact that in many problems of orbital dynamics a large number of solutions of this equation must be calculated. Consequently even a litle shortening of the computation is important. Some of the most recent works on Kepler's equation include the explicit solutions found by Siewert and Burniston (1972) and Neutsch and Schüfer (1986), and the systematic study of iterative methods by Burkhart and Danby (1983) and Danby and Burkhart (1983). In a very recent paper Odell and Gooding (1986) suggest an algorithm which, utilizing a carefully chosen starter, converges in two iterations by means of a high order correction formula.

What may be the best method for solving Kepler's equation? Assuming that a good estimate is not known one must first obtain a starter value which is then to be improved by some iterative process. Because it is computational economy that we are pursuing, the initial estimate should be computationally cheap and accurate enough to not require many

Celestial Mechanics **40** (1987), 329–334.
© 1987 *by D. Reidel Publishing Company.*

© Kluwer Academic Publishers • Provided by the NASA Astrophysics Data System

1987CeMec..40..329M

iterative corrections. We believe that our method is actually close to an optimal one: it requires the solution of a cubic equation while the improvement (to a typical precision of 15 or more figures) can be done by using a high order iteration formula just once.

2. General

Consider the solution of the elliptic and hyperbolic Kepler's equations

$$E - e\,\sin(E) = M$$

$$e\,\sinh(E) - F = M. \tag{1}$$

The main difficulty in solving these is essentially due to the behavior of the equations at and near the point $e = 1, M = 0$. Power series expansion gives

$$(1-e)E + \frac{1}{6}eE^3 = M + ..$$

$$(e-1)F + \frac{1}{6}eF^3 = M + .. \tag{2}$$

One easily finds that the partial derivatives of the solution with respect to both e and M are singular at $e = 1, M = 0$. This suggests that an efficient starter must be based on the solution of an equation which coincides with (2) in the small-M limit. However, $\sin(E)$ (or $\sinh(F)$) can not be approximated to a high enough accuracy by a third degree polynomial over the entire region $|E| \leq \pi$ (or $|F| \leq \infty$). The way out we have found is based on the use of the auxiliary variable $s = \sin(E/3)$ (or $s = \sinh(F/3)$). The advantage of this is that $\sin(E)$ is a third degree polynomial in this variable. In the following we derive the solution separately for the elliptic and hyperbolic cases. We also present a unifying formulation useful for all types of orbits.

3. Elliptic case

With the substitution

$$s = \sin(E/3) \tag{3}$$

the new form of Kepler's equation reads

$$3\arcsin(s) - e(3s - 4s^3) = M. \tag{4}$$

Reduction of M to the interval $-\pi \leq M \leq \pi$ is assumed to make s as small as possible. If the series

$$\arcsin(s) = s + \frac{1}{6}s^3 + \frac{3}{40}s^5 + .. \tag{5}$$

© Kluwer Academic Publishers • Provided by the NASA Astrophysics Data System

1987CeMec..40..329M

is truncated after the third order term we obtain the approximate equation

$$3(1-e)s + (4e + \frac{1}{2})s^3 = M. \tag{6}$$

The error is largest at $M = \pi$, however, adding the simple correction term

$$ds = -0.078s^5/(1+e) \tag{7}$$

improves the estimate so that the maximum error is an order of magnitude smaller. This crude estimate for the error is obtained by using (5), the derivative of (4) at $M = \pi$ and adjusting the constant on front of (7). More accurate (and more complicated) correction terms may be easily found but the functioning of the method is not affected.

After obtaining the value of s by solving eq.(6) and adding the correction (7) we can evaluate the eccentric anomaly by means of the formula

$$E = M + e(3s - 4s^3). \tag{8}$$

This starting approximation, which has the maximum relative error $|dE/E| \leq .002$, we thus obtained without a single evaluation of trigonometric functions! However, the cubic equation (6) has to be solved and this requires the calculation of one square root and one cubic root. In fact, if we write

$$\alpha = (1-e)/(4e + \frac{1}{2}), \quad \beta = \frac{1}{2}M/(4e + \frac{1}{2}) \tag{9a}$$

and

$$z = (\beta \pm \sqrt{\beta^2 + \alpha^3})^{1/3} \tag{9b}$$

then

$$s = z - \alpha/z. \tag{9c}$$

The sign of the square root is to be chosen to be the sign of β.

4. Hyperbolic case

The hyperbolic form of Kepler's equation differs in behavior from the elliptic one. However, near $e = 1, M = 0$ it is still much the same and the same basic approximation leads to good results here too. Thus we write

$$s = \sinh(F/3) \tag{10}$$

and have

$$e(3s + 4s^3) - 3\ln(s + \sqrt{1+s^2}) = M. \tag{11}$$

Using the series

$$\ln(s + \sqrt{1+s^2}) = s - \frac{1}{6}s^3 + .. \tag{12}$$

© Kluwer Academic Publishers • Provided by the NASA Astrophysics Data System

1987CeMec..40..329M

we obtain the approximation

$$3(e-1)s+(4e+\frac{1}{2})s^3=M. \tag{13}$$

Here the approximation (12) completely breaks down for large values of s; however, this do not apply to (13) because the third order term $4es^3$ arising from expressing $\sinh(F)$ in terms of s dominates. Thus, for large s the main spurious effect due to the use of (12) is that it replaces the correct asymptotic relation $4es^3 \sim M$ by $(4e+\frac{1}{2})s^3 \sim M$, which is equivalent to a relative error of $\approx \frac{1}{8e}$ in M. This effect can, however, be corrected by properly selecting a correction term (see below) and consequently the break-down of (12) for large s is harmless, while for small s the error is of order s^5. This is the main reason to adopt the auxiliary variable (10) similarly to the elliptic case instead of using directly $\sinh(F)$, which would lead to failure for large values of M and thus would require the use of different formulae in different regions.

To make the correction corresponding to (7), we are obliged to use a more complicated equation than in the elliptic case. By taking into account the error of (12) for both small and large values of s we constructed a simple rational function for the error estimate

$$ds = 0.071s^5/[(1+0.45s^2)(1+4s^2)e]. \tag{14}$$

After adding this correction to s we evaluate the value of F by the formula

$$F = 3\ln(s+\sqrt{1+s^2}). \tag{15}$$

which now gives a precision very similar to that in the elliptic case. The equivalent to (8) i.e. $F = e(3s+4s^3)-M$ is, of course, correct but would increase the error of F due to error in s.

5. Unifying formulation

Let us write

$$y=\sqrt{|a|}s \tag{16}$$

where a is the semi-major axis (negative for hyperbola). A substitution to the elliptic and hyperbolic equations ((6) and (13)) gives in both cases the equation

$$3qy+(4e+\frac{1}{2})y^3=T. \tag{17}$$

Here $q=a(1-e)$ is the pericenter distance and T is the time measured from pericenter passage (in units in which the gravitational constant and the total mass are unity). Expressions for the parameters α and β in eq. (9a) are now $\alpha = q/(4e+\frac{1}{2})$ and $\beta = \frac{1}{2}T/(4e+\frac{1}{2})$, while the solution (9c) is now y instead of s. The correction terms, however, remain different for different orbital types. For the ellipse we have

$$dy = -0.078ys^4/(1+e) \tag{18}$$

© Kluwer Academic Publishers • Provided by the NASA Astrophysics Data System

1987CeMec..40..329M

where, as in the following, $s^2 = y^2/|a|$. For the hyperbola

$$dy = 0.071ys^4/[(1+0.45s^2)(1+4s^2)e]. \tag{19}$$

The unifying anomaly (see eg. Herrick 1972, p.118-120)

$$X = \int_{t_0}^{t} dt/r \tag{20}$$

may be obtained as follows:
(a) for the ellipse:

$$X = \sqrt{a}E = T/a + ey(3 - 4s^2)$$

(b) for the hyperbola:

$$\begin{aligned} X &= \sqrt{-a}F = 3\sqrt{-a}\ln(s + \sqrt{1+s^2}) \\ &= 3y(1 - \frac{1}{6}s^2 + \frac{3}{40}s^4 + ..)(\text{for} \quad s^2 \ll 1) \end{aligned} \tag{21}$$

while the unified form of Kepler's equation reads

$$qX + eX^3c_3(X^2/a) = T \tag{22}$$

where c_3 is a Stumpff-function (see e.g. Stiefel and Scheifele 1971, p.42-51).

As we can see, the corrections are zeros for a parabola and actually (17) is equivalent to (22) in this case (with $X = 3y$).

6. Final correction

The above approximate solutions are accurate to about

$$|dX/X| \leq 2 \times 10^{-3}. \tag{23}$$

This result applies to both elliptic and hyperbolic equations. In the elliptic case, the maximum error occurs near $e = 1, M = 1.5$ and the corresponding figures for the hyperbola are similar. Due to the smallness of the maximal error and because it occurs in a region far from pericenter, it is possible to correct the approximation to a high precision by using a high order formula just once. We adopted the method suggested by Burkhart and Danby (1983): Write the equations for the correction dx in the form $f + f'dx + \frac{1}{2}f''(dx)^2 + .. = 0$; then one may calculate successively higher order approximations from

$$dx_{n+1} = -f/\sum_{\nu=1}^{n+1} \frac{f^{(\nu)}}{\nu!}(dx_n)^{\nu-1}. \tag{24}$$

Accepting dx_4 as the final solution, we always have a relative precision of the order of 10^{-15} or better, while dx_5 gives three more correct figures. One should note that the

© Kluwer Academic Publishers • Provided by the NASA Astrophysics Data System

1987CeMec..40..329M

trigonometric functions were evaluated only once to obtain the solution. In any practical sense, the method thus found is not iterative but a direct method for solving Kepler's equation.

Acknowledgement

This work has been supported by an NSERC (Canada) operating grant to K.A. Innanen.

References

Burkhart, T.M. and Danby, J.M.A.: (1983), *Cel. Mech.*, **31**, 317-328.

Danby, J.M.A. and Burkhart, T.M.: (1983), *Cel. Mech.*, **31**, 95-107.

Herrick, S.: (1972) *Astrodynamics*, Vol.II, Van Nostrand Reinhold Company, London.

Neutch, W. and Schüfer, E.: (1986), *Astroph. and Space Sci.*, **125**, 77-83.

Odell, A.W. and Gooding, R.H.: (1986), *Cel. Mech.*, **38**,307-334.

Stiefel, E.L. and Scheifele, G., (1971), *Linear and Regular Celestial Mechanics*, Springer,Berlin.

Siewert, C.E. and Burniston, E.E.: (1972), *Cel. Mech.*, **6**, 294-304.

© Kluwer Academic Publishers • Provided by the NASA Astrophysics Data System