1995121346

N95- 27767

Kepler Equation Solver

F. Landis Markley
Guidance and Control Branch, Code 712
NASA Goddard Space Flight Center
Greenbelt, MD 20771

Abstract

Kepler's Equation is solved over the entire range of elliptic motion by a fifth-order refinement of the solution of a cubic equation. This method is not iterative, and requires only four transcendental function evaluations: a square root, a cube root, and two trigonometric functions. The maximum relative error of the algorithm is less than one part in 10^{18}, exceeding the capability of double-precision computer arithmetic. Roundoff errors in double-precision implementation of the algorithm are addressed, and procedures to avoid them are developed.

Introduction

Numerical solution of the two-body problem for orbital motion is heavily dependent on efficient solution of Kepler's Equation

$$M = E - e \sin E \tag{1}$$

for the eccentric anomaly E in terms of the eccentricity e and mean anomaly M [1, 2]. Most methods involve choosing a starting formula and then improving this using an iterative refinement method. Many of these methods have difficulties in the critical region where eccentricity is close to one and mean anomaly is close to zero; some iterative methods even fail to converge in this region [3]. There have been several comparisons of numerical methods for solving Kepler's Equation, with conflicting claims for the accuracy and efficiency of the various algorithms [3-8].

This paper presents a new algorithm using a starting formula resulting from solution of a cubic equation based on a Padé approximation to the sine function. This starting formula has smaller errors than any previously considered [3-8]. A single application of a fifth-order method is used to refine the starting estimate, rather than iteration of a lower-order method to satisfy a convergence criterion. The latter procedure would require at least two more trigonometric function evaluations, and perhaps many more. Odell and Gooding, among others, have emphasized the advantages of refinement using a fixed number of iterations [3]. As pointed out by Mikkola [7], such a method is not really iterative, but is actually a direct solution of Kepler's Equation to the desired accuracy.

The method has errors that are less than the least significant bit of double-precision floating-point numbers over the entire range of elliptic motion, $0 \le e \le 1$. The limit of unit eccentricity is not really elliptic motion, and the solution of Kepler's Equation is not useful there, but the eccentricity can be arbitrarily close to this limit. We present a derivation of the algorithm and display contour plots of its errors as functions of the eccentricity and eccentric anomaly. Numerical problems arising in double-precision implementation of the algorithm are also discussed and resolved.

Starting Formula

Our method starts with a Padé approximation for sin E, depending on a parameter α:

$$\sin E \approx \sigma(\alpha,E) \equiv \frac{6\alpha-(\alpha-3)E^2}{6\alpha+3E^2}E. \tag{2}$$

It is assumed that E and M have been reduced by multiples of 2π to have absolute value less than or equal to π. The Taylor series expansion of this approximation at $E = 0$ is

$$\sigma(\alpha,E) = E - \frac{E^3}{6} + \frac{E^5}{12\alpha} - \dots . \tag{3}$$

This expansion is exact through terms of order E^3, which is crucial for good performance in the critical region with e near unity and M near zero. The series is exact through terms of order E^5 for $\alpha = 10$, and the approximation for sin E is exact at $E = \pm\pi$ for $\alpha = 3\pi^2/(\pi^2-6) = 7.65$. The precise specification of the parameter α for our method will be considered below.

Substitution of equation (2) into Kepler's Equation gives the cubic equation

$$[3(1-e)+\alpha e]E^3 - 3ME^2 + 6\alpha(1-e)E - 6\alpha M = 0. \tag{4}$$

Defining

$$d \equiv 3(1-e)+\alpha e \tag{5}$$

and

$$z \equiv dE \tag{6}$$

and multiplying through by d^2 gives the standard form for the cubic equation [9]

$$z^3 + a_2 z^2 + a_1 z + a_0 = 0, \tag{7}$$

with

$$a_2 = -3M, \tag{8a}$$

$$a_1 = 6\alpha d(1-e), \tag{8b}$$

and

$$a_0 = -6\alpha d^2 M. \tag{8c}$$

We define the auxiliary quantities

$$q \equiv a_1/3 - a_2^2/9 = 2\alpha d(1-e) - M^2 \tag{9}$$

and

$$r \equiv (a_1 a_2 - 3a_0)/6 - a_2^3/27 = 3\alpha d(d-1+e)M + M^3 \tag{10}$$

and note that equation (7) has a unique real root if $q^3 + r^2 > 0$. It is not difficult to see that this condition is satisfied for all positive α, which covers the range of interest, except in the case that $e = 1$ and $M = 0$, in which case the root $z = 0$ is threefold degenerate. In any case, the desired root of equation (7) is given by

$$z = s_1 + s_2 - a_2/3 = s_1 + s_2 + M \tag{11}$$

where

$$s_1 \equiv \left(r + \sqrt{q^3 + r^2} \right)^{\frac{1}{3}} \tag{12a}$$

and

$$s_2 \equiv \left(r - \sqrt{q^3 + r^2} \right)^{\frac{1}{3}}. \tag{12b}$$

The accurate evaluation of equation (11) for small M requires cancellations between s_1 and s_2. We avoid this numerical problem by employing a trick attributed to Karl Stumpff by Battin [2]. Write

$$s_1 + s_2 = \frac{(s_1 + s_2)\left(s_1^2 - s_1 s_2 + s_2^2\right)}{s_1^2 - s_1 s_2 + s_2^2} = \frac{s_1^3 + s_2^3}{s_1^2 - s_1 s_2 + s_2^2} = \frac{2r}{s_1^2 - s_1 s_2 + s_2^2}. \tag{13}$$

We next multiply the numerator and denominator by

$$w \equiv \left(|r| + \sqrt{q^3 + r^2} \right)^{\frac{2}{3}} \tag{14}$$

and use the fact that $s_1 s_2 = -q$ to simplify this expression. The absolute value of r is used in equation (14) to avoid cancellations of positive and negative quantities. Inserting the result into equation (11) and using equation (6) gives the first-order solution to Kepler's equation

$$E_1 = \frac{1}{d}\left(\frac{2rw}{w^2 + wq + q^2} + M \right) \tag{15}$$

Since r is proportional to M as the latter quantity goes to zero, this expression does not depend on cancellations in that limit. The solution is singular when both $e = 1$ and $M = 0$ simultaneously, but this is not of concern since equation (1) shows that $E = 0$ whenever $M = 0$. Thus a numerical solution of Kepler's Equation is unnecessary in this case.

This solution is the starting formula for our algorithms. It requires two transcendental function evaluations; the square root and cube root in equation (14). The cube root actually involves two transcendental function evaluations if it is implemented as a logarithm and an exponential.

Specification of α

The criterion for choosing α is the minimization of the relative errors in the starting formula. These errors are computed on a grid of 201 values of e between 0 and 1 and 251 values of E between 0 and π. The exact value of M is calculated at each grid point from equation (1), and then E_1 is obtained from equation (15). The relative error at each grid point is then

$$error = (E_1 - E)/E. \tag{16}$$

These errors were computed using quadruple precision floating point numbers with 112 bits in the mantissa. The contours of constant errors for $\alpha = 10$ and $\alpha = 3\pi^2/(\pi^2 - 6)$ are shown in Figures 1 and 2, respectively. The contours are linearly spaced with an increment of 0.001 between contours. The errors in Figure 1 are all negative, with a minimum value of -0.040. The errors in Figure 2 are all positive and significantly smaller in magnitude, having a maximum value of 0.013. The errors in both figures are monotonically increasing functions of the eccentricity, which suggests that an optimal α could be found by minimizing the errors for eccentricity equal to unity.

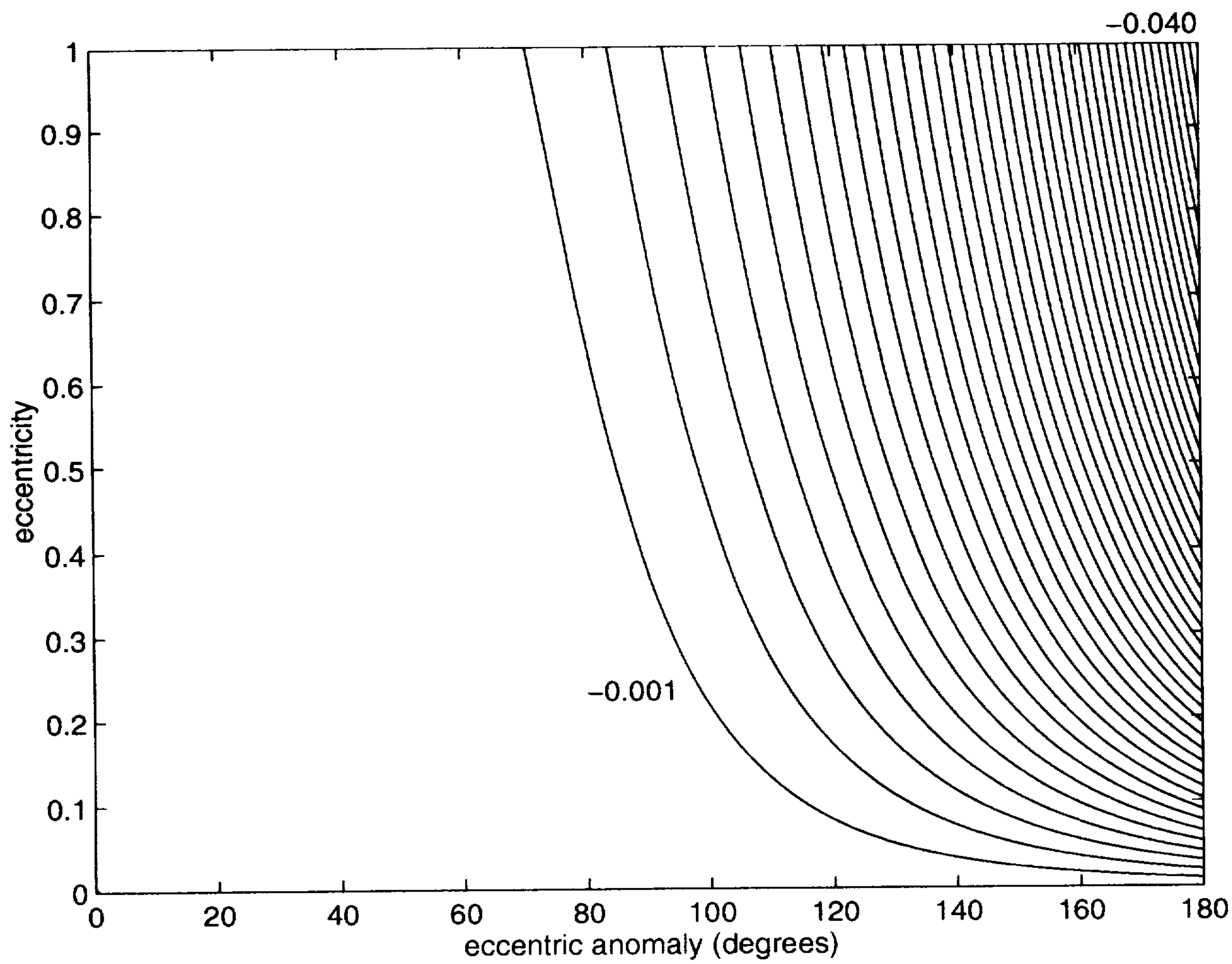

Figure 1. Relative Errors in Starting Formula E_1 for $\alpha = 10$
Equal-error contours with 0.001 linear contour spacing

It is not necessary that α be a constant parameter; we can choose it to be a function of e and M. Equation (2) would be exact if α could be chosen to satisfy

$$\alpha_{\text{ideal}} = \frac{3E^2(E - \sin E)}{E^3 - 6(E - \sin E)} . \tag{17}$$

This equation clearly not useful as it stands, since E is not known until Kepler's Equation has been solved. However, it is possible to find a usable function $\alpha(e, M)$ that is a good approximation to equation (17). The right side of this equation is an even function of E, and so $\alpha(e, M)$ must be an even function of M. It is also useful to choose a form that takes the value $\alpha = 3\pi^2/(\pi^2 - 6)$ when $E = M = \pm\,\pi$, since this will assure the continuity of the solutions over the extended range of these variables outside $[-\pi, \pi]$. This is not really important, since we intend to refine the result to the full accuracy of machine arithmetic, but Figure 2 shows that this condition is likely to lead to a good starting formula.

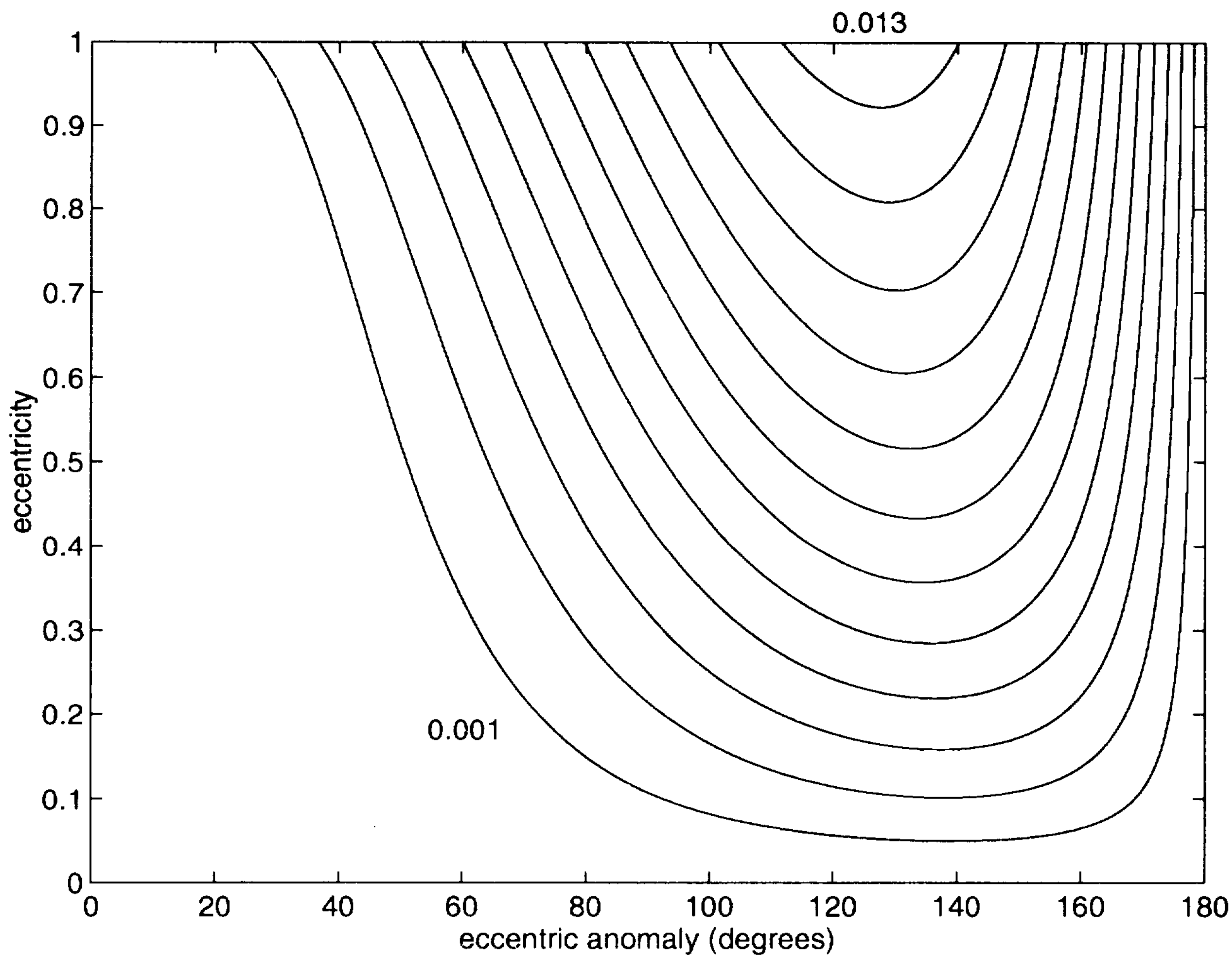

Figure 2. Relative Errors in Starting Formula E_1 for $\alpha = 3\pi^2/(\pi^2 - 6)$
Equal-error contours with 0.001 linear contour spacing

As discussed above, it is most important to find an accurate form for $\alpha(e,M)$ on the boundary $e = 1$. Figure 3 shows α_{ideal} and the straight line fit

$$\alpha(1,M) = \frac{3\pi^2 + 0.8\pi(\pi - |M|)}{\pi^2 - 6} , \tag{18}$$

where M is given by equation (1) with $e = 1$. This formula can be extended to all eccentricities by noting that equations (17) and (18) are functions of E and M, respectively. We obtain the correct dependence of the slope of the straight line at the right end point by making use of

$$\left.\frac{\partial M}{\partial E}\right|_{E=\pi} = 1 + e. \tag{19}$$

The resulting approximation for all eccentricities and mean anomalies is given by

$$\alpha(e,M) = \frac{3\pi^2 + 1.6\pi(\pi - |M|)/(1+e)}{\pi^2 - 6} . \tag{20}$$

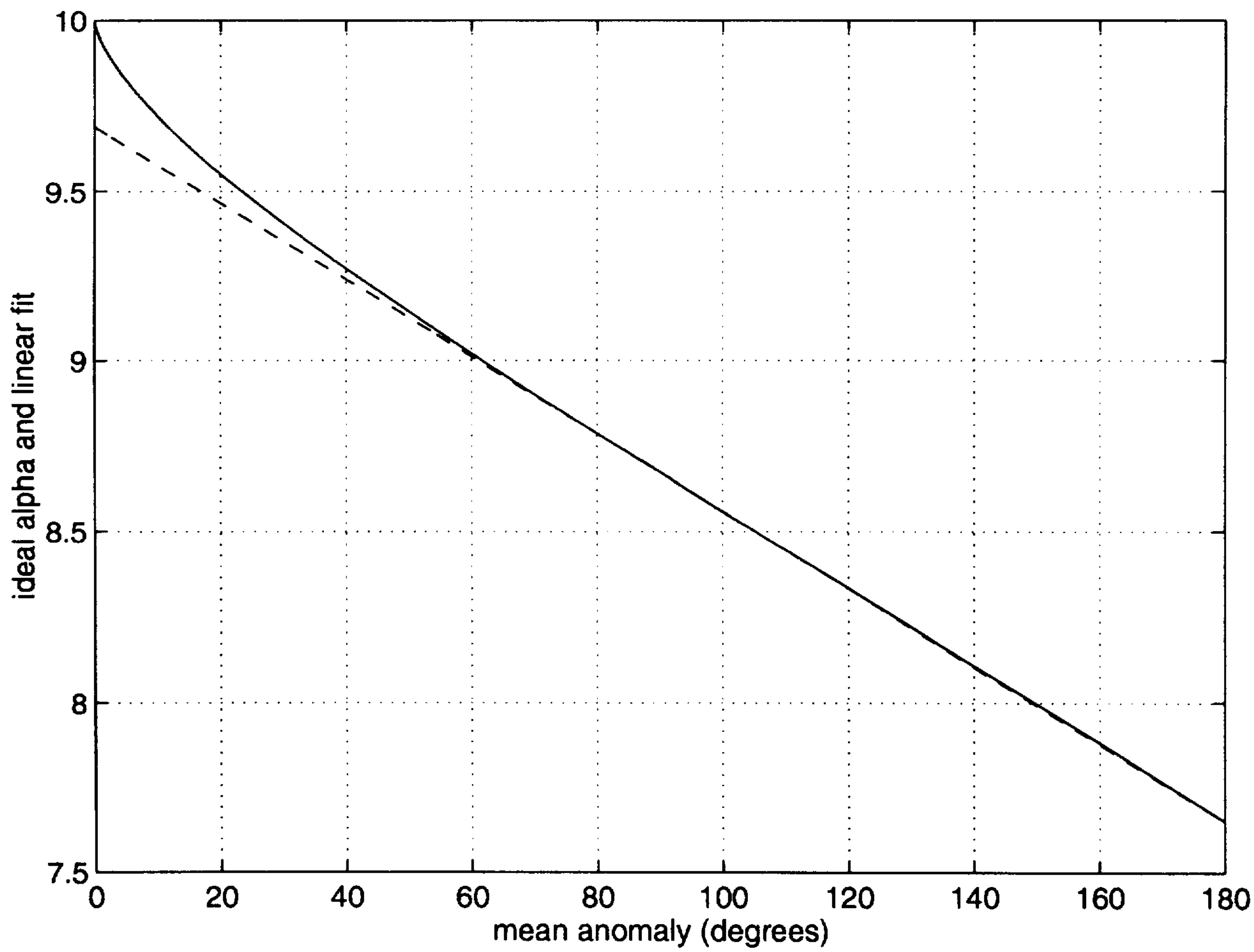

Figure 3. The Parameter α_{ideal} (solid curve) and Straight Line Fit (dotted line)

Our starting formula is given by the solution of the cubic for this form of α. The error contours for this starting formula are shown in Figure 4, with linear contour spacing of 2×10^{-5}. This is fifty times finer than the spacing of the contours in Figures 1 and 2. The values of these errors are between -2.3×10^{-4} and 2.8×10^{-4}, which are smaller than the errors of any other starting formula known to the author. The starting formula does involve a fair amount of computation, but Mikkola [7] has also proposed a starting formula that requires solution of a cubic equation and has maximum errors seven times as large as those of the method proposed here.

Refinement

The iterative refinements are based on finding roots of polynomial approximations of

$$f(E) = E - e\sin E - M. \tag{21}$$

The aim of this paper is to find a computational method yielding errors are smaller than the least significant bit of double-precision floating point numbers with 52 bits in the mantissa, which is

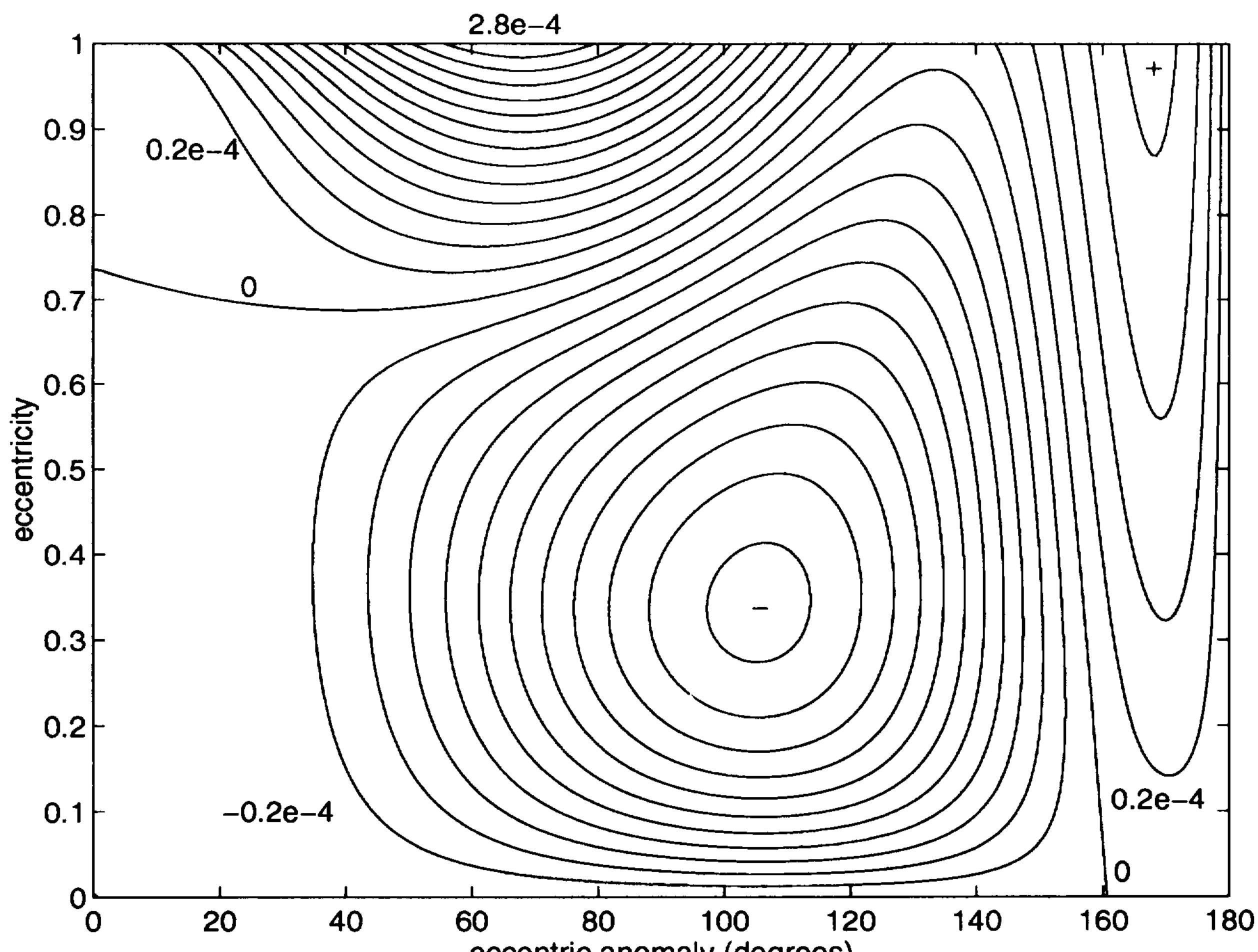

Figure 4. Relative Errors in Starting Formula E_1 for α Given by Equation (20)
Equal-error contours with 2×10^{-5} linear contour spacing

about 10^{-16}. A fifth-order refinement of the starting formula is expected to have relative errors less of $(2.8 \times 10^{-4})^5 = 1.7 \times 10^{-18}$, which is adequate to achieve double-precision accuracy. The third-order (Halley) and higher-order corrections are given by

$$\delta_3(E) = -\frac{f(E)}{f'(E) - \frac{1}{2} f(E) f''(E) / f'(E)}, \tag{22}$$

$$\delta_4(E) = -\frac{f(E)}{f'(E) + \frac{1}{2}\delta_3(E) f''(E) + \frac{1}{6}\delta_3^2(E) f'''(E)}, \tag{23}$$

and

$$\delta_5(E) = -\frac{f(E)}{f'(E) + \frac{1}{2}\delta_4(E) f''(E) + \frac{1}{6}\delta_4^2(E) f'''(E) + \frac{1}{24}\delta_4^3(E) f''''(E)}, \tag{24}$$

where the subscripts denote the order of the correction. The required partial derivatives are:

$$f'(E) = 1 - e\cos E, \tag{25}$$

$$f''(E) = e\sin E. \tag{26}$$

$$f'''(E) = e\cos E = 1 - f'(E), \tag{27}$$

and

$$f''''(E) = -e\sin E = -f''(E). \tag{28}$$

where the second forms of equations (27) and (28) are used to avoid additional trigonometric function evaluations. The fifth-order refined estimate is given by

$$E_5 = E_1 + \delta_5(E_1). \tag{29}$$

The equal-error contours of E_5 are plotted in Figure 5 with logarithmic spacing of the contours, which is to say that the errors on adjacent contours differ by a factor of ten. The zero contour is not plotted because of numerical roundoff; this contour is not significant since the magnitude of the errors is more important than their sign. The magnitude of the relative errors is less than 10^{-24} over large areas of the figure, and the maximum error magnitude over the entire range of elliptic motion is 7.35×10^{-19}. This is about half the naive prediction based on the order of the correction employed. This method requires only two trigonometric function evaluations in addition to the transcendental function evaluations need to solve the cubic equation for the starting formula.

Numerical Considerations

The errors of our method are smaller than the least significant bit of double-precision floating point numbers. Special care must be taken when this method is actually implemented in double-precision arithmetic, however. A naive implementation yields unacceptably large errors exceeding 5×10^{-14}. Plotting error contours shows that all errors with magnitudes in excess of 5×10^{-16} are in the region $e > 0.75$ and $E < 45°$. This effect was discussed by Odell and Gooding [3], who attribute it to cancellations in the computation of $f(E)$ and $f'(E)$ by equations (21) and (25),

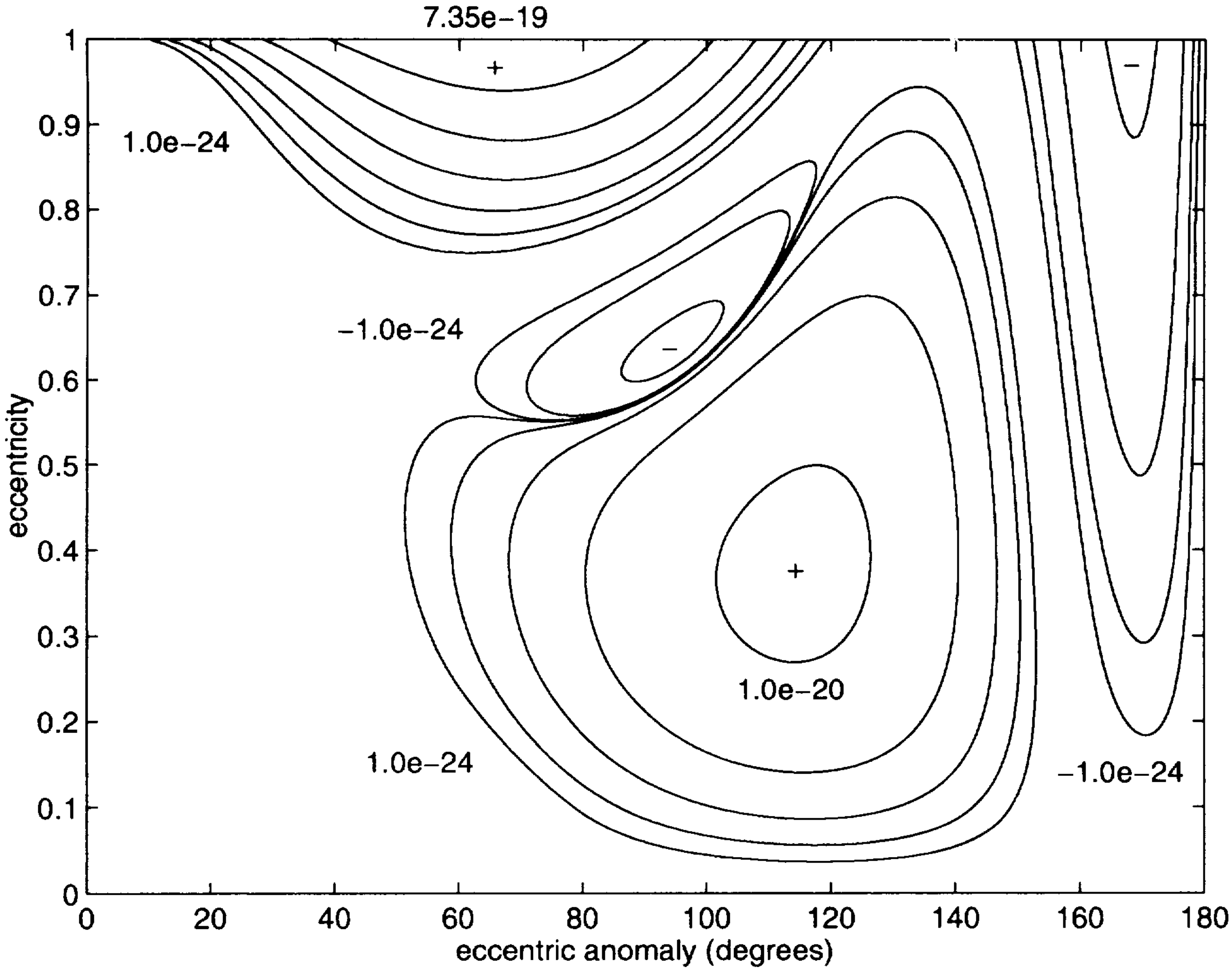

Figure 5. Relative Errors in Refined Eccentric Anomaly E_9
Equal-error contours with $\times$ 10 logarithmic contour spacing

respectively. The solution to this problem is to modify the computation of these quantities. The revision to the first derivative computation is straightforward; equation (25) is replaced by

$$f'(E) = 1 - e + 2e\sin^2(E/2). \tag{30}$$

The fix to equation (21) is not so simple. The method employed here is to replace equation (21) by the equivalent form

$$f(E) = M^*(e,E) - M, \tag{31}$$

where

$$M^*(e,E) \equiv E - e\sin E \tag{32}$$

outside the range $e > 0.5$ and $E_1 < 1$ radian. Inside this range, $M^*(e,E)$ is given by the Padé approximant:

$$M^*(e,E) \equiv (1-e)E + eE^3 \frac{numerator}{denominator}\ , \tag{33}$$

with

$$numerator \equiv -1.7454287843856404 \times 10^{-6} E^6 + 4.1584640418181644 \times 10^{-4} E^4 - 3.0956446448551138 \times 10^{-2} E^2 + 1 \tag{34a}$$

and

$$denominator \equiv 1.7804367119519884 \times 10^{-8} E^8 + 5.9727613731070647 \times 10^{-6} E^6 + 1.0652873476684142 \times 10^{-3} E^4 + 1.1426132130869317 \times 10^{-1} E^2 + 6. \tag{34b}$$

The second derivative is computed as

$$f''(E) = E - M^*(e, E) \tag{35}$$

over the entire range of eccentricity and eccentric anomaly. The third and fourth derivatives are computed as usual. No additional transcendental function evaluations are required by these fixes, although the Padé approximant is probably more expensive to evaluate than the sine function.

The resulting double precision implementation of the algorithm has relative errors less than 4×10^{-16} over the entire range of elliptic motion. Inspection of a plot of the error contours revealed no systematic pattern, confirming that the errors are due solely to unavoidable machine arithmetic roundoff.

Discussion

The algorithm developed in this paper has been shown to have errors well within the inherent limitations of double-precision computer arithmetic, over the entire range of elliptic orbital motion. Among algorithms with this property, this method is at least as efficient as any proposed previously, requiring only four transcendental function evaluations: a square root, a cube root, and two trigonometric functions. The method is singular only when the eccentricity is unity and the mean anomaly is simultaneously zero. There are two reasons why it is never necessary to solve Kepler's Equation in this case: first, unit eccentricity is not really elliptic motion, and second, the eccentric anomaly is known to be zero when the mean anomaly is zero, making numerical solution unnecessary. Special procedures to handle double-precision roundoff errors near this singular point have been developed and tested. The resulting algorithm is well suited for implementation in orbit propagation systems.

References

[1] Peter Colwell, *Solving Kepler's Equation Over Three Centuries*, Willman-Bell, Richmond, VA (1993)

[2] Richard H. Battin, *An Introduction to the Mathematics and Methods of Astrodynamics*, American Institute of Aeronautics and Astronautics, New York (1987)

[3] A. W. Odell and R. H. Gooding,"Procedures for Solving Kepler's Equation," Celestial Mechanics, **38** (1986) pp 307-334

[4] J. M. A. Danby and T. M. Burkardt, "The Solution of Kepler's Equation, I," Celestial Mechanics, **31** (1983) pp 95-107

[5] T. M. Burkardt and J. M. A. Danby , "The Solution of Kepler's Equation, II," Celestial Mechanics, **31** (1983) pp 317-328

[6] J. M. A. Danby, "The Solution of Kepler's Equation, III," Celestial Mechanics, **40** (1987) pp 303-312

[7] Seppo Mikkola, "A Cubic Approximation for Kepler's Equation," Celestial Mechanics, **40** (1987) pp 329-334

[8] L. G. Taff and T. A. Brennan, "On Solving Kepler's Equation," Celestial Mechanics and Dynamical Astronomy, **46** (1989) pp 163-176

[9] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions*, Dover, New York (1965)