An Elegant Lambert Algorithm

Richard H. Battin* and Robin M. Vaughan†

The Charles Stark Draper Laboratory Inc., Cambridge, Massachusetts

Of the many techniques extant for solving the two-body, two-point, time-constrained orbital boundary-value problem, commonly known today as Lambert's problem, none is more conceptually elegant than the classical method devised by Gauss. The simplicity of Gauss' method would certainly have been attractive to the modern astrodynamicist except for two major flaws — the method is singular for a transfer angle of 180 degrees and the convergence rate is extremely slow when that angle is not very small. In this paper a new algorithm is described which exactly parallels both the mechanics and the elegant simplicity of the classical one but is completely devoid of the two basic faults of the original. The equations of the new method are universal and not singular for the 180 degree transfer. (They are singular for a complete revolution through 360 degrees but this should not be cause for great alarm.) Furthermore, convergence is both remarkably rapid and almost uniform as well as being essentially independent of the initial guess. It should further be emphasized that all of the advantages of Gauss' method are inherent in the new method — notably, the preservation of numerical accuracy for small transfer angles (of the order of 2 or 3 degrees, for example).

Introduction

THE DETERMINATION of an orbit, having a specified flight time and connecting two position vectors, frequently referred to as Lambert's problem, is fundamental in astrodynamics. A variety of methods of dealing with this problem has been discussed over the years by many writers — a reasonably comprehensive list of references can be established from those given in two of the first author's papers.[1,2] Most of these earlier methods have been characterized by a particular formulation of the time of flight equation and a particular independent variable to be used in a Newton-Raphson style of iteration. In Ref. (2) the first author was successful in devising a successive substitution algorithm — actually, a set of algorithms — which had some of the features of Gauss' method but, like his, failed to converge for some hyperbolic orbits and had slow convergence properties in many other cases.

The Lambert algorithm described in this paper represents a major improvement over Gauss' method[3], or any other technique known to the authors, and is made possible by exploiting a new principle of which Gauss and his followers were probably not aware (even though it is a fundamental property of two-body orbits) together with a new wrinkle on an idea that Gauss himself invented in developing an iterative solution of Kepler's equation. The new principle[2] is the invariance of the mean point [or normal point as it was called in Ref. (2)] under a certain geometric transformation of the boundary-value problem which brings the mean point radius into coincidence with an orbital apse. The result is that the transformed problem can then be simply described using the elementary form of Kepler's equation.

The second innovation is the introduction of a free parameter in Kepler's equation. The new twist is to choose this free parameter (not to be a constant as Gauss did in his application) but rather to insure rapid convergence over the entire range of problems. Indeed, it is truly startling to observe just how rapid the convergence is when the iterated variable is anywhere near the solution value. It is not uncommon for this variable to improve by as many as four or more significant decimal places in a single iteration step!

Presented as Paper IAF-83-325 at the 34th Congress of the International Astronautical Federation, Budapest, Hungary, Oct. 10-15, 1983; submitted Oct. 26, 1983; revision received May 11, 1984.

*Associate Department Head, Charles Stark Draper Laboratory, Inc. & Adjunct Professor of Aeronautics and Astronautics, Massachusetts Institute of Technology. Fellow AIAA.

†MIT graduate student and Draper fellow.

Editor's note: This paper was prepared by the authors using the computer typesetting program called TEX which was invented by Prof. Donald E. Knuth of Stanford University. The editors will welcome any comments from the readers.

Gauss' Method

Denote the initial and target points by P_1 and P_2 and the corresponding radial positions by r_1 and r_2 as shown in Fig. 1. Let c be the line of sight distance between P_1 and P_2 and let θ be the transfer angle. The problem is to determine the Keplerian orbit for a body which is at P_1 at time t_1 and must arrive at P_2 at the specified time t_2.

For notational convenience in the sequel, define the dimensionless parameter λ as

$$\lambda s = \sqrt{r_1 r_2}\cos\tfrac{1}{2}\theta = \pm\sqrt{s(s-c)} \tag{1}$$

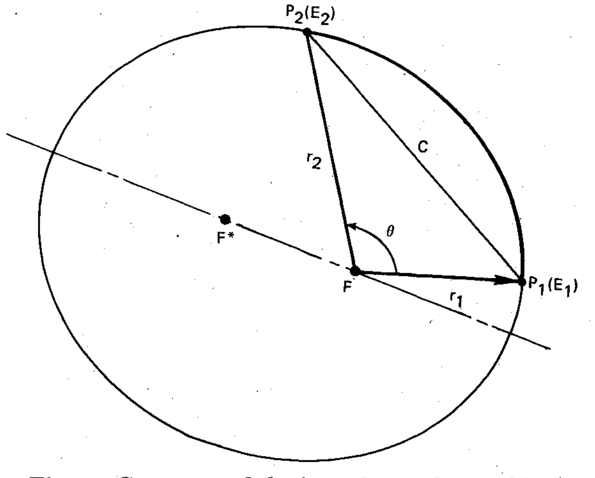

Fig. 1 Geometry of the boundary-value problem.

where s is the semiperimeter of the triangle P_1P_2F, i.e.

$$s = \tfrac{1}{2}(r_1 + r_2 + c)$$

so that λ will decrease monotonically from $+1$ to -1 as θ increases from 0 to 2π with the value $\lambda = 0$ corresponding to $\theta = \pi$.

With great insight and ingenuity, Gauss formulated the problem as two equations

$$y^2 = \frac{m}{\ell + \sin^2 \frac{1}{2}E} \tag{2}$$

and

$$y^3 - y^2 = m\frac{2E - \sin 2E}{\sin^3 E} \tag{3}$$

to be solved simultaneously for the variables y and E where 1) y is the well-known sector-triangle ratio — i.e., the ratio of the area of the sector contained between the radii r_1 and r_2 and the orbit and the area of the triangle P_1P_2F and 2) E is defined as

$$E = \tfrac{1}{2}(E_2 - E_1) \tag{4}$$

where E_1 and E_2 are the eccentric anomalies of the orbit at the terminals P_1 and P_2. The quantities ℓ and m are constants which depend only on the geometry, the flight time $t_2 - t_1$ and the gravitational constant μ. Specifically,

$$\ell = \frac{(1-\lambda)^2}{4\lambda} \quad \text{and} \quad m = \frac{\mu(t_2 - t_1)^2}{8s^3\lambda^3}. \tag{5,6}$$

Before reviewing Gauss' method of solution, it is important to point out that ℓ cannot be accurately computed from Eq. (5) when θ is small (of the order of 2 or 3 degrees) or, equivalently, when λ is nearly one — precisely the case which Gauss had in mind for his application. To preserve accuracy for small θ, Gauss showed that ℓ could be computed from

$$\ell = \frac{\sin^2 \frac{1}{4}\theta + \tan^2 2\omega}{\cos \frac{1}{2}\theta} \tag{7}$$

after ω is first determined from

$$\tan(\tfrac{1}{4}\pi + \omega) = \left(\frac{r_2}{r_1}\right)^{\frac{1}{4}}. \tag{8}$$

Indeed, Gauss' choice of variables in Eqs. (2) and (3) was primarily motivated by the need to have the elements of the solution orbit calculable without loss of precision. Thus, when y and E are found, the semimajor axis a and the orbital parameter p are determined from the formulas*

$$a = \frac{\mu(t_2 - t_1)^2}{4r_1r_2y^2 \sin^2 E \cos^2 \frac{1}{2}\theta} \tag{9}$$

and

$$p = \frac{r_1^2 r_2^2 y^2 \sin^2 \theta}{\mu(t_2 - t_1)^2} \tag{10}$$

which involve only products and quotients.

Return now to the discussion of the method of solution of Gauss' equations — (2) and (3). The classical memoir by Gauss on hypergeometric functions and their continued fraction expansions was published some three years after *Theoria Motus* so his development of the right hand side of Eq. (3) appeared to be ad hoc and somewhat enigmatic. Later, he would have written

$$\frac{2E - \sin 2E}{\sin^3 E} = \frac{4}{3}F(3, 1; \tfrac{5}{2}; \sin^2 \tfrac{1}{2}E) \tag{11}$$

where F is a hypergeometric function† which admits of a continued fraction expansion. As a consequence of this relation, the quantity

$$x = \sin^2 \tfrac{1}{2}E \tag{12}$$

can replace E as one of the two unknowns in Eqs. (2) and (3).

After some clever algebraic manipulations, Gauss was able to replace Eqs. (2) and (3) with

$$x = \frac{m}{y^2} - \ell \quad \text{and} \quad y^3 - y^2 - hy - \frac{h}{9} = 0 \tag{13,14}$$

where

$$h = \frac{m}{\frac{5}{6} + \ell + \xi} \tag{15}$$

and $\xi(x)$, which is of order x^2, determined from the continued fraction

$$\xi = \cfrac{\frac{2}{35}x^2}{1 + \frac{2}{35}x - \cfrac{\frac{40}{63}x}{1 - \cfrac{\frac{4}{99}x}{1 - \cdots}}}. \tag{16}$$

*Gauss also developed error-free formulas for computing the orbital eccentricity. The interested reader can find these in his book *Theoria Motus*.

†The notation for the general form of the hypergeometric function is $F(\alpha, \beta; \gamma; x)$.

In this way, his objective of designing a rapidly convergent algorithm was neatly accomplished. For a reasonably small transfer angle θ (and, hence, a correspondingly small value of x) we may first assume that $x = \xi = 0$. Then h is determined from Eq. (15) and y obtained as the positive real root of the cubic equation (14). Having now a trial value for y, a new value of x is obtained from Eq. (13) with which an improved value of h is found. The process is repeated until y ceases to change by a preassigned amount — usually two or three iterations being sufficient. (This method of successive substitutions was a favorite technique of Gauss.)

The equations of the algorithm are universal as Gauss also demonstrated. By extending the definition of x so that

$$x = \begin{cases} \sin^2 \frac{1}{4}(E_2 - E_1) & \text{ellipse} \\ 0 & \text{parabola} \\ -\sinh^2 \frac{1}{4}(H_2 - H_1) & \text{hyperbola} \end{cases} \tag{17}$$

where H is the hyperbolic analog of the eccentric anomaly, and allowing, thereby, values of x to range from $-\infty$ to $+1$, all types of orbits are included. Furthermore, and fortunately, the continued fraction (16) converges over this extended range.

Gauss, however, chose to develop separate equations for $0 < \theta < \pi$ and $\pi < \theta < 2\pi$. Indeed, had we rigidly followed his notation, we would have used m^2 in place of m. Instead, it seems more appropriate to let ℓ and m be negative for negative values of λ and thus be spared the unnecessary burden of addressing the two cases separately.

Obviously, Gauss knew that his method was singular for $\theta = \pi$ but he judiciously avoided ever mentioning it as if it were just a minor annoyance — a small flaw in an otherwise beautiful scheme. Indeed, he said "*The equations ... possess so much neatness, that there may seem nothing more to be desired.*" This flaw, however, coupled with convergence difficulties when θ is not very small, has rendered the method unattractive to the modern astrodynamicist who is concerned with a more general range of orbit determination problems than Gauss could ever have imagined.

The authors of the present paper are delighted to have discovered a new algorithm which exactly parallels the elegant simplicity of the classical one but is completely devoid of the two basic faults of the original. In the next two sections, we shall separately 1) remove the singularity at $\theta = \pi$ and 2) drastically improve the convergence for the entire range of transfer angles $0 < \theta < 2\pi$ without compromising the elegance of Gauss' scheme.

Removing the Singularity

The singularity at $\theta = \pi$ in Gauss' method may be removed using the transformation described in Ref. (2). According to Lambert's theorem, if the terminals P_1 and P_2 are held fixed, the shape of the orbit may be altered by moving the occupied and vacant foci F and F^* without altering the flight time, provided that

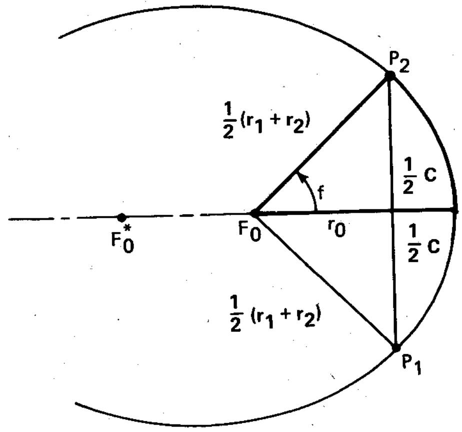

Fig. 2 Transformation of Lambert's problem.

$r_1 + r_2$ and a are unchanged in the process. For the purpose at hand, we choose the transformed orbit so that its major axis is perpendicular to P_1P_2 as shown in Fig. 2.

The time to travel from pericenter of the new orbit to the point P_2 is just one half of the time to traverse the original orbit from P_1 to P_2. The pericenter radius is r_0, the terminal radius is $\frac{1}{2}(r_1 + r_2)$, and the true anomaly f is related to the original central angle θ according to

$$\cos f = \frac{\sqrt{r_1 r_2}\cos\frac{1}{2}\theta}{\frac{1}{2}(r_1 + r_2)}. \tag{18}$$

The time of flight equation is then the elementary form of Kepler's equation, which, for the ellipse, is simply

$$\frac{1}{2}\sqrt{\frac{\mu}{a^3}}(t_2 - t_1) = E - e_0 \sin E \tag{19}$$

where E is the eccentric anomaly of the point P_2 and e_0 is the eccentricity of the transformed orbit.

As discussed in Ref. (2), there are a number of invariants of this orbital transformation. The so-called *mean point* radius — the radius to that point* in the orbit at which the tangent is parallel to the chord P_1P_2 — is one such invariant and is precisely the pericenter radius of the transformed orbit. The difference between the eccentric anomalies at P_1 and P_2 is also an invariant — half of this difference is just the eccentric anomaly E in Kepler's equation (19). Obviously, the eccentricity e_0 is **not** invariant.

Let r_{0p} be the radius to the mean point of the parabola connecting P_1 and P_2. Although it is not explicitly given in Ref. (2), the exceedingly pleasant relationship

$$r_0 = r_{0p}\begin{cases} \sec^2 \frac{1}{2}E & \text{ellipse} \\ \operatorname{sech}^2 \frac{1}{2}H & \text{hyperbola} \end{cases} \tag{20}$$

*Because the eccentric anomaly of this point is the arithmetic mean of the eccentric anomalies of the end points, we have decided to change the name from "normal point," as it was called in Ref. (2), to "mean point."

can be easily demonstrated using the similar triangles discussed there. We can further show from the properties of the parabola that

$$r_{0p} = \tfrac{1}{4}(r_1 + r_2 + 2\sqrt{r_1 r_2}\cos\tfrac{1}{2}\theta). \tag{21}$$

Our objective now is to convert Kepler's equation, if possible, to a form resembling Gauss' equations (2) and (3). To this end, write (19) as

$$\frac{1}{2}\sqrt{\frac{\mu}{a^3}}(t_2 - t_1) = E - \sin E + (1 - e_0)\sin E \tag{22}$$

or, alternately, as

$$\frac{1}{2}\sqrt{\frac{\mu}{8r_{0p}^3}}(t_2 - t_1)\left(\frac{2r_{0p}}{a}\right)^{\frac{3}{2}} = E - \sin E + \left(\frac{2r_{0p}}{a}\right)\tan\tfrac{1}{2}E$$

where we have replaced $1 - e_0$ by

$$1 - e_0 = \frac{r_0}{a} = \frac{r_{0p}}{a}\sec^2\tfrac{1}{2}E.$$

Now, from the classical formula relating the true and eccentric anomalies f and E

$$\tan^2\tfrac{1}{2}f = \frac{1 + e_0}{1 - e_0}\tan^2\tfrac{1}{2}E$$

we obtain

$$1 - e_0 = \frac{2\tan^2\frac{1}{2}E}{\tan^2\frac{1}{2}f + \tan^2\frac{1}{2}E}. \tag{23}$$

Hence

$$\frac{2r_{0p}}{a} = \frac{4\tan^2\frac{1}{2}E}{(1 + \tan^2\frac{1}{2}E)(\tan^2\frac{1}{2}f + \tan^2\frac{1}{2}E)} \tag{24}$$

and, as a consequence, Kepler's equation takes the form

$$\sqrt{\frac{\mu}{8r_{0p}^3}}(t_2 - t_1)\frac{4\tan^3\frac{1}{2}E}{\left[(\tan^2\frac{1}{2}f + \tan^2\frac{1}{2}E)(1 + \tan^2\frac{1}{2}E)\right]^{\frac{3}{2}}} = E - \sin E + \frac{4\tan^3\frac{1}{2}E}{(\tan^2\frac{1}{2}f + \tan^2\frac{1}{2}E)(1 + \tan^2\frac{1}{2}E)}.$$

The analogy with Gauss' equations is now readily made. At the possible risk of confusing the reader, we will use the same notation as in Gauss' method but the symbols will have different meanings. The advantage will be the ease of comparison of the two. Thus, if we define

$$x = \tan^2\tfrac{1}{2}E, \qquad \ell = \tan^2\tfrac{1}{2}f, \tag{25,26}$$

$$m = \frac{\mu(t_2 - t_1)^2}{8r_{0p}^3}, \tag{27}$$

the analog of the first equation of Gauss is had by defining y as*

$$y^2 = \frac{m}{(\ell + x)(1 + x)}. \tag{28}$$

Substituting into Kepler's equation, we obtain the analog of the second of Gauss' equations

$$y^3 - y^2 = m\frac{E - \sin E}{4\tan^3\frac{1}{2}E}. \tag{29}$$

Furthermore, using Eqs. (18) and (21) together with a little trigonometry, it is not difficult to show that†

$$\ell = \left(\frac{1 - \lambda}{1 + \lambda}\right)^2 \quad \text{and} \quad m = \frac{8\mu(t_2 - t_1)^2}{s^3(1 + \lambda)^6}. \tag{30,31}$$

Then, since

$$\begin{aligned}\frac{E - \sin E}{4\tan^3\frac{1}{2}E} &= \frac{1}{2\tan^2\frac{1}{2}E}\left[\frac{\frac{1}{2}E}{\tan\frac{1}{2}E} - \frac{1}{1 + \tan^2\frac{1}{2}E}\right]\\ &= \frac{1}{2x}\left[\frac{\tan^{-1}\sqrt{x}}{\sqrt{x}} - \frac{1}{1 + x}\right]\\ &= -\frac{d}{dx}\frac{\tan^{-1}\sqrt{x}}{\sqrt{x}}\end{aligned}$$

we have

$$\frac{E - \sin E}{4\tan^3\frac{1}{2}E} = -\frac{d}{dx}F(\tfrac{1}{2}, 1; \tfrac{3}{2}; -x) \tag{32}$$

so that Eqs. (28) and (29) are also functions of x and y just as in Gauss' equations. The hypergeometric function satisfies the necessary requirement to be expanded as a continued fraction. Specifically,

$$F(\tfrac{1}{2}, 1; \tfrac{3}{2}; -x) = \cfrac{1}{1 + \cfrac{x}{3 + \cfrac{4x}{5 + \cfrac{9x}{7 + \cdots}}}}. \tag{33}$$

It can be shown that these equations are also universal when the definition of x is extended to include the other conics,

$$x = \begin{cases} \tan^2\frac{1}{4}(E_2 - E_1) & \text{ellipse} \\ 0 & \text{parabola} \\ -\tanh^2\frac{1}{4}(H_2 - H_1) & \text{hyperbola} \end{cases} \tag{34}$$

with its values ranging now from -1 to $+\infty$.

The orbital elements are as easily calculated and error-free as for Gauss' method; formulas for the semimajor axis and the parameter are, indeed,

$$a = \frac{ms(1 + \lambda)^2}{8xy^2} \tag{35}$$

*Note that in this case y is **not** the sector-triangle ratio.

†As can be seen from the new definitions of ℓ and m, we have eliminated the singularity at $\theta = \pi$. However, there is now a singularity at $\lambda = -1$, corresponding to $\theta = 2\pi$, which is more tolerable.

and

$$p = \frac{2r_1r_2y^2(1+x)^2\sin^2\frac{1}{2}\theta}{ms(1+\lambda)^2}. \tag{36}$$

The formula for a is readily obtained using Eqs. (24) and (27). However, the orbital parameter is not an invariant so that we must first establish the parameter p_0 for the transformed orbit. By adapting the fairly standard expression for the parameter in the boundary-value problem

$$p = \frac{r_1r_2\sin^2\frac{1}{2}(f_2-f_1)}{a\sin^2\frac{1}{2}(E_2-E_1)}$$

to the case at hand, we obtain

$$p_0 = \frac{c^2(1+x)^2}{16ax}.$$

Then, using the equation derived in Ref. (2),

$$p = \frac{4r_1r_2\sin^2\frac{1}{2}\theta}{c^2}p_0$$

which relates the parameters of the original and the transformed orbits, the verification of Eq. (36) is immediate.

Finally, we observe that the equation (30) for ℓ is not appropriate when θ is small — just as was the case for Gauss' definition of ℓ. But, as before, we can derive the following alternate expression:

$$\ell = \frac{\sin^2\frac{1}{4}\theta + \tan^2 2\omega}{\sin^2\frac{1}{4}\theta + \tan^2 2\omega + \cos\frac{1}{2}\theta} \tag{37}$$

with ω determined from Eq. (8). Indeed, all of the precision-preserving techniques that Gauss so carefully introduced in his method exist also in this new formulation.

Improving the Convergence

In *Theoria Motus*, Gauss developed an extremely efficient technique for solving the elementary form of Kepler's equation in the case of near parabolic orbits. The key was the introduction of a parameter specifically selected to accelerate the convergence of his successive substitution algorithm. Since our time equation is also the simple form of Kepler's equation, we are tempted to introduce a free parameter in this instance too.

For this purpose, with β as yet unspecified, write Eq. (22) as

$$\frac{1}{2}\sqrt{\frac{\mu}{a^3}}(t_2-t_1) = [1+\beta(1-e_0)]P + (1-e_0)Q$$

where

$$P = E - \sin E$$

and

$$Q = \sin E - \beta(E - \sin E).$$

Now, from Eqs. (23) and (32), together with a trigonometric identity for the sine function, we have

$$1 - e_0 = \frac{2x}{\ell + x},$$

$$P = -4\tan^3\tfrac{1}{2}E\,\frac{dF}{dx},$$

$$\sin E = \frac{2\tan\frac{1}{2}E}{1+x},$$

which are used to develop the following expressions:

$$[1+\beta(1-e_0)]P = -4\tan^3\tfrac{1}{2}E\left[\frac{dF}{dx} + \frac{h_1}{(\ell+x)(1+x)}\right]$$

$$\begin{aligned}(1-e_0)Q &= \frac{4\tan^3\frac{1}{2}E}{(\ell+x)(1+x)}(1+h_1)\\ &= \frac{4y^2}{m}\tan^3\tfrac{1}{2}E\,(1+h_1)\end{aligned}$$

with the quantity h_1 defined as

$$h_1 = 2\beta x(1+x)\frac{dF}{dx}. \tag{38}$$

But, we also have

$$\frac{m}{2}\sqrt{\frac{\mu}{a^3}}(t_2-t_1) = 4y^3\tan^3\tfrac{1}{2}E$$

so that Kepler's equation can be written as

$$y^3 - (1+h_1)y^2 + m\left[\frac{dF}{dx} + \frac{h_1}{(\ell+x)(1+x)}\right] = 0. \tag{39}$$

Clearly, if $\beta = 0$, then $h_1 = 0$ and Eq. (39) reduces to (29). Otherwise, β can have any value whatsoever — **not necessarily constant.**

Now that we have this extra degree of freedom, how can we best use it? To decide, consider the general problem of the simultaneous solution of two equations by successive substitutions. In Fig. (3a) we have plotted two arbitrary functions $y_1(x)$ and $y_2(x)$; the intersection of these two curves is the solution point. To find the solution by successive substitution, we start by choosing an initial value x_0 and then calculate the corresponding value of $y_1(x_0)$. Next, $y_1(x_0)$ is used to obtain a new value of x by locating the point where $y_2(x) = y_1(x_0)$. The new value of x is again used to calculate y_1 which is, in turn, used to find a new value for x. This process is represented in the figure by the horizontal and vertical dotted lines. Clearly, the curvature of the two functions greatly influences the number of iterations required to reach the solution. In the extreme, suppose that $y_1(x)$ is a constant as shown in Fig. (3b). Then the solution would be reached in just one iteration step since for any x_0, we have $y_1(x_0)$ exactly equal to $y_2(x)$ at the solution point.

Using this argument, it appears that we should choose our free parameter β so that dy/dx will be zero

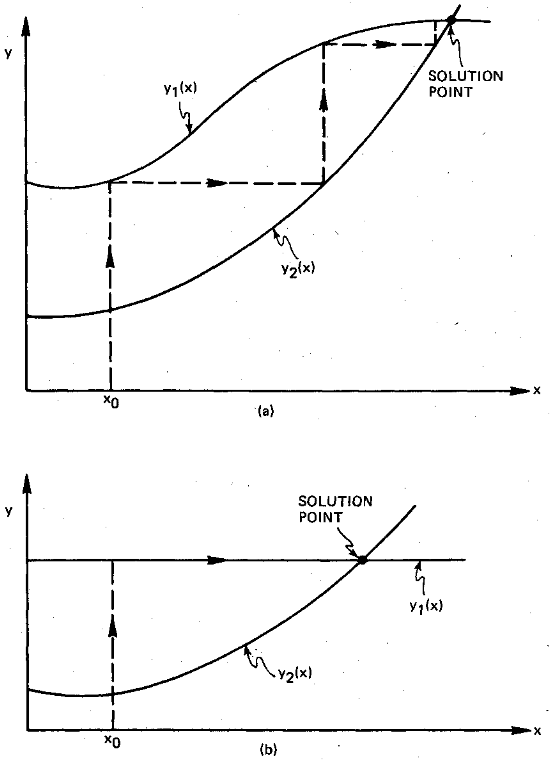

Fig. 3 Graphics of successive substitutions.

at the solution point. Of course, we don't know the location of this point — if we did, no problem would exist in the first place. For the moment, we will ignore this seemingly crucial matter, calculate the derivative of Eq. (39),

$$3y^2\frac{dy}{dx} - 2(1+h_1)y\frac{dy}{dx} - y^2\frac{dh_1}{dx} + m\frac{d^2F}{dx^2} + \frac{m}{(\ell+x)(1+x)}\frac{dh_1}{dx} + mh_1\frac{d}{dx}\frac{1}{(\ell+x)(1+x)} = 0$$

and examine the terms one by one. At the solution point, we know that Eq. (28) is satisfied so that the terms involving dh_1/dx cancel. Hence, for dy/dx to be zero at this point, the sum of the fourth and sixth terms must vanish. Thus,

$$(\ell+x)^2(1+x)^2\frac{d^2F}{dx^2} - h_1(1+2x+\ell) = 0 \qquad (40)$$

and this is the equation we shall use to determine β. This parameter will be a function of x which, of course, should be evaluated at the solution point wherever that point may be. Since we don't know its location, we will use the function instead of its value— knowing full well that this is the right value at the solution point. It will be almost correct near this point and, hopefully, won't cause serious problems otherwise.*

To evaluate the free parameter β from Eq. (40), we first note that since $F = F(\frac{1}{2}, 1; \frac{3}{2}; -x)$ is a hypergeometric function, it must satisfy Gauss' differential equation†

$$2x(1+x)\frac{d^2F}{dx^2} + (3+5x)\frac{dF}{dx} + F = 0$$

which can be used to eliminate the second derivative from Eq. (40) to obtain

$$(\ell+x)^2(1+x)\left[(3+5x)\frac{dF}{dx} + F\right] + 2x(1+2x+\ell)h_1 = 0. \qquad (41)$$

Also, from the derivation of Eq. (32), we have an expression for the first derivative

$$\frac{dF}{dx} = \frac{1}{2x}\left[\frac{1}{1+x} - F\right]; \qquad (42)$$

but, since x can vanish, we must somehow eliminate the indeterminacy. To this end, reference Eq. (33) and define

$$F = \frac{1}{1+xG} \quad \text{where} \quad G = \cfrac{1}{3 + \cfrac{4x}{5 + \cfrac{9x}{7+\cdots}}} \qquad (43, 44)$$

to obtain, thereby,

$$\frac{dF}{dx} = -\frac{(1-G)F}{2(1+x)}.$$

In the interest of simplifying the final result, it happens that we should also define ξ by

$$G = \cfrac{1}{3 + \cfrac{4x}{\xi}} \quad \text{where} \quad \xi = 5 + \cfrac{9x}{7 + \cfrac{16x}{9+\cdots}} \qquad (45, 46)$$

so that

$$\frac{dF}{dx} = -\frac{(2x+\xi)FG}{\xi(1+x)} = -\frac{2x+\xi}{(1+x)[4x+\xi(3+x)]}.$$

When we substitute in Eq. (41) and make appropriate reductions, we obtain the following equation‡ for h_1

$$h_1 = \frac{(\ell+x)^2(1+3x+\xi)}{(1+2x+\ell)[4x+\xi(3+x)]}. \qquad (47)$$

Finally, we have no difficulty in also deriving

$$h_2 = \frac{m(x-\ell+\xi)}{(1+2x+\ell)[4x+\xi(3+x)]} \qquad (48)$$

where we have introduced the notation $-h_2$ for the last term in Eq. (39).

*Gauss achieved a similar flattening of his cubic equation (14) for values of x near $x = 0$ since he was able to insure that his coefficient h had only a second order variation with x.

†In general: $x(1-x)\frac{d^2F}{dx^2} + [\gamma - (\alpha+\beta+1)x]\frac{dF}{dx} - \alpha\beta F = 0$.

‡Since h_1 is proportional to β, we have, in fact, determined the free parameter.

Table 1 Continued Fraction Levels $[\xi(x),\ \xi(\eta)]$

$T=$	0.3	0.5	0.7	0.9	1.0	3.0	5.0	7.0	9.0	11.0
$\lambda=-0.9$	81, 19	50, 15	36, 13	27, 11	24, 10	12, 6	46, 9	81, 9	105, 10	124, 10
-0.7	60, 17	37, 13	26, 10	19, 9	17, 8	16, 7	29, 8	38, 9	45, 9	50, 9
-0.5	44, 14	26, 11	18, 9	13, 7	11, 6	15, 7	22, 8	26, 8	30, 8	33, 8
-0.3	31, 12	19, 9	13, 7	10, 6	8, 5	12, 6	17, 7	20, 7	23, 8	25, 8
-0.1	20, 9	13, 7	10, 6	8, 5	7, 5	11, 6	14, 7	17, 7	19, 7	20, 7
0.0	15, 8	11, 6	9, 6	7, 5	6, 5	10, 6	13, 6	15, 7	17, 7	19, 7
0.1	11, 6	9, 6	8, 5	7, 5	6, 4	9, 5	12, 6	14, 7	16, 7	17, 7
0.3	7, 5	7, 5	6, 4	5, 4	5, 4	8, 5	10, 6	12, 6	13, 6	14, 7
0.5	5, 4	5, 4	5, 4	4, 3	4, 3	7, 5	9, 5	11, 6	12, 6	13, 6
0.7	4, 3	4, 3	3, 3	3, 2	3, 3	6, 4	8, 5	9, 5	10, 6	11, 6
0.9	2, 2	3, 2	3, 3	3, 3	4, 3	6, 4	7, 5	8, 5	9, 5	10, 6

In summary, then, the analogs of Gauss' two equations (13) and (14) are

$$x=\sqrt{\left(\frac{1-\ell}{2}\right)^2+\frac{m}{y^2}}-\frac{1+\ell}{2} \tag{49*}$$

and

$$y^3-y^2-h_1y^2-h_2=0 \tag{50}$$

with the coefficients determined from Eqs. (47) and (48) and the function $\xi(x)$ calculated from the continued fraction

$$\xi=5+\cfrac{\frac{9}{7}x}{1+\cfrac{\frac{16}{63}x}{1+\cfrac{\frac{25}{99}x}{1+\cdots}}} \qquad -1\le x<\infty \tag{51}$$

which is equivalent to Eq. (46).

Fine Points of the New Algorithm

In both Gauss' method and the new algorithm, the two operations — evaluation of the continued fraction and solution of the cubic equation — are the most time consuming. In *Theoria Motus*, Gauss prepared tables for this purpose which span the region over which his method is useful and valid. In lieu of tables, we shall present efficient computational techniques for these two aspects of our algorithm.

In Ref. (1) a top-down method is described for evaluating continued fractions, which we can use to determine $\xi(x)$, and will not be considered further here. However, we must remark that, although the continued fraction representation of $\xi(x)$ in Eq. (51) converges rapidly for small values of x, the number of levels required increases significantly as x becomes large.

A substantial improvement in the rate of convergence can be had at the expense of a little extra preliminary computation. For if we define

$$\eta=\frac{x}{(\sqrt{1+x}+1)^2} \qquad -1<\eta<1, \tag{52}$$

then it is shown in the Appendix that

$$\xi(x)=\cfrac{8(\sqrt{1+x}+1)}{3+\cfrac{1}{5+\eta+\cfrac{\frac{9}{7}\eta}{1+\cfrac{\frac{16}{63}\eta}{1+\cfrac{\frac{25}{99}\eta}{1+\cdots}}}}} \tag{53}$$

In Table 1 is a comparison of the number of continued fraction levels required to compute $\xi(x)$ to eight significant digits for various values of λ and

$$T\equiv\sqrt{\frac{8\mu}{s^3}}(t_2-t_1). \tag{54}$$

using both Eqs. (51) and (53).

Turning our attention now to solving the cubic equation (50), we observe that Eq. (29) has only one positive real root since the right hand side of that equation is positive for all orbits. Furthermore, that root must exceed unity in magnitude. This is the solution to our problem and, of course, must be a root of Eq. (50) also.

It is not difficult to show that $h_1(x)$, defined in Eq. (47), is always positive, but $h_2(x)$, defined in Eq. (48), can have either sign. Hence, there can be more than one positive real root of Eq. (50) and the question of which is the proper choice must be answered.

For parabolic orbits with $\lambda>(1-\sqrt{5})/(1+\sqrt{5})$, it is easy to verify that h_2 is positive and, consequently, for that case Eq. (50) has exactly one positive real root. Now suppose that λ and T vary continuously; then a simple continuity argument will suffice to demonstrate that, when multiple roots are present, the largest is always the correct choice.

The classical explicit formulas for obtaining the roots of a cubic can, of course, be used for solving Eq. (50). They are somewhat cumbersome in that different formulas are required depending on the number of real roots. However, there is an extremely attractive formula, utilizing continued fractions, which guarantees to produce always the correct root.

*Equation (49) is obtained from Eq. (28) by solving for x using the quadratic formula. The choice of sign for the radical is governed by the specified range $-1<x<\infty$.

Table 2 Number of Iterations (Gauss' Method, New Method)

$T=$	0.3	0.5	0.7	0.9	1.0	3.0	5.0	7.0	9.0	11.0
$\lambda=-0.9$	†, 4	†, 4	†, 5	†, 5	†, 5	67, 8	210, 7	190, 6	167, 5	151, 5
−0.7	†, 4	†, 5	†, 5	†, 5	†, 5	44, 5	52, 5	52, 4	49, 4	48, 3
−0.5	†, 4	†, 4	†, 4	†, 4	†, 5	24, 4	26, 4	19, 3	29, 4	32, 4
−0.3	†, 4	†, 4	†, 4	29, 4	17, 4	14, 4	14, 3	19, 4	21, 4	23, 5
−0.1	†, 4	16, 4	10, 4	9, 4	8, 4	10, 4	12, 4	14, 5	15, 5	17, 5
0.0	‡, 4	‡, 4	‡, 4	‡, 4	‡, 4	‡, 4	‡, 4	‡, 5	‡, 5	‡, 5
0.1	5, 4	6, 4	6, 4	5, 4	5, 4	7, 4	9, 4	11, 5	13, 5	13, 5
0.3	3, 4	4, 4	4, 4	4, 4	4, 4	6, 4	8, 5	10, 5	10, 5	11, 5
0.5	3, 3	3, 3	3, 3	3, 3	3, 3	5, 4	6, 5	8, 5	8, 5	9, 5
0.7	3, 3	3, 3	3, 3	3, 3	3, 3	4, 4	5, 4	7, 5	7, 5	8, 5
0.9	2, 2	2, 2	2, 3	2, 3	3, 3	4, 4	5, 4	6, 5	6, 5	7, 5

†Gauss' method does not converge. ‡Gauss' method is singular for $\theta=180$ degrees.

For this purpose, we first calculate

$$u=-\frac{B}{2(\sqrt{1+B}+1)} \tag{55}$$

where

$$B=\frac{27h_2}{4(1+h_1)^3}. \tag{56}$$

Then, the largest positive real root of Eq. (50) is found from*

$$y=\frac{1+h_1}{3}\left[2+\frac{\sqrt{1+B}}{1-2uK(u)^2}\right] \tag{57}$$

where $K(u)$ may be calculated from the continued fraction

$$K(u)=\cfrac{\frac{1}{3}}{1-\cfrac{\frac{4}{27}u}{1-\cfrac{\frac{8}{27}u}{1-\cfrac{\frac{2}{9}u}{1-\cdots}}}}. \tag{58}$$

In general, γ_{2n+1} and γ_{2n}, the odd and even numbered coefficients of u in the continued fraction can be obtained from

$$\gamma_{2n+1}=\frac{2(3n+2)(6n+1)}{9(4n+1)(4n+3)}$$

and

$$\gamma_{2n}=\frac{2(3n+1)(6n-1)}{9(4n-1)(4n+1)}.$$

The quantity $1+B$ can be shown to be always positive so that no difficulty is encountered in the square root. Indeed, unless this term is positive, there will not exist a positive real root of Eq. (50).

Of course, we can always resort to the Newton-Raphson method for finding the root but we must be careful that we converge to the correct root. A little analysis confirms that an appropriate starting value y_0 is either zero or $\frac{2}{3}(1+h_1)$ according as $h_2/(1+h_1)^3$ is or is not less than $-\frac{4}{27}$, respectively. Also, of course, on each subsequent cycle in the iteration it is sensible to use the value of y calculated during the previous cycle.

*The derivation of this formula can be found in Ref. (4). It is too complex to be reproduced here in limited space.

Comparing the Two Methods

Although the derivations of Gauss' method and the new method for solving Lambert's problem are different, the final equations and mechanics of the algorithms are quite similar. The dimensionless parameters λ and T are the inputs to both methods. The new method requires somewhat more algebra for each iteration since there are two coefficients to be found for the cubic equation and Eq. (28) is quadratic in x. There is no need, however, to test the value of λ as in Gauss' method since the new equations are valid for $-1<\lambda\leq 1$. The efficiency of either procedure is the number of iterations necessary to compute x to a given accuracy.

In Table 2, the two methods are compared to contrast the number of iterations required to compute x to eight significant digits. [Many additional results are presented in Ref. (4).] For Gauss' method, the initial values of x selected to generate this table were

$$x_0=\begin{cases} 0 & \text{parabola, hyperbola} \\ \dfrac{\ell}{1+2\ell} & \text{ellipse} \end{cases}$$

where, for the ellipse, x_0 defines a circular orbit. Note the rapid convergence of Gauss' method in the lower left corner of the table. In this region, x is nearly zero and the transfer angle θ is small. Gauss designed his method for problems of this type. The quantity $\xi(x)$ was ingeniously constructed to be of order x^2 so that it would be very small for small x — the result being that h, the coefficient in the cubic, is nearly independent of x so that y is nearly constant.

The two major disadvantages are the singularity at $\lambda=0$ and the convergence properties over the range of λ and T considered. Although Gauss' method converged for all elliptic cases considered, it sometimes required more than 100 iterations to do so. For most of the hyperbolic cases, it did not converge at all.

On the other hand, the new method was designed to converge rapidly for any case independent of the value of x. The nearly uniform convergence behaviour of the new method is seen in Table 2.

The initial value strategy chosen for x was identical to that used for Gauss' method. In this case, those values are

$$x_0 = \begin{cases} 0 & \text{parabola, hyperbola} \\ \ell & \text{ellipse.} \end{cases}$$

The rapid convergence in the lower left corner of the table is retained using the new method, and there is no significant difference in the rate of convergence for positive or negative values of λ. Although not shown in the table, a striking advantage of the new method is that only one more iteration step is necessary to obtain four more significant figures in all the cases considered!

Because the new method is singular, it is instructive to investigate its behaviour for values of λ approaching -1. Table 3 gives the required number of iterations as λ varies from -0.90 to -0.99. The uniformity of convergence persists except for a narrow region near $T = 5$. The increase in iteration steps is not the result of a poor initial guess for x_0; indeed, it turns out that the number of steps approaches a maximum when the solution of the problem approaches the minimum energy orbit — that orbit for which the semimajor axis is $a_m = \frac{1}{2}s = \frac{1}{4}(r_1 + r_2 + c)$ — when the transfer angle is close to 360 degrees. This same region causes difficulties when using a Newton-Raphson technique as was first noted by Lancaster and Blanchard[5]. The time of flight graph experiences a change in curvature, which necessitates abandonment of the Newton-Raphson method, when λ is near -1. Our method takes a little longer than usual but it still converges without modification.

Table 3 Iterations near 360 degrees

$T =$	1.0	3.0	5.0	7.0	9.0	11.0
$\lambda = -0.99$	5	10	14	8	6	5
-0.98	5	9	12	8	6	5
-0.97	5	9	11	7	6	5
-0.96	5	9	10	7	6	5
-0.95	5	9	9	7	5	5
-0.94	5	8	9	6	5	5
-0.92	5	8	8	6	5	5
-0.90	5	8	7	6	5	5

Conclusions

The method of solving Lambert's problem, as presented in this paper, closely parallels Gauss' classical method but has significant advantages over the latter in terms of generality and performance. The new algorithm is designed to converge rapidly for any given geometry and time of flight. It retains all the original advantages of Gauss' method and is applicable to a wide range of problems of practical importance. The convergence properties of the iteration process are nearly uniform and only a small increase in the number of steps will increase dramatically the precision of the solution.

For most space guidance applications, this method appears to be an extremely attractive alternative to Newton-Raphson schemes since one need not be concerned with determining appropriate starting values for different input parameters. Also, when the problem is to be solved repetitively as part of a guidance cycle with the previous solution used as the initial approximation, one might justifiably anticipate that satisfactory convergence could be had with just a single iteration step.

Appendix

From the classic identity

$$\frac{\frac{1}{2}E}{\tan\frac{1}{2}E} = F(1, \tfrac{1}{2}; \tfrac{3}{2}; -\tan^2\tfrac{1}{2}E)$$

we can deduce

$$\frac{\frac{1}{4}E}{\tan\frac{1}{4}E} = F(1, \tfrac{1}{2}; \tfrac{3}{2}; -\tan^2\tfrac{1}{4}E).$$

Then, since

$$\tan^2\tfrac{1}{4}E = \frac{\sec\frac{1}{2}E - 1}{\sec\frac{1}{2}E + 1} = \frac{\sqrt{1+x}-1}{\sqrt{1+x}+1} \equiv \eta,$$

it follows that

$$F(x) \equiv F(1, \tfrac{1}{2}; \tfrac{3}{2}; -x) = \frac{2\tan\frac{1}{4}E}{\tan\frac{1}{2}E} F(1, \tfrac{1}{2}; \tfrac{3}{2}; -\eta)$$
$$= \frac{2F(\eta)}{\sqrt{1+x}+1}.$$

Therefore, Eq. (43) becomes

$$F(x) = \frac{1}{1 + xG(x)} = \frac{2}{\sqrt{1+x}+1}\left[\frac{1}{1+\eta G(\eta)}\right],$$

from which we obtain

$$G(x) = \frac{1 + G(\eta)}{2(\sqrt{1+x}+1)}.$$

Furthermore, from Eq. (45), we have

$$G(x) = \frac{1}{3 + 4x/\xi(x)}$$
$$= \frac{1}{2(\sqrt{1+x}+1)}\left[1 + \frac{1}{3 + 4\eta/\xi(\eta)}\right]$$

and, in a similar manner, we find that

$$\xi(x) = \frac{8(\sqrt{1+x}+1)}{3 + \dfrac{1}{\eta + \xi(\eta)}}.$$

The verification of Eq. (53) is now complete.

References

[1]Battin, R.H., "Lambert's Problem Revisited," *AIAA Journal*, Vol. 15, May 1977, pp. 707–713.

[2]Battin, R.H., Fill, T.J., and Shepperd, S.W., "A New Transformation Invariant in the Orbital Boundary-Value Problem," *Journal of Guidance and Control*, Vol. 1, Jan.-Feb. 1978, pp. 50–55.

[3]Gauss, K.F., *Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, a translation of *Theoria Motus* by C.H. Davis, Dover Publications, Inc., New York, 1963.

[4]Vaughan, R.M., "An Improvement of Gauss' Method for Solving Lambert's Problem," Charles Stark Draper Laboratory, Inc., Cambridge, Mass., Report T-813, May 1983.

[5]Lancaster, E.R. and Blanchard, E.C., "A Unified Form of Lambert's Theorem," NASA TN D-5368, Sept. 1969.