[6] Eckstrom, C. V. and Preisser, J. S., "Flight Test of a 40-Foot-Nominal Diameter Disk-Gap-Band Parachute Deployed at a Mach Number of 2.72 and a Dynamic Pressure of 9.7 Pounds Per Square Foot," TM X-1623, 1968, NASA.

[7] Preisser, J. S., Eckstrom, C. V., and Murrow, H. N., "Flight Test of a 31.2-Foot-Diameter Modified Ringsail Parachute Deployed at a Mach Number of 1.39 and a Dynamic Pressure of 11.0 Pounds Per Square Foot," TM X-1414, 1967, NASA.

[8] Preisser, J. S. and Eckstrom, C. V., "Flight Test of a 30-Foot-Nominal-Diameter Cross Parachute Deployed at a Mach Number of 1.57 and a Dynamic Pressure of 9.7 Pounds Per Square Foot," TM X-1542, 1968, NASA.

[9] Whitlock, C. H., Henning, A. B., and Coltrane, L. C., "Performance of a 16.6-Meter-Diameter Modified Ringsail Parachute in a Simulated Martian Environment," TM X-1550, 1968, NASA.

[10] Lundstrom, R. R., Darnell, W. L., and Coltrane, L. C., "Performance of a 16.6-Meter-Diameter Cross Parachute in a Simulated Martian Environment," TM X-1543, 1968, NASA.

[11] Berndt, R. J., "Experimental Determination of Parameters for the Calculation of Parachute Filling Times," *Jahrbuck 1964 der WGLR*, Vieweg and Sohn, Braunschweig, Germany, 1965, pp. 299–316.

[12] Ames Research Staff, "Equations, Tables, and Charts for Compressible Flow," Rept. 1135, 1953, NACA.

[13] Maynard, J. D., "Aerodynamic Characteristics of Parachutes at Mach Numbers From 1.6 to 3," TN D-752, 1961, NASA.

Normalization and Convergence Acceleration for Indirect Optimization Methods

S. D. Williams*
Lockheed Electronics Company, Houston Aerospace Systems Division, Houston, Texas

THE mathematical descriptions of various optimal physical processes can be formulated as two-point boundary value problems.[1,2] In recent years, there has been a vigorous effort in defining the mathematical process and refining the solution techniques.[3] To determine the optimal trajectory of a continuously thrusting space vehicle, the differential equations of motion are usually nonlinear and, hence, require numerical techniques for obtaining a solution. The methods used are classified as either indirect methods (based on the conditions necessary to achieve mathematical optimality) or direct methods (which use only the differential equations governing the state and the specified boundary conditions). In both methods, successive changes in the initial solution are made in an attempt to satisfy the boundary conditions while extremizing a performance index. In this discussion, only indirect methods are considered.

Previously, only two factors were considered to affect the convergence characteristics of the indirect methods: the numerical integration procedure and the correction procedure that modified the unknown boundary conditions. Now, it is shown that the region in which the initial data are embedded can also have a significant effect upon the rate of convergence for an iteration process and the size of the convergence envelope.

Received July 22, 1969. This study was conducted under contract NAS 9-5384. The author wishes to thank the Theory and Analysis Office for supporting this investigation and G. C. Stuckum, Lockheed Electronics Company, for preparing the illustrations.

* Scientific Programming Specialist, Engineering Applications Section, Information Processing Branch.

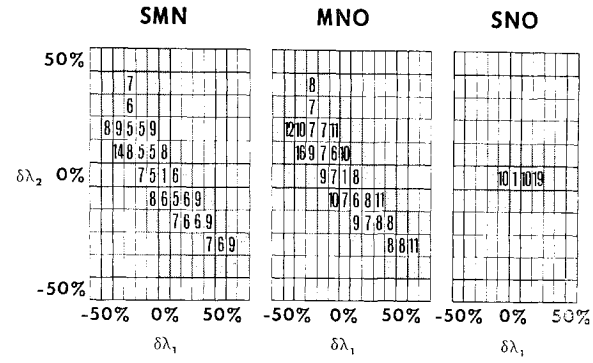

Fig. 1 Convergence envelope for MPF, final time optimal.

Computational Range

Much literature for solving problems on a computer is devoted to techniques for improving stability and reducing the effect of round off and truncation. The equations used are not restricted to a numerical reference frame, whereas the computer used to solve or express these equations is constrained to a fixed word length. The most suitable interval for performing arithmetic operations is the interval $[-1,1]$. This may be stated more formally as follows.

Theorem 1. The unit interval $[0,1]$ is more dense than any other unit positive interval for a fixed word length computer. That this is the case can be seen immediately if we consider that the computer has a mantissa represented by m bits and an exponent has e bits. Then the unit interval $[0,1]$ has 2^{m+e-1} different numbers or representations as an upper bound, whereas the interval $[a, a+1]$ would have $2^{m-n}+1$ numbers, where the representation for a requires n bits, $n>0$. As a direct corollary, we have:

Corollary 2. The most dense biunit interval on a computer is the interval $[-1,1]$.

The direct implication that we would like to make is that the more bits we have to represent a number the more accurate the representation is. This leads to the following result:

Lemma 3. Let A and B be two computer words of length m and n, respectively, where $m>n$. Then a number can be represented more accurately in A than in B.

This lemma can be proved by considering the error between the two representations for an unending fraction. By choosing "nice" numbers the error can be the same, but for all other numbers the error is reduced by using the greater word length.

Normalization of the State Data

In solving problems on a computer, the units of measurement which are convenient in terms of physics or engineering may not be convenient for accuracy, as was shown in the previous section.

A convenient choice of normalized units for an Earth-Mars orbital transfer by Tapley and Lewallen[3] was the unit mass as the mass of the space vehicle, unit length as the Earth's orbital distance from the sun, and initial velocity as unity which established a time quantity and made the gravitational constant unity. This choice yields the fundamental quantities as mass, length, and velocity; all other units are derived from these.

Optimization Problem

In general the optimization problem can be stated as follows: determine the trajectory of the variables that are capable of controlling a dynamic system such that the functional

$$\hat{I} = \phi(x_0,x_f,t_f) + \int_{t_0}^{t_f} \theta(x,u,t)dt \qquad (1)$$

is extremized subject to the constraints, $\Omega(x_0,t_0) = 0$, $\psi(x_f,t_f) = 0$, and $\dot{x} = f(x,u,t)$, where Ω is a p vector of initial constraints; ψ is a q vector of terminal constraints; and $\dot{x}$ is the set of n first-order, nonlinear differential equations that describe the trajectory, x is a state vector, u is an m vector of variables that control the trajectory; t is time; and subscripts 0 and f represent initial and final values.

In the formal process of determining the extremal value of Eq. (1), the constraints are usually adjoined by Lagrange multipliers, and the new functional to be extremized is

$$I = \phi + \nu^T\Omega + \mu^T\psi + \int_{t_0}^{t_f} [\theta - \lambda^T(f - \dot{x})]dt \quad (2)$$

where ν and μ are constant Lagrange multipliers and λ is a time dependent Lagrange multiplier.

The first necessary conditions for an extremum of Eq. (2) may be derived by taking the first variation of I and setting it to zero. When this is done, several transversality conditions must be satisfied, in addition to those already mentioned. At times t_0 and t_f are:

$$\Omega(x_0,t_0) = 0, \quad [\lambda^T + P_x]_{t_0}dx_0 = 0 \quad (3)$$

$$\psi(x_f,t_f) = 0, \quad [P_x - \lambda^T]_{t_f}dx_f = 0 \quad (4)$$

$$[P_t + H]_{t_f}dt_f = 0 \quad (5)$$

where $P(x_0,x_f,t_0,t_f) = \phi(x_0,x_f,t_f) + \nu^T\Omega(x_0,t_0) + \mu^T\psi(x_f,t_f)$, and $H(x,\lambda,u,t) = \theta + \lambda^T f$. In addition, the equations that must be solved everywhere along the trajectory are

$$\dot{x} = H_\lambda{}^T, \quad \dot{\lambda} = -H_x{}^T, \quad 0 = H_u \quad (6)$$

For a local minimium to exist, we also require

$$H_{uu} \geq 0 \quad (7)$$

The optimization problem is usually now reduced to a two-point boundary value problem. The multipliers ν and μ are usually eliminated from the boundary conditions, and the control variables are usually eliminated from the first two equations in Eqs. (6) by use of the third equation. The problem then becomes

$$\dot{z} = \begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} H_\lambda{}^T \\ -H_x{}^T \end{bmatrix} = F(z,t) \quad (8)$$

with boundary conditions, $g(z_0,t_0) = 0$, $h(z_f,t_f) = 0$, which specifies p initial boundary conditions and $q + 1$ terminal conditions, $p + q = 2n$.

In the usual methods used to obtain a numerical solution to the two-point boundary value problem, an initial estimate is made for the q unknown elements of z_0 [these usually correspond to the Lagrange multipliers in Eqs. (6)] and t_f. In general, convergence $h(z_f,t_f) = 0$ will not be achieved. Thus, to achieve convergence, changes must be made for the unknown quantities in z_0 and t_f.[3,4]

Normalization of the Lagrange Multipliers

Hestenes[2] has shown that the Lagrange multipliers in Eq. (2) exist and are unique. It is not difficult to show that if Eq. (2) is mapped by the transformation $F: I \rightarrow I/a$, where a is some positive constant, the two problems are equivalent. If the choice for a is made to be $|\lambda_{i_0}|$ where λ_{i_0} is the initial value of the ith Lagrange multiplier, then one of the terminal boundary conditions can be exchanged for an initial condition.

This transformation then serves several purposes. It places the Lagrange multipliers into a more desirable computational region; it reduces the dimensionality of the problem to be solved, and it requires one less integration of the perturbation equations

$$\delta\dot{z} = \begin{bmatrix} H_{\lambda x} & H_{\lambda\lambda} \\ -H_{xx} & -H_{x\lambda} \end{bmatrix} [\delta z] \quad (9)$$

which are used in determining corrections in the initial estimates for the q unknown parameters in Eq. (6). Breakwell[5] normalized the Lagrange multipliers in solving an interceptor problem as did Jazwinski[6] and McGill and Kenneth[7] in solving an Earth-Mars orbital transfer problem; however, no justification was given for its use.

The primary difficulty encountered when attempting to normalize the Lagrange multipliers is in assigning the proper sign to the multiplier, i.e., ± 1. Care must also be taken to choose a multiplier that does not vanish initially. These problems usually can be resolved satisfactorily by examining the transversality conditions in Eqs. (3–5).

Discussion of the Numerical Results

The minimum time Earth-Mars transfer problem may be stated as follows: Determine the control history $\gamma(t)$ such that a spacecraft may transfer from initial conditions corresponding to those of Earth to terminal conditions corresponding to those of Mars in minimum time.

The differential equations of motion are

$$\dot{x}_1 = \dot{u} = v^2/r - GM/r^2 + (T/m)\sin\gamma$$

$$\dot{x}_2 = \dot{v} = -uv/r + (T/m)\cos\gamma$$

$$\dot{x}_3 = \dot{r} = u, \; \dot{x}_4 = \dot{\theta} = v/r$$

where u, v, r, and θ are the radial velocity, tangential velocity, radial position, and angular position, respectively; the control variable γ is the thrust orientation to the local horizontal; GM is the gravitational constant of the sun; and $m = m_0 - \dot{m}t$ is the vehicle mass.

When the optimization process is applied, γ is eliminated and four additional equations (Euler-Lagrange equations) are to be satisfied;

$$\dot{\lambda}_1 = (v/r)\lambda_2 - \lambda_3$$

$$\dot{\lambda}_2 = -(2v/r)\lambda_1 + (u/r)\lambda_2 - (1/r)\lambda_4$$

$$\dot{\lambda}_3 = (v^2/r^2 - 2GM/r^3)\lambda_1 - (uv/r^2)\lambda_2 + (v/r^2)\lambda_4$$

$$\dot{\lambda}_4 = 0$$

The optimality condition $H_\gamma = 0$ leads to

$$\sin\gamma = \lambda_1/[\pm(\lambda_1{}^2 + \lambda_2{}^2)^{1/2}] \text{ and } \cos\gamma = \lambda_2/[\pm(\lambda_1{}^2 + \lambda_2{}^2)^{1/2}],$$

and from $H_{\gamma\gamma} > 0$ for a minimum, the minus sign is chosen for the radical.

The Earth-Mars time optimal orbital transfer is completely defined when the initial Lagrange multipliers and t_f have been determined; λ_{4_0} is determined to be zero. In an effort to measure the sensitivity of a method to the initially assumed values, a number of cases have been investigated. Two indirect methods were used to solve numerically the orbital transfer problem: the Method of Perturbation Functions (MPF)[3] and the Modified Quasilinearization Method (MQM).[8] The MPF requires assumed values for the Lagrange multipliers and requires final time for the first iteration; the MQM requires an assumed solution in addition to those parameters.

The numerical results of the sensitivity study are best illustrated by building envelopes of convergence (Fig. 1), the boundaries of which are illustrated by the last convergent trial. Attempts beyond this region do not result in a converged solution and are represented by the shaded region. The percentage numbers on the axes represent the percent deviation in the initial trial values from the initial values which yield an optimal solution. For simplicity, only one correction procedure was used, the normal correction procedure (NCP).[4]

Three different convergence envelopes were generated for both MPF and MQM: 1) the initial state data and the Lagrange multipliers were normalized (SMN), 2) the initial state data were normalized (SNO), and 3) the initial Lagrange

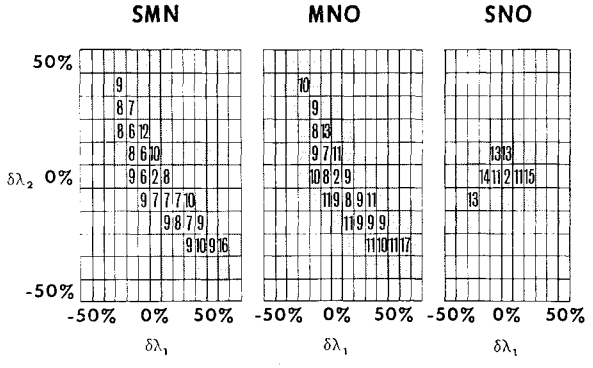

Fig. 2 Convergence envelope for MQM, final time optimal.

multipliers were normalized (MNO). It was unnecessary to consider a case with no normalization since this information could be inferred from the other three cases. In all cases, the value for the final time chosen initially was optimal as was the value for λ_{4_0}. The Lagrange multipliers were normalized by λ_{3_0}, and the optimal value for λ_{3_0} was used for the case when only the initial state data were normalized. For the unnormalized initial state data, an M-K-S system of units was used.

If the initially assumed values (Fig. 1) of λ_{1_0} and λ_{2_0} corresponding to different trial trajectories are presented such that λ_{1_0} is the abscissia and λ_{2_0} is the ordinate, a convergence envelope is formed about the optimal values of λ_{1_0} and λ_{2_0}. The coordinates for the envelopes are the deviations in percent from the optimal values of λ_{1_0} and λ_{2_0}.

Figure 1 illustrates the envelopes of convergence for MPF for the SMN, MNO, and SNO cases. It is seen that normalizing the state data reduces the number of iterations required to achieve convergence, but does not effectively change the size of the convergence envelope. By contrast, normalizing the multipliers not only reduces the number of iterations required to achieve convergence, but also increases the size of the convergence envelope. Normalizing both the state data and the multipliers produces the most favorable results. Figure 2 presents similar envelopes of convergence for MQM. The trend for MQM is the same as for MPF; however, the convergence envelope for the SNO case is slightly larger for MQM than for MPF. This is apparently the result of establishing an effective metric to be used as a convergence criteria for MQM. The initial reference solution used for MQM was the nonlinear differential equations. The physical significance of the convergence envelopes for SMN and MNO cases is explained in detail in Ref. 3.

Normalizing the state and the multipliers caused an increase in size of the convergence envelope by a factor of 700% for MPF and over 300% for MQM. In addition, the computational efficiency was improved by about 50% in terms of the number of iterations required for convergence, and a 20% reduction in computer time per iteration was achieved by not integrating the additional perturbation vector.

Conclusions

The computational advantages of embedding the initial data into the interval $[-1,1]$ have been investigated for the indirect methods of trajectory optimization. The normalization of both the state data and the Lagrange multipliers produced the most significant results, and normalizing only the multipliers gave more favorable results than normalizing only the state data.

Based on the results of this investigation, normalizing the state data and the multipliers is a significant step toward reducing the sensitivity of the initially assumed parameters for convergence characteristics of the indirect optimization methods. This improvement combined with a more sophisticated correction scheme greatly improves the possibility of convergence in one computer run. In addition, this normalization takes full advantage of the numerical procedure and could be used with equal success for direct methods.

References

[1] Bliss, G. A., *Lectures on the Calculus of Variations,* The Univ. of Chicago Press, Chicago, Ill., 1946.

[2] Hestenes, M. R., *Calculus of Variations and Optimal Control Theory,* Wiley, New York, 1966.

[3] Tapley, B. D. and Lewallen, J. M., "Comparison of Several Numerical Optimization Methods," *Journal of Optimization Theory and Applications,* Vol. 1, No. 1, 1967.

[4] Lewallen, J. M., Tapley, B. D., and Williams, S. D., "Iteration Procedures for Indirect Trajectory Optimization Methods," *Journal of Spacecraft and Rockets,* Vol. 5, No. 3, March 1968, pp. 321–327.

[5] Breakwell, J. V., "The Optimization of Trajectories," *Journal of the Society for Industrial and Applied Mathematics,* Vol. 7, No. 2, 1959.

[6] Jazwinski, A. H., "Optimal Trajectories and Linear Control of Nonlinear Systems," *AIAA Journal,* Vol. 2, No. 8, Aug. 1964, pp. 1371–1379.

[7] McGill, R. and Kenneth, P , "Solution of Variational Problems by Means of a Generalized Newton-Raphson Operator," *AIAA Journal,* Vol. 2, No. 10, Oct. 1964, pp. 1761–1766.

[8] Lewallen, J. M., "A Modified Quasi-Linearization Method for Solving Trajectory Optimization Problems," *AIAA Journal,* Vol. 5, No. 5, May 1967, pp. 962–965.

Simplified Technique for Estimating Navigational Accuracy of Interplanetary Spacecraft

LOUIS KINGSLAND JR.* AND WILLARD E. BOLLMAN†
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

THE principal way to determine the position and velocity of a spacecraft on an interplanetary trajectory is by observing the Doppler shift in a coherent, two-way radio signal between Earth tracking stations and the spacecraft. By combining observations over a period of time, it is possible to estimate the trajectory of the spacecraft with considerable accuracy. It is also possible to estimate the navigational accuracy that might be obtained on potential missions by performing statistical studies with real-time orbit determination programs, but the process is very costly and time-consuming, largely because of the requirement to integrate the equations of spacecraft motion.

Attempts have been made to model the planet-spacecraft geometry by means of perfect hyperbolic motion, and while reasonably successful in reproducing the near-planet trajectory geometry, this approach has not been adequate to model the "transitional" phase of the trajectory, where the sun and the planet significantly affect the motion of the spacecraft. Much important tracking data are taken during this period, and experience has shown that a trajectory model that assumes perfectly hyperbolic spacecraft-planet motion fails to

Received August 13, 1969; presented as Paper 69-899 at the AIAA/AAS Astrodynamics Conference, Princeton, N. J., August 20–22, 1969; revision received October 6, 1969. This paper represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract NAS 7-100, sponsored by NASA.

* Manager for Mission Analysis and Engineering, Viking Orbiter Office.

† Member of Technical Staff, Systems Analysis Section.