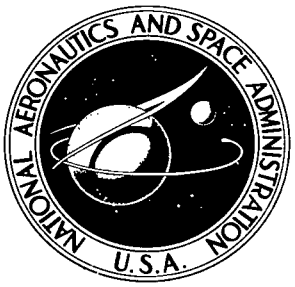
NASA TECHNICAL NOTE

NASA TN D-4542

C. 1



NASA TN D-4542

LOCAL-ERROR CONTROL AND ITS EFFECTS ON THE OPTIMIZATION OF ORBITAL INTEGRATION

by C. E. Velez

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JUNE 1968

NASA TN D-4542

LOCAL-ERROR CONTROL AND ITS EFFECTS ON THE OPTIMIZATION OF ORBITAL INTEGRATION

By C. E. Velez

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

ABSTRACT

In the numerical integration of orbits by a multistep process, it has been suggested that the application of a local error control may increase the efficiency of the integration without any significant loss of accuracy. This study develops methods of controlling the local error and examines the effects of these controls on the integration. It is shown that for various orbit types, controlling the local error can optimize the integration while maintaining the desired accuracy.

CONTENTS

LOCAL ERROR CONTROL AND ITS EFFECTS ON THE OPTIMIZATION OF ORBITAL INTEGRATION

by
C. E. Velez
Goddard Space Flight Center

INTRODUCTION

Many computing systems used to determine the orbits of artificial earth satellites require numerical integration. Due to recent advances in the theory of perturbations and the concomitant increase in the complexity of the mathematical model, highly efficient integration techniques are required. Such a technique is described in this paper, based on the concept of "local error control."

Consider the orbital equations

$$\ddot{x} = \frac{-\mu x}{|x|^3} + P \qquad x(t_0) = x_0 \,, \tag{1}$$

in three space variables, where $|x| = (x^2 + y^2 + z^2)^{1/2}$, P is the perturbation function, and μ is a constant. We assume throughout that P is fairly complicated, so that the efficiency of the integration is proportional to the total number of derivative evaluations. Also, for simplicity, we assume that P is independent of the first derivative $\dot{x}$, i.e., $P = P(t, x)$.

The numerical method under consideration is a multistep process, i.e., a method that approximates the solution of Equation 1 by using formulas of the form

$$\sum_{i=0}^{k} \alpha_i x_{n-i} = h^2 \sum_{i=0}^{k} \beta_i \ddot{x}_{n-i} \qquad n = k, k+1, \cdots \,, \tag{2}$$

where h is the stepsize, $x_i = x(t_0 + ih)$, and α_i, β_i are constants. Formulas of this type define a linear k-step method (Reference 1); they are usually derived by integrating a polynomial approximation of the derivative $\ddot{x}$.

Associated with Equation 2 is a local truncation error of the form

$$R_n = R(t_n) = Ch^{p+2} x^{(p+2)}(\xi) \,, \tag{3}$$

where ξ is contained in $[t_{n-k}, t_n]$, C is a constant, and p is an integer called the "order" of the method, all depending on k, the number of "backpoints" used.

Soar has suggested (Reference 2) controlling the local error by varying the order and stepsize while integrating Equation 1 with this object: It is clear (for example) that the magnitude of R_n is sensitive to variations in p or h. Suppose that during the integration the magnitude of R_n becomes insignificant relative to the required accuracy of the calculations being performed. Then increasing h or decreasing p, separately or in combination, may increase integration efficiency without sacrificing accuracy. The purpose of this study, then, is to develop automatic methods of controlling the magnitude of R_n during integration by varying the parameters p and h, and to examine the effects* of these controls on efficiency and accuracy. Before discussing methods of estimating and controlling the local error, we formulate a commonly used integration model that was used to obtain the conclusions given under "Numerical Results."

THE INTEGRATION FORMULAS

In Equation 2, if $\beta_0 \neq 0$, then knowledge of the solution x_n is required on both sides of the equation and cannot, in general, be explicitly solved. However, equations of this type (closed form) have smaller associated truncation errors as well as desirable stabilizing characteristics (Reference 3). The well-known predictor-corrector algorithm uses formulas of this type by first computing an initial (predicted) approximation of the solution using a similar formula with $\beta_0 = 0$, then using the closed form iteratively until convergence is achieved. It can be shown that for sufficiently small h, the successive corrected values obtained by this process converge to the unique solution of the closed-form equation, provided that the function being integrated is sufficiently smooth.

Consider now the predictor-corrector formulas that are derivable from Newton's backward difference interpolation polynomial (Reference 1, pp. 291-293—see also Appendix A):

$$\text{Predictor: } x_n + 2x_{n-1} + x_{n-2} = \nabla^2 x_n = h^2 \left[1 + \frac{1}{12}\nabla^2 + \frac{1}{12}\nabla^3 + \frac{19}{240}\nabla^4 + \frac{3}{40}\nabla^5 + \cdots + \sigma_q \nabla^q\right] \ddot{x}_{n-1} \quad , \tag{4}$$

$$\text{Corrector: } x_n - 2x_{n-1} + x_{n-2} = \nabla^2 x_n = h^2 \left[1 - \nabla + \frac{1}{12}\nabla^2 - \frac{1}{240}\nabla^4 - \frac{1}{240}\nabla^5 + \cdots + \sigma_q^* \nabla^q\right] \ddot{x}_n \quad . \tag{5}$$

These are the Störmer-Cowell formulas expressed in terms of backward differences. The local truncation errors associated with the formulas are given by Equation 3, where $p = q + 1$ and $C = \sigma_{q+1}$ or σ^*_{q+1}. If the parameter q is fixed (and hence the order p), then Equations 4 and 5 can be written in the form of Equation 2 if the backward differences are expressed in terms of the ordinates $\ddot{x}_i$.

*Local error control does not generally yield a quantitative measure of the accumulated error. However, a qualitative control is possible and in most cases sufficient. For a discussion on accumulated error estimations, see Reference 1.

In particular, for $q = 5$ there are the Störmer-Cowell sixth-order formulas:

$$\nabla^2 x_n = \frac{h^2}{240}\left[317\ddot{x}_{n-1} - 266\ddot{x}_{n-2} + 374\ddot{x}_{n-3} - 276\ddot{x}_{n-4} + 109\ddot{x}_{n-5} - 3\ddot{x}_{n-6}\right], \quad (4)'$$

$$\nabla^2 x_n = \frac{h^2}{240}\left[18\ddot{x}_n + 209\ddot{x}_{n-1} + 4\ddot{x}_{n-2} + 14\ddot{x}_{n-3} - 6\ddot{x}_{n-4} + \ddot{x}_{n-5}\right]. \quad (5)'$$

The coefficients in Equations 4' and 5' clearly depend on the choice of order, so that varying the order during integration would mean producing a new set of coefficients for each p. Therefore, formulas such as Equations 4 and 5 (where variations of the order can be made simply by varying the number of terms retained) were used in this study.* Before any of the above formulas could be used, a set of "starting" values of the solution must be computed.† For example, for Equations 4' and 5', if the values x_i (and hence $\ddot{x}_i$), $i = 0, 1, \cdots, 5$ are known, then Equation 4' could be used to obtain a "predicted" value x_6^P ($n = 6$) and Equation 5' to obtain the successive corrections $x_6^{C_j}$, $j = 1, 2, 3, \cdots$, until convergence to the same criterion δ is achieved, i.e., until

$$\left|x_6^{C_{j+1}} - x_6^{C_j}\right| \leq \delta .$$

The process could then be repeated with $n = 7, 8, \cdots$.

LOCAL ERROR ESTIMATION

Any considerations concerning the control of the local error depend on the capability of obtaining a reasonable estimate of Equation 3 during the integration. A widely used approximation is based on the fact that if $f(x)$ is an n-times continuously differentiable function, then there exists a quantity θ, $0 < \theta < 1$, such that

$$\frac{\Delta^n f(x)}{(\Delta x)^n} = \frac{d^n f}{dx^n}\left[x + n\theta\,(\Delta x)\right],$$

i.e., if Δx is sufficiently small, we can approximate a high-order derivative by a high-order difference. In particular, we can write

$$R_n = R(t_n) = Ch^{p+2} x^{(p+2)}(\xi) = Ch^{p+2}\ddot{x}^{(p)}(\xi) \cong Ch^2\nabla^p\ddot{x}_n . \quad (6)$$

*In actual practice, a modification of Equations 4 and 5 known as the "summed" form of the integration formulas was used. This modification is formulated in Appendix A.

†Experimentation indicates that recently developed high-order Runge-Kutta type formulas are particularly suited for such purposes (Reference 4).

If suitable predictor-corrector formulas are used, another technique is available: comparing the predicted value x_n^p with the finally accepted "corrected" value x_n^c. In particular, it can be proven (Reference 1) that, for Equations 4 and 5,

$$R_n \cong K\left(x_n^c - x_n^p\right) , \tag{7}$$

where K is a constant. This technique is known as Milne's method of estimating the local error.

LOCAL ERROR CONTROL

Consider now a specific technique which makes it possible to control the magnitude of the local error during integration by varying parameters p and h. Let T_1 and T_2 be tolerances specifying upper and lower bounds on the local error so that, for any n, the local error must satisfy

$$T_2 \leq \left|R_n\right| \leq T_1 , \quad \text{where} \quad T_2 \leq T_1 . \tag{8}$$

Controlling the local error by varying the order can then be accomplished by determining whether this condition is satisfied for each value of n; if for some n, $R_n < T_2$ then decrease the order; if $R_n > T_1$, then increase the order. However, varying the order alone is generally not enough for effective control. For example, if, for some n, $R_n > T_1$ and h is large, it may not be possible to satisfy Equation 8 with any p; in which case the stepsize must be decreased. (Also, the danger of numerical instability increases considerably with large p; see References 5 and 6.) On the other hand, $R_n < T_2$ and p small would indicate that a larger stepsize could be used.

Controlling the local error by varying both stepsize and order can be accomplished as follows: Let L_1 and L_2 be limits specifying the desired upper and lower bounds on the order, i.e.,

$$L_2 \leq p \leq L_1 \quad \text{where} \quad L_2 \leq L_1 . \tag{9}$$

Then, if at some point during the integration, $p > L_1$, decrease the stepsize; if $p < L_2$, increase the stepsize.

In a local-error control so designed, the parameters T_1, T_2, L_1, L_2 completely govern the *degree* and *type* of control; i.e., T_1 and T_2 can be selected so as to effect any degree from no control to continuous step-by-step control; likewise, L_1 and L_2 can be selected so that the control involves varying the order alone, varying the step alone, or varying both. (For example, if $L_1 = L_2$, the control depends solely on stepsize variations.)

Changes in stepsize magnitude during integration (unlike changes in the order) are usually hard to accomplish, since they need a "memory" of equally spaced points at every step during integration. For this reason a common technique is "halving-doubling," where an increase or decrease in

stepsize is half or double the current stepsize. Then increasing the stepsize presents no problem if a sufficient number of backpoints are retained. Decreasing is usually accomplished by interpolation or by a single-step method.

OPTIMIZATION OF STEPSIZE

Restricting the variations of stepsize h by a constant factor does not generally yield "optimal" stepsizes;* hence the initial choice of interval could have a substantial effect on the total number of integration steps. For example, suppose the halving-doubling method is being used and the order is held fixed. If for some n, $R_n > T_1$, the stepsize must be decreased. Let h_{opt} be the optimum (largest) value of h for which Equation 8 is satisfied. It might then happen that $h/2 < h_{opt} < h$. But in this method, $h/2$ would be used, although this would necessitate more integration steps than if the optimum stepsize were used. Hence a variation of h that would better approximate its optimum value is desirable. One technique available (see Reference 7 for details and applications in the case of single-step methods) is to compute the stepsize using the local-error estimate, i.e., letting σ be the "allowable" local error for each step, where $T_2 \leq \sigma \leq T_1$. Variations in h can then be computed from the relationship between σ and $R_n \cong Ch^{p+2} x^{(p+2)}(\xi)$:

$$h_{opt} \cong \left[\frac{\sigma}{Cx^{(p+2)}(\xi)}\right]^{1/(p+2)} \cong \left[\frac{\sigma h^{p+2}}{R_n}\right]^{1/(p+2)} , \tag{10}$$

so that, if $\sigma < R_n$, the stepsize is decreased; if $R_n < \sigma$ the stepsize is increased; the variation being approximately optimal with respect to the choice of σ.

OPTIMIZATION OF ORDER

Varying the order by a constant factor, like varying the stepsize, need not yield the "optimal" order.† For example, suppose that a variable-order, constant-step control is being used and the order is varied by ±1. If, for some n, $|R_n| < T_2$, the order must be decreased. Let p_{opt} be the optimum (smallest) value of p for which Equation 8 is satisfied. Then the following situation could exist;

$$p_{opt} < p - 1 < p .$$

In this method, $p - 1$ would be used, although this would result in more calculations than are required at the optimum order. Also, using a variable-step control in addition to varying the order may cause:

$$p_{opt} < L_2 \leq p - 1 ,$$

*The largest stepsize that allows a prescribed local error at a given point.
†The smallest order that allows a prescribed local error.

in which case the stepsize would be increased if p_{opt} were used. One method to obtain the optimum order is to test the local error for various orders until the optimum is established. For example, if the local error estimate used is given by Equation 6, then $Ch^2 \nabla^p \ddot{x}_n$ can be tested for various p until the smallest value satisfying Equation 8 is found.

GENERAL EFFICIENCY CONSIDERATIONS

This section discusses some possible effects of the above mentioned controls on the efficiency of any integration, which will generally depend on attaining the following objectives:

A. Minimizing the number of integration steps,

B. Minimizing the number of corrector iterations required at each step,

C. Minimizing other computational efforts required by the integration formulas being used, such as retaining only the significant terms in Equations 4 or 5 during integration,

D. Minimizing the effort and time involved in computations to control the local error, such as producing the required "memory" when changing the stepsize.

Three types of control are now considered.

1. A variable-order, fixed-step control:

Since varying the order involves only varying the number of terms retained, this control has little effect on attaining A. Equation 7 however, expressing the relationship between the local error and the "predicted-corrected" difference, indicates that controlling the local error may substantially affect the total number of corrector iterations. In particular, an increase in order at some point during integration where the local error is increasing could minimize any increase in the number of required iterations. This control also has an obvious effect on attaining C, and also that given a stepsize, the order required to satisfy a given criterion on the local error is automatically selected.

2. A fixed-order, variable-step control:

This control affects the number of corrector iterations for the same reason as control 1. It evidently affects the attainment of A or D, so one of them must be sacrificed. The gain in efficiency from larger stepsizes must be weighed against the cost in changing the stepsize. Furthermore, there are the two methods for varying the stepsize: halving-doubling, and step computation using the local-error estimate. In both cases the necessary "memory" could be obtained by interpolation, but in the second method all the backpoints must be computed when decreasing and when increasing the stepsize, as compared with opposed half, or none, for the first method. However, the second method may be the more effective of the two in minimizing the number of integration steps. This

control has the further advantage that, given an order, it automatically selects the stepsize (possibly optimum) required to satisfy a given local error criterion.

3. A variable-order, variable-step control:

The control combines the effects of controls 1 and 2. It has one possible disadvantage: in allowing the order to vary before changing the step, it may cause smaller stepsizes, thus opposing objective A. But this control has one clear advantage: it automatically selects both the stepsize and order required to satisfy a given local error criterion.

EFFECTIVE ERROR CONTROL IN ORBITAL INTEGRATION

In general, the effectiveness of a local error control during the integration of a particular orbit depends on the degree and rate of change of the derivative $\ddot{x}^{(p+2)}(\xi)$ (and hence the local error), during a revolution for a fixed p and h. This change is governed by the orbital parameters a (the semi-major axis) and e (the eccentricity) (Reference 2). In particular, for orbits with eccentricity near zero, the local error variation will probably be small over a revolution, and as the eccentricity becomes bounded away from zero, this variation becomes more pronounced. Moreover, the larger the semi-major axis for a particular satellite, the slower the rate of change of the derivative.

Suppose that during the computation of a particular orbit, some local error criterion is to be satisfied over the entire range of integration, which may involve many revolutions. That is, suppose that the local error is to be bounded from above, so as to restrict (at least qualitatively) the propagation of truncation error. Suppose also that the integration method used is of some fixed order (as is generally the case when formulas such as Equations 4' and 5' are used), and that the stepsize is to be specified. The error criterion could then be satisfied, without serious loss of efficiency, by integration over a single revolution with various stepsizes—choosing the largest stepsize that satisfies the criterion over the whole revolution. For orbits with $e \cong 0$, this process could yield the optimal stepsize for the given order. On the other hand, for orbits with $e \geq \epsilon > 0$, the stepsize selected in this manner would mean bounding the local error at its maximum value (e.g., at perigee), and could cause needless computation where the local error is smaller or at its minimum (e.g., near apogee).

Consider the effects of a continuous variable-step control during integration. In both cases cited above (i.e., $e \cong 0$, or $e \geq \epsilon > 0$) an initial stepsize (possibly optimum) satisfying the local error criterion would be computed automatically; furthermore, the stepsize would vary according to variations in the derivative. The numerical results show that in both cases, under suitable conditions, such a control has a considerable effect on the efficiency. In particular, if a sufficiently large range (T_1, T_2) is used in Equation 8 (so that the local error is allowed to range between these two limits before any interval modification is performed), the gains in efficiency due to the larger stepsizes will overshadow any loss due to changing the stepsize.

If, in addition to the above, the order to be used could be specified (as it generally can when formulas such as Equations 4 or 5 are used), then it may be possible to obtain an optimal stepsize

(or stepsizes)—order combination by integrating over a single revolution with various orders and letting the stepsize vary during the revolution. In particular, we could find that order which—together with a continuous step modification—gives the most efficient integration. Examples of such a procedure will be given in the numerical results.

Finally, consider a continuous variable-order control. Again, since we are assuming that the number of evaluations of the derivative governs the overall efficiency, the effects of this control on C (in the previous section), and its possible adverse effects on A when used in conjunction with a variable step control, make this control ineffective from this point of view. If, however, a fixed-step method is used, the effect of this control on the number of corrector iterations can result in considerable efficiency gains.

NUMERICAL RESULTS

This section discusses the results obtained by applying the various local error controls during the integration of three selected orbit types. Appendix B formulates the actual equations of motion (Equation 1); a description of the computer program can be found in Reference 8.

For the sake of simplicity, instead of an approximation for R_n (given by Equation 6) in controlling the error, a bound on the local error given by

$$U_n = Ch^2 \nabla^{k-2} \ddot{x}_n$$

was used, where ∇^{k-2} is the last backward difference retained in the computations (see Appendix A). Since $p = k + 1$, therefore

$$|R_n| \leq |U_n|$$

Thus using U_n gives the same qualitative results as using R_n, and the same quantitative results could be obtained by using R_n with a smaller T_1.

The number of derivative evaluations for each integration was obtained by multiplying the total number of steps taken by the average number of predictor-corrector iterations.

The error estimates were obtained by integrating Equation 1 with $P = 0$, and comparing the results with the Kepler solution. Since the perturbation function P generally has only a small effect on the accumulation of error, the error estimated in this manner can be considered close to the actual error.

All computations were performed on the Univac 1108 computer in double precision. The following units were used:

Unit of length = 6378.388 km,
Unit of time = 13.447 min.

Table 1 gives the results obtained by integrating the orbits with formulas of orders 7, 9, 11, and 13. These integrations were carried out with various stepsizes, where in each case the stepsize was increased until U_n failed to satisfy the inequality

$$|U_n| \leq T_1 \tag{12}$$

over the entire range of integration (chosen arbitrarily as 4000 minutes). The table gives only the last 2 stepsizes tested, indicating the largest stepsize that passed the criterion, and the first stepsize failed.

Table 1

Fixed-Order, Fixed-Step Control.

$T_1 = 0.5 \times 10^{-8}$ $\delta = 0.1 \times 10^{-10}$ (predictor-corrector tolerance)

Semi-major Axis, a	Eccentricity e	Order	Stepsize (min)	No. of Derivative Evaluations	Estimated Error
6.7	0.003	7	5.0	798	4×10^{-10}
			7.0*	569	4×10^{-9}
		9	15.0	262	3×10^{-9}
			17.0*	231	8×10^{-9}
		11	24.0	160	9×10^{-10}
			26.0*	147	2×10^{-9}
		13	22.0	173	3×10^{-12}
			24.0*	158	9×10^{-11}
1.15	0.075	7	0.4	9998	8×10^{-8}
			0.5*	7998	4×10^{-7}
		9	0.9	4440	2×10^{-7}
			1.0*	3996	5×10^{-7}
		11	1.2	3327	5×10^{-8}
			1.3*	3070	1×10^{-7}
		13	1.5	3081	1×10^{-9}
			1.6*	3043	1×10^{-8}
8.5	0.878	7	0.30	13344	1×10^{-7}
			0.35*	11483	4×10^{-7}
		9	0.40	10005	4×10^{-8}
			0.45*	8902	1×10^{-7}
		11	0.30	13340	9×10^{-11}
			0.35*	11433	4×10^{-10}
		13	0.20	19992	4×10^{-12}
			0.25*	15992	3×10^{-12}

*Integration with this stepsize failed the criterion (12) for some n.

The results in Table 1 suggest the following observations:

For a given local error criterion, increasing the order does not generally imply a larger "allowable" stepsize. Since the propagation of error (both truncation and round-off) influences the "smoothness" of the higher-order differences (and hence our local error estimate), this behavior is to be expected, both from the inaccuracies in the differences and from an unstable p and h combination. Also, accuracy increases with the higher orders. Reference 6 includes a table demonstrating this behavior (for the case $e \cong 0$).

The stepsizes resulting from the local error criterion during the integrations correspond to bounding the local error at its maximum (at perigee), although the local error for most of the integration, particularly for the case $e = 0.87$, was considerably smaller than the upper bound.

Table 2 gives the results obtained by integrating the test orbits with a variable-step control. In particular, the stepsize was varied during each integration, forcing Un to satisfy

$$T_1 \leq |Un| \leq T_2 \tag{13}$$

for all n. The stepsize-modification techniques used were halving-doubling and the optimum-step computation where the "allowable" local error is designated by σ; i.e., the step computation is given by

$$h_{opt} = \left[\sigma h^{p+2}/Un\right]^{1/p+2} .$$

Except for those cases indicated by an asterisk, all integrations were performed with an initial stepsize (chosen arbitrarily) of 0.42 min = $1/2^5$ internal units. The sixth column indicates the stepsize or stepsize range that occurred during the multistep integration as a result of the local error control.

Comparing the corresponding results in Tables 1 and 2 leads to the following observations:

For orbits of low eccentricity, (0.003, 0.075), the stepsizes selected in Table 2 as a result of the variable-step control are comparable with the "optimum" stepsizes found by trial and error in Table 1.

Table 2 indicates by asterisks those cases in which the initial step was selected as a result of foreknowledge of the "optimum" step (the initial step used was exactly half the stepsize indicated). These cases demonstrate the dependence of the efficiency on the initial choice of stepsize in the halving-doubling type of step modification.

For orbits of high eccentricity (0.87), the larger stepsizes effected significant gains in efficiency Stepsize modification causes no loss in efficiency, as shown by the results concerning the degree of control in Table 2. The apparent loss in accuracy, (especially for the higher orders) was expected

Table 2

Fixed-Order, Variable-Step Control.

$T_1 = 0.5 \times 10^{-8}$ $T_2 = 0.5 \times 10^{-13}$ $\delta = 0.1 \times 10^{-10}$

Semi-major Axis, a	Eccentricity e	Order p	Type of Step Modification	σ	Stepsize (Range) (min)	No. of Derivative Evaluations	Estimated Error
6.7	0.003	7	H/D		1.7	2377	2×10^{-13}
			OPT	1×10^{-10}	5.4	744	6×10^{-10}
			OPT	1×10^{-11}	3.0	1320	1×10^{-11}
		9	H/D		6.7	590	2×10^{-12}
			H/D*		14.0	281	1×10^{-9}
			OPT	1×10^{-10}	12.4	322	5×10^{-10}
			OPT	1×10^{-11}	8.4	471	1×10^{-11}
		11	H/D		13.4	291	2×10^{-12}
			H/D*		23.0	167	6×10^{-10}
			OPT	1×10^{-10}	20.1	196	1×10^{-10}
			OPT	1×10^{-11}	15.1	260	7×10^{-12}
		13	H/D		13.4 - 26.9	174	6×10^{-9}
			H/D*		23.0	165	5×10^{-12}
			OPT	1×10^{-10}	19.8 - 30.0	155	8×10^{-9}
			OPT	1×10^{-11}	13.9 - 26.8	159	7×10^{-9}
1.15	0.075	7	H/D		0.42	9516	1×10^{-7}
			OPT	1×10^{-10}	0.42	9516	1×10^{-7}
			OPT	1×10^{-11}	0.42	9516	1×10^{-7}
		9	H/D		0.42	9514	2×10^{-10}
			OPT	1×10^{-10}	0.61	6514	6×10^{-9}
			OPT	1×10^{-11}	0.42	9490	2×10^{-10}
		11	H/D		0.84	4781	1×10^{-9}
			OPT	1×10^{-10}	1.04	3849	8×10^{-9}
			OPT	1×10^{-11}	0.78	5131	4×10^{-10}
		13	H/D		0.84 - 1.68	4257	4×10^{-7}
			OPT	1×10^{-10}	1.20	3314	8×10^{-10}
			OPT	1×10^{-11}	0.96	4171	5×10^{-10}
8.5	0.87	7	H/D		0.2 - 3.4	3180	7×10^{-9}
			OPT	1×10^{-10}	0.2 - 9.5	2415	2×10^{-8}
			OPT	1×10^{-11}	0.1 - 6.9	3131	3×10^{-9}
		9	H/D		0.2 - 6.7	1374	7×10^{-8}
			OPT	1×10^{-10}	0.2 - 19.0	1137	7×10^{-8}
			OPT	1×10^{-11}	0.2 - 14.6	1331	1×10^{-9}
		11	H/D		0.2 - 13.4	875	6×10^{-8}
			OPT	1×10^{-10}	0.2 - 29.1	788	2×10^{-8}
			OPT	1×10^{-11}	0.2 - 14.0	907	5×10^{-9}
		13	H/D		0.2 - 26.9	710	1×10^{-7}
			OPT	1×10^{-10}	0.2 - 22.4	661	3×10^{-8}
			OPT	1×10^{-11}	0.1 - 25.6	775	1×10^{-8}

H/D = Halving-doubling
OPT = Optimum-Step Computation

since the error control bounded the error from below as well as from above, whereas for the corresponding results in Table 1 the local error was insignificant for most of the integration.

In all cases, the optimum-step computation-type interval modification yielded fewer derivative evaluations than halving-doubling, with little or no loss of accuracy. The parameter magnitude affected the integration as expected: the small value increased the accuracy at the cost of smaller stepsizes.

In all cases, automatic selection of a stepsize to satisfy the local error criterion for the given order is clearly an advantage. Also, the "best" order (i.e., the one yielding the fewest derivative evaluations) to use with the variable-step control may not be the same as the best order to use for a fixed-order, fixed-step integration.

Table 3 gives the results obtained by integrating the test orbits with a variable-order, variable-step control. The order was allowed to vary between the limits L_1 and L_2 before any step modification was effected. In all cases, the "smallest" order satisfying Equation 12 in any given stepsize was used. All integrations were performed with an initial stepsize of 0.42 mins. and an initial order of $L_1 + (L_2 - L_1)/2$, forcing the initial order p to be in the interval (L_1, L_2). Columns 5 and 6 indicate the stepsize and order ranges that occurred during integration as a result of the error control.

Table 3

Variable-Order, Variable-Step Control.

$T_1 = 0.5 \times 10^{-8}$ $T_2 = 0.5 \times 10^{-13}$ $\delta = 0.1 \times 10^{-10}$ $\sigma = 0.1 \times 10^{-9}$

Semi-major Axis, a	Eccentricity e	$L_1 - L_2$	Type of Step Modification	Stepsize (Range) (min)	Order (Range)	No. of Derivative Evaluations	Estimated Error
6.7	0.003	7 - 13	H/D	3.3	8	1184	9×10^{-14}
			OPT	8.8	10	454	5×10^{-13}
		9 - 13	H/D	6.7	9	588	2×10^{-12}
			OPT	18.4	11	217	5×10^{-11}
		11 - 15	H/D	13.4	11	289	2×10^{-12}
			OPT	24.6	11 - 15	175	2×10^{-9}
1.15	0.075	7 - 13	H/D	0.42	8 - 10	9513	9×10^{-10}
			OPT	0.84	10	4767	5×10^{-9}
		9 - 13	H/D	0.42 - 0.84	9 - 10	4818	6×10^{-9}
			OPT	1.04	11	3849	8×10^{-9}
		11 - 15	H/D	0.84	12 - 13	4751	1×10^{-10}
			OPT	1.2	13	3314	8×10^{-10}
8.5	0.87	7 - 13	H/D	0.2 - 3.3	7 - 13	2744	4×10^{-9}
			OPT	0.2 - 10.8	7 - 13	2269	1×10^{-8}
		9 - 13	H/D	0.2 - 13.4	9 - 13	1002	6×10^{-8}
			OPT	0.2 - 13.2	9 - 13	1067	3×10^{-8}
		11 - 15	H/D	0.1 - 26.8	11 - 15	665	3×10^{-7}
			OPT	0.1 - 20.0	11 - 15	742	2×10^{-7}

Comparing Table 3 with Tables 1 and 2 suggests the following observations:

For orbits of low eccentricity, in Table 3, the stepsizes attained for the various orders selected were comparable to the stepsizes obtained for the corresponding orders in Table 2.

In all cases, (especially $e = 0.87$), allowing the order to vary before changing the stepsize resulted in a smaller "mean" stepsize and thus a larger number of derivative evaluations; so that if the "best" order to use with a variable step control is known, this control gives a more efficient integration than the variable-order, variable-step control.

In all cases, automatic selection of a stepsize and order to satisfy the given local error criterion is clearly an advantage.

Table 4 shows the effects of a variable-order control alone on the average number of predictor-corrector iterations (and hence the total number of derivative evaluations) and on the total error. The order or order range obtained as a result of the control is indicated.

Table 4

Variable-Order, Fixed-Step and Fixed-Order, Fixed-Step Control.

$T_1 = 0.5 \times 10^{-8}$ $T_2 = 0.5 \times 10^{-13}$ $\delta = 0.1 \times 10^{-10}$

Semi-major Axis, a	Eccentricity e	Mode	Stepsize	Order	No. of Steps	Average Number of p-c Iterations	Estimated Error
6.7	0.003	Vary order	25.0	11	154	1.00	1×10^{-9}
		Fixed order	25.0	7	158	1.80	4×10^{-6}
		Fixed order	25.0	9	156	1.00	2×10^{-7}
1.15	0.075	Vary order	1.2	11	3327	1.00	5×10^{-8}
		Fixed order	1.2	7	3331	1.94	1×10^{-5}
		Fixed order	1.2	9	3329	1.28	2×10^{-6}
8.5	0.87	Vary order	0.5	5 - 10	8000	1.00	2×10^{-8}
		Fixed order	0.5	7	7998	1.01	5×10^{-6}

These results show the advantages of automatic selection of order, from the predictor-corrector point of view and from the accumulated error that results from local error control.

Finally, consider the effects of varying the "allowable range" (T_1, T_2) in the local error control. In general, the smaller the interval (T_1, T_2) is made, the greater is the frequency of interval and/or order modification. It would seem (particularly in the case of a variable-step control), that efficiency gains due to the larger stepsizes due to the control could be offset by the cost in changing the stepsize—should such changes occur too frequently during integration.

Table 5 gives the results of integrating one of the test orbits with a fixed T_1 and various values for T_2 (approaching T_1 as a limiting value). A variable-step control (halving-doubling) was used, and the variations in the stepsize, along with the number of step changes and computation time (in

Table 5

Variable-Step Control—Range Modification

$T_1 = 0.5 \times 10^{-8}$ $\delta = 0.1 \times 10^{-10}$

Semi-major Axis, a	Eccentricity e	T_2	Stepsize (Range) (min)	No. of Derivative Evaluations	Estimated Error	Computation Time (min)
8.5	0.87	5×10^{-16}	0.42 - 6.7	1810	7×10^{-8}	0.094
		5×10^{-14}	0.42 - 6.7	1309	7×10^{-8}	0.073
		5×10^{-12}	0.42 - 13.4	886	6×10^{-8}	0.059
		1×10^{-11}	0.42 - 13.4	849	9×10^{-8}	0.058
		2×10^{-11}	0.42 - 13.4	832	1×10^{-7}	0.260
		2.5×10^{-11}	0.42 - 26.9	856	1×10^{-7}	0.810
		3×10^{-11}*	0.42 - 26.9	892	1×10^{-7}	2.10*

minutes) are presented for each integration. An asterisk indicates that integration in which a stepsize modification occurred approximately at each step, rendering the error control completely ineffective, since the entire computation time was governed by the backpoint computation.

These results show that as T_2 approaches T_1 there are efficiency gains at first, because of the larger stepsizes, but, as the range (T_1, T_2) becomes smaller, these gains are offset by the cost in changing the step. Finally, as expected, the diminishing range causes the control to be completely ineffective. Note, however, that the equations of motion (see Appendix B) are simpler than what may be used in actual practice, and in a more realistic situation, a smaller interval (T_1, T_2) may be possible before the computing time is completely governed by the frequency of step modification.

It may be admitted that only a particular model of the perturbation function P has been used. However, variations in this function (such as the inclusion of higher-order gravitational effects) should not affect the general qualitative behavior of the local error.

CONCLUSIONS

The problem of achieving an efficient and accurate orbital integration by a multistep process, using the concept of controlling the local error during integration, has been studied. It has been shown that during the integration, the parameters p and h can be used to control the local error in such a way that the efficiency of the process is improved with no significant loss in accuracy.

In particular, if a sufficiently large range (T_1, T_2) in the local error is allowed, and an order p is given, a variable-step control can automatically yield a good approximation of the optimal initial stepsize (with respect to the given order) and can significantly improve the efficiency of the process by varying this step during integration. Moreover, if a good approximation of the "best" order to use with a variable-step, fixed-order control is known, this control effects the most efficient integration, as compared with the other controls considered. On the other hand, a

variable-order, variable-step control automatically determines both the order and step required to satisfy the given local error criterion, and also gives a reasonably efficient integration. Finally, even a variable-order, fixed-step control, although it does not minimize the number of integration steps, can give a more efficient integration than no control at all, by minimizing the number of predictor-corrector iterations.

We have considered only a "local" optimization problem, in the sense that optimal stepsize is defined on a step-by-step basis as being the largest stepsize satisfying a given *local* error criterion at any given point. A more significant consideration would be optimal stepsize (and/or order) distribution over the entire range of integration, defined on the basis of a criterion imposed on the accumulated truncation error. D. Morrison, under some restrictive conditions (among others, limitation to a single differential equation), considers the problem of optimizing the mesh distribution (Reference 9). A basic difficulty in applying such techniques is the requirement that a "memory" of equally spaced points be available at each step during the integration. One solution worth considering is the integration of divided-difference interpolation polynomials (which do not require equally spaced points), for use in numerical integration.

Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, Maryland, January 22, 1968
311-07-21-01-51

REFERENCES

1. Henrici, P., "Discrete Variable Methods in Ordinary Differential Equations," New York: John Wiley & Sons, 1962.

2. Soar, W. S., "Local Error Control in Numerical Integration Through Optimizing the Order of the Integrator," NASA Goddard Space Flight Center Technical Memorandum X-55-140, 1964.

3. Hull, T., and Creemer, A., "Efficiency of Predictor-Corrector Procedures," *J. Assoc. for Computing Machinery,* 10:291-301, 1963.

4. Shanks, E., "Solutions of Differential Equations by Evaluations of Functions," *Math. of Comput.,* 20:21-38, 1966.

5. Ralston, A., "Relative Stability in the Numerical Solution of Ordinary Differential Equations," *SIAM Review,* 7:114-125, 1965.

6. Sheldon, J., Zondek, B., and Friedman, M., "On the Time Step to be Used for the Computation of Orbits by Numerical Integration," *Math Tables Aids Comput.,* 11:181-189, 1957.

7. Martin, W., Paulson, K., and Sashkin, L., "A General Method of Systematic Interval Computation for Numerical Integration of Initial Value Problems," *Comm. Association for Computing Machinery,* 9:754-757, 1966.

8. Cigarski, G. A., and Velez, C. E., "Cowell Orbit Generator Error and Time Analysis Program," NASA Goddard Space Flight Center Document X-553-67-491, August 1967.

9. Morrison, D., "Optimal Mesh Size in the Numerical Integration of an Ordinary Differential Equation," *J. A. C. M.*, 9:98-103, 1962.

10. Henrici, P., "The Propagation of Round-off Error in the Numerical Solution of Initial Value Problems Involving Ordinary Differential Equations of the Second Order," Symposium, Provisional International Computation Centre, 1960, 275-291.

Appendix A

Derivation of the "Summed" Form of Integration Formulas

By integrating Newton's backward difference interpolation polynomial, the following multistep formulas for the numerical integration of Equation 1 can be derived,* (including formulas for the velocity in the event that P contains $\dot{x}$):

Equations for $\ddot{x} \rightarrow x$

$$\text{Predictor: } x_{m+1} - 2x_m + x_{m-1} = h^2 \left[\sigma_0 \ddot{x}_m + \sigma_1 \nabla\ddot{x}_m + \sigma_2 \nabla^2 \ddot{x}_m + \sigma_3 \nabla^3 \ddot{x}_m + \cdots + \sigma_k \nabla^k \ddot{x}_m\right] , \tag{A1}$$

$$\text{Corrector: } x_{m+1} - 2x_m + x_{m-1} = h^2 \left[\sigma_0^* \ddot{x}_{m+1} + \sigma_1^* \nabla\ddot{x}_{m+1} + \sigma_2^* \nabla^2 \ddot{x}_{m+1} + \cdots + \sigma_k^* \nabla^k \ddot{x}_{m+1}\right] , \tag{A2}$$

Equations for $\ddot{x} \rightarrow \dot{x}$

$$\text{Predictor: } \dot{x}_{m+1} - \dot{x}_m = h\left[\gamma_0 \ddot{x}_m + \gamma_1 \nabla\ddot{x}_m + \cdots + \gamma_k \nabla^k \ddot{x}_m\right] , \tag{A3}$$

$$\text{Corrector: } \dot{x}_{m+1} - \dot{x}_m = h\left[\gamma_0^* \ddot{x}_{m+1} + \gamma_1^* \nabla\ddot{x}_{m+1} + \cdots + \gamma_k^* \nabla^k \ddot{x}_{m+1}\right] , \tag{A4}$$

where the coefficients σ, σ^*, γ, γ^* are given by the following recurrence relationships: letting

$$h_m = \sum_{j=1}^{m} \frac{1}{j} ,$$

$$\sigma_0 = 1 , \qquad \sigma_0^* = 1 , \qquad \gamma_0 = 1 , \qquad \gamma_0^* = 1 ,$$

$$\sigma_m = 1 - \sum_{j=1}^{m} \frac{2h_{j+1}}{j+2} \sigma_{m-j}$$

$$\sigma_m^* = - \sum_{j=1}^{m} \frac{2h_{j+1}}{j+2} \sigma_{m-j}^*$$

*Reference 1, pp. 192-195 and 291-293.

$$\gamma_m = 1 - \sum_{j=1}^{m} \frac{\gamma_{m-j}}{j+1}$$

$$\gamma_m^* = - \sum_{j=1}^{m} \frac{\alpha_{m-j}^*}{j+1}$$

$$m = 1, 2, 3, \cdots, k ,$$

and

$$\nabla \ddot{x}_m = \ddot{x}_m - \ddot{x}_{m-1} ,$$

$$\nabla^2 \ddot{x}_m = \nabla\left(\nabla \ddot{x}_m\right) = \nabla\left(\ddot{x}_m - \ddot{x}_{m-1}\right) = \ddot{x}_m - 2\ddot{x}_{m-1} + \ddot{x}_{m-2} ,$$

$$\nabla^k \ddot{x}_m = \nabla\left(\nabla^{k-1} \ddot{x}_m\right) .$$

It has been established (References 1 and 10) that an algebraic equivalent of Equations A1 through A4, known as the "summed" form of the integration formulas, considerably reduces the propagation of round-off error. Formally, one can obtain this equation modification by applying the inverse difference operators ∇^{-1}, ∇^{-2} defined by $\nabla^{-1}\nabla = I$, $\nabla^{-2}\nabla^2 = I$, (I the identity), to both sides of these equations. In particular, applying ∇^{-2} to both sides of Equations A1 and A2 gives

$$\text{Predictor: } x_{m+1} = h^2\left[\sigma_0 \nabla^{-2} \ddot{x}_m + \sigma_1 \nabla^{-1} \ddot{x}_m + \sigma_2 \ddot{x}_m + \sigma_3 \nabla \ddot{x}_m + \cdots + \sigma_k \nabla^{k-2} \ddot{x}_m\right] ,$$

$$\text{Corrector: } x_{m+1} = h^2\left[\sigma_0^* \nabla^{-2} \ddot{x}_{m+1} + \sigma_1^* \nabla^{-1} \ddot{x}_{m+1} + \sigma_2^* \ddot{x}_{m-1} + \sigma_3^* \nabla \ddot{x}_{m+1} + \cdots + \sigma_k^* \nabla^{k-2} \ddot{x}_{m+1}\right] .$$

Defining ${}^{I}S_m, {}^{II}S_m$ by

$${}^{I}S_m = \nabla^{-1} \ddot{x}_m$$

$${}^{II}S_m = \nabla^{-1}\left({}^{I}S_m\right) = \nabla^{-2} \ddot{x}_m$$

gives

$$\text{Predictor: } x_{m+1} = h^2\left[\sigma_0 \, {}^{II}S_m + \sigma_1 \, {}^{I}S_m + \sigma_2 \ddot{x}_m + \cdots + \sigma_k \nabla^{k-2} \ddot{x}_m\right] , \qquad \text{(A5)}$$

and since

$$ {}^{I}S_{m+1} = \ddot{x}_{m+1} + {}^{I}S_{m} , $$

and

$$ {}^{II}S_{m+1} = {}^{II}S_{m} + {}^{I}S_{m} + \ddot{x}_{m+1} , $$

therefore

$$ \text{Corrector: } x_{m+1} = h^2 \left[\sigma_0^* \, {}^{II}S_m + \left(\sigma_0^* + \sigma_1^* \right) {}^{I}S_m + \left(\sigma_0^* + \sigma_1^* + \sigma_2^* \right) \ddot{x}_{m+1} + \sigma_3^* \nabla \ddot{x}_{m+1} + \cdots + \sigma_k^* \nabla^{k-2} \ddot{x}_{m+1} \right] . \quad \text{(A6)} $$

Likewise for Equations A3 and A4, applying the operator ∇^{-1} to both sides gives

$$ \text{Predictor: } \dot{x}_{m+1} = h \left[\gamma_0 \nabla^{-1} \ddot{x}_m + \gamma_1 \ddot{x}_m + \gamma_2 \nabla \ddot{x}_m + \cdots + \gamma_k \nabla^{k-1} \ddot{x}_m \right] , $$

$$ \text{Corrector: } \dot{x}_{m+1} = h \left[\gamma_0^* \nabla^{-1} \ddot{x}_{m+1} + \gamma_1^* \ddot{x}_{m+1} + \gamma_2^* \nabla \ddot{x}_{m+1} + \cdots + \gamma_k^* \nabla^{k-1} \ddot{x}_{m+1} \right] . $$

Using the definition of ${}^{I}S_m$ as above gives

$$ \text{Predictor: } \dot{x}_{m+1} = h \left[\gamma_0 \, {}^{I}S_m + \gamma_1 \ddot{x}_m + \gamma_2 \nabla \ddot{x}_m + \cdots + \gamma_k \nabla^{k-1} \ddot{x}_m \right] , \quad \text{(A7)} $$

$$ \text{Corrector: } \dot{x}_{m+1} = h \left[\gamma_0^* \, {}^{I}S_m + \left(\gamma_0^* + \gamma_1^* \right) \ddot{x}_{m+1} + \gamma_2^* \nabla \ddot{x}_{m+1} + \cdots + \gamma_k^* \nabla^{k-1} \ddot{x}_{m+1} \right] . \quad \text{(A8)} $$

These are the "summed" forms of the integration formulas. (For details concerning the computational usage of these formulas, see Reference 8.)

Note, again, that Equations A1, A2 and A5, A6 are algebraically equivalent, so that for any fixed k, any solution of A1, A2 is a solution of A5, A6, and conversely. In particular, the local truncation errors associated with these formulas are the same. A similar equivalence exists for Equations A3, A4, and A7, A8.

Appendix B

Equations of Motion

The equations of motion used to obtain the numerical results are given by

$$\ddot{\bar{x}} = -\frac{\mu\bar{x}}{R^3} + \bar{F}_1 + \bar{F}_2 ,$$

where $\bar{F}_1 + \bar{F}_2 = P$ (see Equation 1), and $\bar{x} = (x, y, z)$, $R = |\bar{x}| = \left(x^2 + y^2 + z^2\right)^{1/2}$, μ is a constant, $\bar{F}_1$ is the perturbation due to the non-sphericity of the earth, and $\bar{F}_2$ is the perturbation due to drag. $\bar{F}_1 = \left(F_{1x}, F_{1y}, F_{1z}\right)$ is given by

$$F_{1x} = \frac{\mu x}{R^6}\left[JR^2(5s-1) + Hz(7s-3) + \frac{K}{6}\left(-63s^2 + 42s - 3\right)\right] ,$$

$$F_{1y} = \frac{\mu y}{R^6}\left[JR^2(5s-1) + Hz(7s-3) + \frac{K}{6}\left(-63s^2 + 42s - 3\right)\right] ,$$

$$F_{1z} = \frac{\mu z}{R^6}\left[JR^2(5s-1) + \frac{K}{6}\left(-63s^2 + 70s - 15\right)\right] + \frac{\mu H}{5R^5}\left(35s^2 - 30s + 3\right) ,$$

where $s = (z/R)^2$; and J, H, and K are the second, third, and fourth harmonics of the earth's potential field, respectively. $\bar{F}_2$ is of the form

$$\frac{C_D A}{2M}\rho|\bar{V}_r|\bar{V}_r ,$$

where V_r is the relative velocity of the satellite; ρ is the atmospheric density at the satellite position; and A, M and C_D are the cross-sectional area, mass, and drag coefficient of the satellite respectively. The actual numerical values of these constants that were used to obtain the results can be found in Reference 8.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546
OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546