THE USE OF INTEGRALS IN NUMERICAL INTEGRATIONS OF THE N-BODY PROBLEM

PAUL E. NACOZY

The University of Texas at Austin

Abstract. The numerical integration of systems of differential equations that possess integrals is often approached by using the integrals to reduce the number of degrees of freedom or by using the integrals as a partial check on the resulting solution, retaining the original number of degrees of freedom.

Another use of the integrals is presented here. If the integrals have not been used to reduce the system, the solution of a numerical integration may be constrained to remain on the integral surfaces by a method that applies corrections to the solution at each integration step. The corrections are determined by using linearized forms of the integrals in a least-squares procedure.

The results of an application of the method to numerical integrations of a gravitational system of 25-bodies are given. It is shown that by using the method to satisfy exactly the integrals of energy, angular momentum, and center of mass, a solution is obtained that is more accurate while using less time of calculation than if the integrals are not satisfied exactly. The relative accuracy is ascertained by forward and backward integrations of both the corrected and uncorrected solutions and by comparison with more accurate integrations using reduced step-sizes.

1. Introduction

This paper presents a method to efficiently utilize the integrals in the numerical integration of gravitational systems. The method yields solutions of a higher accuracy while using less time of calculation than conventional procedures of numerical integration that do not use the integrals directly.

A gravitational system of n-bodies that has p integrals may be described uniquely by $(6n-p)$ position and velocity variables in the phase space. The integrals, such as energy, angular momentum, or center of mass, may be regarded as conditions of constraint imposed upon the $6n$ variables. The p integrals constrain the variables of a solution to remain on the intersection of p hypersurfaces, each of $(6n-1)$ dimensions. The intersection is a hypersurface of $(6n-p)$ dimensions.

It is common practice to integrate numerically the full, $6n$th order system of the equations of motion and to employ the integrals only as partial checks on the accuracy of the calculations. But the errors indicated by the integrals are often somewhat misleading. Since the satisfaction of the integrals is only necessary but not sufficient to guarantee accuracy of the solution, a computed solution of a gravitational system frequently has a larger error in the solution than in the integrals.

Moreover, if the error introduced by not satisfying the integrals remains in the solution during a process of numerical integration, and if the system is unstable, the solution with the error will diverge from a solution without the error. Since gravitational systems are very often unstable in the Liapunov sense, the small errors introduced by not satisfying the integrals will often become unbounded with time. Deeper and more

Astrophysics and Space Science **14** (1971) 40–51. *All Rights Reserved*

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

extensive discourses on this concept have been given by Miller (1964) and Szebehely (1968).

The question is whether or not the additional time of calculation necessary to satisfy the integrals is worth the resulting increase in accuracy. That is, could the increased accuracy be obtained in less time of calculation by merely decreasing the truncation error of the numerical integration and not satisfying the integrals exactly?

The answer appears to depend upon which method is used to satisfy the integrals. The integrals may be used to reduce the order of the system to $(6n-p)$, but the integration of the reduced system is often expensive in calculation time. The resulting equations of motion of the reduced system may be much more complex than the original system of order $6n$ due, for example, to the non-linearity introduced by the integral of energy. Also, the equations of order $(6n-p)$ may have lost the symmetry of the original system.

In this paper, it will be shown that the integrals may be satisfied without introducing additional complexity nor losing symmetry. The method introduces constraints on the solution of a numerical integration of the full system of order $6n$. During the computations, corrections are calculated and applied to the $6n$ variables to satisfy the integrals. The corrections are determined by the method of least-squares such that the sum of the squares of the corrections is minimized. The corrections generally are small and hence the integrals may be linearized, eliminating much of the complexity and reducing significantly the calculation time. The corrections, determined in this manner, modify those variables that are most in error so as to greatly increase the effectiveness of the method. This point will be discussed later.

The idea of using corrections of least-squares to satisfy the integrals has a geometrical interpretation. During an integration, errors in the calculation may cause the solution to leave the hypersurface of $(6n-p)$ dimensions defined by the integrals. The least-squares corrections to the $6n$ variables return the solution to the surface *along the normal* to the surface. By continually correcting the variables, the solution remains on the original integral hypersurface during the numerical integration.

Constraining the computed solutions of gravitational systems of n-bodies to remain on the proper integral surfaces has been performed previously by Aarseth (1966) and Miller (presented elsewhere in this volume). Aarseth corrects the integrals, the positions, and the velocities of the computed solution to account for the removal of escaping bodies from the system. Miller compares a corrected solution of the system with a similar, but uncorrected, solution. He finds that the two solutions diverge from each other – indicating the instability of the gravitational system. Neither of the two studies proposes to satisfy the integrals in order to produce a more accurate and efficient integration procedure.

In this paper, the method of satisfying the integrals will be derived and discussed, and its application outlined. The results of applying the method to several dynamical systems are presented. It is shown that solutions of gravitational systems that are corrected by the method are considerably more accurate and require less time of calculation than uncorrected solutions.

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

2. The Equations of Constraint

The equations that constrain the solution to remain on the original integral hypersurface are derived by the use of Lagrangian multipliers. A presentation of the method of Lagrangian multipliers with an excellent motivation for the present discussion is given by Lanczos (1949). A geometrical approach is given by Forsyth (1930).

To find extrema of a function of two variables, $f(x, y)$, subject to the constraining condition

$$g(x, y) = \text{constant}, \tag{1}$$

the two equations

$$\frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0, \quad \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0, \tag{2}$$

are to be solved with Equation (1), to determine the quantities x, y, and λ. Here, λ is the Lagrangian multiplier.

For a dynamical system with two degrees of freedom, let $x = [x_1, x_2, x_3, x_4]$ be the state vector in the phase space, where x_1 and x_2 are the coordinates and x_3 and x_4 are the corresponding velocity components. Let

$$g(x) = 0, \tag{3}$$

be an integral of the system. Equation (3) defines a hypersurface of three dimensions imbedded in the phase space of four dimensions.

During a process of numerical integration of the system, a computed solution is obtained at time t:

$$\eta = \eta(t) = [\eta_1, \eta_2, \eta_3, \eta_4],$$

where η_1 and η_2 are the computed position components and η_3 and η_4 are the computed velocity components. Due to errors in the computational procedure, the integral may not be satisfied exactly but

$$g(\eta) = \varepsilon, \tag{4}$$

where ε is a small quantity. The solution has left the integral surface defined by Equation (3) and is on the surface defined by Equation (4). It is desired to make corrections $\Delta\eta = [\Delta\eta_1, \Delta\eta_2, \Delta\eta_3, \Delta\eta_4]$ to the computed vector η to obtain the vector

$$x = \eta + \Delta\eta,$$

such that

$$g(x) = 0. \tag{5}$$

The square of the magnitude of the correction vector $\Delta\eta$ may be written as

$$f(\Delta\eta) = \sum_{i=1}^{4} (\Delta\eta_i)^2. \tag{6}$$

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

The corrections are uniquely chosen so that the function f of Equation (6) is minimized, subject to the constraint of Equation (5). The solution may be obtained by extending Equations (2) to four dimensions, yielding

$$\Delta\eta_i - \lambda \frac{\partial g}{\partial \eta_i} = 0, \quad i = 1, 2, 3, 4. \tag{7}$$

Equations (5) and (7) may be solved for the five unknowns λ, and $\Delta\eta_i$, $i = 1, 2, 3$, and 4. Unless the integral given by Equation (5) is a simple function of the variables (for instance linear), the solution of the system may be complex (or perhaps not obtainable). This is the case when the integral is the integral of energy of a gravitational system. The solution may be simplified by an expansion of the integral in powers of the corrections. The expansion becomes

$$g(x) = g(\eta) + \sum_{i=1}^{4} \frac{\partial g}{\partial \eta_i} \Delta\eta_i + \cdots. \tag{8}$$

Since the errors of the computation and hence the necessary corrections, $\Delta\eta_i$, are generally small, second and higher-order terms may be neglected.

Solving Equations (7) and (8) for the corrections $\Delta\eta_i$, with $g(x) = 0$ and $g(\eta) = \varepsilon$, yields

$$\Delta\eta_i = \frac{-\varepsilon \dfrac{\partial g}{\partial \eta_i}}{\sum_{j=1}^{4} \left(\dfrac{\partial g}{\partial \eta_j}\right)^2}, \quad i = 1, 2, 3, 4. \tag{9}$$

The correction vector $\Delta\eta$ is added to the computed state vector η to obtain a new state vector x which satisfies the integral $g(x) = 0$, with an error of order $|\Delta\eta|^2$. Geometrically, minimizing Equation (6) subject to Equation (8) causes the vector $\Delta\eta$ of Equation (9) to be normal to a three-dimensional plane which is approximately tangent to the surface $g = 0$ at the point x. The equation of the plane is given by Equation (8), neglecting the second and higher-order terms.

The result of Equation (9) may be generalized to a dynamical system of order $6n$ having p integrals. Denote the state vector of the system by x, where x is a column vector in the phase space with components x_j, $j = 1, 2, \ldots, 6n$. Denote the configuration or position vector of the system by R and the velocity vector by V. The state vector x may be written as

$$x = \begin{pmatrix} x_1 \\ x_3 \\ \vdots \\ x_{6n} \end{pmatrix} = \begin{pmatrix} R \\ V \end{pmatrix}. \tag{10}$$

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

The equations of motion of the system are

$$\dot{x} = \begin{pmatrix} \dot{R} \\ \dot{V} \end{pmatrix} = \begin{pmatrix} V \\ F \end{pmatrix}. \tag{11}$$

where the vector F is, in general, a function of the vector x and the time and will be defined explicitly later. The p integrals of the system may be written as

$$e_j(x) = 0, \quad j = 1, 2, \ldots, p. \tag{12}$$

The functions $e_j(x)$ are the components of a column vector E, so that $E(x) = 0$, where 0 is a null vector.

Equation (11) may be solved by numerical integration yielding a computed solution with a column state vector η at time t. The partial derivatives of the integrals of Equation (12) with respect to the components of the computed state vector η are the elements of a matrix E', having p rows and $6n$ columns. That is,

$$E' = \begin{bmatrix} \frac{\partial e_1}{\partial \eta_1} & \frac{\partial e_1}{\partial \eta_2} & \cdots & \frac{\partial e_1}{\partial \eta_{6n}} \\ \frac{\partial e_2}{\partial \eta_1} & & & \vdots \\ \vdots & & & \vdots \\ \frac{\partial e_p}{\partial \eta_1} & \cdots & & \frac{\partial e_p}{\partial \eta_{6n}} \end{bmatrix}.$$

At time t, due to errors in the computation, some or all of the p components of the vector E may be nonzero. That is,

$$E(\eta) = \varepsilon \neq 0,$$

where ε is an error vector whose elements are small quantities.

It is desired to compute a correction vector $\Delta\eta$ so that the vector

$$x = \eta + \Delta\eta,$$

will satisfy the equation

$$E(x) = 0.$$

The vector $\Delta\eta$ is chosen so that the quantity

$$\Delta\eta^T W \Delta\eta \tag{13}$$

is minimized. Here, W is a weighting matrix and the T superscript indicates matrix transpose.

As in Equation (8), each element of the vector E is expanded in powers of the vector $\Delta\eta$. The expansion becomes

$$E(x) = E(\eta) + E'\Delta\eta + \cdots.$$

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

Neglecting second and higher-order terms, with $E(x)=0$ and $E(\eta)=\varepsilon$, the expansion reduces to:

$$0 = \varepsilon + E'\Delta\eta . \tag{14}$$

Extending the solution given by Equation (7) yields

$$W\Delta\eta - E'^T\lambda = 0 , \tag{15}$$

where λ is a column vector whose p components are the Lagrangian multipliers. Equations (14) and (15) are $(6n+p)$ equations to be solved for $(6n+p)$ unknowns: the components of the two vectors $\Delta\eta$ and λ. Solving Equation (15) for $\Delta\eta$ and substituting the result into Equation (14) gives

$$\varepsilon + E'W^{-1}E'^T\lambda = 0 .$$

Solving for λ and substituting the result into Equation (15), the solution for the correction vector $\Delta\eta$, is

$$\Delta\eta = - W^{-1}E'^T(E'W^{-1}E'^T)^{-1}\varepsilon . \tag{16}$$

The matrix $(E'\ W^{-1}\ E'^T)$ is a $p \times p$, symmetrix matrix. If the matrix E' has rank p, the matrix $(E'\ W^{-1}\ E'^T)$ is positive definite and non-singular.

For gravitational systems, the vector F of Equation (11) is given by

$$F = \nabla_R U , \tag{17}$$

where $\nabla_R U$ denotes a column vector whose $3n$ components are $\partial U/\partial x_i$, $i=1, 2, \ldots, 3n$. The quantity x_i is a component of the vector R defined earlier. The function U is the negative of the potential energy of the system and is defined as

$$U = k^2 \sum_{1 \leqq i \leqq j \leqq n} \frac{m_i m_j}{|r_{ij}|} ,$$

where k is the Gaussian gravitational constant, m_i is the mass of the ith body and,

$$|r_{ij}| = \left(\sum_{p=0}^{2} (x_{3i-p} - x_{3j-p})^2 \right)^{1/2} .$$

The energy integral of the system may be written in terms of U and the state vector x as

$$\tfrac{1}{2} \sum_{i=1}^{n} m_i \sum_{k=0}^{2} (x_{3n+3i-k})^2 - U - C = 0 , \tag{18}$$

where C is the value of the energy for a set of initial conditions. Denote the energy integral as $e_1(x)=0$; the angular momentum integrals as $e_j(x)=0, j=2, 3, 4$; and the center of mass integrals as $e_j(x)=0, j=5, 6, \ldots, 10$. The functions $e_j(x)$ are the components of the column vector E, where

$$E(x) = 0 .$$

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

A numerical integration of the system of Equation (11) with F defined by Equation (17), yields the solution vector η at time t. The errors in the computation may cause some or all of the components of E to be nonzero:

$$E(\eta) = \varepsilon,$$

where the components of the vector ε are the errors of the corresponding integrals. The correction vector $\Delta\eta$ may be calculated by using Equation (16), so that $E(\eta+\Delta\eta)=0$. For the calculation, the quantities ε, E' and W^{-1} are needed.

During many procedures of numerical integration, the errors of the integrals, ε, are computed at various times as a check on the accuracy of the calculation. Since the correction vector $\Delta\eta$ is calculated and applied only at various times, as will be discussed later, little extra calculation is required to obtain ε. Moreover, in the calculation of the force vector F, the individual terms of the potential energy may be calculated as an intermediate step. One may save these calculated terms and compute the energy with a minimum of added effort.

The quantity E' is the partial derivative of E with respect to x and is easily computed. The partial derivative of e_1 with respect to V is equal to V multiplied by a mass. The partial derivative of e_1 with respect to R is simply minus the force F, pre-computed during the integration step. The partial derivative of e_j with respect to x is linear in x for $j=2, 3, 4$, and is equal to a mass or to zero for $j=5, 6, \ldots, 10$.

The quantity W is a diagonal weighting matrix to be included in the least-squares solution of Equation (16). In most variable stepsize numerical integration techniques, an error vector is computed at each integration step. The elements of the error vector correspond to an estimated truncation error in the calculation of each of the elements of the state vector x. The elements of the error vector might be the last differences of a finite difference integration method or the differences between the predictor and corrector of an integration step. The error vector is placed along the diagonal of the matrix W^{-1} in Equation (16). The weighting matrix allows the solution to correct some state variables proportionately more if there is a larger truncation error associated with the calculation of those state variables than of other variables. The concept of corrections that are weighted by the estimated truncation error is due to Gottlieb (1970).

The matrix multiplications of Equation (16) may be grouped efficiently as follows. The multiplication of the quantity $W^{-1}E'^T$ is performed and stored in E'^T. The multiplication $E'E'^T$ is performed and the product inverted. The multiplication and inversion are simplified since $(E'W^{-1}E'^T)$ is symmetric. The inverted matrix is pre- and postmultiplied by E'^T and $-\varepsilon$, respectively, to form $\Delta\eta$, which is then added to the computed state vector.

Finally, one advantageous property of the equation of corrections (Equation (16)) should be noted. In the numerical integration of a gravitational system, the largest errors in the computations often arise in the coordinates and velocities of the bodies that are closest to one another – the binaries. This fact may easily be seen if one looks at the estimated truncation error vector during a numerical integration. Hence, one

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

would like to correct the coordinates and velocities of a binary or a triple system more than those of the other bodies. This is precisely the result achieved by Equation (16) (regardless of whether or not the weighting matrix is used). To verify this point, note that for the integral of energy, the top row of E' or $\partial e_1/\partial x$ contains the two vectors F and V. The largest elements of both of these vectors are those corresponding to the closest bodies. If one examines Equation (16), with only the energy integral present, he will notice that the state vectors of the bodies closest to one another are recipients of the largest corrections while the other bodies receive smaller corrections. A similar analysis for the other integrals yields the same result. In other words, the corrections to the state vector x are proportional to the partial derivatives of the integrals with respect to x. And the partials are larger the closer two bodies become.

3. Evaluation of the Method

The method presented here was applied to the numerical integration of several dynamical systems to determine its practical value. The systems considered were the harmonic oscillator, the gravitational system of two-bodies, and the gravitational system of 25-bodies.

The method was applied to the harmonic oscillator and to the system of two-bodies by the following procedure. Two sets of solutions were obtained by numerical integration with various initial conditions. One set of solutions did not utilize the integrals while the other set introduced corrections determined by the method. The corrections were applied to the state vector after each integration step. All of the integrations were compared with the true solutions of the systems to determine the relative accuracies of the uncorrected and corrected solutions. The solutions of the harmonic oscillator were obtained using a fourth-order Runga-Kutta integration routine with constant step size. Both uncorrected and corrected solutions of the harmonic oscillator used the same step-size and the same number of integration steps. The solutions of the system of two-bodies were obtained using a fourth-order, predicator-corrector integration routine with variable step-size. Both uncorrected and corrected solutions of the system of two-bodies were integrated simultaneously with the same step-sizes and with the same number of integration steps.

The application of the method to the harmonic oscillator in a phase space of two dimensions showed no noticeable differences in accuracy between the corrected and uncorrected solutions. The reason for this negative result will be discussed later.

The application of the method to the system of two-bodies in a phase space of four dimensions over a range of initial conditions showed a large difference in accuracy between corrected and uncorrected solutions. The corrected solutions were about three orders of magnitude more accurate than the uncorrected solutions. Some results are given in Table I.

In the table, two solutions are presented: one for eccentricity $e=0.1$ and the other for $e=0.6$. Both solutions have semi-major axes of $a=2.0$ and were integrated for a duration of 55 orbital periods. The column denoted by $|\Delta R|$ gives the magnitude of

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

TABLE I

Two-body system

Solution		$\lvert\Delta R\rvert$	$\lvert\Delta V\rvert$
$a = 2.0$ $e = 0.1$ $T = 55$ rev.	Uncorrected	2.2×10^{-2}	7.5×10^{-3}
	Corrected	3.1×10^{-5}	9.4×10^{-6}
Solution		$\lvert\Delta R\rvert$	$\lvert\Delta V\rvert$
$a = 2.0$ $e = 0.6$ $T = 55$ rev.	Uncorrected	2.4×10^{-1}	7.9×10^{-2}
	Corrected	1.4×10^{-4}	2.2×10^{-5}

the error in position and the column denoted by $\lvert\Delta V\rvert$ gives the magnitude of the error in velocity. The row denoted by 'Uncorrected' gives the errors at the final time for the numerical integration without corrections. The row denoted by 'Corrected' gives the errors at the final time for the numerical integration that performs corrections after each integration step to satisfy the integrals of energy and of angular momentum. The results of solutions with other initial conditions were similar to the results shown in Table I. The results of the application given in Table I indicate that numerical integrations of the gravitational system of two-bodies that identically satisfy the integrals are more accurate than integrations that do not. The demonstration of the overall accuracy and efficiency of the method by comparison of times of calculation as well as accuracy is given later in the application of the method to the system of 25-bodies.

The different results obtained for the harmonic oscillator and the system of two-bodies offers an explanation of when and why the method appears to be of value. Two points were mentioned in the introduction of this paper: (1) the errors in the integrals are generally less than the errors in the computed solution; and (2) if a dynamical system is unstable, the solution with an error will diverge from a solution without the error. With these points, the following conclusions may be given. The errors in the integral of the harmonic oscillator are small compared to the error in the state variables of the solution. Since the harmonic oscillator is a stable system, a solution with a small error will not diverge from a system without the error. This would explain the result indicating no difference between the corrected and the uncorrected solutions for the harmonic oscillator. The errors in the integrals of the system of two-bodies are also small relative to the state variables of the solution. But the system of two-bodies is unstable in the Liapunov sense and hence the system with the errors will diverge from the system without the errors. And, as seen in Table I, the corrected solution lies several orders of magnitude closer to the true solution than the uncorrected solution.

The method was applied to a gravitational system of 25-bodies using the standard initial conditions as given by Lecar (1968). A highly-accurate, uncorrected numerical integration of the system was performed at the outset. The integration technique employed methods of regularization and was developed by Szebehely and Bettis (described by them elsewhere in this volume). The truncation error of the integration was lowered to the limit of the computer capability. The system was integrated forward and backward in time, and the accuracy verified. This solution was taken as an accurate, standard solution with which other, less accurate solutions were compared.

The numerical integration routine that was used to evaluate the correction method presented here is a 7th-order, Runga-Kutta-Fehlberg, variable-step method (Fehlberg, 1966), applied to the problem of 25-bodies. Two sets of solutions were obtained with the numerical integration. First the system of order $6n$ was integrated without using the integrals. Then the system of order $6n$ was integrated and all or various combinations of the ten integrals of the system were satisfied. Various truncation error tolerances were allowed and all integrations were extended to time equal five units. This time is after a very close encounter of two-bodies and just before the total collapse of the system. All solutions were compared with the more accurate standard solution described above and also were integrated in reverse, from time equal five units to time equal zero. The comparisons and the reversals showed similar accuracies, hence only comparisons of the various solutions with the standard solution are presented here.

Some results of the comparisons are shown in Tables II and III.

TABLE II

25-Body problem – accuracy comparison

	Mean error	Time of calculation (seconds)
Uncorrected	1.6×10^{-1}	178
Corrected (1)	5.1×10^{-3}	179
Corrected (2)	1.0×10^{-3}	178

TABLE III

25-Body problem – Time of calculation comparison

	Mean error	Time of calculation (seconds)
Uncorrected	5.0×10^{-3}	228
Corrected (1)	5.1×10^{-3}	179
Corrected (2)	5.3×10^{-3}	166

The first column of Tables II and III gives the various solutions that were obtained. The first solution, denoted as 'Uncorrected', is a numerical integration of the system of 25-bodies without corrections to satisfy the integrals. The second solution, denoted as 'Corrected (1)', is the integration of the system with corrections, satisfying all ten inte-

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

grals: energy, angular momentum, and center of mass. The third solution, denoted as 'Corrected (2)', is the integration with corrections, satisfying only the energy integral. Both solutions, Corrected (1) and Corrected (2), performed unweighted, least-squares corrections determined by Equation (16) given above. The second column in Tables II and III gives the error of each solution. The state vector of each solution at the final time of five units, containing 150 components of the positions and the velocities of the 25 bodies, was compared with the final state vector of the accurate, standard solution described above. Denoting the 150 differences between the standard and less accurate solutions by ε_i, $i=1, 2, \ldots, 150$; the mean error is computed by

$$\sigma = \left(\frac{1}{150} \sum_{i=1}^{150} \varepsilon_i^2\right)^{1/2}.$$

The quantity σ is given in the second column of the Tables II and III, denoted by 'Mean Error'. In addition to the calculation of the σ's, all numerical integrations were reversed. The errors indicated by the forward and backward integrations for all of the various solutions are consistent with the errors ε_i. Also, the dispersion of the quantities ε_i about the mean error σ is similar for all solutions. Hence, only the mean error σ is given in the tables. The third column of the tables, denoted by 'Time of Calculation', gives the execution time that a CDC 6600 computer required to generate each solution to a time equal to five units and with the accuracy given in the second column of the tables.

The results shown in Tables II and III indicate that the method presented here yields a more efficient numerical integration process. In Table II, a greater accuracy is obtained with the method while using the same time of calculation. And in Table III, the same accuracy is obtained with the method while using less time of calculation.

It may be seen from Tables II and III that satisfying only the energy integral ('Corrected (2)') produces more efficient numerical integrations than the integrations satisfying all ten integrals ('Corrected (1)'). The reasons for this are: (1) The energy error was about 10^5 times larger than the errors in the angular momentum and center of mass integrals; (2) The satisfaction of all ten integrals requires the inversion of a 10×10 matrix of Equation (16) as well as various matrix multiplication operations with 10×150 matrices. Whereas satisfaction of only the energy integral requires just a scalar division for the inversion and dot products of vectors of 150 dimensions. Hence, the results of Tables II and III show that, for the gravitational system and initial conditions of this application, and probably for many other systems and initial conditions, the added improvement of the corrections due to the inclusion of angular momentum and center of mass is small at a large computational expense.

Several time-saving techniques have been incorporated into the correction procedure. The most important is to calculate corrections not at every integration step but only when the corrections become significant. In the application presented here, if the error of the integral of energy increased to a value of approximately 100 times less than the desired or requested truncation error, only then were corrections calculated to satisfy the integrals.

https://doi.org/10.1017/S0252921100028487 Published online by Cambridge University Press

The results shown in Tables I, II, and III, were not obtained using a weighting matrix in the calculation for the corrections. Some numerical integrations were obtained using weighted, least-squares corrections, where the weights were the truncation error vectors described in the preceding section. No appreciable difference in accuracy was noticed between the solutions using weighted corrections and solutions using unweighted corrections. This result is preliminary since too few solutions have been obtained using weighted, least-squares corrections to form a definite conclusion.

The method presented here may be applied to the numerical solution of any system of differential equations that possesses integrals. For gravitational systems, this could include the equations of motion of the restricted problem of three bodies and the equations of motion of a particle under the attraction of a non-spherical solid body. The equations of motion of the system may also be formulated in a set of regularized variables, as long as the variables are constrained by one or more integral relations.

Acknowledgments

The motivation for this study came from a paper by Szebehely (1968). For the many illuminating discussions during the preparation of this study, I am most grateful to Professor Szebehely. To D. Bettis and R. Gottlieb goes my warmest thanks for many discussions and contributions. In particular, I thank Mr. Gottlieb for his contribution of the weighting matrix so that the estimated truncation errors may be included into the corrections. I thank R. Trevino for his diligence in performing many of the numerical comparisons, C. Weiss for his study of the magnitude of the corrections, and A. Rios-Neto for applying the method to the harmonic oscillator. The financial support of the National Science Foundation, Grant No. GP-27369, and the Office of Naval Research, Grant No. N00014-67-A-0126-0007, is gratefully acknowledged.

References

Aarseth, S. J.: 1966, *Monthly Notices Roy. Astron. Soc.* **132**, 35.
Fehlberg, E.: 1966, *Z. Angew. Math. Mech.* **46**, 1.
Forsyth, A. R.: 1930, *Geometry of Four Dimensions*, **I**, Cambridge, p. 80.
Gottlieb, R.: 1970, private communication.
Lanczos, C.: 1949, *The Variational Principles of Mechanics*, Toronto, p. 43.
Lecar, M.: 1968, *Bull. Astron.* **3**, 91.
Miller, R. H.: 1964, *Astrophys. J.* **140**, 250.
Szebehely, V. G.: 1968, *Bull. Astron.* **3**, 33.