# Achieving Brouwer's law with high-order Störmer multistep methods 

K. R. Grazier* W. I. Newman ${ }^{\dagger}$ James M. Hyman ${ }^{\ddagger}$ Philip W. Sharp ${ }^{\S}$ David J. Goldstein ${ }^{\S}$

(Received 28 October 2004, revised 13 July 2005)


#### Abstract

The integration of Newton's equations of motion for self-gravitating systems, particularly in the context of our Solar System's evolution, remains a paradigm for complex dynamics. We implement Störmer's multistep method in backward difference, summed form and perform


[^0]arithmetic according to what we call 'significance ordered computation.' We achieve results where the local truncation error of our order thirteen integrator resides below machine (double) precision and roundoff error accumulation is strictly random and not systematic.

## Contents

1 Introduction ..... C787
2 Truncation error ..... C791
3 Implementation ..... C796
4 Numerical tests ..... C798
5 Discussion ..... C801
ReferencesC802

## 1 Introduction

Many large $N$-body problems emerging in nature obey Newton's laws of motion. Their simulation is important to understanding the origin and evolution of planetary systems, as well as macromolecules in biological applications, the behavior of charged particles in plasmas, and other applications.

The speed of modern computers is such that simulations over large intervals of time are possible. For example, one-billion year simulations of the Solar System are now routine [5, 6, 2, 10, 11, e.g.].

Here we consider $N$-body simulations of the Solar System where the position of the bodies, including the phase angle, must be calculated accurately.


Figure 1: The cosine of the phase error for the order two, explicit StormerVerlet method applied to the two-body problem with zero eccentricity: blue (solid) line, stepsize of $2 \pi / 100$; red (dashed) line, stepsize of $2 \pi / 200$.

While imprecise phase information may suffice for simulations intended to gain qualitative insight, precise phase information about the bodies, particularly the planets, is essential to computing their mutual effect on each other and determining how resonances shape their orbital evolution. Precise phase information is also required to ensure close approaches between planets and small bodies are correctly handled.

The simulations can be done using symplectic methods but typically a small stepsize is required to ensure the phase error is sufficiently small. This
point is illustrated by the results depicted in Figure 1. The figure gives the cosine of the phase error $e_{\theta}$ for the well known order two, explicit StormerVerlet method applied to the two-body problem with zero eccentricity. The problem was implemented so the period $T$ was $2 \pi$. We observe for a stepsize of $T / 100$ that $\cos e_{\theta}$ oscillates between 1 and -1 with a period of approximately $760 T$. Halving the stepsize to $T / 200$ gives a four-fold increase in the period of oscillation of $\cos e_{\theta}$, a result we could expect for an order two method. We confirmed this behaviour by reducing the stepsize to $T / 1000$ when the period of the oscillation became approximately $76,000 T$. Even 76,000 orbits is small compared with the number of orbits for the planets in a one-billion year simulation.

The phase error is reduced using higher order symplectic methods. This comes at the expense of more evaluations of the acceleration per step. An alternative is to use a Störmer (multistep) method of high order. Such a method is well suited to this task because it provides accurate solutions using one evaluation of the acceleration per step. The raison d'etre of this paper is the development of a methodology that will compute the most accurate solution that is possible for a given problem on a given computer, an approach explored in depth by Goldstein [4]. In this sense, our methodology meets the criterion of Feng and Qin [3] for an a posteriori symplectic scheme. Put another way, we seek to suppress the truncation error from our computed trajectories and minimize the roundoff error.

While this approach requires more CPU time than symplectic methods used with a large stepsize, it assures trajectories are as accurate as possible.

Assuming the underlying problem is not manifestly chaotic, Brouwer [1] showed for a fixed stepsize that the growth of random roundoff error will proceed as $t^{1 / 2}$ for conserved quantities such as the energy, and as $t^{3 / 2}$ for other dynamical variables such as the position, where $t$ is the integrated time.

Our philosophy is to employ a high-order Störmer method [8, 7, e.g.] with a formal truncation error that lies below machine precision and that
minimizes the accumulation of arithmetic roundoff error. Truncation error is generally systematic, while roundoff error is rendered random (that is, not systematic) by taking suitable precautions. By keeping truncation error below machine precision, the solution obtained will be as accurate as is possible for that machine. We exploit a set of procedures that are well-known in the numerical analysis community [9, e.g.] to help assure that roundoff accumulation is not systematic. The increased computational expense for this increased reliability and accuracy is often within reach of current fast workstations.

As can be seen from the above description, our method is not in principle restricted to $N$-body simulations of the Solar System. However, to apply the method we require the local truncation error be below machine precision. This requirement clearly depends on the ordinary differential equations being solved. Thus to illustrate our method we need to specify the problem being solved. We use the two-body problem as it has important applications and is a commonly used test problem.

The results reported are for the order thirteen Störmer method. We experimented with Störmer methods of other orders and found the order thirteen method was the best suited for our purposes.

We begin with an investigation of the $L_{2}$ norm of the principal term in the local truncation for the two-body problem with eccentricities up to $1 / 2$ (the maximum of the orbital eccentricities of the gas giants is 0.06 , whereas that of the nine planets is 0.25 ). We show for the order thirteen Störmer method with a stepsize of $h=2 \pi / 1000$ that the norm is less than machine precision in IEEE double precision arithmetic. In Section 3 we describe how to implement Störmer's method to minimize the accumulation of roundoff error. Then in Section 4 we present two sets of numerical results for our method applied to the two-body problem. The first set illustrates how the global error depends on the stepsize at a fixed $t$. The second set is for two ten-million orbit integrations and demonstrates our ability to match the 1937 prediction of Brouwer. We end in Section 5 with a brief discussion.

## 2 Truncation error

Let

$$
\begin{equation*}
\ddot{x}=f(x(t), t), \quad t>0, \quad x(0)=x_{0}, \quad \dot{x}(0)=\dot{x}_{0}, \tag{1}
\end{equation*}
$$

denote the initial value problem describing Newton's laws of motion for an $N$-body simulation. Störmer's method of order $q$ applied to (1) is

$$
\begin{equation*}
x_{n+1}-2 x_{n}+x_{n-1}=h^{2} \sum_{i=0}^{q-1} \alpha_{i} f_{n-i}, \quad n \geq q-1 \tag{2}
\end{equation*}
$$

where $x_{j}$ is the numerical approximation to $x(t)$ at $t=j h, j=0,1, \ldots$, and $f_{j} \equiv f\left(x_{j}, t_{j}\right), j=0,1, \ldots$ The starting values $x_{j}, j=1, \ldots, q-1$, are assumed sufficiently accurate.

Write method (2) in the backward difference form

$$
\begin{equation*}
x_{n+1}-2 x_{n}+x_{n-1}=h^{2} \sum_{m=0}^{q-1} \beta_{m} \nabla^{m} f_{n} \tag{3}
\end{equation*}
$$

where $\nabla^{0} f_{n}=f_{n}$ and $\nabla^{j+1} f_{n}=\nabla^{j} f_{n}-\nabla^{j} f_{n-1}, j=0,1, \ldots$.
The coefficients $\alpha_{i}$ in (2) for high order methods are large in magnitude. This, combined with their alternation in sign, induces substantial growth in roundoff error on long $N$-body simulations of the Solar System [12]. In contrast, the coefficients $\beta_{i}$ in method (2) are all positive and decrease with $i$, leading to far smaller roundoff error.

As outlined in Section 1, an important aspect of our implementation is choosing the stepsize so the norm of the local truncation error (LTE) is below machine precision relative. We now show the principal term in the LTE satisfies this requirement for the order thirteen Störmer method with stepsize $2 \pi / 1000$ applied to the two-body problem with eccentricities up to $1 / 2$.

The equations of motion for the two-body problem are

$$
\begin{equation*}
\ddot{x}=-\frac{x}{\|x\|_{2}^{3}}, \quad x(0)=[1-\epsilon, 0]^{T}, \quad \dot{x}(0)=\left[0,\left((1+\epsilon)(1-\epsilon)^{-1}\right)^{1 / 2}\right]^{T}, \tag{4}
\end{equation*}
$$

where $\epsilon$ is the eccentricity. The true solution is $x(t)=[\cos u-\epsilon,(1-$ $\left.\left.\epsilon^{2}\right)^{1 / 2} \sin u\right]^{T}$ where $u$ satisfies Kepler's equation $u-\epsilon \sin u-t=0$.

Let $e_{q}(t)$ denote the principal term in the LTE at $t$ for the order $q$ Störmer method applied to (4). We have, see p. 296 of Henrici [8] for example,

$$
\begin{equation*}
e_{q}(t)=C_{q+2} h^{q+2} x^{(q+2)}(t) . \tag{5}
\end{equation*}
$$

For $q=13, C_{q+2}=73399737279 / 15!(\approx 0.0561299809)$.
The minimum of $\|x(t)\|_{2}$ is $1-\epsilon$. Hence the norm of the principal term is below machine precision if $\left\|e_{q}(t)\right\|_{2} \leq u(1-\epsilon)$, where $u$ is the unit roundoff. In IEEE double precision arithmetic, $u \approx 2.22 \times 10^{-16}$. A lower bound on $u(1-\epsilon)$ for $0 \leq \epsilon \leq 1 / 2$ is then $10^{-16}$ and in the analysis below we require

$$
\begin{equation*}
\left\|C_{q+2} h^{q+2} x^{(q+2)}(t)\right\|_{2} \leq 10^{-16} \tag{6}
\end{equation*}
$$

To investigate $e_{13}(t)$, we require the derivative $x^{(15)}(t)$. This is found either by recursion, see for example p.107-8 of Roy [14], or by direct differentiation. Both approaches give

$$
\begin{equation*}
\left\|x^{(15)}(t)\right\|_{2}^{2}=(1-\epsilon w)^{-56} \sum_{j=0}^{26} \sum_{k=0}^{[j / 2]} b_{j k} w^{2 k+\sigma_{j}} \epsilon^{j} \tag{7}
\end{equation*}
$$

where $w=\cos u, b_{j k}$ are integers, $[j / 2]$ denotes the integer part of the division $j / 2$, and $\sigma_{j}$ is zero if $j$ is even and one if $j$ is odd. The $b_{j k}$ are available from the authors.

To gain insight for our analysis, Figure 2 plots $\left\|x^{(15)}(t)\right\|_{2}^{2}$ as a function of $w$ and $\epsilon,-1 \leq w \leq 1,0 \leq \epsilon \leq 1 / 2$ : we normalised $\left\|x^{(15)}(t)\right\|_{2}^{2}$ by its value
for $w=1, \epsilon=1 / 2$. Observe there is a marked peak at $w=1, \epsilon=1 / 2$. As we show below, this peak complicated our analysis.

We also evaluated $\left\|x^{(15)}(t)\right\|_{2}^{2}$ for a large number of values of $w$ and $\epsilon$. For a fixed $\epsilon$, we found the maximum of $\left\|x^{(15)}(t)\right\|_{2}^{2}$ was at $w=1$ and the maximum increased monotonically with $\epsilon$. These two results are in accord with physical reasoning. Unfortunately, the maximum of the summation in (7) did not in general occur at $w=1$ and the summation contained terms of opposite sign and similar magnitude. These features combined with the high degree polynomials in $w$ and $\epsilon$ eliminated standard techniques for bounding bi-variate rational polynomials. If the maximum of $\left\|x^{(15)}(t)\right\|_{2}^{2}$ for fixed $\epsilon$ was assumed to occur at $w=1$, our analysis would simplify to a few lines.

For $-1 \leq w \leq 0,\left\|x^{(15)}(t)\right\|_{2}^{2}$ is bounded above by

$$
\begin{equation*}
\sum_{j=0}^{26} \sum_{k=0}^{[j / 2]}\left|b_{j k}\right| \epsilon^{j} \tag{8}
\end{equation*}
$$

This bound increases monotonically with $\epsilon$. When $\epsilon=1 / 2$ and $h=2 \pi / 1000$, the resulting bound on $\left\|e_{13}(t)\right\|_{2}$ is $9.38 \times 10^{-23}$, six orders of magnitude smaller than required. This result permits us to restrict $w$ to $[0,1]$ for the rest of our analysis.

For $0 \leq w \leq 1,\left\|x^{(15)}(t)\right\|_{2}^{2}$ can be bounded above by omitting the terms in (7) with $b_{j k}<0$. This bound increases monotonically with $w$ and $\epsilon$. If the bound is used for $\left\|x^{(15)}(t)\right\|_{2}^{2}$ and $h=2 \pi / 1000$, (6) is satisfied for $\{(w, \epsilon)$ : $0 \leq w \leq 81 / 100,0 \leq \epsilon \leq 1 / 2\}$ and $\{(w, \epsilon): 0 \leq w \leq 1,0 \leq \epsilon \leq 42 / 100\}$. This leaves the region $\{(w, \epsilon): 81 / 100 \leq w \leq 1,42 / 100 \leq \epsilon \leq 1 / 2\}$ to investigate. This region includes the peak of Figure 2.

Let

$$
\alpha_{j}(w)=\sum_{\substack{k=0 \\ b_{j k}>0}}^{[j / 2]} b_{j k} w^{2 k+\sigma_{j}}, \quad \beta_{j}(w)=\sum_{\substack{k=0 \\ b_{j k}<0}}^{[j / 2]} b_{j k} w^{2 k+\sigma_{j}} .
$$



Figure 2: $\left\|x^{(15)}(t)\right\|_{2}^{2}$ for $-1 \leq w \leq 1,0 \leq \epsilon \leq 1 / 2$. The value of $\left\|x^{(15)}(t)\right\|_{2}^{2}$ has been normalised by its value for $w=1, \epsilon=1 / 2$.

The expression for $\left\|x^{(15)}(t)\right\|_{2}^{2}$ can then be written as

$$
\begin{equation*}
(1-\epsilon w)^{-56}\left(\sum_{j=0}^{26} \alpha_{j} \epsilon^{j}+\sum_{j=0}^{26} \beta_{j} \epsilon^{j}\right) . \tag{9}
\end{equation*}
$$

If we set $w_{0}=81 / 100$ and $w_{1}=w_{0}+\delta w, \delta w>0$, the concavity of the expressions for $\alpha_{j}$ and $\beta_{j}$ imply

$$
\alpha_{j}(w) \leq \bar{\alpha}_{j}(w) \equiv \alpha_{j}\left(w_{0}\right)+\frac{\alpha_{j}\left(w_{1}\right)-\alpha_{j}\left(w_{0}\right)}{\left(w_{1}-w_{0}\right)}\left(w-w_{0}\right), \quad w_{0} \leq w \leq w_{1},
$$

$$
\beta_{j}(w) \leq \bar{\beta}_{j}(w) \equiv \beta_{j}\left(w_{0}\right)+\frac{d \beta_{j}\left(w_{0}\right)}{d w}\left(w-w_{0}\right), \quad w \geq w_{0}
$$

where $d \beta_{j}\left(w_{0}\right) / d w \leq 0$. If we now set $\epsilon_{0}=42 / 100$, we have

$$
\begin{equation*}
\left\|x^{(15)}(t)\right\|_{2}^{2} \leq(1-\epsilon w)^{-56} P_{1} \equiv(1-\epsilon w)^{-56}\left(\sum_{j=0}^{26} \bar{\alpha}_{j} \epsilon^{j}+\sum_{j=0}^{26} \bar{\beta}_{j} \epsilon_{0}^{j}\right) \tag{10}
\end{equation*}
$$

for $w_{0} \leq w \leq w_{1}, \epsilon \geq \epsilon_{0} . P_{1}$ is linear in $w$ and it is easily shown by evaluation that $\partial P_{1} / \partial w>0$ for $w=w_{0}, \epsilon=\epsilon_{0}$. Hence the upper bound $(1-\epsilon w)^{-56} P_{1}$ on $\left\|x^{(15)}(t)\right\|_{2}^{2}$ increases monotonically with $w$ and $\epsilon$ for $w_{0} \leq w \leq w_{1}, \epsilon \geq \epsilon_{0}$. This observation suggests the following scheme.

Set $\epsilon=\epsilon_{1}$ where $\epsilon_{1}=\epsilon_{0}+\delta \epsilon, \epsilon_{0}<\epsilon_{1} \leq 1 / 2$. Then check if (6) is satisfied for $w=w_{1}$. If it is, set $w_{0}=w_{1}$ and check by evaluation if $\partial P_{1} / \partial w>0$ at $w=w_{0}$. If it is not, set $w_{1}=w_{0}+\delta w$ and repeat the process. Stop when $w_{1}=1$ or $\partial P_{1} / \partial w<0$. If $w_{1}=1$, set $\epsilon_{1}=\epsilon_{1}+\delta \epsilon$ and repeat the process from $w_{0}=81 / 100$. Continue until $\epsilon_{1}=1 / 2$ and $w_{1}=1$, or $\partial P_{1} / \partial w<0$.

Although there is freedom in the choice of $\delta w$ and $\delta \epsilon$, they must not be too large. Otherwise, $\bar{\alpha}_{j}$ and $\bar{\beta}_{j}$ will be poor approximations to $\alpha_{j}$ and $\beta_{j}$ and $\left[(1-\epsilon w)^{-56} P_{1}\right]^{1 / 2}$ might not satisfy (6) when the true value of $\left\|x^{(15)}(t)\right\|_{2}$ does. We found for $\delta w=1 / 100, \delta \epsilon=2 / 100$ that (6) was satisfied for the region $\{(w, \epsilon): 81 / 100 \leq w \leq 1,42 / 100 \leq \epsilon \leq 1 / 2\}$.

We confirmed the above analysis by implementing the order thirteen Störmer method in 50 digit arithmetic, with the starting values calculated from the true solution. With $h=2 \pi / 1000$, the norm of the lTE error differed by no more than $2.5 \%$ from the norm of the principal term on a grid of representative values of $\epsilon$ and $w$.

It is instructive to examine $\left\|x^{(15)}(t)\right\|_{2}^{2}$ under the assumption its maximum for fixed $\epsilon$ occurs at $w=1$. The expression for $\left\|x^{(15)}(t)\right\|_{2}^{2}$ with $w=1$ is

$$
\begin{equation*}
(1+\epsilon)(1-\epsilon)^{-43}\left(\sum_{j=0}^{6} a_{j} \epsilon^{j}\right)^{2} \tag{11}
\end{equation*}
$$

where coefficients $a_{0}=1, a_{1}=16362, a_{2}=5020623, a_{3}=243313164$, $a_{4}=3274844175, a_{5}=14427513450$ and $a_{6}=18261468225$. Since the coefficients $a_{j}$ are all positive, (11) increases monotonically with $\epsilon$. If $\epsilon=1 / 2$, $\left\|e_{13}\right\|_{2} \approx 1.86 \times 10^{-19}$ for $h=2 \pi / 1000$. In addition, (6) is satisfied for $h \leq 2 \pi / 658$. Thus $h=2 \pi / 658$ serves as an upper bound on the permissible stepsize and indicates $h=2 \pi / 1000$ is not unduly pessimistic.

## 3 Implementation

Roundoff error has long been known to be a major limiting factor in long-term Solar System integrations. Brouwer [1] pioneered this field and many others have since contributed. Especially noteworthy are Quinn and Tremaine [13] and Quinlan [12] who also employed variants of Störmer's method. Using conventional computers (in contrast to special purpose machines with extended precision), they achieved Brouwer's Law with energy error $\propto t^{1 / 2}$ and position error $\propto t^{3 / 2}$ over $10^{7}$ and $10^{8}$ timesteps, respectively. Grazier et al. $[5,6]$ achieved these limits over $10^{11}$ timesteps.

Quinlan [12] appreciated the importance of using backward differences and "summed form." However, Quinlan did not observe the benefit of forcing truncation error to reside below the precision of the computer, thereby making roundoff error the remaining major issue. As an illustration of how significant issues relating to roundoff error have become, dynamical astronomers [13, e.g.] largely abandoned CRAY X-MP computers since they did not abide by the ANSI/IEEE Standard 754 for floating point arithmetic and produced unacceptably large errors (primarily due to biased rounding in hardware).

We now describe the procedure that we developed and call "significance ordered computation." It is based on two principles.

First, we implement Störmer's method so that all series that must be
summed utilize coefficients of comparable magnitude. We have already remarked that the backwards difference form for the Störmer methods has this benefit. The important point here is that the explicit conversion of terms in $\nabla^{m} f_{n}$ into summations over $f_{n-m}$ results in binomial coefficients with alternating signs that will vary over several orders of magnitude. This alone can cause, as we observed in long-time simulations, in the loss of four or more significant digits. A related issue emerges from writing $x_{n+1}=$ $2 x_{n}-x_{n-1}+h^{2} \times \cdots$. The coefficient 2 multiplying $x_{n}$ is sufficient to exacerbate rounding effects. We avoid this using the "leap-frog" formulation $v_{n+1 / 2}=\left(x_{n+1}-x_{n}\right) / h$ wherein we write in place of our original equation $x_{n}=x_{n-1}+h v_{n-1 / 2}$ and $v_{n+1 / 2}=v_{n-1 / 2}+h \times \cdots$. This recasting of the original expressions is generally called the "summed form."

Second, we take particular care in evaluating

$$
\begin{equation*}
\sum_{m=0}^{q-1} \beta_{m} \nabla^{m} f_{n-m} \tag{12}
\end{equation*}
$$

Higham [9] defines what he calls the insertion method wherein elements $y_{i}$ of a series that must be summed are sorted by increasing magnitude and summed pairwise. Since we select our stepsize $h$ to assure that the truncation error resides below machine precision, the summation that we must evaluate, when placed in reverse order, has exactly this effect. This follows because each successive term is often more than an order of magnitude smaller than the previous one. Our numerical test verified this procedure yields two or more additional significant figures and, most importantly, assures that our roundoff errors have zero mean, that is, they do not have a systemic bias. The combination of Higham's insertion method with the use of the summed form and backward differences constitutes what we call significance ordered computation.


Figure 3: Global error versus stepsize after $10^{5}$ periods for the two-body problem with eccentricity 0.05 .

## 4 Numerical tests

Our first set of tests applied our order thirteen Störmer method with a range of stepsizes to the two-body problem with eccentricity of 0.05 (close to that of Jupiter). For each stepsize, we performed sixteen integrations with different, randomly chosen, initial conditions. The starting values and the reference solution were found from the true solution.

Figure 3 gives the root mean square (RMS) of the global error. The figure reveals two asymptotic scalings and a transition (blue) regime. The truncation dominated (black) regime shows a power-law growth with an exponent of 10.8 ; this is in reasonable agreement with the expected value of 13 given the order was estimated from just two points. More importantly, the
roundoff dominated (red) regime shows a power-law decay with an exponent of -0.514 which is very close to the expected $-1 / 2$. In addition, stepsizes below $10^{-3}$ times the period (or, more correctly, the shortest timescale in the problem) bring us into the regime where roundoff dominates.

To test the scaling properties of the error in our integration method, we performed two sets of ten million orbit integrations for the two-body problem. One set was for an eccentricity of 0.05 and the other set was for an eccentricity of 0.50 , an "oscillatory stiff" problem containing a wide distribution of time scales when compared to nearly circular planetary orbits. Each set consisted of sixteen integrations with different, randomly chosen, initial conditions. The stepsize for all integrations was $2 \pi / 1000$.

Figure 4 gives the RMS of the global error and relative energy error for each set of sixteen integrations. The RMS errors were calculated at one thousand evenly spaced points over the ten-million orbits and are of particular importance should the integrator exhibit error growth properties consistent with Brouwer's Law.

The RMS of the global error after ten million periods was $7.1 \times 10^{-4}$ and $1.3 \times 10^{-3}$ for $\epsilon$ equal to 0.05 and 0.50 respectively, and the RMS of the relative energy error was $9.7 \times 10^{-12}$ and $1.3 \times 10^{-11}$ respectively. The good agreement between the two numbers in each pair of errors is further evidence that roundoff error is dominating the total error.

We used linear least squares to fit a power law to each curve in Figure 4. For $\epsilon=0.05$, the exponent in the power law for the global error was 1.54 and that for the relative energy error 0.52 , in very good agreement with the expected values of $3 / 2$ and $1 / 2$ respectively. For $\epsilon=0.50$, the exponents were 1.30 and 0.48 respectively. An exponent of 1.30 is noticeably smaller than expected. A closer examination of the upper curves in Figure 4 suggests the presence of transitory behaviour for $\epsilon=0.50$. The exponent for a power law fit on the first $10^{6}$ orbits is 1.03 whereas that on the remaining orbits is 1.44 .


Figure 4: The RMS of the global error (top) and relative energy error (bottom) for the two-body problem with $\epsilon=0.05$ and $\epsilon=0.5$.

A very good estimate of the RMS of the relative energy error at the endpoint is obtained in the following simple but effective way. As in Henrici [8, p. 35 et seq.], we define by $y$ the true or exact value of a number, and $\tilde{y}$ as its correctly rounded computer representation. By that we mean

$$
|y-\tilde{y}| \leq \frac{1}{2} u
$$

where $u$ is the unit roundoff for the computer. Now let $\Delta E_{n}$ represent the error in the computed estimate of the energy (or any other conserved quantity) at $t=t_{n}$. Given our mode of significance ordered computation, it follows that this error grows according to

$$
\Delta E_{n+1}=E_{n}+\xi_{n}, \quad n=0,1, \ldots,
$$

where $\xi_{n}$ is an individual independent deviate which satisfies

$$
\left\langle\xi_{n}\right\rangle=0 \quad \text { and } \quad\left\langle\xi_{m} \xi_{n}\right\rangle=\delta_{m, n}\left(\frac{1}{2} u\right)^{2}, \quad m, n=0,1, \ldots,
$$

where the operator $\langle\cdot\rangle$ provides the expected value for the random variable. Assuming $\left\langle E_{0}\right\rangle=0$, it follows that

$$
\left\langle\Delta E_{n}^{2}\right\rangle^{1 / 2}=\sqrt{n}\left(\frac{1}{2} u\right) .
$$

Thus, for $u \approx 2 \times 10^{-16}$ and $n=10^{10}$ ( $10^{7}$ orbits and $10^{3}$ steps per orbit), the estimated Rms of the relative energy error is $10^{5} \times 10^{-16}=10^{-11}$, which compares very favourably with our numerical results.

## 5 Discussion

We developed a methodology predicated on producing the most accurate numerical solution possible to Newton's laws of motion on a given computer.

We used the order thirteen Störmer method with the stepsize sufficiently small to assure the truncation error resided below machine precision, and took steps to mitigate and eliminate any systematic influences of roundoff error. While this procedure is more costly than other schemes, it produces results that are orders of magnitude more accurate.

We illustrated our method using the two-body problem. For eccentricities of 0.05 and 0.5 , we obtained a power-law growth in the global error and relative energy error over ten-million orbits close to that expected when the error is dominated by random round-off errors.

Acknowledgment: We thank the two referees for their careful reading of the submitted paper and their suggestions about improving it, particularly Sections 2 and 3. We are particularly grateful to Ferenc Varadi for much discussion. This research was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, and was supported in part by nasa grant nagw 3132 to UCLA.

## References

[1] D. Brouwer. On the accumulation of errors in numerical integration. Astron. J. 46:149-153, 1937. C789, C796
[2] M. Duncan, T. Quinn, and S. Tremaine. The long-term evolution of orbits in the Solar System - A mapping approach. Icarus, 82:402-419, 1989. C787
[3] K. Feng and M. Z. Qin. The symplectic methods for the computation of Hamiltonian Equations. Proc. Conf. Num. Methods for PDE's.
Berlin: Springer-Verlag, 1-37, 1987. C789
[4] D. J. Goldstein. The Near Optimality of Störmer Methods for Long Time Integration of $y^{\prime \prime}=g(y)$. Ph. D. Thesis, Department of Mathematics, University of California, Los Angeles, 1996. C789
[5] K. R. Grazier, W. I. Newman, W. M. Kaula, and J. M. Hyman. Dynamical evolution of planetesimals in the outer Solar System. I. The Jupiter/Saturn zone. Icarus 140:341-352, 1999. C787, C796
[6] K. R. Grazier, W. I. Newman, W. M. Kaula, and J. M. Hyman. Dynamical devolution of planetesimals in the outer Solar System. II. The Saturn/Uranus and Uranus/Neptune zones. Icarus 140:353-368, 1999. C787, C796
[7] E. Hairer, S. P. Nørsett, and G. Wanner. Solving Ordinary Differential Equations, second revised edition. Berlin: Springer-Verlag. C789
[8] P. Henrici. Discrete Variable Methods in Ordinary Differential Equations. New York: John Wiley \& Sons, 1962. C789, C792, C801
[9] N. J. Higham. Accuracy and Stability of Numerical Algorithms. Philadelphia: Society for Industrial and Applied Mathematics, 1996. C790, C797
[10] M. J. Holman. A possible long-lived belt of objects between Uranus and Neptune. Nature. 387:785-788, 1997. C787
[11] T. Ito and K. Tanikawa. Long-term integrations and stability of planetary orbits in our Solar System. MNRAS 336:483-500, 2003. C787
[12] G. D. Quinlan. Round-off error in long-term orbital integrations using multistep methods. Celest. Mech. Dyn. Astron. 58:339-351, 1994. C791, C796
[13] T. Quinn and S. Tremaine. Roundoff error in long-term planetary orbit integrations. Astron. J. 99:1016-1023, 1990. C796
[14] A. E. Roy. Orbital motion. 3 ed., Adam Hilger, ISBN 0-85274-229-0, 1988. C792


[^0]:    *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. mailto:krg@anlashok.jpl.nasa.gov
    ${ }^{\dagger}$ Departments of Earth \& Space Sciences, Physics \& Astronomy, and Mathematics, University of California, Los Angeles, CA 90095-1567, USA. mailto:win@ucla.edu
    ${ }^{\ddagger}$ Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. mailto:mac@t7.lanl.gov
    ${ }^{\S}$ Department of Mathematics, Private Bag 92019, Auckland, New Zealand. mailto:sharp@math.auckland.ac.nz
    © Consultant, Culver City, CA, USA. mailto: acwdjg@comcast.net
    See http://anziamj.austms.org.au/V46/CTAC2004/Graz for this article, © Austral. Mathematical Soc. 2005. Published August 23, 2005. ISSN 1446-8735

