SIAM J. SCI. COMPUT.
Vol. 14, No. 4, pp. 936-952, July 1993

010

THE DEVELOPMENT OF VARIABLE-STEP SYMPLECTIC INTEGRATORS, WITH APPLICATION TO THE TWO-BODY PROBLEM*

M. P. CALVO† AND J. M. SANZ-SERNA†

Abstract. The authors develop and test variable step symplectic Runge-Kutta-Nyström algorithms for the integration of Hamiltonian systems of ordinary differential equations. Numerical experiments suggest that, for symplectic formulae, moving from constant to variable stepsizes results in a marked decrease in efficiency. On the other hand, symplectic formulae with constant stepsizes may outperform available standard (nonsymplectic) variable-step codes. For the model situation consisting in the long-time integration of the two-body problem, our experimental findings are backed by theoretical analysis.

Key words. symplectic integration, Kepler's problem, Runge-Kutta-Nyström methods

AMS subject classifications. 65L05, 70H05, 70F05

1. Introduction. In mechanics, optics, chemistry, etc., situations where dissipation does not play a significant role may be modelled by means of Hamiltonian systems of ordinary differential equations (ODEs) or partial differential equations (PDEs) [2]. Hamiltonian systems of ODEs are of the form

$$\dot{p}^I = -\partial H/\partial q^I, \quad \dot{q}^I = \partial H/\partial p^I, \quad 1 \leqq I \leqq d, \tag{1.1}$$

where the integer d is the number of degrees of freedom, the Hamiltonian $H = H(\mathbf{p}, \mathbf{q}) = H(p^1, \ldots, p^d, q^1, \ldots, q^d)$ is a sufficiently smooth, real function of $2d$ real variables, and a dot represents differentiation with respect to t (time). There has been much recent interest in the numerical integration of (1.1) by means of so-called symplectic or canonical integrators, starting with the work of Ruth [12], Feng [7], and Channell and Scovel [4]. An extensive list of references can be found in the survey [14].

In order to explain in simple terms the meaning and relevance of symplecticness, it is advisable to consider first the question of how to tell, from the knowledge of the *solutions* of a system of ODEs, whether the *system* is of Hamiltonian form or otherwise. More precisely, let $\mathcal{S}$ be an autonomous system of ODEs for the dependent variables $(\mathbf{p}, \mathbf{q})$, and let us introduce the $\mathbb{R}^{2d}$-valued function $\varphi_t(\mathbf{p}_0, \mathbf{p}_0)$ such that, for fixed $\mathbf{p}_0$ and $\mathbf{q}_0$ and varying t, $(\mathbf{p}(t), \mathbf{q}(t)) = \varphi_t(\mathbf{p}_0, \mathbf{q}_0)$ is the solution of $\mathcal{S}$ with initial condition $\mathbf{p}(0) = \mathbf{p}_0$, $\mathbf{q}(0) = \mathbf{q}_0$. If we now see t as a parameter and $\mathbf{p}_0, \mathbf{q}_0$ as variables, $\varphi_t(\mathbf{p}_0, \mathbf{q}_0)$ defines a transformation in the space $\mathbb{R}^{2d}$ (the phase space). This transformation is the *flow* of the differential system $\mathcal{S}$. If we were given φ_t and at the same time $\mathcal{S}$ were concealed from us, could we tell whether $\mathcal{S}$ is a Hamiltonian system or otherwise? The answer to this question is affirmative. The system $\mathcal{S}$ is Hamiltonian *if and only if*, for each t, φ_t *is a symplectic transformation.* Now a transformation $\mathcal{T}$ in phase space is said to be symplectic [2] if for any bounded two-dimensional surface D in phase space, the sum of the two-dimensional (signed) areas of the d projections of D onto the planes (p^I, q^I) is the same as the sum of the two-dimensional (signed) areas of the d projections of $\mathcal{T}(D)$ onto the planes (p^I, q^I). Thus the symplectic character of

* Received by the editors December 18, 1991; accepted for publication (in revised form) April 30, 1992. This research was supported by Junta de Castilla y León under project 1031-89 and by Dirección General de Investigación Científica y Técnica under project PB89-0351.

† Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain (Maripaz@cpd.uva.es and sanzserna@cpd.uva.es).

[16] D. W. WALKER, P. H. WORLEY, AND J. B. DRAKE, *Parallelizing the spectral transform method—Part II*, TM-11855, Oak Ridge National Laboratory, 1991; Concurrency: Practice and Experience, submitted.

[17] Partial results were presented as a contributed paper at the SIAM Conf. on Parallel Processing for Scientific Computing, Houston, TX, March 25-27, 1991. Similar algorithms are being developed by R. Sweet, private communication.

the flow is the hallmark of Hamiltonian systems. Hamiltonian problems have many specific features not shared by other systems of differential equations. All such specific features (absence of attractors, recurrence, etc.) directly derive from the symplecticness of the corresponding flow [2].

A one-step numerical method used with steplength h defines a transformation in phase space $\psi_h(\mathbf{p}_0, \mathbf{q}_0)$ that advances the solution h units of time, starting from $(\mathbf{p}_0, \mathbf{q}_0)$. Of course, $\psi_h(\mathbf{p}_0, \mathbf{q}_0)$ is an approximation to $\varphi_h(\mathbf{p}_0, \mathbf{q}_0)$, and the numerical method approximates $\varphi_{nh} = \varphi_h^n$ by iterating n times ψ_h. For Hamiltonian problems integrated by classical methods, such as explicit Runge-Kutta methods, the transformation ψ_h turns out to be *nonsymplectic*. Then the numerical method misses the important specific features associated with symplectic transformations. However, there are *symplectic* methods for which ψ_h is guaranteed to be symplectic for Hamiltonian problems.

Numerical experiments have shown that for Hamiltonian problems, symplectic integrators may well be an improvement over their nonsymplectic counterparts. However, the development of symplectic methods has so far been confined to *constant stepsize* formulae and, accordingly, numerical tests have used as reference algorithms constant stepsize implementations of classical methods. Such implementations are, by modern numerical ODE standards, very naive, and the question arises of whether, for Hamiltonian problems, a symplectic method with constant stepsizes, may actually be more efficient than a modern variable-step code. Before we carried out the experiments reported in this paper, we felt that the answer to that question would be no. On the other hand, we suspected that for Hamiltonian problems, variable stepsize symplectic algorithms would improve on standard variable stepsize algorithms. Accordingly, we decided to develop variable stepsize symplectic algorithms.

In this paper we report on our experience with the construction and assessment of *variable-step, symplectic, explicit Runge-Kutta-Nyström algorithms*. We used Runge-Kutta-Nyström (RKN) methods rather than Runge-Kutta methods because all symplectic Runge-Kutta formulae are implicit [13]. It appears that both our guesses above were wrong: *constant stepsize symplectic methods may beat standard variable stepsize codes, but variable stepsize symplectic codes are not more advantageous than standard variable stepsize codes.*

Section 2 is devoted to the construction of the symplectic RKN code. The results of the numerical experiments are presented in § 3, where we use as a test problem the well-known two-body (Kepler) problem. In § 4 we analyze our experimental findings. In particular, we provide a complete theoretical study of the performance of general one-step numerical methods in the integration of the two-body problem. Finally, in § 5, we present our conclusions.

2. Construction of a symplectic RKN code.

2.1. RKN methods. We restrict our attention to systems of the special form

$$\dot{\mathbf{p}} = \mathbf{f}(\mathbf{q}), \qquad \dot{\mathbf{q}} = \mathbf{p} \tag{2.1}$$

(i.e., to second-order systems $\ddot{\mathbf{q}} = \mathbf{f}(\mathbf{q})$). If $\mathbf{f}$ is the gradient of a scalar function $-V(\mathbf{q})$, then (2.1) is a Hamiltonian system with

$$H = H(\mathbf{p}, \mathbf{q}) = T(\mathbf{p}) + V(\mathbf{q}), \qquad T(\mathbf{p}) = \tfrac{1}{2}\mathbf{p}^T\mathbf{p}.$$

In mechanics, the $\mathbf{q}$ variables represent Lagrangian coordinates, the $\mathbf{p}$ variables the corresponding momenta, $\mathbf{f}$ the forces, T is the kinetic energy, V the potential energy, and H the total energy [2].

An explicit RKN method for (2.1) takes the form [5], [8]

$$\mathbf{Q}_i = \mathbf{q}_n + h\gamma_i \mathbf{p}_n + h^2 \sum_{j<i} \alpha_{ij} \mathbf{f}(\mathbf{Q}_j),$$

$$\mathbf{p}_{n+1} = \mathbf{p}_n + h \sum_{i=1}^{s} b_i \mathbf{f}(\mathbf{Q}_i), \tag{2.2}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\mathbf{p}_n + h^2 \sum_{i=1}^{s} \beta_i \mathbf{f}(\mathbf{Q}_i),$$

where we assume, unless otherwise stated, that the following well-known condition [5], [8] holds:

$$\beta_i = b_i(1-\gamma_i), \qquad 1 \leqq i \leqq s. \tag{2.3}$$

As in [5], we consider first same as last (FSAL) methods, i.e., methods with

$$\gamma_1 = 0, \qquad \gamma_s = 1, \tag{2.4a}$$

$$\alpha_{sj} = \beta_j, \qquad 1 \leqq j \leqq s-1. \tag{2.4b}$$

Note that (2.4a) implies, via (2.3), that $\beta_s = 0$, and then the last stage $\mathbf{Q}_s$ of the current step coincides with $\mathbf{q}_{n+1}$, which, in turn, is the first stage of the next step. Therefore, a step of an FSAL s-stage method requires only $s-1$ evaluations of $\mathbf{f}$.

The method (2.2) is symplectic if [21], [11], [3], [14]

$$\alpha_{ij} = b_j(\gamma_i - \gamma_j), \qquad i > j. \tag{2.5}$$

For symplectic methods with s stages, we have s coefficients b_i and s coefficients γ_i as free parameters; the coefficients β_i and α_{ij} are determined by (2.3) and (2.5), respectively. On the other hand (2.3), (2.4a), and (2.5) imply (2.4b), so that a symplectic FSAL method has s coefficients b_i and $s-2$ coefficients γ_i, $2 \leqq i \leqq s-1$, as free parameters.

2.2. Derivation of a fourth-order, symplectic, FSAL RKN method. To construct a variable-step symplectic code, we decided to begin with a fourth-order formula. While higher-order formulae are expected to be more efficient, they are also more difficult to construct. For a method (2.2)-(2.3) to have order four, the coefficients should satisfy *seven* order conditions [8]. However, for symplectic methods, not all order conditions are independent [1], [3], [14], [15] and, in fact, it turns out [3] that it is sufficient to impose only *six* of them. For FSAL symplectic methods, four stages furnish six free coefficients, and after imposing order four, no room is left for "tuning" the formula. We then settle for five-stage FSAL symplectic methods, for which a two-parameter family of order-four methods exists. Following a standard practice (see [5] and [6]) we choose among the members of this family the method with "smallest" truncation error.

For smooth problems the **p**-truncation and **q**-truncation errors of an RKN method, respectively, possess Taylor expansions of the form [5], [8]

$$\sum_{i=0}^{\infty} h^{i+1} \sum_j c_j'^{(i+1)} \mathbf{F}_j^{(i+1)} \tag{2.6a}$$

and

$$\sum_{i=1}^{\infty} h^{i+1} \sum_k c_k^{(i+1)} \mathbf{F}_k^{(i)}, \tag{2.6b}$$

where the $\mathbf{F}_j^{(i)}$ are the elementary differentials that only depend on the problem (2.1) being integrated, and the $c_j'^{(i+1)}$ and $c_k^{(i+1)}$ are polynomials in the method coefficients α_{ij}, γ_i, β_i, b_i. In (2.6a), the sum in j is extended to all special Nyström trees with $i+1$ nodes, while in (2.6b), the sum in k is extended to all special Nyström trees with i nodes. For fourth-order methods, $c_j'^{(i)}$ and $c_k^{(i)}$ vanish for $i \leqq 4$ and we try to minimize $c_j'^{(5)}$ and $c_k^{(5)}$. We proceed as follows. Let us denote by $\mathbf{c}'^{(i)}$ and $\mathbf{c}^{(i)}$, respectively, the vectors with components $c_j'^{(i)}$ and $c_k^{(i)}$, and set

$$A'^{(5)} = \|\mathbf{c}'^{(5)}\|, \qquad A^{(5)} = \|\mathbf{c}^{(5)}\|. \tag{2.7}$$

(The norm is the standard Euclidean norm.) We consider $\phi = (A'^{(5)})^2 + (A^{(5)})^2$ as a function of the eight free coefficients γ_i, $2 \leqq i \leqq 4$, and b_j, $1 \leqq j \leqq 5$, and use the NAG subroutine E04UCF to minimize ϕ subject to the six equality constraints that impose order four and subject to bounds $-1.5 \leqq \gamma_i, b_j \leqq 1.5$. Of course, the minimization subroutine requires an initial guess for the minimum and converges only to a local minimum that depends on the initial guess. A thousand random initial guesses (subject to $-1.5 \leqq \gamma_i, b_j \leqq 1.5$) were taken, and we kept the local minimum with the smallest value of ϕ. The method thus obtained does not satisfy to machine precision the conditions for order four, because the NAG routine fails in exactly enforcing the equality constraints. We then kept the values b_1 and b_2 provided by the minimization routine and determined γ_i, $2 \leqq i \leqq 4$, and b_j, $3 \leqq j \leqq 5$, by solving the six equations for order four by means of Newton's method in quadruple precision. This of course resulted in a solution that, while being close to that provided by the minimization procedure, satisfies the order conditions to a very high precision. The coefficients are given by

$$\begin{aligned}
&\gamma_1 = 0, && b_1 = 0.061758858135626325,\\
&\gamma_2 = 0.205177661542286386, && b_2 = 0.338978026553643355,\\
&\gamma_3 = 0.608198943146500973, && b_3 = 0.614791307175577566,\\
&\gamma_4 = 0.487278066807586965, && b_4 = -0.140548014659373380,\\
&\gamma_5 = 1, && b_5 = 0.125019822794526133,
\end{aligned} \tag{2.8}$$

along with (2.3) and (2.5).

For this method the quantities in (2.7) are $A'^{(5)} = 0.00067$ and $A^{(5)} = 0.00071$. As a reference method for the numerical tests, we employ the fourth-order, FSAL, nonsymplectic formula of Dormand, El-Mikkawy, and Prince [5, Table 3]. This has four stages (three evaluations) and $A'^{(5)} = 0.0018$, $A^{(5)} = 0.00046$. Thus, per step, the reference method achieves an accuracy comparable to that of the symplectic method (2.8), but is cheaper by a factor of 3/4. In general, symplectic integrators require, for the same accuracy, more work than their nonsymplectic counterparts since, to impose symplecticness, free parameters are sacrificed that could otherwise be directed at achieving accuracy.

2.3. Error estimation. The standard way [5], [8] of estimating the errors in a pth-order RKN method (2.2) is to supplement (2.2) with formulae

$$\begin{aligned}
&\hat{\mathbf{p}}_{n+1}\mathbf{p}_n + h \sum_{i=1}^{s} \hat{b}_i \mathbf{f}(\mathbf{Q}_i),\\
&\hat{\mathbf{q}}_{n+1} = \mathbf{q}_n + h\mathbf{p}_n + h^2 \sum_{i=1}^{s} \hat{\beta}_i \mathbf{f}(\mathbf{Q}_i)
\end{aligned} \tag{2.9}$$

in such a way that $(\mathbf{p}_n, \mathbf{q}_n) \mapsto (\hat{\mathbf{p}}_{n+1}, \hat{\mathbf{q}}_{n+1})$ is an RKN method of order $q < p$ (usually $q = p-1$ or $q = p-2$). Of course, the computation of $(\hat{\mathbf{p}}_{n+1}, \hat{\mathbf{q}}_{n+1})$ employs *the same* function evaluations $\mathbf{f}(\mathbf{Q}_i)$ that are used to compute $(\mathbf{p}_{n+1}, \mathbf{q}_{n+1})$. The difference between the low-order $(\hat{\mathbf{p}}_{n+1}, \hat{\mathbf{q}}_{n+1})$ and high-order $(\mathbf{p}_{n+1}, \mathbf{q}_{n+1})$ results is then taken to be an approximation to the local error at the step $n \mapsto n+1$. For (2.8), we take the order q of the embedded method to be 3.

The weights $\hat{b}_i$, $1 \leqq i \leqq 5$, must satisfy four equations for the local error in $\hat{\mathbf{p}}_{n+1}$ to be $O(h^4)$. These equations are linear in the $\hat{b}_i$'s and it is a simple matter to express $\hat{b}_i$, $1 \leqq i \leqq 4$, in terms of $\hat{b}_5$, which remains a free parameter. The value of $\hat{b}_5$ is chosen according to a procedure suggested by Dormand and Prince. The quantities

$$C'^{(5)} = \frac{\|\mathbf{c}'^{(5)} - \hat{\mathbf{c}}'^{(5)}\|}{\|\hat{\mathbf{c}}'^{(4)}\|}, \qquad B'^{(5)} = \frac{\|\hat{\mathbf{c}}'^{(5)}\|}{\|\hat{\mathbf{c}}'^{(4)}\|} \tag{2.10}$$

should be made as small as possible (letters with a hat refer, of course, to the lower-order method). The Taylor expansion of the $\mathbf{p}$-component of the error estimator $\mathbf{p}_{n+1} - \hat{\mathbf{p}}_{n+1}$ has coefficients $-\hat{c}_j'^{(4)}$ in the order $O(h^4)$ terms and coefficients $c_j'^{(5)} - \hat{c}_j'^{(5)}$ in the order $O(h^5)$ terms (cf. (2.6a)). Thus, a small $C'^{(5)}$ ensures that, in the ($\mathbf{p}$-component of) the error estimator, the leading $O(h^4)$ term dominates over the next, $O(h^5)$, term of the Taylor expansion. This is beneficial, since the mechanism for stepsize selection assumes an $O(h^4)$ behaviour in the estimator. On the other hand, a small $B'^{(5)}$ ensures that the third-order formula used for estimation is sufficiently different from the fourth-order formula used for timestepping. (Note that as the third-order formula comes closer to the fourth-order formula, the denominator in $B'^{(5)}$ tends to 0 and hence $B'^{(5)}$ tends to infinity.) We minimize the function $\phi(\hat{b}_5) = (B'^{(5)})^2 + (C'^{(5)})^2$ by the simple procedure of evaluating ϕ at uniformly spaced values of $\hat{b}_5$ (the spacing used was 0.01). This yields $\hat{b}_5 = 0.2$.

The coefficients $\hat{\beta}_i$, $1 \leqq i \leqq 5$, are seen as free parameters, i.e., they are not derived from $\hat{b}_i$ through (2.3). For the local error in $\hat{\mathbf{q}}_{n+1}$ to be $O(h^4)$, the $\hat{\beta}_i$, $1 \leqq i \leqq 5$, must satisfy two (linear) equations; this leaves three free parameters. We arbitrarily set $\hat{\beta}_5 = 0$ and expressed $\hat{\beta}_1$ and $\hat{\beta}_2$ in terms of $\hat{\beta}_3$ and $\hat{\beta}_4$. The free $\hat{\beta}_3, \hat{\beta}_4$ are now chosen to minimize $(B^{(5)})^2 + (C^{(5)})^2$, where

$$C^{(5)} = \frac{\|\mathbf{c}^{(5)} - \hat{\mathbf{c}}^{(5)}\|}{\|\hat{\mathbf{c}}^{(4)}\|}, \qquad B^{(5)} = \frac{\|\hat{\mathbf{c}}^{(5)}\|}{\|\hat{\mathbf{c}}^{(4)}\|}. \tag{2.11}$$

The minimization was again performed by sampling the objective function on a grid with 0.01×0.01 spacing. The weights of the third-order formula (2.9) embedded in (2.5) are as follows:

$$\begin{aligned}
&\hat{b}_1 = -0.127115143890665440, && \hat{\beta}_1 = 0.110014238746029571,\\
&\hat{b}_2 = 0.698831995430764851, && \hat{\beta}_2 = 0.189985761253970428,\\
&\hat{b}_3 = 0.375269477646788521, && \hat{\beta}_3 = 0.25,\\
&\hat{b}_4 = -0.146986329186887931, && \hat{\beta}_4 = -0.05,\\
&\hat{b}_5 = 0.2, && \hat{\beta}_5 = 0.
\end{aligned} \tag{2.12}$$

With this choice the quantities in (2.10) and (2.11) are

$$C'^{(5)} = 1.06, \quad B'^{(5)} = 1.06, \quad C^{(5)} = 0.47, \quad B^{(5)} = 0.25.$$

For the fourth-order nonsymplectic scheme used as a reference method, Dormand, El-Mikkawy, and Prince [5] provide an embedded formula with

$$C'^{(5)} = 1.19, \quad B'^{(5)} = 1.20, \quad C^{(5)} = 1.02, \quad B^{(5)} = 1.03.$$

This shows that the minimizations we carried out above are as successful as those in [5].

2.4. Implementation. The embedded pair (2.8), (2.12) and the reference-embedded pair of Dormand, El-Mikkawy, and Prince were implemented in a standard way following very closely the code DOPRIN in [8].

3. Numerical results. Several test problems, including integrable and non-integrable Hamiltonians, were used. The main conclusions as to the relative merit of the various algorithms do not greatly depend on the particular test problem, and hence we only report on the results corresponding to the Newton potential [2] $V(q^1, q^2) = -1/\|\mathbf{q}\|$ with initial condition

$$p^1 = 0, \quad p^2 = \sqrt{\frac{1+e}{1-e}}, \quad q^1 = 1-e, \quad q^2 = 0.$$

Here e is a parameter $0 \leqq e < 1$. The solution is 2π-periodic and its projection onto the (configuration) $\mathbf{q}$-space is an ellipse with eccentricity e and major semiaxis 1. Initially, the moving mass is at the pericentre of the ellipse (i.e., the closest it can be to the coordinate origin). After half a period (apocentre), its distance r to the origin is $1+e$. Thus $r_{\max}/r_{\min} = (1+e)/(1-e)$, which is large for large eccentricities. Moreover, the ith derivatives of the force $\mathbf{f}$ behave like $r^{-(i+2)}$, so that, for large eccentricities, the elementary differentials of high order may vary by several orders of magnitude along the orbit. In fact, this well-known test problem with large e (say, $e = 0.9$) is often taken as a "severe test for the stepsize control procedure" of ODE algorithms [6].

The test problem was integrated by combining each of the eccentricities 0.1, 0.3, 0.5, 0.7, and 0.9 with each of the final times $10\times 2\pi$, $30\times 2\pi$, $90\times 2\pi$, $270\times 2\pi$, $810\times 2\pi$, $2430\times 2\pi$, $7290\times 2\pi$, and $21870\times 2\pi$. We were particularly interested in long time intervals, as it is in this sort of simulation that the advantages of symplecticness should be felt (see [14]). For short time intervals, the local error of the formula is of paramount importance, and it is as the time interval gets larger that advantages derived from a better qualitative behaviour become more prominent. In celestial mechanics very long time integrations are often required with potentials that are small perturbations of the two-body potential considered here.

In the tests we used the symplectic variable-step code (SV), the nonsymplectic variable-step code (NSV), and also fixed-step implementations of the symplectic formulae (SF) and nonsymplectic formulae (NSF). The variable-step codes were tried with absolute error tolerances of $10^{-4}, 10^{-5}, \ldots, 10^{-11}$, and the fixed-step algorithms were run with stepsizes $2\pi/16, 2\pi/32, \ldots, 2\pi/2048$. Errors were measured in the Euclidean norm of $\mathbb{R}^4$.

Figure 1 gives, for $e = 0.5$ and a final time of 21870 periods, the final error against the computational effort measured by the number of $\mathbf{f}$-evaluations. The figure contains information for the runs that yielded errors in the 10^{-1} to 10^{-4} range, namely,

(i) SV with tolerances 10^{-10}, 10^{-11} (plus signs joined by a dashed line);

(ii) NSV with tolerances 10^{-9}, 10^{-10}, 10^{-11} (circles joined by a solid line);

(iii) SF with timestep $2\pi/256$, $2\pi/512$, $2\pi/1024$ (stars joined by a dashed-dotted line);

(iv) NSF with timestep $2\pi/2048$ ($a \times$ sign).

Let us first compare the results of SF and NSF. Recall that these RKN formulae have error constants of roughly the same size, but SF has four evaluations per step against three evaluations per step in NSF. Thus, on *local error considerations alone*,

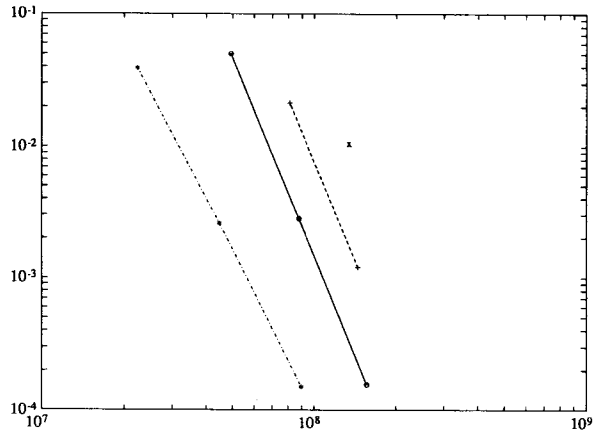

FIG 1. *Error against number of function evaluations, after* 21870 *periods,* $e = 0.5$.

we would expect that for the same global error, the numbers of evaluations of the NSF and SF would be in a ratio 3/4. On the contrary, the experimental results show that, for the same error, the symplectic formula is *four times* less expensive than the nonsymplectic process (ratio 4/1). This shows that there is something in the error propagation mechanism of the symplectic algorithm that gives it a clear advantage over its nonsymplectic counterpart. In the next section we prove rigorously that in the asymptotic expansion of the global error of the symplectic formula, the coefficient of the powers h^4, h^5, h^6, and h^7 grows linearly with the integration time t. Thus in the symplectic formula, for small errors, we need h to be small with respect to $t^{-1/4}$. On the other hand, for the nonsymplectic formula, the coefficient of the leading h^4 term of the global error also increases linearly with t, but the h^5, h^6, and h^7 terms possess coefficients that grow like t^2. If t is large, for small errors, h should be small with respect to $t^{-2/5}$. This is to be compared with $h \ll t^{-1/4}$ for the symplectic case. This shows that for large t symplecticness pays. In fact, when $e = 0.5$, SF improves on NSF if t_{final} is larger than, say, 30 periods.

Turning now to a comparison between NSF and NSV, we observe that for the formula of Dormand, El-Mikkawy, and Prince the use of variable-stepsizes results in a gain in efficiency by a factor of two. In the apocentre, the variable-stepsize code takes stepsizes about seven times as large as those it takes near the pericentre, with the result that, as expected, NSV saves on function evaluations for a given error. Note that the line which joins the NSV points has slope -5 in spite of the method having order four. This is again due to the fact that the coefficient in the leading h^4 term in the global error grows linearly with t, while the coefficients of the subsequent terms grow like t^2; for t large, $t^2h^5 \gg th^4$ and the method behaves as if its order were five (see § 3).

On the other hand, for the symplectic formula, going from fixed to variable-stepsizes results in a *decrease* in efficiency. We will return to this point later. For the present, let us note that, with variable stepsizes, the line joining the points of the symplectic algorithm are in agreement with fifth-order behaviour of the error. In fact, SV and NSV show very similar behaviour. The only difference between them lies in the fact that, for a given error, the costs of NSV and SV are in a ratio 3/4, i.e., in the ratio we would have anticipated from a study of the local errors without taking symplecticness into account.

Like Fig. 1, Fig. 2 corresponds to a final time $21870 \times 2\pi$, but now $e = 0.3$. Again we have displayed the results corresponding to runs for which the errors lie in the 10^{-1} to 10^{-4} range. These are the following:

(i) SV with tolerances 10^{-10}, 10^{-11} (plus signs joined by a dashed line);

(ii) NSV with tolerances 10^{-9}, 10^{-10} (circles joined by a solid line);

(iii) SF with timestep $2\pi/128$, $2\pi/256$, $2\pi/512$ (stars joined by a dashed-dotted line);

(iv) NSF with timestep $2\pi/1024$, $2\pi/2048$ ($\times$ sign, dotted line).

We see that the overall pattern is not changed by changing the eccentricity. The NSV and SV algorithms have efficiencies that are still in the predicted 3/4 ratio. On the other hand, with $e = 0.3$, the advantages of NSV over NSF are less marked, as we would have expected. In fact, for $e = 0.3$, both variable-step codes only vary the stepsize along the orbit by a factor of 3. The NSF points, which for $e = 0.5$ were to the right of the SV dashed line, are now exactly on this SV line.

Figure 3 corresponds to the same final time with $e = 0.7$. The following runs are represented (results for NSF are not reported for stepsizes used, as errors below 10^{-1} could not be obtained):

(i) SV with tolerances 10^{-10}, 10^{-4} (plus signs joined by a dashed line);

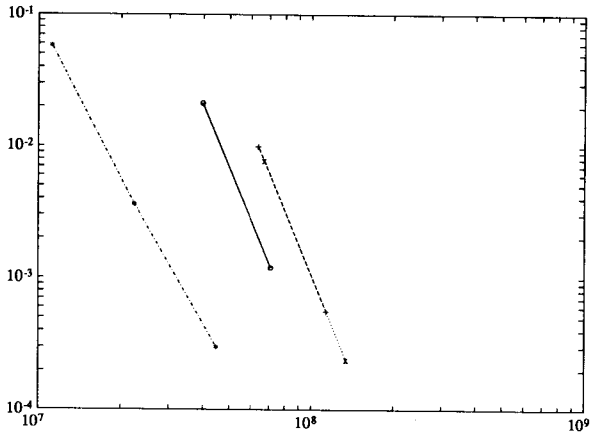

FIG. 2. *Error against number of function evaluations, after* 21870 *periods,* $e = 0.3$.

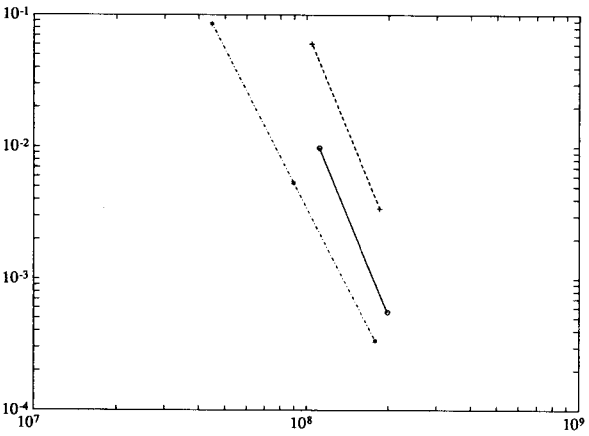

FIG. 3. *Error against number of function evaluations, after* 21870 *periods,* $e = 0.7$.

(ii) NSV with tolerances 10^{-10}, 10^{-11} (circles joined by a solid line);

(iii) SF with time step $2\pi/512, 2\pi/1024, 2\pi/2048$ (stars joined by a dashed-dotted line).

Now SV and NSV become more efficient and change h by a factor of 22. Nevertheless, SF is still the most efficient method: the advantages of symplecticness are not offset by the disadvantages of constant h.

For the smaller values of t_{final} that we tried, the picture is very much the same, except if t_{final} is not large and e is large, SF is the most efficient method, NSV is second, and SV is 4/3 times worse than NSV. For fixed t_{final}, as e approaches 1, the benefits of variable steps become more prominent and NSV improves on SF. For fixed e, as t_{final} increases, the benefits of symplecticness dominate and SF improves on NSV.

Figure 4 gives, for $e=0.5$, error against time for SV (tolerance 10^{-10}), NSV (tolerance 10^{-9}), SF ($h=2\pi/1024$), and NSF ($h=2\pi/2048$). For SF the error shows a linear behaviour with respect to t, as stated earlier. For the other methods the error grows like t^2.

Figure 5 displays, for $e=0.5$, the error in energy $|H(\mathbf{p}_n, \mathbf{q}_n) - H(\mathbf{p}(t_n), \mathbf{q}(t_n))|$ against time. Of course, the theoretical solution preserves energy $H(\mathbf{p}(t), \mathbf{q}(t)) = H(\mathbf{p}(0), \mathbf{q}(0))$ and consequently the error in energy equals the energy growth

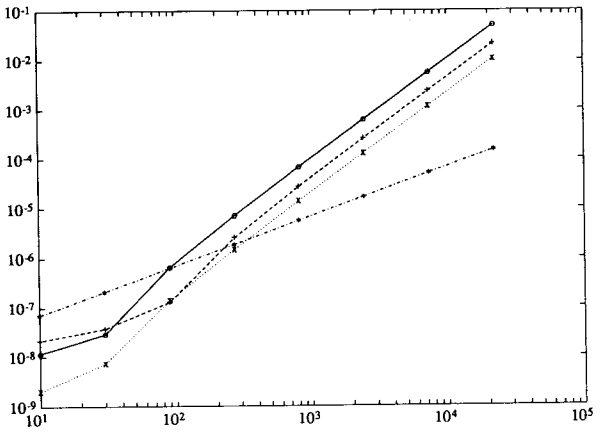

FIG. 4. *Error against time in periods, $e=0.5$.*

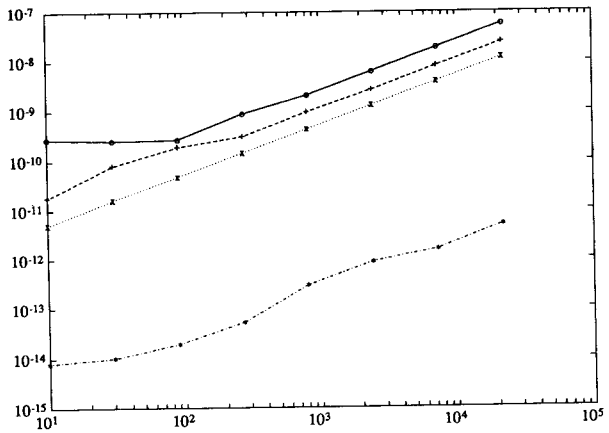

FIG. 5. *Error in energy against time in periods, $e=0.5$.*

$|H(\mathbf{p}_n, \mathbf{q}_n) - H(\mathbf{p}(0), \mathbf{q}(0))|$. The runs displayed are similar to those shown in Fig. 4. In SV, NSV, and NF the error in energy grows linearly with t. This is somewhat surprising since for SV, NSV, and NF, the errors $|(\mathbf{p}_n, \mathbf{q}_n) - (\mathbf{p}(t_n), \mathbf{q}(t_n))|$ grow like t^2. Also note that SF fails in exactly conserving energy, but its energy errors are much smaller than those associated with the remaining algorithms.

4. Integration of Kepler's problem by one-step methods. We now investigate theoretically some of the experimental findings presented above. Our analysis is not restricted to the RKN case and covers general one-step methods.

4.1. Some remarks on Kepler's problem. Let us begin by rewriting Kepler's problem in the compact form

$$\dot{\mathbf{Y}} = \mathbf{F}(\mathbf{Y}), \tag{4.1}$$

where $\mathbf{Y} = [p^1, p^2, q^1, q^2]^T$ and $\mathbf{F} = [\mathbf{f}^T, \mathbf{p}^T]^T$, with $\mathbf{f} = \mathbf{f}(\mathbf{q})$ the force (cf. (2.1)). The notation $\mathbf{G} = \mathbf{G}(\mathbf{Y})$ will be used to refer to the gradient ∇H of the Hamiltonian H with respect to $\mathbf{Y}$. Note that $\mathbf{F}$ and $\mathbf{G}$ are orthogonal at each point $\mathbf{Y}$ because H is an invariant quantity for (4.1).

We consider (4.1) in the region Ω of $\mathbf{Y}$-space covered by elliptic motions, i.e., the region where the energy H is less than 0 (so that escape to ∞ is not possible), and the angular momentum does not vanish (thus avoiding the case where the trajectory in $\mathbf{q}$-space degenerates into a straight segment). All solutions in Ω are periodic with a period

$$T = T(H) = 2\pi/\sqrt{(2|H|)^3}, \tag{4.2}$$

which only depends on the energy H. A reference for Kepler's problem is, e.g., [2, § 8E].

Let us, once and for all, fix an initial condition $\mathbf{Y}_0 \in \Omega$ and set $\mathbf{F}_0 = \mathbf{F}(\mathbf{Y}_0)$, $\mathbf{G}_0 = \mathbf{G}(\mathbf{Y}_0)$. We denote by Φ the one-period map φ_{T_0}, $T_0 = T(H(\mathbf{Y}_0))$. The analysis to follow relies heavily on the properties of the differential Φ_0' of Φ at $\mathbf{Y}_0$ (this is sometimes referred to as the monodromy operator of the periodic solution that goes through $\mathbf{Y}_0$).

LEMMA 1. *The differential Φ_0' is a rank-one modification of the identity given by*

$$\Phi_0' = I + \mathbf{W}_0 \mathbf{G}_0^T, \tag{4.3}$$

with $\mathbf{W}_0 = T'(H(\mathbf{Y}_0))\mathbf{F}_0$ a nonzero vector in $\mathbb{R}^4$ tangent at $\mathbf{Y}_0$ to the solution of Kepler's problem being investigated. Equivalently, Φ_0' is the linear operator in $\mathbb{R}^4$ such that, for vectors $\mathbf{V}$ orthogonal to $\mathbf{G}_0$,

$$\Phi_0' \mathbf{V} = \mathbf{V} \tag{4.4}$$

and

$$\Phi_0' \mathbf{G}_0 = \mathbf{G}_0 + (\mathbf{G}_0^T \mathbf{G}_0)\mathbf{W}_0. \tag{4.5}$$

Proof. We present two different proofs. The first is analytic and is due to R. D. Skeel. Set

$$\Phi(\mathbf{Y}) = \varphi_\tau(\varphi_{T(H(\mathbf{Y}))}(\mathbf{Y})) = \varphi_\tau(\mathbf{Y}),$$

where $\tau = T_0 - T(H(\mathbf{Y}))$ is a function of $\mathbf{Y}$. Then

$$\begin{aligned}\Phi'(\mathbf{Y}) = &= \varphi_\tau'(\mathbf{Y}) + \left(\frac{d}{d\tau}\varphi_\tau(\mathbf{Y})\right)(\nabla\tau)^T \\ &= \varphi_\tau'(\mathbf{Y}) - \left(\frac{d}{d\tau}\varphi_\tau(\mathbf{Y})\right) T'(H(\mathbf{Y}))\mathbf{G}(\mathbf{Y})^T,\end{aligned}$$

and, since φ_t as a function of t satisfies (4.1),

$$\Phi'(\mathbf{Y}) = \varphi_\tau'(\mathbf{Y}) - \mathbf{F}(\varphi_\tau(\mathbf{Y}))\, T'(H(\mathbf{Y}))\mathbf{G}(\mathbf{Y})^T.$$

Now evaluation at $\mathbf{Y} = \mathbf{Y}_0$ leads to $\tau = 0$ and hence to φ_τ equal to the identity map so that $\varphi_\tau'(\mathbf{Y}_0) = I$ and (4.3) follows.

The second proof is more geometric. Consider a vector $\mathbf{V}$ orthogonal to $\mathbf{G}_0$ and consider the new initial condition $\tilde{\mathbf{Y}}_0 = \mathbf{Y}_0 + \varepsilon\mathbf{V}$ with ε small. Since the increment $\varepsilon\mathbf{V}$ is orthogonal to the energy gradient $\mathbf{G}_0$, the new energy $H(\tilde{\mathbf{Y}}_0)$ equals the old energy $H(\mathbf{Y}_0)$ and the new period $T(H(\tilde{\mathbf{Y}}_0))$ equals the old period T_0 (see (2.1)). Here and later "equal" is understood to mean "equal except for $O(\varepsilon^2)$ terms." Hence $\Phi(\tilde{\mathbf{Y}}_0) = \varphi_{T_0}(\tilde{\mathbf{Y}}_0)$ equals $\tilde{\mathbf{Y}}_0$, which implies, by the definition of differential, that $\Phi_0'(\varepsilon\mathbf{V}) = \varepsilon V$. This proves (4.4). Assume now that the new initial condition is chosen to be $\tilde{\mathbf{Y}}_0 = \mathbf{Y}_0 + \varepsilon\mathbf{G}_0$. Now the new energy is in excess of $H(\mathbf{Y}_0)$ by an amount $\varepsilon(\mathbf{G}_0^T\mathbf{G}_0)$ and, accordingly, the new period is in excess from T_0 by an amount $\delta = \varepsilon T'(H(\mathbf{Y}_0))(\mathbf{G}_0^T\mathbf{G}_0)$. Hence after T_0 units of time $\mathbf{Y}$ has not had time to return to its initial position and rather lags *behind* by a vector $\delta\mathbf{F}_0$, because $\mathbf{F}_0$ is the velocity of the flow at $\mathbf{Y}_0$. This proves (4.5).

After (4.3) it is a simple matter to compute the Nth power of Φ_0'. This is given by

$$\Phi_0' = I + N\mathbf{W}_0\mathbf{G}_0^T. \tag{4.6}$$

This formula essentially says (cf. the second proof of the lemma above) that if $\tilde{\mathbf{Y}}_0$ is an initial condition of the form $\mathbf{Y}_0 + \varepsilon\mathbf{V}$, ε small, then after N time increments of length T^0, the solution $\varphi_{NT^0}(\tilde{\mathbf{Y}}_0)$ that starts at $\tilde{\mathbf{Y}}_0$ differs from the solution $\varphi_{NT^0}(\mathbf{Y}_0) = \mathbf{Y}_0$ by terms $\varepsilon\mathbf{V} + \varepsilon N(\mathbf{G}_0^T\mathbf{V})\mathbf{W}_0$. The difference grows linearly with N; this growth is in the direction of $\mathbf{W}_0$, tangent at $\mathbf{Y}_0$ to the solution curve, and furthermore only depends on the initial deviation $\varepsilon\mathbf{V}$ through its component $\varepsilon(\mathbf{G}_0^T\mathbf{V})$ in the direction of $\mathbf{G}_0$.

4.2. Basic error estimate. Let us consider a smooth one-step method ψ_h for the numerical integration of Kepler's problem. This method is assumed to be convergent of order p, i.e., $\psi_h^n(\mathbf{Y}_0) - \varphi_h^n(\mathbf{Y}_0) = O(h^p)$ as $h \to 0$ with nh in a bounded time interval. Furthermore, we assume that the differentials (Jacobian matrices) $(\psi_h^n)'(\mathbf{Y})$ also converge with order p to the differential of the flow, i.e., $(\psi_h^n)'(\mathbf{Y}) - (\varphi_h^n)'(\mathbf{Y}) = O(h^p)$, $h \to 0$, nh bounded. This is automatically satisfied by most standard methods, including Runge-Kutta and Runge-Kutta-Nyström methods.

For simplicity, we only consider the case where the steplength h is of the form T_0/ν, with ν a positive integer, and look at the difference $\mathbf{E}_N$ between the numerical $\psi_h^{\nu N}(\mathbf{Y}_0)$ and theoretical $\varphi_h^{\nu N}(\mathbf{Y}_0) = \varphi_{NT_0}(\mathbf{Y}_0) = \mathbf{Y}_0$ after N periods of the motion. The extension to general values of h and times which are not whole multiples of T_0 is possible but messy, and provides no further insight.

Set $\Psi_h = \psi_h^\nu$ so that Ψ_h is the mapping that advances the numerical solution T_0 units of time. We can then write

$$\begin{aligned}
\mathbf{E}_N &= \Psi_h^N(\mathbf{Y}_0) - \mathbf{Y}_0 = \Psi_h(\Psi_h^{N-1}(\mathbf{Y}_0)) - \Psi_h(\mathbf{Y}_0) + \mathbf{E}_1 \\
&= \Psi_h'\mathbf{E}_{N-1} + O(\|\mathbf{E}_{N-1}\|^2) + \mathbf{E}_1 \\
&= \Psi_h'\mathbf{E}_{N-1} + \mathbf{E}_1 + O(h^{2p}) \\
&= \Phi_0'\mathbf{E}_{N-1} + \mathbf{E}_1 + O(h^{2p}).
\end{aligned}$$

Here and later, the constant implied in the O symbol depends on N. By induction,

$$\mathbf{E}_N = [I + \Phi_0' + \cdots + \Phi_0'^{N-1}]\mathbf{E}_1 + O(h^{2p}).$$

We apply (4.6) to conclude the following theorem.

THEOREM 1. *Under the hypothesis above,*

$$\mathbf{E}_N = N\mathbf{E}_1 + \tfrac{1}{2}(N^2 - N)(\mathbf{G}_0^T\mathbf{E}_1)\mathbf{W}_0 + O(h^{2p}). \tag{4.7}$$

In other words, except for $O(h^{2p})$ terms, the error $\mathbf{E}_N$ after N periods of the solution have been computed grows quadratically with N. The leading N^2 growth is in the direction tangent to the solution at $\mathbf{Y}_0$, corresponding to a phase error along the trajectory. After taking the inner product of (4.7) and the energy gradient $\mathbf{G}_0$, we conclude, in view of the orthogonality of $\mathbf{G}_0$ and $\mathbf{F}_0$, that the energy error after N periods is, except for $O(h^{2p})$ terms, N times the energy error after the first period (cf. Fig. 5).

4.3. The symplectic case. In this subsection we look at the situation where ψ_h is symplectic. The key fact for the analysis [14] is that, given an arbitrarily large positive integer q, it is possible to construct a modified autonomous Hamiltonian function $\tilde{H}_h = H + O(h^p)$ such that ψ_h is consistent of order q with the Hamiltonian problem with Hamiltonian $\tilde{H}_h$, i.e., $\psi_h - \varphi_{h,\tilde{H}_h} = O(h^{q+1})$, where $\varphi_{h,\tilde{H}_h}$ is the h-flow of the problem with Hamiltonian $\tilde{H}_h$. In other words, the mapping ψ_h we are using to integrate Kepler's problem can be seen (except for $O(h^{q+1})$ terms) as the exact flow of a nearby Hamiltonian problem with Hamiltonian $\tilde{H}_h$. Here we choose $q = 2p$. The computed points $\mathbf{Y}_n$, which are $O(h^p)$ away from the solution $\mathbf{Y}(t_n)$ of Kepler's problem, are only $O(h^{2p})$ away from the solution through $\mathbf{Y}_0$ of the modified problem. In particuular,

$$\Psi_h(\mathbf{Y}_0) - \varphi_{T_0,\tilde{H}_h}(\mathbf{Y}_0) = O(h^{2p})$$

and, by implication,

$$\tilde{H}_h(\Psi_h(\mathbf{Y}_0)) - \tilde{H}_h(\varphi_{T_0,\tilde{H}_h}(\mathbf{Y}_0)) = O(h^{2p}). \tag{4.8}$$

On the other hand, $\tilde{H}_h$ is a conserved quantity for the flow $\varphi_{T_0,\tilde{H}_h}$ so that (4.8) can be rewritten

$$\tilde{H}_h(\Psi_h(\mathbf{Y}_0)) - \tilde{H}_h(\mathbf{Y}_0) = O(h^{2p}). \tag{4.9}$$

Taylor expansion of the left-hand side of (4.9) yields

$$\tilde{H}_h(\Psi_h(\mathbf{Y}_0)) - \tilde{H}_h(\mathbf{Y}_0) = \mathbf{G}_0^{h\,T}\mathbf{E}_1 + O(\|\mathbf{E}_1\|^2) = \mathbf{G}_0^{h\,T}\mathbf{E}_1 + O(h^{2p}), \tag{4.10}$$

where $\mathbf{G}_0^h$ is the gradient of $\tilde{H}_h$ at $\mathbf{Y}_0$. Comparison of (4.9) with (4.10) shows that $\mathbf{G}_0^{h\,T}\mathbf{E}_1 = O(h^{2p})$; the error after one period $\mathbf{E}_1$ is "almost" orthogonal to the gradient $\mathbf{G}_0^h$ of the modified energy $\tilde{H}_h$. Finally,

$$|\mathbf{G}_0^T\mathbf{E}_1| = |(\mathbf{G}_0 - \mathbf{G}_0^h)^T\mathbf{E}_1 + \mathbf{G}_0^{h\,T}\mathbf{E}_1| \leq \|\mathbf{G}_0 - \mathbf{G}_0^h\|\,\|\mathbf{E}_1\| + |\mathbf{G}_0^{h\,T}\mathbf{E}_1| = O(h^{2p}).$$

Here we have used the fact that the derivatives of $\tilde{H}_h$ approximate the derivatives of H to the same order, $O(h^p)$, to which $\tilde{H}_h$ approximates H; see [14]. The last bound implies that for a symplectic method the component $\mathbf{G}_0^T\mathbf{E}_1$ of the error $\mathbf{E}_1$ is $O(h^{2p})$, so that in view of (4.7), we may state (cf. the dashed-dotted lines in Figs. 4 and 5) the following theorem.

THEOREM 2. *For a constant-stepsize symplectic method,*

$$\mathbf{E}_N = N\mathbf{E}_1 + O(h^{2p}).$$

Furthermore, the energy error satisfies

$$H(\Psi_h^N(\mathbf{Y}_0)) - H(\mathbf{Y}_0) = O(h^{2p}).$$

4.4. The nonsymplectic case. In this section we explain the experimental fact that the fourth-order method NSF employed in § 3 behaves as if its order were five. We

consider a general one-step method as in § 4.3, but now assume that the order p is *even* ($p \geq 2$) and, furthermore, that ψ_h possesses some symmetries. First, we assume that

$$(\mathbf{p}, \mathbf{q}) = \psi_h(\mathbf{p}_0, \mathbf{q}_0) \Rightarrow (-\mathbf{p}, \mathbf{q}) = \psi_{-h}(-\mathbf{p}_0, \mathbf{q}_0). \tag{4.11}$$

This symmetry holds for the flow of any problem of the form (2.1). Note that according to (2.2), RKN methods obey (4.11). Second, we assume that $(\mathbf{p}, \mathbf{q}) = \psi_h(\mathbf{p}_0, \mathbf{q}_0)$ inherits from φ_t the rotational symmetry of Kepler's problem. Again, this is true for standard methods.

Let us denote by $\mathbf{M}(t)$ the coefficient of the leading $O(h^p)$ term in the asymptotic expansion of the global error of the method ψ_h. It is well known that $\mathbf{M}$ satisfies the variational equation

$$\frac{d\mathbf{M}(t)}{dt} = J(t)\mathbf{M}(t) + \mathbf{L}(t), \tag{4.12}$$

where $J(t)$ is the Jacobian of the vector field evaluated at the theoretical solution $\varphi_t(\mathbf{Y}_0)$, and $\mathbf{L}(t)$ is the coefficient of the leading $O(h^{p+1})$ term in the expansion of the local error. We need the following lemma.

LEMMA 2. *Let $dy/dt = f(y)$ be any smooth differential system with a conserved quantity H. Let*

$$\frac{dm(t)}{dt} = J(t)m(t) + l(t) \tag{4.13}$$

be the corresponding variational equation at a solution of the system. Then,

$$\frac{d}{dt}((\nabla H)^T m) = (\nabla H)^T l.$$

Proof. From (4.13),

$$\frac{d}{dt}(\nabla H)^T m = (\nabla H)^T J(t) m + (\nabla H)^T l + \left(\frac{d}{dt}\nabla H\right)^T m.$$

Differentiation with respect to y of the identity $(\nabla H)^T f \equiv 0$ and evaluation of the result at the solution of $dy/dt = f(y)$ leads to $(\nabla H)^T J + (d\nabla H/dt)^T = 0$, and the proof is complete.

The application of Lemma 2 to (4.12) reveals that

$$\frac{d}{dt}\mathbf{G}^T\mathbf{M} = \mathbf{G}^T\mathbf{L}.$$

We integrate over one period to get

$$\mathbf{G}_0^T\mathbf{M}(T_0) = \int_0^{T_0} \mathbf{G}^T\mathbf{L}\, dt. \tag{4.14}$$

Now let us express $\mathbf{G}^T\mathbf{L}$ in terms of polar coordinates r, $\dot{r}$, θ, $\dot{\theta}$. The rotational symmetry assumed above implies that $\mathbf{G}^T\mathbf{L}$ does not depend on θ. By conservation of angular momentum along the solution, $\dot{\theta}$ can be expressed in terms of r. Thus, in (4.14) the integrand is a function of r and $\dot{r}$. Furthermore, $\mathbf{G}^T\mathbf{L}$ must be *odd* in $\dot{r}$. This is because $h^{p+1}\mathbf{G}^T\mathbf{L}$ is the leading term in energy error after one step, and by (4.11) such an error remains invariant when h is changed into $-h$ and $\dot{r}$ into $-\dot{r}$, while keeping r constant. Now, as t increases from 0 to T_0, the solution takes each value of r twice with opposite values of $\dot{r}$. (This occurs when the moving point passes through points in Kepler's

ellipse that are symmetric with respect to the major axis.) Hence, the integral in (4.14) vanishes and $\mathbf{G}_0^T\mathbf{M}(T_0)=0$, i.e., $\mathbf{G}_0^T\mathbf{E}_1=O(h^{p+1})$. When this information is taken into (4.7) we see that $\mathbf{E}_N$ contains $O(Nh^p)$ and $O(N^2h^{p+1})$ terms. Assume that N is very large. The terms Nh^p, N^2h^{p+1} are of the same size when $h=1/N$, which is unrealistically small since for $h=1/N$, $Nh^p=N^2h^{p+1}=N^{-p+1}\ll 1$. Hence, for realistic choices of h, N^2h^{p+1} is much larger than Nh^p, and the method behaves as an order $p+1$ method with a large N^2 error constant.

4.5. Variable steps. Let us now study the situation for the variable-step integrators. In the experiments we only used *one* initial condition $\mathbf{Y}_0$. The choice of initial condition and tolerance determines the sequence of stepsizes $h_1, h_2, \ldots$ used in the integration. In a "thought experiment," let us imagine that even if other neighbouring initial conditions had been used, we would have still employed the same sequence $h_1, h_2, \ldots$ used for $\mathbf{Y}_0$, rather than letting the step-changing mechanism dictate the choices of stepsizes. This is actually a recommended procedure that ensures that the output of an automatic code is a smooth function of the initial data [8, § II.5]. In our context it also ensures that, if a symplectic formula ψ_h is used, then the transformation $\psi_{h_m}\cdots\psi_{h_1}$, which advances the solution from time $t=0$ to time $t_m=h_1+\cdots+h_m$, is indeed a symplectic transformation. In extending the analysis above to the variable-step experiments, we encounter some difficulties. Previously, we used the fact that we advanced the solution from time $t=0$ to time $t=NT_0$ by iterating N times an operator Ψ_h that advances the solution T_0 units of time. This is not quite true now; it is possible for T_0 not to be a steppoint t_m. But even if it is a steppoint, the sequence of stepsizes employed to go around the orbit in the second, third, ... period is likely to be slightly different from the sequence used in the first period. These difficulties will be ignored for the analysis: we assume that T_0 is a steppoint t_m, and that the sequence of stepsizes used to cover the nth period $(n-1)T_0\leqq t\leqq nT_0$ is just a duplicate of the sequence used to cover the first period $0\leqq t\leqq T_0$. These assumptions are "almost" satisfied for small tolerances; see [20], where it is rigorously shown that, essentially, variable-step algorithms employ a steplength that only depends on the current point in phase space so that, for periodic problems, stepsizes repeat themselves periodically. With our assumptions, the solution after N orbits is given by $\Psi_h^N(\mathbf{Y}_0)$, where $\Psi_h=\psi_{h_m}\cdots\psi_{h_1}$. Then the analysis in § 4.2 leading to Theorem 1 holds with h, the maximum stepsize. Furthermore, the cancellation described in § 4.4 also holds. The functions $\mathbf{M}$ and $\mathbf{L}$ still make sense [16] provided that the stepsizes satisfy

$$h_n=\gamma(t_n)h+O(h^2),$$

with γ a stepsize function; (4.12) must be replaced by [16]

$$\frac{d\mathbf{M}(t)}{dt}=J(t)\mathbf{M}(t)+\gamma(t)\mathbf{L}(t),$$

and (4.14) becomes

$$\mathbf{G}_0^T\mathbf{M}(T_0)=\int_0^{T_0}\gamma^p\mathbf{G}^T\mathbf{L}\,dt.$$

From here we conclude that $\mathbf{G}_0^T\mathbf{E}_1=O(h^{p+1})$ under the extra hypothesis that γ takes the same value as the moving body passes through points in configuration space that are mutually symmetric with respect to the major axis. This hypothesis is certainly reasonable.

On the other hand, the material in § 4.3 does not appear to be extensible to the variable-step situation. Indeed, the experiments indicate that it *cannot* be extended. When trying to extend the analysis in § 4.3 to variable steps, we encounter the difficulty that the modified Hamiltonian $\tilde{H}_h$ depends on the steplength. The computed $(\mathbf{p}_1, \mathbf{q}_1)$ is close to the solution S_1 of the system with Hamiltonian $\tilde{H}_{h_1}$ which at $t=0$ passes through $(\mathbf{p}_0, \mathbf{q}_0)$. The computed $(\mathbf{p}_2, \mathbf{q}_2)$ is close to the solution S_2 of the system associated with $\tilde{H}_{h_2}$ that at $t=h_1$ passes through $(\mathbf{p}_1, \mathbf{q}_1)$, etc. Clearly, $S_1 \neq S_2$ (unless $\tilde{H}_{h_1} = \tilde{H}_{h_2}$) and we do not have a single trajectory near which the computed points stay. There is not a single pseudoenergy $\tilde{H}_h$ "almost" conserved by the numerical points, and nothing can be said of the projection $\mathbf{G}_0^T \mathbf{E}_1$, whose smallness is the key to the success of the symplectic constant-stepsize integrators.

4.6. Remarks and extensions. The fact that for Kepler's problem standard integrators lead to quadratic error growth while symplectic constant-stepsize integrators lead to linear error growth has been noted before in the literature; see Kinoshita, Yoshida, and Nakay [10]. In [22], Yoshida studies the energy error in the symplectic integration of Kepler's problem. His analysis is only formal and, like ours, resorts to a modified Hamiltonian $\tilde{H}_h$. Yoshida assumes that the modified Hamiltonian can be chosen to satisfy

$$\psi_h - \phi_{h,\tilde{H}_h} = 0, \tag{4.15}$$

i.e., that a modified Hamiltonian problem exists so that the computed points exactly solve the modified problem. However, it is known that for nonlinear problems, while it is possible to construct a divergent formal power series for $\tilde{H}_h$ fulfilling (4.15), no actual function $\tilde{H}_h$ can satisfy (4.15). Therefore, the analysis in [22] is only of heuristic value. (Note that rather than (4.15), we assumed only that $\psi_h - \phi_{h,\tilde{H}_h} = O(h^{2p})$.)

On the other hand, the ideas used in the analysis in this section are not restricted to Kepler's problem. For instance, in §§ 4.2–4.4, it is enough to assume that all solutions of the problem being integrated are periodic with a period that only depends on the value of the energy H (and actually changes with H). These assumptions are satisfied by all nonlinear one-degree-of-freedom oscillators, such as the well-known pendulum equation. Therefore, the conclusions in §§ 4.2–4.4 hold for such oscillators. This proves Conjecture 3 in § IV.6 D of Stoffer's thesis [18], which states that for nonlinear oscillators, standard methods have quadratic error growth, and that symplectic methods produce errors that only grow linearly. (Nonlinearity is essential to guarantee a nontrivial dependence of the period on the energy, leading to $\mathbf{W}_0 \neq \mathbf{0}$ in (4.6).)

5. Conclusions. Let us summarize our findings.

(i) The experiments with Kepler's problem reported above and experiments with other Hamiltonian problems (not reported in this paper) reveal that constant-stepsize symplectic integrators can be more efficient than variable-step codes. This provides motivation for the further study of symplectic integration. Comparisons between symplectic and nonsymplectic formulae presented so far in the literature (see [14] for references) have concentrated on constant stepsizes. Our experiments indicate that it is reasonable to expect that, in the future, symplectic software can be developed which outperforms, on Hamiltonian problems, standard variable-step codes. The paper by Herbst and Ablowitz [9], written after the present work was completed, provides a dramatic example of a simple symplectic algorithm outperforming NAG library software.

(ii) The advantages of using symplectic formulae are lost when these formulae are used in a variable-stepsize environment. This came as a surprise to us. However,

after completing this work, we discovered that in 1988, Stoffer [19] had argued that symplectic integrators should not be used with variable stepsizes. His argument is as follows. Integrating a system of ODEs $d\mathbf{y}/dt = \mathbf{f}(\mathbf{y})$ with a variable-step algorithm is "equivalent" [20] to integrating, with constant stepsizes, a transformed problem $d\mathbf{y}/d\tau = r(\mathbf{y})\mathbf{f}(\mathbf{y})$, where the new time τ is related to the old time by $dt/d\tau = r(\mathbf{y})$. The transformed system is not Hamiltonian, even if the original system is, so that the advantages of symplecticness are lost in the transformation. An alternative argument to justify the failure of variable-step symplectic algorithms has been put forward in [14]. A key property of symplectic formulae ψ_h for a Hamiltonian problem with Hamiltonian H is the existence of a modified Hamiltonian $\tilde{H}_h$ in such a way that ψ_h "almost" coincides with the h-flow $\phi_{h,\tilde{H}_h}$ of $\tilde{H}_h$. "Almost" means that, given any large integer q, $\tilde{H}_h$ can be found in such a way that $\psi_h - \phi_{h,\tilde{H}_h} = O(h^{q+1})$. With constant stepsizes, it is possible to interpret the error in a "backward" way: a numerically calculated solution corresponding to H is "almost" an exact solution of a neighbouring Hamiltonian $\tilde{H}_h$. In § 4.5 we saw how such a backward-error analysis interpretation fails in a variable-stepsize situation. An additional reference useful in connection with variable steps for symplectic integrators is [17].

(iii) For the particular cases of Kepler's problem and nonlinear one-degree-of-freedom oscillators, a complete analysis has been presented of the performance of symplectic and nonsymplectic integrators. It has been shown that the advantages of symplectic integrators include not only better qualitative behaviour, but also better quantitative properties in the error growth mechanism.

REFERENCES

[1] L. ABIA AND J. M. SANZ-SERNA, *Partitioned Runge-Kutta methods for separable Hamiltonian problems*, Math. Comput., 1993, to appear.

[2] V. I. ARNOLD, *Mathematical Methods of Classical Mechanics*, 2nd ed., Springer-Verlag, New York, 1989.

[3] M. P. CALVO AND J. M. SANZ-SERNA, *Order conditions for canonical Runge-Kutta-Nyström methods*, BIT, 3 (1992), pp. 131-142.

[4] P. J. CHANNELL AND C. SCOVEL, *Symplectic integration of Hamiltonian systems*, Nonlinearity, 3 (1990), pp. 231-259.

[5] J. R. DORMAND, M. E. A. EL-MIKKAWY, AND P. J. PRINCE, *Families of Runge-Kutta-Nyström formulae*, IMA J. Numer. Anal., 7 (1987), pp. 235-250.

[6] J. R. DORMAND AND P. J. PRINCE, *A family of embedded Runge-Kutta formulae*, J. Comput. Appl. Math., 6 (1980), pp. 19-268.

[7] K. FENG, *Difference schemes for Hamiltonian formalism and symplectic geometry*, J. Comput. Math., 4 (1986), pp. 279-289.

[8] E. HAIRER, S. P. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations* I, *Nonstiff Problems*, Springer-Verlag, Berlin, 1987.

[9] B. M. HERBST AND M. J. ABLOWITZ, *Numerical homoclinic instabilities in the sine-Gordon equation*, Quaestiones Mathematicae, 15 (1992), pp. 345-363.

[10] H. KINOSHITA, H. YOSHIDA, AND H. NAKAY, *Symplectic integrators and their application to dynamical astronomy*, Celestial Mech., 50 (1991), pp. 59-71.

[11] D. OKUNBOR AND R. D. SKEEL, *An explicit Runge-Kutta-Nyström method is canonical if and only if its adjoint is explicit*, SIAM J. Numer. Anal., 19 (1992), pp. 521-527.

[12] R. RUTH, *A canonical integration technique*, IEEE Trans. Nuclear Sci., 30 (1983), pp. 2669-2671.

[13] J. M. SANZ-SERNA, *Runge-Kutta schemes for Hamiltonian systems*, BIT, 28 (1988), pp. 877-883.

[14] ———, *Symplectic Integrators for Hamiltonian problems: An overview*, Acta Numerica, 1 (1992), pp. 243-286.

[15] J. M. SANZ-SERNA AND L. ABIA, *Order conditions for canonical Runge-Kutta schemes*, SIAM J. Numer. Anal., 28 (1991), pp. 1081-1096.

[16] L. F. SHAMPINE, *The step sizes used by one-step codes for* ODE's, Appl. Numer. Math., 1 (1985), pp. 95-106.

[17] R. D. SKEEL AND C. W. GEAR, *Does variable step size ruin a symplectic integrator?*, Phys. D, 60 (1992), pp. 311-313.

[18] D. M. STOFFER, *Some geometric and numerical methods for perturbed integrable systems*, Ph.D. thesis, Eidgenoessische Technische Hochschule (ETH), Zürich, 1988.

[19] ———, *On reversible and canonical integration methods*, Res. Rep. No. 88-05, Applied Mathematics, Eidgenoessische Technische Hochschule (ETH) Zürich, 1988.

[20] D. M. STOFFER AND K. NIPP, *Invariant curves for variable step size integrators*, BIT, 31 (1991), pp. 169-180.

[21] Y. B. SURIS, *Canonical transformations generated by methods of Runge-Kutta type for the numerical integration of the system* $x'' = -\partial U/\partial x$, Zh. Vychisl, Mat. i Mat. Fiz., 29 (1989), pp. 202-211. (In Russian.)

[22] H. YOSHIDA, *Conserved quantities of symplectic integrators for Hamiltonian systems*, preprint.