Numerische Mathematik manuscript No.
(will be inserted by the editor)

Symmetric multistep methods over long times

Ernst Hairer[1], **Christian Lubich**[2]

[1] Dept. de Mathématiques, Univ. de Genève, CH-1211 Genève 24, Switzerland, e-mail: `Ernst.Hairer@math.unige.ch`
[2] Mathematisches Institut, Univ. Tübingen, D-72076 Tübingen, Germany, e-mail: `Lubich@na.uni-tuebingen.de`

Received: April 2003 / Revised version: date

Summary For computations of planetary motions with special linear multistep methods an excellent long-time behaviour is reported in the literature, without a theoretical explanation. Neither the total energy nor the angular momentum exhibit secular error terms. In this paper we completely explain this behaviour by studying the modified equation of these methods and by analyzing the remarkably stable propagation of parasitic solution components.

Keywords. Linear multistep method, Hamiltonian system, N-body system, energy conservation, conservation of angular momentum, symplecticity, invariant tori, linear error growth, backward error analysis, modified differential equation, modulated Fourier expansion.

Mathematics Subject Classification (1991): 65L06, 65L80

1 Introduction

We are concerned with the long-time integration of second order ordinary differential equations

$$M\ddot{q} = -\nabla U(q), \qquad q(0) = q_0, \quad \dot{q}(0) = v_0, \tag{1.1}$$

with a potential $U(q)$ and a positive definite mass matrix M. Typical examples are N-body problems such as those arising in astronomy or in molecular dynamics.

As numerical integrator we consider linear multistep methods for second order differential equations $\ddot{q} = f(q)$ (for $f(q) = -M^{-1}\nabla U(q)$). They are given by a formula of the form

$$\sum_{i=0}^{k} \alpha_i \, q_{n+i} = h^2 \sum_{i=0}^{k} \beta_i \, f(q_{n+i}). \tag{1.2}$$

The simplest but very important special case is

$$q_{n+1} - 2\, q_n + q_{n-1} = h^2 f(q_n), \tag{1.3}$$

which, nowadays, is called the Störmer–Verlet method. Explicit methods of the form (1.2), where the left hand expression is the same as for (1.3), have first been considered by Störmer [18] for computations concerning the aurora borealis. A general convergence theory has been developed by Dahlquist [3], see also Henrici [14, Chapter 6] and Hairer, Nørsett and Wanner [13, Section III.10]. Let us briefly recall some important facts.

It is usual to denote the generating polynomials of the coefficients of the linear multistep method (1.2) by

$$\rho(\zeta) = \sum_{i=0}^{k} \alpha_i \, \zeta^i, \qquad \sigma(\zeta) = \sum_{i=0}^{k} \beta_i \, \zeta^i. \tag{1.4}$$

We assume throughout this article that $\rho(\zeta)$ and $\sigma(\zeta)$ have no common zeros. Method (1.2) is *stable* if all zeros of $\rho(\zeta)$ satisfy $|\zeta| \le 1$, and if the zeros of modulus one have multiplicity not exceeding two. It is of *order* p if the coefficients are such that

$$\frac{\rho(\zeta)}{(\log \zeta)^2} - \sigma(\zeta) = \mathcal{O}((\zeta - 1)^p) \qquad \text{for} \quad \zeta \to 1. \tag{1.5}$$

In particular, 1 must be a double root of $\rho(\zeta)$. Stability and order $p \ge 1$ imply convergence of the numerical method, more precisely, the global error satisfies the estimate (for $t = nh$)

$$\|q_n - q(t)\| \le C_1 (h + t) e^{\omega t} \delta + C_2 t^2 e^{\omega t} h^p, \tag{1.6}$$

where C_1, C_2 are generic constants, ω is proportional to the square root of the Lipschitz constant of $f(q)$, and the starting approximations are assumed to satisfy $q_j - q(jh) = \mathcal{O}(h\delta)$ for $j = 0, \ldots, k-1$.

The methods of Störmer have $\rho(\zeta) = (\zeta - 1)^2 \zeta^{k-2}$ and the polynomial $\sigma(\zeta)$ of degree $k-1$ (hence $\beta_k = 0$) is determined such that (1.5) holds with $p = k$.

It is proved by Dahlquist [3] that the order of a stable multistep method (1.2) cannot exceed $k+2$ (first Dahlquist barrier), and that stable methods of maximal order $p = k + 2$ have even k and are *symmetric*, i.e., they satisfy

$$\alpha_i = \alpha_{k-i}, \quad \beta_i = \beta_{k-i} \quad \text{for all} \quad i. \tag{1.7}$$

For stable symmetric multistep methods all roots of $\rho(\zeta)$ are on the unit circle, and the order p is even. Dahlquist considers the application of such methods to the test equation $\ddot{q} = aq$ and notices the following [3, p. 43f.]:

"Suppose that ζ_j is a *simple* root of unit modulus. Then, the corresponding root of $\rho(\zeta) - ah^2\sigma(\zeta) = 0$ is of the form $\zeta_{jh} = \zeta_j(1 + \mathcal{O}(h^2))$, whence $|\zeta_{jh}^n| = (1 + \mathcal{O}(h^2))^n \sim 1$ ($h \to 0$, $nh = x$), and hence there is no weak instability. If ζ_j is a *double* root, however, then $|\zeta_{jh}^n|$ may, asymptotically, have an exponential growth."

After Dahlquist's work, symmetric multistep methods did not receive much attention over many years. Lambert and Watson [15] took up again this investigation. They found that only for symmetric methods the numerical solution can remain close to a periodic orbit of the linear test equation, and they noted that methods without multiple roots of $\rho(\zeta)$ other than 1 have this property for sufficiently small step size. Only with the article of Quinlan and Tremaine [17], where an excellent performance of symmetric multistep methods for simulations of the outer solar system is reported, the research on the long-time behaviour of these methods for nonlinear problems started. We mention the papers of Tang [19] and of Hairer and Leone [9], where the non-symplecticity of these methods is shown, and the work of Cano and Sanz-Serna [2], where the linear error growth for problems with periodic solution is studied. A lot of attention is paid to symmetric multistep methods in the astronomical literature, e.g., Fukushima [5, 6] and Evans and Tremaine [4].

2 Main results and numerical observations

Our results concern the long-time behaviour of symmetric linear multistep methods (1.2) of order $p \geq 2$. As a stronger condition than mere stability, we shall need the following crucial property throughout (cf. the above citation of Dahlquist):

Definition 1 *A symmetric multistep method (1.2) is called s-stable if, apart from the double root at 1, all zeros of $\rho(\zeta)$ are simple and of modulus one (the letter "s" stands for "simple roots").*

We remark that k is always even for symmetric methods. Otherwise they would be reducible, because (1.7) implies $\rho(-1) = \sigma(-1) =$

0 for odd k. Furthermore, -1 cannot be a root of $\rho(\zeta)$, because complex roots appear as pairs.

The multistep method (1.2) is complemented with a difference formula for approximations of the velocity:

$$v_n = \frac{1}{h} \sum_{j=-l}^{l} \delta_j \, q_{n+j}. \tag{2.1}$$

The v_n are computed *a posteriori* and do not enter the propagation of the numerical solution. We assume that this difference formula is also of order p, that is, it gives the exact derivative for polynomials up to order p.

Instead of the velocities we often consider the momenta $p = Mv$ (no confusion with the order p shall arise), and we set

$$p_n = M v_n. \tag{2.2}$$

To start the multistep method, starting values $q_0, q_1, \ldots, q_{k-1}$ are needed. We assume that their errors are $\mathcal{O}(h^{p+1})$, as they would be if they are obtained from a pth order one-step method:

$$q_j - q(jh) = \mathcal{O}(h^{p+1}) \qquad \text{for} \qquad j = 0, 1, \ldots, k-1. \tag{2.3}$$

Finally we assume that the numerical solution values q_n stay in a fixed compact subset of the domain on which the potential $U(q)$ is smooth, and that the velocity approximations v_n are bounded by a constant. In view of Theorem 1 below, this is for example satisfied if the level sets $\{q : U(q) \le \mu\}$ are compact. The above assumptions are made throughout this section without further mention.

2.1 Energy conservation

The total energy

$$H(q,p) = \frac{1}{2} p^T M^{-1} p + U(q) \tag{2.4}$$

is conserved along solutions of the differential equation (1.1). One way of seeing this is by multiplying the differential equation by $\dot q^T$: $0 = \dot q^T M \ddot q + \dot q^T \nabla U(q) = (d/dt)(\frac{1}{2} \dot q^T M \dot q + U(q)) = (d/dt) H(q,p)$. A related, though more elaborate argument will later be used for showing that the total energy is nearly preserved over very long times along numerical solutions.

Theorem 1 *The total energy is conserved up to $\mathcal{O}(h^p)$ over times $\mathcal{O}(h^{-p-2})$ along numerical solutions obtained by the s-stable symmetric multistep method:*

$$H(q_n, p_n) = H(q_0, p_0) + \mathcal{O}(h^p) \qquad \textit{for} \quad nh \le h^{-p-2}.$$

The constant symbolized by $\mathcal{O}$ is independent of n, h with $nh \le h^{-p-2}$.

Remark 1 The time scales in Theorem 1 and in Theorem 2 below can be further extended if either non-resonance conditions on the roots of $\rho(\zeta)$ are satisfied or if the starting approximations are carefully computed:

- If no root of $\rho(\zeta)$ other than 1 can be written as the product of two other roots, then the conservation up to $\mathcal{O}(h^p)$ holds even over times $\mathcal{O}(h^{-2p-3})$.
- If the starting values are computed such that the numerical solution is "smooth", i.e., the values $z_\ell(0)$ of Lemma 1 below are very small, say of size $\mathcal{O}(h^s)$ with $s > p+1$, the time scales are further increased.

For symplectic one-step methods it is known that the total energy is preserved up to $\mathcal{O}(h^p)$ on exponentially long time intervals $nh \le Ce^{c/h}$ [1]. However, the time scales of Theorem 1 and Remark 1 are already long enough for practical computations. In contrast to the result for one-step methods, symplecticity plays no role in the proof of Theorem 1.

Example 1 For our numerical experiment we consider the Kepler problem which is of the form $\ddot{q} = -\nabla U(q)$ with

$$U(q_1, q_2) = -(q_1^2 + q_2^2)^{-1/2}.$$

We choose initial values $q_1(0) = 1 - e$, $q_2(0) = 0$, $\dot{q}_1(0) = 0$, $\dot{q}_2(0) = \sqrt{(1+e)/(1-e)}$, such that the solution is an ellipse with eccentricity $e = 0.2$, and we apply the following three symmetric methods with constant step size $h = 0.04$ on an interval of length $2\pi \cdot 10^5$:

$$\begin{aligned} &(\mathrm{A}) \qquad \rho(\zeta) = (\zeta - 1)^2 \zeta^6 && (\text{Störmer}) \\ &(\mathrm{B}) \qquad \rho(\zeta) = (\zeta^4 - 1)^2 && \\ &(\mathrm{C}) \qquad \rho(\zeta) = (\zeta - 1)(\zeta^7 - 1) && (\texttt{gni\_lmm2}) \end{aligned} \tag{2.5}$$

and the polynomial $\sigma(\zeta)$ of degree 7 is defined by (1.5) with $p = 8$. All these methods are stable and of order 8, the methods (B) and (C) are symmetric, but only the method (C) is s-stable. Fortran and

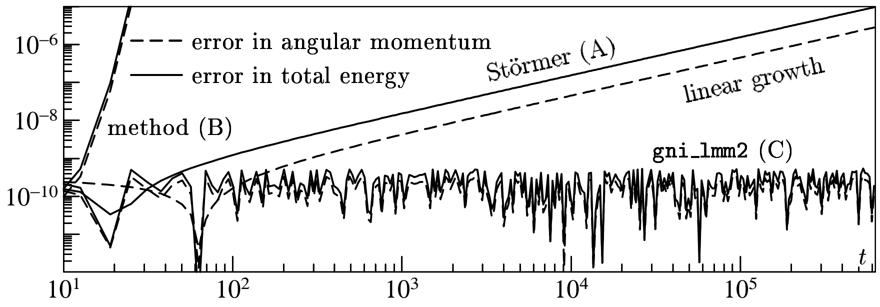

Fig. 1. Energy and angular momentum conservation of the three linear multistep methods given in (2.5)

Matlab versions of the code `gni_lmm2` can be downloaded from the Internet at http://www.unige.ch/math/folks/hairer/ (see also [8]).

The error in the total energy is plotted for all three methods in Fig. 1. In agreement with Theorem 1, the error of method (C) remains bounded of size $\mathcal{O}(h^8)$ on the whole interval. The error of the symmetric method (B), which has double roots of $\rho(\zeta) = 0$ different from 1, shows an exponential error growth which agrees with the classical error estimate (1.6). The non-symmetric method (A) shows an error behaviour of the form $\mathcal{O}(h^8) + \mathcal{O}(th^9)$.

For all methods, the error in the angular momentum behaves in the same way as that for the total energy. This is in contrast to symplectic one-step methods which exactly conserve quadratic first integrals.

2.2 Conservation of angular momentum

N-body systems with rotational symmetry preserve the total angular momentum. More generally, the invariance property

$$U(e^{\tau A} q) = U(q) \qquad \text{for all} \quad \tau, q \tag{2.6}$$

with a matrix A such that AM^{-1} is skew-symmetric, implies, as a special case of Noether's theorem, that the differential equation has the first integral

$$L(q,p) = p^T A q. \tag{2.7}$$

Theorem 2 *Quadratic first integrals of the form (2.7) are conserved up to $\mathcal{O}(h^p)$ over times $\mathcal{O}(h^{-p-2})$ along numerical solutions obtained by the s-stable symmetric multistep method:*

$$L(q_n, p_n) = L(q_0, p_0) + \mathcal{O}(h^p) \qquad \textit{for} \quad nh \le h^{-p-2}.$$

The constant symbolized by $\mathcal{O}$ is independent of n, h with $nh \le h^{-p-2}$.

2.3 Integrable systems: linear error growth and near-invariant tori

The differential equation (1.1) written as

$$\dot{q} = v \ , \qquad \dot{v} = -M^{-1}\nabla U(q) \tag{2.8}$$

is a *reversible system* in the sense that inverting the direction of the initial velocity does not change the solution trajectory, but it inverts the direction of motion. The flow φ_t thus satisfies that

$$\varphi_t(q,v) = (\widehat{q},\widehat{v}) \qquad \text{implies} \qquad (q,-v) = \varphi_t(\widehat{q},-\widehat{v}).$$

The system (2.8) is an *integrable reversible system* if there exists a transformation

$$(q,v) = \psi(a,\theta) \tag{2.9}$$

to *action-angle variables* (a,θ), defined for actions $a = (a_1,\ldots,a_d)$ in some open set of $\mathbf{R}^d$ and for angles $\theta = (\theta_1,\ldots,\theta_d)$ on the whole torus $\mathbf{T}^d = \mathbf{R}^d/(2\pi\mathbf{Z}^d) = \{(\theta_1,\ldots,\theta_d) : \theta_i \in \mathbf{R} \bmod 2\pi\}$, such that the transformation preserves reversibility, that is,

$$(q,v) = \psi(a,\theta) \qquad \text{implies} \qquad (q,-v) = \psi(a,-\theta) \ ,$$

and the system (2.8) is transformed to the form

$$\dot{a} = 0 \ , \qquad \dot{\theta} = \omega(a) \tag{2.10}$$

with *frequencies* $\omega = (\omega_1,\ldots,\omega_d)$. For every a, the torus $\{(a,\theta) : \theta \in \mathbf{T}^d\}$ is thus invariant under the flow. We write the inverse transform of (2.9) as

$$(a,\theta) = (I(q,v), \Theta(q,v))$$

and note that the components of $I = (I_1,\ldots,I_d)$ are first integrals of the system (2.8).

The effect of a perturbation of an integrable reversible system is well under control in subsets of the phase space where the frequencies satisfy the *diophantine condition*

$$|k\cdot\omega| \geq \gamma |k|^{-\nu} \qquad \text{for all} \ \ k \in \mathbf{Z}^d \tag{2.11}$$

for some positive constants γ and ν; see, e.g., [11, Ch. XI], [16].

The following result shows linear error growth and near-preservation of invariant tori over long times.

Theorem 3 *Consider applying the s-stable symmetric multistep method to an integrable reversible system (2.8) with real-analytic potential U. Suppose that $\omega^* \in \mathbf{R}^d$ satisfies the diophantine condition (2.11). Then, there exist positive constants C, c and h_0 such that the following holds for all step sizes $h \le h_0$: every numerical solution (q_n, v_n) starting with frequencies $\omega_0 = \omega(I(q_0, v_0))$ such that $\|\omega_0 - \omega^*\| \le c|\log h|^{-\nu-1}$, satisfies*

$$\begin{aligned} \|(q_n, v_n) - (q(t), v(t))\| &\le C\, t\, h^p \\ \|I(q_n, v_n) - I(q_0, v_0)\| &\le C\, h^p \end{aligned} \qquad \textit{for} \quad 0 \le t = nh \le h^{-p}.$$

The constants h_0, c, C depend on d, γ, ν and on bounds of the potential.

Example 2 We consider the Kepler problem with initial data as in Example 1 and we apply the three methods of (2.5). Figure 2 shows their global error as a function of time. In agreement with Theorem 3, method (C) shows a linear error growth. For the strictly stable Störmer method (A), we would expect a quadratic error growth proportional to h^p. We observe, however, a growth like $\mathcal{O}(th^8) + \mathcal{O}(t^2h^9)$. This can be explained with the results of Section 3 below: the dominant term of the local error is, up to a constant factor, the same for all multistep methods of order eight. Consequently, the error will be a superposition of that of a symmetric method of order 8 with that of a non-symmetric method of order 9. The exponential error growth of method (B) is the behaviour of classical estimates like that of (1.6).

Notice that the estimates of Theorem 3 are confirmed for the Kepler problem, although this problem does not satisfy the diophantine condition (2.11), because here the two frequencies are identical.

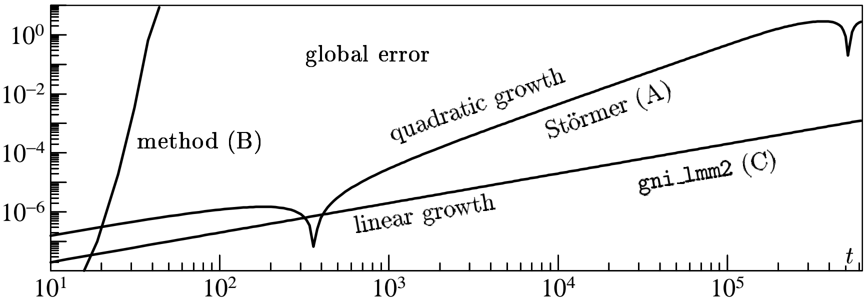

Fig. 2. Global error of the three linear multistep methods given in (2.5) applied to the Kepler problem

Remark 2 The linear error growth and the long-time near-preservation of tori remain valid if the s-stable symmetric multistep method is applied to a perturbed integrable system

$$M\ddot{q} = -\nabla U_0(q) - \epsilon \nabla U_1(q)$$

with integrable $M\ddot{q} = -\nabla U_0(q)$ and $\epsilon = \mathcal{O}(h^\alpha)$ for some $\alpha > 0$ (cf. [11, p. 354]).

Example 3 (Symplecticity) We consider the derivatives of $q(t)$ and $p(t) = M\dot{q}(t)$ with respect to the initial values (q_0, p_0),

$$dq(t) = \frac{\partial q(t)}{\partial(q_0, p_0)} \ , \qquad dp(t) = \frac{\partial p(t)}{\partial(q_0, p_0)} = M d\dot{q}(t) \ ,$$

which are the solution of the variational equation

$$M d\ddot{q} = -\nabla^2 U(q)\, dq. \tag{2.12}$$

The flow of the differential equation (1.1) is *symplectic*, that is, the matrix-valued function

$$S(dq, dp) = dq^T dp - dp^T dq \tag{2.13}$$

is conserved: $S(dq(t), dp(t)) = S(dq(0), dp(0))$ for all t.

For the numerical solution, we assume that the starting values $q_0, \ldots, q_{k-1}$ are given by a one-step method, so that (q_n, p_n) can be considered as a function of (q_0, p_0). We denote by dq_n and dp_n the derivative matrices of q_n and p_n with respect to (q_0, p_0). They are obtained by applying the multistep method to the system (1.1) augmented by the variational equation (2.12), which is of the form $\ddot{Q} = F(Q)$ (with $Q = (q, dq)$) but no longer Hamiltonian.

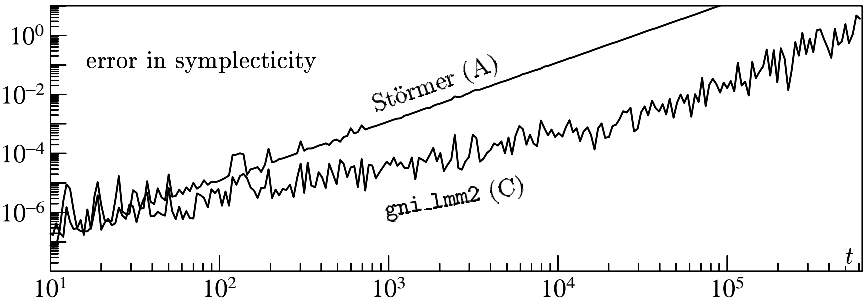

Fig. 3. Error in the symplecticity of the linear multistep methods (A) and (C) given in (2.5) applied to the Kepler problem

As in Example 1 we consider the Kepler problem and the methods of (2.5). Figure 3 shows the Frobenius norm of the error $S(dq_n, dp_n) - S(dq_0, dp_0)$ as a function of time $t = nh$. For the Störmer method we observe quadratic error growth, for the s-stable symmetric multistep method there is a linear growth for a long time which turns finally into a quadratic growth.

We remark that this error behaviour corresponds to the linear growth of the derivatives $d\theta(t)$ of the angle variables. For non-integrable systems with positive Lyapunov exponents we expect the error in the symplecticity to grow exponentially for both methods.

Although the entries of $S(dq, dp)$ are quadratic first integrals of the augmented system, Theorem 2 does not apply because on the one hand the derivatives dq_n and dp_n do not remain bounded, and on the other hand the augmented system is not Hamiltonian.

3 Backward error analysis for smooth numerical solutions

In this section we study the exceptional case of numerical solutions (q_n) for which

$$q_n = y(nh) + \mathcal{O}(h^N) \qquad \text{for a } \textit{smooth} \text{ function } y(t), \tag{3.1}$$

where $N \gg p$ and smoothness is understood to mean that all derivatives of $y(t)$ are bounded independently of h. (Strictly speaking, this refers to families of functions $y(t)$ parametrized by h.) The situation (3.1) is met only for very special starting values, whereas general numerical solutions contain oscillatory terms which correspond to powers of the roots of $\rho(\zeta)$ other than 1 and of their products (parasitic solution components). Nevertheless, the idealized situation of no parasitic terms gives already much insight into the conservation properties of the method, in a technically simpler framework than the general case.

For the remainder of the paper it is convenient to assume that the mass matrix is the identity matrix, $M = I$. This causes no loss of generality, since the substitution $M^{1/2}q \to q$ changes M to I. The multistep method is invariant under this linear transformation.

3.1 Modified differential equation

Smooth functions $y(t)$ with (3.1) satisfy a modified second-order differential equation.

Theorem 4 *There exist unique h-independent functions $f_j(q,v)$ such that, for every truncation index N, every solution of*

$$\ddot{y} = f(y) + hf_1(y,\dot{y}) + \ldots + h^{N-1} f_{N-1}(y,\dot{y}) \tag{3.2}$$

satisfies

$$\sum_{i=0}^{k} \alpha_i y(t+ih) = h^2 \sum_{i=0}^{k} \beta_i f(y(t+ih)) + \mathcal{O}(h^{N+2}). \tag{3.3}$$

If the linear multistep method is of order p, then $f_j = 0$ for $j < p$. If the method is symmetric, then $f_j = 0$ for all odd j, and $f_j(q,-v) = f_j(q,v)$ for all even j so that the flow of (3.2) is reversible.

Proof We denote by D time differentiation and correspondingly by e^{hD} the shift operator. The equation (3.3) can be written as

$$\rho(e^{hD})y = h^2 \sigma(e^{hD}) f(y) + \mathcal{O}(h^{N+2}).$$

With the expansion $x^2\sigma(e^x)/\rho(e^x) = 1 + \mu_1 x + \mu_2 x^2 + \ldots$ this becomes equivalent to

$$\ddot{y} = (1 + \mu_1 hD + \mu_2 h^2 D^2 + \ldots) f(y) + \mathcal{O}(h^N) \tag{3.4}$$

provided that $y(t)$ is a smooth function in the sense specified above. Now, $Df(y) = f'(y)\dot{y}$, which gives us $f_1(q,v) = f'(q)v$. We express the second derivative of y in $D^2 f(y) = f''(y)(\dot{y},\dot{y}) + f'(y)\ddot{y}$ again by the differential equation (3.4) to obtain a formula for f_2. Continuing in this way for the higher time derivatives and collecting equal powers of h determines recursively the functions $f_3, f_4, \ldots$.

If the method is of order p, then $\mu_j = 0$ for $j < p$. If the method is symmetric, then $\mu_j = 0$ for all odd j. This implies the result. □

The defect of a solution $y(t)$ of the truncated modified differential equation (3.2) is of size $\mathcal{O}(h^{N+2})$, whereas that of a solution $q(t)$ of $\ddot{q} = f(q)$ is $\mathcal{O}(h^{p+2})$. Consequently, the classical convergence proof (with $q(t)$ replaced by $y(t)$) yields the following result: if the multistep method is stable and of order p, then for every truncation index N and for $t = nh$ we have

$$\|q_n - y(t)\| \leq C_1(h+t)e^{\omega t}\delta + C_N t^2 e^{\omega t} h^N, \tag{3.5}$$

where ω is proportional to the square root of the Lipschitz constant of $f(q)$, and δ is such that the starting approximations satisfy $q_j - y(jh) = \mathcal{O}(h\delta)$ for $j = 0, \ldots, k-1$. Compared to (1.6), we have improved the second term in the error estimate.

3.2 Modified energy

In the case of a symmetric multistep method, the modified differential equation (3.2) for $f(q) = -\nabla U(q)$ has a formal first integral close to the total energy $H(q,p) = \frac{1}{2}p^T p + U(q)$.

Proposition 1 *For a symmetric multistep method of order p, there exists a formal modified energy*

$$\widetilde{H}(q,p) = H(q,p) + h^p H_p(q,p) + h^{p+2} H_{p+2}(q,p) + \ldots$$

such that its truncation at the $\mathcal{O}(h^N)$ term satisfies

$$\frac{d}{dt}\,\widetilde{H}(y(t),\dot{y}(t)) = \mathcal{O}(h^N)$$

along solutions of the modified differential equation (3.2).

We remark that Theorem 4 and Proposition 1 imply, for smooth numerical solutions (3.1) and their pth order momentum approximations (2.1),

$$H(q_n,p_n) = H(q_0,p_0) + \mathcal{O}(h^p) + \mathcal{O}(th^N).$$

Proof The proof is based on the ideas of the second proof of long-time energy conservation of the Störmer-Verlet method in [12], which uses only the symmetry of the method. Similar to the previous proof, with the expansion $\rho(e^x)/(x^2\sigma(e^x)) = (1 + \gamma_p x^p + \gamma_{p+2}x^{p+2} + \ldots)$, we write the equation (3.3) as

$$(1 + \gamma_p h^p D^p + \gamma_{p+2} h^{p+2} D^{p+2} + \ldots)\,\ddot{y} = -\nabla U(y) + \mathcal{O}(h^N), \quad (3.6)$$

where we note that the left-hand side contains only *even*-order derivatives of y thanks to the symmetry of the method. We multiply both sides of (3.6) with $\dot{y}^T$, so that on the right-hand side we have the total derivative $(d/dt)U(y)$. On the left-hand side we note $\dot{y}^T\ddot{y} = \frac{1}{2}\frac{d}{dt}(\dot{y}^T\dot{y})$ and similarly for higher even-order derivatives

$$\dot{y}^T y^{(2m)} = \frac{d}{dt}\Big(\dot{y}^T y^{(2m-1)} - \ddot{y}^T y^{(2m-2)} + \ldots \pm \frac{1}{2}\, y^{(m)T} y^{(m)}\Big). \quad (3.7)$$

On the left-hand side we thus obtain the time derivative of an expression in which the appearing second and higher derivatives of y can be substituted as functions of $(y,\dot{y})$ via the modified differential equation (3.2). Putting this together, the equation (3.6) multiplied by $\dot{y}^T$ becomes of the form

$$\frac{d}{dt}\Big(\frac{1}{2}\,\dot{y}^T\dot{y} + h^p H_p(y,\dot{y}) + h^{p+2} H_{p+2}(y,\dot{y}) + \ldots\Big) = -\frac{d}{dt}U(y) + \mathcal{O}(h^N),$$

which is the stated result. $\square$

3.3 Modified angular momentum and quadratic first integrals

Whenever we have a quadratic first integral of the form (2.7), for example the total angular momentum in N-body systems, then the modified differential equation has a formal first integral close to it.

Proposition 2 *Suppose that $\ddot{q} = f(q)$ has $L(\dot{q}, q) = \dot{q}^T A q$ as first integral, i.e., A is a skew-symmetric matrix and $f(q)^T A q = 0$. For a symmetric multistep method of order p, there then exists a formal modified first integral*

$$\widetilde{L}(q,p) = p^T A q + h^p L_p(q,p) + h^{p+2} L_{p+2}(q,p) + \ldots$$

such that its truncation at the $\mathcal{O}(h^N)$ term satisfies

$$\frac{d}{dt}\,\widetilde{L}(y(t), \dot{y}(t)) = \mathcal{O}(h^N)$$

along solutions of the modified equation (3.2).

Together with Theorem 4 this implies

$$p_n^T A q_n = p_0^T A q_0 + \mathcal{O}(h^p) + \mathcal{O}(t h^N)$$

for smooth numerical solutions (3.1) with (2.1).

Proof The proof is very similar to the preceding proof. We now take the inner product of (3.6) with Ay. By assumption we have $f(y)^T A y = 0$. Since A is skew-symmetric, we have $\ddot{y}^T A y = \frac{d}{dt}(\dot{y}^T A y)$ and similarly for the higher even-order derivatives

$$y^{(2m)T} A y = \frac{d}{dt}\Big(y^{(2m-1)T} A y - y^{(2m-2)T} A \dot{y} + \ldots \pm y^{(m)T} A y^{(m-1)} \Big).$$

Hence the left-hand side becomes a total derivative, and the right-hand side is of size $\mathcal{O}(h^N)$. Expressing the higher derivatives of y as functions of $(y, \dot{y})$ via the modified differential equation then gives the result. □

3.4 Integrable systems

If the differential equation $\dot{q} = v, \dot{v} = f(q)$ is an integrable reversible system, then we can use reversible perturbation theory to study the behaviour of solutions of the reversible modified differential equation (3.2). In particular, Lemma XI.2.1 of [11] (used as in the proof of Theorem X.3.1) yields the following.

Proposition 3 *Under the conditions of Theorem 3, the solution of the modified differential equation (3.2) of a symmetric multistep method of order p, starting with* (q_0, v_0)*, satisfies*

$$\begin{aligned} \|(y(t),\dot y(t)) - (q(t),\dot q(t))\| &\le C\,t\,h^p \\ \|I(y(t),\dot y(t)) - I(q_0,v_0)\| &\le C\,h^p \end{aligned} \qquad \textit{for} \quad 0 \le t = nh \le h^{-p}.$$

This yields the estimates of Theorem 3 for smooth numerical solutions.

4 Backward error analysis for general numerical solutions, part I

We now consider general numerical solutions obtained by the symmetric multistep method. We derive the modified equations for the principal and the parasitic solution components, study their Hamiltonian-like structure and derive long-term bounds for the parasitic solution components.

4.1 Principal and parasitic modified equations

The results in this subsection are analogues of results in [7] concerning multistep methods for first order differential equations. Here we consider general second order problems $\ddot q = f(q)$ and we assume that $f(q)$ is real-analytic in the considered region.

Let $\zeta_0 = 1$ be the double root of the characteristic polynomial $\rho(\zeta)$ and $\zeta_{\pm 1}, \ldots, \zeta_{\pm(k/2-1)}$ the simple roots on the unit circle, ordered such that $\zeta_{-\ell} = \overline{\zeta}_\ell$. We enumerate the set of all possible products of roots,

$$\{\zeta_\ell\}_{\ell\in\mathcal{I}} = \Big\{ \prod_{|j|<k/2} \zeta_j^{m_j} \;\Big|\; m_j \text{ integer} \Big\},$$

again with $\zeta_{-\ell} = \overline{\zeta}_\ell$. The set of subscripts $\mathcal{I}$ can be finite or infinite. We let $\mathcal{I}^* = \mathcal{I} \setminus \{0\}$.

We aim at writing general solutions q_n of the multistep method (with $q_n - q_{n-1} = \mathcal{O}(h)$) in the form

$$q_n = y(nh) + \sum_{\ell\in\mathcal{I}^*} \zeta_\ell^n \, z_\ell(nh)$$

where $y(t)$ and $z_\ell(t)$ are *smooth* functions (that is again, with all derivatives bounded independently of h). The *principal* solution component $y(t)$ satisfies a second order differential equation close to (1.1) and the *parasitic* components $z_\ell(t)$ for $\ell \neq 0$ are small and are determined by first order differential equations for ℓ corresponding to a root ζ_ℓ of $\rho(\zeta)$, and by algebraic equations otherwise.

The following result extends Theorem 4 in giving the system of modified differential equations for both the principal and parasitic components.

Theorem 5 *Consider an s-stable symmetric multistep method (1.2). There exist unique h-independent functions $f_{\ell,j}(q,v,\mathbf{z}^*)$ with $\mathbf{z}^* := (z_\ell)_{0<|\ell|<k/2}$ such that, for every truncation index N and for every solution of*

$$\begin{aligned}\ddot{y} &= f_{0,0}(y,\dot{y},\mathbf{z}^*) + hf_{0,1}(y,\dot{y},\mathbf{z}^*) + \ldots + h^{N-1}f_{0,N-1}(y,\dot{y},\mathbf{z}^*)\\ \dot{z}_\ell &= hf_{\ell,1}(y,\dot{y},\mathbf{z}^*) + \ldots + h^N f_{\ell,N}(y,\dot{y},\mathbf{z}^*) \quad \textit{if } \rho(\zeta_\ell)=0 \qquad (4.1)\\ z_\ell &= h^2 f_{\ell,2}(y,\dot{y},\mathbf{z}^*) + \ldots + h^{N+1}f_{\ell,N+1}(y,\dot{y},\mathbf{z}^*) \quad \textit{else,}\end{aligned}$$

the function

$$x(t) = y(t) + \sum_{\ell\in\mathcal{I}^*} \zeta_\ell^{t/h}\, z_\ell(t) \qquad (4.2)$$

satisfies

$$\sum_{i=0}^k \alpha_i\, x(t+ih) = h^2 \sum_{i=0}^k \beta_i f(x(t+ih)) + \mathcal{O}(h^{N+2}). \qquad (4.3)$$

For $\mathbf{z}^*=0$ the functions $f_{0,j}(y,\dot{y},0)$ are identical to those of Theorem 4. In particular, $f_{0,j}(y,\dot{y},0)=0$ for $0<j<p$, if the method is of order p. Moreover, the solutions of (4.1) satisfy $z_{-\ell}(t)=\overline{z_\ell}(t)$ for all $\ell\in I$ if this relation holds for the initial values.

Proof We insert (4.2) into (4.3) and note, with $z_0(t)=y(t)$,

$$\begin{aligned}\sum_{i=0}^k \alpha_i\, x(t+ih) &= \sum_{i=0}^k \alpha_i \sum_{\ell\in\mathcal{I}} \zeta_\ell^{(t+ih)/h}\, e^{ihD} z_\ell(t)\\ &= \sum_{\ell\in\mathcal{I}} \zeta_\ell^{t/h} \sum_{i=0}^k \alpha_i \zeta_\ell^i e^{ihD} z_\ell(t) = \sum_{\ell\in\mathcal{I}} \zeta_\ell^{t/h} \rho(\zeta_\ell e^{hD}) z_\ell(t).\end{aligned}$$

We expand $f(x(t))$ into a Taylor series around $y(t)$,

$$\begin{aligned}f(x(t)) &= \sum_{m\geq 0} \frac{1}{m!} f^{(m)}(y(t))\Big(\sum_{\ell_1\in\mathcal{I}^*} \zeta_{\ell_1}^{t/h} z_{\ell_1}(t),\ldots,\sum_{\ell_m\in\mathcal{I}^*} \zeta_{\ell_m}^{t/h} z_{\ell_m}(t)\Big)\\ &= \sum_{\ell\in\mathcal{I}} \zeta_\ell^{t/h} \sum_{m\geq 0} \frac{1}{m!} \sum_{\zeta_{\ell_1}\cdots\zeta_{\ell_m}=\zeta_\ell} f^{(m)}(y(t))(z_{\ell_1}(t),\ldots,z_{\ell_m}(t))\ .\end{aligned}$$

This gives, as above,

$$\sum_{i=0}^{k} \beta_i f(x(t+ih)) = \sum_{\ell\in\mathcal{I}} \zeta_\ell^{t/h} \sigma(\zeta_\ell e^{hD}) \sum_{m\ge 0} \frac{1}{m!} \sum_{\zeta_{\ell_1}\cdots\zeta_{\ell_m}=\zeta_\ell} f^{(m)}(y(t))(z_{\ell_1}(t),\dots,z_{\ell_m}(t)).$$

Comparing coefficients of $\zeta_\ell^{t/h}$ we obtain for all $\ell \in \mathcal{I}$

$$\rho(\zeta_\ell e^{hD}) z_\ell = h^2 \sigma(\zeta_\ell e^{hD}) \sum_{m\ge 0} \frac{1}{m!} \sum_{\zeta_{\ell_1}\cdots\zeta_{\ell_m}=\zeta_\ell} f^{(m)}(y)(z_{\ell_1},\dots,z_{\ell_m}) \tag{4.4}$$

(for $\ell = 0$ and $m = 0$ the sum is understood to include the term $f(y)$). With the expansion $x^\kappa \sigma(\zeta_\ell e^x)/\rho(\zeta_\ell e^x) = \mu_{\ell,0} + \mu_{\ell,1} x + \mu_{\ell,2} x^2 + \dots$ with $\mu_{\ell,0} = \sigma(\zeta_\ell)\kappa!/(\zeta_\ell^\kappa \rho^{(\kappa)}(\zeta_\ell))$ for a κ-fold zero ζ_ℓ of $\rho(\zeta)$ ($\kappa = 2$ for $\ell = 0$, $\kappa = 1$ for $0 < |\ell| < k/2$, and $\kappa = 0$ else), this equation becomes

$$z_\ell^{(\kappa)} = h^{2-\kappa} \Big(\mu_{\ell,0} + \mu_{\ell,1} hD + \mu_{\ell,2} h^2 D^2 + \dots \Big) \sum_{m\ge 0} \frac{1}{m!} \sum_{\zeta_{\ell_1}\cdots\zeta_{\ell_m}=\zeta_\ell} f^{(m)}(y)(z_{\ell_1},\dots,z_{\ell_m}). \tag{4.5}$$

Like in the proof of Theorem 4, this leads to the differential equations (4.1). □

We note that all functions $f_{\ell,j}$ are sums of infinitely many terms. For example,

$$f_{0,0}(y,\dot y,\mathbf{z}^*) = f(y) + \sum_{m\ge 2} \frac{1}{m!} \sum_{\zeta_{\ell_1}\cdots\zeta_{\ell_m}=1} f^{(m)}(y)(z_{\ell_1},\dots,z_{\ell_m})$$

where the second sum is over ℓ_j satisfying $0 < \ell_j < k/2$. As explained in detail in [7], these sums converge absolutely whenever y remains in a compact set K, $f(y)$ is analytic in a neighbourhood of K, and the $\|z_\ell\|$ are bounded by a sufficiently small (h-independent) constant. This constant is proportional to the distance of K and the boundary of the set where f is analytic.

It is possible to avoid infinite sums in the coefficient functions of the modified equation, if we work with variables w_ℓ defined by $z_\ell = h w_\ell$. This leads to a truncated system similar to (4.1) with functions depending on $y, \dot y$ and $\mathbf{w}^* := (w_\ell)_{0<|\ell|<k/2}$, which now are finite sums. Moreover, only finitely many w_ℓ are non-zero.

Thanks to the assumption that all roots of $\rho(\zeta)$ other than 1 are simple, the differential equations for the z_ℓ corresponding to the parasitic roots are *first order* differential equations, with the additional bonus of a factor h on the right-hand side:

$$\dot z_\ell = h\frac{\sigma(\zeta_\ell)}{\zeta_\ell\rho'(\zeta_\ell)}\, f'(y)z_\ell + \text{higher order terms.}$$

Here, "higher order terms" means that they contain either an additional factor h or an additional factor z_j.

Lemma 1 *Consider an s-stable symmetric multistep method (1.2). To every set of starting values $q_0,\ldots,q_{k-1}$ satisfying $q_j - q(jh) = \mathcal{O}(h^s)$ $(j = 0,\ldots,k-1)$ with $1 \le s \le p+2$ there exist (locally) unique initial values $y(0)$, $h\dot y(0)$, $z_\ell(0)$ $(0 < |\ell| < k/2)$ for the system (4.1) such that its solution exactly satisfies*

$$q_j = y(jh) + \sum_{\ell\in\mathcal{I}^*} \zeta_\ell^j z_\ell(jh) \qquad \textit{for}\ \ j = 0,\ldots,k-1. \tag{4.6}$$

These initial values satisfy $z_{-\ell}(0) = \overline{z_\ell}(0)$ and

$$y(0) - q(0) = \mathcal{O}(h^s), \quad h\dot y(0) - h\dot q(0) = \mathcal{O}(h^s), \quad z_\ell(0) = \mathcal{O}(h^s). \tag{4.7}$$

Proof We rewrite (4.6) as

$$\begin{aligned} y(0) + jh\dot y(0) + \sum_{0<|\ell|<k/2} \zeta_\ell^j z_\ell(0) &= q_j + \Big(y(0) + jh\dot y(0) - y(jh)\Big) \\ &+ \sum_{0<|\ell|<k/2} \zeta_\ell^j \Big(z_\ell(0) - z_\ell(jh)\Big) - \sum_{|\ell|\ge k/2} \zeta_\ell^j z_\ell(jh) \end{aligned}$$

with $y(t)$ and $z_\ell(t)$ the solutions of (4.1) for initial values $y(0)$, $h\dot y(0)$, $z_\ell(0)$ for $0 < |\ell| < k/2$. This defines a convergent fixed-point iteration for the initial values, with a contraction factor of $\mathcal{O}(h)$ (after solving the confluent Vandermonde system arising on the left-hand side). If we start the iteration with $(q(0), h\dot q(0), 0,\ldots,0)$, then the first increment is of size $\mathcal{O}(h^s)$, and consequently (4.7) holds. $\square$

If we replace the exact solution $q(t)$ by $y(t) + \sum_{\ell\in\mathcal{I}^*} \zeta_\ell^{t/h} z_\ell(t)$ of Theorem 5 in the classical convergence proof, then we get for s-stable symmetric methods (1.2) that (for $t = nh$)

$$q_n = y(t) + \sum_{\ell\in\mathcal{I}^*} \zeta_\ell^n z_\ell(t) + \mathcal{O}(t^2 e^{\omega t} h^N), \tag{4.8}$$

where ω is proportional to the square root of the Lipschitz constant of $f(q)$. Compared to (3.5) this gives a precise description of the propagation of perturbations in the starting approximations.

4.2 Hamiltonian of the full modified system

The key to proving long-time estimates for the symmetric multistep method is the observation that much of the Hamiltonian structure of the differential equation $\ddot{q} = -\nabla U(q)$ is conserved in the modified equations (4.1). The results and techniques of this subsection are closely related to those of [11, Sect. XIII.6.3] and [10, Sect. 4.2] developed for numerical methods for oscillatory differential equations.

We define $\mathcal{U}(\mathbf{z})$ for $\mathbf{z} = (z_\ell)_{\ell\in\mathcal{I}}$ as

$$\mathcal{U}(\mathbf{z}) = U(z_0) + \sum_{m\geq 1} \frac{1}{m!} \sum_{\zeta_{\ell_1}\cdots\zeta_{\ell_m}=1} U^{(m)}(z_0)(z_{\ell_1},\ldots,z_{\ell_m}), \tag{4.9}$$

where the second sum is over all indices $\ell_1 \in \mathcal{I}^*,\ldots,\ell_m \in \mathcal{I}^*$ (that is, $\ell_j \neq 0$) with $\zeta_{\ell_1}\ldots\zeta_{\ell_m} = 1$, and the first sum actually starts with $m = 2$. With this notation it follows from (4.4) with $f(q) = -\nabla U(q)$ that every solution of the truncated modified equation (4.1) satisfies

$$\rho(\zeta_\ell e^{hD}) z_\ell = -h^2 \sigma(\zeta_\ell e^{hD}) \nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z}) + \mathcal{O}(h^{N+2}) \tag{4.10}$$

(for all $\ell \in \mathcal{I}$) as long as

$$y \in K, \quad \|\dot{y}\| \leq M, \quad \|z_\ell\| \leq \delta \ \text{ for } \ 0 < |\ell| < k/2, \tag{4.11}$$

where K is a compact subset of the domain of analyticity of $U(q)$, $M > 0$ some bound on the derivative, and $\delta > 0$ is a sufficiently small constant (note that this implies $\|z_\ell\| \leq \delta$ for all $\ell \in \mathcal{I}^*$ if the third relation of (4.1) is satisfied and if h is sufficiently small).

For ease of presentation, we assume for the moment that $\sigma(\zeta_\ell) \neq 0$ for all $\ell \in \mathcal{I}$ (in any case we know that this holds for $|\ell| < k/2$, that is, for the roots ζ_ℓ of $\rho(\zeta)$). We apply the operator $\sigma^{-1}(\zeta_\ell e^{hD})$ to both sides of (4.10) and divide by h^2:

$$h^{-2}\Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{hD}) z_\ell = -\nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z}) + \mathcal{O}(h^N). \tag{4.12}$$

We multiply with $\dot{z}_{-\ell}^T$ and sum over all ℓ. This gives

$$h^{-2} \sum_{\ell\in\mathcal{I}} \dot{z}_{-\ell}^T \Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{hD}) z_\ell + \frac{d}{dt}\mathcal{U}(\mathbf{z}) = \mathcal{O}(h^N). \tag{4.13}$$

We now show that also the first expression on the left-hand side is a total derivative of a function depending on $\mathbf{z}$ and its time derivatives. For this we note that

$$\Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{ix}) = \sum_{j\geq 0} c_{\ell,j} x^j \qquad \text{with } \textit{real} \text{ coefficients } \ c_{\ell,j} = (-1)^j c_{-\ell,j}. \tag{4.14}$$

This holds because the symmetry of the multistep method yields $(\rho/\sigma)(1/\zeta) = (\rho/\sigma)(\zeta)$ and hence, for real x,

$$\Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{ix}) = \Big(\frac{\rho}{\sigma}\Big)(\overline{\zeta_\ell e^{ix}}) = \overline{\Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{ix})}.$$

With this expansion we obtain

$$\Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{hD}) z_\ell = \sum_{j=0}^{N+1} c_{\ell,j} (-ih)^j z_\ell^{(j)} + \mathcal{O}(h^{N+2}). \tag{4.15}$$

On the other hand, we have the relations

$$\dot y^T y^{(2m)} \;=\; \frac{d}{dt}\Big(\dot y^T y^{(2m-1)} - \ddot y^T y^{(2m-2)} + \ldots \pm \frac{1}{2}\,(y^{(m)})^T y^{(m)}\Big)$$

for the real function $y = z_0$ and for z_ℓ corresponding to $\zeta_\ell = -1$, while for the complex-valued functions $z = z_\ell$, with complex conjugate $\overline{z} = z_{-\ell}$, we have

$$\begin{aligned}
\operatorname{Re} \dot{\overline{z}}^T z^{(2m)} \;&=\; \operatorname{Re} \frac{d}{dt}\Big(\dot{\overline{z}}^T z^{(2m-1)} - \ddot{\overline{z}}^T z^{(2m-2)} + \ldots \pm \frac{1}{2}\,(\overline{z}^{(m)})^T z^{(m)}\Big)\\
\operatorname{Im} \dot{\overline{z}}^T z^{(2m+1)} \;&=\; \operatorname{Im} \frac{d}{dt}\Big(\dot{\overline{z}}^T z^{(2m)} - \ddot{\overline{z}}^T z^{(2m-1)} + \ldots \mp (\overline{z}^{(m)})^T z^{(m+1)}\Big).
\end{aligned}$$

Together with (4.15) these relations show that the terms

$$\begin{aligned}
&\dot z_{-\ell}^T \Big(\frac{\rho}{\sigma}\Big)(\zeta_\ell e^{hD}) z_\ell + \dot z_\ell^T \Big(\frac{\rho}{\sigma}\Big)(\zeta_{-\ell} e^{hD}) z_{-\ell}\\
&\qquad = \sum_{j=0}^{N+1} c_{\ell,j}\, 2\operatorname{Re}\Big((-ih)^j\, \dot{\overline{z}}_\ell^{\,T} z_\ell^{(j)}\Big) + \mathcal{O}(h^{N+2})
\end{aligned}$$

give a total derivative (up to the remainder term). Hence the left-hand side of (4.13) can be written as the time derivative of a function which depends on z_ℓ, $\ell \in \mathcal{I}$, and on their derivatives. Using the modified equation (4.1) we eliminate all z_ℓ corresponding to ζ_ℓ with $\rho(\zeta_\ell) \neq 0$ and their derivatives, the first and higher derivatives of z_ℓ (for $0 < |\ell| < k/2$), and the second and higher derivatives of $y = z_0$. We thus get a function

$$\mathcal{H}(y, \dot y, \mathbf{z}^*) = H_0(y, \dot y, \mathbf{z}^*) + \ldots + h^{N-1} H_{N-1}(y, \dot y, \mathbf{z}^*) \tag{4.16}$$

such that

$$\frac{d}{dt}\mathcal{H}(y(t), \dot y(t), \mathbf{z}^*(t)) = \mathcal{O}(h^N), \tag{4.17}$$

along solutions of (4.1) that stay in a set defined by (4.11). The function $\mathcal{H}$ is therefore an almost-invariant of the system (4.1).

If, however, $\sigma(\zeta)$ does have a zero ζ_ℓ, then we omit the corresponding term from the sum in (4.13). Hence the term $\dot z_{-\ell}^T \nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z})$ is missing from $(d/dt)\mathcal{U}(\mathbf{z})$ and must therefore be compensated in the remainder term. Since ζ_ℓ is a product of no fewer than two zeros of $\rho(\zeta)$, it follows from (4.5) with $\kappa = 0$ and from $\mu_{\ell,0} = 0$ that $z_\ell = \mathcal{O}(h^3\delta^2)$, as long as $\|z_j\| \le \delta$ for $0 < |j| < k/2$. We further have $\nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z}) = \mathcal{O}(\delta^2)$, so that the remainder term in (4.17) is augmented by $\mathcal{O}(h^3\delta^4)$.

We summarize the above considerations as follows.

Theorem 6 *Every solution of the truncated modified equation (4.1) satisfies, with $\mathcal{H}$ from (4.16),*

$$\mathcal{H}(y(t), \dot y(t), \mathbf{z}^*(t)) = \mathcal{H}(y(0), \dot y(0), \mathbf{z}^*(0)) + \mathcal{O}(th^N) + \mathcal{O}(th^3\delta^4) \tag{4.18}$$

as long as the solution stays in the set defined by (4.11). Moreover,

$$\mathcal{H}(y, \dot y, \mathbf{z}^*) = H(y, \dot y) + \mathcal{O}(h^p) + \mathcal{O}(h\delta^2). \tag{4.19}$$

The closeness to the Hamiltonian $H(y, \dot y) = \frac{1}{2}\|\dot y\|^2 + U(y)$ follows also directly from the above construction. For $\mathbf{z}^* = 0$ we have $\mathcal{H}(y, \dot y, 0) = \widetilde{H}(y, \dot y)$, where $\widetilde{H}$ is the modified energy from Proposition 1.

We will use Theorem 6 in Section 5 to infer the long-time near-conservation of the Hamiltonian along numerical solutions. Before that we need to bound the parasitic components.

4.3 Long-time bounds for parasitic solution components

The modified equations have further almost-invariants which are close to the squares of the norms of the parasitic components that correspond to the roots of $\rho(\zeta)$. We derive them here and use them to show that all parasitic solution components remain small over very long times. The techniques used in this subsection are similar to those in [11, Sects. XIII.6 and XIII.7].

We consider ℓ with $0 < |\ell| < k/2$ for which ζ_ℓ is a *simple* root of $\rho(\zeta)$ and $\sigma(\zeta_\ell) \neq 0$. The dominant term on the left-hand side of (4.12) is $-c_{\ell,1} i h^{-1} \dot z_\ell$. Since

$$\frac{d}{dt}\|z_\ell\|^2 = z_{-\ell}^T \dot z_\ell + z_\ell^T \dot z_{-\ell}, \tag{4.20}$$

we multiply (4.12) with $z_{-\ell}^T$ and the equation for $-\ell$ with z_ℓ^T and form the difference, so that the dominant term on the left-hand side becomes $-c_{\ell,1} i h^{-1} \frac{d}{dt}\|z_\ell\|^2$ (note $c_{-\ell,1} = -c_{\ell,1}$). Dividing by $-c_{\ell,1} i h^{-1}$ gives

$$\begin{aligned} \frac{i}{c_{\ell,1} h}\Big(z_{-\ell}^T \frac{\rho}{\sigma}(\zeta_\ell e^{hD}) z_\ell - z_\ell^T \frac{\rho}{\sigma}(\zeta_{-\ell} e^{hD}) z_{-\ell} \Big) \\ = \frac{ih}{c_{\ell,1}} \Big(-z_{-\ell}^T \nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z}) + z_\ell^T \nabla_{z_\ell} \mathcal{U}(\mathbf{z}) \Big). \end{aligned} \tag{4.21}$$

We first estimate the right-hand expression. Since

$$\nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z}) = \nabla^2 U(z_0) z_\ell + \mathcal{O}(\delta^2),$$

as long as (4.11) is satisfied, we obtain from the symmetry of the Hessian that the right-hand side of (4.21) is of size $\mathcal{O}(h\delta^3)$. The dominant $\mathcal{O}(h\delta^3)$ term is present only if $\zeta_{-\ell}$ can be written as the product of two roots of $\rho(\zeta)$ other than 1. If this is not the case, the expression (4.21) is of size $\mathcal{O}(h\delta^4)$.

Using the expansion (4.15) on the left-hand side of (4.21) and the relations (for $z = z_\ell$)

$$\begin{aligned} \mathrm{Re}\, \overline{z}^T z^{(2m+1)} &= \mathrm{Re}\, \frac{d}{dt}\Big(\overline{z}^T z^{(2m)} - \dot{\overline{z}}^T z^{(2m-1)} \ldots \mp \frac{1}{2} (\overline{z}^{(m)})^T z^{(m)} \Big) \\ \mathrm{Im}\, \overline{z}^T z^{(2m+2)} &= \mathrm{Im}\, \frac{d}{dt}\Big(\overline{z}^T z^{(2m+1)} - \dot{\overline{z}}^T z^{(2m)} + \ldots \pm (\overline{z}^{(m)})^T z^{(m+1)} \Big) \end{aligned}$$

we obtain that (4.21) is, up to $\mathcal{O}(h^N)$, the total derivative of a function depending on $\mathbf{z}$ and its derivatives.

By construction the dominant term is $\frac{d}{dt}\|z_\ell\|^2$. The following terms have at least one more power of h and at least one derivative which by (4.1) gives rise to an additional factor h. Eliminating higher derivatives with the help of (4.1), we arrive at a function of the form

$$\mathcal{K}_\ell(y, \dot{y}, \mathbf{z}^*) = \|z_\ell\|^2 + h^2 K_{\ell,2}(y, \dot{y}, \mathbf{z}^*) + \ldots + h^{N-1} K_{\ell,N-1}(y, \dot{y}, \mathbf{z}^*). \tag{4.22}$$

As we have seen, its total derivative is of size $\mathcal{O}(h\delta^3)$ or smaller. We summarize these considerations in the following theorem.

Theorem 7 *Along every solution of the truncated modified equation (4.1) the function $\mathcal{K}_\ell(y, \dot{y}, \mathbf{z}^*)$ satisfies for $0 < |\ell| < k/2$*

$$\mathcal{K}_\ell(y(t), \dot{y}(t), \mathbf{z}^*(t)) = \mathcal{K}_\ell(y(0), \dot{y}(0), \mathbf{z}^*(0)) + \mathcal{O}(th^N) + \mathcal{O}(th\delta^3) \tag{4.23}$$

as long as the solution stays in the set defined by (4.11). The second error term is replaced by $\mathcal{O}(th\delta^4)$ if no root of $\rho(\zeta)$ other than 1 is the product of two other roots. Moreover,

$$\mathcal{K}_\ell(y, \dot{y}, \mathbf{z}^*) = \|z_\ell\|^2 + \mathcal{O}(h^2\delta^2). \tag{4.24}$$

This result does not yet directly give information about the numerical solution, since the remainder term in (4.8) can still grow exponentially in time. Nevertheless, it allows us to write the numerical solution in a form that is suitable for deriving long-time error estimates. Let us first collect the necessary assumptions:

(A1) the multistep method (1.2) is symmetric, s-stable, of order p;
(A2) the potential function $U(q)$ of (1.1) is defined and analytic in an open neighbourhood of a compact set K;
(A3) the starting approximations $q_0, \ldots, q_{k-1}$ are such that the initial values for (4.1) obtained from Lemma 1 satisfy $y(0) \in K$, $\|\dot{y}(0)\| \le M$, and $\|z_\ell(0)\| \le \delta/2$ for $0 < |\ell| < k/2$;
(A4) the numerical solution $\{q_n\}$ stays for $0 \le nh \le T$ in a compact set K_0 which has a positive distance to the boundary of K.

Theorem 8 *Assume (A1)–(A4). For sufficiently small h and δ and for a fixed truncation index N (large enough such that $h^N = \mathcal{O}(\delta^4)$), there exist functions $y(t)$ and $z_\ell(t)$ on an interval of length*

$$T = \mathcal{O}((h\delta)^{-1})$$

such that

- $q_n = y(nh) + \sum_{\ell \in \mathcal{I}^*} \zeta_\ell^n z_\ell(nh)$ *for* $0 \le nh \le T$*;*
- *on every subinterval $[jh, (j+1)h)$ the functions $y(t), z_\ell(t)$ are a solution of the system (4.1);*
- *the functions $y(t), z_\ell(t)$ have jump discontinuities of size $\mathcal{O}(h^{N+2})$ at the grid points jh;*
- $\|z_\ell(t)\| \le \delta$ *for* $0 \le t \le T$.

If no root of $\rho(\zeta)$ other than 1 is the product of two other roots, all these estimates are valid on an interval of length $T = \mathcal{O}((h\delta^2)^{-1})$.

Proof To define the functions $y(t), z_\ell(t)$ on the interval $[jh, (j+1)h)$ we consider the k consecutive numerical solution values $q_j, q_{j+1}, \ldots, q_{j+k-1}$. We compute initial values for (4.1) according to Lemma 1, and we let $y(t), z_\ell(t)$ be a solution of (4.1) on $[jh, (j+1)h)$. Because of (4.8) such a construction yields jump discontinuities of size $\mathcal{O}(h^{N+2})$ at the grid points.

It follows from Theorem 7 that $\mathcal{K}_\ell(y(t), \dot{y}(t), \mathbf{z}^*(t))$ remains constant up to an error of size $\mathcal{O}(h^2\delta^3)$ on the interval $[jh, (j+1)h)$. Taking into account the jump discontinuities, we find that

$$\mathcal{K}_\ell(y(t), \dot{y}(t), \mathbf{z}^*(t)) \le \mathcal{K}_\ell(y(0), \dot{y}(0), \mathbf{z}^*(0)) + C_1 th\delta^3 + C_2 th^{N+1} \tag{4.25}$$

as long as $\|z_\ell(t)\| \le \delta$. By (4.24) this then implies

$$\|z_\ell(t)\|^2 \le \|z_\ell(0)\|^2 + C_1 th\delta^3 + C_2 th^{N+1} + C_3 h^2\delta^2. \tag{4.26}$$

The assumption $\|z_\ell(t)\| \le \delta$ is certainly satisfied as long as $C_1 th\delta \le 1/4$, $C_2 th^{N+1} \le \delta^2/4$, and $C_3 h^2 \le 1/4$, so that the right-hand side of (4.26) is bounded by δ^2. This proves not only the estimate for $\|z_\ell(t)\|$, but at the same time it guarantees recursively that the above construction of the functions $y(t), z_\ell(t)$ is feasible. $\square$

Notice that for initial values computed by a sufficiently accurate one-step method the constant δ can be chosen as small as $\mathcal{O}(h^{p+2})$ where p is the order of the multistep method (cf. Lemma 1). The above estimates are therefore valid on very long time intervals.

Example 4 To illustrate the long-time behaviour of the parasitic terms z_ℓ we consider the pendulum equation $\ddot q = -\sin q$, and we apply the symmetric multistep methods with generating polynomials

$$\begin{aligned} &\text{(S)} \qquad \rho(\zeta) = (\zeta-1)^2(\zeta^2+1), \qquad \sigma(\zeta) = \tfrac{1}{6}(7\zeta - 2\zeta^2 + 7\zeta^3), \\ &\text{(T)} \qquad \rho(\zeta) = (\zeta-1)^2(\zeta+1)^2, \qquad \sigma(\zeta) = \tfrac{4}{3}(\zeta + \zeta^2 + \zeta^3). \end{aligned}$$

Both methods are explicit and of order 4. The starting values are chosen far from a smooth solution, so that the propagation of the parasitic terms in the numerical solution can be better observed.

The parasitic roots of method (S) are $\pm i$ and both are simple. The numerical solution is therefore of the form

$$q_n = y(nh) + i^n z_1(nh) + (-i)^n \overline{z_1(nh)} + (-1)^n z_2(nh).$$

One observes in Fig. 4 that the functions $z_j(t)$ not only remain bounded and small, but they stay nearly constant over the considered interval.

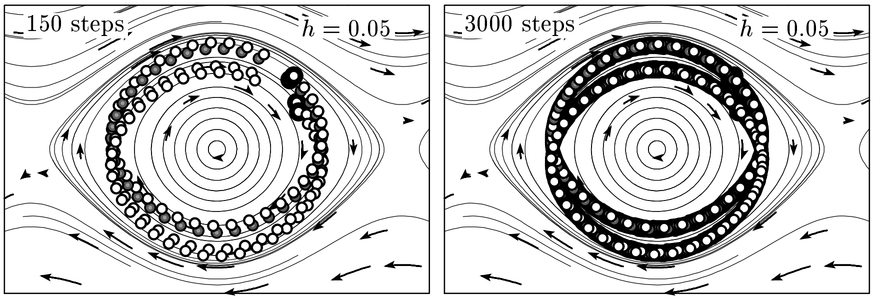

Fig. 4. Stable propagation of perturbations in the starting values, method (S)

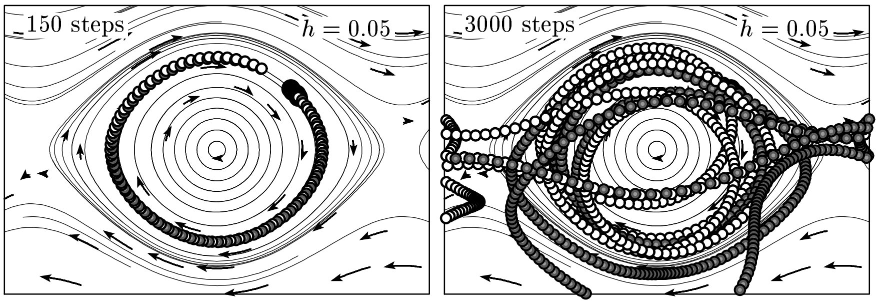

Fig. 5. Unstable propagation of perturbations in the starting values, method (T)

Method (T) has a double parasitic root at -1 and, therefore, is not s-stable. Its numerical solution behaves like

$$q_n = y(nh) + (-1)^n z(nh).$$

In Fig. 5 every second approximation is drawn in grey. One sees that the numerical solution stays on two smooth curves $y(t) + z(t)$ and $y(t) - z(t)$ which, however, do not remain close to each other for method (T).

5 Backward error analysis for general numerical solutions, part II

The results of the previous section enable us to finally prove the theorems of Section 2.

5.1 Conservation of energy

The energy conservation is now a direct consequence of Theorems 6 and 8. We shall use the representation of q_n in terms of functions $y(t), z_\ell(t)$ as in Theorem 8. Taking into account the jump discontinuities of these functions, Theorem 6 yields

$$\mathcal{H}(y(t), \dot y(t), \mathbf{z}^*(t)) = \mathcal{H}(y(0), \dot y(0), \mathbf{z}^*(0)) + \mathcal{O}(th^3\delta^4) + \mathcal{O}(th^{N+1}).$$

We have $\delta = \mathcal{O}(h^{p+1})$ if the starting approximations are computed by a pth order one-step method. If N is chosen sufficiently large, this together with (4.19) implies that

$$H(y(t), \dot y(t)) = H(y(0), \dot y(0)) + \mathcal{O}(h^p) \qquad \text{for} \quad 0 \le t \le T = \mathcal{O}(h^{-p-2}).$$

If the velocity approximation $p_n = v_n$ (for identity mass matrix) is given by a pth order finite difference formula (2.1), it follows from

Theorem 8 that $p_n = \dot{y}(nh) + \mathcal{O}(h^p)$ provided the truncation index N is sufficiently large. This gives the statement of Theorem 1. If no root of $\rho(\zeta)$ other than 1 is a product of two other roots, the statement holds on intervals of length $\mathcal{O}(h^{-2p-3})$.

5.2 Conservation of angular momentum and quadratic first integrals

The invariance property (2.6) implies, for $\mathcal{U}$ of (4.9),

$$\mathcal{U}(e^{\tau A}\mathbf{z}) = \mathcal{U}(\mathbf{z}) \qquad \text{for all} \quad \tau, \mathbf{z}.$$

Along solutions $\mathbf{z}(t)$ of the modified equations (4.10) we therefore have up to terms of size $\mathcal{O}(h^N)$

$$\begin{aligned} 0 &= \frac{d}{d\tau}\Big|_{\tau=0} \mathcal{U}(e^{\tau A}\mathbf{z}) = \sum_{\ell\in\mathcal{I}} z_{-\ell}^T\, A\, \nabla_{z_{-\ell}} \mathcal{U}(\mathbf{z}) \\ &= \sum_{\ell\in\mathcal{I}} h^{-2} z_{-\ell}^T\, A \left(\frac{\rho}{\sigma}\right)(\zeta_\ell e^{hD}) z_\ell. \end{aligned} \tag{5.1}$$

If $\sigma(\zeta)$ has a root ζ_ℓ, then the corresponding term is omitted from the last sum, leading to a remainder term which in the worst case is $\mathcal{O}(h^3\delta^4)$, as in Theorem 6. Like in the previous proofs, the last sum is, for skew-symmetric A, the total derivative of a function

$$\mathcal{L}(y, \dot{y}, \mathbf{z}^*) = L_0(y, \dot{y}, \mathbf{z}^*) + \ldots + h^{N-1} L_{N-1}(y, \dot{y}, \mathbf{z}^*)$$

which satisfies (under the same assumptions as in Theorem 6)

$$\mathcal{L}(y(t), \dot{y}(t), \mathbf{z}^*(t)) = \mathcal{L}(y(0), \dot{y}(0), \mathbf{z}^*(0)) + \mathcal{O}(th^3\delta^4) + \mathcal{O}(th^{N+1})$$

and

$$\mathcal{L}(y, \dot{y}, \mathbf{z}^*) = L(y, \dot{y}) + \mathcal{O}(h^p) + \mathcal{O}(\delta^2/h). \tag{5.2}$$

The statement of Theorem 2 thus follows in exactly the same way as that for Theorem 1 in Section 5.1.

5.3 Integrable systems

Assume that the differential equation $\ddot{q} = -\nabla U(q)$ is an integrable reversible system (see Section 2.3). By Theorem 8, the numerical solution can be written as $q_n = y(nh) + \sum_{\ell\in\mathcal{I}^*} \zeta_\ell^n z_\ell(nh)$, where (at least locally) $y(t)$ is the solution of a modified differential equation (first equation of (4.1))

$$\ddot{y} = f_{0,0}(y, \dot{y}, \mathbf{z}^*) + h f_{0,1}(y, \dot{y}, \mathbf{z}^*) + \ldots + h^{N-1} f_{0,N-1}(y, \dot{y}, \mathbf{z}^*) \tag{5.3}$$

which, for $\mathbf{z}^* = 0$ becomes the modified differential equation (3.2). We now consider (5.3) as a differential equation for y only with $\mathbf{z}^*(t)$ as a given function. Since $z_j(t) = \mathcal{O}(\delta)$ (see Theorem 8) and since $\mathbf{z}^*$ appears at least quadratically in (5.3), this equation is a $\mathcal{O}(\delta^2)$ perturbation of (3.2). We now apply the same transformation as for the proof of Proposition 3. The additional (non-reversible) perturbation of size $\mathcal{O}(\delta^2)$ in the differential equation (5.3) produces an error term of size $\mathcal{O}(t\delta^2)$ in the action variables and of size $\mathcal{O}(t^2\delta^2)$ in the angle variables. If $\delta = \mathcal{O}(h^{p+1})$, these terms are negligible with respect to those already appearing in Proposition 3. The errors due to the jump discontinuities (Theorem 8) are also negligible. We have thus proved the statement of Theorem 3.

Acknowledgement. The authors are grateful to Gerhard Wanner and to the participants of the numerical analysis seminar in Geneva for stimulating discussions on the subject of this paper.

References

1. G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms. *J. Statist. Phys.*, 74:1117–1143, 1994.
2. B. Cano and J. M. Sanz-Serna. Error growth in the numerical integration of periodic orbits by multistep methods with application to reversible systems. *IMA J. Numer. Anal.*, 18:57–75, 1998.
3. G. Dahlquist. Stability and error bounds in the numerical integration of ordinary differential equations. *Trans. of the Royal Inst. of Techn., Stockholm, Sweden*, 130, 1959.
4. N. W. Evans and S. Tremaine. Linear multistep methods for integrating reversible differential equations. *Astron. J.*, 118:1888–1899, 1999.
5. T. Fukushima. Symmetric multistep methods revisited. In *30th Symposium on Celestial Mechanics*, pages 229–247, 1998.
6. T. Fukushima. Symmetric multistep methods revisited: II. Numerical experiments. In *173rd colloquium of the International Astronomical Union*, pages 309–314, 1999.
7. E. Hairer. Backward error analysis for multistep methods. *Numer. Math.*, 84:199–232, 1999.
8. E. Hairer and M. Hairer. GniCodes – Matlab programs for geometric numerical integration. *In: Frontiers in Numerical Analysis (Durham 2002), Springer, Berlin*, 2003.
9. E. Hairer and P. Leone. Order barriers for symplectic multi-value methods. *In: Numerical analysis 1997, Proc. of the 17th Dundee Biennial Conference 1997, D. F. Griffiths D. J. Higham & G. A. Watson eds. Pitman Research Notes in Mathematics Series*, 380:133–149, 1998.
10. E. Hairer and C. Lubich. Long-time energy conservation of numerical methods for oscillatory differential equations. *SIAM J. Numer. Anal.*, 38:414–441, 2000.

11. E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations.* Springer Series in Computational Mathematics 31. Springer, Berlin, 2002.
12. E. Hairer, C. Lubich, and G. Wanner. Geometric numerical integration illustrated by the Störmer–Verlet method. *Acta Numerica*, 12, 2003.
13. E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems.* Springer Series in Computational Mathematics 8. Springer, Berlin, 2nd edition, 1993.
14. P. Henrici. *Discrete Variable Methods in Ordinary Differential Equations.* John Wiley & Sons Inc., New York, 1962.
15. J. D. Lambert and I. A. Watson. Symmetric multistep methods for periodic initial value problems. *J. Inst. Maths. Applics.*, 18:189–202, 1976.
16. J. Moser. Stable and random motions in dynamical systems. *Annals of Mathematics Studies. No. 77*, 1973.
17. G. D. Quinlan and S. Tremaine. Symmetric multistep methods for the numerical integration of planetary orbits. *Astron. J.*, 100:1694–1700, 1990.
18. C. Störmer. Sur les trajectoires des corpuscules électrisés. *Arch. sci. phys. nat. Genève*, 24:5–18, 113–158, 221–247, 1907.
19. Y.-F. Tang. The symplecticity of multi-step methods. *Computers Math. Applic.*, 25:83–90, 1993.