Journal of Computational Mathematics, Vol.20, No.5, 2002, 449–460.

BACKWARD ERROR ANALYSIS OF SYMPLECTIC INTEGRATORS FOR LINEAR SEPARABLE HAMILTONIAN SYSTEMS*

Peter Görtz
(*Institut für Praktische Mathematik, Universität Karlsruhe, Englerstraße 2, 76128 Karlsruhe, Germany*)

Abstract

Symplecticness, stability, and asymptotic properties of Runge–Kutta, partitioned Runge–Kutta, and Runge–Kutta–Nyström methods applied to the simple Hamiltonian system $\dot{p} = -\nu q, \dot{q} = \kappa p$ are studied. Some new results in connection with P–stability are presented. The main part is focused on backward error analysis. The numerical solution produced by a symplectic method with an appropriate stepsize is the exact solution of a perturbed Hamiltonian system at discrete points. This system is studied in detail and new results are derived. Numerical examples are presented.

Key words: Hamiltonian systems, Backward error analysis, Symplectic integrators.

1. Introduction

In the area of symplectic integration of Hamiltonian systems of the form

$$\dot{u} = -J \nabla H(u),$$

where

$$u = \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} p^{(1)} & \dots & p^{(n)} & q^{(1)} & \dots & q^{(n)} \end{bmatrix}^T, \ J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix},$$

$H \in C^{(\infty)}(\mathcal{M})$ is the Hamiltonian, $\mathcal{M} \subseteq \mathbb{R}^{2n}$ open is the phase space,

$$\nabla H = \begin{bmatrix} \frac{\partial H}{\partial p^{(1)}} & \dots & \frac{\partial H}{\partial q^{(n)}} \end{bmatrix}^T,$$

backward error analysis plays an important role. The idea is to interprete the numerical solution produced by a symplectic one–step method as the exact solution of a perturbed Hamiltonian system. In general, this is only formally possible; the perturbed Hamiltonian system is given as a power series which is usually divergent (Feng [4], Hairer [9], Tang [15], Yoshida [17]; cf. Hairer, Nørsett, Wanner [10], Sanz–Serna, Calvo [14]). If the Hamiltonian system is linear, i.e., the Hamiltonian is a quadratic form, then the perturbed Hamiltonian system can be expressed by the logarithm of a matrix. Conditions exists which guarantee the existence of a logarithm of the relevant matrix (Wang [16]).

Often the Hamiltonian system is linear and separable as follows:

$$\left.\begin{aligned} &\begin{bmatrix} \dot{p} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} 0 & -N \\ K & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}, \ K, N \in \mathbb{R}^{n \times n} \ \textit{symmetric}, \\ &\textit{where a nonsingular matrix } W \in \mathbb{R}^{n \times n} \textit{ exists with} \\ &\quad W^{-1} K W^{-T} = \operatorname{diag}(\kappa_1, \dots, \kappa_n), \ W^T N W = \operatorname{diag}(\nu_1, \dots, \nu_n). \end{aligned}\right\} \quad (1)$$

* Received February 23, 2000; Final revised December 31, 2000.

The Hamiltonian splits into the sum of two quadratic forms, $H(p,q) = \frac{1}{2}p^T Kp + \frac{1}{2}q^T Nq$. The most occuring case is that K is positive definite, then a matrix W exists with $W^{-1}KW^{-T} = I$. This is evident from the fact that with K also K^{-1} is symmetric and positive definite, and therefore by theorems about the principal axis transformation there exists a nonsingular matrix W such that $W^T K^{-1} W$ is equal to I and $W^T NW$ is a diagonal matrix. The situation K and N positive definite arise for example in connection with small oscillation approximations for nonlinear mechanical systems near stable equilibrium points (cf. Abraham, Marsden [1], Arnold [2]).

For the numerical integration of (1) Runge–Kutta (RK) methods, partitioned Runge–Kutta (PRK) methods, and Runge–Kutta–Nyström (RKN) methods can be used (cf. Hairer, Nørsett, Wanner [10], Sanz–Serna, Calvo [14]; see also [7]), which are summarized as Runge–Kutta type (RKT) methods. After a symplectic transformation of coordinates (1) decomposes into n Hamiltonian systems of the form

$$\begin{bmatrix} \dot{p} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} 0 & -\nu \\ \kappa & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}, \kappa, \nu \in \mathbb{R}, \tag{2}$$

with $H(p,q) = \frac{1}{2}\kappa p^2 + \frac{1}{2}\nu q^2$. Methods that are symplectic for all systems of type (1) are called ls–symplectic. Stability properties are studied in detail in [6] and [8]. In this paper a backward error analysis of ls–symplectic RKT methods is presented. First, in section 2 the main results concerning ls–symplecticness and stability are summarized and some new results are given. In section 3 the backward error analysis is developed. If $\kappa\nu > 0$ in (2), then the solution to given initial conditions describes an ellipse in the phase plane. The numerical solution of an ls–symplectic RKT method with an admissible step size is the exact solution of a perturbed Hamiltonian system and lies also on an ellipse; the perturbed system is formulated, the shape of the ellipse is studied. Further, the conservation of the Hamiltonian is investigated, a lower and an upper bound for the error are given. In section 4 numerical examples are presented. All the results can easily be generalized to the integration of (1).

Note that after a further symplectic transformation of coordinates system (2) reduces in the case $\kappa\nu > 0$ to $\dot{p} = -\omega q$, $\dot{q} = \omega p$ with $\omega > 0$. For only studying the stability of RKT methods this symplification reduces the amount of work, but the results are also valid for $\kappa \neq \nu$. For backward error analysis on the other side there is no real benefit from $\kappa = \nu$. So, there is no need for this further symplification here. Especially, some early investigations are not restricted to that (Feng, Qin [5]).

2. Basic results

The symplecticness and stability of RKT methods for linear separable Hamiltonian systems of type (1) are studied in detail in [6] and [8]. In this section a short summary and some new results are given which are close related to the theory of P–stability (van der Houwen, Sommeijer [11], [12]).

2.1 ls–symplecticness and stability

A one–step method is called *ls–symplectic* if it is symplectic for all systems of type (1). The basis for the investigation of ls–symplecticness of RKT methods is that such a method applied to (1) with initial condition $u(0) = u_0$ reduces to

$$u_{m+1} = G(hK, hN)u_m, \quad m = 0, 1, 2, \ldots,$$

where for square matrices X, Y of the same size

$$G(X,Y) = \begin{bmatrix} \Gamma_{11}(YX) & -Y\Gamma_{12}(XY) \\ X\Gamma_{21}(YX) & \Gamma_{22}(XY) \end{bmatrix}.$$

The $\Gamma_{ij}, i,j = 1,2$, are rational functions of the form $\frac{\Psi_{ij}}{\Psi}$, where $\Psi_{11}, \ldots, \Psi_{22}$, Ψ are polynomials with real coefficients that are determined by the parameters of the method. For explicit methods it holds $\Psi \equiv 1$, for RK methods it holds $\Gamma_{11} = \Gamma_{22}$ and $\Gamma_{12} = \Gamma_{21}$. The matrix $G(hK, hN)$ exists if

$$\begin{bmatrix} h\kappa_l \\ h\nu_l \end{bmatrix} \in \mathcal{D} := \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : \Psi(xy) \neq 0 \right\} \text{ for } l = 1, \ldots, n. \tag{3}$$

It is assumed that condition (3) is satisfied. This is the case at least for sufficiently small $h > 0$, and it is a fact that *a RKT method is ls–symplectic if and only if* $\det G(x, y) = 1$ *for all* $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathcal{D}$

2.2 Symplectic stability and dispersion

Test equation (2) is stable, but not asymptotically stable, if $\kappa\nu > 0$ or $\kappa = \nu = 0$, it is unstable in all other cases. If $\kappa = \nu = 0$, then RKT methods produce of course the exact solution, so of interest are only the cases $\kappa, \nu > 0$ and $\kappa, \nu < 0$. The solution of

$$\text{(2) with } \kappa\nu > 0 \text{ and initial conditions } p(0) = p_0, q(0) = q_0 \tag{4}$$

is given as

$$\begin{bmatrix} p(t) \\ q(t) \end{bmatrix} = \begin{bmatrix} \cos(\sqrt{\kappa\nu}\, t) & \mp\sqrt{\frac{\nu}{\kappa}} \sin(\sqrt{\kappa\nu}\, t) \\ \pm\sqrt{\frac{\kappa}{\nu}} \sin(\sqrt{\kappa\nu}\, t) & \cos(\sqrt{\kappa\nu}\, t) \end{bmatrix} \begin{bmatrix} p_0 \\ q_0 \end{bmatrix}, \ \kappa, \nu \gtrless 0, \tag{5}$$

with

$$\frac{p^2(t)}{p_0^2 + \frac{\nu}{\kappa} q_0^2} + \frac{q^2(t)}{\frac{\kappa}{\nu} p_0^2 + q_0^2} \equiv 1. \tag{6}$$

Taking the signs of the sine terms in (5) into account (6) leads to the following result.

Theorem 1. *The solution of (4) describes in the phase plane the ellipse with semiaxis* $\sqrt{p_0^2 + \frac{\nu}{\kappa} q_0^2}$ *in p–direction and semiaxis* $\sqrt{\frac{\kappa}{\nu} p_0^2 + q_0^2}$ *in q–direction. In the case* $\kappa, \nu > 0$ *the ellipse is passed through in mathematical positive direction and in the case* $\kappa, \nu < 0$ *in mathematical negative direction.*

An ls–symplectic RKT method applied to (4) reduces to the discrete linear system $u_{m+1} = G(h\kappa, h\nu) u_m, m = 0, 1, 2, \ldots$, where $G(h\kappa, h\nu)$ is symplectic, i.e., $\det G(h\kappa, h\nu) = 1$. The stability condition is $|\mathrm{tr}\, G(h\kappa, h\nu)| < 2$; the stable cases $G(h\kappa, h\nu) = \pm I$ are not taken into account. So, for an ls–symplectic RKT method the set $\gamma := \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathcal{D} : |\mathrm{tr}\, G(x, y)| < 2 \right\}$ is called *symplectic stability region*, and the method is called *symplectically stable* if $\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : xy > 0 \right\} \subseteq \gamma$.

Theorem 2. *([6]) An ls–symplectic RKT method applied to (4) reduces for* $\begin{bmatrix} h\kappa \\ h\nu \end{bmatrix} \in \gamma$ *to*

$$u_m = Z(mh) u_0, \ m = 1, 2, 3, \ldots,$$

where $Z(t) = [z_{ij}(t)]_{i,j=1,2}$ *with*

$$\begin{aligned}
z_{11}(t) &= \cos\left(\frac{\varphi(\kappa\nu h^2)}{h}\, t\right) - \frac{\delta(\kappa\nu h^2)}{2\sigma(\kappa\nu h^2)} \sin\left(\frac{\varphi(\kappa\nu h^2)}{h}\, t\right), \\
z_{12}(t) &= \frac{-\nu h \Gamma_{12}(\kappa\nu h^2)}{\sigma(\kappa\nu h^2)} \sin\left(\frac{\varphi(\kappa\nu h^2)}{h}\, t\right), \\
z_{21}(t) &= \frac{\kappa h \Gamma_{21}(\kappa\nu h^2)}{\sigma(\kappa\nu h^2)} \sin\left(\frac{\varphi(\kappa\nu h^2)}{h}\, t\right), \\
z_{22}(t) &= \cos\left(\frac{\varphi(\kappa\nu h^2)}{h}\, t\right) + \frac{\delta(\kappa\nu h^2)}{2\sigma(\kappa\nu h^2)} \sin\left(\frac{\varphi(\kappa\nu h^2)}{h}\, t\right), \\
\varphi(\kappa\nu h^2) &= \arccos\left(\tfrac{1}{2}\Gamma_{11}(\kappa\nu h^2) + \tfrac{1}{2}\Gamma_{22}(\kappa\nu h^2)\right),
\end{aligned}$$

$$\begin{aligned}\delta(\kappa\nu h^2) &= \Gamma_{22}(\kappa\nu h^2) - \Gamma_{11}(\kappa\nu h^2),\\ \sigma(\kappa\nu h^2) &= \sqrt{1-\tfrac{1}{4}\Big(\Gamma_{11}(\kappa\nu h^2)+\Gamma_{22}(\kappa\nu h^2)\Big)^2}.\end{aligned}$$

A comparison between the numerical solution in Theorem 2 and the exact solution (5) at $t = mh$, $h > 0$, shows that $\varphi(\kappa\nu h^2)$ is an approximation to $\sqrt{\kappa\nu}h$. The difference $\phi(\sqrt{\kappa\nu}\,h) := \sqrt{\kappa\nu}h - \varphi(\kappa\nu h^2)$ is called *dispersion* or *phase error* of the method $\Big($at the point $\begin{bmatrix} h\kappa \\ h\nu \end{bmatrix}\Big)$. The method has *order of dispersion* d if d is the greatest integer such that $\phi(z) = O(z^{d+1})$ for $z \to 0_+$; the limit $\lim\limits_{z\to 0_+} \dfrac{\phi(z)}{z^{d+1}}$ is called *error constant.*

2.3 Asymptotic relations

For every RKT method in $\Gamma_{11}(\kappa\nu h^2)$, $\Gamma_{22}(\kappa\nu h^2)$ only even powers of h appear, and in $-\nu h\Gamma_{12}(\kappa\nu h^2)$, $\kappa h\Gamma_{21}(\kappa\nu h^2)$ only odd powers of h appear. Hence, if $\kappa\nu > 0$, then for an r–th order RKT method $\Gamma_{11}(\kappa\nu h^2)$, $\Gamma_{22}(\kappa\nu h^2)$ are approximations of order $2[\frac{r}{2}]+1$ to $\cos(\sqrt{\kappa\nu}h)$ and $-\nu h\Gamma_{12}(\kappa\nu h^2)$, $\kappa h\Gamma_{21}(\kappa\nu h^2)$ are approximations of order $2[\frac{r+1}{2}]$ to sine expressions. With this and the next lemma it is possible to investigate for an ls–symplectic RKT method the order of dispersion and the asymptotic behaviour of the fractions in front of the sine expressions in Theorem 2.

Lemma 3. *For an r–th order ls–symplectic RKT method the following holds:*

a) For real x, y with xy positive and sufficiently small

$$\tfrac{1}{2}\Big(\Gamma_{11}(xy)+\Gamma_{22}(xy)\Big) = \cos(\sqrt{xy}) + g(\sqrt{xy}),$$

where for $|z|$ sufficiently small

$$g(z) = e_2 z^{2\rho+2} + e_4 z^{2\rho+4} + e_6 z^{2\rho+6} + \ldots$$

with $\rho \geq [\frac{r+1}{2}]$, and $e_2, e_4, e_6, \ldots \in \mathbb{R}, e_2 \neq 0$.

b) With the notations in a) for sufficiently small $z > 0$ let

$$f(z) = \frac{\cos(z)+\beta(z)}{\sqrt{\Big(1-(\cos(z)+\beta(z))^2\Big)^3}}\, g(z),$$

where β is a not necessarily continuous function such that for every z the value $\beta(z)$ lies between 0 and $g(z)$. Then

$$f(z) = O\Big(z^{2\rho-1}\Big) \text{ for } z \to 0_+.$$

Proof. a) is proved by some straight forward calculations using power series extensions.

b) It has to be distinguished whether e_2 is positive or negative, i.e., $g(z) > 0$ for all sufficiently small $z > 0$ or $g(z) < 0$ for all sufficiently small $z > 0$. Note that $\cos(z) > 0$, $\sin(z) > 0$, and $|\cos(z) + g(z)| < 1$ if $z > 0$ sufficiently small.

- $e_2 > 0$: For $z > 0$ sufficiently small it holds

$$\begin{aligned} 0 \;<\; f(z) &< \frac{\cos(z)+g(z)}{\sqrt{\Big(1-(\cos(z)+g(z))^2\Big)^3}}\, g(z)\\ &\leq \frac{1}{z^3\sqrt{\Big(1+f_2z^2+f_4z^4+\ldots)\Big)^3}}\, C\, z^{2\rho+2}\\ &\leq 2\,C\,z^{2\rho-1} \text{ where } C \geq e_2 \text{ and } f_2, f_4, \ldots \in \mathbb{R}.\end{aligned}$$

- $e_2 < 0$: For $z > 0$ sufficiently small it holds

$$\begin{aligned} |f(z)| &< \frac{\cos(z) + |g(z)|}{\sqrt{\Big(1-\cos^2(z)\Big)^3}}\,|g(z)| \\ &< \frac{2}{z^3\Big(1-\frac{1}{3!}z^2+\frac{1}{5!}z^4-+\ldots\Big)^3}\,C\,z^{2\rho+2} \\ &< 2\,C\,z^{2\rho-1} \text{ where } C \geq |e_2| \qquad \square \end{aligned}$$

With the notations and abbreviations in Theorem 2 and Lemma 3 important asymptotic relations of RKT methods can be formulated now; compare (5).

Theorem 4. *An r–th order ls–symplectic RKT method satisfies the following asymptotic relations:*

a) The order of dispersion is 2ρ and the error constant is e_2, i.e., the order of dispersion is at least $2[\frac{r+1}{2}]$.

b) Let $\kappa\nu > 0$. Then for $h \to 0_+$ it is

$$\begin{aligned} \frac{\delta(\kappa\nu h^2)}{\sigma(\kappa\nu h^2)} &= O\Big(h^{2[\frac{r}{2}]+1}\Big), \\ \frac{-\nu h\Gamma_{12}(\kappa\nu h^2)}{\sigma(\kappa\nu h^2)} &= \mp\sqrt{\tfrac{\nu}{\kappa}} + O\Big(h^{2[\frac{r+1}{2}]}\Big), \ \text{if } \kappa,\nu \gtrless 0, \\ \frac{\kappa h\Gamma_{21}(\kappa\nu h^2)}{\sigma(\kappa\nu h^2)} &= \pm\sqrt{\tfrac{\kappa}{\nu}} + O\Big(h^{2[\frac{r+1}{2}]}\Big), \ \text{if } \kappa,\nu \gtrless 0. \end{aligned}$$

Proof. Let $z > 0$ be sufficiently small, i.e., such that for all $\tilde{z} \in (0, z]$ the inequalities $0 < \cos(\tilde{z}) + g(\tilde{z}) < 1, \cos(\tilde{z}) > 0, \sin(\tilde{z}) > 0$, and $g(\tilde{z}) > 0$ respectively $g(\tilde{z}) < 0$ hold; further let

$$\mathcal{F}_z(t) = \arccos\Big(\cos(z) + t\Big), \quad \mathcal{G}_z(t) = \frac{1}{\sqrt{1-(\cos(z)+t)^2}}$$

for $t \in \mathcal{I}_z := \Big(-1-\cos(z), 1-\cos(z)\Big)$.

a) For every $t \in \mathcal{I}_z$ the Taylor formula implies the representation

$$\begin{aligned} \mathcal{F}_z(t) &= \mathcal{F}_z(0) + \mathcal{F}_z'(0)t + \tfrac{1}{2}\mathcal{F}_z''(\xi_t)t^2 \\ &= z - \frac{1}{\sin(z)}\,t - \frac{\cos(z)+\xi_t}{2\sqrt{\Big(1-(\cos(z)+\xi_t)^2\Big)^3}}\,t^2 \end{aligned}$$

with ξ_t between 0 and t. With Lemma 3 this leads in the special case $t = g(z)$ to

$$\begin{aligned} & z - \arccos\Big(\cos(z) + g(z)\Big) \\ &= \frac{1}{\sin(z)}\,g(z) + \frac{\cos(z)+\xi_{g(z)}}{2\sqrt{\Big(1-(\cos(z)+\xi_{g(z)})^2\Big)^3}}\,g^2(z) \\ &= e_2 z^{2\rho+1} + O\Big(z^{2\rho+3}\Big) + O\Big(z^{2\rho-1}\Big)\,O\Big(z^{2\rho+2}\Big) \\ &= e_2 z^{2\rho+1} + O\Big(z^{2\rho+3}\Big) \text{ for } z \to 0_+. \end{aligned}$$

b) For every $t \in \mathcal{I}_z$ the Taylor formula implies the representation

$$\begin{aligned}\mathcal{G}_z(t) &= \mathcal{G}_z(0) + \mathcal{G}_z'(\xi_t)t \\ &= \frac{1}{\sin(z)} + \frac{\cos(z)+\xi_t}{\sqrt{\left(1-(\cos(z)+\xi_t)^2\right)^3}}\, t\end{aligned}$$

with ξ_t between 0 and t. With Lemma 3 this leads in the special case $z = \sqrt{\kappa\nu}h$, $t = g(z)$ to

$$\begin{aligned}\frac{1}{\sigma(\kappa\nu h^2)} &= \frac{1}{\sin(\sqrt{\kappa\nu}h)} + \\ &\quad \frac{\cos(\sqrt{\kappa\nu}h) + \xi_{g(\sqrt{\kappa\nu}h)}}{\sqrt{\left(1-\left(\cos(\sqrt{\kappa\nu}h)+\xi_{g(\sqrt{\kappa\nu}h)}\right)^2\right)^3}}\, g(\sqrt{\kappa\nu}h) \\ &= \frac{1}{\sin(\sqrt{\kappa\nu}h)} + O\Big((\sqrt{\kappa\nu}h)^{2\rho-1}\Big) \\ &= \frac{1}{\sin(\sqrt{\kappa\nu}h)} + O\Big(h^{2\rho-1}\Big)\end{aligned}$$

for $h \to 0_+$. The statements now result from

$$\Gamma_{22}(\kappa\nu h^2) - \Gamma_{11}(\kappa\nu h^2) = O\Big(h^{2[\frac{r}{2}]+2}\Big),$$

$$-\nu h\Gamma_{12}(\kappa\nu h^2) = \mp\sqrt{\tfrac{\nu}{\kappa}}\sin(\sqrt{\kappa\nu}h) + O\Big(h^{2[\frac{r+1}{2}]+1}\Big), \text{if } \kappa,\nu \gtrless 0,$$

$$\kappa h\Gamma_{21}(\kappa\nu h^2) = \pm\sqrt{\tfrac{\kappa}{\nu}}\sin(\sqrt{\kappa\nu}h) + O\Big(h^{2[\frac{r+1}{2}]+1}\Big), \text{if } \kappa,\nu \gtrless 0,$$

for $h \to 0_+$. □

3. Backward error analysis

The perturbed Hamiltonian system that is solved exactly by an ls–symplectic RKT method applied to (4) is studied. Note that $u = \begin{bmatrix} p \\ q \end{bmatrix}, u_0 = \begin{bmatrix} p_0 \\ q_0 \end{bmatrix}, \bar{u} = \begin{bmatrix} \bar{p} \\ \bar{q} \end{bmatrix}$, and so on. The results can easily be generalized to the integration of (1).

3.1 Perturbed Hamiltonian system

The following result is obvious when looking at the eigenvalues.

Lemma 5. *For every symplectic matrix $L = [\ell_{ij}]_{i,j=1,2} \in \mathbb{R}^{2\times 2}$ (i.e., $\det L = 1$) with $|\mathrm{tr}\, L| < 2$ it is $\ell_{12}\ell_{21} < 0$.*

Now the main result can be formulated; the notations of Theorem 2 and $[\ell_{ij}]_{i,j=1,2} := G(h\kappa, h\nu)$ are used.

Theorem 6. *Let*

- *$u_0 \neq 0$ in (4),*
- *for a given ls–symplectic RKT method the stepsize $h > 0$ chosen in such a way that $\begin{bmatrix} h\kappa \\ h\nu \end{bmatrix} \in \gamma$ and $(u_m)_{m\in\mathbb{N}}$ the numerical solution for (4) produced by the method,*
- $$S(h\kappa, h\nu) = \frac{\varphi(\kappa\nu h^2)}{h\sigma(\kappa\nu h^2)} \begin{bmatrix} \kappa h\Gamma_{21}(\kappa\nu h^2) & \frac{1}{2}\delta(\kappa\nu h^2) \\ \frac{1}{2}\delta(\kappa\nu h^2) & \nu h\Gamma_{12}(\kappa\nu h^2) \end{bmatrix},$$

- *$\hat{u}(t)$ the solution with value u_0 at $t_0 = 0$ of the Hamiltonian system*
$$\dot{\bar{u}} = -JS(h\kappa, h\nu)\bar{u} \tag{7}$$
with Hamiltonian $\bar{H}(\bar{u}) = \frac{1}{2}\bar{u}^T S(h\kappa, h\nu)\bar{u}$.

Then
$$\hat{u}(t) = Z(t)u_0, \text{ i.e., } u_m = \hat{u}(mh),\ m = 1,2,3,\ldots,$$
and $\hat{u}(t)$ describes in the phase plane for

- *$\ell_{11} = \ell_{22}$ the ellipse with semiaxis $\sqrt{p_0^2 - \frac{\ell_{12}}{\ell_{21}}q_0^2}$ in $p-$direction and semiaxis $\sqrt{-\frac{\ell_{21}}{\ell_{12}}p_0^2 + q_0^2}$ in $q-$direction,*
- *$\ell_{11} \neq \ell_{22}$ the ellipse with semiaxis*
$$\sqrt{\frac{\ell_{21}p_0^2 + (\ell_{22} - \ell_{11})p_0q_0 - \ell_{12}q_0^2}{\frac{1}{2}\left((\ell_{21} - \ell_{12}) \pm \sqrt{(\ell_{12} + \ell_{21})^2 + (\ell_{22} - \ell_{11})^2}\right)}}, \quad \ell_{22} - \ell_{11} \gtrless 0,$$
in $p-$direction and semiaxis
$$\sqrt{\frac{\ell_{21}p_0^2 + (\ell_{22} - \ell_{11})p_0q_0 - \ell_{12}q_0^2}{\frac{1}{2}\left((\ell_{21} - \ell_{12}) \mp \sqrt{(\ell_{12} + \ell_{21})^2 + (\ell_{22} - \ell_{11})^2}\right)}}, \quad \ell_{22} - \ell_{11} \gtrless 0,$$
in $q-$direction rotated in mathematical positive direction about
$$\alpha = \tfrac{1}{2}\operatorname{arccot}\frac{\ell_{12} + \ell_{21}}{\ell_{22} - \ell_{11}}.$$

If $\begin{bmatrix}\tau\kappa \\ \tau\nu\end{bmatrix} \in \gamma$ for all $\tau \in (0, h]$, then in the case $\kappa, \nu > 0$ the ellipse is passed trough in mathematical positive direction and in the case $\kappa, \nu < 0$ in mathematical negative direction.

Proof. The identity $\hat{u}(t) = Z(t)\,u_0$ is given by differentiation and some simple algebraic manipulations; $u_m = Z(mh)u_0, m = 1,2,3,\ldots$ is obvious. The proof of the shape of the ellipse is rather extensive:

<u>$\ell_{11} = \ell_{22}$</u>

- It is
$$\bar{H}(\hat{u}(t)) = \tfrac{1}{2}\hat{u}^T(t)S(h\kappa, h\nu)\hat{u}(t) \equiv \bar{H}(u_0),$$
i.e., $\ell_{21}\hat{p}^2(t) - \ell_{12}\hat{q}^2(t) \equiv \ell_{21}p_0^2 - \ell_{12}q_0^2$. Equivalent to this equation is
$$\frac{\hat{p}^2(t)}{p_0^2 - \frac{\ell_{12}}{\ell_{21}}q_0^2} + \frac{\hat{q}^2(t)}{-\frac{\ell_{21}}{\ell_{12}}p_0^2 + q_0^2} \equiv 1.$$
Because of Lemma 5 the fractions $\frac{\ell_{12}}{\ell_{21}}$, $\frac{\ell_{21}}{\ell_{12}}$ are negative, so $p_0^2 - \frac{\ell_{12}}{\ell_{21}}q_0^2$ and $-\frac{\ell_{21}}{\ell_{12}}p_0^2 + q_0^2$ are positive. This proves the statement about the size of the ellipse.
- With Theorem 4 b) for sufficiently small $\tau \in \mathcal{I} := (0, h]$ it is
$$\beta_{12}(\tau\kappa, \tau\nu) := \frac{-\nu\tau\Gamma_{12}(\kappa\nu\tau^2)}{\sigma(\kappa\nu\tau^2)} \lessgtr 0 \text{ for } \kappa, \nu \gtrless 0,$$
$$\beta_{21}(\tau\kappa, \tau\nu) := \frac{\kappa\tau\Gamma_{21}(\kappa\nu\tau^2)}{\sigma(\kappa\nu\tau^2)} \gtrless 0 \text{ for } \kappa, \nu \gtrless 0.$$
$\beta_{12}(\tau\kappa, \tau\nu)\beta_{21}(\tau\kappa, \tau\nu)$ is a continuous function of τ on the interval $\mathcal{I}$ and because of Lemma 5 always negative. Hence, $\beta_{12}(\tau\kappa, \tau\nu)$ and $\beta_{21}(\tau\kappa, \tau\nu)$ cannot change their signs on $\mathcal{I}$, i.e.,
$$\beta_{12}(\tau\kappa, \tau\nu) \lessgtr 0, \quad \beta_{21}(\tau\kappa, \tau\nu) \gtrless 0 \quad \text{on } \mathcal{I}, \quad \kappa, \nu \gtrless 0.$$
This proves the statement about the direction in which the ellipse is passed through.

$\underline{\ell_{11} \neq \ell_{22}}$
Note that $\alpha \in (0, \frac{\pi}{2})$.

- Because of $\cot(2\alpha) = \frac{\ell_{12}+\ell_{21}}{\ell_{22}-\ell_{11}}$ there is the relation
$$(\ell_{22} - \ell_{11})\cos(2\alpha) = (\ell_{12} + \ell_{21})\sin(2\alpha). \tag{8}$$
Let $\mathcal{T} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$, then $\mathcal{T}$ rotates a vector of $\mathbb{R}^2$ in mathematical positive direction about α. With (8) and the addition theorems for sine and cosine some simple algebraic manipulations give
$$\mathcal{T}^T S(h\kappa, h\nu)\mathcal{T} \quad = \quad \frac{\varphi(h\kappa, h\nu)}{h\sigma(\kappa\nu h^2)}\, \mathrm{diag}(y_{11}, y_{22}),$$
where
$$\begin{aligned} y_{11} &= \ell_{21}\cos^2\alpha + (\ell_{22} - \ell_{11})\cos\alpha\sin\alpha - \ell_{12}\sin^2\alpha, \\ y_{22} &= -\ell_{12}\cos^2\alpha - (\ell_{22} - \ell_{11})\cos\alpha\sin\alpha + \ell_{21}\sin^2\alpha. \end{aligned}$$
This decomposition of $S(h\kappa, h\nu)$ has the following consequences:
  - Because of $\mathcal{T}^T = \mathcal{T}^{-1}$ the eigenvalues of $S(h\kappa, h\nu)$ are given as
  $$\lambda_1 = \frac{\varphi(h\kappa, h\nu)}{h\sigma(\kappa\nu h^2)}\, y_{11}, \quad \lambda_2 = \frac{\varphi(h\kappa, h\nu)}{h\sigma(\kappa\nu h^2)}\, y_{22}.$$
  The fraction $\frac{\varphi(h\kappa,h\nu)}{h\sigma(\kappa\nu h^2)}$ is positive; the determinante of $S(h\kappa, h\nu)$ which is the product of λ_1 and λ_2 is given as $(\frac{\varphi(h\kappa,h\nu)}{h})^2$, i.e., is also positive. Therefore:
  $$\begin{gathered} \text{Either } y_{11}, y_{22} > 0 \text{ and } S(h\kappa, h\nu) \text{ positive definite,} \\ \text{or } y_{11}, y_{22} < 0 \text{ and } S(h\kappa, h\nu) \text{ negative definite.} \end{gathered} \tag{9}$$
  - With $\bar{u}^* := \mathcal{T}^T\hat{u}$ it is
  $$\begin{aligned} \bar{H}(\hat{u}(t)) &= \tfrac{1}{2}\hat{u}^T(t) S(h\kappa, h\nu)\hat{u}(t) \\ &= \frac{\varphi(h\kappa, h\nu)}{2h\sigma(\kappa\nu h^2)}\Big(y_{11}\bar{p}^{*^2}(t) + y_{22}\bar{q}^{*^2}(t)\Big) \;\equiv\; \bar{H}(u_0) \\ &= \frac{\varphi(h\kappa, h\nu)}{2h\sigma(\kappa\nu h^2)}\; \underbrace{\Big(\ell_{21}p_0^2 + (\ell_{22} - \ell_{11})p_0 q_0 - \ell_{12}q_0^2\Big)}_{=:E_0}, \end{aligned}$$
  i.e.,
  $$\frac{\bar{p}^{*^2}(t)}{\frac{E_0}{y_{11}}} + \frac{\bar{q}^{*^2}(t)}{\frac{E_0}{y_{22}}} \equiv 1. \tag{10}$$

Due to (9) the fractions $\frac{E_0}{y_{11}}$ and $\frac{E_0}{y_{22}}$ are positive. Because of $\hat{u} = \mathcal{T}\bar{u}^*$ the identity (10) shows that $\hat{u}(t)$ describes the ellipse with semiaxis $\sqrt{\frac{E_0}{y_{11}}}$ in p–direction and semiaxis $\sqrt{\frac{E_0}{y_{22}}}$ in q–direction rotated in mathematical positive direction about α.

- Algebraic computations based on the addition theorems of sine and cosine show that
$$\begin{aligned} y_{11} &= \tfrac{1}{2}\Big((\ell_{21} - \ell_{12}) \pm \sqrt{(\ell_{12} + \ell_{21})^2 + (\ell_{22} - \ell_{11})^2}\Big), \\ y_{22} &= \tfrac{1}{2}\Big((\ell_{21} - \ell_{12}) \mp \sqrt{(\ell_{12} + \ell_{21})^2 + (\ell_{22} - \ell_{11})^2}\Big) \end{aligned}$$
for $\ell_{22} - \ell_{11} \gtrless 0$.

- For the statement about the direction in which the ellipse is passed through the proof in the case $\ell_{11} = \ell_{22}$ is also valid. $\square$

There are some important consequences of this theorem.

Remark 7.

- *Taking into account that the solution of (7) to the initial condition* $\bar{u}(0) = u_0$ *is given as* $\bar{u}(t) = \exp(-tJS(h\kappa, h\nu))u_0$*, there is the relation* $Z(t) = \exp(-tJS(h\kappa, h\nu))$.
- *Let* $E = [e_+, e_-]$*, where* e_+, e_- *are eigenvectors of* $G(h\kappa, h\nu)$ *to the eigenvalues* $\frac{1}{2}(\Gamma_{11}(\kappa\nu h^2) + \Gamma_{22}(\kappa\nu h^2)) \pm i\sigma(\kappa\nu h^2)$*, and let*
$$log\, G(h\kappa, h\nu) := E \begin{bmatrix} i\varphi(h\kappa, h\nu) & 0 \\ 0 & -i\varphi(h\kappa, h\nu) \end{bmatrix} E^{-1} \tag{11}$$
a logarithm of $G(h\kappa, h\nu)$*. Then some extensive algebra shows that*
$$-JS(h\kappa, h\nu) = \tfrac{1}{h} log\, G(h\kappa, h\nu). \tag{12}$$
This means the perturbed Hamiltonian system can also be written as
$$\dot{\bar{u}} = \tfrac{1}{h} log\, G(h\kappa, h\nu)\, \bar{u} \tag{13}$$
(cf. Sanz–Serna, Calvo [14, p. 132]).
- *Because of* $\frac{\varphi(\kappa\nu h^2)}{h} = \sqrt{\kappa\nu} + O(h^d)$ *for* $h \to 0_+$ *and Theorem 4 it is*
$$\begin{aligned} S(h\kappa, h\nu) &= \begin{bmatrix} \kappa & 0 \\ 0 & \nu \end{bmatrix} + D(h), \ \textit{where} \\ D(h) &= \begin{bmatrix} O(r^{2[\frac{r+1}{2}]}) & O(r^{2[\frac{r}{2}]+1}) \\ O(r^{2[\frac{r}{2}]+1}) & O(r^{2[\frac{r+1}{2}]}) \end{bmatrix} \ \textit{for } h \to 0_+. \end{aligned} \tag{14}$$
- *If in* $Z(t), S(h\kappa, h\nu)$*, (11) instead of* $\varphi(\kappa\nu h^2)$ *the expression* $\varphi(\kappa\nu h^2) + 2k\pi$ *with* $k \in \mathbb{Z} \setminus \{0\}$ *is used, then Theorem 2, Theorem 6, (12), and (13) are also valid. But the frequency* $\frac{\varphi(\kappa\nu h^2)+2k\pi}{h}$ *is not in accordance with the frequency* $\sqrt{\kappa\nu}$ *in the solution of (4), i.e.,* $\lim_{h\to 0_+} \frac{\varphi(\kappa\nu h^2)+2k\pi}{h} \neq \sqrt{\kappa\nu}$.

An ls–symplectic RK method applied to (4) with a stepsize $h > 0$ such that $\begin{bmatrix} h\kappa \\ h\nu \end{bmatrix} \in \gamma$ is H–conserving, i.e., $H(u_m) = H(u_0)$ for all $m \in \mathbb{N}$. A RKN method that is not *induced by a RK method* (cf. Sanz–Serna, Calvo [14, p. 36–37]) does not conserve energy. But with the notations of Theorem 2 and (14) there are the following estimations.

Theorem 8. *If an r–th order RKN method that is not induced by a RK method is applied to (4) with a stepsize* $h > 0$ *such that* $\begin{bmatrix} h\kappa \\ h\nu \end{bmatrix} \in \gamma$*, then*
$$H(u_m) = H(u_0) + \tfrac{1}{2} u_0^T D(h) u_0 - \tfrac{1}{2} u_m^T D(h) u_m$$
and
$$\begin{gathered} \tfrac{1}{2} u_0^T D(h) u_0 - \tfrac{1}{2} \max_{0 \le t \le \frac{2\pi h}{\varphi(\kappa\nu h^2)}} u_0^T Z^T(t) D(h) Z(t) u_0 \\ \le\ H(u_m) - H(u_0)\ \le \\ \tfrac{1}{2} u_0^T D(h) u_0 - \tfrac{1}{2} \min_{0 \le t \le \frac{2\pi h}{\varphi(\kappa\nu h^2)}} u_0^T Z^T(t) D(h) Z(t) u_0 \end{gathered}$$
for all $m \in \mathbb{N}$.

Proof. The statement results from Theorem 6 by taking into account that

$$\bar{H}(u_m) = \bar{H}(u_0) = H(u_0) + \tfrac{1}{2}u_0^T D(h)u_0, \quad \bar{H}(u_m) = H(u_m) + \tfrac{1}{2}u_m^T D(h)u_m$$

for all $m \in \mathbb{N}$. $\square$

3.2 The Case $G(h\kappa, h\nu) = -I$

The stable case $G(h\kappa, h\nu) = -I$ is not included in the definition of symplectic stability, because of some problems. For the 2–stage Gauss method (cf. Dekker, Verwer [3, p. 64]) simple algebraic manipulations yield

$$G(x,y) = \begin{bmatrix} \dfrac{1-\frac{5}{12}yx+\frac{1}{144}(yx)^2}{1+\frac{1}{12}yx+\frac{1}{144}(yx)^2} & -y\dfrac{1-\frac{1}{12}yx}{1+\frac{1}{12}yx+\frac{1}{144}(yx)^2} \\ x\dfrac{1-\frac{1}{12}yx}{1+\frac{1}{12}yx+\frac{1}{144}(yx)^2} & \dfrac{1-\frac{5}{12}yx+\frac{1}{144}(yx)^2}{1+\frac{1}{12}yx+\frac{1}{144}(yx)^2} \end{bmatrix},$$

with

$$\mathcal{D} = \mathbb{R}^2 \quad \text{and} \quad \gamma = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : xy \in (0,12) \cup (12,\infty) \right\}.$$

For $h\kappa h\nu = 12$ it is $G(h\kappa, h\nu) = -I$ and $S(h\kappa, h\nu)$ is not defined. Taking into account that

- $\Gamma_{11} = \Gamma_{22}$,
- $$\frac{1-\frac{1}{12}z}{\sqrt{1-\left(\frac{1-\frac{5}{12}z+\frac{1}{144}z^2}{1+\frac{1}{12}z+\frac{1}{144}z^2}\right)^2}} = \frac{1}{\sqrt{z}}\frac{1-\frac{1}{12}z}{|1-\frac{1}{12}z|}\left(1+\tfrac{1}{12}z+\tfrac{1}{144}z^2\right)$$
  for $z > 0$, $z \neq 12$,
- $\arccos\left(\frac{1-\frac{5}{12}z+\frac{1}{144}z^2}{1+\frac{1}{12}z+\frac{1}{144}z^2}\right) = \pi$ for $z = 12$,

there are the relations

$$\begin{aligned} \lim_{h\nearrow\sqrt{\frac{12}{\kappa\nu}}} S(h\kappa, h\nu) &= \begin{bmatrix} \kappa\frac{3\pi}{\sqrt{12}} & 0 \\ 0 & -\nu\frac{3\pi}{\sqrt{12}} \end{bmatrix} \\ &\neq \begin{bmatrix} -\kappa\frac{3\pi}{\sqrt{12}} & 0 \\ 0 & \nu\frac{3\pi}{\sqrt{12}} \end{bmatrix} = \lim_{h\searrow\sqrt{\frac{12}{\kappa\nu}}} S(h\kappa, h\nu). \end{aligned}$$

This means, $S(h\kappa, h\nu)$ can not be defined in a convenient way for $\kappa\nu h^2 = 12$. Hence, backward error analysis is not possible for this value.

4. Numerical Examples

The harmonic oscillator

$$\dot{p} = -q, \qquad \dot{q} = p \tag{15}$$

with Hamiltonian $H(p,q) = \frac{1}{2}p^2 + \frac{1}{2}q^2$ and a solution with a frequency of 1, i.e., a 2π–periodic solution, is first integrated with the implicit midpoint rule (cf. Dekker, Verwer [3, p. 64]), here the frequency of the solution of the perturbed Hamiltonian system is investigated, second with the 1–stage PRK method of Ruth ([13]), here the degeneration of the orbit is investigated.

4.1 Implicit midpoint rule

The implicit midpoint rule also called 1–stage Gauss method is symplectic for every Hamiltonian system and it is

$$G(x,y) = \begin{bmatrix} \frac{1-\frac{1}{4}yx}{1+\frac{1}{4}yx} & -y\frac{1}{1+\frac{1}{4}xy} \\ x\frac{1}{1+\frac{1}{4}yx} & \frac{1-\frac{1}{4}xy}{1+\frac{1}{4}xy} \end{bmatrix}, \qquad \mathcal{D} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : xy \neq -4 \right\}$$

with $\gamma = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : xy > 0 \right\}$, i.e., the method is symplectically stable, it has order of dispersion of 2 and an error constant of $\frac{1}{12}$.

The movement of the circle with center $p = 1$, $q = 1$ in the p/q–plane (x in the figure) and radius $\frac{1}{5}$ is investigated. "After 12 steps of length $\frac{2\pi}{12}$ the circle should have returned to its original location", as Sanz–Serna and Calvo state ([14, p. 71]). But Theorem 2 and Theorem 6 show that the numerical solution produced by the method is the exact solution of a perturbed Hamiltonian system at discrete points and this solution has a frequency of $\frac{12}{2\pi} \arccos \frac{1-\frac{\pi^2}{144}}{1+\frac{\pi^2}{144}} = 0.978049...$, i.e., has a period of $4\pi^2 \Big/ \Big(12 \arccos \frac{1-\frac{\pi^2}{144}}{1+\frac{\pi^2}{144}}\Big) = 2\pi \cdot 1.02244....$ So, the size of the gap between the o and the x in Figure 1 is $2\pi\Big(2\pi \Big/ \Big(12 \arccos \frac{1-\frac{\pi^2}{144}}{1+\frac{\pi^2}{144}}\Big) - 1\Big) = 2\pi \cdot 0.02244...$ which means an angle of approximately 8°.

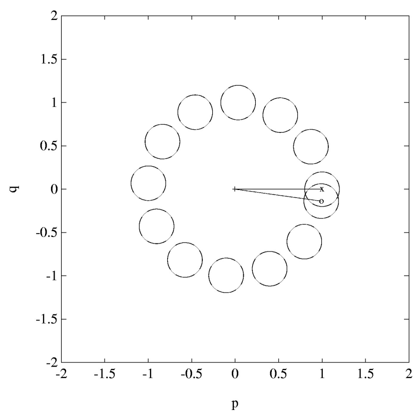

Figure 1

4.2 1–stage Ruth method

The first order method introduced by Ruth ([13]) is a PRK method which is symplectic for every separable Hamiltonian system and it is

$$G(x,y) = \begin{bmatrix} 1 & -y \\ x & 1-xy \end{bmatrix}, \qquad \mathcal{D} = \mathbb{R}^2$$

with $\gamma = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : xy \in (0,4) \right\}$, it has order of dispersion of 2 and an error constant of $-\frac{1}{24}$.

The solution of (15) with initial values $p_0 = 0.5$ and $q_0 = 0$ describes in the p/q–plane the circle with centre 0 and radius 0.5. According to [5] 2000 steps of the method are performed, with step size 0.1 and with steps size 1.9. Feng and Qin describe the shape of the ellipse for $h = 0.1$ as "the orbit appears as an ellipse close to the circle" and for $h = 1.9$ as "the orbit appears as a tilted oblate ellipse". With the previous theory precise statements are possible: For $h = 0.1$ the numerical solution lies on the ellipse with semiaxis 0.5129... in p–direction and semiaxis 0.4879... in q–direction and in the case $h = 1.9$ on the ellipse with semiaxis 2.2360... in p–direction and 0.3580... in q–direction. In both cases the ellipse is rotated in mathematical positive direction about 45° and passed through also in mathematical positive direction.

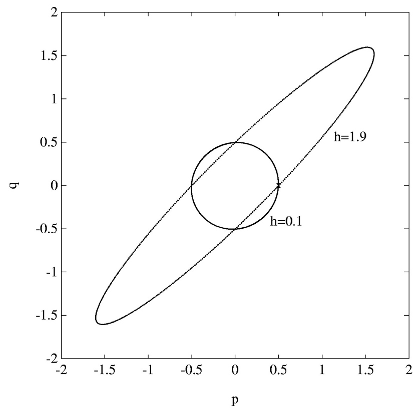

Figure 2

References

[1] R. Abraham, J. E. Marsden. Foundations of Mechanics. 2nd Edition. Addison–Wesley Publishing Company, Redwood City, 1978.

[2] V. I. Arnold. Mathematical Methods of Classical Mechanics. 2nd Edition. Springer–Verlag, New–York, 1989.

[3] K. Dekker, J. G. Verwer. Stability of Runge–Kutta Methods for Stiff Nonlinear Differential Equations. North–Holland, Amsterdam, 1984.

[4] K. Feng. The calculus of generating functions and the formal energy for Hamiltonian algorithms. Collected Works of Feng Kang (II) (1995), 284–302. National Defence Industry Press, Beijing.

[5] K. Feng, M. Z. Qin. Hamiltonian algorithms for Hamiltonian systems and a comparative numerical study. *Comput. Phys. Comm.* **65** (1991), 173–187.

[6] P. Görtz. Symplektische Stabilitätstheorie zur numerischen Integration von Hamilton–Systemen. Ph.D. Thesis, Karlsruhe, 1995.

[7] P. Görtz, R. Scherer. Reducibility and characterization of symplectic Runge–Kutta methods. *ETNA* **2** (1994), 194–204.

[8] P. Görtz, R. Scherer. Hamiltonian systems and symplectic integrators. *Nonlinear Analysis, Theory, Methods & Applications* **30** (1997), 1887–1892.

[9] E. Hairer. Backward analysis of numerical integrators and symplectic methods. *Ann. Numer. Math.* **1** (1994), 107–132.

[10] E. Hairer, S. P. Nørsett, G. Wanner. Solving Ordinary Differential Equations I: Nonstiff Problems. 2nd Edition. Springer–Verlag, Berlin, 1993.

[11] P. J. van der Houwen, B. P. Sommeijer. Explicit Runge–Kutta (–Nyström) methods with reduced phase error for computing oscillating solutions. *SIAM J. Numer. Anal.* **24** (1987), 595–617.

[12] P. J. van der Houwen, B. P. Sommeijer. Diagonally implicit Runge–Kutta–Nyström methods for oscillatory problems. *SIAM J. Numer. Anal.* **26** (1989), 414–429.

[13] R. D. Ruth. A canonical integration technique. *IEEE Transactions on Nuclear Science* **NS–30** (1983), 2669–2671.

[14] J. M. Sanz–Serna, M. P. Calvo. Numerical Hamiltonian Problems. Chapman & Hall, London, 1994.

[15] Y. F. Tang. Formal energy of a symplectic scheme for Hamiltonian systems and its application (I). *Computers Math. Applic.* **27**(7) (1994), 31–39.

[16] D. Wang. Some aspects of Hamiltonian systems and symplectic algorithms. *Physica D* **73** (1994), 1–16.

[17] H. Yoshida. Recent progress in the theory and application of symplectic integrators. *Celes. Mech.* **56** (1993), 27–43.