Journal of Computational Mathematics
Vol.29, No.2, 2011, 167–184.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1009-m3108

FINITE ELEMENT METHOD WITH SUPERCONVERGENCE FOR NONLINEAR HAMILTONIAN SYSTEMS*

Chuanmiao Chen
College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China
Email: cmchen@hunnu.edu.cn
Qiong Tang
Department of Information and Computation, Hunan University of Technology, Zhuzhou 412008, China
Email: zzgxysx@163.com
Shufang Hu
College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China
Email: shufang-hu163@163.com

Abstract

This paper is concerned with the finite element method for nonlinear Hamiltonian systems from three aspects: conservation of energy, symplicity, and the global error. To study the symplecticity of the finite element methods, we use an analytical method rather than the commonly used algebraic method. We prove optimal order of convergence at the nodes t_n for mid-long time and demonstrate the symplecticity of high accuracy. The proofs depend strongly on superconvergence analysis. Numerical experiments show that the proposed method can preserve the energy very well and also can make the global trajectory error small for long time.

Mathematics subject classification: 65N30.
Key words: Nonlinear Hamiltonian systems, Finiteelement method, Superconvergence, Energy conservation, Symplecticity, Trajectory.

1. Introduction

We consider the nonlinear Hamiltonian systems

$$z_t = -JH_z, \quad z(0) = z_0, \tag{1.1}$$

where

$$H_z = \begin{pmatrix} H_p \\ H_q \end{pmatrix}, \quad J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{1.2}$$

$z = (p,q)^T = (p_1, \cdots, p_n; q_1, \cdots, q_n)^T$, $H(z) = H(p,q)$ is a real-valued smooth function and J is a skew-symmetric matrix of order $2n$. Obviously, $J^T = J^{-1} = -J, \quad J^2 = -I_{2n}$. In application, the Hamiltonian $H(z)$ is often the total energy. Hamiltonian systems have two important properties: conservation and symplecticity. These properties are the hallmark of Hamiltonian systems.

* Received April 25, 2009 / Revised version received April 14, 2010 / Accepted April 30, 2010 / Published online November 20, 2010 /

Symplectic geometry in phase space R^{2n} is the mathematical foundation to study Hamiltonian systems. Let $x = (x_1, \cdots, x_n, x_{n+1}, \cdots, x_{2n})^T \in R^{2n}$. Then symplectic structure is defined by a skew-symmetric bilinear inner product

$$[x, y] = (x, Jy) = \sum_{j=1}^{n} (x_j y_{n+j} - x_{n+j} y_j), \qquad x, y \in R^{2n}. \tag{1.3}$$

Hence $[x, x] = (x, Jx) = 0$. In the symplectic space, a linear operator A is symplectic iff $A^T JA = J$. All solutions $z(t)$ of (1.1) form a symplectic group with one-parameter. These solutions have an important symplecticity property (see Section 5)

$$\left(\frac{Dz(t)}{Dz_0}\right)^T J \left(\frac{Dz(t)}{Dz_0}\right) = J, \qquad 0 \le t < \infty. \tag{1.4}$$

Moreover, multiplying Eq. (1.1) by J and z_t, we have the energy conservation

$$0 = \int_0^t J(z_t + JH_z) z_t dt = -\int_0^t H_z z_t dt = -H\big(z(t)\big)\Big|_0^t. \tag{1.5}$$

It is important to construct discrete algorithms which preserve these basic properties. Ruth [1] and Feng [2] have originally proposed the sympletic geometry algorithms which preserve the global symplectic structure and have tracking ability over long times. Feng and his co-authors then published several important works afterwards, see, e.g., [3-6]. Later on many symplectic schemes are studied by Chinese scholars, such as the partitioned algorithm (Sun [7]), multi-step algorithm (Tang [8]), volume-preserving algorithm (Shang [10,11]). Recent work can be found in [9,12,21] and a review [14].

Under the influence of Feng's work, several new symplectic algorithms are developed. For example, the symplectic Runge-Kutta method (SRK) is proposed by Sanz-Serna, Lasagni and Suris (see [15-18]). Later on the symplectic algorithms are also generalized to deal with partial differential systems.

Many scholars pointed out that the energy conservation is more important at certain times, see, e.g., Stuart *et al.* [19] (pp.583-584,642-644) and Hairer *et al.* [20] (p.12). So we turn to the finite element method (FEM). It is found that the continuous FEM always preserves the energy, and is approximately symplectic [22,23]. FEM is an exact orthogonal projection, which makes it possible to explore its refined properties, such as superconvergence, long-time error, approximate symplecticity and so on. These properties describe another kind of the structure different from the symplectic algorithms. Besides, the spectrum algorithm is also an orthogonal projection, see Tang and Xu [13]. It should be pointed out that the symplectic collocation method and SRK are equivalent under some conditions (see [20], p.27), which may be considered to be the approximately orthogonal projection based on some fixed quadrature. This quadrature makes the symplectic collocation method and SRK to possess the symplecticity, and to approximately preserve the energy. Therefore it is suggested that both SRK and FEM belong to the same setting of the orthogonal projection, but only with different quadratures.

In Table 1.1, we compare three properties of three algorithms: SFD (symplectic finite difference algorithm), SRK (symplectic Runge-Kutta method), and FEM (continuous finite element method).

In addition to preserving the symplecticity and energy, there is a third criterion to evaluate an algorithm, i.e., small deviations of computational trajectory after long times, which is possibly more important in applications. We now give a proposition as follows:

Table 1.1: Comparison of three algorithms.

Three Properties	SFD	SRK	FEM
Energy conservation	approx.	approx.	exactly
Symplecticity(linear case)	exactly	exactly	exactly
Symplecticity(nonlinear case)	exactly	exactly	approx.
Long-time deviation of trajectory	small	smaller	smaller

Proposition 1.1. *A good algorithm of $2m$-order accuracy for the Hamiltonian system should be of the optimal error at node t_n*

$$|(z-Z)(t_n)| \le Ct_n h^{2m},$$

for the long-time $t_n \le ch^{-2m}$, where the constant C is independent of h, t_n.

Here $T = ch^{-2m}$ is called the long-time, because at this time the deviation $|(z-Z)(T)| \approx cC$ is already the quantity independent of h, and then further computation is meaningless. Numerical experiments show that three algorithms mentioned above satisfy this proposition for long-time (although their errors are different), however its proof is quite difficult and challenging. Up to now, most of studies are confined in a short time. An excellent long-time result of SRK for Newton system under Siegel's diophantine condition is included in Hairer [19]. Unfortunately, this condition is not satisfied by the Kepler system, see [20], p.354. In this paper, we focus on the m-degree finite elements method and it is proved that FEM for nonlinear Hamiltonian system under some reasonable conditions satisfies this proposition for a mid-long time $t_n \le ch^{-m}$ (Theorem 2.1) and is essentially symplectic (Theorem 5.1). Besides, FEM for linear system satisfies the proposition for long-times and is symplectic (Theorem 4.1). To study these properties, we have used an analytical method rather than an algebraic method.

We close this section by mentioning that Karakashian and Makridakis [24] studied the space-time continuous finite element methods for nonlinear Schrödinger system and analyzed superconvergence at time node (for short-times). They proposed a remarkable KM-trick to cancel the influence of Laplacian operator. Of course, this scheme also preserves the energy.

2. Basic Assumption and Main Results

Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a partition of $G = (0, T)$, with $K_j = (t_j, t_{j+1})$, $h_j = t_{j+1} - t_j$. Assume that the partition is uniform($h_j = h$). Denote by S^h the m-degree continuous finite element space. Each m-degree polynomial in K_j has $m+1$ parameters, but only m freedoms, as its starting value at point t_j is given. We define the finite element solution $Z = (P, Q)^T \in S^h$ of (1.1) satisfying the orthogonal condition in K_j

$$\int_{K_j} (Z_t + JH_z(Z))\xi dt = 0, \qquad \xi \in P_{m-1}, \tag{2.1}$$

where P_{m-1} is a set of $m-1$-degree polynomials. Taking $\xi = Z_t$ in (2.1), we get

$$0 = \int_{K_j} J(Z_t + JH_z(Z))Z_t dt = -\int_{K_j} H_z(Z)Z_t dt = -H(Z(t))\Big|_{t=t_j}^{t=t_{j+1}}. \tag{2.2}$$

Lemma 2.1. (Energy Conservation [22,23]). *Continuous finite element solutions at node t_n always preserve the energy for nonlinear Hamiltonian systems.*

This suggests that the shape of trajectory $Z(t_n)$ in phase plane is always unchanged. This is the most important advantage of FEM.

In this paper we shall introduce a linearized equation of (1.1),

$$w_t = Bw, \qquad w(0) = w_0, \qquad B = -JH_{zz}(z(t)). \tag{2.3}$$

To study the long-time behavior of FE, we make the following

Assumption 2.1. (Basic Assumption). *The solution $z(t)$ of nonlinear system* (1.1) *and the solution $w(t)$ of* (2.3) *are uniformly bounded for any t,*

$$|z(t)| \le C_0(z_0), \quad |w(t)| \le C_1|w(0)|, \qquad 0 \le t < \infty. \tag{2.4}$$

Under the assumption, the energy surface $H(z) = H(z_0)$ is close and the solution $z(t)$ is periodic. For linear system, $H(p, q) = C$ should be an elliptic surface rather than a parabolic or hyperbolic one. For nonlinear system, for example, the Heissen matrix H_{zz} is positive definite. Example 7.2 shows that local loss of the convexity of $H(z)$ is admissible.

The assumption directly derives that all high order derivatives are uniformly bounded,

$$|D_t^k z(t)| \le C_k(z_0), \qquad |D_t^k w(t)| \le C_k|w(0)|. \tag{2.5}$$

Our main result can be stated as the following theorem.

Theorem 2.1. (Deviation of Trajectory). *Under the basic assumption for nonlinear system* (1.1)*, the deviation of m-degree FE solution $Z(t)$ at node t_n has optimal superconvergence*

$$|Z(t_n) - z(t_n)| \le Ct_n h^{2m}, \tag{2.6}$$

which is valid for a mid-long time $t_n \le ch^{-m}$ and the constants C is independent of h, t_n.

Remark 2.1. Theorem 2.1 shows that the error $|e(t_n)|$ grows linearly with t_n. When t_n is large enough, the curve of $Z(t)$ will be far away from that of the true solution $z(t)$, although its shape is similar to $z(t)$. See Figs. 7.2,7.3,7.6-7.8.

In Section 5, we shall propose an equality for the symplecticity of FEM and prove the essential symplecticity (Theorem 5.1).

3. The Orthogonal Projections in an Element

We use the linear transformation $t = hs$, which maps $E = (-1, 1)$ onto $K = (-h, h)$. Then $g(t)$ becomes $g(hs)$, which is still denoted by $g(s)$. Obviously, $D_s^i g = h^i D_t^i g = \mathcal{O}(h^i)$.

Introduce the Legendre polynomials in E

$$l_0 = 1, \quad l_1 = s, \quad l_2 = \frac{1}{2}(s^2 - 1), \quad l_3 = \frac{1}{2}(5s^3 - 3s), \quad \cdots, \quad l_n = \gamma_n D_s^n (s^2 - 1)^n,$$

where $\gamma_n = 1/(2n)!!$. It is known that the inner product $(l_i, l_j) = 0, i \ne j$, and $(l_j, l_j) = 2/(2j+1) = c_{j+1}$.

Integrating the Legendre polynomials over $(-1, s)$, we get the M-type polynomials

$$M_0 = 1, \quad M_1 = s, \quad M_2 = (s^2 - 1)/2, \cdots,$$

$$M_{n+1}(s) = \int_{-1}^{s} l_n(s)ds = \gamma_n D_s^{n-1}(s^2-1)^n,$$

which are quasi-orthogonal, i.e., $(M_i, M_j) \neq 0$ if $j - i = 0, \pm 2$, else $(M_i, M_j) = 0$. Obviously $M_n(\pm 1) = 0, n \geq 2$.

To construct an L-type projection, we expand $w(s)$ as a Legendre series:

$$w(s) = \sum_{j=0}^{\infty} b_j l_j(s), \quad b_j = j_1(w, l_j)_E, \qquad j_1 = j + \frac{1}{2}, \tag{3.1}$$

where the coefficient are given by integration by parts for $0 \leq i \leq j$:

$$b_j = j_1 \gamma_j (-1)^i \big(D_s^i w, D_s^{j-i}(s^2-1)^j\big)_E = \mathcal{O}(h^i)|D_t^i w|_K, \qquad |w|_K = \max_{t\in K}|w(t)|.$$

The sum of the first $(m-1)$-degree and the remainder are given by

$$w_L \equiv L_h w_L = \sum_{j=0}^{m-1} b_j l_j(s), \qquad r = w - w_L = \sum_{j=m}^{\infty} b_j l_j(s) \perp P_{m-1}(s), \tag{3.2}$$

respectively. By the Bramble-Hilbert lemma, we have the maximum norm estimate

$$|w - w_L|_K = \max_{t\in K}\big|(w - w_L)(t)\big| \leq Ch^m |D_t^m w|_K.$$

Define an integral operator S,

$$S_t w(t) = \int_{-h}^{t} w(t)dt = h\int_{-1}^{s} w(s)ds = hS_s w(s), \qquad |S_t w|_K \leq 2h|w|_K.$$

For the remainder r of the L-type projection, we have

$$\begin{aligned} &S_t^i r(t) = h^i \sum_{j=m}^{\infty} b_j \gamma_j \partial_s^{j-i}(s^2-1)^j \perp P_{m-1-i}, && 0 \leq i \leq m-1,\\ &|S_t^i r(t)|_K \leq Ch^i\big|r(t)\big|_K \leq Ch^{m+i}|D_t^m w|_K, \ \ S_t^i r(\pm h) = 0, \ \ 0 \leq i \leq m. \end{aligned}$$

Secondly expanding $u_s(s)$ as a L-type series, integrating in s and taking $b_0 = (u(-1)+u(1))/2$, we get a M-type series [26,27]:

$$u(s) = \sum_{j=0}^{\infty} b_j M_j(s), \quad b_{j+1} = j_1(u_s, l_j)_E = \mathcal{O}(h^i)|D_t^i u|_K, \qquad 1 \leq i \leq j+1. \tag{3.3}$$

The sum of the first m terms and the remainder are

$$u_m = Q_h u = \sum_{j=0}^{m} b_j M_j(s), \qquad R = u - u_m = \sum_{j=m+1}^{\infty} b_j M_j(s) \perp P_{m-2}(s), \tag{3.4}$$

respectively. Because $b_0 = (u(1) + u(-1))/2$ and $b_1 = (u_t, l_0)/2 = (u(1) - u(-1))/2$, we have $u_m(\pm 1) = u(\pm 1)$ which guarantees that the piecewise m-degree polynomial u_m constructed in each element is continuous in the interval $G = (0, t_N)$.

The remainder $R = u - u_m$ has the following properties:

$$R(\pm 1) = 0; \quad D_s R \perp P_{m-1}, \ m \geq 1; \quad R \perp P_{m-2}, \qquad m \geq 2, \tag{3.5}$$

$$S_t^i R = h^i \sum_{j=m+1}^{\infty} b_j \gamma_{j-1} D_s^{j-2-i}(s^2-1)^{j-1} \perp P_{m-2-i}, \quad 0 \le i \le m-2,$$
$$|S_t^i R|_K \le Ch^{m+1+i}|D_t^{m+1}u|_K, \quad S_t^i R(\pm 1) = 0, \qquad 0 \le i \le m-1.$$

Two projection operators L_h and Q_h, and their orthogonality play an important role in the study of superconvergence.

4. Proof of Theorem 2.1

Assume that z and $Z \in S^h$ are the exact solution of (1.1) and m-degree finite element solution of (2.1), respectively. The error $e = z - Z$ satisfies the orthogonal relation in $K = (t_j, t_{j+1})$

$$(e_t, \xi)_K = -J\big(H_z(z) - H_z(Z), \xi\big)_K, \qquad e(0) = 0, \quad \xi \in P_{m-1}. \tag{4.1}$$

Set $z^s = Z + se, 0 \le s \le 1$, $z^0 = Z, z^1 = z$ and $\phi(s) = -JH_z(z^s)$. Then

$$\phi'(s) = -JH_{zz}(z^s)e, \qquad \phi'(1) = Be, \qquad B = -JH_{zz}(z(t)).$$

Using

$$\phi(1) - \phi(0) = \int_0^1 \phi'(s)ds = \phi'(1) - \int_0^1 \phi''(s)sds,$$

we get

$$-J\big(H_z(z^1) - H_z(z^0)\big) = B(z)e + b(t)e^2, \qquad b(t) = -\int_0^1 JH_{zzz}(z^s)sds, \qquad \big|b(t)\big| \le C.$$

Consequently, we can get an error equation

$$(e_t, \xi)_K = (Be, \xi)_K + (be^2, \xi)_K, \qquad e(0) = 0, \quad \xi \in \ P_{m-1}. \tag{4.2}$$

Denoting by Z_I the m-degree M-projection of z and decomposing the error as $e = z - Z = (z - Z_I) - (Z - Z_I) = R - \theta$. Then $\theta = z - z_I \in S^h$ satisfies

$$(\theta_t, \xi)_K = (B\theta, \xi)_K - (BR, \xi)_K + (be^2, \xi)_K, \qquad \xi \in P_{m-1}, \quad \theta(0) = 0, \tag{4.3}$$

where $R_t \perp P_{m-1}$ is used.

Our new idea is to use the orthogonality correction technique proposed by Chen in 1999 (also see [27,28]). We decompose $\theta = u + v$, where u is the local error and v is the global error. Firstly define the local correction in K

$$u = \sum_{j=2}^{m} a_j M_j(s), \qquad u(t_j) = u(t_{j+1}) = 0, \quad t = \bar{t}_j + hs \in K, \quad s \in E,$$

satisfying

$$(u_t - Bu, \xi)_K = -(BR, \xi)_K, \qquad \xi = l_{i-1}(t), \quad i = 2, 3, \cdots, m, \tag{4.4}$$

in order to cancel the terms in BR as much as possible . It remains that the global error $v \in S^h$ satisfies in $G = (0, t_n)$

$$(v_t - Bv, \xi)_K = r_K(\xi) + (be^2, \xi)_K, \quad \xi = l_{i-1}(t), \qquad 1 \le i \le m, \quad v(0) = 0. \tag{4.5}$$

The remainder

$$r_K(\xi) = -(u_t, \xi)_K + (Bu + BR, \xi)_K \tag{4.6}$$

is simplified to $r_K(l_{i-1}) = 0, 2 \le i \le m$, and

$$r_K(l_0) = h\sum_{j=2}^{m} a_j(B, M_j)_E + h\sum_{j=m+1}^{\infty} b_j(B, M_j)_E = \mathcal{O}(h^{2m+1})|D_t^{m+1}u|_K, \tag{4.7}$$

which will be proved later. The proof of Theorem 2.1 consists of three parts.

4.1. Local error u and orthogonality correction

Taking $\xi = u_t$ in (4.4) and deducing $|u_t|_K$, we have maximum norm estimate

$$|u_t|_K \le C|u|_K + C|R|_K \le C|u|_K + Ch^{m+1}.$$

Using $u(t) = \int_{t_j}^{t} u_t dt$ yields

$$|u|_K \le Ch|u_t|_K \le Ch|u|_K + Ch^{m+2}, \qquad t \in K.$$

When h is small enough, we can cancel $Ch|u|$ on the right side and get an optimal estimate

$$|u|_K \le Ch^{m+2}. \tag{4.8}$$

To get more refined estimate of u, transforming K into $E = (-1, 1)$, (4.4) becomes a linear algebraic system

$$\sum_{j=2}^{m} k_{ij}a_j \equiv c_i a_i - h\sum_{j=2}^{m} a_j\big(BM_j(s), l_{i-1}(s)\big)_E = \eta_i, \qquad i = 2, 3, \cdots, m, \tag{4.9}$$

where $R \perp P_{m-2}$ and integrating by part is used. Observe

$$\eta_i = h(R, B^T l_{i-1}) = (-1)^{m-1}h\big(S_s^{m-1}R, D_s^{m-1}(B^T l_{i-1})\big) = \mathcal{O}(h^{2m+2-i}).$$

In virtue of the quasi-orthogonality of M_j, $(M_j, l_i) \ne 0,\ j = i, i-2$, else 0, (4.9) is absolutely diagonally dominant, i.e., $k_{ii} = c_i + \mathcal{O}(h^2)$ positive and other elements $k_{ij} = \mathcal{O}(h^{|i-j|}), i \ne j$. For sufficiently small h, we can get the useful estimates

$$a_i = \mathcal{O}(h^{2m+2-i}), \qquad i = 2, 3, \cdots, m. \tag{4.10}$$

4.2. Global error v.

Taking $\xi = v_t$ in (4.5), we have the maximum norm estimate

$$|v_t|_K \le C|v|_K + Ch^{2m+1} + C|e|_K^2.$$

Using

$$v(t) = v_j - \int_t^{t_n} v_t dt \quad \text{and} \quad |e|_K = |R - u - v|_K \le Ch^{m+1} + |v|_K,$$

we have

$$|v|_K \le |v_j| + h|v_t|_K \le |v_j| + Ch|v|_K + Ch^{2m+2} + Ch|v|_K^2,$$

which gives, when h is suitably small, that

$$|v|_K \le C|v_j| + Ch^{2m+2} + Ch|v|_K^2.$$

Assuming that $|v|_K \le C$, we get an important estimate

$$|v_j| \le |v|_K \le C|v_j| + Ch^{2m+2}, \tag{4.11}$$

i.e., $|v|_K$ and v_j are of the same order.

4.3. Nodal error $|v_n|$.

Assume that $|v_n| = \max_{1\le j\le n} |v_j|$. Construct a conjugate problem

$$w_t + B^T w = 0, \qquad t \in G = (0, t_n), \quad w(t_n) = v_n. \tag{4.12}$$

By the basic assumption, we have the uniform bounds

$$|D_t^l w(t)|_G \le C|v_n|, \qquad l = 0, 1, 2, \cdots .$$

Denote by w_L the $(m-1)$-degree L-type projection of w in K, whose remainder

$$r = w - w_L \perp P_{m-1}, \quad |r|_K \le Ch^m |D_t^m w|_K \le Ch^m |v_n|, \qquad t \in K.$$

Integrating by parts leads to

$$I = (v_t - Bv, w)_G = (vw)(t_n) - (v, w_t + B^T w)_G = |v_n|^2.$$

On the other hand, using $r \perp v_t$ and Eq. (4.5) yields

$$|v_n|^2 = (v_t - Bv, r)_G + (v_t - Bv, w_L)_G = -(Bv, r)_G + r_G(w_L) + (be^2, w_L)_G, \tag{4.13}$$

where

$$\big|(be^2, w_L)_G\big| \le Ct_m |e|_G^2 |v_n|, \qquad |e|_G \le Ch^{m+1} + |v|_G,$$
$$|r_G(w_L)| \le Ch \sum_{i=1}^{n} \bigg(\sum_{j=2}^{m} h^{2m+2-j} \big|(B, M_j)\big| + h^{m+1} \big|(B, M_{m+1})\big| \bigg) \le Ct_n h^{2m}.$$

In virtue of Theorem 6.1 in Section 6, we get a long-time estimate

$$\big|(Bv, r)_G\big| \le Ct_n h^{2m} |v|_G \ |v_n| + Ct_n \Big(h^{2m} + |v|_G^2 \Big) |v_n|,$$

(in general, only $|(Bv, r)_G| \le Ct_n |v|_G \ |r|_G$). Reducing $|v_n|$, we have

$$|v_n| \le Ct_n h^{2m} |v|_G + Ct_n h^{2m} + Ct_n |v|_G^2.$$

and, using (4.11),

$$|v|_G \le C|v_n| + Ch^{2m+2} \le Ct_n h^{2m} |v|_G + Ct_n h^{2m} + Ct_n |v|_G^2.$$

If confining the maximum time t_n (called the long-time) such that $\gamma = Ct_n h^{2m} < \frac{1}{2}$, we can cancel $Ct_n h^{2m} |v|_G$ on the right side and get a long-time estimate

$$|v_n| \le |v|_G \le Ct_n h^{2m} + Ct_m |v|_G^2, \quad t \in G = (0, t_n). \tag{4.14}$$

We need the following simple estimate (also see Remark 2.1),

Lemma 4.1. *Assume that $y \geq 0$ satisfies $y \leq a + by^2, \quad a, b > 0, \quad 4ab \leq 1$. Then $y \leq 2a$.*

From (4.14) we directly get

$$|v_n| \leq |v|_G \leq 2Ct_nh^{2m}, \qquad \text{if } \; C^2t_n^2h^{2m} < \frac{1}{4}, \tag{4.15}$$

which is valid only for a mid-long time $t_n \leq ch^{-m}$.

Finally, noting that $e_n = R_n - u_n - v_n = -v_n$ at nodes t_n, we get $|e_n| = |v_n| \leq Ct_nh^{2m}$. Hence, Theorem 2.1 is proved.

Remark 4.1. Lemma 2.1 can be proved by a monotone increasing and bounded iterative sequence $y_0 = a, \cdots, y_{n+1} = a + by_n^2, \cdots$, whose limit Y satisfies $Y = a + bY^2$ and a smaller root $Y = 2a/(1 + \sqrt{1-4ab}) \leq 2a$. We recall that to deduce the term $C|v|^2$ in $|v| \leq a + b|v|^2$, Frehse-Rannacher [25] have once used the continuation method, a more complicated argument. Obviously the Lemma 2.1 is direct and simple.

Note that for linear system, the term $|v|^2$ in (4.14) disappears and (4.15) is valid for long-time $t_n \leq ch^{-2m}$. So we get an interesting result as follows:

Theorem 4.1. *For linear system, the m-degree continuous finite element $Z(t_n)$ at node $t = t_n$ is symplectic and $|(z - Z)(t_n)| \leq Ct_mh^{2m}$ for the long-time $t_m \leq ch^{2m}$.*

Proof. It is sufficient to discuss the symplecticity. For linear system $z_t = Bz$ with $B = -JL$, where L is a symmetric positive definite matrix of order $2n$, its exact solution is written as $z(t) = e^{Bt}z_0$. The m-degree finite element $Z(t)$ in the first node $t_1 = h$ can be expressed by the Cramer law

$$Z(h) = Q_m(hB)^{-1}P_m(hB)z_0,$$

where P_m, Q_m are m-degree polynomials of the matrix hB. On the other hand, we have the highest order superconvergence at the first node $t_1 = h$

$$|z(h) - Z(h)| = |e(h)| \leq C(h|B|)^{2m+1}|z_0|, \tag{4.16}$$

which shows that $Z(h)$ is $2m$-order diagonally Pade approximation to e^{Bh}. Consequently, $Z(h)$ is symplectic. $\square$

5. The Essential Symplecticity

We recall the proof of (1.4). Setting the derivatives $z'(t) = \frac{Dz(t)}{Dz_0}$, by (1.1) we have

$$z'_t = -JH_{zz}z', \qquad z'(0) = I_{2n}, \qquad A(z) = H_{zz}(z),$$

where $A(z)$ is a symmetrical $2n \times 2n$ square matrix. Direct calculation gives

$$\begin{aligned} D_t(z'^TJz') &= (z'_t)^TJz' + z'^TJz'_t = (-JAz')^TJz' + z'^TJ(-JAz') \\ &= -z'^TA^TJ^TJz' - z'^TJ^2Az' = -z'^TAz' + z'^TAz' = 0. \end{aligned}$$

Then we get $z'^TJz' = J$ for any t.

To study the symplecticity of the discrete scheme, we follow Feng's idea to investigate whether the partial derivatives $Z' = \frac{DZ_n}{Dz_0}$ at nodes t_n are symplectic. Differentiating (2.1) leads to

$$(Z'_t, \xi) = \big(- JH_{zz}(Z)Z', \xi\big), \qquad Z'(0) = I_{2n}, \qquad A(Z) = H_{zz}(Z), \tag{5.1}$$

where its transposition is similar, i.e.,

$$\big((Z'_t)^T, \xi\big) = \Big(\big(- JH_{zz}(Z)Z'\big)^T, \xi\Big), \qquad Z'(0) = I_{2n}.$$

We investigate the following integral

$$\omega_n = (Z'^T JZ')(t_n) - J = \int_0^{t_n} D_t(Z'^T JZ')dt = (Z_t'^T, JZ') + (Z'^T J, Z'_t).$$

The matrix $Z'(t)$ can be expressed by the Legendre expansion in the element K_j

$$Z'(s) = \sum_{j=0}^{m} F_j l_j(s), \quad Y = \sum_{j=0}^{m-1} F_j l_j(s) \in P_{m-1}, \quad r = Z' - Y = F_m l_m(s) \perp P_{m-1}.$$

Using Eq. (5.1) of Z', we have

$$\begin{aligned} \omega_n &= \big((Z'_t)^T, JY\big) + (Y^T J, Z'_t) = \big((-JAZ')^T, JY\big) + (Y^T J, -JAZ') \\ &= (-Z'^T A^T J^T, JY) + (Y^T J, -JAZ') = -(Z'^T A, Y) + (Y^T, AZ') \\ &= -(r^T A, Y) + (Y^T, Ar) - (Y^T A, Y) + (Y^T, AY), \end{aligned}$$

where the last two terms disappear. So we get an important equality

$$\omega_n = Z'(t_n)^T JZ'(t_n) - J = \sum_{l=1}^{n} \sum_{j=0}^{m-1} h\Big(- F_m^T A_j F_j + F_j^T A_j F_m\Big)\Big|_{K_l}, \tag{5.2}$$

where the square matrix is given by

$$A_{mj} = \int_E A(s) l_m(s) l_j(s) ds, \qquad j = 0, 1, \cdots, m-1.$$

For linear system, A is a constant matrix. Then $A_j = 0$ and $\omega_n = 0$ by (5.2). This is the simplest proof of the symplecticity.

For nonlinear system, the $A(Z)$ is variable. The finite element Z in general is not symplectic, but is of symplecticity of high accuracy (or called essentially symplectic). For the finite element solutions Z and Z', we make the following assumption.

Assumption 5.1. *In an interval $G = (0, T)$, the finite element solutions $Z(t), Z'(t)$ and their derivatives are uniformly bounded*

$$|D_t^l Z(t)| \le C_0, \quad |D_t^l Z'(t)| \le C_1, \quad l = 0, 1, \cdots, m, \quad 0 \le t \le T \le ch^{-2m}. \tag{5.3}$$

Remark 5.1. By the basic assumption, the finite element $Z(t)$ preserves the energy at node t_n, $H(Z(t_n)) = H(Z_0)$, then $Z(t_n)$ and $Z(t)$ are uniformly bounded. Furthermore, we can prove that their derivatives $D_t^l Z(t)$ are also uniformly bounded. Besides, if A is constant, symmetrical, and positive definitive, we can prove that $D_t^l Z'(t)$ are uniformly bounded. But in the case of variable A, its proof is very difficult. Hence, in this paper we temporarily accept Assumption 5.1.

Theorem 5.1. (Essential Symplecticity). *Assume that the finite element solutions $Z(t)$ and its derivative $Z'(t)$ satisfy Assumption* 5.1 *for long time $T = ch^{-2m}$. Then there is the symplecticity deviation of high accuracy for long time*

$$|Z'^T(t_n)JZ'(t_n) - J| \le CC_1^2 t_n h^{2m}, \quad 0 \le t_n \le T. \tag{5.4}$$

Proof. Under Assumption 5.1, we have in each element K_l

$$|F_j| \le CC_1 h^j, \quad |A_j| \le Ch^j, \qquad 0 \le j \le m.$$

So (5.2) directly leads to the following estimate

$$|\omega_n| \le CC_1^2 t_n h^{2m}, \qquad t_n = nh,$$

which completes the proof of the theorem. □

It should be pointed out that the proof of Theorem 5.1 only depends on the finite element solution Z, Z', and independent of the true trajectory z, z' and is Theorem 2.1,whereas the proof of the latter may be the most difficult one.

6. A Refined Estimate for the Global Error

Returning to the global error, $v \in S^h$ satisfies (4.5) and (4.13), i.e.,

$$(v_t, \xi)_G = (Bv, \xi)_G + r_G(\xi) + (be^2, \xi)_G, \qquad |v_n|^2 = -(Bv, r)_G + r_G(w_L) + (be^2, w_L)_G,$$

where w is defined in (4.12), $r = w - w_L \perp P_{m-1}$,

$$|r| \le Ch^m |D_t^m w| \le Ch^m |v_n| \quad \text{and} \quad \big|r_G(w_L)\big| \le Ct_n h^{2m}.$$

We prove a refined estimate as follows:

Theorem 6.1. *For $v \in S^h$ satisfying* (4.5) *and $r = w - w_L \perp P_{m-1}$, there is a refined estimate in any interval $G = (0, t_n)$*

$$\big|(Bv, r)_G\big| \le Ct_n\big(|v|_G + h^{2m} + |e|_G^2\big)h^m |r|_G, \qquad |r|_G \le Ch^m |v_n|, \tag{6.1}$$

where constant C is independent of h, t_n.

Proof. We shall repeatedly use the following techniques: integrating by parts to get $S_t^m r = \mathcal{O}(h^m)|r|$ and substituting the derivatives v_t by the original equation (4.5). Actually the integral operator S_t is used in each element K_j. The proof is completed in the whole interval $G = (0, t_n)$. For simplicity the low-index G is omitted.

For $m \ge 1$, using S_t and orthogonality of r, we can transform

$$(Bv, r) = -(B_t v, S_t r) - (Bv_t, S_t r) = I_1 + I_2. \tag{6.2}$$

Obviously

$$|I_1| = \big|(B_t v, S_t r)\big| \le Ct_n |v| Ch|r|.$$

To estimate I_2, set $F_1 = B^T S_t r$ and its $(m-1)$-degree L-type projection $f_1 = L_h(F_1)$. Obviously $r_1 = F_1 - f_1 \perp P_{m-1}$ and

$$|f_1| \le C|F_1| \le C|S_t r| \le Ch|r|, \qquad |r_1| \le C|F_1| \le Ch|r|.$$

By the original equation (4.5), we have

$$
\begin{aligned}
-I_2 &= (Bv_t, S_t r) = (v_t, F_1) = (v_t, f_1) = (Bv, f_1) + r_G(f_1) + (be^2, f_1) \\
&= (Bv, r_1) + (Bv, F_1) + r_G(f_1) + (be^2, f_1), \\
|I_2| &\le Ct_n|v|Ch|r| + Ct_n\big(h^{2m} + |e|^2\big)|f_1|, \quad |f_1| \le Ch|r| < Ch^{m+1}|v_n|.
\end{aligned} \tag{6.3}
$$

So (6.1) for $m = 1$ is valid.

If $m \ge 2$, we can repeat above treatments. By (6.2) we have

$$
I_1 = -(B_t v, S_t r) = (B_{tt} v, S_t^2 r) + (B_t v_t, S_t^2 r) =: I_{11} + I_{12}, \quad |S_t^2 r| \le Ch^2|r|.
$$

Obviously $|I_{11}| \le Ct_n h^2|r|$ and treat the term I_{12} in the same way used in (6.3). By (6.3), we have

$$
\begin{aligned}
I_2 =& -(Bv, r_1) - (B^2 v, S_t r) - r_G(f_1) - (be^2, f_1) \\
=&(v_t, B^T S_t r_1) + (B_t v, S_t r_1) + (B^2 v_t + (B^2)_t v, S_t^2 r) - r_G(f_1) - (be^2, f_1).
\end{aligned}
$$

Setting $F_2 = B^T S_t r_1, f_2 = L_h(F_2), r_2 = F_2 - f_2$, obviously

$$
|f_2| \le C|F_2| \le C|S_t r_1| \le Ch|r_1| \le Ch^2|r|, \quad |r_2| \le C|F_2| \le Ch^2|r|.
$$

The first term in I_2 is transformed to

$$
\begin{aligned}
(v_t, B^T S_t r_1) =&(v_t, F_2) = (v_t, f_2) = (Bv, f_2) + r_G(f_2) + (be^2, f_2), \\
\le& Ct_n|v|Ch^2|r| + Ct_n\big(h^{2m} + |e|^2\big)|f_2|, \quad |f_2| \le Ch^2|r|.
\end{aligned}
$$

Similarly treat $(B^2 v_t, S_t^2 r)$ in I_2. The estimates of other terms in I_2 is simple. We have

$$
|(Bv, r)| \le Ct_n|v|h^2|r| + \big(Ct_n h^{2m} + Ct_n|e|^2\big)h^2|r|, \quad |r| \le Ch^m|v_n|,
$$

and (6.1) for $m = 2$ is valid.

This argument can be repeated m times and then Theorem 6.1 is established. □

7. Numerical Experiments

Example 7.1. Consider nonlinear Hamiltonian system (see [4], p.143)

$$
H(p, q) = \frac{1}{2}(p^2 + 4q^2 + 4q^4/3), \qquad p_0 = 0.5, \quad q_0 = 0.25, \tag{7.1}
$$

where the canonical system is $q' = p,\ p' = -(4q + 8q^3/3)$. We shall compare three algorithms: the quadratic finite element (2FE) with five-point Gauss quadrature, fourth-order symplectic Runge-Kutta method (4SRK) and fourth-order symplectic difference scheme (4SS) [4] based on an expansion at a middle point Z^*:

$$
Z^{k+1} = Z^k - hJ\nabla H(Z^*) + \frac{h^3}{24} J\nabla\{(\nabla H)^T J H_{zz} J\nabla H\}(Z^*), \quad Z^* = \frac{1}{2}(Z^{k+1} + Z^k).
$$

Take $h = 0.1, N \le 10^5$. Three computational trajectories in phase plane are close each other (see Fig. 7.1), but in fact 2FEM preserves the energy, whereas the energy for 4SRK (and 4SS) is of the larger error (see Table 7.1).

Table 7.1: The error $H_h(t) - H$ at nodes for 2FEM and 4SRK, $h = 0.1$.

	$t = 1$	$t = 10$	$t = 100$	$t = 1000$	$t = 10000$
$2FE$	-1.110e-16	1.110e-16	-8.326e-16	-4.773e-15	-3.447e-14
$4SRK$	5.876e-8	5.760e-10	1.210e-9	2.735e-8	7.330e-8

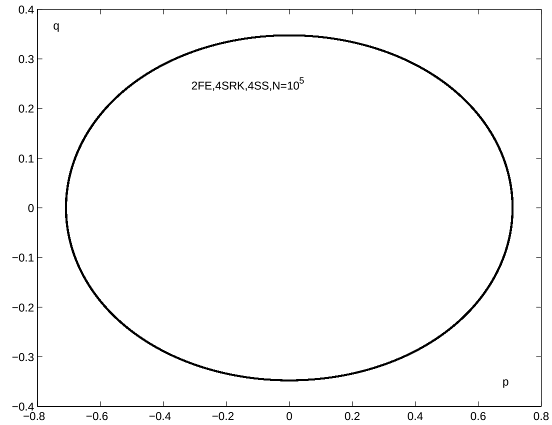

Fig. 7.1. Trajectories in phase plane, 2FEM and 4SRK, $h = 0.1$, step number is $N = 10^5$.

In the following, we turn to the trajectory curves $P(t)$ and $Q(t)$ in the physical plane. It is observed that three sets of trajectory $P(t), Q(t)$ for 2FEM, 4SRK and 4SS are close to each other when $t \leq T = Nh = 10^2$ (the left in Fig. 7.2). When $t = Nh = 10^4$, two trajectories for 2FEM and 4SRK are still close to each other (the right in Fig. 7.2), but that for 4SS already has the deviation of a half period. Hence, we will not discuss the 4SS case anymore. When $t \geq Nh = 10^6$, the trajectories for 2FEM and 4SRK are also of the larger deviations (Fig. 7.3).

To investigate the error for $P(t), Q(t)$, we have computed the exacter solution $p(t), q(t)$ by 2FEM with smaller step-length $h = 0.01$. The corresponding errors $e_p(t) = p - P, e_q(t) = q - Q$ are listed in Table 7.2. We see that these errors are close to each other and grow linearly in time.

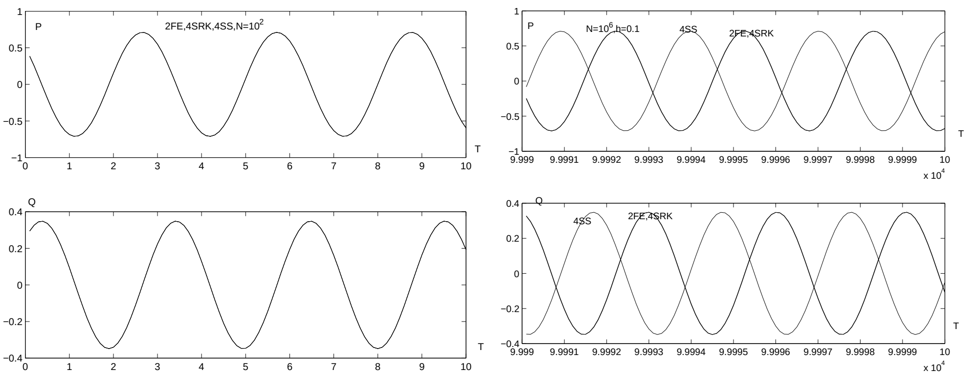

Fig. 7.2. $P(t), Q(t)$ for 2FEM, 4SRK, 4SS, $h = 0.1$,$T = 10$ (left),$T = 10^5$ (right).

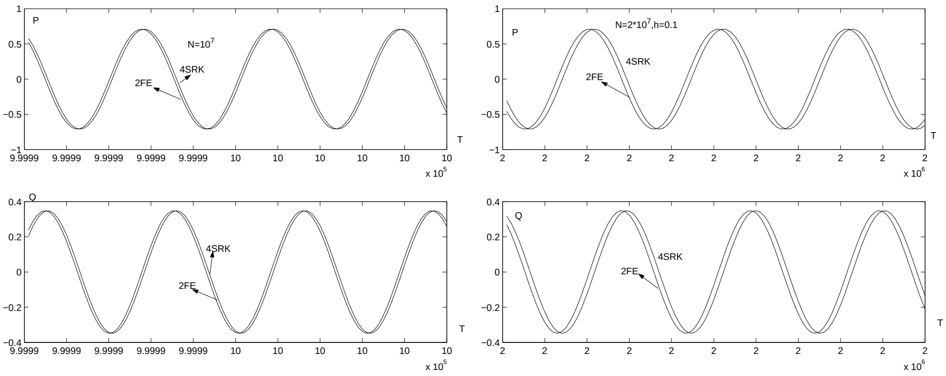

Fig. 7.3. $P(t), Q(t)$ for 2FEM and 4SRK, $h = 0.1$,$T = 10^6$ (left),$T = 2 * 10^6$ (right).

Table 7.2: The errors $e_p(t), e_q(t)$ at nodes for 2FEM and 4SRK.

$e_p(t)$	$t=1$	$t=10$	$t=100$	$t=1000$	$t=10000$
$2FE$	1.147e-6	2.230e-5	2.157e-4	1.607e-3	3.692e-2
$4SRK$	1.051e-6	2.270e-5	2.203e-4	1.642e-3	3.772e-2
$e_q(t)$	$t=1$	$t=10$	$t=100$	$t=1000$	$t=10000$
$2FE$	2.069e-6	1.685e-5	1.686e-4	1.824e-3	9.229e-3
$4SRK$	2.051e-6	1.715e-5	1.722e-4	1.864e-3	9.443e-3

Example 7.2. Consider the nonlinear Huygens system (see [3,4]):

$$p' = 2q - 4q^3, \quad q' = 2p; \quad p_0 = 0, \ q_0 = 1.1; \quad H = p^2 + q^4 - q^2. \tag{7.2}$$

We do not know if (7.2) has an analytical solution, but its solution can be computed by 2FEM $(P(t), Q(t))$ with smaller step-length h. Taking a smaller $h = 0.01$ and computing $n = 450$ steps, we obtain $\omega(P) = \omega(Q) = 4.02065$ as the periods of P and Q.

When the step-length $h = 0.2$ and $t \in (0, 5)$, the curves $P(t), Q(t)$ for three algorithms are close to each other. If taking $h = 0.4$, the 2FEM and 4SRK (real lines) still perform well, but 4SS (dot lines) deviates much larger, see Fig. 7.4. Note that the Heissen matrix

$$H_{zz} = 2\begin{pmatrix} 1 & 0 \\ 0 & 6q^2 - 1 \end{pmatrix}$$

is positive definite for $|q| > 1/\sqrt{6} \approx 0.408$, but for $|q| < 1/\sqrt{6}$ its sign changes in two small pieces in Fig. 7.5, which yields that the curve $P(t)$ in Fig. 7.4 has two smaller peaks (or valleys) in each large peak (or valley).

Next their energy errors with step-length $h = 0.2$ are listed in Table 7.3. We see that 2FE can preserve the energy very well, whereas 4SS and 4SRK cannot, whose accuracy is not good even if in the starting period.

Table 7.3: The energy errors $H - H_h$ for 2FE, 4SRK and 4SS, $h = 0.2$.

$t=$	0.2	2	20	200	2000	$2*10^4$	$2*10^5$
2FE	1.11e-16	4.44e-16	2.44e-15	7.49e-15	1.70e-14	9.40e-14	3.58e-13
4SS	2.95e-4	7.79e-6	3.04e-4	1.37e-6	8.23e-5	2.59e-3	8.85e-4
4SRK	-4.06e-6	-2.73e-4	-5.43e-4	-5.72e-6	-3.34e-4	-5.00e-4	-

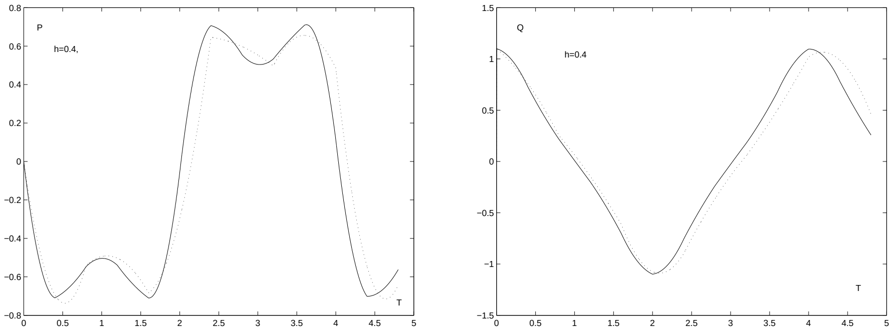

Fig. 7.4. $h = 0.4$, curves $P(t), Q(t)$, 2FE, 4SRK (real) and 4SS (dot).

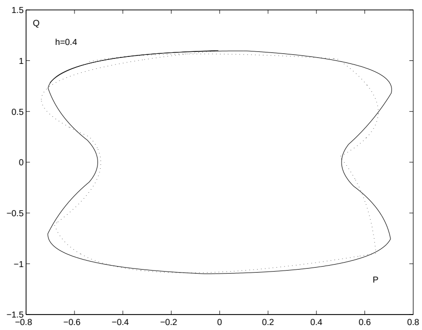

Fig. 7.5. $h = 0.4$, (P, Q)-phase plane, 2FE, 4SRK (real) and 4SS (dot).

Finally we investigate the computational trajectories with $h = 0.1$ and depict four curves in two periods at the final time $t = 10^3, 5 \times 10^3, 10^4$, respectively. The numbers in Figs. 7.6-7.9 stand for: 1. The exacter 2FE with $h = 1/40$; 2. 2FE ; 3. 4SRK; 4. 4SS, We see that 2FE and 4SRK are better, whereas 4SS moves to the right over one half period (Fig. 7.7, $t = 5000$) and one period (Fig. 7.8, $t = 10^4$). Fig. 7.9 ($t \leq 2000$) shows the error oscillations in corresponding sector domains. Note that the true solution satisfies $|z| < 1$. Then, the error $|e| > 0.2$ is already meaningless. Moreover, all three errors grow linearly in time.

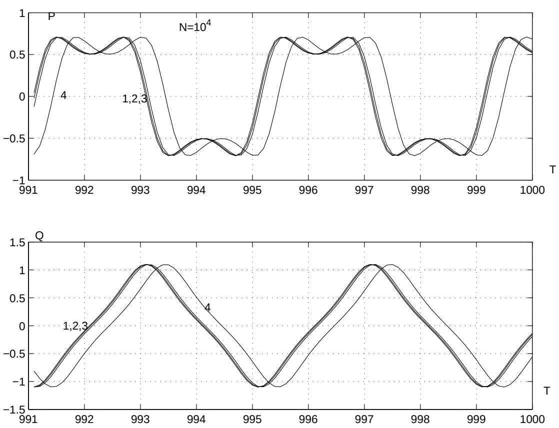

Fig. 7.6. $h = 0.1$, $T = 10^3$, curves $P(t)$, $Q(t)$ remove (in two periods).

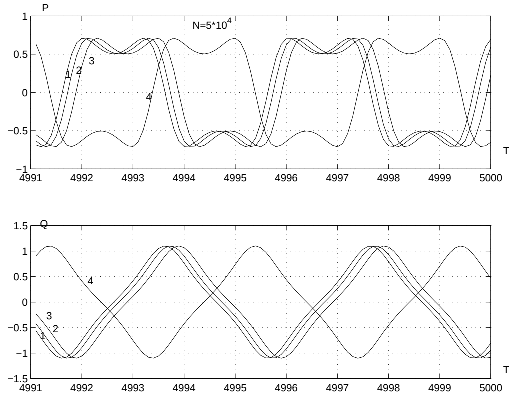

Fig. 7.7. $h = 0.1$, $T = 5 * 10^3$, curves $P(t)$, $Q(t)$ remove. Over a half period for 4SS.

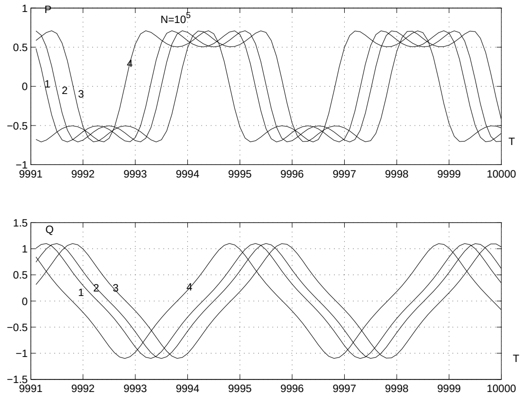

Fig. 7.8. $h = 0.1$, $T = 10^4$, curves $P(t)$, $Q(t)$ remove. Over one period for 4SS.

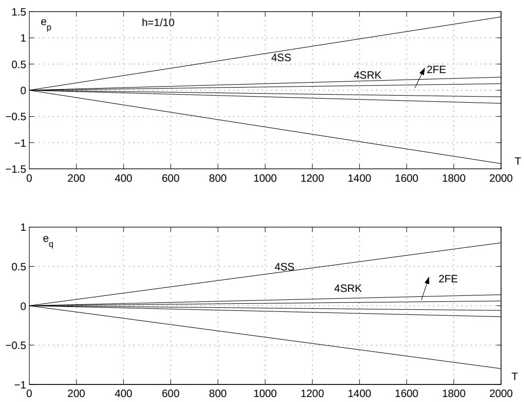

Fig. 7.9. Oscillation sectors of errors e_P, e_Q, $h = 0.1$, $T = 2 * 10^3$.

These numerical experiments show that the simplecticity, energy conservation and trajectory deviation for long-time are three different important properties in computations for the nonlinear Hamilton system.

Acknowledgments. The work was supported by the National Natural Science Foundation of China (No. 10771063) and the Key Laboratory of High Performance Computation and Stochastic Information Processing of Ministry of Education. The authors would like to thank the referees for their valuable suggestions.

References

[1] R. Ruth, A canonical integration technique, *IEEE T. Nucl. Sci.*, **30** (1983), 2669-2671.

[2] K. Feng, On ifference schemes and symplectic geometry, *Proceedings of the 5-th Inter., Symposium of Differential Geometry and Differential Equations*, Beijing, 1984, 42-58

[3] K. Feng's Collection of Works, Vol. 2, Beijing: National Defence Industry Press, 1995.

[4] K. Feng, M.Z. Qing, Sympletic Geometric Algorithms for Hamilton System, Hangzhou: Zhejiang Science and Techniques Press, 2003.

[5] K. Feng, D.L. Wang, Symplectic difference schemes for Hamiltonian systems in general symplectic structures, *J. Comput. Math.*, **9** (1991), 86-96.

[6] K. Feng, H.M. Wu, M.Z. Qin, D.L. Wang, Construction of canonical difference schemes for Hamiltonian formalism via generating functions, *J. Comput. Math.*, **7** (1989), 71-96.

[7] G. Sun, Symplectic partitioned Runge-Kutta methods, *J. Comput. Math.*, **11** (1993b), 365-372.

[8] Y.F. Tang, The symplecticity of multi-step methods, *Comput. Math. Appl.*, **25** (1993), 83-90.

[9] Y.F. Tang, Symplectic computation of Hamilton systems (I), *J. Comput. Math.*, **20** (2002), 267-276.

[10] Z.J. Shang, Construction of volume-preserving difference scheme for source-free systems via generating function, *J. Comput. Math.*, **12**:3 (1993), 265-272.

[11] K. Feng, Z.J. Shang, Volume-preserving algorithms for source-free dynamical systems. *Numer. Math.*, **71** (1995), 451-463.

[12] Z.J. Shang, Volume-preserving maps, source-free systems and their local structures. *J. Phys. A-Math. Gen.*, **29** (2006), 5601-5615.

[13] T. Tang, X. Xu, Accuracy enhancement using spectral postprocessing for differential equations and integral equations, *Commun. Comput. Phys.*, **5** (2009),779-792.

[14] X.S. Liu, Y.Y. Qi, J.H. He, P.Z. Ding, Recent progress in symplectic algorithms for use in quantum systems. *Commun. Comput. Phys.*, **2** (2007), 1-53.

[15] J.M. Sanz-Serna, Runge-Kutta schemes for Hamiltonian systems, *BIT*, **28** (1988), 877-883.

[16] F. Lasagni, Canonical Runge-Kutta methods, *ZAMP*, **39** (1988), 952-953.

[17] Y.B. Suris, The canonicity of mappings generated by Runge-Kutta type methods when integrating the system $X'' = -\partial v/\partial x$, *USSR Comput. Math. Math. Phys.*, **29:1** (1989), 138-144.

[18] J.M. Sanz-Serna, M.P. Calvo, Numerical Hamiltonian Problems, London: Chamman and Hall, 1994.

[19] A.M. Stuart, A.R. Humphries, Dynamical Systems and Numerical Analysis, Cambridge University Press, 1996.

[20] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Springer, 2003.

[21] E. Hairer, Conjugate-symplecticity of linear multistep methods, *J. Comput. Math.*, **26** (2008), 657-659.

[22] Q. Tang, C.M. Chen, Energy conservation and symplectic properties of continuous finite element methods for Hamiltonian systems, *Appl. Math. Comput.*, **181** (2006), 1357-1368

[23] Q. Tang, C.M. Chen, L.H. Liu, Continuous finite element methods for Hamiltonian systems, *Appl. Math. Mech-Engl.*, **28**:8 (2007), 1071-1080.

[24] O. Karakashian, C. Makridakis, A space-time finite element method for the nonlinear Schrodinger equation: The continuous Galerkin method, *SIAM J. Numer. Anal.*, **36**:6 (1999), 1779-1807.

[25] J. Frehse, R. Rannacher, Asymptotic L^{∞}-error estimates for linear finite element approximations of quasilinear boundary value problems, *SIAM J. Numer. Anal.*, **15** (1978), 418-431.

[26] C.M. Chen, Orthogonality correction technique in superconvergence analysis, Inter., *J. Numer. Anal. Model.*, **2** (2005), 31-42.

[27] C.M. Chen, Structure Theory of Superconvergence for Finite Elements. Changsha: Hunan Science and Techniques. 2001.