IMA Journal of Numerical Analysis (2000) **20**, 235–261

The numerical integration of relative equilibrium solutions. The nonlinear Schrödinger equation

A. DURÁN[†] AND J. M. SANZ-SERNA[‡]

Departamento de Matemática Aplicada y Computación, Universidad de Valladolid, Valladolid, Spain

[Received 24 April 1998 and in revised form 11 January 1999]

We analyse different error propagation mechanisms for conservative and nonconservative time-integrators of nonlinear Schrödinger equations. We use a geometric approach based on interpreting waves as relative equilibria.

1. Introduction

This paper is devoted to a detailed study of the error growth of numerical time-integrators for solitary waves of nonlinear Schrödinger equations. It is shown that schemes that preserve conserved quantities of the equation possess better error propagation mechanisms than their nonconservative counterparts.

The recent literature includes several analyses (Calvo & Hairer, 1995; Calvo & Sanz-Serna, 1993; Calvo *et al.*, 1998; Cano & Sanz-Serna, 1997, 1998; Estep & Stuart, 1995; Hairer & Lubich, 1997) of the *quantitative* advantages of so-called geometric integrators (Sanz-Serna, 1997), i.e. of numerical methods that take into account geometric properties of the system of differential equations being integrated. The references above deal with *ordinary* differential equations; partial differential equations have been considered in Frutos & Sanz-Serna (1997), a paper that takes the Korteweg–de Vries equation as a case study.

The present paper, following on from Durán (1997), complements the work in Frutos & Sanz-Serna (1997) in several ways. The most obvious difference is that here we deal with nonlinear Schrödinger equations rather than with the Korteweg–de Vries equation. A more significant difference is that we build upon the geometric study in Durán & Sanz-Serna (1998), thereby obtaining a better insight than was possible in Frutos & Sanz-Serna (1997), where the geometric mechanisms leading to favourable error propagation were not clearly identified. Furthermore some technical results on asymptotic expansions of the error, that were taken for granted without proof in Frutos & Sanz-Serna (1997), are investigated here.

The paper is structured as follows. Section 2 contains background material on the problem being integrated. The main issues discussed are the interpretation of travelling waves as Hamiltonian relative equilibria and variational equations. Section 3 includes the main results and Section 4 is devoted to numerical experiments. As in Durán & Sanz-Serna (1998), we point out that we have preferred not to include several extensions relative to

[†] Email: angel@mac.cie.uva.es

[‡] Email: sanzserna@cpd.uva.es

higher-order terms in the expansion of the error or to solutions more general than solitary waves; in particular for multisoliton solutions of the cubic nonlinear Schrödinger equation, results similar to those in Frutos & Sanz-Serna (1997), Section 5.3 apply. This paper only considers discretization in time; fully discrete schemes may be the subject of a future study.

2. The nonlinear Schrödinger equation

2.1 *Hamiltonian structure and conserved quantities*

We consider nonlinear Schrödinger equations of the form

$$\mathrm{i}u_t + u_{xx} + f(u) = 0, \quad -\infty < x < \infty, \quad t > 0, \tag{1}$$

where $u = u(x,t)$ is complex-valued and f is a complex-valued function of a complex variable. Most of our results concern the particular case

$$\mathrm{i}u_t + u_{xx} + |u|^{2\sigma}u = 0, \quad -\infty < x < \infty, \quad t > 0, \quad \sigma > 0, \tag{2}$$

which, as is well known, appears in various physical applications, especially when $\sigma = 1, 3$ (Kelley, 1965; Zakharov, 1972). We shall not discuss such aspects of (1) as the existence and uniqueness of solutions of the corresponding initial value problem (cf. Ginibre & Velo, 1979, 1985; Kato, 1987). For our purposes, however, we need to comment on the *Hamiltonian structure* of (1). It is expedient to handle the unknown u in (1) by means of its real and imaginary parts, v and w respectively. Then, the *phase space* Ω for the Hamiltonian formalism consists of pairs (v, w) of sufficiently smooth real functions of the spatial variable x such that v, w and their derivatives decrease sufficiently fast at infinity. If the function f satisfies

$$f(z) = \frac{\partial V}{\partial \overline{z}}, \quad z \in \mathcal{C}, \tag{3}$$

(where $\overline{z}$ denotes the complex conjugate) for some real-valued potential V, equation (1) can be written as an infinite-dimensional Hamiltonian system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} v \\ w \end{pmatrix} = \Xi\,\delta H, \quad \Xi = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

where δ denotes the variational derivative and the energy H is given by the functional

$$H = \frac{1}{2}\int_{-\infty}^{\infty} (v_x^2 + w_x^2 - V(v, w))\,\mathrm{d}x. \tag{4}$$

Note that in the case of (2), the potential is $V(z) = |z|^{2\sigma+2}/(\sigma + 1)$.

In the subsequent analysis we shall make use of the *symmetry groups* (Olver, 1993) of equation (1). If, in addition to (3), we assume that f satisfies the conditions

$$f(\overline{z}) = \overline{f(z)}, \quad z \in \mathcal{C}, \tag{5}$$

$$f(\omega z) = \omega f(z), \quad \omega, z \in \mathcal{C}, \quad |\omega| = 1, \tag{6}$$

then (1) admits the two-parameter Abelian (i.e. commutative) group of gauge transformations and translations

$$\begin{aligned} &G_{(\alpha,\beta)}(v(x), w(x)) \\ &\quad = (v(x-\beta)\cos\alpha - w(x-\beta)\sin\alpha, v(x-\beta)\sin\alpha + w(x-\beta)\cos\alpha), \end{aligned} \tag{7}$$

as a symmetry group. Here, α and β are real parameters: α measures the angle of rotation in the complex u-plane and β governs the translation along the x-axis. Note that the one-parameter group of rotations $G_{(\alpha,0)}$ is the flow of the Hamiltonian vector field $g_1(u) = \Xi\delta I_1(u)$ associated with the Hamiltonian function

$$I_1 = -\frac{1}{2}\int_{-\infty}^{\infty}(v^2 + w^2)\,\mathrm{d}x,$$

and the one-parameter group of translations $G_{(0,\beta)}$ is the flow of the Hamiltonian vector field $g_2(u) = \Xi\delta I_2(u)$ with

$$I_2 = \frac{1}{2}\int_{-\infty}^{\infty}(vw_x - wv_x)\,\mathrm{d}x.$$

Furthermore, I_1 and I_2 are conserved quantities of (1). These quantities satisfy the involution conditions

$$\{I_i, H\} = 0, \quad i = 1, 2, \tag{8}$$

$$\{I_1, I_2\} = 0, \tag{9}$$

where $\{\cdot,\cdot\}$ denotes the *Poisson bracket* (Marsden & Ratiu, 1994; Olver, 1993),

$$\{F, G\} = \int_{-\infty}^{\infty}\delta F\,\Xi\,\delta G\,\mathrm{d}x.$$

The condition (8) means that each quantity I_i is a first integral of (1) and implies in particular that $\{I_i, H\} = \text{constant}$, $i = 1, 2$, which is the condition for H to possess the symmetry group (7) generated by g_1 and g_2. Similarly (9) means that I_1 (resp. I_2) is a first integral of the Hamiltonian system with Hamiltonian function I_2 (resp. I_1) and in particular the flows of $\Xi\delta I_1$ and $\Xi\delta I_2$ commute, leading to (7) being Abelian.

2.2 *Relative equilibria. Solitary waves*

Equation (2) possesses a family of *solitary wave solutions* (Whitham, 1974) that, when $\sigma \neq 2$, can be completely interpreted by means of the symmetry group (7) and the corresponding *relative equilibria* (Arnold, 1989; Olver, 1993). The following reduction has been described in detail in Durán & Sanz-Serna (1998) for finite-dimensional Hamiltonian systems and can be formally applied to the present case. The phase space Ω is foliated by level sets $\{I_1 = c_1, I_2 = c_2\}$ of the first integrals I_1, I_2. These level sets are manifolds invariant by the flow of (2), i.e. an initial condition on $\{I_1 = c_1, I_2 = c_2\}$ leads to a solution of (2) that remains on this level set for all values of t. Furthermore, each orbit $\{G_{(\alpha,\beta)}(v, w), \alpha, \beta \in \mathcal{R}\}$ of the symmetry group is contained in a level set

$\{I_1 = c_1, I_2 = c_2\}$ (see (9)). Thus, each of these level sets is foliated by orbits of the group so that we can construct the corresponding quotient space or *reduced phase space*, whose points are orbits in the original phase space. On each reduced phase space (specified by the values c_1, c_2), the original Hamiltonian system gives rise to a new Hamiltonian system, the *reduced system* (Arnold, 1989; Olver, 1993). The reduced system governs the evolution in time of the group orbits, i.e. describes the behaviour of u modulo translations in x and gauge transformations.

Restricting our attention to a fixed level set $\{I_i = c_i, i = 1, 2\}$ we look for relative equilibria $u_0 = (v_0, w_0) \in \Omega$ (Arnold, 1989)

$$\delta H(u_0) - \lambda_0^1 \delta I_1(u_0) - \lambda_0^2 \delta I_2(u_0) = 0, \tag{10}$$

$$I_i(u_0) = c_i, \quad i = 1, 2, \tag{11}$$

where λ_0^i are real numbers; that is, we look for stationary points $u_0 \in \Omega$ of H restricted to the level set. The group orbit through a relative equilibrium is an equilibrium of the reduced system; furthermore the solution of (2) with initial value u_0 is simply $G_{(t\lambda_0^1, t\lambda_0^2)}(u_0)$ (Durán & Sanz-Serna, 1998) and therefore the time evolution of the initial profile is given by a translation coupled with a gauge transformation where the parameters α and β that govern rotation and translation vary linearly with t.

For the case at hand, (10) reads

$$u_0'' + |u_0|^{2\sigma} u_0 - \lambda_0^1 u_0 - \mathrm{i}\lambda_0^2 u_0' = 0.$$

After setting $u_0(x) = \rho(x) \exp(\mathrm{i}\theta(x))$ with real functions ρ, θ and integrating, we obtain

$$\rho(x) = (a(\sigma + 1))^{1/2\sigma} (\operatorname{sech} \sigma \sqrt{a} x)^{1/\sigma}, \quad a = \lambda_0^1 - \frac{(\lambda_0^2)^2}{4},$$
$$\theta(x) = \frac{\lambda_0^2}{2} x.$$

Furthermore, the constraint (11) leads to

$$\lambda_0^1 = \left(\frac{c_1}{L_\sigma}\right)^{2\sigma/(2-\sigma)} + \frac{c_2^2}{c_1^2}, \quad \lambda_0^2 = \frac{2c_2}{c_1}, \quad L_\sigma = \int_{-\infty}^{\infty} (\sigma + 1)^{1/\sigma} (\operatorname{sech} \sigma x)^{2/\sigma} \, \mathrm{d}x.$$

Along with this solution u_0, (10)–(11) possess the family of solutions

$$\varphi(x, x_0, \theta_0) = G_{(\theta_0, x_0)}(u_0) = \rho(x - x_0) \exp(\mathrm{i}\theta(x - x_0) + \mathrm{i}\theta_0); \tag{12}$$

these are the group orbit through $u_0 = \rho \exp(\mathrm{i}\theta)$ and they project onto the same equilibrium in the reduced phase space. Hence, we obtain the solutions of (2) given by

$$\begin{aligned}\psi(x, t, a, c, x_0, \theta_0) &= G_{(t\lambda_0^1, t\lambda_0^2)}(\varphi) \\ &= \rho(x - ct - x_0) \exp(\mathrm{i}(\tfrac{1}{2}c(x - ct - x_0) + \theta_0)) \\ &\quad \times \exp(i(a + \tfrac{1}{4}c^2)t),\end{aligned} \tag{13}$$

($c = \lambda_0^2$). This is a four-parameter family of solitary wave solutions with parameters x_0, θ_0, c_1, c_2 (or, equivalently, x_0, θ_0, a, c). The parameter a determines the amplitude of the wave and c its velocity, while x_0 controls the initial location and θ_0 the initial phase.

2.3 *Linearization of relative equilibria*

In order to study the time behaviour of the errors in the numerical integration of solitary wave problems for (2), we analyse the properties of solutions of the variational equation of (2) near the relative equilibrium solution ψ in (13). Denoting by δ the perturbation of ψ, the variational equation is

$$\mathrm{i}\delta_t + \delta_{xx} + \sigma|\psi|^{2(\sigma-1)}\psi^2\overline{\delta} + (\sigma+1)|\psi|^{2\sigma}\delta = 0. \tag{14}$$

The change of variables $\delta = G_{(t(a+c^2/4),tc)}\Delta$ (as in Durán & Sanz-Serna, 1998, Lemma 2.2) transforms (14) into

$$\mathrm{i}\Delta_t + \Delta_{xx} + \sigma|\varphi|^{2(\sigma-1)}\varphi^2\overline{\Delta} + (\sigma+1)|\varphi|^{2\sigma}\Delta - \mathrm{i}c\Delta_x - (a+\tfrac{1}{4}c^2)\Delta = 0. \tag{15}$$

(Observe that (15) results from linearizing $u_t = \Xi(\delta H(u) - \lambda_0^1\delta I_1(u) - \lambda_0^2\delta I_2(u))$ around its equilibrium φ.) If we look for Δ in the form

$$\Delta(\xi,T) = E(\xi,T)\exp\left(\mathrm{i}\frac{c}{2\sqrt{a}}\xi + \mathrm{i}\theta_0\right),$$

with $\xi = \sqrt{a}(x - ct - x_0)$, $T = at$, we can write (15) as ($E = F + \mathrm{i}G$)

$$\begin{pmatrix} F \\ G \end{pmatrix}_T = L\begin{pmatrix} F \\ G \end{pmatrix}, \quad L = \begin{pmatrix} 0 & L_1 \\ -L_2 & 0 \end{pmatrix}, \tag{16}$$

with the operators L_1, L_2 given by

$$L_1 = -\frac{\partial^2}{\partial\xi^2} + 1 - R_\sigma(\xi)^{2\sigma}, \quad L_2 = -\frac{\partial^2}{\partial\xi^2} + 1 - (2\sigma+1)R_\sigma(\xi)^{2\sigma},$$

where

$$R_\sigma(x) = (\sigma+1)^{1/2\sigma}(\mathrm{sech}(\sigma x))^{1/\sigma}.$$

We are also interested in the analysis of the nonhomogeneous version of (14) with source terms that possess $\{G_{(\alpha,\beta)}, \alpha, \beta \in \mathcal{R}\}$ (or at least $\{G_{(t(a+c^2/4),tc)}, t \in \mathcal{R}\}$) as a symmetry group. The changes of variables above lead, in the nonhomogeneous case, to an equation of the form

$$\begin{pmatrix} F \\ G \end{pmatrix}_T = L\begin{pmatrix} F \\ G \end{pmatrix} + \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}, \tag{17}$$

with S_1, S_2 constant. For the analysis of (16) we consider the space $H^1 \times H^1$ of pairs of real functions $(V, W)^T$, $V \in H^1$, $W \in H^1$ and norm

$$\|(V,W)^T\|_{H^1\times H^1} = (\|V\|_{H^1}^2 + \|W\|_{H^1}^2)^{1/2}.$$

The next result, due to Weinstein (1985), describes the spectral properties of the operator L.

LEMMA 2.1 0 is an eigenvalue of L with geometric multiplicity 2 and algebraic multiplicity 4. The generalized kernel of L is spanned by the functions

$$\begin{aligned}
\Phi_1(\xi) &= \frac{1}{\gamma_1}(0, R_\sigma(\xi))^T,\\
\Phi_2(\xi) &= \frac{-1}{2\gamma_1}\left(\frac{1}{\sigma}R_\sigma(\xi) + \xi R'_\sigma(\xi), 0\right)^T,\\
\Phi_3(\xi) &= \frac{1}{\gamma_2}(R'_\sigma(\xi), 0)^T,\\
\Phi_4(\xi) &= \frac{1}{2\gamma_2}(0, \xi R_\sigma(\xi))^T,
\end{aligned}$$

where

$$\gamma_1 = \left(\frac{1}{\sigma} - \frac{1}{2}\right)c_\sigma, \quad \gamma_2 = \frac{1}{2}c_\sigma, \quad c_\sigma = \int_{-\infty}^{\infty} R_\sigma^2(\xi)\,\mathrm{d}\xi.$$

More precisely, $\Phi_1, \Phi_3 \in \operatorname{Ker} L$ and $L\Phi_2 = -\Phi_1,\ L\Phi_4 = -\Phi_3$.

Thus, (16) has solutions of the form

$$c_1\Phi_1 + c_2(\Phi_2 - T\Phi_1) + c_3\Phi_3 + c_4(\Phi_4 - T\Phi_3),$$

that, when expressed in terms of the original variables u, x, t, are seen to be linear combinations of the partial derivatives of the solitary wave ψ with respect to the parameters a, c, x_0, θ_0 and therefore represent perturbations of the solitary wave that can be seen as the result of changing the parameter values. More precisely, Φ_1, Φ_3 correspond basically to the group generators g_1, g_2 evaluated at the solitary wave, and therefore induce changes in the location x_0 and phase θ_0 of the wave. On the other hand, Φ_2, Φ_4 correspond to changes in the solitary wave due to changes in the value of I_1 and I_2, determined by a and c (cf. Frutos & Sanz-Serna, 1997).

Another result, also due to Weinstein, is the key to the analysis of solutions of (17).

LEMMA 2.2 Suppose that $0 < \sigma < 2$ and let $\mathcal{M}$ be the subspace $\mathcal{M} = H^1 \times H^1 \bigcap (\operatorname{Ker}_g L^*)^\perp$, where L^* is the adjoint operator of L with respect to the inner product

$$\left\langle \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}, \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \right\rangle = \int_{-\infty}^{\infty} (V_1(\xi)W_1(\xi) + V_2(\xi)W_2(\xi))\,\mathrm{d}\xi.$$

Then $H^1 \times H^1 \simeq \operatorname{Ker}_g L \oplus \mathcal{M}$. Furthermore, if $(F(T), G(T))^T$ is a solution of (17) with initial condition in $\mathcal{M}$, then $(F(T), G(T))^T \in \mathcal{M}$ for all T and there exists $C > 0$ such that

$$\|(F(T), G(T))^T\|_{H^1 \times H^1} \leqslant C\|(F(0), G(0))^T\|_{H^1 \times H^1}. \tag{18}$$

Since the symmetry group (7) consists of isometries in $H^1 \times H^1$ (rotations and translations), the growth with time of solutions of (14) is identical to that of solutions of (16). Due to (18), the only source of growth with time of these solutions comes from its component in the generalized kernel of L.

As far as (17) is concerned, if $S = (S_1, S_2)^T \in H^1 \times H^1$, using Lemma 2.2, we can decompose

$$S = S_{\mathrm{Ker}_g L} + S_{\mathcal{M}},$$

with $S_{\mathrm{Ker}_g L} \in \mathrm{Ker}_g L$, $S_{\mathcal{M}} \in \mathcal{M}$. Taking $S_{\mathrm{Ker}_g L} = S^{(1)} + S^{(2)}$ with $S^{(1)} \in \mathrm{Ker}\, L$ and $S^{(2)}$ in a supplement of $\mathrm{Ker}\, L$ in $\mathrm{Ker}\, L^2$, and due to Lemma 2.1 and the corresponding projection onto $\mathrm{Ker}_g L$, we have

$$S^{(1)} = \alpha_1 \Phi_1 + \alpha_3 \Phi_3, \qquad S^{(2)} = \alpha_2 \Phi_2 + \alpha_4 \Phi_4,$$

with

$$\alpha_i = \langle S, \Psi_i \rangle, \qquad 1 \leqslant i \leqslant 4, \tag{19}$$

where Ψ_i, $1 \leqslant i \leqslant 4$ form a basis of the generalized kernel of the adjoint L^* chosen in such a way that $\langle \Phi_i, \Psi_j \rangle = \delta_{ij}$. Explicitly

$$\begin{aligned}
\Psi_1(\xi) &= \left(0, \frac{1}{\sigma} R_\sigma(\xi) + \xi R'_\sigma(\xi)\right)^T, \\
\Psi_2(\xi) &= (-2R_\sigma(\xi), 0)^T, \\
\Psi_3(\xi) &= (-\xi R_\sigma(\xi), 0)^T, \\
\Psi_4(\xi) &= (0, -2R'_\sigma(\xi))^T.
\end{aligned}$$

Observe that Ψ_2, Ψ_4 are essentially the gradients of the invariants I_1, I_2 evaluated at the solitary wave (see Durán & Sanz-Serna, 1998).

In this situation, we have the following result.

LEMMA 2.3 If $0 < \sigma < 2$ and $S \in H^1 \times H^1$, the solution of (17) with zero initial condition is

$$\begin{aligned}
(F(T), G(T))^T = {}& \alpha_1 T \Phi_1(\xi) + \alpha_2 (T \Phi_2(\xi) - \tfrac{1}{2} T^2 \Phi_1(\xi)) \\
&+ \alpha_3 T \Phi_3(\xi) + \alpha_4 (T \Phi_4(\xi) - \tfrac{1}{2} T^2 \Phi_3(\xi)) \\
&+ \int_0^T \exp((T - \tau) L) S_{\mathcal{M}} \, \mathrm{d}\tau,
\end{aligned} \tag{20}$$

where α_i, $1 \leqslant i \leqslant 4$ is given by (19). Furthermore, the integral in (20) remains, for all T, bounded in the $H^1 \times H^1$ norm.

Proof. By using Duhamel's principle and the preceding remarks, we can write

$$(F(T), G(T))^T = T S^{(1)} + (T I + \tfrac{1}{2} T^2 L) S^{(2)} + \int_0^T \exp((T - \tau) L) S_{\mathcal{M}} \, \mathrm{d}\tau,$$

where I is the identity operator. Replacing $S^{(1)}$ and $S^{(2)}$ by their values we obtain the first part of the lemma. On the other hand, since L is a closed operator (Weinstein, 1985) and $\mathcal{M}$

is invariant by L, there exists a unique $L^{-1}S_{\mathcal{M}} \in \mathcal{M}$, such that we can write the integral in the form

$$\int_0^T \exp((T-\tau)L)S_{\mathcal{M}}\,\mathrm{d}\tau = (\exp(TL)-I)L^{-1}S_{\mathcal{M}},$$

and we use Lemma 2.2 to conclude the proof. □

Note that the first part of (20) corresponds to the projection of the solution on $\mathrm{Ker}_g\,L$, with its components in $\mathrm{Ker}\,L$ (including the dominant term $(T^2/2)LS^{(2)}$) lying in the direction given by the group generators evaluated at the solitary wave solution; the component in $\mathrm{Ker}_g\,L$ but not in $\mathrm{Ker}\,L$ corresponds to the variation of the relative equilibrium with respect to the parameters a, c that govern the level manifold. Thus, this first part represents perturbations in the wave parameters that grow quadratically with time in the direction of $\mathrm{Ker}\,L$ (i.e. we have quadratic perturbations in the group parameters x_0, θ_0) and linearly in the level set parameters a, c. If the source term S is orthogonal to the gradients of the invariants evaluated at the solitary wave, then $S^{(2)} = 0$ and the growth is linear. On the other hand, the bounded behaviour in time of the second part of (20) shows, for $0 < \sigma < 2$, that the perturbations that do not represent changes in wave parameters can be controlled in the H^1 norm uniformly in time (Weinstein, 1985).

3. Numerical integration of solitary waves

3.1 *Error propagation*

The preceding results can be applied to analyse the behaviour of approximations to a solitary wave of the family (13). We focus on semidiscrete (discrete t, continuous x) one-step integrators for the initial value problem for (2). Such one-step integrators take the form

$$U^{n+1} = \chi_{\Delta t}(U^n), \tag{21}$$

where Δt denotes the time step, $U^n = U^n(x)$ is a numerical solution at time level $t_n = n\Delta t, n = 0, 1, \ldots$ and $\chi_{\Delta t}$ approximates the flow of the equation. Thus, if $U^0 = u_0$, then U^n is an approximation to the value $u(t_n)$ of the solution u of (2) with initial condition u_0.

The *local error* at a state $u \in \Omega$ is, by definition, the difference between the true Δt-flow at u and $\chi_{\Delta t}(u)$. If r is the order of the method, then the local error is $O(\Delta t^{r+1})$ as $\Delta t \to 0$. Taking u_0 as the relative equilibrium φ given by (12) and if the solution ψ in (13) is approximated by (21) with $U^0 = \varphi$, we assume that the local error has an asymptotic expansion of the form

$$\Delta t^{r+1} l(\psi) + \Delta t^{r+1} R(\psi, \Delta t), \tag{22}$$

where l, R are mappings defined in Ω with values in Ω, l is independent of Δt and $\|R(\cdot, \Delta t)\|_{H^1} \to 0$ as $\Delta t \to 0$. We also assume that the mapping $\chi_{\Delta t}$ is invariant by the one-parameter group $\{G_{((a+c^2/4)t, ct)} : t \in \mathcal{R}\}$ so that l admits this group as a symmetry group. Note that this condition is not restrictive, because this group consists of rotations and translations and most standard integrators are invariant with respect to linear transformations (Stoffer, 1995).

As far as the *global error* $U^n(x) - \psi(x, t_n)$ is concerned, we suppose that this possesses an expansion of the form

$$U^n(x) - \psi(x, t_n) = \Delta t^r e(x, t_n) + \Delta t^r q(x, t_n, \Delta t), \quad -\infty < x < \infty, \tag{23}$$

where the function e is independent of Δt and satisfies the corresponding nonhomogeneous variational equation with source term $-\mathrm{i}l(\psi(t))$ and zero initial condition (Cano & Sanz-Serna, 1997). Moreover, q is a remainder that, for fixed t, tends to zero in the H^1 norm as $\Delta t \to 0$.

Using Theorem 3.2 of Durán & Sanz-Serna (1998) and the results obtained in Lemma 2.3 we now state a theorem that describes the asymptotic behaviour in time of the above approximation to the solitary wave (13) given by (21). The proof is similar to that of Durán & Sanz-Serna (1998).

THEOREM 3.1 Suppose that (22), (23) hold and $\chi_{\Delta t}$ is invariant by the one-parameter group $\{G_{((a+c^2/4)t,ct)} : t \in \mathcal{R}\}$. If $\sigma < 2$, then we have

$$U^n(x) = \psi(x, t_n, \tilde{a}, \tilde{c}, \tilde{x}_0, \tilde{\theta}_0) + \Delta t^r \rho(x, t_n) + \Delta t^r Q(x, t_n, \Delta t), \tag{24}$$

with

$$\begin{aligned}
\tilde{a} &= a - \frac{\alpha_2}{\gamma_1 a^{-2+1/2\sigma}} \Delta t^r t_n, \\
\tilde{c} &= c + \frac{\alpha_4}{\gamma_2 a^{-3/2+1/2\sigma}} \Delta t^r t_n, \\
\tilde{x}_0 &= x_0 - \frac{\alpha_3}{\gamma_2 a^{-1/2+1/2\sigma}} \Delta t^r t_n - \frac{\alpha_4}{2\gamma_2 a^{-3/2+1/2\sigma}} \Delta t^r t_n^2, \\
\tilde{\theta}_0 &= \theta_0 + \frac{\alpha_1}{\gamma_1 a^{-1+1/2\sigma}} \Delta t^r t_n - \frac{c\alpha_3}{2\gamma_2 a^{-1/2+1/2\sigma}} \Delta t^r t_n \\
&\quad - \frac{\alpha_4 c}{4\gamma_2 a^{-3/2+1/2\sigma}} \Delta t^r t_n^2 + \frac{\alpha_2}{2\gamma_1 a^{-2+1/2\sigma}} \Delta t^r t_n^2,
\end{aligned} \tag{25}$$

and $\alpha_j = \langle -\mathrm{i}l, \Psi_j \rangle$, $j = 1(1)4$ (see (19)). The function ρ is independent of Δt and is bounded in the H^1 norm uniformly in time. The function Q is a remainder such that, for fixed t, $\|Q(\cdot, t, \Delta t)\|_{H^1} \to 0$ as $\Delta t \to 0$.

Moreover, if the method (21) satisfies the conditions

$$\langle -\mathrm{i}l, \nabla I_j(\varphi) \rangle = 0, \quad j = 1, 2, \tag{26}$$

then $\alpha_2 = \alpha_4 = 0$.

Thus, this theorem shows that the numerical solution consists of three components:

(i) First, we have a new solitary wave, the so-called *modified solitary wave* (Frutos & Sanz-Serna, 1997), whose amplitude and velocity $\tilde{a}$, $\tilde{c}$ differ from the corresponding parameters a, c of the original wave in terms that grow linearly with time, while the perturbation of the two other parameters, location and phase, grow quadratically (see $\tilde{x}_0$, $\tilde{\theta}_0$ in (25)). Note that, if the scheme (21) conserves the quantities I_1 and I_2 then, in particular, it satisfies the conditions (26) and the quadratic behaviour in time of the parameters of the modified solitary wave is not present. This is due to the fact that for a scheme preserving I_i, the leading term of the local error has to be orthogonal to the gradient of I_i at the original solitary wave and this implies that the coefficients α_2, α_4 vanish (recall the remarks preceding Lemma 2.3). In such a 'conservative' case, the modified solitary wave keeps the original amplitude and the parameter $\tilde{c}$ coincides with the original c; the errors $\tilde{x}_0 - x_0$, $\tilde{\theta}_0 - \theta_0$ grow linearly with time. It is also useful to point out that, in the case of a method that preserves the Hamiltonian (4) and only one of the quantities I_1 or I_2, a slightly different argument may be used to reach the same conclusions as in the case where both I_1 and I_2 are preserved. Here, the leading term of the local error is orthogonal to the gradients of H and the quantity I_1 or I_2 evaluated at the solitary wave; since this wave is a relative equilibrium (see (10)–(11)), then the gradient of the energy H at the wave is a linear combination of the gradients of I_1 and I_2; hence, the leading term of the local error satisfies (26).

(ii) The second term, a *complementary term*, $\Delta t^r \rho$ represents errors of leading order $O(\Delta t^r)$ that cannot be interpreted as changes in the parameters of the solitary wave. This term corresponds to the component of the function e that lies in $\mathcal{M}$ (see Lemma 2.2) and therefore is bounded in the H^1 norm uniformly in time (cf. Frutos & Sanz-Serna, 1997).

(iii) Finally, the third term is a remainder of higher order ($o(\Delta t^r)$).

3.2 *The implicit midpoint rule*

In this section we prove that the standard midpoint rule satisfies the hypotheses that were assumed for the analysis in Section 3.1. The material in this section improves in different ways upon standard analyses (Robinson *et al.*, 1993; Sanz-Serna, 1984) of time discretizations of nonlinear Schrödinger equations. For instance, we prove convergence in H^1 and the existence of an asymptotic expansion of the error.

First, we define the numerical method. Suppose that in (1) the function f satisfies

$$f \in C^2(\mathcal{R}^2), \quad f(0) = 0, \tag{27}$$

(note that for (2), the first condition holds for $\sigma > 1/2$) and consider the initial value problem given by (1) with $0 < t \leqslant t_{\max}$ and the initial condition u_0. If $U^n = U^n(x)$ denotes a numerical approximation to the solution u at time level $t_n = n\Delta t, n = 0, 1, \dots, N = \lfloor t_{\max}/\Delta t \rfloor$, the (semidiscrete) implicit midpoint rule is given by

$$\begin{aligned} \frac{U^{n+1} - U^n}{\Delta t} &= \mathrm{i}\partial_{xx} U^{n+1/2} + \mathrm{i} f(U^{n+1/2}), \\ U^{n+1/2} &= \frac{U^n + U^{n+1}}{2}, \quad n = 0, 1, \dots, N-1, \end{aligned} \tag{28}$$

where we take $U^0 = u_0$. The formulae (28) require the terms $U^{n+1/2}$ to be in the domain of the operator $\mathrm{i}\partial_{xx}$. Therefore, we say that a vector $\mathbf{U} = (U^0, U^1, U^2, \dots, U^N)$ is a numerical solution of (28) if $\mathbf{U}$ satisfies the corresponding equations, with $U^n \in H^1$ for $n \in \{0, 1, \dots, N\}$ and $U^{n+1/2} \in H^3$ for $n \in \{0, 1, \dots, N-1\}$. Sometimes (Sanz-Serna

& Calvo, 1994) it is convenient to rewrite (28) in terms of the average $Z = U^{n+1/2}$, that is

$$Z = U^n + \mathrm{i}\frac{\Delta t}{2}\partial_{xx}(Z) + \mathrm{i}\frac{\Delta t}{2}f(Z), \qquad n = 0, 1, \dots, N-1;$$

at each step, we solve an implicit Euler equation for Z with step size $\Delta t/2$ and obtain the next level approximation via the extrapolation $U^{n+1} = 2Z - U^n$ (Sanz-Serna & Calvo, 1994).

In order to obtain the main convergence result we employ an argument of consistency and nonlinear stability (López-Marcos & Sanz-Serna, 1988).

Consistency

For the analysis of the consistency, we define the truncation errors $\mathcal{T}^n$, $n = 0, 1, \dots, N$, as the residuals in the theoretical solution $u(t_n)$,

$$\mathcal{T}^{n+1} = \frac{u(t_{n+1}) - u(t_n)}{\Delta t} - \mathrm{i}\partial_{xx}\frac{u(t_{n+1}) + u(t_n)}{2} - \mathrm{i}f\left(\frac{u(t_{n+1}) + u(t_n)}{2}\right).$$

The conditions on the nonlinear term in (27) imply that f is locally Lipschitz as a mapping in H^1 and, using Taylor expansions, it can be easily seen that if the functions $u, u_{tt}, u_{ttt}, \partial_{xx}u_{tt}$ are bounded in the H^1 norm uniformly in time, then

$$\|\mathcal{T}^{n+1}\|_{H^1} = O((\Delta t)^2), \qquad \Delta t \to 0. \tag{29}$$

Nonlinear stability

To analyse the stability of (28), we consider vectors $\mathbf{V} = (V^0, V^1, \dots, V^N)^T$ and $\mathbf{W} = (W^0, W^1, \dots, W^N)^T$ for $V^j,\ W^j \in H^1$, $j = 0, 1, \dots, N$, $V^{j+1/2}, W^{j+1/2} \in H^3$, $j = 0, 1, \dots, N-1$ and residuals $\boldsymbol{\rho} = (\rho^0, \rho^1, \dots, \rho^N)^T$, $\boldsymbol{\sigma} = (\sigma^0, \sigma^1, \dots, \sigma^N)^T$ defined by

$$\begin{aligned}
\rho^0 &= V^0 - u_0,\\
\rho^{n+1} &= \frac{V^{n+1} - V^n}{\Delta t} - \mathrm{i}\partial_{xx}V^{n+1/2} - \mathrm{i}f(V^{n+1/2}),\\
\sigma^0 &= W^0 - u_0,\\
\sigma^{n+1} &= \frac{W^{n+1} - W^n}{\Delta t} - \mathrm{i}\partial_{xx}W^{n+1/2} - \mathrm{i}f(W^{n+1/2}),
\end{aligned} \tag{30}$$

with $n = 0, 1, \dots, N-1$. Here, $\mathbf{V}$ and $\mathbf{W}$ can be seen as solutions of (28) under perturbations given by $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ respectively. The stability analysis aims at estimating the size of $\mathbf{V} - \mathbf{W}$ in terms of the size of $\boldsymbol{\rho} - \boldsymbol{\sigma}$ (López-Marcos & Sanz-Serna, 1988).

THEOREM 3.2 Given $R > 0$, if

$$\max_{0 \leqslant n \leqslant N} \|V^n\|_{H^1} \leqslant R, \qquad \max_{0 \leqslant n \leqslant N} \|W^n\|_{H^1} \leqslant R, \tag{31}$$

then, for Δt sufficiently small, there exists $C = C(t_{\max}, R) > 0$ such that

$$\max_{0\leqslant n\leqslant N} \|V^n - W^n\|_{H^1} \leqslant C\left(\|\rho^0 - \sigma^0\|_{H^1} + \Delta t \sum_{k=1}^{N} \|\rho^k - \sigma^k\|_{H^1}\right). \tag{32}$$

Proof. From (30) we have

$$\begin{aligned}\frac{V^{n+1} - W^{n+1}}{\Delta t} - \frac{V^n - W^n}{\Delta t}\\ &= \mathrm{i}\partial_{xx}(V^{n+1/2} - W^{n+1/2}) + \mathrm{i}(f(V^{n+1/2}) - f(W^{n+1/2}))\\ &\quad +(\rho^{n+1} - \sigma^{n+1}), \quad n = 0, 1, \dots, N-1.\end{aligned} \tag{33}$$

We define $e^n = V^n - W^n, n = 0, 1, \dots, N$. Taking the L^2-inner product of $V^{n+1/2} - W^{n+1/2} - \partial_{xx}(V^{n+1/2} - W^{n+1/2})$ with (33), integrating by parts and taking the real part of each term in the resulting expression, we obtain

$$\begin{aligned}&\frac{1}{2\Delta t}(\|e^{n+1}\|_{H^1}^2 - \|e^n\|_{H^1}^2)\\ &= \mathrm{Re}\Big[\mathrm{i}\langle f(V^{n+1/2}) - f(W^{n+1/2}), V^{n+1/2} - W^{n+1/2}\rangle\\ &\quad + \mathrm{i}\langle\partial_x(f(V^{n+1/2}) - f(W^{n+1/2})), \partial_x(V^{n+1/2} - W^{n+1/2})\rangle\Big]\\ &\quad +\mathrm{Re}\Big[\langle\rho^{n+1} - \sigma^{n+1}, V^{n+1/2} - W^{n+1/2}\rangle\\ &\quad +\langle\partial_x(\rho^{n+1} - \sigma^{n+1}), \partial_x(V^{n+1/2} - W^{n+1/2})\rangle\Big].\end{aligned} \tag{34}$$

Since (31) holds and f is locally Lipschitz in H^1, there exists $L = L(R)$ such that

$$\|f(V^{n+1/2}) - f(W^{n+1/2})\|_{H^1} \leqslant L\|V^{n+1/2} - W^{n+1/2}\|_{H^1},$$

which, along with the inequality

$$\|V^{n+1/2} - W^{n+1/2}\|_{H^1} \leqslant \tfrac{1}{2}(\|e^{n+1}\|_{H^1} + \|e^n\|_{H^1}),$$

can be used in (34) to obtain

$$\frac{1}{\Delta t}(\|e^{n+1}\|_{H^1} - \|e^n\|_{H^1}) \leqslant L(\|e^{n+1}\|_{H^1} + \|e^n\|_{H^1}) + \|\rho^{n+1} - \sigma^{n+1}\|_{H^1},$$

that is

$$\|e^{n+1}\|_{H^1} \leqslant \frac{1 + L\Delta t}{1 - L\Delta t}\|e^n\|_{H^1} + \frac{\Delta t}{1 - L\Delta t}\|\rho^{n+1} - \sigma^{n+1}\|_{H^1}.$$

Finally, if we restrict our attention to the values of Δt for which $L(R)\Delta t < 1/2$, we can estimate

$$\|e^{n+1}\|_{H^1} \leqslant 2\mathrm{e}^{2L\Delta t}\left(\|e^n\|_{H^1} + \Delta t\|\rho^{n+1} - \sigma^{n+1}\|_{H^1}\right), \quad 0 \leqslant n \leqslant N-1,$$

which leads to (32). □

Existence of a numerical solution and convergence

By using the preceding results, we can now state the convergence of the method (28) in the H^1 norm. The next theorem also solves the problem of the existence of a numerical solution.

THEOREM 3.3 Assume the conditions for which (29) holds, set

$$M = \max_{0\leqslant t\leqslant t_{\max}} \|u(t)\|_{H^1}$$

and take $R > 0$ with $R > 4M$. Then

(i) (Existence of numerical solution) For Δt sufficiently small, there exists a unique solution $\mathbf{U}$ of (28) satisfying $\max_{0\leqslant n\leqslant N} \|U^n\|_{H^1} < R$.

(ii) (Convergence) If u^n denotes the theoretical solution at time level $t_n, n = 0, 1, \ldots, N$, then

$$\max_{0\leqslant n\leqslant N} \|U^n - u^n\|_{H^1} = O(\Delta t^2), \quad \Delta t \to 0. \tag{35}$$

Proof. Note first that the operator $A = (I - \mathrm{i}(\Delta t/2)\partial_{xx})^{-1}$ exists in H^1, maps H^1 on H^3 and its norm is less than or equal to one. In terms of the average $Z = U^{n+1/2}$ we can write (28) as

$$Z = A\left[U^n + \mathrm{i}\frac{\Delta t}{2}f(Z)\right], \qquad n = 0, 1, \ldots, N-1.$$

Our proof is based on a fixed point argument and recurrence. We take $U^0 = u_0 \in H^1$ and consider the mapping $G_0 : H^1 \to H^1$,

$$Z \longmapsto G_0(Z) = A\left[U^0 + \mathrm{i}\frac{\Delta t}{2}f(Z)\right].$$

Since the norm of A is $\leqslant 1$ and f is locally Lipschitz in H^1, then, if $\|Z_1\|_{H^1} \leqslant R$, $\|Z_2\|_{H^1} \leqslant R$, we have

$$\|G_0(Z_1) - G_0(Z_2)\|_{H^1} \leqslant \frac{\Delta t}{2}L(R)\|Z_1 - Z_2\|_{H^1},$$

where $L(R)$ is a Lipschitz constant of f; hence, for Δt sufficiently small, G_0 is a contractive mapping and since

$$\|G_0(0)\|_{H^1} \leqslant \|U^0\|_{H^1} = \|u_0\|_{H^1} \leqslant M < \frac{R}{2},$$

(recall $f(0) = 0$) there is a unique $Z^0 \in H^1$ with $\|Z^0\|_{H^1} \leqslant R$ such that $G_0(Z^0) = Z^0$. In fact, $Z^0 \in H^3$ (because A maps H^1 on H^3) so that taking $U^1 = 2Z^0 - U^0$, then $Z^0 = U^{1/2}$ and we complete the first part of the recurrence process. Now, suppose we have obtained $U^0, U^1, \dots, U^n$ with $U^j \in H^1$, $0 \leqslant j \leqslant n$, $U^{j+1/2} \in H^3$, $0 \leqslant j \leqslant n-1$, $\|U^j\|_{H^1} < R/2$, $0 \leqslant j \leqslant n$ and U^j satisfying (28). The above argument applied to the mapping

$$Z \longmapsto G_n(Z) = A\left[U^n + \mathrm{i}\frac{\Delta t}{2} f(Z)\right],$$

is valid to obtain the next level of approximation U^{n+1}. The uniqueness of the numerical solution is a consequence of that of the averages $U^{n+1/2}$. This completes the proof of (i).

On the other hand, the inequality (32) with $V^n = U^n$, $W^n = u^n$ leads to ($U^0 = u_0$)

$$\max_{0 \leqslant n \leqslant N} \|U^n - u^n\|_{H^1} \leqslant C(R, t_{\max}) \Delta t \sum_{k=1}^{n} \|\mathcal{T}^k\|_{H^1},$$

for Δt sufficiently small and where $\mathcal{T}^k$ is the kth component of the local truncation error, so that the consistency result (29) proves (35). □

Asymptotic expansion

We complete the study of (28) with a proof of the existence of an asymptotic expansion of the numerical solution. This result requires a lemma that investigates the local error in more detail and that can be easily proved.

LEMMA 3.4 Suppose that $f \in C^2(\mathcal{R}^2)$, $f(0) = 0$ and the solution u of (1) satisfies that u_{tttt}, u_{ttttt} and $\partial_{xx} u_{tttt}$ exist and are bounded in the H^1 norm uniformly in $[0, t_{\max}]$. Then the local truncation error possesses an expansion of the form

$$\mathcal{T}^{n+1} = \Delta t^2 l(u(t_n)) + \Delta t^4 R(t_n, \Delta t), \quad 0 \leqslant n \leqslant N-1, \tag{36}$$

where the function l does not depend on Δt and R is bounded in the H^1 norm uniformly in $[0, t_{\max}]$.

The description of the asymptotic behaviour of the numerical solution needs more regularity conditions on the nonlinear term f which, in the case of (2), restrict the range of values of σ to $\sigma \geqslant 1$.

THEOREM 3.5 Under the conditions of Lemma 3.4 we additionally assume that $f \in C^3(\mathcal{R}^2)$ and let $e \in H^1$ be the solution of the initial value problem for the nonhomogeneous variational equation,

$$\begin{aligned} \mathrm{i}e_t + e_{xx} + f'(u)e &= -\mathrm{i}l, \quad 0 < t \leqslant t_{\max}, \\ e(0) &= 0, \end{aligned} \tag{37}$$

where l satisfies (36). If the functions e_{tt}, e_{ttt}, $\partial_{xx} e_{tt}$ exist and are bounded in the H^1 norm uniformly in t, $0 \leqslant t \leqslant t_{\max}$, then for $\varDelta t$ sufficiently small, the numerical solution of (28) admits an asymptotic expansion of the form

$$U^n(x) = u(x, t_n) + (\varDelta t)^2 e(x, t_n) + (\varDelta t)^2 q(x, t_n, \varDelta t), \quad 0 \leqslant n \leqslant N,$$

where the function q is a remainder that, for fixed t, satisfies

$$\|q(\cdot, t, \varDelta t)\|_{H^1} \to 0,$$

as $\varDelta t \to 0$.

Proof. By setting $M = \max_{0 \leqslant t \leqslant t_{\max}} \|u(t)\|_{H^1}$ and $R > 0$ with $R > 4M$, we consider the numerical solution $\mathbf{U}$ given by Theorem 3.3 along with the vector $\mathbf{W}$ with components

$$W^n = u(t_n) + (\varDelta t)^2 e(t_n), \quad 0 \leqslant n \leqslant N.$$

If $\varDelta t$ is small, $\|W^n\|_{H^1} < R, \;\; 0 \leqslant n \leqslant N$ and the estimate (32) applied to this case leads to

$$\|U^n - W^n\|_{H^1} \leqslant C \varDelta t \sum_{k=1}^{N} \|\sigma^k\|_{H^1}, \quad 0 \leqslant n \leqslant N,$$

where $\boldsymbol{\sigma}$ is the residual associated with $\mathbf{W}$,

$$\begin{aligned}
\sigma^{n+1} &= \frac{W^{n+1} - W^n}{\varDelta t} - \mathrm{i}\partial_{xx} W^{n+1/2} - \mathrm{i} f(W^{n+1/2}) \\
&= \frac{u(t_{n+1}) - u(t_n)}{\varDelta t} + (\varDelta t)^2 \left(\frac{e(t_{n+1}) - e(t_n)}{\varDelta t} \right) \\
&\quad -\mathrm{i}\partial_{xx} \left(\frac{u(t_{n+1}) + u(t_n)}{2} + (\varDelta t)^2 \left(\frac{e(t_{n+1}) + e(t_n)}{2} \right) \right) \\
&\quad -\mathrm{i} f \left(\frac{u(t_{n+1}) + u(t_n)}{2} + (\varDelta t)^2 \left(\frac{e(t_{n+1}) + e(t_n)}{2} \right) \right),
\end{aligned}$$

with $0 \leqslant n \leqslant N - 1$. Using (36), (37) and Taylor expansions, we can write

$$\begin{aligned}
\sigma^{n+1} &= -\mathrm{i} f \left(\frac{u(t_{n+1}) + u(t_n)}{2} + (\varDelta t)^2 \left(\frac{e(t_{n+1}) + e(t_n)}{2} \right) \right) \\
&\quad +\mathrm{i} f \left(\frac{u(t_{n+1}) + u(t_n)}{2} \right) + \mathrm{i} (\varDelta t)^2 f'(u(t_{n+1/2})) e(t_{n+1/2}) \\
&\quad +(\varDelta t)^4 \left[R(t_n, \varDelta t) + \frac{1}{8} \int_0^1 (1-s)^2 e_{ttt} \left(t_{n+1/2} + \tfrac{1}{2} \varDelta t s \right) \mathrm{d}s \right. \\
&\quad \left. -\frac{\mathrm{i}}{4} \int_0^1 (1-s) \partial_{xx} e_{tt} \left(t_{n+1/2} + \tfrac{1}{2} \varDelta t s \right) \mathrm{d}s \right]. \qquad (38)
\end{aligned}$$

TABLE 1

Errors with respect to the solitary wave at $t = 20$

Δt	IMR	SDIRK3	MCN
2·5E−02	6·9732E−02	1·7030E−01	8·3516E−02
1·25E−02	1·7407E−02	2·2096E−02	2·0847E−02
6·25E−03	4·3498E−03	2·7892E−03	5·2097E−03
3·125E−03	1·0873E−03	3·4962E−04	1·3023E−03

TABLE 2

Errors with respect to the modified solitary wave at $t = 20$

Δt	IMR	SDIRK3	MCN
2·5E−02	3·1608E−03	8·1238E−03	3·4323E−03
1·25E−02	7·8869E−04	2·9002E−04	8·5595E−04
6·25E−03	1·9710E−04	1·3656E−05	2·1389E−04
3·125E−03	4·9264E−05	1·2315E−06	5·3463E−05

Now, observe that the part of (38) involving f and f' can be written as

$$
\begin{aligned}
& f\left(\frac{u(t_{n+1})+u(t_n)}{2} + (\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\right) - f\left(\frac{u(t_{n+1})+u(t_n)}{2}\right)\\
&\quad -(\Delta t)^2 f'(u(t_{n+1/2}))e(t_{n+1/2})\\
&= (\Delta t)^2 \int_0^1 \left(f'\left[\frac{u(t_{n+1})+u(t_n)}{2} + \tau(\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\right]\right.\\
&\quad \times \left.\left(\frac{e(t_{n+1})+e(t_n)}{2}\right) - f'(u(t_{n+1/2}))e(t_{n+1/2})\right)\mathrm{d}\tau \qquad\qquad (39)\\
&= (\Delta t)^2 \int_0^1 \left(f'\left[\frac{u(t_{n+1})+u(t_n)}{2} + \tau(\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\right]\right.\\
&\quad \left. - f'(u(t_{n+1/2}))\right) e(t_{n+1/2})\,\mathrm{d}\tau\\
&\quad -\frac{(\Delta t)^4}{4}\int_0^1 f'\left[\frac{u(t_{n+1})+u(t_n)}{2} + \tau(\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\right]\\
&\quad \times \left(\int_0^1 (1-s)e_{tt}\left(t_{n+1/2}+\tfrac{1}{2}\Delta t s\right)\mathrm{d}s\right)\mathrm{d}\tau,
\end{aligned}
$$

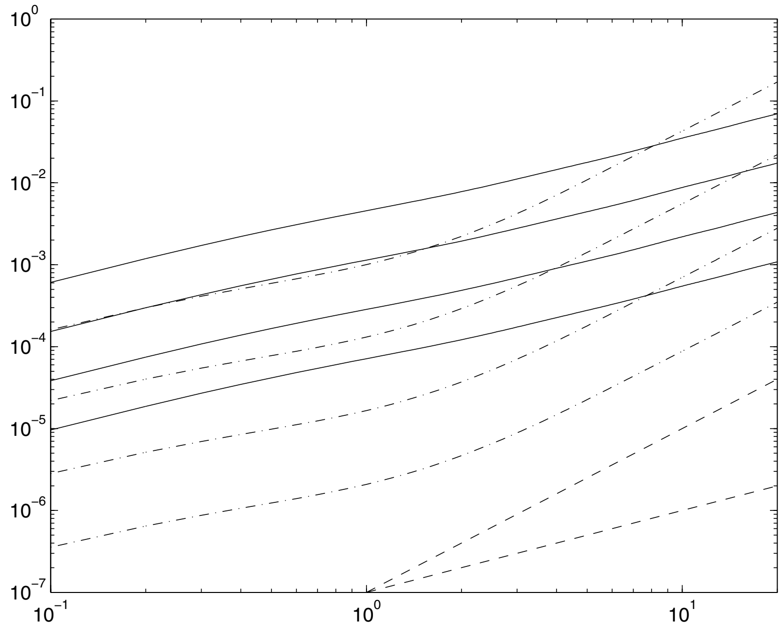

FIG. 1. L^2 error against t. Solid line: midpoint rule; chain line: SDIRK3. The time steps are $\Delta t = 1/40, 1/80, 1/160, 1/320$. The dashed lines at the bottom show the slopes for linear and quadratic growth in time.

and the coefficient of Δt^4 is, by hypothesis, bounded in the H^1 norm. Finally, the integrand of the coefficient of Δt^2 in (39) can be expressed in the form

$$\begin{aligned}
&\Bigg(f'\left[\frac{u(t_{n+1})+u(t_n)}{2}+\tau(\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\right] \\
&\quad - f'(u(t_{n+1/2}))\Bigg)e(t_{n+1/2}) \\
&\quad = \int_0^1 f''\Bigg(\sigma u(t_{n+1/2})+(1-\sigma)\Bigg(\frac{u(t_{n+1})+u(t_n)}{2} \\
&\qquad +\tau(\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\Bigg)\Bigg) \qquad (40) \\
&\qquad \times\left[\frac{u(t_{n+1})+u(t_n)}{2}+\tau(\Delta t)^2\left(\frac{e(t_{n+1})+e(t_n)}{2}\right)\right. \\
&\qquad \left. -u(t_{n+1/2}), e(t_{n+1/2})\right]\mathrm{d}\sigma,
\end{aligned}$$

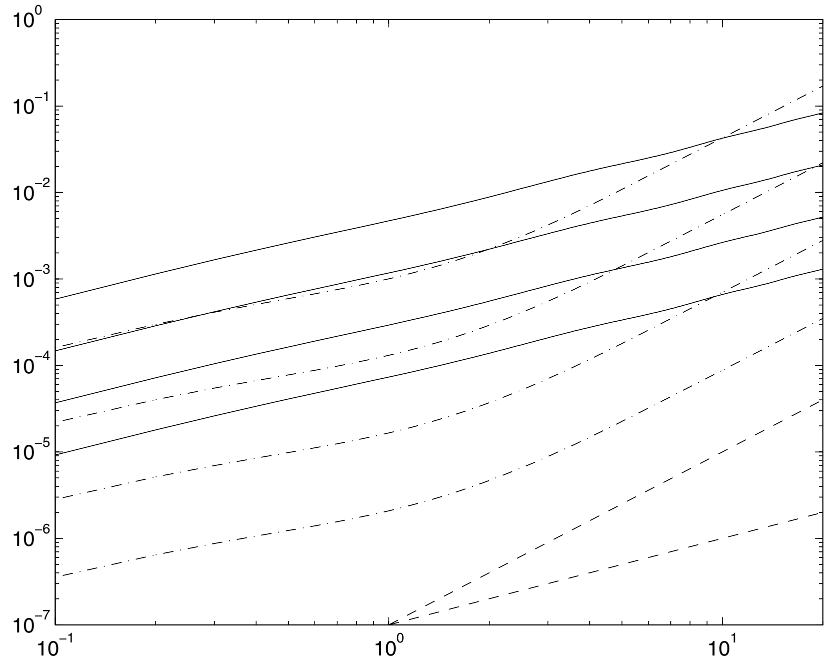

FIG. 2. L^2 error against t. Solid line: MCN; chain line: SDIRK3. $\Delta t = 1/40, 1/80, 1/160, 1/320$. The dashed lines at the bottom show the slopes for linear and quadratic growth in time.

where the expansion of $(u(t_{n+1}) + u(t_n))/2$ in the brackets on the right-hand side of (40) and the regularity condition $f \in C^3$ lead us to write σ^{n+1} as an $O(\Delta t^4)$ term. □

4. Numerical experiments

4.1 *Conservation properties and numerical methods*

We consider the well known cubic Schrödinger equation (equation (2) with $\sigma = 1$, see e.g. Robinson *et al.*, 1993; Sanz-Serna, 1984) and three numerical methods. Two of the schemes considered belong to the family of simply diagonally implicit Runge–Kutta methods (SDIRK)

$$\begin{array}{|rc} \gamma & 0 \\ 1-2\gamma & \gamma \\ \hline 1/2 & 1/2 \end{array}$$

and correspond to the following values of the parameter γ:

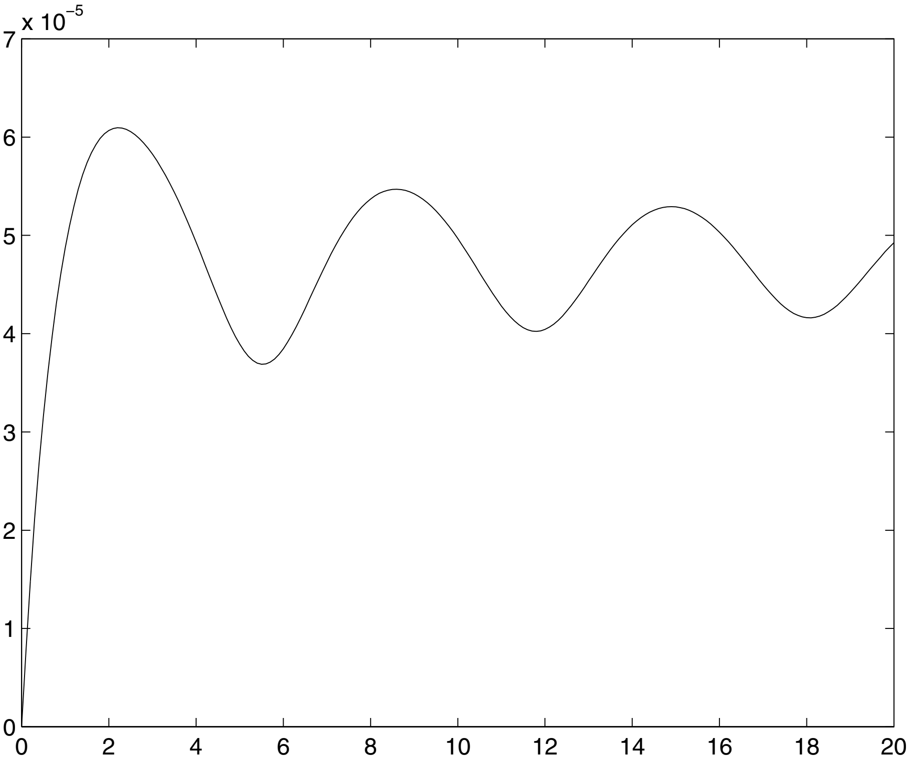

FIG. 3. L^2 error with respect to the modified solitary wave against t. IMR with $\Delta t = 1/320$.

[IMR] $\gamma = 1/2$; that is, the implicit midpoint rule, analysed in Section 3.2, with order two; this method conserves quadratic invariants of the system being integrated (Sanz-Serna & Calvo, 1994).

[SDIRK3] $\gamma = (3 + \sqrt{3})/6$. This value of γ, due (independently) to Nørsett and Crouzeix (see Hairer *et al.*, 1993), gives rise to a third-order method, that is not conservative.

We also consider a third scheme (of order two), which we call MCN, that can be written as

[MCN]

$$\begin{aligned} U^{n+1} &= U^n + \mathrm{i}\Delta t[\partial_{xx}(U^{n+1/2}) + F(U^n, U^{n+1})U^{n+1/2}], \\ U^{n+1/2} &= \frac{U^n + U^{n+1}}{2}, \\ F(U^n, U^{n+1}) &= \frac{|U^n|^2 + |U^{n+1}|^2}{2}. \end{aligned}$$

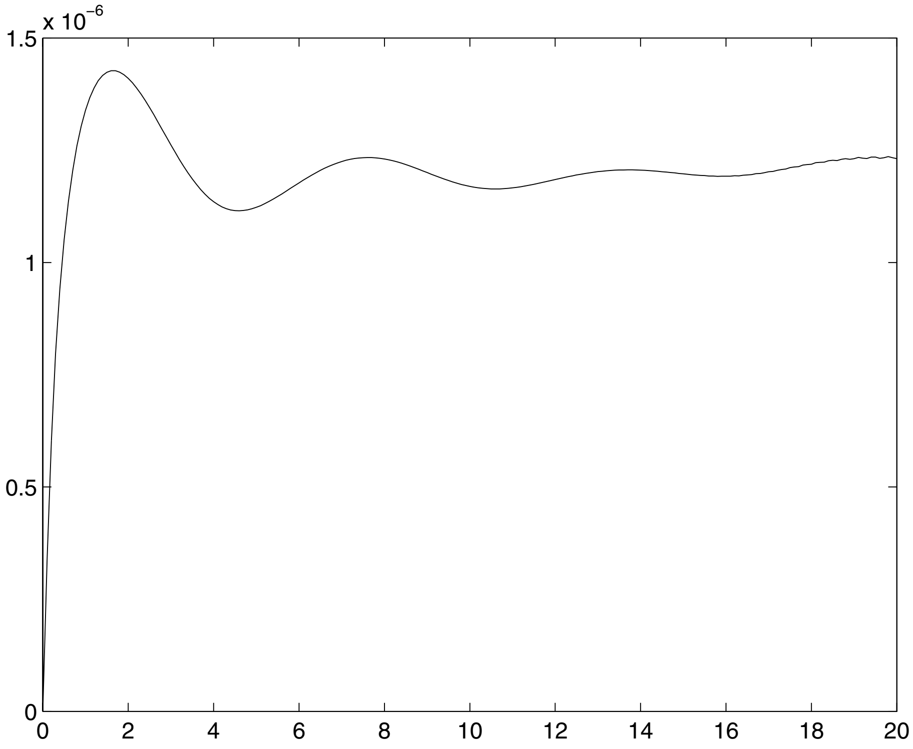

FIG. 4. L^2 error with respect to the modified solitary wave against t. SDIRK3 with $\Delta t = 1/320$.

The construction of this method is based on ensuring conservation of the Hamiltonian (4) (Itoh & Abe, 1988; Strauss & Vázquez, 1978).

The implicit midpoint rule preserves the invariants I_1, I_2 (even in the case of (1) with hypotheses (3), (5) and (6)) but not the Hamiltonian (4). On the other hand, MCN conserves I_1 and H but not I_2 and, finally, SDIRK3 does not conserve any of the three quantities. This different behaviour with regard to conservation of invariants determined our choice of the methods, since our aim is to illustrate differences in the long-time behaviour of the approximations in connection with conservation properties. A comparision between the efficiency of the methods is outside the scope of our work.

4.2 *Numerical results*

We present results concerning approximations to the solitary wave (13) with parameters $a = 1$ (amplitude $\sqrt{2}$), $c = 3$, $x_0 = -10$ and $\theta_0 = \pi/4$. To implement the methods, we use a fully aliased pseudospectral spatial discretization in such a way that, virtually, errors obtained correspond only to the time dicretization. This was achieved, as in Frutos & Sanz-

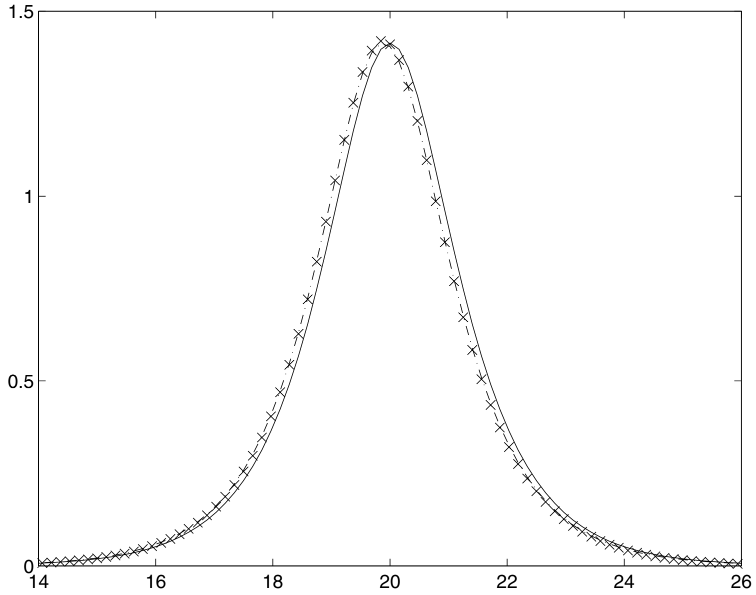

FIG. 5. Midpoint rule with $\Delta t = 1/20$ at $t = 10$. Modulus of the original solitary wave (solid line), modified wave (chain line) and numerical solution (crosses).

Serna (1997), by succesively doubling the number of spatial grid-points until a grid was found for which no further error reduction was possible.

The validity of the hypotheses of Theorem 3.1 was proved in Section 3.2 for the case of IMR. We have no doubt that similar (but even more tedious) analyses can show that the other two methods also fulfil the requisite hypotheses.

We first analyse the behaviour in time of the global error of each method. Figure 1 gives, in a log-log scale, the L^2 norm of the global error as a function of time up to $t_{\max} = 20$, with the solid lines corresponding to IMR and the chain lines to SDIRK3. Similarly, Fig. 2 compares the time behaviour of global errors corresponding to MCN (solid lines) and SDIRK3 (chain lines). The step sizes (for the two figures) are $\Delta t = 1/40, 1/80, 1/160, 1/320$. Observe that the distance between parallel lines corresponding to a given method agrees nicely with the order of convergence, $r = 2$ for IMR and MCN, and $r = 3$ for SDIRK3. This can also be observed in Table 1, which gives L^2 errors at the final time $t_{\max} = 20$.

Recall that, since IMR conserves I_1 and I_2, and MCN preserves I_1 and H, both methods satisfy (26) and therefore the perturbations of the parameters of the original solitary wave grow at most linearly with time. SDIRK3 does not conserve any of the quantities and in fact, for the leading term l of its local error, $\langle -\mathrm{i}l, \nabla I_j(\varphi)\rangle \neq 0$, $j = 1, 2$, so that the corresponding error propagation must be quadratic in time. The slopes of the lines in Figs 1 and 2 show that for the two second-order methods, IMR and MCN, errors

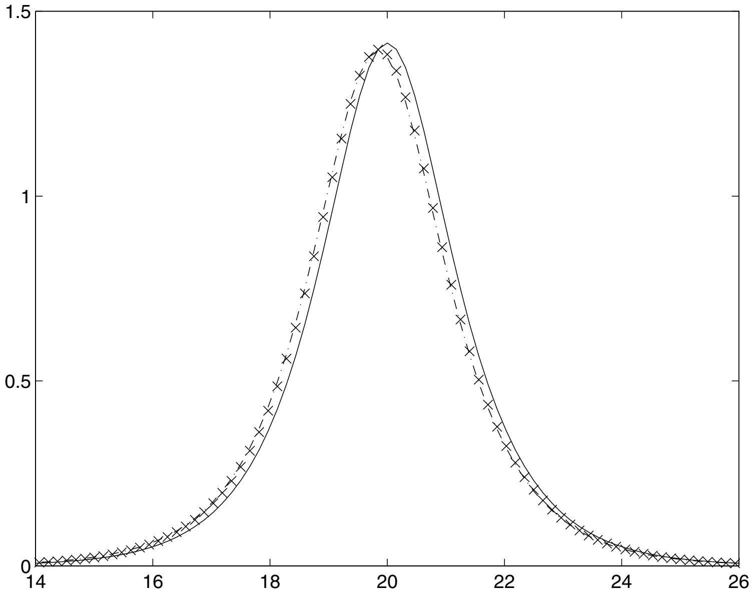

FIG. 6. SDIRK3 with $\Delta t = 1/20$ at $t = 10$. Modulus of the original solitary wave (solid line), modified wave (chain line) and numerical solution (crosses).

grow like t, while for SDIRK3, they grow like t^2 (compare with the lines plotted in the lower right-hand corner of the figures). This confirms our theoretical results.

Modified errors

To study the structure of the error in more detail, we analyse modified errors, i.e. the errors of each method with respect to the modified solitary wave given by (24). We have analytically computed the parameters $\tilde{a}$, $\tilde{c}$, $\tilde{x}_0$, $\tilde{\theta}_0$ of the modified solitary wave and measured the differences $U^n - \psi(t_n, \tilde{a}, \tilde{c}, \tilde{x}_0, \tilde{\theta}_0)$, so that we are measuring the size of the complementary term $\Delta t^r \rho(\cdot, t_n)$ plus the remainder $\Delta t^r Q(\cdot, t_n, \Delta t)$. Table 2 shows, for the three schemes considered, the modified error at final time $t_{\max} = 20$. Comparing Tables 1 and 2, we observe two properties shared by the three schemes: first, errors with respect to the modified solitary wave are much smaller than errors with respect to the original wave; thus the major component of the error corresponds to the errors in the solitary wave parameters that determine the modified solitary wave. On the other hand, Figs 3 and 4 illustrate the evolution in time of the modified error for IMR and SDIRK3 (the case of MCN is similar to that of IMR and will not be discussed as it provides no additional information) and we can see the bounded behaviour of the complementary term and the remainder.

Returning to Table 2, note that in the case of IMR, modified errors behave as $O(\Delta t^2)$, which suggests that, for the values of Δt considered, the complementary term dominates

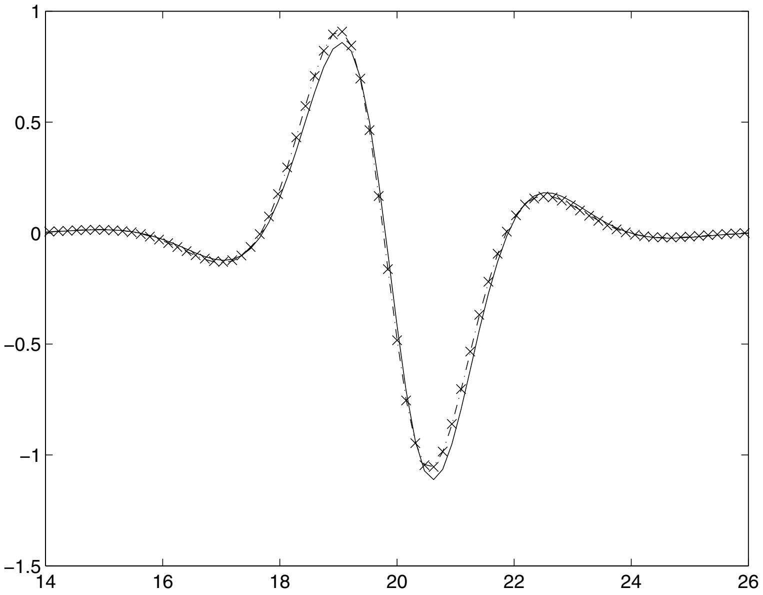

FIG. 7. Midpoint rule with $\Delta t = 1/20$ at $t = 10$. Real part of the original solitary wave (solid line), modified wave (chain line) and numerical solution (crosses).

over the remainder; in fact, the remainder is expected to be especially small $O(\Delta t^4)$ (see (36)) due to the time reversibility of IMR; in contrast, modified errors corresponding to SDIRK3 do not have an $O(\Delta t^3)$ behaviour so that the complementary term is negligible both when compared with the whole error $U^n - \psi(t_n, a, c, x_0, \theta_0)$ and when compared with the remainder.

The behaviour we have described above can also be observed by plotting the solitary waves. For instance, Fig. 5 shows, at $t = 10$, the modulus of the true solitary wave (solid line), the numerical solution (crosses) and the modified solitary wave, with a time step $\Delta t = 1/20$, for IMR. Figure 6 displays the same information but this time for SDIRK3. Observe that, in the case of IMR, the modulus of the modified solitary wave keeps exactly the original profile and travels with a new propagation velocity $c - \alpha_3 \Delta t^2/2$ and a new velocity of the phase, that differs by $(\alpha_1 \sqrt{a}/2 - c\alpha_3/4)\Delta t^2$ from the original one. The numerical solution essentially behaves like this new solitary wave (compare Tables 1 and 2 again) except for a small change in the amplitude, due to the complementary term. On the other hand, Figure 6 displays the changes in the amplitude and velocity of the modified solitary wave for the nonconservative method SDIRK3 and shows the smaller amplitude ($\alpha_2 > 0$) and velocity of propagation of the modified wave.

It is also of interest to study the evolution of the phase of the wave. Figures 7 and 8 respectively display the real part of three waves (original, modified and ‘numerical’) for IMR and SDIRK3 and show the variation of the modified wave in terms of the velocity of

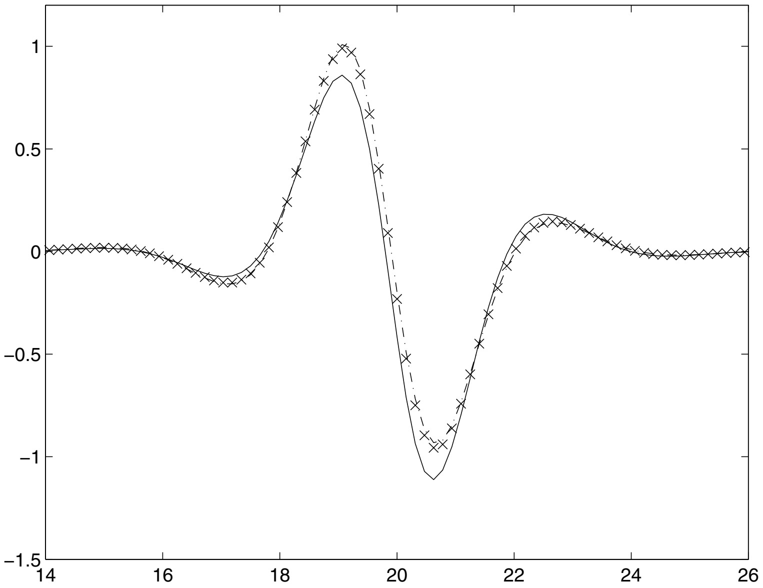

FIG. 8. SDIRK3 with $\Delta t = 1/20$ at $t = 10$. Real part of the original solitary wave (solid line), modified wave (chain line) and numerical solution (crosses).

the phase. In the case of IMR, the perturbation associated with the phase is negative and therefore the velocity of the phase of the modified solitary wave is smaller than that of the true one. Note that, since the modified velocity of propagation is smaller than the original one, the motion of the real part of the modified wave is slightly delayed with respect to that of the true wave; the modified motion is slightly displaced to the left with respect to the true motion. In contrast, in the case of SDIRK3 (Fig. 8) the displacement is to the right, due to the velocity of the phase being larger than for the exact case.

Finally, we can also compare the behaviour, in the numerical integration, of the conserved quantities considered (Frutos & Sanz-Serna, 1997). Figure 9 shows the evolution of the difference

$$H(U^n) - H(\psi(t_n)) = H(U^n) - H(U^0), \tag{41}$$

between the value of the energy (4) at the solitary wave and the discrete version of H at the numerical solution, from $t = 0$ to $t = 10$. Since (4) is a conserved quantity, this difference also estimates the evolution of H in the numerical integration. Observe that, since IMR satisfies (26) and the solitary wave is a relative equilibrium (see (10)), the leading term of its local error is orthogonal to the gradient of H at the wave; this condition and the conservation of the Hamiltonian by ψ show that the leading term of the global error, determined by (37), is also orthogonal to that gradient. Therefore (see Frutos & Sanz-Serna, 1997, Section 2.2) the errors (41) behave as $O(\Delta t^4)$, which can be seen in Fig. 9. In

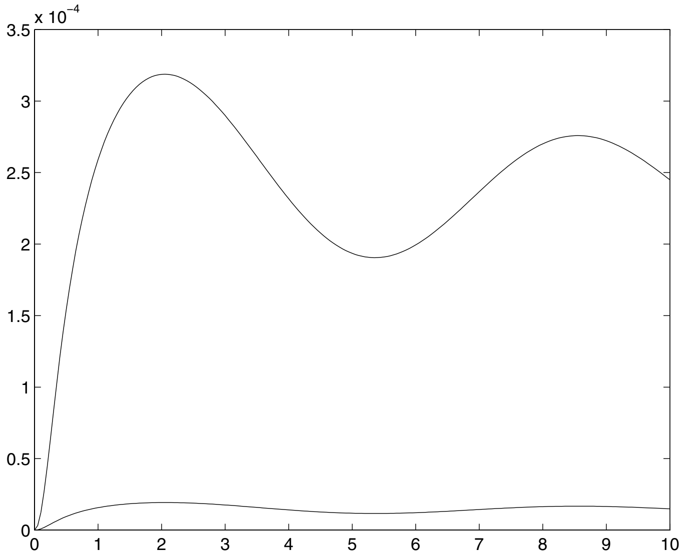

FIG. 9. Energy error against t. IMR with $\Delta t = 1/20, 1/40$.

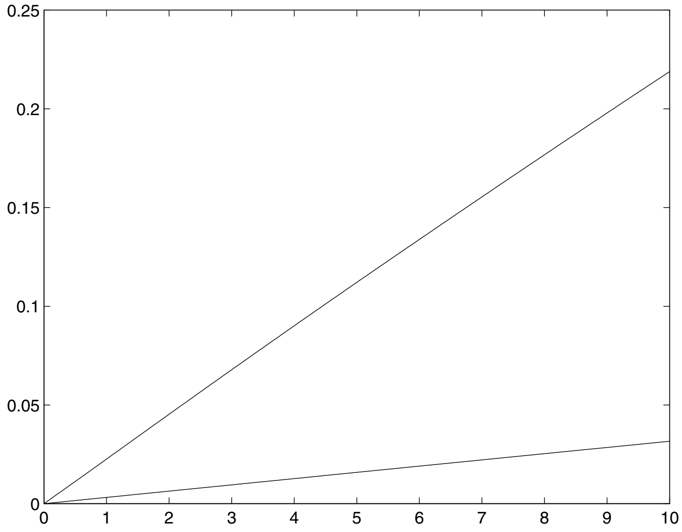

FIG. 10. Energy error against t. SDIRK3 with $\Delta t = 1/20, 1/40$.

the case of SDIRK3, this additional orthogonality condition is not satisfied and, as Fig. 10 shows, the difference (41) has an $O(\Delta t^3)$ behaviour.

Acknowledgements

This research has been supported by project PB95-705.

REFERENCES

ARNOLD, V. I. 1989 *Mathematical Methods of Classical Mechanics*. Berlin: Springer.

CALVO, M. P. & HAIRER, E. 1995 Accurate long-term integration of dynamical systems. *Appl. Numer. Math.* **18**, 95–105.

CALVO, M. P., LOPEZ-MARCOS, M. A., & SANZ-SERNA, J. M. 1998 Variable step implementation of geometric integrators. *Appl. Numer. Math.* **28**, 1–16.

CALVO, M. P. & SANZ-SERNA, J. M. 1993 The development of variable step symplectic integrators, with application to the two-body problem. *SIAM J. Sci. Comput.* **14**, 936–952.

CANO, B. & SANZ-SERNA, J. M. 1997 Error growth in the numerical integration of periodic orbits, with application to Hamiltonian and reversible systems. *SIAM J. Numer. Anal.* **34**, 1391–1417.

CANO, B. & SANZ-SERNA, J. M. 1998 Error growth in the numerical integration of periodic orbits by multistep methods, with application to reversible systems. *IMA J. Numer. Anal.* **18**, 57–75.

DURAN, A. 1997 Propagación del error en la integración numérica de la ecuación no lineal de Schroedinger. *PhD Thesis*, Universidad de Valladolid, Spain.

DURAN, A. & SANZ-SERNA, J. M. 1998 The numerical integration of relative equilibrium solutions. Geometric theory. *Nonlinearity* **11**, 1547–1567.

ESTEP, D. J. & STUART, A. M. 1995 The rate of error growth in Hamiltonian conserving integrators. *Z. Angew. Math. Phys.* **46**, 407–418.

FRUTOS, J. DE & SANZ-SERNA, J. M. 1997 Accuracy and conservation properties in numerical integration: the case of the Korteweg–de Vries equation. *Numer. Math.* **75**, 421–445.

GINIBRE, J. & VELO, G. 1979 On a class of nonlinear Schrödinger equations, I. General case. *J. Func. Anal.* **32**, 1–32.

GINIBRE, J. & VELO, G. 1985 The global Cauchy problem for the nonlinear Schrödinger equation revisited. *Ann. Inst. Henri-Poincaré* **2**, 309–327.

HAIRER, E. & LUBICH, CH. 1997 The life-span of backward error analysis for numerical integrators. *Numer. Math.* **76**, 441–462.

HAIRER, E., NØRSETT, S. P., & WANNER, W. G. 1993 *Solving Ordinary Differential Equations I, Nonstiff Problems* 2nd edn. Berlin: Springer.

ITOH, T. & ABE, K. 1988 Hamiltonian-conserving dicrete canonical equations based on variational difference quotients. *J. Comput. Phys.* **77**, 85–102.

KATO, T. 1987 On nonlinear Schrödinger equations. *Ann. Inst. Henri-Poincaré, Physique Théorique* **46**, 113–129.

KELLEY, P. L. 1965 Self focusing of optical beams. *Phys. Rev. Lett.* **15**, 1005–1008.

LOPEZ-MARCOS, J. C. & SANZ-SERNA, J. M. 1988 A definition of stability for nonlinear problems. *Numerical Treatments of Differential Equations* (Strehmel, ed). Leipzig: Teubner, pp 216–226.

MARSDEN, J. E. & RATIU, T. S. 1994 *Introduction to Mechanics and Symmetry.* New York: Springer.

OLVER, P. J. 1993 *Applications of Lie Groups to Differential Equations* 2nd edn. New York: Springer.

ROBINSON, M. P., FAIRWEATHER, G., & HERBST, B. M. 1993 On the numerical solution of the cubic Schrödinger equation in one space variable. *J. Comput. Phys.* **104**, 277–284.

SANZ-SERNA, J. M. 1984 Methods for the numerical solution of the nonlinear Schrödinger equation. *Math. Comput.* **43**, 21–27.

SANZ-SERNA, J. M. 1997 Geometric integration. *The State of the Art in Numerical Analysis* (I. S. Duff and G. A. Watson, eds). Oxford: Clarendon, pp 121–143.

SANZ-SERNA, J. M. & CALVO, M. P. 1994 *Numerical Hamiltonian Problems*. London: Chapman & Hall.

STOFFER, D. 1995 Variable steps for reversible integration methods. *Computing* **55**, 1–22.

STRAUSS, W. A. & VÁZQUEZ, L. 1978 Numerical solution of a nonlinear Klein–Gordon equation. *J. Comput. Phys.* **28**, 271–278.

WEINSTEIN, M. I. 1985 Modulational stability of ground states of nonlinear Schrödinger equations. *SIAM J. Math. Anal.* **16**, 473–491.

WHITHAM, G. B. 1974 *Linear and Nonlinear Waves.* New York: Wiley.

ZAKHAROV, V. E. 1972 Collapse of Langmuir waves. *Sov. Phys.–JETP* **35**, 908–922.