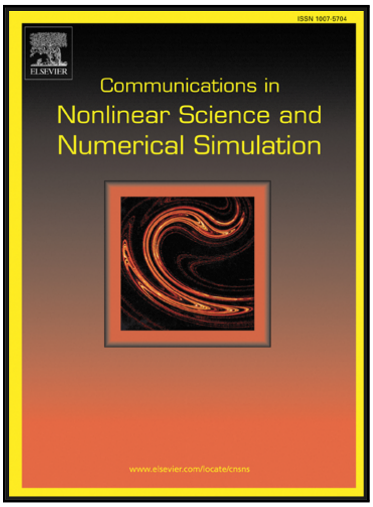

A class of energy-conserving Hamiltonian boundary value methods for nonlinear Schrödinger equation with wave operator

Luigi Brugnano, Chengjian Zhang, Dongfang Li

PII: S1007-5704(17)30440-9
DOI: 10.1016/j.cnsns.2017.12.018
Reference: CNSNS 4409

To appear in: *Communications in Nonlinear Science and Numerical Simulation*

Received date: 17 October 2017
Revised date: 26 December 2017
Accepted date: 30 December 2017

Highlights

- A conservation property for the nonlinear Schrodinger equation with wave operator is studied.
- A spectrally accurate space semi-discretization is considered.
- A class of energy-conserving boundary value methods for the induced large-size Hamiltonian ODE problem is considered.
- The efficient implementation of the methods is studied and numerical tests are reported.

A class of energy-conserving Hamiltonian boundary value methods for nonlinear Schrödinger equation with wave operator[☆]

Luigi Brugnano[a], Chengjian Zhang[b,c,*], Dongfang Li[b,c]

[a]*Dipartimento di Matematica e Informatica “U. Dini”, Università di Firenze, Viale Morgagni 67/A, I-50134 Firenze, Italy*
[b]*School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China*
[c]*Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Science and Technology, Wuhan 430074, China*

Abstract

In this paper, we study the efficient solution of the nonlinear Schrödinger equation with wave operator, subject to periodic boundary conditions. In such a case, it is known that its solution conserves a related functional. By using a Fourier expansion in space, the problem is at first casted into Hamiltonian form, with the same Hamiltonian functional. A Fourier-Galerkin space semi-discretization then provides a large-size Hamiltonian ODE problem, whose solution in time is carried out by means of energy-conserving methods in the HBVM class (Hamiltonian Boundary Value Methods). The efficient implementation of the methods for the resulting problem is also considered and some numerical examples are reported.

Keywords: Nonlinear Schrödinger equation, Hamiltonian problem, Wave operator, Energy-conserving methods, Hamiltonian boundary value methods
MSC: 65P10, 65N40.

1. Introduction

In this paper, we deal with the numerical solution of the following nonlinear Schrödinger equation with wave operator:

$$u_{tt}(x,t) - c^2 u_{xx}(x,t) + 2\mathrm{i}\alpha u_t(x,t) + \beta(x) f'(|u(x,t)|^2)u(x,t) = 0, \qquad (x,t) \in \Omega := [a,b] \times [0,\infty), \tag{1.1}$$

where, as is usual, the subscript denotes the partial derivative w.r.t. the given variable. Moreover, i is the imaginary unit, α and $c \neq 0$ are real constants, and β and f are real functions, with f' the derivative of f. Equations of this type have many different applications in Physics, such as nonrelativistic limit of the Klein-Gordon equation [25, 26, 28], Langmuir wave envelope approximation in plasma physics [4], model of planar light bullets [2, 34] and so forth. For this reason, it has been subject of investigation, both from a theoretical (see, e.g., [21]) and, more recently, also from a numerical point of view (see, e.g., [1, 19, 20, 23, 24, 29–33]). We here consider the case where the equation (1.1) is completed with the initial conditions:

$$u(x,0) = u_0(x), \qquad u_t(x,0) = v_0(x), \qquad x \in [a,b], \qquad \text{and} \quad \text{periodic b.c.} \tag{1.2}$$

Consequently, both u_0 and v_0 will be assumed to be periodic functions, regular enough (as a periodic function). We shall also assume β to be periodic and suitably regular, even though the periodicity would

[☆] Work supported by Università di Firenze (project “Risoluzione numerica di problemi Hamiltoniani ed applicazioni”) and NSFC (Grant No. 11571128).

[*] Corresponding author.

Email addresses: `luigi.brugnano@unifi.it` (Luigi Brugnano), `cjzhang@mail.hust.edu.cn` (Chengjian Zhang), `dfli@mail.hust.edu.cn` (Dongfang Li)

be not strictly needed (see Remark 3.2 below). Also f is assumed to be suitably regular. It is known by the following Theorem 2.1 that the solution of problem (1.1)–(1.2) conserves the functional:

$$\mathcal{H}[u](t) = \frac{1}{2}\int_a^b |u_t(x,t)|^2 + c^2|u_x(x,t)|^2 + \beta(x)f(|u(x,t)|^2)\,\mathrm{d}x, \tag{1.3}$$

so that

$$\mathcal{H}[u](t) = \mathcal{H}[u](0), \qquad \forall t \geq 0. \tag{1.4}$$

Hence, the conservation property (1.3)–(1.4) is important for the correct numerical simulation of such problem. As an example, when $\beta(x), f(\xi) > 0$, $x \in [a,b]$, $\xi \geq 0$, the conservation of (1.3) implies the boundedness of the partial derivatives of the solution. This, in turn, implies the boundedness of the solution, upon regularity assumptions on β, f, u_0, v_0 are made (see, e.g., [33] or Corollary 2.4 below). In addition to this, the conservation of the Hamiltonian functional has proved to confer more robustness on the numerical solution (see, e.g., [3, 6], for the nonlinear Schrödinger equation and the semilinear wave equation, respectively). For this reason, in this paper we are concerned with the numerical solution of problem (1.1)–(1.2), while exactly conserving an arbitrarily high-order approximation to (1.3). We would like to emphasize that we shall here consider the case where in (1.1) $x \in [a,b]$ (i.e., the 1D case), even though the arguments can be naturally extended to the case where $x \in [a_1,b_1]\times\cdots\times[a_d,b_d]$, with $d \geq 1$ (in which case, u_{xx} becomes Δu).

With this premise, the structure of the paper is as follows: in Section 2 we cast the problem into real form, also verifying the conservation property (1.3)–(1.4), moreover, we recast the problem into Hamiltonian form, by considering a Fourier-type expansion in space; next, in Section 3 we consider a semi-discrete problem, which amounts to a large-size Hamiltonian system of ODEs; in Section 4 we sketch the basic facts about Hamiltonian Boundary Value Methods (HBVMs), which we shall use to solve the problem in time while conserving the energy, and also explaining the details about their efficient implementation for the problem at hand; in Section 5 we collect some test problems; at last, in Section 6 we report a few concluding remarks.

2. Fourier expansion in space

To begin with, let us pose the problem (1.1)–(1.2) in real form. By setting

$$u(x,t) = \varphi(x,t) + \mathrm{i}\psi(x,t), \qquad u_0(x) = \varphi_0(x) + \mathrm{i}\psi_0(x), \qquad v_0(x) = \varphi_1(x) + \mathrm{i}\psi_1(x), \tag{2.1}$$

the real and imaginary parts of the involved functions, we see that (1.1) can be rewritten as

$$\begin{array}{rcll} \varphi_{tt} - c^2\varphi_{xx} - 2\alpha\psi_t + \beta(x)f'(\varphi^2+\psi^2)\varphi &=& 0, & \\ \psi_{tt} - c^2\psi_{xx} + 2\alpha\varphi_t + \beta(x)f'(\varphi^2+\psi^2)\psi &=& 0, & (x,t)\in\Omega. \end{array} \tag{2.2}$$

Hereafter, for sake of brevity, we often avoid to explicitly mention the arguments (x,t). Finally, the initial conditions (1.2) become

$$\varphi(x,0) = \varphi_0(x), \quad \psi(x,0) = \psi_0(x), \quad \varphi_t(x,0) = \varphi_1(x), \quad \psi_t(x,0) = \psi_1(x), \qquad x \in [0,1], \tag{2.3}$$

with periodic boundary conditions. In a similar way, the functional (1.3) becomes

$$\mathcal{H}[\varphi,\psi](t) = \frac{1}{2}\int_a^b \varphi_t(x,t)^2 + \psi_t(x,t)^2 + c^2[\varphi_x(x,t)^2 + \psi_x(x,t)^2] + \beta(x)f(\varphi(x,t)^2+\psi(x,t)^2)\,\mathrm{d}x. \tag{2.4}$$

In the sequel, when not necessary we shall also omit the arguments (x,t) for the functions appearing in the functional $\mathcal{H}$, for sake of brevity. By using (2.2) one proves the conservation of the fucntional $\mathcal{H}$.

Theorem 2.1. *The functional (2.4) is conserved along the solution of problem (2.2)–(2.3).*

Proof. In fact, one has, by using (2.2), integration by parts, taking into account the periodic boundary conditions, and denoting, as is usual, with the dot the time derivative:

$$\begin{aligned}\dot{\mathcal{H}}[\varphi,\psi](t) &= \int_a^b \varphi_t\varphi_{tt}+\psi_t\psi_{tt}+c^2[\varphi_x\varphi_{xt}+\psi_x\psi_{xt}]+\beta(x)f'(\varphi^2+\psi^2)(\varphi\varphi_t+\psi\psi_t)\,\mathrm{d}x\\ &= \int_a^b \varphi_t\varphi_{tt}+\psi_t\psi_{tt}-c^2[\varphi_t\varphi_{xx}+\psi_t\psi_{xx}]+\beta(x)f'(\varphi^2+\psi^2)(\varphi\varphi_t+\psi\psi_t)\,\mathrm{d}x\\ &= \int_a^b \varphi_t\left[\varphi_{tt}-c^2\varphi_{xx}+\beta(x)f'(\varphi^2+\psi^2)\varphi\right]+\psi_t\left[\psi_{tt}-c^2\psi_{xx}+\beta(x)f'(\varphi^2+\psi^2)\psi\right]\mathrm{d}x\\ &= \int_a^b \varphi_t[2\alpha\psi_t]+\psi_t[-2\alpha\varphi_t]\,\mathrm{d}x \;=\; 0.\end{aligned}$$

This implies the theorem holds. □

Next, because of the periodic boundary conditions, we expand the functions φ and ψ in space, by using the following orthonormal basis for periodic functions in $L^2[a,b]$,

$$c_j(x)=\sqrt{\frac{2-\delta_{j0}}{b-a}}\cos\left(2\pi j\frac{x-a}{b-a}\right),\qquad j\ge 0,\qquad s_j(x)=\sqrt{\frac{2}{b-a}}\sin\left(2\pi j\frac{x-a}{b-a}\right),\qquad j\ge 1, \tag{2.5}$$

with δ_{j0} the Kronecker delta, such that for all allowed values of i and j:

$$\int_a^b c_i(x)\,c_j(x)\mathrm{d}x=\delta_{ij}=\int_a^b s_i(x)\,s_j(x)\mathrm{d}x,\qquad \int_a^b c_i(x)\,s_j(x)\mathrm{d}x=0. \tag{2.6}$$

Consequently, for suitable time dependent coefficients $\gamma_j(t),\eta_j(t),\alpha_j(t),\beta_j(t)$, one has the expansions:

$$\begin{aligned}\varphi(x,t) &= c_0(x)\gamma_0(t)+\sum_{j\ge1}c_j(x)\gamma_j(t)+s_j(x)\eta_j(t),\\ \psi(x,t) &= c_0(x)\alpha_0(t)+\sum_{j\ge1}c_j(x)\alpha_j(t)+s_j(x)\beta_j(t).\end{aligned} \tag{2.7}$$

Thus, the periodic boundary conditions result to be fulfilled. The expansions (2.7) can be cast in a more compact form, by defining the infinite vectors

$$\boldsymbol{w}(x)=\begin{pmatrix}c_0(x)\\c_1(x)\\s_1(x)\\c_2(x)\\s_2(x)\\\vdots\end{pmatrix},\qquad \boldsymbol{q}_1(t)=\begin{pmatrix}\gamma_0(t)\\\gamma_1(t)\\\eta_1(t)\\\gamma_2(t)\\\eta_2(t)\\\vdots\end{pmatrix},\qquad \boldsymbol{q}_2(t)=\begin{pmatrix}\alpha_0(t)\\\alpha_1(t)\\\beta_1(t)\\\alpha_2(t)\\\beta_2(t)\\\vdots\end{pmatrix}, \tag{2.8}$$

as follows:

$$\varphi(x,t)=\boldsymbol{w}(x)^\top\boldsymbol{q}_1(t),\qquad \psi(x,t)=\boldsymbol{w}(x)^\top\boldsymbol{q}_2(t). \tag{2.9}$$

In so doing, we can easily compute the partial derivatives:

$$\begin{aligned}\varphi_t(x,t)&=\boldsymbol{w}(x)^\top\dot{\boldsymbol{q}}_1(t), &\qquad \varphi_{tt}(x,t)&=\boldsymbol{w}(x)^\top\ddot{\boldsymbol{q}}_1(t),\\ \varphi_x(x,t)&=\boldsymbol{w}'(x)^\top\boldsymbol{q}_1(t), &\qquad \varphi_{xx}(x,t)&=\boldsymbol{w}''(x)^\top\boldsymbol{q}_1(t),\end{aligned} \tag{2.10}$$

and similarly for ψ. The following result holds true.

Lemma 2.2. *Let us define the matrices*

$$J_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = -J_2^\top = -J_2^{-1}, \qquad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.11}$$

and the following infinite matrices:

$$\tilde{D} = \frac{2\pi}{b-a} \begin{pmatrix} 0 & & & \\ & 1 \cdot J_2 & & \\ & & 2 \cdot J_2 & \\ & & & \ddots \end{pmatrix} = -\tilde{D}^\top, \tag{2.12}$$

$$D = \tilde{D}^\top \tilde{D} \equiv \left(\frac{2\pi}{b-a}\right)^2 \begin{pmatrix} 0 & & & \\ & 1^2 \cdot I_2 & & \\ & & 2^2 \cdot I_2 & \\ & & & \ddots \end{pmatrix}. \tag{2.13}$$

Then

$$\boldsymbol{w}'(x) = \tilde{D}\boldsymbol{w}(x), \qquad \boldsymbol{w}''(x) = -D\boldsymbol{w}(x). \tag{2.14}$$

Proof. The first equality in (2.14) follows by observing that (see (2.8) ans (2.11)–(2.12))

$$\tilde{D}\boldsymbol{w}(x) = \begin{pmatrix} 0 \\ -s_1(x) \\ c_1(x) \\ -2s_2(x) \\ 2c_2(x) \\ \vdots \end{pmatrix} \equiv \boldsymbol{w}'(x).$$

The second equality then also derives, by observing that (see (2.13)) $-D = \tilde{D}^2$. □

As a consequence of the previous Lemma, the space derivatives in (2.10) can be expressed as

$$\varphi_x(x,t) = (\tilde{D}\boldsymbol{w}(x))^\top \boldsymbol{q}_1(t), \qquad \varphi_{xx}(x,t) = -(D\boldsymbol{w}(x))^\top \boldsymbol{q}_1(t), \tag{2.15}$$

and similarly for ψ_x and ψ_{xx}. By considering that, because of the orthogonality conditions (2.6),

$$\int_a^b \boldsymbol{w}(x)\boldsymbol{w}(x)^\top \mathrm{d}x = I, \tag{2.16}$$

the identity operator, one easily derives that (2.2) can be recast in the "frequency" space as,

$$\begin{aligned} \ddot{\boldsymbol{q}}_1 + c^2 D\boldsymbol{q}_1 - 2\alpha\dot{\boldsymbol{q}}_2 + \int_a^b \boldsymbol{w}\beta f'((\boldsymbol{w}^\top\boldsymbol{q}_1)^2 + (\boldsymbol{w}^\top\boldsymbol{q}_2)^2)(\boldsymbol{w}^\top\boldsymbol{q}_1)\mathrm{d}x &= 0, \\ \ddot{\boldsymbol{q}}_2 + c^2 D\boldsymbol{q}_2 + 2\alpha\dot{\boldsymbol{q}}_1 + \int_a^b \boldsymbol{w}\beta f'((\boldsymbol{w}^\top\boldsymbol{q}_1)^2 + (\boldsymbol{w}^\top\boldsymbol{q}_2)^2)(\boldsymbol{w}^\top\boldsymbol{q}_2)\mathrm{d}x &= 0, \qquad t > 0, \end{aligned} \tag{2.17}$$

where, for sake of brevity, we also skip the argument x for the functions β and $\boldsymbol{w}$. In order to pose the problem into first order form, let us define the infinite vectors

$$\boldsymbol{p}_1(t) = \dot{\boldsymbol{q}}_1(t) - \alpha\boldsymbol{q}_2(t), \qquad \boldsymbol{p}_2(t) = \dot{\boldsymbol{q}}_2(t) + \alpha\boldsymbol{q}_1(t), \tag{2.18}$$

so that (2.17) is rewritten as

$$\begin{aligned}
\dot{\boldsymbol{q}}_1 &= \boldsymbol{p}_1 + \alpha \boldsymbol{q}_2, \\
\dot{\boldsymbol{q}}_2 &= \boldsymbol{p}_2 - \alpha \boldsymbol{q}_1, \\
\dot{\boldsymbol{p}}_1 &= -c^2 D\boldsymbol{q}_1 + \alpha(\boldsymbol{p}_2 - \alpha\boldsymbol{q}_1) - \int_a^b \boldsymbol{w}\beta f'((\boldsymbol{w}^\top \boldsymbol{q}_1)^2 + (\boldsymbol{w}^\top \boldsymbol{q}_2)^2)(\boldsymbol{w}^\top \boldsymbol{q}_1)\mathrm{d}x, \\
\dot{\boldsymbol{p}}_2 &= -c^2 D\boldsymbol{q}_2 - \alpha(\boldsymbol{p}_1 + \alpha\boldsymbol{q}_2) - \int_a^b \boldsymbol{w}\beta f'((\boldsymbol{w}^\top \boldsymbol{q}_1)^2 + (\boldsymbol{w}^\top \boldsymbol{q}_2)^2)(\boldsymbol{w}^\top \boldsymbol{q}_2)\mathrm{d}x, \qquad t > 0,
\end{aligned} \tag{2.19}$$

with the initial conditions:

$$\begin{aligned}
\boldsymbol{q}_1(0) &= \int_a^b \boldsymbol{w}(x)\varphi_0(x)\mathrm{d}x, & \boldsymbol{q}_2(0) &= \int_a^b \boldsymbol{w}(x)\psi_0(x)\mathrm{d}x, \\
\boldsymbol{p}_1(0) &= \int_a^b \boldsymbol{w}(x)(\varphi_1(x) - \alpha\psi_0(x))\mathrm{d}x, & \boldsymbol{p}_2(0) &= \int_a^b \boldsymbol{w}(x)(\psi_1(x) + \alpha\varphi_0(x))\mathrm{d}x.
\end{aligned} \tag{2.20}$$

The following result then holds true.

Theorem 2.3. *The system of equations (2.19) is Hamiltonian with Hamiltonian*

$$\begin{aligned}
H(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{p}_1, \boldsymbol{p}_2) &= \frac{1}{2}\big[(\boldsymbol{p}_1 + \alpha\boldsymbol{q}_2)^\top(\boldsymbol{p}_1 + \alpha\boldsymbol{q}_2) + (\boldsymbol{p}_2 - \alpha\boldsymbol{q}_1)^\top(\boldsymbol{p}_2 - \alpha\boldsymbol{q}_1) \\
&\quad + c^2[\boldsymbol{q}_1^\top D\boldsymbol{q}_1 + \boldsymbol{q}_2^\top D\boldsymbol{q}_2] + \int_a^b \beta f((\boldsymbol{w}^\top\boldsymbol{q}_1)^2 + (\boldsymbol{w}^\top\boldsymbol{q}_2)^2)\mathrm{d}x\big].
\end{aligned} \tag{2.21}$$

Moreover H is equivalent to the functional $\mathcal{H}$ defined in (2.4).

Proof. The first part of the proof is straightforward, since one easily realizes that

$$\dot{\boldsymbol{q}}_i = \frac{\partial H}{\partial \boldsymbol{p}_i}, \qquad \dot{\boldsymbol{p}}_i = -\frac{\partial H}{\partial \boldsymbol{q}_i}, \qquad i = 1, 2.$$

In order to prove that H is equivalent to the functional $\mathcal{H}$ defined in (2.4), it is enough to observe that, by virtue of (2.9)–(2.16) and (2.19):

$$\begin{aligned}
(\boldsymbol{p}_1 + \alpha\boldsymbol{q}_2)^\top(\boldsymbol{p}_1 + \alpha\boldsymbol{q}_2) &= \dot{\boldsymbol{q}}_1^\top\dot{\boldsymbol{q}}_1 = \int_a^b (\boldsymbol{w}^\top\dot{\boldsymbol{q}}_1)^\top(\boldsymbol{w}^\top\dot{\boldsymbol{q}}_1)\mathrm{d}x = \int_a^b \varphi_t^2\mathrm{d}x, \\
(\boldsymbol{p}_2 - \alpha\boldsymbol{q}_1)^\top(\boldsymbol{p}_2 - \alpha\boldsymbol{q}_1) &= \dot{\boldsymbol{q}}_2^\top\dot{\boldsymbol{q}}_2 = \int_a^b (\boldsymbol{w}^\top\dot{\boldsymbol{q}}_2)^\top(\boldsymbol{w}^\top\dot{\boldsymbol{q}}_2)\mathrm{d}x = \int_a^b \psi_t^2\mathrm{d}x, \\
\boldsymbol{q}_1^\top D\boldsymbol{q}_1 = \boldsymbol{q}_1^\top\tilde{D}\tilde{D}^\top\boldsymbol{q}_1 &= \int_a^b \boldsymbol{q}_1^\top\tilde{D}\boldsymbol{w}\boldsymbol{w}^\top\tilde{D}^\top\boldsymbol{q}_1\mathrm{d}x = \int_a^b \varphi_x^2\mathrm{d}x, \\
\boldsymbol{q}_2^\top D\boldsymbol{q}_2 = \boldsymbol{q}_2^\top\tilde{D}\tilde{D}^\top\boldsymbol{q}_2 &= \int_a^b \boldsymbol{q}_2^\top\tilde{D}\boldsymbol{w}\boldsymbol{w}^\top\tilde{D}^\top\boldsymbol{q}_2\mathrm{d}x = \int_a^b \psi_x^2\mathrm{d}x,
\end{aligned}$$

and

$$\int_a^b \beta f((\boldsymbol{w}^\top\boldsymbol{q}_1)^2 + (\boldsymbol{w}^\top\boldsymbol{q}_2)^2)\mathrm{d}x = \int_a^b \beta f(\varphi^2 + \psi^2)\mathrm{d}x.$$

Consequently, from (2.21) one obtains that

$$H(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{p}_1, \boldsymbol{p}_2) = \frac{1}{2}\int_a^b \left[\varphi_t^2 + \psi_t^2 + c^2(\varphi_x^2 + \psi_x^2) + \beta f(\varphi^2 + \psi^2)\right]\mathrm{d}x.$$

This completes the proof. □

As anticipated in the introduction, the conservation of (2.21) has relevant implications in the solution of problem (2.19)–(2.20), as stated in the following result.

Corollary 2.4. *Assume that for problem (2.19)–(2.20) one has* $\beta, f > 0$ *and, moreover,*

$$H(\boldsymbol{q}_1(0), \boldsymbol{q}_2(0), \boldsymbol{p}_1(0), \boldsymbol{p}_2(0)) < \infty.$$

Then, the solution of the problem is uniformly bounded.

Proof. The statement easily follows from the conservation of (2.21), which implies that $\|\boldsymbol{q}_i\|_2^2$ and, therefore, $\|\boldsymbol{p}_i\|_2^2$, $i = 1, 2$, are bounded. □

3. Fourier-Galerkin space semi-discretization

In order for problem (2.19)–(2.20) to be solvable on a computer, one needs to truncate the infinite expansions (2.7) to finite sums. Therefore, having fixed a conveniently large value $N \gg 1$, one approximates (2.7) as

$$\begin{aligned} \varphi(x,t) \;\approx\; \hat\varphi(x,t) &= c_0(x)\gamma_0(t) + \sum_{j=1}^N c_j(x)\gamma_j(t) + s_j(x)\eta_j(t), \\ \psi(x,t) \;\approx\; \hat\psi(x,t) &= c_0(x)\alpha_0(t) + \sum_{j=1}^N c_j(x)\alpha_j(t) + s_j(x)\beta_j(t). \end{aligned} \qquad (3.1)$$

The truncated expansions (3.1) can be cast in a vector form similar to (2.9), by replacing the infinite vectors and matrices (2.8) and (2.12)–(2.13), respectively by

$$\boldsymbol{w}(x) = \begin{pmatrix} c_0(x) \\ c_1(x) \\ s_1(x) \\ \vdots \\ c_N(x) \\ s_N(x) \end{pmatrix}, \quad \boldsymbol{q}_1(t) = \begin{pmatrix} \gamma_0(t) \\ \gamma_1(t) \\ \eta_1(t) \\ \vdots \\ \gamma_N(t) \\ \eta_N(t) \end{pmatrix}, \quad \boldsymbol{q}_2(t) = \begin{pmatrix} \alpha_0(t) \\ \alpha_1(t) \\ \beta_1(t) \\ \vdots \\ \alpha_N(t) \\ \beta_N(t) \end{pmatrix} \in \mathbb{R}^{2N+1}, \qquad (3.2)$$

and

$$\tilde D = \frac{2\pi}{b-a} \begin{pmatrix} 0 &&& \\ & 1\cdot J_2 && \\ && \ddots & \\ &&& N\cdot J_2 \end{pmatrix} = -\tilde D^\top \in \mathbb{R}^{2N+1\times 2N+1}, \qquad (3.3)$$

$$D = \tilde D^\top \tilde D \equiv \left(\frac{2\pi}{b-a}\right)^2 \begin{pmatrix} 0 &&& \\ & 1^2\cdot I_2 && \\ && \ddots & \\ &&& N^2\cdot I_2 \end{pmatrix} \in \mathbb{R}^{2N+1\times 2N+1}, \qquad (3.4)$$

where we continue to use the same notation for the infinite vectors and matrices and the corresponding truncated versions, in order not to complicate the notations, even though, hereafter, they will denote the finite ones. Consequently, one obtains the expansions

$$\hat\varphi(x,t) = \boldsymbol{w}(x)^\top \boldsymbol{q}_1(t), \qquad \hat\psi(x,t) = \boldsymbol{w}(x)^\top \boldsymbol{q}_2(t), \qquad (3.5)$$

in place of (2.8)–(2.9). Moreover, expressions similar to (2.10) hold true for the partial derivatives of $\hat\varphi$ and $\hat\psi$, as well as the result of Lemma 2.2 continues formally to hold for the truncated vectors (3.2). As a result, equations (2.14)–(2.16) continue formally to hold for the finite approximations, even though now the functions (3.5) don't satisfy the equations (2.2) anymore. However, in the spirit of Galerkin methods, by requiring the residual be orthogonal to the functional space

$$\mathcal{V}_N = \operatorname{span}\{c_0(x), c_1(x), s_1(x), \dots c_N(x), s_N(x)\}, \tag{3.6}$$

to which the approximations (3.5) belong for all t, one formally obtains again the equations (2.19), with the initial conditions formally still given by (2.20). Clearly, (2.19) is Hamiltonian, with Hamiltonian formally still given by (2.21), even though, this latter is now only an approximation to the functional (2.4). Nevertheless, it is known from the theory of Fourier methods [17] that, under regularity assumptions on $\varphi_0, \psi_0, \varphi_1, \psi_1$, β, and f, one has that the truncated approximations to φ, ψ, and $\mathcal{H}$ converge more than exponentially to them, as $N \to \infty$ (this fact is usually referred to as *spectral accuracy*).

Remark 3.1. *A criterion for the choice of N is to check that the residuals (see (2.20))*

$$\|\varphi_0 - \boldsymbol{w}^\top \boldsymbol{q}_1(0)\|, \quad \|\psi_0 - \boldsymbol{w}^\top \boldsymbol{q}_2(0)\|, \quad \|(\varphi_1 - \alpha\psi_0) - \boldsymbol{w}^\top \boldsymbol{p}_1(0)\|, \quad \|(\psi_1 + \alpha\varphi_0) - \boldsymbol{w}^\top \boldsymbol{p}_2(0)\|, \tag{3.7}$$

corresponding to the initial conditions, are small enough, and, moreover, the difference of the values of

$$H(\boldsymbol{q}_1(0), \boldsymbol{q}_2(0), \boldsymbol{p}_1(0), \boldsymbol{p}_2(0))$$

is within round-off error level, when passing from N to $N+1$.

Finally, in order to obtain a full space semi-discretization, one needs to conveniently approximate the integrals appearing in (2.19). For this purpose, as observed in [6], one can use a composite trapezoidal rule, evaluated at the abscissae,

$$x_i = a + i\frac{b-a}{m}, \qquad i = 0, \dots, m, \tag{3.8}$$

with m suitably large (see, e.g., [18, Th. 5.1.4] and [22, Th. 1.1]). Hence, the truncated problem (2.19), having dimension $4(2N+1)$, with the integrals approximated via the composite trapezoidal rule at the abscissae (3.8), define the semi-discrete problem in space to be integrated in time. The corresponding semi-discrete Hamiltonian is then given formally by (2.21), with the integral appearing in it approximated via the composite trapezoidal rule based at the abscissae (3.8).

Remark 3.2. *We observe that when the function $\beta(x)$ in (1.1) is not periodic in $[a, b]$, then a different quadrature has to be used to compute the integrals in (2.19), in place of the composite trapezoidal rule based at the abscissae (3.8). E.g., a high-order Gaussian formula.*

Remark 3.3. *As is clear, the proposed Fourier-Galerkin space semi-discretization is tailored for the case of periodic boundary conditions. For sake of completeness, we mention that for general boundary conditions different space semi-discretizations should be considered and, moreover, the Hamiltonian functional (2.4) may be no longer conserved (even though its variation can still be correctly reproduced, as is shown, e.g., in [6] for the semi-linear wave equation). It must be also emphasized that the used space semi-discretization greatly affects the efficient implementation of the fully discrete method, as is shown in the next section. In fact, the chosen space semi-discretization will result in an approximate Jacobian with diagonal blocks, which allows for a very efficient implementation of the method.*

4. Hamiltonian Boundary Value Methods

In order to obtain a fully discrete method, we now need to integrate the Hamiltonian problem (2.19)–(2.20) with the vectors $\boldsymbol{w}, \boldsymbol{q}_1, \boldsymbol{q}_2$ and matrix D defined by (3.2)–(3.4). The fact of obtaining a Hamiltonian semi-discrete ODE problem, from a PDE with Hamiltonian structure, is important, as observed in [27], if one uses a suitable *geometric integrator*, able to take advantage of this property. For this reason, we shall here consider the energy-conserving Runge-Kutta methods named *Hamiltonian Boundary Value Methods (HBVMs)* for numerically solving (2.19)–(2.20). Such methods have been studied in a series of papers

[5, 8, 9, 11, 12] and have been generalized along several directions, including the application to Hamiltonian PDEs [3, 6] (see also the recent monograph [7], for a thorough introduction to such methods). In more detail, a HBVM(k,s) method is the k-stage Runge-Kutta method with Butcher tableau

$$\begin{array}{c|c} \boldsymbol{c} & \mathcal{I}_s\mathcal{P}_s^\top\Omega \\ \hline & \boldsymbol{b}^\top \end{array}, \qquad \boldsymbol{b} = \begin{pmatrix} b_1 & \dots & b_k \end{pmatrix}^\top, \quad \boldsymbol{c} = \begin{pmatrix} c_1 & \dots & c_k \end{pmatrix}^\top, \tag{4.1}$$

where, by setting $\{P_j\}_{j\ge0}$ the Legendre polynomial basis, orthonormal on $[0,1]$,

$$P_i \in \Pi_i, \qquad \int_0^1 P_i(x)P_j(x)\mathrm{d}x = \delta_{ij,}, \qquad \forall i,j = 0,1,\dots,$$

(b_i, c_i) are the weights and abscissae of the Gauss-Legendre quadrature formula of order $2k$ (i.e., $P_k(c_i) = 0$, $i = 1,\dots,k$), and

$$\mathcal{P}_s = \begin{pmatrix} P_{j-1}(c_i) \end{pmatrix}, \; \mathcal{I}_s = \begin{pmatrix} \int_0^{c_i} P_{j-1}(x)\mathrm{d}x \end{pmatrix} \in \mathbb{R}^{k\times s}, \qquad \Omega = \mathrm{diag}(\boldsymbol{b}) \in \mathbb{R}^{k\times k}. \tag{4.2}$$

It is also known that (see, e.g., [7]) a HBVM(k,s) method applied for solving the ODE-IVPs

$$\dot{y} = g(y), \qquad t\in[0,h], \qquad y(0) = y_0,$$

with h the considered stepsize, defines a polynomial approximation $\sigma \in \Pi_s$ such that

$$\sigma(0) = y_0, \qquad y_1 := \sigma(h) \approx y(h), \qquad \dot{\sigma}(ch) = \sum_{j=0}^{s-1} P_j(c)\hat{\gamma}_j, \quad c\in[0,1], \tag{4.3}$$

with

$$\hat{\gamma}_j := \sum_{i=1}^k b_i P_j(c_j) g(\sigma(c_i h)) \;\equiv\; \int_0^1 P_j(c)g(\sigma(ch))\mathrm{d}c + \Delta_j(h), \qquad j = 0,\dots,s-1, \tag{4.4}$$

and the quadrature error

$$\Delta_j(h) = \begin{cases} 0, & \text{if } \; g(\sigma)\in\Pi_\nu, \; \text{ with } \nu \le 2k-1-j, \\ O(h^{2k-j}), & \text{otherwise.} \end{cases} \tag{4.5}$$

On the basis of the previous statements, the following result holds true.

Theorem 4.1. *For all $k \ge s$, the k-stage Runge-Kutta HBVM(k,s) method (4.1):*

- *is symmetric and has order $2s$;*
- *when $k = s$ it reduces to the (symplectic) s-stage Gauss collocation method;*
- *it is energy-conserving for problem (2.19)–(2.21) when f is a polynomial of degree*

$$\tilde{\nu} \le k/s; \tag{4.6}$$

- *in the non polynomial case, by setting, with reference to the Hamiltonian function defined at (2.21),*

$$\boldsymbol{y} := \begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \end{pmatrix} \in \mathbb{R}^{4(2N+1)}, \qquad H(\boldsymbol{y}) := H(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{p}_1, \boldsymbol{p}_2), \tag{4.7}$$

$\boldsymbol{y}_0 := \boldsymbol{y}(0)$, and $\boldsymbol{y}_1 \approx \boldsymbol{y}(h)$ the new approximation, with h the used stepsize, one has that

$$H(\boldsymbol{y}_1) - H(\boldsymbol{y}_0) = O(h^{2k+1}). \tag{4.8}$$

Proof. For the first two points, we refer, e.g., to [7, 11]. Concerning the last two points, let us rewrite (2.19)–(2.20), by using the more compact notation (4.7) for the vector of the unknowns, as

$$\dot{\boldsymbol{y}} = J\nabla H(\boldsymbol{y}), \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0, \qquad J = \begin{pmatrix} O_{4N+2} & I_{4N+2} \\ -I_{4N+2} & O_{4N+2} \end{pmatrix}, \tag{4.9}$$

with O_{4N+2} and I_{4N+2} the $4N+2 \times 4N+2$ zero and identity matrices, respectively. Consequently, by taking into account of (4.3)–(4.5), the method will induce a polynomial approximation $\boldsymbol{\sigma} \in \Pi_s$ such that

$$\boldsymbol{\sigma}(0) \;=\; \boldsymbol{y}_0, \qquad \boldsymbol{\sigma}(h) \;=:\; \boldsymbol{y}_1, \qquad \dot{\boldsymbol{\sigma}}(ch) \;=\; \sum_{j=0}^{s-1} P_j(c)\hat{\boldsymbol{\gamma}}_j, \quad c \in [0,1],$$

with

$$\begin{aligned} \hat{\gamma}_j \;&:=\; \sum_{i=1}^{k} b_i P_j(c_j) J\nabla H(\boldsymbol{\sigma}(c_i h)) \;\equiv\; \boldsymbol{\gamma}_j(\boldsymbol{\sigma}) + \boldsymbol{\Delta}_j(h), \qquad\qquad (4.10)\\ \boldsymbol{\gamma}_j(\boldsymbol{\sigma}) \;&=\; \int_0^1 P_j(c) J\nabla H(\boldsymbol{\sigma}(ch))\mathrm{d}c,\\ \boldsymbol{\Delta}_j(h) \;&=\; \begin{cases} 0, & \text{if } \; J\nabla H(\boldsymbol{\sigma}) \in \Pi_\nu, \; \text{ with } \nu \le 2k-1-j, \\ O(h^{2k-j}), & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, by taking into account that (see (4.9)) $J^\top J = I$, the identity matrix of dimension $4(2N+1)$, one has that

$$\begin{aligned} H(\boldsymbol{y}_1) - H(\boldsymbol{y}_0) \;&=\; H(\boldsymbol{\sigma}(h)) - H(\boldsymbol{\sigma}(0)) \;=\; h\int_0^1 \nabla H(\boldsymbol{\sigma}(ch))^\top \dot{\boldsymbol{\sigma}}(ch)\mathrm{d}c \\ =\; h\int_0^1 \nabla H(\boldsymbol{\sigma}(ch))^\top \dot{\boldsymbol{\sigma}}(ch)\mathrm{d}c \;&=\; h\int_0^1 \nabla H(\boldsymbol{\sigma}(ch))^\top \sum_{j=0}^{s-1} P_j(c)\left[\boldsymbol{\gamma}_j(\boldsymbol{\sigma}) + \boldsymbol{\Delta}_j(h)\right]\mathrm{d}c \\ =\; h\sum_{j=0}^{s-1}\left[\int_0^1 P_j(c) J\nabla H(\boldsymbol{\sigma}(ch))\mathrm{d}c\right]^\top J\left[\boldsymbol{\gamma}_j(\boldsymbol{\sigma}) + \boldsymbol{\Delta}_j(h)\right] \;&=\; h\sum_{j=0}^{s-1} \boldsymbol{\gamma}_j(\boldsymbol{\sigma})^\top J\left[\boldsymbol{\gamma}_j(\boldsymbol{\sigma}) + \boldsymbol{\Delta}_j(h)\right] \\ =\; h\sum_{j=0}^{s-1} \boldsymbol{\gamma}_j(\boldsymbol{\sigma})^\top J\boldsymbol{\Delta}_j(h) \;&=\; \begin{cases} 0, & \text{if } \; J\nabla H(\boldsymbol{\sigma}) \in \Pi_\nu, \; \text{ with } \nu \le 2k-s, \\ O(h^{2k+1}), & \text{otherwise.} \end{cases} \end{aligned}$$

In the second case, (4.8) follows. In the former case, one has that the Hamiltonian is conserved, provided that $J\nabla H(\boldsymbol{\sigma}) \in \Pi_\nu$ with $\nu \le 2k-s$, i.e. (see (2.21)), $f \in \Pi_{\tilde{\nu}}$ with

$$(\tilde{\nu} - 1)2s + s \le 2k - s,$$

from which (4.6) follows. □

Remark 4.1. *As is clear from (4.6), by choosing k large enough, one can always gain an* exact *energy conservation, in the polynomial case. However, also in the non-polynomial case one can still obtain a* practical *energy conservation by choosing k large enough, since it suffices to make the error (4.8) fall within the round-off error level.*

The use of a large value of k, in turn, doesn't make the implementation of the Runge-Kutta method (4.1) too much costly, since, as we are going to sketch below, the discrete problem generated by its application has (block) dimension s, *independently* of k. As matter of fact, by considering the formulation (4.9) of problem (2.19)–(2.20), denoting $\boldsymbol{e} = (1,\dots,1)^\top \in \mathbb{R}^k$, and setting

$$Y \equiv \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \in \mathbb{R}^{4k(2N+1)}, \qquad \nabla H(Y) := \begin{pmatrix} \nabla H(Y_1) \\ \vdots \\ \nabla H(Y_k) \end{pmatrix}, \tag{4.11}$$

the stage vector for the method (4.1) applied for solving (4.9), and ∇H evaluated at the stages, respectively, one obtains the nonlinear set of k vector equations

$$Y = \boldsymbol{e} \otimes \boldsymbol{y}_0 + h\mathcal{I}_s\mathcal{P}_s^\top \Omega \otimes J\,\nabla H(Y). \tag{4.12}$$

However, by further setting the vector

$$\hat{\boldsymbol{\gamma}} \equiv \begin{pmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_{s-1} \end{pmatrix} := \mathcal{P}_s^\top \Omega \otimes J\,\nabla H(Y), \tag{4.13}$$

containing the vector coefficients (4.10), it follows that (4.12) can be written as

$$Y = \boldsymbol{e} \otimes \boldsymbol{y}_0 + h\mathcal{I}_s \otimes I\,\hat{\boldsymbol{\gamma}}. \tag{4.14}$$

By plugging (4.14) into (4.13), one then obtains the equation

$$G(\hat{\boldsymbol{\gamma}}) := \hat{\boldsymbol{\gamma}} - \mathcal{P}_s^\top \Omega \otimes J\,\nabla H\left(\boldsymbol{e} \otimes \boldsymbol{y}_0 + h\mathcal{I}_s \otimes I\hat{\boldsymbol{\gamma}}\right) = \boldsymbol{0}, \tag{4.15}$$

whose (block) dimension is s, independently of k. Once the discrete problem (4.15) has been solved, the new approximation is then given by

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 + h\hat{\gamma}_0.$$

In fact, taking into account (4.2), (4.10), (4.13), setting $\boldsymbol{e}_1 \in \mathbb{R}^s$ the first unit vector, and considering that $P_0(x) \equiv 1$, one has

$$\hat{\gamma}_0 = \boldsymbol{e}_1^\top \mathcal{P}_s^\top \Omega \otimes J\nabla H(Y) = \boldsymbol{e}^\top \Omega \otimes J\,\nabla H(Y) \equiv \sum_{i=1}^k b_i J\nabla H(Y_i),$$

(we refer to, e.g., [7, 11] for full details). Consequently, the complexity for solving the discrete problem (4.15) generated by the application of the HBVM(k,s) method (4.1) is greatly simplified, w.r.t. solving the stage equation (4.12). In addition to this, by taking into account that

$$\mathcal{P}_s^\top \Omega \mathcal{I}_s = X_s := \begin{pmatrix} \xi_0 & -\xi_1 & & \\ \xi_1 & 0 & \ddots & \\ & \ddots & \ddots & -\xi_{s-1} \\ & & \xi_{s-1} & 0 \end{pmatrix} \in \mathbb{R}^{s\times s}, \qquad \xi_i = \frac{1}{2\sqrt{|4i^2-1|}}, \tag{4.16}$$

one has that the simplified Newton iteration for solving (4.15) reads

$$\begin{array}{rcl} \text{FOR } r & = & 0,1,\dots: \\ \text{solve} & & \left[I_s \otimes I - hX_s \otimes J\nabla^2 H(\boldsymbol{y}_0)\right]\Delta^r = -G(\hat{\boldsymbol{\gamma}}^r) \\ \text{set} & & \hat{\boldsymbol{\gamma}}^{r+1} = \hat{\boldsymbol{\gamma}}^r + \Delta^r \\ \text{END} & & \end{array} \tag{4.17}$$

starting, e.g., from $\hat{\boldsymbol{\gamma}}^0 = \mathbf{0}$. We observe that the coefficient matrix of the linear system in (4.17) has dimension s times larger than that of the continuous problem (2.19), and we need to factor it at each integration step. However, we can simplify this procedure along two directions, as below explained.

Firstly, by setting I_ℓ the identity matrix of dimension

$$\ell := 2N + 1, \tag{4.18}$$

and considering matrix D defined at (3.4), one has

$$\nabla^2 H(\boldsymbol{y}_0) = \begin{pmatrix} (c^2 D + \alpha^2 I_\ell + F_{11}) & F_{12} & & -\alpha I_\ell \\ F_{12} & (c^2 D + \alpha^2 I_\ell + F_{22}) & \alpha I_\ell & \\ & \alpha I_\ell & I_\ell & \\ -\alpha I_\ell & & & I_\ell \end{pmatrix},$$

with

$$F_{ij} = \int_a^b \beta \boldsymbol{w}\boldsymbol{w}^\top \left[2f''((\boldsymbol{w}^\top q_1)^2 + (\boldsymbol{w}^\top q_2)^2)(\boldsymbol{w}^\top \boldsymbol{q}_i)(\boldsymbol{w}^\top \boldsymbol{q}_j) + \delta_{ij} f'((\boldsymbol{w}^\top q_1)^2 + (\boldsymbol{w}^\top q_2)^2))\right] \mathrm{d}x.$$

Hence, when $N \gg 1$ (see (3.4)), one may assume

$$\left(\frac{2\pi c}{b-a} N\right)^2 \gg \|\beta\|(\|f''\| + \|f'\|).$$

Consequently, by setting

$$D(c,\alpha) := c^2 D + \alpha^2 I_\ell, \tag{4.19}$$

we can consider the approximate Hessian matrix

$$\nabla^2 H(\boldsymbol{y}_0) \approx \begin{pmatrix} D(c,\alpha) & & & -\alpha I_\ell \\ & D(c,\alpha) & \alpha I_\ell & \\ & \alpha I_\ell & I_\ell & \\ -\alpha I_\ell & & & I_\ell \end{pmatrix} =: M, \tag{4.20}$$

which is *constant*.

Secondly, in place of the simplified Newton iteration (4.17) with the simplified Hessian (4.20), we consider a "splitting-Newton" *blended iteration*. This iteration, at first devised in [13], has been implemented in the computational code `BiM` [15] (which is available at the *Test Set for IVP Solvers* [35]), and has also been considered for HBVMs [5, 9], proving to be very efficient when applied to Hamiltonian PDEs, as is shown in [6] for the semi-linear wave equation, and in [3] for the nonlinear Schrödinger equation. We here sketch the main facts for the solution of problem (2.19), since each PDE has its own structural properties to be exploited in order to make efficient the nonlinear iteration. In more details, the iteration (4.17) is replaced by the following one:

```
FOR $r$ = $0, 1, \dots$ :                                                   (4.21)
    set    $\boldsymbol{\eta}^r = -G(\hat{\boldsymbol{\gamma}}^r)$
    set    $\boldsymbol{\eta}_1^r = \rho_s X_s^{-1} \otimes I\, \boldsymbol{\eta}^r$
    set    $\Delta^r = I_s \otimes \Sigma\, [\boldsymbol{\eta}_1^r + I_s \otimes \Sigma\, (\boldsymbol{\eta}^r - \boldsymbol{\eta}_1^r)]$
    set    $\hat{\boldsymbol{\gamma}}^{r+1} = \hat{\boldsymbol{\gamma}}^r + \Delta^r$
END
```

starting, e.g., from $\hat{\boldsymbol{\gamma}}^0 = \mathbf{0}$. Here X_s is the matrix defined at (4.16),

$$\rho_s = \min_{\lambda \in \sigma(X_s)} |\lambda|, \tag{4.22}$$

(a few values of the parameter ρ_s are listed in Table 1), and (see (4.9), (4.18), and (4.20))

$$\Sigma := (I - h\rho_s JM)^{-1} \in \mathbb{R}^{4\ell\times 4\ell}. \tag{4.23}$$

This latter matrix, having the same size as that of the continuous problem (2.19), is *constant* and, therefore, needs to be computed only once. Moreover, by considering that (see (4.19))

$$\Sigma^{-1} = \begin{pmatrix} I_\ell & -\delta I_\ell & -\varepsilon I_\ell & O_\ell \\ \delta I_\ell & I_\ell & O_\ell & -\varepsilon I_\ell \\ \varepsilon D(c,\alpha) & O_\ell & I_\ell & -\delta I_\ell \\ O_\ell & \varepsilon D(c,\alpha) & \delta I_\ell & I_\ell \end{pmatrix}, \qquad \text{with} \qquad \varepsilon := h\rho_s,\ \delta := \alpha\varepsilon, \tag{4.24}$$

and O_ℓ the zero $\ell\times\ell$ matrix, has a block diagonal structure, the following result holds true.

Theorem 4.2. *Let define the permutation matrix P of dimension $4\ell \equiv 4(2N+1)$ such that*

$$P\begin{pmatrix} 1 \\ 2 \\ \vdots \\ 4\ell \end{pmatrix} = (\,1,\ \ell+1,\ 2\ell+1,\ 3\ell+1,\ 2,\ \ell+2,\ 2\ell+2,\ 3\ell+2,\ \dots,\ \ell, 2\ell,\ 3\ell,\ 4\ell\,)^\top .$$

Then

$$\Sigma = P^\top \begin{pmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_\ell \end{pmatrix} P,$$

where, by setting as is usual $\lfloor x \rfloor$ the largest integer less than or equal to x,

$$\Sigma_i = \begin{pmatrix} 1 & -\delta & -\varepsilon & 0 \\ \delta & 1 & 0 & -\varepsilon \\ \xi_i & 0 & 1 & -\delta \\ 0 & \xi_i & \delta & 1 \end{pmatrix}^{-1}, \quad \xi_i := \varepsilon\left[\left(\frac{2\pi c\lfloor i/2 \rfloor}{b-a}\right)^2 + \alpha^2\right], \quad i = 1,\dots,\ell. \tag{4.25}$$

Proof. From (4.19) and (4.24), one has that

$$\Sigma^{-1} = \varepsilon c^2 F + G \otimes I_\ell,$$

with

$$F = \begin{pmatrix} O_\ell & O_\ell & O_\ell & O_\ell \\ O_\ell & O_\ell & O_\ell & O_\ell \\ D & O_\ell & O_\ell & O_\ell \\ O_\ell & D & O_\ell & O_\ell \end{pmatrix}, \qquad G = \begin{pmatrix} 1 & -\delta & -\varepsilon & 0 \\ \delta & 1 & 0 & -\varepsilon \\ \varepsilon\alpha^2 & 0 & 1 & -\delta \\ 0 & \varepsilon\alpha^2 & \delta & 1 \end{pmatrix}.$$

The statement then follows by considering that $P^\top(G\otimes I_\ell)P = I_\ell \otimes G$ and, by virtue of (3.4),

$$P^\top F P = \begin{pmatrix} F_1 & & \\ & \ddots & \\ & & F_\ell \end{pmatrix},$$

with

$$F_i = \left(\frac{2\pi}{b-a}\left\lfloor \frac{i}{2} \right\rfloor\right)^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad i = 1,\dots,\ell.$$

Table 1: Parameter defined at (4.22).

s	1	2	3	4	5	6
ρ_s	0.5000	0.2887	0.1967	0.1475	0.1173	0.0971

This completes the proof. □

As a consequence of Theorem 4.2, in order to perform the blended iteration (4.21), one needs to compute (see (4.18)) the ℓ 4×4 matrices in (4.25). Actually, approximately half of them, since from (4.25) one easily realizes that

$$\xi_{2i} = \xi_{2i+1} \quad \Rightarrow \quad \Sigma_{2i} = \Sigma_{2i+1}, \qquad i = 1, \dots, N \equiv (\ell - 1)/2.$$

In conclusion, one obtains that the linear algebra cost for performing the iteration (4.21) is *linear* in the dimension of the problem (2.19) to be solved, both in terms of required operations and memory requirements.

5. Numerical examples

In this section we provide a couple of numerical examples, aimed at confirming the conservation properties and accuracy of the proposed method.

The first problem is

$$u_{tt} - u_{xx} + \mathrm{i}u_t + \sin(2|u|^2)u = 0, \qquad (x,t) \in [0,1] \times [0,T], \tag{5.1}$$

for which the function f (see (1.1)) is non-polynomial, with the initial conditions

$$u(x,0) = \exp(2\pi \mathrm{i}x), \qquad u_t(x,0) = 2\pi\mathrm{i}\exp(2\pi\mathrm{i}x), \qquad x \in [0,1]. \tag{5.2}$$

The corresponding functional (1.3) is then given by

$$\mathcal{H}[u](t) = \frac{1}{2}\int_0^1 |u_t|^2 + |u_x|^2 + \sin^2(|u|^2)\mathrm{d}x. \tag{5.3}$$

Moreover, the solution of problem (5.1)–(5.2) turns out to be in the form

$$u(x,t) = \rho(t)\exp(2\pi\mathrm{i}(x + \omega(t)),$$

with $\rho(t) \approx 1$ and $\omega(t)$ real and smooth functions, as is shown in Figure 1. Consequently, choosing $N = 1$ in (3.1) and $m = 3$ in (3.8) is enough to have an exact representation of the solution and of the equations. The corresponding value of the Hamiltonian functional (5.3) turns out to be given by

$$\mathcal{H}[u] \equiv H_0 = 4\pi^2 + \frac{\sin^2(1)}{2} \approx 39.83. \tag{5.4}$$

In Table 2 we list the obtained result by using HBVM(k,s), with $s = 1$ and $k = 1, 2, 3$, and 10, when fixing $T = 2$ and using a time-step $h = 2^{-n}$, $n = 1, \dots, 10$. Similarly, in Table 3 we list the obtained result by using HBVM(k,s), with $s = 2$ and $k = 2, 3, 4$, and 10, when fixing $T = 2$ and using a time-step $h = 2^{-n}$, $n = 1, \dots, 8$. In both tables, err_u denotes the maximum error in the computed solution, which has been numerically estimated, whereas err_H is the error in the numerical Hamiltonian, whose value is known to be given by (5.4). We also list the corresponding estimated convergence rates (** means that the round-off error level has been reached), along with the mean number of blended iterations (4.21) per step. From the figures in the two tables, one easily deduces that the HBVM(k,s) method, according to Theorem 4.1:

- has the prescribed order $2s$;

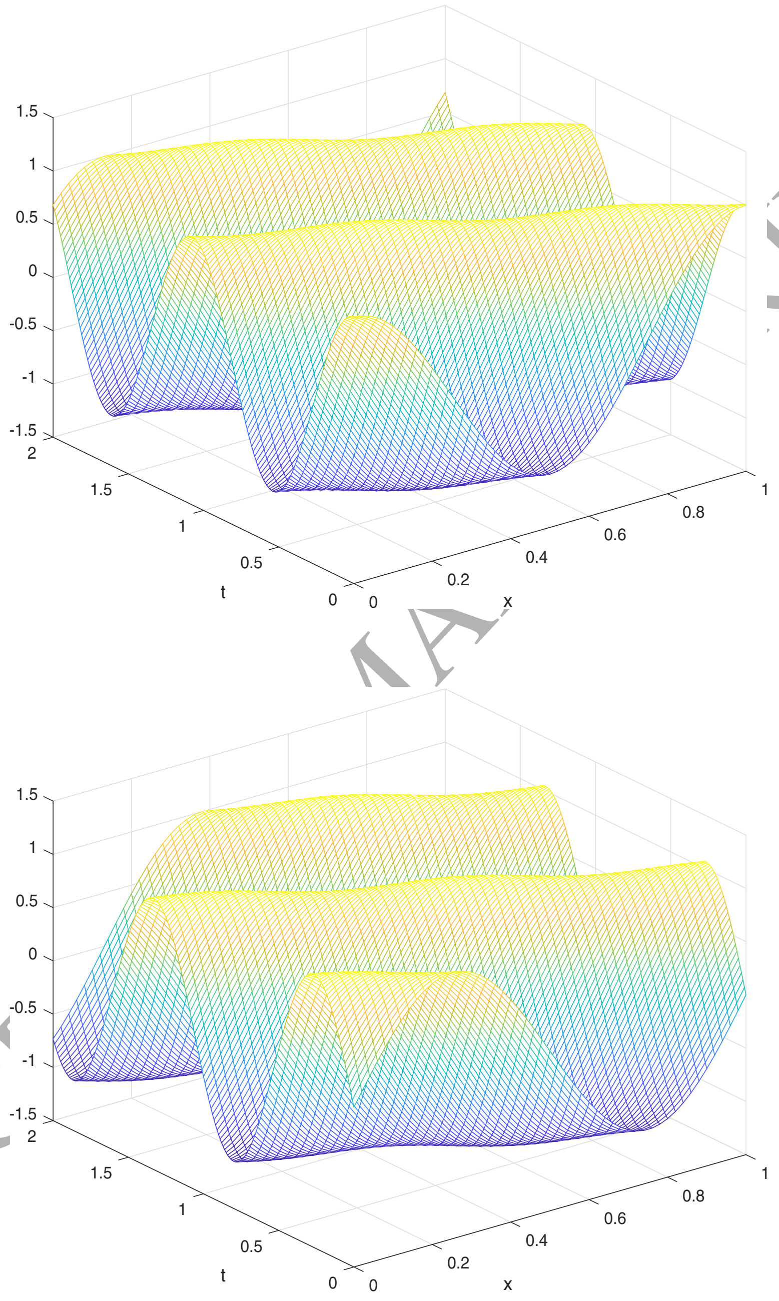

Figure 1: Real (upper plot) and imaginary (lower plot) parts of the solution of problem (5.1)–(5.2).

- the Hamiltonian error decreases with order $2k$, for the smaller values of k, whereas one obtains a practical energy conservation, when using $k = 10$ and time-steps smaller that $h = 0.25$;
- the mean number of blended iterations decreases with the time-step (as is obviously expected) and, moreover, it is remarkably independent of k, for s fixed.

We observe that, for this problem, the solution error is approximately the same, for h and s fixed, independently of the value of k used. However, this could be no more the case, when one has particular solutions, which may vary, depending on the value of the Hamiltonian functional, as it has been clearly shown in [6] for the sine-Gordon equation. Unfortunately, we are not aware of similar situations for the considered equation (1.1). Nevertheless, the conservation of the Hamiltonian functional may be useful, as the next example shows.

Let us consider the following problem,

$$u_{tt} - u_{xx} + 2.2\mathrm{i}u_t + \left(|u|^{18} - 8|u|^6\right) u = 0, \qquad (x,t) \in [0,1] \times [0,T], \tag{5.5}$$

for which the function f (see (1.1)) is a polynomial of degree $\nu = 10$, with the initial conditions

$$u(x,0) = 0, \qquad u_t(x,0) = 9\sin^{10}(\pi x), \qquad x \in [0,1]. \tag{5.6}$$

The corresponding functional (1.3) is now given by

$$\mathcal{H}[u](t) = \frac{1}{2}\int_0^1 |u_t|^2 + |u_x|^2 + \frac{|u|^{20}}{10} - 2|u|^8 \mathrm{d}x, \tag{5.7}$$

whose constant value is now given by

$$\mathcal{H}[u] \equiv H_0 = \frac{81}{2}\int_0^1 \sin^{20}(\pi x) \approx 7.136. \tag{5.8}$$

The real and imaginary parts of the solution of problem (5.5)–(5.6) are plotted in Figure 2, for $T = 6$. Choosing $N = 50$ in (3.1) and $m = 201$ in (3.8) is now appropriate for obtaining an accurate numerical solution: as matter of fact, both the initial value of the Hamiltonian and the initial data turns out to be approximated to full machine accuracy. We now fix the time-step $h = 0.03$, thus performing 200 integration steps, by using the 4-th order methods HBVM(2,2) (i.e., the symplectic 2-stage Gauss method) and HBVM(20,2). The latter method, according to Theorem 4.1, is energy conserving. This is confirmed by the computed numerical solution, for which the maximum energy error is ≈ 0.41 for HBVM(2,2) and $1.33 \cdot 10^{-14}$ for HBVM(20,2). In the upper plots in Figure 3 we plot the errors in the real part of the computed solution, whereas in the lower plots are the errors in the imaginary part. Moreover, the plots on the left concern the HBVM(20,2) method, with maximum errors $\approx 5 \cdot 10^{-2}$ and $7 \cdot 10^{-2}$, respectively. Similarly, the plots on the right concern the HBVM(2,2) method, with maximum errors $\approx 7 \cdot 10^{-1}$ and $6 \cdot 10^{-1}$, respectively. One then concludes that the solution provided by the energy-conserving method is pretty more accurate (about one order of magnitude), w.r.t. the one provided by the symplectic 2-stage Gauss method. Consequently, at least in this case, conserving the energy seems to confer more reliability on the computed numerical solution.

6. Conclusions

In this paper we have considered the numerical solution of the nonlinear Schrödinger equation with wave operator equipped with periodic boundary conditions. The problem has been, at first, cast into Hamiltonian form, by means of a Fourier-Galerkin space semi-discretization. Energy-conserving Runge-Kutta methods in the HBVMs class have then been used for the time integration, while conserving the energy of the system. The efficient implementation of such methods has been also studied, showing that their computational complexity per step is *linear* in the dimension of the semi-discrete problem, and their effectiveness has been evaluated on a couple of test problems.

Table 2: problem (5.1)–(5.2), with $T = 2$, solved by unsing $N = 1, m = 3$ for th espace semi-discretization, and HBVM(k,1) in time with stepsize $h = 2^{-n}$.

$k = s = 1$					
n	err_u	rate	err_H	rate	it
1	1.979e+0	—	1.406e-02	—	10.0
2	1.072e+0	0.9	8.089e-03	0.8	9.1
3	3.711e-01	1.5	6.902e-03	0.2	7.1
4	9.936e-02	1.9	2.325e-03	1.6	6.3
5	2.521e-02	2.0	6.261e-04	1.9	6.0
6	6.324e-03	2.0	1.595e-04	2.0	5.0
7	1.583e-03	2.0	4.006e-05	2.0	5.0
8	3.957e-04	2.0	1.003e-05	2.0	4.0
9	9.894e-05	2.0	2.508e-06	2.0	4.0
10	2.474e-05	2.0	6.269e-07	2.0	4.0
$k = 2, s = 1$					
1	1.970e+0	—	2.809e-02	—	11.0
2	1.066e+0	0.9	7.653e-03	1.9	9.0
3	3.740e-01	1.5	8.458e-04	3.2	7.6
4	1.009e-01	1.9	6.162e-05	3.8	6.4
5	2.565e-02	2.0	4.000e-06	3.9	6.0
6	6.439e-03	2.0	2.524e-07	4.0	5.0
7	1.611e-03	2.0	1.581e-08	4.0	5.0
8	4.030e-04	2.0	7.913e-10	4.3	4.0
9	1.008e-04	2.0	2.132e-14	**	4.0
10	2.519e-05	2.0	2.132e-14	**	4.0
$k = 3, s = 1$					
1	1.972e+0	—	4.375e-04	—	10.0
2	1.067e+0	0.9	1.598e-04	1.5	9.0
3	3.744e-01	1.5	1.235e-05	3.7	7.8
4	1.009e-01	1.9	3.048e-07	5.3	6.4
5	2.565e-02	2.0	5.353e-09	5.8	6.0
6	6.439e-03	2.0	1.421e-14	**	5.0
7	1.611e-03	2.0	1.421e-14	**	5.0
8	4.030e-04	2.0	1.421e-14	**	4.0
9	1.008e-04	2.0	1.421e-14	**	4.0
10	2.519e-05	2.0	1.421e-14	**	4.0
$k = 10, s = 1$					
1	1.972e+0	—	2.202e-11	—	10.3
2	1.067e+0	0.9	7.105e-15	11.6	9.0
3	3.744e-01	1.5	7.105e-15	**	8.0
4	1.009e-01	1.9	7.105e-15	**	6.9
5	2.565e-02	2.0	7.105e-15	**	6.0
6	6.439e-03	2.0	7.105e-15	**	5.0
7	1.611e-03	2.0	7.105e-15	**	5.0
8	4.030e-04	2.0	7.105e-15	**	4.0
9	1.008e-04	2.0	1.421e-14	**	4.0
10	2.519e-05	2.0	1.421e-14	**	4.0

Table 3: problem (5.1)–(5.2), with $T = 2$, solved by unsing $N = 1, m = 3$ for th espace semi-discretization, and HBVM(k,2) in time with stepsize $h = 2^{-n}$.

$k = s = 2$					
n	err_u	rate	err_H	rate	it
1	6.417e-01	—	5.007e-03	—	19.3
2	6.317e-02	3.3	1.818e-03	1.5	16.6
3	4.307e-03	3.9	1.570e-04	3.5	13.2
4	2.760e-04	4.0	9.807e-06	4.0	11.0
5	1.736e-05	4.0	6.157e-07	4.0	9.1
6	1.087e-06	4.0	3.853e-08	4.0	8.0
7	6.794e-08	4.0	2.409e-09	4.0	7.0
8	4.247e-09	4.0	1.506e-10	4.0	6.0
$k = 3, s = 2$					
1	6.438e-01	—	4.284e-03	—	19.0
2	6.330e-02	3.3	1.871e-04	4.5	16.8
3	4.386e-03	3.9	4.260e-06	5.5	13.6
4	2.812e-04	4.0	6.826e-08	6.0	11.0
5	1.769e-05	4.0	1.073e-09	6.0	9.3
6	1.107e-06	4.0	1.421e-14	**	8.0
7	6.922e-08	4.0	1.421e-14	**	7.0
8	4.327e-09	4.0	7.105e-15	**	6.0
$k = 4, s = 2$					
1	6.427e-01	—	3.375e-04	—	19.3
2	6.345e-02	3.3	2.068e-05	4.0	16.8
3	4.388e-03	3.9	1.174e-07	7.5	13.6
4	2.812e-04	4.0	5.108e-10	7.8	11.0
5	1.769e-05	4.0	1.421e-14	**	9.3
6	1.107e-06	4.0	7.105e-15	**	8.0
7	6.922e-08	4.0	7.105e-15	**	7.0
8	4.327e-09	4.0	7.105e-15	**	6.0
$k = 10, s = 2$					
1	6.431e-01	—	1.958e-08	—	19.5
2	6.344e-02	3.3	1.421e-14	**	17.4
3	4.388e-03	3.9	7.105e-15	**	14.2
4	2.812e-04	4.0	7.105e-15	**	11.2
5	1.769e-05	4.0	7.105e-15	**	9.8
6	1.107e-06	4.0	1.421e-14	**	8.0
7	6.922e-08	4.0	7.105e-15	**	7.1
8	4.327e-09	4.0	1.421e-14	**	6.0

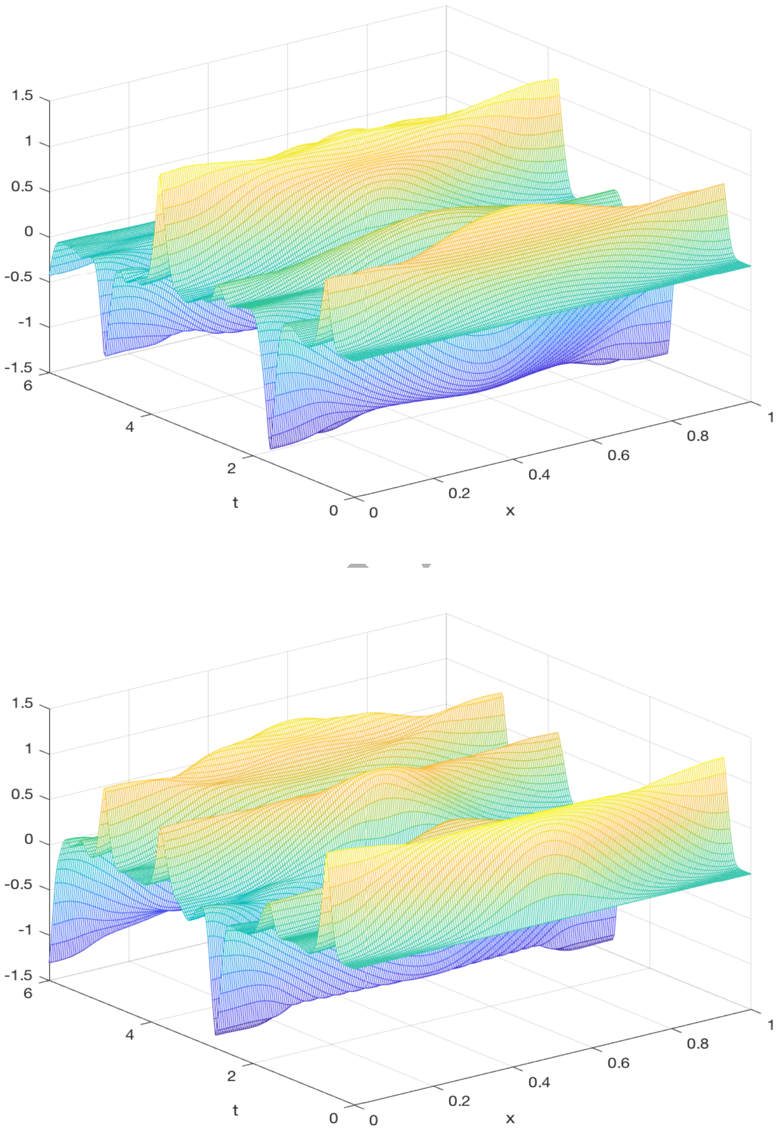

Figure 2: Real (upper plot) and imaginary (lower plot) parts of the solution of problem (5.5)–(5.6).

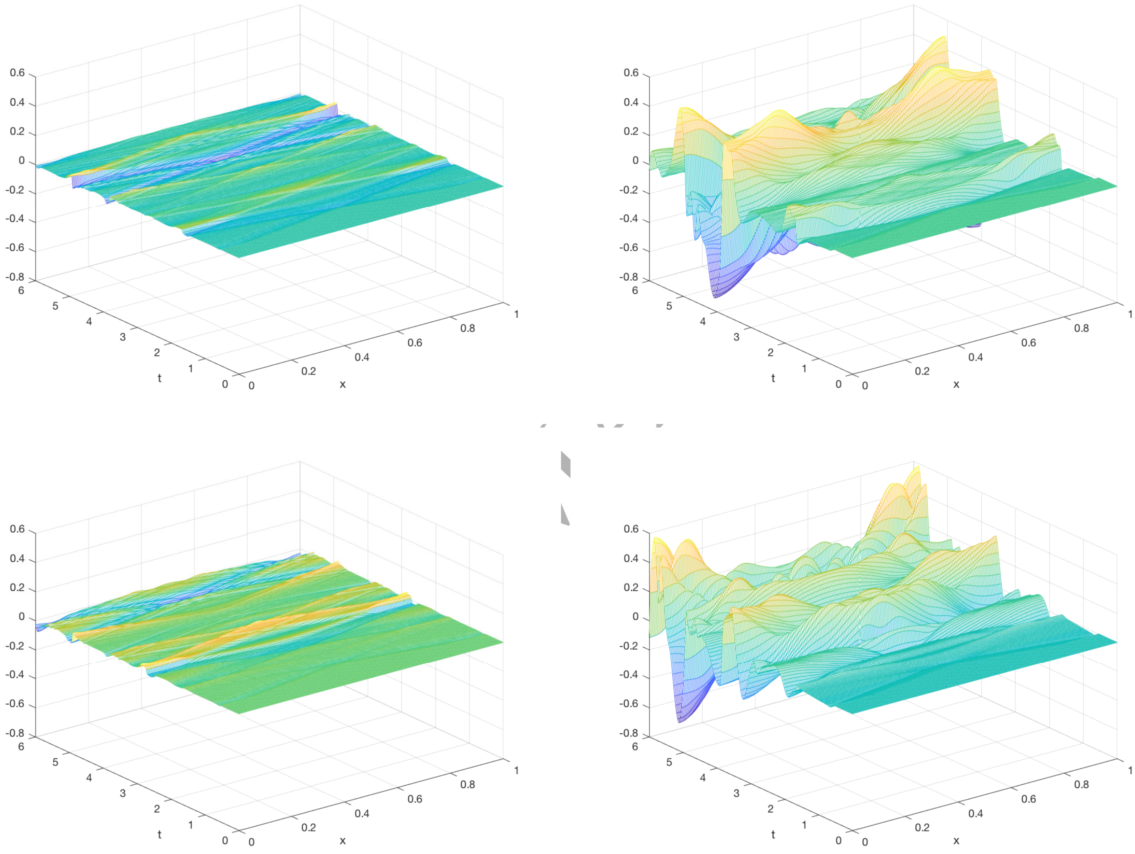

Figure 3: Computed errors in the real (upper plots) and imaginary parts (lower plots) of the solution of problem (5.5)–(5.6) by using HBVM(20,2) (left plots) and HBVM(2,2) (right plots) with time-step $h = 0.03$.

References

[1] W.Z. Bao, Y.Y. Cai. Uniform error estimates of finite difference methods for the nonlinear Schrödinger equation with wave operator. *SIAM J. Numer. Anal.* **20** (2014) 492–521.

[2] W. Bao, X. Dong, J. Xin. Comparisons between sine-Gordon and perturbed nonlinear Schrödinger equations for modeling light bullets beyond critical collapse. *Physica D* **239** (2010) 1120–1134.

[3] L. Barletti, L. Brugnano, G. Frasca Caccia, F. Iavernaro. Energy-conserving methods for the nonlinear Schrödinger equation. *Appl. Math.Comput.* **318** (2018) 3–18.

[4] L. Bergé, T. Colin. A singular perturbation problem for an envelope equation in plasma physics. *C. R. Acad. Sci. Paris Sér. I Math.* **320**, no. 1 (1995) 31–34.

[5] L. Brugnano, G. Frasca Caccia, F. Iavernaro. Efficient implementation of Gauss collocation and Hamiltonian boundary value methods. *Numer. Algorithms* **65** (2014) 633–650.

[6] L. Brugnano, G. Frasca Caccia, F. Iavernaro. Energy conservation issues in the numerical solution of the semilinear wave equation. *Appl. Math.Comput.* **270** (2015) 842–870.

[7] L. Brugnano, F. Iavernaro. *Line Integral Methods for Conservative Problems.* Chapman et Hall/CRC, Boca Raton, FL, 2016.

[8] L. Brugnano, F. Iavernaro, D. Trigiante. Hamiltonian boundary value methods (energy preserving discrete line integral methods). *J. Numer. Anal. Ind. Appl. Math.* **5**, No. 1-2 (2010), 17–37.

[9] L. Brugnano, F. Iavernaro, D. Trigiante. A note on the efficient implementation of Hamiltonian BVMs. *J. Comput. Appl. Math.* **236** (2011) 375–383.

[10] L. Brugnano, F. Iavernaro, D. Trigiante. Energy and quadratic invariants preserving integrators based upon Gauss collocation formulae. *SIAM J. Numer. Anal.* **50**, No. 6 (2012) 2897–2916.

[11] L. Brugnano, F. Iavernaro, D. Trigiante. A simple framework for the derivation and analysis of effective one-step methods for ODEs. *Appl. Math.Comput.* **218** (2012) 8475–8485.

[12] L. Brugnano, F. Iavernaro, D. Trigiante. Analisys of Hamiltonian boundary value methods (HBVMs): a class of energy-preserving Runge-Kutta methods for the numerical solution of polynomial Hamiltonian systems. *Commun. Nonlinear Sci. Numer. Simul.* **20** (2015) 650–667.

[13] L. Brugnano, C. Magherini. Blended implementation of block implicit methods for ODEs. *Appl. Numer. Math.* **42** (2002) 29–45.

[14] L. Brugnano, C. Magherini. Recent advances in linear analysis of convergence for splittings for solving ODE problems. *Appl. Numer. Math.* **59** (2009) 542–557.

[15] L. Brugnano, C. Magherini. The BiM code for the numerical solution of ODEs. *J. Comput. Appl. Math.* **164-165** (2004) 145–158.

[16] L. Brugnano, Y. Sun. Multiple invariants conserving Runge-Kutta type methods for Hamiltonian problems. *Numer. Algorithms* **65** (2014) 611–632.

[17] C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang. *Spectral Methods in Fluid Dynamics*, Springer-Verlag, New York, 1988.

[18] G. Dahlquist, Å. Bijörk. *Numerical Methods in Scientific Computing, vol. 1.* SIAM, Philadelphia, 2008.

[19] L. Guo, Y. Xu. Energy conserving local discontinuous Galerkin methods for the nonlinear Schrödinger equation with wave operator. *J. Sci. Comput.* **65** (2015) 622–647.

[20] H. Hu, Y. Chen. A conservative difference scheme for two-dimensional nonlinear Schrödinger equation with wave operator. *Numer. Methods Partial Differential Equations* **32**, no. 3 (2016) 862–876.

[21] A. Komech, B. Vainberg. On asymptotic stability of stationary solutions to nonlinear wave and Klein-Gordon equations. *Arch. Rational Mech. Anal.* **134**, no. 3 (1996) 227–248.

[22] A. Kurganov, J. Rauch, et al. The order of accuracy of quadrature formulae for periodic functions. In: A. Bove, et al. (Eds.), *Advances in Phase Space Analysis of Partial Differential Equations*, Birkhäuser, Boston, 2009.

[23] X. Li, L. Zhang, S. Wang. A compact finite difference scheme for the nonlinear Schrödinger equation with wave operator. *Appl. Math. Comput.* **219** (2012) 3187–3197.

[24] X. Li, L. Zhang, T. Zhang. A new numerical scheme for the nonlinear Schrödinger equation with wave operator. *J. Appl. Math. Comput.* **54** (2017) 109–125.

[25] S. Machihara, K. Nakanishi, T. Ozawa. Nonrelativistic limit in the energy space for nonlinear Klein-Gordon equations. *Math. Ann.* **322** (2002) 603–621.

[26] B. Najman. The nonrelativistic limit of the nonlinear Klein-Gordon equation. *Nonlinear Anal.* **15**, no. 3 (1990) 217–228.

[27] J.M. Sanz-Serna, M.P. Calvo. *Numerical Hamiltonian problems.* Chapman & Hall, London, 1994.

[28] A.Y. Schoene. On the nonrelativistic limits of the Klein-Gordon and Dirac equations. *J. Math. Anal. Appl.* **71**, no. 1 (1979) 36–47.

[29] S.-W. Vong, Q.-J. Meng, S.-L. Lei. On a discrete-time collocation method for the nonlinear Schrödinger equation with wave operator. *Numer. Methods Partial Differential Equations* **29**, no. 2 (2013) 693–705.

[30] J. Wang. Multisymplectic Fourier pseudospectral method for the nonlinear Schrödinger equation with wave operator. *J. Comput. Math.* **25**, no. 1 (2007) 31–48.

[31] L. Wang, L. Kong, L. Zhang, W. Zhou, X. Zheng. Multi-symplectic preserving integrator for the Schrödinger equation with wave operator. *Appl. Math. Model.* **39** (2015) 6817–6829.

[32] S. Wang, L. Zhang, R. Fan. Discrete-time orthogonal spline collocation methods for the nonlinear Schrödinger equation with wave operator. *J. Comput. Appl. Math.* **235** (2011) 1993–2005.

[33] T. Wang, L. Zhang. Analysis of some new conservative schemes for nonlinear Schrödinger equation with wave operator. *Appl. Math. Comput.* **182** (2006) 1780–1794.

[34] J.X. Xin. Modeling light bullets with the two-dimensional sine-Gordon equation. *Physica D 135* (2000) 345–368.

[35] `https://archimede.dm.uniba.it/~testset/testsetivpsolvers/`