



High Order Conservative Schemes for the Generalized Benjamin–Ono Equation on the Unbounded Domain

Kai Yang^{1,2}

Received: 27 September 2021 / Revised: 20 March 2023 / Accepted: 3 May 2023 /
Published online: 15 June 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

We propose a novel class of arbitrarily high-order mass or energy conservative numerical schemes for the generalized Benjamin–Ono equation on the whole real line. The spatial discretization is achieved by the pseudo-spectral method with the rational basis functions. By reformulating the spatial discretized system into the different equivalent forms, either the spatial semi-discretized mass or energy can be preserved exactly under the continuous time flow. Combined with the symplectic Runge–Kutta, with or without the scalar auxiliary variable reformulation, the fully discrete (space-time discrete sense) energy or mass conservative schemes can be constructed with arbitrarily high-order temporal accuracy, respectively. Our numerical results show the conservation of the proposed schemes, and also the superior accuracy and stability to the non-conservative (Leap-frog) scheme.

Keywords gBO · Rational basis functions · High order conservative schemes · Hamiltonian system · Unbounded domain

1 Introduction

This paper considers the numerical methods for solving the generalized Benjamin–Ono (gBO) equation

$$\begin{cases} u_t = -(-\mathcal{H}u_x + \frac{1}{m}u^m)_x, & x \in \mathbb{R}, t > 0, m \in \mathbb{Z}^+, \\ u(x, 0) = u_0, \end{cases} \quad (1.1)$$

where $u \in L^1(\mathbb{R})$ is the velocity potential, and the Hilbert transform \mathcal{H} is defined by

$$\mathcal{H}f(x) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{f(y)}{x-y} dy, \quad (1.2)$$

✉ Kai Yang
yangkai10@cqu.edu.cn

¹ College of Mathematics and Statistics, Chongqing University, Chongqing, China

² Key Laboratory of Nonlinear Analysis and its Applications (Chongqing University), Ministry of Education, Chongqing, China

where p.v. stands for the principle value of the indefinite integral, or equivalently, $\widehat{\mathcal{H}f}(\xi) = -i\text{sgn}(\xi)\widehat{f}(\xi)$ on the Fourier frequency side. When $m = 2$, it is the well-known Benjamin–Ono (BO) equation

$$u_t - \mathcal{H}u_{xx} + u_x u = 0, \tag{1.3}$$

derived by Benjamin [5] in 1967 and Ono [54] in 1975. This Eq. (1.3) models the one-dimensional waves in deep water. The BO equation is closely related to the Korteweg-de Vries (KdV) equation, where the Hilbert transform term $\mathcal{H}u_{xx}$ is replaced by u_{xxx} . The KdV equation models the one-dimensional shallow water waves. Both equations, BO and KdV, are completely integrable, and the Lax pair can be constructed as described in e.g., [2, 3, 27, 34, 52]. Without loss of generality, we can consider other nonlinearities. One example is to consider the power nonlinearities as of in Eq. (1.1). They are relevant in various other models of water waves, e.g., see [1, 8, 9, 22]. When $m = 3$, the Eq. (1.1) is typically referred to as the modified Benjamin–Ono (mBO) equation. When $m \geq 3$, the Eq. (1.1) is typically referred to as the generalized Benjamin–Ono (gBO) equation. In general, the gBO equation (1.1) conserves the following three quantities

$$I[u(t)] \stackrel{\text{def}}{=} \int u(x, t) dx = I[u_0]; \tag{1.4}$$

$$M[u(t)] \stackrel{\text{def}}{=} \int [u(x, t)]^2 dx = M[u_0]; \tag{1.5}$$

$$E[u(t)] \stackrel{\text{def}}{=} \int \left[\frac{1}{2} \left((\mathcal{H}\partial_x)^{\frac{1}{2}} u(x, t) \right)^2 - \frac{1}{m(m+1)} (u(x, t))^{m+1} \right] dx = E[u_0]. \tag{1.6}$$

The first one is called the momentum, or the conservation of the first integral, or the hyperbolic conservation law, and the last two are often called mass and energy (Hamiltonian), respectively.

Besides its physical applications, the gBO equation attracts great interest in investigating from the mathematical point of view, since it is a good example to study the fractional partial differential equations. The well-posedness theory for the Cauchy problem has been discussed initially in [37, 63]. Further improvements on the well-posedness questions were done in [16, 17, 35, 41, 50, 51, 67, 70]. The gBO equation possesses the scaling invariance property, i.e., suppose $u(x, t)$ is the solution to (1.1), then, $u_\lambda = \lambda^{\frac{2}{m-1}} u(\lambda x, \lambda^2 t)$ is also a solution to (1.1) for some constant λ . Moreover, when considering the homogeneous Sobolev norm $\|u\|_{\dot{H}^{s_c}} = \|u_\lambda\|_{\dot{H}^{s_c}}$ (where $\|u\|_{\dot{H}^{s_c}} = \| |\xi|^{s_c} \widehat{u}(\xi) \|_{L^2}$), we have the relation $s_c = \frac{1}{2} - \frac{1}{m-1}$. When $s_c < 0$ ($m < 3$), it refers to the L^2 -subcritical case; when $s_c = 0$ ($m = 3$), it refers to the L^2 -critical case; and when $s_c > 0$ ($m > 3$), it refers to the L^2 -supercritical case. The soliton resolution conjecture are one of the most interesting topics in the L^2 -subcritical case; and for the L^2 -critical and L^2 -supercritical cases, there may exist blow-up solutions. This was numerically observed in [10] and our recent paper [59]. Besides the blow-up solutions, there are still many open questions, such as the soliton stability and the dispersion limit. These kind of questions have been studied both numerically and analytically. Compared with the (generalized) KdV equation (e.g., [11, 30, 31], etc.), however, the gBO equation is less well studied (e.g., [49, 56] and review [62]). Therefore, a stable, efficient and accurate numerical algorithm would be desired the future study.

Numerical investigations on the BO equation have been started some time ago. Related articles can be found in [10, 24, 68] for the domain truncation approach; [12, 13, 71] for the computation of the Hilbert transform on \mathbb{R} ; [15, 32] for the pseudo-spectral method with the rational basis functions; and [14] for a comparison between the domain truncation and the

pseudo-spectral method on \mathbb{R} . Despite some years of investigations, there are still far less studies about numerical methods for the gBO equations than the gKdV equations. To our best knowledge, there are no results concerning the conservative schemes for the gBO equation on the whole real line \mathbb{R} so far. On the other hand, the conservative schemes are always preferable in simulating the PDE's with conserved quantities, especially for studying the long time solution behavior, since it generally possesses good accuracy and stability. One possible reason is the numerical approximation of the Hilbert transform on \mathbb{R} , which is not as well studied as on a finite domain. However, if considering the conventional domain truncation spatial discretization strategy (e.g., the finite difference or Fourier spectral methods), the Hilbert transform usually leads to a slow decaying function, and consequently, to a relatively large domain truncation error.

The purpose of this paper is to construct the conservative schemes for the gBO equation (1.1) on the whole real line \mathbb{R} , with arbitrarily high order accuracy in time. The spatial discretization is achieved by the rational basis functions with the pseudo-spectral approach from [32]. We first notice that the pseudo-spectral discretization from the rational basis functions will result in the operators with the "SBP" property discussed in [57]. Next, we prove that by reformulating into the different forms, and applying the Hermitian or anti-Hermitian properties of the resulting spatial semi-discretized system, either the spatial semi-discretized mass or energy will be preserved. For the temporal discretization, the Crank–Nicolson method with the conventional reformulation of the nonlinear potential term (e.g., see [25] for the nonlinear Schrödinger (NLS) equation case) will lead to the conservation of the energy in the discrete time flow. Furthermore, the high order mass and energy conservative schemes can be constructed from the scalar auxiliary variable (SAV) approach (see [21, 65, 66] for applications to dispersive PDEs). By using the symplectic Runge–Kutta (SRK) method, the three invariant quantities (1.4)–(1.6) [with proper modifications for energy (1.6)] will be preserved exactly in the discrete time flow. However, due to the limitation of the spatial discretization, we can only conserve either the discrete mass or the discrete energy in the space-time fully discrete sense. In fact, this strategy is universal. By a similar space-time discretization, it is easy to construct the conservative schemes for the gKdV equations and the structure-preserving schemes (the discrete mass and energy are preserved exactly at the same time) for the NLS equations, as well as their high dimensional generalization by applying the tensor product. This will be useful in studying the long time behavior of the solutions for those equations, as well as the slow decaying solutions, since the traditional domain truncation strategy (e.g., [29, 45, 46, 73]) requires large computational domain, and consequently, it results in large number of nodes in spatial discretization.

This paper is organized as follows. In Sect. 2, we introduce the pseudo-spectral spatial discretization strategy from the rational basis functions. Then, we define the discrete inner product with respect to the collocation points from such rational basis functions. Finally, we give the mass-conservative or the energy-conservative spatial semi-discretized form of the gBO equation (1.1). In Sect. 3, we first introduce the Crank–Nicolson types of temporal discretization. We show that the Crank–Nicolson method with its conventional modification on the nonlinear term, will preserve the energy exactly in the discrete time flow. Combining with the previous results in Sect. 2, we give two fully discretized schemes for the gBO equation, which conserve either the discrete mass or the discrete energy. Next, we consider the high-order conservative schemes achieved by the symplectic Runge–Kutta method with the SAV reformulation. From the classical argument on the symplectic Runge–Kutta method (e.g. [20, 60]), we show that the reformulated system preserves the quantities (1.4)–(1.6) exactly in the discrete time flow. Again, combined with the spatial discretization results in Sect. 2, we give two fully discretized schemes with high order temporal accuracy—one

conserves the discrete mass and the other conserves the discrete energy. In Sect. 4, we illustrate the numerical examples. As a comparison, we also show the numerical results obtained from the non-conservative semi-implicit Leap-Frog scheme. Our numerical results show that the proposed schemes preserve the designate quantities based on the type of conservative scheme we choose. Compared with the non-conservative scheme, these conservative schemes also possess better accuracy in most cases, especially in simulating the soliton type solutions, which is of great physical interest. Additionally, the error from the temporal discretization decreases on the order as expected (second order for the IRK2 and Leap-Frog schemes, and fourth order for the IRK4 schemes). These results show the validity and efficiency of the numerical methods proposed.

2 Spatial Discretization

In this section, we describe the rational basis functions in \mathbb{R} used for the spatial discretization. The review of basis functions can be found in [19, 71]. One advantage of this discretization is that it can easily represent the Hilbert transform. Then, we define the discrete inner product corresponding to the collocation points from the rational basis functions. Finally, we introduce two types of the spatial discretization for the gBO equation (1.1): one is mass-conservative and the other one is energy-conservative.

2.1 Rational Basis Functions

Consider the rational basis functions on the whole real line \mathbb{R} , which comes from the Fourier transform of the Laguerre functions (functions of parabolic cylinder), and behaves as x^{-1} at infinity, i.e.,

$$u(x, t) = \sum_{k=-\infty}^{\infty} \hat{u}_k(t) \rho_k(x), \quad \rho_k(x) = \frac{(\alpha + ix)^k}{(\alpha - ix)^{k+1}}, \tag{2.1}$$

where i is the complex number, and α is a mapping parameter that we will describe later. In [19], it is shown that $\{\rho_k(x)\}_{k=-\infty}^{\infty}$ form a complete orthogonal basis in $L^2(-\infty, \infty)$ with the following orthogonality

$$\int_{-\infty}^{\infty} \rho_j(x) \overline{\rho_k(x)} dx = \begin{cases} \pi/\alpha, & j = k \\ 0, & j \neq k. \end{cases} := \frac{\pi}{\alpha} \delta_{j,k}. \tag{2.2}$$

Therefore, we have

$$\hat{u}_k(t) = \frac{\alpha}{\pi} \int_{-\infty}^{\infty} u(x, t) \rho_k(x) dx.$$

From the rational expansion (2.1), the Hilbert transform can be easily calculated [71] by

$$\mathcal{H}(u(t, x)) = \sum_{k=-\infty}^{\infty} -i \hat{u}_k(t) \operatorname{sgn}(k) \rho_k(x), \tag{2.3}$$

with $\text{sgn}(k) = 1$ when $k = 0$. Meanwhile, the derivatives of $u(x, t)$ can be computed by using the relation $\rho_k + \rho_{k-1} = \frac{2\alpha(\alpha+ix)^{n-1}}{(\alpha-ix)^{n+1}}$, and consequently,

$$u_x(x, t) = \sum_{k=-\infty}^{\infty} \frac{i}{2\alpha} [k\hat{u}_{k-1} + (2k + 1)\hat{u}_k + (k + 1)\hat{u}_{k+1}]\rho_k(x), \tag{2.4}$$

Similar to find the higher order derivatives for Chebyshev or Legendre basis approximation in [64, Chapter 3], the second (or higher) order derivative of $u(x, t)$ can be computed iteratively as follows:

$$u_{xx}(x, t) = \sum_{k=-\infty}^{\infty} -\frac{1}{4\alpha^2} [k(k - 1)\hat{u}_{k-2} + 4k^2\hat{u}_{k-1} + (6k^2 + 6k + 2)\hat{u}_k + 4(k + 1)^2\hat{u}_{k+1} + (k + 2)(k + 1)\hat{u}_{k+2}]\rho_k(x). \tag{2.5}$$

In numerical computations, a truncation of N -term interpolation function $I_N u$ are used to approximate the function $u(x)$ using the basis $\{\rho_k(x)\}_{-N/2}^{N/2-1}$ to interpolate from the node values at $u(x_{-N/2}), \dots, u(x_{N/2-1})$, i.e.,

$$u(x, t) \approx I_N u := \hat{\mathbf{u}}^T \vec{\rho} := \sum_{k=-N/2}^{N/2-1} \hat{u}_k(t)\rho_k(x),$$

where $\hat{\mathbf{u}} = (\hat{u}_{-N/2}, \hat{u}_{-N/2+1}, \dots, \hat{u}_{N/2-1})^T$ is the vector of the truncated coefficients, and $\vec{\rho} = (\rho_{-N/2}(x), \dots, \rho_{N/2-1}(x))^T$ is the vector function of $\rho_k(x)$. This leads to the sparse matrix forms

$$u_x \approx [\mathbf{S}_1 \hat{\mathbf{u}}]^T \vec{\rho}, \quad u_{xx} \approx [\mathbf{S}_2 \hat{\mathbf{u}}]^T \vec{\rho}, \quad \mathcal{H}u \approx [\mathbf{H}\hat{\mathbf{u}}]^T \vec{\rho}, \tag{2.6}$$

where \mathbf{S}_1 and \mathbf{S}_2 are given in (2.4) and (2.5) via the coefficients of $\{\hat{u}_k\}$, and

$$\mathbf{H} = -i \text{diag}(\text{sgn}(-N/2 + 0.5), \dots, \text{sgn}(N/2 - 0.5))$$

is the diagonal matrix representing the approximation of the Hilbert transform in (2.3).

From (2.4) to (2.6), it is easy to see that the matrices \mathbf{S}_1 and \mathbf{H} are anti-Hermitian, and the matrix \mathbf{S}_2 is real and symmetric. The anti-Hermitian and Hermitian properties for \mathbf{S}_1 and \mathbf{S}_2 are crucial in constructing the conservative schemes.

Now, consider the change of variable

$$x = \alpha \tan \frac{\theta}{2}, \quad \text{or equivalently,} \quad e^{i\theta} = \frac{\alpha + ix}{\alpha - ix}, \quad -\pi \leq \theta \leq \pi,$$

and a spatial discretization $x_j = \alpha \tan \frac{\theta_j}{2}$, $\theta_j = jh$, $h = 2\pi/N$, $j = -N/2, \dots, N/2 - 1$, where the α is the mapping parameter indicating that $N/2$ collocation points are located in the interval $[-\alpha, \alpha]$. Notice that

$$u(x_j) = \sum_{k=-N/2}^{N/2-1} \hat{u}_k \rho_k(x_j) \Rightarrow u(x_j)(\alpha - ix_j) = \sum_{k=-N/2}^{N/2-1} \hat{u}_k e^{ik\theta_j}, \tag{2.7}$$

hence, the Fast Fourier transform (FFT) can be applied to obtain the coefficients \hat{u}_k . We note that the above discretization in space is not uniform in x , but uniform in θ , and the singularity

at $x_{-N/2} = -\infty$ can be removed by imposing the boundary condition $u(-\infty) = 0$, i.e., setting $u_{-N/2} = 0$.

We denote the matrix \mathbf{F} to be the standard Fast Fourier transform (FFT) matrix with $\{k\theta_j\}$, i.e.,

$$\mathbf{F}_{kj} = \frac{1}{N}e^{-ik\theta_j}, \quad \mathbf{F}_{jk}^{-1} = e^{ik\theta_j}, \quad -N/2 \leq j, k \leq N/2 - 1.$$

Note that instead of writing explicitly, the matrices \mathbf{F} and \mathbf{F}^{-1} can be computed by FFT, (see, e.g., [64, Chapter 2] and [69, Chapter 3]). Denote the diagonal matrix $\mathbf{P} = \text{diag}(\alpha - ix_{-N/2}, \dots, \alpha - ix_{N/2-1})$ to be the weight matrix, which comes from (2.7). The coefficients \hat{u}_k in the vector form can be represented by

$$\hat{\mathbf{u}} = \mathbf{F}\mathbf{P}\mathbf{u}, \tag{2.8}$$

where $\mathbf{u} = (u_{-N/2}, \dots, u_{N/2-1})^T$ and $u_j = u(x_j)$.

Now, we define the discrete inner product with respect to the rational basis function. Denote the inner product between the two functions $u(x)$ and $v(x)$ on \mathbb{R} by

$$\langle u(x), v(x) \rangle := \int_{\mathbb{R}} u(x)\bar{v}(x)dx.$$

Recall that the interpolation function $I_N u$ is the approximation of $u(x) \approx I_N u = \sum_{k=-N/2}^{N/2-1} \hat{u}_k \rho_k(x)$, then, the approximation of the inner product for functions u and v will be

$$\begin{aligned} \langle u, v \rangle &\approx \int_{\mathbb{R}} I_N u \overline{I_N v} dx = \frac{\pi}{\alpha} \sum_{k=-N/2}^{N/2-1} \hat{u}_k \bar{\hat{v}}_k \\ &= \frac{\pi}{\alpha} (\mathbf{F}\mathbf{P}\mathbf{v})^T (\mathbf{F}\mathbf{P}\mathbf{u}) = \frac{\pi}{\alpha} \bar{\mathbf{v}}^T \bar{\mathbf{P}}^T \mathbf{F}^T \mathbf{F}\mathbf{P}\mathbf{u} = \frac{\pi}{\alpha N} \bar{\mathbf{v}}^T \bar{\mathbf{P}}\mathbf{P}\mathbf{u}, \end{aligned} \tag{2.9}$$

from the orthogonal property (2.2), the relation (2.8), and $\bar{\mathbf{F}}^T = \frac{1}{N}\mathbf{F}^{-1}$ (e.g., see [64, Chapter 2]). Denote the diagonal matrix $\mathbf{W} = \bar{\mathbf{P}}\mathbf{P} = \text{diag}(\alpha^2 + x_{-N/2}^2, \dots, \alpha^2 + x_{N/2-1}^2)$ to be the product of the two diagonal matrices \mathbf{P} and $\bar{\mathbf{P}}$. According to (2.9), we can define the discrete inner product with respect to the collocation points $\{x_j\}$ from the rational basis function:

$$\langle \mathbf{u}, \mathbf{v} \rangle_h := \frac{\pi}{N\alpha} \bar{\mathbf{v}}^T \mathbf{W}\mathbf{u} = \frac{\pi}{N\alpha} \sum_{j=-N/2}^{N/2-1} w_j u_j \bar{v}_j, \tag{2.10}$$

where $w_j = \alpha^2 + x_j^2$ can be considered as the weights for the quadrature rule in (2.10).

Now, for simplicity, we define the first order and second order differential matrices $\mathbf{D}_{1,2} = \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_{1,2}\mathbf{F}\mathbf{P}$. Then, $u_x(x_j) \approx \mathbf{D}_1\mathbf{u}(j)$, the j th element of the vector $\mathbf{D}_1\mathbf{u}$, and similarly for its second order derivative $u_{xx}(x_j) \approx \mathbf{D}_2\mathbf{u}(j)$.

We notice that for $u, v \in \mathbb{V}_N$, where \mathbb{V}_N is the space spanned by $\{\rho_k(x)\}_{-N/2}^{N/2-1}$, the operator \mathbf{D}_1 satisfies the summation by part property, i.e.,

$$\int u_x v dx = \langle \mathbf{D}_1\mathbf{u}, \bar{\mathbf{v}} \rangle_h = -\langle \mathbf{u}, \mathbf{D}_1\bar{\mathbf{v}} \rangle_h = -\int uv_x dx. \tag{2.11}$$

Indeed, from (2.9), we have

$$\int u_x v dx = \langle \mathbf{D}_1\mathbf{u}, \bar{\mathbf{v}} \rangle_h = \frac{\pi}{\alpha N} \bar{\mathbf{v}}^T \mathbf{W}\mathbf{D}_1\mathbf{u} = \frac{\pi}{\alpha N} (\mathbf{F}^T \mathbf{S}_1^T (\mathbf{F}^{-1})^T \bar{\mathbf{P}}\mathbf{v})^T \mathbf{P}\mathbf{u}$$

$$\begin{aligned}
 &= \frac{\pi}{\alpha N} \overline{(\mathbf{P}^{-1} \bar{\mathbf{F}}^T \bar{\mathbf{S}}_1^T (\bar{\mathbf{F}}^{-1})^T \mathbf{P} \bar{\mathbf{v}})^T} \bar{\mathbf{P}} \mathbf{P} \mathbf{u} = -\frac{\pi}{\alpha N} \overline{(\mathbf{F} \mathbf{P} \mathbf{D}_1 \bar{\mathbf{v}})^T} \mathbf{F} \mathbf{P} \mathbf{u} \\
 &= -\langle \mathbf{u}, \mathbf{D}_1 \bar{\mathbf{v}} \rangle_h = -\int uv_x dx.
 \end{aligned}$$

Unfortunately, the operator \mathbf{D}_1 does not satisfy the chain rule or the integration property. Indeed, for example, let $u \in \mathbb{V}_N$, we consider the power nonlinearity u^3 . For any test function $v \in \mathbb{V}_N$, we want $\int (u^3)_x \bar{v} dx = \int 3u^2 u_x \bar{v} dx$. However,

$$\int (u^3)_x \bar{v} dx \approx \int I_N((u^3)_x) \bar{v} dx = \langle \mathbf{D}_1 \mathbf{u}^3, \bar{\mathbf{v}} \rangle_h \neq \langle \text{diag}(3\mathbf{u}^2) \mathbf{D}_1 \mathbf{u}, \bar{\mathbf{v}} \rangle_h = \int I_N(3u^2 u_x) \bar{v} dx.$$

Similarly, $\langle \mathbf{D}_1 \mathbf{u}, \mathbf{1} \rangle_h \neq 0$ (which implies $\int u_x dx$), since 1 is not in the function space \mathbb{V}_N unless $N \rightarrow \infty$. This implies unlike the Fourier spectral method, the rational basis function spectral method here does not preserve the first integral.

2.2 Conservative Spatial Discretization

To discuss the conservative spatial discretizations, we first define the spatial semi-discretized L^1 -type integral, mass and energy from (1.4)–(1.6). Let $\mathbf{1} = (1, 1, \dots, 1)^T$ be the $N \times 1$ vector. For simplicity, we also denote by $\mathbf{u}^m = (u_{-N/2}^m, \dots, u_{N/2-1}^m)^T$ to be the pointwise power of the vector \mathbf{u} . Then, the spatial semi-discretized L^1 -type integral, mass and energy are defined as follows

$$I_h = \langle \mathbf{u}, \mathbf{1} \rangle_h; \tag{2.12}$$

$$M_h = \langle \mathbf{u}, \mathbf{u} \rangle_h; \tag{2.13}$$

$$E_h = \frac{1}{2} \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}, \mathbf{u} \rangle_h - \frac{1}{m(m+1)} \langle \mathbf{u}^m, \mathbf{u} \rangle_h. \tag{2.14}$$

It is easy to see that if $\mathbf{u} \in \mathbb{R}^N$, then $\frac{d}{dt} I_h = \langle \mathbf{u}_t, \mathbf{1} \rangle_h$, and $\frac{d}{dt} M_h = 2 \langle \mathbf{u}_t, \mathbf{u} \rangle_h$ from (2.10). We also note that

$$\begin{aligned}
 \frac{d}{dt} E_h &= \frac{1}{2} \left(\langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}_t, \mathbf{u} \rangle_h + \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}, \mathbf{u}_t \rangle_h \right) \\
 &\quad - \frac{1}{m(m+1)} \left(\langle (\mathbf{u}^m)_t, \mathbf{u} \rangle_h + \langle \mathbf{u}^m, \mathbf{u}_t \rangle_h \right) \\
 &= \text{Re} \left(\langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}, \mathbf{u}_t \rangle_h \right) - \frac{1}{m} \langle \mathbf{u}^m, \mathbf{u}_t \rangle_h.
 \end{aligned} \tag{2.15}$$

The last equality is obtained from the following argument. We first noticed that the matrix $\mathbf{H} \mathbf{S}_1$ is real and symmetric, and \mathbf{P} is diagonal. Consequently, using (2.10) and $\bar{\mathbf{F}}^T = \frac{1}{N} \mathbf{F}^{-1}$, we obtain

$$\begin{aligned}
 \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}, \mathbf{u}_t \rangle_h &= \frac{\pi}{N\alpha} \bar{\mathbf{u}}_t^T \mathbf{P} \bar{\mathbf{P}} \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u} \\
 &= \frac{\pi}{N\alpha} \mathbf{u}^T \mathbf{P}^T \mathbf{F}^T \mathbf{H} \mathbf{S}_1 (\mathbf{F}^{-1})^T \bar{\mathbf{P}} \bar{\mathbf{u}}_t = \frac{\pi}{N\alpha} \mathbf{u}^T \mathbf{P} \bar{\mathbf{P}} \bar{\mathbf{P}}^{-1} \bar{\mathbf{F}}^{-1} \overline{\mathbf{H} \mathbf{S}_1 \bar{\mathbf{F}}} \bar{\mathbf{P}} \bar{\mathbf{u}}_t \\
 &= \frac{\pi}{N\alpha} \mathbf{u}^T \mathbf{P} \bar{\mathbf{P}} \bar{\mathbf{P}}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}_t = \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}_t, \mathbf{u} \rangle_h.
 \end{aligned}$$

Similarly, we have

$$\langle (\mathbf{u}^m)_t, \mathbf{u} \rangle_h = m \langle \mathbf{u}^{m-1} \mathbf{u}_t, \mathbf{u} \rangle_h = m \langle \mathbf{u}^m, \mathbf{u}_t \rangle_h.$$

Recall that the energy conservation (1.6) is obtained by taking the inner product of $(-\mathcal{H}u_x + \frac{1}{m}u^m)$ on both sides in the gBO equation (1.1). On the other hand, the mass conservation (1.5) of the gBO equation (1.1) is obtained by first rewriting it in the form

$$u_t = \mathcal{H}u_x - \frac{1}{m+1} \left((u^m)_x + u^{m-1}u_x \right),$$

and then, taking the inner product with u on both sides. When the equation is discretized with our rational basis functions, we noticed that $u_x \approx \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}$. Thus, we can consider the matrix $\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}$ to be the approximation of the first order derivative operator ∂_x . Similarly, we have the approximation of $\mathcal{H}\partial_x, \partial_{xx}$ from $\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}$ and $\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_2\mathbf{F}\mathbf{P}$, respectively. Using this idea, we propose the following proposition for the spatial conservative discretizations.

Proposition 2.1 *The following spatial semi-discretized equation to the gBO equation (1.1)*

$$\mathbf{u}_t = -\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P} \left(-\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} + \frac{1}{m}\mathbf{u}^m \right) \tag{2.16}$$

conserves the spatial semi-discretized energy, i.e.,

$$\frac{d}{dt} E_h = 0. \tag{2.17}$$

On the other hand, the following spatial semi-discretized equation to the gBO equation (1.1)

$$\mathbf{u}_t = \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_2\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m+1} \left(\text{diag}(\mathbf{u}^{m-1})\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} + \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^m \right) \tag{2.18}$$

conserves the spatial semi-discretized mass, i.e.,

$$\frac{d}{dt} M_h = 0. \tag{2.19}$$

Proof Putting the Eq. (2.16) in (2.15) yields

$$\begin{aligned} \frac{d}{dt} E_h &= \text{Re} \left(\langle \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m, \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m) \rangle_h \right) \\ &= \frac{\pi}{\alpha N} \text{Re} \left[\overline{\left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m) \right)^T} \mathbf{W} \left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m \right) \right] \\ &= \frac{\pi}{\alpha N} \text{Re} \left[\overline{\left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m \right)^T} \overline{\left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P} \right)^T} \mathbf{W} \left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m \right) \right] \\ &= \frac{\pi}{\alpha N} \text{Re} \left[\overline{\left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m \right)^T} \left(\tilde{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P} \right)^T \left(\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} - \frac{1}{m}\mathbf{u}^m \right) \right] = 0, \end{aligned} \tag{2.20}$$

since $\mathbf{W} = \tilde{\mathbf{P}}\mathbf{P}$, and notice that the matrix $\frac{1}{N}\tilde{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}$ is anti-Hermitian, which possesses the property regarding to the quadratic form $\tilde{\mathbf{f}}^T\mathbf{D}\mathbf{f} = 0$ if \mathbf{D} is an anti-Hermitian matrix.

Similarly, we first note that $\mathbf{W}\text{diag}(\mathbf{u}^{m-1}) = \text{diag}(\mathbf{u}^{m-1})\mathbf{W}$, since both of them are diagonal matrices. Then,

$$\frac{d}{dt} M_h = 2\langle \mathbf{u}_t, \mathbf{u} \rangle = 2\langle \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_2\mathbf{F}\mathbf{P}\mathbf{u}, \mathbf{u} \rangle_h$$

$$\begin{aligned}
 & - \frac{2}{m+1} \langle (\text{diag}(\mathbf{u}^{m-1})\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} + \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^m), \mathbf{u} \rangle_h \\
 & = 0 - \frac{2\pi}{(m+1)\alpha N} \left((\mathbf{u}^m)^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} + \mathbf{u}^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^m \right) = 0. \tag{2.21}
 \end{aligned}$$

The identity holds because the first part $2\langle \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_2\mathbf{F}\mathbf{P}\mathbf{u}, \mathbf{u} \rangle_h = 0$ in (2.21) results from the quadratic form $\frac{2\pi}{\alpha N} \bar{\mathbf{u}}^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}$ according to the definition (2.10), and the matrices $\frac{1}{N} \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_2\mathbf{F}\mathbf{P}$ is anti-Hermitian. For the second part $(\mathbf{u}^m)^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} + \mathbf{u}^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^m = 0$ in (2.21), we first notice that the identity

$$\mathbf{f}^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{g} = N(\bar{\mathbf{F}}\bar{\mathbf{P}}\mathbf{f})^T \mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{g}$$

when \mathbf{f} and \mathbf{g} are real vectors. Then, taking $\mathbf{f} = \mathbf{u}^m$ and $\mathbf{g} = \mathbf{u}$ yields $(\mathbf{u}^m)^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u} + \mathbf{u}^T \bar{\mathbf{P}}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^m = 0$, since \mathbf{S}_1 is anti-Hermitian. □

3 Temporal and Full Discretization

In this section, we first discuss the temporal discretization, and then, the space-time full discretization of the gBO equation (1.1). We start with the most commonly used Crank–Nicolson-type scheme. After that, we consider the high order conservative schemes. This is achieved by the symplectic Runge–Kutta method, such as the Gauss–Legendre Runge–Kutta method. When considering the energy conservation, the scalar auxiliary variable (SAV) approach from [21, 65] will be incorporated. With this kind of approach, one can easily construct the conservative numerical scheme with arbitrarily high order accuracy in time.

3.1 Crank–Nicolson-Type Scheme

We first introduce the notations. Assume that our simulation is on the finite time interval $t \in [0, T]$. Define τ to be the time step and $t_n = n\tau$ to be the time at the n th time step. Denote $u^n \approx u(x, t_n)$ to be the semi-discretization in time. Denote $I^n = I[u^n]$, $M^n = M[u^n]$ and $E^n = E[u^n]$ to be the momentum, mass and energy from (1.4)–(1.6) at time $t = t_n$. For convenience, the half-time step is denoted as $u^{n+\frac{1}{2}} = \frac{1}{2}(u^n + u^{n+1})$ from the linear interpolation. We also denote the full discretization by $u_j^n \approx u(x_j, t_n)$, and the column vector $\mathbf{u}^n \approx u(\mathbf{x}, t_n)$. Now, we define the discrete first integral, mass and energy as follows:

$$I_h^n = \langle \mathbf{u}^n, \mathbf{1} \rangle_h; \tag{3.1}$$

$$M_h^n = \langle \mathbf{u}^n, \mathbf{u}^n \rangle_h; \tag{3.2}$$

$$E_h^n = \frac{1}{2} \langle \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^n, \mathbf{u}^n \rangle_h - \frac{1}{m(m+1)} \langle (\mathbf{u}^n)^m, \mathbf{u}^n \rangle_h. \tag{3.3}$$

We propose the following Crank–Nicolson-type *mass-conservative* scheme as follows:

Theorem 3.1 *The scheme*

$$\begin{aligned}
 \frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\tau} & = \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_2\mathbf{F}\mathbf{P}\mathbf{u}^{n+\frac{1}{2}} \\
 & - \frac{1}{m+1} \left(\text{diag} \left((\mathbf{u}^{n+\frac{1}{2}})^{m-1} \right) \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^{n+\frac{1}{2}} + \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}(\mathbf{u}^{n+\frac{1}{2}})^m \right) \tag{3.4}
 \end{aligned}$$

conserves the discrete mass (3.2) exactly in time, i.e.,

$$M_h^{n+1} = M_h^n.$$

Proof The proof is straightforward. Equipping the Eq. (3.4) with the discrete inner product (2.10) with the vector $\mathbf{u}^{n+\frac{1}{2}}$, and using the identity (2.21) in Proposition 2.1 yields the result. \square

For the Crank–Nicolson-type *energy-conservative* scheme, by modifying of the nonlinear term, we have the following theorem.

Theorem 3.2 *The scheme*

$$\begin{aligned} \frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\tau} &= \mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\left[\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^{n+\frac{1}{2}} \right. \\ &\quad \left. - \frac{1}{m(m+1)}\text{diag}\left(\frac{(\mathbf{u}^{n+1})^{m+1} - (\mathbf{u}^n)^{m+1}}{(\mathbf{u}^{n+1})^2 - (\mathbf{u}^n)^2}\right)\mathbf{u}^{n+\frac{1}{2}}\right] \end{aligned} \tag{3.5}$$

conserves the discrete energy (3.3) exactly in time, i.e.,

$$E_h^{n+1} = E_h^n.$$

Proof Following the same idea as in Theorem 3.1, we equip the Eq. (3.5) with the discrete inner product (2.10) with the vector

$$\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{H}\mathbf{S}_1\mathbf{F}\mathbf{P}\mathbf{u}^{n+\frac{1}{2}} - \frac{1}{m(m+1)}\text{diag}\left(\frac{(\mathbf{u}^{n+1})^{m+1} - (\mathbf{u}^n)^{m+1}}{(\mathbf{u}^{n+1})^2 - (\mathbf{u}^n)^2}\right)\mathbf{u}^{n+\frac{1}{2}}.$$

Then, using the identity (2.20) in Proposition 2.1 yields the result. \square

The construction for the conservative schemes in Theorems 3.1 and 3.2 are standard. If we only consider the semi-discretization in time, the scheme (3.4) is the midpoint rule, or the implicit 2nd order Runge–Kutta method (IRK2), which is also known as the symplectic (quadratic preserving) Runge–Kutta method, e.g., see [63], and thus, the quadratic quantity (mass), is conserved. We split the potential into the form

$$\frac{1}{m}(u^m)_x = \frac{1}{m+1}[u^{m-1}u_x + (u^m)_x]$$

for the purpose of creating the symmetry for the mass conservations in the spatial discretization, which we discussed in the previous section. For the energy conservation, we need to reformulate the potential part, which is widely used in literature, see e.g., [23, 43, 48] for the NLS case.

We next discuss the numerical schemes with higher order temporal accuracy. For simplicity and conciseness, we only consider the semi-discretization in time. The space-time full discretization results can be easily generalized together with the results from Sect. 2.

3.2 High Order Conservative Schemes

The high order temporal conservative schemes can be achieved by the symplectic Runge–Kutta (SRK) method. We first briefly review the RK method before showing our results. Consider the problem

$$u_t = f(u). \tag{3.6}$$

Table 1 Butcher’s the s -stage Gaussian–Legendre collocation Runge–Kutta methods with $s = 1, 2$

(A) IRK2		
$\frac{1}{2}$		$\frac{1}{2}$
		1
(B) IRK4		
$\frac{1}{2} - \frac{1}{6}\sqrt{3}$	$\frac{1}{4}$	$\frac{1}{4} - \frac{1}{6}\sqrt{3}$
$\frac{1}{2} + \frac{1}{6}\sqrt{3}$	$\frac{1}{4} + \frac{1}{6}\sqrt{3}$	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$

From the time $t = t_n$ to $t = t_{n+1}$, let $b_i, a_{ij}(i, j = 1, \dots, s)$ be real numbers, and $c_i = \sum_{j=1}^s a_{ij}$ be the collocation points. Denote the intermediate values U_i to be the solution satisfying (3.6) at the intermediate time $t^i = t_n + \tau c_i$. Then, the intermediate values U_i ’s are calculated by

$$U_i = u^n + \tau \sum_{j=1}^s a_{ij} f_j, \tag{3.7}$$

where $f_i = f(U_i)$. The solution u^{n+1} is updated by

$$u^{n+1} = u^n + \tau \sum_{j=1}^s b_j f_j. \tag{3.8}$$

We usually write the coefficients $\mathbf{A} = (a_{ij})$, $\mathbf{b} = (b_1, b_2, \dots, b_s)$ and $\mathbf{c} = (c_1, c_2, \dots, c_s)^T$ in the Butcher’s Tableaus ([18]):

$$\begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b} \end{array}$$

For example, we list two commonly used Runge–Kutta methods in the Butcher’s Tableaus in Table 1. They are the s -stage Runge–Kutta methods with $s = 1, 2$, respectively. These methods are coming from the Gaussian–Legendre quadrature, known as the IRK2 and IRK4 methods, since the temporal accuracy is on the order of 2 and 4, respectively. We use these methods in our numerical simulations in the next section. There are many other types of Runge–Kutta methods as well, we refer the interested reader to [6, 20, 28, 60, 61].

We prove the following theorem for the mass-conservative scheme.

Theorem 3.3 *The s -stage symplectic (quadratic preserving) Runge–Kutta method, which satisfies*

$$b_i a_{ij} + b_j a_{ji} = b_i b_j, \quad \text{for } i, j = 1, \dots, s, \tag{3.9}$$

conserves the discrete mass exactly in time for the spatial discretized gBO equation (2.18), i.e.,

$$M_h^{n+1} = M_h^n.$$

Proof The proof is standard. From the standard RK theory (e.g., [20, 60]), we can show that the temporal semi-discretized scheme conserves the mass exactly in the discrete time flow.

Indeed, the RK theory shows that

$$M^{n+1} - M^n = 2\tau \sum_{i=1}^s b_i \langle U_i, f(U_i) \rangle + \tau^2 \sum_{i,j=1}^s (b_i a_{ij} + b_j a_{ji} - b_i b_j) \langle f(U_i), f(U_j) \rangle = 0,$$

since $\langle U_i, f(U_i) \rangle = 0$ by putting $f(U)$ in the form of (1.1).

When considering the space-time full discretization, we have

$$M_h^{n+1} - M_h^n = 2\tau \sum_{i=1}^s b_i \langle \mathbf{U}_i, f(\mathbf{U}_i) \rangle_h + \tau^2 \sum_{i,j=1}^s (b_i a_{ij} + b_j a_{ji} - b_i b_j) \langle f(\mathbf{U}_i), f(\mathbf{U}_j) \rangle_h = 0,$$

where the vector \mathbf{U}_i is the discretized version of the intermediate value U_i for $i = 1, \dots, s$, and $\langle \mathbf{U}_i, f(\mathbf{U}_i) \rangle_h = 0$ by using the relation (2.21). □

The symplectic Runge–Kutta method cannot preserve the discrete energy. In order to construct the energy-preserving scheme, we need to reformulate the potential term in the same idea as in (3.5). This is achieved by using the scalar auxillary approach from [21, 65]. We reformulate the Eq. (1.1) into an equivalent system as follows:

$$\begin{cases} u_t = - \left(-\mathcal{H}u_x + \frac{1}{m} \frac{u^m v}{\sqrt{(u^m, u) + C_0}} \right)_x, \\ v_t = \frac{m+1}{2\sqrt{(u^m, u) + C_0}} \langle u^m, u_t \rangle, \end{cases} \tag{3.10}$$

with the initial condition

$$u(x, 0) = u_0, \quad v_0 = \sqrt{\langle u_0^m, u_0 \rangle + C_0}.$$

Then, the energy to the system (1.1) is modified into the equivalent form

$$E[u(t), v(t)] := \frac{1}{2} \langle \mathcal{H}u_x, u \rangle - \frac{1}{m(m+1)} (v^2 - C_0) \equiv E[u_0, v_0]. \tag{3.11}$$

Here, we slightly abuse the notation $E[u, v]$ to represent the modified energy for convenience, since it is equivalent to the energy $E[u]$ in (1.6) in the continuous sense. The C_0 is a constant to make sure that the term $\langle u^m, u \rangle + C_0$ is positive for all time $t \in [0, T]$. In the actual computation, the C_0 is adjustable during the time evolution, and thus, we only need to choose the constant C_0 such that the term $\langle u^m, u \rangle + C_0 > 0$ in the time interval $t \in [t_n, t_{n+1}]$. This is easily fulfilled, since we only consider the solution smooth in time. We will discuss the C_0 adjustment process at the end of this subsection.

Denote $v^n \approx v(t_n)$ to be the semi-discretization of v in time, and also $v_h \approx \mathbf{v}(\mathbf{u})$ to be the semi-discretization of v in space. We write the space-time full discretization of v as $v_h^n \approx \mathbf{v}(\mathbf{u}^n, t_n)$. The reformulated equation system (3.11) can be discretized by the rational basis functions into the following form

$$\begin{cases} \mathbf{u}_t = -\mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{S}_1 \mathbf{F} \mathbf{P} \left(-\mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u} + \frac{1}{m} \frac{\mathbf{u}^m v_h}{\sqrt{\langle \mathbf{u}^m, \mathbf{u} \rangle_h + C_0}} \right) := f(\mathbf{u}, v_h), \\ (v_h)_t = \frac{m+1}{2\sqrt{\langle \mathbf{u}^m, \mathbf{u} \rangle_h + C_0}} \langle \mathbf{u}^m, \mathbf{u}_t \rangle_h := g(\mathbf{u}, v_h), \end{cases} \tag{3.12}$$

with the initial conditions

$$\mathbf{u}^0 = u(\mathbf{x}, 0), \quad v_h^0 = \sqrt{\langle (\mathbf{u}^0)^m, \mathbf{u}^0 \rangle_h + C_0}.$$

The fully discrete modified energy is defined as follows

$$E_h^n = \frac{1}{2} \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}^n, \mathbf{u}^n \rangle_h - \frac{1}{m(m+1)} ((v_h^n)^2 - C_0). \tag{3.13}$$

Next, we prove the following theorem for the high order energy-conservative schemes.

Theorem 3.4 *The s-stage symplectic Runge–Kutta method, which satisfies (3.9), conserves the L^1 -type integral (1.4), mass (1.5) and modified energy (3.11) in the discrete time flow for the reformulated gBO equation system (3.10), i.e.,*

$$I^{n+1} = I^n, \quad M^{n+1} = M^n, \quad \text{and} \quad E^{n+1} = E^n. \tag{3.14}$$

Furthermore, the symplectic Runge–Kutta method preserves the discrete energy (3.13) for the spatial semi-discretized system (3.12), i.e.,

$$E_h^n = E_h^{n-1} = \dots = E_h^0. \tag{3.15}$$

Proof The proof for the conservation of the temporal semi-discretized first integral, mass and energy (3.14) is standard, e.g., see [20, 47, 60, 72].

For the proof of the discrete energy conservation (3.15), substituting the inner product into the discrete sense, straightforward calculations yield

$$\begin{aligned} E_h^{n+1} &= \frac{1}{2} \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \mathbf{u}^{n+1}, \mathbf{u}^{n+1} \rangle_h - \frac{1}{m(m+1)} ((v_h^{n+1})^2 - C_0) \\ &= \frac{1}{2} \langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} (\mathbf{u}^n + \tau \sum_{i=1}^s b_i f(\mathbf{U}_i, V_i)), \mathbf{u}^n + \tau \sum_{i=1}^s b_i f(\mathbf{U}_i, V_i) \rangle_h \\ &\quad - \frac{1}{m(m+1)} \left((v_h^n + \tau \sum_{i=1}^s b_i g(\mathbf{U}_i, V_i))^2 - C_0 \right) \\ &= E_h^n + \tau \sum_{i=1}^s b_i \left(\langle \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} f(\mathbf{U}_i, V_i), \mathbf{U}_i \rangle_h - \frac{2}{m(m+1)} V_i g(\mathbf{U}_i, V_i) \right) \\ &\quad + \tau^2 \sum_{i,j=1}^s (b_i a_{ij} + b_j a_{ji} - b_i b_j) (\langle f(\mathbf{U}_i, V_i), f(\mathbf{U}_j, V_j) \rangle_h + g(\mathbf{U}_i, V_i) g(\mathbf{U}_j, V_j)) \\ &= E_h^n \end{aligned}$$

by using the relation (2.20), (3.12) and (3.9), where V_i is defined as the intermediate value of v_h^n for (3.12) similar to (3.7) and (3.8). □

Remark 3.1 Comparing with the proof in [72, Theorem 3.1], the symplectic Runge–Kutta method can preserve all the three quantities in the discrete time flow for the reformulated system (3.10). However, due to the limitation of the spatial discretization, only the discrete energy will be preserved in the fully discrete sense. The conservation of the discrete momentum (1.4) for the gKdV equations in [72] is obtained by using the circulant and anti-symmetric property of the first order differential matrix from the Fourier pseudo-spectral discretization which we note that the rational basis functions here do not possess such property.

The adjustment process for the constant C_0 from [72] can be adapted here. Suppose at $t = t_n$, as the time evolves, the solution u^n leads to the term $\int (u^n)^{m+1} dx + C_0 < Tol$, where Tol is a given positive number (e.g., $Tol = 5$). Then, we need to choose \tilde{C}_0 to ensure

$\int (u^{n+1})^{m+1} dx + \tilde{C}_0 > 0$. The constant \tilde{C}_0 can be chosen such that $\int (u^n)^{m+1} dx + \tilde{C}_0 > Tol$. For example, we can take $\tilde{C}_0 = 10 - \int (u^n)^{m+1} dx$, which leads to our new $\tilde{v}^n \approx \sqrt{10}$ (and generally ensure the positivity for $\int (u^{n+1})^{m+1} dx + \tilde{C}_0$). Then, by using $E[u^n, v^n] = E[u^n, \tilde{v}^n]$ from (3.11), we have our new \tilde{v}^n

$$\tilde{v}^n = \sqrt{(v^n)^2 + \tilde{C}_0 - C_0}. \tag{3.16}$$

Finally, we substitute the v^n and C_0 in (3.10) with \tilde{v}^n and \tilde{C}_0 , and then, continue with the time evolution for $t = t_{n+1}, t_{n+2}, \dots$.

Remark 3.2 Note that $v^2 = \int u^{n+1} dx + C_0$ holds only at the collocation points $t = t_n + \tau c_i$ for each $i = 1, 2, \dots, s$ in $t \in [t_n, t_{n+1}]$. However, the constant c_i may not necessarily be equal to 0 or 1, e.g., see Table 1. This means $v^2 = \int u^{n+1} dx + C_0$ does not hold at t_n in the discrete time flow. Therefore, the new \tilde{v}^n can only be evaluated by (3.16) to keep the discrete energy (3.11) invariant.

4 Numerical Results

In this section, we list our numerical examples for the proposed schemes. Before discussing the examples, we first describe a type of a fixed point iteration solver from [21, 72], which can be easily adapted here for solving the resulting nonlinear system from the IRK methods with the total computational cost on the order of $\mathcal{O}(N \log(N))$ from FFT. Since the solution to (1.1) is continuous in time, u^n is supposed to be close enough to u^{n+1} for sufficiently small time step τ . Thus, u^n is a good initial guess for the fixed point iteration to start with. While we don't give an analytic proof on the convergence or conservation properties during the fixed point iteration process, the solution always converges to the given accuracy level in our numerical experiments (usually less than 10^{-12} , which is our fixed point tolerance), and thus, the mass or energy are preserved from u^n to u^{n+1} . We also refer to a recent paper [7] for interesting readers, which studies the convergence speed and hyperbolic conservation laws for different nonlinear algebraic system solvers resulting from the hyperbolic equations. Now, we take the IRK4 ($s = 2$) case as an example to describe our fixed point iteration solver. Recall $\mathbf{D}_1 = \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{S}_1 \mathbf{F} \mathbf{P}$, and denote $\mathbf{D}_{\mathbf{H1}} = \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P}$. The IRK4 scheme to (3.12) are rewritten into the system

$$\begin{cases} \mathbf{U}_i = \mathbf{u}^n + \tau \sum_{j=1}^2 a_{ij} \mathbf{f}_j, \\ \Phi_i = \frac{(\mathbf{U}_i)^m}{\sqrt{((\mathbf{U}_i)^m, \mathbf{U}_i)_h + C_0}}, \quad g_i = \frac{m+1}{2} \langle \Phi_i, \mathbf{f}_i \rangle_h, \\ V_i = v_h^n + \tau \sum_{j=1}^2 a_{ij} g_j, \quad i = 1, 2, \end{cases} \tag{4.1}$$

and

$$\begin{cases} \mathbf{f}_1 = -\mathbf{D}_1 \left(\mathbf{D}_{\mathbf{H1}} (\mathbf{u}^n + \tau \sum_{j=1}^2 a_{1j} \mathbf{f}_j) + \frac{1}{m} \Phi_1 V_1 \right), \\ \mathbf{f}_2 = -\mathbf{D}_1 \left(\mathbf{D}_{\mathbf{H1}} (\mathbf{u}^n + \tau \sum_{j=1}^2 a_{2j} \mathbf{f}_j) + \frac{1}{m} \Phi_2 V_2 \right). \end{cases} \tag{4.2}$$

Then, \mathbf{u}^{n+1} and v_h^{n+1} can be updated by (3.8).

The system (4.1) and (4.2) can be solved by the fixed point iteration. At the l th iteration, we have

$$\begin{cases} (\mathbf{I} + \tau a_{11} \mathbf{D}_{\mathbf{H2}}) \mathbf{f}_1^{l+1} + \tau a_{12} \mathbf{D}_{\mathbf{H2}} \mathbf{f}_2^{l+1} = -(\mathbf{D}_{\mathbf{H2}} \mathbf{u}^m - \frac{1}{m} \mathbf{D}_1 (\Phi_1^l V_1^l)), \\ \tau a_{21} \mathbf{D}_3 \mathbf{f}_1^{l+1} + (\mathbf{I} + \tau a_{22} \mathbf{D}_{\mathbf{H2}}) \mathbf{f}_2^{l+1} = -(\mathbf{D}_{\mathbf{H2}} \mathbf{u}^m - \frac{1}{m} \mathbf{D}_1 (\Phi_2^l V_2^l)), \end{cases} \tag{4.3}$$

where \mathbf{I} is the identity matrix, and $\mathbf{D}_{\mathbf{H2}} = \mathbf{D}_1 \mathbf{D}_{\mathbf{H1}} = \mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_2 \mathbf{F} \mathbf{P}$. The system (4.3) can be first transformed into the sparse matrix by multiplying the diagonal matrix \mathbf{P} , and then implement the matrix operation \mathbf{F} by FFT. Finally, the standard sparse matrix solvers can be applied, since $\mathbf{H} \mathbf{S}_2$ is a sparse matrix. After obtaining \mathbf{f}_i^{l+1} , we can use (4.1) to obtain the values of other variables U_i^{l+1} , Φ_i^{l+1} , g_i^{l+1} , and V_i^{l+1} . The total computational cost remains on the same order.

We set $\mathbf{f}_1^0 = \mathbf{f}_2^0 = f(\mathbf{u}^n, v_h^n)$ to start the iteration and the fixed point iteration (4.3) converges in our numerical simulations, since the values from previous time step are usually close enough to the actual solution. The iteration terminates when

$$\max_i \|\mathbf{f}_i^{l+1} - \mathbf{f}_i^l\|_\infty < \epsilon,$$

where $i = 1, 2$, and ϵ is typically set to be $\epsilon = 10^{-12}$ in our simulations.

We denote by the IRK2-MC and IRK4-MC the mass conservative schemes for solving (2.18) by using the 1st and 2nd stage RK methods with Gauss–Legendre collocation points which are 2nd and 4th order accuracy in time, respectively from Table 1. We also denote by the IRK2-EC and IRK4-EC the energy conservative schemes, which are still 2nd and 4th order accuracy in time for solving the reformulated system (3.12). As a comparison, we use the commonly used 2nd order non-conservative semi-implicit Leap-Frog scheme as follows

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^{n-1}}{2\tau} = -\mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{S}_1 \mathbf{F} \mathbf{P} \left(\mathbf{P}^{-1} \mathbf{F}^{-1} \mathbf{H} \mathbf{S}_1 \mathbf{F} \mathbf{P} \left(\frac{\mathbf{u}^{n+1} + \mathbf{u}^{n-1}}{2} \right) + \frac{1}{m} (\mathbf{u}^n)^{\mathbf{m}} \right), \tag{4.4}$$

denoted as Leap-Frog. It is worth noting that the Leap-Frog scheme (4.4) may not be unconditionally stable. Indeed, if we consider the linear simplified model

$$u_t = ik^2 u - \lambda iku,$$

where ik^2 can be considered as the eigenvalues of the operator $\mathcal{H} \partial_{xx}$ ranging from $k = -N/2, \dots, N/2 - 1$, and similarly, ik is the eigenvalue of the operator ∂_x , and λ is the constant for approximating the term $\frac{1}{m} u^{m-1}$. The Leap-Frog scheme of this linear problem yields

$$\frac{u^{n+1} - u^{n-1}}{2\tau} - \frac{1}{2} ik^2 (u^{n+1} + u^{n-1}) + \lambda iku^n = 0.$$

By taking $z = k\tau$, its characteristic polynomial yields

$$(1 - izk)\xi^2 + 2iz\lambda\xi - (1 + izk) = 0,$$

which gives two roots with respect to ξ , i.e.,

$$\xi_{1,2} = \frac{1}{1 - izk} \left(iz\lambda \pm \sqrt{-z^2\lambda^2 + (1 + z^2k^2)} \right).$$

The stability condition is that $|\xi| \leq 1$. One sufficient condition for $|\xi| \leq 1$ is

$$-z^2\lambda^2 + (1 + z^2k^2) \geq 0, \tag{4.5}$$

which yields $|\xi|^2 = 1$ from direct calculation.

Thus, for each $k = k_0$, when $k_0^2 \geq \lambda^2$, (4.5) is automatically fulfilled. However, when $k_0^2 \leq \lambda^2$, we need $z \leq \frac{1}{\sqrt{\lambda^2 - k_0^2}}$, and this yields the restriction on the time step:

$$\tau \leq \frac{1}{|k_0| \sqrt{\lambda^2 - k_0^2}}.$$

Or equivalently, note that $k_0 \sqrt{\lambda^2 - k_0^2}$ reaches the maximum when $k_0 = \frac{\lambda}{\sqrt{2}}$, and thus,

$$\tau \leq 2/\lambda^2.$$

This indicates that when λ (or $\|u^m\|_{L^\infty}$ in the nonlinear case we consider) increases, we need to shrink the time step τ by λ^{-2} to ensure the stability. In other words, the instability occurs at the constant λ from the nonlinear part instead of the high frequency level (large values of $|k|$) from the linear part.

We track the following quantities at $t = t_n$ to check the accuracy:

$$\mathcal{E}^n = \|u^n_{exact} - \mathbf{u}^n\|_\infty; \tag{4.6}$$

$$\mathcal{E}_I^n = \max_{l < n} |I_h^l - I_h^0|; \tag{4.7}$$

$$\mathcal{E}_M^n = \max_{l < n} |M_h^l - M_h^0|; \tag{4.8}$$

$$\mathcal{E}_E^n = \max_{l < n} |E_h^l - E_h^0|. \tag{4.9}$$

When the SAV approach is not applied (IRK2-MC, IRK4-MC and Leap-Frog), the discrete energy E_h^n is computed from (3.3); and when the SAV approach is applied, the discrete energy E_h^n is computed from the modified version (3.13). We mention here that it is easy to see the equivalence between the Crank–Nicolson scheme (3.4) and the IRK2-MC scheme. The energy-conservative Crank–Nicolson (CNEC) scheme, which (3.5) considers reformulating the potential, also shares the same idea as in the IRK2-EC scheme.

Now, we are ready to illustrate examples for our numerical simulations.

Example 1. Our first example considers the soliton solution for the BO ($m = 2$) equation, $u(x, t) = \frac{4c}{1+c^2(x-x_0-ct)^2}$. These type of solutions come from the smooth, positive, decaying at infinity solitary wave solution to the profile equation

$$\mathcal{H}Q_x + cQ - \frac{1}{m}Q^m = 0, \tag{4.10}$$

where c is a constant that indicates the speed of the traveling waves as well as the magnitude. The solutions are expected to travel to the right as the solitons, for example from the spectral stability result in [4, 36] and the inverse scattering theory [27, 33].

In our numerical simulations, we take $\alpha = 25$ with $N = 1024$. We take the traveling speed $c = 2$ and starting point at $x_0 = -20$. The time step τ is taken to be $\tau = \frac{1}{20}$ for all the four IRK type methods as well as the Crank–Nicolson method, and $\tau = \frac{1}{40}$ for the Leap-Frog scheme (4.4), since taking the $\tau = \frac{1}{20}$ will lead to the numerical instability in our numerical computations for the Leap-Frog scheme. We stop our numerical simulation at $T = 20$.

Figure 1 shows the solution profile obtained from the IRK4-EC scheme. The left subplot is the initial condition u_0 , the right subplot is the time evolution. One can see that the solution travels in the solitary wave manner, which is previously observed in [32, 59] and also as expected.

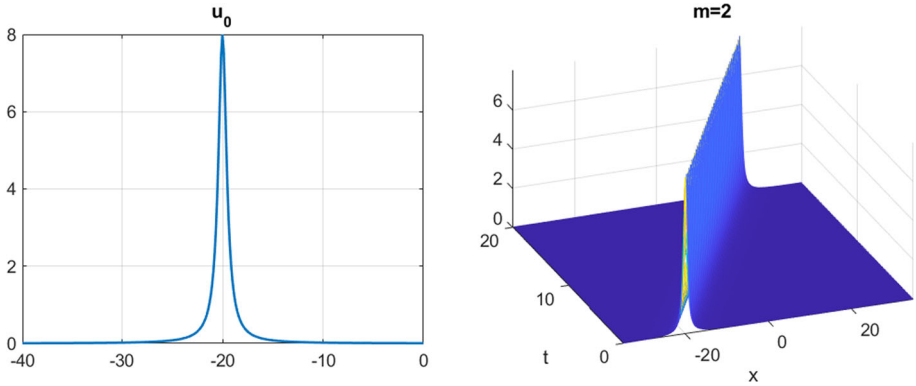


Fig. 1 The solution profile for Example 1 from the IRK4-EC. Left: u_0 . Right: $u(x, t)$

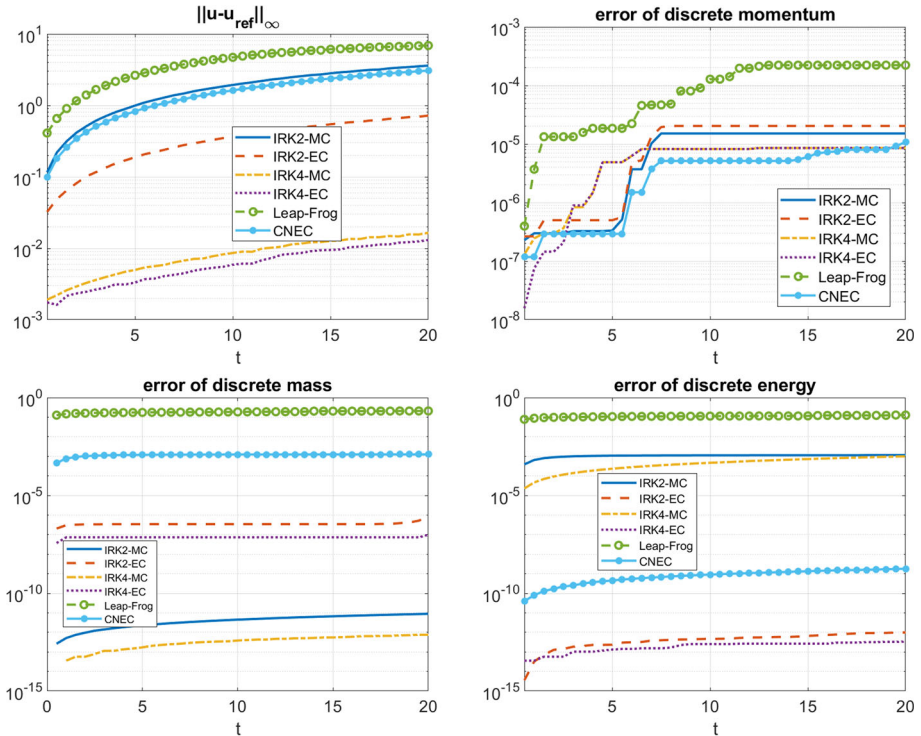


Fig. 2 The errors in Example 1 by different time integrators: IRK2-MC (solid blue); IRK2-EC(dash red); IRK4-MC (dash dot orange); IRK4-EC (dot purple); Leap-Frog (circle green); CNEC (doc blue). Top left: $\|u - u_{exact}\|_\infty$. Top right: discrete momentum error. Bottom left: discrete mass error. Bottom right: discrete energy error (Color figure online)

Figure 2 tracks the results obtained from the different time integrators. The top left subplot in Fig. 2 shows $\|u^n - u_{exact}\|_\infty$ with respect to time, where $u_{exact} = u(x, t_n)$ is the exact solution. One can see that the Leap-Frog scheme (4.4) has the largest error (see the green circle line). Moreover, the 4th order schemes (IRK4-MC and IRK4-EC) own the better

Table 2 The convergence rates of Leap-Frog, IRK2-MC, IRK2-EC IRK4-MC and IRK4-EC in Example 1

τ	Leap-Frog		IRK2-MC		IRK2-EC		IRK4-MC		IRK4-EC	
	Error	Rate	Error	Rate	Error	Rate	Error	Rate	Error	Rate
$\frac{1}{10}$	NA	NA	7.05	NA	3.45	NA	0.35	NA	0.61	NA
$\frac{1}{20}$	NA	NA	3.60	1.96	0.72	4.80	$1.6e-2$	21.6	$1.3e-2$	46.8
$\frac{1}{40}$	6.88	NA	1.02	3.53	0.18	4.05	$8.9e-4$	18.6	$5.6e-4$	23.2
$\frac{1}{80}$	1.92	3.59	0.26	3.93	0.44	4.03	$6.7e-5$	13.2	$5.1e-5$	10.9

Table 3 The comparison between the IRK4-MC, RK4 and the RRK4-MC method

	τ_{\max}	\mathcal{E}_{∞}	\mathcal{E}_M	\mathcal{E}_E	T	CPU time
IRK4-MC	0.05	0.016	$7e-12$	$1e-3$	20	30.28s
RK4	$1.6e-3$	$3e-5$	$1e-7$	$9e-10$	20	7.1s
RRK4-MC	$1.6e-3$	$3e-5$	$8e-14$	$1e-7$	$20 - (4e - 8)$	7.8s

accuracy than the second order schemes (IRK2-MC, IRK2-EC and Leap-Frog), which is as expected, since they have higher order temporal accuracy. Moreover, we found that the IRK2-EC scheme performs better than the CNEC scheme and the IRK2-MC scheme. Furthermore, we observe that the energy-conservative schemes (dash red line for IRK2-EC and dot purple line for IRK4-EC) perform better than the mass-conservative schemes (blue solid line for IRK2-MC and orange dash-dot line for IRK4-EC).

The top right subplot in Fig. 2 tracks the error of the discrete first integral (4.7) at different times. One can see that the more accurate the time integrators are, the better preservation of this quantity will be, though they are not conserved exactly according to the Rational Basis Function spatial discretization.

The bottom two subplots in Fig. 2 track the error of discrete mass (4.8) and energy (4.9), respectively. The mass-conservative schemes (IRK2-MC and IRK4-MC) keep the error of discrete mass below at the level of 10^{-12} , which is the tolerance of the fixed iteration in solving the resulting nonlinear system from the implicit Runge–Kutta method. On the other hand, the discrete mass error for the other types of schemes is relatively large, especially the Leap-Frog scheme (thus, referred to as least accurate).

The bottom right subplot tracks the error of the discrete energy. It shows that the energy-conservative schemes (IRK2-EC and IRK4-EC) keep the error of discrete mass below the level of 10^{-12} . This justifies the validity of our schemes. Similarly, for the other time integrators, the Leap-Frog scheme performs the worst even with a smaller time step τ ; the other two mass-conservative schemes keep the error of discrete energy around the level of 10^{-4} .

We also list the L^{∞} error \mathcal{E}^n at $t = T$ with different time step τ for these five time integrators in Table 2. Denote the error at time $t = T$ for the time step τ as $\mathcal{E}(\tau) = \|u(\mathbf{x}, T) - \mathbf{u}(T)\|_{\infty}$, and the rate defined as rate = $\frac{\mathcal{E}(2\tau)}{\mathcal{E}(\tau)}$. One can see that the Leap-Frog, IRK2-MC and IRK2-EC decrease with the ratio around 4, which are as expected, since they are of the second order schemes (since $4 = 2^2$). On the other hand, the ratio of the 4th order methods IRK4-MC and IRK4-EC is around 2^4 , since they are 4th order methods. When the time step τ is small, the decay rate is slightly below 16 (see the last row in Table 2). This is probably because the temporal error becomes comparable to the spatial discretization error, and consequently, affects the ratio.

We also want to compare the numerical efficiency between these implicit methods with explicit methods, such as the standard explicit RK4 and the recently proposed mass conservative relaxation Runge–Kutta methods from [38], e.g., the relaxation RK4 (RRK4-MC) method. These explicit methods need to take the time step τ small enough to satisfy the CFL condition ($\tau \sim 1/N^2$). Thus, usually, these explicit methods will have better accuracy compared with the proposed implicit methods, since the small time step is the necessity. On the other hand, smaller time step means more time evolution, resulting in more computational time. In Table 3, we denote the τ_{\max} meaning that the time step τ we use, and if we take the time step to be 2τ , the numerical simulation will fail. Table 3 shows the numerical tests for the Example 1. It shows that we can take $\tau = 1.6e - 3$, and obtain more accurate results in shorter computational time compared with the implicit methods (e.g., IRK4-MC). This shows that explicit time integrators are also an option for the simulation of the gBO type equations, since the linear operator $\mathcal{H}\partial_{xx}$ is not as stiff as the ∂_{xxx} in the KdV case. The RRK methods can also preserve the mass, or energy if we change a little about the scheme (see [57, 58] for details). It is more efficient than the implicit schemes such as the IRK4-MC scheme in this BO case. However, besides changing the time step τ when changing the nodes N from the CFL restriction, we also observe that the severe instability issue for the explicit methods will occur when the nonlinear power gets higher (see the case for $m = 5$ in Example 3), and thus, the implicit methods will become more efficient in those cases. Another concern about the RRK4 scheme is taken the non-uniform time step $\gamma_n\tau$ (γ_n is the constant to make the mass or energy to be preserved) at each time step. This may loose one order of accuracy if we want to interpret the solution as the given time $t_n = n\tau$, which is called the incremental direction technique (IDT) method in this case. In summary, both of the RRK schemes and the IRK schemes are good conservative schemes that may be useful in the future numerical investigation of gBO type equations. The choice of the time integrator will depend on case by case.

Example 2. We next consider the scattering solution for the BO equation ($m = 2$) with the initial condition $u_0 = -2 \operatorname{sech}^2(x)$. Its KdV version has been studied for questions on dispersion limit, see, e.g. [31, 40]. Here, we expect that a similar solution behavior may happen, since the BO equation only changes the dispersion term u_{xxx} from the KdV equation to $\mathcal{H}u_{xx}$ (less amount of dispersion if viewed on the Fourier frequency side). Note that a negative value for $\int u^3 dx$ may occur, and thus, the C_0 adjustment process in (3.16) will make $v(t)$ stay positive, and will keep the algorithm applicable for all time. The exact solution is not explicitly given, since due to the negative sign in the initial condition and coefficients chosen.

In this example, we still take the $N = 1024$ and $\alpha = 25$ for the spatial discretization. The time step $\tau = \frac{1}{400}$ ($\tau = \frac{1}{800}$ for the Leap-Frog) and the stopping time $T = 2$. We compute the reference solution u_{ref} by both IRK4-MC and IRK4-EC methods independently with an ultimately small time step ($\tau = 1/6400$), denoted as $u_{\text{ref-MC}}$ and $u_{\text{ref-EC}}$, respectively. Since we intend to track the convergence rate with respect to time, to minimize the influence from the spatial discretization error, we use the $u_{\text{ref-MC}}$ to compute the L^∞ error $\|\mathbf{u}^n - u_{\text{ref}}\|_\infty$ when the \mathbf{u}^n is obtained by the mass-conservative schemes (IRK2-MC and IRK4-MC), and use the $u_{\text{ref-EC}}$ to compute the L^∞ error $\|\mathbf{u}^n - u_{\text{ref}}\|_\infty$ when \mathbf{u}^n is obtained by the energy-conservative schemes (IRK2-EC and IRK4-EC) and the Leap-Frog scheme. The spatial discretization error accumulates as the time evolves, see Fig. 5. From Fig. 5, the difference increases to the level of 10^{-6} between these two solutions by the time we terminate the simulation.

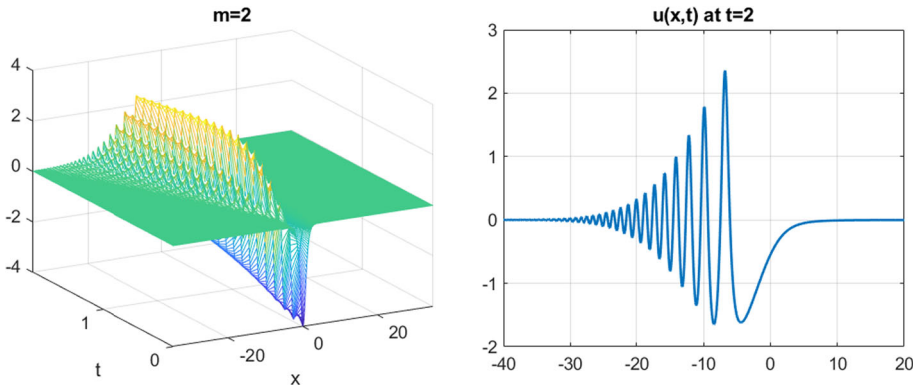


Fig. 3 The solution profile in Example 2 from IRK4-SAV. Left: $u(x, t)$. Right: $u(x, t)$ at $t = 2$

Table 4 The convergence rates of Leap-Frog, IRK2-MC, IRK2-EC, IRK4-MC and IRK4-EC in Example 2

τ	Leap-Frog		IRK2-MC		IRK2-EC		IRK4-MC		IRK4-EC	
	Error	Rate	Error	Rate	Error	Rate	Error	Rate	Error	Rate
$\frac{1}{200}$	0.104	NA	0.027	NA	0.027	NA	$4.4e-5$	NA	$4.4e-5$	NA
$\frac{1}{400}$	0.026	4.00	$6.7e-3$	3.99	$6.7e-3$	3.99	$2.8e-6$	15.8	$2.8e-6$	15.8
$\frac{1}{800}$	$6.5e-3$	3.99	$1.7e-3$	4.02	$1.7e-3$	4.02	$1.7e-7$	16.0	$1.7e-7$	16.0
$\frac{1}{1600}$	$1.6e-3$	4.00	$4.2e-4$	3.99	$4.2e-4$	3.99	$1.1e-8$	16.0	$1.1e-8$	16.0

Figure 3 shows the solution profile obtained from the IRK4-EC method. The left subplot shows the solution profile at different times t . The right plot shows the solution at the terminal time $t = 2$. We can see that the solution radiates to the right with fast oscillations. On the other hand, compared with the similar type of solutions to the KdV case (e.g., in [39, 72]), the frequency is smaller. This indicates that the lower order dispersion ($\mathcal{H}\partial_{xx}$ compared with ∂_{xxx}) generates slower oscillations.

Figure 4 tracks the first integral-error, error of discrete first integral, mass and energy with respect to time. One can see that the results are similar to the previous example, and also agree with our analysis in Sect. 2 and 3.

Table 4 shows the L^∞ error at $t = T$ with respect to the different time step τ . The decay rate is on the 2nd order for the 2nd order schemes (Leap-Frog, IRK2-MC and IRK2-EC), and on the 4th order for the 4th order schemes (IRK4-MC and IRK4-EC). Surprisingly, the IRK types of schemes (IRK2-MC and IRK2-EC, IRK4-MC and IRK4-EC) generate almost the same error (up to the decimals that we report) from the different reference solutions ($u_{\text{ref-MC}}$ and $u_{\text{ref-EC}}$). This implies that some possible cancellations may occur between the spatial discretization errors.

Example 3. Our final example considers the mBO ($m = 3$) and the gBO ($m = 4$) cases. We take the initial condition $u_0 = 0.99Q$, where Q is the soliton solution from (4.10) with $c = 1$. In these cases, while there is no explicit form for Q , the profile of Q can be obtained numerically, e.g., by the Petviashvili iteration from [55, 59], and its convergence analysis in [42, 44, 53]. From [26], when $u_0 = 0.99Q$, which indicates that the solution is below the mass-energy threshold, the solution is proven to exist globally in time. Recent numerical study in [59] shows that the solution blows up when $u_0 = 1.01Q$. In this paper, we consider

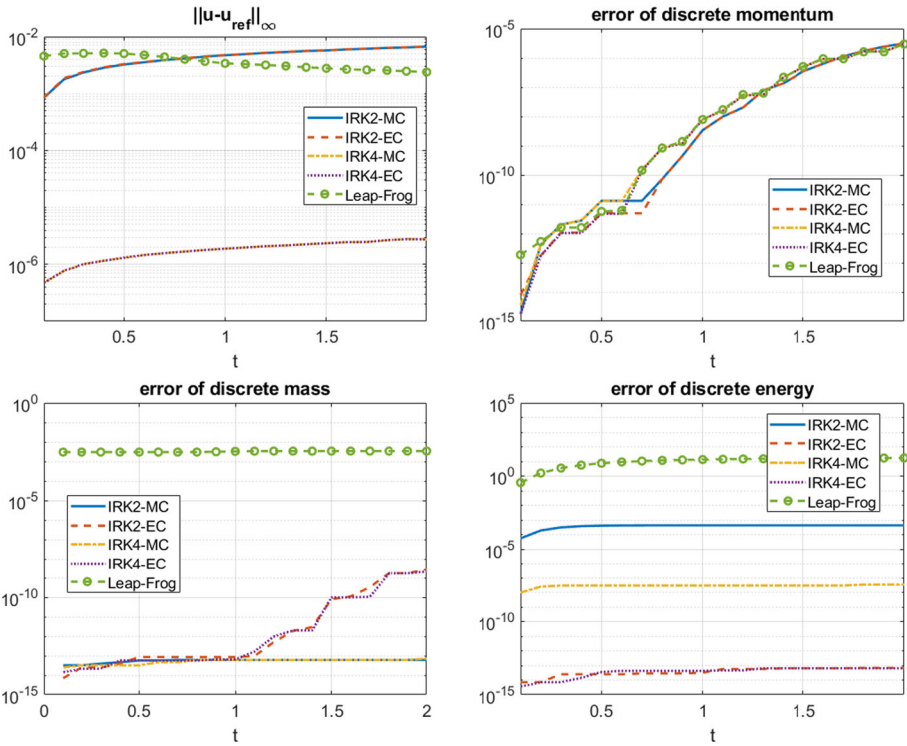
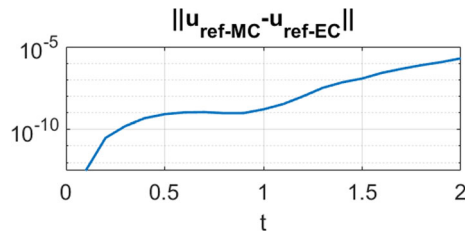


Fig. 4 The errors in Example 2 by different time integrators: IRK2-MC (solid blue); IRK2-EC(dash red); IRK4-MC (dash dot orange); IRK4-EC (dot purple); Leap-Frog (circle green). Top left: $\|u^n - u_{ref}\|_\infty$. Top right: discrete momentum error. Bottom left: discrete mass error. Bottom right: discrete energy error

Fig. 5 The difference of the reference solution obtained by the IRK4-EC scheme and IRK4-MC scheme. One can see the difference (mainly caused by the spatial discretization) keeps increasing to the level 10^{-6}



the globally existing solutions, and thus, we take $u_0 = 0.99Q$ in our example. We take $N = 1024, \alpha = 25, \tau = 0.02$ ($\tau = 0.01$ for the Leap-Frog scheme due to the stability issue) in our simulation. We run until $T = 10$ for $m = 3$, and $T = 5$ for $m = 4$.

We also test the $m = 5$ case. In this case, we take $N = 512, \alpha = 10, T = 1$ and $\tau = 0.001$ for the IRK schemes and the Crank–Nicolson scheme. It is worth noting that these implicit schemes still suffer the numerical instability. Indeed, when taking $N = 1024$, we cannot obtain the solution even for $\tau = 1e - 4$, and the solution starts behave normally from taking $\tau = 1e - 5$ in this case. This is an interesting question but unfortunately, to the authors’ best knowledge, no nonlinear stability analysis is considered for this type of questions so far. On the other hand, for given $N = 512$ and $\alpha = 10$, the RRK4 method still fails even

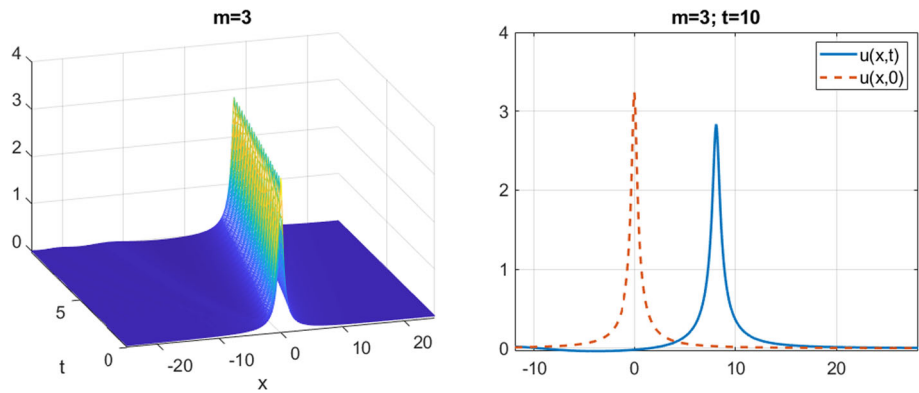


Fig. 6 The solution profile in Example 3 from IRK4-EC with $m = 3$. Left: $u(x, t)$. Right: $u(x, t)$ at $t = 10$

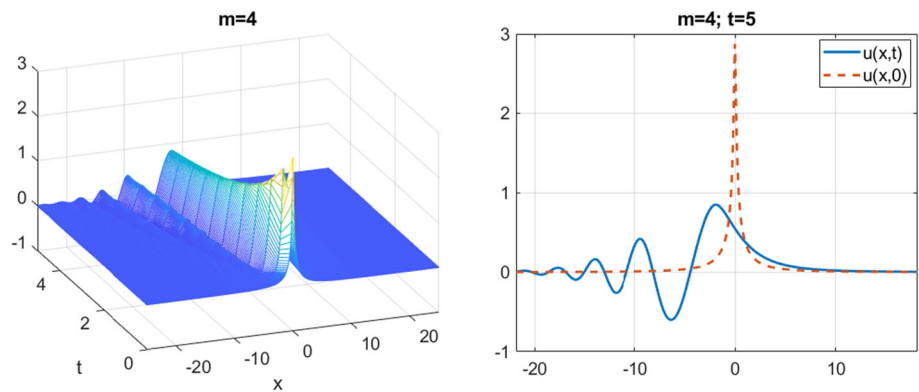


Fig. 7 The solution profile in Example 3 from IRK4-EC with $m = 4$. Left: $u(x, t)$. Right: $u(x, t)$ at $t = 5$

when we take $\tau = 1e - 6$. Therefore, in the high nonlinearity case, the implicit schemes are recommended with extra care.

Figures 6, 7 and 8 show the solution profiles obtained from the scheme IRK4-EC for $m = 3, m = 4$ and $m = 5$, respectively. The left subplots show the solution profiles $u(x, t)$ at different times. The right subplots show the solution at the final time T (blue solid line) and their comparison with the initial condition u_0 (red dash line). For $m = 3$, the solution travels to the right with some radiation parts scattering to the left. However, for the $m = 4, 5$ cases, the solution completely radiates to the left. Indeed, $m = 3$ is the L^2 -critical case, while $m = 4, 5$ are L^2 -supercritical cases. They fall into different category despite of the similar nonlinearity form. These different scattering solution behavior are worth to be further studied from both numerical and analytic point of view.

Figures 9, 10 and 11 track the error of discrete first integral (left subplot), mass (middle subplot) and energy (right subplot) at different times for $m = 3, 4, 5$, respectively. Again, the discrete mass or energy can be preserved by choosing the mass-conservative scheme or energy-conservative scheme, respectively, which agrees with the analysis in Sects. 2 and 3.

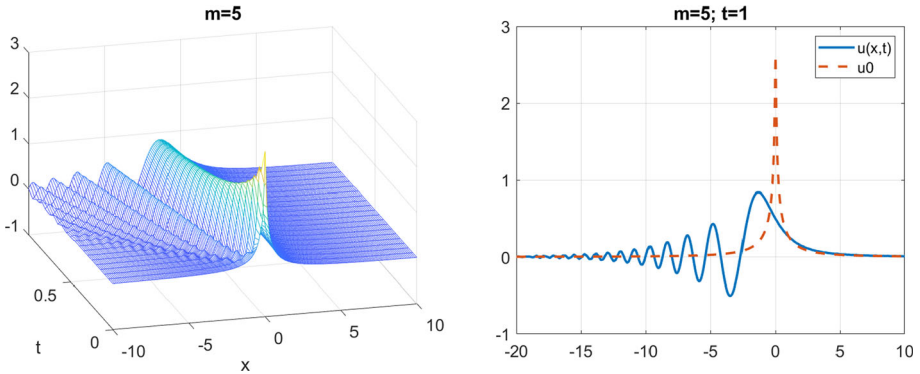


Fig. 8 The solution profile in Example 3 from IRK4-EC with $m = 5$. Left: $u(x, t)$. Right: $u(x, t)$ at $t = 1$

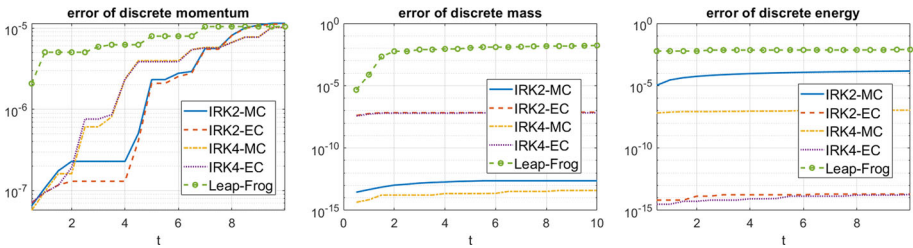


Fig. 9 The errors in Example 3 ($m = 3$) by different time integrators: IRK2-MC (solid blue); IRK2-EC (dashed red); IRK4-MC (dash dot orange); IRK4-EC (dot purple); Leap-Frog (circle green). Left: discrete momentum error. Middle: discrete mass error. Right: discrete energy error (Color figure online)

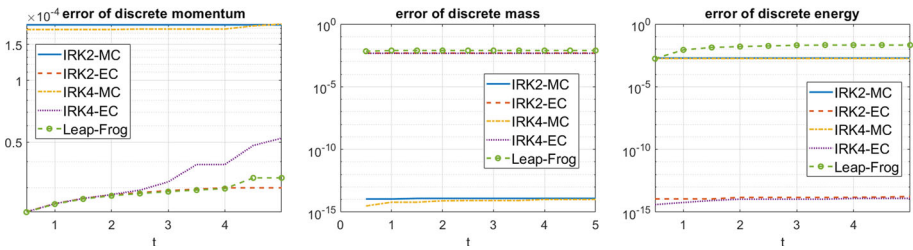


Fig. 10 The errors in Example 3 ($m = 4$) by different time integrators: IRK2-MC (solid blue); IRK2-EC (dashed red); IRK4-MC (dash dot orange); IRK4-EC (dot purple); Leap-Frog (circle green). Left: discrete momentum error. Middle: discrete mass error. Right: discrete energy error (Color figure online)

5 Conclusion and Other Discussion

The rational basis functions are powerful basis functions for spectral spatial discretization. Besides being the eigenfunctions of the Hilbert transform, the accessibility of the fast Fourier transform and the resulting sparse differential matrices, we find that the spatial conservative schemes can be constructed from its (anti-)Hermitian properties. The quadrature rule of this spectral method are also proposed.

Combined with the conservative time integrators, such as the classical symplectic Runge–Kutta schemes, the SRK schemes with the scalar auxiliary variable reformulation, or the Relaxation Runge–Kutta schemes, arbitrarily high order mass-conservative or energy-

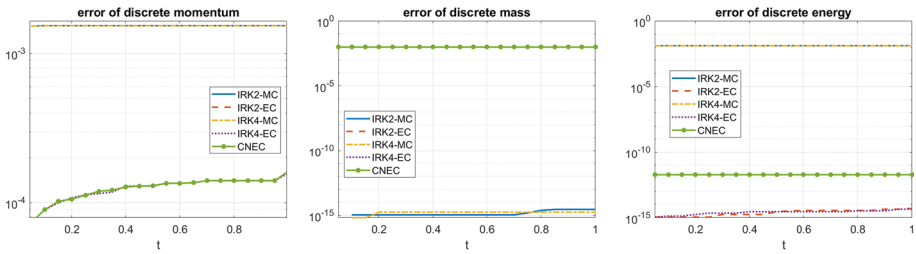


Fig. 11 The errors in Example 3 ($m = 5$) by different time integrators: IRK2-MC (solid blue); IRK2-EC (dash red); IRK4-MC (dash dot orange); IRK4-EC (dot purple); CNEC (dot green). Left: discrete momentum error. Middle: discrete mass error. Right: discrete energy error (Color figure online)

conservative numerical schemes can be constructed. In the meanwhile, the explicit Runge–Kutta schemes can be good conservative time integrators for the BO ($m = 2$) equations, but when the nonlinear power gets higher, the numerical observed nonlinear instability problem suggests us to take the implicit schemes.

This pseudo-spectral approach can be extend to construct the the conservative schemes for the gKdV equations, i.e.,

$$u_t = -u_{xxx} - \frac{1}{m}(u^m)_x;$$

and also the mass-energy conservative (structure-preserving) schemes for the NLS equations, i.e.,

$$u_t = i(u_{xx} + |u|^{m-1}u).$$

For example, the corresponding spatial semi-discretized form of the Eq. (3.12) for the gKdV equation will be

$$\begin{cases} \mathbf{u}_t = -\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_1\mathbf{F}\mathbf{P}\left(-\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_2\mathbf{F}\mathbf{P}\mathbf{u} + \frac{1}{m}\frac{\mathbf{u}^m v_h}{\sqrt{\langle \mathbf{u}^m, \mathbf{u} \rangle_h + C_0}}\right), \\ (v_h)_t = \frac{m+1}{2\sqrt{\langle \mathbf{u}^m, \mathbf{u} \rangle_h + C_0}}\langle \mathbf{u}^m, \mathbf{u}_t \rangle_h. \end{cases}$$

Similarly, the semi-discretized form for the NLS equation will be

$$\begin{cases} \mathbf{u}_t = i\left(-\mathbf{P}^{-1}\mathbf{F}^{-1}\mathbf{S}_2\mathbf{F}\mathbf{P}\mathbf{u} + \frac{|\mathbf{u}|^{m-1}\mathbf{u}v_h}{\sqrt{\langle |\mathbf{u}|^{m-1}\mathbf{u}, \mathbf{u} \rangle_h + C_0}}\right), \\ (v_h)_t = \frac{m}{2\sqrt{\langle |\mathbf{u}|^{m-1}\mathbf{u}, \mathbf{u} \rangle_h + C_0}}\text{Re}\left(\langle |\mathbf{u}|^{m-1}\mathbf{u}, \mathbf{u}_t \rangle_h\right). \end{cases}$$

The straightforward adaption of proof in Theorem 3.4 will show the energy-conservative result for the gKdV equations, and the mass and energy preserving result for the NLS equations. These results can also be easily extended to higher dimensions (e.g., Zakharov-Kuznetsov equation or the d -dimensional NLS equation) by applying the tensor product in the spatial discretization. We omit the proofs and numerical examples here for conciseness.

In summary, by applying the rational basis functions, the above illustrated conservative schemes will increase the computational efficiency significantly, especially, in tracking the solution’s long time behavior, or the slow decaying solutions (e.g., $u_0 = \frac{1}{1+x^2}$), since far less number of nodes are needed compared with the traditional domain truncation approaches.

Acknowledgements The author is thankful for Dr. Roudenko's helpful discussion, reading and remarks on the paper.

Funding The author is partially supported by the star-up grant from Chongqing University (PI: Kai Yang) and the NSF grant DMS-1927258 (PI: Svetlana Roudenko).

Data Availability Enquiries about data availability should be directed to the authors.

Declarations

Competing interests The authors have not disclosed any competing interests.

References

1. Abdelouhab, L., Bona, J., Felland, M., Saut, J.-C.: Nonlocal models for nonlinear, dispersive waves. *Physica D* **40**(3), 360–392 (1989)
2. Ablowitz, M.J., Ablowitz, M., Clarkson, P., Clarkson, P.A.: *Solitons, Nonlinear Evolution Equations and Inverse Scattering*, vol. 149. Cambridge University Press, Cambridge (1991)
3. Ablowitz, M.J., Fokas, A.S., Anderson, R.L.: The direct linearizing transform and the Benjamin–Ono equation. *Phys. Lett. A* **93**(8), 375–378 (1983)
4. Albert, J., Bona, J.L., Henry, D.: Sufficient conditions for stability of solitary-wave solutions of model equations for long waves. *Physica D* **24**, 343–366 (1987)
5. Benjamin, T.B.: Internal waves of permanent form in fluids of great depth. *J. Fluid Mech.* **29**(3), 559–592 (1967)
6. Benjamin, T.B., Bona, J.L., Mahony, J.J.: Model equations for long waves in nonlinear dispersive systems. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Sci.* **272**(1220), 47–78 (1972)
7. Birken, P., Linders, V.: Conservative iterative methods for implicit discretizations of conservation laws (2021)
8. Bock, T.L., Kruskal, M.D.: A two-parameter Miura transformation of the Benjamin–Ono equation. *Phys. Lett. A* **74**(3–4), 173–176 (1979)
9. Bona, J., Kalisch, H.: Models for internal waves in deep water. *Discrete Contin. Dyn. Syst. A* **6**(1), 1–20 (2000)
10. Bona, J., Kalisch, H.: Singularity formation in the generalized Benjamin–Ono equation. *Discrete Contin. Dyn. Syst.* **11**, 27–46 (2004)
11. Bona, J.L., Souganidis, P.E., Strauss, W.A.: Stability and instability of solitary waves of Korteweg–de Vries type. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **411**(1841), 395–412 (1987)
12. Boyd, J.P.: Spectral methods using rational basis functions on an infinite interval. *J. Comput. Phys.* **69**(1), 112–142 (1987)
13. Boyd, J.P.: The orthogonal rational functions of Higgins and Christov and algebraically mapped Chebyshev polynomials. *J. Approx. Theory* **61**(1), 98–105 (1990)
14. Boyd, J.P., Xu, Z.: Comparison of three spectral methods for the Benjamin–Ono equation: fourier pseudospectral, rational Christov functions and Gaussian radial basis functions. *Wave Motion* **48**(8), 702–706 (2011)
15. Boyd, J.P., Xu, Z.: Numerical and perturbative computations of solitary waves of the Benjamin–Ono equation with higher order nonlinearity using Christov rational basis functions. *J. Comput. Phys.* **231**(4), 1216–1229 (2012)
16. Burq, N., Planchon, F.: Smoothing and dispersive estimates for 1D Schrödinger equations with BV coefficients and applications. *J. Funct. Anal.* **236**(1), 265–298 (2006)
17. Burq, N., Planchon, F.: On the well-posedness of the Benjamin–Ono equation. *Math. Ann.* **340**, 497–542 (2008)
18. Butcher, J.C.: Implicit Runge–Kutta processes. *Math. Comput.* **18**(85), 50–64 (1964)
19. Christov, C.L.: A complete orthonormal system of functions in $L^2(-\infty, \infty)$ space. *SIAM J. Appl. Math.* **42**(6), 1337–1344 (1982)
20. Cooper, G.: Stability of Runge–Kutta methods for trajectory problems. *IMA J. Numer. Anal.* **7**(1), 1–13 (1987)
21. Cui, J., Wang, Y., Jiang, C.: Arbitrarily high-order structure-preserving schemes for the Gross–Pitaevskii equation with angular momentum rotation. *Comput. Phys. Commun.* **261**, 107767 (2021)

22. Davis, R.E., Acrivos, A.: Solitary internal waves in deep water. *J. Fluid Mech.* **29**(3), 593–607 (1967)
23. Debussche, A., Di Menza, L.: Numerical simulation of focusing stochastic nonlinear Schrödinger equations. *Physica D* **162**(3), 131–154 (2002)
24. Dutta, R., Holden, H., Koley, U., Risebro, N.H.: Convergence of finite difference schemes for the Benjamin–Ono equation. *Numer. Math.* **134**(2), 249–274 (2016)
25. Duyckaerts, T., Merle, F.: Dynamics of threshold solutions for energy-critical wave equation. *Int. Math. Res. Pap. IMRP*, pages Art ID rpn002, 67 (2008)
26. Farah, L.G., Linares, F., Pastor, A.: Global well-posedness for the k -dispersion generalized Benjamin–Ono equation. *Differential Integral Equations*, **27**(7/8), 601–612, 07 (2014)
27. Fokas, A.S., Ablowitz, M.J.: The inverse scattering transform for the Benjamin–Ono equation—a pivot to multidimensional problems. *Stud. Appl. Math.* **68**(1), 1–10 (1983)
28. Geng, S.: Construction of high order symplectic Runge–Kutta methods. *J. Comput. Math.* **11**(3), 256–260 (1993)
29. Gong, Y., Wang, Q., Wang, Y., Cai, J.: A conservative Fourier pseudo-spectral method for the nonlinear Schrödinger equation. *J. Comput. Phys.* **328**, 354–370 (2017)
30. Grava, T., Klein, C.: Numerical solution of the small dispersion limit of Korteweg–de Vries and Whitham equations. *Commun. Pure Appl. Math. A J. Issued Courant Inst. Math. Sci.* **60**(11), 1623–1664 (2007)
31. Grava, T., Klein, C.: A numerical study of the small dispersion limit of the Korteweg–de Vries equation and asymptotic solutions. *Physica D* **241**(23–24), 2246–2264 (2012)
32. James, R., Weideman, J.: Pseudospectral methods for the Benjamin–Ono equation. *Adv. Comput. Methods Partial Differ. Equ.* **7**, 371–377 (1992)
33. Kaup, D.J., Lakoba, T.I., Matsuno, Y.: Complete integrability of the Benjamin–Ono equation by means of action-angle variables. *Phys. Lett. A* **238**(2–3), 123–133 (1998)
34. Kaup, D.J., Matsuno, Y.: The inverse scattering transform for the Benjamin–Ono equation. *Stud. Appl. Math.* **101**(1), 73–98 (1998)
35. Kenig, C., Koenig, K.: On the local well-posedness of the Benjamin–Ono and modified Benjamin–Ono equations. *Math. Res. Lett.* **10**, 879–895 (2003)
36. Kenig, C., Martel, Y.: Asymptotic stability of solitons for the Benjamin–Ono equation. *Rev. Mat. Iberoam.* **25**, 909–970 (2009)
37. Kenig, C.E., Ponce, G., Vega, L.: On the generalized Benjamin–Ono equation. *Trans. Am. Math. Soc.* **342**, 155–172 (1994)
38. Ketcheson, D.I.: Relaxation Runge–Kutta methods: conservation and stability for inner-product norms. *SIAM J. Numer. Anal.* **57**(6), 2850–2870 (2019)
39. Klein, C., Saut, J.-C.: IST versus PDE, a comparative study. In: *Hamiltonian Partial Differential Equations and Applications*. Fields Institute Communications, vol. 75, pp. 383–449 (2015)
40. Klein, C., Peter, R.: Numerical study of blow-up and dispersive shocks in solutions to generalized Korteweg–de Vries equations. *Phys. D* **304**(305), 52–78 (2015)
41. Koch, H., Tzvetkov, N.: On the local well-posedness of the Benjamin–Ono equation on $H^s(\mathbb{R})$. *Int. Math. Res. Not.* **26**, 1449–1464 (2003)
42. Lakoba, T., Yang, J.: A generalized Petviashvili iteration method for scalar and vector Hamiltonian equations with arbitrary form of nonlinearity. *J. Comput. Phys.* **226**(2), 1668–1692 (2007)
43. Landoulsi, O., Roudenko, S., Yang, K.: Interaction with an obstacle in the 2d focusing nonlinear Schrödinger equation. *Adv. Comput. Math.* (2023) (**accepted**)
44. Le, U., Pelinovsky, D.E.: Convergence of Petviashvili’s method near periodic waves in the fractional Korteweg–de Vries equation. *SIAM J. Math. Anal.* **51**(4), 2850–2883 (2019)
45. Liu, H., Yan, J.: A local discontinuous Galerkin method for the Korteweg–de Vries equation with boundary effect. *J. Comput. Phys.* **215**(1), 197–218 (2006)
46. Liu, H., Yi, N.: A Hamiltonian preserving discontinuous Galerkin method for the generalized Korteweg–de Vries equation. *J. Comput. Phys.* **321**, 776–796 (2016)
47. Liu, Z., Zhang, H., Qian, X., Song, S.: Mass and energy conservative high order diagonally implicit Runge–Kutta schemes for nonlinear Schrödinger equation in one and two dimensions. *arXiv preprint arXiv:1910.13700* (2019)
48. Millet, A., Rodriguez, A.D., Roudenko, S., Yang, K.: Behavior of solutions to the 1d focusing stochastic nonlinear Schrödinger equation with spatially correlated noise. *Stoch. Partial Differ. Equ. Anal. Comput.* **9**, 1031–1080 (2021)
49. Miloh, T., Prestin, M., Shtilman, L., Tulin, M.: A note on the numerical and N-soliton solutions of the Benjamin–Ono evolution equation. *Wave Motion* **17**(1), 1–10 (1993)
50. Molinet, L., Ribaud, F.: Well-posedness results for the generalized Benjamin–Ono equation with arbitrary large initial data. *Int. Math. Res. Not.* **70**, 3757–3795 (2004)

51. Molinet, L., Ribaud, F.: Well-posedness results for the generalized Benjamin–Ono equation with small initial data. *J. Math. Pures Appl.* **83**, 277–311 (2004)
52. Nakamura, A.: A direct method of calculating periodic wave solutions to nonlinear evolution equations. I. Exact two-periodic wave solution. *J. Phys. Soc. Jpn.* **47**(5), 1701–1705 (1979)
53. Olson, D., Shukla, S., Simpson, G., Spirn, D.: Petviashvili’s method for the Dirichlet problem. *J. Sci. Comput.* **66**(1), 296–320 (2016)
54. Ono, H.: Algebraic solitary waves in stratified fluids. *J. Phys. Soc. Jpn.* **39**(4), 1082–1091 (1975)
55. Pelinovsky, D.E., Stepanyants, Y.A.: Convergence of Petviashvili’s iteration method for numerical approximation of stationary solutions of nonlinear wave equations. *SIAM J. Numer. Anal.* **42**(3), 1110–1127 (2004)
56. Pelloni, B., Dougalis, V.: Numerical solution of some nonlocal, nonlinear dispersive wave equations. *J. Nonlinear Sci.* **10**(1), 1–22 (2000)
57. Ranocha, H., Mitsotakis, D., Ketcheson, D.I.: A broad class of conservative numerical methods for dispersive wave equations. *Commun. Comput. Phys.* **29**(4), 979–1029 (2021)
58. Ranocha, H., Sayyari, M., Dalcin, L., Parsani, M., Ketcheson, D.I.: Relaxation Runge–Kutta methods: fully discrete explicit entropy-stable schemes for the compressible Euler and Navier–Stokes equations. *SIAM J. Sci. Comput.* **42**(2), A612–A638 (2020)
59. Roudenko, S., Wang, Z., Yang, K.: Dynamics of solutions in the generalized Benjamin–Ono equation: a numerical study. *J. Comput. Phys.* **445**, 110570 (2021)
60. Sanz-Serna, J.: Runge–Kutta schemes for Hamiltonian systems. *BIT Numer. Math.* **28**(4), 877–883 (1988)
61. Sanz-Serna, J., Abia, L.: Order conditions for canonical Runge–Kutta schemes. *SIAM J. Numer. Anal.* **28**(4), 1081–1096 (1991)
62. Saut, J.: Benjamin–Ono and intermediate long wave equations: modeling, IST and PDE. In: Miller, P., Perry, P., Saut, J.C., Sulem, C. (eds.) *Nonlinear Dispersive Partial Differential Equations and Inverse Scattering*. Fields Institute Communications, vol. 83, pp. 95–160. Springer, New York (2018)
63. Saut, J.-C.: Sur quelques généralisations de l’ équation de Korteweg–de Vries. *J. Math. Pures Appl.* **58**, 21–61 (1979)
64. Shen, J., Tang, T., Wang, L.-L.: *Spectral Methods*, volume 41 of Springer Series in Computational Mathematics Algorithms, Analysis and Applications. Springer, Heidelberg (2011)
65. Shen, J., Xu, J., Yang, J.: The scalar auxiliary variable (SAV) approach for gradient flows. *J. Comput. Phys.* **353**, 407–416 (2018)
66. Shen, J., Xu, J., Yang, J.: A new class of efficient and robust energy stable schemes for gradient flows. *SIAM Rev.* **61**(3), 474–506 (2019)
67. Tao, T.: Global well-posedness of the Benjamin–Ono in $H^1(\mathbb{R})$. *J. Hyperbolic Differ. Eq.* **1**(1), 27–49 (2004)
68. Thomée, V., Murthy, A.V.: A numerical method for the Benjamin–Ono equation. *BIT Numer. Math.* **38**(3), 597–611 (1998)
69. Trefethen, L.N.: *Spectral Methods in MATLAB*, volume 10 of Software, Environments, and Tools. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2000)
70. Vento, S.: Sharp well-posedness results for the generalized Benjamin–Ono equation with high nonlinearity. *Differ. Integr. Equ.* **22**(5–6), 425–446 (2009)
71. Weideman, J.A.C.: Computing the Hilbert transform on the real line. *Math. Comp.* **64**(210), 745–762 (1995)
72. Yang, K.: Arbitrarily high-order conservative schemes for the generalized korteweg–de vries equation. *SIAM J. Sci. Comput.* **44**(4), A2709–A2733 (2022)
73. Yi, N., Huang, Y., Liu, H.: A direct discontinuous Galerkin method for the generalized Korteweg–de Vries equation: energy conservation and boundary effect. *J. Comput. Phys.* **242**, 351–366 (2013)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.