



**HAL**  
open science

# Adaptation and assesement of Projected Nesterov accelerated gradient flow to compute stationary states of nonlinear Schrödinger equations

Xavier Antoine, Chorouq Bentayaa, Jérémie Gaidamour

## ► To cite this version:

Xavier Antoine, Chorouq Bentayaa, Jérémie Gaidamour. Adaptation and assesement of Projected Nesterov accelerated gradient flow to compute stationary states of nonlinear Schrödinger equations. International Journal of Computer Mathematics, 2023, 101 (1), pp.21-36. 10.1080/00207160.2023.2294688 . hal-04270784v2

**HAL Id: hal-04270784**

**<https://hal.science/hal-04270784v2>**

Submitted on 10 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Adaptation and assesement of Projected Nesterov accelerated gradient flow to compute stationary states of nonlinear Schrödinger equations

Xavier ANTOINE<sup>a</sup>, Chorouq BENTAYAA<sup>a</sup>, Jérémie GAIDAMOUR<sup>a</sup>

<sup>a</sup> *Université de Lorraine, CNRS, Inria, IECL, F-54506 Vandoeuvre-lès-Nancy Cedex, France*

---

## Abstract

The aim of the paper is to derive minimization algorithms based on the Nesterov accelerated gradient flow [23, 24, 25] to compute the ground state of nonlinear Schrödinger equations, which can potentially include a fractional laplacian term. A comparison is developed with standard gradient flow formulations showing that the Nesterov accelerated gradient flow has some interesting properties but at the same time finds also some limitations due to the nature of the problem. A few simulations are finally reported to understand the behavior of the algorithms and open the path to further complicate questions that require more advanced studies concerning the application of the Nesterov accelerated gradient flow to nonlinear Schrödinger equations.

*Keywords:* Bose-Einstein condensation; nonlinear space fractional Schrödinger equation; stationary states; gradient flow; Nesterov accelerated gradient flow ; pseudo-spectral method

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Projected gradient flow formulation and its discretization</b>	<b>3</b>
<b>3</b>	<b>Projected Nesterov accelerated gradient flow formulation</b>	<b>5</b>
3.1	Time discretization of the CN-NAGF formulation (3.15) as a second-order equation	6
3.2	Time discretization of the CN-NAGF formulation (3.15) as a first-order system	10
<b>4</b>	<b>Pseudospectral spatial discretization</b>	<b>11</b>
<b>5</b>	<b>Numerical examples</b>	<b>11</b>
5.1	The integer case ( $\alpha = 1$ )	11
5.2	The general fractional case ( $0 < \alpha < 1$ )	13
<b>6</b>	<b>Conclusion</b>	<b>14</b>

## 1. Introduction

In the present paper, we consider the problem of computing a stationary state of the  $d$ -dimensional ( $d = 1, 2, 3$ ) NonLinear space Fractional Schrödinger Equation (NLFSE) written as

---

*Email addresses:* [xavier.antoine@univ-lorraine.fr](mailto:xavier.antoine@univ-lorraine.fr) (Xavier ANTOINE),  
[chorouq.bentayaa@univ-lorraine.fr](mailto:chorouq.bentayaa@univ-lorraine.fr) (Chorouq BENTAYAA), [jeremie.gaidamour@univ-lorraine.fr](mailto:jeremie.gaidamour@univ-lorraine.fr)  
(Jéréemie GAIDAMOUR)

the constrained energy minimization problem

$$\begin{cases} \text{Find } \phi \in L^2(\mathbb{R}^d) \text{ such that} \\ \phi \in \arg \min_{\|\phi\|_2=1} \mathcal{E}^\alpha(\phi). \end{cases} \quad (1.1)$$

The  $L^2(\mathbb{R}^d)$ -norm of a function  $\phi$  is given by

$$\|\phi\|_2^2 = \int_{\mathbb{R}^d} |\phi|^2 d\mathbf{x} := \langle \phi, \phi \rangle.$$

We also introduce the hermitian inner-product

$$\forall (u, v) \in L^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d), \quad \langle u, v \rangle := \int_{\mathbb{R}^d} uv^* d\mathbf{x},$$

setting  $v^*$  as the complex conjugate of a function  $v$ . For (1.1), we denote by  $\mathcal{E}^\alpha$  the total fractional energy functional of the dimensionless NLFSE

$$\mathcal{E}^\alpha(\phi) = \int_{\mathbb{R}^d} \frac{1}{2} (-\nabla^2)^\alpha \phi \phi^* d\mathbf{x} + \int_{\mathbb{R}^d} V(\mathbf{x}) |\phi|^2 d\mathbf{x} + \int_{\mathbb{R}^d} F(|\phi|^2) d\mathbf{x} := \mathcal{E}_k^\alpha(\phi) + \mathcal{E}_p(\phi) + \mathcal{E}_i(\phi).$$

In 3D, the fractional Laplace operator is  $(-\Delta)^\alpha \phi = (-\nabla^2)^\alpha \phi$ . The gradient operator is defined by  $\nabla := (\partial_x, \partial_y, \partial_z)^t$  for  $\mathbf{x} = (x, y, z)^t \in \mathbb{R}^3$  (similarly, for the two-dimensional, we set  $\nabla := (\partial_x, \partial_y)^t$  and  $\mathbf{x} = (x, y)^t \in \mathbb{R}^2$ ). Based on Fourier integral operators ( $0 \leq \alpha \leq 2$ ), we have the representation

$$(-\nabla^2)^\alpha \phi = \int_{\mathbb{R}^d} \hat{\phi}(\mathbf{k}) |\mathbf{k}|^{2\alpha} e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{k}, \quad (1.2)$$

setting  $|\mathbf{k}| = (k_x^2 + k_y^2)^{1/2}$  in 2D for  $\mathbf{k} := (k_x, k_y)$ , where the Fourier transform is

$$\hat{\phi}(\mathbf{k}) = \int_{\mathbb{R}^d} \phi(\mathbf{x}) e^{-i\mathbf{k} \cdot \mathbf{x}} d\mathbf{x}.$$

The case  $0 < \alpha < 1$  corresponds to subdispersion in the NLSFE while  $1 < \alpha < 2$  leads to superdispersion [6]. We define  $V$  as the confining potential. The nonlinearity is given through the smooth real-valued function  $f(\Psi) := F'(\Psi)$ , where the density function is  $\Psi = |\phi|^2$ . A very common case is the cubic nonlinearity

$$F(s) = \beta s^2 / 2. \quad (1.3)$$

Therefore, we have  $f(s) = \beta s$ , defining  $\beta$  as the nonlinearity strength (see [3, 5, 6, 7, 12] for other examples). The energy gradient is:  $\nabla \mathcal{E}^\alpha(\phi) = 2\mathcal{H}_\phi^\alpha \phi$ , with  $\mathcal{H}_\phi^\alpha = \frac{1}{2}(-\Delta)^\alpha + V + f(|\phi|^2)$ . Let  $\mathcal{S} = \{\phi \in L^2(\mathbb{R}^d), \|\phi\|_2 = 1\}$  be the unit spherical manifold for the normalization constraint. For  $\phi \in \mathcal{S}$  given, we define the tangent space at  $\phi$  as:  $T_\phi \mathcal{S} = \{h \in L^2(\mathbb{R}^d), \text{Re} \langle \phi, h \rangle = 0\}$ . The orthogonal projection  $M_\phi$  onto this space is such that we have  $M_\phi h = h - \text{Re} \langle \phi, h \rangle \phi$ . Then, the Euler-Lagrange equation related to our constraint problem at a minimum  $\phi \in \mathcal{S}$  is  $M_\phi \nabla \mathcal{E}^\alpha(\phi) = 0$ , which is equivalent to the nonlinear fractional eigenvalue problem:  $\mathcal{H}_\phi^\alpha \phi = \lambda^\alpha \phi$ . The parameter  $\lambda^\alpha := \lambda^\alpha(\phi) = \langle \mathcal{H}_\phi^\alpha \phi, \phi \rangle$  is the Lagrange multiplier associated to the spherical constraint (chemical potential for  $\alpha = 1$ ).

Up to now, many studies were directed towards the development of minimization algorithms for the computation of the ground state (and other excited states) of the nonlinear Schrödinger equation for  $\alpha = 1$ . For example, and without being exhaustive, we refer to

[1, 4, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17, 19, 20, 21, 22, 30] where various methods and some extensions are analyzed and evaluated. Less studies were devoted to the fractional case (see e.g. [5, 6, 7]). The aim of the present paper is to focus on the possibility to design new minimization algorithms based on advanced gradient technique. Indeed, one of the most standard and powerful method for the computing of the ground state is based on the gradient flow formulation [8, 11]. Basically, the discrete standard schemes are related to a projected gradient flow formulation combined with suitable explicit or semi-implicit discretization schemes. Recently, with the incredible growth of deep learning techniques, more robust gradient techniques were introduced to improve the convergence rate of the basic steepest descent method. This is the case of the widely used Nesterov accelerated gradient method [23, 24, 25] which finds many impressive applications. Nevertheless, the (constraint and unconstraint) minimization problems that are solved with this algorithm [23, 24, 25] and its variants [26, 28, 29] usually involve simpler energy functionals and constraints than for the ground state computation associated to the nonlinear Schrödinger equation. The goal of this paper is to understand if one can design Nesterov accelerated gradient flow based schemes for our problem and to test the efficiency and robustness of the resulting methods. To this end, we extend the standard methods designed for the gradient flow formulation to the Nesterov accelerated gradient flow formulation to develop some new schemes that lead to minimization algorithms. Supported by some numerical examples, this analysis shows that getting efficient minimization schemes based on the Nesterov accelerated gradient flow formulation is a nontrivial task and is maybe even not possible in some cases like the fractional case.

The organization of the paper is the following. After the introduction in Section 1, Section 2 is devoted to basic properties of the projected gradient flow formulation and its discretization. Section 3 proposes the design of a few Nesterov accelerated gradient flow based schemes which are discretized in the spatial pseudo-spectral framework in Section 4. Numerical examples are reported in Section 5. We finally end by a conclusion in Section 6.

## 2. Projected gradient flow formulation and its discretization

In this section, we recall the standard Gradient Flow with Discrete Normalization (GF-DN) [8, 11] for computing the stationary states of the NLFSE. In particular, we straightforwardly extend the Gradient Flow with Lagrange Multipliers (GF-LM) studied in [22] which has recently been proven to be accurate for  $\alpha = 1$ .

We introduce some discrete times  $t_k \geq 0$ ,  $k \in \mathbb{N}$ , with  $t_0 = 0$ . We designate by  $\rho_k := t_k - t_{k-1} > 0$  the  $k$ -th local time step of the gradient flow under consideration. In the case of a constant step method with  $\rho_k = \rho$ ,  $k \in \mathbb{N}$ , then the discrete times are  $t_k = k\rho > 0$ . The GF-DN reads [8, 11] as follows

$$\left\{ \begin{array}{l} \text{Find } \phi := \phi(\mathbf{x}, t) \text{ such that} \\ \partial_t \phi = -\frac{1}{2} \nabla \mathcal{E}^\alpha(\phi), \quad t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, t_{k+1}) = \phi(\mathbf{x}, t_{k+1}^+) = \frac{\phi(\mathbf{x}, t_{k+1}^-)}{\|\phi(\mathbf{x}, t_{k+1}^-)\|_2}, \quad k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d. \end{array} \right. \quad (2.4)$$

In the above system, we set:  $\phi(\mathbf{x}, t_k^\pm) = \lim_{t \rightarrow t_k^\pm} \phi(\mathbf{x}, t)$ . The initial data  $\phi_0$ , such that  $\|\phi_0\|_2 = 1$ , is the initial guess of the ground state, given for example through an ansatz of the ground state (e.g. the Thomas-Fermi approximation of the ground state or a gaussian function). To simplify the presentation, let us now assume that the nonlinearity is cubic. Then, the GF-DN (2.4) can

be viewed as a first-order splitting method for the following Continuous Normalized Gradient Flow (CN-GF) [22]

$$\begin{cases} \text{Find } \phi := \phi(\mathbf{x}, t) \text{ such that} \\ \partial_t \phi = -\left(\frac{1}{2}\nabla \mathcal{E}^\alpha(\phi) - \lambda_\phi^\alpha(t)\phi\right), \quad t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \end{cases} \quad (2.5)$$

where the energy is such that

$$\mathcal{E}^\alpha(\phi) := \int_{\mathbb{R}^d} \left[ \frac{1}{2}(-\nabla^2)^\alpha \phi \phi^* + V|\phi|^2 + \frac{\beta}{2}|\phi|^4 \right] d\mathbf{x}, \quad (2.6)$$

and the Lagrange multiplier is

$$\lambda^\alpha = \lambda_\phi^\alpha(t) = \frac{\langle \mathcal{H}_\phi^\alpha(\cdot, t), \phi(\cdot, t) \rangle}{\|\phi(\cdot, t)\|_2^2} = \frac{1}{\|\phi\|_2^2} \int_{\mathbb{R}^d} \left[ \frac{1}{2}(-\nabla^2)^\alpha \phi \phi^* + V|\phi|^2 + \beta|\phi|^4 \right] d\mathbf{x}.$$

It is proven that the CN-GF (2.5) is normalization conserving as well as energy diminishing [22] for  $\alpha = 1$ . Therefore, when  $t \rightarrow +\infty$ , the solution to the CN-GF (2.5) converges to the ground state if the initial data  $\phi_0$  is well-chosen.

Let us now consider the time discretization of (2.5). Following [22], four discretizations are analyzed. Here, we restrict our study to the two following schemes a) (explicit) Forward Euler (FE) scheme and b) semi-implicit Backward Euler (BE) scheme. The Gradient Flow with Forward Euler (GF<sub>FE</sub>) discretization and Lagrange multiplier is given by

$$\begin{cases} \text{Compute } \phi^{k+1}, k \geq 0, \text{ such that} \\ \frac{\tilde{\phi}^{k+1} - \phi^k}{\rho_{k+1}} = -\left(\frac{1}{2}\nabla \mathcal{E}^\alpha(\phi^k) - \lambda^\alpha(\phi^k)\phi^k\right), \quad \mathbf{x} \in \mathbb{R}^d, \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \quad \mathbf{x} \in \mathbb{R}^d, \\ \phi^0 = \phi_0, \quad \mathbf{x} \in \mathbb{R}^d, \end{cases} \quad (2.7)$$

where we have used the first-order approximation of the time derivative

$$\partial_t \phi(t_{k+1}, \mathbf{x}) \simeq \frac{\tilde{\phi}^{k+1} - \phi^k}{\rho_{k+1}}$$

and we remark that  $\tilde{\phi}^{k+1}$  needs to be normalized to get the correct solution  $\phi^{k+1}$ .

Let us now introduce the following modified total energy

$$\mathcal{E}_{\phi^k}^\alpha(\phi) := \int_{\mathbb{R}^d} \left[ \frac{1}{2}(-\nabla^2)^\alpha \phi \phi^* + V|\phi|^2 + \frac{\beta}{2}|\phi^k|^2|\phi|^2 \right] d\mathbf{x}, \quad (2.8)$$

with gradient

$$\frac{1}{2}\nabla \mathcal{E}_{\phi^k}^\alpha(\tilde{\phi}^{k+1}) = \mathcal{H}_{\phi^k}^\alpha \tilde{\phi}^{k+1} = \frac{1}{2}(-\Delta)^\alpha \tilde{\phi}^{k+1} + V\tilde{\phi}^{k+1} + \beta|\phi^k|^2\tilde{\phi}^{k+1}. \quad (2.9)$$

Then, the Gradient Flow with semi-implicit Backward Euler (GF<sub>BE</sub>) discretization and Lagrange multiplier is

$$\begin{cases} \text{Compute } \phi^{k+1}, k \geq 0, \text{ such that} \\ \frac{\tilde{\phi}^{k+1} - \phi^k}{\rho_{k+1}} = -\left(\frac{1}{2}\nabla \mathcal{E}_{\phi^k}^\alpha(\tilde{\phi}^{k+1}) - \lambda^\alpha(\phi^k)\phi^k\right), \quad \mathbf{x} \in \mathbb{R}^d, \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \quad \mathbf{x} \in \mathbb{R}^d, \\ \phi^0 = \phi_0, \quad \mathbf{x} \in \mathbb{R}^d, \end{cases} \quad (2.10)$$

which leads to solving a linear variable coefficients PDE at each step  $k \geq 0$ , following

$$\begin{cases} \text{Compute } \phi^{k+1}, k \geq 0, \text{ such that} \\ \left( I + \rho_{k+1} \mathcal{H}_{\phi^k}^\alpha \right) \tilde{\phi}^{k+1} = (I + \rho_{k+1} \lambda^\alpha(\phi^k)) \phi^k, & \mathbf{x} \in \mathbb{R}^d, \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, & \mathbf{x} \in \mathbb{R}^d, \\ \phi^0 = \phi_0, & \mathbf{x} \in \mathbb{R}^d. \end{cases} \quad (2.11)$$

Let us remark that the Lagrange multiplier is treated explicitly here accordingly to [22].

Now, let us recall the important energy decaying properties of both schemes.

**Proposition 1.** *Let us assume that  $V(\mathbf{x}) \geq 0$ ,  $\alpha = 1$ ,  $\beta \geq 0$  and that  $\phi^n$  is sufficiently smooth for the GF<sub>FE</sub> scheme (2.7). Then, for  $k \geq 0$ , there exists  $\rho_{k+1}^{\text{FE}} > 0$  such that for any step  $0 < \rho_{k+1} \leq \rho_{k+1}^{\text{FE}}$ , we have the energy decreasing property*

$$\mathcal{E}^\alpha(\phi^{k+1}) \leq \mathcal{E}^\alpha(\phi^k). \quad (2.12)$$

If again  $V(\mathbf{x}) \geq 0$ ,  $\alpha = 1$  and  $\beta \geq 0$ , then for any step  $\rho_{k+1} > 0$ , the solution  $\phi^{k+1}$  to the GF<sub>BE</sub> scheme (2.10) satisfies the modified energy decaying property

$$\mathcal{E}_{\phi^k}^\alpha(\phi^{k+1}) \leq \mathcal{E}_{\phi^k}^\alpha(\phi^k) = \lambda^\alpha(\phi^k). \quad (2.13)$$

### 3. Projected Nesterov accelerated gradient flow formulation

Recently, steepest descent techniques were improved, in particular to design some more efficient minimization algorithms for deep learning problems. More specifically, we propose to design some Nesterov Accelerated type algorithms [23, 24, 25, 26, 28, 29] for computing the ground state of the NLSFE and to compare their performance and robustness with the schemes based on the standard gradient flow formulation.

To this end, let us introduce the following Nesterov Accelerated Gradient Flow (NAGF) formulation with Discrete Normalization (NAGF-DN)

$$\begin{cases} \text{Find } \phi := \phi(\mathbf{x}, t) \text{ such that} \\ \partial_t^2 \phi + \gamma(t) \partial_t \phi = -\frac{1}{2} \nabla \mathcal{E}^\alpha(\phi), & t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, t_{k+1}) = \phi(\mathbf{x}, t_{k+1}^-) = \frac{\phi(\mathbf{x}, t_{k+1}^-)}{\|\phi(\mathbf{x}, t_{k+1}^-)\|_2}, & k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^d, \\ \partial_t \phi(\mathbf{x}, 0) = \phi_1(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^d, \end{cases} \quad (3.14)$$

which can also be written as the CN-NAGF system with Lagrange multiplier

$$\begin{cases} \text{Find } \phi := \phi(\mathbf{x}, t) \text{ such that} \\ \partial_t^2 \phi + \gamma(t) \partial_t \phi = -\left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\phi) - \lambda^\alpha(\phi) \phi \right), & t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^d, \\ \partial_t \phi(\mathbf{x}, 0) = \phi_1(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^d. \end{cases} \quad (3.15)$$

The term  $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is a positive real-valued momentum function that we will fix later. The initial guess  $\phi_0$  can be taken as for the GF-DN approach. Concerning the choice of the initial velocity  $\phi_1$ , this is less clear since it is not given *a priori*. We may consider here the possibilities  $\phi_1 = 0$  and  $\phi_1 = \phi_0$  which are relatively standard, but also the case where  $\phi_1$  is given randomly. From numerical simulations (not reported here), it appears that  $\phi_1 = \phi_0$  is a suitable choice that we keep for the paper.

### 3.1. Time discretization of the CN-NAGF formulation (3.15) as a second-order equation

Let us now derive some time discretizations of the CN-NAGF. Classically, we can use the first-order approximation (2) as well as the following discretization of the second-order time derivative at time  $t_k$  for non uniform time steps  $\rho_k$ , for  $k \geq 1$ ,

$$\partial_t^2 \phi(t_k, \mathbf{x}) \simeq \frac{1}{\rho_{k+1}\rho_{k+1/2}} \phi^{k+1} - \frac{2}{\rho_k \rho_{k+1}} \phi^k + \frac{1}{\rho_k \rho_{k+1/2}} \phi^{k-1}, \quad (3.16)$$

with  $\rho_{k+1/2} = (\rho_k + \rho_{k+1})/2$ . Then, setting  $\gamma_k = \gamma(t_k)$ , we obtain the implicit relation for the first equation of (3.15) written at time  $t_k$

$$\phi^{k+1} = \phi^k + \frac{\rho_{k+1}}{\rho_k} (1 - \gamma_k \rho_{k+1/2}) (\phi^k - \phi^{k-1}) - \rho_{k+1} \rho_{k+1/2} \left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\phi^k) - \lambda^\alpha(\phi^k) \phi^k \right). \quad (3.17)$$

Now, let us introduce the auxiliary function  $\psi^k$  such that

$$\psi^k := \phi^k + \frac{\rho_{k+1}}{\rho_k} (1 - \gamma_k \rho_{k+1/2}) (\phi^k - \phi^{k-1}). \quad (3.18)$$

Therefore, we have the equivalent writing of (3.17)

$$\phi^{k+1} = \psi^k - \rho_{k+1} \rho_{k+1/2} \left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\phi^k) - \lambda^\alpha(\phi^k) \phi^k \right). \quad (3.19)$$

The gradient term in equation (3.19) is explicit and taken at point  $\phi^k$ . Because of (3.18), an alternative and better choice to  $\nabla \mathcal{E}^\alpha(\phi^k)$  consists in using  $\nabla \mathcal{E}^\alpha(\psi^k)$ . Indeed, since  $\|\phi^k - \phi^{k-1}\|_2$  is supposed to be small, we have

$$\mathcal{E}^\alpha(\psi^k) \simeq \mathcal{E}^\alpha(\phi^k) + \frac{\rho_{k+1}}{\rho_k} (1 - \gamma_k \rho_{k+1/2}) \nabla \mathcal{E}^\alpha(\phi^k) (\phi^k - \phi^{k-1}). \quad (3.20)$$

In addition, if we assume that the momentum function  $\gamma$  is such that:  $(1 - \gamma_k \rho_{k+1/2}) \geq 0$  and noticing that

$$\phi^k - \phi^{k-1} \simeq -\rho_k \nabla \mathcal{E}^\alpha(\phi^k), \quad (3.21)$$

we deduce

$$\mathcal{E}^\alpha(\psi^k) \simeq \mathcal{E}^\alpha(\phi^k) - \rho_{k+1} (1 - \gamma_k \rho_{k+1/2}) \|\nabla \mathcal{E}^\alpha(\phi^k)\|_2^2 \leq \mathcal{E}^\alpha(\phi^k). \quad (3.22)$$

Therefore choosing  $\psi^k$  instead of  $\phi^k$  in (3.19) provides a solution with lower total energy. As a consequence, we consider the following explicit two-stages semi-discrete system approximating (3.15)

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, \text{ for } k \geq 1, \text{ such that} \\ \psi^k := \phi^k + \frac{\rho_{k+1}}{\rho_k} (1 - \gamma_k \rho_{k+1/2}) (\phi^k - \phi^{k-1}), \\ \tilde{\phi}^{k+1} = \psi^k - \rho_{k+1} \rho_{k+1/2} \left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\psi^k) - \lambda^\alpha(\phi^k) \phi^k \right), \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho_1 \phi_1}{\|\phi_0 + \rho_1 \phi_1\|_2}. \end{array} \right. \quad (3.23)$$

To obtain  $\phi^1$ , we use the approximation  $\tilde{\phi}^1 = \phi_0 + \rho_1 \phi_1$  and normalize it. This scheme is called NAGF<sub>FE</sub><sup>2</sup> scheme (the exponent 2 refers to the fact that the equation is discretized as a second-order Ordinary Differential Equation (ODE)). The above scheme corresponds to the

NAGF-DN method, with variable step size and Lagrange multiplier, to compute the minimum of the total energy functional  $\mathcal{E}^\alpha$ . Now, to simplify the presentation, let us assume that we have uniformly sampled discretization time points  $t_k = k\rho$ . Then, the constant step NAGF-DN writes

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \psi^k := \phi^k + (1 - \gamma_k \rho)(\phi^k - \phi^{k-1}), \\ \tilde{\phi}^{k+1} = \psi^k - \rho^2 \left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\psi^k) - \lambda^\alpha(\phi^k) \phi^k \right), \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho \phi_1}{\|\phi_0 + \rho \phi_1\|_2}. \end{array} \right. \quad (3.24)$$

Let us now look into the details on how to choose the momentum function  $\gamma$ . As already said, we have to check that we have the property:  $\gamma_k \rho \leq 1$ . Inspired by the work by Candès *et al.* [28], we can choose  $\gamma(t) = 3/t$ . Then, one gets  $\gamma_k \rho \geq 1$  for  $k \geq 3$ . In addition, we obtain:  $1 - \gamma_k \rho = (1 - 3/k)$ , since  $t_k = k\rho$ . An alternative and better choice is to rather consider the following slight modification  $\gamma(t) = \gamma_\rho(t) = 3/(t + 2\rho)$ . Then, in this case, we have  $\gamma(t_k) := \gamma_{k,\rho} = 3\rho^{-1}/(k + 2)$  and  $1 - \gamma_k \rho = (k - 1)/(k + 2)$ , which is the well-known choice initially introduced by Nesterov [28]. Another popular momentum choice [28] is given by

$$\gamma_k = \theta_k(\theta_{k-1}^{-1} - 1), \text{ with } \theta_{k+1} = \frac{1}{2}(\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2), \theta_0 = 1.$$

For the function  $\gamma_\delta$  and a constant step size  $\rho > 0$ , the algorithm simplifies to

$$\left\{ \begin{array}{l} \text{compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \psi^k := \phi^k + \frac{k-1}{k+2}(\phi^k - \phi^{k-1}), \\ \tilde{\phi}^{k+1} = \psi^k - \rho^2 \left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\psi^k) - \lambda^\alpha(\phi^k) \phi^k \right), \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho \phi_1}{\|\phi_0 + \rho \phi_1\|_2}. \end{array} \right. \quad (3.25)$$

In the previous derivation, the first-order derivative was approximated at first-order. For a constant step  $\rho$ , a second-order approximation is

$$\partial_t \phi(t_k, \mathbf{x}) \simeq \frac{\phi^{k+1} - \phi^{k-1}}{2\rho}, \quad (3.26)$$

leading to the second-order formulation

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \psi^k := \phi^k + \frac{2k+1}{2k+7}(\phi^k - \phi^{k-1}), \\ \tilde{\phi}^{k+1} = \psi^k - \rho^2 \left( \frac{1}{2} \nabla \mathcal{E}^\alpha(\psi^k) - \lambda^\alpha(\phi^k) \phi^k \right), \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho \phi_1}{\|\phi_0 + \rho \phi_1\|_2}. \end{array} \right. \quad (3.27)$$



The two above schemes are explicit. Alternatively, we can also consider a semi-implicit Backward Euler discretization which leads to the NAGF<sub>BE</sub><sup>2</sup> scheme

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \frac{1}{\rho_{k+1}\rho_{k+1/2}}\tilde{\phi}^{k+1} - \frac{2}{\rho_k\rho_{k+1}}\phi^k + \frac{1}{\rho_k\rho_{k+1/2}}\phi^{k-1} + \gamma_k\frac{(\phi^k - \phi^{k-1})}{\rho_k} \\ \quad \quad \quad = -\left(\frac{1}{2}\nabla\mathcal{E}_{\phi^k}^\alpha(\tilde{\phi}^{k+1}) - \lambda^\alpha(\phi^k)\phi^k\right), \\ \phi^k = \frac{\tilde{\phi}^k}{\|\tilde{\phi}^k\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho\phi_1}{\|\phi_0 + \rho\phi_1\|_2}. \end{array} \right. \quad (3.28)$$

As a consequence, each time step of (3.28) requires the solution to a linear partial differential equation written in the framework of NA as

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \psi^k = \phi^k + \frac{\rho_{k+1}}{\rho_k}(1 - \rho_{k+1/2}\gamma_k)(\phi^k - \phi^{k-1}), \\ (I + \rho_{k+1/2}\rho_{k+1}\mathcal{H}_{\phi^k}^\alpha)\tilde{\phi}^{k+1} = \psi^k + \rho_{k+1/2}\rho_{k+1}\lambda^\alpha(\phi^k)\phi^k, \\ \phi^k = \frac{\tilde{\phi}^k}{\|\tilde{\phi}^k\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho\phi_1}{\|\phi_0 + \rho\phi_1\|_2}. \end{array} \right. \quad (3.29)$$

If  $\rho$  is constant, the scheme (3.28) simplifies as follows

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \frac{\tilde{\phi}^{k+1} - 2\phi^k + \phi^{k-1}}{\rho^2} + \gamma_k\frac{(\phi^k - \phi^{k-1})}{\rho} = -\left(\frac{1}{2}\nabla\mathcal{E}_{\phi^k}^\alpha(\tilde{\phi}^{k+1}) - \lambda^\alpha(\phi^k)\phi^k\right), \\ \phi^k = \frac{\tilde{\phi}^k}{\|\tilde{\phi}^k\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho\phi_1}{\|\phi_0 + \rho\phi_1\|_2}, \end{array} \right. \quad (3.30)$$

and (3.29) is

$$\left\{ \begin{array}{l} \text{Compute } \phi^{k+1}, k \geq 1, \text{ such that} \\ \psi^k = \phi^k + (1 - \rho\gamma_k)(\phi^k - \phi^{k-1}), \\ (I + \rho^2\mathcal{H}_{\phi^k}^\alpha)\tilde{\phi}^{k+1} = \psi^k + \rho^2\lambda^\alpha(\phi^k)\phi^k, \\ \phi^k = \frac{\tilde{\phi}^k}{\|\tilde{\phi}^k\|_2}, \\ \phi^0 = \phi_0, \\ \phi^1 = \frac{\phi_0 + \rho\phi_1}{\|\phi_0 + \rho\phi_1\|_2}. \end{array} \right. \quad (3.31)$$

For the constant step case, let us note that, by combining (3.20) with (3.21), and using a Taylor's expansion of  $\mathcal{E}^\alpha(\phi^{k-1})$ , we obtain :  $\mathcal{E}^\alpha(\phi^k) \leq \mathcal{E}^\alpha(\phi^{k-1})$ ,  $k \geq 1$  if  $1 - \gamma_k\rho \geq 0$ . In the

case where the momentum is too large, then this energy decay property can be violated and oscillations may appear. Choosing an adaptive step  $\rho_k = 1/\gamma_k$ , for large values of  $\gamma_k$ , would slow down the convergence of the algorithm. Even for the linear case  $\beta = 0$  and without normalization constraint, the Nesterov accelerated algorithm is known to be non energy diminishing [28], unlike the standard gradient flow approach (see Proposition 1). In particular, if one tries to adapt the proof of [11] for the energy decay of the gradient flow for  $\alpha = 1$ , we would have to prove that

$$\frac{d\mathcal{E}^\alpha}{dt}(\phi) = -2 \int_{\mathbb{R}^d} \partial_t \phi (\partial_t^2 \phi + \gamma(t) \partial_t \phi) d\mathbf{x} \quad (3.32)$$

is negative, which may be wrong in the general case. At the continuous level, the partial differential equation solved on each interval  $[t_k, t_{k+1})$  is given by a nonlinear damped Klein-Gordon equation [10, 14] following the first equation of (3.15)

$$\partial_t^2 \phi + \gamma(t) \partial_t \phi - \frac{1}{2} \Delta \phi + V \phi + \beta |\phi|^2 \phi - \lambda(\phi) \phi = 0, \quad t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \quad (3.33)$$

which can indeed lead to oscillations as a nonlinear wave equation (but with an associated decaying energy [10, 14, 27] for  $\alpha = 1$ ). To solve this problem of the fluctuations related to the large momentum case, restarting strategies of  $\gamma_k$  can be easily included into the algorithm. The first approach is the so-called adaptive restarting [26] (ARNAG) based on the i) function scheme

$$\mathcal{E}^\alpha(\phi^k) > \mathcal{E}^\alpha(\phi^{k-1}), \quad (3.34)$$

or the ii) gradient scheme

$$\nabla \mathcal{E}^\alpha(\psi^{k-1})(\phi^k - \phi^{k-1}) > 0. \quad (3.35)$$

A second approach is based on the scheduled restart SRNAG [29] using the update formula for the velocity

$$\psi^k := \phi^k + \frac{\text{mod}(k, r_m)}{\text{mod}(k, r_m) + 3} (\phi^k - \phi^{k-1}), \quad (3.36)$$

where  $r_m \in \mathbb{N}$  defines the restart frequency parameter. In particular, a fixed restart  $r_m = r_f$  can be considered, where  $r_f$  is set *a priori*. These restarting strategies have all been implemented in the Nesterov-type algorithm for our problem. From the numerical experiments, it appears that our situation, restarting is not very efficient and can sometimes penalize the convergence towards the ground state.

Usually, the introduction of the second-order ODE in the Nesterov gradient formulation (i.e. the first equation of (3.15)) is related to a mechanical system interpretation (e.g. heavy ball formulation). Nevertheless, an alternative viewpoint is the following: let us first assume that  $\gamma$  is a constant function and let us define the Laplace transform

$$\mathcal{L}(X)(\tau) = \int_0^{+\infty} e^{-\tau t} X(t) dt, \quad \tau > 0,$$

where we set  $X = \partial_t \phi$ . Then, Laplace transforming the ODE and using the first-order initial condition  $X_0 = \phi_1$  leads to

$$(\tau + \gamma) \mathcal{L}(X) = -\mathcal{L}\left(\frac{1}{2} \nabla \mathcal{E}^\alpha(\phi) - \lambda^\alpha(\phi) \phi\right) + \phi_1, \quad (3.37)$$

or equivalently to the convolution formulation (using the  $\star$  operator)

$$X = \partial_t \phi = -\mathcal{L}^{-1}\left(\frac{1}{\tau + \gamma}\right) \star \left(\left(\frac{1}{2} \nabla \mathcal{E}^\alpha(\phi) - \lambda^\alpha(\phi) \phi\right) + \phi_1\right), \quad (3.38)$$

where we know that

$$\mathcal{L}^{-1}\left(\frac{1}{\tau + \gamma}\right) = e^{-\gamma t} Y(t),$$

denoting by  $Y$  the Heaviside's function and by  $\mathcal{L}^{-1}$  the inverse Laplace transform [18]. Therefore, based on (3.37) or (3.38), we can see that the effect of the explicit convolution, or equivalently solving implicitly the local ODE, amounts to precondition the gradient flow in view of a faster convergence. It is not clear *a priori* which operator should be considered. For  $\gamma(t) = 3/t$ , we have a similar result but with the property that

$$\mathcal{L}(\gamma(t)X)(\tau) := \mathcal{L}\left(\frac{3X}{t}\right)(\tau) = 3 \int_{\tau}^{+\infty} \mathcal{L}(X)(u) du.$$

### 3.2. Time discretization of the CN-NAGF formulation (3.15) as a first-order system

The NA algorithms derived in Section 3.1 are based on discretizations of the first- and second-order derivatives of the second-order ODE. A direct alternative is to rather rewrite (3.15) as a first-order ODE system and then discretize. More concretely, (3.15) can be written as

$$\left\{ \begin{array}{l} \text{Find } (\phi, \psi) \text{ such that} \\ (\partial_t + \gamma(t))\psi = -\left(\frac{1}{2}\nabla\mathcal{E}^\alpha(\phi) - \lambda^\alpha(\phi)\phi\right), \quad t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \partial_t\phi = \psi, \quad t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \\ \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \\ \partial_t\phi(\mathbf{x}, 0) = \phi_1(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d. \end{array} \right. \quad (3.39)$$

Then the explicit Forward Euler discretization gives the NAGF<sub>FE</sub><sup>1</sup> scheme

$$\left\{ \begin{array}{l} \text{Compute } (\phi^{k+1}, \psi^{k+1}), k \geq 0, \text{ such that} \\ \psi^{k+1} = (1 - \rho_{k+1}\gamma_k)\psi^k - \rho_{k+1}\left(\frac{1}{2}\nabla\mathcal{E}^\alpha(\phi^k) - \lambda^\alpha(\phi^k)\phi^k\right), \\ \tilde{\phi}^{k+1} = \phi^k + \rho_{k+1}\psi^{k+1}, \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \\ \phi^0 = \phi_0, \\ \psi^0 = \phi_1. \end{array} \right. \quad (3.40)$$

An alternative first-order semi-implicit discretization leads to

$$\left\{ \begin{array}{l} \text{Compute } (\phi^{k+1}, \psi^{k+1}), k \geq 0, \text{ such that} \\ (I + \rho_{k+1}^2 \mathcal{H}_{\phi^k}^\alpha)\tilde{\phi}^{k+1} = \phi^k + \rho_{k+1}(1 - \rho_{k+1}\gamma_k)\psi^k + \rho_{k+1}^2 \lambda^\alpha(\phi^k)\phi^k, \\ \psi^{k+1} = \frac{\tilde{\phi}^{k+1} - \phi^k}{\rho_{k+1}}, \\ \phi^{k+1} = \frac{\tilde{\phi}^{k+1}}{\|\tilde{\phi}^{k+1}\|_2}, \\ \phi^0 = \phi_0, \\ \psi^0 = \phi_1. \end{array} \right. \quad (3.41)$$

Concerning the initial data  $\phi_1$ , it can be chosen as  $\phi_0$ , randomly or as an approximation through the gradient flow since  $\phi_1 = \partial_t\phi$  at  $t = 0$ . Here, we will use  $\phi_1 = \phi_0$  which appears as adapted from the numerical experiments.

## 4. Pseudospectral spatial discretization

We consider now a pseudo-spectral discretization scheme using the Fast Fourier Transforms (FFTs) [1, 2, 3, 11]. In the 2D case (the extension to the 3D case is direct), we introduce the truncated wave function  $\phi$  in a square domain  $\mathcal{D} := [-a_x, a_x] \times [-a_y, a_y]$  ( $a_x$  and  $a_y$  are both positive). We impose periodic boundary conditions, and we discretize the wave function  $\phi$  by using even numbers  $n_x$  and  $n_y$  grid points in the  $x$ - and  $y$ -directions, respectively. Let  $M := (n_x, n_y)$ . Then we consider the uniformly sampled grid  $\mathcal{D}_M := \{\mathbf{x}_{k_1, k_2} = (x_{k_1}, y_{k_2})\}_{(k_1, k_2) \in \mathcal{O}_M}$ . Here, we set:  $\mathcal{O}_M := \{0, \dots, n_x - 1\} \times \{0, \dots, n_y - 1\}$ ,  $x_{k_1+1} - x_{k_1} = h_x$  and  $y_{k_2+1} - y_{k_2} = h_y$ , for the mesh sizes  $h_x = 2a_x/n_x$  and  $h_y = 2a_y/n_y$ . We also need the  $N \times N$  Hermitian matrix(-free) operators acting from  $\mathbb{C}^N$  ( $N = n_x n_y$  in 2D) to  $\mathbb{C}$  such that  $\llbracket (-\Delta)^\alpha \rrbracket := \llbracket (-\partial_x^2 - \partial_y^2)^\alpha \rrbracket$ . Therefore, we can derive the discrete gradient of the energy

$$(\nabla \mathcal{E}^\alpha)(\phi) \approx (\llbracket \nabla \rrbracket E^\alpha)(\phi) = 2\llbracket \mathcal{H}_\phi^\alpha \rrbracket \phi,$$

where we define

$$\llbracket \mathcal{H}_\phi^\alpha \rrbracket := \frac{1}{2} \llbracket (-\Delta)^\alpha \rrbracket + \llbracket V \rrbracket + \llbracket f(|\phi|^2) \rrbracket.$$

Let  $\phi := (\tilde{\phi}(\mathbf{x}_{k_1, k_2}))_{(k_1, k_2) \in \mathcal{O}_M}$  ( $\tilde{\phi}$  designates the approximation of  $\phi$ ) be the unknown vector in  $\mathbb{C}^N$ . An array representation  $\phi$  in the vector space of 2D complex-valued arrays  $\mathcal{M}_{n_x, n_y}(\mathbb{C})$  (stored according to the 2D grid) is identified with the reshaped vector  $\phi \in \mathbb{C}^N$ . For simplification, we do not precise the brackets  $\llbracket A \rrbracket$ , which means that  $A$  is the matrix operator related to a continuous operator  $A$ , considering the FFT approximation. This leads to  $\mathcal{O}(N \log N)$  operations to compute the application of the discrete operator since we use a 2D FFT.

For a discrete given function  $\phi \in \mathbb{C}^N$ , the total discrete energy  $E(\phi)$  is such that

$$E^\alpha(\phi) = E_k^\alpha(\phi) + E_p(\phi) + E_i(\phi), \quad (4.42)$$

where

$$E_k(\phi) := \frac{1}{2} \langle (-\nabla)^\alpha \phi, \phi \rangle, E_p(\phi) := \langle V \phi, \phi \rangle, \quad E_i(\phi) := \langle F(\rho), \mathbf{1} \rangle. \quad (4.43)$$

We set  $\mathbf{1} \in \mathbb{C}^N$  as the vector with components 1, while  $\|\phi\|_2$  is the discretization of the  $L^2(\mathbb{R}^d)$ -norm on the uniform grid using  $\langle \mathbf{u}, \mathbf{v} \rangle$ . Finally,  $\mathbf{u}$  and  $\mathbf{v}$  are two complex-valued functions on  $\mathcal{D}_M$ .

## 5. Numerical examples

### 5.1. The integer case ( $\alpha = 1$ )

We first start by considering the standard integer case ( $\alpha = 1$ ) with quadratic potential

$$V(\mathbf{x}) := \frac{1}{2}(x^2 + y^2), \quad (5.44)$$

and a cubic nonlinearity (1.3) with nonlinearity strength  $\beta = 5000$ . We fix  $a_x = a_y = 20$ , defining then the computational domain  $\mathcal{D} = [-20; 20]^2$ . We discretize in space with  $M = (256, 256)$  points. For all the numerical simulations, we consider the stopping criterion [4]:

$$\mathcal{E}_{\text{err}}^{\alpha, k} := |E^\alpha(\phi^{k+1}) - E^\alpha(\phi^k)| \leq \varepsilon_{\text{err}},$$

with  $\varepsilon_{\text{err}} = 10^{-10}$ . In the following, we denote by  $k_{\text{err}}$  the value of  $k$  for which the stopping criterion is fulfilled. The initial guess is chosen as the Thomas-Fermi (TF) approximation

$$\phi_0 = \frac{\phi_{\beta}^{\text{TF}}}{\|\phi_{\beta}^{\text{TF}}\|_2}, \text{ with } \phi_{\beta}^{\text{TF}} = \begin{cases} \sqrt{\frac{\lambda_{\beta}^{\text{TF}} - V(\mathbf{x})}{\beta}}, & \text{if } \lambda_{\beta}^{\text{TF}} > V(\mathbf{x}), \\ 0, & \text{otherwise,} \end{cases} \quad (5.45)$$

where the eigenvalue approximation  $\lambda_{\beta}^{\text{TF}}$  is given by  $\lambda_{\beta}^{\text{TF}} = \sqrt{\beta/\pi}$ . This leads to a numerical converged energy equals to 26.6372. For solving the linear systems arising in the semi-implicit schemes (related to (2.11), (3.29) and (3.41) after discretization by the pseudospectral scheme), we use the iterative BiCGSTAB (BS) solver. To this end, we force the BiCGSTAB tolerance to  $\epsilon_{\text{BS}} = 10^{-11}$  for a maximal number of BiCGSTAB iterations equal to 1000. For a given iteration  $k$ , the BiCGSTAB solver will need  $n_{\text{BS}}^k (\leq 1000)$  iterations. In the next Figures and Tables, we report the number of iterations of the various algorithms to converge according to the stopping criterion for  $\varepsilon_{\text{err}}$ . For the explicit algorithms (FE), the global number of iterations is given by  $k_{\text{err,g}} = k^{\text{err}}$ . However, for the semi-implicit schemes (BE), each iteration  $k$  requires a certain number of inner iterations  $n_{\text{BS}}^k$  of BiCGSTAB. Since we are using a pseudospectral scheme, then the global computational cost of the whole procedure is given :  $k_{\text{err,g}} = \sum_{k=0}^{k_{\text{err}}} n_{\text{BS}}^k$ . In addition, each iteration of BiCGSTAB corresponds to a matrix vector product computed in a matrix-free way *via* the FFT/iFFT.

We report in Table 1 the number of global iterations  $k_{\text{err,g}}$  to converge as well as  $k_{\text{err}}$  for the semi-implicit schemes with respect to the constant step  $\rho$ . For the explicit schemes, we observe that the Nesterov accelerated schemes (NAGF<sub>FE</sub><sup>2</sup> and NAGF<sub>FE</sub><sup>1</sup>) converge for larger steps  $\rho$  than the standard gradient flow (NC=Not Converged in the Table). This is probably due to the conditional stability properties of the schemes which should be investigated more. Even if we do not prospect this question here, this remark should open the path to using well-adapted time stepping techniques to increase the convergence rates of the explicit schemes. For the semi-implicit schemes, as it is known, their stability is better and we observe a convergence even for large steps, where the behavior of the NAGF<sub>BE</sub><sup>2</sup> scheme is slightly better than the standard GF<sub>BE</sub> scheme as long as the step is large. For smaller steps, we observe a deterioration of NAGF<sub>BE</sub><sup>2</sup> which is maybe due to a conditioning problem, in particular of the linear systems to resolve. This however should be possible to improve it in both schemes based on the use of adapted preconditioners [2]. Moreover, using regular grid finite-difference schemes may avoid this problem since then efficient direct solvers for structured meshes could be used for the linear systems solution. All the results in Table 1 are obtained without any restarting strategy. Indeed, we implemented all the previous restarting strategies but we have observed that they usually slow down the convergence rate of the algorithms. Therefore, we recommend to not using restarting strategies for the GPE. For the quadratic potential, the damping function  $\gamma(t)/3t$  is a well-suited choice (but any other choice given in Section 3.1 also leads to similar convergence results). We report on Figure 1 the behavior of the energy at each time step  $k$  of the algorithms ( $\rho = 10^{-2}$  except for GF<sub>FE</sub> where we have to consider  $\rho = 10^{-3}$  to get a converged solution). We observe in particular that the energy may not decay even if the convergence to the stationary state is obtained. These results (and others not reported here for the sake of conciseness) show that accelerated algorithms provide some very specific behavior for the GPE but still need to be understood, from the point of view of the properties but also based on the discretization schemes.

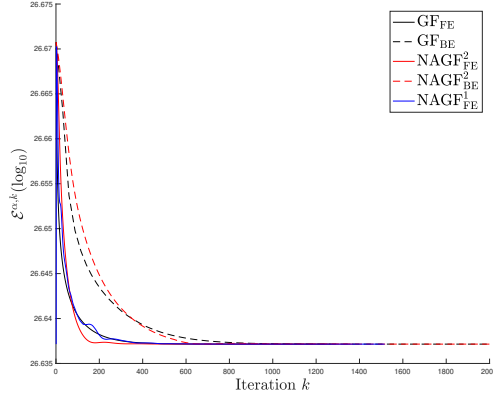


Figure 1: Example 1 ( $\alpha = 1$ ): energy convergence curves.

Scheme	GF <sub>FE</sub>	GF <sub>BE</sub>		NAGF <sub>FE</sub> <sup>2</sup>	NAGF <sub>BE</sub> <sup>2</sup>		NAGF <sub>FE</sub> <sup>1</sup>
Iter.	$k_{\text{err,g}}$	$k_{\text{err}}$	$k_{\text{err,g}}$	$k_{\text{err,g}}$	$k_{\text{err}}$	$k_{\text{err,g}}$	$k_{\text{err,g}}$
$\rho = 1 \times 10^{-0}$	NC	65	1701	NC	64	1670	NC
$\rho = 5 \times 10^{-1}$	NC	69	1745	NC	66	1617	NC
$\rho = 1 \times 10^{-1}$	NC	85	1828	NC	69	853	NC
$\rho = 5 \times 10^{-2}$	NC	91	1773	NC	111	700	351
$\rho = 1 \times 10^{-2}$	NC	257	2372	1371	1395	3399	1515
$\rho = 5 \times 10^{-3}$	NC	448	2808	2842	2097	4643	2859
$\rho = 1 \times 10^{-3}$	1266	1958	5369	24618	18403	26390	11020

Table 1: Number of iterations to converge for the various methods with constant steps  $\rho$ .

### 5.2. The general fractional case ( $0 < \alpha < 1$ )

In the case of a fractional exponent  $0 < \alpha < 1$ , it appears that only the methods based on the standard gradient flow (i.e. GF<sub>FE</sub> and GF<sub>BE</sub>) are able to provide a converged solution to the NLFSE. To illustrate this claim, we report on Figure 2 an example for the fractional exponent  $\alpha = 0.6$ . For both the GF<sub>FE</sub> (left) and NAGF<sub>FE</sub><sup>2</sup> (right) schemes, we fix the following parameters:  $\beta = 1000$ ,  $\mathcal{D} = [-20; 20]^2$ ,  $M = (256, 256)$  points and  $\Delta t = 0.5 \times 10^{-3}$ . On the left figure, we observe that the evolving solution converges towards a stationary solution after 51357 time steps. For the NAGF<sub>FE</sub><sup>2</sup> scheme, after enough iterations (about 40000 here), the solution diverges (see right figure). This appears even when refining the time step, and happens for any Nesterov-type scheme. More generally, based on Nesterov formulation, which leads to the nonlinear space-fractional wave equation

$$\partial_t^2 \phi + \gamma(t) \partial_t \phi + \frac{1}{2} (-\Delta)^\alpha \phi + V \phi + \beta |\phi|^2 \phi - \lambda(\phi) \phi = 0, \quad t \in [t_k, t_{k+1}), k \geq 0, \mathbf{x} \in \mathbb{R}^d, \quad (5.46)$$

we always obtain a solution with an energy  $\mathcal{E}^\alpha$  which tends to infinity, even when considering very small space and time discretization steps.

Based only on simulations, it is difficult to certify the origin of the behavior of the schemes and if they may be adapted to the fractional case. Among the possible reasons leading to the current restriction, one may think to the behavior of the solution of the wave equation (5.46) with normalization step, to unadapted time-discretization schemes or/and finally to the choice of the tuning parameters in Nesterov approach (initial guess, restarting or not, damping function) that may be helpful but are difficult to fix. Finally, in the case of standard gradient

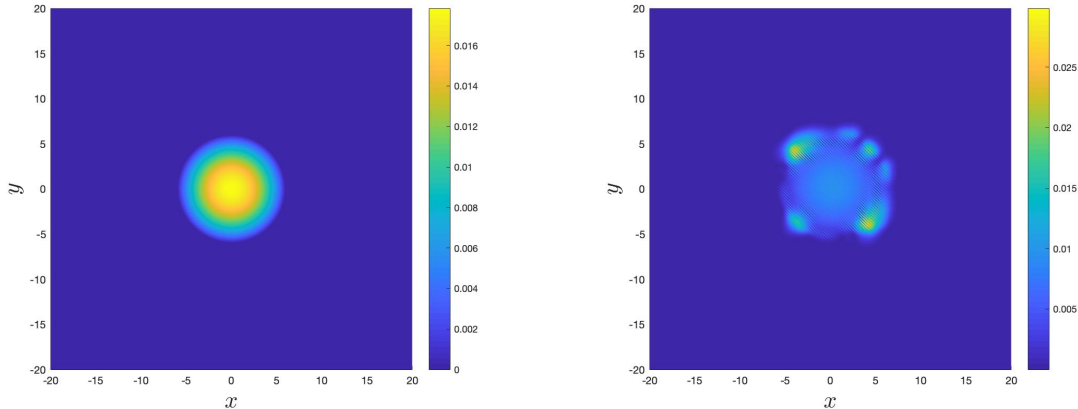


Figure 2: Example 2 ( $\alpha = 0.6$ ): left : converged density function  $|\phi|^2$  when using the explicit standard gradient flow. Right: divergence of the evolving density function  $|\phi|^2$  after enough time steps of the explicit Nesterov gradient flow for the same parameters as on the left figure.

flows, we remarked that it would be important to design robust preconditioners for the solution of the linear system arising in  $\text{GF}_{\text{BE}}$  since the conditioning and computational effort is strongly affected by the choice of the time step which needs to be taken small when  $\alpha$  tends to decay.

## 6. Conclusion

This paper presented the adaptation and assessment of the Nesterov accelerated gradient method to the computation of the ground state of the nonlinear Schrödinger equation, including the space fractional Schrödinger equation. Even if the algorithms are as simple as for the standard gradient flow method and show so interesting properties, many questions remain open concerning the possibility of improving their robustness in the case of strongly nonlinear quantum problems, which can in addition involve a space fractional operator that translates some nonlocal spatial interactions into the quantum system. In particular, for the fractional case, it was shown that one has to solve a nonlinear damped fractional Klein-Gordon equation with an additional projection step. The physical behavior of such a system that is derived from the Nesterov gradient flow formulation may be not adapted if the system has a growing energy. At this step, the proposed algorithms still need further developments following a few directions that were sketched in the paper, the most crucial being the statement of the system properties (energy behavior), the consistent and stable discretization of the resulting system and finally the choice of the tuning parameter related to the Nesterov gradient flow formulation.

**Acknowledgments.** X. ANTOINE and C. BENTAYAA sincerely thank the funding support of the University of Lorraine under the program LUE-UL.

## References

- [1] X. Antoine and R. Duboscq. GPELab, a Matlab toolbox to solve Gross-Pitaevskii equations I: Computation of stationary solutions. *Computer Physics Communications*, 185(11):2969–2991, 2014.
- [2] X. Antoine and R. Duboscq. Robust and efficient preconditioned Krylov spectral solvers for computing the ground states of fast rotating and strongly interacting Bose-Einstein condensates. *Journal of Computational Physics*, 258:509–523, 2014.

- [3] X. Antoine and R. Duboscq. Modeling and Computation of Bose-Einstein Condensates: Stationary States, Nucleation, Dynamics, Stochasticity. In Besse, C. and Garreau, J.C., editor, *Nonlinear Optical and Atomic Systems: at the Interface of Physics and Mathematics*, volume 2146 of *Lecture Notes in Mathematics*, pages 49–145. 2015.
- [4] X. Antoine, A. Levitt, and Q. Tang. Efficient spectral computation of the stationary states of rotating Bose-Einstein condensates by preconditioned nonlinear conjugate gradient methods. *Journal of Computational Physics*, 343:92 – 109, 2017.
- [5] X. Antoine, Q. Tang, and J. Zhang. On the numerical solution and dynamical laws of nonlinear fractional Schrödinger/Gross-Pitaevskii equations. *International Journal of Computer Mathematics*, 95(6-7):1423–1443, 2018.
- [6] X. Antoine, Q. Tang, and Y. Zhang. On the ground states and dynamics of space fractional nonlinear Schrödinger/Gross-Pitaevskii equations with rotation term and nonlocal nonlinear interactions. *Journal of Computational Physics*, 325:74–97, 2016.
- [7] X. Antoine, Q. Tang, and Y. Zhang. A preconditioned conjugated gradient method for computing ground states of rotating dipolar Bose-Einstein condensates via kernel truncation method for dipole-dipole interaction evaluation. *Communications in Computational Physics*, 24(4):966–988, 2018.
- [8] W. Bao and Y. Cai. Mathematical theory and numerical methods for Bose-Einstein condensation. *Kinetic and Related Models*, 6(1):1–135, 2013.
- [9] W. Bao, I. Chern, and F. Lim. Efficient and spectrally accurate numerical methods for computing ground and first excited states in Bose-Einstein condensates. *Journal of Computational Physics*, 219(2):836–854, DEC 10 2006.
- [10] W. Bao and X. Dong. Analysis and comparison of numerical methods for the Klein–Gordon equation in the nonrelativistic limit regime. *Numerische Mathematik*, 120:189–229, 2012.
- [11] W. Bao and Q. Du. Computing the ground state solution of Bose-Einstein condensates by a normalized gradient flow. *SIAM Journal on Scientific Computing*, 25(5):1674–1697, 2004.
- [12] W. Bao, S. Jiang, Q. Tang, and Y. Zhang. Computing the ground state and dynamics of the nonlinear Schrödinger equation with nonlocal interactions via the nonuniform FFT. *Journal of Computational Physics*, 296:72–89, 2015.
- [13] W. Bao and W. Tang. Ground-state solution of Bose-Einstein condensate by directly minimizing the energy functional. *Journal of Computational Physics*, 187(1):230–254, 2003.
- [14] W. Bao and X. Zhao. Comparison of numerical methods for the nonlinear Klein-Gordon equation in the nonrelativistic limit regime. *Journal of Computational Physics*, 398:108886, 2019.
- [15] D. Baye and J.M. Sørensen. Resolution of the Gross-Pitaevskii equation with the imaginary-time method on a Lagrange mesh. *Physical Review E*, 82(5), 2010.
- [16] M. Caliari, A. Ostermann, S. Rainer, and M. Thalhammer. A minimisation approach for computing the ground state of Gross-Pitaevskii systems. *Journal of Computational Physics*, 228(2):349–360, 2009.



- [17] M. Caliari and S. Rainer. GSGPEs: A MATLAB code for computing the ground state of systems of Gross-Pitaevskii equations. *Computer Physics Communications*, 184(3):812–823, 2013.
- [18] J.L. Chiff. *The Laplace transform: Theory and Applications*. Springer, Undergraduate Texts in Mathematics, 1999.
- [19] I. Danaila and P. Kazemi. A new Sobolev gradient method for direct minimization of the Gross-Pitaevskii energy with rotation. *SIAM Journal on Scientific Computing*, 32(5):2447–2467, 2010.
- [20] I. Danaila and B. Protas. Computation of ground states of the Gross-Pitaevskii functional via Riemannian optimization. *SIAM Journal on Scientific Computing*, 39(6):B1102–B1129, 2017.
- [21] C.M. Dion and E. Cancès. Ground state of the time-independent Gross-Pitaevskii equation. *Computer Physics Communications*, 177(10):787–798, NOV 15 2007.
- [22] W. Liu and Y. Cai. Normalized gradient flow with Lagrange multiplier for computing ground states of Bose-Einstein condensates. *SIAM Journal on Scientific Computing*, 43(1):219–242, 2021.
- [23] Y. Nesterov. Gradient methods for minimizing composite objective function. Core discussion paper, 2007. [http://www.ecore.be/DPs/dp\\_1191313936.pdf](http://www.ecore.be/DPs/dp_1191313936.pdf).
- [24] Y. Nesterov. A method of solving a convex programming problem with convergence rate  $\mathcal{O}(1/k^2)$ . In *Doklady Akademii Nauk*, volume 269, pages 543–547. Russian Academy of Sciences, 1983.
- [25] Y. Nesterov. *Introductory Lectures on Convex Optimization: a basic course*. Kluwer Academic Publishers, Massachusetts, 2004.
- [26] B. O’Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15:715–732, 2013.
- [27] J. Royer. Energy decay for the Klein-Gordon equation with highly oscillating damping. *Annales Henri Lebesgues*, 120:297–312, 2018.
- [28] W. Su, S. Boyd, and E.J. Candès. A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *Journal of Machine Learning Research*, 17:1–43, 2016.
- [29] B. Wang, T.M. Nguyen, A.L. Bertozzi, R.G. Baraniuk, and S.J. Osher. Scheduled restart momentum for accelerated stochastic gradient descent. *arXiv: 2002. 10583*, 2020.
- [30] X. Wu, Z. Wen, and W. Bao. A regularized Newton method for computing ground states of Bose-Einstein condensates. *Journal of Scientific Computing*, 73:303–329, 2017.