



**HAL**  
open science

## The Steepest Slope toward a Quantum Few-body Solution

Paolo Recchia, Debabrota Basu, Mario Gattobigio, Christian Miniatura,  
Stéphane Bressan

► **To cite this version:**

Paolo Recchia, Debabrota Basu, Mario Gattobigio, Christian Miniatura, Stéphane Bressan. The Steepest Slope toward a Quantum Few-body Solution. *Few-Body Systems*, 2024, 65 (4), pp.102. 10.1007/s00601-024-01965-7 . hal-04672894

**HAL Id: hal-04672894**

**<https://hal.science/hal-04672894v1>**

Submitted on 19 Aug 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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# The Steepest Slope toward a Quantum Few-body Solution: Gradient Variational Methods for the Quantum Few-body Problem

Paolo Recchia<sup>1</sup>, Debabrota Basu<sup>2</sup>, Mario Gattobigio<sup>3</sup>, Christian Miniatura<sup>3, 4</sup>, and  
Stéphane Bressan<sup>1, 5</sup>

<sup>1</sup>School of Computing, National University of Singapore, 117417, Singapore

<sup>2</sup>Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 - CRISTAL, F-59000 Lille, France

<sup>3</sup>Université Côte d'Azur, CNRS, Institut de Physique de Nice, 17 rue Julien Lauprêtre, 06200 Nice, France

<sup>4</sup>Centre for Quantum Technologies, National University of Singapore, 117543, Singapore

<sup>5</sup>International Research Laboratory on Artificial Intelligence, CNRS International Laboratory, 2955, Singapore

## Abstract

Quantum few-body systems are deceptively simple. Indeed, with the notable exception of a few special cases, their associated Schrödinger equation cannot be solved analytically for more than two particles. One has to resort to approximation methods to tackle quantum few-body problems. In particular, variational methods have been proposed to ease numerical calculations and obtain precise solutions. One such method is the Stochastic Variational Method, which employs a stochastic search to determine the number and parameters of correlated Gaussian basis functions used to construct an ansatz of the wave function. Stochastic methods, however, face numerical and optimization challenges as the number of particles increases.

We introduce a family of gradient variational methods that replace stochastic search with gradient optimization. We comparatively and empirically evaluate the performance of the baseline Stochastic Variational Method, several instances of the gradient variational method family, and some hybrid methods for selected few-body problems. We show that gradient and hybrid methods can be more efficient and effective than the Stochastic Variational Method. We discuss the role of singularities, oscillations, and gradient optimization strategies in the performance of the respective methods.

**Keywords:** Stochastic Variational Method, Gradient-Based Optimization, Few-body systems

## 1 Introduction

The instances of the Schrödinger equation that govern the dynamics of quantum few-body systems typically lack analytical solutions when more than two particles are involved, even when the interaction potentials are known. Physicists have devised numerical approximation methods to address quantum few-body problems. Among these methods, the most prevalent

are variational methods [Cohen-Tannoudji et al., 2019], which leverage the Variational Principle [Griffiths and Schroeter, 2018], also referred to as Ritz theorem [Suzuki et al., 1998]. The Variational Principle states that the expectation value of the Hamiltonian for any state of a quantum system is greater than or equal to the ground state energy. The ground state energy is precisely the expectation value of the Hamiltonian in the ground state. This expectation value can be expressed as the Rayleigh quotient involving the Hamiltonian and the wave function of the state.

One state-of-the-art variational method is the Stochastic Variational Method [Suzuki et al., 1998]. As its name suggests, this variational method employs a stochastic search to determine the parameters of correlated Gaussian basis functions used to construct an ansatz of the wave function. However, as the number of particles increases, the Stochastic Variational Method and its variants, such as stochastic variational methods using other families of basis functions, face numerical and optimization challenges.

Yasuyuki Suzuki and Kálmán Varga, discussing the stochastic search of the Stochastic Variational Method in their 1998 reference textbook [Suzuki et al., 1998], wrote “*A skeptic may say that one should, instead of the above gambling method, try a deterministic parameter search [...]*”.

We propose revisiting this claim in light of modern gradient descent methods, like the Adam algorithm [Kingma and Ba, 2014], developed for training machine learning models. Specifically, we introduce a family of gradient variational methods that replace stochastic search with gradient optimization. We comparatively and empirically evaluate the performance of the baseline Stochastic Variational Method, several instances of the gradient variational method family, and some hybrid methods for selected few-body problems. We show that gradient and hybrid methods can be more efficient and effective than the state-of-the-art Stochastic Variational Method. We discuss the role of singularities, oscillations, and gradient optimization strategies in the performance of the respective methods.

This work primarily focuses on few-body systems composed of identical bosons. Under certain conditions, these systems exhibit universal properties, such as the intriguing Efimov effect [Efimov, 1970]. Moreover, systems of identical bosons are effectively represented by Gaussian-shaped potentials [Recchia et al., 2022a, Kievsky et al., 2021]. Specifically, given that clusters of helium atoms serve as natural candidates for Efimov Physics [Kolganova et al., 2011], we use these systems for testing purposes.

The paper is organized as follows. In Section 2, we delineate the core principles of the Stochastic Variational Method as initially introduced in references [Kukulín and Krasnopol’sky, 1977, Suzuki et al., 1998]. Moving on to Section 3, and after a short review of the current literature about methods to tackle few-body systems, we illustrate alternative optimization techniques suitable for nonlinear parameters. Section 4 introduces our proposal to embed certain gradient-based optimization approaches into the framework of the Stochastic Variational Method. Lastly, in Section 5, we compare and illustrate the performance of these optimization schemes with the original Stochastic Variational Method. We then conclude and offer some perspectives.

## 2 Background: the Stochastic Variational Method

In this section, we outline the key features, advantages, and drawbacks of the state-of-the-art Stochastic Variational Method. Throughout the paper, we follow the notations and definitions provided in [Suzuki et al., 1998] unless otherwise specified.

## 2.1 The variational method for solving the Schrödinger equation

Let us consider few-body systems of interacting particles with two- and three-body interactions. The Schrödinger equation governing such systems can be formulated as a linear algebra problem by introducing the linear and Hermitian Hamiltonian operator (1), where  $\hbar$  is the Planck constant,  $N$  is the number of particles,  $a$ ,  $b$ , and  $c$  are the indices of the different particles,  $m_a$  is the mass of a particle,  $V$  and  $W$  are the two-body and a three-body potentials, respectively, and  $\mathbf{r}_a$  is the position of a particle

$$H = \sum_{a=1}^N -\frac{\hbar^2}{2m_a} \frac{\partial^2}{\partial \mathbf{r}_a^2} + \sum_{b>a=1}^N V(\mathbf{r}_a, \mathbf{r}_b) + \sum_{c>b>a=1}^N W(\mathbf{r}_a, \mathbf{r}_b, \mathbf{r}_c), \quad (1)$$

The Hilbert space of the Hamiltonian is infinite-dimensional with both a continuous and discrete spectrum of eigenvalues. However, we focus on finding the bound states of the Hamiltonian, which allows us to consider the Schrödinger equation as the eigenvalue problem (2) with a discrete spectrum of eigenvalues  $E_n$  and  $\Phi_n$  eigenfunctions,  $E_1$  and  $\Phi_1$  are the ground state energy and wavefunction respectively.

$$H\Phi_n = E_n\Phi_n. \quad (2)$$

The Ritz theorem [Suzuki et al., 1998], also called variational principle [Cohen-Tannoudji et al., 2019, Griffiths and Schroeter, 2018], states that, for an arbitrary function  $\phi$  of the state space, the expectation value of  $H$  in the state  $\phi$  is larger than or equal to the ground state energy, as expressed in Equation (3). In addition, the theorem states that  $E[\phi]$  equals  $E_1$  if and only if  $\phi$  is an eigenstate of  $H$  with eigenvalue  $E_1$

$$E[\phi] = \frac{\langle \phi | H | \phi \rangle}{\langle \phi | \phi \rangle} \geq E_1. \quad (3)$$

Ritz variational method [Suzuki et al., 1998] leverages Ritz theorem. To find the lowest possible value approaching the ground state energy  $E_1$ , we need to minimize the functional in (3) with respect to the trial function  $\phi$ , often called *ansatz*. The minimization is achieved by imposing the stationary condition of (3). Typically, the ansatz takes the form of a sum over basis elements  $\psi_k$  as expressed

$$\psi \approx \phi = \sum_{k=1}^K c_k \psi_k. \quad (4)$$

With this ansatz and in conjunction with the stationary condition, the Schrödinger Equation (2) transforms into a linear system of  $K$  equations in terms of the coefficients  $c_k$  (more details are in Appendix C). This system can be solved as a generalized finite eigenvalue problem, often referred to as the secular equation

$$\mathcal{H}\mathcal{C} = \mathcal{S}\mathcal{C}E, \quad (5)$$

where  $\mathcal{H}ij = \langle \psi_i | H | \psi_j \rangle$  and  $\mathcal{S}ij = \langle \psi_i | \psi_j \rangle$  represent the matrix elements of the Hamiltonian and overlap on the chosen basis, respectively. Here,  $E = \text{diag}(E_i)_{i=1, \dots, K}$  represents the eigenvalues (in particular, the lowest  $E_1$  is the ground state energy approximation we are interested in), and  $\mathcal{C}$  denotes the matrix of eigenvectors with columns being the coefficients  $c_k$  for each eigenvalue.

Several important aspects of variational methods should be emphasized. Firstly, the approximation given by (4) reduces the eigenvalue problem (5) to a finite-dimensional form.

Another crucial theorem asserts that increasing the number of basis elements enhances the accuracy of computed eigenvalues and consequently the ground state energy [Suzuki et al., 1998]. Secondly, various diagonalization techniques can be applied to solve the secular equation (5). In this study, we adopt the method detailed in [Piela, 2006], as described in Appendix C, simplifying the calculation by converting the generalized eigenvalue problem into a standard one. Thirdly, the effectiveness of variational methods heavily relies on the choice of the ansatz, particularly the family of basis functions denoted as  $\psi_k$ .

For the sake of completeness, it is important to note that the generalized Ritz theorem empowers the variational methods for the approximation of the excited states of the system as well. Specifically, solving the secular equation (5) provides an estimate of the energy spectrum of the system under consideration [Suzuki et al., 1998].

## 2.2 Stochastic optimization and Gaussian basis in the Stochastic Variational Method

Like any other variational technique, the Stochastic Variational Method relies on the Ritz theorem. A key feature of the Stochastic Variational Method is the choice of the ansatz, which takes the form of a sum over a system of non-orthonormal Gaussian basis functions in configuration space that is parametrized by a finite number of continuous nonlinear  $\alpha_{ab}^{(k)}$  parameters

$$\langle \mathbf{r}_1, \dots, \mathbf{r}_N | \psi_k \rangle = \exp \left[ -\frac{1}{2} \sum_{b>a=1}^N \left( \frac{\mathbf{r}_a - \mathbf{r}_b}{\alpha_{ab}^{(k)}} \right)^2 \right]. \quad (6)$$

We refer to this choice as the *Gaussian basis*. To minimize the functional (3), we observe, based on (4) and (6), that two types of parameters need to be determined: the set of linear variational parameters represented as  $c_k$ , and the set of nonlinear variational parameters denoted as  $\alpha_{ab}^{(k)}$  (these are the coefficients in the Gaussian exponents). Determining the linear parameters  $c_k$  is a straightforward task involving solving a linear problem, specifically the eigenvalue problem described in (5). On the other hand, finding the nonlinear parameters can be quite challenging, especially in cases where various local minima influence the parameter space. The Stochastic Variational Method adopts a stochastic search to address this challenge. Indeed, introducing random steps in the process allows for exploring the entire parameter space, including regions with local minima. The fundamental concept behind the original Stochastic Variational Method is to perform multiple trial evaluations at each step when a new basis element is added. Subsequently, only the trial returning the lowest eigenvalue is chosen [Kukulín and Krasnopol'sky, 1977].

The effectiveness of the Gaussian basis in describing various types of few-body systems can be attributed to two main reasons. First, under given conditions, *the Gaussian basis exhibits the property of being overcomplete*, meaning that removing a countable subset of functions does not affect the completeness of the system (more details about the requirements for completeness are discussed and proved in [Klahn and Bingel, 1977, Shaw, 2020]). This characteristic enhances the method's robustness and reliability; in fact, rejecting a countable subset of basis elements is not a significant issue. Secondly, the Gaussian basis simplifies the computation of certain integrals arising during the evaluation of matrix elements in the eigenvalue problem (5), as a significant number of Gaussian integrals can be analytically evaluated.

However, the overcompleteness in the Gaussian basis introduces a significant drawback. When dealing with the generalized eigenvalue problem (5), the computation of the inverse of the

overlap matrix  $\mathcal{S}$  becomes necessary (see appendix C for further details). Given that the basis vectors are overcomplete, some vectors may be linearly dependent on each other, resulting in very small eigenvalues of the overlap matrix  $\mathcal{S}$ . This, in turn, *contributes to numerical instabilities when attempting to calculate the inverse of  $\mathcal{S}$* . In the Stochastic Variational Method, these singularities are eliminated whenever a newly proposed basis yields them. Subsequently, a fresh random basis is suggested. We emphasize that various stopping criteria can be suggested to stop the iterations in the Stochastic Variational Method. For example, the algorithm can conclude with a given number of basis  $K$ , or when a specified CPU runtime is reached or when the relative difference between the last two energy values attained is below a given threshold  $\varepsilon$

$$\left| \frac{E_k - E_{k-1}}{E_k} \right| \leq \varepsilon. \quad (7)$$

In the next sections, as we mainly focus on a performance comparative analysis, we usually stop the Stochastic Variational Method after a given CPU runtime is reached. Another important aspect of the baseline Stochastic Variational Method pertains to the interval  $(0, \alpha_{max}]$  within which the nonlinear variational parameters are randomly chosen. As depicted in Equation (6), the variational parameters  $\alpha_{ab}^{(k)}$  share the same dimensions as the interparticle distance. Consequently, we can deduce a reasonable range for selecting the nonlinear parameters based on a preliminary understanding of the system under consideration. Nevertheless, this information is often unavailable beforehand, presenting an additional challenge in implementing the Stochastic Variational Method.

The algorithm of the Stochastic Variational Method, as originally proposed in [Kukulín and Krasnopol’sky, 1977], is summarized in Algorithm 1, in Appendix A.

## 3 Related Work

### 3.1 Overview of methods for solving the quantum few-body problem

Numerous methods have been developed to solve quantum few-body systems. The Stochastic Variational Method is one among them. Despite their variety, we can classify them into two big categories: *non-variational methods* and *variational methods*. These categories can be further divided into subcategories. Non-variational methods include the *exact method* and *numerical integration*. Variational methods can be further divided into *variational methods with an orthonormal basis*, *variational methods with a non-orthonormal basis*, and *Variational Monte Carlo methods*.

The *Faddeev Method* [Faddeev, 1965] provides an exact technique for solving the Schrödinger equation in three-body systems. Initially developed for studying nuclear physics scattering problems, this method is also applicable in identifying bound states within such systems. With specific classes of interactions, the Faddeev method solves three coupled single-particle Schrödinger equations, termed Faddeev components or equations, rather than the full Schrödinger equation. The appropriate boundary conditions applied to each Faddeev component yield the corresponding energy eigenvalue for the entire three-body system. Extensions of the Faddeev equations to systems beyond three-body are referred to as Faddeev-Yakubovsky methods [Faddeev and Merkuriev, 1993]. However, scalability remains a primary challenge for this approach. The number of Faddeev components, and thus the differential equations to solve, grows factorially with the number of particles, making numerical computations quickly

impractical. Presently, the Faddeev-Yakubovsky method has found applications in the literature for four-body systems [Lazauskas, 2018, Braaten and Hammer, 2006].

Among the *numerical integration* techniques, the most common is the Numerov integration [Numerov, 1927]. It is a numerical method to solve ordinary differential equations of second order in which the first-order term does not appear, like the Schrödinger equation. Though it allows solving the Schrödinger equation with a wide variety of interactions, its application is restricted to low-dimensional cases: the current literature extends the Numerov integration in a grid of up four dimensions [Kuenzer et al., 2016, Kuenzer and Hofer, 2019].

*Variational methods* require a choice of basis functions, which can be either orthonormal or non-orthonormal. *Hyperspherical Harmonics* are a widely studied and applied example of orthonormal and complete basis functions. Although this basis was first introduced in 1935 by Zernike and Brinkman [Zernike and Brinkman, 1935] and reintroduced 25 years later by Delves [Delves, 1958] and Smith [Smith, 1962], it has only been in recent decades that it has seen widespread use and development [Kievsky et al., 2008]. The primary advantage of Hyperspherical Harmonics is their property of being eigenfunctions of the kinetic energy operator. Additionally, because Hyperspherical Harmonics form an orthonormal basis, the overlap matrix  $S$  in (5) becomes the identity matrix, simplifying the generalized eigenvalue problem in (5) to a standard eigenvalue problem. This eliminates the need to calculate the inverse of the overlap matrix, thereby avoiding the singularity issues discussed in Section 2.2. However, unlike the Gaussian basis, the matrix elements of the potential term of the Hamiltonian with Hyperspherical Harmonics cannot be evaluated analytically, necessitating numerical integration techniques. For instance, calculations based on Chebyshev and Laguerre weights and points [Abramowitz and Stegun, 1964] are used in [Kievsky et al., 2008]. Moreover, the number of Hyperspherical Harmonics required for convergence is typically larger than that of Gaussians [Marcucci et al., 2020]. This large number of basis elements makes the diagonalization of the Hamiltonian matrix numerically challenging, and standard diagonalization techniques are often inadequate. In the Hyperspherical Harmonics method, the iterative Lanczos algorithm is widely used to address this issue [Lanczos, 1950, Marcucci et al., 2020].

In contrast to the variational methods with orthonormal basis, the other family of variational methods, such as the Stochastic Variational Method [Suzuki et al., 1998], compute the nonlinear parameters of the chosen non-orthonormal basis via random trial and error. In literature, the Gaussian basis is considered state-of-the-art for the Stochastic Variational Method, and other bases have hardly been considered. Alternative families of bases, like exponentials and wavelets, are currently under investigation. For optimising the nonlinear parameters in those bases, the same arguments of the Stochastic Variational Method apply.

The *Variational Monte Carlo* method always relies on the variational principle, particularly focusing on the functional described by equation (3). By defining the local energy  $E_{\text{loc}}(\mathbf{r}) = \frac{H\phi(\mathbf{r})}{\phi(\mathbf{r})}$ , we can express the functional (3) as the average of  $E_{\text{loc}}(\mathbf{r})$  where the normalized square of the ansatz  $|\phi(\mathbf{r})|^2 / \int d\mathbf{r} |\phi(\mathbf{r})|^2$  represents a probability distribution. We can then calculate the functional (3) via Monte Carlo techniques where the points for the integration are sampled from the probability distribution  $|\phi(\mathbf{r})|^2 / \int d\mathbf{r} |\phi(\mathbf{r})|^2$ . Various sampling techniques can be employed, such as Metropolis sampling [Metropolis et al., 1953, Hastings, 1970], importance sampling through simulation of the Fokker-Planck differential equation [Hammond et al., 1994], and correlated sampling [Bär and Bopp, 1988]. Once the most suitable sampling technique is identified, the variational parameters of the ansatz  $\phi$  are primarily optimized by minimizing the expectation value of the local energy. The stochasticity due to the sampling introduces a statistical uncertainty in the energy expectation value, which can be higher compared to other



variational techniques. Although this uncertainty can be mitigated by increasing the sampling points, it may limit the method’s performance.

## 3.2 Nonlinear optimization in the Stochastic Variational Methods

Besides the singularities’ issue introduced in Section 2.2, another common issue of the Stochastic Variational Method concerns its computational cost, especially when the number  $N$  of particles increases. Firstly, the number of non-linear parameters  $\alpha_{ab}$  in the ansatz in Equation (6) increases quadratically with  $N$ , like  $N(N-1)/2$ . Secondly, to efficiently and effectively converge as the space of parameters increases, a larger number  $K$  of basis functions is also required. Thirdly, to solve the secular equation (5), one must calculate the determinant and the inverse of a  $K \times K$  matrix, which scales like  $O(K^3)$  (Gauss-Jordan). Lastly, wavefunction symmetrization is required for systems of identical Bosons and Fermions, and  $N!$  of the previously mentioned operations is necessary. The main bottleneck regards the nonlinear optimization of the variational parameters, which is hard to accomplish, especially when the space is affected by multiple local minima. Nonlinear optimization is a broad field [Aragón et al., 2019], and a comprehensive review of all existing approaches is beyond the scope of this paper. In this brief section, we offer an overview of two major categories: *gradient-free optimization* and *gradient-based optimization*.

### 3.2.1 Gradient-free optimization methods

Gradient-free optimization methods do not require information about the gradient of the objective function. They can be useful when this information is unavailable or the functions are not smooth. We can classify them into three subcategories.

*Deterministic derivative-free optimizations*, such as Nelder-Mead [Nelder and Mead, 1965], Powell [Powell, 1964], that move on the slope of the function following a given strategy. These approaches have different drawbacks. First, their high computational cost makes them not scalable for systems with increasing particles. Second, they are sensitive to a given starting point, always reaching the same minimum from the same initial condition. Third, the solutions found by these approaches may not be the global ones since they can easily get stuck in local minima, especially when the size of the Hilbert space increases.

*Metaheuristic approaches*, for example, Genetic Algorithm [Fogel, 1998] and Simulated Annealing [Kirkpatrick et al., 1983], are other advanced derivative-free techniques to find the global optimum of a given function. Even though they have never been appropriately tested for optimizing the nonlinear variational parameters, both these methods were born to handle tasks on a discrete space [Bianchi et al., 2009]. In addition, both suffer from the local minima issues as the deterministic approaches.

While the abovementioned optimization methods may seem more sophisticated than the *stochastic optimization*, state-of-the-art research indicates that random trials often yield greater success by circumventing the common occurrence of getting stuck in local minima [Suzuki et al., 1998].

### 3.2.2 Gradient-based optimization methods

In the current literature, no systematic benchmarking of *gradient-based optimizations* applied to variational methods for quantum few-body problems exists. However, in past years, gradient-based optimization methods have become widespread due to the increasing popularity of machine learning models. When the objective function to optimize is smooth, the gradient-based methods



proved to be reliable techniques for efficiently finding the minima of the function thanks to the calculation of its gradient. Gradient Descent [Curry, 1944] is the foundational optimization algorithm. It updates the parameters in the opposite direction of the gradient of the objective function. Some variants of gradient descent help speed the algorithm’s convergence in different contexts. Gradient descent with momentum [Rumelhart et al., 1986] stores the past gradients at each iteration and determines the next update as a linear combination of these past gradients. Gradient descent with Root Mean Square Propagation (RMSProp) [Hinton et al., 2012] adjusts the learning rate by dividing it by a factor that depends on the moving average of the past squared gradients. Gradient descent with Adaptive Momentum Estimation (Adam) [Kingma and Ba, 2014] combine the idea in Momentum and RMSProp. It adapts the learning rates by employing an exponential forgetting factor to compute running averages for both gradients and the squared of gradients. This latter gradient optimization method, in particular, has demonstrated superior performance compared to previous ones across different datasets and various machine learning models, as indicated in [Kingma and Ba, 2014] (for a more detailed outline about the gradient methods, refer to [Ruder, 2016]). The current paper aims to evaluate and compare the performance of gradient-based optimization techniques in the framework of variational methods applied to quantum few-body problems. Following initial testing, we have identified Adam as the most efficient gradient-based optimization. This finding suggests that Adam could serve as a promising alternative for exploring the nonlinear parameters in the Gaussian basis of the Stochastic Variational Method.

It is crucial to recognize that during the approach to a local or global minimum, gradient-based optimizations can lead to oscillations in the loss function (representing the system energy in this context). Such oscillations are well-known across various gradient methods, including Adam [Goodfellow et al., 2016, Ruder, 2016]. Recent research has delved deeper into their underlying causes. For instance, in studies such as those [Choo et al., 2020, Arora et al., 2022, Ahn et al., 2022], the authors have investigated various neural network models, loss functions, and datasets. They have highlighted the relationship between the shape of the loss function near a local minimum and the selection of the learning rate as significant factors contributing to oscillations and potential divergences. Specifically, when the product of the learning rate and the sharpness of the loss function (the largest eigenvalue of its Hessian) exceeds two, gradient optimization becomes unstable. This instability manifests as oscillations in the loss function, which can either lead to divergence or transition into what [Ahn et al., 2022] terms the “unstable convergence” (or “edge of stability” as described in [Choo et al., 2020]). At this edge of stability, although the loss function exhibits non-monotonic behavior and oscillates, optimization can still consistently decrease over a long timescale, and as a net result, gradient descent effectively continues to optimize the loss function.

Understanding the causes of these oscillations requires insight into the shape of the loss function, particularly its sharpness, which is not readily known beforehand, especially in higher-dimensional spaces. This type of analysis exceeds the scope of the current paper. However, as further discussed in Section 5.5, in our context, oscillations primarily stem from selecting a high and multiplicatively increasing learning rate. This choice is often made to accelerate convergence and has proven beneficial in various machine learning models [Smith and Topin, 2019]. When the learning rate is high, the product of the sharpness and the learning rate may exceed two, potentially causing the optimization to overshoot the minimum, resulting in oscillations, unstable convergence, and eventual divergences. Several strategies can mitigate these issues, such as engineering a decreasing learning rate as optimization approaches the minimum, gradient clipping [Mikolov et al., 2012, Pascanu et al., 2013], cyclic learning rates

as proposed in [Smith, 2017], or employing a decaying learning rate with cosine annealing as suggested in [Loshchilov and Hutter, 2016]. However, it’s important to acknowledge that while these techniques can prevent oscillation, they may also potentially slow the optimization process.

## 4 Methodology

In this work, we integrate gradient-based optimization (specifically Adam) within the variational methods framework. In practice, we consider the Gaussian basis as in the Stochastic Variational Method. The objective function to minimize is the ground state energy of the Hamiltonian in the secular equation (5), with respect to the nonlinear parameters in the Gaussian basis (6). The gradient-based optimization enables us to devise a family of optimization strategies, broadly categorized into *gradient variational methods* and *hybrid methods* where random search is mixed with gradient optimization. Modifying the hyperparameters within the gradient variational methods yields a spectrum of different optimization schemes. Following a general overview of the proposed gradient variational methods, we describe their particular instances with specific hyperparameters in the subsequent sections.

### 4.1 Gradient variational methods

In contrast to the baseline Stochastic Variational Method, which optimizes individual basis elements separately, in our proposed gradient variational methods, we can fix the basis size (referred to as a batch) and optimize the entire batch of basis elements using Adam. Following the optimization of this batch, two viable options are possible.

The first, which we refer to as the *static gradient variational methods*, consists in iteratively adding and optimizing batches of basis functions of predefined sizes while committing to the optimized values of the previous batches. The iterative nature of this method is reminiscent of the Stochastic Variational Method iterative addition and optimization of one basis function. The static gradient variational methods are outlined in Algorithm 2 in Appendix A. The second option removes the constraint of committing the previous batches. It adds the flexibility of optimizing all the previous batches simultaneously with the new proposed batch. The approach is outlined in Algorithm 3 in Appendix A; we call them *dynamic gradient variational methods*.

Similar to the Stochastic Variational Method, various stopping criteria can be proposed to halt the proposed gradient variational methods (for instance, CPU time, the difference between the energies computed in the last batches (7), etc.). The primary goal of this paper is to compare the performances of various methods; therefore, we usually conclude the algorithm once a specified CPU runtime is reached.

As in the Stochastic Variational Method, singularities are not eliminated in gradient variational methods. To address the ill conditions within these methods, we observe that a slight modification of the calculated gradient allows us to bypass singularities without significantly jeopardizing the optimization process. We chose to modify the gradient by generating a new one from a normal distribution with a mean value equal to the previously calculated gradient and a standard deviation of  $\sigma = 0.1$ .

In gradient variational methods the sizes of the batches are predetermined, and we must consider them as hyperparameters. Changing the sizes of the batches leads to different instances of the gradient variational methods. In the present paper, we evaluate and test three instances of the gradient variational methods by selecting specific values for the batches’ size.

The *Baseline Gradient Variational Method (GVM)* is the simplest gradient variational method, achieved by selecting to optimize one fixed batch of  $K$  basis elements. The pseudocode is outlined in Algorithm 4 in Appendix A.

Within the family of static gradient variational methods, the simplest involves setting each batch size equal to one. Following our chosen terminology, we refer to this approach as the *Static Gradient Variational Method (Static-GVM)*, detailed in Algorithm 5 in Appendix A. This method closely resembles the baseline Stochastic Variational Method, wherein each basis element is optimized individually as they are proposed, with the only distinction that the optimization is performed by Adam rather than stochastically. This approach introduces a new hyperparameter, the fixed value of the gradient optimization iterations.

Also within the dynamic gradient variational methods, we opt for testing the simplest where the batches’ size is set to one. The same arguments applied to the previous method are relevant here as well. First, we decided to designate it the *Dynamic Gradient Variational Method (Dynamic-GVM)*. Second, the number of optimization iterations must be predetermined and regarded as a hyperparameter (details in Algorithm 6 in Appendix A).

## 4.2 Hybrid method SVM-GVM

This method combines the baseline Stochastic Variational Method with the baseline Gradient Variational Method. Initially, we employ the Stochastic Variational Method for a specified number of basis elements (considered as a warm-up procedure for the basis set). Subsequently, we optimize the entire obtained set as in the Gradient Variational Method with Adam. This method is detailed in Algorithm 7 in Appendix A.

Similar to the Gradient Variational Method (GVM), the hyperparameter in this method is the basis size, and among the various stopping criteria, the primary choice is the CPU runtime.

# 5 Performance evaluation and analysis

## 5.1 Exeprimental setup

### 5.1.1 The physical systems

To evaluate the convergence and performance of the methods detailed in the previous Section 4, we focus on the ground state energy of few-body systems composed of identical Bosons, particularly clusters of He up to six-body. These systems have been extensively studied and are well-replicated with potentials of Gaussian shape, as indicated in previous works [Bazak et al., 2020, Gattobigio et al., 2011, Recchia et al., 2022a, Recchia et al., 2022b]. In temperature units, the energy scale associated to Helium atom is  $\hbar^2/(ma_0^2) = 43.281\,31\text{ K}$ , where  $a_0$  is the Bohr radius,  $\hbar$  the reduced Planck’s constant and  $m$  the Helium atom mass. For the dimer  $\text{He}_2$ , we use a two-body Gaussian potential

$$V(r) = V_0 e^{-(r/r_0)^2}, \quad (8)$$

where  $r$  indicates the interparticle distance, the parameters  $V_0 = -1.227\,171\text{ K}$  and  $r_0 = 10.030\,19\,a_0$  are chosen in order to reproduce the two-body observables. For systems involving three, four, five, and six-body, we consider the following two- and three-body Gaussian-shaped

potentials

$$\text{two-body potential: } V(r) = V_0 e^{-(r/r_0)^2} + V_1 \left(\frac{r}{r_0}\right)^2 e^{-(r/r_0)^2} \quad (9)$$

$$\text{three-body potential: } W(\rho) = W_0 e^{-(\rho/\rho_0)^2} + W_1 \left(\frac{\rho}{\rho_0}\right)^2 e^{-(\rho/\rho_0)^2}, \quad (10)$$

here,  $\rho^2$  represents the hyperradius, defined as  $\rho^2 = r_{12}^2 + r_{13}^2 + r_{23}^2$ . The potential’s parameters are specified in Table 1, selected to align with three-body observables and match those obtained from the widely utilized Helium–Helium interaction model, namely, the LM2M2 potential [Aziz and Slaman, 1991].

$r_0(\text{a}_0)$	$V_0(\text{K})$	$V_1(\text{K})$	$W_0(\text{K})$	$W_1(\text{K})$	$\rho_0(\text{a}_0)$
13.5	-0.9840318	0.3201161	0.7815301	-0.4	14.0

Table 1: Parameters of the potential (9) and (10).

### 5.1.2 The framework and the computational resources

Deep learning practitioners widely use the Adam optimizer, which has various high-performing and well-established implementations in different frameworks. We use Python within the PyTorch framework (version 2.1.0) for our implementations. It is worth noting that the Stochastic Variational Method could have been implemented with different and potentially more performing libraries. However, we opted to implement it consistently using the PyTorch framework for a fair comparison.

All the code was executed on the Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz. Despite having multiple cores available on the same machine, we ran the codes on a single CPU for benchmarking purposes to accurately record the CPU runtime.

## 5.2 Overall performance evaluation

In this section, we compare the overall performances of the baseline Stochastic Variational Method (SVM), the baseline Gradient Variational Method (GVM) and the hybrid method SVM-GVM. More details about the other methods and the chosen hyperparameters are illustrated in the next Section 5.3.

In Table 2, we compare the average energy, the standard deviation, and the lowest and highest energy values obtained by the original baseline Stochastic Variational Method with the baseline Gradient Variational Method and the hybrid SVM-GVM across all  $n$ -body systems under investigation. The stopping criteria for collecting data in this table is a given CPU runtime.

An initial observation from Table 2 indicates that the original Stochastic Variational Method consistently fails to predict the energy value for a two-body system. The variance is excessively high, and the energy values achieved are physically meaningless across multiple runs. As further explored in Section 5.4, this system presents significant challenges, and it is prone to produce several singularities, thereby rendering unreliable calculations in various experiments. The proposed gradient-based methods mitigate this issue by yielding more stable and reliable results.

$n$ -body	Methods	Av. energy (K)	Std. dv. (K)	Max. energy (K)	Min. energy(K)
2	SVM	$-3.726876 \times 10^{-3}$	$7.29 \times 10^{-3}$	$-1.291354 \times 10^{-3}$	$-2.55869 \times 10^{-2}$
	GVM	$-1.293587 \times 10^{-3}$	$2.38 \times 10^{-7}$	$-1.226184 \times 10^{-3}$	$-1.302777 \times 10^{-3}$
	<b>SVM-GVM</b>	<b><math>-1.302786 \times 10^{-3}</math></b>	<b><math>4.89 \times 10^{-9}</math></b>	<b><math>-1.302777 \times 10^{-3}</math></b>	<b><math>-1.302794 \times 10^{-3}</math></b>
3	SVM	<b>-0.1263994</b>	<b><math>1.10 \times 10^{-7}</math></b>	<b>-0.1263991</b>	-0.1263995
	GVM	-0.1263990	$1.25 \times 10^{-7}$	-0.1263953	-0.1263995
	SVM-GVM	-0.1263992	$5.00 \times 10^{-7}$	-0.1263981	-0.1263995
4	SVM	-0.5665653	$3.10 \times 10^{-5}$	-0.566511	-0.566605
	GVM	-0.5666512	$3.69 \times 10^{-5}$	-0.566571	-0.566703
	<b>SVM-GVM</b>	<b>-0.5667353</b>	<b><math>4.56 \times 10^{-6}</math></b>	<b>-0.566727</b>	<b>-0.566741</b>
5	SVM	-1.345126	$1.64 \times 10^{-4}$	-1.344697	-1.345298
	GVM	-1.345787	$3.90 \times 10^{-4}$	-1.344813	-1.346191
	<b>SVM-GVM</b>	<b>-1.346138</b>	<b><math>8.03 \times 10^{-5}</math></b>	<b>-1.346035</b>	<b>-1.346227</b>
6	SVM	-2.484672	$1.92 \times 10^{-3}$	-2.481843	-2.487954
	GVM	-2.504021	$8.51 \times 10^{-4}$	-2.502001	-2.504860
	<b>SVM-GVM</b>	<b>-2.504888</b>	<b><math>4.06 \times 10^{-4}</math></b>	<b>-2.503834</b>	<b>-2.505333</b>

Table 2: Comparison of performance between the original Stochastic Variational Method (SVM), the baseline GVM, and the hybrid SVM-GVM for 2, 3, 4, 5, and 6-body systems. Each method was executed ten times and terminated upon reaching a specified CPU runtime limit.

In the case of a three-body system, where the occurrence of ill conditions is less probable, Table 2 suggests that the original Stochastic Variational Method outperforms the gradient variational methods proposed. The average and standard deviation values are marginally superior to those of the hybrid SVM-GVM method. In this case, stochastic optimization remains a viable option compared to gradient optimization. However, as further elucidated in Section 5.5, the average energy obtained at the end of the optimization procedure does not represent the lowest average energy value achieved during the optimization process. As mentioned in Section 3.2.2, during the optimization, the gradient optimization may result in energy oscillations and a non-optimal energy value. For systems comprising more than three bodies, it is safe to state that the proposed Gradient Variational Method and the hybrid SVM-GVM surpass the original Stochastic Variational Method. Both energy values and standard deviations demonstrate improvement.

## 5.3 Detailed performance evaluation

### 5.3.1 Hyperparameters and performance evaluation in the two-body system

For a two-body system, accurate experimental data exist [Grisenti et al., 2000, Zeller et al., 2016], and its spectrum is well-known. Table 3 details the selected hyperparameters for the methods investigated (the CPU time for stopping, the number of trials in the Stochastic Variational Method, the number of basis elements, the number of iterations for the gradient variational methods and the learning rate with the eventual scheduler chosen for Adam). It is important to note that the hyperparameters in this table have been fine-tuned to achieve good performance across all methods. Notably, for the Adam optimizer, we opted for a learning rate value greater than one with a linear multiplicative scheduler (the other hyperparameters are fixed as proposed in the original paper [Kingma and Ba, 2014]). This choice is made to speed up the search for the minimum, particularly as we approach it. As anticipated in Section 3.2.2, using a dynamic and higher learning rate value has benefitted diverse machine learning applications [Smith and

Topin, 2019].

Methods	Max CPU time (s)	# Trials	# Bases	# Optim. iter.	LR	Linear scheduler
SVM	20	100	Time stop	Not app.	Not app.	Not app.
GVM	20	Not app.	10	Time stop	3.0	1.003
Static-GVM	20	Not app.	Time stop	15/basis	1.2	1.01
Dynamic-GVM	20	Not app.	Time stop	100/basis	1.2	1.003
SVM-GVM	20	100	10	Time stop	1.2	1.003

Table 3: Hyperparameters in the proposed methods for a two-body system.

Some relevant outputs of the methods are detailed in Table 4 (the average number of basis in the Stochastic Variational Method, the average number of iterations in the gradient variational methods, the average ground state energy, and the minimum and maximum value of the energies obtained in ten different runs). Figure 1 illustrates the energy of the run returning the lowest value over ten runs by all methods.

Methods	Av. bases	Av. optim. iter.	Av. last energy (K)	Std. dv. (K)	Success rate
<b>SVM</b>	15	Not app.	$-1.303\ 022 \times 10^{-3}$	$1.03 \times 10^{-7}$	7%
GVM	Not app.	170	$-1.293\ 587 \times 10^{-3}$	$2.38 \times 10^{-5}$	100%
Static-GVM	6	Not app.	$-0.356\ 871 \times 10^{-3}$	$7.24 \times 10^{-4}$	100%
Dynamic-GVM	5	Not app.	$-1.279\ 790 \times 10^{-3}$	$1.95 \times 10^{-5}$	100%
SVM-GVM	Not app.	180	$-1.302\ 786 \times 10^{-3}$	$4.88 \times 10^{-9}$	100%

Table 4: Some relevant outputs in the proposed methods for the chosen two-body system.

The primary conclusion drawn from Figure 1 and Table 4 is that the Static-GVM essentially fails to converge, or it takes an impractical amount of time to approach the ground energy value. The random search for parameters is a significant part of this method: each new basis is randomly proposed into the set and then optimized by Adam. Given that the gradient calculation is computationally expensive compared to the random search, only a limited number of optimization iterations can be set in this method (15 for each basis, as reported in Table 3). The number of gradient optimization iterations is significantly smaller than the other gradient methods and the number of trials in the Stochastic Variational Method. The gradient optimization slowly converges with each new basis, necessitating more basis elements to reach the minimum. This, in turn, increases the likelihood of encountering singularities and instabilities in calculating the eigenvalues. For these reasons, Static-GVM is not included from the subsequent experiments involving more than two-body.

By examining Figure 1 and Table 4, another observation regards the method Dynamic-GVM. Despite its convergence, it appears slower than the other proposed methods. This behavior is indeed expected. Each basis optimization occurs multiple times. The multiple optimization of the same basis inevitably slows the overall optimization procedure. However, as seen in Table 4, the method approaches the global minimum with only five bases. Beyond being a high-performing technique, the Dynamic-GVM method is a promising tool for gaining insights about the minimum Hilbert space needed to describe a given few-body system. In the next sections, we no longer consider this method for experiments involving more than two bodies. We focus on the performance of the baseline Gradient Variational Method (GVM) and the hybrid SVM-GVM compared to the baseline Stochastic Variational Method.

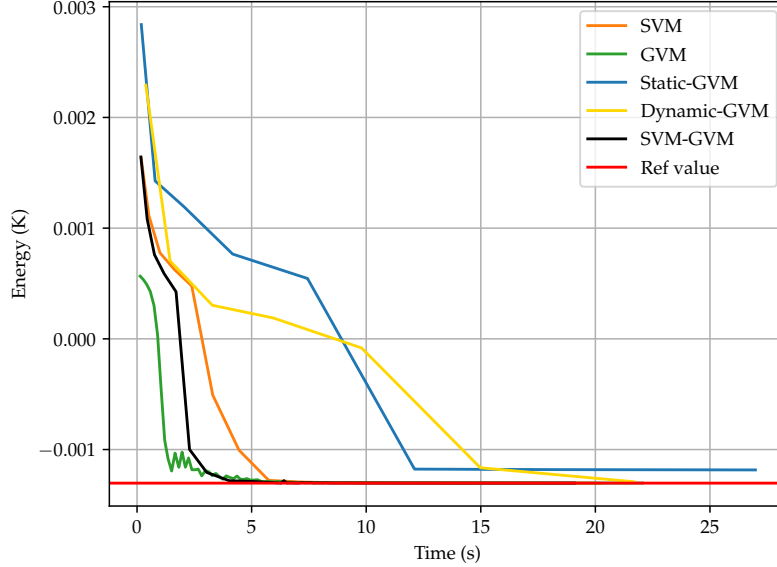


Figure 1: The energy of the best run of the proposed methods with the CPU time and the reference value for a two-body system.

We can infer a final important conclusion from the third and fourth columns of Table 4. The original Stochastic Variational Method outperforms the gradient optimization techniques for the two-body system. As anticipated, the random search in a small parameter space, such as the one in the two-body system, is very fast, while calculating gradients becomes a bottleneck for the other gradient-based methods. However, it is important to underline that we had to run this algorithm multiple times to collect valuable runs due to the high occurrence of singularities in the Stochastic Variational Method (as better detailed in the next Section 5.4). The success rate of each method is indicated in the last column of Table 4. We define a run as unsuccessful if the number of singularities exceeds the number of trials or if the calculation of the ground state eigenvalue yields meaningless results, such as an energy value lower than the actual ground state energy. As shown in this table, the success rate with the Stochastic Variational Method is only 7%, implying that approximately 140 runs of the algorithm are required to obtain ten valuable results. Despite the Stochastic Variational Method being faster, the need to run it multiple times to achieve successful and useful results significantly slows down its overall performance. The gradient-based methods do not face the same challenges. Singularities are rare occurrences, and each technique succeeds in every performed run. This, in turn, makes these methods more stable and still a valuable alternative to the baseline Stochastic Variational Method, even in the smaller parameter space of the two-body systems.



### 5.3.2 Hyperparameters and performance evaluation in the three-body system

Similar to the two-body system, the three-body molecule of  $\text{He}_3$  has been extensively examined from a theoretical perspective [Roudnev and Yakovlev, 2000, Barletta and Kievsky, 2001, Braaten and Hammer, 2006], and experimentally, it has been demonstrated to exhibit an Efimov-like spectrum [Kunitski et al., 2015]. However, in contrast to the two-body scenario, the parameter space becomes more complex, with three nonlinear parameters in the Gaussian basis (6). This increased complexity reduces the likelihood of encountering singularities. In all tests, singularities were never encountered with more than two bodies. Since our current focus is evaluating the methods’ performance, we opt to stop them once a specified CPU time is reached. Similar to the fine-tuning process employed for the two-body system, the hyperparameters for the three-body have been chosen to ensure optimal performance, particularly the high and linear multiplicative learning rate (as indicated in column 5 and six of Table 5.

Methods	Max CPU time (s)	# Trials	# Bases	# Optim. iter.	LR	Linear scheduler
SVM	1500	100	Time stop	Not app.	Not app.	Not app.
GVM	1500	Not app.	40	Time stop	3.0	1.003
SVM-GVM	1500	100	40	Time stop	1.2	1.003

Table 5: Hyperparameters in the proposed methods for a three-body system.

Table 6 shows relevant outputs for the chosen three-body system. Figure 2 depicts the energy along the CPU time of the run returning the lowest value obtained over ten runs.

Methods	Av. bases	Av. optim. iter.	Av. last energy (K)	Std. dv. (K)
<b>SVM</b>	123	Not app.	<b>-0.1263994</b>	<b><math>1.10 \times 10^{-7}</math></b>
GVM	Not app.	270	-0.1263990	$1.25 \times 10^{-6}$
SVM-GVM	Not app.	200	-0.1263992	$5.00 \times 10^{-7}$

Table 6: Some relevant outputs in the proposed methods for the chosen three-body system.

When considering the average energy achieved at the end of the optimization process, the third and fourth column in Table 6 leads to the same conclusion as in the two-body system: the Stochastic Variational Method continues to outperform gradient optimization techniques in the three-body system as well, both in term of average energy and variance. Nevertheless, the average energy values obtained, particularly in the Gradient Variational Method and SVM-GVM, are very close to those obtained by the Stochastic Variational Method, with only 0.003‰ and 0.0002‰ of separation, respectively. Moreover, as elucidated in Section 5.3.2, it is essential to highlight that these energy values represent the mean of the energy obtained at the end of the optimization procedure. As mentioned in Section 3.2.2, gradient optimization methods may result in energy oscillation and unstable convergence; therefore, the final energy is not always the minimum energy. This, according to the variational principle, cannot happen in the Stochastic Variational Method since by adding a new basis element, the new energy is always lower than the previous one. The phenomenon of energy oscillation is further explored in Section 5.5.

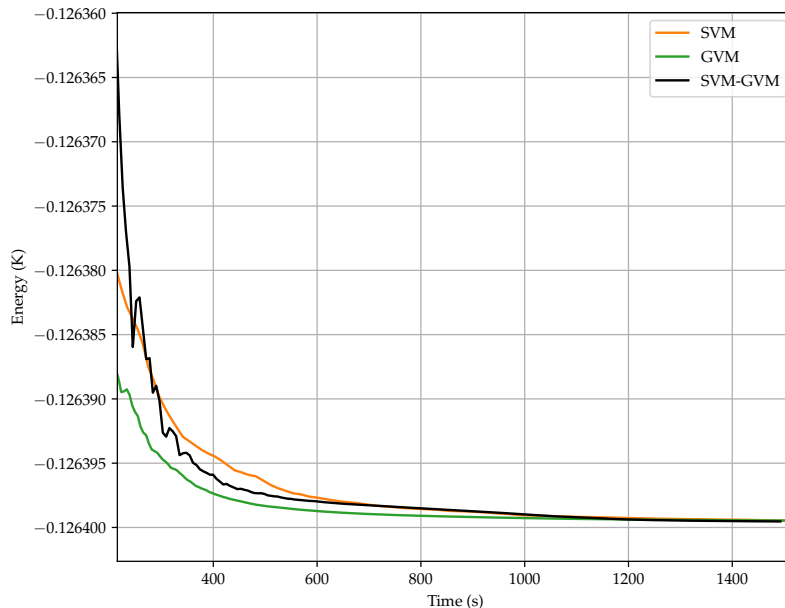


Figure 2: The energy of the best run of the proposed methods with the CPU time for a three-body system. The tail of the graph has been magnified for better clarity.

### 5.3.3 Hyperparameters and performance evaluation with four-, five-body and six-body

Very seldom do experimental data exist for more than three-body. The potential we use in four-, five- and six-body is the same as in Equation (9) and (10), and the parameters are the same as in Table 1. This potential has been shown to accurately approach the same ground energy values’ obtained by the Helium–Helium LM2M2 interaction [Recchia et al., 2022b].

As chosen in the previous sections, we mainly focus on the methods’ performance. We stop the algorithms once a specified CPU time is reached. The hyperparameters of the methods are presented in Table 7. Unlike the two- and three-body systems, there is no need for the same fine-tuning of the reported hyperparameters: with four-, five- and six-body systems, we observe fast convergence of energy with less sensitivity of the choice of hyperparameters.

Table 8 shows some relevant outputs, and Figures 3, 4 and 5 illustrate the energy of the run resulting in the lowest value of SVM, GVM, and SVM-GVM over ten runs in four-, five- and six-body systems respectively.

As evident in Table 8 and Figures 3, 4 and 5 the gradient variational methods outperform the original Stochastic Variational Method in four-, five-body and six-body systems in each conducted runs. Notably, the hybrid SVM-GVM method is the most efficient, converging faster with a smaller variance. From Table 8, we observe that the improvement in average energy is approximately 0.3‰ in four-body, 0.8‰ in five-body and 8‰ in six-body. *As expected, the gradient methods exhibit enhanced performance compared to random search as the dimension of*

<b>4-body</b>						
Methods	Max CPU time (s)	# Trials	# Bases	# Optim. iter.	LR	Linear scheduler
SVM	2500	100	Time stop	Not app.	Not app.	Not app.
GVM	2500	Not app.	25	Time stop	0.8	1.0
SVM-GVM	2500	100	25	Time stop	0.8	1.0
<b>5-body</b>						
Methods	Max CPU time (s)	# Trials	# Bases	# Optim. iter.	LR	Linear scheduler
SVM	10000	100	Time stop	Not app.	Not app.	Not app.
GVM	10000	Not app.	25	Time stop	0.8	1.0
SVM-GVM	10000	100	25	Time stop	0.8	1.0
<b>6-body</b>						
Methods	Max CPU time (s)	# Trials	# Bases	# Optim. iter.	LR	Linear scheduler
SVM	19000	100	Time stop	Not app.	Not app.	Not app.
GVM	19000	Not app.	25	Time stop	0.8	1.0
SVM-GVM	19000	100	25	Time stop	0.8	1.0

Table 7: Hyperparameters in the SVM, GVM, and SVM-GVM for four-, five- and six-body systems.

<b>4-body</b>				
Methods	Average bases	Average optim. iter.	Av. energy (K)	Std. dv. (K)
SVM	93	Not app.	-0.5665653	$3.10 \times 10^{-5}$
GVM	Not app.	390	-0.5666512	$3.69 \times 10^{-5}$
<b>SVM-GVM</b>	Not app.	360	<b>-0.5667353</b>	<b><math>4.56 \times 10^{-6}</math></b>
<b>5-body</b>				
Methods	Average bases	Average optim. iter.	Av. last energy (K)	Std. dv. (K)
SVM	125	Not app.	-1.345126	$1.64 \times 10^{-4}$
GVM	Not app.	700	-1.345787	$3.90 \times 10^{-4}$
<b>SVM-GVM</b>	Not app.	650	<b>-1.346138</b>	<b><math>8.03 \times 10^{-5}</math></b>
<b>6-body</b>				
Methods	Average bases	Average optim. iter.	Av. last energy (K)	Std. dv. (K)
SVM	125	Not app.	-2.484672	$1.92 \times 10^{-3}$
GVM	Not app.	850	-2.504021	$8.51 \times 10^{-4}$
<b>SVM-GVM</b>	Not app.	750	<b>-2.504888</b>	<b><math>4.06 \times 10^{-4}</math></b>

Table 8: Some relevant outputs in the SVM, GVM, and SVM-GVM in four-, five- and six-body systems.

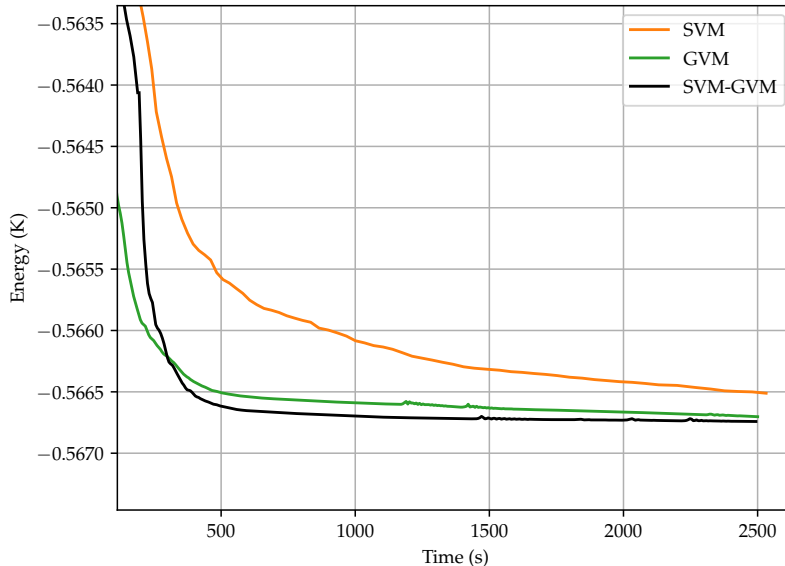


Figure 3: The energy of the best run of the SVM, GVM, and SVM-GVM with the CPU time in a four-body system. The tail of the graphs has been magnified for better clarity.

*the search space increases.*

#### 5.4 Singularities in the Stochastic Variational Methods and challenging systems

The study of the two-body system merits particular attention for several reasons. Firstly, only one nonlinear parameter must be chosen in the Gaussian basis (6), making the random choice in the Stochastic Variational Method efficient. However, the space of nonlinear parameters is very small, and as mentioned in Section 2.2, the occurrence of singularities is more likely to happen in the random choice. From a physics perspective, this system is very shallow, and its ground state energy is close to zero, indicating weak binding. This property makes it computationally challenging to find the ground state eigenvalue of the Hamiltonian, further increasing the likelihood of encountering singularities. For all these reasons, the two-body system is the ideal candidate to assess the stability of the proposed gradient variational methods with the Stochastic Variational Method.

In this section, we compare the amount of singularities occurrences between the Stochastic Variational Method and the Dynamic-GVM. Both methods are stopped after reaching  $K = 20$  basis elements. To ensure a fair comparison, we set the number of trials  $L$  in the Stochastic Variational Method to be the same as the number of gradient optimization iterations  $L$  in the Dynamic-GVM, with both set to one hundred. We repeated the two methods one hundred times and recorded the count of singularity occurrences on each basis. Figure 6 illustrates the average count of singularities at each basis.

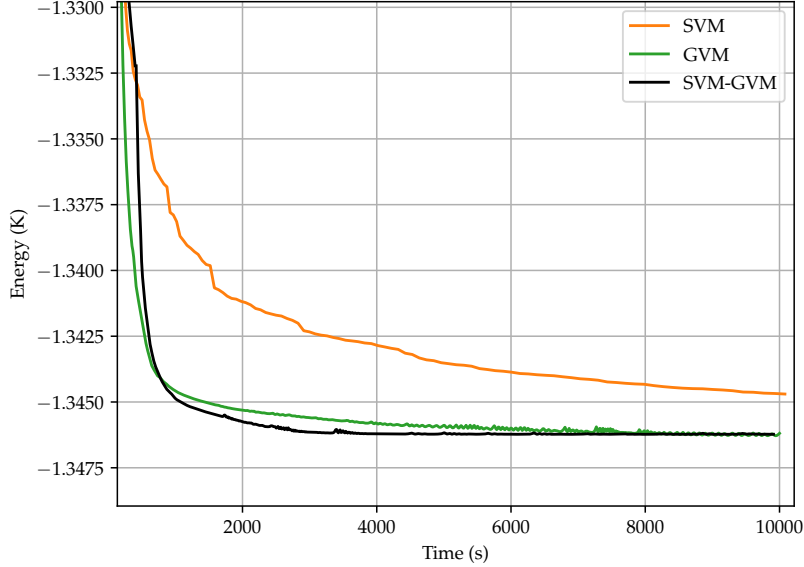


Figure 4: The energy of the best run of the SVM, GVM, and SVM-GVM with the CPU time in a five-body system. The tail of the graph has been magnified for better clarity.

Two essential considerations arise from Figure 6. First, we see that the number of singularities with the Dynamic-GVM is less than that of the original Stochastic Variational Method. Second, we observe that the amount of singularities increases as the basis size and then as we approach the ground state energy value. It is worth noting that, in the Dynamic-GVM, Adam attempts to find the global minimum supported by a given number of bases. As we approach the minimum, we can expect a higher chance of encountering singularities. Nevertheless, the gradient method remains reasonably stable even in this scenario, further reinforcing the assertion about its improved stability compared to the Stochastic Variational Method.

Theoretical investigation into why gradient variational methods are less prone to singularities is beyond the scope of the current paper. However, a potential explanation lies in the search method employed. Randomly jumping into a space affected by singularities is “riskier” compared to smoothly exploring it by following the steepest slope of the energy landscape.

## 5.5 Oscillations in gradient variational methods

As anticipated in Section 3.2.2, the value of the learning rate plays a crucial role in the neighborhood of a minimum. Specifically, if the product of the loss function’s sharpness (the highest eigenvalue of its Hessian) and the learning rate exceeds two, the optimization process oscillates. This can lead to either divergences or a phase known as unstable convergence (or the edge of stability), where the optimization converges non-monotonically over a long timescale [Choo et al., 2020, Ahn et al., 2022]. In optimising the quantum few-body ground state energy via gradient variational methods, we also encounter divergences and unstable

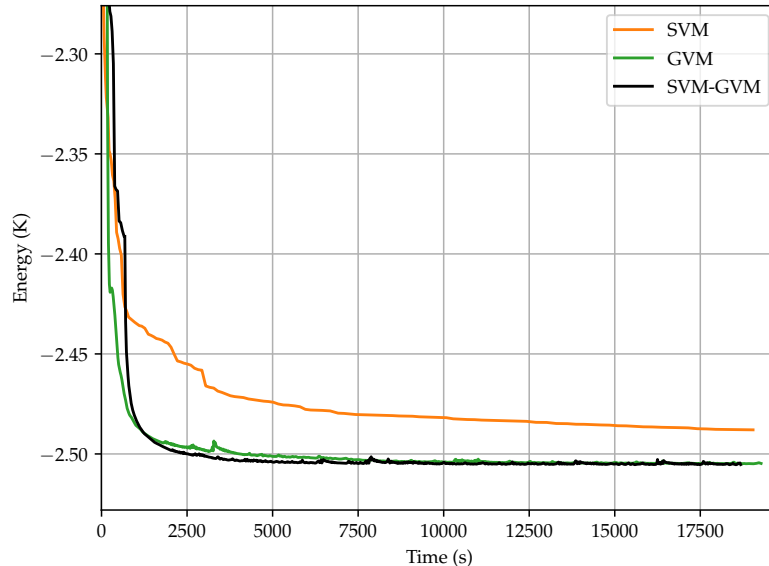


Figure 5: The energy of the best run of the SVM, GVM and SVM-GVM with the CPU time in a six-body system. The tail of the graph has been magnified for better clarity.

convergence instances.

We encounter divergences when the gradient approaches the region of the minimum near the end of the optimization procedure. In this scenario, if CPU runtime is used as the stopping criterion, the optimization continues even after reaching the optimum. As noted in Section 5.3.2, Table 6 suggests that, when considering the average energy at the end of the optimization, the Stochastic Variational Method outperforms the gradient variational methods GVM and SVM-GVM. However, the final energy obtained using gradient variational methods is not always the minimum energy reached during the optimization. Figure 7 shows the energies obtained in the worst runs, specifically those converging to the highest energy values, for the baseline Stochastic Variational Method and the gradient variational methods GVM and SVM-GVM.

As observed in Figure 7, these gradient variational methods outperform the original Stochastic Variational Method, and approaching the global minimum, they tend to overshoot it, resulting in the energy’s oscillation. In our case, this phenomenon is primarily attributed to the high and multiplicative increasing learning rate (as seen with the linear multiplicative learning rate chosen for the three-body system in Table 5), which can result in a high value of the sharpness multiplied by the learning rate, causing the oscillations and divergences. To mitigate the increasing learning rate different workarounds have been proposed in the literature (for the sake of examples, we can cite cyclic learning rates [Smith, 2017] or a decaying learning rate with cosine annealing [Loshchilov and Hutter, 2016], clipping the gradient [Mikolov et al., 2012, Pascanu et al., 2013]). However, it is important to note that while these techniques may mitigate the oscillations, they potentially slow the optimization process. For the current analysis, in the first column of Table 9, we present the average minimum energy obtained by each method.

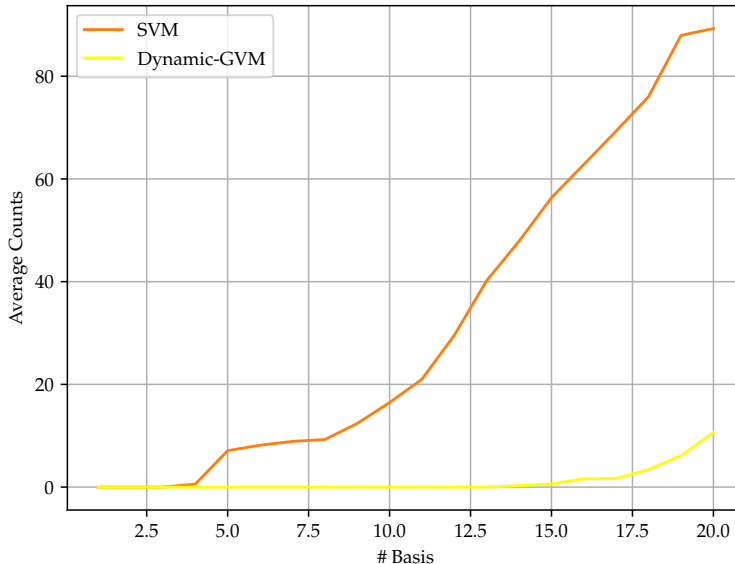


Figure 6: Average counts of singularities occurrences over 100 runs of the SVM and the Dyanamic-GVM.

The values closely align with those obtained using the baseline Stochastic Variational Method. Specifically, the SVM-GVM method slightly outperforms the original Stochastic Variational Method, with a comparable variance. Regardless of the oscillation around the global minimum, we can state that the hybrid SVM-GVM method slightly outperforms the baseline Stochastic Variational Method in a three-body system as well.

Methods	Av. min. energy (K)	Std. dv. (K)
SVM	-0.1263994	$1.10 \times 10^{-7}$
GVM	-0.1263994	$1.37 \times 10^{-7}$
<b>SVM-GVM</b>	<b>-0.1263995</b>	$1.18 \times 10^{-7}$

Table 9: The average and the standard deviation of the minimum energy obtained throughout the optimization process in the proposed methods for the chosen three-body system.

Besides the divergences observed near the global minimum, the Figures 1, 2, 3, 4, and 5 also show energy oscillations during the optimization process. However, these oscillations do not compromise the optimization, and although the energy decreases non-monotonically, it converges over a long timescale. Understanding the primary cause of these oscillations requires knowledge of the energy landscape, particularly whether it is affected by different local minima and their sharpness. Exploring the entire space of nonlinear parameters to identify local minima and evaluate their sharpness is computationally challenging. For our current analysis, we only investigate the presence of local minima within a grid in selected cases where the space is small



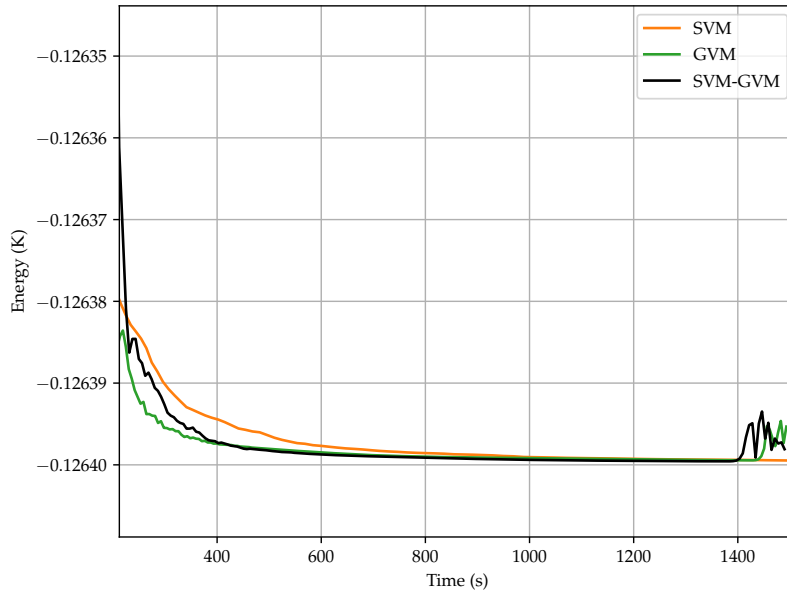


Figure 7: The energies of the methods SVM, GVM, and SVM-GVM in the worst over ten runs, with the corresponding CPU time and the reference value for a three-body system.

enough to make this assessment feasible. Specifically, for the two-body system with one and two basis elements, we did not encounter local minima within the chosen grid, apart from a few cases, due to numerical error. Figure 8 depicts the energy landscape concerning the two nonlinear parameters for a two-body system with two bases. The landscape shown in the figure appears smooth, with two distinct symmetric wide valleys (the global minima represented by the red dots in the figure). This simple analysis suggests that the oscillations may be attributed again to the high learning rate of the gradient optimization step rather than the presence of multiple local minima or sharp minima. Nevertheless, we cannot conclusively assert that the space of nonlinear parameters is free of multiple local minima or sharp minima. In higher-dimensional space, the deepness of the minima may differ in different directions, and verifying the presence of local minima and evaluating their sharpness becomes computationally infeasible.

## 6 Conclusion

The Stochastic Variational Method is a stochastic search method leveraging the variational principle. It is an efficient, accurate, state-of-the-art method for solving various few-body quantum systems. It is, however, challenged by singularities when the number of particles is small and by the size of the search space when the number of particles increases.

We investigate the opportunity to replace stochastic search with gradient-based optimization. We propose and compare three gradient-based methods (named gradient variational methods)

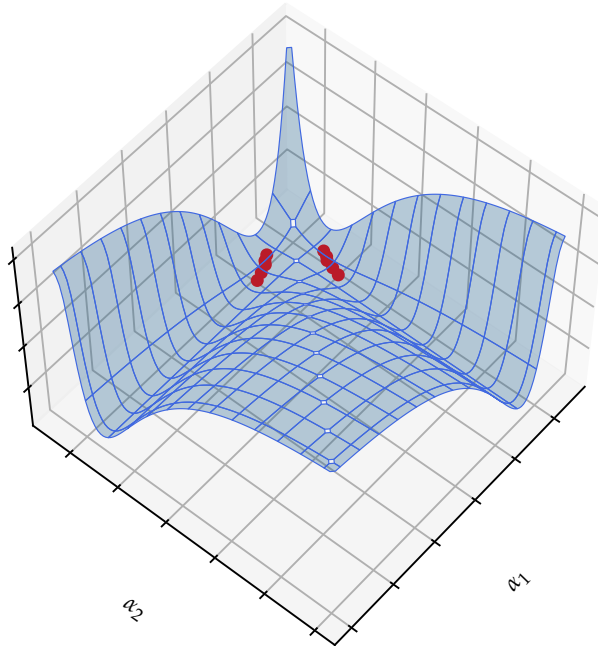


Figure 8: The landscape of the energy in case of two nonlinear parameters with two-body and two bases. The red dots indicate the point where all the surrounding points in the grid are higher (the global minima).

and a hybrid method that mixes the Stochastic Variational Method with a gradient variational method. We evaluate the strengths and weaknesses of these algorithms for the computation of the ground state energy of two-, three-, four-, five and six-body systems containing identical bosons, specifically clusters of helium He, which can be effectively described using Gaussian-shaped potentials.

Our results for the two-body system show that the gradient variational methods exhibit improved stability, resulting in better performance. Indeed, fewer singularities occur during the optimization process, suggesting a potential advantage in incorporating gradient-based techniques into studying challenging physics systems.

The Stochastic Variational Method outperforms the proposed gradient variational methods for the computation of the energy of a three-body system. However, when employing gradient variational methods, we observe occasional oscillations and divergences in the energy value, particularly when approaching the minima. This highlights the importance of finely tuning the learning rate in the gradient descent algorithm to prevent its increasing. Nevertheless, upon excluding the oscillations from our analysis, the gradient variational methods, particularly the hybrid SVM-GVM method, show competitive performance that surpasses the results obtained

by the Stochastic Variational Method.

In exploring four-, five- and six-body systems, the tested gradient variational methods consistently outperform the Stochastic Variational Method, demonstrating their efficiency in higher-dimensional spaces.

The outcomes of this study suggest that while the original Stochastic Variational Method remains powerful, gradient variational methods offer promising alternatives. These techniques show potential in addressing the computational challenges encountered in the Stochastic Variational Method, paving the way for more efficient solutions to the quantum few-body problem. The observed success of gradient variational methods in higher-dimensional systems implies their scalability, encouraging further exploration in even more complex physics scenarios.

For the sake of completeness, we should acknowledge the potential further enhancement in both the Stochastic Variational Method and our proposed gradient variational methods. For example, Stochastic Variational Methods lend themselves to parallelization, where multiple trials can be conducted concurrently, with the minimum selected from among these trials, while gradient variational methods can be refined by incorporating analytical expressions of gradients whenever possible.

We are now exploring alternative basis sets with nonlinear parameters, such as wavelets as well as neural network ansätze. The interplay between random and gradient-based search strategies could contribute to refining these methods for broader use in quantum few-body physics.

## Acknowledgements

This research is supported by Singapore Ministry of Education, grant MoE-T1251RES2302, and by the National Research Foundation, Prime Minister’s Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) programme as part of the programme Descartes. D. Basu acknowledges the support of the ANR JCJC project REPUBLIC (ANR-22-CE23-0003-01), and the PEPR project FOUNDRY (ANR23-PEIA-0003).

## A List of Algorithms

In this Appendix, we report the pseudocodes of the algorithms described and tested in the paper:

1. Stochastic Variational Method (SVM)
2. Static gradient variational methods
3. Dynamic gradient variational methods
4. Baseline Gradient Variational Method (GVM)
5. Static Gradient Variational Method (Static-GVM)
6. Dynamic Gradient Variational Method (Dynamic-GVM)

---

**Algorithm 1:** Stochastic Variational Method (SVM)

---

**Data:**  $K$  = basis size,  $L$  = number of trials,  $H$  = Hamiltonian

```
1 for basis  $k = 1$  to  $K$  do
2   for number of trials  $l = 1$  to  $L$  do
3     Pick as the number of pair  $\frac{N(N-1)}{2}$  random nonlinear parameters
4      $\alpha_{ab}^{(k)} \sim U(0, \alpha_{max}]$ 
5     Make up the basis element  $\psi_k^{(l)}$ 
6     for each index  $j \geq i = 1$  to  $k$  do
7       Being  $\psi_k = \psi_k^{(l)}$ 
8       Calculate  $\mathcal{S}_{ij}^{(l)} = \langle \psi_i | \psi_j \rangle$  and  $\mathcal{S}_{ji}^{(l)} = \mathcal{S}_{ij}^{(l)}$ 
9       Calculate  $\mathcal{H}_{ij}^{(l)} = \langle \psi_i | H | \psi_j \rangle$  and  $\mathcal{H}_{ji}^{(l)} = \mathcal{H}_{ij}^{(l)}$ 
10    end
11    Solve the eigenvalue problem  $\mathcal{H}^{(l)} \mathcal{C} = E^{(l)} \mathcal{S}^{(l)} \mathcal{C}$ 
12    if  $\mathcal{S}^{(l)}$  is singular then
13      go to 3;
14    end
15    end
16    Set  $l = \underset{l}{\operatorname{argmin}}[E_1^{(l)}]$ 
17    Set  $\psi_k = \psi_k^{(l)}$ 
18    return  $E_1 = E_1^{(l)}$ 
19 end
```

---

---

**Algorithm 2:** Static gradient variational methods

---

**Data:**  $K = K_1, K_2, \dots$  = batches' size,  $H$  = Hamiltonian

- 1 Epoch  $m = 1$
- 2 Initialize  $K_0 = 0$
- 3 **repeat**
- 4     Pick at random a batch of  $\frac{N(N-1)}{2} \times K_m$  nonlinear parameters  $\alpha_{K_m}$
- 5     From  $\alpha_{K_m}$  make up the batch  $\mathbf{A}_{K_m} = \{\psi_k\}_{k=K_{m-1}+1, \dots, K_{m-1}+K_m}$
- 6      $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A}_{K_m})$
- 7     Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = \text{ESC}$
- 8     **if**  $\mathcal{S}$  is singular **then**
- 9         | **go to 4;**
- 10     **end**
- 11     **repeat**
- 12         | Being the lowest eigenvalue the loss function  $E_1(\alpha_{K_m})$
- 13         | Calculate the gradient of the loss function  $\nabla E_1(\alpha_{K_m})$
- 14         | Update  $\alpha_{K_m}$  based on Adam optimizer
- 15         | From the new  $\alpha_{K_m}$  make up the batch  $\mathbf{A}_{K_m} = \{\psi_k\}_{k=K_{m-1}+1, \dots, K_{m-1}+K_m}$
- 16         |  $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A}_{K_m})$
- 17         | Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = \text{ESC}$
- 18         | **if**  $\mathcal{S}$  is singular **then**
- 19             |     Modify the gradient extracting it from a normal distribution
- 20             |      $\nabla E_1(\alpha_{K_m}) \sim N(\nabla E_1(\alpha_{K_m}), 0.1)$
- 21             |     **go to 14;**
- 22         | **end**
- 23     **until** A given condition is met;
- 24     Next epoch  $m = m + 1$
- 25 **until** A given condition is met;
- 26 **return**  $E_1$
- 27 **Function**  $\text{FillOverlapHamiltonian}(\mathbf{A}_{K_m})$ :
- 28     **for** each index  $j \geq i = K_{m-1} + 1$  **to**  $K_{m-1} + K_m$  **do**
- 29         | Calculate  $\mathcal{S}_{ij} = \langle \psi_i | \psi_j \rangle$  and  $\mathcal{S}_{ji} = \mathcal{S}_{ij}$
- 30         | Calculate  $\mathcal{H}_{ij} = \langle \psi_i | H | \psi_j \rangle$  and  $\mathcal{H}_{ji} = \mathcal{H}_{ij}$
- 31     **end**
- 32     **return**  $\mathcal{S}, \mathcal{H}$

---

---

**Algorithm 3:** Dynamic gradient variational methods

---

**Data:**  $K = K_1, K_2, \dots$  = batches' size,  $H$  = Hamiltonian

- 1 Epoch  $m = 1$
- 2 Initialize  $K_0 = 0$
- 3 **repeat**
- 4     Pick at random a batch of  $\frac{N(N-1)}{2} \times K_m$  nonlinear parameters  $\alpha_{K_m}$
- 5     Append to the entire set  $\alpha \leftarrow \alpha_{K_m}$
- 6     From  $\alpha$  make up the batch  $\mathbf{A} = \{\psi_k\}_{k=0, \dots, K_{m-1} + K_m}$
- 7      $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A})$
- 8     Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = ESC$
- 9     **if**  $\mathcal{S}$  is singular **then**
- 10         | **go to 4;**
- 11     **end**
- 12     **repeat**
- 13         | Being the lowest eigenvalue the loss function  $E_1(\alpha)$
- 14         | Calculate the gradient of the loss function  $\nabla E_1(\alpha)$
- 15         | Update  $\alpha$  based on Adam optimizer
- 16         | From the new  $\alpha$  make up the basis  $\mathbf{A} = \{\psi_k\}_{k=0, \dots, K_{m-1} + K_m}$
- 17         |  $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A})$
- 18         | Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = ESC$
- 19         | **if**  $\mathcal{S}$  is singular **then**
- 20             | Modify the gradient extracting it from a normal distribution
- 21             |  $\nabla E_1(\alpha) \sim N(\nabla E_1(\alpha), 0.1)$
- 22             | **go to 15;**
- 23         | **end**
- 24     **until**  $A$  given condition is met;
- 25     Next epoch  $m = m + 1$
- 26 **until**  $A$  given condition is met;
- 27 **return**  $E_1$
- 28 **Function**  $\text{FillOverlapHamiltonian}(\mathbf{A})$ :
- 29     **for** each index  $j \geq i = 1$  **to**  $K_{m-1} + K_m$  **do**
- 30         | Calculate  $\mathcal{S}_{ij} = \langle \psi_i | \psi_j \rangle$  and  $\mathcal{S}_{ji} = \mathcal{S}_{ij}$
- 31         | Calculate  $\mathcal{H}_{ij} = \langle \psi_i | H | \psi_j \rangle$  and  $\mathcal{H}_{ji} = \mathcal{H}_{ij}$
- 32     **end**
- 33     **return**  $\mathcal{S}, \mathcal{H}$

---

---

**Algorithm 4:** Gradient Variational Method (GVM)

---

**Data:**  $K$  = basis size,  $H$  = Hamiltonian

- 1 Pick  $K \times \frac{N(N-1)}{2}$  random nonlinear parameters  $\alpha$
- 2 From  $\alpha$  make up the set of basis elements  $\mathbf{A} = \{\psi_k\}_{k=1,\dots,K}$
- 3  $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A})$
- 4 Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = E\mathcal{C}$
- 5 **if**  $\mathcal{S}$  is singular **then**
- 6 |   **go to 1**;
- 7 **end**
- 8 **repeat**
- 9 |   Being the lowest eigenvalue the loss function  $E_1(\alpha)$
- 10 |   Calculate the gradient of the loss function  $\nabla E_1(\alpha)$
- 11 |   Update  $\alpha$  based on Adam optimizer
- 12 |   From the new  $\alpha$  make up the basis elements  $\mathbf{A} = \{\psi_k\}_{k=1,\dots,K}$
- 13 |    $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A})$
- 14 |   Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = E\mathcal{C}$
- 15 |   **if**  $\mathcal{S}$  is singular **then**
- 16 |   |   Modify the gradient extracting it from a normal distribution  
17 |   |    $\nabla E_1(\alpha) \sim N(\nabla E_1(\alpha), 0.1)$
- 17 |   |   **go to 11**;
- 18 |   **end**
- 19 **until**  $A$  given condition is met;
- 20 **return**  $E_1$
- 21 **Function**  $\text{FillOverlapHamiltonian}(\mathbf{A})$ :
- 22 |   **for** each index  $j \geq i = 1$  to  $k$  **do**
- 23 |   |   Calculate  $\mathcal{S}_{ij} = \langle \psi_i | \psi_j \rangle$  and  $\mathcal{S}_{ji} = \mathcal{S}_{ij}$
- 24 |   |   Calculate  $\mathcal{H}_{ij} = \langle \psi_i | H | \psi_j \rangle$  and  $\mathcal{H}_{ji} = \mathcal{H}_{ij}$
- 25 |   |   **return**  $\mathcal{S}, \mathcal{H}$
- 26 |   **end**

---



---

**Algorithm 5:** Static Gradient Variational Method (Static-GVM)

---

**Data:**  $L$  = number of optimization iterations,  $H$  = Hamiltonian

- 1 Epoch  $m = 1$
- 2 **repeat**
- 3     Pick as the number of pairs  $\frac{N(N-1)}{2}$  random nonlinear parameters  $\alpha_{ab}^{(m)}$
- 4     **for** *number of step*  $l = 1$  **to**  $L$  **do**
- 5         Calculate the corresponding basis vector  $\psi_m$
- 6         **for** *each index*  $j = 1$  **to**  $m$  **do**
- 7             Calculate  $\mathcal{S}_{mj} = \langle \psi_m | \psi_j \rangle$  and  $\mathcal{S}_{jm} = \mathcal{S}_{mj}$
- 8             Calculate  $\mathcal{H}_{mj} = \langle \psi_m | H | \psi_j \rangle$  and  $\mathcal{H}_{jm} = \mathcal{H}_{mj}$
- 9         **end**
- 10         Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = E\mathcal{S}\mathcal{C}$
- 11         **if**  $\mathcal{S}$  is singular &  $l = 0$  **then**
- 12             **go to 3;**
- 13         **end**
- 14         **if**  $\mathcal{S}$  is singular &  $l > 0$  **then**
- 15             Modify the gradient extracting it from a normal distribution
- 16              $\nabla E_1(\alpha_{ab}^{(m)}) \sim N(\nabla E_1(\alpha_{ab}^{(m)}), 0.1)$
- 17             **go to 20;**
- 18             **end**
- 19             Being the lowest eigenvalue the loss function  $E_1(\alpha_{ab}^{(m)})$
- 20             Calculate the gradient of the loss function  $\nabla E_1(\alpha_{ab}^{(m)})$
- 21             Update  $\alpha_{ab}^{(m)}$  based on Adam optimizer
- 22         **end**
- 23         Increase epoch  $m = m + 1$
- 24 **until** *Until a condition is met;*
- 25 **return**  $E_1$

---

---

**Algorithm 6:** Dynamic Gradient Variational Method (Dynamic-GVM)

---

**Data:**  $L$  = number of optimization iterations,  $H$  = Hamiltonian

- 1 Epoch  $m = 1$
- 2 **repeat**
- 3     Pick as the number of pairs  $\frac{N(N-1)}{2}$  random nonlinear parameters  $\alpha_{ab}^{(m)}$
- 4     Append to the set of nonlinear parameters  $\alpha \leftarrow \alpha_{ab}^{(m)}$
- 5     Make up the basis element  $\psi_m$
- 6     Append to the set of bases  $\mathbf{A} \leftarrow \psi_m$
- 7      $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A})$
- 8     Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = E\mathcal{S}\mathcal{C}$
- 9     **if**  $\mathcal{S}$  is singular **then**
- 10         | **go to** 3;
- 11     **end**
- 12     **for** number of optimization iterations  $l = 1$  **to**  $L$  **do**
- 13         Being the lowest eigenvalue the loss function  $E_1(\alpha)$
- 14         Calculate the gradient of the loss function  $\nabla E_1(\alpha)$
- 15         Update  $\alpha$  based on Adam optimizer
- 16         From the new  $\alpha$  make up the basis set  $\mathbf{A} = \{\psi_k\}_{k=1, \dots, m}$
- 17          $\mathcal{S}, \mathcal{H} = \text{FillOverlapHamiltonian}(\mathbf{A})$
- 18         Solve the eigenvalue problem  $\mathcal{H}\mathcal{C} = E\mathcal{S}\mathcal{C}$
- 19         **if**  $\mathcal{S}$  is singular **then**
- 20             | Modify the gradient extracting it from a normal distribution
- 21             |  $\nabla E_1(\alpha) \sim N(\nabla E_1(\alpha), 0.1)$
- 22             | **go to** 15;
- 23         **end**
- 24     **end**
- 25     Increase epoch  $m = m + 1$
- 26 **until** A given condition is met;
- 27 **return**  $E_1$
- 28 **Function**  $\text{FillOverlapHamiltonian}(\mathbf{A})$ :
- 29     **for** each index  $j \geq i = 0$  **to**  $m$  **do**
- 30         | Calculate  $\mathcal{S}_{ij} = \langle \psi_i | \psi_j \rangle$  and  $\mathcal{S}_{ji} = \mathcal{S}_{ij}$
- 31         | Calculate  $\mathcal{H}_{ij} = \langle \psi_i | H | \psi_j \rangle$  and  $\mathcal{H}_{ji} = \mathcal{H}_{ij}$
- 32         | **return**  $\mathcal{S}, \mathcal{H}$
- 33     **end**

---

---

**Algorithm 7:** SVM-GVM

---

**Data:**  $K$  = basis size,  $H$  = Hamiltonian

// Warm up with Stochastic Variational Method

- 1 Perform the Algorithm 1 for  $K$  basis elements
- 2 Return the set of nonlinear parameters  $\alpha = \{\alpha_{ab}^{(k)}\}_{k=1, \dots, K}$  and the corresponding basis set  $\mathbf{A} = \{\psi_k\}_{k=1, \dots, K}$
- // Optimization with the Gradient Variational Method
- 3 Perform the Algorithm 4 with the initialized set of nonlinear parameters  $\alpha$  and corresponding basis set  $\mathbf{A}$

---

## B The secular equation

Let us approximate the eigenvector  $\psi_l$  with the ansatz  $\phi_l$  as in (4)

$$\psi_l \approx \phi_l = \sum_{k=1}^K c_{kl} \psi_k. \quad (11)$$

For the  $l$ th eigenvalue, the functional (3) can be written as

$$E_l[\phi_l] = \frac{\langle \sum_{k=1}^K c_{kl} \psi_k | H | \sum_{k=1}^K c_{kl} \psi_k \rangle}{\langle \sum_{k=1}^K c_{kl} \psi_k | \sum_{k=1}^K c_{kl} \psi_k \rangle} = \frac{\sum_{i=1}^K \sum_{j=1}^K c_{il}^* c_{jl} \mathcal{H}_{ij}}{\sum_{i=1}^K \sum_{j=1}^K c_{il}^* c_{jl} \mathcal{S}_{ij}}, \quad (12)$$

where

$$\mathcal{H}_{ij} = \langle \psi_i | H | \psi_j \rangle \quad (13)$$

$$\mathcal{S}_{ij} = \langle \psi_i | \psi_j \rangle, \quad (14)$$

are the matrix elements of the Hamiltonian and overlap matrices, respectively, in the chosen basis. According to the Ritz theorem, the parameters  $c_{kl}$  in (11) can be determined by the stationary condition

$$\delta E_l[\phi_l] = 0. \quad (15)$$

The variation of the functional (12) with respect to the parameters  $c_{kl}$  in (11), leads to the stationary condition

$$\frac{\partial E_l[\phi_l]}{\partial c_{kl}^*} = \frac{\sum_{j=1}^K c_{jl} (\mathcal{H}_{kj} - E_l \mathcal{S}_{kj})}{\sum_{i=1}^K \sum_{j=1}^K c_{il}^* c_{jl} \mathcal{S}_{ij}} = 0 \implies \sum_{j=1}^K \mathcal{H}_{kj} c_{jl} = E_l \sum_{j=1}^K \mathcal{S}_{kj} c_{jl}. \quad (16)$$

Introducing the eigenvector matrix  $\mathcal{C}$  and the diagonal matrix of eigenvalues  $E = \text{diag}(E_l)_{l=1, \dots, K}$ , Equation (16) becomes

$$(\mathcal{H}\mathcal{C})_{kl} = (\mathcal{S}\mathcal{C}E)_{kl}, \quad (17)$$

which represents a linear system of  $K$  equations in  $c_{kl}$ . For each eigenvalue  $E_l$  this system can be solved as a generalized finite eigenvalue problem, often referred to as the secular equation (5)

$$\mathcal{H}\mathcal{C} = \mathcal{S}\mathcal{C}E. \quad (18)$$

## C Solving the secular equation

The primary goal is to reduce the task of solving a generalized eigenvalue problem as the one formulated in (5) in a standard eigenvalue problem. We could achieve this purpose by using the symmetric orthogonalization of the basis.

Instead of the old basis set  $\mathbf{A} = \{\psi_k\}_{k=1,\dots,K}$  in which the matrices  $\mathcal{H}ij = \langle \psi_i | H | \psi_j \rangle$  and  $Sij = \langle \psi_i | \psi_j \rangle$  are calculated, we could use an orthogonal basis set defined as

$$\mathbf{A}' = \mathcal{S}^{-\frac{1}{2}} \mathbf{A}. \quad (19)$$

The matrix  $\mathcal{S}^{-\frac{1}{2}}$  is computed as follow

1. We diagonalize  $\mathcal{S}$  with a unitary matrix

$$\mathcal{S}_{\text{diag}} = U^\dagger \mathcal{S} U. \quad (20)$$

2. Since the matrix  $\mathcal{S}$  is symmetric positive definite, the eigenvalues of  $\mathcal{S}$  are all positive. We can safely calculate the square roots of the diagonal elements in  $\mathcal{S}_{\text{diag}}$  that, in turn, allow us to define the symbol  $\mathcal{S}_{\text{diag}}^{\frac{1}{2}}$ .

3. The matrix  $\mathcal{S}^{\frac{1}{2}}$  is then defined as

$$\mathcal{S}^{\frac{1}{2}} = U \mathcal{S}_{\text{diag}}^{\frac{1}{2}} U^\dagger. \quad (21)$$

4. Finally the matrix  $\mathcal{S}^{-\frac{1}{2}}$  is defined as

$$\mathcal{S}^{-\frac{1}{2}} = (\mathcal{S}^{\frac{1}{2}})^{-1} = U \mathcal{S}_{\text{diag}}^{-\frac{1}{2}} U^\dagger. \quad (22)$$

By multiplying both left side of (5) by  $\mathcal{S}^{-\frac{1}{2}}$  defined in Equation (22) and considering that  $\mathcal{S}^{-\frac{1}{2}} \mathcal{S} = \mathbb{I}$ , after some calculation we obtain

$$\mathcal{S}^{-\frac{1}{2}} \mathcal{H} \mathcal{C} = E \mathcal{S}^{-\frac{1}{2}} \mathcal{S} \mathcal{C} \implies \tilde{\mathcal{H}} \tilde{\mathcal{C}} = E \tilde{\mathcal{C}} \quad (23)$$

which can be rewritten as a standard eigenvalue problem

$$\tilde{\mathcal{H}} \tilde{\mathcal{C}} = E \tilde{\mathcal{C}} \quad (24)$$

where

$$\tilde{\mathcal{H}} = \mathcal{S}^{-\frac{1}{2}} \mathcal{H} \mathcal{S}^{-\frac{1}{2}} \quad (25)$$

$$\tilde{\mathcal{C}} = \mathcal{S}^{\frac{1}{2}} \mathcal{C}. \quad (26)$$

The standard eigenvalue problem (24) has the same spectrum of eigenvalues as the generalized eigenvalue problem (5).

## References

- [Abramowitz and Stegun, 1964] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office.
- [Ahn et al., 2022] Ahn, K., Zhang, J., and Sra, S. (2022). Understanding the unstable convergence of gradient descent. In *International Conference on Machine Learning*, pages 247–257. PMLR.

- [Aragón et al., 2019] Aragón, F. J., Goberna, M. A., López, M. A., and Rodríguez, M. M. (2019). *Nonlinear optimization*. Springer.
- [Arora et al., 2022] Arora, S., Li, Z., and Panigrahi, A. (2022). Understanding gradient descent on the edge of stability in deep learning. In *International Conference on Machine Learning*, pages 948–1024. PMLR.
- [Aziz and Slaman, 1991] Aziz, R. A. and Slaman, M. J. (1991). An examination of ab initio results for the helium potential energy curve. *J. Chem. Phys.*, 94(12):8047.
- [Bär and Bopp, 1988] Bär, H.-J. and Bopp, P. (1988). *Mh kalos and pa whitlock: Monte carlo methods, volume i: Basics*, john wiley and sons, new york, chichester, brisbane, toronto and singapore 1986, library of congress qa298. k35 1986. 186 seiten mit einem index, preis:£ 29.00.
- [Barletta and Kievsky, 2001] Barletta, P. and Kievsky, A. (2001). Variational description of the helium trimer using correlated hyperspherical harmonic basis functions. *Physical Review A*, 64(4):042514.
- [Bazak et al., 2020] Bazak, B., Valiente, M., and Barnea, N. (2020). Universal short-range correlations in bosonic helium clusters. *Phys. Rev. A*, 101(1).
- [Bianchi et al., 2009] Bianchi, L., Dorigo, M., Gambardella, L. M., and Gutjahr, W. J. (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8(2):239–287.
- [Braaten and Hammer, 2006] Braaten, E. and Hammer, H.-W. (2006). Universality in few-body systems with large scattering length. *Physics Reports*, 428(5-6):259–390.
- [Choo et al., 2020] Choo, K., Mezzacapo, A., and Carleo, G. (2020). Fermionic neural-network states for ab-initio electronic structure. *Nature Comm.*, 11:2368.
- [Cohen-Tannoudji et al., 2019] Cohen-Tannoudji, C., Diu, B., and Laloë, F. (2019). *Quantum mechanics, volume 3: fermions, bosons, photons, correlations, and entanglement*. John Wiley & Sons.
- [Curry, 1944] Curry, H. B. (1944). The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258–261.
- [Delves, 1958] Delves, L. (1958). Tertiary and general-order collisions. *Nuclear Physics*, 9(3):391–399.
- [Efimov, 1970] Efimov, V. (1970). Energy levels arising from resonant two-body forces in a three-body system. *Physics Letters B*, 33(8):563–564.
- [Faddeev and Merkuriev, 1993] Faddeev, L. and Merkuriev, S. (1993). Quantum scattering theory for several particle systems, ser. *Mathematical physics and applied mathematics*. Dordrecht: Kluwer Academic, 11.
- [Faddeev, 1965] Faddeev, L. D. (1965). *Mathematical aspects of the three-body problem in the quantum scattering theory*, volume 69. Israel Program for Scientific Translations.
- [Fogel, 1998] Fogel, D. B. (1998). *Evolutionary computation: the fossil record*. Wiley-IEEE Press.

- [Gattobigio et al., 2011] Gattobigio, M., Kievsky, A., and Viviani, M. (2011). Spectra of helium clusters with up to six atoms using soft-core potentials. *Phys. Rev. A*, 84(5):052503.
- [Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- [Griffiths and Schroeter, 2018] Griffiths, D. J. and Schroeter, D. F. (2018). *Introduction to quantum mechanics*. Cambridge university press.
- [Grisenti et al., 2000] Grisenti, R., Schöllkopf, W., Toennies, J. P., Hegerfeldt, G. C., Köhler, T., and Stoll, M. (2000). Determination of the bond length and binding energy of the helium dimer by diffraction from a transmission grating. *Physical Review Letters*, 85(11):2284.
- [Hammond et al., 1994] Hammond, B. L., Lester, W. A., and Reynolds, P. J. (1994). *Monte Carlo methods in ab initio quantum chemistry*, volume 1. World Scientific.
- [Hastings, 1970] Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.
- [Hinton et al., 2012] Hinton, G., Srivastava, N., and Swersky, K. (2012). Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on*, 14(8):2.
- [Kievsky et al., 2021] Kievsky, A., Gattobigio, M., Girlanda, L., and Viviani, M. (2021). Efimov physics and connections to nuclear physics. *Annual Review of Nuclear and Particle Science*, 71:465–490.
- [Kievsky et al., 2008] Kievsky, A., Rosati, S., Viviani, M., Marcucci, L. E., and Girlanda, L. (2008). A high-precision variational approach to three- and four-nucleon bound and zero-energy scattering states. *J. Phys. G*, 35(6):063101.
- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [Kirkpatrick et al., 1983] Kirkpatrick, S., Gelatt Jr, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.
- [Klahn and Bingel, 1977] Klahn, B. and Bingel, W. A. (1977). The convergence of the rayleigh-ritz method in quantum chemistry: II. investigation of the convergence for special systems of slater, gauss and two-electron functions. *Theoretica chimica acta*, 44:27–43.
- [Kolganova et al., 2011] Kolganova, E. A., Motovilov, A. K., and Sandhas, W. (2011). The 4he trimer as an efimov system. *Few-Body Systems*, 51(2-4):249–257.
- [Kuenzer and Hofer, 2019] Kuenzer, U. and Hofer, T. S. (2019). A four-dimensional numerov approach and its application to the vibrational eigenstates of linear triatomic molecules—the interplay between anharmonicity and inter-mode coupling. *Chemical Physics*, 520:88–99.
- [Kuenzer et al., 2016] Kuenzer, U., Sorarù, J.-A., and Hofer, T. S. (2016). Pushing the limit for the grid-based treatment of schrödinger’s equation: a sparse numerov approach for one, two and three dimensional quantum problems. *Physical Chemistry Chemical Physics*, 18(46):31521–31533.

- [Kukulin and Krasnopol'Sky, 1977] Kukulin, V. and Krasnopol'Sky, V. (1977). A stochastic variational method for few-body systems. *Journal of Physics G: Nuclear Physics*, 3(6):795.
- [Kunitski et al., 2015] Kunitski, M., Zeller, S., Voigtsberger, J., Kalinin, A., Schmidt, L. P. H., Schöffler, M., Czasch, A., Schöllkopf, W., Grisenti, R. E., Jahnke, T., et al. (2015). Observation of the efimov state of the helium trimer. *Science*, 348(6234):551–555.
- [Lanczos, 1950] Lanczos, C. (1950). An iteration method for the solution of the eigenvalue problem of linear differential and integral operators.
- [Lazauskas, 2018] Lazauskas, R. (2018). Solution of the n- he 4 elastic scattering problem using the faddeev-yakubovsky equations. *Physical Review C*, 97(4):044002.
- [Loshchilov and Hutter, 2016] Loshchilov, I. and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- [Marcucci et al., 2020] Marcucci, L. E., Dohet-Eraly, J., Girlanda, L., Gnech, A., Kievsky, A., and Viviani, M. (2020). The hyperspherical harmonics method: a tool for testing and improving nuclear interaction models. *Frontiers in Physics*, 8:69.
- [Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.
- [Mikolov et al., 2012] Mikolov, T. et al. (2012). Statistical language models based on neural networks. *Presentation at Google, Mountain View, 2nd April*, 80(26).
- [Nelder and Mead, 1965] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.
- [Numerov, 1927] Numerov, B. (1927). Note on the numerical integration of  $d^2x/dt^2 = f(xt)$ . *Astronomische Nachrichten*, 230:359.
- [Pascanu et al., 2013] Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318. Pmlr.
- [Piela, 2006] Piela, L. (2006). *Ideas of quantum chemistry*. Elsevier.
- [Powell, 1964] Powell, M. J. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162.
- [Recchia et al., 2022a] Recchia, P., Kievsky, A., Girlanda, L., and Gattobigio, M. (2022a). Gaussian parametrization of efimov levels: remnants of discrete scale invariance. *Few-Body Systems*, 63(1):1–10.
- [Recchia et al., 2022b] Recchia, P., Kievsky, A., Girlanda, L., and Gattobigio, M. (2022b). Subleading contributions to n-boson systems inside the universal window. *Physical Review A*, 106(2):022812.
- [Roudnev and Yakovlev, 2000] Roudnev, V. and Yakovlev, S. (2000). Investigation of 4he3 trimer on the base of faddeev equations in configuration space. *Chemical Physics Letters*, 328(1-2):97–106.

- [Ruder, 2016] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.
- [Shaw, 2020] Shaw, R. A. (2020). The completeness properties of gaussian-type orbitals in quantum chemistry. *International Journal of Quantum Chemistry*, 120(17):e26264.
- [Smith, 1962] Smith, F. T. (1962). A symmetric representation for three-body problems. i. motion in a plane. *Journal of Mathematical Physics*, 3(4):735–748.
- [Smith, 2017] Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 464–472. IEEE.
- [Smith and Topin, 2019] Smith, L. N. and Topin, N. (2019). Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, volume 11006, pages 369–386. SPIE.
- [Suzuki et al., 1998] Suzuki, Y., Suzuki, M., and Varga, K. (1998). *Stochastic variational approach to quantum-mechanical few-body problems*, volume 54. Springer Science & Business Media.
- [Zeller et al., 2016] Zeller, S., Kunitski, M., Voigtsberger, J., Kalinin, A., Schottelius, A., Schober, C., Waitz, M., Sann, H., Hartung, A., Bauer, T., et al. (2016). Imaging the he2 quantum halo state using a free electron laser. *Proceedings of the National Academy of Sciences*, 113(51):14651–14655.
- [Zernike and Brinkman, 1935] Zernike, F. and Brinkman, H. (1935). Hypersphärische funktionen und die in sphärischen bereichen orthogonalen polynome. In *Proc. Akad. Wet. Amsterdam*, volume 38, pages 161–170.