



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Role of shift constant in Energy Shifted SAV for Hamiltonian Systems

### Citation for published version:

Zama, F, Ducceschi, M & Bilbao, S 2024, Role of shift constant in Energy Shifted SAV for Hamiltonian Systems. in *12th International Conference on Mathematical Modeling in Physical Sciences (IC-MSQUARE 2023) 28/08/2023 - 31/08/2023 Belgrade, Serbia*. vol. 2701, 012089, Journal of Physics: Conference Series, IOP Publishing, pp. 1-10, 12th International Conference on Mathematical Modeling in Physical Sciences, Belgrade, Serbia, 28/08/23. <https://doi.org/10.1088/1742-6596/2701/1/012089>

### Digital Object Identifier (DOI):

[10.1088/1742-6596/2701/1/012089](https://doi.org/10.1088/1742-6596/2701/1/012089)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

12th International Conference on Mathematical Modeling in Physical Sciences (IC-MSQUARE 2023) 28/08/2023 - 31/08/2023 Belgrade, Serbia

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



PAPER • OPEN ACCESS

## Role of shift constant in Energy Shifted SAV for Hamiltonian Systems

To cite this article: F Zama *et al* 2024 *J. Phys.: Conf. Ser.* **2701** 012089

View the [article online](#) for updates and enhancements.

You may also like

- [The impact of graphene oxide particles on viscosity stabilization for diluted polymer solutions using in enhanced oil recovery at HTHP offshore reservoirs](#)  
Ba Dung Nguyen, Trung Kien Ngo, Truong Han Bui et al.
- [A model and simulation of lattice vibrations in a superabundant vacancy phase of palladium–deuterium](#)  
M R Staker
- [White light diffraction phase microscopy for imaging of red blood cells for different storage times](#)  
Özlem Kocahan, Nesrin Çelebiolu and Merve Uyank



The Electrochemical Society

Advancing solid state & electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research



# Role of shift constant in Energy Shifted SAV for Hamiltonian Systems

F Zama<sup>1</sup>, M Ducceschi<sup>2</sup>, S Bilbao<sup>3</sup>

<sup>1</sup> Department of Mathematics, University of Bologna, Bologna, Italy

<sup>2</sup> Department of Industrial Engineering, University of Bologna, Bologna, Italy

<sup>3</sup> Acoustics and Audio Group, University of Edinburgh, 12 Nicolson Square, Edinburgh, United Kingdom EH8 9DF

E-mail: [fabiana.zama@unibo.it](mailto:fabiana.zama@unibo.it)

**Abstract.** In recent years, considerable work has been devoted to the design of energy-stable numerical methods, a class of geometrical integration technique in which the discrete energy or pseudo-energy remains conserved. This property finds practical applications in systems for which the energy is bounded from below since the growth of the solutions is then itself bounded, yielding a form of stability. Such a property may be further exploited to transform the energy overall into a quadratic form, representing the core of recent numerical techniques such as the invariant energy quadratization (IEQ) and the scalar auxiliary variable (SAV) approaches. These methods have been applied to a large class of problems due to their remarkable efficiency. Yet, several aspects of such techniques have seen little investigation. In this work, the role of the “shift” constant in the expression of the potential energy in Hamiltonian systems is investigated. This is a positive gauge constant that may be used to augment the expression of the energy. Since Hamilton’s equations are given in terms of gradients of the Hamiltonian, such a constant has no influence on the resulting dynamics in the continuous system. However, empirical evidence has suggested that the convergence properties of the associated SAV approaches are affected by the magnitude of the shift factor. In this work, the behaviour of the relative global error of SAV schemes in the cases of the harmonic and Duffing oscillators is numerically investigated. The results reveal an optimal shift factor that increases the convergence rate. Using the optimal shift factor, the proposed method displays variable order accuracy ranging from second to twelfth order.

## 1. Introduction

Hamiltonian systems are widespread in physics, describing a large class of dynamical systems. Under time-invariant or autonomous conditions, the Hamiltonian corresponds to the physical energy of a given system, and is conserved. In the simulation setting, mimicking this property through a conserved numerical energy (or pseudo-energy) conservation is, thus, a desirable feature. In many numerical designs, the enforcing of such a conservation property is achieved by means of fully implicit schemes requiring the iterative solution of nonlinear algebraic equations at each time step, or via the exact evaluation of continuous-time integrals [1, 2]. In the former case, the iterative nature of the time-stepping routine often represents a computational bottleneck, while in the latter, exact numerical energy conservation is achieved in the few cases where the integrals can be computed exactly. Recently, fully explicit, energy-stable numerical designs for a class of Hamiltonian systems have been proposed [3], overcoming all such limitations. These



designs are part of a larger class of techniques rooted in energy quadratisation, and appearing under various guises such as the Invariant Energy Quadratisation (IEQ) [4, 5], and the Scalar Auxiliary Variable (SAV) methods [6]. These methods have seen a growing body of applications, particularly in gradient flow systems (see e.g. [7, 8]), and including diagonally-implicit Runge-Kutta methods [9].

Hamilton's dynamical equations are derived from the gradients of the Hamiltonian, and thus the equations are gauge-invariant: that is, Hamiltonians differing by a constant yield the same equations. Numerically, however, gauge constants have been shown to influence the behaviour of the numerical error, as in the Navier-Stokes simulation routines presented in [10]. The influence of the gauge constant in SAV-like methods for Hamiltonian systems will be the subject of this article, building on previous work by the authors [3]. A rule for selecting the shift constant is tested here on two scalar problems: the simple harmonic oscillator and the Duffing oscillator (a particularly simple nonlinear oscillator). The experimental analysis reveals that an optimal shift factor exists for a given potential, which can improve the local convergence rate up to several orders, and which may be obtained as a fraction of the system's energy.

The manuscript is organized as follows: Section 2 describes the proposed method, applied to the test cases described in Section 3. Section 4 includes the numerical tests and the estimates of the optimal gauge constant's values.

## 2. Energy-shifted Hamiltonian systems

Hamilton's equations are a set of ordinary differential equations (ODEs) giving the rate of change of the displacements and momenta of a system of particles. In this work, a single particle with position  $q(t)$  and momentum  $p(t)$  is considered, for which:

$$\dot{q} = \frac{\partial H}{\partial p} \quad \dot{p} = -\frac{\partial H}{\partial q}. \quad (1)$$

Here,  $H = H(q, p) : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the Hamiltonian, restricted here to the form:

$$H(p, q) = \frac{1}{2}p^2 + V(q), \quad (2)$$

where  $V(q)$  is the potential energy of the particle. Here, and in the following, the Hamiltonian is scaled by the particle's mass for the sake of conciseness. It is furthermore assumed that:

$$V(q) \geq 0 \quad \forall q \in \mathbb{R}, \quad (3)$$

implying  $H \geq 0 \quad \forall (p, q)$ . The non-negativity of the potential  $V$  is verified in many Hamiltonian systems, though not all – amongst others, the gravitational potential is an exception. When (3) holds, the potential energy may be written as [3]:

$$V = \frac{1}{2}\psi^2, \quad (4)$$

where  $\psi \in \mathbb{R}_0^+$  is referred to as the “auxiliary variable” [3, 4]. Consequently, Hamilton's equations take the form:

$$\dot{q} = p \quad \dot{p} = -g\psi. \quad (5)$$

with  $g := \frac{\partial \psi}{\partial q}$ .

The equations in (5) can be combined through differentiation, and the rate of change of the auxiliary variable may itself be obtained by a simple application of the chain rule. Together, these yield the following system, forming the basis for SAV-like approaches:

$$\ddot{q} = -g\psi \quad \dot{\psi} = g\dot{q}. \quad (6)$$

Initialisation is provided by:

$$q(0) = q_0 \quad \dot{q}(0) = p_0 \quad \psi(0) = \sqrt{2V(q_0)}, \quad (7)$$

where  $q_0, p_0$  are constants.

Note that the same equations can be obtained by replacing  $V$  with  $V_\epsilon$  in (2), a shifted potential defined as:

$$V_\epsilon := V + \frac{\epsilon}{2} \quad \epsilon > 0,$$

where the non-negative shift  $\epsilon$  has the interpretation of a gauge constant.

### 2.1. Time stepping routine

System (6) is now integrated in time using the method proposed in [3]. This is a finite difference scheme discretizing the time domain into uniformly-spaced grid intervals at the times  $t_n = nk$ , where  $n \in \mathbb{N}$  is the time index, and  $k$  is time step. This grid is used to compute the discrete-time position  $q^n \approx q(t_n)$ . Moreover, an interleaved grid centered at the midpoints  $t_n + \frac{k}{2}$  is employed to approximate the auxiliary variable  $\psi^{n+\frac{1}{2}} \approx \psi(t_n + \frac{k}{2})$ . An approximation to (6) is obtained as:

$$q^{n+1} = 2q^n - q^{n-1} - \frac{k^2}{2} g_\epsilon^n \left( \psi^{n+\frac{1}{2}} + \psi^{n-\frac{1}{2}} \right) \quad \psi^{n+\frac{1}{2}} = \psi^{n-\frac{1}{2}} + \frac{1}{2} g_\epsilon^n (q^{n+1} - q^{n-1}), \quad (8)$$

where:

$$g_\epsilon^n := \frac{\partial \psi}{\partial q} \Big|_{q=q^n} = \frac{1}{\sqrt{2V(q^n) + \epsilon}} \frac{\partial V}{\partial q} \Big|_{q=q^n}. \quad (9)$$

All finite difference approximations in the scheme are centered, making the scheme reversible and (at least) second-order accurate. Furthermore, it possesses a conserved, non-negative energy of the form:

$$\mathfrak{h}^{n+\frac{1}{2}} = \frac{(q^{n+1} - q^n)^2}{2k^2} + \frac{(\psi^{n+\frac{1}{2}})^2}{2} = \mathfrak{h}^{\frac{1}{2}} \geq 0, \quad (10)$$

from which bounds on  $|q^n - q^{n-1}|$  and  $|\psi^{n-\frac{1}{2}}|$  in terms of the initial energy  $\mathfrak{h}^{\frac{1}{2}}$  are easily derived, leading to unconditional stability.

The numerical initialisation of scheme (8) is given in terms of the exact solution  $q(t)$ , as:

$$q^0 = q(0) \quad q^1 = q(k) \quad \psi^{\frac{1}{2}} = \sqrt{2V(q(k/2)) + \epsilon}. \quad (11)$$

Whilst  $q$  is here a scalar, scheme (8) may be generalised to the vector case; an explicit update exists, making the scheme particularly attractive from the standpoint of real-time simulation [3, 11–13].

### 3. Test Cases

Two test problems are considered, with closed-form solution. For both problems, in order to simplify the analysis, null velocity initial conditions are considered, such that  $p_0 = 0$  in (7).

*Simple Harmonic Oscillator* The first test case is the classical simple harmonic oscillator, for which:

$$V(q) = \frac{\nu^2 q^2}{2}, \quad (12)$$

where  $\nu > 0$  is the oscillator's angular frequency in radians per second. An exact solution to (6) exists as:

$$q(t) = q_0 \cos(\nu t). \quad (13)$$

Substituting the definition of  $V(q)$  as per (12) in (9), one has:

$$g_\epsilon^n = \frac{\nu^2 q^n}{\sqrt{\nu^2 (q^n)^2 + \epsilon}}. \quad (14)$$

*Duffing Oscillator* The second test case is the Duffing oscillator, a nonlinear oscillator with a mixed linear/cubic restoring force. It models various oscillatory systems, such as the simple pendulum under a moderately large vibration amplitude [14] and, in the distributed case, it governs the dynamics of plates, shells and strings [12, 15]. In the scalar case considered here, one has:

$$V(q) = \frac{\nu^2 q^2}{2} + \frac{\gamma q^4}{4}. \quad (15)$$

The sign of  $\gamma$  may be positive or negative, leading to a ‘‘hardening’’ or a ‘‘softening’’ nonlinearity, respectively. Because (3) is required to hold, only the case  $\gamma \geq 0$  is considered here. An analytic solution exists, generalising (13), as:

$$q(t) = q_0 \operatorname{cn} \left( \sqrt{\nu^2 + \gamma q_0^2} t; \frac{\gamma q_0^2}{2\gamma q_0^2 + 2\nu^2} \right). \quad (16)$$

Here,  $\operatorname{cn}(z; m)$  is the Jacobi elliptic function of argument  $z$  and parameter  $m$ . Using (15) in (9), the nonlinear gradient is given in this case by:

$$g_\epsilon(q^n) = \frac{\nu^2 q^n + \gamma (q^n)^3}{\sqrt{\nu^2 (q^n)^2 + \gamma \frac{q^4}{2} + \epsilon}}. \quad (17)$$

#### 4. Numerical Results

To investigate the impact of the shift constant on the convergence properties of scheme (8), the influence of  $\epsilon$  on the relative error is first assessed. The relative error is here defined as:

$$\delta_\epsilon^N := \frac{|q^N - q(t_N)|}{|q(t_N)|}, \quad (18)$$

that is, the relative difference between the output of scheme (8) at the final time step  $N$  and the corresponding exact solution  $q(t_N)$ . Note that  $q^n$ , from (8), depends on  $\epsilon$  via  $g_\epsilon^n$ , and, hence, so does  $\delta_\epsilon^N$ . Then, define:

$$\epsilon^* := \arg \min_{\epsilon} \delta_\epsilon^N \quad \epsilon_{\min} \leq \epsilon \leq \epsilon_{\max}, \quad (19)$$

provided this minimum exists. Here,  $\epsilon^*$  will be sought as a fraction of  $\mathfrak{h}_0$  the total numerical energy in the absence of shift, such that:

$$\epsilon := \eta \mathfrak{h}_0 \quad \eta_{\min} \leq \eta \leq \eta_{\max}, \quad (20)$$

with

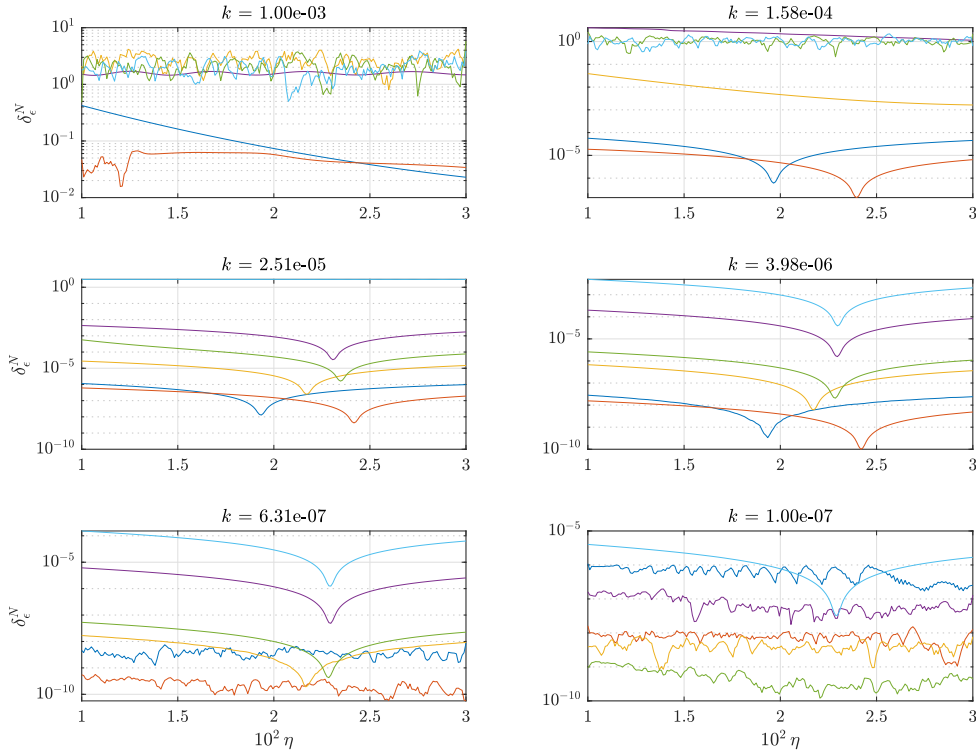
$$\mathfrak{h}_0 := \frac{(q(k) - q(0))^2}{2k^2} + V(q(k/2)). \quad (21)$$

The total energy  $\mathfrak{h}_0$  provides in this case a useful scaling factor. Hence, the global minimum will be defined in terms of the scaled shift, as:

$$\epsilon^* := \eta^* \mathfrak{h}_0. \quad (22)$$

#### 4.1. Simple Harmonic Oscillator

The behaviour of the relative error  $\delta_\epsilon^N$  is first checked as a function of  $\eta$ , for a set of frequencies  $\Omega = \{\nu \mid \nu \in (1, 2, 5, 10, 20, 50) \cdot 10^2\}$ , and under various choices of the time step  $k$ . It is convenient to perform a check on a coarse grid along the  $\eta$  axis, and then to restrict the search to a useful range. The coarse search reveals the presence of minima in the range  $0.01 \leq \eta^* \leq 0.03$ , as seen in Figure 1. An analysis of this figure reveals that a single (and, hence, global) minimum



**Figure 1.** Coarse behaviour of the relative error  $\delta_\epsilon^N$  as a function of the scaled shift  $\eta$ , under various choices of the time step  $k$ , given on top of each panel, and linear frequency  $\nu$ , given by different colours. Colour scheme:  $\nu = 100$  (blue), 200 (red), 500 (yellow), 1000 (purple), 2000 (green), 5000 (cyan). Here,  $N = \text{floor}(0.3/k)$ , and the  $\eta$  axis is sampled using 500 linearly spaced grid points between 0.01 and 0.03. When  $10^{-6} \lesssim k \lesssim 10^{-5}$ , a single, global minimum is detected  $\forall \nu$ . For larger values of  $k$ , a single minimum is detected only for frequencies such that  $\nu k \lesssim 0.1$  whereas for smaller values of  $k$ , the effects of round-off are visible as a noise-like distortion. For all tests, the system is initialised with  $q_0 = 1$ ,  $p_0 = 0$ .

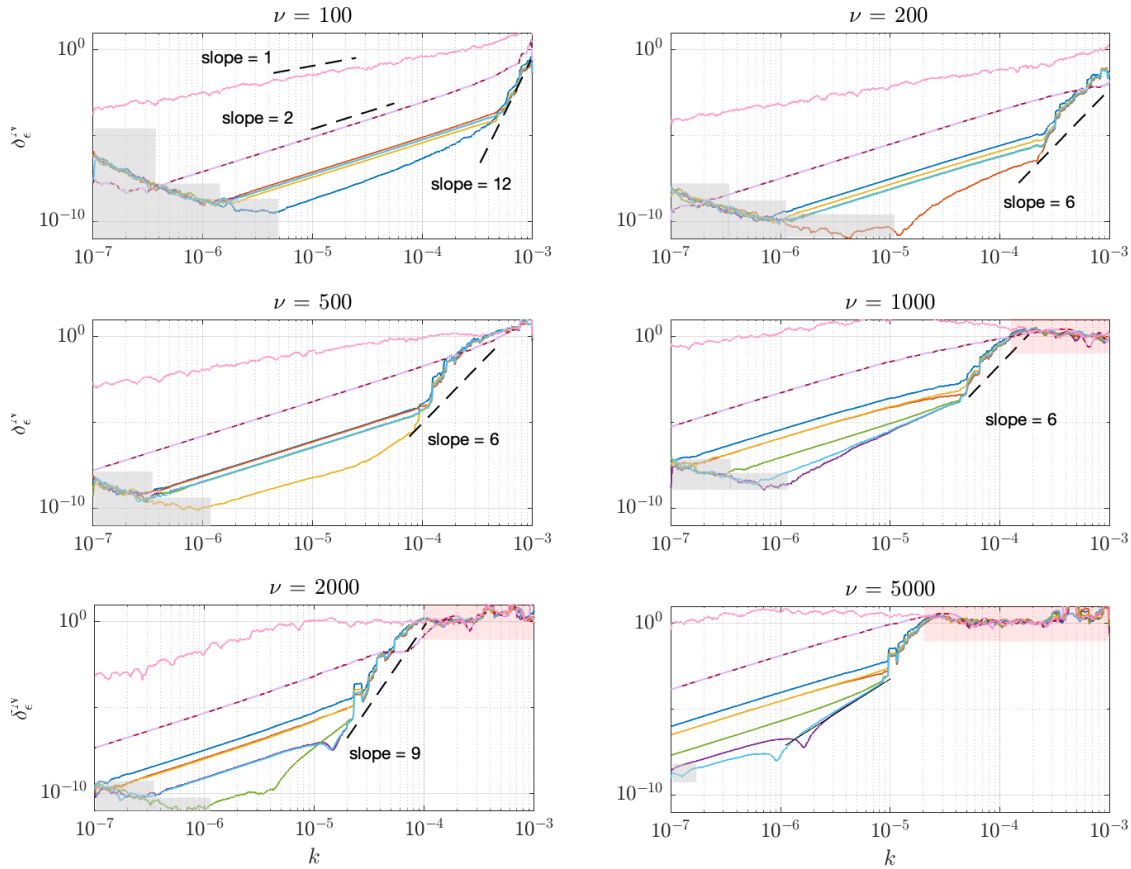
exists for  $\delta_\epsilon^N$  when the time step  $k$  becomes sufficiently small compared to the radian frequency  $\nu$ , such that  $\nu k \lesssim 0.1$ . When  $10^{-6} \lesssim k \lesssim 10^{-5}$ , a clear global minimum appears for all frequencies. The definition of the scaled shift  $\eta$  as per (20) results appropriate from the inspection of this figure, where the minimum points  $\eta^*$  appear to be more or less aligned for all frequencies, though a small dependency of  $\eta^*$  on  $\nu$  is nonetheless observed. Interestingly, for smaller values of the time step  $k$ , the behaviour of  $\delta_\epsilon^N$  is dominated by round-off errors appearing as a noise-like distortion.

This qualitative analysis suggests retrieving  $\eta^*$  numerically, using an appropriate value for the time step  $k$  and a very fine grid along  $\eta$ . The results are summarised in Table 1. Note

$\nu$	100	200	500	1000	2000	5000
$10^2 \eta^*$	1.931	2.420	2.171	2.293	2.283	2.292

**Table 1.** Numerically-computed global minimum points from Figure 1. For the calculation,  $k = 3 \cdot 10^{-6}$  was used for  $\nu = \{100, 200, 500\}$ , and  $k = 10^{-6}$  for  $\nu = \{1000, 2000, 5000\}$ . In both cases, the  $\eta$  axis was scanned between 0.01 and 0.03 using 5000 grid points.

that, while Figure 1 displays the case  $q_0 = 1$ , the minimum points  $\eta^*$  are independent of the magnitude of the initial condition  $q_0$  for fixed  $\nu$ . Furthermore, such values for  $\eta^*$  are independent of  $k$  for fixed  $\nu$ , provided that  $k$  is small enough to yield sufficient resolution, but large enough to keep away from round-off errors. In practice, the optimal scaled shift value appears to be a



**Figure 2.** Convergence curves for the simple harmonic oscillator, under various choices of the natural frequency  $\nu$  as indicated on top of each panel, and scaled shift constant  $\eta$ , where  $\eta \in W \cup \{0, 10^{12}\}$ . The leapfrog scheme is also used. Colour scheme:  $\eta = 1.9193 \cdot 10^{-2}$  (blue),  $2.4191 \cdot 10^{-2}$  (red),  $2.1692 \cdot 10^{-2}$  (yellow),  $2.2931 \cdot 10^{-2}$  (purple),  $2.2839 \cdot 10^{-2}$  (green),  $2.2923 \cdot 10^{-2}$  (cyan), 0 (pink),  $10^{12}$  (dashed ilac), leapfrog (burgundy). Grey-shaded areas represent portions of the plane dominated by round-off errors, whilst red-shaded areas correspond to time steps yielding insufficient resolution (no convergence).

function of the radian frequency only, so that  $\eta^* = \eta^*(\nu)$ , and is thus independent of  $k$  and  $q_0$ . Let  $W = \{\eta \mid \eta = \eta^*(\nu), \nu \in \Omega\}$  denote the set of the computed minimum points from Table



1. The convergence rate of scheme (8) as a function of the time step  $k$  is now checked, as per Figure 2. Each panel considers one frequency from the set  $\Omega$ , and the nine coloured lines correspond to the eight values of the shift constant obtained from  $W \cup \{0, 10^{12}\}$  (that is, from the set  $E = \{\epsilon \mid \epsilon \in (W \cup \{0, 10^{12}\}) \cap \mathfrak{h}_0\}$ ), plus the second-order leapfrog scheme, used as a reference. For all panels, as expected, the value of  $\delta_\epsilon^N$  is considerably smaller in correspondence of  $\epsilon^*$ . When  $k \approx 10^{-5}$  (a typical value of the time step in acoustics simulation), the relative error is several orders of magnitude smaller for the energy-shifted SAV scheme compared to the leapfrog scheme. These results suggest computing  $\epsilon^*$  with a precision at least 4 significant digits, since a smaller precision yields rather large amplifications of the relative error  $\delta_\epsilon^N$ . This conclusion is supported by the behaviour of the curves in Figure 1, all presenting a very steep descent in the neighbourhood of  $\eta^*$ .

The dependence of the convergence rate on  $\epsilon$  is another evident feature of the curves in Figure 2, and one of the most interesting findings of this work. Whilst scheme (8) remains, generally, second-order convergent, the rate of convergence seems to be largely affected by the magnitude of the shift factor. Locally, slopes range from 3 to 12 when  $\epsilon = \epsilon^*$ . When  $\epsilon$  is chosen in some neighbourhood around  $\epsilon^*$ , some improvements are still observed for larger time steps. This suggests that increasing the precision on the estimate of  $\epsilon^*$  may improve the convergence rates further, though this behaviour can only be understood by a formal analysis of the structure of the error, left as future work.

The limiting cases  $\epsilon = \{0, \infty\}$  are also worth commenting. It is remarked that the somewhat “natural” choice  $\epsilon = 0$  yields a very slow convergence rate, if any: for larger values of the oscillator frequency  $\nu$ , the convergence of the scheme seems to be completely impaired. On the other hand, a very large value of  $\epsilon$  produces a scheme virtually indistinguishable from the leapfrog algorithm, as one can show immediately from (8) (in this case,  $g_\epsilon^n \approx (\sqrt{\epsilon})^{-1} \partial V / \partial q$ ,  $\psi^{n+\frac{1}{2}} \approx \sqrt{\epsilon}$ , from which the leapfrog is recovered).

Before proceeding, it is worth remarking that the case  $\epsilon = 0$  may be treated differently. First, note that, according to definition (14),  $g_\epsilon^n = \nu \text{sign}(q^n)$  when  $\epsilon = 0$ . Whilst  $\psi$ , from (4) was restricted to be a positive constant, it may well change sign according to  $\psi = \pm \sqrt{2V + \epsilon}$ . Selecting the sign accordingly, one may then get  $g_\epsilon^n = \nu$ , yielding a different three-step scheme. The assessment of this scheme is left as future work.

#### 4.2. Duffing oscillator

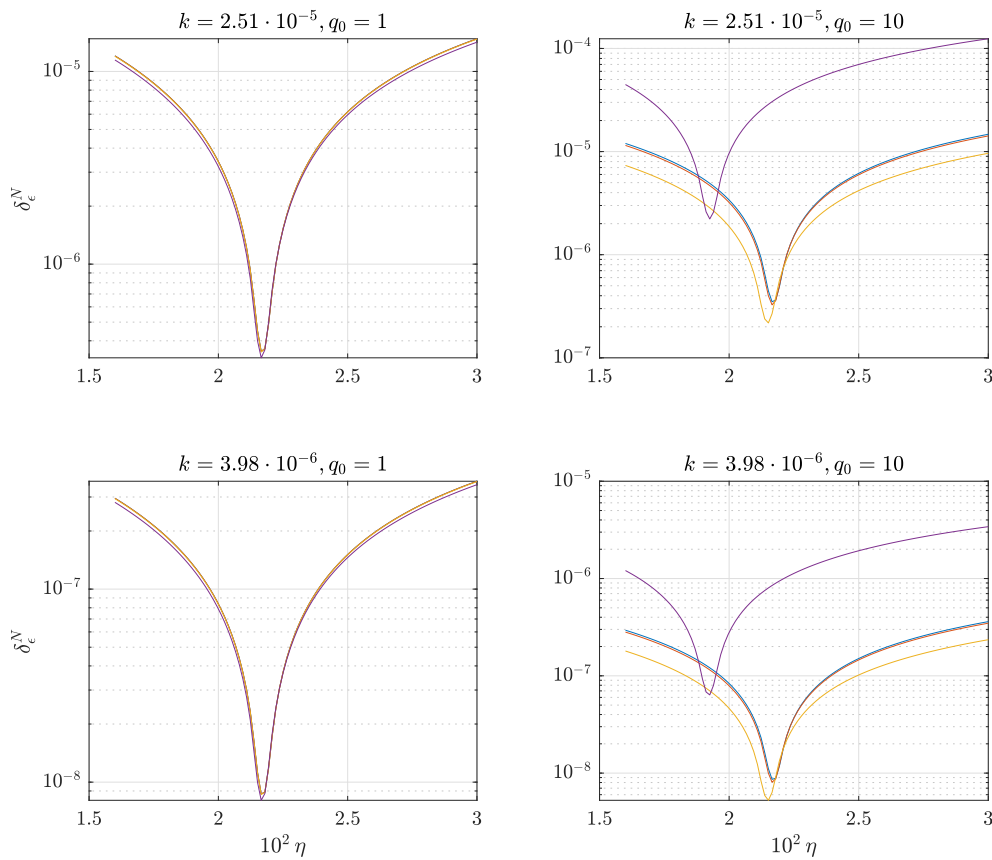
The analysis of the Duffing oscillator may be done analogously. It is convenient in this case to keep the linear frequency  $\nu$  fixed, and to study the behaviour of the scheme for varying  $(\gamma, q_0)$ : both these parameters influence the amplitude-dependent nonlinearity of the system. In the following,  $\gamma \in \{0.1, 1, 10, 100\}$  and  $q_0 \in \{0.1, 1, 10\}$ .

First, a coarse check of the  $\eta$  axis is performed, as per Figure 3. In the figure, plots of  $\delta_\epsilon^N$  under two different choices of the initial condition  $q_0$  and time step  $k$  are given. Whilst a single, global minimum is recovered for fixed  $\nu, \gamma, k$ , the value of  $\eta^*$  now clearly depends on the initial condition  $q_0$ . Thus,  $\eta^* = \eta^*(q_0, \nu, \gamma)$ . Table 2 reports values of the minimum points  $\eta^*$  under various choices for  $\gamma, q_0$ . Note that, as expected, for small values of the nonlinear parameter  $\gamma$  and  $q_0$ , the same value as from Table 1 is recovered ( $\eta^* = 0.02171$  for  $\nu = 500$ ).

Let now  $V = \{\eta \mid \eta = \eta^*(\nu, \gamma, q_0), \nu = 500, \gamma \in \{0.1, 1, 10, 100\}, q_0 \in \{0.1, 1, 10\}\}$  be the set of the optimal scaled shift values from Table 2. In Figure 4, the behaviour of the relative error  $\delta_\epsilon^N$  is assessed against the time step  $k$ , using the values of the shift constant from  $V \cup \{0\}$ , plus the leapfrog scheme. Much like the simple harmonic oscillator case, it can be appreciated that the convergence rate improves drastically whenever  $\epsilon = \epsilon^*$ . Values in the neighbourhood of  $\epsilon^*$  also lead to a considerably faster convergence for larger  $k$  compared to the leapfrog. Again, the no-shift case ( $\epsilon = 0$ ) leads to poor or no convergence.

$\gamma$	0.1	1.0	10	100
$10^2 \eta^*$ , $q_0 = 0.1$	2.171	2.171	2.171	2.171
$10^2 \eta^*$ , $q_0 = 1.0$	2.171	2.171	2.171	2.168
$10^2 \eta^*$ , $q_0 = 10$	1.919	2.168	2.150	1.921

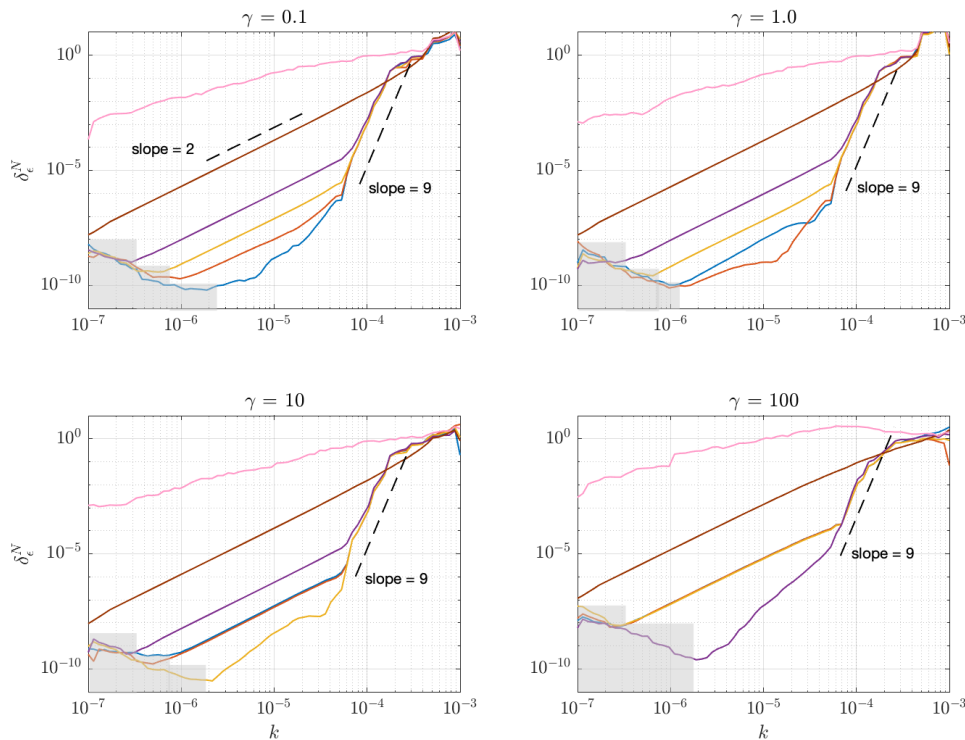
**Table 2.** Numerically-computed global minimum points for the Duffing oscillator, with linear frequency  $\nu = 500$ . Nonlinear parameter  $\gamma$  and initial condition  $q_0$  as indicated. For the numerical search,  $k = 3 \cdot 10^{-6}$  was used, and the  $\eta$  axis was scanned between 0.015 and 0.03 using 5000 grid points.



**Figure 3.** Coarse behaviour of the relative error  $\delta_\epsilon^N$  for the Duffing oscillator as a function of the scaled shift  $\eta$ , under various choices of the time step  $k$ , given on top of each panel, and  $\gamma$ , given by different colours. Colour scheme:  $\gamma = 0.1$  (blue), 1.0 (red), 10 (yellow), 100 (purple). Here,  $N = \text{floor}(0.3/k)$ , and the  $\eta$  axis is sampled using 500 linearly spaced grid points between 0.015 and 0.03. For the tests, the system is initialised with  $q_0$  as indicated, showing the dependence of  $\eta^*$  on the initial condition.

## 5. Conclusions

In this work, an assessment of the gauge constant in energy-shifted SAV methods for Hamiltonian system was given. The scalar cases of the simple harmonic oscillator and the Duffing oscillator



**Figure 4.** Convergence curves for the Duffing oscillator, under various choices of the nonlinear coefficient  $\gamma$  as indicated on top of each panel, and scaled shift constant  $\eta$ , where  $\eta \in V \cup \{0\}$ . The leapfrog scheme is also used. Colour scheme:  $\eta = 1.919 \cdot 10^{-2}$  (blue),  $2.168 \cdot 10^{-2}$  (red),  $2.150 \cdot 10^{-2}$  (yellow),  $1.921 \cdot 10^{-2}$  (purple), 0 (pink), leapfrog (burgundy). Grey-shaded areas represent portions of the plane dominated by round-off errors. For all panels,  $q_0 = 10$ .

were treated in detail. In both cases, the value of the shift affects the convergence rate, and numerically computed optimal values were given as a fraction of the system's energy. For the harmonic oscillator, a linear problem, such optimal value depends exclusively on the linear oscillator frequency, whilst for the Duffing oscillator, a nonlinear problem, such optimal value depends on the amplitude of the initial condition as well as the linear frequency and nonlinear parameter. When the shift is chosen in the neighbourhood of the numerically computed optimal values, local changes in the convergence rate up to twelfth order are observed, as well as a large reduction of the magnitude of the relative error compared to standard algorithms such as the leapfrog. On the other hand, when the schemes are run with no shift, the error does not decrease with decreasing time step, and hence convergence is halted. These findings suggest studying the behaviour of relative error as a function of the shift analytically, which is left as future work.

- [1] Quispel G and McLaren D 2008 *Journal of Physics A: Mathematical and Theoretical* **41** 1–7
- [2] Brugnano L, Iavernaro F and Trigiante D 2012 *Computer Physics Communications* **183** 1860–1868
- [3] Bilbao S, Ducceschi M and Zama F 2023 *Journal of Computational Physics* **472** 111697 ISSN 0021-9991
- [4] Yang X and Han D 2016 *Journal of Computational Physics* **330** 1116–1134
- [5] Yang X, Zhao J and Wang Q 2017 *Journal of Computational Physics* **333** 104–127
- [6] Shen J, Xu J and Yang J 2018 *Journal of Computational Physics* **353** 407–416
- [7] Gong Y and Zhao J 2019 *Applied Mathematics Letters* **94** 224–231
- [8] Liu Z and Li X 2022 *Numerical Algorithms* 1–22
- [9] Zhang H, Qian X and Song S 2020 *Applied Mathematics Letters* **102** 1–9
- [10] Lin L, Yang Z and Dong S 2019 *Journal of Computational Physics* **388** 1–22
- [11] Ducceschi M, Bilbao S and Webb C 2023 Real-time modal synthesis of nonlinearly interconnected networks

- Proceedings of the 26th International Conference on Digital Audio Effects (DAFx-23)* (Copenhagen, Denmark)
- [12] Bilbao S, Webb C, Wang Z and Ducceschi M 2023 Real-time gong synthesis *Proceedings of the 26th International Conference on Digital Audio Effects (DAFx-23)* (Copenhagen, Denmark)
- [13] Ducceschi M, Hamilton M and Russo R 2023 Simulation of the snare-membrane collision in modal form using the scalar auxiliary variable (sav) method *Proceedings of the 10th Convention of the European Acoustics Association (Forum Acusticum 2023)* (Turin, Italy)
- [14] Kovacic I and Brennan M J 2011 *The Duffing Equation* (John Wiley & Sons)
- [15] Ducceschi M, Bilbao S and Webb C 2022 Real-time simulation of the struck piano string with geometrically exact nonlinearity via a scalar quadratic energy method *Proceedings of the 10th European Nonlinear Dynamics Conference (ENOC2020)* (Lyon, France)