Commun Nonlinear Sci Numer Simulat 137 (2024) 108168

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns

Adaptive parameters tuning based on energy-preserving splitting integration for Hamiltonian Monte Carlo Method

Cristiano Tamborrino [a], Fasma Diele [b,*], Carmela Marangi [b], Cristina Tarantino [c]

[a] Department of Computer Science, University of Bari, Via Edoardo Orabona, 4, 70125 Bari BA, Italy
[b] Istituto per Applicazioni del Calcolo 'M.Picone', National Research Council (CNR), via Amendola 122/D, 70126 Bari, Italy
[c] Institute of Atmospheric Pollution Research (IIA), National Research Council (CNR), c/o Interateneo Physics Department, Via Amendola 173, 70126 Bari, Italy

ARTICLE INFO

Keywords:
Hamiltonian Monte Carlo
Energy-preserving splitting methods
Gaussian distributions

ABSTRACT

Splitting schemes, a class of numerical integrators for Hamiltonian problems, offer a favorable alternative to the Störmer–Verlet method in Hamiltonian Monte Carlo (HMC) methodology. However, the performance of HMC is highly sensitive to the adopted step size. In this paper, we propose a novel approach for selecting the step size h for advancing with the method defined by the free parameter b, within the family of one-parameter second order splitting procedures. Our methodology utilizes a designated function h_b of the parameter b to determine the step size, i.e. $h = h_b(b)$. By appropriately restricting the domain of h_b to a suitable interval I, the pairs (b, h_b) with $b \in I$ ensure both stability and Hamiltonian preservation when sampling from Gaussian distributions. As a result, our technique never rejects a sample within the HMC process, and this characteristic is the key factor behind its superior performance compared to similar methods recently introduced in other studies. Additionally, we assess the effectiveness of the methods defined by the pairs (b, h_b) for general not-Gaussian distribution sampling. In this case we also present a technique based on an adaptive selection of the b parameter for improving the HMC performance. The effectiveness of the proposed approach is evaluated through benchmark examples from literature and experiments involving the Log-Gaussian Cox process and Bayesian Logistic Regression.

1. Introduction

The Hamiltonian Monte Carlo (HMC) algorithm is a Markov chain Monte Carlo (MCMC) method designed to iteratively generate random samples that progressively approximate a target probability distribution. Originally known as Hybrid Monte Carlo in the late 1980s, it found its initial application in the field of Lattice Quantum Chromodynamics [1]. Several years later, Radford Neal recognized its potential in the field of applied statistics within his work on Bayesian neural networks [2]. However, it was his seminal paper [3] that ushered this approach into the mainstream of statistical computing. HMC merges the principles of molecular dynamics and computational statistics by employing a molecular dynamics-inspired approach based on Hamiltonian formalism [4,5] to guide the Monte Carlo sampling process, enhancing the efficiency and effectiveness of statistical sampling.

At each step of the Markov chain in HMC, a numerical integration of a Hamiltonian system is required. Traditionally, the Störmer–Verlet or Leapfrog algorithm, a second-order splitting method, has been widely used for this purpose (see, for example, [6]).

* Corresponding author.
E-mail addresses: cristiano.tamborrino@uniba.it (C. Tamborrino), fasma.diele@cnr.it (F. Diele), carmela.marangi@cnr.it (C. Marangi), cristina.tarantino@iia.cnr.it (C. Tarantino).

https://doi.org/10.1016/j.cnsns.2024.108168
Received 23 August 2023; Received in revised form 9 May 2024; Accepted 17 June 2024
Available online 28 June 2024

However, researchers have been actively exploring alternative algorithms that can offer improved efficiency. Several studies (such as [7–9], and references therein) have addressed the challenge of developing more effective integration schemes.

A common thread among recent research efforts proposing alternatives to the Störmer–Verlet algorithm is the analysis of their effectiveness when applied to Gaussian distributions. It is important to note, as highlighted in [10], that it is not practically meaningful to use a Markov chain algorithm to sample from a Gaussian distribution, just as it does not make sense to numerically integrate the equations of a simple harmonic oscillator. However, it is a common practice to evaluate the performance of algorithms on simple problems, such as sampling from Gaussian distributions, as they serve as benchmarks for assessing their performance in more complex situations.

In this context, we introduce an innovative approach that relies on the functional form $h_b(b) = \sqrt{\frac{4\,b^2 - 6\,b + 1}{b^2\,(2\,b-1)}}$ to assign a suitable step size $h = h_b(b)$ to each parameter b that defines the one-parameter family of second-order splitting procedures provided in [10]. Our proposed criterion, when applied to sampling from Gaussian distributions, ensures that all methods within the splitting family exhibit a zero expectation value for the energy error random variable. This approach builds upon energy-preserving splitting methods for Hamiltonian dynamics introduced in [11], which we adapt to the context of HMC. Instead of fixing the step size h and selecting the parameter b that minimizes the expectation of the energy error as generally done in literature (see e.g. [10]), we reverse and improve the procedure: first, we choose a specific method within the one-parameter family of second-order splitting procedure by fixing the parameter b in the domain of definition of the function h_b. Then, we implement the selected method using $h = h_b(b)$ as step size. This guarantees that the expected value for the energy error is nullified.

As mentioned, the initial step of our approach involves the selection of the parameter b in the definition domain of the function h_b. The adopted criterion ensures that the range of the function h_b falls within the stability interval of the one-parameter family of second-order splitting methods as established in [10] where the univariate Gaussian distribution is used as test model. This process determines the open interval $I =]\frac{3-\sqrt{5}}{4}, \frac{1}{4}]$ from which we ultimately make our selection of b.

Given the significant impact of energy distortion on calculated energy averages and the resulting high percentage of rejections in the HMC algorithm, preserving energy as much as possible is crucial, especially in high-dimensional problems [12]. Consequently, this paper aims to investigate whether the use of the couples (b, h_b), with $b \in I$, which effectively nullifies the energy error for Gaussian test distributions, can also reduce the number of rejection steps when applied within HMC for sampling from arbitrary distributions.

As an additional advancement, we introduce a novel variation of the HMC algorithm that includes an adaptive adjustment of the parameter b within the range of I. We start with an initial value of b sufficiently large and we refine it based on the performance of the HMC algorithm. During the sampling process, each time a sample is rejected (or each time we reach a given percentage of rejections) we reduce the value of b. As a result, at each step of the Markov chain, our adaptive approach selects a different splitting method for the numerical integration of the Hamiltonian system within the HMC algorithm. Each of these methods is implemented with the corresponding h_b as the step size.

This adaptive choice of b allows us to adapt the splitting method to the specific characteristics of the target distribution and optimize the sampling performance. It is worth noting that the adaptive selection of the parameter b from the one-parameter family of second-order splitting procedures has been adopted in HMC literature only recently [13]. In their paper the authors extend their technique into the domain of computational statistics, building upon concepts previously introduced within the framework of molecular dynamics [14]. The authors *suggest an algorithm that, once the system to be integrated has been specified and the user has chosen a value of the step size, it identifies the 'best' b.* In their approach, the best b corresponds to minimize a bound for the expectation of the energy error for the univariate Gaussian distribution. As mentioned earlier, we reverse and improve the procedure: once the value of $b \in I$ has been selected, the step size is uniquely determined by the function h_b. By setting the step size $h = h_b(b)$ the expectation of the energy error for Gaussian distributions, both univariate and multivariate, is nullified.

To assess the efficacy of our approach, we perform various tests, one of which involves the Log-Gaussian Cox model. This model is commonly employed as a point process to represent the presence of invasive species. It allows us to study the spatial distribution and occurrence of these species by considering the intensity of their presence as a random field. By applying our technique to this model, we can evaluate its performance and suitability for capturing the characteristics of invasive species distributions. This model plays a significant role in understanding and predicting the spread of such species, and it serves as a valuable statistical tool in related research areas [15–18].

The paper's structure is organized as follows: Section 2 provides an overview of the general framework for sampling from a target distribution using the HMC algorithm. It includes a brief discussion of theoretical and practical implementation details. Section 3 introduces the maps constructed using the splitting of the Hamiltonian vector field. The key finding, outlined in Theorem 3.2, showcases how to determine the interval I so that the couple (b, h_b), with $b \in I$ leads to a stable and energy-preserving approximation for univariate Gaussian distributions. Additionally, we extend this result to multivariate Gaussian distributions in Section 3.3 (Theorem 3.4). Drawing from the aforementioned theorem, we present a novel two-step splitting method in Section 4 for sampling within the HMC algorithm from Gaussian distributions. For sampling from generic distributions, in Section 5, we first propose a classical implementation of the proposed splitting procedure that consists in a fixed method (identified by a single chosen b) which advances with the step size $h = h_b$. Subsequently, we implement the adaptive reduction of the parameter b by employing Algorithm 5.1, which corresponds to the utilization of different splitting methods at each step of the Markov chain in the HMC algorithm. Section 6 presents numerical experiments to validate the theoretical results. Finally, in Section 7, we provide concluding remarks and discuss potential future developments of the proposed approach. As supplementary information in Appendix, we analyze the energy-preserving linear maps generated by applying a volume-preserving and momentum flip-reversible integrator within the HMC method for both univariate and multivariate Gaussian distributions.

2. The Hamiltonian Monte Carlo algorithm

The HMC algorithm, as described below, follows the steps outlined in [3]. Let us consider a dataset X, and our objective is to sample the variable $\mathbf{q} \in \mathbf{R}^d$ from a probability distribution of interest $\mathcal{P}(\mathbf{q})$ with prior density $\pi(\mathbf{q})$ and likelihood function $L(\mathbf{q}|X)$, i.e., $\mathcal{P}(\mathbf{q}) = \pi(\mathbf{q})L(\mathbf{q}|X)$. The first step is to associate a scalar function $U(\mathbf{q})$ to $\mathbf{q}$ through the canonical distribution. We define it as follows:

$$U(\mathbf{q}) = -\log\left[\mathcal{P}(\mathbf{q})\right] - \log(Z), \quad Z > 0,$$

which implies that $\mathcal{P}(\mathbf{q}) \propto \exp(-U(\mathbf{q}))$ with Z chosen as the normalizing constant needed for the probability to integrate to one. Next, we introduce auxiliary variables $\mathbf{p} \in \mathbf{R}^d$, independent of $\mathbf{q}$, and specify their distribution using another scalar function $K(\mathbf{p})$. In HMC, a common practice is to use a quadratic functional dependence on $\mathbf{p}$ given by:

$$K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T D_\beta^{-1}\mathbf{p},$$

where, without loss of generality, we assume that the components of $\mathbf{p}$ are independent. In this case, D_β is a diagonal matrix with entries β_i^2, representing the variances of the ith component p_i of $\mathbf{p}$. The canonical distribution $\mathcal{P}(\mathbf{p}) = \exp(-K(\mathbf{p}))$ results in a zero-mean multivariate Gaussian distribution. We denote the function for the joint variable of $\mathbf{q}$ and $\mathbf{p}$ as

$$H(\mathbf{q},\mathbf{p}) = K(\mathbf{p}) + U(\mathbf{q}). \tag{2.1}$$

This function defines a joint canonical distribution satisfying:

$$\mathcal{P}(\mathbf{q},\mathbf{p}) = \frac{1}{Z}\exp(-H(\mathbf{q},\mathbf{p})) = \frac{1}{Z}\exp(-U(\mathbf{q}))\exp(-K(\mathbf{p})) = \mathcal{P}(\mathbf{q})\mathcal{P}(\mathbf{p}). \tag{2.2}$$

We observe that the joint canonical distribution for $\mathbf{q}$ and $\mathbf{p}$ factorizes, indicating that the two variables are independent.
The concept behind HMC involves interpreting the variables $(\mathbf{q},\mathbf{p})$ as the position and momentum of a system of particles, each with masses equal to the variances β_i^2. These particles move in a potential $U(\mathbf{q})$ with a corresponding kinetic energy $K(\mathbf{p})$. The sum of kinetic and potential energy is expressed by the function (2.1) which represents the Hamiltonian of this system. The Hamilton's equations of motion are defined by the following equations

$$\frac{d\mathbf{q}}{dt} = \nabla_{\mathbf{p}} H(\mathbf{q},\mathbf{p}) = \nabla_{\mathbf{p}} K(\mathbf{p}), \qquad \frac{d\mathbf{p}}{dt} = -\nabla_{\mathbf{q}} H(\mathbf{q},\mathbf{p}) = -\nabla_{\mathbf{q}} U(\mathbf{q}). \tag{2.3}$$

It can be easily proved that the Hamiltonian is a first integral of (2.3) so that its invariance corresponds to the physical principle of conservation of total energy (2.1). By defining $\mathbf{y} = (q_1,\dots,q_d,p_1,\dots p_d)^T \in \mathbf{R}^{2d}$ we write (2.3) in the form

$$\frac{d\mathbf{y}}{dt} = J\,\nabla H(\mathbf{y}), \qquad J = \begin{pmatrix} 0_d & \mathcal{I}_d \\ -\mathcal{I}_d & 0_d \end{pmatrix}$$

where $\mathcal{I}_d$ is the identity matrix of dimension d. It can be demonstrated that the flow map $\Phi_t(\mathbf{y})$ is a volume preserving map, as indicated by the fact that its Jacobian, denoted as $\Phi_t'(\mathbf{y})$, has a determinant of 1 for all values of t. Moreover, Hamiltonian flows exhibit an additional significant property, often referred to as simplecticity, expressed as $(\Phi_t'(\mathbf{y}))^T J\,\Phi_t'(\mathbf{y}) = J$.

Given that (2.2) shows that the variables $\mathbf{q}$ are independent of $\mathbf{p}$, we can use Hamiltonian dynamics (2.3) to sample from the joint canonical distribution $\mathcal{P}(\mathbf{q},\mathbf{p})$ and disregard the momentum contributions. The introduction of the auxiliary variable $\mathbf{p}$ allows the Hamiltonian dynamics to operate effectively [3].

Starting with the generation of an initial position state $\mathbf{q}^{(i)} \sim \pi(\mathbf{q})$ for $i = 0,\dots,L$, each iteration of the HMC algorithm consists of two steps. In the first step, the initial momentum $\mathbf{p}^{(i)}$ is chosen by randomly drawing values from a zero-mean multivariate Gaussian distribution $\mathcal{N}(0, D_\beta)$. In the second step, starting at $t = 0$ with initial states $\mathbf{Q}(0) = \mathbf{q}^{(i)}$ and $\mathbf{P}(0) = \mathbf{p}^{(i)}$, the Hamiltonian dynamics are solved for a fixed interval $t \in (0, T^*]$, given by the equations:

$$\frac{d\mathbf{Q}}{dt} = \nabla_{\mathbf{P}} K(\mathbf{P}) = D_\beta^{-1}\mathbf{P}, \qquad \frac{d\mathbf{P}}{dt} = -\nabla_{\mathbf{Q}} U(\mathbf{Q}), \qquad t \in (0, T^*]. \tag{2.4}$$

where the Hamiltonian function is defined as:

$$H(\mathbf{Q},\mathbf{P}) = \frac{1}{2}\mathbf{P}^T D_\beta^{-1}\mathbf{P} + U(\mathbf{Q}). \tag{2.5}$$

After simulating the Hamiltonian dynamics, the state of the position at the end of the simulation, $\mathbf{Q}(T^*)$, is used as the next state of the Markov chain by setting $\mathbf{q}^{(i+1)} = \mathbf{Q}(T^*)$.

The selection of the fictitious final time T^* in the Hamiltonian dynamics of the HMC algorithm plays a crucial role and should be carefully chosen to preserve ergodicity. Ergodicity ensures that the HMC algorithm explores the entire state space of the target distribution. In an HMC iteration, the momentum variables can influence the position variables in arbitrary ways. However, if the chosen value of T^* leads to exact periodicity for some function of the state, ergodicity can fail. For instance, when $q^{(i)} \sim \mathcal{N}(0,1)$ and $p^{(i)} \sim \mathcal{N}(0,1)$, the Hamiltonian dynamics for Q and P correspond to the equations of a harmonic oscillator:

$$\frac{dQ}{dt} = P, \qquad \frac{dP}{dt} = -Q. \tag{2.6}$$

The solutions of these equations are periodic with a period of 2π. If $T^* = 2\pi$ is chosen, the trajectory will return to the same position coordinate, and the HMC algorithm will not be ergodic. To address this potential issue of non-ergodicity, one approach is to randomly choose the value of T^* and repeat this process regularly. By introducing randomness in the selection of T^*, the trajectory explores different regions of the state space, enhancing the ergodicity of the HMC algorithm.

2.1. Practical implementation of the HMC algorithm

In practical implementation of HMC algorithm, we need to numerically integrate the Hamiltonian system (2.4) using a map Ψ_h that approximates the theoretical flow. Given the initial states $\mathbf{Q}_0 = \mathbf{Q}(0)$ and $\mathbf{P}_0 = \mathbf{P}(0)$, the map iteratively updates the states as $(\mathbf{Q}_{n+1}, \mathbf{P}_{n+1}) = \Psi_h(\mathbf{Q}_n, \mathbf{P}_n)$, for $n = 0, \dots, N$, where N and the step size h satisfy $N\,h = T^*$.

To ensure the validity of the approximation, the chosen map Ψ_h must satisfy two important properties (see, e.g [19]). As the theoretical flow the numerical map should be volume-preserving, which means that, for all n the determinant of its Jacobian matrix $\Psi_h'(\mathbf{Q}_n, \mathbf{P}_n)$ is equal to 1:

$$\det(\Psi_h'(\mathbf{Q}_n, \mathbf{P}_n)) = 1.$$

Second, the map should be "momentum flip" reversible, according to the definition in [20]:

$$\Psi_h(\mathbf{Q}_n, \mathbf{P}_n) = (\mathbf{Q}_{n+1}, \mathbf{P}_{n+1}) \quad \Longleftrightarrow \quad \Psi_h(\mathbf{Q}_{n+1}, -\mathbf{P}_{n+1}) = (\mathbf{Q}_n, -\mathbf{P}_n),$$

for $n = 0, \dots, N$. This symmetry ensures that the dynamics can be reversed, allowing for correct exploration of the target probability distribution $\pi(\mathbf{q})$.

The proposed variables $\mathbf{q}^* = \mathbf{Q}(T^*)$ and $\mathbf{p}^* = \mathbf{P}(T^*)$ at the end of the simulation are used, and their acceptance is determined using an update rule similar to the Metropolis acceptance criterion. Specifically, this proposed state is accepted as the next state of the Markov chain with probability

$$\alpha = \min\left(1, \exp\left(H(\mathbf{q}^{(i)}, \mathbf{p}^{(i)}) - H(\mathbf{q}^*, \mathbf{p}^*)\right)\right) > u, \quad u \sim \mathcal{U}(0,1)$$

and $\mathbf{q}^{(i+1)} = \mathbf{q}^*$. If the proposed state is not accepted, the next state is the same as the current state $\mathbf{q}^{(i+1)} = \mathbf{q}^{(i)}$. Combining these steps, sampling random momentum, followed by Hamiltonian dynamics and Metropolis acceptance criterion, defines the HMC Algorithm 2.1 for drawing L samples from a target distribution. This process ensures that the proposed states are accepted or rejected based on their relative probabilities, allowing for exploration of the target distribution.

Algorithm 2.1 HMC algorithm

Draw $\mathbf{q}^{(1)} \sim \pi(\mathbf{q})$, $\mathbf{q}^{(1)} \in \mathbb{R}^d$, $L \geq 1$, set $i = 0$
while $i < L$ **do**
  i=i+1
  Draw $\mathbf{p}^{(i)} \sim \mathcal{N}(0, D_\beta)$
  Set $(\mathbf{Q}_0, \mathbf{P}_0) = (\mathbf{q}^{(i)}, \mathbf{p}^{(i)})$,
  Randomly choose $T^* > 0$
  Set $h > 0$ and $N = \lfloor \frac{T^*}{h} \rfloor$
  Evaluate $(\mathbf{Q}_{n+1}\,\mathbf{P}_{n+1}) = \Psi_h(\mathbf{Q}_n, \mathbf{P}_n)$, for $n = 0, \dots N-1$
  Set $(\mathbf{q}^*, \mathbf{p}^*) = (\mathbf{Q}_N, \mathbf{P}_N)$
  Calculate $\alpha = \min\left(1, \exp\left(H(\mathbf{q}^{(i)}, \mathbf{p}^{(i)}) - H(\mathbf{q}^*, \mathbf{p}^*)\right)\right)$
  Draw $u \sim \mathcal{U}(0,1)$
  Update: if $\alpha > u$ then $\mathbf{q}^{(i+1)} = \mathbf{q}^*$; otherwise $\mathbf{q}^{(i+1)} = \mathbf{q}^{(i)}$
end while
return Markov chain $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(L)}$

Indeed, it has been shown in [10] that for a momentum-flip reversible volume-preserving transformation, the phase space can be divided into two regions of equal volume. One region corresponds to negative energy errors, where the energy at the end of the simulation $H(\mathbf{Q}_N, \mathbf{P}_N)$ is less than the energy at the beginning $H(\mathbf{Q}_0, \mathbf{P}_0)$, and the other region corresponds to flipping the momentum with positive energy errors, where the energy increases. If the map used for approximation is energy-preserving, i.e., $H(\mathbf{Q}_N, \mathbf{P}_N) = H(\mathbf{Q}_0, \mathbf{P}_0)$, then all proposals will be accepted because there are no positive energy errors. However, if the map is not energy-preserving, there will be proposals with positive energy errors, leading to potential rejections. In such cases, the acceptance of proposals depends on the Metropolis acceptance criterion, comparing the probabilities of the joint distributions at the end of the simulation and the initial state. This highlights the importance of selecting a map that is as energy-preserving as possible to minimize the number of rejections and improve the efficiency of the HMC algorithm.

3. Splitting methods

The Störmer–Verlet method is indeed a commonly used map within the HMC algorithm. It belongs to the class of symmetric splitting methods, which have been extensively studied in the context of Hamiltonian dynamics.

Symmetric splitting methods, including Störmer–Verlet, are based on the idea of splitting the Hamiltonian flow into two or more subflows and constructing a map as a composition of these subflows. Each subflow is a Hamiltonian flow, which means it is volume-preserving and reversible. The composition of volume-preserving maps preserves volume, ensuring that the resulting map is also volume-preserving. Furthermore, since the semiflows are reversible and the composition is symmetric, the overall splitting map is reversible as well. A detailed proof of this property can be found in the Ref. [10].

While the Störmer–Verlet method is widely used, there have been efforts in the literature to develop more accurate maps within the same class of symmetric splitting methods. The focus of these developments is on improving the performance of the map within the HMC algorithm rather than on the accuracy of approximating the dynamical flow. Several references, such as [9,12], explore the design of optimal symmetric splitting methods with improved accuracy and efficiency properties. It is worth noting that the accuracy of the map within the HMC algorithm refers to its ability to generate proposals with minimal energy errors, leading to a higher acceptance rate and improved sampling efficiency.

3.1. Störmer–Verlet method

The Störmer–Verlet method is based on the splitting of the flow in two semiflows and is built as a symmetric composition of semiflows. Setting $\mathbf{Y} = [\mathbf{Q}, \mathbf{P}]^T \in \mathbb{R}^{2d}$, it can be useful to denote the Hamiltonian dynamics (2.4) in vector form as

$$\frac{d\mathbf{Y}}{dt} = f(\mathbf{Y}) := \left[D_\beta^{-1}\mathbf{P}, -\nabla_{\mathbf{Q}} U(\mathbf{Q})\right]^T.$$

Let $\varphi_t^{[\mathbf{P}]}$ and $\varphi_t^{[\mathbf{Q}]}$ represent the exact flows associated to the dynamics $\frac{d\mathbf{Y}}{dt} = f^{[\mathbf{P}]}(\mathbf{Y})$ and $\frac{d\mathbf{Y}}{dt} = f^{[\mathbf{Q}]}(\mathbf{Y})$, where $f = f^{[\mathbf{P}]} + f^{[\mathbf{Q}]}$ and

$$f^{[\mathbf{P}]}(\mathbf{Y}) := [D_\beta^{-1}\mathbf{P}, \mathbf{0}_d]^T, \qquad f^{[\mathbf{Q}]}(\mathbf{Y}) := [\mathbf{0}_d, -\nabla_{\mathbf{Q}} U(\mathbf{Q})]^T.$$

The map $\mathbf{Y}_{n+1} = \Psi_h^{(SV)}(\mathbf{Y}_n)$, with

$$\Psi_h^{(SV)} := \varphi_{h/2}^{[\mathbf{Q}]} \circ \varphi_h^{[\mathbf{P}]} \circ \varphi_{h/2}^{[\mathbf{Q}]}, \tag{3.1}$$

defines the (*velocity*) Störmer–Verlet method.[1]

When a Störmer–Verlet step is applied to the linear test problem described by a harmonic oscillator, as defined in the following equations:

$$\frac{dQ}{dt} = \frac{P}{\beta^2}, \qquad \frac{dP}{dt} = -\frac{Q}{\alpha^2}, \tag{3.2}$$

it leads to a linear transformation represented in matrix form as $\mathbf{Y}_{n+1} = \mathcal{M}_2^{(h,\sigma)}\mathbf{Y}_n$ where

$$\mathcal{M}_2^{(h,\sigma)} = \begin{bmatrix} \mathrm{p}_h & \mathrm{e}_h + \sigma^{-1}\,\mathrm{q}_h \\ \mathrm{e}_h - \sigma\,\mathrm{q}_h & \mathrm{p}_h \end{bmatrix}, \tag{3.3}$$

and

$$\mathrm{p}_h = 1 - \frac{h_\sigma^2}{2}, \quad \mathrm{e}_h = \frac{\sigma\, {h_\sigma}^3}{4(\sigma^2+1)}, \quad \mathrm{q}_h = h_\sigma - \sigma\,\mathrm{e}_h$$

with $\sigma = \frac{\beta}{\alpha}$ and $h_\sigma = \frac{h}{\alpha\beta}$. For $h < 2\alpha\beta$ it results $h_\sigma \le 2$, and then the trajectories are stable as it results $\mathrm{p}_h^2 - 1 < 0$ (see Appendix).

Since $\mathrm{e}_h \neq 0$, the Störmer–Verlet integrator cannot preserve the energy when applied to the linear test model (3.2). The expectation of the random variable $\Delta_2^{(N)} := H(\mathbf{Y}_N) - H(\mathbf{Y}_0)$ is given by

$$\mathbb{E}(\Delta_2^{(N)}) = \frac{N}{2}\left(\sigma + \frac{1}{\sigma}\right)^2 \left(\frac{\sigma\, {h_\sigma}^3}{4(\sigma^2+1)}\right)^2 = \frac{N}{32} h_\sigma^6 = T^* \left(\frac{h_\sigma}{2}\right)^5.$$

(see Theorem A.2 in Appendix).

3.2. One-parameter family of second order splitting methods

The literature provides a range of alternatives to the Störmer–Verlet method to enhance the performance of the numerical integrator used in the HMC algorithm. A strategy involves adjusting specific free parameters within a particular class of methods. In this paper, we conduct this exploration within the context of a one-parameter family of second-order splitting schemes featured by the desired properties of volume-preservation and reversibility [10]. The family is defined by the composition equation:

$$\Psi_h^{(b)} := \varphi_{b\,h}^{[\mathbf{Q}]} \circ \varphi_{h/2}^{[\mathbf{P}]} \circ \varphi_{(1-2\,b)h}^{[\mathbf{Q}]} \circ \varphi_{h/2}^{[\mathbf{P}]} \circ \varphi_{b\,h}^{[\mathbf{Q}]}, \qquad b \in \mathbb{R}, \quad b \neq 0, 1/2. \tag{3.4}$$

When $b = 0, 1/2$, the method reduces to the classical velocity and position Störmer–Verlet integrators.

Common strategies for selecting the parameter b in (3.4) are directed towards improving the numerical performance of the chosen method when applied to linear test models (3.2). One approach focuses on maximizing the stability interval length while simultaneously minimizing specific error constants, as proposed in [7]. Another strategy centers on enhancing energy conservation properties, thus reducing the expected energy error, as investigated in [10]. With the more ambitious objective of completely

[1] We mention that the *position* Störmer–Verlet method starts the integration by solving the semiflow $f^{[P]}$ so that $\Psi_h^{(SV)} := \varphi_{h/2}^{[P]} \circ \varphi_h^{[Q]} \circ \varphi_{h/2}^{[P]}$.

nullifying the expected energy error for the linear test model (3.2), we turn to the energy-preserving integrators introduced in [11] for the numerical solution of Hamiltonian systems. Our goal is to evaluate their effectiveness within the statistical computing framework of the HMC algorithm.

The class of second-order splitting methods can be expressed as $\mathbf{Y}_{n+1} = \Psi_h^{(b)}(\mathbf{Y}_n)$, where $\Psi_h^{(b)}$ is given by Eq. (3.4). When applied to the model test system (3.2), these methods can be written as a linear map in matrix form: $\mathbf{Y}_{n+1} = \mathcal{M}_2^{(h,\sigma)}\mathbf{Y}_n$, where $\mathcal{M}_2^{(h,\sigma)}$ is given by Eq. (3.3). The coefficients p_h, q_h are defined as follows:

$$\mathrm{p}_h = 1 - \frac{h_\sigma^2}{2} + \frac{h_\sigma^4}{4}\, b\,(1-2\,b),$$
$$\mathrm{q}_h = \frac{b^2(1-2\,b)}{4\,(\sigma^2+1)}\, {h_\sigma}^5 + \frac{4\,b^2 + 2\,b\,\sigma^2 - 4\,b - \sigma^2}{4(\sigma^2+1)}\, {h_\sigma}^3 + h_\sigma.$$

where $\sigma = \dfrac{\beta}{\alpha}$, $h_\sigma = \dfrac{h}{\alpha\,\beta}$. The expression for e_h is given by

$$\mathrm{e}_h = \mathrm{e}_h(b) = \frac{h_\sigma^3\sigma}{4(\sigma^2+1)}\left(2\,b^3\,h_\sigma^2 - b^2\,h_\sigma^2 - 4\,b^2 + 6\,b - 1\right). \tag{3.5}$$

The stability interval can be determined from the result in [10], i.e.

$$0 < h_\sigma = \frac{h}{\alpha\,\beta} < \min\left\{\sqrt{\frac{2}{b}}, \sqrt{\frac{2}{1/2 - b}}\right\}, \qquad 0 < b < \frac{1}{2}. \tag{3.6}$$

The expectation of the random variable $\Delta_2^{(N)}$

$$\mathbb{E}(\Delta_2^{(N)}) = T^*\left(\frac{h_\sigma}{2}\right)^5 \left(2\,b^3\,{h_\sigma}^2 - b^2\,{h_\sigma}^2 - 4\,b^2 + 6\,b - 1\right)^2$$

(see Theorem A.2 in Appendix) which can be nullified exploiting the following result which generalizes Theorem 1 given in [11].

Theorem 3.1. *For all b, $h > 0$ define*

$$R(b,h) := 2\left(\frac{h}{\alpha\,\beta}\right)^2 b^3 - \left(4 + \left(\frac{h}{\alpha\,\beta}\right)^2\right) b^2 + 6\,b - 1. \tag{3.7}$$

Fix $h > 0$ and consider b_h a real root of the third degree polynomial (3.7) in the variable b; then the scheme $\mathbf{Y}_{n+1} = \Psi_h^{(b_h)}(\mathbf{Y}_n)$, with $\Psi_h^{(b_h)}$ given in (3.4) is energy-preserving for the test model (3.2).

Proof. Write $\mathrm{e}_h(b)$ in (3.5) as $\mathrm{e}_h(b) = \dfrac{\sigma}{4(\sigma^2+1)}\left(\dfrac{h}{\alpha\,\beta}\right)^3 R(b,h)$; then, from $R(b_h,h) = 0$ it follows $\mathrm{e}_h(b_h) = 0$. From Theorem A.1, the result follows. □

In the HMC framework it can be more useful to adopt a different perspective:

Theorem 3.2. *Consider the open interval $I = \left]\frac{3-\sqrt{5}}{4}, \frac{1}{4}\right] \subset \mathbb{R}$ and the function $h_b : I \to \mathbb{R}_+$ defined as :*

$$b \in I \longrightarrow h_b = \sqrt{\frac{4\,b^2 - 6\,b + 1}{b^2\,(2\,b - 1)}} \in \mathbb{R}_+. \tag{3.8}$$

Fix $b \in I$ and consider the symplectic, reversible scheme in (3.4) given by $\mathbf{Y}_{n+1} = \Psi_h^{(b)}(\mathbf{Y}_n)$ which advances with step size $h := \alpha\,\beta\,h_b$; then it provides a stable, energy-preserving approximation of the test model (3.2).

Proof. Fix $b \in I$. When the scheme $\mathbf{Y}_{n+1} = \Psi_h^{(b)}(\mathbf{Y}_n)$ is applied to the linear test model (3.2) with step size $h := \alpha\,\beta\,h_b$, then $h_\sigma = \dfrac{\alpha\,\beta\,h_b}{\alpha\,\beta} = h_b$ and

$$\begin{aligned}\mathrm{e}_h(b) &= \frac{\sigma h_b^3}{4(\sigma^2+1)}\,(2\,h_b^2\,b^3 - (4 + h_b^2)\,b^2 + 6b - 1)\\ &= \frac{\sigma h_b^3}{4(\sigma^2+1)}\left(b^2\,(2\,b-1)h_b^2 - (4b^2 - 6b + 1)\right) = 0.\end{aligned}$$

From Theorem A.1, the conservation of energy follows. Moreover, since $\frac{3-\sqrt{5}}{4} < b \le \frac{1}{4}$, it results

$$0 < h_b \le \sqrt{\frac{4}{1-2b}} = \min_{b\in I}\left\{\sqrt{\frac{2}{b}}, \sqrt{\frac{2}{1/2 - b}}\right\}$$

so that $h_\sigma = h_b$ satisfies the stability condition (3.6). □

An important consequence which will be useful to extend the described result to the multivariate case, is the following.

Theorem 3.3. *With the notations used above, fix $b \in I$ and consider*

$$\tilde{\Psi}_h^{(b)} := \tilde{\varphi}_{bh}^{[Q]} \circ \tilde{\varphi}_{h/2}^{[P]} \circ \tilde{\varphi}_{(1-2\,b)h}^{[Q]} \circ \tilde{\varphi}_{h/2}^{[P]} \circ \tilde{\varphi}_{bh}^{[Q]} \tag{3.9}$$

where $\tilde{\varphi}_t^{[P]}$ and $\tilde{\varphi}_t^{[Q]}$ represent the exact flows of the dynamics $\frac{d\mathbf{Y}}{dt} = [\sigma^{-1} P, \mathbf{0}]^T$ and $\frac{d\mathbf{Y}}{dt} = [\mathbf{0}, -\sigma\, Q]^T$, respectively. Then, the symplectic and reversible scheme $\mathbf{Y}_{n+1} = \tilde{\Psi}_{h_b}^{(b)}(\mathbf{Y}_n)$, with h_b defined in (3.8), provides a stable, energy-preserving approximation of the test model (3.2).

Proof. It is enough to observe that the scheme $\mathbf{Y}_{n+1} = \tilde{\Psi}_{h_b}^{(b)}(\mathbf{Y}_n)$ is equivalent to the scheme (3.4) given by $\mathbf{Y}_{n+1} = \Psi_h^{(b)}(\mathbf{Y}_n)$ with $h = \alpha\,\beta\,h_b$. □

3.3. Generalization to multivariate Gaussian distributions

Theorem 3.3 presents the construction for a symplectic, reversible, and energy-preserving scheme for the jth oscillator

$$\frac{dQ_j}{dt} = \frac{P_j}{\beta_j^2}, \qquad \frac{dP_j}{dt} = -\frac{Q_j}{\alpha_j^2}, \qquad \text{for} \quad j = 1, \dots d, \tag{3.10}$$

where $j = 1, \dots, d$. Indeed, as before, fix $b \in I$ and consider

$$\tilde{\Psi}_h^{(b)} := \tilde{\varphi}_{bh}^{[Q_j]} \circ \tilde{\varphi}_{h/2}^{[P_j]} \circ \tilde{\varphi}_{(1-2\,b)h}^{[Q_j]} \circ \tilde{\varphi}_{h/2}^{[P_j]} \circ \tilde{\varphi}_{bh}^{[Q_j]}. \tag{3.11}$$

where the flows $\tilde{\varphi}_t^{[P_j]}$ and $\tilde{\varphi}_t^{[Q_j]}$ correspond to the exact flows of the following differential equations:

$$\frac{d\mathbf{Y}^{(j)}}{dt} = [\sigma_j^{-1}\, P_j,\, 0]^T \quad \text{and} \quad \frac{d\mathbf{Y}^{(j)}}{dt} = [0,\, -\sigma_j\, Q_j]^T,$$

where $\mathbf{Y}^{(j)} := [Q_j,\, P_j]$ and $\sigma_j = \dfrac{\beta_j}{\alpha_j}$ for $j = 1, \dots, d$. Then, the symplectic, reversible scheme $\mathbf{Y}_{n+1}^{(j)} = \tilde{\Psi}_{h_b}^{(b)}(\mathbf{Y}_n^{(j)})$ where $\tilde{\Psi}_h^{(b)}$ is defined in (3.11) is a stable integrator for the jth oscillator (3.10) which preserves the jth Hamiltonian $H_j(Q_j, P_j) = \dfrac{1}{2}\left(\dfrac{Q_j^2}{\alpha_j^2} + \dfrac{P_j^2}{\beta_j^2}\right) = \dfrac{1}{2\,\alpha_j\,\beta_j}\left(\sigma_j\, Q_j^2 + \dfrac{P_j^2}{\sigma_j}\right)$.

Now we have all the instruments to provide symplectic, reversible, energy-preserving schemes for the d-dimensional test model

$$\frac{d\mathbf{Q}}{dt} = D_\beta^{-1}\mathbf{P}, \qquad \frac{d\mathbf{P}}{dt} = -D_\alpha^{-1}\mathbf{Q}. \tag{3.12}$$

Set D_α and D_β as $d \times d$ diagonal matrices with entries α_j^2 and β_j^2, respectively, for $j = 1, \dots, d$ and define $\Sigma := D_\beta^{1/2}\, D_\alpha^{-1/2}$, we can give the following result

Theorem 3.4. *Fix b in the interval $I = \left]\frac{3-\sqrt{5}}{4}, \frac{1}{4}\right[\subset \mathbb{R}$ and, with the notations used above, consider*

$$\tilde{\Psi}_h^{(b)} := \tilde{\varphi}_{b\,h}^{[\mathbf{Q}]} \circ \tilde{\varphi}_{h/2}^{[\mathbf{P}]} \circ \tilde{\varphi}_{(1-2\,b)h}^{[\mathbf{Q}]} \circ \tilde{\varphi}_{h/2}^{[\mathbf{P}]} \circ \tilde{\varphi}_{b\,h}^{[\mathbf{Q}]}, \tag{3.13}$$

where $\tilde{\varphi}_t^{[\mathbf{P}]}$ and $\tilde{\varphi}_t^{[\mathbf{Q}]}$ represent the exact flows of

$$\frac{d\mathbf{Y}}{dt} = [\Sigma^{-1}\,\mathbf{P},\, \mathbf{0}_d]^T, \qquad \frac{d\mathbf{Y}}{dt} = [\mathbf{0}_d,\, -\Sigma\,\mathbf{Q}]^T.$$

Then, the symplectic and reversible method $\mathbf{Y}_{n+1} = \tilde{\Psi}_{h_b}^{(b)}\,\mathbf{Y}_n$ with h_b defined in (3.8) provides a stable approximation for the system (3.12) which preserves the Hamiltonian

$$H(\mathbf{Y}) = \frac{1}{2}\,\mathbf{Y}^T\, \mathcal{D}_{2d}^{-1}\,\mathbf{Y}, \quad \mathcal{D}_{2d} := \begin{bmatrix} D_\alpha & \mathbf{0}_d \\ \mathbf{0}_d & D_\beta \end{bmatrix}. \tag{3.14}$$

Proof. The method $\mathbf{Y}_{n+1} = \tilde{\Psi}_{h_b}^{(b)}\,\mathbf{Y}_n$ when applied to approximate system (3.12) can be expressed as $\mathbf{Y}_{n+1} = \mathcal{M}_{2d}^{(h_b,\Sigma)}\,\mathbf{Y}_n$ where

$$\mathcal{M}_{2d}^{(h_b,\Sigma)} = \begin{bmatrix} \mathrm{P}_{h_b} & \mathrm{E}_{h_b} + \Sigma^{-1}\,\mathrm{Q}_{h_b} \\ \mathrm{E}_{h_b} - \Sigma\,\mathrm{Q}_{h_b} & \mathrm{P}_{h_b} \end{bmatrix}, \tag{3.15}$$

where the matrix E_{h_b} has, by construction, extra-diagonal entries $\mathrm{E}_{h_b}(i,j) = 0$ and diagonal ones

$$\mathrm{E}_{h_b}(j,j) = \frac{h_b^3 \sigma_j}{4(\sigma_j^2+1)}\left(2\,b^3\,{h_b}^2 - b^2\,{h_b}^2 - 4\,b^2 + 6\,b - 1\right) = 0,$$

for the choice of h_b. Consequently $\mathrm{E}_{h_b} = \mathbf{0}_{d\times d}$. The error in energy at each step is given by

$$\Delta_{2d}^{(n,h_b)} = H(\mathbf{Y}_{n+1}) - H(\mathbf{Y}_n) = \frac{1}{2}\,\mathbf{Y}_n^T\,\mathcal{E}_{2d}^{(h_b)}\,\mathbf{Y}_n \tag{3.16}$$

where $\mathcal{E}_{2d}^{(h_b)} := \mathcal{K}_{2d}^{(h_b)^T}\,\mathcal{K}_{2d}^{(h_b)} - \mathcal{D}_{2d}^{-1}$ and $\mathcal{K}_{2d}^{(h_b)} = \mathcal{D}_{2d}^{-1/2}\,\mathcal{M}_{2d}^{(h_b,\Sigma)}$. As $\mathrm{E}_{h_b} = \mathbf{0}_{d\times d}$ then the matrix $\mathcal{E}_{2d}^{(h_b)}$ results to have the following expression

$$\mathcal{E}_{2d}^{(h_b)} = \begin{pmatrix} \mathrm{P}_{h_b} D_\alpha \mathrm{P}_{h_b} + \mathrm{Q}_h D_\alpha \mathrm{Q}_{h_b} - D_\alpha & \mathbf{0}_d \\ \mathbf{0}_d & \mathrm{P}_{h_b} D_\beta \mathrm{P}_{h_b} + \mathrm{Q}_h D_\beta \mathrm{Q}_{h_b} - D_\beta \end{pmatrix}.$$

with P_{h_b}, Q_{h_b}, D_α and D_β diagonal matrices which satisfy the relation $\mathrm{P}_{h_b}^2 + \mathrm{Q}_{h_b}^2 = \mathrm{I}_d$, then $\mathcal{E}_{2d}^{(h_b)} = \mathbf{0}_{2d\times 2d}$. □

As a corollary, we provide the following result

Theorem 3.5. *Fix b in the interval $I = \left]\frac{3-\sqrt{5}}{4}, \frac{1}{4}\right[\subset \mathbb{R}$ and consider the method $\mathbf{Y}_{n+1} = \Psi_{h_b}^{(b)}(\mathbf{Y}_n)$, where $\Psi_h^{(b)}$ is defined in (3.4) and the step size h_b is given in (3.8). It provides a symplectic, reversible, stable approximation for the system*

$$\frac{d\mathbf{Q}}{dt} = \mathbf{P}, \qquad \frac{d\mathbf{P}}{dt} = -\mathbf{Q},$$

which preserves the Hamiltonian $H(\mathbf{Q},\mathbf{P}) = \frac{1}{2}\,\mathbf{P}^T\,\mathbf{P} + \frac{1}{2}\,\mathbf{Q}^T\,\mathbf{Q}$.

Proof. It is enough to notice that $D_\alpha \equiv D_\beta = \mathcal{I}_d$ and $\Sigma = \mathcal{I}_d$; then the method $\mathbf{Y}_{n+1} = \Psi_{h_b}^{(b)}(\mathbf{Y}_n)$, where $\Psi_h^{(b)}$ is defined in (3.4) reduces to $\mathbf{Y}_{n+1} = \tilde{\Psi}_{h_b}^{(b)}(\mathbf{Y}_n)$, where $\tilde{\Psi}_h^{(b)}$ is defined in (3.13). □

4. Novel splitting two-step method for sampling from Gaussian distributions

Theorem 3.4 constitutes a powerful tool for proposing an effective alternative to the Störmer–Verlet integrator to sample, with HMC algorithm, the variable $\mathbf{q} \in \mathbf{R}^d$ from a Gaussian distribution $\mathcal{N}(0, D_\alpha)$ where variance α_i^2 of the ith component q_i of $\mathbf{q}$ are the entries of the diagonal matrix D_α. The auxiliary variables $\mathbf{p} \in \mathbf{R}^d$ are taken randomly by drawing from a Gaussian distribution $\mathcal{N}(0, D_\beta)$ with variances β_i^2. To approximate the Hamiltonian dynamics, given by Eqs. (3.12), we propose to use the splitting method $\mathbf{Y}_{n+1} = \tilde{\Psi}_h^{(b)}(\mathbf{Y}_n)$, with $\tilde{\Psi}_h^{(b)}$ in (3.13) with b in the interval $I = \left]\frac{3-\sqrt{5}}{4}, \frac{1}{4}\right[\subset \mathbb{R}$ and the step size $h = h_b$ given in (3.8). The novel method is denoted as nSP2S (new splitting two-step method for Gaussian distribution). Within HMC, the proposed procedure can be implemented without performing the Metropolis test because the Hamiltonian is kept constant. This is an immediate consequence of Theorem 3.4.

5. Novel splitting two-step method for sampling from generic distributions

As mentioned in the Introduction, recent research efforts seeking alternatives to the Störmer–Verlet algorithm within HMC to sample from generic distributions have consistently focused on evaluating their performance when applied to Gaussian distributions taken as test models. In line with the same efforts, in this section, we introduce a criterion for selecting the step size h in HMC processes when using the splitting method $\mathbf{Y}_{n+1} = \Psi_h^{(b)}(\mathbf{Y}_n)$, with $\Psi_h^{(b)}$ in (3.4), to sample from generic distributions. We propose to fix b in the interval $I = \left]\frac{3-\sqrt{5}}{4}, \frac{1}{4}\right[\subset \mathbb{R}$ and to adopt the step size $h = h_b$ given in (3.8). With abuse of notation, the novel method is denoted as nSP2S (new splitting two-step method for generic distribution).

The rationale of our proposal lies in the fact that, in contrast to the Störmer–Verlet method, the one-parameter splitting (3.4), with $b \in I$ and the appropriately chosen step size $h = h_b$, provides a method which shows an optimal performance when applied to Gaussian distributions with unitary variances and zero correlations because it assures the acceptance of all proposals. This is an immediate consequence of Theorem 3.5.

5.1. Selection of the b parameter

Each value of the b parameter in the open interval $I = \left]\frac{3-\sqrt{5}}{4}, \frac{1}{4}\right[$ detects a stable method in the class of the splitting methods (3.4). Hence, we may wonder about what is the 'best' choice and, consequently, the 'best method' to adopt. Classical criteria might be followed:

1. choose $b = 0.25$ to enlarge h_b as much as possible. Consequently, $h_b \approx 2.828$. This is a very large step which can be used, in practice, only for Gaussian distributions and for very low-dimensional non-stiff problems;
2. enlarge h_b as much as possible increasing b but taking into account that stability decreases when we approach the roots of $\mathrm{p}_{h_b}^2 - 1 = 0$. The best choice corresponds to b such that $\mathrm{p}_{h_b}^2 = 0$. We find $b = b_{stab} \approx 0.2008$ and $h_{b_{stab}} \approx 1.3432$;

3. choose b to minimize the leading error term $k_{3,1}^2 + k_{3,2}^2$ with $k_{3,1} = \frac{12\,b^2-12\,b+2}{24}$ and $k_{3,2} = \frac{-6\,b+1}{24}$. In this case, the optimal choice corresponds to $b = b_{ML} \approx 0.1932$ (see [21]) and the resulting step is $h_{b_{ML}} \approx 0.6573$;

For fair numerical comparisons with different procedures, it is essential to maintain a fixed step size among integrators. Therefore, another possible choice of b corresponds to values of h_b equal to those employed as step sizes (h) in the benchmark tests. This is achieved by setting $b = h_b^{-1}(h)$, where the function $h_b^{-1} :]0,\, h_{b_{\max}}] \longrightarrow I$ represents the inverse function of h_b in Eq. (3.8) and it is defined as the smallest real root of the third order polynomial $2\,h^2\,b^3 - (4+h^2)\,b^2 + 6b - 1$.

5.2. Adaptive selection of the b parameter

Here, we are also proposing a promising strategy for an adaptive choice of the method. Starting from b_{max} set at one of the classical choices described above, we decrease this value (according to a fixed percentage) each time a sample is rejected. The HMC algorithm with a 99.7% reduction is described in Algorithm 5.1. Another variant involves reducing the parameter b each time a specified target rejection rate is exceeded. In our simulations, we will also test the performance of the proposed integrator built on the described adaptive procedures.

Algorithm 5.1 Novel Adaptive HMC algorithm

Draw $\mathbf{q}^{(1)} \sim \pi(\mathbf{q})$, $\mathbf{q}^{(1)} \in \mathbb{R}^d$, set $L \geq 1$, $T_{max} > 0$
set $red = 0.997$, $b_{min} = \frac{3-\sqrt{5}}{4}$, choose $b_{max} \in \left]\frac{3-\sqrt{5}}{4},\, \frac{1}{4}\right]$,
set $factor = b_{max} - b_{min}$, set $b = b_{max}$, $h = \sqrt{\frac{4\,b^2-6\,b+1}{b^2\,(2\,b-1)}}$
set $i = 0$
while $i < L$ **do**
  i=i+1
  Draw $\mathbf{p}^{(i)} \sim \mathcal{N}(0, D_\beta)$
  Set $(\mathbf{Q}_0,\, \mathbf{P}_0) = (\mathbf{q}^{(i)},\, \mathbf{p}^{(i)})$,
  Draw $u \sim \mathcal{U}(0,1)$, set $T^* = h + (T_{max} - h)\cdot u$, $N = \lfloor \frac{T^*}{h} \rfloor$
  Evaluate $(\mathbf{Q}_{n+1}\,\mathbf{P}_{n+1}) = \Psi_h^{(b)}(\mathbf{Q}_n,\, \mathbf{P}_n)$, for $n = 0, \dots N-1$
  Set $(\mathbf{q}^*,\, \mathbf{p}^*) = \left(\mathbf{Q}_N,\, \mathbf{P}_N\right)$
  Calculate $\alpha = \min\left(1, \exp\left(H(\mathbf{q}^{(i)}, \mathbf{p}^{(i)}) - H(\mathbf{q}^*, \mathbf{p}^*)\right)\right)$
  Update: if $\alpha > u$ then $\mathbf{q}^{(i+1)} = \mathbf{q}^*$;
  otherwise $\mathbf{q}^{(i+1)} = \mathbf{q}^{(i)}$,
  set $factor = red \cdot factor$, $b = b_{min} + factor$, $h = \sqrt{\frac{4\,b^2-6\,b+1}{b^2\,(2\,b-1)}}$;
end while
return Markov chain $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(L)}$

6. Numerical examples

6.1. Bivariate Gaussian distributions

As an initial example, we consider a simple two-dimensional test from [3] to demonstrate the energy-preserving property of the proposed splitting technique. The objective is to sample two position variables $\mathbf{X} = [X_1, X_2]^T$ from a bivariate distribution. The position variables have zero means, standard deviations of 1, and a correlation of 0.95. Additionally, we introduce two momentum variables $\mathbf{P} = [P_1, P_2]$, which follow a Gaussian distribution with zero means, standard deviations of 1, and zero correlation. We then define the Hamiltonian as

$$U(\mathbf{X}) + K(\mathbf{P}) = \frac{1}{2}\,\mathbf{X}^T\,S_{95}^{-1}\,\mathbf{X} + \frac{1}{2}\,\mathbf{P}^T\,\mathbf{P}, \qquad S_{95} = \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}.$$

It is important to highlight that the use of our novel integrator may necessitate suitable preconditioning of the dynamics, similar to other methods proposed in the literature (see e.g. [22,23]). Indeed, in order to describe the above problem with notations suitable for the application of the proposed procedure, we diagonalize the symmetric matrix $S_{95} = V^T\,D_\alpha\,V$ with V unitary matrix of eigenvectors. In doing so, the Hamiltonian can be written as

$$U(\mathbf{Q}) + K(\mathbf{P}) = \frac{1}{2}\,\mathbf{Q}^T\,D_\alpha^{-1}\,\mathbf{Q} + \frac{1}{2}\,\mathbf{P}^T\,\mathbf{P}, \qquad D_\alpha = \begin{pmatrix} 0.05 & 0 \\ 0 & 1.95 \end{pmatrix}$$

with $\mathbf{Q} = V\,\mathbf{X}$.

To illustrate the basic functionality of the nSP2S method $\mathbf{Y}_{n+1} = \tilde{\Psi}_{h_b}^{(b)}\,\mathbf{Y}_n$ with $\tilde{\Psi}_h^{(b)}$ defined in (3.13) applied with $\Sigma = D_\alpha^{-1/2}$ we compare it with the Störmer–Verlet method $\mathbf{Y}_{n+1} = \Psi_h^{(SV)}\mathbf{Y}_n$ with $\Psi_h^{(SV)}$ in (3.1) (SV-method), and with the two stages and three

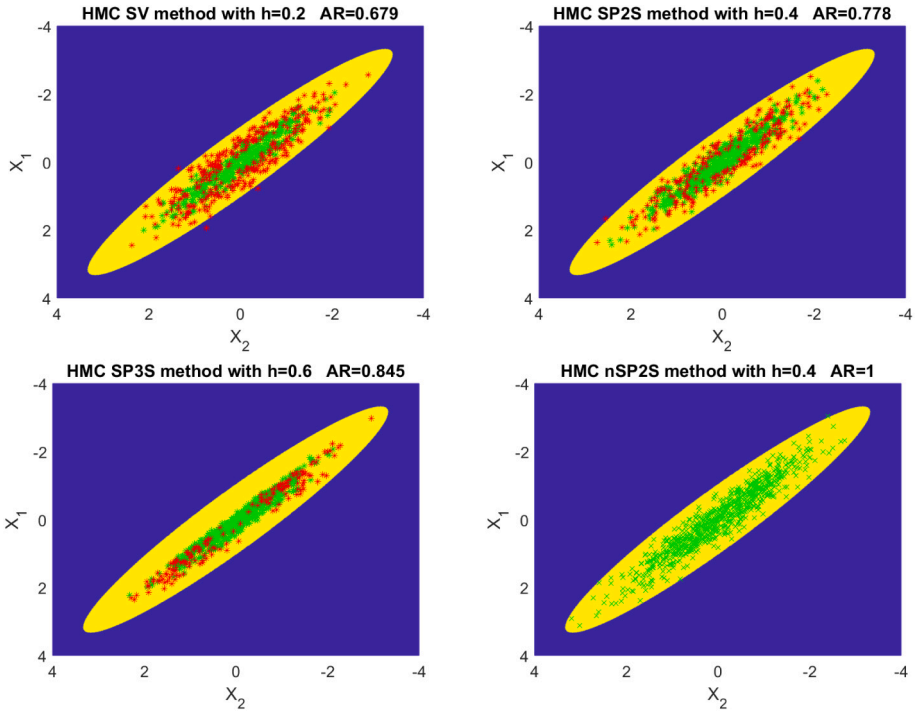

Fig. 1. Accepted (green) and rejected (red) samples of HMC method from the bivariate distribution with covariance matrix S_{95} given in Section 6.1 with different integrators. Target probability density evaluated with built-in Matlab function mvnpdf. Strictly positive values in yellow, null values in blue. Parameters $T^* = 5$, initial state $\mathbf{X}^{(1)} = [0, 2]^T$.

stages methods presented in [10], (hereafter denoted with SP2S and SP3S). The SP2S method, characterized by the composition in (3.4) applied with $b = b_{BCS} = \frac{3-\sqrt{3}}{6}$ is defined in [10] within the category of one-parameter second-order splitting methods. Similarly, within the family of two-parameter second-order splitting methods:

$$\Psi_h^{(b,a)} := \varphi_{b\,h}^{[\mathbf{Q}]} \circ \varphi_{a\,h}^{[\mathbf{P}]} \circ \varphi_{(1/2-b)h}^{[\mathbf{Q}]} \circ \varphi_{(1-2a)\,h}^{[\mathbf{P}]} \circ \varphi_{(1/2-b)h}^{[\mathbf{Q}]} \circ \varphi_{a\,h}^{[\mathbf{P}]} \circ \varphi_{b\,h}^{[\mathbf{Q}]}, \tag{6.1}$$

the SP3S method (also known in the literature as the BlCaSa method [24]) corresponds to the values $b = 0.11888010966548$ and $a = 0.29619504261126$. We include a comparison with SP3S method as it has been shown to be the most effective method when considering the case of multivariate Gaussian distributions [12].

It is worth emphasizing that the comparison methods we have chosen for evaluation do not necessitate any preconditioning. Therefore, it may be crucial to consider the computational cost associated with the initial diagonalization step when comparing our procedure to these methods. Nevertheless, for problems with low dimensions, the inclusion of this added cost does not have a substantial impact on the results.

We set a sample number $L = 1000$ and we run the experiment using a fixed path length of $T^* = 5$ for all integrators. The selection of step sizes was made to equalize the computational costs with respect to the SV scheme. Specifically, we considered one step of the three-stage method SP3S to be equivalent in cost to three steps of SV. Similarly, one step of the two-stage methods SP2S and nSP2S was considered equivalent in terms of cost to the composition of two steps of SV. For this purpose, we utilize a step size of $h_{SP3S} = 0.6$ for the three-stage SP3S method. Subsequently, we employ a step size of $h_{SV} = \frac{1}{3}\, h_{SP3S} = 0.2$ for the SV scheme and $h_{SP2S} = 2\, h_{SV} = 0.4$ for both two-stage schemes SP2S and nSP2S methods, respectively. In our approach, the method nSP2S runs with $b = h_b^{-1}(0.4) \approx 0.191795$. Fig. 1 we depict accepted and rejected samples in the plane of the target probability density function for all of the methods considered. The ellipse presented corresponds to region where the theoretical probability density function, evaluated with built-in Matlab function mvnpdf, is strictly positive, with the area outside the ellipse representing zero probability. This visualization makes it evident that nSP2S facilitates a more effective exploration of the entire state space. Moreover, as theoretically predicted, nSP2S maintains the maximum acceptance rate $AR = 1$.

6.2. Multivariate Gaussian distribution

In a more general case, we consider a multivariate Gaussian distribution as the target distribution [3,10] defined as follows:

$$\pi(\mathbf{q}) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{d} j^2 q_j^2\right),$$

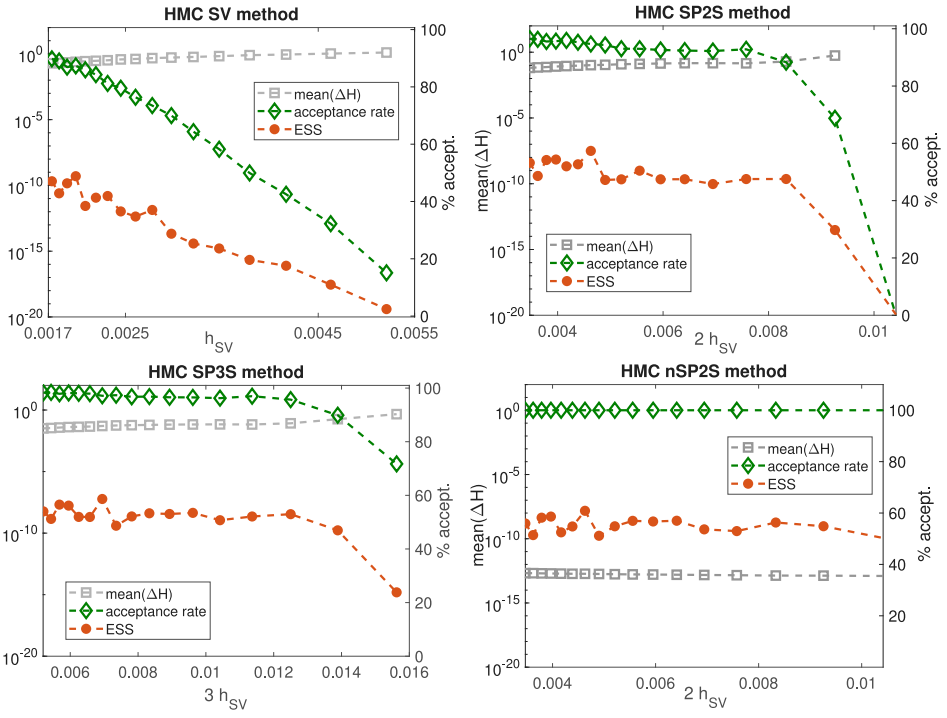

Fig. 2. Multivariate Gaussian target with $L = 5000$ samples of dimension $d = 256$; acceptance rate percentage (green diamonds), mean of ΔH (grey squares) and ESS percentage (orange points) comparison for SV, SP2S, SP3S and nSP2S integrators, measuring with different step size h (for details, refer to the text).

where $\mathbf{q} = [q_1, \dots, q_d]^T$ represents the position variables. The potential energy function $U(\mathbf{Q})$ is given by $\frac{1}{2}\,\mathbf{Q}^T\, D_\alpha^{-1}\,\mathbf{Q}$, where $\mathbf{Q}$ is the vector of independent variables with zero mean and standard deviations $\alpha_j = \frac{1}{j}$ for $j = 1, \dots, d$. The kinetic energy function $K(\mathbf{P})$ is set to $\frac{1}{2}\,\mathbf{P}^T\,\mathbf{P}$, where $\mathbf{P}$ represents the momentum variables.

As in the previous example, we compared our algorithm nSP2S with SV, SP2S, and SP3S methods. We did not consider further improvements of these methods, as provided in [9], since, like the previous methods, they do not preserve the energy of Gaussian distributions.

To perform the experiments we used the same parameters used in [12]. First, in Fig. 2, we present the results of our experiments with $d = 256$, setting the number of samples to $L = 5000$ and choosing a number of burn-in samples equal to 1000. The initial $\mathbf{q}^{(1)}$ is drawn from the target $\pi(\mathbf{q})$. For the chosen methods, we fix the path length to $T^* = 5$. To ensure a fair comparison in terms of computational cost, the SP3S method advances with step size $h_{SP3S} = 5/(320 + 40\,l)$ for $l = 0, 1 \dots 16$. Then, we run SV method with step size $h_{SV} := \frac{1}{3}\,h_{SP3S}$ and we adopt $h_{SP2S} = 2\,h_{SV} = \frac{2}{3}\,h_{SP3S}$ for both SP2S and nSP2S methods. To ensure our method running with a step size of $\frac{2}{3}h_{SP3S}$, we set $b = h_b^{-1}(\frac{2}{3}\,h_{SP3S})$. For all the experiments we have measured the acceptance rate, the mean of 4000 samples of the energy errors $\Delta H^{(i)}(\mathbf{q}^{(i)}, \mathbf{p}^{(i)}) = H(\mathbf{q}^*, \mathbf{p}^*) - H(\mathbf{q}^{(i)}, \mathbf{p}^{(i)})$ and the effective sample size ESS of the first component $q_1^{(i)}$ of $\mathbf{q}^{(i)}$ which corresponds to the component with largest standard deviation $\alpha_1 = 1$ (see [25]).

On the left vertical axis, the mean energy errors (grey squares) are plotted on a logarithmic scale for different step sizes h. On the right vertical axis, we show the acceptance percentage rate (green diamonds) and the effective sample size (ESS) (orange circles) in linear scale for SV (top-left), SP2S (top-right), SP3S (down-left) and the novel nSP2S method (down-right). As expected, our approach achieves a 100% acceptance rate with mean energy errors of the order of 10^{-16}. Moreover, the ESS remains above 50%. It is important to highlight that, unlike the SP2S method, which encounters stability issues with sufficiently large step sizes as depicted in the top-right section of Fig. 2, the nSP2S method is capable of running with larger step sizes. This results in a lower computational cost without compromising its performance. It is worth noting that Theorem 3.4 guarantees the stability of the approximations even when we increase the dimensionality d of the problem. This is because the stability condition is inherently satisfied for each oscillator, without the need for any additional adjustment of the time step, as required by the SP2S method [10,14]. To illustrate this point, we provide an example in Fig. 3 with $d = 500$ and we choose the time step h varying in a large interval from 0.004 to 0.2. The same Figure includes (on the left) plots for the energy error and acceptance rates for both SP2S and nSP2S methods. Additionally, we have included (on the right) the values of b used by the nSP2S and SP2S methods during the integration in the same interval of h.

Furthermore, in Fig. 4, we present the results of the conducted qualitative comparative analysis by estimating the mean and standard deviation of the samples for nSP2S and SP3S, using the Gaussian target distribution with $d = 256$ and setting a sample number $L = 1000$. We set $T^* = 5$ and we run SP3S with $h_{SP3S} = 0.011$ and nSP2S with $h_{SP2S} = \frac{2}{3}\,h_{SP3S} = 0.007\overline{3}$ (to which

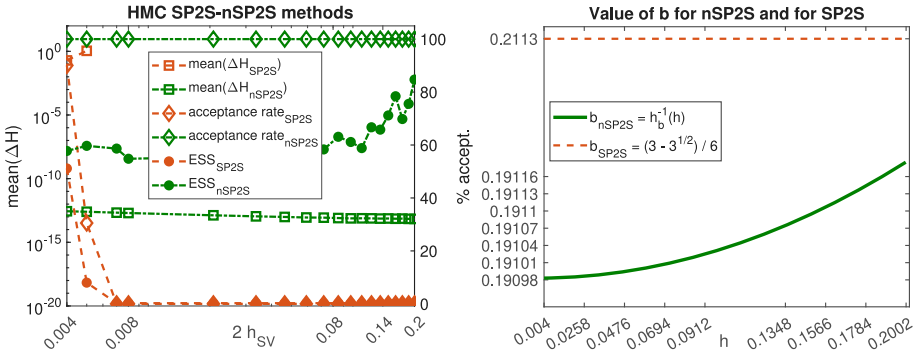

Fig. 3. Multivariate Gaussian target with $L = 5000$ samples of dimension $d = 500$; acceptance rate percentage, mean of ΔH and ESS percentage comparison for SP2S and nSP2S integrators, measuring with step size $h \in [0.004, 0.2]$ (left), Values of stepsizes $h \in [0.004, 0.2]$ against values of the corresponding b's for SP2S and nSP2S integrators (right).

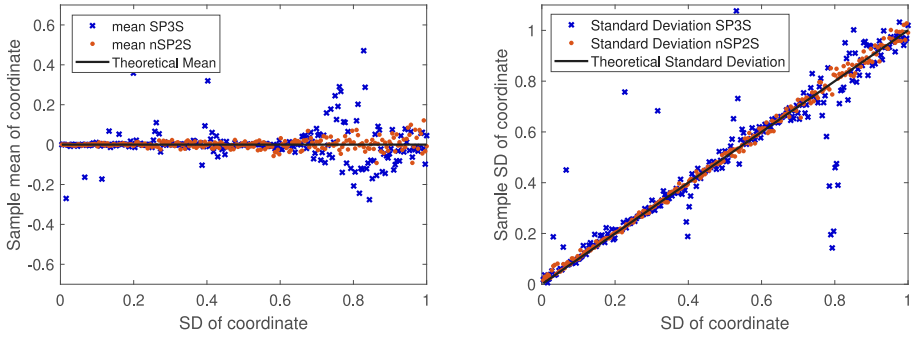

Fig. 4. Multivariate distribution. Comparison between the SP3S method, running with $h_{SP3S} = 0.011$, and the nSP2S method, running with $h_{SP2S} = \frac{2}{3} h_{SP3S} = 0.007\overline{3}$. Estimates of means (left) and standard deviations (right) for the $d = 256$ dimensional example. On the x-axes are reported the standard deviations α_j of each variable q_j for $j = 1, \dots, d$, on the y-axes the estimated values of means and standard deviations, each evaluated from $L = 1000$ iterations, are reported.

corresponds $b \approx 0.1909833$). The estimation of means and standard deviations from the obtained samples is compared to the theoretical values of the distribution, i.e., zero mean and standard deviations α_j for $j = 1, \dots, d$. One can notice that the estimates generated by the HMC algorithm using trajectories evaluated with nSP2S show a greater accuracy, compared to the SP3S method which rejects only 6 proposals.

6.3. Perturbed Gaussian models

To highlight the effectiveness of our approach, we consider a generic potential energy function $U(\mathbf{Q})$, as the specific case $\varepsilon = 1$ of a potential energy of a perturbed Gaussian class defined by

$$U_\varepsilon(\mathbf{Q}) = \frac{1-\varepsilon}{2}\mathbf{Q}^T\mathbf{Q} + \varepsilon\, U(\mathbf{Q}), \qquad 0 \le \varepsilon \le 1.$$

It is trivial to check that $U_1(\mathbf{Q}) = U(\mathbf{Q})$. We associate a kinetic energy function $K(\mathbf{P}) = \frac{1}{2}\mathbf{P}^T\mathbf{P}$ so that the Hamiltonian can be written as

$$K(\mathbf{P}) + U_\epsilon(\mathbf{Q}) = \frac{1}{2}\mathbf{P}^T\mathbf{P} + \frac{1-\varepsilon}{2}\mathbf{Q}^T\mathbf{Q} + \varepsilon\, U(\mathbf{Q})$$

and the Hamiltonian system is given by

$$\frac{d\mathbf{Q}}{dt} = \mathbf{P}, \qquad \frac{d\mathbf{P}}{dt} = -\nabla_{\mathbf{Q}} U_\epsilon(\mathbf{Q}) = (\varepsilon - 1)\mathbf{Q} - \varepsilon\, \nabla_{\mathbf{Q}} U(\mathbf{Q}).$$

We apply the method $\mathbf{Y}_{n+1} = \Psi^{(b)}_{h_b}\mathbf{Y}_n$ with $\Psi^{(b)}_h$ defined in (3.4) which provides an energy preserving method for the above Hamiltonian system when $\varepsilon = 0$. We investigate the performance of the algorithm with respect to the acceptance rates and to the error in energy at $\varepsilon = 1$ for two different specific models: the Log-Gaussian Cox model and the logistic regression.

6.3.1. Log-Gaussian Cox model

As first example we considered, as target, the Log-Gaussian Cox distribution [26]. For this model the data set is organized in the vector $\mathbf{X} = \left[X_{1,1}, \dots, X_{1,d}, X_{2,1}, \dots, X_{2,d}, \dots X_{d,1}, \dots X_{d,d}\right]^T$, representing the number of points $X_{i,j}$ in each cell (i, j) of a $d \times d$

dimensional grid in $[0,1]\times[0,1]$. The purpose is to sample the variable $\mathbf{Q} = \left[Q_{1,1},\dots,Q_{1,d},Q_{2,1},\dots,Q_{2,d},\dots Q_{d,1},\dots Q_{d,d}\right]^T$ from the probability distribution given by

$$\mathcal{P}(\mathbf{Q}) = \prod_{i,j=1}^{d} \exp(X_{i,j}Q_{i,j} - m\exp(Q_{i,j}))\, \exp\left(-\frac{1}{2}(\mathbf{Q}-\mu\mathbf{1})^T S^{-1}(\mathbf{Q}-\mu\mathbf{1})\right)$$

where $m = 1/d^2$ represents the area of each cell and the matrix S is given

$$S = \begin{pmatrix} T_1 & T_2 & T_3 & \dots & \dots & T_d \\ T_2 & T_1 & T_2 & \dots & & \vdots \\ T_3 & T_2 & \dots & \dots & \dots & \dots \\ \vdots & \dots & \dots & \dots & T_2 & T_3 \\ \vdots & & \dots & T_2 & T_1 & T_2 \\ T_d & \dots & \dots & T_3 & T_2 & T_1 \end{pmatrix}, \quad T_i(k,j) = \sigma^2 e^{-\frac{\sqrt{(1-i)^2+(k-j)^2}}{\beta d}}, \quad i,\, k,\, j = 1,\dots d.$$

with σ^2, β, are fixed parameters and $\mu = \log\left(\sum_{i,j=1}^{d} X_{i,j}\right) - \sigma^2/2$. The potential energy function is defined as

$$\begin{aligned} U(\mathbf{Q}) = -\log[\mathcal{P}(\mathbf{Q})] &= \sum_{i,j=1}^{d} m\exp(Q_{i,j}) - X_{i,j}Q_{i,j} + \frac{1}{2}(\mathbf{Q}-\mu\mathbf{1})^T S^{-1}(\mathbf{Q}-\mu\mathbf{1}) \\ &= \frac{1}{2}(\mathbf{Q}-\mu\mathbf{1})^T S^{-1}(\mathbf{Q}-\mu\mathbf{1}) + m\,\mathbf{1}^T\exp(\mathbf{Q}) - \mathbf{X}^T\mathbf{Q}. \end{aligned}$$

with

$$\nabla_{\mathbf{Q}} U(\mathbf{Q}) = S^{-1}(\mathbf{Q}-\mu\mathbf{1}) + m\exp(\mathbf{Q}) - \mathbf{X}.$$

The Log-Gaussian Cox model is particularly relevant as a point process for modeling presence-only species distribution [27], such as the case of Scots pines in Eastern Finland [26,28] or the spread of the invasive species Eucalyptus sparsifolia in Australia [27]. In our study, we address a similar problem related to alien plants, specifically focusing on the highly competitive invasive species *Ailanthus Altissima (Mill.) Swingle* thriving in the Murgia Alta Natura 2000 protected area and National Park in southern Italy [15]. For the purpose of our tests, we used a dataset consisting of a high-resolution (2 m) mapping of Ailanthus Altissima presence, obtained through multi-temporal remote sensing satellite data and machine learning techniques using a two-stage hybrid classification process [29]. The satellite images used in the dataset were provided by the European Space Agency (ESA) under the Data Warehouse 2011–2014 policy within the FP7-SPACE BIO_SOS project (www.biosos.eu) and the European LIFE project (www.ispacnr.it/progetto-life-alta-murgia).

From the complete dataset, we extracted a small area, scaled in the square $[0,1]\times[0,1]$ containing 185 trees plotted in Fig. 7 (top-left). To calculate the parameters of the Log Gaussian Cox model we followed the moment-based estimation methodology described in [30], resulting in $\beta = 0.127$, $\sigma^2 = 3.5881$, and $\mu = \log(185) - \sigma^2/2$. We collected a Markov chain of $L = 5000$ samples of dimension $d^2 = 400$ with HMC, discarding the first 1000 burn-in samples.

In Fig. 5, we present the comparison results with competitors based on acceptance percentage and mean energy errors. We fix the path length $T = 3$ and we use $h_{SP3S} = \frac{3}{n}$ with $n = 60, 40, 30, 20, 15, 12, 10$ for the SP3S method. Then we run SV method with $h_{SV} = \frac{1}{3}\,h_{SP3S}$. For both SP2S and nSP2S methods, we adopt $h_{SP2S} = 2\,h_{SV} = \frac{2}{3}\,h_{SP3S}$. To ensure that our method nSP2S runs with the correct step size, we set $b = h_b^{-1}(\frac{2}{3}h_{SP3S})$. In Fig. 5 the horizontal axis of each plot represents the different step sizes. Even if our approach achieves higher acceptance rates compared to the SP2S, which is considered optimal in literature within the class of second-order two-stage splitting schemes, these results do not show clearly the superiority of our procedure when applied without the proposed adaptive approach. Furthermore, compared to the SP3S method, recognized as optimal in the class of second-order three-stage splitting schemes, our proposed nSP2S procedure demonstrates only slightly better acceptance rates and, as regards efficiency, for several choices of h the ESS of SP3S is even greater.

To show the potential advantage deriving from the adaptive approach of our method (Algorithm 5.1) compared with the non-adaptive one, we expanded the step size range $h_{SP3S} = \frac{3}{n}$, adding values for n of 9, 8, 7, 6, 5, 4, for the SP3S method. Subsequently, in Fig. 6, we compare its performance with nSP2S method, applied both with and without the adaptive procedure. The adaptive procedure starts by setting $b_{\max}$ to $h_b^{-1}(h_{SP2S})$, where $h_{SP2S} = \frac{2}{3}h_{SP3S}$ and the reduction of b is determined by the parameter $red = 0.99$. The results of this comparison are presented in Fig. 6, where a noticeable enhancement is shown with the novel adaptive procedure.

Regarding the above comparison, it is important to note that directly comparing the adaptive algorithm to other non-adaptive methods presents challenges. The provided results, for example, do not consider the increased computational cost resulting from reductions in b coefficient, which in turn lead to corresponding reductions in the step size h_b (refer to Table 1). However, as mentioned earlier, here we limit on demonstrating the potential benefits of adaptively selecting the pairs $(b,\, h_b)$ within the considered family of two-step second-order splitting procedures.

Finally, in Fig. 7 (top-right) we display the estimated intensity map of the Ailanthus trees obtained using the novel adaptive procedure (Algorithm 5.1) in correspondence of the higher dimension $d^2 = 4096$. Setting $T_{max} = 3$, starting from the value $b_{max} = b_{BCS} \approx 0.2113$ and a reduction factor of $red = 0.997$, the method quickly converges to the shown configuration with acceptance rate $AR = 0.6443$. In Fig. 7 (down-left) we depict the convergence of values of b to the optimal value $b = b_{opt} = 0.1909970$ and their corresponding values of h_b approaching $h_{b_{opt}} = 0.0527835$, with respect to the number of sample iterations L.

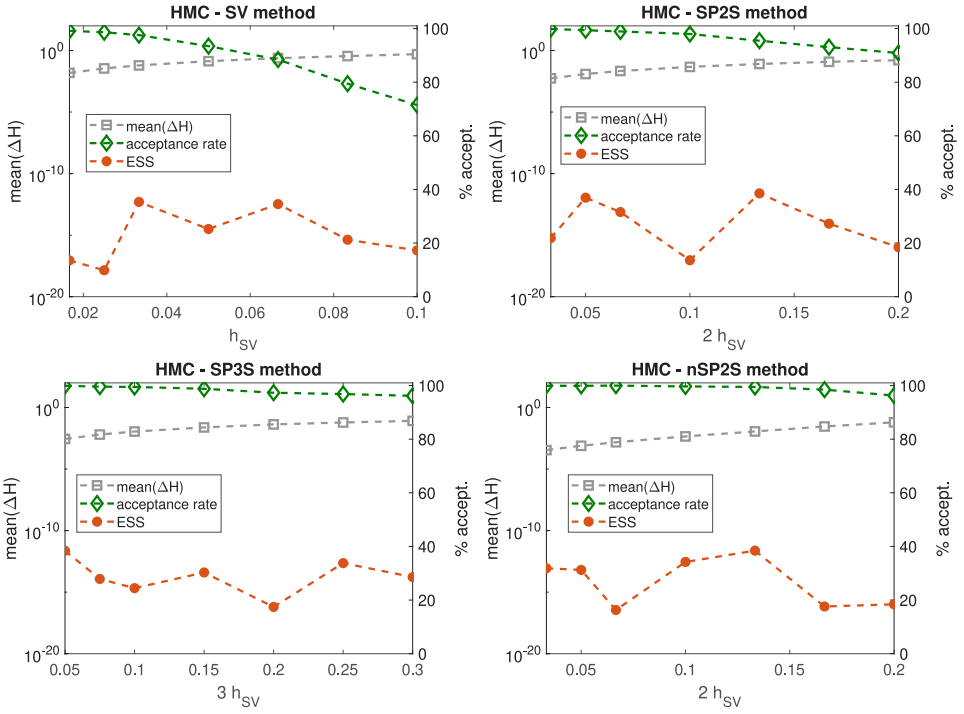

Fig. 5. Log-Gaussian Cox target with $L = 5000$ samples of dimension $d^2 = 400$; acceptance rate percentage (green diamonds), mean of ΔH (grey squares) and ESS percentage (orange points) comparison for SV, SP2S, SP3S and nSP2S integrators, measuring with different step size h (for details, refer to the text).

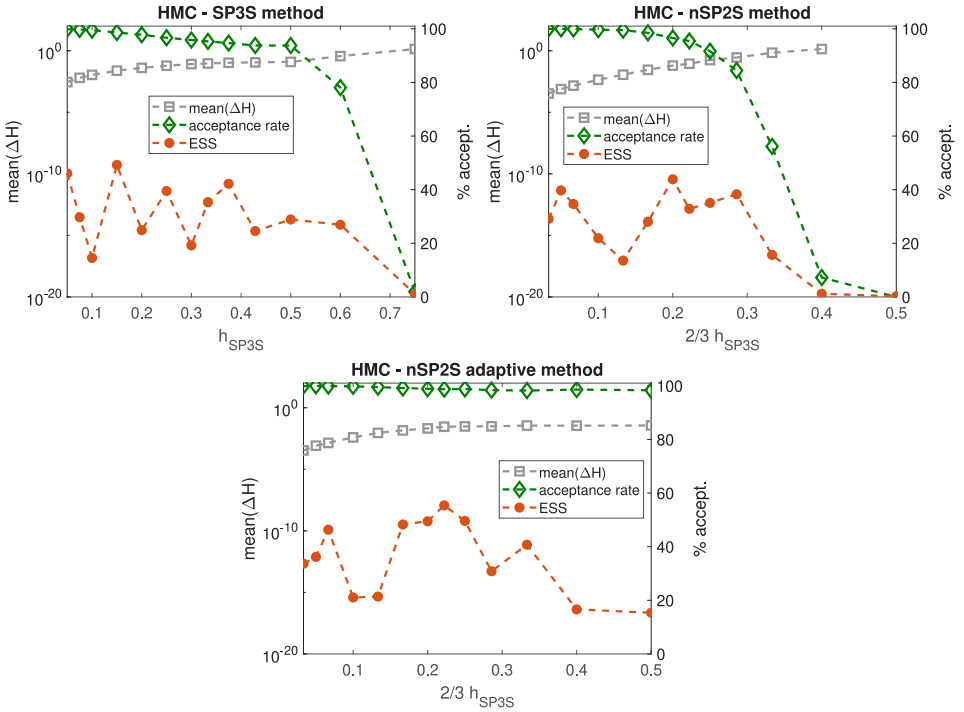

Fig. 6. Log-Gaussian Cox target with $L = 5000$ samples of dimension $d^2 = 400$; comparison for SP3S and nSP2S integrators, and the nSP2S method with adaptive procedure and reduction parameter $red = 0.99$. Acceptance rate percentage (green diamonds), mean of ΔH (grey squares) and ESS percentage (orange points).

Table 1
Log-Gaussian Cox target with $L = 5000$ samples of dimension $d^2 = 400$. Range of b coefficients and related h_b values for the adaptive nSP2S procedure in Fig. 6.

h_{SP2S}	b_{max}	Range of b	Range of h_b
0.033	0.1909886	$b \in [0.1909884, b_{max}]$	$h_b \in [0.0328, h_{SP2S}]$
0.050	0.1909956	$b \in [0.1909955, b_{max}]$	$h_b \in [0.0497, h_{SP2S}]$
0.066	0.1910054	$b \in [0.1910045, b_{max}]$	$h_b \in [0.0653, h_{SP2S}]$
0.100	0.1910334	$b \in [0.1910282, b_{max}]$	$h_b \in [0.0946, h_{SP2S}]$
0.133	0.1910727	$b \in [0.1910521, b_{max}]$	$h_b \in [0.1170, h_{SP2S}]$
0.166	0.1911232	$b \in [0.1910687, b_{max}]$	$h_b \in [0.1303, h_{SP2S}]$
0.200	0.1911850	$b \in [0.1910829, b_{max}]$	$h_b \in [0.1407, h_{SP2S}]$
0.222	0.1912324	$b \in [0.1910935, b_{max}]$	$h_b \in [0.1479, h_{SP2S}]$
0.250	0.1912989	$b \in [0.1911034, b_{max}]$	$h_b \in [0.1544, h_{SP2S}]$
0.285	0.1913959	$b \in [0.1910915, b_{max}]$	$h_b \in [0.1466, h_{SP2S}]$
0.333	0.1915456	$b \in [0.1910912, b_{max}]$	$h_b \in [0.1464, h_{SP2S}]$
0.400	0.1917948	$b \in [0.1911095, b_{max}]$	$h_b \in [0.1583, h_{SP2S}]$
0.500	0.1922562	$b \in [0.1911030, b_{max}]$	$h_b \in [0.1542, h_{SP2S}]$

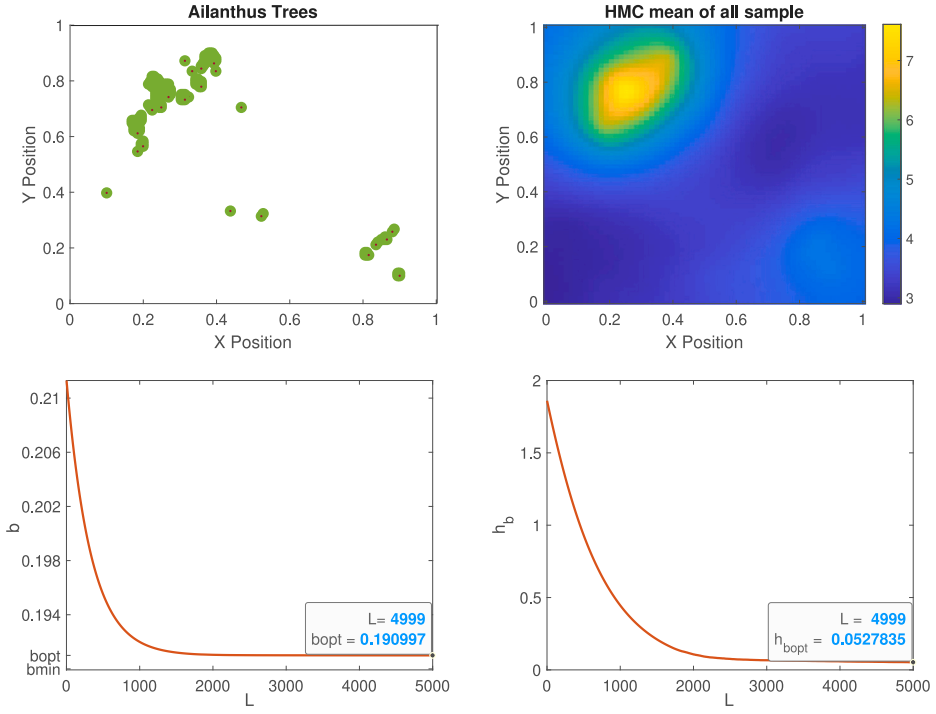

Fig. 7. Presence of Ailanthus trees (top-left), position of the trees (green points) suitably scaled between 0 and 1. Estimated intensity map of Ailanthus obtained accordingly to the Algorithm 5.1 (top-right). Adaptive choice of b against sample iteration L (down-left), corresponding values of h_b against sample iteration L (down-right).

6.3.2. Logistic regression model

As a second test, we consider a Bayesian Logistic regression model, following the notation used in [31]. We denote the n-dimensional vector of labels associated with the instances matrix $X \in \mathbb{R}^{n\times(d+1)}$ as $\mathbf{Y} = [Y_1, \dots, Y_n]^T$. Let $\mathbf{x}_k = [X_{k,0}, \dots, X_{k,d}]^T$ the $(d+1)$-dimensional (column) vector corresponding to the kth row of the matrix X, for $k = 1, \dots, n$. The regression coefficients for the d covariates and the intercept are collected in the vector $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_d]^T$. To specify a prior distribution for $\boldsymbol{\beta}$, we assume a multivariate normal distribution with covariance matrix $D = \sigma^2 I$, where I is the $(d+1)$-dimensional identity matrix and σ^2 is a freely chosen variance. The goal is to sample the parameters $\boldsymbol{\beta}$ from the resulting distribution, which can be expressed as follows:

$$\begin{aligned}\mathcal{P}(\boldsymbol{\beta}) \quad &\propto \quad \exp\Big(\boldsymbol{\beta}^T X^T (\mathbf{Y} - \mathbf{1_n}) - \sum_{j=1}^{n}\Big[\log(1+\exp(-\mathbf{x}_j^T\boldsymbol{\beta}))\Big]\Big)\exp\Big(-\frac{1}{2}\boldsymbol{\beta}^T D^{-1}\boldsymbol{\beta}\Big)\\ &= \quad \exp\Big(\boldsymbol{\beta}^T X^T \mathbf{Y} - \sum_{j=1}^{n}\Big[\log(1+\exp(\mathbf{x}_j^T\boldsymbol{\beta}))\Big]\Big)\exp\Big(-\frac{1}{2}\boldsymbol{\beta}^T D^{-1}\boldsymbol{\beta}\Big)\end{aligned}$$

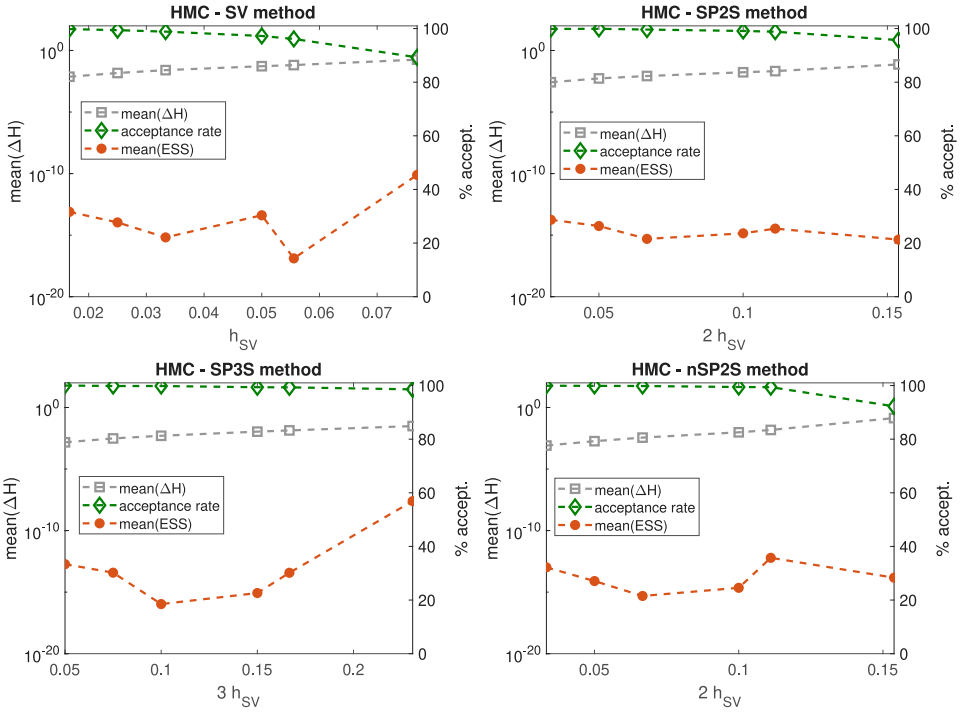

Fig. 8. Logistic regression model with $L = 5000$ samples of dimension $d + 1 = 8$; acceptance rate percentage (green diamonds), mean of ΔH (grey squares) and mean of ESS percentage (orange points) comparison for SV, SP2S, SP3S and nSP2S integrators, measuring with different step sizes h (for details, refer to the text).

The potential energy function, is defined as:

$$U(\boldsymbol{\beta}) = -\log[\mathcal{P}(\boldsymbol{\beta})] = \frac{1}{2}\boldsymbol{\beta}^T D^{-1}\boldsymbol{\beta} + \sum_{k=1}^{n}\left[\log(1+\exp(\mathbf{x}_k^T\boldsymbol{\beta}))\right] - \boldsymbol{\beta}^T X^T \mathbf{Y}$$

Without loosing of generality, we can set $\sigma = 1$ as the same results are obtained on the scaled dataset $\tilde{X} = X/\sigma$. The gradient of the potential energy function is given by

$$\nabla_{\boldsymbol{\beta}} U(\boldsymbol{\beta}) = \boldsymbol{\beta} - X^T\left(\mathbf{Y} - \left[\frac{\exp(\mathbf{x}_1^T\boldsymbol{\beta})}{1+\exp(\mathbf{x}_1^T\boldsymbol{\beta})}, \frac{\exp(\mathbf{x}_2^T\boldsymbol{\beta})}{1+\exp(\mathbf{x}_2^T\boldsymbol{\beta})}, \ldots, \frac{\exp(\mathbf{x}_n^T\boldsymbol{\beta})}{1+\exp(\mathbf{x}_n^T\boldsymbol{\beta})}\right]^T\right).$$

For this experiment, we have tested the method on the logistics regression model using the Pima Indian, Ripley, Heart, German credit, and Australian credit datasets from the UCI repository [32], which includes various matrices of instances and labels. We report the results obtained with the Pina dataset, since the tests performed with the other datasets have no significant differences in terms of performance.

Following standard practice, we applied a normalization procedure to the dataset, ensuring that it has a mean of 0 and a standard deviation of 1. Prior to normalization, we performed a scaling procedure on the data, taking into account the chosen value of σ.

We collected a Markov chain of $L = 5000$ samples of dimension $d + 1 = 8$ with HMC, discarding the first 1000 burn-in samples. In Fig. 8, we present the comparison results with competitors based on acceptance percentage and mean energy errors. We fix the path length $T = 3$ and we use $h_{SP3S} = \frac{3}{n}$ with $n = 60, 40, 30, 20, 18, 13$, for the SP3S method. Then we run SV method with $h_{SV} := \frac{1}{3}h_{SP3S}$. For SP2S we adopt $h_{SP2S} = 2\,h_{SV} = \frac{2}{3}h_{SP3S}$ and we set $b = h_b^{-1}(\frac{2}{3}h_{SP3S})$ for nSP2S. The horizontal axis of each plot represents the different step sizes used. While our method demonstrates superior performance compared to SV and similar performance to other competitors for lower values of the step size, it exhibits poor performance for larger step sizes. It is worth noting that, although both two-stage integrators, SP2S and nSP2S, frequently experience failures with the large step sizes making it challenging to conclusively determine the superior two-stage scheme, the SP3S scheme always exhibits the best performance.

To give evidence of the improved behavior of our proposed procedure within the adaptive Algorithm 5.1 we expanded the step size range $h_{SP3S} = \frac{3}{n}$, adding values for n of 12, 10, 9, 8, 7, 6, 5, 4, for the SP3S method. In Fig. 9, we compare its performance with non-adaptive and adaptive nSP2S methods. As in the previous comparison, the adaptive procedure starts by setting $b_{\max} = h_b^{-1}(h_{SP2S})$, where $h_{SP2S} = \frac{2}{3}h_{SP3S}$. The reduction of b is then determined by parameter $red = 0.98$. The results of this comparison are presented in Fig. 9, where, once again, the obtained enhancement with the novel adaptive procedure is demonstrated. The range of values assumed by the b coefficient, along with the corresponding ranges for the step size h_b, is provided in Table 2.

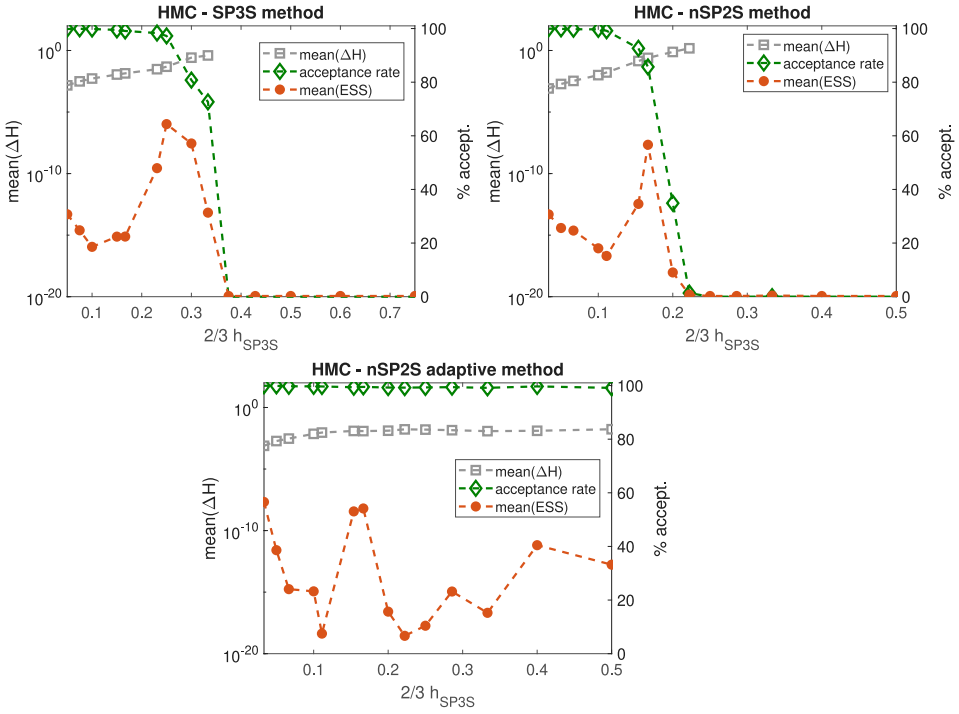

Fig. 9. Logistic regression model with $L = 5000$ samples of dimension $d + 1 = 8$; acceptance rate percentage (green diamonds), mean of ΔH (grey squares) and mean of ESS percentage (orange points). Comparison for SP3S, and nSP2S integrators in correspondence of $red = 0.98$.

Table 2
Logistic regression model with $L = 5000$ samples of dimension $d + 1 = 8$. Range of b coefficients and related h_b values for the adaptive nSP2S procedure in Fig. 9.

h_{SP2S}	b_{max}	Range of b	Range of h_b
0.033	0.1909886	$b \in [0.1909882, b_{max}]$	$h_b \in [0.0320, h_{SP2S}]$
0.050	0.1909956	$b \in [0.1909942, b_{max}]$	$h_b \in [0.0470, h_{SP2S}]$
0.066	0.1910054	$b \in [0.1910006, b_{max}]$	$h_b \in [0.0590, h_{SP2S}]$
0.100	0.1910334	$b \in [0.1910188, b_{max}]$	$h_b \in [0.0842, h_{SP2S}]$
0.111	0.1910334	$b \in [0.1910221, b_{max}]$	$h_b \in [0.0880, h_{SP2S}]$
0.153	0.1910727	$b \in [0.1910256, b_{max}]$	$h_b \in [0.0919, h_{SP2S}]$
0.166	0.1911232	$b \in [0.1910273, b_{max}]$	$h_b \in [0.0937, h_{SP2S}]$
0.200	0.1911850	$b \in [0.1910231, b_{max}]$	$h_b \in [0.0891, h_{SP2S}]$
0.222	0.1912324	$b \in [0.1910219, b_{max}]$	$h_b \in [0.0878, h_{SP2S}]$
0.250	0.1912989	$b \in [0.1910266, b_{max}]$	$h_b \in [0.0930, h_{SP2S}]$
0.285	0.1913959	$b \in [0.1910287, b_{max}]$	$h_b \in [0.0951, h_{SP2S}]$
0.333	0.1915456	$b \in [0.1910221, b_{max}]$	$h_b \in [0.0880, h_{SP2S}]$
0.400	0.1917948	$b \in [0.1910300, b_{max}]$	$h_b \in [0.0965, h_{SP2S}]$
0.500	0.1922562	$b \in [0.1910241, b_{max}]$	$h_b \in [0.0902, h_{SP2S}]$

Furthermore, in order to show that the estimated samples β obtained with the adaptive approach are consistent with the frequency estimates calculated using the generalized linear model (glm) [33] the Algorithm 5.1 has been run by setting, the path length $T_{max} = 3$, $b_{max} = 0.1932$ (case 3 of Section 5.1), $red = 0.98$ and we plot the result in Fig. 10 where the red line representing the frequentist estimates that falls within the central location of the histograms of $\beta_0, \beta_1, \ldots, \beta_7$. In Fig. 11 we show how the parameter b (on the left left) and the value of h_b (on the right) change with respect to sample iterate L converging to the value $b_{opt} = 0.1910212$ to which corresponds $h_{b_{opt}} = 0.0870547$.

7. Conclusions

The recent research literature focuses on the search for efficient volume-preserving and reversible integrators that can replace the Störmer Verlet method in practical implementations of the HMC method. The effectiveness of these numerical algorithms is measured by their ability to reduce the expectation of the energy error when applied to univariate and multivariate Gaussian distributions [10,12].

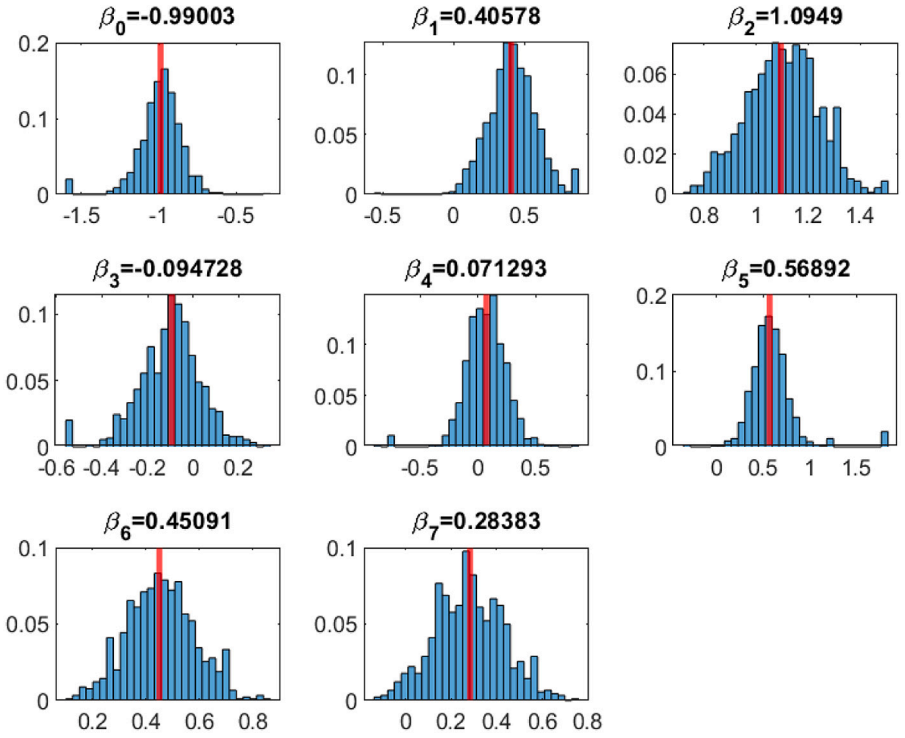

Fig. 10. Exploration of logistic regression using the Pima Indian dataset. Examination of posterior estimates through Hamiltonian Monte Carlo sampling with nSP2S (histograms), complemented by a comparison with frequentist estimates (depicted of a vertical lines) using Generalized Linear Model (glm).

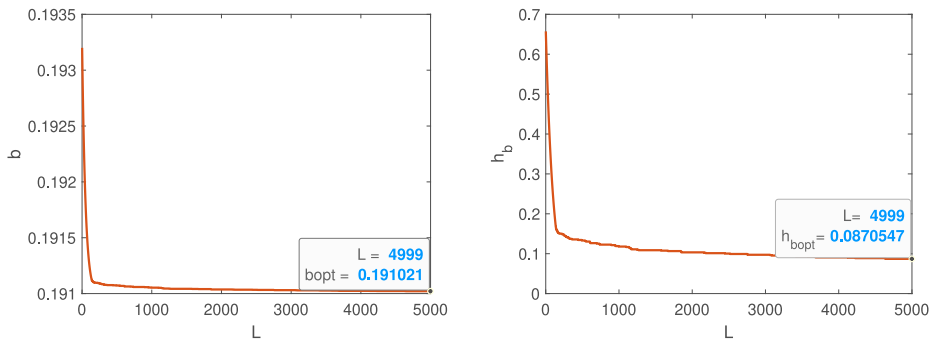

Fig. 11. Logistic regression target: adaptive choice of b (on the left) and resulting h_b (on the right) against the sample iterate L in correspondence of $red = 0.98$.

In this paper, we have reversed and improved the adaptive selection of the parameter b from the one-parameter family of second-order splitting procedures, appeared in HMC literature only recently [13,14]. Once the value of b in a suitable interval I has been selected to ensure the stability of the chosen method within the family of integrators, the step size used to advance in time is uniquely determined by the function $h_b(b) = \sqrt{\frac{4b^2-6b+1}{b^2(2b-1)}}$. By setting the step size $h = h_b(b)$ the expectation of the energy error for Gaussian distributions, both univariate and multivariate, is nullified. With all proposals being accepted by construction, our proposed method outperforms the competitors mentioned in [9,10,12]. Consequently, the above competitors cannot be considered optimal for Gaussian distributions as they do not preserve energy, and they are not more cost-effective than the methods examined in this study.

Moreover, we propose to use the same the couple (b, h_b) with $b \in I$ for sampling with HMC from generic distributions. To assess the effectiveness of our proposed approach, we conducted tests on general classes of target distributions, such as the Log-Gaussian Cox model and Bayesian logistic regression. Our criterion proves to be effective in achieving good performance both in terms of acceptance percentage rate and the effective sample size (ESS). In comparison to the SV scheme, our proposed procedure demonstrates superior performance for both Log-Gaussian Cox and Bayesian logistic regression models, aligning with the second-order scheme identified as optimal in two-stage integrators. Compared to three-stage second-order splitting procedures, especially

in the case of the Bayesian logistic regression model, our method demonstrates a lower performance. In our future research, we will attempt to extend our approach to investigate whether the SP3S method can be surpassed by an energy-preserving integrator (if any), properly choosing the triplets (a, b, h) in the class of three-stages second order splitting scheme in (6.1).

Additionally, we propose to enhance the performance of the novel integrator nSP2S by incorporating an adaptive selection of the parameter b, which identifies a specific method within the family of one-parameter splitting integrators. In this adaptive approach, we begin with a sufficiently large initial value for b, and we decrease its value each time a sample is not accepted. Another possible variant involves reducing the parameter b each time a specified target rejection rate is exceeded. For both examples considered, we verified the effectiveness in addressing high dimensionality problems and an observed ability to enhance the performance compared to non-adaptive procedures. As a future research direction we intent to exploit the use of more robust criteria for an adaptive method's selection in one-parameter families of integrators having the objective of reducing the computational cost.

Availability of codes

All codes were developed in MATLAB. For implementing algorithms with fixed parameters, we made slight modifications to the original codes, which were kindly provided to the authors by Prof. Mari Paz Calvo upon their request. We have released our codes implementing both the fixed and the novel adaptive procedure on GitHub repository https://github.com/CnrIacBaGit/HMC_adaptive.

CRediT authorship contribution statement

Cristiano Tamborrino: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Investigation, Data curation. **Fasma Diele:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Carmela Marangi:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Cristina Tarantino:** Writing – original draft, Visualization, Software, Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors express sincere gratitude to the editor and the anonymous reviewer for their valuable comments and constructive feedback. Their insightful suggestions and attention to detail have significantly contributed to improving the quality and clarity of this manuscript.

The authors Fasma Diele and Cristiano Tamborrino are members of the Gruppo Nazionale Calcolo Scientifico-Istituto Nazionale di Alta Matematica (GNCS-INdAM). Fasma Diele and Carmela Marangi are funded under the National Recovery and Resilience Plan (NRRP), Italy, Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Award Number: Project code CN 00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, Italy, CUP B83C22002930006, Project title "National Biodiversity Future Center - NBFC. The authors thanks Mr. Cosimo Grippa for his valuable technical support.

Cristiano Tamborrino acknowledge the support of the PNRR project FAIR - Future AI Research, Italy (PE00000013), Spoke 6 - Symbiotic AI, Italy (CUP H97G22000210007) under the NRRP MUR program funded by the NextGenerationEU. Cristina Tarantino is grateful to LIFE Preparatory Project NewLife4Drylands (G.A. LIFE20PRE/IT/000007) and to Alta Murgia National Park Authority for the remarkable ecological support.

Appendix. Energy-preserving linear maps for Gaussian distributions

The chosen step size $h > 0$ is crucial in the implementation of Algorithm 2.1. Too small a step size will waste computation time as it will require a large N in order to reach the final step $T^* = N\,h$. Too large a step size will increase bounded oscillations in the value of the Hamiltonian, which would be constant if the trajectory were simulated by an energy-preserving map. Moreover, when values for h are chosen above the critical stability threshold, which is characteristic of each approximating map Ψ_h, then the Hamiltonian grows without bound, resulting to an extremely low acceptance rate for states proposed by simulated trajectories.

Hence the selection of the step size h should obey to stability constraints. The issue of stability is traditionally faced by means of a test problem; for HCM flows, it is represented by the problem defined by a Gaussian zero-mean distribution for both q and p. Firstly, we account for the one-dimensional problem and then we extend the analysis to the multi-dimensional case.

A.1. Univariate case

We generalize the approach in both [3,10] by considering generic standard deviations, α for q and β for p, with zero correlation. The Hamiltonian dynamics for Q and P define the equations

$$\frac{dQ}{dt} = \frac{P}{\beta^2}, \qquad \frac{dP}{dt} = -\frac{Q}{\alpha^2}. \tag{A.1}$$

Setting $\mathbf{Y} = [Q, P]^T$, the Hamiltonian can be expressed as $H(\mathbf{Y}) = \frac{1}{2}\mathbf{Y}^T \mathcal{D}_2^{-1}\mathbf{Y} = \frac{1}{2}\left(\frac{Q^2}{\alpha^2} + \frac{P^2}{\beta^2}\right)$ where $\mathcal{D}_2 := \begin{bmatrix}\alpha^2 & 0\\ 0 & \beta^2\end{bmatrix}$. Starting from Q_0, P_0, the theoretical solution at $t_n = n\,h$ is represented as a linear map $\mathbf{Y}(t_n) = \mathcal{F}^{(n\,h_\sigma,\sigma)}\,\mathbf{Y}_0$, where

$$\mathcal{F}^{(n\,h_\sigma,\sigma)} := \begin{bmatrix}\cos(n\,h_\sigma) & \sigma^{-1}\sin(n\,h_\sigma)\\ -\sigma\sin(n\,h_\sigma) & \cos(n\,h_\sigma)\end{bmatrix}, \quad \sigma := \frac{\beta}{\alpha}, \quad h_\sigma := \frac{h}{\alpha\,\beta}.$$

Notice that Hamiltonian can be expressed as $H(\mathbf{Y}) = \frac{1}{2\,\alpha\,\beta}\left(\sigma\,Q^2 + \frac{P^2}{\sigma}\right)$.

We mentioned that the numerical map used to replace the theoretical solution with an approximation should be volume-preserving (here equivalent to symplectic) and momentum be flip-reversible. Both characteristics direct our attention to the class of integrators that, when applied to the test problem (A.1), can be expressed as

$$\mathbf{Y}_{n+1} = \mathcal{M}_2^{(h,\sigma)}\,\mathbf{Y}_n$$

where $\mathcal{M}_2^{(h,\sigma)}(1,1) = \mathcal{M}_2^{(h,\sigma)}(2,2)$ and $\det(\mathcal{M}_2^{(h,\sigma)}) = 1$.
Setting $\mathrm{p}_h = \mathcal{M}_2^{(h,\sigma)}(1,1) = \mathcal{M}_2^{(h,\sigma)}(2,2)$, $\mathrm{q}_h = \frac{\sigma}{\sigma^2+1}(\mathcal{M}_2^{(h,\sigma)}(1,2) - \mathcal{M}_2^{(h,\sigma)}(2,1))$ and $\mathrm{e}_h = \frac{1}{\sigma^2+1}(\sigma^2\mathcal{M}_2^{(h,\sigma)}(1,2) + \mathcal{M}_2^{(h,\sigma)}(2,1))$, the matrix $\mathcal{M}_2^{(h,\sigma)}$ can be written as

$$\mathcal{M}_2^{(h,\sigma)} = \begin{bmatrix}\mathrm{p}_h & \mathrm{e}_h + \sigma^{-1}\,\mathrm{q}_h\\ \mathrm{e}_h - \sigma\,\mathrm{q}_h & \mathrm{p}_h\end{bmatrix}, \tag{A.2}$$

and, from $\det(\mathcal{M}_2^{(h,\sigma)}) = 1$, the following relation holds

$$\mathrm{p}_h^2 - (\mathrm{e}_h + \sigma^{-1}\,\mathrm{q}_h)(\mathrm{e}_h - \sigma\,\mathrm{q}_h) = 1. \tag{A.3}$$

The stability of the trajectories depends on eigenvalues of $\mathcal{M}_2^{(h,\sigma)}$ which solve the polynomial

$$\lambda^2 - 2\,\mathrm{p}_h\,\lambda + 1 = 0.$$

When $\mathrm{p}_h^2 - 1 \geq 0$ then the eigenvalues are real with at least one having absolute value greater than one, hence the trajectories are unstable. When $\mathrm{p}_h^2 - 1 < 0$ the eigenvalues are complex with modulus equal to one, hence the trajectories are stable.

The key consideration for what follows is that integrators for which it results $\mathrm{e}_h = 0$ are energy preserving. Indeed, the error in energy at each step is given by

$$\begin{aligned}\Delta_2^{(n,h)} &:= H(\mathbf{Y}_{n+1}) - H(\mathbf{Y}_n) = \frac{1}{2}\mathbf{Y}_{n+1}^T\mathcal{D}_2^{-1}\mathbf{Y}_{n+1} - \frac{1}{2}\mathbf{Y}_n^T\mathcal{D}_2^{-1}\mathbf{Y}_n\\ &= \frac{1}{2}\mathbf{Y}_n^T\mathcal{M}_2^{(h,\sigma)^T}\mathcal{D}_2^{-1}\mathcal{M}_2^{(h,\sigma)}\mathbf{Y}_n - \frac{1}{2}\mathbf{Y}_n^T\mathcal{D}_2^{-1}\mathbf{Y}_n\\ &= \frac{1}{2}\mathbf{Y}_n^T\left(\mathcal{M}_2^{(h,\sigma)^T}\mathcal{D}_2^{-1}\mathcal{M}_2^{(h,\sigma)} - \mathcal{D}_2^{-1}\right)\mathbf{Y}_n\\ &= \frac{1}{2}\mathbf{Y}_n^T\left(\mathcal{K}_2^{(h)^T}\mathcal{K}_2^{(h)} - \mathcal{D}_2^{-1}\right)\mathbf{Y}_n\end{aligned}$$

where $\mathcal{K}_2^{(h)} = \mathcal{D}_2^{-1/2}\mathcal{M}_2^{(h,\sigma)} = \begin{pmatrix}\frac{\mathrm{p}_h}{\alpha} & \frac{\mathrm{e}_h}{\alpha} + \frac{\mathrm{q}_h}{\beta}\\ \frac{\mathrm{e}_h}{\beta} - \frac{\mathrm{q}_h}{\alpha} & \frac{\mathrm{p}_h}{\beta}\end{pmatrix}$.

Let us evaluate

$$\mathcal{E}_2^{(h)} = \mathcal{K}_2^{(h)^T}\mathcal{K}_2^{(h)} - \mathcal{D}_2^{-1} = \begin{pmatrix}\frac{\mathrm{p}_h^2-1}{\alpha^2} + \left(\frac{\mathrm{e}_h}{\beta} - \frac{\mathrm{q}_h}{\alpha}\right)^2 & \left(\frac{1}{\alpha^2} + \frac{1}{\beta^2}\right)\mathrm{e}_h\,\mathrm{p}_h\\ \left(\frac{1}{\alpha^2} + \frac{1}{\beta^2}\right)\mathrm{e}_h\,\mathrm{p}_h & \frac{\mathrm{p}_h^2-1}{\beta^2} + \left(\frac{\mathrm{e}_h}{\alpha} + \frac{\mathrm{q}_h}{\beta}\right)^2\end{pmatrix} \tag{A.4}$$

so that $\Delta_2^{(n,h)} = \frac{1}{2}\mathbf{Y}_n^T\mathcal{E}_2^{(h)}\mathbf{Y}_n$, for $n = 0, \dots, N$ and

$$\Delta_2^{(N)} := H(\mathbf{Y}_N) - H(\mathbf{Y}_0) = \sum_{n=0}^{N}\Delta_2^{(n,h)} = \frac{1}{2}\sum_{n=0}^{N}\mathbf{Y}_n^T\mathcal{E}_2^{(h)}\mathbf{Y}_n. \tag{A.5}$$

Theorem A.1. *Consider the Hamiltonian test problem (3.2) and a symplectic and momentum flip - reversible integrator, expressed as* $\mathbf{Y}_{n+1} = \mathcal{M}_2^{(h,\sigma)} \mathbf{Y}_n$ *with* $\mathcal{M}_2^{(h,\sigma)}$ *defined in (A.2), when applied to (A.1). If it results that* $\mathrm{e}_h = 0$*, then the integrator preserves the Hamiltonian.*

Proof. It is enough to observe that, whenever $\mathrm{e}_h = 0$, the matrix $\mathcal{E}_2^{(h)}$ in (A.4) has null entries on the right-left diagonal. From relation (A.3), it follows that on the principal diagonal $\mathcal{E}_2^{(h)}(1,1) = \dfrac{\mathrm{p}_h^2 + \mathrm{q}_h^2 - 1}{\alpha^2} = \mathcal{E}_2^{(h)}(2,2) = \dfrac{\mathrm{p}_h^2 + \mathrm{q}_h^2 - 1}{\beta^2} = 0$ which completes the proof. □

Theorem A.2. *Assume that* Q_0*,* P_0 *are two random variables with Gaussian zero-mean distribution, standard deviations* α *and* β *respectively and zero correlation. Suppose that the Hamiltonian dynamics (A.1) is approximated by means of a linear map* $\mathbf{Y}_{n+1} = \mathcal{M}_2^{(h,\sigma)} \mathbf{Y}_n$ *with* $\mathcal{M}_2^{(h,\sigma)}$ *given in (A.2). Then, the expectation of the random variable* $\Delta_2^{(N)}$ *in (A.5) is given by*

$$\mathbb{E}(\Delta_2^{(N)}) = \frac{N}{2} \left(\sigma + \frac{1}{\sigma}\right)^2 \mathrm{e}_h^2$$

and, consequently, $\mathbb{E}(\Delta_2^{(N)}) = 0$ *iff* $\mathrm{e}_h = 0$*.*

Proof. From $\Delta_2^{(n,h)} = \dfrac{1}{2} \mathbf{Y}_n^T \mathcal{E}_2^{(h)} \mathbf{Y}_n$, we can evaluate

$$\begin{aligned} 2\,\Delta_2^{(n,h)} \;=\; & \left[\frac{\mathrm{p}_h^2 - 1}{\alpha^2} + \left(\frac{\mathrm{e}_h}{\beta} - \frac{\mathrm{q}_h}{\alpha}\right)^2\right] Q_0^2 + 2\,\mathrm{e}_h\,\mathrm{p}_h \left(\frac{1}{\alpha^2} + \frac{1}{\beta^2}\right) Q_0\,P_0 \\ + \; & \left[\frac{\mathrm{p}_h^2 - 1}{\beta^2} + \left(\frac{\mathrm{e}_h}{\alpha} + \frac{\mathrm{q}_h}{\beta}\right)^2\right] P_0^2 \end{aligned}$$

for $n = 0, \dots, N$. From $\mathbb{E}(Q_0^2) = \alpha^2$, $\mathbb{E}(P_0^2) = \beta^2$, $\mathbb{E}(Q_0\,P_0) = 0$, it results that

$$2\,\mathbb{E}(\Delta_2^{(n,h)}) = 2\left(\mathrm{p}_h^2 - 1\right) + \left(\sigma^{-1}\,\mathrm{e}_h - \mathrm{q}_h\right)^2 + \left(\sigma\,\mathrm{e}_h + \mathrm{q}_h\right)^2 = \left(\sigma\,\mathrm{e}_h + \sigma^{-1}\,\mathrm{e}_h\right)^2$$

and the statement trivially follows. □

The following result generalizes Proposition 4.3 in [10] for Gaussian zero-mean distributions with generic standard deviations α and β:

Theorem A.3. *Under the same hypothesis of Theorem A.1, assuming* $|\mathcal{M}_2^{(h,\sigma)}| < 1$*, the expectation of the random variable* $\Delta_2^{(N)}$ *in (A.5) can be expressed as*

$$\mathbb{E}(\Delta_2^{(N)}) = N \sin^2(h_{\chi_h})\,\rho(h), \qquad \rho(h) = \frac{1}{2}\left(\tilde{\chi}_h - \frac{1}{\tilde{\chi}_h}\right)^2, \qquad \tilde{\chi}_h := \sigma\,\chi_h^{-1}.$$

Proof. Under the assumption $|\mathcal{M}_2^{(h,\sigma)}| < 1$ then $|\mathrm{p}_h| < 1$ and we can define $h_{\chi_h} = \arccos \mathrm{p}_h$, $h_{\chi_h} \in [0, \pi]$ and $\sin h_{\chi_h} = \sqrt{1 - \mathrm{p}_h^2}$. From $\sin^2(h_{\chi_h}) = 1 - \mathrm{p}_h^2$, and exploiting the relations

$$\frac{1}{\chi_h} = \frac{\mathrm{e}_h + \sigma^{-1}\,\mathrm{q}_h}{\sqrt{1 - \mathrm{p}_h^2}}, \qquad \chi_h = \frac{\sigma\,\mathrm{q}_h - \mathrm{e}_h}{\sqrt{1 - \mathrm{p}_h^2}},$$

which are both satisfied from (A.3), we can prove that

$$2\,\sin^2(h_{\chi_h})\,\rho(h) = \sin^2(h_{\chi_h}) \left(\frac{\sigma}{\chi_h} - \frac{\chi_h}{\sigma}\right)^2 = \left(\sigma + \frac{1}{\sigma}\right)^2 \mathrm{e}_h^2.$$

From Theorem A.2 the result follows. □

A.2. Multivariate case

The motion of d oscillators

$$\frac{dQ_j}{dt} = \frac{P_j}{\beta_j^2}, \qquad \frac{dP_j}{dt} = -\frac{Q_j}{\alpha_j^2}, \qquad \text{for} \quad j = 1, \dots d, \tag{A.6}$$

can be represented as an Hamiltonian system

$$\frac{d\mathbf{Q}}{dt} = D_\beta^{-1}\mathbf{P}, \qquad \frac{d\mathbf{P}}{dt} = -D_\alpha^{-1}\mathbf{Q} \tag{A.7}$$

where D_α and D_β are $d \times d$ diagonal matrices with entries α_j^2 and β_j^2, respectively, for $j = 1, \dots, d$ and Hamiltonian function $\frac{1}{2}\mathbf{Q}^T D_\alpha^{-1}\mathbf{Q} + \frac{1}{2}\mathbf{P}^T D_\beta^{-1}\mathbf{P}$. Setting $\mathbf{Y} = [\mathbf{Q}, \mathbf{P}]^T$, $\mathcal{D}_{2d} := \begin{bmatrix} D_\alpha & \mathbf{0}_d \\ \mathbf{0}_d & D_\beta \end{bmatrix}$ the Hamiltonian can be written as

$$H(\mathbf{Y}) = \frac{1}{2}\mathbf{Y}^T \mathcal{D}_{2d}^{-1}\mathbf{Y}. \tag{A.8}$$

Define $\Sigma = D_\beta^{1/2} D_\alpha^{-1/2}$; then a symplectic and momentum flip - reversible integrator for the d-dimensional system (A.7) can be expressed as $\mathbf{Y}_{n+1} = \mathcal{M}_{2d}^{(h,\Sigma)}\mathbf{Y}_n$ where

$$\mathcal{M}_{2d}^{(h,\Sigma)} = \begin{bmatrix} \mathrm{P}_h & \mathrm{E}_h + \Sigma^{-1}\mathrm{Q}_h \\ \mathrm{E}_h - \Sigma\,\mathrm{Q}_h & \mathrm{P}_h \end{bmatrix}, \tag{A.9}$$

where P_h, Q_h and E_h are d dimensional diagonal matrices satisfying

$$\mathrm{P}_h^2 - (\mathrm{E}_h + \Sigma^{-1}\mathrm{Q}_h)(\mathrm{E}_h - \Sigma\,\mathrm{Q}_h) = I_d.$$

Similarly to the univariate case, the error in energy at each step is given by

$$\Delta_{2d}^{(n,h)} = H(\mathbf{Y}_{n+1}) - H(\mathbf{Y}_n) = \frac{1}{2}\mathbf{Y}_n^T \left(\mathcal{K}_{2d}^{(h)\,T}\mathcal{K}_{2d}^{(h)} - \mathcal{D}_{2d}^{-1} \right)\mathbf{Y}_n \tag{A.10}$$

where $\mathcal{K}_{2d}^{(h)} := \mathcal{D}_{2d}^{-1/2}\mathcal{M}_{2d}^{(h,\sigma)} = \begin{pmatrix} D_\alpha^{-1/2}\mathrm{P}_h & D_\alpha^{-1/2}\mathrm{E}_h + D_\beta^{-1/2}\mathrm{Q}_h \\ D_\beta^{-1/2}\mathrm{E}_h - D_\alpha^{-1/2}\mathrm{Q}_h & D_\beta^{-1/2}\mathrm{P}_h \end{pmatrix}$.

When $\mathrm{E}_h = \mathbf{0}_{d\times d}$ then the matrix $\mathcal{E}_{2d}^{(h)} := \mathcal{K}_{2d}^{(h)\,T}\mathcal{K}_{2d}^{(h)} - \mathcal{D}_{2d}^{-1}$ is given by

$$\mathcal{E}_{2d}^{(h)} := \begin{pmatrix} \mathrm{P}_h D_\alpha \mathrm{P}_h + \mathrm{Q}_h D_\alpha \mathrm{Q}_h - D_\alpha & \mathbf{0}_d \\ \mathbf{0}_d & \mathrm{P}_h D_\beta \mathrm{P}_h + \mathrm{Q}_h D_\beta \mathrm{Q}_h - D_\beta \end{pmatrix}. \tag{A.11}$$

Since P_h, Q_h, D_α and D_β are diagonal matrices and $\mathrm{P}_h^2 + \mathrm{Q}_h^2 = \mathrm{I}_d$, then $\mathcal{E}_{2d}^{(h)} = \mathbf{0}_{2d\times 2d}$.

References

[1] Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. Phys Lett B 1987;195(2):216–22.
[2] Neal RM, Neal RM. Monte Carlo implementation. Bayes Learn Neural Netw 1996;55–98.
[3] Neal RM, et al. MCMC using Hamiltonian dynamics. Handb Markov Chain Monte Carlo 2011;2(11):2.
[4] Alder BJ, Wainwright TE. Studies in molecular dynamics. I. General method. J Chem Phys 1959;31(2):459–66.
[5] Sanz-Serna J-M, Calvo M-P. Numerical Hamiltonian problems, vol. 7, Courier Dover Publications; 2018.
[6] Leimkuhler B, Reich S. Simulating Hamiltonian dynamics, no. 14, Cambridge University Press; 2004.
[7] Predescu C, Lippert RA, Eastwood MP, Ierardi D, Xu H, Jensen MØ, Bowers KJ, Gullingsrud J, Rendleman CA, Dror RO, et al. Computationally efficient molecular dynamics integrators with improved sampling accuracy. Mol Phys 2012;110(9–10):967–83.
[8] Takaishi T, De Forcrand P. Testing and tuning symplectic integrators for the hybrid Monte Carlo algorithm in lattice QCD. Phys Rev E 2006;73(3):036706.
[9] Blanes S, Calvo MP, Casas F, Sanz-Serna JM. Symmetrically processed splitting integrators for enhanced Hamiltonian Monte Carlo sampling. SIAM J Sci Comput 2021;43(5):A3357–71.
[10] Blanes S, Casas F, Sanz-Serna J. Numerical integrators for the hybrid Monte Carlo method. SIAM J Sci Comput 2014;36(4):A1556–80.
[11] Pace B, Diele F, Marangi C. Splitting schemes and energy preservation for separable Hamiltonian systems. Math Comput Simulation 2015;110:40–52.
[12] Calvo MP, Sanz-Alonso D, Sanz-Serna J. HMC: Reducing the number of rejections by not using leapfrog and some results on the acceptance rate. J Comput Phys 2021;437:110333.
[13] Nagar L, Fernández-Pendás M, Sanz-Serna JM, Akhmatskaya E. Adaptive multi-stage integration schemes for Hamiltonian Monte Carlo. 2023, arXiv preprint arXiv:2307.02096.
[14] Fernández-Pendás M, Akhmatskaya E, Sanz-Serna J. Adaptive multi-stage integrators for optimal energy conservation in molecular simulations. J Comput Phys 2016;327:434–49.
[15] Baker CM, Blonda P, Casella F, Diele F, Marangi C, Martiradonna A, Montomoli F, Pepper N, Tamborrino C, Tarantino C. Using remote sensing data within an optimal spatiotemporal model for invasive plant management: the case of Ailanthus altissima in the Alta Murgia National Park. Sci Rep 2023;13(1):14587.
[16] Baker CM, Diele F, Lacitignola D, Marangi C, Martiradonna A. Optimal control of invasive species through a dynamical systems approach. Nonlinear Anal Real World Appl 2019;49:45–70.
[17] Baker CM, Diele F, Marangi C, Martiradonna A, Ragni S. Optimal spatiotemporal effort allocation for invasive species removal incorporating a removal handling time and budget. Nat Resour Model 2018;31(4):e12190.
[18] Lacitignola D, Diele F, Marangi C. Dynamical scenarios from a two-patch predator–prey system with human control–Implications for the conservation of the wolf in the Alta Murgia National Park. Ecol Model 2015;316:28–40.
[19] Bou-Rabee N, Sanz-Serna JM. Geometric integrators and the Hamiltonian Monte Carlo method. Acta Numer 2018;27:113–206.
[20] Hairer E, Lubich C, Wanner G, et al. Geometric numerical integration illustrated by the Stormer–Verlet method. Acta Numer 2003;12(12):399–450.
[21] McLachlan RI. On the numerical integration of ordinary differential equations by symmetric composition methods. SIAM J Sci Comput 1995;16(1):151–68.
[22] Shahbaba B, Lan S, Johnson WO, Neal RM. Split Hamiltonian Monte Carlo. Stat Comput 2014;24:339–49.
[23] Casas F, Sanz-Serna JM, Shaw L. Split Hamiltonian Monte Carlo revisited. Stat Comput 2022;32(5):86.
[24] Campos CM, Sanz-Serna J. Palindromic 3-stage splitting integrators, a roadmap. J Comput Phys 2017;346:340–55.
[25] Martino L, Elvira V, Louzada F. Effective sample size for importance sampling based on discrepancy measures. Signal Process 2017;131:386–401.
[26] Mø ller J, Syversveen AR, Waagepetersen RP. Log Gaussian Cox processes. Scand J Stat 1998;25(3):451–82.
[27] Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI. Point process models for presence-only analysis. Methods Ecol Evol 2015;6:366–79.

[28] Christensen O, Roberts G, Rosenthal J. Scaling limits for the transient phase of local Metropolis-Hastings algorithms. J R Statist Soc Ser B 2003;67.
[29] Tarantino C, Casella F, Adamo M, Lucas R, Beierkuhnleind C, Blonda P. Ailanthus altissima mapping from multi-temporal very high resolution satellite images. ISPRS J Photogramm Remote Sens 2019;147:90–103.
[30] Diggle P, Moraga P, Rowlingson B, Taylor B. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. Statist Sci 2013;28.
[31] Thomas S, Tu W. Learning Hamiltonian Monte Carlo in R. Amer Statist 2021;75(4):403–13.
[32] Bache K, Lichman M. UCI machine learning repository. 2013.
[33] Nelder JA, Wedderburn RWM. Generalized linear models. J R Statist Soc Ser A (Gener) 1972;135(3):370–84.