A PROBABILISTIC RUNGE-KUTTA METHOD WITH RANDOM TIME STEPS FOR CHAOTIC AND GEOMETRIC INTEGRATION

ASSYR ABDULLE[*] AND GIACOMO GAREGNANI[†]

Abstract. A novel probabilistic numerical method for quantifying the uncertainty induced by the time integration of ordinary differential equations (ODEs) is introduced. Departing from the classical strategy to randomize ODE solvers by adding a random forcing term, we show that a probability measure over the numerical solution of ODEs can be obtained by introducing suitable random time-steps in a classical time integrator. This intrinsic randomization allows for the conservation of geometric properties of the underlying deterministic integrator such as mass conservation, symplecticity or conservation of first integrals. Weak and mean-square convergence analysis are derived. We also analyze the convergence of the Monte Carlo estimator for the proposed random time step method and show that the measure obtained with repeated sampling converges in mean-square sense independently of the number of samples. Numerical examples including chaotic Hamiltonian systems, chemical reactions and Bayesian inferential problems illustrate the accuracy, robustness and versatility of our probabilistic numerical method.

1. Introduction. A variety of methods for integrating ordinary differential equations (ODEs) has been studied in the last decades, [7, 8, 6], with an emphasis on building accurate and stable deterministic approximations of the exact solution. However, these methods often provide unreliable solutions when applied to chaotic nonlinear systems of ODEs, i.e., equations for which a small perturbation in the initial condition results in a possibly large variation of the exact solution. In this case, even if the initial condition used for numerical integration is exact, or close to exact in floating-point precision, the deviation given by numerical errors can lead to completely wrong solutions of the underlying equation. A classic example of these problems is the Lorenz system [13], a simple model of atmospheric convection, which for certain parameter values presents a typical chaotic behavior, which is nonetheless fully deterministic. Systems resulting in deterministic chaos arise in many practical applications, with the most notable example represented by chaotic chemical reactions. There are furthermore cases of dynamical systems possessing geometric features but nonetheless exhibiting a chaotic behavior. An example of these problems is given by the Hénon-Heiles equation, an Hamiltonian system for which the solution tends to a strange attractor while retaining the property of symplecticity of the flow [9].

In order to overcome this issue, probabilistic numerical integrators of ODEs have recently been proposed [5, 12]. These methods have the purpose of accounting for the numerical error in a statistical manner, rather than with deterministic estimates. In particular, the common underlying idea is to furnish the solution in terms of a probability distribution instead of a punctual value, as in classic solvers for ODEs. One of the most remarkable efforts is presented in Conrad et al. [5], who construct probabilistic solutions with a slight modification of one-step deterministic methods, as, for example, integrators belonging to the Runge-Kutta class. In particular, they propose to perturb the deterministic numerical solution with an additive noise contribution at each time step. Scaling opportunely the noise term with respect to the time step employed for integration, they manage to explore the phase space of chaotic dynamical systems without altering the convergence of the underlying deterministic scheme. An example of this favourable behavior for chaotic system is depicted in Figure 1, where the thin grey lines represent the approximate evolution in time of the first component of the Lorenz system given by the probabilistic method we propose in this work, while the thick black line is given by a classic Runge-Kutta method. The approximate density given by 2000 realisations of the numerical solution at different times are shown together with the time evolution.

The main danger of adding a noise term to deterministic numerical solutions as proposed in [5] is that the random contribution can produce disruptive effects on favourable geometric features of the deterministic scheme. A direct example of this non-robust behavior is given by ODEs for which the solution is supposed to stay positive and small. In this case, the addition of a random contribution

---

[*]Mathematics Section, École Polytechnique Fédérale de Lausanne (assyr.abdulle@epfl.ch)

[†]Mathematics Section, École Polytechnique Fédérale de Lausanne (giacomo.garegnani@epfl.ch)
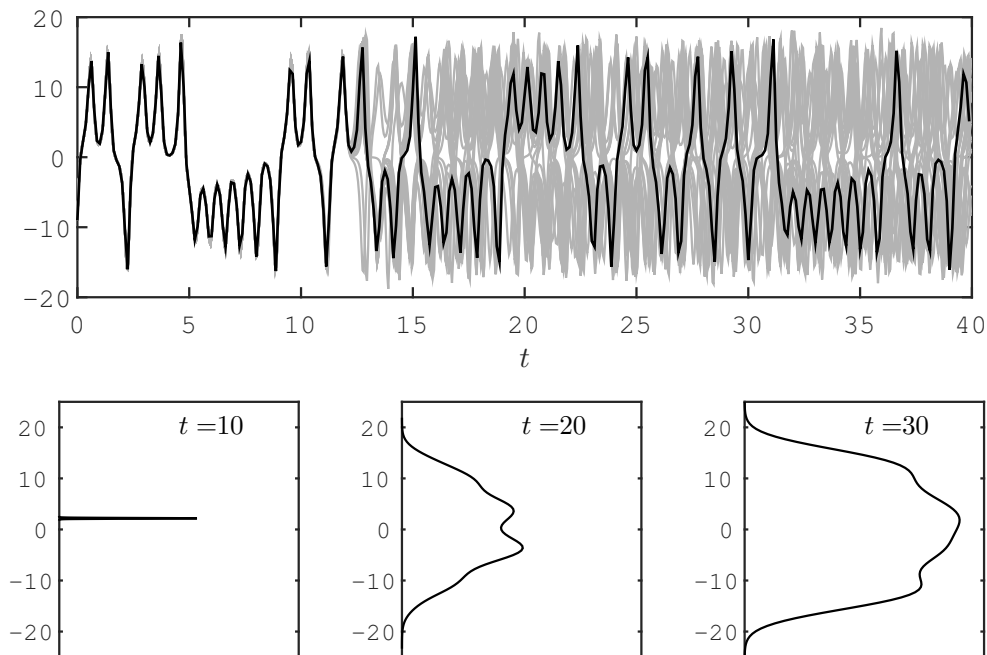
Fig. 1: Time evolution of the first component of the Lorenz system in the chaotic regime and its density estimated with 2000 trajectories of RTS-RK at three different times.

could force the solution in the negative plane, hence the numerical solution could be physically mean-
ingless. Chemical reactions with small population size for one species at some time of the evolution
is a typical physical example of such a situation. Moreover, many of these equations present strong
instabilities in case any of the components turns negative, which makes it even numerically inadvisable
to have negative values. An additive noise contribution yields a non-zero probability to obtain negative
solutions, which can become an issue in case the magnitude of one component is small with respect to
the time step.

Motivated by this undesirable property, we present in this work a new probabilistic method for
ODEs based on a random selection of the time steps, hence turning the stochastic component of the
numerical solution from additive to intrinsic to the scheme itself. In this way, favorable geometric
properties of deterministic integrators are inherited by our probabilistic integrator.

For this new robust probabilistic integrator, we are able to prove strong and weak convergence
towards the exact solution of the underlying ODE. Precisely, setting the variance of the random time
steps to be proportional to some power of a deterministic time step allows to retrieve the rates of the
underlying Runge-Kutta integrator.

It has been pointed out by Kersting and Hennig [12] that the performances of a probabilistic
method strongly depend on the overall quality of a set of values drawn from the numerical solution.
Indeed, the numerical solution given by a probabilistic integrator is a random variable and its numerical
evaluation relies on a sampling strategy, e.g., a Monte Carlo technique. Hence, the relation between
sample size and step size has to be studied in the context of probabilistic integrators. Following the
notation introduced by Kersting and Hennig, the numerical solution is identified by a probability
measure Q_h which converges to the Dirac measure concentrated on the exact solution as $h \to 0$.
However, the quantity we have access to is another measure, denoted by $Q_h(M)$, which is obtained
after M repeated samplings of Q_h. Hence, it is expected that

$$ (1) \qquad\qquad Q_h(M) \overset{M \to \infty}{\Longrightarrow} Q_h \overset{h \to 0}{\Longrightarrow} \delta_y. $$

In this work we show that in a mean-square sense the quality of $Q_h(M)$ is independent of M, and

therefore the computational cost due to repeated sampling can be considered as negligible. We can show a similar property for the additive noise method proposed in [5].

A large variety of dynamical systems is characterised by geometrical properties of their flow map [6]. Most notably, Hamiltonian systems, which are employed for modelling a variety of physical phenomena, are endowed with the property of symplecticity, i.e., the flow of the ODE conserves areas in the state space. While geometric properties of Runge-Kutta schemes have been analysed extensively in the deterministic case, they have not been considered yet for probabilistic numerical methods. It is nonetheless possible to find examples of Hamiltonian systems which present a chaotic behaviour [9], and for which it is therefore relevant to provide a probabilistic quantification of the numerical error. Precisely, in case the considered ODE is Hamiltonian, approximating the trajectory with a geometric integrator yields accurate numerical approximations over long-time. A probabilistic scheme inheriting this favourable property would provide with reliable uncertainty quantification over the same time spans. The method we present in this work, being only an intrinsic modification of a Runge-Kutta integrator, fully preserves all geometric properties. In particular, we show theoretically and with numerical experiments that

- any Runge-Kutta method with random time steps conserves linear first integrals,
- if the Runge-Kutta method with fixed time steps conserves quadratic first integrals, then the same method with random time steps conserves quadratic first integrals,
- if the Runge-Kutta method with fixed time steps is symplectic, then the same method with random time steps is also symplectic.

Moreover, the aforementioned robustness property of positivity of the probabilistic solution can be considered as the simplest geometric property conserved by our numerical method. Interestingly, all the properties above are true for each trajectory of the random time-stepping integrator. We thus say that they are verified in the strong sense. For a method with addive noise as presented in [5], only simplecticity is automatically inherited from the underlying deterministic scheme in the strong sense. In contrast, linear first integrals are only conserved in expectation (or weakly), while quadratic first integrals are never conserved neither in the strong nor in the weak sense.

Probabilistic numerical methods revealed themselves to be particularly effective when applied in the frame of Bayesian inverse problems for parameterised differential equations. In general, any Bayesian inverse problems can be formulated as

$$\text{Find } \pi(u \mid \mathcal{Y}), \text{ with } u \in X, \mathcal{Y} \in Y, \text{ given observations } \mathcal{Y} = \mathcal{G}(u) + \varepsilon,$$

where X and Y are Banach spaces, $\mathcal{G} \colon X \to Y$ is called the forward operator, ε is an additive noise contribution and the object of interest $\pi(u \mid \mathcal{Y})$ is called the posterior distribution. In our setting, u is a parameter driving an ODE, $\mathcal{Y}$ is an observed quantity of interest computable from the solution, and $\mathcal{G}$ is an operator mapping the parameter to the quantity of interest. The action of the operator $\mathcal{G}$ is only available in approximate form when the solution of the ODE does not admit a closed-form expression. It has been shown that employing a deterministic numerical scheme, especially with a coarse discretisation, can lead to overly confident posterior distributions [5, 4]. This issue can be solved by computing the posterior distribution employing a probabilistic approximation of the forward operator. In this work, we show empirically via a numerical experiment that our novel method is efficient in approximating the posterior distribution.

The paper is organised as follows. In section 2 we introduce the setting for probabilistic numerics and present our novel numerical scheme. We then show in sections 3 and 4 the properties of weak and mean square convergence of the numerical solution towards the exact solution of the ODE. In section 5 we analyse the accuracy of Monte Carlo estimators drawn from the numerical solution. The geometric properties of the numerical scheme are presented in section 6, while in section 7 we introduce Bayesian inverse problems in the ODE setting, and show how our method can be integrated in existing sampling strategies. Finally, we show a variety of numerical experiments confirming our theoretical results in section 8.

2. Random time step Runge-Kutta method. Let us consider a Lipschitz function $f \colon \mathbb{R}^d \to \mathbb{R}^d$ and the ODE

$$(2) \qquad\qquad y' = f(y), \quad y(0) = y_0 \in \mathbb{R}^d.$$
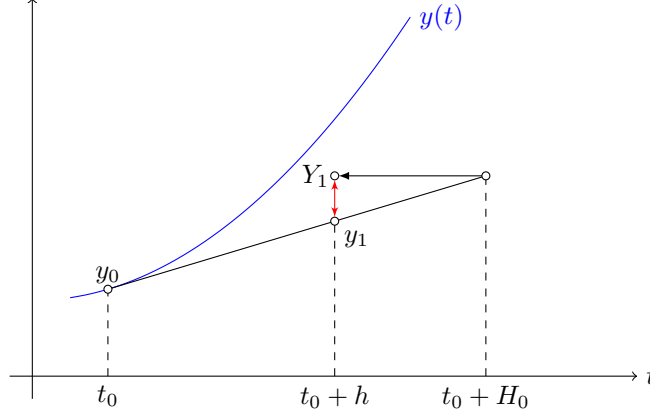
Fig. 2: Graphical representation of one step of the RTS-RK method with $\Psi_h(y) = y + hf(y)$. The red arrow is the stochastic contribution due to random time-stepping.

In the following, we will write for simplicity the solution $y(t)$ of (2) in terms of the flow of the ODE. In particular, we consider the family $\{\varphi_t\}_{t\geq 0}$ of functions $\varphi_t \colon \mathbb{R}^d \to \mathbb{R}^d$ such that

$$(3) \qquad\qquad y(t) = \varphi_t(y_0).$$

Given a time step h, let us consider a Runge-Kutta method which approximates deterministically the solution $\varphi_t(y_0)$ of (2). In particular, we can write the numerical solution y_k approximating $\varphi_{t_k}(y_0)$, with $t_k = kh$ in terms of the numerical flow $\{\Psi_t\}_{t\geq 0}$, with $\Psi_t \colon \mathbb{R}^d \to \mathbb{R}^d$, which is uniquely determined by the coefficients of the method, as

$$(4) \qquad\qquad y_{k+1} = \Psi_h(y_k), \quad k = 0, 1, \ldots.$$

In order to provide a probabilistic interpretation of the numerical solution rather than a series of punctual values, Conrad et al. propose the scheme defined by

$$(5) \qquad\qquad Y_{k+1} = \Psi_h(Y_k) + \xi_k(h), \quad k = 0, 1, \ldots,$$

where Y_k is a random variable approximating $y(t_k)$ with $Y_0 = y_0$, and $\xi_k(h)$ are opportunely scaled independent and identically distributed (i.i.d.) random variables with values in $\mathbb{R}^d$. Maintaining the same notation as in (5), in this work we propose a random time-stepping Runge-Kutta method (RTS-RK), i.e., the scheme defined defined by the recurrence relation

$$(6) \qquad\qquad Y_{k+1} = \Psi_{H_k}(Y_k), \quad k = 0, 1, \ldots,$$

where Y_k is still a random variable approximating $y(t_k)$ and the time steps H_k are locally given by a sequence of i.i.d. random variables with values in $\mathbb{R}^+$. A graphical representation of one step of the RTS-RK method is given in Figure 2.

Remark 2.1. In terms of computational cost simulating the two methods (6) and (5) are almost equivalent, as they imply the same number of function evaluations as the underlying deterministic solver Ψ_h. Nonetheless, the random time-stepping method has the slight advantage that the random variable that has to be drawn at each time step takes values in $\mathbb{R}_+$, while for the additive noise ξ_k takes values in $\mathbb{R}^d$.

2.1. Assumptions and notations. We now present the main assumptions and notations which are used throughout the rest of this work. Firstly, we have to consider the possible values taken by the random step sizes, which have to satisfy restrictions that are necessary not to spoil the properties of deterministic methods.

Assumption 2.2. The i.i.d. random variables H_k satisfy for all $k = 0, 1, \ldots$
(i) $H_k > 0$ a.s.,

(ii) there exists $h > 0$ such that $\mathbb{E}\, H_k = h$,

(iii) there exists $p \geq 1$ such that the scaled random variables $Z_k := H_k - h$ satisfy

$$(7) \qquad \qquad \mathbb{E}\, Z_k^2 = Ch^{2p},$$

which is equivalent to $\mathbb{E}\, H_k^2 = h^2 + Ch^{2p}$.

The class of random variable satisfying the hypotheses above is general. However, it is practical for an implementation point of view to have examples of these variables.

Example 2.3. Let us consider the random variables $\{H_k\}_{k \geq 0}$ such that

$$(8) \qquad \qquad H_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}(h - h^p, h + h^p), \quad 0 < h < 1, \quad p \geq 1.$$

We easily verify that all the assumptions (i), (ii) and (iii) are satisfied, as

(i) $H_k > 0$ a.s. trivially since $h < 1$,

(ii) $\mathbb{E}\, H_k = h$ since

$$(9) \qquad \qquad \mathbb{E}\, H_k = \frac{1}{2}(h + h^p + h - h^p) = h.$$

(iii) The random variables $Z_k = H_k - h$ are $Z_k \sim \mathcal{U}(-h^p, h^p)$. Therefore

$$(10) \qquad \qquad \mathbb{E}\, Z_k^2 = \frac{4h^{2p}}{12} = \frac{1}{3}h^{2p}.$$

Another choice of random variables could simply be

$$(11) \qquad \qquad H_k \overset{\text{i.i.d.}}{\sim} \log \mathcal{N}\left(\log h - \log \sqrt{1 + h^{2p-2}}, \log(1 + h^{2p-2})\right),$$

for which the properties above are trivially verified (with $C = 1$), provided $p > 1$.

We secondly introduce an assumption on the deterministic method underlying the RTS-RK scheme, identified by its numerical flow Ψ_h.

Assumption 2.4. The Runge-Kutta method defined by the numerical flow $\{\Psi_t\}_{t \geq 0}$ satisfies the following properties.

(i) For h small enough, there exists a constant $C > 0$ such that

$$(12) \qquad \qquad \|\Psi_h(y) - \varphi_h(y)\| \leq Ch^{q+1}, \quad \forall y \in \mathbb{R}^d,$$

i.e., the deterministic solver has order q.

(ii) The map $t \mapsto \Psi_t(y)$ is of class $C^2(\mathbb{R}^+, \mathbb{R}^d)$ and Lipschitz continuous of constant L_Ψ, i.e.,

$$(13) \qquad \qquad \|\Psi_t(y) - \Psi_s(y)\| \leq L_\Psi |t - s|, \quad \forall t, s > 0.$$

3. Weak convergence analysis. The first property of the RTS-RK method we wish to analyze is its weak convergence, which gives an indication about the behavior of the numerical solution (6) in the mean sense. Let us define the weak order of convergence.

DEFINITION 3.1. *The numerical method* (6) *has weak order r for* (2) *if for any function $\Phi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with all derivatives bounded uniformly on $\mathbb{R}^d$ there exists a constant $C > 0$ independent of h such that*

$$(14) \qquad \qquad |\mathbb{E}\, \Phi(Y_k) - \Phi(y(kh))| \leq Ch^r,$$

for all $k = 1, 2, \ldots, N$.

Let us introduce the Lie derivative of the flow $\mathcal{L} = f \cdot \nabla$, which allows us to adopt the semi-group notation for the exact solution of (2) and write for any smooth function Φ

$$(15) \qquad \Phi(\varphi_h(y)) = e^{h\mathcal{L}}\Phi(y).$$

Moreover, let us remark that the probabilistic numerical solution $\{Y_k\}_{k\geq 0}$ forms an homogeneous Markov chain, and hence there exists an operator $\mathcal{P}_t$, the infinitesimal generator, such that

$$(16) \qquad \mathbb{E}\left(\Phi(Y_k) \mid Y_0 = y\right) = (\mathcal{P}_{t_k}\Phi)(y).$$

The family of operators $\{\mathcal{P}_t\}_{t\geq 0}$ forms a Markov semi-group together with the associativity operation

$$(17) \qquad (\mathcal{P}_{t+s}\Phi)(y) = \mathcal{P}_t(\mathcal{P}_s\Phi)(y).$$

In order to have an analogy with the notation (15), we adopt the exponential form of the infinitesimal generator and denote in the following $\mathcal{P}_t = e^{t\mathcal{L}^h}$, where we explicitly write the dependence of the Markov generator on the step size h. We can now state a result of local weak convergence of the probabilistic numerical solution.

LEMMA 3.2 (Weak local order). *Let Assumptions 2.2 and 2.4 hold. If $\mathbb{E}\,|H_0^4| < \infty$, there exists a constant $C > 0$ independent of h such that for any function $\Phi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with all derivatives bounded uniformly on $\mathbb{R}^d$*

$$(18) \qquad |\mathbb{E}(\Phi(Y_1) \mid Y_0 = y) - \Phi(\varphi_h(y))| \leq Ch^{\min\{2p, q+1\}}.$$

Proof. Let us expand the functional Φ computed on the numerical solution as

$$
(19) \quad
\begin{aligned}
\Phi(Y_1) &= \Phi(\Psi_{H_0}(Y_0)) \\
&= \Phi\left(\Psi_h(Y_0) + (H_0 - h)\partial_t\Psi_h(Y_0) + \frac{1}{2}(H_0 - h)^2\partial_{tt}\Psi_h(Y_0) + \mathcal{O}(|H_0 - h|^3)\right) \\
&= \Phi(\Psi_h(Y_0)) + \left((H_0 - h)\partial_t\Psi_h(Y_0) + \frac{1}{2}(H_0 - h)^2\partial_{tt}\Psi_h(Y_0)\right) \cdot \nabla\Phi(\Psi_h(Y_0)) \\
&\quad + \frac{1}{2}(H_0 - h)^2\partial_t\Psi_h(Y_0)\partial_t\Psi_h(Y_0)^\top : \nabla^2\Phi(\Psi_h(Y_0)) + \mathcal{O}(|H_0 - h|^3),
\end{aligned}
$$

where we denote by $\nabla^2\Phi$ the Hessian matrix of Φ, and by $:$ the inner product on matrices induced by the Frobenius norm on $\mathbb{R}^d$, i.e., $A : B = \mathrm{tr}(A^\top B)$. Taking the conditional expectation with respect to $Y_0 = y$ and applying Assumption 2.2 we get

$$
(20) \quad
\begin{aligned}
e^{h\mathcal{L}^h}\Phi(y) - \Phi(\Psi_h(y)) &= \frac{1}{2}Ch^{2p}\partial_{tt}\Psi_h(y) \cdot \nabla\Phi(\Psi_h(y)) \\
&\quad + \frac{1}{2}Ch^{2p}\partial_t\Psi_h(y)\partial_t\Psi_h(y)^\top : \nabla^2\Phi(\Psi_h(y)) + \mathcal{O}(h^{3p}),
\end{aligned}
$$

where we exploited Hölder inequality for the last term. Moreover, expanding Φ in y we get

$$
(21) \quad
\begin{aligned}
\Phi(\Psi_h(y)) &= \Phi\left(\Psi_0(y) + h\partial_t\Psi_0(y) + \mathcal{O}(h^2)\right) \\
&= \Phi(y) + \mathcal{O}(h).
\end{aligned}
$$

which implies

$$
(22) \quad
\begin{aligned}
e^{h\mathcal{L}^h}\Phi(y) - \Phi(\Psi_h(y)) &= \frac{1}{2}Ch^{2p}\partial_{tt}\Psi_h(y) \cdot \nabla\Phi(y) \\
&\quad + \frac{1}{2}Ch^{2p}\partial_t\Psi_h(y)\partial_t\Psi_h(y)^\top : \nabla^2\Phi(y) + \mathcal{O}(h^{2p+1}).
\end{aligned}
$$

Let us remark that thanks to Assumption 2.4.(i) we have

$$(23) \qquad e^{h\mathcal{L}}\Phi(y) - \Phi(\Psi_h(y)) = \mathcal{O}(h^{q+1}).$$

Combining (23) and (22) we have the one-step weak error of the probabilistic method on the original ODE, i.e.,

$$e^{h\mathcal{L}}\Phi(y) - e^{h\mathcal{L}^h}\Phi(y) = \mathcal{O}(h^{\min\{2p, q+1\}}), \tag{24}$$

which proves the desired result. □

In order to obtain a result on the global order of convergence we need a further stability assumption, which is the same as Assumption 3 in [5].

Assumption 3.3. The function f is such that the operator $e^{h\mathcal{L}^h}$ satisfies for all functions $\Phi \in \mathbb{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with all derivatives uniformly bounded in $\mathbb{R}^d$ and a positive constant L

$$\sup_{u \in \mathbb{R}^d} |e^{h\mathcal{L}^h}\Phi(u)| \leq (1 + Lh) \sup_{u \in \mathbb{R}^d} |\Phi(u)|. \tag{25}$$

Remark 3.4. Given the assumptions on f and Φ above, it is true for the exact solution that

$$\sup_{u \in \mathbb{R}^d} |e^{h\mathcal{L}}\Phi(u)| \leq (1 + Lh) \sup_{u \in \mathbb{R}^d} |\Phi(u)|. \tag{26}$$

We can now state the main result on weak convergence. Let us remark that the theorem and its proof are similar to Theorem 2.4 in [5].

THEOREM 3.5 (Weak order). *Let the assumptions of Lemma 3.2 and Assumption 3.3 hold. Then, there exists a constant $C > 0$ independent of h such that for all functions $\Phi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with all derivatives bounded in $\mathbb{R}^d$*

$$|\mathbb{E}\,\Phi(Y_k) - \Phi(y(kh)))| \leq Ch^{\min\{2p-1, q\}}, \tag{27}$$

for all $k = 1, 2, \ldots, N$.

Proof. Let us introduce the following notation

$$\begin{aligned} w_k(u) &= \Phi(\varphi_{t_k}(u)), \\ W_k(u) &= \mathbb{E}(\Phi(Y_k) \mid Y_0 = u). \end{aligned} \tag{28}$$

By the triangular inequality and the associativity property of $\exp(t\mathcal{L}^h)$, we have
$$\sup_{u \in \mathbb{R}^d} |W_k(u) - w_k(u)| \leq \sup_{u \in \mathbb{R}^d} |e^{h\mathcal{L}}w_{k-1}(u) - e^{h\mathcal{L}^h}w_{k-1}(u)| + \sup_{u \in \mathbb{R}^d} |e^{h\mathcal{L}^h}w_{k-1}(u) - e^{h\mathcal{L}^h}W_{k-1}(u)|. \tag{29}$$

We then apply Lemma (3.2) to the first term and Assumption 3.3 to the second, thus obtaining

$$\sup_{u \in \mathbb{R}^d} |W_k(u) - w_k(u)| \leq Ch^{\min\{2p, q+1\}} + (1 + Lh) \sup_{u \in \mathbb{R}^d} |w_{k-1}(u) - W_{k-1}(u)|. \tag{30}$$

Proceeding iteratively on the index k and noticing that $w_0 = W_0$, we obtain

$$\begin{aligned} \sup_{u \in \mathbb{R}^d} |w_k(u) - W_k(u)| &\leq Ckh^{\min\{2p, q+1\}} \\ &\leq CTh^{\min\{2p-1, q\}}, \end{aligned} \tag{31}$$

which proves the result for any chosen initial condition y_0 in $\mathbb{R}^d$, as

$$|\mathbb{E}\,\Phi(Y_k) - \Phi(y(kh)))| \leq \sup_{u \in \mathbb{R}^d} |W_k(u) - w_k(u)|. \tag{32}$$

□

Remark 3.6. In [5], Conrad et al. define ordinary and stochastic modified equations in order to prove a result of weak convergence applying techniques of backward error analysis. In particular, they show that their probabilistic solver approximates in the weak sense a stochastic differential equation (SDE) where the deterministic part is given by the original ODE. For our probabilistic solver, it is possible to prove that the numerical solutions approximates in the weak sense the solution of an SDE where the diffusion term depends on the derivative of the map $t \mapsto \Psi_t(y)$.

4. Mean square convergence analysis. The second property of (6) we analyze is its mean square order of convergence, which gives an indication on the path-wise distance between each realisation of the numerical solution and the exact solution of (2). Let us define the mean square order of convergence.

DEFINITION 4.1. *The numerical method* (6) *has mean square order of convergence r for* (2) *if there exists a constant $C > 0$ independent of h such that*

$$(33) \qquad \left(\mathbb{E}\|Y_k - y(kh)\|^2 \right)^{1/2} \leq Ch^r$$

for all $k = 1, 2, \ldots, N$.

Remark 4.2. Let us remark that if a numerical method has mean square order of convergence r, it will have the same strong order of convergence, i.e.,

$$(34) \qquad \mathbb{E}\|Y_k - y(kh)\| \leq Ch^r.$$

This comes trivially from Jensen's inequality, as

$$(35) \qquad \mathbb{E}\|Y_k - y(kh)\| \leq \left(\mathbb{E}\|Y_k - y(kh)\|^2 \right)^{1/2}.$$

We start by analysing how the method converges with respect to the mean step size h in the local sense, i.e., after one step of the numerical integration.

LEMMA 4.3 (Local mean square convergence). *Under Assumptions 2.2 and 2.4 the numerical solution Y_1 given by one step of the RTS-RK method* (6) *satisfies*

$$(36) \qquad \left(\mathbb{E}\|Y_1 - y(h)\|^2 \right)^{1/2} \leq Ch^{\min\{p, q+1\}},$$

where C is a real positive constant independent of h and the coefficients p, q are given in the assumptions.

Proof. By triangular and Young's inequalities we have for all $y \in \mathbb{R}^d$

$$(37) \qquad \mathbb{E}\|\Psi_{H_0}(y) - \varphi_h(y)\|^2 \leq 2\,\mathbb{E}\|\Psi_{H_0}(y) - \Psi_h(y)\|^2 + 2\|\Psi_h(y) - \varphi_h(y)\|^2.$$

We now consider Assumption 2.4.(ii) and 2.4.(i), thus getting

$$(38) \qquad \mathbb{E}\|\Psi_{H_0}(y) - \varphi_h(y)\|^2 \leq 2L_\Psi^2\,\mathbb{E}|H_0 - h|^2 + 2C_1 h^{2(q+1)}.$$

We now apply Assumption 2.2 to obtain

$$(39) \qquad \begin{aligned} \mathbb{E}\|\Psi_{H_0}(y) - \varphi_h(y)\|^2 &= 2L_\Psi^2 C_2 h^{2p} + 2C_1 h^{2(q+1)} \\ &\leq C^2 h^{2\min\{q+1, p\}}, \end{aligned}$$

which is the desired result with $C = \max\{2L_\Psi^2 C_2, 2C_1\}^{1/2}$. □

As a consequence of the one-step convergence, we can prove a result of global mean square convergence.

THEOREM 4.4 (Global mean square convergence). *Let $t_k = kh$ for $k = 1, 2, \ldots, N$, where $Nh = T$. Then, under the assumptions of Lemma 4.3 the numerical solution given by* (6) *satisfies*

$$(40) \qquad \sup_{k=1,2,\ldots,N} \left(\mathbb{E}\|Y_k - y(t_k)\|^2 \right)^{1/2} \leq Ch^{\min\{p-1/2, q\}},$$

where C is a real positive constant independent of h.

In order to prove this result, let us introduce the following standard lemmas.

LEMMA 4.5. *Given the ODE* (2) *with* f *globally Lipschitz, then for any* y *and* w *in* $\mathbb{R}^d$ *and* $h < 1$ *we have*

$$\|\varphi_h(y) - \varphi_h(w)\| \leq (1 + Ch)\|y - w\|, \tag{41}$$
$$\|\varphi_h(y) - \varphi_h(w) - (y - w)\| \leq Ch\|y - w\|, \tag{42}$$

where C *is a positive constant independent of* h.

Proof. We have trivially

$$\varphi_h(y) - \varphi_h(w) = y - w + \int_0^h \big(f(\varphi_t(y)) - f(\varphi_t(w))\big)\mathrm{d}t. \tag{43}$$

Taking the norm on both sides, applying the triangular inequality and since f is Lipschitz continuous with constant L, we have

$$\|\varphi_h(y) - \varphi_h(w)\| \leq \|y - w\| + L\int_0^h \|\varphi_t(y) - \varphi_t(w)\|\mathrm{d}t. \tag{44}$$

Applying Gronwall's inequality and assuming $h < 1$ we have

$$\|\varphi_h(y) - \varphi_h(w)\| \leq e^{Lh}\|y - w\| \leq (1 + \hat{C}h)\|y - w\|, \tag{45}$$

where $\hat{C} = e^L - 1$. Hence, defining $Z := \varphi_h(y) - \varphi_h(w) - (y - w)$, we have

$$\begin{aligned}
\|Z\| &\leq L\int_0^h \|\varphi_t(y) - \varphi_t(w)\|\mathrm{d}t \\
&\leq hL(1 + \hat{C}h)\|y - w\|.
\end{aligned} \tag{46}$$

which proves the result with $C = L(1 + \hat{C})$. □

LEMMA 4.6. *Suppose that for arbitrary* N *and* $k = 0, \ldots, N$ *we have*

$$u_k \leq (1 + Ah)u_{k-1} + Bh^r, \tag{47}$$

where $h = T/N$, $A > 0$, $B \geq 0$, $r \geq 1$ *and* $u_k \geq 0, k = 0, \ldots, N$. *Then*

$$u_k \leq e^{AT}u_0 + \frac{B}{A}(e^{AT} - 1)h^{r-1}. \tag{48}$$

The proof of this lemma is based on a recurrence argument and is omitted in this work. We can now prove the main result on mean square convergence.

Proof of Theorem 4.4. In the following, we denote by C a generic positive constant. Let us define $e_k^2 := \mathbb{E}\|Y_k - y(t_k)\|^2$. Adding and subtracting the exact flow applied to the numerical solution, we obtain

$$\begin{aligned}
e_{k+1}^2 &= \mathbb{E}\|\Psi_{H_k}(Y_k) - \varphi_h(Y_k)\|^2 + \mathbb{E}\|\varphi_h(Y_k) - \varphi_h(y(t_k))\|^2 \\
&\quad + 2\mathbb{E}\Big(\varphi_h(Y_k) - \varphi_h(y(t_k)), \Psi_{H_k}(Y_k) - \varphi_h(Y_k)\Big),
\end{aligned} \tag{49}$$

where we denote by $(\cdot, \cdot)$ the inner product in $\mathbb{R}^d$. Let us consider the three terms in (49) separately. For the first term, we have by Lemma 4.3

$$\mathbb{E}\|\Psi_{H_k}(Y_k) - \varphi_h(Y_k)\|^2 \leq Ch^{\min\{2p, 2(q+1)\}}. \tag{50}$$

For the second term, since φ_h is Lipschitz with constant $(1 + Ch)$, we have

$$\mathbb{E}\|\varphi_h(Y_k) - \varphi_h(y(t_k))\|^2 \leq (1 + Ch)^2 e_k^2. \tag{51}$$

Let us now define $Z = \varphi_h(Y_k) - \varphi_h(y(t_k)) - (Y_k - y(t_k))$. Then we, can rewrite the inner product as

$$
(52) \quad \mathbb{E}\left(\varphi_h(Y_k) - \varphi_h(y(t_k)), \Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) = \mathbb{E}\left(Y_k - y(t_k), \Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right)
$$
$$
+ \mathbb{E}\left(Z, \Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right).
$$

We bound the two terms in (52) separately. For the first term, by the law of total expectation, we have

$$
(53) \quad \mathbb{E}\left(\left(Y_k - y(t_k) \right)^\top \left(\Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) \right) = \mathbb{E}\,\mathbb{E}\left(\left(Y_k - y(t_k) \right)^\top \left(\Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) \mid Y_k \right)
$$
$$
= \mathbb{E}\left(\left(Y_k - y(t_k) \right)^\top \mathbb{E}\left(\Psi_{H_k}(Y_k) - \varphi_h(Y_k) \mid Y_k \right) \right).
$$

Applying Cauchy-Schwarz inequality to the outer expectation we get

$$
(54) \quad \mathbb{E}\left(\left(Y_k - y(t_k) \right)^\top \left(\Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) \right) \le \left(\mathbb{E}\| \mathbb{E}\left(\Psi_{H_k}(Y_k) - \varphi_h(Y_k) \mid Y_k \right) \|^2 \right)^{1/2} e_k
$$
$$
\le C h^{\min\{2p,q+1\}} e_k,
$$

where we applied Lemma 3.2. We now consider the second term in (52). By Cauchy-Schwarz inequality we have

$$
(55) \quad \mathbb{E}\left(Z, \Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) \le \left(\mathbb{E}\|Z\|^2 \right)^{1/2} \left(\mathbb{E}\|\Psi_{H_k}(Y_k) - \varphi_h(Y_k)\|^2 \right)^{1/2}.
$$

We now apply Lemma 4.5 and Lemma 4.3 to obtain

$$
(56) \quad \mathbb{E}\left(Z, \Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) \le C h^{\min\{p+1,q+2\}} e_k.
$$

We can hence bound the scalar product in (52) with Young's inequality and assuming $h < 1$ as

$$
(57) \quad \mathbb{E}\left(\varphi_h(Y_k) - \varphi_h(y(t_k)), \Psi_{H_k}(Y_k) - \varphi_h(Y_k) \right) \le C(h^{\min\{p+1,q+2\}}) e_k
$$
$$
\le \frac{h e_k^2}{2} + C \frac{h^{\min\{2p+1,2q+1\}}}{2}.
$$

Combining (50), (51) and (57), we have

$$
(58) \quad e_{k+1}^2 \le C h^{\min\{2p,2q+1\}} + (1 + Ch) e_k^2,
$$

which implies the desired result by Lemma 4.6 and since $e_0 = 0$.   □

Remark 4.7. Let us remark that the difference between global and local orders of convergence is not exactly one, as it usually is in the purely deterministic case. In fact, thanks to the independence of the random variables there is only a $1/2$ loss in the random part of the exponent, while the natural loss of one order is verified in the deterministic component.

Remark 4.8. The result of mean square convergence suggests that a reasonable choice for the noise scale p is to fix it to $q + 1/2$, where q is the order of the Runge-Kutta method Ψ_h. In this way, the properties of convergence of the underlying deterministic method are not spoiled, nonetheless getting a probabilistic interpretation of the numerical solution.

5. Monte Carlo estimators. The third property we analyze is the mean-square convergence of Monte Carlo estimators drawn from the random time-stepping Runge-Kutta method. Let us consider a function $\Phi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with Lipschitz constant L_Φ and a final time $T > 0$. In general, the quantity $\Phi(Y_k)$ is not accessible, and we have to replace it by its Monte Carlo estimator

$$
(59) \quad \hat{Z} = M^{-1} \sum_{i=1}^M \Phi(Y_N^{(i)}),
$$

where $T = hN$ is the final time, M is the number of trajectories and we denote by $\{Y_N^{(i)}\}_{i=1}^M$ a set of realizations of the numerical solution. Hence, we are interested in studying the mean square error of the Monte Carlo estimator, which is defined as

$$\text{MSE}(\hat{Z}) = \mathbb{E}(Z - \hat{Z})^2, \tag{60}$$

where $Z = \Phi(y(T))$. In the following result, we prove that this quantity converges to zero independently of the number of trajectories M.

THEOREM 5.1. *Under Assumptions 2.2 and 2.4, the Monte Carlo estimator $\hat{Z}$ satisfies*

$$\text{MSE}(\hat{Z}) \leq C\Big(h^{2\min\{2p-1,q\}} + \frac{h^{2\min\{p-1/2,q\}}}{M}\Big), \tag{61}$$

where C is a positive constant independent of h and M.

Proof. Thanks to the classic decomposition of the MSE, we have

$$\text{MSE}(\hat{Z}) = \text{Var}\,\hat{Z} + \big(\mathbb{E}(\hat{Z} - Z)\big)^2. \tag{62}$$

Applying Theorem 3.5 to the second term, we have

$$\text{MSE}(\hat{Z}) \leq \text{Var}\,\hat{Z} + Ch^{2\min\{2p-1,q\}}. \tag{63}$$

The variance of the estimator can be trivially bounded exploiting the Lipschitz continuity of Φ and the independence of the samples by

$$\text{Var}\,\hat{Z} \leq M^{-1}L_\Phi^2\,\mathbb{E}\|Y_N - y(T)\|^2. \tag{64}$$

Applying Theorem 4.4 we get

$$\text{Var}\,\hat{Z} \leq M^{-1}L_\Phi^2 Ch^{2\min\{q,p-1/2\}}, \tag{65}$$

which proves the desired result. □

Let us remark that with the choice $p = q + 1/2$, which is the minimum p for which the order of convergence of the underlying deterministic method is not affected by the probabilistic setting, we have $\text{MSE}(\hat{Z}) \leq Ch^{2q}$ with $M = 1$.

Remark 5.2. This result is of critical importance as it implies that the Monte Carlo estimators drawn from (6) converge in the mean square sense independently of the number of samples M in (59). Moreover, if we impose that the two terms in (61) are balanced, we get that the number of trajectories M has to satisfy

$$M = \begin{cases} \mathcal{O}(1) & \text{if } q \leq p - 1/2, \\ \mathcal{O}\big(h^{-(2q-(2p-1))}\big) & \text{if } p - 1/2 < q \leq 2p - 1, \\ \mathcal{O}\big(h^{-(2p-1)}\big) & \text{if } q > 2p - 1. \end{cases} \tag{66}$$

Fixing M as above, we have that the MSE satisfies

$$\text{MSE}(\hat{Z}) = \begin{cases} \mathcal{O}\big(h^{2q}\big) & \text{if } q \leq 2p - 1, \\ \mathcal{O}\big(h^{2(2p-1)}\big) & \text{if } q > 2p - 1. \end{cases} \tag{67}$$

Let us introduce as a measure of error the square root of the MSE and impose a fixed tolerance ε, i.e.,

$$\text{MSE}(\hat{Z})^{1/2} = \mathcal{O}(\varepsilon), \tag{68}$$

we have that the step size h has to satisfy

$$h = \begin{cases} \mathcal{O}\big(\varepsilon^{1/q}\big) & \text{if } q \leq 2p - 1, \\ \mathcal{O}\big(\varepsilon^{1/(2p-1)}\big) & \text{if } q > 2p - 1. \end{cases} \tag{69}$$

Therefore, the computational cost to attain such a tolerance ε is given by

$$\text{cost} = \mathcal{O}(Mh^{-1}) = \begin{cases} \mathcal{O}(\varepsilon^{-1/q}) & \text{if } q \leq p - 1/2, \\ \mathcal{O}\big(\varepsilon^{-2+2\frac{p-1}{q}}\big) & \text{if } p - 1/2 < q \leq 2p - 1, \\ \mathcal{O}\big(\varepsilon^{-\frac{2p}{2p-1}}\big) & \text{if } q > 2p - 1. \end{cases} \tag{70}$$

6. Geometric properties. Numerical methods for ODEs are often studied in terms of their geometric properties [6]. In particular, we investigate here whether the random choice of time steps in (6) spoils the properties of the underlying deterministic Runge-Kutta method. Let us recall the definition of first integral for an ODE.

DEFINITION 6.1. *Given a function $I\colon \mathbb{R}^d \to \mathbb{R}$, then $I(y)$ is a first integral of (2) if $I'(y)f(y) = 0$ for all $y \in \mathbb{R}^d$.*

In particular, if there exists $v \in \mathbb{R}^d$ such that $I(y) = v^\top y$, we say that I is a linear first integral of (2). Moreover, if there exists a symmetric matrix $S \in \mathbb{R}^{d \times d}$ such that $I(y) = y^\top S y$, then we say that I is a quadratic first integral of (2). If this property of the ODE is conserved by a numerical integrator, i.e., if for the any $y \in \mathbb{R}^d$ it is true that $I(\Psi_h(y)) = I(y)$, then we say that the numerical method conserves the first integral. In particular, this implies that the invariant is conserved along the trajectory of the numerical solution, i.e., $I(y_k) = I(y_0)$ for all $k \geq 0$.

Example 6.2. To illustrate this concept we first discuss the case of linear first integrals, which can be seen as a general case of the conservation of mass in physical systems. Let us consider a linear first integral $I(y) = v^\top y$ and any Runge-Kutta method with coefficients $\{b_i\}_{i=1}^s$, $\{a_{ij}\}_{i,j=1}^s$. Then, we have for a time step $H_0 > 0$

$$(71) \qquad I(Y_1) = v^\top y_0 + H_0 \textstyle\sum_{i=1}^s b_i v^\top f(y_0 + H_0 \sum_{j=1}^s a_{ij} K_j),$$

where $\{K_i\}_{i=1}^s$ are the internal stages of the Runge-Kutta method. Since $I(y)$ is a first integral, $v^\top f(y) = 0$ for any $y \in \mathbb{R}^d$. Hence $I(Y_1) = I(y_0)$ and iteratively $I(Y_k) = I(y_0)$ for all $k \geq 0$ along the numerical trajectory. The equality above shows that any RTS-RK method conserves linear first integrals path-wise, or in the strong sense.

It is known that no Runge-Kutta method can conserve any polynomial invariant of order $n \geq 3$ [6]. Nonetheless, for some particular problems there exist tailored Runge-Kutta methods which can conserve polynomial invariants of higher order. We therefore can state the following general result.

THEOREM 6.3. *If the Runge-Kutta scheme defined by Ψ_h conserves an invariant $I(y)$ for an ODE, then the numerical method (6) conserves $I(y)$ for the same ODE.*

Proof. If $I(\Psi_h(y)) = I(y)$ for any h, then $I(\Psi_{H_0}(y)) = I(y)$ for any value that H_0 can assume. □

We now consider quadratic first integrals, i.e., first integrals of the form $I(y) = y^\top C y$ with C a symmetric matrix, which are conserved by Runge-Kutta methods that satisfy the hypotheses of Cooper's theorem [6]. The conservation of quadratic first invariants is of the utmost importance, e.g., for Hamiltonian systems, as it implies the symplecticity of the scheme. It is known [6] that all Gauss methods conserve quadratic first integrals. The simplest member of this class of methods is the implicit midpoint rule, which is a one-stage method defined by coefficients $b_1 = 1$ and $a_{11} = 1/2$.

COROLLARY 6.4. *If the Runge-Kutta scheme defined by Ψ_h conserves quadratic first integrals then the numerical method (6) conserves quadratic first integrals.*

This result is a direct consequence of Theorem 6.3.

6.1. The additive noise method. The properties above for the RTS-RK method are not satisfied by the additive noise method presented in [5]. In particular, let us remark that the conservation of first integrals is exact for any trajectory of the RTS-RK method, and is not an average property. In other words, we can say that (6) conserves linear first integrals in the strong sense. For the additive noise numerical method (5), we have

$$(72) \qquad \begin{aligned} I(Y_1) &= v^\top y_0 + h\textstyle\sum_{i=1}^s b_i v^\top f(y_0 + h\sum_{j=1}^s a_{ij} K_j) + v^\top \xi_0(h), \\ &= v^\top (y_0 + \xi_0(h)). \end{aligned}$$

If the random variable ξ_0 is zero-mean, then $\mathbb{E}\,I(Y_1) = I(y_0)$ and iteratively along the solution $\mathbb{E}\,I(Y_{k+1}) = \mathbb{E}\,I(Y_k) = I(y_0)$. The linear first invariants are therefore conserved in average, but not in a path-wise fashion.

For quadratic first integrals, we have instead that the additive noise method does not conserve them neither path-wise nor in the weak sense, as we have

$$
(73) \qquad
\begin{aligned}
I(Y_1) &= (\Psi_h(y_0) + \xi_0(h))^\top S(\Psi_h(y_0) + \xi_0(h)) \\
&= I(y_0) + 2\xi_0(h)^\top S\Psi_h(y_0) + \xi_0(h)^\top S\xi_0(h).
\end{aligned}
$$

If the random variables are zero-mean and if there exists a matrix Q such that $\mathbb{E}\,\xi_0(h)\xi_0(h)^\top = Qh^{2p+1}$ for some $p \geq 1$ (Assumption 1 in [5]) we then have

$$
(74) \qquad \mathbb{E}\,I(Y_1) = I(y_0) + Q : Sh^{2p+1}.
$$

Hence, along the trajectories of the solution a bias is introduced in the first integral which persists even in the mean sense. In general, Theorem 6.3 is not valid for the additive noise method, as the random contribution drives the first integral far from its true value at each time step. This could produce in practice large deviations of the numerical approximation from the true solution, especially in the long time regime.

6.2. Hamiltonian systems. A class of dynamical systems of particular interest for their geometric properties are the Hamiltonian systems. Given a function $E(p, q)$, called the Hamiltonian, with p and q in $\mathbb{R}^d$, they can be written as

$$
(75) \qquad y' = J^{-1}\nabla E(y), \quad y(0) = y_0,
$$

where $y = (p, q)^\top$ and the matrix $J \in \mathbb{R}^{2d \times 2d}$ is defined as

$$
(76) \qquad J = \begin{pmatrix} \mathbf{0} & I \\ -I & \mathbf{0} \end{pmatrix},
$$

and where I and $\mathbf{0}$ are respectively the identity and the zero matrices in $\mathbb{R}^{d \times d}$. It is well-known that the flow $\varphi_t \colon \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ of any system of the form (75) is symplectic, i.e.,

$$
(77) \qquad \partial_y \varphi_t(y)^\top J \partial_y \varphi_t(y) = J, \quad \forall y \in \mathbb{R}^d,
$$

where we denote by $y \in \mathbb{R}^{2d}$ the vector $(p, q)^\top$. Condition (77) is equivalent in a geometric sense to saying that the flow φ_t of the system conserves areas in the phase space. In a natural manner, the flow Ψ_h of a one-step numerical method for (75) is called symplectic it satisfies

$$
(78) \qquad \partial_y \Psi_h(y)^\top J \partial_y \Psi_h(y) = J.
$$

It has been pointed out [16, 6] that applying an adaptive step size technique to a symplectic method can destroy its simplecticity. Therefore, Skeel and Gear [16] write any adaptive technique in function of a map $\tau(y, h)$ such that the k-th time step h_k is selected as $h_k = \tau(y_k, h)$, where h is a base value for the time step. Hence, in order to have again a symplectic method, the new condition to be satisfied is

$$
(79) \qquad V^\top J V = J, \quad V = \partial_y \Psi_{\tau(y,h)}(y) + \partial_t \Psi_{\tau(y,h)}(y)\partial_y \tau(y, h)^\top.
$$

In view of this condition, we can state a result of symplecticity for our numerical scheme.

THEOREM 6.5. *If the flow Ψ_h of the deterministic integrator is symplectic, then the flow of the random time-stepping probabilistic method (6) is symplectic.*

Proof. We can consider the random mapping such that on the nodes $\tau(y, h) = h\Theta_k$ where $\{\Theta_k\}_{k \geq 0}$ is a sequence of i.i.d. random variables such that $\Theta_k > 0$ a.s., $\mathbb{E}\,\Theta_k = 1$ and $\mathrm{Var}\,\Theta_k = Ch^{2p-2}$ for any k. In this way, at each time step we have $\tau(y, h) = \tau(h) = H_k$, which are random time steps satisfying Assumption 2.2. Then the probabilistic method (6) can be written as $Y_{k+1} = \Psi_{h\Theta_k}(Y_k)$. Since the mapping $\tau(y, h)$ is independent of y, we have

$$
(80) \qquad V := \partial_y \Psi_{h\Theta_k}(y) + h\partial_h \Psi_{h\Theta_k}(y)\partial_y \vartheta(y, h)^\top = \partial_y \Psi_{h\Theta_k}(y).
$$

Hence, we apply the symplecticity of the deterministic method and get

$$
(81) \qquad V^\top J V = J,
$$

which is the desired result. $\qquad\qquad\square$

Remark 6.6. Let us remark that the symplecticity of a deterministic numerical method $y_1 = \Psi_h(y)$ is a consequence of the conservation of quadratic first integrals. It is though interesting to note that the additive noise method (5), which does not conserve quadratic first integrals nevertheless can be proved to be symplectic if the underlying deterministic method is symplectic. In fact, if the flow of the deterministic integrator is $\Psi_h(y)$, the flow of the probabilistic method is given by $\tilde{\Psi}_h(y) = \Psi_h(y) + \xi_k(h)$. Since $\xi_k(h)$ is independent of y, we have $\partial_y \tilde{\Psi}_h(y) = \partial_y \Psi_h(y)$.

7. Bayesian inference inverse problems. It has been recently shown [5] that probabilistic methods for ordinary and partial differential equations guarantee good results in the context of Bayesian inverse problems. In this section, we briefly introduce a Bayesian inverse problem in the ODE setting and how the RTS-RK method can be employed in this framework.

Let us consider a function $f_\vartheta \colon \mathbb{R}^d \to \mathbb{R}^d$ which depends on a real parameter $\vartheta \in \Theta$ where Θ is an open subset of $\mathbb{R}^n$ and the ODE

$$(82) \qquad\qquad y'_\vartheta = f_\vartheta(y), \quad y_\vartheta(0) = y_0 \in \mathbb{R}^d.$$

In order to simplify the notation, we consider y_0 to be a fixed initial condition. In general, y_0 could depend itself on ϑ. In the classical setting of numerical analysis, the main problem of interest could be summarised as

$$(83) \qquad \text{given the parameter } \vartheta \in \Theta, \text{ determine (or approximate) the solution } y \in \mathbb{R}^d.$$

In the following, we will refer to (83) as the forward problem. Inverse problems can be in general stated as

$$(84) \quad \text{given information on the solution } y \in \mathbb{R}^d, \text{ determine (or approximate) the parameter } \vartheta \in \Theta.$$

In the Bayesian setting, the aim is not to determine exactly the parameter ϑ or to give a punctual approximation to its value, but to establish a probability measure, the posterior measure, given observations and prior knowledge available on ϑ.

It is not always possible to observe the solution $y_\vartheta(t)$ continuously in time, but we nonetheless consider a quantity derived from $y_\vartheta(t)$ to be accessible. We denote the observable quantity by $\mathcal{Y}$ and we consider it to be a vector of $\mathbb{R}^m$ with $m \in \mathbb{N}$. We summarise the action of the forward problem with a function $\mathcal{G} \colon \Theta \to \mathbb{R}^m$ directly mapping the parameter to the observations. Moreover, in real applications measurements can not be exact, hence we consider observations to be always corrupted by an additive source of noise. In particular, we model noise as a Gaussian random variable $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon \in \mathbb{R}^{m \times m}$ is the covariance matrix, independent of ϑ and taking values in $\mathbb{R}^m$. Given these assumptions, we can write observations as

$$(85) \qquad\qquad \mathcal{Y} = \mathcal{G}(\vartheta) + \varepsilon.$$

Given a new value of ϑ, it is possible to compute the function $\mathcal{G}(\vartheta)$, i.e., solve the forward problem, and obtain the likelihood of the observations as

$$(86) \qquad\qquad \pi(\mathcal{Y} \mid \vartheta) = e^{-V_{\mathcal{Y}}(\vartheta)},$$

where the potential function $V_{\mathcal{Y}} \colon \Theta \to \mathbb{R}$ is defined in the Gaussian case as

$$(87) \qquad\qquad V_{\mathcal{Y}}(\vartheta) = \frac{1}{2} \big(\mathcal{G}(\vartheta) - \mathcal{Y}\big)^\top \Sigma_\varepsilon^{-1} \big(\mathcal{G}(\vartheta) - \mathcal{Y}\big).$$

The second building block of Bayesian inverse problems is the prior distribution, which we denote by $\pi_0(\vartheta)$. The prior encodes all the knowledge on the parameter that is known before observations are provided. In the following, we adopt a common abuse of notation, confounding measures and their probability density function.

Once the likelihood model and the prior distribution are established, it is possible to compute the posterior distribution $\pi(\vartheta \mid \mathcal{Y})$ via Bayes' theorem, i.e.,

$$(88) \qquad\qquad \pi(\vartheta \mid \mathcal{Y}) = \frac{1}{Z(\mathcal{Y})} \pi(\mathcal{Y} \mid \vartheta) \pi_0(\vartheta),$$

where $Z(\mathcal{Y})$ is a normalising constant independent of ϑ given by

$$(89) \qquad Z(\mathcal{Y}) = \int_{\Theta} \pi(\mathcal{Y} \mid \vartheta) \pi_0(\vartheta) \, \mathrm{d}\vartheta.$$

In case the forward model $\mathcal{G}$ involves the solution of an ODE, the likelihood function will not be computed exactly. In particular, we denote by $\mathcal{G}^h(\vartheta)$ the result of the forward model computed with a Runge-Kutta method with time step h, and consequently with $V_{\mathcal{Y}}^h(\vartheta)$ and $\pi^h(\vartheta \mid \mathcal{Y})$ the potential and the likelihood function obtained replacing $\mathcal{G}(\vartheta)$ with $\mathcal{G}^h(\vartheta)$. We can then define the approximated posterior distribution $\pi^h(\vartheta \mid \mathcal{Y})$ as

$$(90) \qquad \pi^h(\vartheta \mid \mathcal{Y}) = \frac{1}{Z^h(\mathcal{Y})} \pi^h(\mathcal{Y} \mid \vartheta) \pi_0(\vartheta),$$

where the normalising constant $Z^h(\mathcal{Y})$ is given by

$$(91) \qquad Z^h(\mathcal{Y}) = \int_{\Theta} \pi^h(\mathcal{Y} \mid \vartheta) \pi_0(\vartheta) \, \mathrm{d}\vartheta.$$

In [17, Theorem 4.6], Stuart proves that the posterior distribution $\pi^h(\vartheta \mid \mathcal{Y})$ converges to $\pi(\vartheta \mid \mathcal{Y})$ with respect to h with the same rate as $V_{\mathcal{Y}}^h(\vartheta)$ converges to $V_{\mathcal{Y}}(\vartheta)$. Convergences is proved in the Hellinger distance, which is defined for probability density functions as

$$(92) \qquad d_{\mathrm{Hell}}\big(\pi^h(\vartheta \mid \mathcal{Y}), \pi(\vartheta \mid \mathcal{Y})\big)^2 = \frac{1}{2} \int_{\Theta} \left(\sqrt{\pi^h(\vartheta \mid \mathcal{Y})} - \sqrt{\pi(\vartheta \mid \mathcal{Y})} \right)^2 \mathrm{d}\vartheta.$$

Hence, when there is no restriction in computational resources and it is possible to choose h small, the approximated posterior distribution can be pushed arbitrarily close to the true posterior.

In this work we consider the case when h is fixed, and in particular when discretization provides with a coarse approximation of $\mathcal{G}(\vartheta)$. More precisely, we are interested in the case where the numerical error dominates the noise contribution. It has been shown empirically [5, 4] that in this small noise limit the approximated posterior distributions can be overly confident on the value of the parameter. In particular, the expectation of ϑ computed under the posterior distribution presents a bias with respect to the true value, which is not highlighted by higher order moments. This undesirable phenomenon can be corrected employing a probabilistic method, as the one presented by Conrad et al. in [5] or the RTS-RK method, to approximate the potential $V_{\mathcal{Y}}(\vartheta)$. Let us denote by $\xi \in \mathcal{X}$ the auxiliary random variable introduced by the probabilistic method. In the case of RTS-RK, we have $\xi = (H_0, H_1, H_1, \ldots, H_{N-1})^\top$ and $\mathcal{X} \subset \mathbb{R}_+^N$. The likelihood function, denoted as $\pi_{\mathrm{pr}}^h(\mathcal{Y} \mid \vartheta)$ is then approximated as

$$(93) \qquad \pi_{\mathrm{pr}}^h(\mathcal{Y} \mid \vartheta) = \int_{\mathcal{X}} e^{-V_{\mathcal{Y}}^{h,\xi}(\vartheta)} \, \mu(\mathrm{d}\xi),$$

where $V_{\mathcal{Y}}^{h,\xi}$ is approximation of the potential function given by the probabilistic method and μ is the probability measure induced by the auxiliary variables. The corresponding posterior distribution π_{pr}^h is then obtained applying Bayes' rule. Modifying the posterior in this manner allows to obtain qualitatively better results, which account for the uncertainty introduced by the numerical solver.

7.1. Sampling strategies. In the discussion above we presented how to theoretically construct posterior distributions on the parameter ϑ from observations. Since the posterior could be complex and high-dimensional, it is necessary in practice to rely on a Monte Carlo method in order to obtain a sample $\{\vartheta^{(i)}\}_{i=1}^M$ from this distribution. In order to fulfil this purpose, we choose the Markov chain Monte Carlo method (MCMC), and in particular the Metropolis Hastings method (MH) [11]. The main idea of MCMC methods is to build a Markov chain $\{\vartheta^{(i)}\}$ that serves as a sample from the posterior distribution.

The MH algorithm forms the Markov chain through a Markov transition function $q \colon \Theta \times \Theta \to \mathbb{R}$, which is usually referred to as the proposal distribution, which is measurable in its first argument, i.e., for any $B \subset \Theta$ the function

$$(94) \qquad \int_B q(x, y) \mathrm{d}y$$

is measurable in x, and is for any fixed $x \in \Theta$ a probability density function in its second argument, i.e., $q(x, y) \geq 0$ for any $y \in \Theta$ and

$$(95) \qquad \int_{\Theta} q(x, y) \mathrm{d}y = 1.$$

A common choice for q is a Gaussian distribution in the second argument, i.e., for any $x \in \Theta$ the function $q(x, \cdot)$ is the probability density function of a random variable distributed as $\mathcal{N}(x, \Sigma_q)$. In this case, the MH is referred to in literature as random walk Metropolis (RWM), and its performances strongly depend on the choice of the variance Σ_q. Assuming that it is possible to evaluate the posterior distribution and given an initial guess $\vartheta^{(i)}$, the MH algorithm proceeds as follows

  (i) compute $\pi(\vartheta^{(1)} \mid \mathcal{Y})$ and set $i = 1$;
  (ii) propose a new value ϑ^* sampling from $q(\vartheta^{(i)}, \cdot)$;
  (iii) with probability

$$(96) \qquad \alpha(\vartheta^{(i)}, \vartheta^*) = \min\left\{ \frac{\pi(\vartheta^* \mid \mathcal{Y})q(\vartheta^*, \vartheta^{(i)})}{\pi(\vartheta^{(i)} \mid \mathcal{Y})q(\vartheta^{(i)}, \vartheta^*)}, 1 \right\},$$

  set $\vartheta^{(i+1)} = \vartheta^*$, otherwise set $\vartheta^{(i+1)} = \vartheta^{(i)}$;
  (iv) if $i < M$, set $i \leftarrow i + 1$ and return to step (ii).

It is possible to show that the invariant distribution of the Markov chain $\{\vartheta^{(i)}\}_{i \geq 0}$ generated by the MH algorithm is in fact the posterior distribution $\pi(\vartheta \mid \mathcal{Y})$.

In the probabilistic case, it is not possible to compute exactly the integral in (93), and hence the algorithm above has to be slightly modified. In particular, we employ the pseudo-marginal MH (PMMH) method proposed by Andrieu and Roberts [3], which allows to incorporate auxiliary variables when generating the sample. It is possible to show that if the posterior is replaced by an unbiased estimator of its value which is then used to compute the acceptance probability (96), the PMMH targets exactly the distribution $\pi_{\mathrm{pr}}(\vartheta \mid \mathcal{Y})$. In order to guarantee this property of exactness, it is nonetheless necessary that the estimator from the previous value is recycled in the next iteration. We note that such a procedure can trigger a sticky behaviour of the Markov chain. A remedy is to use the Monte Carlo within Metropolis method (MCWM), where the estimator of the posterior for the previous value is recomputed. This algorithm does not present a sticky behaviour but it loses the property of exactness of PMMH [14]. The PMMH algorithm can be schematically summarised as follows

  (i) compute an unbiased estimator $\hat{\pi}^{(1)}$ of $\pi_{\mathrm{pr}}(\vartheta^{(1)} \mid \mathcal{Y})$ and set $i = 1$;
  (ii) propose a new value ϑ^* sampling from $q(\vartheta^{(i)}, \cdot)$;
  (iii) compute an unbiased estimator $\hat{\pi}^*$ of $\pi_{\mathrm{pr}}(\vartheta^* \mid \mathcal{Y})$;
  (iv) with probability

$$(97) \qquad \alpha(\vartheta^{(i)}, \vartheta^*) = \min\left\{ \frac{\hat{\pi}^* q(\vartheta^*, \vartheta^{(i)})}{\hat{\pi}^{(i)} q(\vartheta^{(i)}, \vartheta^*)}, 1 \right\},$$

  set $\vartheta^{(i+1)} = \vartheta^*$ and $\hat{\pi}^{(i+1)} = \hat{\pi}^*$, otherwise set $\vartheta^{(i+1)} = \vartheta^{(i)}$ and $\hat{\pi}^{(i+1)} = \hat{\pi}^{(i)}$;
  (v) if $i < M$, set $i \leftarrow i + 1$ and return to step (ii).

Let us finally remark that the estimator of the posterior can be computed with a plain Monte Carlo method, which guarantees unbiasedness and hence the exactness of the PMMH method.

8. Numerical experiments. In this section, we present a series of numerical experiments that illustrate the versatility and usefulness of our new random time stepping method. These experiments also corroborate the theoretical results presented in the previous sections.

Method	ET					RK4				
q			2					4		
p	1	1.5	2	2.5	3	3	3.5	4	4.5	5
$\min\{q, p-1/2\}$	0.5	1	1.5	2	2	2.5	3	3.5	4	4
strong order	0.51	1.02	1.54	2.01	2.01	2.50	3.01	3.56	4.02	4.01

Table 1: Mean square order of convergence for the random time-stepping explicit trapezoidal (ET) and fourth-order Runge-Kutta (RK4) as a function of the value of p of Assumption 2.2.

Method	ET			RK4				
q		2				4		
p	1	1.5	2	1	1.5	2	3	4
$\min\{q, 2p-1\}$	1	2	2	1	2	3	4	4
weak order	0.98	2.06	2.12	0.90	1.96	3.01	3.97	4.08

Table 2: Weak order of convergence for the random time-stepping explicit trapezoidal (ET) and fourth-order Runge-Kutta (RK4) as a function of the value of p of Assumption 2.2.

8.1. Mean square order of convergence. In order to verify the result predicted in Theorem 4.4, we consider the FitzHug-Nagumo equation, which is defined as

$$
(98) \qquad
\begin{aligned}
y_1' &= c\left(y_1 - \frac{y_1^3}{3} + y_2\right), & y_1(0) &= -1, \\
y_2' &= -\frac{1}{c}(y_1 - a + by_2), & y_2(0) &= 1,
\end{aligned}
$$

where a, b, c are real parameters with values $a = 0.2$, $b = 0.2$, $c = 3$. We integrate the equation from time $t_0 = 0$ to final time $T = 1$. The reference solution is generated with an high order method on a fine time scale. We consider as deterministic solvers the explicit trapezoidal rule and the classic fourth order Runge-Kutta method, which verify Assumption 2.4 with $q = 2$ and $q = 4$ respectively. Moreover, we consider random time steps as in Example 2.3, where we vary p in order to verify the order of convergence predicted in Theorem 4.4. We vary the mean time step h taken by the random time steps H_n in the range $h_i = 0.01 \cdot 2^{-i}$, with $i = 0, 1, \ldots, 4$. Then, we simulate 10^3 realizations of the numerical solution Y_{N_i}, with $N_i = T/h_i$ for $i = 0, 1, \ldots, 4$, and compute the approximate mean square order of convergence for each value of h with a Monte Carlo mean. Results (Table 1) show that the orders predicted theoretically by Proposition 4.4 are confirmed numerically.

8.2. Weak order of convergence. We now verify the weak order of convergence predicted in Theorem 3.5. For this experiment we consider the ODE (98) as well, with the same time scale and parameters as above. The reference solution at final time is generated in this case as well with an high-order method on a fine time scale. The deterministic integrators we choose in this experiment are the explicit trapezoidal rule and the classic fourth-order Runge-Kutta method. The mean time step varies in the range $h_i = 0.1 \cdot 2^{-i}$ with $i = 0, 1, \ldots, 5$, and we vary the value of p in Assumption 2.2 in order to verify the theoretical result of Theorem 3.5. The function $\Phi \colon \mathbb{R}^d \to \mathbb{R}$ of the solution we consider is defined as $\varphi(x) = x^\top x$. Finally, we consider 10^6 trajectories of the numerical solution in order to approximate the expectation with a Monte Carlo sum. Results (Table 2) show that the order of convergence predicted theoretically is confirmed by numerical experiments.

8.3. Monte Carlo estimator. We shall now verify numerically the validity of Theorem 5.1. We consider the ODE (98), with final time $T = 1$ and the same parameters as above. In this case as well, we consider the explicit trapezoidal rule and the fourth-order explicit Runge-Kutta method with random time steps having mean $h_i = 0.25 \cdot 2^{-i}$ with $i = 0, 1, \ldots, 8$. For the explicit trapezoidal rule, we fix $M = 10^3$ and $p = 1.5$, so that for bigger values of h the first term in the bound presented in Theorem 5.1 dominates, while in the regime of small h, the higher order of the first term makes the second term larger in magnitude. This behaviour results in the change of slope in the convergence plot
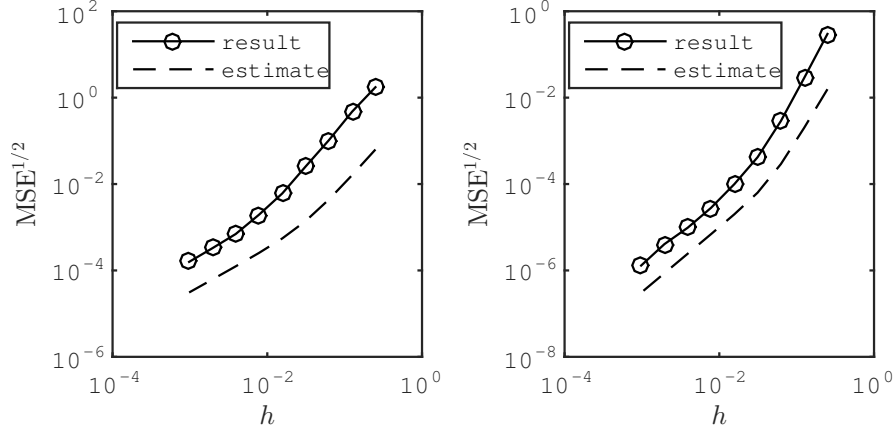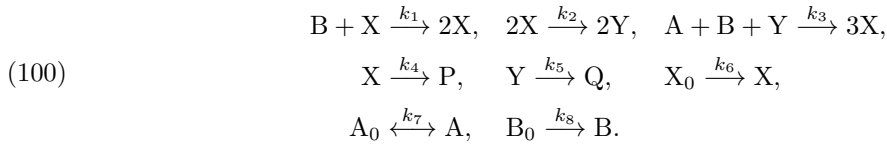
Fig. 3: Convergence of the MSE of the Monte Carlo estimator for the random time-stepping explicit trapezoidal (ET) (left figure) and fourth-order Runge-Kutta (RK4) (right figure). The dashed line corresponds to the order predicted in Theorem 5.1 with $M = 10^3$ for ET and $M = 10^4$ for RK4.

which can be observed in Figure 3, both in the theoretical estimate and in the numerical results. We perform the same experiment using the fourth-order explicit Runge-Kutta method, fixing $M = 10^4$ and $p = 2$, thus obtaining a numerical confirmation of the theoretical result.

8.4. Robustness. In this numerical experiment we verify the robustness of RTS-RK when applied to chemical reactions. Let us consider the Peroxide-Oxide chemical reaction, which is macroscopically defined by the following balance equation

$$(99) \qquad\qquad O_2 + 2NADH + 2H^+ \rightarrow 2H_2O + 2NAD^+.$$
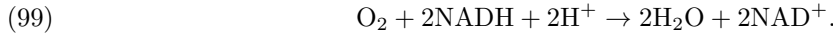
This reaction has to be catalyzed by an enzyme to take place, which reacts with the reagents to create intermediate products of the reaction. A successful model [15] to describe the time-evolution of the chemical system is the following

$$
\begin{aligned}
(100) \qquad & B + X \xrightarrow{k_1} 2X, \quad 2X \xrightarrow{k_2} 2Y, \quad A + B + Y \xrightarrow{k_3} 3X, \\
& X \xrightarrow{k_4} P, \quad\; Y \xrightarrow{k_5} Q, \quad\;\; X_0 \xrightarrow{k_6} X, \\
& A_0 \xleftrightarrow{k_7} A, \quad B_0 \xrightarrow{k_8} B.
\end{aligned}
$$

Here, A and B are respectively $[O_2]$ and $[NADH]$, P, Q are the products and X, Y are intermediates results of the reaction process. It is therefore possible to model the time evolution of the reaction with the following system of nonlinear ODEs

$$
\begin{aligned}
(101) \qquad
& A' = k_7(A_0 - A) - k_3 ABY, & A(0) &= 6, \\
& B' = k_8 B_0 - k_1 BX - k_3 ABY, & B(0) &= 58, \\
& X' = k_1 BX - 2k_2 X^2 + 3k_3 ABY - k_4 X + k_6 X_0, & X(0) &= 0, \\
& Y' = 2k_2 X^2 - k_5 Y - k_3 ABY, & Y(0) &= 0,
\end{aligned}
$$

where $A_0 = 8$, $B_0 = 1$, $X_0 = 1$ and the real parameters k_i, $i = 1, \dots, 8$ representing the reaction rates take values

$$
\begin{aligned}
(102) \qquad
& k_1 = 0.35, \quad k_2 = 250, \quad\; k_3 = 0.035, \quad k_4 = 20, \\
& k_5 = 5.35, \quad k_6 = 10^{-5}, \quad k_7 = 0.1, \quad\;\; k_8 = 0.825.
\end{aligned}
$$

It has been shown [15] that for these values of the parameters the system exhibits a chaotic behavior. In particular, at long time the trajectories are captured in a strange attractor, and the system shows a strong sensitivity to perturbations on the initial condition.
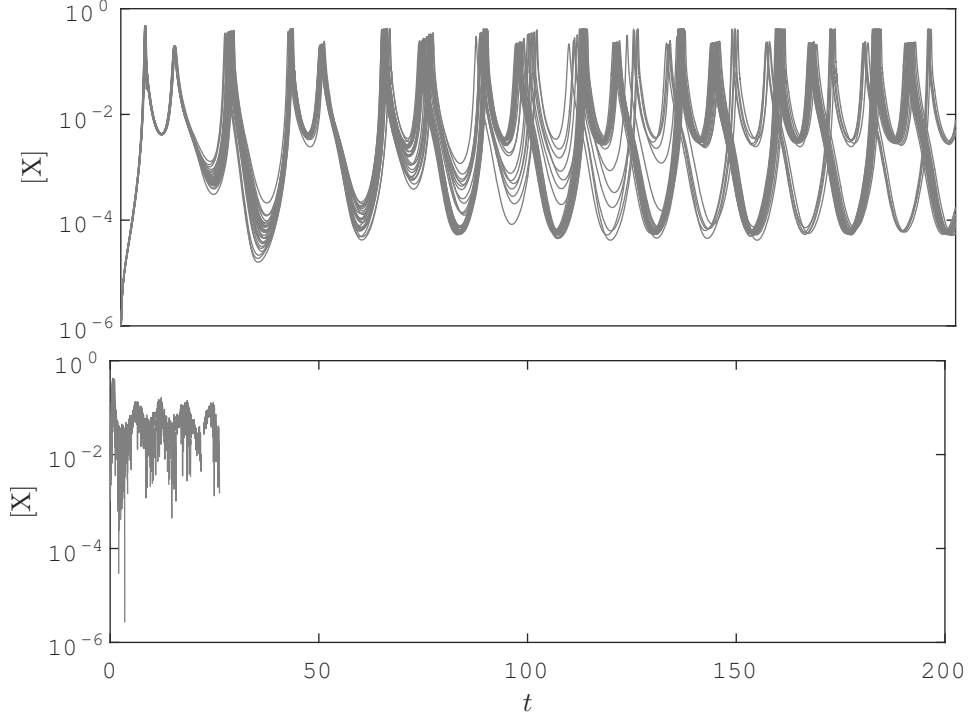
Fig. 4: Thirty trajectories of the numerical value of the concentration of the X species for the random time-stepping and additive noise methods (above and below respectively).

Since the components of the solution represent the concentration of chemicals, we require the numerical solution to be positive. Apart from physical considerations, numerically we observe that if one of the components takes negative values, the solution shows strong instabilities. Let us consider the maximum time step $h_{\max}$ for which the deterministic numerical solution stays positive and thus is stable. The probability that the RTS-RK method presents instabilities before time $T = hN$ depends on the distribution of the variables H_k. In some situations this probability can be computed explicitly. For example, if they are drawn from a uniform distribution as in Example 2.3, we can just set $h + h^p < h_{\max}$ thus having a null probability of being unstable. In contrast, for the additive noise method we can have disruptive effects even with small time steps if the solution has a small magnitude. To illustrate this issue, let us consider a one-dimensional problem solved by the additive noise method (5) with $\xi_k(h)$ i.i.d. random variables $\mathcal{N}(0, h^{2p+1})$. With this choice of random variables, even in case the underlying deterministic method with time step h would give a positive value, the probability of the probabilistic numerical solution to turn negative is always greater than zero. In particular, if the magnitude of the numerical solution Y_k is close to zero then the probability of the event $(Y_{k+1} < 0)$ is significant. Hence, integrating over long time an equation whose components should stay positive with a small magnitude employing the additive noise method likely produces instabilities regardless of the chosen time step.

Let us apply the additive noise method (5) and the random time-stepping scheme (6) to equation (101). We choose $h = 0.05$ as the mean time steps for (6) and as the time step for (5), while we employ the Runge-Kutta-Chebyshev method (RKC) [18] as deterministic integrator. The RKC method is a stabilized numerical integrator of first order, which allows to obtain stable solutions with an explicit scheme. Higher order explicit stabilized methods such as ROCK2 or ROCK4 [2, 1] could also be used as deterministic solvers for the RTS-RK method. It can be seen in Figure 4 that the RTS-RK method conserves the positivity of the numerical solution while capturing the chaotic nature of the chemical reaction, while the additive noise scheme produces negative values, thus showing strong instabilities in the long-time behavior.
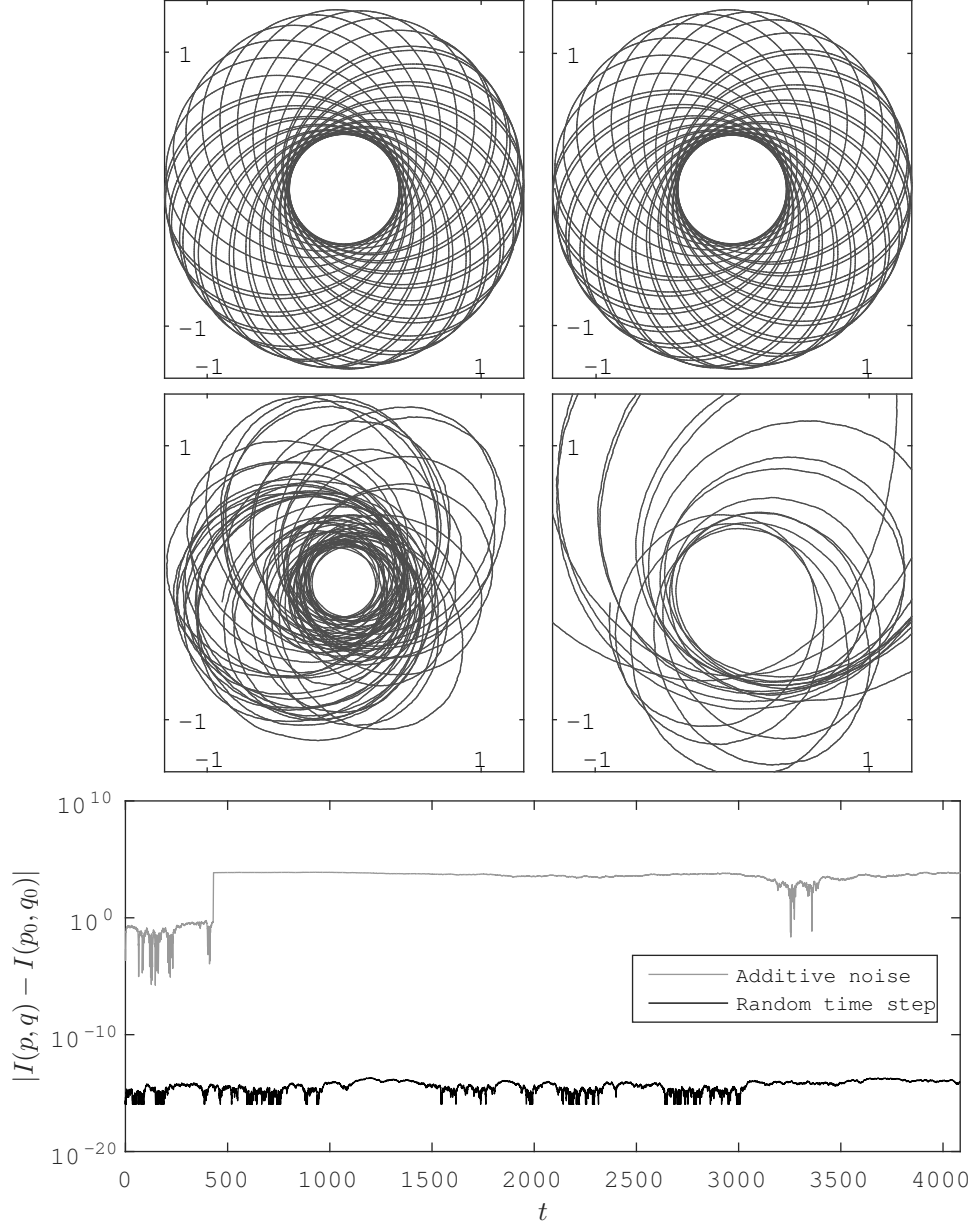
Fig. 5: Trajectories of (103) given by the RTS-RK method (6) for $0 \leq t \leq 200$ and $3800 \leq t \leq 4000$ (first row, left and right), and by the additive noise method (5) for $0 \leq t \leq 200$ and $200 \leq t \leq 400$ (second row, left and right). Error on the angular momentum for $0 \leq t \leq 4000$ given by the two methods.

8.5. Conservation of quadratic first integrals. A simple model for the two-body problem in celestial mechanics is the Kepler system with a perturbation, , which reads

$$
\begin{aligned}
(103) \quad & q_1' = p_1, \quad p_1' = -\frac{q_1}{\|q\|^3} - \frac{\delta q_1}{\|q\|^5}, \\
& q_2' = p_2, \quad p_2' = -\frac{q_2}{\|q\|^3} - \frac{\delta q_2}{\|q\|^5},
\end{aligned}
$$

where p_1, p_2 are the two components of the velocity and q_1, q_2 are the two components of the position. We assume the perturbation parameter δ to be equal to 0.015 and the initial condition to be

$$(104) \qquad q_1(0) = 1 - e, \quad q_2(0) = 0, \quad p_1(0) = 0, \quad p_2(0) = \sqrt{(1+e)/(1-e)},$$

where $e = 0.6$ is the eccentricity. It is well-known that this equation has the Hamiltonian and the angular momentum as quadratic first integrals. In particular, we focus here on the angular momentum, which reads

$$(105) \qquad I(p,q) = q_1 p_2 - q_2 p_1.$$

We consider the simplest Gauss collocation method, namely the implicit midpoint rule, as the deterministic Runge-Kutta method. It is known that Gauss collocation methods conserve quadratic first integrals. According to Theorem 6.3, we expect therefore that the random time-stepping method (6) implemented with Ψ_h given by the implicit midpoint rule conserves also quadratic first integrals. We therefore integrate (103) with mean time step $h = 0.01$ from time $t = 0$ to time $t = 4000$ which corresponds to approximately 636 revolutions of the system (long-time behavior). Moreover, we consider the additive noise method (5) with $h = 0.01$, expecting that the first integral will not be conserved. We observe in Figure 5 that the method (6) conserves the angular momentum, while for the method (5) the approximate conservation of the quadratic first integral shown in Section 6.1 is lost when integrating (103) over long time.

8.6. Chaotic Hamiltonian systems. Probabilistic methods for differential equations are of particular interest when applied to chaotic systems. Hamiltonian systems can present chaotic features in the long-time behavior, while conserving geometric properties. It is therefore relevant to study whether both the geometric features and the chaotic nature are captured by the probabilistic integrator. An example of this class of system is given by the Hénon-Heiles problem [9], whose Hamiltonian (total energy) is given by

$$(106) \qquad H(p,q) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|q\|^2 + q_1^2 q_2 - \frac{1}{3}q_2^3,$$

where p and q are vectors of $\mathbb{R}^2$ representing velocity and position respectively. The corresponding system of ODEs can then be written as

$$(107) \qquad \begin{aligned} p' &= -\nabla_q H(p,q), \quad p(0) = p_0 \in \mathbb{R}^2, \\ q' &= \nabla_p H(p,q), \qquad q(0) = q_0 \in \mathbb{R}^2. \end{aligned}$$

It is well-known that this systems presents typical deterministic chaos for values $H > 1/8$. We pick as deterministic integrator the s-stage trapezoidal rule [10]. It has been proved that this numerical scheme conserves polynomial Hamiltonians of degree less than s for s even, or $s+1$ for s odd. Thanks to Theorem 6.3, if we choose as the s-stage trapezoidal rule, with $s \geq 4$, as the deterministic component Ψ_h of (6), we expect the Hamiltonian (106) to be conserved exactly for all the numerical trajectories. However, we do not expect the additive noise method (5) to conserve the Hamiltonian in this case.

For this numerical experiment, we assume the initial condition to be

$$(108) \qquad q_1(0) = 0.5, \quad q_2(0) = 0, \quad p_1(0) = 0, \quad p_2(0) = 0.1,$$

so that the Hamiltonian is $H = 0.13 > 1/8$ and (107) is in the chaotic regime. We then integrate the equation for time $t \in [0, 600]$ with the two probabilistic versions of the 5-stage trapezoidal rule choosing as mean time step $h = 0.01$ for (6) and fixing $h = 0.01$ for (5). We compute $M = 10$ trajectories for both the numerical methods. The results given in Figure 6 show that each trajectory given by the random time-stepping technique conserves the Hamiltonian function, nonetheless capturing the chaotic behavior of the system at long time (Figure 6, second row). In fact, after a short time where all trajectories approximately coincide, the attractor in the phase space is explored fully by the set of trajectories. In contrast, the additive noise method does not conserve the Hamiltonian, with all trajectories presenting an energy deviation from the initial value. Due to this deviation, all trajectories present an unstable behavior after a short time, while the random time-stepping technique is stable for the same value of h.
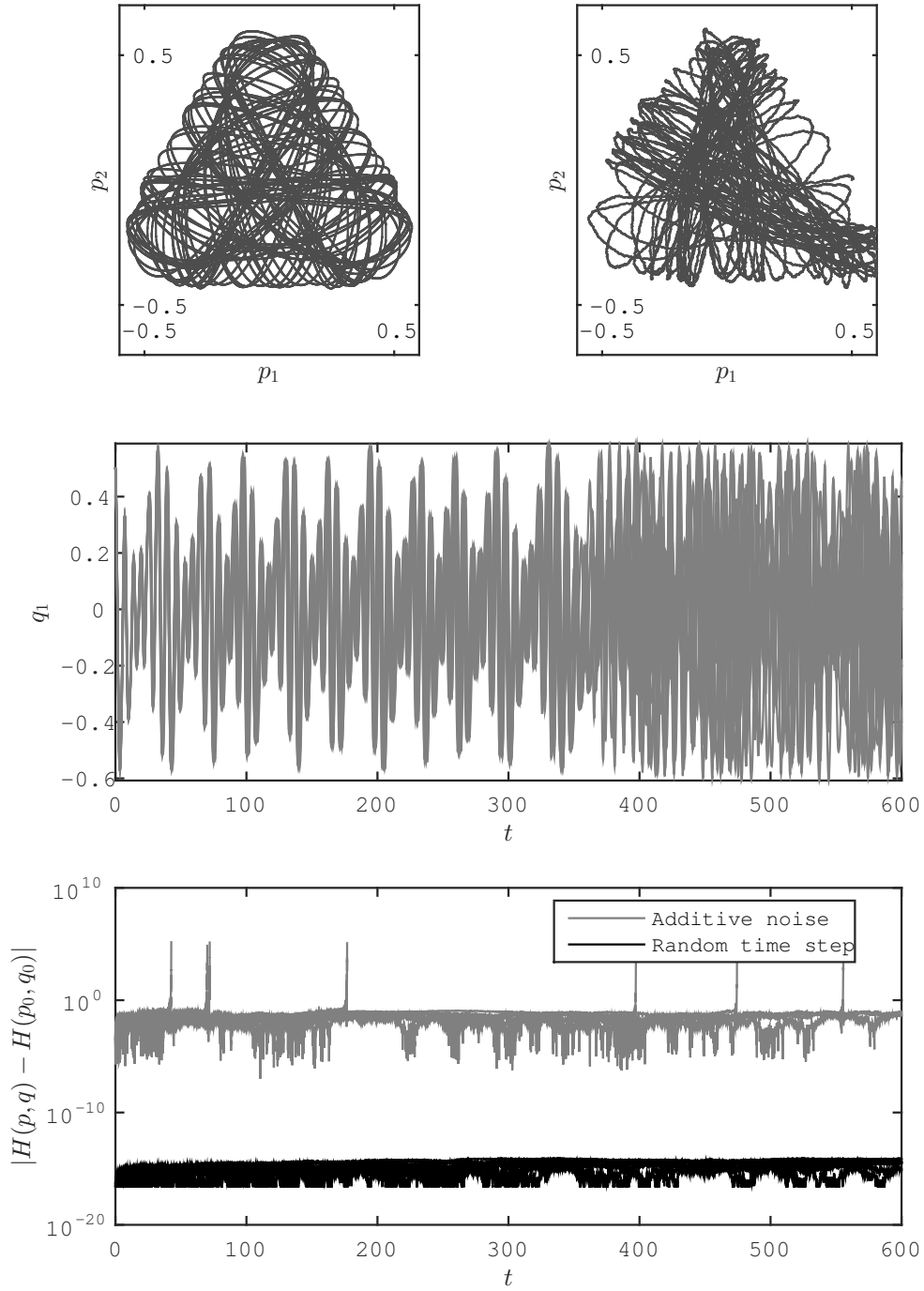
Fig. 6: A trajectory of (107) given by the RTS-RK method (6) for $0 \leq t \leq 350$ (first row, left), and by the additive noise method (5) (first row, right). Value of q_1 for the whole set of realizations of (6) for $0 \leq t \leq 600$. Error on the Hamiltonian $H(p_0, q_0) = 0.13$ for $0 \leq t \leq 600$.
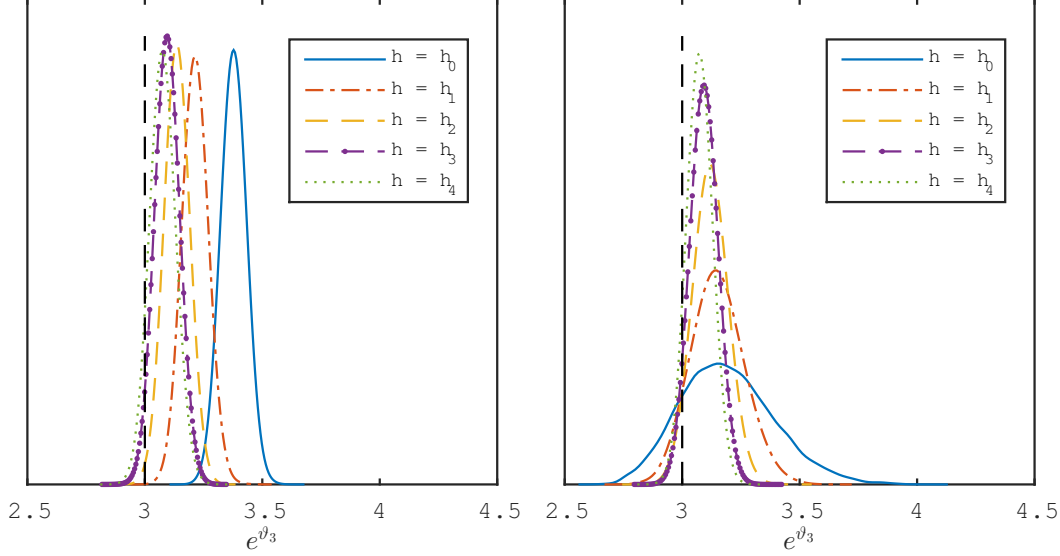
Fig. 7: Posterior distribution for $\exp(\vartheta_3)$ obtained with the standard random walk Metropolis and the explicit Euler method (left picture) and the pseudo-marginal Metropolis-Hastings and RTS-RK explicit Euler method (right picture). In the plot, $h_i = 0.1 \cdot 2^{-i}$. The dashed vertical line indicates the true value of the parameter $\exp(\vartheta_3^*)$.

8.7. Bayesian inferential problems. For the last numerical experiment we revisit the FitzHug-Nagumo equation,

$$
\begin{aligned}
(109) \qquad & y'_{\vartheta,1} = c\big(y_{\vartheta,1} - \frac{y_{\vartheta,1}^3}{3} + y_{\vartheta,2}\big), \quad y_{\vartheta,1}(0) = -1, \\
& y'_{\vartheta,2} = -\frac{1}{c}(y_{\vartheta,1} - a + b y_{\vartheta,2}), \quad y_{\vartheta,2}(0) = 1,
\end{aligned}
$$

where in the spirit of Bayesian inverse problems we consider $a, b, c \in \mathbb{R}$ to be unknown parameters. We recall that their true value is given by $a = 0.2$, $b = 0.2$ and $c = 3$, see equation (98). The test is then to infer the value of the parameter $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3)^\top = (\log a, \log b, \log c)^\top \in \mathbb{R}^3$ from observations of the solution of (109). The exponential transformation operated on each component of the parameter is taken to ensure positiveness of $(a, b, c)^\top$ for any value of ϑ. We consider the forward problem operator $\mathcal{G} \colon \mathbb{R}^n \to \mathbb{R}^m$ introduced in Section 7 to be defined by

$$
(110) \qquad \mathcal{G}(\vartheta) = \big(y_\vartheta(0.1)^\top, y_\vartheta(0.2)^\top, \dots, y_\vartheta(1)^\top\big)^\top,
$$

thus $n = 3$ and $m = 20$. Synthetic observations are then generated fixing ϑ to its true value $\vartheta^* = (\log(0.2), \log(0.2), \log(3))^\top$ and integrating the equation with a Runge-Kutta solver with a small time step and biased by Gaussian observational noise, i.e.,

$$
(111) \qquad \mathcal{Y} = \mathcal{G}(\vartheta) + \eta, \quad \eta \sim \mathcal{N}\big(0, (0.05)^2 I_{20}\big),
$$

where I_r is the r-dimensional identity matrix. We fix the prior distribution on ϑ to be a standard Gaussian $\mathcal{N}(0, I_3)$. Let us remark that since the amount of noise in the observational model is small, this choice of prior does not yield information on the true value of the parameter. We then sample from the posterior distributions $\pi^h(\vartheta \mid \mathcal{Y})$ and $\pi_{\mathrm{pr}}^h(\vartheta \mid \mathcal{Y})$ employing the sampling methods described in Section 7.1, namely

(i) the standard RWM with the forward map approximated by the explicit Euler method,

(ii) the PMMH with the forward map approximated via RTS-RK implemented with the explicit Euler method.

In both cases, we perform the MCMC algorithm four times, varying the step size h (or mean step size for RTS-RK), precisely choosing $h_i = 0.1 \cdot 2^{-i}$, $i = 0, 1, 2, 3, 4$. We recall that for the PMMH algorithm the stationary distribution of the generated Markov chain is independent of the number of samples chosen to approximate the posterior at each iteration with Monte Carlo. In order to get an initial value for the Markov chain in the PMMH case, we perform a short trial run of the MCWM method to move towards sets with significant values of the posterior density. In Figure 7 we show the marginal posterior for the parameter $\exp(\vartheta_3)$. We observe that while for the standard deterministic approximation of $\mathcal{G}$ the coarser values of h provide an overconfident posterior centred in a biased value, the same computation performed with RTS-RK and PMMH accounts better for the statistical uncertainty introduced by the discretisation error.

REFERENCES

[1] A. ABDULLE, *Fourth order Chebyshev methods with recurrence relation*, SIAM J. Sci. Comput., 23 (2002), pp. 2041–2054.

[2] A. ABDULLE AND A. A. MEDOVIKOV, *Second order Chebyshev methods based on orthogonal polynomials*, Numer. Math., 90 (2001), pp. 1–18.

[3] C. ANDRIEU AND G. O. ROBERTS, *The pseudo-marginal approach for efficient Monte Carlo computations*, Ann. Statist., 37 (2009), pp. 697–725.

[4] J. COCKAYNE, C. OATES, T. SULLIVAN, AND M. GIROLAMI, *Probabilistic numerical methods for PDE-constrained Bayesian inverse problems*, AIP Conference Proceedings, 1853 (2017), p. 060001.

[5] P. R. CONRAD, M. GIROLAMI, S. SÄRKKÄ, A. STUART, AND K. ZYGALAKIS, *Statistical analysis of differential equations: introducing probability measures on numerical solutions*, Stat. Comput., (2016).

[6] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Series in Computational Mathematics 31, Springer-Verlag, Berlin, second ed., 2006.

[7] E. HAIRER, S. P. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations I. Nonstiff Problems*, vol. 8, Springer Verlag Series in Comput. Math., Berlin, 1993.

[8] E. HAIRER AND G. WANNER, *Solving ordinary differential equations II. Stiff and differential-algebraic problems*, Springer-Verlag, Berlin and Heidelberg, 1996.

[9] M. HÉNON AND C. HEILES, *The applicability of the third integral of motion: Some numerical experiments*, Astronom. J., 69 (1964), pp. 73–79.

[10] F. IAVERNARO AND D. TRIGIANTE, *High-order symmetric schemes for the energy conservation of polynomial Hamiltonian problems*, JNAIAM J. Numer. Anal. Ind. Appl. Math., 4 (2009), pp. 87–101.

[11] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences, 160, Springer, 2005.

[12] H. KERSTING AND P. HENNIG, *Active uncertainty calibration inBayesian ODE solvers*, in Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI 2016), AUAI Press, 2016, pp. 309–318.

[13] E. N. LORENZ, *Deterministic nonperiodic flow*, J. Atmos. Sci., 20 (1963), pp. 130–141.

[14] F. J. MEDINA-AGUAYO, A. LEE, AND G. O. ROBERTS, *Stability of noisy Metropolis–Hastings*, Stat. Comput., 26 (2016), pp. 1187–1211.

[15] L. F. OLSEN, *An enzyme reaction with a strange attractor*, Phys. Lett. A, 94 (1983), pp. 454 – 457.

[16] R. D. SKEEL AND C. W. GEAR, *Does variable step size ruin a symplectic integrator?*, Physica, 60 (1992), pp. 311–313.

[17] A. M. STUART, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[18] P. VAN DER HOUWEN AND B. P. SOMMEIJER, *On the internal stage Runge-Kutta methods for large m-values*, Z. Angew. Math. Mech., 60 (1980), pp. 479–485.