A PROBABILISTIC INTEGRATOR OF ORDINARY DIFFERENTIAL EQUATIONS BASED ON RUNGE-KUTTA METHODS WITH RANDOM SELECTION OF THE TIME STEPS

ASSYR ABDULLE* AND GIACOMO GAREGNANI†

Abstract.

1. Introduction. Probabilistic integrators for ordinary differential equations (ODE) are important add a clever introduction. Conrad et al. [1] recently proposed a numerical scheme based on Runge-Kutta methods and an additive noise technique to provide with a probabilistic interpretation of the solution of ODEs. Given a time step $h > 0$, the method simply reads

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

where Y_k is an approximation of the solution $y(t_k)$ at time $t_k = kh$, Ψ_h is the numerical flow of the chosen deterministic Runge-Kutta method and the $\xi_k(h)$ are opportunely scaled random variables.

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

$$(2) \qquad y' = f(y), \quad y(0) = y_0 \in \mathbb{R}^d.$$

We can write the solution $y(t)$ in terms of the flow of the ODE, i.e., the family $\{\varphi_t\}_{t \geq 0}$ of functions $\varphi_t \colon \mathbb{R}^d \to \mathbb{R}^d$ such that

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

Let us consider a Runge-Kutta method for (2). Given a time step h, we can write the numerical solution at y_n approximation of $y(t_n)$, with $t_n = nh$ 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

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

If the ODE is chaotic, integrating the equation with different time steps leads to different numerical solutions. In order to describe the uncertainty of the numerical solution, we choose at each time t_k the time step as the realization of a random variable H_k. Therefore, the numerical solution is given by a discrete stochastic process $\{Y_k\}_{k \geq 0}$ such that

$$(5) \qquad Y_{k+1} = \Psi_{H_k}(Y_k).$$

This probabilistic numerical method shares the basic idea with [1].

3. Strong order analysis. The first property of (5) we analyze is its strong order of convergence, which gives an indication on the path-wise distance between each realization of the numerical solution and the exact solution of (2). Let us define the strong order of convergence.

DEFINITION 3.1. *The numerical method (5) has strong order r for (2) if there exists a constant $C > 0$ independent of h such that*

$$(6) \qquad \mathbb{E}|Y_k - y(kh)| \leq Ch^r$$

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

---
*Mathematics Section, École Polytechnique Fédérale de Lausanne (assyr.abdulle@epfl.ch)
†Mathematics Section, École Polytechnique Fédérale de Lausanne (giacomo.garegnani@epfl.ch)

In order to prove a result of strong convergence, we have to introduce assumptions on the random variables $\{H_k\}_{k\geq 0}$ as well as the deterministic numerical flow Ψ_h. Let us first consider the random time steps.

Assumption 3.2. The i.i.d. random variables H_k satisfy for all $k = 1, 2, \ldots$

(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

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

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 3.3. Let us consider the random variables $H_{kk\geq 0}$ such that

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

All the assumptions above are verified for this choice of the time steps, as

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

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

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

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

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

The following assumption on the numerical flow is needed for proving strong convergence.

Assumption 3.4. 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

$$(11) \qquad |\Psi_h(y) - \varphi_h(y)| \leq Ch^{q+1}, \quad \forall y \in \mathbb{R}^d.$$

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

$$(12) \qquad |\Psi_t(y) - \Psi_s(y)| \leq L_\Psi |t - s|, \quad \forall t, s > 0.$$

(iii) There exists a constant $C > 0$ independent of h such that

$$(13) \qquad |\Psi_h(y) - \Psi_h(z)| \leq (1 + Ch)|y - z|, \quad \forall y, z \in \mathbb{R}^d.$$

The local strong convergence of (5) can now be proved.

THEOREM 3.5 (Strong local order). *Under Assumptions 3.2 and 3.4 the numerical solution Y_1 given by one step of* (5) *satisfies*

$$(14) \qquad \mathbb{E}|Y_1 - y(h)| \leq Ch^{\min\{q+1,p\}},$$

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

Proof. By the triangular inequality we have for all $y \in \mathbb{R}^d$

$$(15) \qquad \mathbb{E}|\Psi_{H_0}(y) - \varphi_h(y)| \leq \mathbb{E}|\Psi_{H_0}(y) - \Psi_h(y)| + \mathbb{E}|\Psi_h(y) - \varphi_h(y)|.$$

We now consider Assumption 3.4.(ii) and 3.4.(i), thus getting

$$(16) \qquad \mathbb{E}|\Psi_{H_0}(y) - \varphi_h(y)| \leq L_\Psi \, \mathbb{E}|H_0 - h| + C_1 h^{q+1}.$$

We now apply Jensen's inequality and Assumption 3.2 to obtain

$$(17) \qquad \begin{aligned} \mathbb{E}|\Psi_{H_0}(y) - \varphi_h(y)| &\leq L_\Psi (\mathbb{E}|H_0 - h|^2)^{1/2} + C_1 h^{q+1} \\ &\leq L_\Psi h^p + C_1 h^{q+1} \\ &\leq C h^{\min\{q+1, p\}}, \end{aligned}$$

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

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

THEOREM 3.6 (Strong global order). *If $t_k = kh$ for $k = 1, 2, \ldots, N$, where $Nh = T$, and under Assumptions 3.2 and 3.4 the numerical solution given by (5) satisfies*

$$(18) \qquad \sup_{k=1,2,\ldots,N} \mathbb{E}|Y_k - y(t_k)| \leq C h^{\min\{q, p-1/2\}},$$

where C is a real positive constant independent of h.

Proof. Let us define $e_k := \mathbb{E}|Y_k - y(t_k)|$. Thanks to the triangular inequality we have

$$(19) \qquad e_k \leq \mathbb{E}|\Psi_{H_{k-1}}(Y_{k-1}) - \Psi_{H_{k-1}}(y(t_{k-1}))| + \mathbb{E}|\Psi_{H_{k-1}}(y(t_{k-1})) - \varphi_h(y(t_{k-1}))|.$$

We then apply Assumption 3.4.(iii) for the first term and Theorem 3.5 for the second term, thus obtaining

$$(20) \qquad e_k \leq (1 + C_1 h)e_{k-1} + C_2 h^{\min\{q+1, p\}}.$$

Hence, iterating over k and noticing that $e_0 = 0$, we get

$$(21) \qquad e_k \leq C_2 h^{\min\{q+1, p\}} \sum_{i=0}^{k-1} (1 + C_1 h)^i.$$

We then remark that

$$(22) \qquad \sum_{i=0}^{k-1} (1 + C_1 h)^i \leq T e^{C_1 T} h^{-1},$$

which proves the desired result with $C = C_2 T e^{C_1 T}$. □

4. Weak order analysis. The second property we wish to analyze is the weak convergence, giving an indication of how the behavior of the numerical solution (5) in the mean sense. Let us define the weak order of convergence.

DEFINITION 4.1. *The numerical method (5) has weak order r for (2) if for any function $\Phi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ there exists a constant $C > 0$ independent of h such that*

$$(23) \qquad |\mathbb{E}\,\Phi(Y_k) - \Phi(y(kh))| \leq C h^r,$$

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

In order to prove a result of weak convergence, we apply techniques of backwards error analysis. In general, these techniques imply the theoretical existence of modified equations such that the order of convergence of the method with respect to these equations is higher than the order with respect to the original equation. Let us introduce the operators $\mathcal{L}$ and $\mathcal{L}^h$ such that, if Y_1 is the first value produced by the numerical scheme described above,

$$\Phi(\varphi_h(y)) = e^{h\mathcal{L}}\Phi(y),$$

(24)

$$\mathbb{E}\,\Phi(Y_1 \mid Y_0 = y) = e^{h\mathcal{L}^h}\Phi(y),$$

for all functions Φ in $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$. In particular, we can write explicitly $\mathcal{L} = f \cdot \nabla$, while no closed form expression for $\mathcal{L}^h$ is available. We now expand the functional of the numerical solution as follows

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

(25)

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., $A : B = \mathrm{tr}(A^T B)$. Taking the conditional expectation with respect to $Y_0 = y$ we get

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

(26)

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

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

(27)

which implies

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

(28)

Consider now the following modified ODE and modified SDE

(29) $$\hat{y}' = f^h(\hat{y}),$$

(30) $$d\tilde{y} = \Big(f^h(\tilde{y}) + \frac{1}{2}Ch^{2p-1}\partial_{tt}\Psi_h(\tilde{y})\Big)dt + \sqrt{Ch^{2p-1}\partial_t\Psi_h(\tilde{y})\partial_t\Psi_h(\tilde{y})^T}\,dW_t,$$

where W_t, with $t \geq 0$, is a standard d-dimensional Wiener process. These equations define the differential operators $\hat{\mathcal{L}}$ and $\tilde{\mathcal{L}}$ such that

(31) $$\Phi(\hat{y}(h) \mid \hat{y}(0) = y) = (e^{h\hat{\mathcal{L}}}\Phi)(y), \quad \mathbb{E}\,\Phi(\tilde{y}(h) \mid \tilde{y}(0) = y) = (e^{h\tilde{\mathcal{L}}}\Phi)(y).$$

The operators $\hat{\mathcal{L}}$ and $\tilde{\mathcal{L}}$ can be written explicitly from (29) and (30) as

$$(32) \qquad \hat{\mathcal{L}} = f^h \cdot \nabla, \quad \tilde{\mathcal{L}} = \left(f^h + \frac{1}{2}Ch^{2p-1}\partial_{tt}\Psi_h\right) \cdot \nabla + \frac{1}{2}Ch^{2p-1}\partial_t\Psi_h\partial_t\Psi_h^T : \nabla^2.$$

Let us remark that the function f^h is of the form

$$(33) \qquad f^h = f + \sum_{i=q}^{q+l} h^i f_i,$$

where q is given in Assumption 3.4.(i) and the terms f_i are chosen such that

$$(34) \qquad e^{h\hat{\mathcal{L}}}\Phi(y) - \Phi(\Psi_h(y)) = \mathcal{O}(h^{q+2+l}).$$

It has been shown [4] that such a function always exists for Runge-Kutta methods. Let us compute the distance between the generator of the SDE and the modified ODE

$$(35) \qquad e^{h\tilde{\mathcal{L}}}\Phi(y) - e^{h\hat{\mathcal{L}}}\Phi(y) = e^{hf^h \cdot \nabla}\left(e^{\frac{1}{2}Ch^{2p}(\partial_{tt}\Psi_h \cdot \nabla + \partial_t\Psi_h\partial_t\Psi_h^T : \nabla^2)} - I\right)\Phi(y).$$

Expanding with Taylor the two factors we get

$$(36) \qquad \begin{aligned} e^{h\tilde{\mathcal{L}}}\Phi(y) - e^{h\hat{\mathcal{L}}}\Phi(y) &= (I + \mathcal{O}(h))\left(\frac{1}{2}Ch^{2p}\partial_{tt}\Psi_h \cdot \nabla + \frac{1}{2}Ch^{2p}\partial_t\Psi_h\partial_t\Psi_h^T : \nabla^2 + \mathcal{O}(h^{4p})\right)\Phi(y) \\ &= \frac{1}{2}Ch^{2p}\partial_{tt}\Psi_h(y) \cdot \nabla\Phi(y) + \frac{1}{2}Ch^{2p}\partial_t\Psi_h(y)\partial_t\Psi_h(y)^T : \nabla^2\Phi(y) + \mathcal{O}(h^{2p+1}) \end{aligned}$$

Now considering (28) and (34), we have

$$(37) \qquad \begin{aligned} e^{h\mathcal{L}^h}\Phi(y) - e^{h\hat{\mathcal{L}}}\Phi(y) &= \frac{1}{2}Ch^{2p}\partial_{tt}\Psi_h(y) \cdot \nabla\Phi(y) + \frac{1}{2}Ch^{2p}\partial_t\Psi_h(y)\partial_t\Psi_h(y)^T : \nabla^2\Phi(y) \\ &\quad + \mathcal{O}(h^{2p+1}) + \mathcal{O}(h^{q+2+l}). \end{aligned}$$

Then, combining (36) and (37) we then get the distance between the probabilistic method and the solution of the SDE as

$$(38) \qquad e^{h\tilde{\mathcal{L}}}\Phi(y) - e^{h\mathcal{L}^h}\Phi(y) = \mathcal{O}(h^{2p+1}) + \mathcal{O}(h^{q+2+l}).$$

Choosing $l = 2p - 1 - q$ we have the balance between the two terms and

$$(39) \qquad e^{h\tilde{\mathcal{L}}}\Phi(y) - e^{h\mathcal{L}^h}\Phi(y) = \mathcal{O}(h^{2p+1}),$$

which is the one-step weak error between the probabilistic numerical method and the modified SDE. For the original equation, let us remark that thanks to Assumption 3.4.(i) we have

$$(40) \qquad e^{h\mathcal{L}}\Phi(y) - \Phi(\Psi(h,y)) = \mathcal{O}(h^{q+1}).$$

Combining (40) and (28) we have the one-step weak error of the probabilistic method on the original ODE, i.e.,

$$(41) \qquad e^{h\mathcal{L}}\Phi(y) - e^{h\mathcal{L}^h}\Phi(y) = \mathcal{O}(h^{\min\{2p, q+1\}}).$$

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 [1].

Assumption 4.2. The function f is in $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and all its derivatives are uniformly bounded on $\mathbb{R}^d$. Furthermore, f is such that the operators $e^{h\mathcal{L}}$ and $e^{h\mathcal{L}^h}$ satisfy, for all functions $\Phi \in \mathbb{C}^\infty(\mathbb{R}^d, \mathbb{R})$ and a positive constant L

$$
(42) \qquad
\begin{aligned}
\sup_{u \in \mathbb{R}^d} |e^{h\mathcal{L}}\Phi(u)| &\le (1 + Lh) \sup_{u \in \mathbb{R}^d} |\Phi(u)|, \\
\sup_{u \in \mathbb{R}^d} |e^{h\mathcal{L}^h}\Phi(u)| &\le (1 + Lh) \sup_{u \in \mathbb{R}^d} |\Phi(u)|.
\end{aligned}
$$

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 [1].

THEOREM 4.3. *Under Assumptions 3.2, 3.4 and 4.2 there exists a constant $C > 0$ such that for all functions $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ in $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$*

$$
(43) \qquad |\varphi(u(T)) - \mathbb{E}\,\varphi(U_N)| \le C h^{\min\{2p-1, q\}},
$$

where $u(t)$ is the solution of (2) *and $T = Nh$. Moreover,*

$$
(44) \qquad |\varphi(\tilde{u}(T)) - \mathbb{E}\,\varphi(U_N)| \le C h^{2p},
$$

where $\tilde{u}(t)$ is the solution of (30).

Proof. Let us introduce the following notation

$$
(45) \qquad
\begin{aligned}
w_k &= \Phi(y(t_k) \mid y(0) = y_0) \\
W_k &= \mathbb{E}\,\Phi(Y_k \mid Y_0 = y_0).
\end{aligned}
$$

Then by the triangular inequality and the Markov property we have

$$
(46) \qquad |w_k - W_k| \le |e^{h\mathcal{L}}w_{k-1} - e^{h\mathcal{L}^h}w_{k-1}| + |e^{h\mathcal{L}^h}w_{k-1} - e^{h\mathcal{L}^h}W_{k-1}|.
$$

Applying (41) to the first term and Assumption 4.2 to the second, we have

$$
(47) \qquad |w_k - W_k| \le C h^{\min\{2p, q+1\}} + (1 + Lh)|w_{k-1} - W_{k-1}|.
$$

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

$$
(48) \qquad
\begin{aligned}
|w_k - W_k| &\le C k h^{\min\{2p, q+1\}} \\
&\le C T h^{\min\{2p-1, q\}},
\end{aligned}
$$

which proves the first inequality. The proof of the second inequality follows the same steps as above considering $w_k = \mathbb{E}\,\Phi(\tilde{y} \mid \tilde{y}(0) = y_0)$ and applying (39). □

5. Monte Carlo estimators. The third property we are interested in studying is the behavior of Monte Carlo estimator drawn from the numerical solution (5). Given a function $\Phi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with Lipschitz constant L_Φ, we consider the mean square error (MSE) of the estimator $\hat{Z}$ defined as

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

where $T = hN$ is the final time, M is the number of trajectories and we denote by $Y_N^{(i)}$ the realizations of the solution for $i = 1, 2, \ldots, M$. It is trivial to remark that

$$
(50) \qquad
\begin{aligned}
\mathrm{MSE}(\hat{Z}) &= \mathbb{E}(\hat{Z} - Z)^2 \\
&= \mathrm{Var}(\hat{Z}) + \left(\mathbb{E}(\hat{Z} - Z)\right)^2.
\end{aligned}
$$

Hence, thanks to the result of Theorem 4.3, we have

$$\text{MSE}(\hat{Z}) = \text{Var}(\hat{Z}) + \mathcal{O}(h^{2\min\{q, 2p-1\}}). \tag{51}$$

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

$$\text{Var}\,\hat{Z} \le M^{-1} L_\Phi^2 \,\text{Var}\, Y_N. \tag{52}$$

We can now prove a bound for the MSE of the Monte Carlo estimator.

THEOREM 5.1. *Under Assumptions 3.2 and 3.4, the MSE of the Monte Carlo estimator satisfies*

$$\text{MSE}(\hat{Z}) \le C h^{\min\{2q, 2p-1\}} + \text{h.o.t.}, \tag{53}$$

where C is a positive constant independent of h.

Proof. Thanks to (51) and (52), we just have to show

$$\text{Var}(Y_k \mid Y_0 = y) \le \hat{C} h^{2p-1}, \tag{54}$$

for a positive constant $\hat{C}$. From the definition of variance, of the operator $e^{h\mathcal{L}^h}$ and from the Markov property we have

$$\text{Var}(Y_k \mid Y_0 = y) = e^{(k-1)h\mathcal{L}^h} \,\mathbb{E}(Y_1^2 \mid Y_0 = y) - \left(e^{(k-1)h\mathcal{L}^h} \,\mathbb{E}(Y_1 \mid Y_0 = y)\right)^2. \tag{55}$$

Using the definition of the numerical method (5) we have

$$\begin{aligned}
\mathbb{E}(Y_1^2 \mid Y_0 = y) &= \mathbb{E}(\Psi_{H_0}(y))^2 \\
&= \mathbb{E}(\Psi_{H_0}(y) - \Psi_h(y))^2 + \Psi_h(y)^2 \\
&\le L_\Psi^2 C h^{2p} + \Psi_h(y)^2,
\end{aligned} \tag{56}$$

where we exploited Assumption 3.2 and 3.4.(ii). Moreover, it is trivial that

$$\mathbb{E}(Y_1 \mid Y_0 = y) = \Psi_h(y). \tag{57}$$

Hence, we can bound the variance at the N-th step as

$$\begin{aligned}
\text{Var}(Y_k \mid Y_0 = y) &\le L_\Psi^2 C h^{2p} + \mathcal{O}(h^{2p+1}) + e^{(k-1)h\mathcal{L}^h} \Psi_h(y)^2 - \left(e^{(k-1)h\mathcal{L}^h} \Psi_h(y)\right)^2 \\
&= L_\Psi^2 C h^{2p} + \mathcal{O}(h^{2p+1}) + \text{Var}(\Psi_h(Y_k) \mid Y_0 = y).
\end{aligned} \tag{58}$$

Thanks to Assumption 3.4.(iii) we now have

$$\text{Var}(Y_k \mid Y_0 = y) \le L_\Psi^2 C h^{2p} + \mathcal{O}(h^{2p+1}) + (1 + \hat{C}h)^2 \,\text{Var}(Y_k \mid Y_0 = y), \tag{59}$$

for some constant $\hat{C} > 0$ independent of h. Hence,

$$\text{Var}(Y_k \mid Y_0 = y) \le \left(L_\Psi^2 C h^{2p} + \mathcal{O}(h^{2p+1})\right) \sum_{i=0}^{k-1} (1 + \hat{C}h)^{2i}. \tag{60}$$

We now remark that the sum is bounded as

$$\sum_{i=0}^{k-1} (1 + \hat{C}h)^{2i} \le k e^{2\hat{C}hk} \le T e^{2CT} h^{-1}, \tag{61}$$

where T is the final time of (2). The desired result is then proved setting $C = T e^{2CT}$. $\qquad\square$

Remark 5.2. This result is of critical importance as it implies that the Monte Carlo estimators drawn from (5) converge in the mean square sense independently of the number of samples M in (49). Hence, we can set $M = \mathcal{O}(1)$ when computing Monte Carlo averages without losing accuracy in the result if h is small enough. In fact, introducing as a measure of error the square root of the MSE, and imposing a fixed tolerance ε, i.e.,

$$\text{(62)} \qquad \qquad \text{MSE}(\hat{Z})^{1/2} = \mathcal{O}(\varepsilon),$$

we have thanks to Theorem 5.1 that the time step has to satisfy

$$\text{(63)} \qquad \qquad h = \mathcal{O}\big(\varepsilon^{1/\min\{q,p-1/2\}}\big),$$

regardless of the number of trajectories M. Therefore, the computational cost to attain such a tolerance ε is given by

$$\text{(64)} \qquad \qquad \text{cost} = \mathcal{O}(Mh^{-1}) = \mathcal{O}(\varepsilon^{\min\{q,p-1/2\}}).$$

6. Geometric properties. Numerical methods for ODEs are often studied in terms of their geometric properties [3]. In particular, we investigate here whether the random choice of time steps in (5) 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 $I\colon \mathbb{R}^n \to \mathbb{R}$, 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 $I(y) = v^T y$, 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 $I(y) = y^T S y$, then we say that I is a quadratic first integral of (2).*

If this property of the ODE is conserved by a numerical method, i.e., if for the first value $y_1 \in \mathbb{R}^d$ given by the numerical method it is true that $I(y_1) = I(y_0)$, then we say that the numerical method conserves the first integral. The first issue we treat is the conservation of linear first integrals, which can be seen as a general case of the conservation of mass in physical systems.

THEOREM 6.2. *The numerical method (5) conserves linear first integrals.*

Proof. For any linear first integral $I(y) = v^T y$ and by the definition of Runge-Kutta methods, we have

$$\text{(65)} \qquad \qquad I(Y_1) = v^T y_0 + H_0 \sum_{i=1}^{s} b_i v^T f(y_0 + H_0 \sum_{j=1}^{s} a_{ij} K_j),$$

where $\{b_i\}_{i=1}^{s}$, $\{a_{ij}\}_{i,j=1}^{s}$ and $\{K_i\}_{i=1}^{s}$ are the coefficients and the internal stages of the Runge-Kutta method respectively. Since $I(y)$ is a first integral, $v^T f(y) = 0$ for any $y \in \mathbb{R}^d$. Hence $I(Y_1) = I(y_0)$. $\square$

Remark 6.3. Let us remark that the conservation of linear first invariants is exact for any trajectory of the numerical method, and is not an average property. In other words, we can say that (5) conserves linear first integrals in the strong sense. For the additive noise numerical method (1), we have

$$\text{(66)} \qquad \begin{aligned} I(Y_1) &= v^T y_0 + h \sum_{i=1}^{s} b_i v^T f(y_0 + h \sum_{j=1}^{s} a_{ij} K_j) + v^T \xi_0(h), \\ &= v^T(y_0 + \xi_0(h)). \end{aligned}$$

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

We now consider quadratic first invariants, which are conserved by Runge-Kutta methods that satisfy the hypotheses of Cooper's theorem [3]. The conservation of quadratic first invariants is of the utmost importance, e.g., for Hamiltonian systems, as it implies the symplecticity of the scheme.

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	0.52	1.01	1.52	2.02	2.01	2.50	2.99	3.55	3.99	3.98

Table 1: Strong 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 3.2.

THEOREM 6.4. *If the Runge-Kutta scheme defined by Ψ_h conserves quadratic first integrals then the numerical method* (5) *conserves quadratic first integrals.*

Proof. Trivially, 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. ☐

Remark 6.5. It has been remarked numerically [2] that performing adaptivity in time steps with techniques of step size selection can spoil the symplectic nature of a numerical integrator. This phenomenon is particularly accentuated in the long-time behavior and in systems for which it is crucial to reduce the time step. We will present numerical experiments in this work showing that for random time steps satisfying Assumption 3.2 there is no such disruptive effect.

Remark 6.6. Even quadratic first integrals are conserved by the each trajectory of the numerical method. Instead, for the additive noise numerical method, we find

$$(67) \quad \begin{aligned} I(Y_1) &= (\Psi_h(y_0) + \xi_0(h))^T S(\Psi_h(y_0) + \xi_0(h))^T \\ &= I(y_0) + 2\xi_0(h)^T S\Psi_h(y_0) + \xi_0(h)^T S\xi_0(h). \end{aligned}$$

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

$$(68) \quad \mathbb{E}\,I(Y_1) = I(y_0) + Qh^{2p+1} : S.$$

Hence, for this method quadratic first integrals are not conserved path-wise, and even in the weak sense they present a bias.

It is known that no Runge-Kutta method can conserve any polynomial invariant of order $n \geq 3$ [3]. 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 6.7. *If the Runge-Kutta scheme defined by Ψ_h conserves an invariant $I(y)$ for an ODE, then the numerical method* (5) *conserves $I(y)$ for the same ODE.*

Proof. The proof is the same as the proof of Theorem 6.4. ☐

7. Numerical experiments. In this section, we show numerical experiments on some classic test cases that confirm the theoretical results presented above.

7.1. Strong order of convergence. In order to verify the strong order of convergence predicted in Theorem 3.6, we consider the FitzHug-Nagumo equation, which is defined as

$$(69) \quad \begin{aligned} y_1' &= c\big(y_1 - \frac{y_1^3}{3} + y_2\big), & y_1(0) &= -1, \\ y_2' &= -\frac{1}{c}(y_1 - a + by_2), & y_2(0) &= 1, \end{aligned}$$

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 3.2.

| Method	ET		RK4			
q		2			4	
p	2	3	2	3	4	5
$\min\{2q, 2p-1\}$	3	4	3	5	7	8
MSE order	3.01	4.05	3.04	5.02	7.08	8.06

Table 3: Convergence of the MSE of the Monte Carlo estimator for the random time-stepping explicit trapezoidal (ET) and fourth-order Runge-Kutta (RK4) with respect to p of Assumption 3.2.

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 3.4 with $q = 2$ and $q = 4$ respectively. Moreover, we consider random time steps as in Example 3.3, where we vary p in order to verify the order of convergence predicted in Theorem 3.6. We vary the mean time step h taken by the random time steps H_n in the range $h_i = 0.01^i$, with $i = 0, 1, \ldots, 4$. Then, we simulate 10^4 realizations of the numerical solution Y_{N_i}, with $N_i = T/h_i$ for $i = 0, 1, \ldots, 4$, and compute the approximate strong order of convergence for each value of h with a Monte Carlo mean. Results (Table 1) show that the orders predicted theoretically by Proposition 3.6 are confirmed numerically.
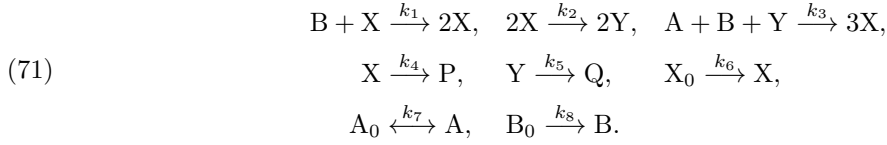
7.2. Weak order of convergence. We now verify the weak order of convergence predicted in Theorem 4.3. For this experiment we consider the ODE (69) as well, with the same time scale and parameters as above. The reference solution at final time is generated in this case as well with an high-order method 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 3.2 in order to verify the theoretical result of Theorem 4.3. The function $\Phi \colon \mathbb{R}^d \to \mathbb{R}$ of the solution we consider is defined as $\varphi(x) = x^T x$. 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.

7.3. Monte Carlo estimator. In the same spirit of the previous numerical experiments, we now verify numerically the validity of Theorem 5.1. We consider the ODE (69), with final time $T = 10$ 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 $h_i = 0.1 \cdot 2^{-i}$ with $i = 0, 1, \ldots, 5$. We thus vary the value of p of Assumption 3.2 and compute the mean order of convergence over 300 repetitions of the experiment. For each repetition, we consider only one trajectory to approximate the Monte Carlo estimator, i.e., $M = 1$. We then compute the error with respect to a reference solution given by a high-order method with a small time step. Results (Table 3) show that the convergence order given in Theorem 5.1 is respected in practice.

7.4. Robustness. Let us consider the Peroxide-Oxide chemical reaction, which is macroscopically defined the following balance equation

$$
(70) \qquad O_2 + 2NADH + 2H^+ \rightarrow 2H_2O + 2NAD^+.
$$
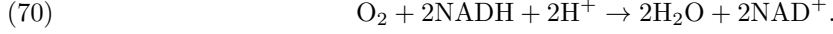
This reaction has to be catalyzed by an enzyme to take place, which reacts with the reagents to create intermediate products of the reaction. A successful model [7] to describe the time-evolution of the chemical system is the following

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

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

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

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

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

It has been shown [7] 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. From the physical point of view, since the components of the solution represent the concentration of chemicals, we require that the numerical solution respects the positivity of the solution and the conservation of mass. Let us apply the additive noise method (1) and the random time-stepping scheme (5) to equation (72). We choose $h = 0.05$ as the mean time steps for (5) and as the time step for (1), while we employ the stabilized explicit Runge-Kutta-Chebyshev method (RKC) [8] as deterministic integrator. Results (Figure 1) show that the method we propose in this work 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.

7.5. Conservation of quadratic first integrals. We consider the Kepler system with perturbation, a simple model for the two-body problems in celestial mechanics, which reads

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

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

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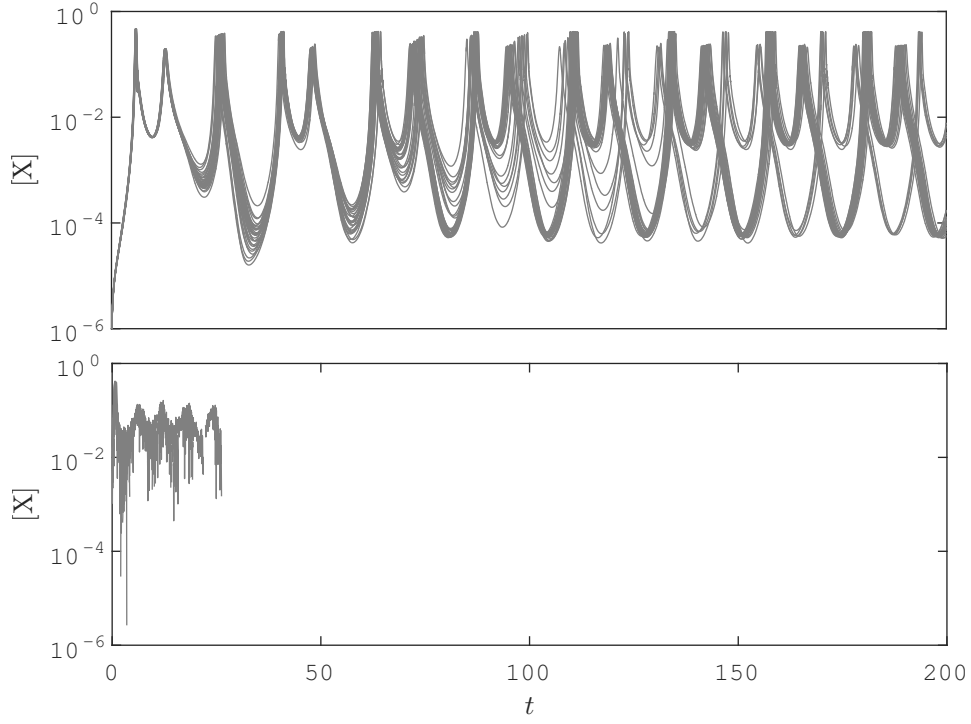$$

Fig. 1: Thirty trajectories of the numerical value of the concentration of the X species for the random time-stepping and additive noise models.

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

$$(76) \qquad\qquad I(p, q) = q_1 p_2 - q_2 p_1.$$

We consider the simplest Gauss collocation method, the implicit midpoint rule, as the deterministic Runge-Kutta method. It is known that all the integrators based on Gauss nodes conserve quadratic first integrals. We expect therefore that the random time-stepping method (5) implemented with Ψ_h given by the implicit midpoint rule conserves this property thanks to Theorem 6.4. We therefore integrate (74) with mean time step $h = 0.01$ from time $t = 0$ to time $t = 4000$ which corresponds to approximately 636 revolutions of the system, thus considering the long-time behavior of the numerical solution. Moreover, we consider the additive noise method (1) with $h = 0.01$, expecting that the first integral will not be conserved. Results (Figure 2) show that the method (5) conserves perfectly the angular momentum, while (1) satisfies the result shown in Remark 6.6 only in the short time, while in the long time the solution does not reflect the physical features of the underlying system.

7.6. 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 [5], whose Hamiltonian is given by

$$(77) \qquad\qquad H(p, q) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|q\|^2 + q_1^2 q_2 - \frac{1}{3}q_2^3,$$
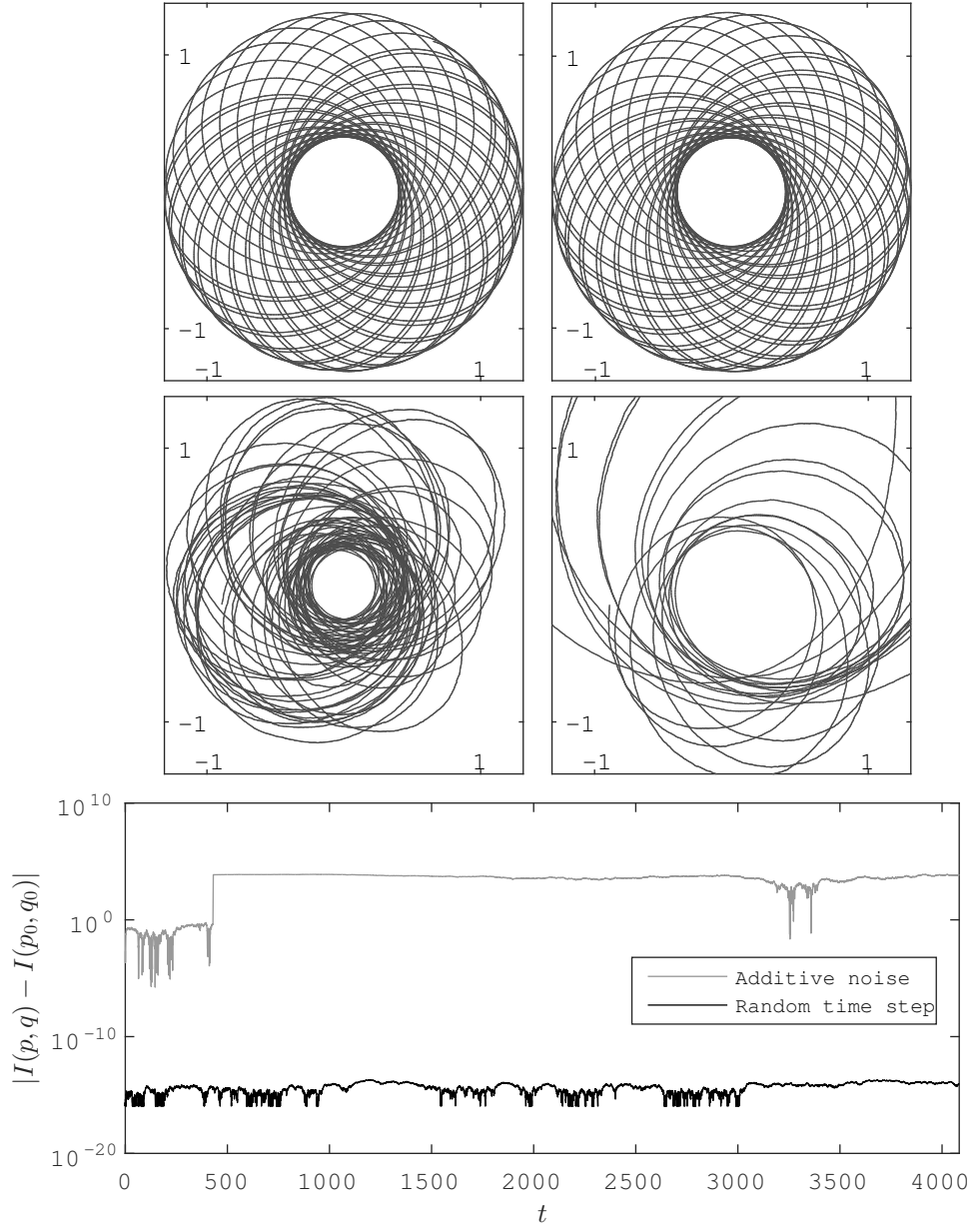
Fig. 2: Trajectories of (74) given by (5) for $0 \leq t \leq 200$ and $3800 \leq t \leq 4000$ (first row, left and right), and by (1) for $0 \leq t \leq 200$ and $200 \leq t \leq 400$ (second row, left and right). Error on the angular momentum for $0 \leq t \leq 4000$ given by the two methods.

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

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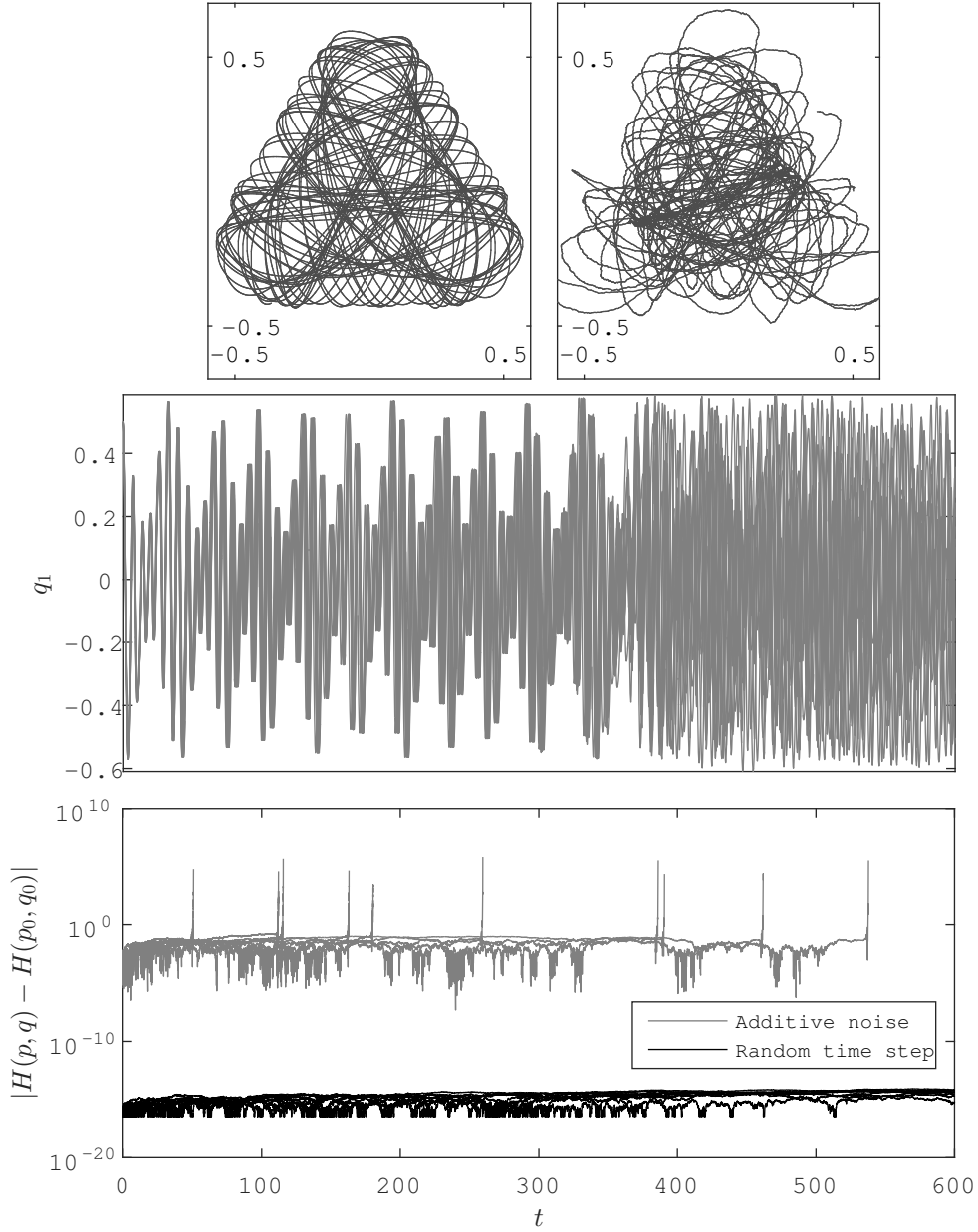\end{aligned}
$$

Fig. 3: Trajectories of (78) given by (5) for $0 \leq t \leq 350$ (first row, left), and by (1) (first row, right). Value of q_1 for the whole set of realizations of (5) for $0 \leq t \leq 600$. Error on the Hamiltonian for $0 \leq t \leq 600$.

It is well-known that this systems presents typical deterministic chaos for values $H > 1/8$. We consider as the deterministic integrator the s-stage trapezoidal rule [6], which is defined by its coefficients $\{b_i\}_{i=1}^s$, $\{c_i\}_{i=1}^s$, with $c_1 = 0$ and $c_s = 1$. The coefficients $\{a_{ij}\}_{i,j=1}^s$ are then determined as $a_{ij} = c_i b_j$. 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 6.7, if we choose as the s-stage trapezoidal rule, with

$s \geq 4$, as the deterministic component Ψ_h of (5), the Hamiltonian (77) should be conserved exactly by all the numerical trajectories. On the other hand, we do not expect the additive noise method (1) to conserve the Hamiltonian in this case.

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

$$(79) \qquad q_1(0) = 0.5, \quad q_2(0) = 0, \quad p_1(0) = 0, \quad p_2(0) = 0.1,$$

so that the Hamiltonian is $H = 0.13 > 1/8$ and (78) is in the chaotic regime. We then integrate the equation for time $t \in [0, 600]$ with the two probabilistic versions of the five-stage trapezoidal rule choosing as mean time step $h = 0.01$ for (5) and fixing $h = 0.01$ for (1). We compute $M = 10$ trajectories for both the numerical methods. Results (Figure 3) 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. 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. On the other hand, the additive noise method does not conserve the Hamiltonian neither path-wise nor in the mean sense. Furthermore, due to this deviation from the correct energy value, all trajectories present an unstable behavior after a short time, while the random time-stepping technique is stable for the same value of h.

REFERENCES

[1] 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).

[2] E. Hairer, *Variable time step integration with symplectic methods*, Appl. Numer. Math., 25 (1997), pp. 219–227.

[3] 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.

[4] 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.

[5] M. Hénon and C. Heiles, *The applicability of the third integral of motion: Some numerical experiments*, Astronom. J., 69 (1964), pp. 73–79.

[6] 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.

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

[8] P. van der Houwen and J. Kok, *Numerical solution of a minimax problem*, tech. report, Mathematical Centre, Amsterdam, 1971. Report TW 124/71.