CONSERVATION HAMILTONIAN RTS-RK

ASSYR ABDULLE* AND GIACOMO GAREGNANI†

1. Mean Hamiltonian. Consider the Hamiltonian $Q\colon \mathbb{R}^d \to \mathbb{R}$ and the ODE

$$y' = J^{-1}\nabla Q(y), \quad y(0) = y_0. \tag{1}$$

Applying a symplectic Runge Kutta method identified by its numerical flow Ψ, we have that the modified equation is still Hamiltonian and there exist functions Q_j, $j = 2, \dots$, such that

$$\tilde{Q}(y) = Q(y) + hQ_2(y) + h^2 Q_3(y) + \dots, \tag{2}$$

where h is the time step. The series in (2) does not converge, hence we consider the truncation after N terms

$$\tilde{Q}(y) = Q(y) + hQ_2(y) + \dots + h^{N-1}Q_N(y). \tag{3}$$

Moreover, if q is the order of convergence for Ψ, we have that $Q_i \equiv 0$ for $i = 2, \dots, q$, hence

$$\tilde{Q}(y) = Q(y) + h^q Q_{q+1}(y) + \dots + h^{N-1}Q_N(y). \tag{4}$$

Let us assume that Q is analytic in a neighbourhood of y_0 and denoting $f = J^{-1}\nabla Q$ that there exist positive constants R and M such that $\|f(y)\| \leq M$ for all $y \in B_{2R}(y_0) \subset \mathbb{R}^d$. Let us moreover introduce the constants μ and κ given by

$$\mu = \sum_{i=1}^s |b_i|, \quad \kappa = \max_{i=1,\dots,s} \sum_{j=1}^s |a_{ij}|, \tag{5}$$

where $\{b_i\}_{i=1}^s$ and $\{a_{ij}\}_{i,j=1}^s$ are the coefficients of the Runge-Kutta method. Finally, let us introduce the constant $\eta = \max\{\kappa, \mu/(2\log 2 - 1)\}$. Denoting by $\tilde{\varphi}_{N,t}(y)$ the exact flow of the equation corresponding to $\tilde{Q}$, we have that the local error satisfies [2, Theorem IX.7.6]

$$\|\Psi_h(y_0) - \tilde{\varphi}_{N,h}(y_0)\| \leq h\gamma M e^{-\kappa/h}, \tag{6}$$

for all $h \leq \kappa/4$, where $\kappa = R/(eM\eta)$ and $\gamma = e(2 + 1.65\eta + \mu)$.

[...]

LEMMA 1.1. *Under the assumption ADD THEM, there exist constants $C_i > 0$, $i = 1, \dots, 4$, such that*

$$\mathbb{E}|\tilde{Q}(Y_n) - \tilde{Q}(y_0)| \leq C_1 e^{-\kappa/2h}\big(1 + C_2 h^{2p-1}\big), \tag{7}$$

$$\mathbb{E}|Q(Y_n) - Q(y_0)| \leq C_1 e^{-\kappa/2h}(1 + C_2 h^{2p-1}) + C_3 h^q + C_4 h^{q+p-1/2}, \tag{8}$$

over exponentially long time intervals $nh \leq e^{\kappa/h}$.

Proof. We exploit the conservation of $\tilde{Q}$ along the trajectories of its corresponding dynamical system, i.e., $\tilde{Q}(\tilde{\varphi}_{N,z}(y)) = \tilde{Q}(y)$ for $y \in \mathbb{R}^d$ and $z > 0$ and employ a telescopic sum to obtain

$$\begin{aligned}\mathbb{E}|\tilde{Q}(Y_n) - \tilde{Q}(y_0)| &\leq \sum_{j=1}^n \mathbb{E}|(\tilde{Q}(Y_j) - \tilde{Q}(Y_{j-1})| \\ &= \sum_{j=1}^n \mathbb{E}|\tilde{Q}(Y_j) - \tilde{Q}(\tilde{\varphi}_{N,H_{j-1}}(Y_{j-1}))| \\ &= \sum_{j=1}^n \mathbb{E}\,\mathbb{E}\left(|\tilde{Q}(Y_j) - \tilde{Q}(\tilde{\varphi}_{N,H_{j-1}}(Y_{j-1}))| \mid H_{j-1}\right),\end{aligned} \tag{9}$$

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

where we applied the total expectation with respect to H_{j-1} for the last equality. Then, as Q is Lipschitz with constant independent of h and under the assumptions on $\{H_i\}_{i\geq 0}$ and (6) we have

$$\begin{aligned}\mathbb{E}|\tilde{Q}(Y_n)-\tilde{Q}(y_0)| &\leq C\sum_{j=0}^{n-1}\mathbb{E}\left(H_j e^{-\kappa/H_j}\right)\\ &= Cn\,\mathbb{E}\left(H_0 e^{-\kappa/H_0}\right),\end{aligned}\tag{10}$$

where the equality is given by the assumption of the random time steps being *i.i.d.* We can now consider the function $g(x)=xe^{-\kappa/x}$ and the bound

$$\begin{aligned}g(x) &\leq g(h)+g'(h)(x-h)+\frac{1}{2}\max_{x>0}g''(x)(x-h)^2\\ &\leq e^{-\kappa/h}\big(h+\frac{h+\kappa}{h}(x-h)\big)+\frac{27}{2\kappa}e^{-3}(x-h)^2,\quad x>0,\end{aligned}\tag{11}$$

which is valid as $\max_{x>0}g''(x)=27e^{-3}/\kappa$. Hence

$$\begin{aligned}\mathbb{E}|\tilde{Q}(Y_n)-\tilde{Q}(y_0)| &\leq Cn\,\mathbb{E}\left(e^{-\kappa/h}\big(h+\frac{h+\kappa}{h}(H_0-h)\big)+\frac{27}{\kappa}e^{-3}(H_0-h)^2\right)\\ &= Cnhe^{-\kappa/h}(1+\frac{27}{\kappa}e^{-3}h^{2p-1})\\ &\leq Ce^{-\kappa/2h}(1+\frac{27}{\kappa}e^{-3}h^{2p-1}),\end{aligned}\tag{12}$$

where the equality is given by the assumptions on H_0. Hence, the first result is proved with $C_1=C$ and $C_2=27e^{-3}/\kappa$. Let us now consider the original Hamiltonian and introduce the notation

$$R(y)=h^{-q}\big(\tilde{Q}(y)-Q(y)\big),\tag{13}$$

i.e., $R(y)=Q_{q+1}(y)+hQ_{q+2}(y)+\ldots+h^{N-q-1}Q_N(y)$. We then have by the triangular inequality

$$\mathbb{E}|Q(Y_n)-Q(y_0)|\leq\mathbb{E}|\tilde{Q}(Y_n)-\tilde{Q}(y_0)|+h^q\,\mathbb{E}|R(Y_n)-R(y_0)|.\tag{14}$$

The first term in the sum above is bounded thanks to (12). For the second term, we add and subtract the function R evaluated at exact solution of the modified equation to obtain

$$\mathbb{E}|R(Y_n)-R(y_0)|\leq\mathbb{E}|R(Y_n)-R(\tilde{\varphi}_{N,nh}(y_0))|+|R(\tilde{\varphi}_{N,nh}(y_0))-R(y_0)|,\tag{15}$$

where the expectation on the second term disappears and there exists $C>0$ independent of h and N such that

$$|R(\tilde{\varphi}_{N,nh}(y_0))-R(y_0)|\leq C.\tag{16}$$

For the first term, as R is Lipschitz with a constant independent of h and N we have

$$\begin{aligned}\mathbb{E}|R(Y_n)-R(\tilde{\varphi}_{N,nh}(y_0))| &\leq C\,\mathbb{E}\|Y_n-\tilde{\varphi}_{N,nh}(y_0)\|\\ &\leq \hat{C}e^{Lhn}h^{\min\{p-1/2,N\}},\end{aligned}\tag{17}$$

where the second bound is given by the strong order of convergence of the RTS-RK when applied to the modified equation, as the deterministic component in this case has order N. Since N is arbitrary, we can assume that $\min\{p-1/2,N\}=p-1/2$. Hence, we have the final decomposition of the error on the original Hamiltonian for positive constants C_i, $i=1,\ldots,4$, i.e.

$$\mathbb{E}|Q(Y_n)-Q(y_0)|\leq C_1e^{-\kappa/2h}(1+C_2h^{2p-1})+C_3h^q+C_4e^{Lhn}h^{q+p-1/2}.\tag{18}$$

□

Alternative proof of an alternative result. By Taylor expansion of the numerical solution and since $\tilde{Q}(y)$ is bounded we have

$$\mathbb{E}\,\tilde{Q}(Y_n) \leq \mathbb{E}\,\tilde{Q}(Y_{n-1}) + C\,\mathbb{E}\,H_{n-1}, \tag{19}$$

hence denoting by $\tilde{\Delta}_n \coloneqq \tilde{Q}(Y_n) - \tilde{Q}(y_n)$ where y_n constant time steps

$$\mathbb{E}\,\tilde{\Delta}_n \leq \mathbb{E}\,\tilde{\Delta}_{n-1} + C\,\mathbb{E}\,H_{n-1}, \tag{20}$$

which implies by Brouwer's argument [1, 3]

$$\mathbb{E}|\tilde{\Delta}_n| \leq Cn^{1/2}h^p. \tag{21}$$

By triangular inequality

$$\begin{aligned} \mathbb{E}|Q(Y_n) - Q(y_0)| \leq{}& \mathbb{E}|Q(Y_n) - \tilde{Q}(Y_n)| + \mathbb{E}|\tilde{Q}(Y_n) - \tilde{Q}(y_n)| \\ &+ |\tilde{Q}(y_n) - Q(y_n)| + |Q(y_n) - Q(y_0)|. \end{aligned} \tag{22}$$

Now $|Q(y) - \tilde{Q}(y)| \leq Ch^q$ for any y and result on fixed time steps

$$\mathbb{E}|Q(Y_n) - Q(y_0)| \leq Ch^q + Cn^{1/2}h^p. \tag{23}$$

□

Remark 1.2. The two results implied by Lemma 1.1 are consistent with the theory of deterministic symplectic integrators. In fact, in the limit $p \to \infty$, we have

$$\mathbb{E}|\tilde{Q}(Y_n) - \tilde{Q}(y_0)| = \mathcal{O}(e^{-\kappa/h}), \tag{24}$$

$$\mathbb{E}|Q(Y_n) - Q(y_0)| = \mathcal{O}(h^q), \tag{25}$$

and the expectation $\mathbb{E}\,Q(Y_n) \to Q(y_n)$, where y_n is the numerical solution given by the deterministic method.

Remark 1.3. In the bound (18) it is possible that $C_3 \ll C_4$, i.e., for large values of h the term corresponding to the randomness of the RTS-RK method can be dominating. On the other hand, the higher order of convergence $q + p - 1/2$ makes this term negligible when h tends to zero. In particular, implementing the reasonable choice $p = q + 1/2$ and disregarding the first term which decreases exponentially with h, we have

$$\mathbb{E}|Q(Y_n) - Q(y_0)| \leq C_3h^q + C_4h^{2q}. \tag{26}$$

1.1. Numerical experiment. Let us consider the Hénon-Heiles system, which is given by the Hamiltonian $Q\colon \mathbb{R}^4 \to \mathbb{R}$ defined by

$$Q(p, q) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|q\|^2 + q_1^2q_2 - \frac{1}{3}q_2^3, \tag{27}$$

where $y = (p, q)^\top \in \mathbb{R}^4$. We consider an initial condition such that $Q(y_0) = 0.13$ and integrate the equation employing the RTS-RK method with on the Gauss collocation method on two stages ($q = 4$) and the noise scale $p = \{2, 4\}$. We vary the mean time step $h_i = 0.2 \cdot 2^{-i}$ for $i = 0, \ldots, 7$ and consider the final time $T = 10^4$ for both values of p. We then compute the value of Q at final time and compare it with $Q(y_0)$ to check numerically the validity of Lemma 1.1. Results are shown in Figure 1, where the dashed and dotted lines are given by (18) disregarding the first term and setting $C_3 = 3 \cdot 10^{-2}$, $C_4 = 2 \cdot 10^{-4}$. It is possible to remark that for small values of h the slope of the error decreases as the asymptotic regime is reached.

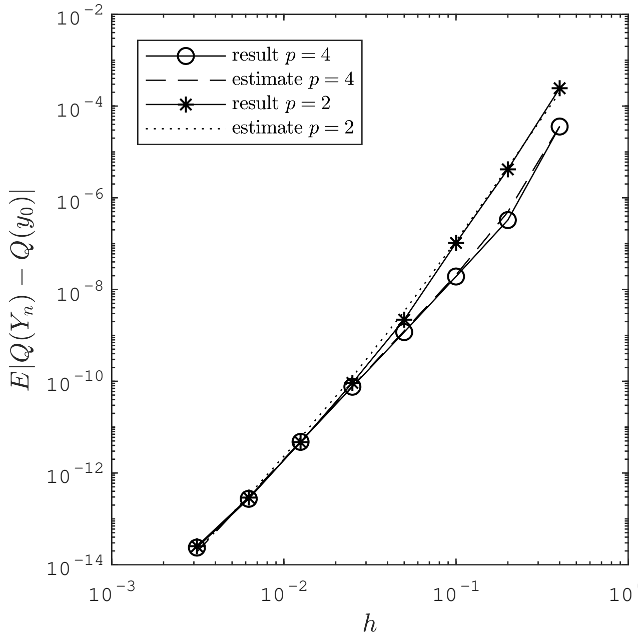

Fig. 1: Convergence of the mean error on the Hamiltonian for the Hénon-Heiles problem.

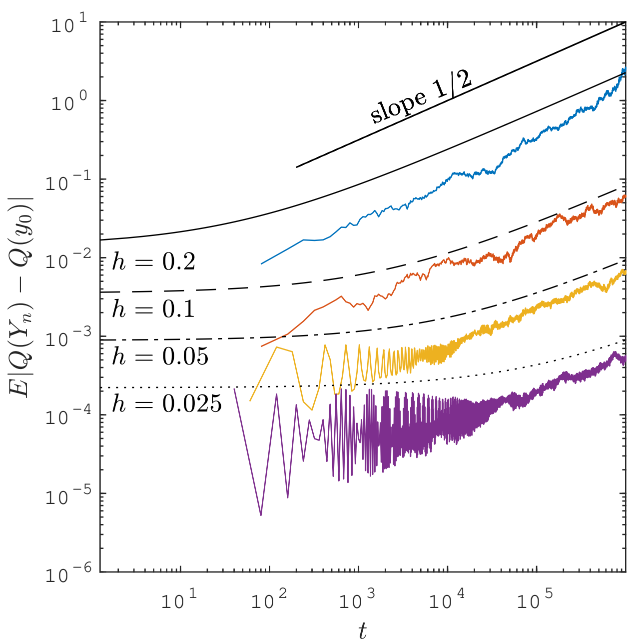

Fig. 2: Time evolution of the mean error, pendulum problem

1.2. Numerical experiment. Let us consider the pendulum problem, which is given by the Hamiltonian $Q\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$Q(p, q) = \frac{p^2}{2} - \cos q, \tag{28}$$

where $y = (p, q)^\top \in \mathbb{R}^2$. We consider the initial condition $(p_0, q_0) = (1.5, -\pi)$ and integrate the equation employing RTS-RK based on the implicit midpoint method ($q = 2$) and the noise scale $p = 2$. We vary the mean time step $h \in \{0.2, 0.1, 0.05, 0.025\}$ and consider the final time $T = 10^6$. We then study the time evolution of the numerical error on the Hamiltonian Q. Results are shown in Figure 2, where it is possible to notice . . . .

REFERENCES

[1] D. Brouwer, *On the accumulation of errors in numerical integration*, The Astronomical Journal, 46 (1937), pp. 149–153.

[2] 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.

[3] E. Hairer, R. McLachlan, and A. Razakarivony, *Achieving Brouwer's law with implicit Runge-Kutta methods*, BIT, 48 (2008), pp. 231–243.