Answer to the referees concerning the review of "Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration"

Assyr Abdulle Giacomo Garegnani

We thank again the reviewers for their comments. Some comments are not directly addressed in this report (i.e., minor mathematical issues, language/syntax issues, typos, references) but have nonetheless been considered in the rewriting of our paper.

As for the first review, all new modifications have been typeset in red in the new version of the paper for easing the review process. In this document, our answers to the two reviewers are divided into two sections.

We are grateful for the outstanding and unusual effort put by the two reviewers on commenting our manuscript. We therefore decided to acknowledge their precious contributions in the manuscript.

1 Reviewer #1

1. **Answer to the following comment:**

   > *Currently, the equation count in the paper is 175, which is rather high. Please consider removing equation numbers, especially if the equations are not referenced later in the paper. I think the reader will be grateful for being able to quickly skip equations that are not numbered, especially when they want to understand a step in a proof and the justification was given many pages earlier. On the other hand, if a mathematical statement is repeatedly used after its first appearance, I strongly recommend putting the statement in the equation environment and labelling the statement for easy reference.*

   We were aware of this issue, and maintained all equations numbered just for easing the review process. In this way, it would have been easier for the reviewers to ask questions about non-referenced equations. In the new version, we made use of the package `autonum`, which ensures that only referenced equations are numbered. The equation count is now 60, which seems reasonable to us given the length of our manuscript.

2. **Answer to the following comment:**

   > *[First comment after the statement of Lemma 4]: "The proof of Lemma 4 follows from the definition of the flow and the Gronwall inequality". The Gronwall inequality is not used here; it is the global Lipschitz continuity of the right-hand side / driving vector field f that is used for (37) ((16) in the new version).*

   It is true the Lipschitz continuity of f plays a fundamental role in proving (16), and we added this detail in the text. Nonetheless, we employed (and omitted) the following short proof, which exploits the Grönwall inequality. First, by definition of the flow and the triangle inequality, we have

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

The Lipschitz continuity of f is now employed to get

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

Now, the Grönwall inequality yields

$$\|\varphi_h(y) - \varphi_h(w)\| \leq \|y - w\|e^{Lh}.$$

We now write

$$e^{Lh} = 1 + h\frac{e^{Lh} - 1}{h},$$

and remark that the ratio $(e^{Lh} - 1)/h$ is a growing function of h. Therefore, replacing $h = 1$ (only in the ratio) we obtain an upper bound, i.e.,

$$e^{Lh} \leq 1 + h(e^L - 1).$$

The proof is completed by setting $C = e^L - 1$.

3. **Answer to the following comment:**

> *[Second comment after the statement of Lemma 4]: "... and the proof of Lemma 2 follows from the discrete Gronwall inequality". This comment should be moved so that it appears immediately after Lemma 2.*

Yes. We missed this detail after the first review.

4. **Answer to the following comment:**

> *Page 12, lines 5-8 [equation and inequalities in (54) (unnumbered in the new version)]: I understand that the second inequality follows by the Lipschitz continuity of Φ. However, the justification for the first inequality is unclear. Please provide a more explicit proof for the first inequality.*

We believe this passage does not need any further explanation. Let X be a random variable and c be a constant. Then, basic properties of the variance imply

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(X - c)\\ &= \mathbb{E}(X - c)^2 - (\mathbb{E}(X - c))^2\\ &\leq \mathbb{E}(X - c)^2.\end{aligned}$$

Therefore, setting $c = \Phi(y(T))$ (which is not random) and $X = \Phi(Y_N)$, we have the first inequality.

5. **Answer to the following comment:**

> *Page 12, lines 20-26 [equation (56) (unnumbered in the new version)]: I think that one must be careful here with the cases when $p < q \leq 2p$ and $2p < q$. Take the case that $2p < q$, for example. In this case we want both terms in (51) to be $\mathcal{O}(h^{4p})$, which means that M^{-1} must be $\mathcal{O}(h^{2p})$. However, the condition that M is $\mathcal{O}(h^{-2p})$ does not imply that M^{-1} is $\mathcal{O}(h^{2p})$.*

This is correct. We switched to the big-Θ notation for the two suboptimal cases, briefly defining what we mean by the big-Θ notation.

6. **Answer to the following comment:**

> *Page 16, lines 13-14 [text after (73) ((31) in the new version)]: "which, in light of (70) ((30) in the new version), satisfy $|\eta_j| \leq CH_je^{-\kappa/H_j}$ almost surely." Since the almost sure inequality on $|\eta_j|$ is important later on - e.g. it is referenced in the proof of (169) (unnumbered in the new version, proof of Lemma 8) - I would write the inequality in an equation environment and refer to the equation whenever it is used.*

We welcome the advice. The inequality is now in an equation environment (number (33)) and is referenced when employed, i.e., in the proof of Lemma 6 and in the proof of Lemma 8.

7. **Answer to the following comment:**

> *Page 16, lines 40-42 [Lemma 6]: The formulation of the lemma needs improvement for clarity and logical order. Furthermore, since the parameter $\bar{h}$ does not assume any particularly interesting values in the paper, it might be simpler to just replace $\bar{h}$ to 1, so that the reader does not need to keep track of another parameter. Moreover, in Theorem 6 h is assumed to satisfy $0 < h \leq 1$. For example: "Suppose that Assumption 1, Assumption 3 and Assumption 6 hold true, and suppose that $0 < h \leq 1$. Then the random variables η_j satisfy ..."*

We welcome the advice. We rephrased the statement of Lemma 6 as suggested, and modified its proof with $\bar{h} = 1$. Only one sentence was modified in the proof, i.e., *and because Mh can be bounded by* ~~$M\bar{h}$~~ M.

8. **Answer to the following comments:**

> *Page 16, lines 52-54 [Lemma 7]: Prior to the statement of Lemma 7, I recommend reminding the reader that the integer parameter N in Lemma 7 is not T/h.*
>
> *Page 17, lines 23-28 [Lemma 8]: The formulation of Lemma 8 needs improvement for clarity and logical order. Although it may seem repetitive to do so, the parameter N should be properly introduced in the statement of the lemma, since Lemma 8 and Lemma 7 are presented as independent assertions.*

We added a sentence before the statement of Lemma 7 and a sentence within the statement of Lemma 8 in order to clarify the use of symbols q and N in both lemmas.

9. **Answer to the following comment:**

> *Page 17, lines 40-45 [Theorem 6]: The statement of the theorem should be reformulated for clarity, e.g. "Let $0 < h \leq 1$. Suppose that Assumption 1 holds with $p \geq 3/2$, and that Assumption 3, Assumption 5, and Assumption 6 hold."*
>
> - *The statement of the theorem is unclear as to what is the difference between the numerical solution Y_n and the solution given by the RTS-RK method.*
> - *I recommend breaking up the long sentence into shorter sentences that are easier to understand.*
> - *I recommend using the full "if ... then ..." construction instead of writing "if ... there exist a constant ...", because this will make it easier for a reader to distinguish the conclusion from the hypotheses.*
> - *I recommend stating the condition that $Y_0 = y_0$.*

We thank the reviewer for these suggestions. We rewrote the statement of Theorem 6 taking into account the comments above. We hope that now the theorem statement is more readable, since the result it presents is key in the context of our work.

10. **Answer to the following comments:**

> *Page 33, lines 58-59 [(156) (unnumbered in the new version)]: Please give a more thorough explanation of why the first inequality in (156) (unnumbered in the new version) is true.*

This follows from the Lipschitz continuity of Q_{k+1}, which is smooth due to Assumption 5 and to the theory of backward error analysis. We added a comment in the text. Remark that a sign of absolute value had to be replaced with a norm.

2 Reviewer #2

1. **Answer to the following comment:**

   > *Page 14, below Eq. (62) (unnumbered in the new version – below (26)): How can Q be both a first integral for (61) ((26) in the new version) (line 14) and not be conserved exactly (line 16–19)? This is particularly confusing in light of Sect. 6. I am guessing this is only a case of confusing wording?*

   The Hamiltonian Q is a first integral for the ODE

   $$y'(t) = J^{-1}\nabla Q(y(t)), \quad y(0) = y_0,$$

   and as such it will be conserved by its exact solution $y(t)$ for $t \geq 0$, i.e., $Q(y(t)) = Q(y_0)$ for $t \geq 0$. In Section 6, we treated the case of simple first integrals (e.g., linear, quadratic, polynomial), for which it is possible to design numerical integrators whose numerical flow conserves exactly the first integral. These systems need not be Hamiltonian, but only admit a first integral. Consider now for example the pendulum case of Section 9.6. In this case, the term $\cos w$ appearing in the Hamiltonian makes it impossible (or at least very hard) to design a class of numerical integrators which conserves exactly the Hamiltonian, even though the exact solution conserves this energy. The theory of symplectic integrators gives the possibility of *almost conserving* the Hamiltonian energy numerically, and to do so over long time spans.

2. **Answer to the following comment:**

   > *In the first submission, I had troubles understanding how the long-time conservation in the Hamiltonian was achieved given the impossibility remarks in the beginning of this section. I now understand that this property only holds in expectation, but not path wise. I am of the opinion that this needs to be established in the numerical experiments section as well. My personal proposal would be to replicate Fig. 9 (Fig. 10 in the new version) only displaying step sizes $h = 0.2$ and $h = 0.05$ and plotting the Hamiltonian over time of a handful of individual sample paths. But maybe the authors find an even more suitable graphical representation.*

   Here lies the main difference between the results presented in Section 6 and the ones presented in Section 7 (i.e., mainly Theorem 6). While for "simple" first integrals, i.e., linear or quadratic, the RTS-RK method can conserve them exactly when the deterministic integrator it is built on does, for Hamiltonian systems and symplectic methods the approximate conservation is maintained in the mean. We thought this would have been clear from the statement of Theorem 6, but we decided to welcome your advice and add a commented plot to the numerical experiment presented in Section 9.6.

   We first tried to plot single trajectories but the resulting figure would have been a bit difficult to interpret. Therefore, we plot the evolution of the mean Hamiltonian ± twice the standard deviation for a family of 100 realizations, which represent an approximate confidence interval on the Hamiltonian (in time). We repeat this plot for both $h = 0.2$ and $h = 0.1$. Comments on this result ought to be found both in the caption of Fig. 9 and in Section 9.6. We hope that this clarifies our argument.

3. **Answer to the following comment:**

   > *Sect. 9.3, mean-square convergence of MC estimators: I am of the strong opinion that there should be at least one experiment showing the variability of the estimator with respect to sample size M. Maybe this could be easiest achieved by plotting multiple curves in Fig. 6 with different sample sizes. Also include error bars.*

   We took into account this comment and included an experiment which shows this property. In particular, if we pick $p = q$ in Theorem 3, the bound reduces to

   $$\operatorname{MSE}(\widehat{Z}_{N,M}) \leq Ch^{2q}\Big(1 + \frac{1}{M}\Big).$$

Therefore, the term within the parenthesis tends to 1 for $M \to \infty$, and we expect that for a fixed value of h the MSE stagnates with respect to M. This is highlighted in the rewriting of our manuscript, and a figure (Fig. 7) is added to corroborate our theoretical result.

Regarding error bars, the MSE gives already a notion of variability of the estimator. Therefore, we believe that these results are complete with respect to this perspective.

4. **Answer to the following comment:**

> *Page 5, Remark 1: I misremembered Conrad et al. to be only valid for explicit integrators. Given the same underlying integrator, I'd argue that the cost of the two methods is actually either the same (the cost of the RNG is probably negligible given the cost of the integrator) or incomparable (it is unclear whether the additive noise makes the implicit system harder or easier to solve). Maybe the editor agrees that this remark could either be removed or restated simply as "have the some computational complexity".*

We believe this is exactly the point we tried to make in Remark 1. Nevertheless, the reviewer's comment simplifies the matter of the exposition.

5. **Answer to the following comment:**

> *Page 6, Assumption 2.(ii): I am still uncertain about this point. Consider Hairer, Norsett and Wanner, Sect. II.2, Theorem 2.11. I take this result as "if the vector field is sufficiently often differentiable, the integrator is sufficiently often differentiable". Wouldn't this result suffice to guarantee 2.(ii), given a sufficiently often differentiable vector field? On the other hand: f needs to be sufficiently often differentiable anyway in order to obtain high order convergence, correct? Or have I forgotten any other corner case?*

We thank the reviewer for this comment. We took out this assumption and modified the statement and proof of Lemma 1 accordingly.

6. **Answer to the following comment:**

> *Page 8, Assumption 4: you have changed notation from y to u. Why? Consider using y for the variable name. Same for Lemma 2.*

In Assumption 4 and Theorem 1, the vector u is the "starting point" of integration. In particular, in Theorem 1 it is an arbitrary initial condition. Therefore, we believe that employing the symbol y could be confusing for the reader. Moreover, Lemma 2 is valid for any sequence u_k. In order to have a more coherent notation, we changed u_k in Lemma 2 to e_k, as we employ this result for bounding error sequences.

7. **Answer to the following comment:**

> *Page 13, Proof of Corollary 1: this is not really a consequence, but analogous, no?*

We believe that the current phrasing is correct. In Theorem 4, we state that if the base Runge–Kutta method conserves a first integral $I(y)$ which could be belonging to any function class, then so does the RTS-RK method built on the latter Runge–Kutta scheme. In Corollary 1, we specify the class of first integrals to be the class of quadratic functions. Therefore, it is a sub-result (i.e., a consequence) of Theorem 4. Corollary 1 could therefore seem superfluous, but given the importance of conservation of quadratic first integrals in literature, we believe that specifying this result explicitly is relevant.

8. **Answer to the following comment:**

> *Page 14, Eqs. (63) through (65): Is the symplecticity of an integrator a property of the integrator alone or are there integrators that are only symplectic for certain Hamiltonian systems? Is there a main argument, why (64) would be the new condition? In light of the first question: couldn't there be a case, where (64) is satisfied for a certain integrator and vector field, but not (65)?*

Thank you for this comment, this is a typo. No, symplecticity of an integrator is not a property of the integrator alone. In fact, citing from Hairer, Lubich and Wanner (Definition VI.3.1): *"A numerical one-step method is called symplectic if the one-step map $y_1 = \Phi_h(y_0)$ is symplectic whenever the method is applied to a smooth Hamiltonian system"*. In our work, this point may have been slightly unclear, and we therefore modified the text below Definition 4.

The definition of symplecticity involves the Jacobian of the flow map. If a time-stepping strategy is employed, this Jacobian is modified as h is not independent of y any more (first equation of Section 7.1). This is the case because time-stepping strategies rely on error indicators which depend on the solution itself. Therefore, even though a flow map which *would have been symplectic with a fixed time step* is employed, a time-stepping strategy could destroy the symplecticity. It is nonetheless possible to build symplecticity-preserving time-stepping strategies (i.e., functions $\tau(y, h)$). For our integrator, the time steps are variable but chosen at random. Therefore, the time step is independent of the solution and the map is symplectic.

9. **Answer to the following comment:**

> *Page 15, lines 16–38: this is introduced with little background for verification and not followed-up upon later. For me, this has generated more distraction than intuition. Consider removing these sections, in particular Eq. (66). Consider simply introducing the modified Hamiltonian (69) and refer to Hairer, Lubich and Wanner for details.*

We believe that introducing the modified Hamiltonian without writing a brief introduction about modified equations would be even more confusing. In particular, the (essential) sentence *"It is possible to prove (see e.g. [1, Section IX.8]) that for a Hamiltonian system (26) and a symplectic integrator the modified equation is still a Hamiltonian system, i.e., there exists a modified Hamiltonian $\widetilde{Q}$ defined as [...]"* would be incomprehensible for a reader who is not acquainted with the theory of backward error analysis and/or of symplectic integration. We agree that it is difficult to introduce clearly a complex topic as backward error analysis in half a page. Therefore, we added a sentence to refer an interested reader to [1, Chapter IX].

10. **Answer to the following comment:**

> *Lemma 7: may I propose to use other variable names for a_{jk}, b_j . These might get easily confused with the parameters of a RK method. Why not use directly δ, η?*

Lemma 7 is self-contained and the quantities involved are introduced in its statement. Therefore, we opt not to modify this notation.

11. **Answer to the following comment:**

> *Theorem 6: is there any example in the literature where the constants can be simply stated? It would be nice to have an example where the minimum in Eq. (83) (unnumbered in the new version) could be resolved to a concrete value.*

Given the degree of complexity of the proof, and the fact that the coefficients depend on a number of different factors, we think this would be relatively impossible. Nevertheless, in case $p \to \infty$ (i.e., p is large enough), Remark 11 gives a simple expression for the minimum.

References

[1] 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.