Reviewer report

Anonymous Referee

April 23, 2019

1 Summary

The paper presents a novel probabilistic method for the numerical solution of initial value problems in ordinary differential equations. The proposed method is to combine a geometric integrator with randomized time steps to yield a probability distribution over a fixed mesh.

The authors study and proof convergence in weak, and mean-squared sense, and analyze the convergence of a Monte Carlo implementation. The authors also analyze the proposed method with respect to geometry-preserving integration over long time integration periods. An experimental section demonstrates the empirical behavior of the asymptotic bounds as well as a motivating example in the application for Bayesian inverse problems.

My recommendation for this work is to publish it after major revisions.

This work represents a major step in the currently very active area of research of probabilistic numerics. Recent advances in this field have an established history with this journal [2, 6]. As the idea for probabilistic numerical methods has only gained momentum in recent years, there exists an urgent need of filling the missing gaps in equivalent methods for the existing literature in numerical analysis. This work proposes probabilistic solvers which maintain favorable geometric properties of the numerical integration [4] while simultaneously providing a thorough convergence analysis. Solvers of this type have thus far been missing in the probabilistic numerical tool chain. Thus, the paper adds much value to the current state of the art in probabilistic numerics.

However, some aspects of the paper need to be addressed prior to publication which warrant a major revision. These are discussed in detail below.

2 Long-term conservation of Hamiltonian Flow

From the way this section is written, it is clear that the authors are very familiar with the details of symplectic integrators. Given the publication history in this journal, e.g., by looking at previous Statistics & Computing submissions, it is reasonable to assume that the journal's audience will be more familiar with the probabilistic aspects and less familiar with the background in geometric integration.

I found it exceptionally hard to work through Section 7, and Section 7.2 in particular. To be honest, I am not confident that I understood all arguments that have been made in this section. Thus, I believe it necessary for this section to provide more detail in the derivations in order to make the manuscript a bit more self contained.

I am uncertain how exactly the authors might achieve this task and I apologize for the lack of constructive feedback. Maybe the following pointers might help:

- Page 14, after Eq. (63): is Q required to be polynomial or ∇Q?
- Page 14, around Eq. (65): the authors reference the paper by Skeel & Gear. Given the main reference throughout the rest of the text is [4], maybe the authors could use the notation thereof and point to specific sections of [4].
- Page 14: Maybe hint at the argument why $\tau(y, h)$ exists and why Θ_k exists without violating Assumption 1.
- Page 14: How can I see from (66) that (65) is satisfied?
- Page 15: Try to give an outline of the argument where κ comes from.
- Page 15: Can you give a specific reference that explains the existence of $\tilde{Q}$ in (70)?
- Page 16: How can I see that (74) can be written thusly?
- Lemma 6: I found the arguments of the proof helpful for following the text. Consider putting it in the main body.
- Proof of Theorem 6: I have not fully understood how this imposing of this terms should work in Eqs. (91) and (95).
- I have not been able to build up an intuition about Remark 10 in light of Theorem 6. How do these two statements relate to each other? What exactly am I to make of this as an user of your method?

I hope this feedback is sufficient enough for the authors to read their manuscript through the eyes of a reader unfamiliar with symplectic integrators to ease the burden of understanding.

3 Convergence of Monte Carlo for $M = 1$

This result is remarkable and warrants some further discussion.

On one hand, the whole point of a probabilistic method is to obtain a posterior distribution to quantify the numerical error. One Monte Carlo sample, however, only defines a posterior Dirac distribution, which could be understood to be nothing else than a deterministic point estimation. So, the question than turns out to be how many samples are required to obtain a characterization of the posterior in the non-limiting case for $h > 0$ fixed.

On the other hand: maybe an user is really not interested in a posterior distribution *per se*, but is required to "hedge against fixed step size risk". Maybe the authors could come up with a scenario, where a certain fixed step size would yield in catastrophic error, but a randomized step size with same expected step size does not?

Maybe the authors could try to run a sensitivity analysis on a hyper-grid (for a short time-span example) and report errors for best, worst and average choices for H_k. Or maybe the authors could run an experiment and report average errors over a grid of decreasing h and increasing M. I would also be open for the authors own preferred choice for dissecting this result in a bit more detail, but I think some discussion is required.

4 Experimenal section

I find Section 9 of numerical experiments, while exhaustive, a bit much to digest and thus stealing a bit the focus from the main points.

Sections 9.1—9.3 seem to vary only in details. I suggest to rework this as a running example for the end of each according Section 3—5. Could the authors also please report standard deviations and explain the rationale for choosing different number of Monte Carlo samples and step sizes. In particular for the mean square convergence, the chosen step sizes seem very small. Could the authors give some analysis whether on this scale the variation of the step sizes or the variation due to randomness has the bigger effect. I suggest that for each step size, the authors also run two deterministic solves with the lower and the upper bound of the uniform distribution (but without adjusting the required steps).

The experiment in Section 9.4 is interesting, but needs some more details. Were the same solvers used for RTS-RK and AN of Conrad et al. [2]? Also, there is a remaining degree-of-freedom in the AN method, the scaling of the posterior variance. The question is how much posterior variance is left once the variance ξ_k is reduced to the point where the solution does not diverge. This scaling could be found efficiently with interval bisection to give a distribution where 95% of samples do not diverge before $t = 100$. How do distributions of RTS and AN compare at different points in time?

Finally, the experiment in Section 9.7 should also compare against Conrad et al. [2]. Also, is the prior $\pi_0 = N(0, I)$ a realistic choice in this setting? From $Q(\varphi_{t_{obs}}(y_0) + \varepsilon)$, it should be possible to concentrate the prior much more strongly on values with a similar Hamiltonian, no?

5 Motivating example: a posteriori error estimator

Figure 2 is currently a source of some confusion. Embedded methods (as described in Hairer et al. [3], Wanner and Hairer [7]) typically estimate and control the *local error* $|\Psi_h(y_n) - \varphi_h(y_n)|$. This quantity seems to be compared to the *global error* $|\Psi_{hn}(y_0) - \varphi_{hn}(y_0)|$. Is this indeed what is depicted in Figure 2?

Secondly, the global errors level off on a quite high value. Could it be that this is simply the average distance between points on the strange attractor? Also, it seems a bit loaded to speak of a true error for a chaotic system.

I am of the opinion that the motivation given without this section would suffice. If the authors want to keep this section, the above points need to be addressed.

6 Minor points and questions

- The related works listed are a bit dated. Readers might benefit from a reference of recent advancements. The authors could could mention the recent review in Oates and Sullivan [5] or Cockayne et al. [1].
- Page 4, line 4: consider adding `\pageref` of Fig. 4 for convenience of the reader
- Page 4 line 20: I am not a chemistry expert. Wouldn't one speak of concentration/masses?
- Remark 1: it should probably be highlighted at this point that many symplectic integrators are implicit which needs to be considered when considering the runtime cost of numerical integration.
- Is assumption 2.(ii) ever violated?
- Why is $T = Nh$ specially mentioned in Definition 1? Also see Theorem 1 and similar.
- Please add a citation for the semi-group notation in (13).
- The transformation from (17) to (18) is a bit much for one equation. I am not sure whether I understood correctly. Please expand.
- When might Assumption 4 be violated?
- Remark 4: please add an appendix sketching the arguments for such a construction.
- The authors could reference specific Equations for Remark 7.
- Remark 8 should be compared to the results of [2].

- The authors should highlight before Theorem 3 that they mean in the case for $h \to 0$.
- Where did M^-1 go from (54) to (55)? Cf. (51)?
- System (122) might be easier to parse written as a second-order system.
- Adding a concluding section might make it easier for the reader to remember the main take-away points of the paper.

References

[1] J. Cockayne, C. Oates, T. Sullivan, and M. Girolami. Bayesian Probabilistic Numerical Methods. *ArXiv e-prints*, stat.ME 1702.03673, Feb. 2017.

[2] P. R. Conrad, M. Girolami, S. Särkkä, A. Stuart, and K. Zygalakis. Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Statistics and Computing*, 27(4):1065–1082, Jul 2017. ISSN 1573-1375. doi: 10.1007/s11222-016-9671-0. URL `https://doi.org/10.1007/s11222-016-9671-0`.

[3] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving ordinary differential equations. 1, Nonstiff problems.* Springer-Vlg, 1991.

[4] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, volume 31. Springer Science & Business Media, 2006.

[5] C. J. Oates and T. J. Sullivan. A Modern Retrospective on Probabilistic Numerics. *arXiv e-prints*, 1901.04457, Jan. 2019.

[6] M. Schober, S. Särkkä, and P. Hennig. A probabilistic model for the numerical solution of initial value problems. *Statistics and Computing*, 29(1): 99–122, Jan 2019. ISSN 1573-1375. doi: 10.1007/s11222-017-9798-7. URL `https://doi.org/10.1007/s11222-017-9798-7`.

[7] G. Wanner and E. Hairer. *Solving ordinary differential equations II.* Springer Berlin Heidelberg, 1996.