

Reviewer 1

Summary

The paper proposes a random time step Runge Kutta (RTS-RK) method and studies certain properties of this method, namely its local and global weak order, its local and global mean square convergence (and thus strong order), the mean square error of associated Monte Carlo estimators, the conservation of first integrals, the long-time conservation of Hamiltonians for symplectic dynamical systems, and the performance of the method on Bayesian inverse problems. Both theoretical and numerical results are given, using nontrivial examples that appear in the literature on symplectic integrators. The theoretical analysis overall appears to be quite thorough. The numerical experiments appear to be well-chosen from the point of view of intrinsic interest, but also for their ability to illustrate the differences between the RTS-RK method and the method proposed by Conrad et al. (2017). In some cases, e.g. the experiments discussed in Section 9.7, the choice of parameters are motivated by brief but sufficient justification and references to the literature.

General comments

- Not all choices of parameters in the numerical experiments are motivated or explained. For each choice of parameter in each numerical experiment, the authors should provide the motivations and justifications for the choices made. Otherwise the choice of parameter may appear arbitrary, thus reducing the illustrative power of each numerical experiment. The authors should formulate the statements of their assumptions, lemmas, propositions, theorems, corollaries, and remarks precisely. In particular, the dependence or independence of 'constants' on other parameters should be explicitly and clearly stated. Quantifiers should be used in the right order. See the detailed comments below.
- Whenever possible, the authors should refer to equations, lemmas, hypotheses etc. explicitly. Doing so has a few advantages. First, it may reduce the probability that the reader is confused by a certain step in a proof. Second, it makes it easier for readers to go back and check the consistency of a claim with statements that appear earlier in the paper. Third, it makes the article a more cohesive whole. The value of a proof is that it enables the reader to verify that what is claimed is true. Therefore, authors should organise and present their proofs so that the reader needs to expend as little effort as possible in order to verify that the proof is correct. This is especially true of proofs that involve tedious calculations, because one is likely to make mistakes when performing tedious calculations. The proof of Theorem 6 - which involves tedious calculations - does not satisfy the requirement described above. This is because the decomposition that occurs over (84), (85), and (86) is not presented in a way that facilitates the reader's task of verifying that the decomposition is correct. Furthermore, the estimates in (87) and (88) are presented without proofs. The authors should demonstrate that the decomposition that occurs over (84), (85), and (86) is correct by showing the intermediate decompositions. For example, the decomposition could be formulated as a lemma using arbitrary quantities, e.g. a_j instead of η_j and $b_{j,k}$ instead of $(H_j^k - h^k)\Delta_{j,k}$, so that the reader can see and verify the correctness of the decomposition, without having to do so in terms of the visually more complicated $(H_j^k - h^k)\Delta_{j,k}$. The estimates in (87) and (88) should also be given in separate lemmas, each with their own proofs.

Detailed comments

- (1) Page 1, Line 26: 'in mean-square sense' → 'in the mean-square sense'
- (2) Page 1, Lines 43-44: 'punctual value' → All instances of 'punctual value' in the article should be replaced with 'point value', because 'punctual' is used only to refer to time.

- (3) Page 2, Lines 29-30, 31-32, 36: A symbol other than σ should be chosen for the standard deviation of the Gaussian random variable ε , because σ is already used as a parameter in the first equation in (1).
- (4) Page 3, Lines 52-53: 'The most employed example to justify ... Bayesian inverse problems.' \rightarrow 'Bayesian inverse problems are most often used to justify the usefulness of probabilistic methods for differential equations.'
- (5) Page 3, Line 57: '... an inferential problem, the numerical error' \rightarrow "... an inverse problem, then the numerical error'
- (6) Page 4, Line 3-4: 'can be corrected employing' \rightarrow 'can be corrected by employing'
- (7) Page 4, Line 11-12: 'where a probabilistic methods for ODEs is presented' \rightarrow 'where a probabilistic method for ODEs is presented'
- (8) Page 4, Line 12: 'consists in perturbing' \rightarrow 'consists of perturbing'
- (9) Page 4, Line 14: 'Scaling opportunely' \rightarrow 'By appropriately scaling' All instances of 'opportunely' in the paper should be replaced with 'appropriately'. 'Opportune' has a precise meaning that does not correspond to what the authors intend.
- (10) Page 4, Line 21-22: 'is a typical physical example of such a situation' \rightarrow 'are typical physical examples.'
- (11) Page 4, Line 22-24: 'An additive random term could force the solution on the negative plane with a non-zero probability, which can become significantly big in case the magnitude of one component is small.' This sentence is unclear, because it is not clear what can become significantly big, and it is not precisely clear what the component is of.
- (12) Page 4, Line 45-46: 'i.e., employing symplectic integrators.' \rightarrow 'i.e., by employing symplectic integrators.'
- (13) Page 5, Lines 47-48, 50-51: The instances of 'punctual values' and 'opportunely' here should be replaced by 'point values' and 'appropriately' (see previous comments).
- (14) Page 6, Lines 15-23: The way Assumption 1 is currently formulated, the authors leave the possibility that the parameters p and C may depend on k . Is this intentional? If not, then Assumption 1 should be rewritten to emphasise that p and C do not depend on k .
- (15) Page 6, Line 24-25: 'The class of random variable' \rightarrow 'The class of random variables'
- (16) Page 6, Line 49-50: 'Lipschitz continuous of constant' \rightarrow 'Lipschitz continuous with constant'
- (17) Page 6, Line 51-52: In equation (11), the authors may want to state that L_Ψ does not depend on y .
- (18) Page 7, Lines 23-25, Definition 1: '... for any function Φ ... with all derivatives bounded uniformly on \mathbb{R}^d .' -From the point of view of probability theory, the class of 'test functions' that one considers is the class of bounded, continuous functions. To avoid any confusion, the authors should state very clearly whether or not the test functions that they consider are bounded. I consider 'all derivatives' to be imprecise, as certain readers may take this to mean that the test functions themselves are bounded, when in fact the authors consider unbounded test functions later, e.g. $\Phi(x) := x^\top x$. -For brevity and simplicity, perhaps the authors could define the appropriate function class in an equation and use a symbol to refer to the function class afterwards, instead of repeatedly writing '... with all derivatives bounded uniformly on \mathbb{R}^d '.
- (19) Page 7, Line 34-36: '... let us remark that ... a homogeneous Markov chain,' -Since this remark was already made in the last two sentences of page 5, I suggest the authors write '... recall that ... a homogeneous Markov chain' instead.
- (20) Page 7, Line 35-36: 'hence given $h > 0$ there exists an operator \mathcal{P}_h such that' -As a service to the reader, provide a reference to a theorem or proposition that justifies the use of 'hence'. I do not expect all readers of this article to be familiar with homogeneous Markov chains.

- (21) Page 7, Line 40-41: 'where we explicitly' → 'where we explicitly'
- (22) Page 7, Line 40-41: 'and denote in the following ... the Markov generator on the step size h .' I am aware that Conrad et al. use \mathcal{L}^h for the infinitesimal generator of the Markov chain. However, I think this notation can be improved for the present article, for the following reasons: 1) \mathcal{L} has already been used for the Lie derivative of the flow, and \mathcal{L} and \mathcal{L}^h are not related by taking 'powers'; 2) I think the infinitesimal generator should depend on the distribution H_0 from which the random time steps are drawn, and this dependence is just as important as the dependence on the deterministic step size h . If the authors wish to emphasise the dependence of the generator on h , then I think they should also emphasise its H_0 -dependence as well, at least for the first instance of the infinitesimal generator. If the authors wish to do so, they can omit the H_0 dependence in subsequent instances to simplify notation, but I think it is advantageous to the authors to remind the reader that the present setting is different from that of Conrad et al.
- (23) Page 7, Line 41-42: 'thanks to the homogeneity ...' - The phrase 'thanks to' is too informal for scientific writing. The authors should change all instances of 'thanks to' in the paper to 'due to'. Also, as a service to the reader, provide a reference to a theorem or proposition that justifies (15). I do not expect all readers of this article to be familiar with homogeneous Markov chains.
- (24) Page 7, Lines 46-51: In Lemma 1, the authors should add that $C > 0$ does not depend on the initial condition y .
- (25) -Deleted comment-
- (26) Page 8, Line 27: In equation (22), it should be observed that if the authors want the constant $C > 0$ of equation (16) to not depend on y , then the constant in the big-O term in equation (22) must not depend on y either. For this to hold, we need $\partial_{tt}\Psi_h(y)$ to be bounded by a constant that does not depend on y . We know that $\partial_t\Psi_h(y)$ is bounded by L_ψ , given Assumption 2(ii), but we do not know that $\partial_{tt}\Psi_h$ is uniformly bounded on \mathbb{R}^d .
- (27) Page 8, Lines 33-37: In Assumption 4, if \mathcal{L}^h depends on the distribution H_0 (which I think it does), then this assumption should also state something about the distribution H_0 , not just the vector field f . At the very least, I would expect 'The function f is such that ...' to be changed to 'The function f and the distribution H_0 are such that ...'. Also, although it may be obvious to the authors, I think it should be stated explicitly as a service to the reader that the positive constant L may depend on f and H_0 , but not on Φ or h .
- (28) Page 8, Lines 38-41: In Remark 3, regarding the phrase 'Given the assumptions on f and Φ above, ...' the authors should state explicitly all the assumptions on f and Φ (and H_0) that are necessary for (24) to hold. Note that in Assumption 3 of the paper of Conrad et al. (reference [6] in the present paper), **both** the analogues of (23) and (24) are stated in the assumption. In contrast, the formulation of Remark 3 of the present paper suggests that (24) follows from the assumptions for (23), and in particular with the same constant L . If this is what the authors intended, then a proof should be given. Alternatively, since neither Remark 3 nor equation (24) are used in the rest of the paper, the authors may choose to remove Remark 3.
- (29) Page 8, Lines 44-47: In Theorem 1, the authors should add that the constant C in equation (25) does not depend on the initial condition. I also think that '... with all derivatives bounded in \mathbb{R}^d ' should be changed to '... with all derivatives uniformly bounded in \mathbb{R}^d ' (so that Lemma 1 can be applied) or to use a symbol (see comment No. 18 above).
- (30) Page 8, Line 56: 'By the triangular inequality' - Replace all instances of 'triangular inequality' with 'triangle inequality'.
- (31) Page 9, Line 2-3: 'We then apply Lemma 1 to the first term and Assumption 4 to the second' - For an application of Lemma 1 to the first term to yield (28), note that the conclusion of Lemma 1 must hold with some C that is independent of the initial condition (this is not stated in the formulation of Lemma 1). Therefore, the formulation of Lemma 1 should be updated accordingly. - Assumption 4 cannot be applied to the second term under the present level of generality. This is because the authors have not assumed that the vector field f is infinitely differentiable. By the

chain rule, w_k is not infinitely differentiable. Therefore Assumption 4 cannot be applied to w_k . The authors should show that for all k , both w_k and W_k belong to the right class of functions (infinitely smooth, with all derivatives uniformly bounded in \mathbb{R}^d , etc.)

- (32) Page 9, Lines 6-11: The authors should provide detailed, step-by-step, rigorously justified steps that proceed from (28) to the first inequality in (29). Do not use vague phrases such as 'Proceeding iteratively'. Note that 'proceeding iteratively' does not yield the first inequality in (29), even after using $w_0 = W_0$. Instead, it yields

$$\sup_{u \in \mathbb{R}^d} |W_k(u) - w_k(u)| \leq Ch^{\min\{2p+1, q+1\}} \sum_{j=0}^{k-1} (1 + Lh)^j.$$

In fact, using Gronwall's inequality for nonnegative sequences yields

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

Since the upper bound of k is $N = T/h$, the exponential will increase to infinity faster than any power of h

- (33) Page 9, Line 25-26: 'is thought as an' \rightarrow 'is thought of as an'
- (34) Page 9, Line 38-44: In Definition 2, the authors should state something about the dependence of C on the initial condition $Y_0 = y_0$ in (2).
- (35) Page 9, Lines 54-60: In Lemma 2, the authors should state something about the dependence of C on the initial condition $Y_0 = y_0$ in (2)
- (36) Page 10, Lines 7-9: For clarity, the authors should specify what C_1 and C_2 are.
- (37) Page 10, Lines 17-23: In Theorem 2, the authors should state something about the dependence of C on the initial condition $Y_0 = y_0$ in (2).
- (38) Page 10, Line 27: 'and $h < 1$ ' \rightarrow 'and $0 < h < 1$ '
- (39) Page 10, lines 47-48: 'we denote by c a generic positive constant' – I suggest replacing the phrase above with constant that does not depend on [insert quantities here] whose value may change from line to line' as this is much clearer.
- (40) Page 10, lines 58-59: 'since φ_h lipschitz $(1 + ch)$ ' \rightarrow Perhaps the authors could refer to equation (37) here
- (41) Page 11, Line 2-3: 'Then we, can rewrite' \rightarrow 'Then we can rewrite'
- (42) Page 11, Lines 11-14: If the transpose notation \top will be used, then the authors would reduce the notational burden of the reader by rewriting all inner products using \top .
- (43) Page 11, Lines 27-28: 'We now apply Lemma 3'-The authors could be more specific and write that they apply inequality (38).
- (44) Page 11, Lines 33-34: The first inequality in (49) is incorrect. The exponent of h should be $\min\{p + 3/2, q + 1\}$ because $q + 1$ (see the second inequality in (46)) is strictly less than $q + 2$ (see (48)). However, the second inequality in (49) is correct.
- (45) Page 11, Lines 48-52, Remark 8:-Exactly the same observation is made in the paper of Conrad et al. (see the comment after Theorem 2.2 in the Conrad et al. paper). I recommend that the authors should inform the reader that the observation of Remark 8 in their paper is not new, by bringing the reader's attention to this part of the Conrad et al. paper
- (46) Page 11, Lines 50-52: '... are not spoiled, nonetheless getting a probabilistic interpretation ...' To preserve the formal tone in scientific writing, I suggest that this part be rewritten as '... are preserved, while yielding a probabilistic interpretation ...'.

- (47) Page 11, Lines 54-55: 'Monte Carlo estimators'-A better title for this section would be 'Mean square convergence of Monte Carlo estimators'
- (48) Page 12, Line 2-3: $Z_N = \Phi(Y_N)' \rightarrow Z_N = \mathbb{E}[\Phi(Y_N)]$ -In (51), perhaps the authors could explicitly indicate the dependence of \hat{Z}_N on M , and update all instances of \hat{Z}_N accordingly, especially in (53), (56), and (57).
- (49) Page 12, Line 7: 'where $T = hN$ is the final time, M is the number of trajectories' - T does not appear in (51) and thus this part of the sentence should be removed.-Replace 'number of trajectories' with 'number of realizations of the numerical solution' for consistency with the last part of the sentence containing the text quoted above.
- (50) Page 12, Lines 12-14: 'In the following result, we prove that this quantity converges to zero independently of the number of trajectories M .' \rightarrow 'In the following result, we prove that this quantity converges to zero independently of the number of trajectories M , in the limit $h \rightarrow 0$.'
- (51) Page 12, Lines 15-16, Theorem 3: 'the Monte Carlo estimator \hat{Z} ' \rightarrow 'the Monte Carlo estimator \hat{Z}_N '-See comment 48 above.
- (52) Page 12, Lines 29-32: As a service to the reader, provide a detailed and completely justified proof of (56). This should require only a few extra lines in addition to what is written. Demonstrating the intermediate steps will be helpful for some readers.
- (53) Page 12, Lines 42-44: 'In the sub-optimal case ... are balanced.' - The authors could provide an explicit formula for M in terms of h , p and q in the case that p 'if for any'
- (54) Page 12, Lines 59-60: 'this implies that the invariant' \rightarrow 'this implies that the first integral I '
- (55) Page 13, Lines 19-21: The theorem statement should be rewritten in such a way that the preceding terminology is used, so that readers who are not familiar with the terminology from symplectic integrators will find the statement less confusing. At the same time, the statement should also remind the reader of the fact that the numerical method (6) is random, by using the appropriate terms from probability theory. An example: 'Let $I(y)$ be a first integral of (2), and let Ψ_h be a Runge-Kutta for (2). If for all $h > 0$ Ψ_h conserves the first integral, then the numerical method (6) conserves the first integral, almost surely.' Note that the proof of Theorem 4 should also be rewritten to use 'almost surely'.
- (56) Page 13, Lines 33-34, Corollary 1: This should be rewritten to use 'almost surely'.
- (57) Page 13, Lines 36-37: 'This result is a direct consequence of Theorem 4.'-For better readability, I suggest putting this sentence in the 'proof' environment.
- (58) Page 13, Lines 53-54, 59-60 (equations (60) and (61)): For consistency with lines 25-26 above, ' S ' should be replaced with ' C '.
- (59) Page 14, Lines 5-6: 'This could produce in practice large deviations ...' \rightarrow 'In practice, this could produce large deviations ...'
- (60) Page 14, Lines 12-13: '...are the Hamiltonian systems' \rightarrow Subject-verb mismatch. Replace with 'is the class of Hamiltonian systems'.
- (61) Page 14, Lines 13-14: '... they can be written as ...' \rightarrow It would be more correct to replace the above text with '... a Hamiltonian system is defined as the solution to ...'
- (62) Page 14, Lines 29-31, Definition 4: 'A differentiable map $g : U \rightarrow \mathbb{R}^{2d}$ (where $U \subset \mathbb{R}^{2d}$ is an open set) is called symplectic ...' \rightarrow 'Let $U \subset \mathbb{R}^{2d}$ be a nonempty open set. A differentiable map $g : U \rightarrow \mathbb{R}^{2d}$ is called symplectic ...'
- (63) Page 14, Lines 44-45: '... any adaptive technique in function of a map ...' \rightarrow 'an adaptive technique in terms of a map ...'
- (64) Page 14, Lines 57-58: '... H_k is selected via a random mapping $\tau(y, h) = \tau(h) = h\Theta_k$, where Θ_k are opportunely scaled ...'-Do the authors mean that $H_k = \tau(y, h) = h\Theta_k$? If so, they should

explicitly state this. I did not find 'is selected via' to be sufficiently clear.-Replace 'opportune' with 'appropriately'

- (65) Page 15, Lines 5-6: '... local symplecticity of the flow map does not imply alone a good conservation ...'→'... local symplecticity of the flow map does not imply good conservation ...'
- (66) Page 15, Line 14: 'Our goal is obtaining a bound ... over long time.'→'Our goal is to obtain a bound ... that holds over long times.' – On a related note, I suggest reformulating the title of this section to 'Long-time conservation of Hamiltonians'
- (67) Page 15, Line 15-16: 'As stated above, ...'-It is not clear where was it stated above.
- (68) Page 15, Line 17: '... introduce the bases of this technique '→'... introduce the basis of this technique'
- (69) Page 15, Line 21: From the point of view of formal scientific writing, it is not appropriate to put '(see e.g. [?])' in the title of an assumption. It would be better to format this as 'see e.g. [?]' and to add it to the end of the preceding sentence: '... on the regularity of the ODE; see e.g. [?].' Use the `\LaTeX`command [?, Section IX.7].
- (70) Page 16, Lines 2-4, Theorem 5: '... if the numerical solution y_n ... is close enough to the initial condition'-From the point of view of formal scientific writing, it is not appropriate to put '(see e.g. [?])' in the title of the theorem. It would be better to format this as 'see e.g. [?]' and to add it to the end of the preceding sentence: 'In particular, we have the following result; see e.g. [?, Theorem IX.8.1].'-The constraint that the numerical solution 'is close enough to the initial condition' is vague. Theorem IX.8.1. of reference 9 is not very helpful: it just specifies some compact subset K of the domain of analyticity without even saying what K is. The authors should provide a more informative and specific constraint on the numerical solution.
- (71) Page 16, Lines 19-23:-Do the authors intend for (74) to be the definition of the η_j 's? If so, then they should explicitly state this, reformulate (74), and use the ':= ' notation if possible to indicate that a definition is being made, instead of an assertion of equality between two objects. – The authors should state the statistical properties of the η_j 's, e.g. whether or not they are independent with respect to certain random variables, whether or not they are identically distributed, etc. They should provide justifications for these assertions. Doing so provides a useful service to readers who wish to verify that the proof of Theorem 6 is correct. – It is not clear why the η_j 's satisfy $|\eta_j| \leq CH_j e^{-\kappa/H_j}$ almost surely. Provide an explanation.
- (72) Page 16, Lines 40-41: 'Moreover, for any $r, s > 1$ such that $r + s < R$, then for any $r, s > 1$ such that $r + s < 1$. this hypothesis should be stated explicitly. in addition, since the authors have not explicitly proof that value of C may change from line to line, inequality is strictly speaking false. problem can resolved by adding a factor 2 or stating at beginning proof.
- (73) page 17, Line 46 (second (90)): for true, we need $t_n \geq 1$. proof.
- (74) 50: 'le us impose ... fixing' – for simplicity and clarity, I recommend replace quoted text with following text: 'using (80), obtain that

$$t_n \geq \mathcal{O}(\min\{\}), \tag{1}$$

- (75) Page 19, 6, first (93): two terms inside parentheses been justified before their appearance here. are requested state justify bounds used.
- (76) 10, last help reader verify correct, write they use $t_n = nh$. an explicit bound \sqrt{r} helps understand why $\mathcal{O}(e^{-\kappa(4mh)})$
- (77) 12-13: '... length given (91). now thanks triangular inequality' → '... of length t_n for t_n given by (80). By the triangle inequality,'
- (78) Page 19, Line 16-17, second inequality of (94): More justification is needed here. Why are the first two terms on the right-hand side of the first inequality of (94) bounded by $C_4 h^q$?

- (79) Page 19, Line 27-28, Remark 9: – From the point of view of probability theory, I do not think it is appropriate to refer to $p \rightarrow \infty$ as a 'deterministic limit'. It may be better to just write 'in the limit as $p \rightarrow \infty$ '. – Explain in more detail why the coefficient M in Assumption 6 tends to 1 as $p \rightarrow \infty$.
- (80) Page 19, Line 31-33: '... which is consistent with the theory of deterministic symplectic integrators.' – Justify this assertion by citing a specific reference.
- (81) Page 19, Lines 34-37, Remark 10: To a reader unfamiliar with symplectic integrators, it is not immediately clear what is the significance of this remark with respect to the preceding results in this paper. If a connection exists, then the authors should state and explain it clearly; otherwise the remark does not serve any meaningful purpose and ought to be removed.
- (82) Page 19, Lines 38-40, Remark 11: 'We introduce ... the remainder \widehat{S}_1 .' – In the proof, the assumption is also used to simplify the terms in \widehat{S}_2
- (83) Page 19, Line 45: 'Bayesian inference inverse problems' – Suggest replacing this with 'Bayesian inverse problems' or 'Bayesian inference'.
- (84) Page 20, Lines 3-4: '... to establish a probability measure on ϑ , the posterior measure, given observations and all the prior knowledge available' \rightarrow '... to establish a probability measure on ϑ known as the 'posterior measure', given observed data and a probability measure known as the 'prior' that captures all available knowledge of the unknown parameter.'
- (85) Page 20, Lines 12-14, equation (99): Random variables should be written using upper-case letters, so equation (99) should be ' $Z = \mathcal{G}(\vartheta) + \varepsilon$ '.
- (86) Page 20, Line 18-19: '... called potential or negative log-likelihood' \rightarrow '... called the potential or negative log-likelihood'
- (87) Page 20, Lines 26-27: 'In the following, we adopt ... their probability density function.' – I disagree with the assertion that notation is commonly abused. The authors are aware of Stuart's Acta Numerica paper; in this paper there is no ambiguity between measures and their probability density functions, or more precisely, between a measure μ that is absolutely continuous with respect to another measure ν defined on the same measurable space and the corresponding Radon–Nikodym density. The authors should rewrite the affected parts of the manuscript
- (88) Page 20, Line 45: 'Convergences is proved' \rightarrow 'There, convergence is shown'
- (89) Page 20, Lines 47-49, equation (104): The expression of the Hellinger distance is incorrect, because the integral on the right-hand side is not taken with respect to the prior. See Definition 6.35 in Stuart's Acta Numerica paper.
- (90) Page 20, Line 51-52: '... can be pushed arbitrarily close to ...' \rightarrow '... can be made arbitrarily close to ...'
- (91) Page 20, Lines 59-60: '... presents a bias ...' \rightarrow '... exhibits a bias ...'
- (92) Page 20, Line 61: '... can be corrected employing ...' \rightarrow '... can be corrected by using ...'
- (93) Page 21, Lines 5-15: – The authors switch between using 'pr' or 'prob' subscripts in these lines. They should use just one subscript throughout the paper; note that the 'pr' subscript is used on page 22. – In 'The likelihood function, denoted as ... is then approximated as', replace 'approximated as' with 'defined by'. – In '...corresponding posterior distribution ... is then obtained as', replace 'obtained as' with 'defined by'.
- (94) Page 21, Line 27: 'thanks to the way' \rightarrow 'due to the manner in which'
- (95) Page 21, Line 30: 'is usually employed' – Provide evidence for this by citing some literature.
- (96) Page 21, Line 35-36: 'In the limited case of linear problems and Gaussian prior ... the posterior distributions' – It may not be sufficiently clear to readers who are not familiar with Bayesian inverse what 'linear problems' means. Furthermore, the adjective 'limited' is not useful. I suggest

writing 'If the forward operator \mathcal{G} is linear, the prior on the unknown parameter is Gaussian, and the negative-log likelihood is given by (101), then there is an explicit formula for the corresponding posterior distribution.'

- (97) Page 21, Section 8.1: To make it explicitly clear how this section is connected to the preceding material (in particular, the introduction to Bayesian inverse problems), I recommend that the authors do the following: 1) Write ' $Z = \varphi_h(y_0^*) + \epsilon$ ' instead of ' $z = \varphi_h(y_0^*) + \epsilon$ '. Upper case letters should be used for random variables and lower case letters should be used for deterministic values. 2) Specify the parameter space Θ , observable, and linear forward operator. 3) As far as possible, use the same symbols as used in (99), (100), (101)
- (98) Page 21, Section 8.1: The symbol ' \rightarrow ' and the phrase 'tends to' are used to suggest convergence. However, the authors do not specify the type of convergence. Thus, the statements should be considered only at an imprecise, heuristic level, and not at the level of rigorous mathematics. The authors should notify the reader of this.
- (99) Page 22, Section 9: The authors should specify all the parameters that they use in their experiments. For example, in some subsections below the value of the parameter p is given, while in other subsections it is not.
- (100) Page 23, Figure 4: It is difficult to distinguish the curves corresponding to different values of σ . I recommend that the authors reduce the range of values of y_0 in each subplot to make it easier for the reader to distinguish the curves for different values of σ
- (101) Page 23, Line 34-35: 'FitzHug–Nagumo' \rightarrow 'FitzHugh–Nagumo'
- (102) Page 23, Line 42-43: 'with an high order' \rightarrow Replace all instances of 'an high order' in the paper with 'a high order'
- (103) Page 24, Table 2: the sub-table for the explicit trapezoidal method should contain as many columns as the number of values of h that were used.
- (104) Page 24, Lines 29-32: 'is defined as $\varphi(x) = x^\top x$.' \rightarrow φ should be replaced by Φ . Different notation should be used to distinguish definitions from equations. Also, Theorem 1 gives weak order when the function Φ is evaluated at the state of the numerical solution for an arbitrary time point. The authors should state at which time point they evaluated the function for the results of Table 2.
- (105) Page 24, Lines 36-37: 'Monte Carlo estimator' – A better title would be 'Mean-square convergence of Monte Carlo estimators'.
- (106) Page 25, Lines 10-11: 'are intermediates results' \rightarrow 'are intermediate results'
- (107) Page 25, Lines 25-26: 'exhibits a chaotic behaviour. In particular, at long time the trajectories are captured in a strange attractor' \rightarrow 'exhibits chaotic behaviour. In particular, in the long-time limit the trajectories lie on a strange attractor'
- (108) Page 25, Line 32-33: 'can be selected (see for example Example 1) ... is zero' \rightarrow 'can be selected ... is zero; see e.g. Example 1.'
- (109) Page 25, Line 39-40: 'We choose $h = 0.05$ as the mean of uniformly distributed time steps' – What value of p was used?
- (110) Page 26, Figure 6: Each of the two subplots should be given their own titles.
- (111) Page 26, Lines 23-24: 'We assume the perturbation ' \rightarrow ' We set the perturbation'
- (112) Page 26, Line 35: 'midpoint rule conserves also quadratic first integrals' \rightarrow 'midpoint rule also conserves quadratic first integrals'
- (113) Page 26, Line 36: 'We therefore integrate (122) with uniformly distributed mean time step' \rightarrow 'We integrated (122) with uniformly distributed mean time step'. – Specify the value of p
- (114) Page 26, Lines 44-45: 'Hamiltonian systems' – A better title for this section would be 'Conservation of Hamiltonians'.

- (115) Page 26, Lines 52-52: 'and then it grows proportionally to the square root of time' – The reader would benefit more if the authors made this observation immediately after Theorem 6. – The authors should explain how they obtain the 'square root of time' observation.
- (116) Page 27, Figure 7:-Each of the subplots should be given their own titles. – I suggest that the error subplot be put in a separate figure environment, in order to avoid spacing issues between the x -axis label of the top left subplot and the y -axis label of the bottom subplot. – The caption for the bottom subplot should specify that $I(\cdot, \cdot)$ is defined in (124).
- (117) Page 27, Lines 42-43: 'presents a chaotic behaviour' → 'exhibits chaotic behaviour'
- (118) Page 27, Lines 43-44: 'through a single observations' → 'through a single observation'133. Page 27, Lines 44-45: 'Noise is then assumed to be' → 'Noise is set to be'-Some readers may assess that the variance of the Gaussian random variable is excessively small (on the order of 10^{-8}) relative to the initial condition. The authors should explain why they chose such a small variance.
- (119) Page 27, Lines 32-33: 'Bayesian inferential problems' – A better title would be 'Concentration of posteriors in Bayesian inference'.135. Page 28, Figure 8: If the authors wish to keep the x -axis label as ' t ', then the y -axis label should be $\mathbb{E}|Q(Y(t)) - Q(y_0)|$.
- (120) Page 28, Lines 35-36: 'in order to observe convergence' → 'in order to study whether the approximate posteriors concentrate'
- (121) Page 28, Lines 40-41: 'when integrating in time the dynamical system' → 'when integrating the dynamical system forward in time'
- (122) Page 28, Lines 41-42: 'Hence, initial conditions with a different energy level with respect to the observation are endowed with a high value of likelihood' – The authors should explicitly state the likelihood that they use and clearly explain how the energy level and the likelihood are connected.
- (123) Page 29, Lines 35-36: 'Under Assumption 6'-The authors could be more specific and write 'Under Assumption 6, we have that $H_j \leq Mh$ almost surely; hence'
- (124) Page 29, Line 46: 'on both sides and as' → 'on both sides and using'
- (125) Page 29, Line 50: 'which proves the desired as $e^{-r\kappa/x}$ is a growing function of x ' → 'which proves the desired inequality. This is because Assumption 6 implies that $M \geq 1$, and because $e^{-r\kappa/x}$ is a growing function of x , so that $e^{-r\kappa/h} \leq e^{-r\kappa/(Mh)}$ '. – The authors need to bound the Mh term by a constant that does not depend on h .
- (126) Page 29, Lines 56-58, equation (134): 'almost surely' should be added.
- (127) Page 29, Line 59: 'which implies the desired result proceeding as above' → 'which implies the desired result by proceeding as above.'
- (128) Page 29, Proof of Lemma 6: The authors only consider the case where $H_j \geq h$. They should discuss the case where H_j 'in the mean-square sense'