Reponse to Editor and Reviewer Comments

To the Editor and Reviewers:

I appreciate the thoughtful comments and recommendations made by the AE and Reviewers and the opportunity to revise my manuscript "A quantum parallel Markov chain Monte Carlo" (JCGS-22-275). I have carefully considered these extremely helpful suggestions and have addressed them to the best of my abilities. I believe that the manuscript is now significantly improved and am thankful for the time and energy put forth by the Editors and Reviewers.

In particular, I have:

- clarified how the present work compares to other quantum Monte Carlo methods;
- expanded the scope of the methods to include discrete target distributions (Section **??**);
- included an application to Bayesian image analysis (Section **??**);
- included a short discussion around challenges reading classical data into quantum machines; and
- responded to all other Reviewer/AE questions and comments.

The document below contains my replies (in **bold**) alongside the AE and Reviewers' comments (normal text). I list changes to the text with indented normal text set off by quotation marks.

With sincere gratitude,

Andrew Holbrook

Editor and Reviewer Comments:

Editor

Dear Andrew,

Two reviews of your manuscript are somewhat split, but they also focus on different aspects of your paper. So the paper cannot be accepted. Rev 2 addresses the mathematical dimension of your study (as well as the awkward juxtaposition of the mathematical presentation and the data example); Rev 1 seems to focus on details of the covid data but also has some queries about the methods. The paper is well outside anything to do with my expertise, but I can certainly identify with the general points of Rev 2, especially the absence of demonstrated method performance on data where you know the answer.

If you think you can revise to overcome reviewer concerns, you are welcome to do so. Any revision will be subjected to re-review with no prejudice toward acceptance. I think that both reviewers are highly appropriate, and given that their expertise far exceeds mine, I cannot apprise you of whether a successful revision is feasible.

Jim

Reviewer 1

Phylogeography inference has been a very difficult problem in the sense of the tree size and the dimen- sionality of states. However, in this manuscript, the authors have shown that, based on the CTMC model, the first-order approximation of the gradient towards the rate matrix can improve the computational efficiency with acceptably bounded errors, which can be hugely confined with correction. The method is applied to SARS-CoV-2 transmission to infer the traveling factors across 33 regions, framed in a mixed-effects model. I am not an expert in Graph Theory and Random Matrix Theory but the manuscript is clear enough for me to understand. I, therefore, am strongly recommending for publication in PNAS, while there are still a few issues that the authors need to be addressed properly:

1. The authors spend 3 out of 7 pages in the main text and three corresponding sections in

the Appendix discussing the refined error and the proof, while postpone the application with the affine-corrected error to future work. I would expect more than a relatively abstract error comparison figure for this issue. For example, a simple numerical simulation similar to the last paragraph in Application to compute the posterior likelihood ratio or posterior density ratio of the corrected and the uncorrected methods, because the authors eventually applied the method within a Bayesian framework.

2. I am curious about the selection of the extra two additional Chinese provinces and 11 additional countries when fitting the mixed-effects model and how the inclusion of the extra regions improve the fitting. Is there any criteria to determine the inclusion?

3. For the mixed-effects model, do the authors assume independent predictors for the fixed effects? It seems to me that, in Fig. 2, the air traffic proximity and the intracontinental proximity are highly correlated, which can be the reason why 2 is not statistically significant. Then what's the justification for the predictor selection?

4. Am I correct that the total number of particles (i.e., samples) in the MCMC framework is 2,000? The ESS for is extremely small, do the authors have any explanations?

5. What does the random effect matrix look like? The authors indicate that air traffic proximity is the only significant fixed-effect predict, then I would wonder how much does the random effect contribute to the model?

6. The ranges for predictor matrices and posterior mean rate matrix are different. However, the color palettes are identical, which makes it extremely hard to compare the impact of each predictor. For example, I cannot see how the sentence "... but one may also see the influence of the Hubei asymmetry in, e.g., the squares corresponding to travel between Guangdong and Hubei provinces." makes sense visually from Fig. 2 and 3. It also doesn't make sense to me conceptually, because air traffic proximity is the only significant predictor.

7. The color palette in Fig. 4 makes it really hard to distinguish Hubei, other provinces in China, and other countries. It will be great if the authors can highlight Hubei in one color, other provinces in another color with gradients, and other countries in a third color with gradients. And why there are only 29 regions?

8. The proposed method is able to include higher-dimension rate matrix compared with Lemey et al. (2009, 2014) [1, 2]. Can the authors then have an external comparison to the results in **?** just visually to see whether the color pattern in the phylogenetic tree is matched?

Reviewer 2

This is a review for the manuscript "On the surprising effectiveness of a simple matrix exponential derivative approximation, with application to global SARS-CoV-2". For the benefit of the authors on how they interpret my review, I am a theoretical/computational statistician with background on theoretical population genetics, interested in Markov chains and its implementations in Monte Carlo methods.

Before content-specific comments let me first clarify that there are issues with the numbering of theorems and remarks in the manuscript that I have received (e.g., theorem 2 is missing, and there are two remarks labeled as Remark 2), and below I stick to the numbering used in the manuscript that I have received for review to avoid any confusion.

The main idea in the manuscript revolves around the three theorems presented and how the approximations provided by these theorems improve computational efficiency in large matrices involved in the continuous time markov chains. The positives are: 1) The theorems are stated well as in solid applied mathematical work, 2) Their proofs seem to hold as far as I can tell. To me it is clear that if there are any gaps or minor incorrect statements they can be remedied with little effort and the theorems and corollaries will hold. It is also clear that a lot of effort and thought was put into these theorems and corollaries and I appreciate that, along with the clear exposition which made following the mathematical reasoning easy.

Unfortunately, the manuscript did not sit well with me, so to the negatives. There are two major negatives for me. The second item is the more important one, but the first motivates the second, and it deeply troubled me.

1) I think the way the manuscript is positioned with respect to the interpretation of the scope of the work is inaccurate. After a first read of the manuscript, I was left with the impression that the work was originally conceived and intended as an applied mathematics or a stochastic process paper, which –for example– with the omission of opening paragraph of the manuscript would stop before section 2. Presented this way, I think its scope is ideal for a specific technical subject-matter journal (e.g., applied math, stochastic process, statistics). With the inclusion of the first paragraph (which I do not find clear, see below specific comments on this), and section 2, I felt that there was an effort to make the paper appealing to a broader audience. I think this effort falls short of its aim and the integration of theoretical results with data looks ad-hoc. In particular, the example data set chosen to demonstrate the results does not have much information about the parameters, and it is as

if it was included just because its dimensionality is large. This tells us nothing about the quality of inference (see 2 below) and also misleads us to believe that the approximation method presented is successful as a statistical computational –in the sense of methods that facilitate inference but they are not inferential methods themselves– tool. I note that the example consists of a real data set in which we do not know whether the assumed probability model generating the data is true, let alone the true parameter values of that model. Even if there was a signal in the data, I would still be hesitant to accept the approximation method presented as a valid statistical computational tool in the context of MCMC, before it was well established with some simulation studies (again, see 2 below). Therefore, there is no way to assess the quality of inference about parameters with this example. The only use of this data set can be in supporting the claim that the approximation presented through the theorems provides a computational advantage at higher dimensions. But then, why use SARS-Cov-2 data for this goal, whereas it can be better established by a simulated data set?

2) To me, a huge barrier to accept that the approximations presented can be successful in aiding statistical inference is the lack of simulation studies to assess the quality of inference. (This is not about the type of simulations shown in Figure 1 as detailed below). For a work like this, before applying to real data, the first test would be to design a large simulation study, where the model and the parameters are known and to test whether the application of the approximation method when implemented within the Hamiltonian Monte Carlo has good statistical inference performance. MCMC makes its own approximations (mixing, convergence to stationarity, etc..) and errors introduced by an approximative method injected to an MCMC sampling might be exacerbated. Even the statistical performance of well-known MCMC algorithms can vary with different signal strengths (e.g., magnitude of the parameters relative to each other in a model), as well as the model size. It would be premature to think that a method is valid based solely on its computational performance (as it is presented in the manuscript) without first assessing its statistical performance of recovering an approximate joint posterior distribution of parameters and it passes rigorous tests. A way to check this is by simulating data from target distributions of interest under a variety of parameter value combinations and then apply the method to see whether the posterior distribution of parameters are recovered successfully (in the case of Bayesian inference for this manuscript). I am surprised that the authors did not see this as necessary especially because they are working with large mixed effect models where things can go awry pretty quickly. As an example, I would have liked to see a comparison of the statistical performance in say, a 3 x 3 simulation experiment design. Variable 1, The method with 3 levels: Standard Hamiltonian Monte Carlo, the naive method of reference (18), and the method of the

manuscript; Variable 2, The model size with 3 levels: Small, Intermediate, Large; Variable 3, The number of parameters with considerable signal with 3 levels: 1 parameter, half of the parameters, all parameters in the model.

Specific Comments:

* Please check numbering of theorems and remarks.

* Regarding the opening paragraph. The first sentence did not make sense to me. What is big data (why is it in double quotes, anyways?), what is modeling schemes, what is structures. The whole paragraph is written vaguely and I felt dismissive of the applications of the work being presented.

* Figure 1 is designed to show how the approximation behaves beyond the assumptions of theorems and corollaries, and support the claim that the results apply more generally by numerical evidence. I am not convinced that this constitutes strong evidence (as it is claimed in the manuscript), however. My feeling is that the claim holds, but it needs to be demonstrated with distributions that are not so docile. I think that deviations from the assumptions of theorems and corollary are not so large for cases a, b, and c. For Figure 1c, I have a specific comment: I am not sure about what the authors mean by "Cauchy entries truncated to be positive", but my understanding is that this is a folded Cauchy distribution at zero with the support of the probability density function defined on the non-negative real line (if this is not correct please add a sentence to describe). Cauchy distribution is a member of the subexponential family and I wonder whether the tail properties of the folded Cauchy are distinct from those of the subexponential family (I do not have an answer to this, so I suggest that the authors check and provide a brief explanation).

* Equations 12-19 seems to follow straightforwardly from the section where equation 2-7 are presented. I agree that writing them explicitly is better, but I felt that they disrupt the flow of the section "Deterministic bounds on approximation error in time". Please consider emulating them to somewhere in 2-7 or following, intermediate steps in 12-19 maybe put in an appendix.

* In presenting the subexponential families in formulae 40-41, notation "exp" is used instead of "e". Please change to "e" to be consistent with the rest of the manuscript. Also, I think this definition comes from Vershynin. There might be some regularity conditions attached to this specific definition, although I cannot remember exactly. Please check. There is a more fundamental definition of subexponential families using the cumulative distribution

functions, maybe that serves better here. Your choice of course.

* Remark 6. The sentence starting with "In general..." is not clear in which sense it is general. "Proxy" is also not clear.

* The paragraph starting with "In addition..." on page 7 is too short for the reader to understand the implications of the analysis. Please consider expanding it.

* Appendix D. Surrogate-trajectory Hamiltonian Monte Carlo is described quite detached from the work presented in the manuscript. Please consider describing it in the context of the work.

* For completeness, it is worth restating that Q is symmetric in Theorem 3. (Or is it not?)

* Remark 2 (iii). "eigenvalue" should be "eigenvalues".

* References (5) is a book, but there is no section or page.

References

Lemey, P., A. Rambaut, T. Bedford, N. Faria, F. Bielejec, G. Baele, C. A. Russell, D. J. Smith, O. G. Pybus, D. Brockmann, et al. 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza h3n2. PLoS pathogens 10:e1003932.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. PLoS computational biology 5:e1000520.