Tuning Hamiltonian Monte Carlo

Jinji Li and Kristopher A. Innanen

ABSTRACT

Tuning the Hamiltonian Monte Carlo (HMC) stands as a crucial endeavor, especially within the realm of geophysical optimization and inversion where the problems are complex and high-dimensional. In the context of this report, we offer an in-depth examination of the functionalities associated with the tunable parameters, such as the integration length, time step, and mass matrix. The minimum-model sampling experiment has shown that assigning the second-order information to the mass matrix can help the sampler reach the target distribution more efficiently, and the HMC-FWI experiment has shown that an adaptive tuning strategy can significantly accelerate the convergence and thus produce more precise solutions within limited sampling attempts. This work contributes to a deeper understanding of the role of tuning in HMC applications and can serve as a valuable reference for future utilization.

INTRODUCTION

Hamiltonian Monte Carlo (HMC) is a powerful and sophisticated Markov chain Monte Carlo (MCMC) algorithm that has gained prominence in the field of Bayesian statistics and probabilistic machine learning. Developed by Duane et al. (1987) in physics, and used in statistics since Neal (1993). HMC offers a more efficient alternative to traditional MCMC methods like the random-walk Metropolis-Hastings algorithm. At its core, HMC leverages concepts from classical mechanics to explore high-dimensional parameter spaces, making it particularly well-suited for complex Bayesian inference problems (Brooks et al., 2011; Beskos et al., 2013; Mangoubi and Smith, 2017).

The key innovation behind Hamiltonian Monte Carlo is its use of Hamiltonian dynamics, a branch of physics that describes the evolution of particles in a conservative system (Hamilton, 1834). In the context of Bayesian inference, the parameters of a probabilistic model are treated as particles in a high-dimensional space, and the posterior distribution over these parameters is modeled as a potential energy landscape. By simulating the trajectories of these particles through the parameter space using Hamiltonian dynamics, HMC can efficiently explore complex posterior distributions, leading to faster convergence and reduced correlation between samples compared to traditional MCMC methods (Brooks et al., 2011; Fichtner et al., 2018).

However, the power of HMC comes at a cost: it is highly dependent on several tuning parameters, and its performance can be sensitive to these choices (Beskos et al., 2013; Homan and Gelman, 2014; Fichtner et al., 2021; Hoffman and Sountsov, 2022). In practice, selecting appropriate values for parameters such as the step size and the number of leapfrog steps can be a non-trivial task. Poorly tuned HMC algorithms may lead to inefficient exploration of the parameter space, causing slow convergence, biased estimates, and unreliable results. Therefore, understanding the necessity of tuning HMC is paramount to harness its full potential and extract accurate inferences from complex probabilistic models.

In the following sections, we will delve into the crucial aspects of HMC tuning, exploring how these parameters impact the algorithm's performance and the strategies to fine-tune them effectively. In the linear problem, we can analytically observe how variations in step size and the number of leapfrog steps directly impact the exploration of the parameter space. The insights gained from this simplified context will serve as a foundation for addressing the intricacies and challenges posed by complex problems like FWI, where posterior distributions can be highly multi-modal, high-dimensional, and challenging to navigate.

THEORY

Integrating the Hamiltonian dynamics

The Hamiltonian dynamics (Hamilton, 1834) can be described by a Hamiltonian equation $H(\mathbf{p}, \mathbf{q})$, such that

$$H(\mathbf{p}, \mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}), \tag{1}$$

where $U(\mathbf{q})$ denotes the potential energy related to the position vector $\mathbf{q}$, and $K(\mathbf{p}) = \mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}/2$ is the kinetic energy represented by the mass matrix $\mathbf{M}$ and momentum vector $\mathbf{p}$ of the mechanic system.

The trajectory of this N-dimensional artificial particle is determined by Hamilton's equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} = \left[\mathbf{M}^{-1}\mathbf{p}\right]_i, \tag{2}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i}, \tag{3}$$

where $i = 1, ..., N$.

In most of the fusion of HMC and Bayesian inference approaches, the potential energy can be regarded as the misfit functions, and the gravitational force is parallel to its gradient. The kinetic energy determines the maximum admissible misfit increase, and the momentum $\mathbf{p}$ serves as a directional tolerance that determines initial directions along which the particle will move towards any alternative combinations in the model space.

The leapfrog approach (ISERLES, 1986), as one of the most prominent integration methods for simulating the Hamiltonian dynamics, is extensively applied. Basically, for a fixed time step δt, the leapfrog integration scheme is in the following form:

(1) $\mathbf{p}\left(t + \frac{\delta t}{2}\right) = \mathbf{p}(t) - \frac{1}{2}\delta t \left[\frac{\partial U}{\partial \mathbf{m}}|_t\right]$.

(2) $\mathbf{m}(t + \delta t) = \mathbf{m}(t) + \delta t \left[\frac{\partial K}{\partial \mathbf{p}}|_{\frac{\delta t}{2}}\right]$.

(3) $\mathbf{p}(t + \delta t) = \mathbf{p}\left(t + \frac{\delta t}{2}\right) - \frac{1}{2}\delta t \left[\frac{\partial U}{\partial \mathbf{m}}|_{t+\delta t}\right]$.

Note that step (2) is repeated when there is a certain integration length L.

Given its preeminent status for HMC-related problems, it is essential to discuss the parameters in the leapfrog integration, it becomes imperative to delve into the key parameters governing the leapfrog integration technique. Specifically, we must scrutinize the tunable parameters δt and L that directly influence the efficacy of the algorithm. Additionally, the selection of the mass matrix $\mathbf{M}$ assumes a significant role, impacting both the efficiency and stability of HMC in many scenarios.

Time step δt

The choice of the time step, often denoted as δt, in the leapfrog method is a critical parameter that significantly influences the behavior and performance of HMC. An inappropriate choice of time step can lead to biased or inaccurate samples, and too large a step size might result in poor exploration of the target distribution. Conversely, overly small step sizes, especially when the trajectory length is small, can lead to slow convergence and increased computational costs (Brooks et al., 2011). Fichtner et al. (2021) also indicated that the energy conservation is relevant to the first order in δt.

Moreover, selecting an overly large time step can lead to numerical instability, causing divergent trajectories or unreliable results. It can be illustrated by a 1-dimensional problem when the mass matrix is set to 1, and the potential energy is represented by the minimum model:

$$H\left(q,p\right)=\frac{q^2}{2\sigma}+\frac{p^2}{2}, \tag{4}$$

where σ is an arbitrary standard deviation to define the Gaussian distribution. For one iteration in HMC, the mapping from the current moment t to the next step $t+\delta t$ can be written in the form below:

$$\begin{bmatrix} q\left(t+\delta t\right) \\ p\left(t+\delta t\right) \end{bmatrix} = \begin{bmatrix} 1-\frac{\delta t^2}{2\sigma^2} & \delta t \\ -\frac{\delta t^2}{\sigma^2}+\frac{\delta t^3}{4\delta t^4} & 1-\frac{\delta t^2}{2\sigma^2} \end{bmatrix} \begin{bmatrix} q\left(t\right) \\ p\left(t\right) \end{bmatrix}. \tag{5}$$

The eigenvalues of the coefficient matrix in a linear system of ODEs are indicative of the system's stability. In this example, the magnitudes of the eigenvalues are:

$$\left(1-\frac{\delta t^2}{2\sigma^2}\right)\pm\sqrt{\frac{\delta t^2}{4\sigma^2}-1}. \tag{6}$$

If the fraction ${}^{\delta t}/_{\sigma}<2$, the eigenvalues are complex, thus making the trajectories stable. In contrast, δt with that fraction to be greater than 2 will lead the trajectory diverge to infinity. Propagating this concept to multidimensional problems, the stepsize might be determined by the square root of the smallest eigenvalue of the model covariance matrix.

Additionally, when the chosen δt exceeds the stability limit for a specific region, the trajectory is unlikely to visit that region, even if it contains substantial probability. This is because, if the trajectory ever enters a region where the chosen of stepsize would produce

instability, it will dwell there for an extended period due to a very low acceptance probability. While HMC maintains the correct distribution overall, it may rarely transition to regions with different stability requirements from those it initially explores. Thus, a random selection of δt from some distribution that also holds the stability could be desirable.

Another consideration upon the choice of stepsize is the acceptance rate of a proposal made by HMC. Previous research, such as Kennedy and Pendleton (2000); Brooks et al. (2011); Beskos et al. (2013) has shown that, the optimal acceptance rate for HMC is around 65%. Consequently, a strategic choice of the time step can be made, taking into account both stability and the desired acceptance rate. While ensuring numerical stability, as indicated by the criterion involving the fraction ${}^{\delta t}/_{\sigma}$, δt can also be adjusted to align with the target acceptance rate. Specifically, δt is decreased when the acceptance rate falls below 65% and increased when the rate surpasses 85%. These two acceptance rate values have been empirically identified as practical benchmarks by Fichtner et al. (2021).

Integration length L

The total length of the trajectory represents the number of leapfrog steps taken during the HMC simulation and impacts the acceptance probability of proposed states and the efficiency of the sampling process. Setting an appropriate value for L is essential for maintaining a reasonable acceptance rate and avoiding excessive computational overhead.

Typically, a longer trajectory explores a larger portion of the parameter space, potentially leading to more thorough exploration of the posterior distribution. In addition, HMC relies on achieving ergodicity, which means the algorithm should be capable of exploring the entire posterior distribution over time. Longer trajectories increase the likelihood of achieving a more representative sample of the posterior. However, when trajectories return to the same position coordinate after a particular number of steps, HMC can be non-ergodic. Setting σ in equation 4 to be 1, and the following analytical form of $q\left(t\right)$ and $p\left(t\right)$ can be derived as:

$$q\left(t\right) = cos\left(t\right), p\left(t\right) = -sin\left(t\right). \tag{7}$$

This means that the Hamiltonian trajectory is clockwise around the origin point. In such a case, HMC may suffer from prolonged exploration times. To mitigate this potential issue, a practical approach involves randomizing the selection of L, often drawn from a modest interval. This measure helps maintain ergodicity by preventing the algorithm from becoming trapped in periodic behaviors.

However, in high-dimensional problems where visualizing the distribution is challenging and periodicity may not be apparent, analogous issues can arise. To tackle these complexities, a strategic approach involves adapting the integration length by examining the autocorrelation among samples. This method adopts a trial-and-error tuning strategy, commencing with an initial estimate of the number of leapfrog steps, running several samples, and assessing the correlation between these samples. If the preliminary runs yield lower correlation coefficients, it signifies the need to increase the integration length accordingly (Brooks et al., 2011). This approach can be applied dynamically in an on-the-fly fashion. By evaluating the coefficients of a subset of samples and referencing them, the integra-

tion length can be flexibly adjusted to optimize the efficiency and effectiveness of HMC in high-dimensional spaces.

Mass matrix M

The mass matrix M plays a role analogous to mass in classical physics. It introduces inertia into the Hamiltonian dynamics, governing how the momentum variables interact with the position variables. Analytically, the mass matrix's elements influence the curvature of the energy landscape and the characteristic time scales of the system. Additionally, the mass matrix can be thought of as a scaling factor for the kinetic energy term in the Hamiltonian. In some Bayesian modeling contexts, the mass matrix is related to the covariance structure of the parameters (Fichtner et al., 2018).

A plausible preconditioning of M can be inspired by the Hessian matrix $\mathbf{H}$ of the misfit function (Fichtner et al., 2021). This can be derived by substituting step (1) into step (2) in the leapfrog scheme:

$$\mathbf{m}\left(t+\delta t\right)=\mathbf{m}\left(t\right)+\delta t\mathbf{M}^{-1}\left[\mathbf{p}\left(t\right)-\frac{1}{2}\delta t\frac{\partial U\left(t\right)}{\partial\mathbf{m}}\right]. \tag{8}$$

Adding equation 8 with its backpropagated version yields:

$$\mathbf{m}\left(t+\delta t\right)=2\mathbf{m}\left(t\right)-\mathbf{m}\left(t-\delta t\right)-\delta t^{2}\mathbf{M}^{-1}\frac{\partial U\left(t\right)}{\partial\mathbf{m}}. \tag{9}$$

By assuming an error term that is continuously propagated in the integration process, and using the first-order approximation ${}^{\partial U\left[\mathbf{m}\left(t\right)+\delta\mathbf{m}\left(t\right)\right]}/_{\partial\mathbf{m}}\approx\mathbf{H}\left(t\right)\delta\mathbf{m}\left(t\right)$, the central finite-difference equation that also takes the current time step t and the next time step $t+\delta t$ into account is:

$$\begin{aligned}\mathbf{m}\left(t+\delta t\right)+\delta\mathbf{m}\left(t+\delta t\right)=2\mathbf{m}\left(t\right)+2\delta\mathbf{m}\left(t\right)-\mathbf{m}\left(t-\delta t\right)-\delta\mathbf{m}\left(t-\delta t\right)\\ -\delta t^{2}\mathbf{M}^{-1}\frac{\partial U\left(t\right)}{\partial\mathbf{m}}-\delta t^{2}\mathbf{M}^{-1}\mathbf{H}\left(t\right)\delta\mathbf{m}\left(t\right).\end{aligned} \tag{10}$$

Consequently, the propagation of errors in HMC predominantly hinges on the second-order information inherent in the potential energy landscape. Notably, various studies, including those by Zhang and Sutton (2011) and Fichtner et al. (2018, 2021), have demonstrated that the incorporation of the Hessian matrix, even when using approximated versions in many instances, can significantly enhance the convergence speed of the HMC process. This underscores the value of leveraging second-order information to streamline and expedite the HMC algorithm's performance.

NUMERICAL EXPERIMENTS

In this section, we present two numerical experiments. The first experiment addresses a sampling problem involving a 100-dimensional distribution, with varying standard deviations. The second experiment is centered around fine-tuning the parameters of HMC-FWI. In the first experiment, our primary focus is on investigating the enhancements achieved through the incorporation of second-order information. In the second experiment, our objective is to analyze the impact of an adaptive tuning strategy applied to both trajectory length and integration step.

100D distribution problem

The model space can be described as a 100-dimensional multivariate Gaussian distribution with mean values set to zero. The standard deviations within this distribution vary according to the value of n, which ranges from 1 to 100. The variables in this distribution are considered to be independent. This framework leads to the potential energy (representing the posterior distribution) being expressed as the minimum model: $\frac{1}{2}\mathbf{m}^T\mathbf{C}_\mathbf{M}^{-1}\mathbf{m}$. We initially assume an identical mass matrix and implement HMC in the first test. Subsequently, we transition to using the inverse covariance matrix as the mass matrix, and we analyze the effects of this change in the second phase of the experiment.

We collect a total of 1000 samples, and for propagating the HMC, we employ a leapfrog step size L drawn from a uniform distribution between 15 and 30, with the time step δt being randomly selected from the range $(0.0104, 0.0156)$. The choice of δt is consistent with the methodology outlined in Brooks et al. (2011). It's worth noting that identical random seeds and correspondingly generated initial models are utilized in both phases of this experiment.

To illustrate the extent to which the Hessian matrix can enhance the sampling process, which also relates to the concept of convergence, we visualize the first 20 walker trajectories projected onto the moderate and least constrained dimensions, namely the $50th$ and $100th$ dimensions. These trajectories are presented in Figure 1 below.

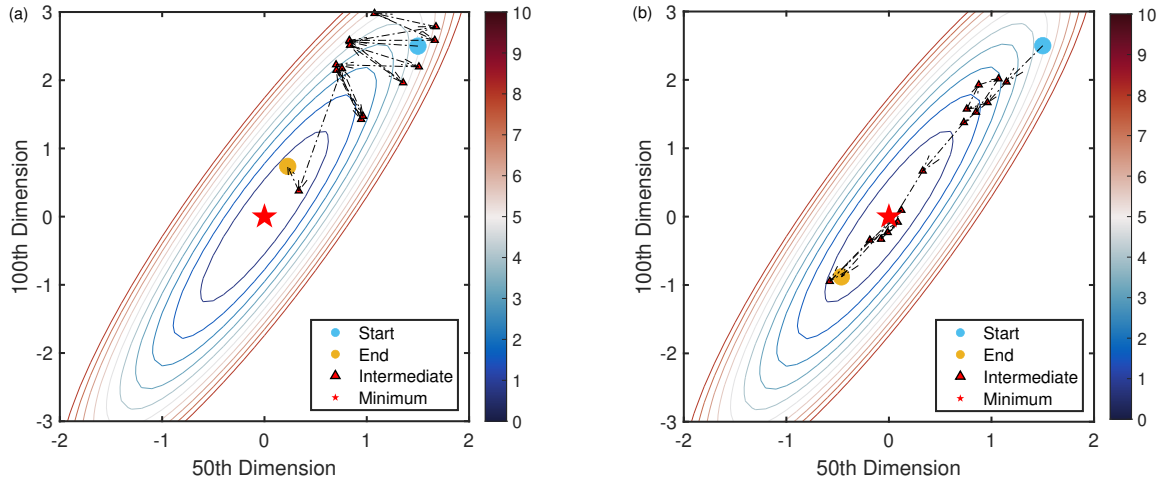

FIG. 1. Initial 20 MC and HMC walker trajectories projected onto the 50th and 100th dimensions. (a) Trajectories of HMC when $\mathbf{M} = \mathbf{I}$. (b) Trajectories of HMC when $\mathbf{M} = \mathbf{C}_\mathbf{M}^{-1}$.

The results and corresponding marginal distributions are depicted in Figure 2 and Figure 3. It becomes apparent that when the Hessian information is factored into the sampling process, it exhibits a structured pattern. In contrast, in the absence of Hessian information, the sampling process displays relatively fewer structural features, yet some subtle random-walk characteristics are still observable. Regarding the marginal distribution, the posterior without the incorporation of Hessian information exhibits a subtle skewness away from the normal distribution. In contrast, the posterior distribution that includes Hessian information aligns more closely with a Gaussian distribution.

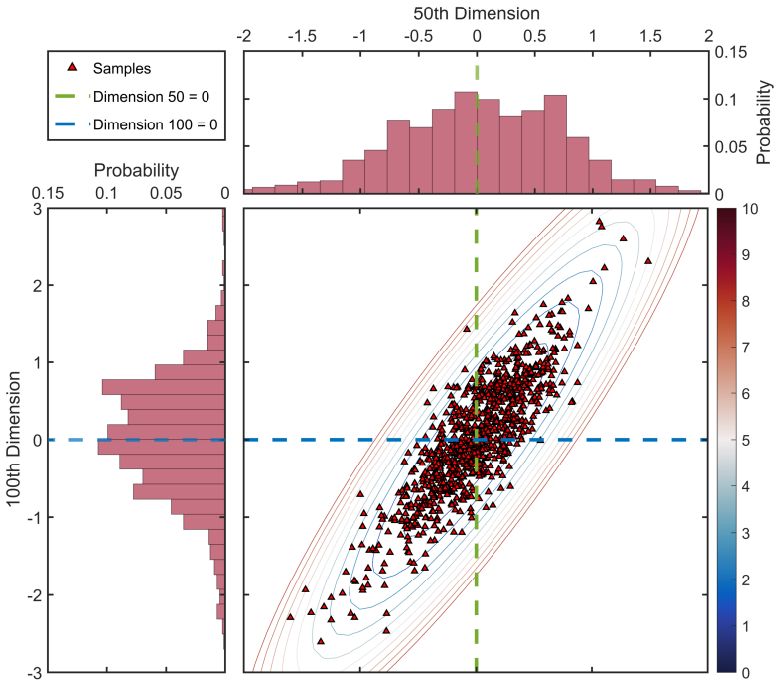

FIG. 2. Accepted samples of MC projected onto the 50th and 100th dimensions, and the marginal distributions.

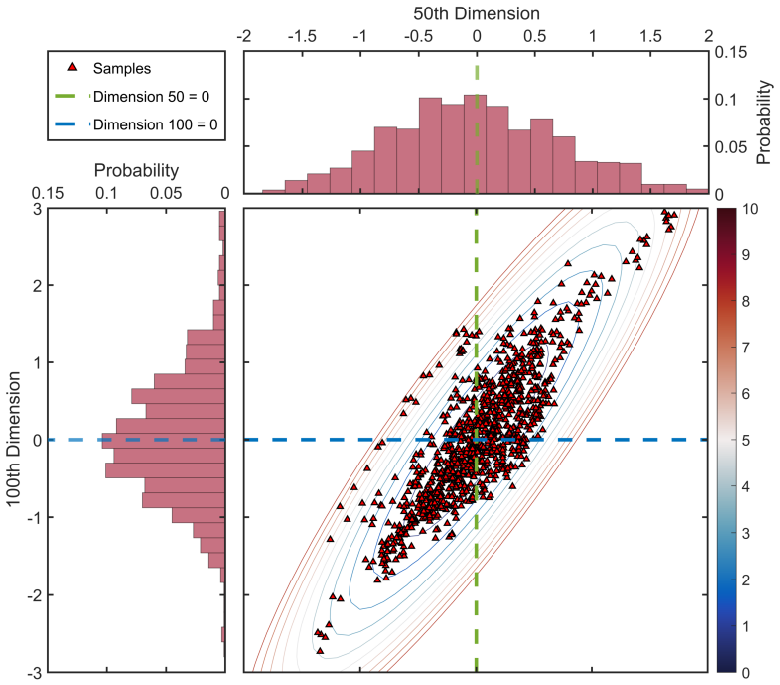

FIG. 3. Accepted samples of HMC projected onto the 50th and 100th dimensions, and the marginal distributions.

Tuning of HMC-FWI

In this section, we endeavor to explore the significance of adaptive tuning in addressing such challenges. In the setup for this experiment, our objective is to fine-tune the HMC for FWI problems. As displayed in Figure 4, the P-wave velocity model we seek to estimate is a Gaussian anomaly located within a homogeneous background model, which also serves as the initial guess. Constant density is used in this experiment.

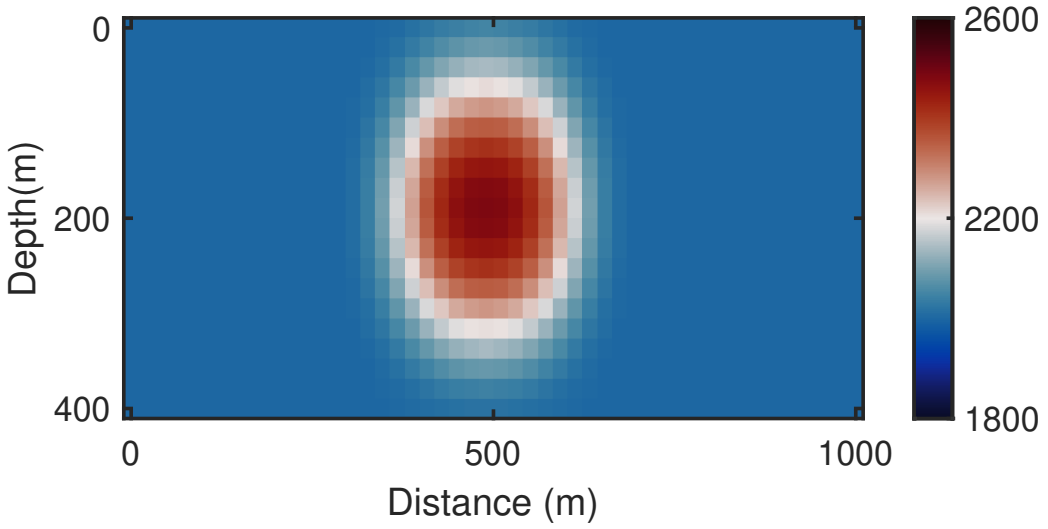

FIG. 4. True model for HMC-FWI

In line with our previous discussions on adaptive tuning, we implement a dynamic integration length, which is randomly chosen from the range of 15 to 30. According to discussions centered on the initial settings of the time step, such as Brooks et al. (2011); Beskos et al. (2013), we set the starting time step, denoted as δt, to 2 times the maximum value within the model vector, and scale it by $L \times dimension^{-1/4}$. The adjustment of the time step is based on the acceptance rate within the sampling subset. If the acceptance rate falls below 65%, we decrease δt by a factor of 0.8. Conversely, when the acceptance rate within a subset exceeds 85%, we increase the time step by reciprocal of 0.8. It's worth noting that the boundaries for the acceptable range of acceptance rates and the specific factor used to modify δt are established empirically and are in line with the findings of Fichtner et al. (2021).

We conducted two HMC-FWI experiments, one without tuning and the other with tuning. The sampling duration was set at $20,000$ steps, divided into 20 subsets. We deploy a transmissive acquisition system that involved 24 sources, evenly spaced at intervals of 2 grid points on both the upper and lower sides of the model. Additionally, there were 48 receivers deployed across the model's surface. In terms of the frequency-domain FWI parameters, we employed a single frequency band and divided it into 8 subfrequencies, evenly ranging from 1 Hz to 15 Hz. The HMC-FWI's acceptance rate without parameter tuning stands at approximately 10.74%, equating to merely 2147 accepted proposals. The mean model and the standard deviation plot in Figure 5 also show results that are distant from the desired ones. In contrast, the acceptance rate for the tuned proposals is notably higher, at

approximately 67.09%, resulting in around $13,418$ proposals being accepted. This significant discrepancy in acceptance rates primarily arises from the untuned parameters in the former scenario, which do not adequately align with the dynamics of the objective function. In this context, the exceptionally low acceptance rate in the initial test suggests that the sampler was trapped in an infeasible region and remained there for an extended period. Initially, the choice of δt may be suitable for simulating the initial steps of Hamiltonian dynamics, but as the search progresses, parameter adjustments based on acceptance become imperative. This adjustment process can also serve as a partial reflection of the evolving dynamics within the objective function.

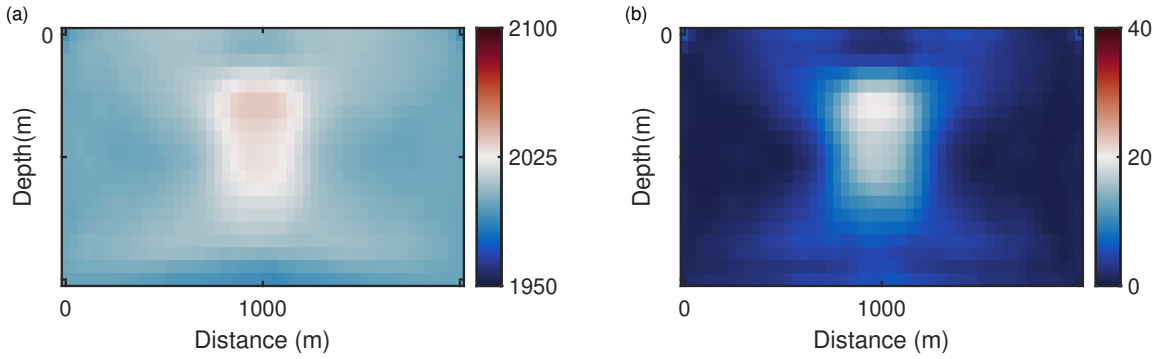

FIG. 5. HMC-FWI results without adaptive tuning. (a) mean of the accepted models. (b) the standard deviation of the accepted models.

The HMC-FWI results after tuning are presented in Figure 6 display a similar anomaly structure as the accurate model, as depicted by the model means. The influence of the acquisition geometry is evident due to the HMC being guided by the gradient, which becomes more pronounced in the upper part of the anomaly. However, uncertainties are observed in the same region, indicated by high standard deviation values, even though it receives priority updates. Two specific positions, $(500, 100)$ and $(500, 200)$, also indicated by stars in Figure 6 (a), have their distributions plotted. At the position of $(500, 100)$, the V_P distribution is approximately centered at its true value indicated by the red dashed vertical line, while at $(500, 200)$, the distribution displays a skewed appearance. This result can also be evidenced through a difference in uncertainty, as Figure 6 (b) shows that the uncertainty of the upper position is significantly less than the lower model position. Taking one vertical profile that crosses the two reference points in Figure 6 (a) and the same position in Figure 5 for comparison in Figure 7, the HMC mean model after tuning is close to the true profile, while the HMC mean model without tuning is merely updated.

The results of the HMC-FWI after parameter tuning, as depicted in Figure 6, exhibit a similar anomaly to that of the accurate model, as illustrated by the model means. The influence of the acquisition geometry is evident due to the HMC being guided by the gradient, which becomes more pronounced in the upper part of the anomaly. However, there are noticeable uncertainties present in this region, as indicated by the high standard deviation values, even though it receives priority updates. To highlight specific positions, $(500, 100)$ and $(500, 200)$, marked with stars in Figure 6 (a), their distributions are plotted. At the $(500, 100)$ position, the distribution of V_P is approximately centered around its true value, as denoted by the red dashed vertical line. In contrast, at $(500, 200)$, the distribution appears

skewed. This variation is further evident in the difference in uncertainty, where Figure 6 (b) reveals that the uncertainty at the upper position is considerably lower than that at the lower model position. By examining a vertical profile that crosses both reference points in Figure 6 (a) and comparing it with the corresponding position in Figure 5, it becomes apparent in Figure 7 that the HMC-FWI mean model, after parameter tuning, closely approximates the true profile. In contrast, the model without tuning is merely updated. The misfit plot depicted in Figure 8 provides a more intuitive illustration of the convergence problem when adaptive tuning is not applied.

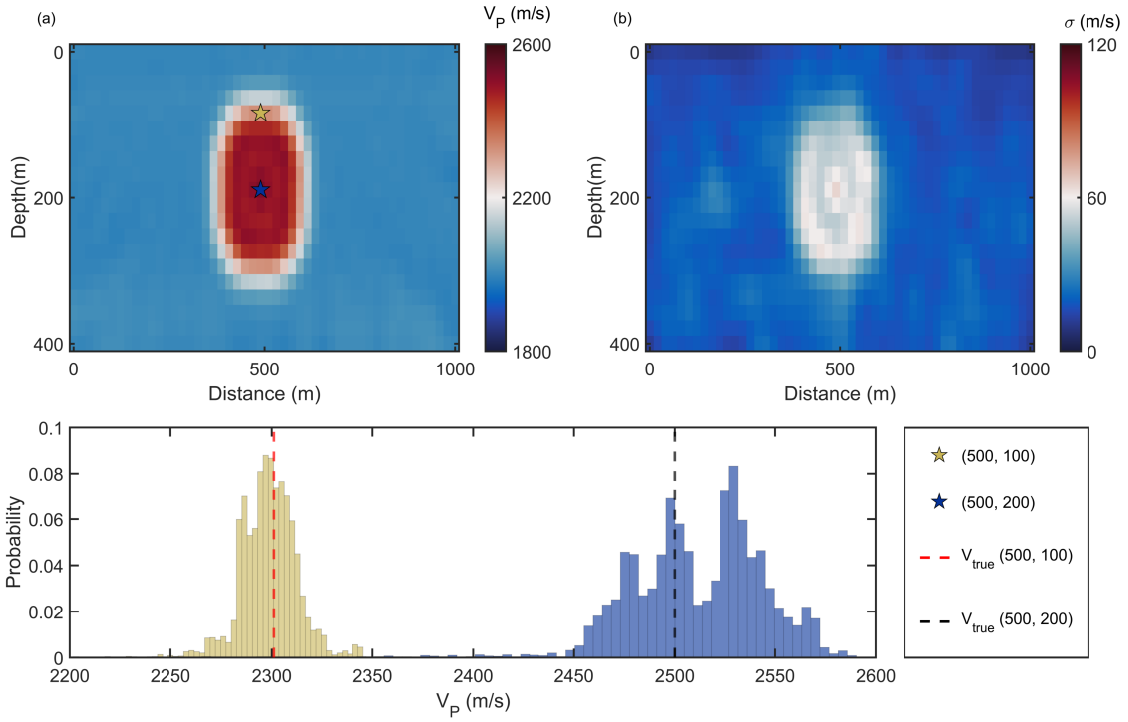

FIG. 6. HMC-FWI results with adaptive tuning. The yellow star denotes the model position at $(500, 100)$, and the blue star denotes the model position at () (a) mean of the accepted models. (b) the standard deviation of the accepted models. (c) probability distribution of models at $(500, 100)$ and $(500, 200)$, and the dashed line shows the true values at these positions.

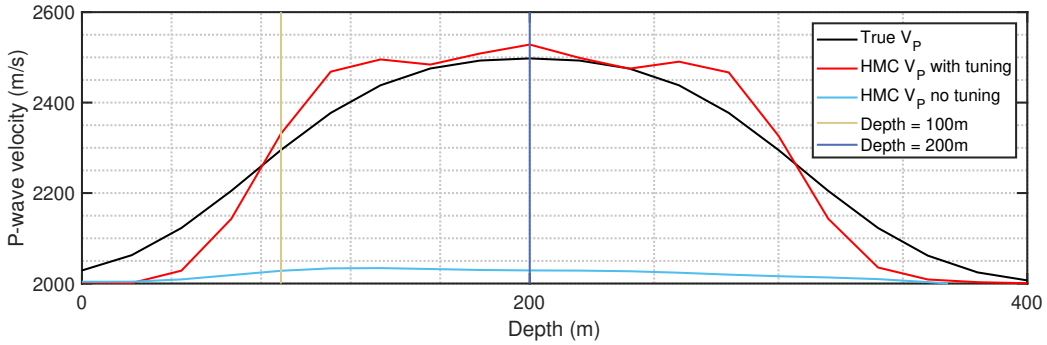

FIG. 7. Vertical profile crossing two reference points.

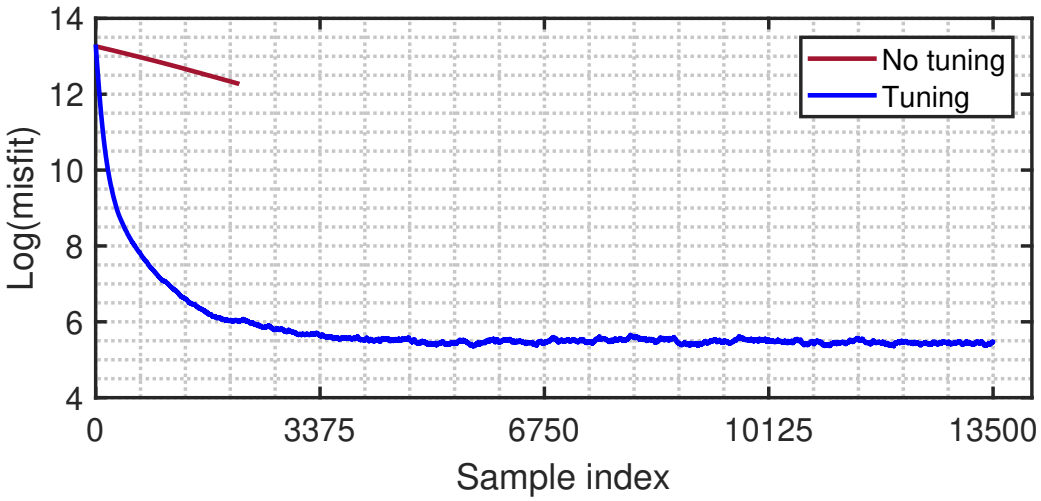

FIG. 8. Misfit's variation with iteration.

DISCUSSION

The analysis of our experiments highlights the pivotal role that tuning plays in HMC applications, especially in the HMC-FWI problems. In the simple sampling problem, the utilization of the second-order information of the potential energy, namely the probability density, assists in more efficiently navigating the particle movement toward the desired distribution. This suggests that the curvature information may also have a positive impact on the HMC-FWI problems, as Hessian is usually positively impactful in FWI. The investigation of incorporating the Hessian matrix into the mass matrix is our next step. Various approaches exist for leveraging second-order information. These approaches may include implementations such as iterative approaches like BFGS and L-BFGS realizations (Zhang et al., 2016; Fichtner et al., 2021; Liang et al., 2023). The development of this adaptation is currently underway and represents our next research endeavor.

Comparing the outcomes of tuned and untuned HMC-FWI attempts, we observe a remarkable enhancement in the structural similarity between the mean model and the true model when tuning is employed. This enhancement becomes particularly evident when we focus on specific model positions. Notably, in cases where untuned HMC-FWI confronts convergence obstacles, preventing the attainment of complete posterior distributions, the adoption of adaptive tuning emerges as a robust remedy. It not only fosters more accurate results that closely align with the true values but also expedites the convergence process. It is important to note that, while various research endeavors have theoretically proposed plausible initial settings for tunable parameters, we suggest not skipping the trial-and-error phase when employing HMC, as this ensures the stability and reliability of the process, as theoretical conclusions may necessitate further investigation, especially within the domain of more intricate models and higher-dimensional problems.

CONCLUSIONS

This study has delved into the HMC problems by discussing the critical role that adaptive tuning plays in such problems. Our research contributes to a deeper understanding of the tunable parameters in HMC and provides a reference point for future research and application in this field. We have theoretically and experimentally demonstrated how adaptive tuning significantly improves the HMC-related optimization problems. The tuned HMC accelerates the total convergence, thus assisting in more precise posterior estimation within a limited amount of samples. In higher-dimensional problems, the implementation of adaptive tuning holds the potential to conserve substantial computational resources. However, as the complexity of models and dimensions increases, the effectiveness of adaptive tuning may warrant continued examination.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

One of the authors of this report was supported by the CSEG Foundation.

REFERENCES

Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M., and Stuart, A., 2013, Optimal tuning of the hybrid monte carlo algorithm: Bernoulli, **19**, No. 5A, 1501–1534, funding by NSF.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., 2011, Handbook of Markov Chain Monte Carlo, CRC press.

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D., 1987, Hybrid monte carlo: Physics Letters B, **195**, No. 2, 216–222.
URL https://www.sciencedirect.com/science/article/pii/037026938791197X

Fichtner, A., Zunino, A., and Gebraad, L., 2018, Hamiltonian Monte Carlo solution of tomographic inverse problems: Geophysical Journal International, **216**, No. 2, 1344–1363, https://academic.oup.com/gji/article-pdf/216/2/1344/41325089/gji_216_2_1344.pdf.
URL https://doi.org/10.1093/gji/ggy496

Fichtner, A., Zunino, A., Gebraad, L., and Boehm, C., 2021, Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration: Geophysical Journal International, **227**, No. 2, 941–968, https://academic.oup.com/gji/article-pdf/227/2/941/39322013/ggab270.pdf.
URL https://doi.org/10.1093/gji/ggab270

Hamilton, W. R., 1834, On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function: Philosophical Transactions of the Royal Society of London, **124**, 247–308.
URL http://www.jstor.org/stable/108066

Hoffman, M. D., and Sountsov, P., 2022, Tuning-free generalized hamiltonian monte carlo, *in* Camps-Valls, G., Ruiz, F. J. R., and Valera, I., Eds., Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, vol. 151 of *Proceedings of Machine Learning Research*, PMLR, 7799–7813.
URL https://proceedings.mlr.press/v151/hoffman22a.html

Homan, M. D., and Gelman, A., 2014, The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo: J. Mach. Learn. Res., **15**, No. 1, 1593–1623.

ISERLES, A., 1986, Generalized Leapfrog Methods: IMA Journal of Numerical Analysis, **6**, No. 4, 381–392, https://academic.oup.com/imajna/article-pdf/6/4/381/2181277/6-4-381.pdf.
URL https://doi.org/10.1093/imanum/6.4.381

Kennedy, A., and Pendleton, B., 2000, Cost of generalised hmc algorithms for free field theory: Nuclear Physics B - Proceedings Supplements, **83-84**, 816–818, proceedings of the XVIIth International Symposium on Lattice Field Theory.
URL https://www.sciencedirect.com/science/article/pii/S0920563200918139

Liang, Z., Wellmann, F., and Ghattas, O., 2023, Uncertainty quantification of geologic model parameters in 3d gravity inversion by hessian-informed markov chain monte carlo: GEOPHYSICS, **88**, No. 1, G1–G18, https://doi.org/10.1190/geo2021-0728.1.
URL https://doi.org/10.1190/geo2021-0728.1

Mangoubi, O., and Smith, A., 2017, Rapid mixing of hamiltonian monte carlo on strongly

log-concave distributions: arXiv: Probability.
URL https://api.semanticscholar.org/CorpusID:56216095

Neal, R. M., 1993, Probabilistic inference using markov chain monte carlo methods.

Zhang, Y., and Sutton, C., 2011, Quasi-Newton Methods for Markov Chain Monte Carlo, vol. 24, Curran Associates, Inc.
URL https://proceedings.neurips.cc/paper_files/paper/2011/file/e702e51da2c0f5be4dd354bb3e295d37-Paper.pdf

Zhang, Y., Wang, X., Chen, C., Henao, R., Fan, K., and Carin, L., 2016, Towards Unifying Hamiltonian Monte Carlo and Slice Sampling: arXiv e-prints, arXiv:1602.07,800, 1602.07800.