

Hamiltonian Monte Carlo methods for uncertainty quantification in waveform inversion

Jinji Li and Kristopher A. Innanen
CREWES, University of Calgary

Summary

It is generally understood that effective use of seismic waveform inversion for modern monitoring applications requires that robust quantification of uncertainties be possible. Currently, there is no consensus on the optimal methodology. Many candidate methods involve sampling in model space in the vicinity of the inverted parameters, but this can represent a dramatic increase in computational cost. Classical Markov Chain Monte Carlo (MCMC) methods, while widely used for Bayesian inference, are inefficient in this sense, particularly when dealing with high-dimensional parameter spaces. To mitigate this, Hamiltonian Monte Carlo (HMC) methods have been introduced. By simulating Hamiltonian dynamics and incorporating a Metropolis acceptance step, HMC avoids many of the unhelpful trials associated with the random walks in MCMC. This increases the acceptance rate, enabling more efficient parameter space exploration. Additionally, generating many plausible model candidates amounts to exploration of the “inversion null space”, which can be particularly valuable in inversion problems such as full-waveform inversion (FWI) where the model space is complex and multidimensional. The use of HMC in practical FWI is not yet common, and so we use this paper to examine its basic behaviour, and flesh out some of its advantages in hydrocarbon recovery or CO₂ storage, using simulated geological reservoir or plume model. Our findings are generally positive about the applicability of HMC in seismic inversion and monitoring, but we identify some areas in which standard drawing of candidates might be improved in light of this specific (waveform inversion) problem.

Theory

Hamiltonian dynamics (Hamilton, 1834) can be conceptually understood by fictitiously visualizing a frictionless particle moving along a 2D surface with varying heights. In this scenario, the system's state is described by the position of this particle \mathbf{q} (2D vector) and its generalized momentum \mathbf{p} (also a 2D vector). While moving, this particle's potential energy can be represented by $U(\mathbf{q})$, and its kinetic energy is given by $\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}$, where \mathbf{M} is the mass matrix of this particle. A generalization of this matrix is making the off-diagonal entries zeros while keeping the diagonal members equal to some preset values, but there are more constrained settings in some more complex problems (Fichtner et al., 2021). In most situations, the mass matrix is positive definite. When the surface is flat, the particle moves at a constant velocity equal to $\mathbf{M}^{-1}\mathbf{p}$. However, if the surface inclines, the momentum allows the particle to continue, resulting in decreased kinetic energy and increased potential energy. Eventually, when the kinetic energy reaches zero, this particle will slide back down the slope, increasing kinetic energy and decreasing potential energy. During the whole process, this particle is governed by a "gravitational" force that is parallel to $-\nabla U(\mathbf{q})$, which is regarded as the steepest descent direction. Such mechanics can be described by a Hamiltonian equation $H(\mathbf{p}, \mathbf{q})$, such that

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

The partial derivatives of the Hamiltonian determine how \mathbf{q} and \mathbf{p} change over time, t , according to Hamilton's equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} = [\mathbf{M}^{-1}\mathbf{p}]_i, \quad (2)$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i}, \quad (3)$$

where $i = 1, 2, \dots, n$ denotes the index in the n -dimensional vector.

The fundamental idea behind HMC involves sampling from an auxiliary distribution in a phase space with twice the original space's dimensions. The auxiliary distribution is known as the canonical distribution (Davey, 2009), and this extended phase space is denoted as (\mathbf{p}, \mathbf{m}) , where \mathbf{p} represents momentum variables and \mathbf{m} , which replaces \mathbf{q} in equation (1), represents the variables of interest (e.g., model parameters in inversion problems). Furthermore, artificial momentum variables are introduced to align with the quantity of position variables. More intuitively, the concept of HMC views a model \mathbf{m} as analogous to a particle in mechanics. This particle moves from its present position to a new position following a Hamiltonian trajectory (Brooks et al., 2011). The geometry of the trajectory is controlled by the misfit that is interpreted as potential energy U , as well as the kinetic energy K and mass \mathbf{M} of the particle, namely the artificially introduced auxiliary quantities (Fichtner et al., 2018).

In many inversion problems including FWI, a plausible assumption of the prior information is the Gaussian uncertainties (e.g., Tarantola, 2005; Dettmer et al, 2010; Dosso et al, 2014; Fichtner et al, 2021). Consequently, the posterior, namely the potential energy in HMC can be defined as the logarithm of the probability density linked to the inversion variables:

$$U(\mathbf{m}) = -\log \rho_{\mathbf{m}}(\mathbf{m}|\mathbf{d}) = \frac{1}{2}(\mathbf{d}_{obs} - \mathbf{d}_{syn})^T \mathbf{C}_D^{-1}(\mathbf{d}_{syn} - \mathbf{d}_{syn}) + \frac{1}{2}\mathbf{m}^T \mathbf{C}_M^{-1}\mathbf{m}, \quad (4)$$

where \mathbf{d}_{obs} and \mathbf{d}_{syn} is the observed and synthetic (real) data; \mathbf{C}_D^{-1} and \mathbf{C}_M^{-1} are the inverse of the data and the model covariance matrices, respectively.

By combining the components elucidated in the preceding sections, we can outline a fundamental workflow for implementing HMC in the context of FWI problems. Irrespective of the dimensionality, the subsurface structures to be estimated can be succinctly represented as a one-dimensional vector, constituting the model vector \mathbf{m} . The momentum vector \mathbf{p} is derived from a multi-dimensional normal distribution with zero means and a covariance matrix \mathbf{M} , expressed as $\mathcal{N}(0, \mathbf{M})$. The adjoint method (Plessix, 2006) is applied to get the gradient of U in equation 3.

Results

In this section, we present an acoustic FWI using HMC in the frequency domain. We artificially create a structure with varying P-wave velocity to be estimated (refer to Figure 1). The horizontal and vertical grids are chosen to be 100 and 50, respectively, with a grid interval of 20 m. Regarding the acquisition system, 48 seismic sources with an interval of 2 grids on the top, 96 receivers with an interval of 1 grids on the model's surface, and 48 receivers with an interval of 1

grids on the model's left and right side. The synthetic data is discretized into a set of 8 frequencies, spanning from 1 to 15 Hz. In our inversion process, we consider the physical parameter to be inverted as independently and identically distributed following a Gaussian distribution. In addition, we assume that the data and the model covariance matrices are identity matrices. The initial model was stochastically generated from a uniform distribution centered at the smoothed model, with boundaries extending from -150 to +150 relative to the smoothed model. 30,000 samples are generated in this problem.

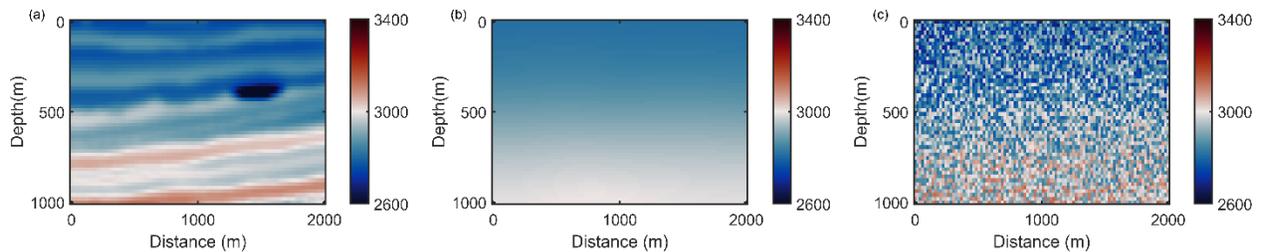


Figure 1. P-wave velocity models for HMC-FWI. (a) – (c) true, smoothed, and the initial model.

An adaptive tuning strategy is applied to enhance the stability and increase the acceptance rate. We implement a dynamic integration length randomly chosen from the range of 30 to 50. 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 a reciprocal of 0.8.

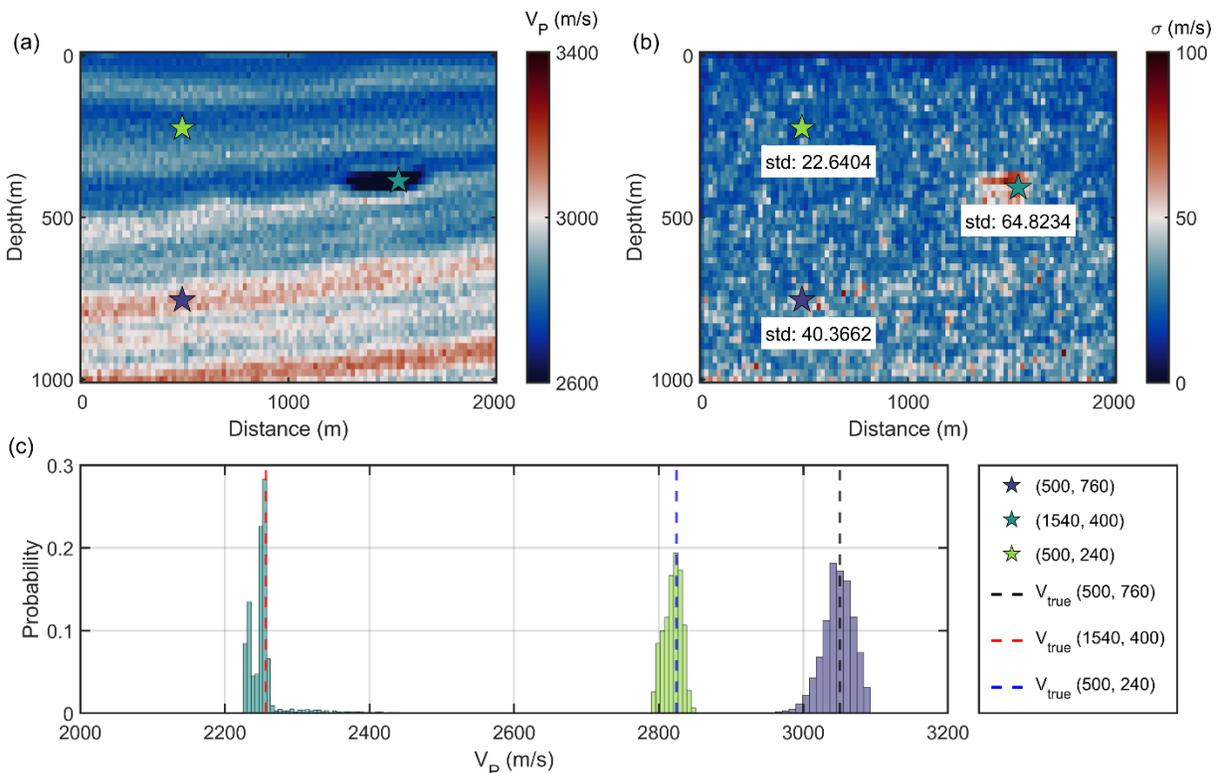


Figure 2. HMC-FWI results. The white star denotes the model position at (1540, 400), the blue star denotes the model position at (500, 760), and the red star denotes the model position at (500, 240). (a) mean of the accepted models. (b) the standard deviation of the accepted models. (c) probability distribution of models at positions denoted by the three stars, and the dashed line shows the true values at these positions.

The overall acceptance rate is approximately 78.02%, elucidating 23,407 plausible model candidates. The model mean depicted in Figure 2 (a) exhibits a similar structure to the true model. Employing the uncertainty plot presented in Figure 2 (b), we conduct a detailed analysis focusing on points corresponding to varying uncertainty levels representing low, moderate, and high uncertainties. This analysis is further elucidated by the distributions shown in Figure 2 (c). For positions characterized by a low standard deviation value, the P-wave velocity reveals a normal distribution centered around the true model value. In positions with a moderate standard deviation value, the model distribution manifests as a near Gaussian distribution, with a slight deviation on the higher-value side but still maintaining a central alignment with the true model value. However, positions characterized by a high standard deviation result in a multi-modal distribution, with the majority of models aligning with the true model value. Examining two vertical profiles intersecting the reference points illustrated in Figure 2 for comparison in Figure 3, it is evident that the HMC mean model closely approximates the true profile.

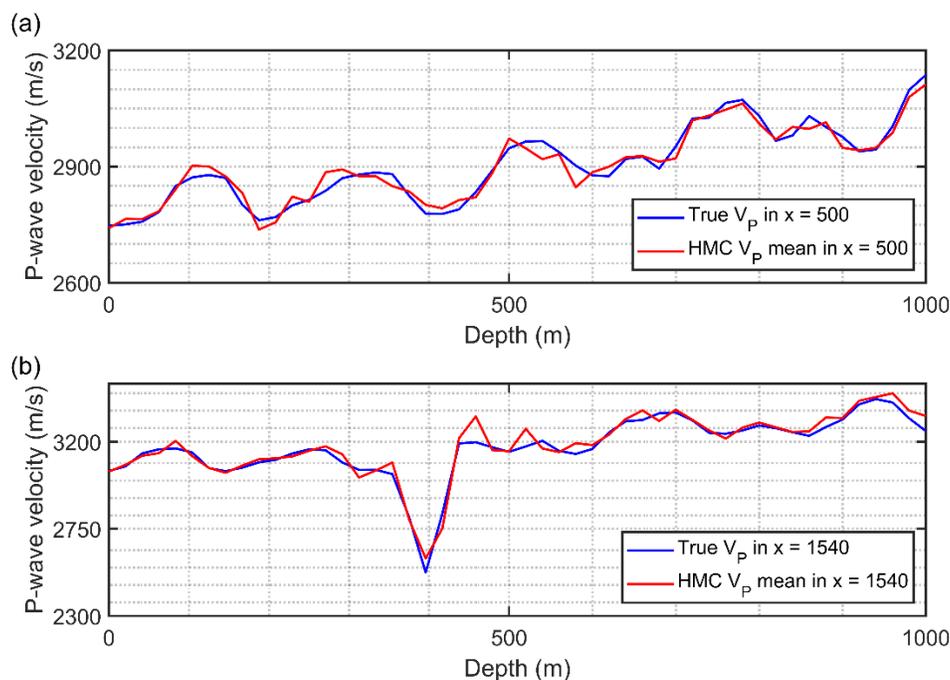


Figure 3. Vertical profile crossing reference points in Figure 2. The blue lines are the true profiles, and the red lines are the profiles of the HMC mean model. (a) profile crossing $x=500$. (b) profile crossing $x=1540$.

While this study does not explicitly conduct a well-defined inversion nullspace analysis as demonstrated by (Fichtner et al., 2021), HMC, as other MC variants, generates a chain consisting of model candidates with associated objective function values. We have set the tolerance to be 0.11% of the initial objective function value, as observed in the small window

where the objective function oscillates around a certain value in Figure 4. This tolerance criterion has yielded approximately 2,228 plausible models, of which 9 are displayed in Figure 5.

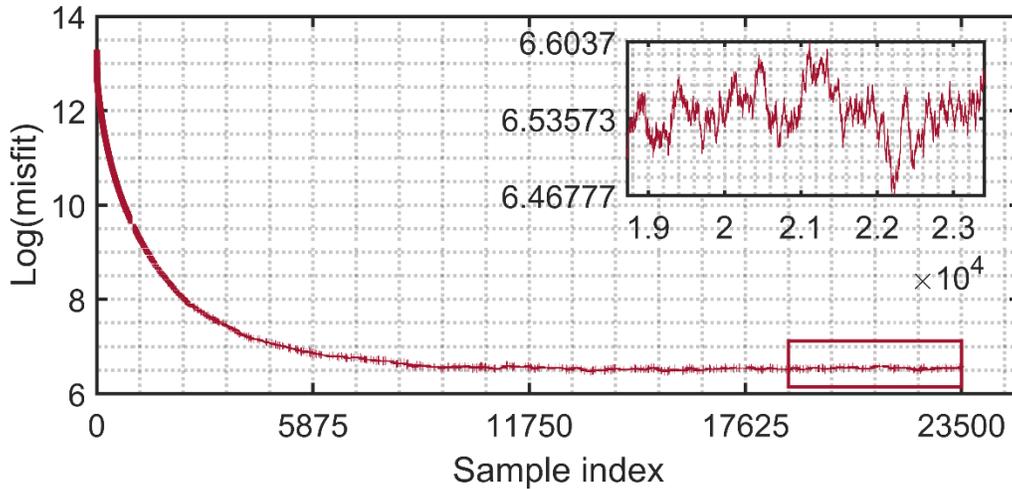


Figure 4. Misfit's log variation with iteration.

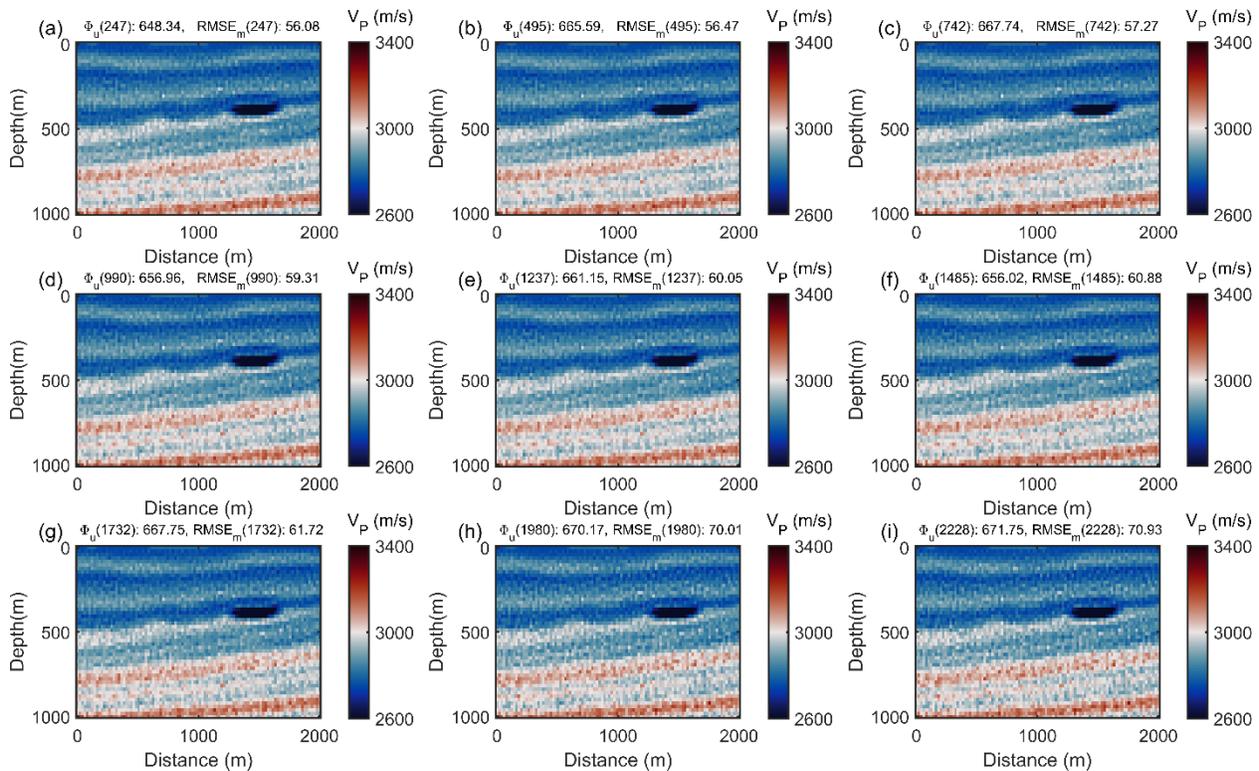


Figure 5. 9 models in the subset with objective function's values lower than 0.11% of the initial values. (a)-(i) are evenly selected models from the subset. The data misfit is shown by Φ , and the model misfit is shown by RMSE on top of each subplot.

In each subplot of Figure 5, the top section presents both objective function values and root-mean-square errors (RMSEs) between the true model and each candidate in the inversion nullspace. A comparison of these values reveals a noteworthy characteristic of the inversion nullspace, namely, the model that results in a lower objective function value is not necessarily more accurate. For instance, the model in Figure 5 (b) yields a relatively larger objective function value compared to the model in Figure 5 (a), yet the accuracy of the former is not correspondingly greater. This observation contributes to a more comprehensive understanding of uncertainty and underscores the need to complement FWI uncertainty delineation with additional information.

Conclusions

This study has presented a thorough overview of HMC, elucidating its conceptual foundation, key components, and numerical implementation. HMC has demonstrated clear advantages over conventional MC methods in the context of FWI. Notably, it exhibits a higher acceptance rate, improved efficiency in exploring null spaces, and the potential for faster convergence. As we move forward, our research focus will be directed toward refining the tuning of the HMC algorithm to enhance its adaptability and efficacy in diverse FWI scenarios. By harnessing the full potential of HMC, we anticipate achieving more reliable and efficient inversion results in seismic imaging applications, ultimately contributing to advancements in subsurface characterization and exploration.

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. The first author was supported by the CSEG Foundation.

References

- 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.
- 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.
- Davey, K., 2009, What is gibbs's canonical distribution?: *Philosophy of Science*, 76.
- Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., 2011, *Handbook of Markov Chain Monte Carlo*, CRC press.
- Fichtner, A., Zunino, A., and Gebraad, L., 2018, Hamiltonian Monte Carlo solution of tomographic inverse problems: *Geophysical Journal International*, 216, No. 2, 1344–1363.
- Tarantola, A., 2005, *Inverse problem theory and methods for model parameter estimation*: SIAM.
- Dettmer, J., Dosso, S. E., and Holland, C. W., 2010, Trans-dimensional geoacoustic inversion: *The Journal of the Acoustical Society of America*, 128, No. 6, 3393–3405.
- Dosso, S. E., Dettmer, J., Steininger, G., and Holland, C. W., 2014, Efficient transdimensional bayesian inversion for geoacoustic profile estimation: *Inverse Problems*, 30, No. 11, 114,018.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, 167, No. 2, 495–503.