Bayesian Elastic Full-Waveform Inversion using Hamiltonian Monte Carlo

Lars Gebraad[1] and Andreas Fichtner[1]

[1]ETH Zürich

November 24, 2022

Abstract

We develop a Hamiltonian Monte Carlo (HMC) sampler which solves a multi-parameter elastic full-waveform inversion (FWI) in a probabilistic setting for the first time. This gives novel access to the full posterior distribution for this type of highly non-linear inverse problem. Typically, FWI has focused on using gradient descent methods with proper regularization to iteratively update models to a minimum misfit value. Non-uniqueness and uncertainties are mostly in this approach. Bayesian inversions offer an alternative by assigning a probability to each model in model space given some data and prior constraints. The drawback is the need to evaluate a very large number of models. Random walks from Markov chains counter this effect by only exploring regions of model space where probability is significant. HMC method additionally incorporates gradient information, i.e. local structure, typically available for numerical waveform tomography experiments. So far, HMC has only been implented for acoustic FWI. We implement HMC for multiple 2D elastic FWI set-ups. Using parallelized wave propagation code, wavefields and kernels are computed on an regular numerical grid and projected onto basis functions. These gradients are subsequently used to explore the posterior space of different target models using HMC. The free parameters in these experiments are P and S velocity, and density. Although simulating Hamiltonian dynamics in the resulting phase space is approximated numerically, the results of the Markov chain are nevertheless very insightful. No prior tuning of kernels, data or model space is required, under the constraint that the sampler is properly tuned. After a burn in phase during which the mass matrix is iteratively optimized, the Markov chain is run on multiple nodes. After approximately 100,000 samples (combined from all nodes) the Markov chain mixes well. The resulting samples give access to the full posterior distribution, including the mean and maximum-likelihood models, conditional probabilities, inter-parameter correlations and marginal distributions.

Bayesian Elastic Full-Waveform Inversion using Hamiltonian Monte Carlo

Lars Gebraad (lars.gebraad@erdw.ethz.ch), Andreas Fichtner
ETH Zürich

ETH zürich SWP SEISMOLOGY & WAVE PHYSICS

Motivation & this study

Traditional **Full-Waveform Inversion** (FWI) is a deterministic procedure that operates by iteratively minimising an objective functional to invert for e.g. material parameters. Bayesian inference offers full non-linear characterisation at the expense of requiring many samples (computations). Typical sampling algorithms only update few parameters or are characterised by slow model space exploration, at least for sufficiently high dimensions.

Hamiltonian Monte Carlo (HMC) is a Markov chain Monte Carlo (McMC) sampling method which for the first time allows us efficiently generate enough samples on such a high-dimensional expensive forward problem. It does so by **incorporating misfit (or posterior) gradient information** in the sampling algorithm; paving the way for powerful non-linear Bayesian inference. The algorithm has been thoroughly described in the literature (e.g. Neal 2011), although its application in geophysical inverse problems is still limited, but spreading (Fichtner & Simutė, 2018; Sen & Biswas, 2017).

This study proves that HMC works for **2D synthetic multi-parameter (elastic) FWI studies** by solving multiple such inverse problems using limited cluster computations. Three different inverse problems (with differing number of model parameters) are investigated illustrating the benefits of high-dimensional posterior information.

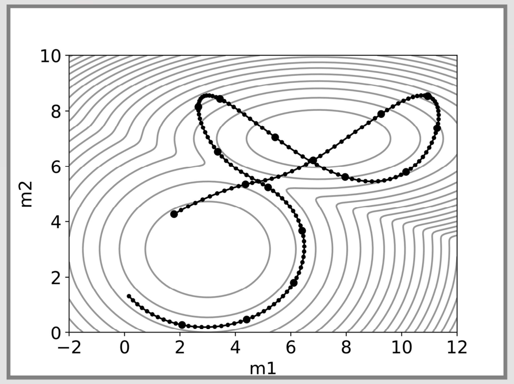
Theory on FWI and HMC

FWI is an inversion method that inverts for medium parameters by modelling complete wavefields. Its target functional is defined by the fit between synthetic and observed seismograms. The adjoint of the elastic wave equation is used to compute the parameter gradient or kernels. In deterministic inversions these gradients are used for gradient-descent style optimisations.

HMC is an McMC sampling method where the proposals are generated by simulating **Hamiltonian dynamics** (the equations of motion) on a potential surface (like planets under the influence of gravity) for some arbitrary time. This surface is given by the misfit function of the inverse problem, as illustrated in **Figure 1**. This way, the gradient of the target determines partially the direction of the new sample, guiding the sampler to interesting models. This is especially **beneficial for highly correlated parameters**.

The two tuning parameters for the HMC algorithm are the simulation time and the particle **mass matrix M**. The particle mass is similar to the mass of an object moving in a potential field, but every dimension can have a different mass, as long as the mass matrix M is positive definite and symmetric. By changing the mass matrix of the algorithm, model space is preferentially explored in different directions. This is illustrated in **Figure 1**.

Combining HMC and FWI is very natural, as the **FWI kernels themselves function as the derivatives needed in HMC!**



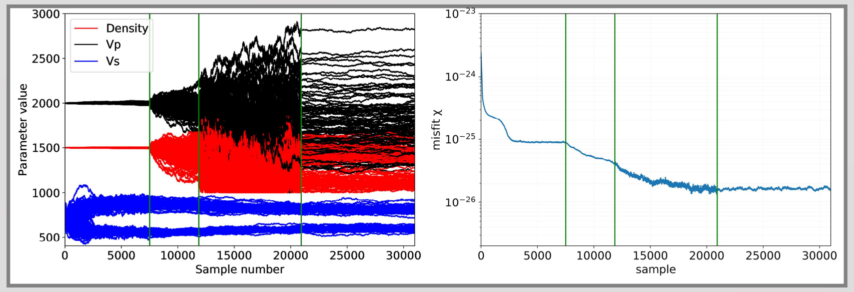
← **Figure 1:** One HMC trajectory through multimodal model space. One region has misfit structure that differs strongly from the other. This way, the sampler will behave differently in both regions. In non-linear problems such as this tuning is locally optimal.

Abstract ID : 363458
Abstract Number: S31D-0525
S31D: Frontiers in Theoretical and Computational Seismology Posters

Methods & results

We perform 3 separate **chequerboard** tests; twice with the same pattern but differing perturbation magnitudes, and once with a finer grained pattern. This way, we are able to **characterise both non-linearity and dimensional scaling** of the method. All three inversions solve for three parameter sets: P-wave velocity (**V_p**), S-wave velocity (**V_s**) and density. The chosen misfit is **L2 distance** between the seismograms. The set-up remains the same throughout al three inversions; 2 sources and 6 receivers placed around the chequerboard. The set-up can be seen in **Figures 3 and 4**.

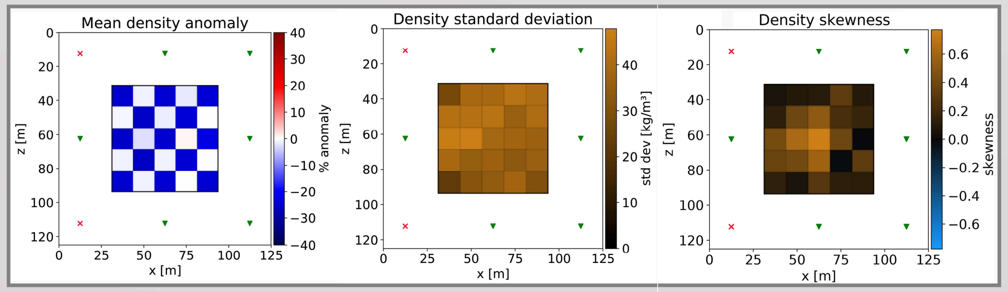
The forward and adjoint simulations are both finite-difference simulations using 4 point staggered grid stencils after Virieux (1986). The kernels computed are not smoothed but projected onto basis functions spanning 5×5 or 10×10 finite-difference grid points (square basis functions).



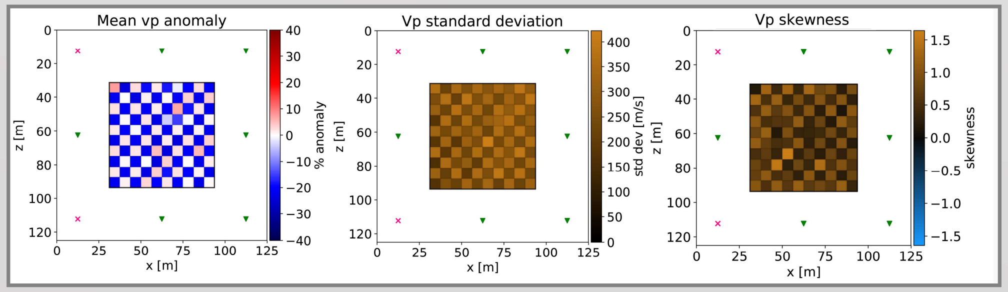
↑ **Figure 2:** Trace plot and misfit plot of one HMC **exploratory** chain. Every green line indicates tuning parameter update.

Tuning of the sampler is done iteratively during **an exploratory chain**. During this chain, first **V_s** is properly tuned, then **density**, then **V_p**. The reason for this is that if one tries to tune the least sensitive parameter first, the effect is drowned out by the burn-in effect of more sensitive parameters. When parameters seem to be **stationary over a few thousand samples**, the mass matrix is updated, until no large change in parameters or misfit is observed after changing masses. This is shown in **Figure 2**, where the different parameter sets reach stable values at different points in the exploratory chain. One parameter set (i.e. density) has only one mass, which is chosen on purpose to save time. The end result of all tuning is used in the main chains (next paragraph).

The samples used for the solution of the inverse problem are generated by **multiple chains**. These chains are started at samples randomly chosen from the last section of the exploratory chain. Each target uses between 10 and 40 chains (coarse vs. fine). The chains of **200K-600K accepted samples (approximately 65% acceptance)** are combined and **thinned to 1%** to speed up post-processing.



↑ **Figure 3:** Statistical moments of the posterior of **density** of a coarse target. The chequerboard target are perturbations of 25%, which are well resolved to within a few percent. Standard deviations differ per medium parameter. Skewness (third statistical moment) is non-zero for many parameters, but strongest in parameters furthest from the experimental array. Experimental set-up influence posterior structure strongly. Red × symbols indicate sources, green ▾ symbols indicate receivers.

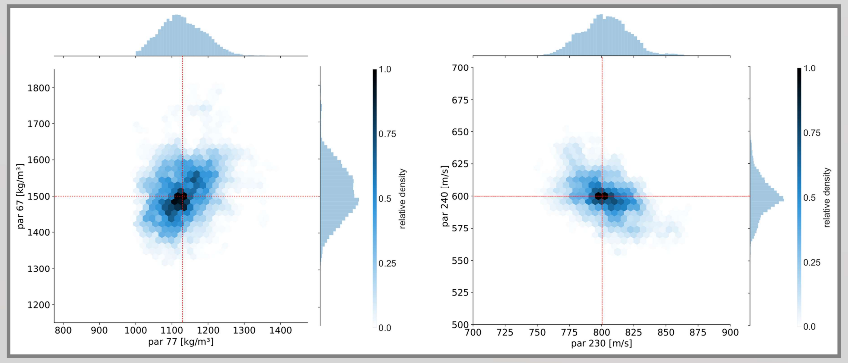


↑ **Figure 4:** Statistical moments of the posterior of **V_p** of a fine target from a chain started at the true model. Even after starting at the correct model, the inverse problem does not resolve the parameter well. Standard deviations and skewnesses are high and varying with position. Red × symbols indicate sources, green ▾ symbols indicate receivers.

↓ **Figure 5:** Marginals for 4 arbitrary parameters of the fine grained target. Correlations and linearity differ for all parameters. Red lines indicate true model.

The statistical moments shown in **Figure 3 and 4** emphasize three important findings of this study:

1. HMC is able to sample FWI posteriors adequately **even without a specific starting model** and **with strong perturbations (where deterministic methods would struggle with non-linearity)**, and;
2. Bayesian inference of FWI problems is able to **invert for density** accurately, and;
3. The posteriors for these FWI problems contain information normally not retrieved by deterministic FWI combined with Gaussian assumptions. **Higher order posterior moments are non-zero.**



All three different parameter sets show different resolvabilities. Parameter uncertainty also differs within the model. S-wave speed is best resolved, with strongest correlations between parameters, across all three targets. This is inherent to L2 misfits, as S-wave amplitudes are largest. The evaluation of the high dimensional chequerboard FWI problem was hindered by the limitations of FWI (subwavelength structure). This was confirmed by re-starting an McMC chain at the true model, as shown in **Figure 4**. Bayesian inference gives us access to complex information of our inverse problems, such as marginals (shown in **Figure 5**) and conditionals (*ask the presenter for a demo!*).

Conclusions

We have shown that Hamiltonian Monte Carlo allows us to perform Bayesian inference on FWI problems, even solving the FWI problem simultaneously for all three elastic parameters without starting models. In addition this study highlights some of the potential benefits Bayesian inference brings to the interpretation of FWI problems, especially in quantifying non-linearity.

We foresee that HMCFWI can be developed further using advanced McMC techniques such as replica exchange, to enhance convergence and model space exploration, in addition to modelling techniques to enhance performance. Bayesian inference of FWI problems could shed light on questions in the field where advanced uncertainty analysis is required.

References

Fichtner, A. & Simutė, S. (2018) - Hamiltonian Monte Carlo Inversion of Seismic Sources in Complex Media Journal of Geophysical Research: Solid Earth, 123 , 2984-2999
Neal, R. M. (2011) - MCMC Using Hamiltonian Dynamics, Handbook of Markov Chain Monte Carlo, p113-p162
Virieux, J. (1986) - P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. Geophysics, 51(4):889–901
Sen, M. K. & Biswas, R. (2017) - Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm, Geophysics, 82(4):R119-R134