Efficient Bayesian inference for full waveform inversion: an overview of modern approaches

*Philippe Marchner *, Univ. Grenoble Alpes, Romain Brossier, Univ. Grenoble Alpes, and Ludovic Métivier, Univ. Grenoble Alpes/CNRS*

Full Waveform Inversion (FWI) is a high-resolution seismic imaging method and a mature technology in exploration geophysics. It is inherently a severely ill-posed inverse problem, typically formulated as the minimization of a least-squares misfit function over the model parameter map $m(x), x \in \mathbb{R}^d$

$$\min_m C(m) = \frac{1}{2}\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{L^2},$$

where d_{obs} are observed seismic traces, and $d_{\text{cal}}[m]$ are simulated traces obtained through the numerical resolution of a wave equation. Gradient-based methods can tackle this problem iteratively, but due to the strong nonlinearity and the large effective null-space of the inverse problem, the recovered model often lacks interpretability. Considering Bayesian statistical inversion (Kaipio and Somersalo, 2006) addresses these limitations. By defining the likelihood probability distribution $\pi(d_{\text{obs}}|m) \propto \exp(-C(m))$, and a prior model $\pi_{\text{prior}}(m)$, we can infer on the posterior distribution through the application of Bayes' theorem $\pi(m|d_{\text{obs}}) \propto \pi(d_{\text{obs}}|m)\pi_{\text{prior}}(m)$. However, FWI's high computational cost, where a single likelihood evaluation may involve solving a wave equation with $\mathcal{O}(10^9)$ degrees of freedom in three-dimensions, renders direct sampling impossible. Inspired by recent FWI studies (Zhang et al., 2023; Hoffmann et al., 2024), we explore strategies to sample from the posterior with a limited number of likelihood evaluations leveraging parallel computing architectures.

TOWARDS MODERN BAYESIAN APPROACHES

Recent advances in computational optimal transport (Chewi et al., 2024; Santambrogio, 2015) have provided new insights for Bayesian inference thanks to the theory of gradient flows in the space of probabilities measures. When the model $M = m(x)$ is treated as a realization of a random variable, the classical FWI gradient descent has a probalistic interpretation in terms of Langevin dynamics for the stochastic process $(M_t)_{t\geq 0}$ in artificial time. The law $\mu_t \sim M_t$ satisfies a Fokker-Planck equation, and follows the gradient of the KL divergence between μ_t and the posterior in the Wasserstein metric. This perspective motivates the design of novel Bayesian inference methods. On one hand, Markov Chain Monte Carlo (MCMC) such as HMC (Gebraad et al., 2020) and Ensemble Kalman methods (EnKF) (Thurin et al., 2019) define algorithms that evolve particle-based representations of $(M_t)_{t\geq 0}$. EnKF is particularly attractive due to its embarrassingly parallel nature, and can perform inference at a reduced cost compared to standard MCMC, but introduces bias in the representation of the posterior. On the other hand, variational inference operates directly in probability space, requiring the posterior to be approximated by a family of simpler distributions for tractability. Common choices include Gaussian families (Kucukelbir et al., 2017) or empirical measures (Liu and Wang, 2016). Motivated by these developements, we analyze the strengh and weaknesses of some Bayesian methods for a simple FWI benchmark, and progressively increase its complexity.

A MOTIVATING EXAMPLE

We consider a one-dimensional FWI problem in the acoustic approximation. The space-time domain is $\Omega = [0, L] \times [0, T]$, with $L = 3$ km and $T = 3$ sec. The model $m(x)$ to be inferred describes the speed of the P-wave, and is a 2-layered medium $m(x) = (c_1, c_2)$, split at $x = 1.5$ km. The wave equation is discretized with a 2nd-order finite-difference explicit scheme with absorbing boundary conditions at $x = 0$ and $x = L$. A ricker source wavelet centered at $f_0 = 5$ Hz is set at $z = 10$m. Three receivers located at $x_r = \{1.5, 2, 2.5\}$ km are used to record the transmitted wave only. The countours of the cost function are shown in Fig. 1 for a reference set at $m^*(x) = (2500, 4000)$ m/s. For the inference, we use a uniform prior distribution with $c_{\text{min}} = 1000$ m/s and $c_{\text{max}} = 6000$ m/s.

We run a MCMC that uses the Langevin dynamics as a proposal distribution (MALA, Roberts and Tweedie (1996)). Fig. 1 shows the last samples of the chain. As a comparison, we superpose samples of a vanilla Ensemble Kalman Inversion (EKI, Iglesias et al. (2013)) with 30 ensemble members after 10 iterations. As expected, EKI is cheap but underestimates uncertainties. For both methods, the quality of the sampling is strongly influenced by the initial model. We pursue further investigations on a n-layered medium, and identify desirable sampling properties based on large-scale FWI requirements and the theory of gradient flows. We play on the ill-posedness of the inversion, and slowly move to more complex situations in two-dimensions.

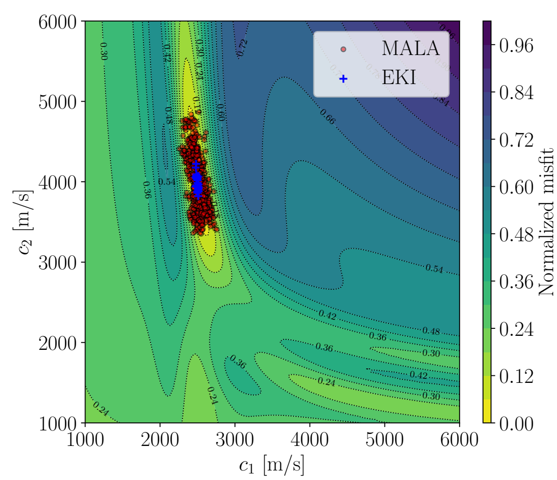

Figure 1: Countours of the normalized FWI misfit function, and samples of a converged MALA chain and EKI.