Bayesian Approaches to a Seismic Source Inversion Problem.

Juan Pablo Madrigal Cianci*, **Sebastian Krumscheid***, **Fabio Nobile***, **and Raúl Tempone†**

* MATH-CSQI, EPFL, Switzerland; † SRI Center for Uncertainty Quantification in Computational Science, KAUST, Thuwal, Saudi Arabia.

Introduction

We focus on implementing sampling strategies for Bayesian inversion, using Markov Chain Monte Carlo (MCMC), to recover the probability distribution of the spatial location of the epicenter of an earthquake in a bounded domain.

Problem Setting

Mathematical Model:

- Physical Domain: $D \subset \mathbb{R}^d$, $d = 1, 2, 3$.
- Time interval: $[0, T]$.
- Random parameters: $\theta = [\theta_s, \theta_e] \in \Omega$.

θ_s, θ_e are the random parameters related to the source location and earth properties respectively.

The seismic wave propagation in time almost surely satisfies the following initial-boundary value problem:

$$\rho(\mathbf{x}, \theta_e)\partial_t^2 \mathbf{s}(\mathbf{x}, \theta, t) - \boldsymbol{\nabla}\cdot\mathbf{T}(\mathbf{s}(\mathbf{x}, \theta, t)) = \mathbf{f}(t, \mathbf{x}, \theta_s), \quad \text{in } D \times (0, T] \times \Omega, \tag{1a}$$

with initial conditions

$$\begin{aligned} \mathbf{s}(\mathbf{x}, \theta, t = 0) &= \mathbf{g}_1(\mathbf{x}), \\ \partial_t \mathbf{s}(\mathbf{x}, \theta, t = 0) &= \mathbf{g}_2(\mathbf{x}), \end{aligned} \quad \text{in } D, \tag{1b}$$

and the free surface boundary condition

$$\hat{\mathbf{n}} \cdot \mathbf{T} = 0, \quad \text{on } \partial D_S, \tag{1c}$$

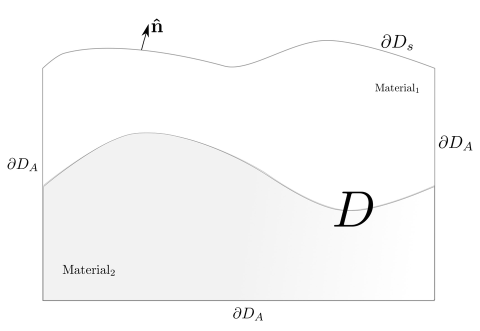


Figure: schematic of the domain.

where $\mathbf{s}$ is the displacement field and $\mathbf{T}(\mathbf{s})$ is the stress tensor, which, for an elastic and anisotropic solid can be written as

$$\mathbf{T} = \mathbf{c} : \nabla \mathbf{s},$$

for the fourth-order elastic tensor $\mathbf{c}$. On the artificial boundary, ∂D_A, absorbing boundary conditions are prescribed to simulate escaping waves. In our particular case we take $\mathbf{g_1}, \mathbf{g_2} = 0$ and $\mathbf{f}(t, \mathbf{x}, \theta_s) = M\nabla\delta(\mathbf{x} - \theta_\mathbf{s})k(t)$, where M is the moment tensor of the source, and $k(t)$ is a function representing the temporal component of the source.

Methodology:

Data $\mathbf{d}_r$ is measured at N_r receivers located at the surface and collected at N_t time instances. We assume $\mathbf{d_r}$ is polluted by some additive Gaussian noise $\mathbf{E}_{r,m}$

$$\mathbf{d}_r(t_m) = \mathbf{s}(\mathbf{x}_r, t_m; \theta) + \mathbf{E}_{r,m}, \quad E_{r,m} \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad r = 1, \dots N_r, \ m = 1, \dots N_t. \tag{2}$$

We are interested in finding the probability distribution of $\pi(\theta|\mathbf{d}_r)$. From Bayes' theorem:

$$\pi^{\text{post}}(\theta|\mathbf{d}_r) \propto \mathcal{L}(\mathbf{d}_r|\theta)\pi^0(\theta), \tag{3}$$

where we denote the posterior, likelihood and prior distributions as π^{post}, $\mathcal{L}$, and π^0, respectively. In the light of (2), the previous equation can be written as:

$$\pi^{\text{post}}(\theta|\mathbf{d}(t)) \propto \exp\left[-\frac{1}{2}\sum_{r=1}^{N_r}\sum_{m=1}^{N_t} \frac{|\mathbf{d}_r(t_m) - \mathbf{s}(\mathbf{x}_r, t_m; \theta)|^2}{\sigma^2}\right]\pi^0(\theta). \tag{4}$$

We use MCMC methods to obtain samples from said posterior.

Standard Metropolis-Hastings

The main idea behind the strategies presented here is the Metropolis-Hastings (MH) algorithm:

function METROPOLIS-HASTINGS$(N, \pi^{\text{post}}, q, \theta^1)$
  for $n = 1, 2, \dots, N-1$ **do**
    Sample $\theta^* \sim q(\cdot; \theta^n)$.
    Set $\theta^{n+1} = \theta^*$ with probability α, where

$$\alpha = \min\left[\frac{\pi^{\text{post}}(\theta^*)q(\theta^n; \theta^*)}{\pi^{\text{post}}(\theta^n)q(\theta^*; \theta)}, 1\right].$$

    Set $\theta^{n+1} = \theta^n$ otherwise.
  end for
  Output $\{\theta^n\}_{n=1}^N$.
end function

A common version of the Metropolis-Hastings algorithm is the Random Walk Metropolis (RWM) in which the proposals are generated by

$$\theta^* = \theta^n + \xi, \quad \xi \sim N(0, \Sigma).$$

We drop the super-index "post" for the remainder of the poster.

More Advanced Samplers

Pseudo-Marginal MH

Assume that $\pi(\theta_s)$ is intractable but that $\pi(\theta_s, \theta_e)$ is not, or that it is easier to evaluate. Since we are only interested in $\pi(\theta_s)$, we marginalize:

$$\pi(\theta_s) = \int \pi(\theta_s, \theta_e)d\theta_e \approx \frac{1}{M}\sum_{m=1}^{M} \pi(\theta_s, \theta_e^m) = \hat{\pi}(\theta_s; \bar{\theta}_e), \ \bar{\theta}_e = [\theta_e^1, \dots, \theta_e^M], \tag{5}$$

for $\theta_e \sim q_{\theta_e}$. We can estimate $\pi(\theta_s)$ by $\hat{\pi}(\theta_s; \bar{\theta}_e)$ since this is a positive and unbiased estimator. The algorithm is given by:

function PSEUDO-MARGINAL$(N, \hat{\pi}, q, q_u, \theta_s^1)$
  for $n = 1, 2, \dots, N-1$ **do**
    Sample $\theta_s^* \sim q(\cdot; \theta_s^n)$.
    Sample $\bar{\theta}_e \sim q_{\theta_e}(\cdot)$.
    Set $\theta_s^{n+1} = \theta_s^*$ with probability α, where

$$\alpha = \min\left[\frac{\hat{\pi}(\theta_s^*; \bar{\theta}_e)q(\theta_s^n; \theta_s^*)}{\hat{\pi}(\theta_s^n; \bar{\theta}_e)q(\theta_s^*; \theta_s^n)}, 1\right].$$

    Set $\theta_s^{n+1} = \theta_s^n$ otherwise.
  end for
  Output $\{\theta_s^n\}_{n=1}^N$.
end function

Notice that $\bar{\theta}_e$ is sampled at each step and it is used to compute the estimators $\hat{\pi}(\theta_s^n; \bar{\theta}_e), \hat{\pi}(\theta_s^*; \bar{\theta}_e)$.

Parallel-Tempering

Denote $\pi(\theta) \propto e^{-E(\theta)}$. When exploring multi-modal distributions it is possible that a RWM-type sampler will get stuck in one of the modes.

- Introduce K temperatures t_k, $1 = t_1 < t_2 < \dots < t_K$ to artificially "smooth" the density. This results in K different densities

$$\pi_k(\theta) \propto e^{-\frac{E(\theta)}{t_k}}.$$

- Run a chain for each t_k and propose to swap states between chains at two consecutive temperatures every N_{swap} iterations. Accept this swap with a Metropolis-Hastings step.

function PARALLEL TEMPERING$(K, \{t_k\}_{k=1}^K, N, N_{\text{swap}}, \pi, q, \theta_k^1)$
  for $n = 1, 2, \dots, N-1$ **do**
    for $k = 1, \dots, K$ **do**
      Sample $\theta_k^* \sim q(\cdot; \theta_k^n)$.
      Set $\theta_k^{n+1} = \theta_k^*$ with probability α_k, where

$$\alpha_k = \min\left[\frac{\pi_k(\theta_k^*)q(\theta_k^n; \theta_k^*)}{\pi_k(\theta_k^n)q(\theta_k^*; \theta_k^n)}, 1\right], \quad \text{with } \pi_k(\cdot) = \pi^{1/t_k}(\cdot).$$

      Set $\theta_k^{n+1} = \theta_k^n$ otherwise.
    end for
    if Mod$(n, N_{\text{swap}}) = 0$ **then**
      for $k = 1, \dots, K-1$ **do**
        Swap θ_k^{n+1} and θ_{k+1}^{n+1} with probability

$$\alpha_k = \min\left[\frac{\pi_k(\theta_{k+1}^{n+1})\pi_{k+1}(\theta_k^{n+1})}{\pi_k(\theta_k^{n+1})\pi_{k+1}(\theta_{k+1}^{n+1})}, 1\right].$$

      end for
    end if
  end for
  Output $\{\theta_1^n\}_{n=1}^N$.
end function

Multi-Level Sampler

We present a multi-level algorithm to accelerate the sampling from the posterior distribution. Define an increasing sequence of levels $\{l\}_{l=0}^L$ and an associated discretization h_l for the mathematical model, $h_l;\ h_0 > h_1 > \dots > h_L$. This induces a sequence of posteriors $\pi_0, \pi_1, \dots, \pi_L$. We run sequential chains for each l, using the information obtained at level $l-1$ to generate the samples at level l.

function MULTI-LEVEL MCMC$(K, L, \{N_l\}_{l=0}^L, \{\pi_l\}_{l=0}^L, q, \theta_0^1)$
  $\chi_0 =$ `Metropolis-Hastings`$(N_0, \pi_0, q, \theta_0^1)$
  for $l = 1, \dots, L$ **do**
    for $n = 1, \dots, N_l - 1$ **do**
      Sample $\hat{\theta}_{l-1} \sim \mu$, where $\mu =$ KDE(χ_{l-1}).
      Generate proposal $\theta^* \sim q(\cdot; \hat{\theta}_{l-1})$.
      Set $\theta_{l,j}^{n+1} = \theta^*$ with probability α_j, where $j = l, l-1$, and

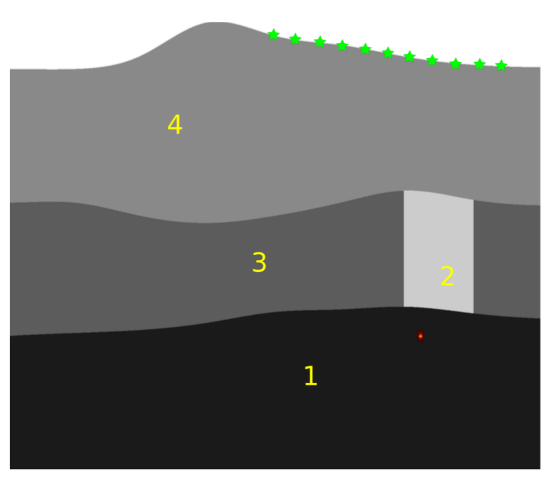
$$\alpha_j = \min\left[\frac{\pi_j(\theta^*)q(\theta_{l,j}^n; \theta^*)\mu(\theta_{l,j}^n)}{\pi_j(\theta_{l,j}^n)q(\theta^*; \theta_{l,j}^n)\mu(\theta^*)}, 1\right].$$

      Set $\theta_{l,j}^{n+1} = \theta_{l,j}^n$ otherwise.
    end for
    Set $\chi_l = \{\theta_{l,l}^n\}_{n=1}^{N_l}$
  end for
  Output $\{\theta_{L,L}^n\}_{n=1}^{N_L}$
end function

Notice that two chains $\{\theta_{l,l}\}_{n=1}^{N_l}$, $\{\theta_{l,l-1}\}_{n=1}^{N_l}$, are constructed at each level l, so that multi-level estimators such as those in [1, 2] can be computed.

Results

We run the experiments in 2 spatial dimensions in the domain shown bellow.



Max height: 3350 m. Width: 4000 m. Receivers: green stars. Source: red dot.

- **Spatial discretization:** Spectral Element Method, $[40 \times 60]$ elements, 5 GLL nodes/element.
- **Time-marching scheme:** Second-order accurate Newmark.
- **Implementation:** SPECFEM2D (see [4]).

(See also the poster *Multilevel Monte Carlo (MLMC) Acceleration of Seismic Wave Propagation under Uncertainty*).

Parameters:

- Source location $\theta_s = [\theta_{s,x}, \theta_{s,y}]$.
- Earth properties $\theta_e^i = [\rho^i, V_p^i, V_s^i]$, for each layer $i = 1, 2, 3, 4$.

ρ^i: material density, V_p^i: compressional velocity, V_s^i: shear wave velocity.

Data is obtained by solving (1), with θ shown below and adding Gaussian noise with $\sigma = 2 \times 10^{-3}$.

True Values			
θ_s			
$\theta_{s,x} = 3100$		$\theta_{s,y} = 1000$	
θ_e			
Layer	ρ	V_p	V_s
1	2700	3000	1732
2	2500	2700	0
3	2200	2500	1443
4	2200	2000	1343

We use uniform priors:

$$\theta_{s,x} \sim \text{U}(700, 3700), \ \theta_{s,y} \sim \text{U}(300, 2000), \ \theta_e^i \sim \text{U}(\theta_e^i - 200, \theta_e^i + 200).$$

We perform the inversion experiments by obtaining 10000 samples from the posterior distribution with each algorithm.

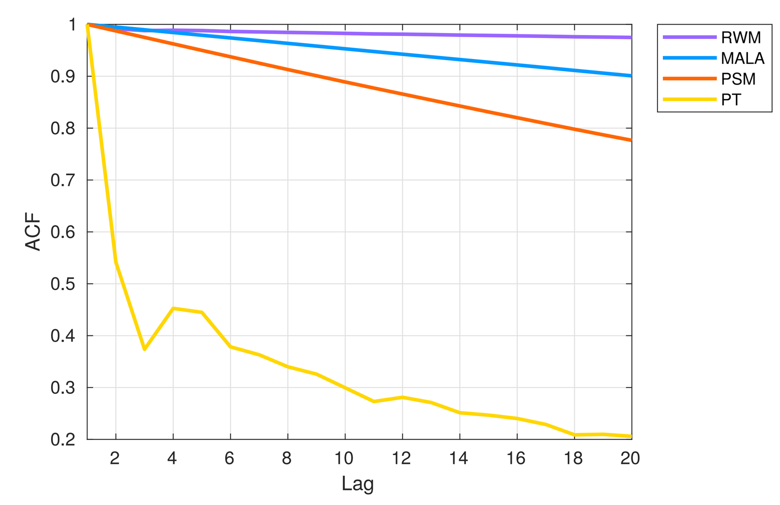


Figure: Autocorrelation function for the x component of the source $\theta_{s,x}$ at lag 20 for different methods; Random Walk Metropolis (RWM) Metropolis-Adjusted Langevin Algorithm (MALA), Pseudo-Marginal (PSM) and Parallel Tempering (PT). The autocorrelation time can be thought of as a measure of the inefficiency of the sampler.

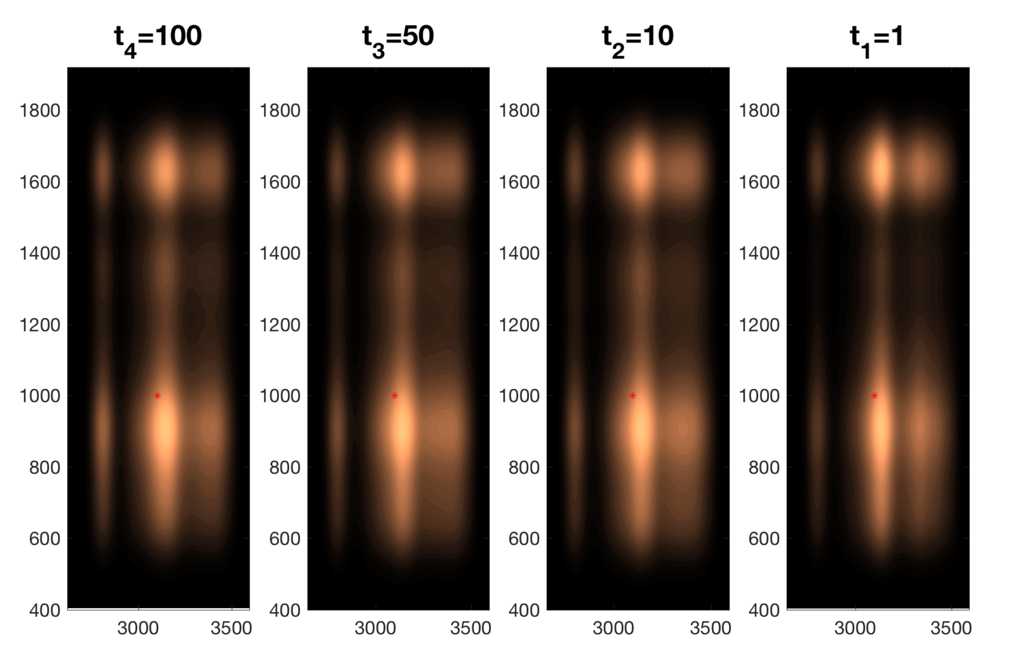


Figure: Comparison of the posterior density estimate for the parallel tempering algorithm for different temperatures.

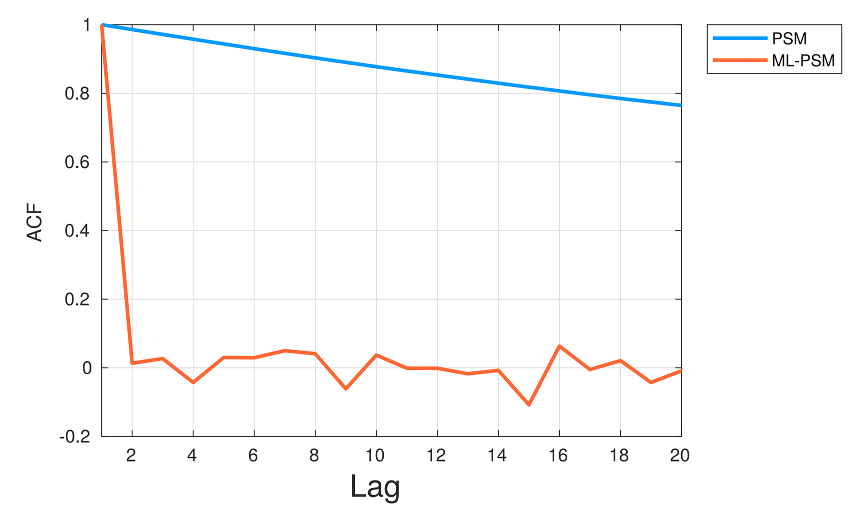


Figure: Comparison of the autocorrelation for the samples of $\theta_{s,x}$ for the pseudo-marginal and a 2-level version of the pseudo-marginal. We can see that the multi-level algorithm results in almost i.i.d samples from the posterior distribution.

Conclusions

- In general, Bayesian seismic source inversion via MCMC methods is a difficult problem to tackle given the high parameter space, the limited a-priori information and the possible multi-modality of the posterior distribution. We can see that using alternative samples strategies can yield better results than RWM or MALA.
- The multilevel algorithm yields promising experimental results as it seems to generate (almost) i.i.d samples faster than other samplers, at the one-time cost of running the chain at the previous levels.

Acknowledgements

The authors acknowledge the support from the KAUST Grant OSR-2015-CRG4-2585-01: "Advanced Multi-Level sampling techniques for Bayesian Inverse Problems with applications to subsurface". The authors also acknowledge the support from the Center for ADvanced MOdeling Science (CADMOS).

References

[1] Giles, Michael B. "Multilevel Monte Carlo path simulation." Operations Research 56.3 (2008): 607-617.

[2] Dodwell, Tim J., et al. "A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow." SIAM/ASA Journal on Uncertainty Quantification 3.1 (2015): 1075-1108.

[3] Earl, David J., and Michael W. Deem. "Parallel tempering: Theory, applications, and new perspectives."Physical Chemistry Chemical Physics 7.23 (2005): 3910-3916.

[4] Tromp, D. Komatitsch, and Q. Liu. "Spectral-element and adjoint methods in seismology". Communications in Computational Physics, 3(1):1?32, 2008.

[5] Andrieu, Christophe, and Gareth O. Roberts. "The pseudo-marginal approach for efficient Monte Carlo computations." The Annals of Statistics 37.2 (2009): 697-725.