ECOLE DOCTORALE
DOCTORAL SCHOOL

PROGRAMME DOCTORALE EN MATHEMATIQUE (EDMA)
DOCTORAL PROGRAM IN MATHEMATICS (EDMA)

Research Plan

Titre provisoire de la thèse / Provisional thesis title

Candidate: Juan Pablo Madrigal Cianci

Advisor: Fabio Nobile

July 2, 2018

Contents

1 Introduction, Setup, and Objectives

ALMOST FINAL DRAFT

1.1 Introduction

This thesis project focuses on the development and implementation of Markov Chain Monte Carlo (MCMC) techniques, as well as their multi-level and multi-index extensions (MLMCMC and MIMCMC respectively) techniques for Bayesian inverse problems (BIP) arising in seismic and geophysical applications; such as seismic inversion and earthquake source estimation. The goal of these type of problems is to determine a set of parameters $\boldsymbol{\theta}$, which identify e.g, the earthquake source characteristics, based on few recorded waveforms at the surface. We will use the elasto-dynamic wave equation to model such seismic event, and will take the recorded waveform to be the time series of the ground displacement[1] at different receivers scattered throughout the surface. In this case, we take the set of unknown parameters $\boldsymbol{\theta}$ to be properties related to the source function (location, set-off time, moment tensor), as well as the material properties of the medium, such as density and Lamé parameters. Traditionally, seismic inversion has been performed using a deterministic approach, by introducing a cost functional $J(\boldsymbol{\theta})$ measuring the misfit between the modeled surface waveforms and the observed ones. This cost functional is then minimized using traditional optimization algorithm to find a minimizer (see, e.g, [7,38]). In statistical terms, this approach can often be interpreted as a maximum likelihood point estimation. However, there is a growing need to quantify the uncertainty in these methods [25], by adopting, for instance, a Bayesian approach and sampling techniques such as MCMC. The aim of this work is to develop accelerated and efficient MCMC algorithms to tackle this problem.

1.2 Objectives

Modern computing facilities and computational techniques are starting to make Bayesian inversion approaches and MCMC feasible for large scale inverse problems involving partial differential equations (PDEs) [37].There is a wealth of literature devoted to those problem arising from elliptic PDEs (see, for example [37]), however, this is otherwise rather scarce for hyperbolic equations. One of the first objectives from this work is then to extend the BIP theory to cover these type of problems. An additional issue with a MCMC computation is its inherent cost. Contrary to plain Monte Carlo, MCMC is sequential in nature and

[1]We remark however that in practice these receiver can also measure other wave related quantities, such as velocity or acceleration of the ground at the recording position.

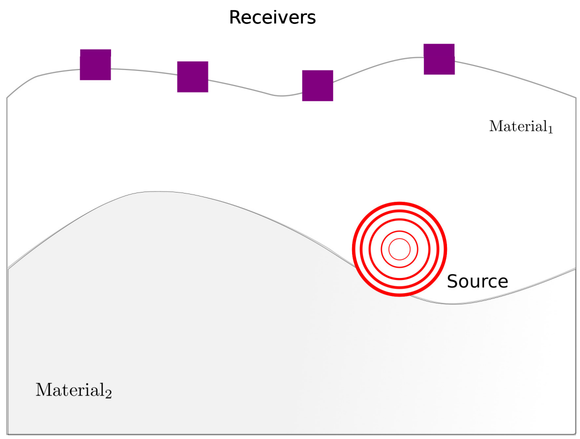


Figure 1 – Experimental setup of the problem. An earthquake generated at source travels through material properties and the displacement of the earth is recorded at the receivers (purple squares). Based on the recording data of the receivers, we employ MCMC techniques to recover the source location.

produces samples that are, in general, correlated. One proposed idea to mitigate this burden is the use of multi-level samplers [8,9,12], for which most of the work is done using a coarse discretization of the PDE, and the sample thus obtained is then "corrected" using more refined discretization levels. there are few results available so far. This is a rather new approach to MCMC techniques and as such, there are rather few results available so far. A further improvement on the multi-level sampler is offered by the multi-index technique [10] there are the multi-index samplers, for which the computational cost can be reduced even further by taking different discretization directions. .

1.3 Problem Set-up

We are interested in determining the earthquake properties of a seismic event. Experimentally, data $\mathbf{y} \in \mathbb{R}^d$, assumed to be polluted by some additive random noise η, is recorded at N_r receivers and at N_t time instances. This receivers can be located, for example, at the surface of the earth or in observation wells. The recorded data can be described by an observation operator $\mathcal{F}$ applied to a realization of a model for the seismic wave propagation, $\mathbf{u} := \mathbf{u}(\boldsymbol{\theta})$, viewed as a map from a parameter space Ω to $\mathbb{R}^d$. Based on this, our aim is then to determine a conditional probability distribution $\pi(\boldsymbol{\theta}|y)$ of a set of parameters $\theta \in \Omega$ such that $\mathcal{F}(\boldsymbol{\theta}) = \mathcal{F}(\mathbf{u}(\boldsymbol{\theta}))$ closely resembles the measured data $\boldsymbol{y}$. In the following subsections we will describe in more detail the model used to describe the seismic motion, the observation operator, and the Bayesian inversion procedure used to obtain $\pi(\boldsymbol{\theta}|y)$.

1.3.1 Forward Problem

We follow the analysis presented in [15, 18, 29, 30]. We denote the spatial variable by $\mathbf{x}$ and the set of unknown parameters by $\boldsymbol{\theta}$. Consider an elastic, non-homogeneous medium in an open, bounded set $D \subset \mathbb{R}^d$, with $\partial D = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, $|\Gamma_D|,\ |\Gamma_N| > 0$, for $d = 2, 3$ and a time interval $I = [0, T]$. Given some complete probability space (Ω, F, P), where Ω is...we can pose the following stochastic initial boundary value problem given by

$$\varrho(\mathbf{x},\boldsymbol{\theta})\mathbf{u}_{tt}(t,\mathbf{x},\boldsymbol{\theta}) - \nabla\cdot\boldsymbol{\sigma}(\mathbf{u}(t,\mathbf{x},\boldsymbol{\theta})) = \mathbf{f}(t,\mathbf{x},\boldsymbol{\theta}) \qquad \text{in } I\times D\times\Omega \tag{1a}$$

$$\mathbf{u}(0,\mathbf{x},\boldsymbol{\theta}) = \mathbf{g_1}(\mathbf{x}), \quad \mathbf{u}_t(0,x,\boldsymbol{\theta}) = \mathbf{g}_2(\mathbf{x}) \qquad \text{on } \times D\times\Omega \tag{1b}$$

$$\boldsymbol{\sigma}(\mathbf{u}(t,\mathbf{x},\boldsymbol{\theta}))\cdot\mathbf{n} = \mathbf{0} \qquad \text{on } I\times\Gamma_N\times\Omega, \tag{1c}$$

$$\mathbf{u}(t,\mathbf{x},\boldsymbol{\theta}) = \mathbf{0} \qquad \text{on } I\times\Gamma_D\times\Omega, \tag{1d}$$

where $\mathbf{u} : I\times D\times\Omega \to \mathbb{R}^d$, $\mathbf{x}\in\mathbb{R}^d, \boldsymbol{\theta}\in\Omega, t\in I$, and with $\boldsymbol{\sigma} = \lambda(\mathbf{x},\boldsymbol{\theta})\nabla\cdot\mathbf{u}I + \mu(\mathbf{x},\boldsymbol{\theta})(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)$. Here we denote by $\mathbf{u}$ the displacement vector, by $\varrho > 0$ the material density, by λ, μ the Lamé parameters, and by $\mathbf{f}$ a (generalized) body force.

Assumption 1. *include comment on assumptionsLet $s \geq 0$ be an integer. We assume that*

$$\mathbf{f}(\cdot,\cdot,\boldsymbol{\theta}) \in \mathbf{L}^2([0,T];\mathbf{H}^s(D)) \text{ for a.e. } \boldsymbol{\theta}\in\Omega, \quad \mathbf{g}_1 \in \mathbf{H}^{s+1}(D), \quad \mathbf{g_2}\in\mathbf{H}^s(D), \quad s\geq 0. \tag{2}$$

Under Assumption 1, problem (1) admits a unique solution for a.e $\boldsymbol{\theta}\in\Omega$. We can then interpret the solution $\mathbf{u}(t,\mathbf{x},\boldsymbol{\theta})$ as a Banach space-valued function on Ω [30],

$$\mathbf{u} = \mathbf{u}(\boldsymbol{\theta}) : \Omega \to \mathbf{V} := \mathbf{C^0}([\mathbf{0},\mathbf{T}];\mathbf{H^{s+1}}(\mathbf{D}))\cap\mathbf{C^1}([\mathbf{0},\mathbf{T}];\mathbf{H^s}(\mathbf{D}))\cap\mathbf{H^2}([\mathbf{0},\mathbf{T}];\mathbf{H^{s-1}}(\mathbf{D})), \ \ \mathbf{s}\geq\mathbf{0}. \tag{3}$$

In turn, (3) is a weak solution to problem (1) provided that at $t = 0$, $\mathbf{u}(\boldsymbol{\theta}) = \mathbf{g}_1$, $\mathbf{u}_t(\boldsymbol{\theta}) = \mathbf{g}_2$ and that for all test functions $\mathbf{v}\in\mathbf{C}_0^\infty([0,T];H^1(D))$ the following weak formulation holds

$$\int_0^T\int_D (\varrho\mathbf{u}_{tt}\cdot\mathbf{v} + \nabla\mathbf{v} : \boldsymbol{\sigma}(\mathbf{u}))\,d\mathbf{x}dt = \int_0^T\int_D \mathbf{f}\cdot\mathbf{v}d\mathbf{x}dt. \tag{4}$$

For the analysis of the well possedness of the Bayesian inverse problem we will need $\mathbf{u}$ to be Lipschitz in Ω. To this end, we quote,recall,...the following theorem, presented in [30]:

Theorem 1. *For the solution of problem (1) with data given by Assumption 1 check on additional assumptions on Motamed's paper and with uniformly elliptic material properties, we have that*

$$\partial_{\theta_n}\mathbf{u}(\boldsymbol{\theta}) \in \mathbf{C}^0([0,T];\mathbf{H}^s(D)), \quad s\geq 0, \tag{5}$$

where θ_n denotes the n^{th} component of $\boldsymbol{\theta}$. Moreover, $\partial_{\theta_n}\mathbf{u}(\boldsymbol{\theta})$ is uniformly bounded in Ω.

Proof. The previous theorem is a particular case of Theorem 3 in [30], for the case $k = 1$, what is k?and $s \geq 0$. □

Corollary 1. *Under the assumptions of the previous theorem, $\mathbf{u}(\boldsymbol{\theta})$ is Lipschitz continuous, when seen as a map from Ω tp $\mathbf{C}^0([0,T];\mathbf{H}^s(D))$.*

For the purpose of this project, system (1) is approximated numerically, using a spectral element method for the space discretization and the leapfrog method for the time marching scheme.

1.3.2 Observation Operator

for the application at hand, data is gathered using an observation operator. As such, we will also need to study the properties of this operator in order to properly formulate the well possedness of the BIP. We define the forward map $\boldsymbol{\mathcal{F}} : \Omega \to \mathbb{R}^{m\times N_{rec}}$ by

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = (\mathcal{O}_1(\mathbf{u}(\boldsymbol{\theta})), \ldots, \mathcal{O}_{N_{rec}}(\mathbf{u}(\boldsymbol{\theta})),$$

where $\mathcal{O}_i \in (\mathbf{V}')^m$ is a linear observation operator at times $\{t_1, \ldots, t_m\}$, associated to the i^{th} receiver. In many problems in seismic inversion, data is only available at discrete points in the spatial domain, which in turn implies that the observation operator lacks enough regularity. To alleviate this, we define our i^{th} linear operator as

$$\mathcal{O}_i(\mathbf{v}) = \left(\frac{1}{|\mathcal{B}(\mathbf{x}_i, r)|}\int_{\mathcal{B}(\mathbf{x}_i,r)} \mathbf{v}(t_1, \mathbf{x}), \ldots, \frac{1}{|\mathcal{B}(\mathbf{x}_i, r)|}\int_{\mathcal{B}(\mathbf{x}_i,r)} \mathbf{v}(t_m, \mathbf{x})\right)^T, \tag{6}$$

for any $\mathbf{v} \in C^0([0,T], L^2(D))$, and where $\mathcal{B}(\mathbf{x}_i, r)$ denotes the ball centered at $\mathbf{x}_i$ with radius r. Let us denote

$$|\boldsymbol{\mathcal{F}}(\boldsymbol{\theta})|^2 = \sum_{i=1}^{N_{rec}} \max_j |\mathcal{O}_{i,j}(\mathbf{u}(\boldsymbol{\theta}))|^2 .$$

Based on this, we can show that $\boldsymbol{\mathcal{F}}$ is Lipschitz continuous:

Proposition 1. *Under the assumption of uniform ellipticity define this or state as an assumptionon the material properties, $\boldsymbol{\mathcal{F}}$ is Lipschitz continuous.*

not too happy with how equation is being displayed

Proof. We begin by noting that

$$\begin{aligned}|\mathcal{O}_{i,j}(\mathbf{v})| &= \left|\frac{1}{|\mathcal{B}(\mathbf{x}_i, r)|}\int_{\mathcal{B}(\mathbf{x}_i,r)} \mathbf{v}(t_j, \mathbf{x}) d\mathbf{x}\right| \\ &\leq \sqrt{\frac{1}{|\mathcal{B}(\mathbf{x}_i, r)|}\|\mathbf{v}(t_j, \cdot)\|^2_{L^2(\mathcal{B}(\mathbf{x}_i,r))}} \leq \sqrt{\frac{1}{|\mathcal{B}(\mathbf{x}_i, r)|}\|\mathbf{v}(\boldsymbol{\theta})\|^2_{C^0([0,T];L^2(\mathcal{B}(\mathbf{x}_i,r)))}}\end{aligned} \tag{7}$$

This in turn implies that $\forall\ \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Omega$,

$$|\mathcal{F}(\boldsymbol{\theta}) - \mathcal{F}(\boldsymbol{\theta}')| \leq \|\mathbf{u}(\boldsymbol{\theta}) - \mathbf{u}(\boldsymbol{\theta}')\|_V .$$

Moreover, by Theorem 1, we have that there exist a $C_L > 0$ such that

$$\|\mathbf{u}(\boldsymbol{\theta}) - \mathbf{u}(\boldsymbol{\theta}')\|_V \leq C_L \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_{L^2(\Omega)},$$

thus,

$$|\mathcal{F}(\boldsymbol{\theta}) - \mathcal{F}(\boldsymbol{\theta}')| \leq C_L \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_{L^2(\Omega)}.$$

□

1.3.3 Bayesian Formulation

We consider a bounded and finite parameter space Ω explain with bounded.We will use the assumption of additive noise, i.e, we consider the case where we are given data y polluted by some noise η, with $\eta, y \in \mathbb{R}^q$ and modeled by some forward map or linear observation operator $\mathcal{F} : \Omega \to \mathbb{R}^q$ such that

$$y = \mathcal{F}(\boldsymbol{\theta}) + \eta, \quad \eta \sim g, \tag{8}$$

where f is some arbitrary probability distribution. Given the noisy, recorded data y, we are interested in determining a set of parameters $\boldsymbol{\theta}$ such that when used as input in 1, we obtain a wave form that closely resembles the one observed. To do so, we adopt a Bayesian approach to the problem of determining $\boldsymbol{\theta}$ from y. Roughly speaking, contrary to the deterministic counterpart, where only point estimates are obtained, the Bayesian approach will lead to the notion of finding a probability measure π^y in Ω, containing information about the relative probability of different states $\boldsymbol{\theta}$, given the data y. Incorporating our prior beliefs in a density π^0 and denoting the conditional probability of $\boldsymbol{\theta}$ given y, by π^y, we have that from Bayes theorem

$$\pi^y \propto g(y - \mathcal{F}(\boldsymbol{\theta}))\pi^0(\boldsymbol{\theta}), \tag{9}$$

Moreover, we choose priors π^0 such that $\pi(\Omega) = 1$, that is, the posterior is absolutely continuous with respect to the prior. Using Theorem 6.31 in [37], we can rewrite the previous expression in terms of the Radon-Nikodym derivative

$$\frac{d\pi^y}{d\pi^0} \propto \exp(-\Phi(\theta, y)), \tag{10}$$

where we are abusing notation and representing both density and distribution with the symbol π, and where we call $\Phi(\boldsymbol{\theta}, y)$ the potential function. For simplicity, we will limit ourselves to the case of additive Gaussian noise case;

$$g = \mathcal{N}(0, \Sigma).$$

ALMOST FINAL DRAFT; DON'T KNOW IF I SHOULD INCLUDE A BRIEF DESCRIPTION OF MCMC

2 State of the Art

We now discuss the current state of the art of the research at hand. There are two main points to address in this respect; the first one relates to the current state of seismic source inversion, and the second one addresses the state of multi-level and multi-index Markov chain Monte Carlo methods. The former has been traditionally dominated by deterministic inversion, with the use of the more robust, MCMC based probabilistic inversion having been recently introduced. The latter tries to extend the ideas of multi-level and multi-index Monte Carlo to an MCMC setting. This in turn is still at a very early stage, with only a handful of papers discussing this ideas.

2.1 Seismic Inversion

Some things in this section borrow form the CRG4 grant proposal...How can I cite that? In earthquake inversion, we try to estimate the material properties (such as velocity, density and Lamé parameters) or the kinematic parameters of the earthquake rupture process that occurs on a geological fault plane of finite spatial extent (finite-fault) within the earth. These parameters can be the point-source location, the spatially variable displacement across the fault surface (slip), the slip direction, the slip duration (which is tied to a poorly known slip-rate function), and the rupture time (time at which each point on the fault starts to slip; this is tied to a rupture velocity). To this end, seismologists have developed numerous techniques over the past decades. Traditionally, a deterministic approach has been employed, on which a misfit functional $J(\boldsymbol{\theta})$ (where $\boldsymbol{\theta}$ represents the vector of parameters of interest) measuring the difference between some recorded data y and some generated data $\mathcal{F}(\boldsymbol{\theta})$ is defined and minimized. Thus the deterministic inversion reads as a constrained optimization problem

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Omega}} \Phi(\boldsymbol{\theta})$$

with

$$\Phi(\boldsymbol{\theta}) = J(\boldsymbol{\theta}) - \frac{\alpha}{2}R, \quad \text{such that (1) holds,}$$

where R is a regularizing term (such as Tykhonov or total variation) and α is some regularization parameter. This approach has been studied extensively [13,19] and it is still a fairly active area of research in the geophysics community, however, these methods are prone to discrepancies [4]. An example can be seen for the 2011 Tohoku earthquake (Japan), where a survey over more than 20 different inversion methods described in various papers yields widely different results for the source location (see, for example, [11,25,36]). This calls for models that can accurately quantify uncertainty.

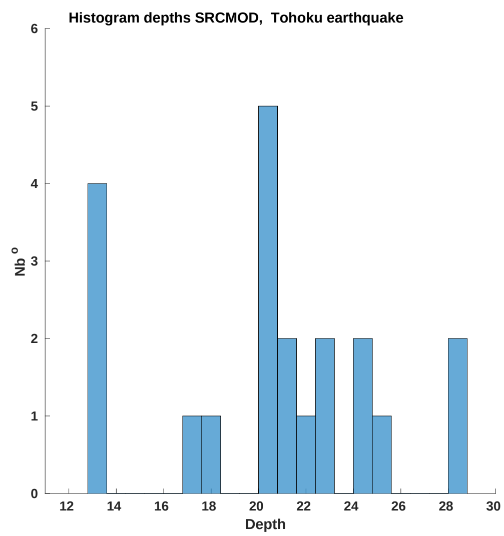


Figure 2 – Histogram of different source locations obtained for the 2011 Mw 9.0 Tohoku earthquake.

More recently, full Bayesian estimation of earthquake properties has been attempted (see, for example [6,27]) although it is still in its early stages and geared mostly towards inversion on the material properties of the medium, rather than on the physical properties of the earthquake. This lack of development of this type of methods is mostly linked to the large cost associated to obtaining numerical simulations of the forward problem of local, regional, or global seismic wave propagation. To mitigate this computational burden, various simplifying assumptions are typically made. Thus, a complete or comprehensive quantification of the uncertainties in the resulting finite-fault rupture models, owing to (a) data errors; (b) unknown earth structure; (c) unknown rupture physics (e.g. detailed fault geometry; poorly known slip-rate function); (d) simplification in computing the response functions is still missing in the literature.

2.2 Markov Chain Monte Carlo

2.3 Muti-level and Multi-index Markov Chain Monte Carlo

The work done on ML/MI-MCMC is still on its very early stages, with only a handful of papers on the topic [8,12,14] , focusing mostly on inverse problems involving elliptic PDEs. To the best of our knowledge, there has not been a rigorous attempt to extend these methods to inverse problems involving hyperbolic PDEs. Moreover, the algorithms discussed therein are open to improvements. To better understand the theory behind these methods, we first discuss briefly the ideas underlying MLMC and MIMC, as ML/MI-MCMC are an extension of these to MCMC. MLMC are algorithms for computing expectations that arise in stochastic simulations in cases in which the stochastic model can not be simulated exactly, rather approximated at different levels of accuracy and as such, at different computational costs. Just as Monte Carlo methods, they rely on repeated random sampling, but these samples are taken on different levels ℓ of accuracy (cheap simulations), in such a way that the majority of the samples M_ℓ are obtained with low accuracy. In the case of PDE driven inverse problems, this can be understood as introducing a hierarchy of discretizations $\{h_\ell\}_{\ell=0}^L$ such that the cost of solving the PDE and the accuracy of the numerical solution increase as ℓ increases. MLMC methods can greatly reduce the computational cost of standard Monte Carlo methods by taking most of the samples at a low accuracy and corresponding low cost, and only very few samples at high accuracy and corresponding high cost. In the standard MLMC approach we denote the resulting approximation of a quantity of interest using mesh size h_ℓ in the simulation of a forward observation operator event by $\mathrm{QoI}_\ell(\boldsymbol{\theta})$. Then, the expected value of the finest approximation, QoI_L, can be expressed as

$$\mathbb{E}[\mathrm{QoI}_L] = \mathbb{E}[\mathrm{QoI}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathrm{QoI}_\ell - \mathrm{QoI}_{\ell-1}],$$

and the MLMC estimator is obtained by approximating the expected values in the telescoping sum by sample averages as

$$\mathcal{A} = \frac{1}{M_0}\sum_{m=1}^{M_0} \mathrm{QoI}_0(\boldsymbol{\theta}_{0,m}) + \sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{m=1}^{M_\ell}\left(\mathrm{QoI}_\ell(\boldsymbol{\theta}_{\ell,m}) - \mathrm{QoI}_{\ell-1}(\boldsymbol{\theta}_{\ell,m})\right), \tag{11}$$

where, for every ℓ, $\{\boldsymbol{\theta}_{\ell,m}\}_{m=1}^{M_\ell}$ denotes independent identically distributed (i.i.d.) realizations of the mesh-independent random variables $\boldsymbol{\theta}$. If W_ℓ is the average cost associated with generating one sample of the difference, $\mathrm{QoI}_\ell - \mathrm{QoI}_{\ell-1}$, or simply QoI_0 if $\ell = 0$, then the cost of the estimator (11) is

$$\mathcal{W} = \sum_{\ell=0}^{L} M_\ell W_\ell. \tag{12}$$

We assume that the work required to generate one sample of mesh size h is proportional to $h^{-\gamma}d$, where d is the dimension of the computational domain and $\gamma > 0$ represents the complexity of generating one sample with respect to the number of degrees of freedom. Thus, we model the average cost on level ℓ as

$$W_\ell \leq h_\ell^{-\gamma}d, \tag{13}$$

and consequently use the representation

$$\mathcal{W} \leq \sum_{\ell=0}^{L} \frac{M_\ell}{h_\ell^\gamma} \tag{14}$$

for the total work to evaluate the MLMC estimator (11). It can be shown that under certain conditions on the ccuracy of the PDE solver and on the variance of the increments, for a given tolerance ϵ, there exist a constant c_{mlmc}

$$\mathbb{E}[\mathcal{W}] \leq \begin{cases} c_{\text{mlmc}}\epsilon^{-2} & \beta > \gamma \\ c_{\text{mlmc}}\epsilon^{-2}\log(\epsilon)^2 & \beta = \gamma \\ c_{\text{mlmc}}\epsilon^{-2-(\gamma-\beta)/\alpha} & \beta < \gamma, \end{cases} \tag{15}$$

where α is related to the numerical accuracy of the PDE solver and β is related to the decay rate of the variance of the increments [9]. A further improvement over the non-optimal regime ($\beta > \gamma$) can be obtained using multi-index Monte Carlo (MIMC), for which different discretization directions are used [10]. These methods have successfully been implemented for a wide variety of problems, including simulation of random paths, option pricing etc. Extending these methods to a MCMC setting is not a trivial task since (i) the MLMCMC algorithm must be constructed in such a way that samples in two continuous chains are correlated (so that the variance of the increment in QoI_ℓ decays as ℓ increases) and the algorithm should, ideally generate samples with a small autocorrelation time within chains, since, as opposed to pure Monte Carlo, MCMC generates samples that are correlated, which will generate a bias on the estimator. As mentioned before, only a few works so far have dealt with the ML-MI extension of MCMC methods for posterior exploration in Bayesian inversion. Of theoretical importance is the work of [12], as it provides a very thorough analysis to a particular version of these methods applied to inverse problems involving elliptic PDEs. However, the method described in that paper generates proposals based on independent samplers (see [3]), and this in turn can become widely inefficient if the proposal generating function is not constructed carefully. The work of [8] proposes a more efficient algorithm and provides an estimate for the total work necessary to obtain a tolerance of ϵ, given by a form similar to (15).

Definition 1. *We define the mean squared error (MSE) associated to a MCMC algorithm* *by*

$$e^2_{MSE}(QoI_\ell) = \mathbb{E}_\vartheta\left[\left(\frac{1}{M_\ell}\sum_i^{M_\ell} QoI^i - \mathbb{E}_\mu(QoI)\right)^2\right],$$

where QoI is some quantity we wish to estimate, following some (unachievable) distribution μ *and where* ϑ *is the distribution from which we are actually able to sample.*

Definition 2. *Suppose each sample at level* ℓ *has an associated cost* $\mathcal{W}_\ell$*. We denote the* ε*-cost, i.e, the cost associated to obtaining a mean squared error of* ε *when estimating QoI by* $\mathcal{W}^\epsilon_\ell(QoI)$*.*

Theorem 2. *Denote by* $Y_\ell =$ *Suppose that all chain have been sufficiently burned-in and that there exist* $\alpha, \beta, \gamma > 0$*, and* $\varepsilon < e^{-1}$*, such that* $\alpha \geq \min(\beta, \gamma)$*. Suppose the following assumptions hold for all* $\ell \geq 0$*:*

i) $\left|\mathbb{E}_{\pi_{\ell,h}}[Q] - \mathbb{E}_{\pi_\ell[Q]}\right| \leq C_1 h^\alpha_\ell$

ii) $\mathbb{V}_{\pi_{\ell,h}}[Y_\ell] \leq C_2 h^\beta_\ell$

iii) $e_{MSE}(Y_\ell) \leq C_3 \sup|\mathbb{V}_\ell[Y_\ell]| M_\ell^{-1/2}$

iv) $W^\epsilon_\ell(QoI) \leq C_4 h_\ell^{-\gamma}$,

then, there exist a number of levels L *and a sequence of* $\{N_\ell\}_{\ell=0}^L$ *such that*

$$e(QoI_L) \leq \varepsilon,$$

and

$$\mathcal{W}^\epsilon_L(QoI) \leq C_5 \begin{cases} \varepsilon^{-2}|\log\varepsilon|, & if\ \ \beta > \gamma, \\ \varepsilon^{-2}|\log\varepsilon|^3, & if\ \ \beta = \gamma, \\ \varepsilon^{-2+(\gamma-\beta)/\alpha}|\log\varepsilon|, & if\ \ \beta < \gamma. \end{cases} \tag{16}$$

WAITING FOR SOME RESULTS FROM THE CLUSTER, WILL UPDATE SOON

3 State of Our Research and First Results

We have focused mainly on three topics that tackle the problem at hand, both theoretically and computationally. More precisely,

i We have studied and proved the well possedness of the BIP for the hyperbolic equation PDE at hand,

ii We have implemented different sampling strategies to a case study inversion problem. These methods include random walk Metropolis (RWM), adaptive Metropolis, delayed rejection adaptive metropolis (DRAM), and parallel tempering.

iii We have proposed a MLMCMC algorithm and tried different sampling strategies.

In addition, we have also started to look into other research directions which shall be discussed in Section 4. We proceed to discuss in more detail the three points explained above.

3.1 On the Well-Possedness of the BIP

We mention the results pertaining the well possedness of the BIP. The results shown herein are an extension to [8, 12, 37], and can be found in more detail in the upcoming work [24]. For simplicity, we omit most of the proofs relating to the minor results.

Definition 3. *Let μ, ν be two distributions with Radon-Nikodym derivative with respect to a common measure λ given by $\frac{d\mu}{d\lambda}$, $\frac{d\nu}{d\lambda}$, respectively. The Hellinger distance between μ and ν is defined as*

$$d^2_{Hell}(\mu, \nu) = \int_X \left(\sqrt{\frac{d\nu}{d\lambda}} - \sqrt{\frac{d\mu}{d\lambda}} \right)^2 d\lambda. \tag{17}$$

Notice that the Hellinger distance is ralated to the total variation (TV) norm by

$$d^2_{\text{Hell}}(\mu, \nu) \leq d_{\text{TV}}(\mu, \nu) \leq \sqrt{2} d_{\text{Hell}}(\mu, \nu),$$

which follows from the relationship between the 1 and 2 norms.

Proposition 2. *Let $\boldsymbol{\theta} \in \Omega$, where Ω is a bounded domain. Moreover, Assume that for each component of $\boldsymbol{\theta}$, $0 < \boldsymbol{\theta}^i_{\min} \leq \boldsymbol{\theta}^i \leq \boldsymbol{\theta}^i_{\max} < \infty$. Then $\mathbf{u}(\boldsymbol{\theta})$ is uniformly bounded in Ω.*

The proof of the preceding theorem can be found in [30].

Given a covariance matrix Σ, we define the potential function (i.e, the negative log-likelihood) $\Phi(\boldsymbol{\theta}; y)$ by

$$\Phi(\boldsymbol{\theta}; y) = \frac{1}{2}|y - \mathcal{F}(\boldsymbol{\theta})|^2_{\Sigma^{-1}}. \tag{18}$$

Analogously, we define the potential arising from a discretized PDE with discretization parameter h_ℓ by $\Phi_\ell(\boldsymbol{\theta}; y)$.

Proposition 3. *We have that, for each level ℓ, the potential function is locally bounded, i.e, there exists $\Phi^M(r)$ such that*

$$0 \leq \Phi_\ell(\boldsymbol{\theta}; y) \leq \Phi^M_\ell(r).$$

Additionally, this potential function is locally Lipschitz with respect to both the data y and the parameter $\boldsymbol{\theta}$, i.e, there is a mapping $G : \mathbb{R} \times \boldsymbol{\theta} \to \mathbb{R}$ and a constant C_{pot} such that for each $r > 0$, $G(r, \cdot) \in L^2(\boldsymbol{\theta})$; and for every $|y|, |y'|$, it holds that

$$|\Phi_\ell(\boldsymbol{\theta}; y) - \Phi_\ell(\boldsymbol{\theta}; y')| \leq G(r, \boldsymbol{\theta})|y - y'|_{\Sigma^{-1}}, \tag{19}$$

and

$$|\Phi_\ell(\boldsymbol{\theta}; y) - \Phi_\ell(\boldsymbol{\theta}'; y)| \leq C_{pot}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_{L^2(\Omega)}. \tag{20}$$

Proposition 4. *For each positive constant r there is a positive constant $C(r)$ such that if $|y|_\Sigma, |y|_\Sigma \leq r$, then*

$$d_{Hell}(\pi^y, \pi^{y'}) \leq C(r)|y - y'|_\Sigma.$$

Proof. The proof follows from that of proposition 25 in [12]. □

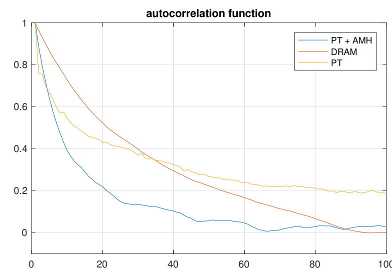
Theorem 3. *Under the preceding propositions, the Bayesian inverse problem is well posed.*

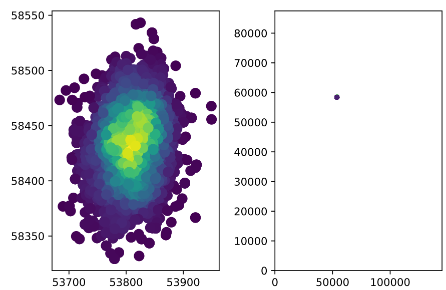
Waiting for some results, need to update the problem description here, as well as including pictures on the domain, etc. I have some older ones that I obtained with the desktop pc, but I re-run the experiments in the cluster.

3.2 Solving the Inverse Problem: The Tanzania Case Study

We now address the results claimed in (ii). In particular, we implemented different MCMC algorithms and applied them to a simplified seismic source inversion problem. More precisely, we test the data on a 2-dimensional model of an earthquake that took place on the Tanzania basin on the 12th of October 2016 at 1:31:53[2]. We are interested in locating the spatial component of the source of an earthquake. Moreover, we consider the material

[2] Other source inversion experiments using this case study are currently being run by other recipients of the KAUST CRG4 grant.





(b) Autocorrelation function for different samplers. As we can see, the most efficient sampler is the combination of parallel tempering and adaptive Metropolis. All samplers took equivalent time.

(a) Here we can see the probability density for the spatial components of the source location. The left figure is a zommed-in version of the right one.

Figure 3 – Inversion Results

properties of the ground ρ, μ, λ to be uncertain. However, we claim to know some of the structure of the medium, such as layer position. The experiment are run for $T = 17$, $\Delta t = 10^{-3}$, on a mesh of 58×34 elements. The spatial part of the wave equation is discretized using a spectral element method with 5 Gauss-Legendre-Lobatto (GLL) nodes per element, and the time marching scheme is done using a leapfrog method. In particular we compare random-walk Metropolis (RWM), delayed rejection adaptive Metropolis (DRAM), and Parallel-Tempering (PT).

same as before, I have some older results on the toy problem with the KDE and some results for a toy problem using the ideal proposal (sampling for normals), but I guess including upcoming results might be better.

3.3 MLMCMC strategy

Only few works so far have dealt with Multi-level extensions of MCMC algorithms for posterior exploration in Bayesian inversion ([12] [8]). Both authors address the case of parameter identification for elliptic PDEs and propose different strategies in order to build Multi-Level Metropolis-Hastings algorithms. As for multigrid methods applied to discretised (deterministic) PDEs, the key is to avoid estimating the expected value of Q_ℓ directly on level ℓ, but instead to estimate the correction with respect to the next lower level. Since in the context of MCMC simulations, the target distribution π depends on the discretization parameter ℓ, the new multilevel MCMC (MLMCMC) estimator has to be defined carefully [8]. We will use the identity

$$\mathbb{E}_L = \sum_{\ell=0}^{L} \mathbb{E}_\ell[Q_\ell] - \mathbb{E}_{\ell-1}[Q_{\ell-1}],$$

where Q is some quantity of interest and where we have used the convention that $\ell = -1$ means the expectation w.r.t the 0 measure. We define the increments $Y_\ell = Q_\ell - Q_{\ell-1}$. There are two main ideas underlying the reduction in computational cost associated with the multilevel estimator:

- Samples at level ℓ are cheaper to estimate
- The Variance of Y_ℓ goes to 0 as ℓ goes to infinity.

To guarantee the reduction in computation effort, we use the following assumptions:

Assumption 2. *The following are assumed to hold:*

1. $\exists \alpha > 0$ *such that* $\|\mathbf{u} - \mathbf{u}_{h_\ell}\|_V \leq C_1 h^{-\alpha\ell}$, *where* u_ℓ *denotes the solution with discretization parameter* h_ℓ.
2. $\exists \beta > 0$ *such that* $\mathbb{V}_\ell[Y_\ell] \leq C_2 h^{-\beta\ell}$, *where* u_ℓ *denotes the solution with discretization parameter* h_ℓ; Y *defined to be the increment as in [8].*
3. $\exists \gamma > 0$ *such that the cost* $\mathcal{C}_\ell$ *associated to evaluating* $\Phi(\boldsymbol{\theta}; y)_\ell$ *behaves as* $\mathcal{C}_\ell \leq C_3 h^{-\gamma}$.
4. *We can generate samples from* $\pi_\ell, \pi_{\ell-1}$ *using an independent sampler MH algorithm.*

Notice that 1 and 3 in Assumption 2 are related to the PDE at hand. An algorithm that

function MLMCMC($\{N_\ell\}_{\ell=0}^L, \{\pi_\ell\}_{\ell=0}^L, q, \theta^0$)
    if $\ell = 0$ **then**
        $\chi_{0,0}$ =`Metropolis-Hastings`$(N_0, \pi_0, q, \theta^0)$
    end if
    for $\ell = 1, \dots, L$ **do**
        for $n = 1, 2, \dots, N_\ell - 1$ **do**
            Sample $\tilde{\theta} \sim \hat{\pi}_{\ell-1}$.
            Set $\theta^* \sim q(\tilde{\theta})$.
            for $j = \ell - 1, \ell$ **do**
                Set $\theta_{\ell,j}^{n+1} = \theta^*$ with acceptance probability α_j:

$$\alpha_j(\theta_{\ell,j}^n, \theta^*) = \min\left[\frac{\pi_j(\theta^*)q(\theta_{\ell,j}^n)\hat{\pi}_{\ell-1}(\theta_{\ell,j}^n)}{\pi_j(\theta_{\ell,j}^n)q(\theta^*)\hat{\pi}_{\ell-1}(\theta^*)}, 1\right].$$

                Set $\theta_{\ell,j}^{n+1} = \theta^n$ otherwise.
            end for
        end for
        Create chains $\chi_{\ell,\ell-1} = \{\theta_{\ell,\ell-1}^n\}_{n=0}^{N_\ell}$ and $\chi_{\ell,\ell} = \{\theta_{\ell,\ell}^n\}_{n=0}^{N_\ell}$
    end for
    Output $\chi_{\ell,\ell}$, and $\chi_{\ell,\ell-1}$ for $\ell = 0, \dots, L$.
end function

Note that in the case where we generate proposals by subsampling from the previous chain we obtain the algorithm presented in [8]. We present some definitions that we will use throughout the rest of the paper.

Definition 4. *We say that at a given level ℓ and iteration n, the chain **diverged** if $\boldsymbol{\theta}_{\ell,\ell}^n \neq \boldsymbol{\theta}_{\ell,\ell-1}^n$.*

3.4 On how to generate $\hat{\pi}_\ell$

One thing that has not been discussed so far is how to choose $\hat{\pi}_\ell$. We now discuss some proposal generating functions for the independent sampler

3.4.1 Ideal Proposal

We first study the ideal, (yet no implementable for the majority of the problems), case for which we are able to do an independent sampler generating proposal from the target distribution at the previous level. That is, suppose that at each level ℓ , $\pi_\ell = \hat{\pi}_\ell$. Thus, given a sequence of target distributions $\{\pi_\ell\}_{\ell=1}^\infty$, we generate candidate states from $\pi_{\ell-1}$ at each iteration, i.e, we generate proposal states by

$$\boldsymbol{\theta}^* \sim \pi_{\ell-1}.$$

Even if this *oracle* algorithm is not implementable in practice for the majority of more *interesting* problems, we can learn some of the properties that belong to the “best case scenario" algorithm.

Proposition 5. *Let $\pi_{\ell-1} = \hat{\pi}_{\ell-1}$, and let there be a C_1 such that $\pi_\ell \leq C_1 \pi_{\ell-1}$. Then, assuming that we are on a bounded state space and that $\forall \boldsymbol{\theta} \in \Omega$ and $\ell \geq 0$, we have that $\pi_\ell(\boldsymbol{\theta}) > 0$, the following holds:*

1. *The acceptance rate $\alpha_\ell \to 1$ as $\ell \to \infty$.*
2. *The integrated autocorrelation time τ goes to 1 as $\ell \to \infty$, assuming that we can generate i.i.d samples from $\pi_{\ell-1}$.*

3.4.2 Sampling From the Prior

there is a result that quantifies the conv. rate for ind. sampler based on bounds of the lielihhod, i'll include it soon

3.4.3 KDE approximation

One possible way of generating $\hat{\pi}_{\ell-1}$ is to create a KDE based on the samples obtained at the previous level.

3.4.4 Laplace's Approximation

Alternatively , we can use Laplace's approximation in order to compute $\hat{\pi}_{\ell-1}$ under the assumptions that (i) it is a smooth function in Ω and (ii) that the distribution is well peaked. Under these assumptions, we have that, denoting $q_{\ell-1}(\boldsymbol{\theta}) = \hat{\pi}_{\ell-1}$ and omitting the dependence on the level for notational clarity, the Taylor expansion of $q(\boldsymbol{\theta})$ around is maximum a posteriori (MAP) $\boldsymbol{\theta}^{MAP}$ is given by

$$
\begin{aligned}
q(\boldsymbol{\theta}) &= q(\boldsymbol{\theta}') + (\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP})^T \nabla q(\boldsymbol{\theta}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP})^T H (\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP}) + \mathcal{O}((\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP})^3), \\
&= q(\boldsymbol{\theta}') + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP})^T H (\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP}) + \mathcal{O}((\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP})^3), \quad \text{since } \nabla q(\boldsymbol{\theta}^{MAP}) = 0, \\
&\approx C - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP})^T H (\boldsymbol{\theta} - \boldsymbol{\theta}^{MAP}), \qquad (21)
\end{aligned}
$$

Despite being an interesting approach and having some desirable properties (see [1]), this approach has two main drawbacks. The first one is the lack of generality, as it is not always the case that a posterior has a unique maximizer. The second one is the computational challenge of computing the Hessian. This however, can be approximated and computed "in line" if an optimization algorithm is used to find the MAP.

3.4.5 Optimization of a Parametric Family

An additional approach to computing $\hat{\pi}_{\ell-1}$ is as based on the minimization of the Hellinger distance between $\pi_{\ell-1}$ and $\hat{\pi}_{\ell-1}$. Formally, we seek to approximate $\pi_{\ell-1}$ using a parametric family of distributions with density $f(\boldsymbol{\theta};\gamma)$, where γ is given by

$$\gamma^* = \min_{\gamma\in\Gamma} \int_\Omega \frac{1}{2}\left(\sqrt{\hat{\pi}_{\ell-1}} - \sqrt{f(\boldsymbol{\theta};\gamma)}\right)^2 d\gamma + \frac{\alpha}{2} R(\gamma), \tag{22}$$

where R is a regularizer with regularization term α. Notice that the previous integral can be rewritten in terms of an expectation with respect to the samples of the previous chain $\chi_{\ell-1,\ell-1}$ as

$$\mathbb{E}_{\pi_{\ell-1}}\left[\left(\sqrt{\pi_{\ell-1}} - \sqrt{f(\boldsymbol{\theta};\gamma)}\right)^2 \frac{1}{\pi_{\ell-1}}\right], \tag{23}$$

thus, the optimization problem can be read as find γ^* such that

$$\gamma^* = \min_{\gamma\in\Gamma} \quad \mathbb{E}_{\pi_{\ell-1}}\left[\left(\sqrt{\pi_{\ell-1}} - \sqrt{f(\boldsymbol{\theta};\gamma)}\right)^2 \frac{1}{\pi_{\ell-1}}\right] + \frac{\alpha}{2} R(\gamma). \tag{24}$$

This expectation can in turn be approximated with error $N_\ell^{1/2}$ by

$$\approx \frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left[\left(\sqrt{\pi_{\ell-1}(\boldsymbol{\theta}^i)} - \sqrt{f(\boldsymbol{\theta}^i;\gamma)}\right)^2 \pi_{\ell-1}^{-1}(\boldsymbol{\theta}^i)\right] + \frac{\alpha}{2} R(\gamma). \tag{25}$$

Moreover, note that it is possible to save the values of the un-normalized posterior $\pi_{\ell-1}$ at the previous level, and as such, this optimization algorithm should represent a lower cost with respect to the solution of the PDE. Moreover, note that each approximation can be improved at the next level, given that each chain is evaluated twice in the algorithm. Alternatively, a similar approach can be proposed minimizing the Kullback-Lieber divergence.

HALF-WAY DONE (wanted to try to include some experiments with the transport maps), but I wonder if I should move sections 3.4.2 to 3.4.5 here?

4 Future planning

- There are some pictures here to be included soon

There are different research directions and projects that would be nice to explore in the future, both from a computational and from a theoretical perspective. We here present an outline of the ideas that we would like to explore during the duration of this project.

- There are still many different ways of proposing MLMCMC algorithms, two of which have been discussed and will be studied in the near future. The first one is to extend the same framework of the parallel tempering algorithm, where multiple chains using different noise levels (so called "temperatures") are run in parallel and set to exchange states using a Metropolis-Hastings acceptance-rejection step every so often, to a multi-level or even multi-index setting, where there would be temperatures and discretization levels to take into account. A similar idea to this has been implemented in the context of sequential Monte Carlo by [21].

- The other interesting research direction arises when combining multi-level ideas with transport maps. Transport maps have been studied in [26, 32, 33] and used as an efficient way of accelerating MCMC.

- Concerning the inversion problem, it would be interesting to move the project into more challenging problems, such as those involving Cartesian three dimensional models, or source inversion at a global scale. It is also important to apply the methods explored herein to inversion of slip faults, rather than "just" point sources, as these represent more realistic models.

- Concerning the inversion problem, it would be interesting to move the project into more challenging problems, such as those involving Cartesian three dimensional models, or source inversion at a global scale. It is also important to apply the methods explored herein to inversion of slip faults, rather than "just" point sources, as these represent more realistic models.

- We are interested in performing Bayesian inversion for the source location. However, in many cases the material properties of the earth are deemed to be uncertain, and as such can be treated as nuisance parameters in the Bayesian inversion.this transition here can be done more smoothly In more technical terms, denoting by $\boldsymbol{\theta}^s, \boldsymbol{\theta}^e$ the (unknown) source and material properties respectively, in such a way that $\boldsymbol{\theta} = (\boldsymbol{\theta}^s, \boldsymbol{\theta}^e)$, we would like to obtain $\pi(\boldsymbol{\theta}^s|y)$ based on $\pi(\boldsymbol{\theta}^s, \boldsymbol{\theta}^e|y)$. Thus,

$$\pi(\boldsymbol{\theta}^s|y) = \int_{\Theta^e} \pi(\boldsymbol{\theta}^s, \boldsymbol{\theta}^e)\pi(\boldsymbol{\theta}^e)d\boldsymbol{\theta}^e \tag{26}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} \pi(\boldsymbol{\theta}^s, \boldsymbol{\theta}_i^e), \quad \boldsymbol{\theta}^e \sim \pi(\boldsymbol{\theta}^e). \tag{27}$$

We can therefore combine two ideas to obtain samples from $\pi(\boldsymbol{\theta}^s|y)$; we can use the pseudo-marginal MCMC [2] to sample from $\pi(\boldsymbol{\theta}^s|y)$ where we can only access $\pi(\boldsymbol{\theta}^s, \boldsymbol{\theta}^e|y)$, and, additionally, we can use a multi-level Monte Carlo for a fast computation of the expected value (27).

- Lastly, it would be a very interesting idea to collect all these sampling strategies, multi-level or not, and implement a library (either in MATLAB or python) that can be used as a "blackbox", so that other research groups can use the methods developed in this work.

This is probably done

5 Miscellaneous

5.1 Conferences

- Uncertainty quantification for complex systems: theory and methodologies, Isaac Newton Institute of Mathematical Sciences, Cambridge, UK., 09.03.18-15.03.18
  Poster: Bayesian Approaches to a Seismic Source Inversion Problem.

5.2 Courses and Workshops Attended

- Gene Golub Summer School 2018: "Inverse Problems: Systematic Integration of Data with Models under Uncertainty." Breckenridge, Colorado. June 17-30, 2018 - *(4 credits)*

- Nobile, Fabio, "Stochastic Simulations". Mathematics, MATH 414. EPFL - (*5 credits*)

- SpeedUp Workshop, University of Bern.

- Parallel computing mini-classes at EPFL (offered by SCITAS).

5.3 Others

- Academic Visit, King Abdullah University of Science and Technology, from 11.17-12.17 and 16.03.18-06.04.18. Stochastic Numerics Research Group, in charge of Prof. Raúl Tempone.

- Co-supervisor of Master thesis, Marc Witkowski *Monte Carlo Methods For Contact Problems with Rough Surfaces.* February 2018.

- Co-supervisor of Master thesis, Mathieu Odobez. *The Zig-Zag Method for Inversion Problems.* July 2018.

References

[1] On the convergence of laplace's approximation and its implications for bayesian computation.

[2] Christophe Andrieu, Gareth O Roberts, et al. The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.

[3] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

[4] Igor A Beresnev. Uncertainties in finite-fault slip inversions: to what extent to believe?(a critical review). *Bulletin of the Seismological Society of America*, 93(6):2445–2458, 2003.

[5] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[6] Tan Bui-Thanh, Omar Ghattas, James Martin, and Georg Stadler. A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523, 2013.

[7] Carsten Burstedde and Omar Ghattas. Algorithmic strategies for full waveform inversion: 1d experiments. *Geophysics*, 74(6):WCC37–WCC46, 2009.

[8] Tim J Dodwell, Chris Ketelsen, Robert Scheichl, and Aretha L Teckentrup. A hierarchical multilevel markov chain monte carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1075–1108, 2015.

[9] Michael B Giles. Multilevel monte carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[10] Abdul-Lateef Haji-Ali, Fabio Nobile, and Raúl Tempone. Multi-index monte carlo: when sparsity meets sampling. *Numerische Mathematik*, 132(4):767–806, 2016.

[11] Gavin P Hayes. Rapid source characterization of the 2011 m w 9.0 off the pacific coast of tohoku earthquake. *Earth, planets and space*, 63(7):4, 2011.

[12] Viet Ha Hoang, Christoph Schwab, and Andrew M Stuart. Complexity analysis of accelerated mcmc methods for bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.

[13] Lars Hörmander. The analysis of linear partial differential operators. iii, volume 274 of grundlehren der mathematischen wissenschaften [fundamental principles of mathematical sciences], 1985.

[14] Ajay Jasra, Kengo Kamatani, Kody JH Law, and Yan Zhou. A multi-index markov chain monte carlo method. *International Journal for Uncertainty Quantification*, 8(1), 2018.

[15] Patrick Joly. Variational methods for time-dependent wave propagation problems. In *Topics in computational wave propagation*, pages 201–264. Springer, 2003.

[16] Boško S Jovanović and Endre Süli. *Analysis of finite difference schemes: for linear partial differential equations with generalized solutions*, volume 46. Springer Science & Business Media, 2013.

[17] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

[18] Uwe Köcher and Markus Bause. Variational space–time methods for the wave equation. *Journal of Scientific Computing*, 61(2):424–453, 2014.

[19] Dimitri Komatitsch and Jeroen Tromp. Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical journal international*, 139(3):806–822, 1999.

[20] Dirk P Kroese, Thomas Taimre, and Zdravko I Botev. *Handbook of monte carlo methods*, volume 706. John Wiley & Sons, 2013.

[21] Jonas Latz, Iason Papaioannou, and Elisabeth Ullmann. Multilevel sequential^2 monte carlo for bayesian inverse problems. *arXiv preprint arXiv:1709.09763*, 2017.

[22] JL Lions and E Magenes. Non-homogeneous boundary value problems, vol. 1, 1972.

[23] Qinya Liu, Jascha Polet, Dimitri Komatitsch, and Jeroen Tromp. Spectral-element moment tensor inversions for earthquakes in southern california. *Bulletin of the Seismological Society of America*, 94(5):1748–1761, 2004.

[24] Juan Pablo Madrigal Cianci, Sebastian Krumscheid, Raúl Tempone, and Fabio Nobile. Multi-level markov chain monte carlo for bayesian inversion involving the elastic wave equation. *Work in progress*, 2018.

[25] Martin Mai and K Thingbaijam. Srcmod: An online database of finite-fault rupture models.

[26] Youssef Marzouk, Tarek Moselhy, Matthew Parno, and Alessio Spantini. An introduction to sampling via measure transport. *arXiv preprint arXiv:1602.05023*, 2016.

[27] SE Minson, M Simons, and JL Beck. Bayesian inversion for finite fault earthquake source models i—theory and algorithm. *Geophysical Journal International*, 194(3):1701–1726, 2013.

[28] Mohammad Motamed and Daniel Appelo. A multiorder discontinuous galerkin monte carlo method for hyperbolic problems with stochastic parameters. *SIAM Journal on Numerical Analysis*, 56(1):448–468, 2018.

[29] Mohammad Motamed, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for the second order wave equation with a discontinuous random speed. *Numerische Mathematik*, 123(3):493–536, 2013.

[30] Mohammad Motamed, Fabio Nobile, and Raúl Tempone. Analysis and computation of the elastic wave equation with random coefficients. *Computers & Mathematics with Applications*, 70(10):2454–2473, 2015.

[31] Saulo Pomponet Oliveira and Stela Angelozi Leite. Error analysis of the spectral element method with gauss–lobatto–legendre points for the acoustic wave equation in heterogeneous media. *Applied Numerical Mathematics*, 129:39–57, 2018.

[32] Matthew Parno, Tarek Moselhy, and Youssef Marzouk. A multiscale strategy for bayesian inference using transport maps. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1160–1190, 2016.

[33] Matthew D Parno and Youssef M Marzouk. Transport map accelerated markov chain monte carlo. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):645–682, 2018.

[34] Pierre-Arnaud Raviart, Jean-Marie Thomas, Philippe G Ciarlet, and Jacques Louis Lions. *Introduction à l'analyse numérique des équations aux dérivées partielles*, volume 2. Dunod Paris, 1998.

[35] Christian P Robert, George Casella, and George Casella. *Introducing monte carlo methods with r*, volume 18. Springer, 2010.

[36] Guangfu Shao, Xiangyu Li, Chen Ji, and Takahiro Maeda. Focal mechanism and slip history of the 2011 m w 9.1 off the pacific coast of tohoku earthquake, constrained with teleseismic body and surface waves. *Earth, planets and space*, 63(7):9, 2011.

[37] Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[38] Jeroen Tromp, Carl Tape, and Qinya Liu. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1):195–216, 2005.

[39] Stephen Wright and Jorge Nocedal. Numerical optimization. *Springer Science*, 35(67-68):7, 1999.