Diouane, Youssef and Calandra, Henri and Gratton, Serge and Vasseur, Xavier A Parallel Evolution Strategy for Acoustic Full-Waveform Inversion. (2014) In: EAGE Workshop on High Performance Computing for Upstream, 7 September 2014 - 10 September 2014 (Chania, Greece).

Introduction

The knowledge of earth internal structures at different scales is of major interest for economy, humans, environment and science. Several methods have been developed for Earth imaging using seismic wave information. The full waveform inversion (FWI) attempts to build images of the subsurface physical parameters using the full wavefield, solved as an optimization procedure. FWI using non-linear least-squares (LS) minimization can be efficient when the starting model is accurate enough (Virieux and Operto, 2009), but otherwise suffers from stalled convergence to spurious local minima due to the oscillatory nature of the data. Thus, the holly grail in seismic imaging is to be able to automate the FWI in an efficient way and without the need for sophisticated a priori knowledge on the starting model. The latter needs to explain the data within a low frequency, a smooth version of the true model can be seen as a good starting model. First-arrival traveltime tomography, Laplace domain inversion or stereotomography are typically the most used methods to generate an accurate initial model, which is then improved upon by the FWI.

In this work, we are trying to propose another alternative to find an initial model for the FWI without any physical knowledge. Motivated by the recent growth of high performance computing (HPC), we will try to tackle the high non-linearity of the problem to minimize, using global optimization methods which are easy to parallelize, in particular, evolution strategies. Global optimization methods have been already used to solve such problem. The first application was through simulated annealing which was used to invert the ocean bottom properties (Collins and Kuperman, 1992). A second application came from Gerstoft, who used genetic algorithms to invert seismoacoustic data (Gerstoft, 1993). The main drawbacks of global optimization methods were that they are very consuming in terms of the cost function evaluation and very depending on the parameterization of the methods. For instance in the prespecified examples (genetic algorithms and simulated annealing), the methods were based on a fast cost function evaluation as well as a very simple parameterization of the model. The first contribution of this work is to adapt evolution strategies to the FWI setting where the cost function evaluation is the most expensive part. The second contribution is the parameterization of the regarded problem, by being able to represent the model, as faithfully as possible, while limiting the number of parameters needed, since each additional parameter is an additional dimension to explore. The last contribution is to propose a highly parallel evolution strategy adapted to the FWI setting. The initial results, obtained in this direction, show that great improvement can be done to automate the FWI.

In the optimization problem under consideration, one is looking for a minimiser of the following cost function: $C(m) = \frac{1}{2}\left||d(m) - d_{obs}|\right|_W = \frac{1}{2}\overline{\Delta d(m)}^{\,T} W \overline{\Delta d(m)}$, where the model m represents some physical parameters of the subsurface discretized over the computational domain. The misfit vector $\Delta d(m) = d(m) - d_{obs}$ of dimension N is computed as the difference at the receiver positions between the recorded seismic data d_{obs} and the modeled seismic one $d(m)$. The latter is related to the modelled seismic wavefield $u(m)$ projected using the operator R, which extracts the values of the wavefield at the receiver positions for each source: $d(m) = Ru(m)$. The weight matrix W is in general used to include a priori data information.

A globally convergent evolution strategy

Evolution strategies (ES) (Rechenberg, 1973) are evolutionary algorithms designed for continuous problems. In this work, we are dealing with a large class of ES, where in each iteration μ models (called parents) are selected as the best in terms of the cost function C of a broader set of λ ($> \mu$) models (called offsprings) corresponding to the notation (μ, λ)-ES. At the k-th iteration of these ES, the new offsprings $\{m_{k+1}^1, \ldots, m_{k+1}^\lambda\}$ are generated around a weighted mean m_k of the previous parents, corresponding to the notation $(\mu / \mu_w, \lambda)$-ES. The generation process is done by $m_{k+1}^i = m_k + \sigma_k^{ES} d_k^i$, $i = 1, \ldots, \lambda$, where d_k^i is drawn from a certain distribution and σ_k^{ES} is a chosen step size.

In (Diouane et al, 2013), it has been shown how to modify the above mentioned class of ES to rigorously achieve a form of global convergence, meaning convergence to stationary models independently of the starting model. The modifications consisted essentially of the reduction of the size of the steps whenever a sufficient decrease condition on the objective function values is not verified. By a sufficient decrease we mean a decrease of the type $C(m_{k+1}) \leq C(m_k) - \rho(\sigma_k)$, where σ_k stands for the step size parameter and $\rho(.)$ obeys some properties, in particular $\rho(\sigma)/\sigma \to 0$ when $\sigma \to 0$. When such a condition is satisfied, the step size can be reset to the step size σ_k^{ES} maintained by the ES themselves, as long as this latter one is sufficiently large. The numerical experiments therein measured the effect of these modifications into a given ES. The overall conclusions were that modifying ES to promote smaller steps when the larger steps are uphill leads to an improvement in the efficiency of the algorithms in the search of a stationary model. The modified form of the original ES algorithm is described as follows:

- ***Initialization*** : Choose positive integers λ and μ such that $\lambda \geq \mu$. Select an initial model m_0, evaluate $C(m_0)$. Choose initial step lengths $\sigma_0, \sigma_0^{ES} > 0$ and initial weights$\{\omega_0^1, \ldots, \omega_0^\mu\}$. Choose constants $\beta_1, \beta_2, d_{min}, d_{max}$ such that $0 < \beta_1 \leq \beta_2 < 1$ and $0 < d_{min} < d_{max}$. Select a forcing function $\rho(.)$ and set $k = 0$.

- ***Until some stopping criterion is satisfied:***

1. ***Offspring Generation***: Compute new sample models $M_{k+1} = \{m_{k+1}^1, \ldots, m_{k+1}^\lambda\}$ such that

$$m_{k+1}^i = m_k + \sigma_k \, d_k^i \,,$$

where d_k^i is drawn from a given distribution and obeys $d_{min} \leq d_k^i \leq d_{max}, i = 1, \ldots, \lambda$.

2. ***Parent Selection***: Evaluate $C(m_{k+1}^i), i = 1, \ldots, \lambda,$ and reorder the offspring models in $M_{k+1} = \{\tilde{m}_{k+1}^1, \ldots, \tilde{m}_{k+1}^\lambda\}$ by increasing order: $C(\tilde{m}_{k+1}^1) \leq \cdots \leq C(\tilde{m}_{k+1}^\lambda)$. Select the new parents as the best μ offspring sample models

$$m_{k+1}^{\text{trial}} = \sum_{i=1}^{\mu} \omega_{k+1}^i \tilde{m}_{k+1}^i.$$

3. ***Imposing Sufficient Decrease***:

   If $C(m_{k+1}^{\text{trial}}) \leq C(m_k) - \rho(\sigma_k)$, then consider the iteration successful, set $m_{k+1} = m_{k+1}^{\text{trial}}$, and $\sigma_{k+1} \geq \sigma_k$ (for example $\sigma_{k+1} = \max(\sigma_k, \sigma_k^{ES})$.

   Otherwise, consider the iteration unsuccessful, set $m_{k+1} = m_k$ and $\sigma_{k+1} = \bar{\beta}_k \sigma_k$, with $\bar{\beta}_k \in \{\beta_1, \beta_2\}$.

4. ***ES Updates***: Update the ES step length σ_{k+1}^{ES}, the chosen distribution, and the weights $\{\omega_{k+1}^1, \ldots, \omega_{k+1}^\mu\}$. Increment k and return to ***Step 1***.

Parameterization of the model

ES can have great success on problems that are known to be computationally hard, good results have been found in terms of the quality of the minimum found. However, most types of ES suffer from the *curse of dimensionality*, meaning that their performance is good on low dimensional problems, but deteriorates as the dimensionality of the search space increases.

The FWI is a method for deriving velocity models from the seismic data. For realistic simulations, the size of the velocity models N in general exceeds 10^6, thus trying to solve directly the problem using ES is out of scope. However, the goal is not to solve the FWI but only to find a good starting model which will be later improved using a local gradient method. The initial model m_0 is thus needed just to represent the general structure of the true model, such representation is generally smooth and can be expressed using only few parameters. Later, the ES method will try to find the parameters that lead to such smooth representation of our model.

Representing the velocity model using fewer model parameters has been investigated in the past, see for example (Oldenburg et al, 1993). The construction of the model using only few parameters in our case, was ensured through the following mechanism: Suppose that we have only few parameters n that we want to magnify to cover all $N \gg n$ parameters model, which then will be used to evaluate the cost function C. This magnification process is ensured, in our case, through an adaptation of the inverse wavelet transform (more precisely a Haar wavelet). The possible discontinuous character of the magnified model was then smoothed through an adaptation of the discrete cosine transform. The whole process (magnification and smoothing) is shown to be equivalent to apply a linear operator on the few parameters to get the model we are interested in.

Evolution strategies adapted to FWI setting

The diagram in ***Figure 1*** shows the structure of our optimizer that is to be adapted to solve our inverse problem. The scenario is as follows:

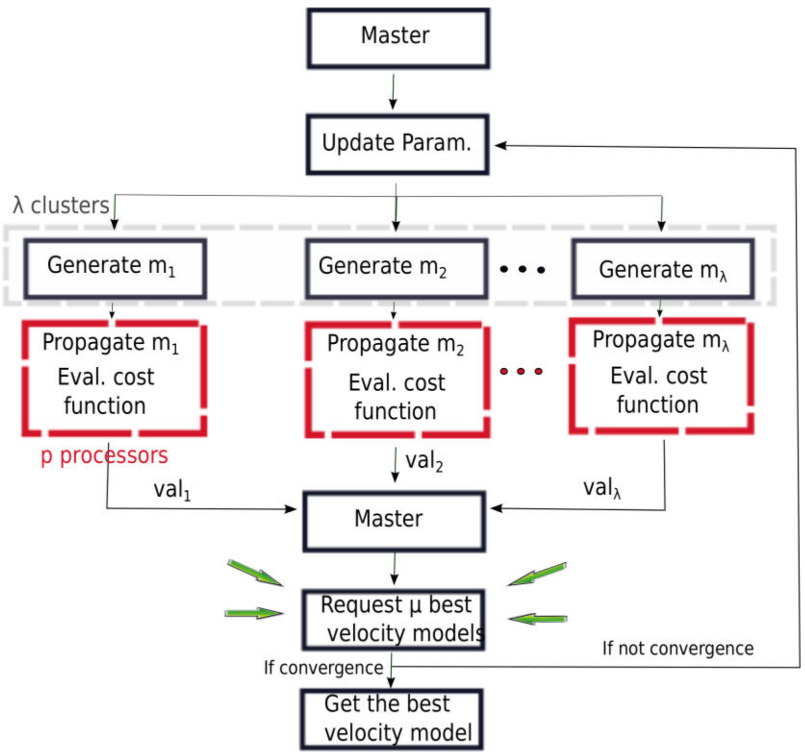

Figure 1 *A parallel evolution strategy for full-waveform inversion.*

The component *Update Param.* will be responsible for updating all the ES parameters (e.g. the distribution, the step length ...). In addition, it will launch asynchronously λ clusters represented in the diagram by the component *Generate* m_i. Each of these clusters generates a reduced velocity model based on the ES parameters and strategies. Once the velocity model is generated, this cluster evokes the component *Propagate* m_i. The wave propagation simulation on each velocity model m_i deals with all the shots (many right hand sides) at once. The *Propagate* m_i component is in fact an MPI process making use of p processors and is responsible for discretizing and building the linear system to be solved and to provide the information needed to evaluate the cost function $C(m_i)$. The last component will just return the value of the cost function to the master. Once the master receives λ results, it will choose the μ best results and return it to the *Update Param*. component to update the ES parameters and repeat the loop until a convergence criterion is achieved.

The interface with ES and the propagator will be as follows. After the generation of the velocity models using the magnification and the smoothing process, the wave propagation simulation starts, the propagator component will build a real velocity model this time of the real problem size (the true dimensions). The propagation itself will behave as a black box process, hiding the complexity of the discretization and the solution of its respective linear system from ES. Also, the flexibility and modularity of the propagator component is a key property, such that changing the chosen solver and/or the discretization nuances will not incur in any rewriting of ES implementation. MPI-2 has been used with the MPI_Comm_spawn interface that allows an MPI process to spawn a number of clusters. Each newly spawned cluster has a new MPI_COMM_WORLD intracommunicator that allows to launch easily the propagation simulations. The proposed ES implementation is portable and the propagator itself can be a standalone server. When the available cluster number is less than λ, one can launch many propagations on the same cluster until we get the λ function evaluations needed.

As a first test scenario, we consider the Salt Dome velocity model (N=676*676*210) using a *garbage-in garbage-out* scheme, meaning that the observed data are generated from the velocity model we are inverting. The purpose is to find a good initial velocity model for the FWI without any a priori information. We worked with a three-dimensional heterogeneous acoustic Helmholtz equation, written in the frequency domain, as a propagator. Optimization tests have been performed using 27 and 64 parameters in a low frequency range (f=1, 2 and 3 Hz) on clusters with up to 2048 cores CPU Sandy Bridge. A gradient constant velocity is chosen as initial model for our ES. Obtained results will be shown and analyzed during the talk.

Conclusions

The first obtained results are encouraging, and show that one can, without any a priori information on the velocity model, predict a dome form on the velocity model. A lot of progress can be made specially if one optimizes the computational cost of the propagation, ongoing works are on this direction. As the gradient information of our cost function C can be available, a possibility under current investigation is to combine local methods (gradient based ones) with ES to speed up the convergence properties and get better solutions.

References

Collins, M. D. and Kuperman, W. A. [1992] Nonlinear inversion for ocean bottom properties: The Journal of the Acoustical Society of America, no. 92, 2770-2782.

Diouane, Y., Vicente, L. N. and Gratton, S. [2013] Globally convergent evolution strategy: under revision for Mathematical Programming Journal.

Gerstoft, P. [1993] Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions: The Journal of the Acoustical Society of America, no. 95, 417-782.

Oldenburg, D. W., McGillivray P. R., and Ellis, R. G. [1993] Generalized subsapce methods for large scale inverse problems: Geophysics Journal International, 12-20.

Rechenberg, I. [1973] Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution. Frommann-Holzboog.

Virieux, J. and Operto, S. [2009] An overview of full-waveform inversion in exploration geophysics: Geophysics, no. 74, WCC1-WCC26.