Progressive matching optimisation method for FWI

Francisco A. Moura *, Suzane A. Silva, João M. de Araújo and Liacir S. Lucena

Departamento de Física, Universidade Federal do Rio Grande do Norte, Natal, RN, 59064-360, Brazil

*Corresponding author: Francisco A. Moura. E-mail: aliriomoura@gmail.com

Received 12 September 2019, revised 26 October 2019
Accepted for publication 17 December 2019

Abstract

To eliminate the dependency on a good initial model of the traditional full waveform inversion (FWI) method, we propose an optimisation method combining a derivative free optimisation method of modified particle swarm with gradient descent search. We worked with the acoustic wave approximation, in two dimensions, with the synthetic Marmousi velocity model as the test case. We were able to obtain a high-precision inversion of this model, comparable to traditional FWI methods, with the distinct advantage of not using an initial model close to the global optimal, as would usually be required. For this result, we used a progressive inversion scheme by consecutive layers, and a modified particle swarm optimisation algorithm where we introduced the gradient of the misfit function as a local search guide, and other regularization terms.

Keywords: Full waveform inversion, particle swarm optimisation, hybrid optimisation method

1. Introduction

Full waveform inversion (FWI), in geophysical exploration, is the process of using the whole information contained in seismograms to infer the properties of the subsurface (Wang 2016; Jones 2019). This is done by simulating the wave propagation in a candidate model for the subsurface and comparing the simulated seismogram with the one measured in the field, using an objective function that measures the difference between them; the most common misfit function is the L2 norm of the difference between observed and calculated data. FWI is therefore, an inverse optimisation problem to find the model with the minimum value for this objective function. We usually call the solution of this inverse problem the inverted model, which is the model that best reproduces the observed data.

In this work, we use the acoustic wave approximation where the subsurface is described by a model of the values of acoustic wave speed at each point. Traditionally, FWI is done by using the adjoint method formulation to calculate the gradient of the misfit function with respect to the model variables (Tarantola 1984), and the optimisation is performed with some gradient descent method that updates the model following the opposite direction of the misfit function gradient. One of the most commonly used methods is the limited memory Broyden–Fletcher–Goldfarb–Shannon algorithm (L-BFGS) (Byrd et al. 1995; Rao & Wang 2017).

These kinds of gradient-based methods are local optimisation algorithms and have a strong dependency on the initial model that must be provided by some other approaches as an input to the FWI (Mulder & Plessix 2008), and there are several techniques based on tomography or other means to try to obtain good initial models (Zhu et al. 1992; Zelt & Barton 1998; Zhang & Toksöz 1998; Taillandier et al. 2009; Wang & Rao 2009). Some of these model building methods are more automatic, others require human interpretation, but all of them are time consuming. There are usually several subsurface models that are located at local minima of the misfit function, and gradient descent methods will converge to the local minimum that is in the basin of attraction of the initial model, not necessarily to the desired global minimum which is the model that best represents the real subsurface structures. That is why the determination of a good initial model is vital to these methods.

To overcome this need, we propose a hybrid optimisation method that uses the best features of a derivative free...
optimisation (DFO) algorithm (Conn et al. 2009), in this case the particle swarm optimisation (PSO) (Eberhart & Kennedy 1995), with the fast local search capacity of the gradient-based method. In this work, we take advantage of the global search performed by the PSO to achieve what can be thought of as simultaneously building the initial model and performing the inversion automatically by layers. Therefore, the proposed method is designed to eliminate the need for an initial model.

In gradient-based FWI, each evaluation of the misfit function and its gradient by the adjoint method entails two wavefield simulations for each source which are costly, and an active area of research is how to reduce the number of simulations needed (Rao & Wang 2017; Rao et al. 2018). Although in purely DFO methods only one propagation is needed, their biggest problem lies in what is called the ‘curse of dimensionality’. To explore the search space, DFO requires many more evaluations of the misfit function than traditional FWI, and since each evaluation involves the simulation of the wave propagation through the model, the computational cost of DFO grows to be prohibitively large with the number of variables being optimised. In our case, this number is proportional to the size and resolution of the velocity model we are trying to invert. However, DFO methods are becoming progressively more attractive with the increasing capacity of modern computers.

Some previous works have already explored the application of the PSO or other DFO methods to FWI. Such methods are usually variations or hybrid versions of traditional DFO algorithms. These applications require initial models with the same degree of accuracy as traditional methods (Shaw & Srivastava 2007, Chen & Wang 2017) or are aimed at generating the initial model itself (Diouane et al. 2016; Mazzotti et al. 2016; Aleardi & Mazzotti 2017; Mojica & Kukreja 2019), to be used as the starting point of traditional FWI. Recent progress at performing FWI without an initial model has also been made by designing the misfit function to use the phase information contained in the signal in frequency or time domain (Luo et al. 2018; Hu et al. 2019).

These DFO approaches to FWI try to address the problem of increasing number of variables by employing different representation spaces for the velocity model. The number of variables necessary to describe the whole model is reduced by applying some transformation, at the cost of a loss in resolution. With the same intent of keeping the dimensionality of the problem low, we have developed an alternative strategy of segmenting the inversion process into layers. We start from the shallow portion of the model and work our way down, using the information acquired during the inversion of one layer to infer the next one. We trade one inversion with a large intractable number of variables for several inversions to be performed in sequence with a much smaller number, in a limited region of space and time, inside which the characteristics are inverted and consolidated to be used in the next step, in a strategy we call progressive matching.

Combined with the PSO base algorithm, the proposed method uses a gradient descent approach to speed up the local search around the current model, while premature convergence to a local minimum is avoided by the PSO ability to explore the search space globally. We then introduce several modifications to the PSO model update equation to get the best out of both gradient-based and gradient-free methods. We introduced constraints to the value of the model variables that are analogous to the usual regularizations used in other FWI methods, a combination of box and total variation constraints (Anagaw & Sacchi 2011), but instead of being applied to the misfit function, they are incorporated directly in the model updating equations of the PSO. They have the effect of forcing the search to remain in the space of physically allowed models.

2. Progressive matching

We adopted a strategy of dividing the inversion into smaller problems to be solved progressively, by slicing the volume of the model into thin horizontal layers, to reduce the number of variables being solved for simultaneously. In 2D, each layer consists of one line of points in the discretised model, which will be solved to its final value, plus the next three lines down, which can also vary during this step but will be solved for again in the next iterations. After each optimisation, we move this spatial window down by one unit and do a new optimisation. Each line of the model will be inverted four times, with the exception of the first and last three lines of the model.

This number of simultaneously inverted lines is a compromise between the number of variables to solve for and the misfit function sensitivity. We may reason about the limiting cases of solving for all lines simultaneously or for just one line at a time. The first case is equivalent to inverting the whole model from no initial model. There are too many variables, no prior information, and the PSO will not succeed. In the second case, each line would be inverted only once, and it would be impossible to isolate the contribution of that single line of velocity values to the error in the objective function, also resulting in a failed inversion.

The chosen value of four lines for simultaneous inversion was the empirically determined smallest number that provided enough opportunity for the algorithm to correct itself between sequential inversions, and comprised a large enough area such that its influence can be observed in the objective function. Using as few lines as possible reduces the number of variables being solved for in each inversion, saving computational time and facilitating the job of the PSO.

Since the model is being inverted from top to bottom, we do not need an initial model to begin the inversion: the very first lines are invertible without prior knowledge and we use...
Figure 1. Schematics of the progressive matching strategy divided in steps. (a) The model is correctly inverted up to line 10. (b) Copy the velocity values from line 10 to line 11. (c) Perform the inversion for lines 8 to 11. (d) Copy the value from line 11 to 12 and invert lines 9 to 12. (e) Invert lines 10 to 13. (f) Invert lines 11 to 14.

the result of previous layers to guide the following ones, removing the necessity of a global initial model. In our synthetic test case, the acquisition was assumed to be maritime with a flat water bottom and we considered the wave speed at the layer of water to be known. The model is truncated to be only as deep as the last line currently being optimised, and we use a perfectly matched layers (PML) (Komatitsch & Martin 2007) border condition in all directions. This domain reduction helps reduce the cost of wave propagation simulation.

This process is illustrated in figure 1, where we show part of a hypothetical inversion of a 15 × 15 points discretised velocity model displaying an inclined layered interface between two regions with different velocities. Assuming we have inverted the model correctly up to a certain depth, we start each new unknown line as a copy of the previous line and run an inversion of the last four lines. As the sequence of panels shows, line 11 will be inverted four times, the first one when the interval being inverted is between lines 8 and 11 and ending at 11 to 14, as the red inversion window progresses downwards. At each of these inversions, more information is incorporated and the line gets closer to the real velocity values.

The misfit function calculation, in this method, will be restricted to a window around the estimated time it takes for a wave to travel from the source to the current line being inverted and back to the receiver, for each source/receiver pair. Assuming we have accurately solved the velocity model up to a depth \( z \), we can use the current inverted model so far to estimate this time using the Eikonal equation in two dimensions (Gray & May 1994):

\[
\left( \frac{\partial T(x, z)}{\partial x} \right)^2 + \left( \frac{\partial T(x, z)}{\partial z} \right)^2 = \frac{1}{v^2(x, z)},
\]

where \( T(x, z) \) is the estimated travel time to the point with coordinates \((x, z)\) in the coordinate system where the source is located at the origin, with boundary condition \( T(0, 0) = 0 \). As before, we limit the domain of the solution to the part of the model above the current depth that has already been inverted for. We do this to avoid, as much as is possible, having to compute in our misfit function the errors caused by events that occur in layers deeper than the one we are trying to invert at that point, since these events would not be captured in the calculated signal and would only serve as sources of error.

For each line of the model, these travel times will have to be calculated only once for each shot and receiver, at the setup of the optimisation problem, to determine the time window.
that will be used to compare the reference and simulated signals. There are standard fast algorithms for doing so (Zhao 2005), and the cost of this step is negligible in relation to the rest of the optimisation. The time window is individually calculated for each source/receiver pair, the lower boundary $T_{0}^{s,r}$ for a source $s$ and receiver $r$ is the smallest sum of the travel time between the source and a point in the current line and back to the receiver, and the window has a fixed width $w$. We used as the misfit function the $L2$ norm of the difference between observed and simulated signals, which results in the definition for the misfit function:

$$\delta = \frac{1}{2} \sum_{s,r} \left( d_{obs}^{s,r}(t) - d_{sim}^{s,r}(t) \right)^2,$$

with $T_{0}^{s,r} < t \leq T_{0}^{s,r} + l$.

where indices $s$ and $r$ indicate sources and receivers.

During the optimisations, we only update the deepest part of the model between wavefield calculations, and we use a second-order finite differences approximation in time to integrate the wave equation. This enables us to make two small improvements in the wavefield calculations. We can store the solution of the wavefield at the two instants before the wavefield in the updated region has any numerically significant amplitude and start the next calculation from these field values, starting at some later time instead of at the initial instant of the shot. Also, since we only need the simulated signal inside the time windows for the misfit function, we can limit the direct wave propagation calculation to the longest time needed. Both of these improvements save computational time in comparison to the full model inversion at once.

### 3. Modified particle swarm optimisation

The algorithm is based on a modified PSO, which utilises a population of $N_{pop}$ particles to represent candidate solutions to the optimisation problem. Each solution is coded as a particle in the search space with associated position and velocity vectors, which evolves in time according to the canonical particle swarm equations:

$$\vec{v}_t = \vec{a} \vec{v}_{t-1} + r_1 \alpha (\vec{g}_t - \vec{m}_t) + r_2 \beta (\vec{g}_s - \vec{m}_t),$$

$$\vec{m}_{t+1} = \vec{m}_t + \vec{v}_t \gamma,$$

where $\vec{a}$ is a friction parameter, $\vec{m}_t$, $\vec{m}_{t+1}$, $\vec{v}_t$ and $\vec{v}_{t+1}$ are the particle position and velocities at a step $t$ and the following one, $r_1$ and $r_2$ are uniformly distributed random numbers on the interval $[0,1]$, $\alpha$ and $\beta$ are parameters that balance the local and global exploration of search space by the particles, $\vec{g}_t$ and $\vec{g}_s$ are the best local and global positions at that step. The best local position is the one with the smallest misfit function visited by this particle so far, and the global best is the one found by any particle.

Besides the problem with the high number of variables, we must deal with the rate of convergence of the PSO and regularisation of the model. Without more guidance, the swarm of particles converges too slowly to the global minimum, forcing a large number of function evaluations, and may become trapped or waste too much time exploring regions of the search space which are not physically meaningful, such as random velocity models consisting of mostly 'noise'. To balance this, we introduced new terms in the time-evolution equation (4) for the PSO. We added a regularisation term that disfavors models that have a very high variation on wave speed between immediately neighbouring points. Usually, in FWI problems, regularisation is treated as penalties on the misfit function for unwanted solutions. In our case, we introduced it directly into the updating procedure of the algorithm with a physically inspired approach. The regularisation was performed by connecting, for each particle, each coordinate to their neighbouring ones by 'critically dampened harmonic oscillators' with spring constant $k$, dependent on the friction $\omega$. The idea is that in the absence of a driving force, when the particle is hovering around a minimum point in search space, the result will relax to a smoothed version of itself, and we are not introducing excessive energy into the swarm by accelerating them. We also add a gradient-based term to complete the algorithm. The resulting time-evolution equation is

$$\vec{v}_t = \vec{v}_t - k_1 (\vec{m}_t - \vec{m}_t) - r_3 \gamma \nabla \delta(\vec{m}),$$

$$k = \frac{(1 - 4\omega)^2}{4},$$

$$\vec{m}_t = \vec{m}_t - \frac{\sum_{i=1}^{8} m_i}{8},$$

$$\vec{m}_{t+1} = \vec{m}_t + \vec{v}_t \gamma$$

where $\vec{v}_t$ is the one defined by equation (4), $r_3$ is a third uniformly distributed random number and $\nabla \delta(\vec{m})$ is the gradient of the misfit function with respect to the parameters of the model currently being inverted. The idea is to allow the algorithm to visit different regions of the search space of models in order to localise different minima of the objective function and compare them, while avoiding getting trapped in the basin of just one local minimum.

The addition of the gradient term guides each particle to the minimum point of the misfit function closest to its current location. This changes the method from a fully derivative free to a hybrid method. There are two main differences between this gradient calculation and the more conventional FWI. First, we are not calculating the full derivative of the misfit function with respect to every variable of the model, only the ones being inverted at that step of the progressive matching, which saves a significant amount of memory on storing the forward and backward propagated wavefields necessary for the calculation of the gradient by the adjoint
wavefield method. Also, we do not calculate a step length in the
gradient direction, we instead rely on stochastic step sizes
and the power of having multiple concurrent particles to find
the optimum.

On the other hand, the gradient term can pull the whole
swarm too fast to the vicinity of a local minimum, taking
away the power of the global exploration of the search space.
To counterbalance that, we added a reset condition to help
spread out the swarm to avoid this premature convergence
caused by the gradient term. If, for a set number $\eta$ of steps,
the best value of the misfit function found so far by a parti-
cle is not improved, the particle position is reset to a random
position within a distance of $2\rho_0$ from the current global best
position.

During the evolution of the particles, we enforce lower and
upper bounds on the wave speed of the model and reverse
the component of particle velocity that would direct it out of
the allowed region in search space. This is analogous to an
elastic collision with imaginary boundary walls when parti-
cles are bound to leave the search space. This procedure was
adopted because simply setting the component of particle ve-
locity that hit the upper or lower boundary of search space to
zero resulted in particles losing energy and swarming at the
boundary values of wave speed.

With the time evolution of the particles well defined, the
last things we need are a method for the selection of the ini-
itial population and stopping criteria to end the optimisation.
Each optimisation problem in the sequence of the progres-
sive matching process is over $N$ variables, where $N$ is the
number of points in the four lines of the discretised model. Since,
at the beginning of each optimisation, the wave speed of the
lines immediately preceding it was already solved for, we use
this information as the starting point for the generation of our
initial population of particles.

We construct a point in search space that corresponds to
the last three lines of the previous optimisation plus a repe-
tition of the last line. The initial population is chosen by uni-
formly selecting points on an $N$-dimensional sphere with ra-
dius $\rho_0$ centered around this point. In essence, we are starting
the search close to what is already our best guess for the first
three lines and approximating the completely unknown new
line by the one immediately above it. Initial velocities are
chosen at random directions with a magnitude between 0 and $\rho_0$.

The boundary values for wave speed at the current opti-
misation were also set based on the minimum and maximum
values found on the previous one, with some margin $\Delta \nu$.
For the $n^{th}$ optimisation, the boundaries $v^{n}_{\text{min}}, v^{n}_{\text{max}}$ were set
as $v^{n}_{\text{min}} = v^{n-1}_{\text{min}} - \Delta \nu$ and $v^{n}_{\text{max}} = v^{n-1}_{\text{max}} + \Delta \nu$, where $v^{n-1}_{\text{min/max}}$
refers to the values actually found on the inversion of the pre-
vious layer.

For the first optimisation, when there is no prior informa-
tion, we start from a point containing a uniform wave speed, in
this reported case 1800 m s$^{-1}$. We adopted a stopping con-
dition of 25 particle swarm iterations without a relative im-
provement of at least 1% of the objective function.

4. Results

We used as a test model the central portion of the well-known
2D Marmousi model, with dimensions of 2.5 km $\times$ 5 km re-
sampled to a discretisation of 20 m in both directions, result-
ing in $125 \times 250$ acoustic wave speed variables for the inver-
sion. This target model is shown on figure 2. There were 62
shots and 250 receivers at 60 m depth evenly horizontally dis-
tributed. During the inversion, values of wave speed for the
80-m depth water layer was kept fixed at 1500 m s$^{-1}$. We used
20 cells, equivalent to 400 m, of PML boundaries in every di-
rection. Source wavelet was a Ricker wavelet with a peak fre-
frequency of 8 Hz. Wavefield modelling was carried out with a
fourth order in space and second order in time finite differ-
ences algorithm, with a time step of $\Delta t = 0.001$ s.

Numerical parameters of the algorithm used on the in-
version, in the order they were defined in the text, were the
following: $l = 2$ s, $N_{\text{pop}} = 200$, $\omega = 0.5$, $\alpha = 0.1$, $\beta = 0.1$,
$\gamma = 0.5$, $\eta = 25$, $\rho_0 = 6000$ m s$^{-1}$ and $\Delta \nu = 250$ m s$^{-1}$.
To test the efficiency of the proposed algorithm in inverting the target model, we compared the inversion with the usual gradient-based, adjoint wavefield FWI method, using the LBFGS-B algorithm for the optimisation of the same misfit function, calculated over the entire model with the whole seismogram as usual. This method requires an initial model, and for that, we used a horizontal gradient, shown on figure 3a, which presents a hard case for the traditional method. Figure 3b shows the result of the inversion after 107 iterations of the LBFGS-B algorithm, when the misfit function had converged to a minimum.

Figure 4 shows the result of the inversion using our proposed method. The model contains 120 unknown lines at the beginning, discounting the water layer at the top, which results in 117 inversion steps for the progressive matching procedure. We needed three full passes from top to bottom in order to arrive at the presented solution, resulting in 351 partial inversions in total. A better imaging was obtained compared to the gradient-based inversion, without the need of an initial model, recovering most of the features of the target model within the precision allowed by the discretisation used, with much better accuracy for the deeper parts, and better position for the reflectors starting from 1400 m deep and below.

To compare quantitatively the results of the inversions by the proposed method and the traditional gradient-based FWI, and evaluate the results by areas of the model, we propose a measure of similarity based on the correlation between inverted and target models. The value $S_{ij}$ at the point of the discretised model with coordinates $(i, j)$ is defined by the equations

$$ S_{ij} = \left( \frac{\sum_{k,l} I_{k,l} T_{k,l}}{\sum_{k,l} I_{k,l}^2 \sum_{k,l} T_{k,l}^2} \right)^2, \tag{10} $$

where $I$ and $T$ are the inverted and target models, and indices $k$ and $l$ are over a square window centered in $(i, j)$ with 200-m sides. $S_{ij}$ is identical to 1 when the region around position $(i, j)$ of the target and inverted model are exactly the same, and less than 1 if there is any dissimilarity between them.

Figure 5 show the values of $S_{ij}$ for the inverted models by the proposed method and by the traditional gradient-based method. Both methods have similar accuracy on the shallow,
well-resolved parts of the model, and suffer a quality decrease around the high wave speed salt deposit regions, with a distinct advantage to the hybrid method overall. One of the differences between them is that because of the regularisation constraints imposed on the particle swarm our method tends to smooth the sharp interfaces between areas of constant wave speed. However, the actual values of the wave speed inside these regions are better recovered than by the traditional FWI.

5. Conclusion

We demonstrated a new approach to FWI based on an optimisation algorithm based on the DFO particle swarm algorithm, and progressive vertical inversion of the model, called progressive matching. The biggest advantage of this method is the elimination of the need for an initial model that is close to the real model by using the previously inverted layers to approximate the next one, and using a spatial-temporal
window over the observed and calculated seismograms to evaluate the misfit function, to isolate the contribution to the data of the region of the model around a given depth. Effectively, the construction of the initial model is done concurrently to the inversion, taking advantage of the flexibility of the modified particle swarm optimisation algorithm that combines local and global exploration of the search space.

We made modifications to the canonical particle swarm optimisation algorithm in order to equate it to the FWI by progressive matching layers, by adding two new terms to the equation for particle position evolution. The first term added, to avoid large nonphysical spatial fluctuations of the inverted wave speed model, and the second added term, to speed up convergence to the minimum of the misfit function.

The proposed method resulted in the final inversion of a wave speed model that is at least as precise as the traditional FWI given a not-ideal initial model, but without the need of an any initial model at all.

Acknowledgements

The authors gratefully acknowledge support from Shell Brasil through the New Methods for Full Waveform Inversion project at Universidade Federal do Rio Grande do Norte and the strategic importance of the support given by ANP through the R&D levy regulation. Authors would like to thank Jorge L. Lopez from Shell for the careful reading and comments on the manuscript. J.M.A. acknowledges Conselho Nacional de Desenvolvimento Científico e Tecnológico for support through Bolsa de Produtividade em Pesquisa (grant no. 313431/2018-3).

Conflict of interest statement. None declared.

References

Aleardi, M. & Mazzotti, A., 2017. Inelastic full-waveform inversion and uncertainty estimation by means of a hybrid genetic algorithm–gibbs sampler approach, Geophysical Prospecting, 65, 64–85.

Anagaw, A.Y. & Sacchi, M.D., 2011. Full waveform inversion with total variation regularization, in 2011 CSGP/CSEG/CWLS GeoConvention, Recovery: Energy Environment Economy, Canadian Society of Petroleum Geologists, Calgary, Canada.

Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C., 1995. A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific Computing, 16, 1190–1208.

Chen, G. & Wang, Z., 2017. Robust full-waveform inversion based on particle swarm optimization, SEG Technical Program Expanded Abstracts, Society of Exploration Geophysicists, 1302–1306.

Conn, A.R., Scheinberg, K. & Vicente, L.N., 2009. Introduction to Derivative-Free Optimization, Vol. 8, SIAM.