FWI without tears: a forward modeling-free gradient

Marcelo Guarido, Laurence R. Lines and Robert Ferguson*
University of Calgary

Summary

Full waveform inversion (FWI) is a machine learning algorithm with the goal to find the Earth's model parameters that minimize the difference of acquired and synthetic shots. In this work, we are introducing a new interpretation of the gradient as the residual impedance inversion of the acquired data. Its estimation is forward modeling and wavelet free, reducing its costs drastically, as the inverted model could be obtained on a personal laptop without the need of parallel processing. The new method was applied, with great success, on the acoustic Marmousi simulation. The inverted model, when using the same starting point, is comparable to the results when using the migrated residuals. This approximation also opened the possibility to change the order of migration and the stack steps, during the gradient estimation, to use a post-stack depth migration, and results are promising. In the end, we are proposing a new FWI approximation that is cheap and stable, and could be applied on a real seismic survey in a processing center that has enough computer power to run a PSDM or even just a post-stack depth migration.

Introduction

Seismic inversion techniques are the ones that use intrinsic information contained in the data to determine rock properties by matching a model that "explains" the data. Some examples are the variation of amplitude per offset, or AVO (Shuey, 1985; Fatti et al., 1994), the traveltime differences between traces, named traveltime tomography (Langan et al., 1984; Bishop and Spongberg, 1984; Cutler et al., 1984), or even by matching synthetic data to the observed data, as it is done in full waveform inversion (Tarantola, 1984; Virieux and Operto, 2009; Margrave et al., 2010; Pratt et al., 1998), among others. These inversions can compute rock parameters as P and S waves velocities, density, viscosity and others. In this paper, we are focusing only on the inversion of the P wave velocity (acoustic).

Full waveform inversion (FWI) is a machine learning based method, with the objective to estimate the Earth's model parameters that minimize the difference between observed (acquired) data and synthetic shots (Margrave et al., 2011). This is accomplished by iteratively updating the starting model with a scaled gradient and then creating new synthetic shots with the new model.

The method was proposed in the early 80's (Pratt et al., 1998) but it was considered too expensive in computational terms. Lailly (1983) and Tarantola (1984) simplified the methodology by using the steepest-descent method (or gradient method) in the time domain to minimize the objective function without explicitly calculate the partial derivatives. They estimated the gradient by backpropagating the residuals using a reverse-time migration (RTM). Pratt et al. (1998) develop a matrix formulation for the full waveform inversion in the frequency domain, and presented efficient strategies to compute the gradient and the inverse of the Hessian matrix, for both the Gauss-Newton and the Newton approximations. The FWI is shown to be more efficient if applied in a multi-scale method, where lower frequencies are inverted first and they are increased on later iterations (Pratt et al., 1998; Virieux and Operto, 2009; Margrave et al., 2010). An overview of the FWI theory and studies are compiled by Virieux and Operto (2009). Lindseth (1979) showed that an impedance inversion from seismic data is not effective due to the lack of low frequencies during the acquisition but could be compensated by the match with a sonic-log profile. Margrave et al. (2010) used a gradient method and matched it with sonic logs profiles to compensate the absence of the low frequency, and to calibrate the model update by computing the step length and a phase rotation (avoiding cycle skipping). They also proposed the use of a PSPI (phase-shift-plus-interpolation) migration (Ferguson and Margrave, 2005) instead of the RTM, so the iterations are done in time domain but only selected frequency bands are migrated, using a deconvolution imaging condition (Margrave et al., 2011; Wenyong et al., 2013) as a better reflectivity estimation, same strategy used by Guarido et al. (2015a;2015b). Guarido et al. (2016) show the need of the application of an impedance inversion step in the gradient and use a band-limited impedance inversion (BLIMP) method using the algorithm implemented by Ferguson and Margrave (1996). Warner and Guasch (2014) use the deviation of the Wiener filters of the real and estimated data as the object function with great results.

We are proposing a new approximation for the FWI, where the gradient is interpreted as a residual impedance of the current model and the impedance inversion of the acquired data. On each iteration, the data is PSPI migrated (Ferguson and Margrave, 2005), with a deconvolution imaging condition, using the current model and applying a BLIMP inversion on the stacked data. A conjugate gradient is also used to improve the quality of the gradient and to reduce the number of iterations (Zhou et al., 1995; Vigh and Starr, 2008). The step length is computed by a least-squares minimization (Pica et al., 1990). To compute the residuals on the standard methodology, a finite difference forward modelling algorithm is used to create the synthetic shots. The results of the new approximation are comparable with the classic method (steepest descent). We went further and inverted the order of the migration and stack processing steps, and computed the gradient using a zero-offset PSPI migration (post-stack). The preliminary tests results are promising, with a huge gap for improvements.

Theory

The objective of the FWI methodology is to minimize an objective function. Here we minimize the residuals **Δd(m)**, that is the difference between observed data $\mathbf{d_0}$ and synthetic data **d(m)**, in the current model model **m** (here P wave velocity):

$$C(\mathbf{m}) = ||\mathbf{d}_0 - \mathbf{d}(\mathbf{m})||^2 = ||\Delta\mathbf{d}(\mathbf{m})||^2 \quad (1)$$

Minimizing the objective function **C(m)** with respect to the model **m**, we can to the steepest-descent formula (Pratt et al.,1998):

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n \mathbf{g}_n \quad (2)$$

where **α** is the step length, **g** is the gradient and **n** is the n-th iteration. This equation shows that a model update can be obtained by adding a scaled gradient to the current model. This routine is kept until stopping criteria is reached. The gradient should be, due to the theory, computed by a reverse time migration of the residuals (Tarantola, 1984; Pratt et al., 1998; Virieux and Operto, 2009), but we decided to use the phase-shift-plus-interpolation (PSPI) migration, by the assumption that FWI is a set of processing tools and any pre-stack depth migration could be used to back-propagate the residuals. Later, the BLIMP algorithm uses the initial model as pilot to apply an impedance inversion of the gradient. The first iterations use only the low frequency on the data while the higher frequencies are included on later iterations.

By interpreting the gradient computation steps (migration, stack, and impedance inversion) as seismic processing tools, equation 2 can be rewritten in terms of the operators **M** for migration, **S** for stacking and **I** for impedance inversion, leading to:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I\left\{S\left[M\left(\mathbf{d}_0 - \mathbf{d_n}\right)\right]\right\} \quad (3)$$

where $\mathbf{d_n}$ is the synthetic shot. Guarido et al. (2016) assume that all three operators are linear (true for migration and stack and approximate for impedance inversion), and the gradient can be interpreted as a residual difference of the processed acquired data and the processed synthetic data (both migrated using the current model). The second one, on a perfect case, is the current model itself (the migrated, stacked and impedance inversion of the synthetic data). This explanation is better visualized in equation 4:

$$\begin{aligned}\mathbf{m}_{n+1} &= \mathbf{m}_n - \alpha_n (I\left\{S\left[M\left(\mathbf{d}_0\right)\right]\right\} - \underbrace{I\left\{S\left[M\left(\mathbf{d_n}\right)\right]\right\}}_{\text{Current model}}) \\ &= \mathbf{m}_n + \alpha_n (I\left\{S\left[M\left(\mathbf{d}_0\right)\right]\right\} - \mathbf{m}_n) \end{aligned} \quad (4)$$

Interpreting the gradient as the residual difference of the processed acquired data and the current model saves us to compute a synthetic data at each shot position. Source estimation is also not required. Two forward modeling are still required on the step length process. But, as it is only an amplitude matching, it is unnecessary to compute the correct wavelet.

We can make the method even cheaper if we invert the order of the migration and stacking operators on equation 4. This would result on using a stacked session as input and a post-stack migration at each iteration:

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n \left(I\left\{M\left[S(\mathbf{d}_0)\right]\right\} - \mathbf{m}_n\right) \quad (5)$$

Two forward modeling are still required to estimate the step length. However, the costs drop significantly.

Examples

Simulations are done on the Marmousi velocity model (figure 1a). Simulated acquired data are generated by a 2D acoustic finite difference code using a Ricker wavelet with 5Hz of dominant frequency (even though the dominant frequency of the data is 12Hz) on 104 different positions. Starting model (figure 1b) is a smoothed version of the real Marmousi. For the classic FWI method (figure1c), forward modeling is done using current model (initial model on first iteration and updated model subsequently) and the same wavelet as the acquired data (we are just applying the FWI on a synthetic simulation. No real data was tested).

First iteration uses only low frequency content, and the initial band is from 4Hz to 6Hz. The same frequency band is repeated until convergence is reached (objective function varies less than 0.001% for three consecutive iterations). Then the frequency band is changed by fixing the minimum frequency in 4Hz and increasing the maximum one by 2Hz. This routine is repeated until the maximum frequency of 30Hz.

Acquired data is backpropagated using a PSPI migration algorithm, then muted and stacked (Guarido et al. 2015b), resulting on a reflection coefficient model, in depth, that has the same size of the velocity model. It represents the usual gradient estimation and it is often assumed to be an equivalent to velocity when multiplied by the step length. So, in many cases, the step length can be interpreted as an impedance inversion operator. We decided to convert the reflection coefficients model to velocity by applying an impedance inversion. As data lack in low frequencies (1 to 3Hz), we use the BLIMP algorithm, assuming the initial model is a good pilot, to fulfill the missing frequency content. This means that the initial model must contain the low frequency (linear trend) of the study are. Later, the step length is estimated.

Figure 1 shows the Marmousi model (a), initial model (b), inverted model based on equation 3 (c), where residuals are computed as the difference between acquired and synthetic data with an impedance inversion applied to the gradient, and, finally, the resulted model based on equation 4 (d), the forward modeling free gradient with a PSDM. For both inversions, the step length is estimated as proposed by Pica et al. (1990). The resulted models are comparable and show great resolution.

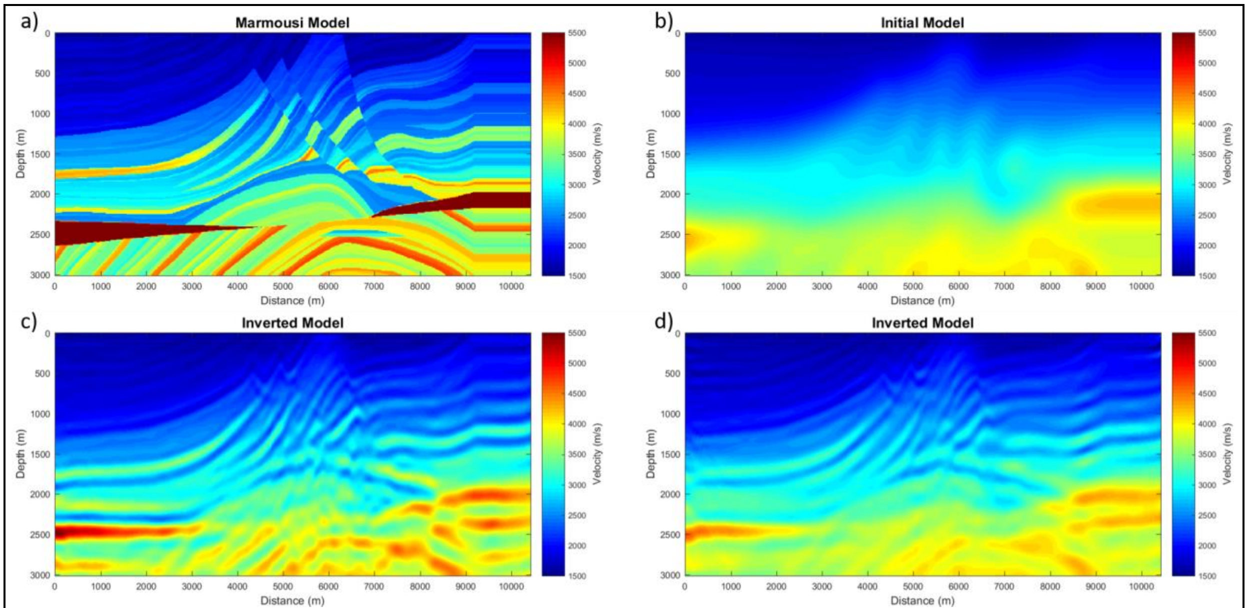

Figure 1: a) true Marmousi model, b) initial model for all runs, c) inverted model with classic FWI and d) inverted model with the forward modeling free gradient method.

Model of figure 1c (using synthetic data to estimate the gradient) has more geological structures than the model of figure 1d (forward modeling free gradient), better noticed on deeper areas. In the shallow and mid depth, the models are comparable. This means that the regular FWI still works better, mostly on higher frequencies, on a synthetic simulation. However, it still requires a very good source estimation so the residuals are stable. The forward modeling free gradient does not require the source, and we believe it would be more stable when applied on real data. Another advantage of the forward modeling free method is the computing requirement and processing time, which is reduced by about 70%.

Figure 2a is the stacked session used as input data for the post-stack FWI method, based on equation 5, and resulted model is shown on figure 2b. There is a loss of resolution if compared to the previous results. However, most of the major layers were correctly inverted and placed. Shallow and mid-depth areas are comparable to previous models. It is also possible to note some borders effects. They are due to the step length be estimated using the central shot as control point for the whole model. This effect could be reduced if more control points, closer to the borders, are included.

The differences between methods are, mostly, the costs associated to each one. For the classic method, to run the full inversion routine, it was required 24 clusters for a parallel processing in MatLab, and total elapsed time was over 48 hours. The forward modeling free gradient method (pre-stack) reduce the costs considerably, and the routine

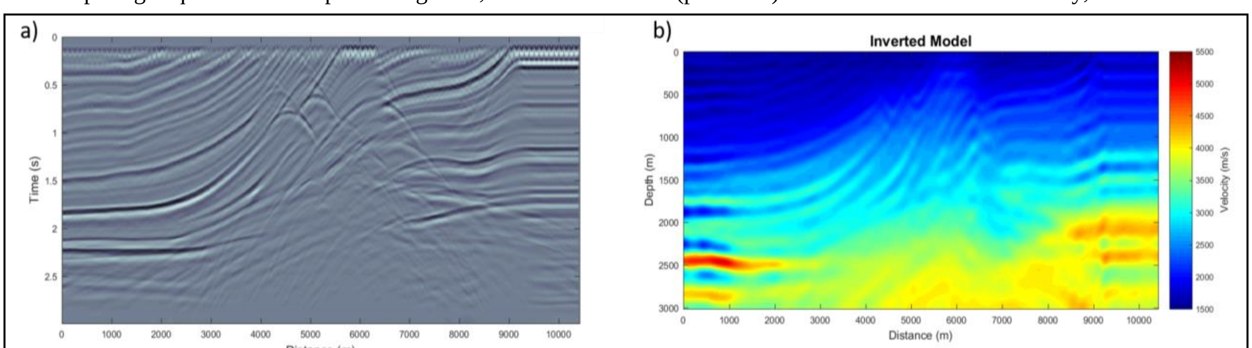

Figure 2: a) stacked section as input data and b) inverted model on the post-stack approximation.

ran on a personal gaming laptop (16Gb of RAM), with no parallel processing, and 8 hours of run time in Octave. The post stack method ran on a tablet with dual core processor (4Gb RAM), where the total elapsed time was around 1 hour only.

with the classic FWI, the results are comparable, but with some loss in resolution as costs become cheaper. However, the cost-benefit trade-off looks to be worthwhile.

A post-stack method with preliminary results were also presented, reducing even more the costs for a FWI run, but

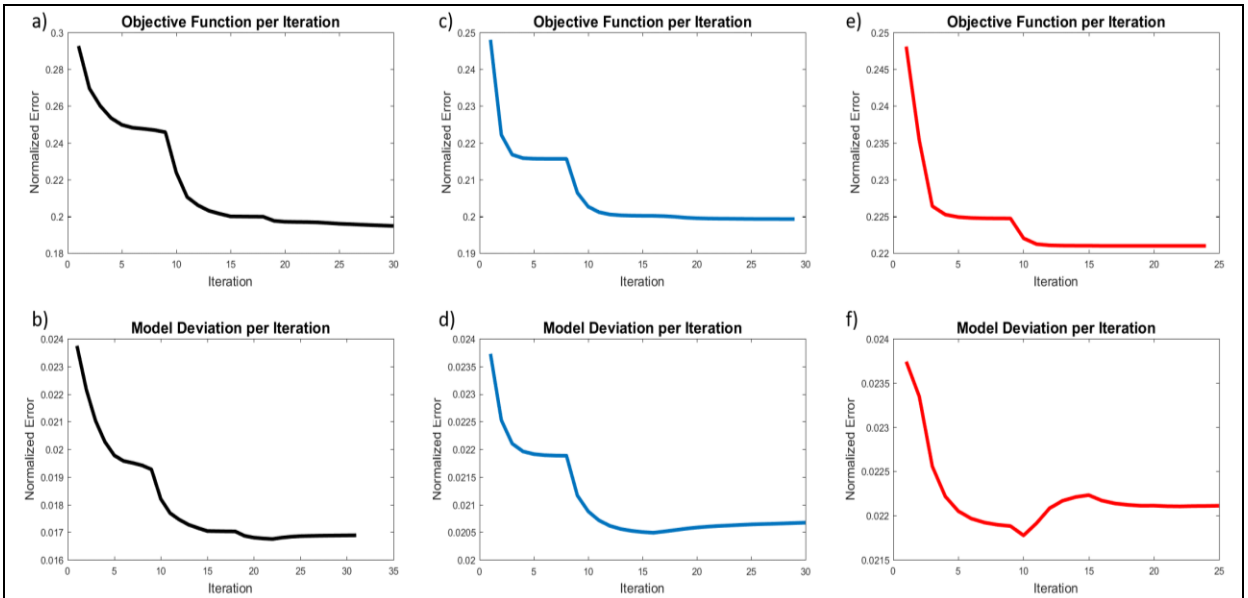

Figure 3: Respectively, shot and model errors of a) and b) classic FWI, c) and d) forward modeling free gradient method and e), and f) post-stack approximation.

Figure 3 compares the objective function and models deviations of the 3 methods. They all show to be stable and we observe the convergence of the objective function. The models deviations show to reach a minimum at some point and then starts to slowly diverge. We believe this is due to the low signal-to-noise ratio at higher frequencies. We also observe a “break” of the curves. This happens when the inversion starts to include the dominant frequency of the data (12Hz) during the migration.

It is safe to say that the resolution of the inverted model decreases the method gets simpler and cheaper. The choice of the method is just a matter of cost and benefit. Better responses will require the highest investments, and no guarantee of stability, as for some surveys the source can be very complicated to estimate. However, we show that a reasonable result, with just a small loss of resolution, can be achieved by a drastically reduction of costs, and a more robust inversion.

Conclusions

We have presented a new FWI method based on interpreting the gradient as a residual difference of the impedance inversion of the acquired data and the current inverted model, removing the need to compute one forward modeling per shot location on every iteration. Comparing also losing some resolution and the addition of border effects. However, we are confident that this is a safe strategy to follow with the goal of applying the FWI on large surveys with reduced computer requirements and gain on stability, as it does not require a source estimation. In the end, the choice of which method to be used will depend on the investment power of the user.

Acknowledgements

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. We also thank Soane Mota dos Santos for the suggestions, tips and productive discussions.