

FWI without tears: a forward modeling free gradient

Marcelo Guarido*, Laurence R. Lines*, Robert J. Ferguson†

ABSTRACT

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. We did a simulation of the methodology on 2D data at the Marmousi model. This report is divided in two parts. On the first part, we applied a band-limited impedance inversion on the migrated residuals to estimate a gradient cheaper and leading to a with higher resolution at deeper areas and more continuity of the geologic features when compared to previous works. On the second part, we introduce a new interpretation of the gradient as a residual impedance inversion of the acquired data. Its estimation is forward modeling and wavelet free, reducing its costs drastically, as the inverted model was obtained on a personal laptop without parallel processing. The new method was successfully applied on the acoustic Marmousi simulation. The inverted model, when using the same starting point, is comparable to the results of using the migrated residuals. A preliminarily test was done by inversion the order of migration and stack and using a post-stack depth migration to estimate the gradient with promising outputs. In the end, we are proposing a new FWI approximation that is cheap and stable and could be used on real data in the same 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 informations 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 traveltimes differences between traces, named traveltimes 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 work we are focused in the inversion of the P wave velocity.

FWI is a least-square based inversion, which objective is to find the model parameters that minimizes the difference between observed (acquired) and synthetic shots (Margrave et al., 2011), or the residuals. This is accomplished in an iterative fit method by linearizing a non-linear problem. It is a machine learning method very similar to a *Ridge Regression* (Chipman, 1999), which minimizes non-linear problems by adding a regularization term to avoid over fitting (to smooth the model). In seismic processing, we regularize the inversion by convolving the model with a 2D Gaussian window (Margrave et al., 2010).

The full waveform inversion was proposed in the early 80's (Pratt et al., 1998) but the

*CREWES - University of Calgary

†University of Calgary

technique 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 calculate, explicitly, the partial derivatives. They compute the gradient by a reverse-time migration (RTM) of the residuals. Pratt et al. (1998) develop a matrix formulation for the full waveform inversion in the frequency domain and present more efficient ways to compute the gradient and the inverse of the Hessian matrix (the sensitive matrix) the Gauss-Newton or 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 is increased as more iterations are done (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. Warner and Guasch (2014) use the deviation of the Weiner filters of the real and estimated data as the object function with great results.

We are applying the FWI methodology using the PSPI migration Margrave et al. (2010); Ferguson and Margrave (2005) with a deconvolution imaging condition to compute the gradient. 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-square minimization (Pica et al., 1990) and is being estimated for individual frequencies. The synthetic data is done by a finite difference forward modeling algorithm.

During previous works (Guarido et al., 2014, 2015) the model was updated directly by a scaled output of the migration (reflection coefficients). We came with 2 different approximations for the gradient and this report is divided in 2 parts.

On the first part, we start from the point where Margrave et al. (2011) suggests some kind of trace integration (impedance inversion) of the gradient must be applied. A first try was executed on Guarido et al. (2015) by a simple exponentiation of the trace integration in time domain but the lack of low frequency shows to be a big issue on impedance inversion (Treitel et al., 1995). Ferguson and Margrave (1996) implemented a simple algorithm for a band-limited impedance inversion (BLIMP) that is a modification of the SAIL (Seismic Approximate Impedance Logo) implementation of Waters (1978). It combines the low frequency impedance of a sonic log or a velocity model (stacking velocity or migration velocity) with the seismic impedance in the frequency domain.

The second part, as a new approximation, the gradient is treated as a residual impedance inversion, relative to the previous iteration, of the acquired data and no forward modeling

is required on its estimation. The results are comparable with the classic methods. We even went further and inverted the migration and stack processing order to estimate the gradient based on a post-stack depth migration. This test is a preliminarily one but is really promising.

THEORY - PART 1

The steepest-descent method

The objective function of the FWI method is, in general:

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

where Δd is the data residual (the difference between acquired and synthetic shots), m is the model (in this work, the P-wave velocity) and $\|\cdot\|$ represents the norm-2 of the array. The minimization is done by calculating the Taylor's expansion of the objective function of the equation 1 around a perturbation δm of the model and taking the derivative equal to zero (Tarantola, 1984; Pratt et al., 1998; Virieux and Operto, 2009). The solution is:

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

where H is the Hessian (or sensitive matrix), g is the gradient computed by back-propagating the data residual and n is the n -th iteration. It is known as the Newton method. For the steepest-descent method, the Hessian matrix can be neglected and be equalized to the identity matrix:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n \mathbf{g}_n \quad (3)$$

where α is the step length (or scale factor). At this part, the gradient is understood as the PSPI migrated residuals with a deconvolution imaging condition (Margrave et al., 2010, 2011; Wenyong et al., 2013; Guarido et al., 2014).

In this work we are proposing to compute the gradient as the stack of the scaled gradient per each frequency. This way, equation 3 can be written as:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \frac{1}{N} \sum_{i=1}^N \alpha_n(\omega_i) \mathbf{g}_n(\omega_i) \quad (4)$$

where ω_i is the i -th frequency in the total of N frequencies used to compute the gradient. The early iterations uses very low frequencies only ($\approx 2 - 4Hz$) and next iterations have the lower frequency fixed and the maximum frequency of the range increased. This means that, for each iterations, N migrations are computed. The step length α will be computed for

each frequency prior the stack. This method was used on a previous work (Guarido et al., 2015). However, the actual work is done using the equation 3 with a conjugate gradient and an impedance inversion of the migrated residuals.

The step length

Pica et al. (1990) computed the step length (scale factor) by minimizing the objective function:

$$C(\mathbf{m}_{n+1}) = C(\mathbf{m}_n + \alpha_n \mathbf{g}_n) = [\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n + \alpha_n \mathbf{g}_n)]^T [\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n + \alpha_n \mathbf{g}_n)] \quad (5)$$

We also have:

$$\mathbf{F}\delta\mathbf{m} = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{d}(\mathbf{m} + \epsilon\delta\mathbf{m}) - \mathbf{d}(\mathbf{m})}{\epsilon} \quad (6)$$

where \mathbf{F} is an operator that takes the derivative of \mathbf{d} at the point \mathbf{m} and $\delta\mathbf{m}$ is the perturbation in the model.

Equation 5 can be written as:

$$C(\mathbf{m}_{n+1}) = [\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n) + \alpha_n \mathbf{F}_n \mathbf{g}_n]^T [\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n) + \alpha_n \mathbf{F}_n \mathbf{g}_n] \quad (7)$$

Minimizing equation 7 relative to α_n (taking the derivative and making it equals to zero), leads to the optimal α_n :

$$\alpha_n = \frac{[\mathbf{F}_n \mathbf{g}_n]^T [\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n)]}{[\mathbf{F}_n \mathbf{g}_n]^T [\mathbf{F}_n \mathbf{g}_n]} \quad (8)$$

This reduce the problem for the step length to a single forward modeling in a perturbation in the velocity model $\mathbf{m} = \mathbf{m}_n + \epsilon \mathbf{g}_n$ and is cheaper than a step length computed by a line search (Guarido et al., 2015).

For our tests, computation of the step of equation 8 is done using a single reference shot (the one in the center of the velocity model).

The conjugate gradient

Equation 3 can be rewritten replacing the gradient \mathbf{g}_n by the conjugate gradient \mathbf{h}_n (Zhou et al., 1995; Vigh and Starr, 2008; Ma et al., 2010):

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n \mathbf{h}_n \quad (9)$$

where

$$\mathbf{h}_0 = \mathbf{g}_0, \beta_n = \frac{\mathbf{g}_n^T (\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{g}_{n-1}^T \mathbf{g}_{n-1}}, \mathbf{h}_n = \mathbf{g}_n + \beta_n \mathbf{h}_{n-1} \quad (10)$$

Band-limited impedance inversion (BLIMP)

Previous works (Guarido et al., 2014, 2015) used the stacked migrated residuals (reflection coefficients) to update the model. However the FWI formulation implies on integrate the trace of the migrated residuals to compute the gradient. In other words, an impedance inversion should be applied (Margrave et al., 2010, 2011).

In Guarido et al. (2015) a simple exponential of the trace integration in time is used as impedance inversion with promising results. Now we used the algorithm of Ferguson and Margrave (1996) for a band-limited impedance inversion (BLIMP) of the migrated residuals. Ferguson and Margrave (1996) adds a well log low frequencies' to inversion. We chose to use the initial or current inverted model (pilot model) as the low frequency content. At this point we are assuming that the linear trend of the initial model is equivalent to the Earth linear trend for that specific area.

BLIMP is applied on time domain. Inputs (initial model and migrated residuals) are converted from depth to time using the current inverted velocity model and lately, after the process, converted back to depth using the same model. The first step of BLIMP is to specify the relationship between the migrated residuals and the pilot model. Thus, the migrated residuals need to be "converted" to acoustic impedance (P-wave velocity) using the normal incidence approach (Treitel et al., 1995):

$$R_i = \frac{\rho_{i+1}V_{i+1} - \rho_iV_i}{\rho_{i+1}V_{i+1} + \rho_iV_i} \quad (11)$$

where R_i is the reflection coefficient on the i th interface, ρ is the density and V is the P wave velocity propagation. The multiplication of density and velocity is the acoustic impedance I_i . For small contrasts of acoustic impedances in geological interfaces (usually < 0.3), equation 11 can be written:

$$R_i = \frac{\Delta I_i}{2I_i} \quad (12)$$

Assuming the reflection sequence as continuous in time, and taking the limit $\Delta_t \rightarrow 0$:

$$R(t) = \frac{1}{2}d(\ln I(t)) \quad (13)$$

Integrating over time and assuming constant density:

$$V(t) = V_0 e^{2 \int_{t_1}^t R(\tau) d\tau} \quad (14)$$

Ferguson and Margrave (1996) replace the reflection coefficients $R(\tau)$ by a scaled reflectivity $S(\tau) = 2R(\tau)/\gamma$. Then 14 becomes:

$$V(t) = V_0 e^{\gamma \int_{t_1}^t S(\tau) d\tau} \quad (15)$$

γ is a scale factor estimated by minimizing the objective function:

$$\Gamma = \sum [V_{pilot}(\omega) * B + (\gamma - 1)V(\omega)]^2 \quad (16)$$

where $V(\omega)$ is the Fourier spectra of equation 15, $V_{pilot}(\omega)$ is the Fourier spectra of the pilot velocity model and B is a low-pass filter.

The output impedance is the inverse Fourier transform of the equation 17:

$$V_{out}(\omega) = V_{pilot}(\omega) * B + \gamma V(\omega) \quad (17)$$

Ferguson and Margrave (1996) use a methodology for the impedance (velocity multiplied by the density). Here it is simplified for acoustic inversion only.

Equation 17 shows that the scaled Fourier spectra of the input migrated residuals converted to velocity is added to the low frequency content of the pilot velocity. This suggests that the missing low frequency on habitual seismic data can be compensated by an initial model that contains a good estimation of the local linear velocity trend. Usually a migration velocity is used for this task.

RESULTS - PART 1

Input data

The synthetic test is done in the Marmousi model (figure 1). Pilot shots (used as the real shots), were created using an acoustic finite difference algorithm. In the total, 102 shots are used to simulate field data. Shot spacing is 100m. Each shot has a maximum of 401 receivers (varying in the edges of the model) totalizing 2000m of maximum offset. The register time is 3s and sample rating of 4ms (figure 2). The dominant frequency of the wavelet is 5Hz.

During the FWI routine, synthetic shots are created with the same finite difference forward modeling algorithm, wavelet and acquisition parameters as pilot shots. They differ from each other by the velocity model used and the routine starts using a guessing model, usually a depth migration velocity (Virieux and Operto, 2009). The residuals (difference

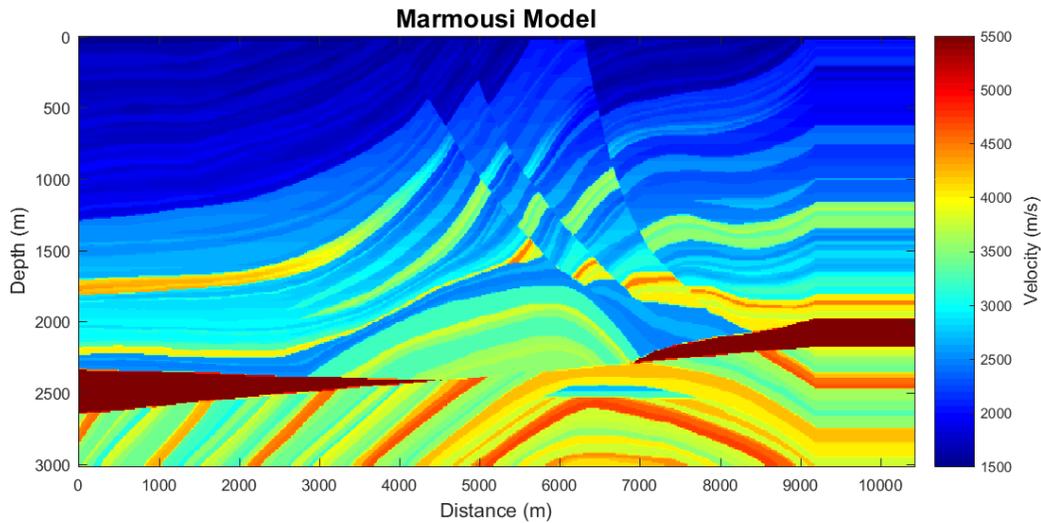


FIG. 1. The Marmousi 2D model. The color bar indicates the wave propagation velocity in m/s.

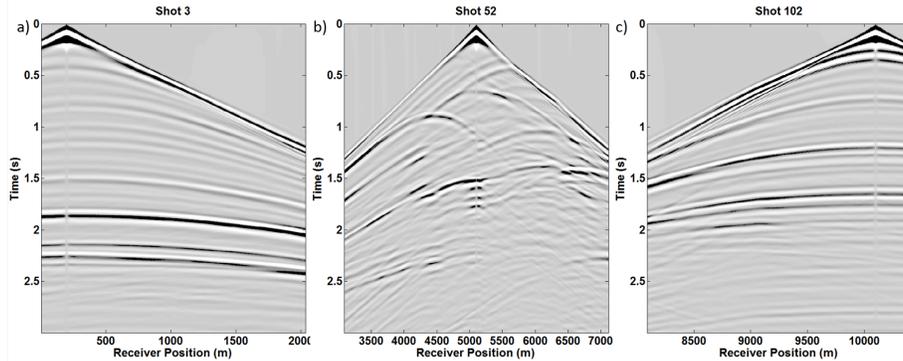


FIG. 2. Three of the 102 shots used in the test. a) shot 3, b) shot 52 and c) shot 102.

between real and synthetic shot) is then migrated with the PSPI migration with a deconvolution imaging condition. For the first iterations, only very low frequencies are used ($4 - 6\text{Hz}$), and the range is increased by 2Hz when errors are stable (variation of the 3 last iterations is less than 0.1%). Each residual is migrated separately using the initial or updated model. A mute is applied before stacking the migrated residuals (Guarido et al., 2014) and the stack is preconditioning by convolving it with a 2D Gaussian window (Margrave et al., 2010; Guarido et al., 2014). Ferguson and Margrave (1996)'s algorithm for impedance inversion is used after the conjugate gradient step. Lately the step length is estimated using equation 8 and the model is updated for the next iteration.

Initial model as well control

BLIMP is applied on the migrated residuals to estimate the gradient and the model update expecting to compensate the lack of low frequencies on the data (seismic lowest frequency used is 4Hz) following those steps:

- Residuals are computed as the difference between synthetic and acquired seismic

shots.

- Each residual is depth migrated using the PSPI algorithm with a deconvolution imaging condition.
- The migrated residuals are muted and stacked to form a pseudo-gradient.
- The pseudo-gradient and current model are stretched from depth to time.
- BLIMP is applied on the pseudo-gradient using the current model's low frequency to form the gradient in time.
- Gradient is stretched from time to depth using current model.
- Step length is computed using equation 8.
- Model is updated using the scaled gradient.

Even if we compute a gradient that points on the opposite direction to the global minimum and estimate a step length to converge faster to the same point, the gradient descent method is known to get "trapped" on a local minimum (Pratt et al., 1998). One way to minimize this effect is to start the inversion with very low frequencies (multi-scale approach). Another way is to start the inversion closer to the global minimum (a good initial model).

We expect a better result for the model closest to the global minimum because it enhances the gradient in two ways: 1) the migrated residuals are in a more correct place with the best model and 2) more details of the low frequency in the model (1 to 3Hz) results on a better impedance inversion of the migrated residuals.

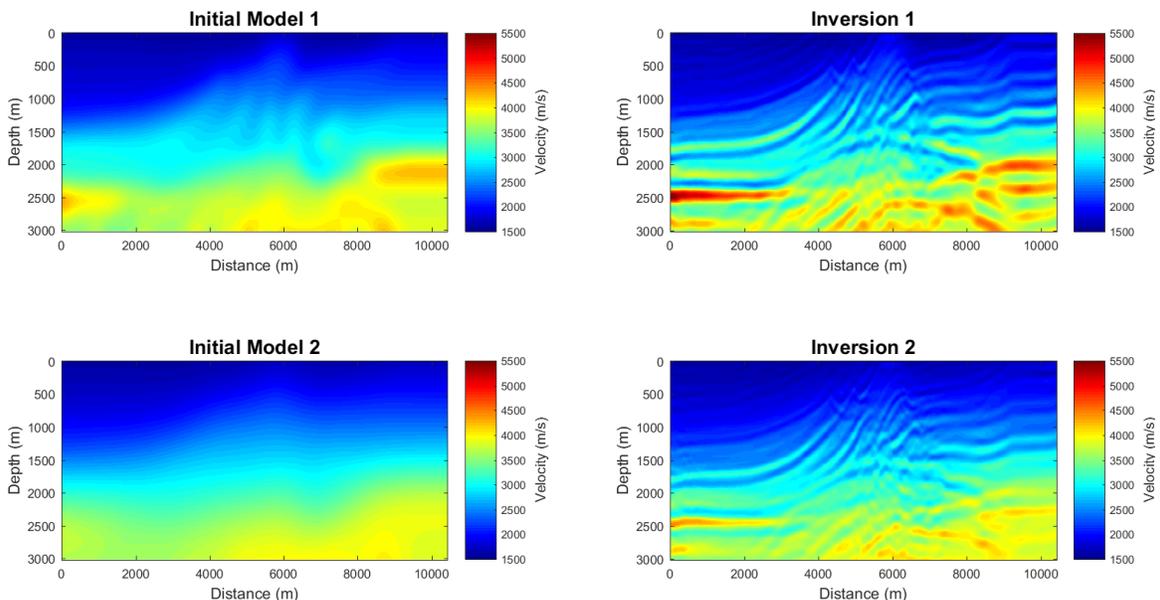


FIG. 3. Inversion as a function of the initial model. The inversion 1 has more details than the inversion 2 due to the initial model be closer to the local minimum.

Figure 3 illustrates the different inversions with varying qualities of the initial model. In both cases, the multi-scale approach is implemented. When the starting point is closer to the global minimum, the inversion leads to a higher resolution resulting model (Guarido et al., 2015). In both cases, it is noticeable of how much information is included to the model. Most of the features from low to mid frequencies were successfully recovered. The faults with high dip angle in the shallow (central portion of the model) are present on both inversions. Even the simulation of a gas anomaly ($x = 6500, z = 2500$) could be interpreted for the inversion with the best initial model.

The algorithm worked successfully for low to mid frequencies (up to 16Hz) and had no updates on higher frequencies (the gradient exists but the step length goes to zero, meaning that the inversion was too close to a minimum point).

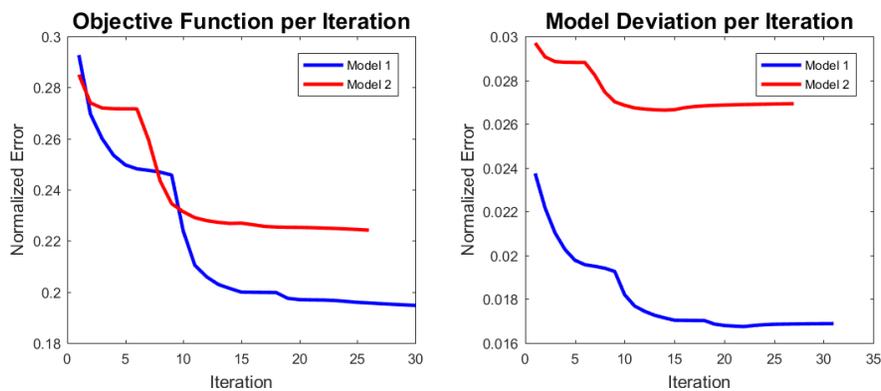


FIG. 4. Plots of the objective function per iteration (left) and model deviation per iteration (right) of models 1 (blue) and 2 (red).

For the validation of the tests, the errors are plotted on figure 4. On the left is the objective function (difference between acquired and synthetic shots) of the two models (1 in blue and 2 in red) and on the right the model deviation (difference between inverted and correct Marmousi models). The most notorious observation is quite obvious: higher resolution inversion is a result of a higher better initial model.

We can observe also another behaviors on figure 4. For both initial model tests we there are two step decrease of the errors: one when we start to update the initial model at very low frequencies ($4 - 6Hz$) and another when the inversion reaches the dominant frequency of the data ($\sim 12Hz$). When comparing the two plots, an odd observation came to our eyes. Even with the objective function decreasing at each iteration, the model deviation has a slight increase on both models. That is the point when the inversion reached frequencies on the data with low signal to noise ratio.

At this point we remember the update is based on equation 3. It is, considering the computer power to calculate it, cheaper the equation 4, method used by Guarido et al. (2015), as it requires only one migration pass per iteration. Figure 5 is a group of images comparing both works. Current inverted model (top left) seems to work better on deeper events, shows more continuity of the geological structures at the model and suffer from less borders effects, while the previous work model (bottom left) did a better work on the shallow part of the model and shows a "cleaner" image. In the end, both inverted models

are a good representation of the Marmousi model, however the costs on computing the gradient are very different. We could change the interpretation of the gradient to make its estimation cheaper and still have a similar model.

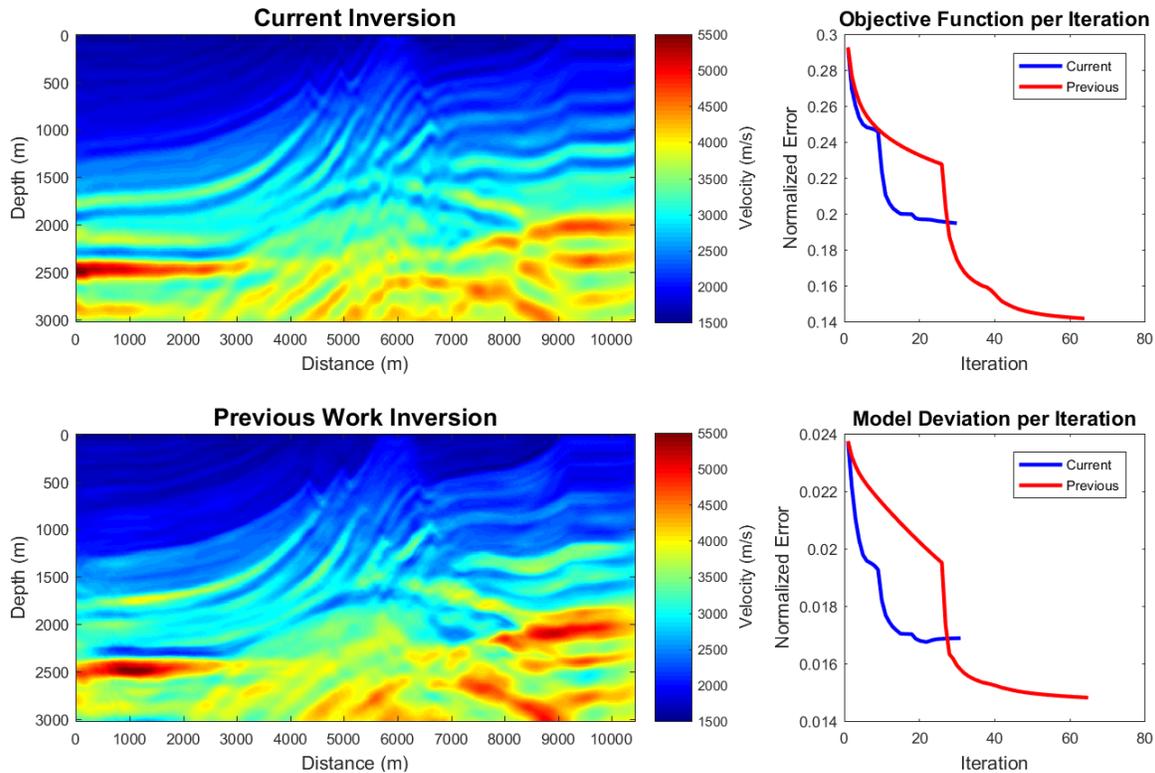


FIG. 5. On the left are the resulting velocity models on current (top) and previous (bottom) work inversions. On the right are the comparison of the objective function (top) and models deviations (bottom) per iteration.

Checking the plots of the objective function behavior (top right) and model deviation (bottom right) per iteration, current work (blue lines) converge faster (less number of iterations) than the previous work (red lines). Actually it converges with less than a half of iterations of the other method. But the errors do not look similar to each other as the models do. The shot differences are much lower on the previous inversion than on the current one. However, the same can not be said about the model deviation. There is still a difference but not so large as the shot differences. The normalized model errors were reduced from 0.024 to 0.017 (current) and 0.015 (previous). In others word, the models can be said to be equivalent, but the costs are not.

The inversion applying the BLIMP algorithm on the migrated residuals showed to be cheaper than the work on Guarido et al. (2015) in two points: the costs on computing the gradient and the number of iterations required to reach convergence.

THEORY - PART 2

New approximation for the gradient

Computing the gradient requires three very known seismic processing steps: PSDM of the residuals, stacking and impedance inversion. Let's understand those steps as the

operators M for migration, S for stacking and I for impedance inversion. Equation 3 can be re-written as:

$$\begin{aligned}\mathbf{m}_{n+1} &= \mathbf{m}_n - \alpha_n \mathbf{g}_n \\ &= \mathbf{m}_n - \alpha_n I \{S [M (\Delta \mathbf{d}(\mathbf{m}_n))]\} \\ &= \mathbf{m}_n - \alpha_n I \{S [M (\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n))]\}\end{aligned}\quad (18)$$

where $\Delta \mathbf{d}(\mathbf{m}_n)$ is the n-th iteration residual $\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_n)$, \mathbf{d}_0 is the acquired data, $\mathbf{d}(\mathbf{m}_n)$ is the synthetic data of the n-th iteration and \mathbf{m}_n is the n-th iteration inverted model. For simplification and easier visualization, let's set $\mathbf{d}(\mathbf{m}_n) = \mathbf{d}_n$. Then equation 18 is:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I \{S [M (\mathbf{d}_0 - \mathbf{d}_n)]\} \quad (19)$$

Considering the linearity property of the migration operator (see appendix), the acquired and synthetic shots can be migrated separately:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I \{S [M (\mathbf{d}_0)] - S [M (\mathbf{d}_n)]\} \quad (20)$$

The next step is to use the linearity property of the stacking operator and equation 20 becomes:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I \{S [M (\mathbf{d}_0)] - S [M (\mathbf{d}_n)]\} \quad (21)$$

For the impedance inversion operator, the first approximation we tried is $I(x_1 - x_2) = I_0 \frac{I(x_1)}{I(x_2)}$. Then equation 21 can be written as:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I_0 \frac{I \{S [M (\mathbf{d}_0)]\}}{\underbrace{I \{S [M (\mathbf{d}_n)]\}}_{\text{Current model}}} \quad (22)$$

where the denominator is the impedance inversion of the stacked depth migrated synthetic shots should be the velocity model on which the forward modeling algorithm is ran into. In others words, it is the model of the current iteration \mathbf{m}_n . Then equation 25 can be simplified to:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I_0 \frac{I \{S [M (\mathbf{d}_0)]\}}{\mathbf{m}_n} \quad (23)$$

Equation 23 shows to be unstable as it can be divided by zero (as we are inverting for lower frequencies greater than 3Hz and not capturing the Earth linear trend, it can result on

zeros and even negative impedance). Trying to avoid this we added a stabilization factor SF:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n I_0 \frac{I \{S [M (\mathbf{d}_0)]\}}{\mathbf{m}_n + SF} \quad (24)$$

Choosing a value for the stabilization factor showed to be challenging. A too small value can lead areas of the gradient to have a huge value compared with its neighbors. A large value (around 1 and above) can change the denominator significantly. And the value needed to be changed at each iteration by trial and error. Avoiding the division of equation 24 is the best tactic at this point. So we took the risk to say that the impedance inversion operator is *approximately* linear (for that, we assume Earth impedance to follow a linear trend and a Taylor expansion of the reflection coefficients is used to estimate the update as a perturbation of Earth's impedance), we end up with another solution for equation 21:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha_n (I \{S [M (\mathbf{d}_0)]\} - \underbrace{I \{S [M (\mathbf{d}_n)]\}}_{\text{Current model}}) \quad (25)$$

Again, on the second hand of the gradient approximation we have the model of the current iteration \mathbf{m}_n . Then equation 25 is simplified to:

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n (I \{S [M (\mathbf{d}_0)]\} - \mathbf{m}_n) \quad (26)$$

Equations 26 and 24 understand the gradient as a residual impedance inversion of the acquired data relative to the current model. This is an impressive result as no forward modeling or source estimation is required to compute the gradient. The objective function is being minimized on the estimation of the step length (that requires 2 synthetic data: one for the current model and one for a gradient perturbation of the same model). On our Marmousi tests, we reduced, per iteration, the need of 105 forward modeling runs to only 2. Actually, the amount of forward modelings required for any project does not depend on how many shots it has, but on how many control points for the step length estimation the processor wants to use.

RESULTS - PART 2

Pre-stack depth migration based gradient

Equations 26 and 24 reduced abruptly the computer power required for the inversion. From now on, all the tests were done on a personal gaming laptop (ASUS Intel Core i7-4700HQ 2.40GHz, 4 cores, 16Gb of Ram memory) in Octave (free scripting software) on a Linux terminal and no parallel processing.

Figure 6 shows the inversion based on equation 26, starting from the model on the top left of figure 3 and with the computer specification cited previously. The result is impressive

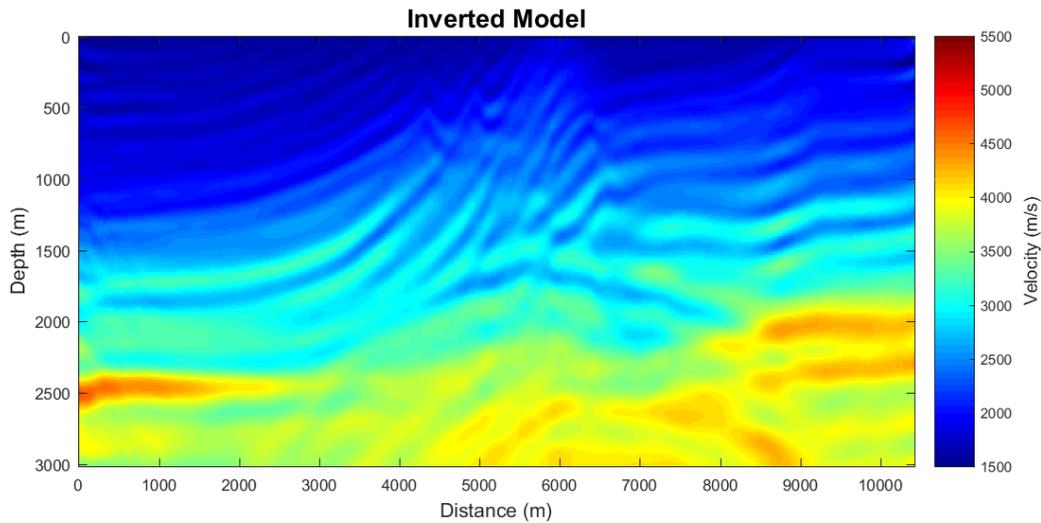


FIG. 6. Inverted model using the initial model on the top left of figure 3 and based on equation 26. The algorithm ran on a personal laptop without the need of parallel processing in about 36 hours.

and the process stopped after 29 iterations in about 36 hours of run time with no parallel processing (it was repeated using the 4 cores of the computer for parallel processing and elapsed time decreased to 18 hours). The method worked well in the whole model but better at shallower areas. Major structures are successfully inverted. Even the fault zones in the center of the model can be interpreted. Using any forward modeling to compute the gradient shows to be a valid approximation. BLIMP provides us an impedance model which includes the perturbation (high frequency) of the impedance to the low frequency model. The difference of the perturbations of current and previous iterations (the proposed gradient) points to the correct minimum but is not optimized to minimize the objective function of equation 1. The step length (that requires 2 forward modeling per iteration to be estimated) works as the minimization operator. The conjugate gradient was not used (we tested with and without it and the ending models were equivalent).

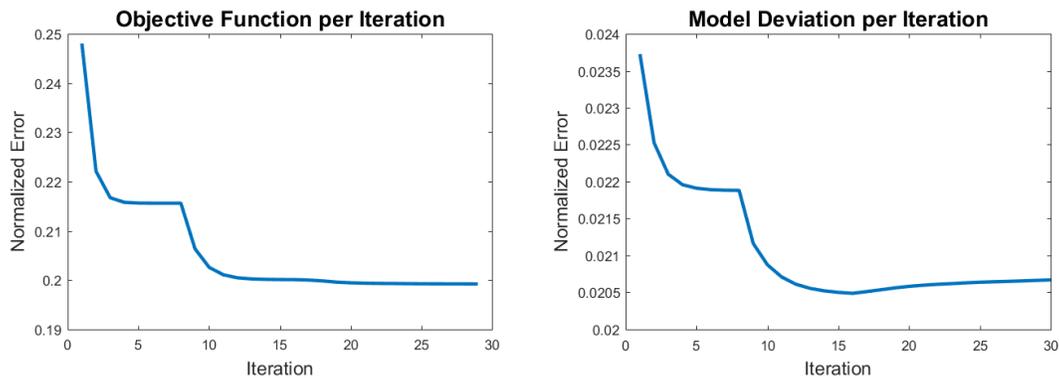


FIG. 7. Plots of the objective function per iteration (left) and model deviation per iteration (right). There was a total of 23 updates in the model but the convergence is reached with only 12.

Objective function and model deviation plots are on figure 7. We ran a total of 29 iterations (the code stops when a maximum frequency, selected by the processor, is reached. And it changes the frequency content of the current iteration when the errors converge or start to diverge). The inversion process has 2 major decrease in errors: one inverting up to

6Hz and another up to 10Hz. We believe those are the dominant frequencies related to the wavelength as same "size" of layers in the model. It is interesting to note that even with the objective function keeps decreasing by iteration, the model deviation can diverge at some point (close to iteration 17). The meaning is that the process is trapped on a local minimum of the objective function that changed direction of the optimized model for it. With all these information in mind, we can say the method is stable and leads to a optimized model gets closer to the real one when minimizing the objective function.

Post-stack depth migration based gradient

Looking again at equation 26, we understand we are estimating the gradient by applying some seismic processing tools. If we think about migration and stack, migrating the data and stacking should have the same effect as doing the inverted process (stack then migrate). Then equation 26 is equivalent to:

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n (I \{M [S (\mathbf{d}_0)]\} - \mathbf{m}_n) \quad (27)$$

This means we can input in the algorithm a stacked section (figure 8) and apply a post-stack depth migration algorithm (we used a zero-offset based PSPI with a cross-correlation imaging condition). The stacking velocity used is the same initial model of the others tests. However it was converted to a rms velocity..

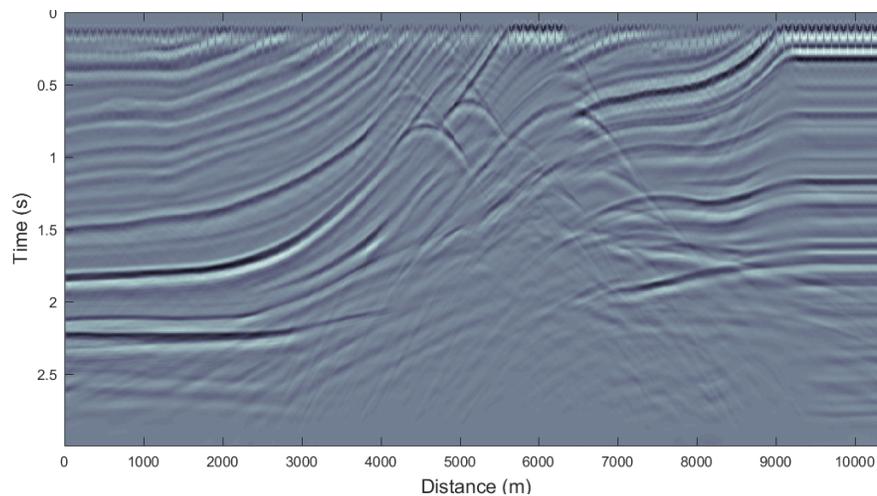


FIG. 8. Stacked section used as input in the post-stack method.

Analyzing the stacked section on figure 8, we noticed the presence of diffractions on the center area of the model caused by the fault zones. We can also observe a strong horizontal event close to the surface and also internal multiples are very evident on the deeper areas. We took note of all those observations and studied how they affected the inversion.

We ran two different preliminarily tests. On the first we pre-conditioned the gradient by convolving it with a 2D Gaussian window. This step were neglected during the second test, were we only scaled the "raw" gradient. In both we used the starting model of the top left of figure 3 and results are on figure 9. The left model is the resulted inversion when the

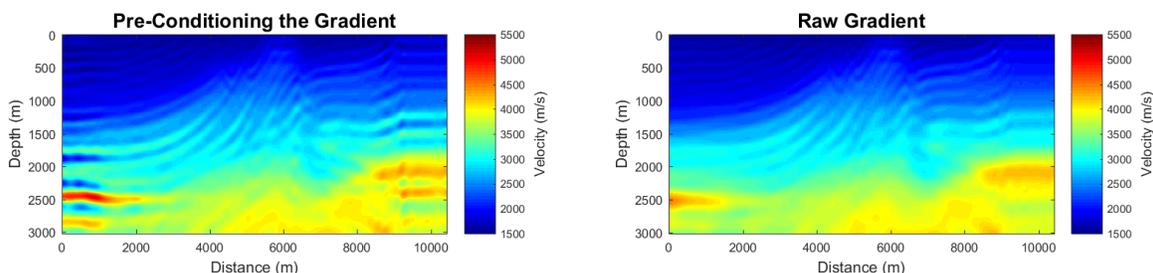


FIG. 9. Inverted models using: pre-conditioned gradient (left) and the "raw" one (right).

gradient is smoothed and on the right the gradient was kept raw. In both cases information is included to the starting model giving a more like Marmousi velocity model, but with a reduced resolution when compared to previous inversions. Inversion got truncated by the horizontal event close to the surface that we observed on figure 8. Pre-conditioning the gradient did a better job when compared with the raw one. However, it suffers with borders effects. Comparing the objective function (figure 10 left) we can say that both are stable with the smoothed gradient (blue line) doing a better job. The borders effects are not notice in the plot as the objective function is computed on a control shot in the center of the survey. Borders effects are more pronounceable on the model deviation plot (figure 10 right) where the smoothed gradient initially reduces the model difference and, from iteration 10, it starts to diverge.

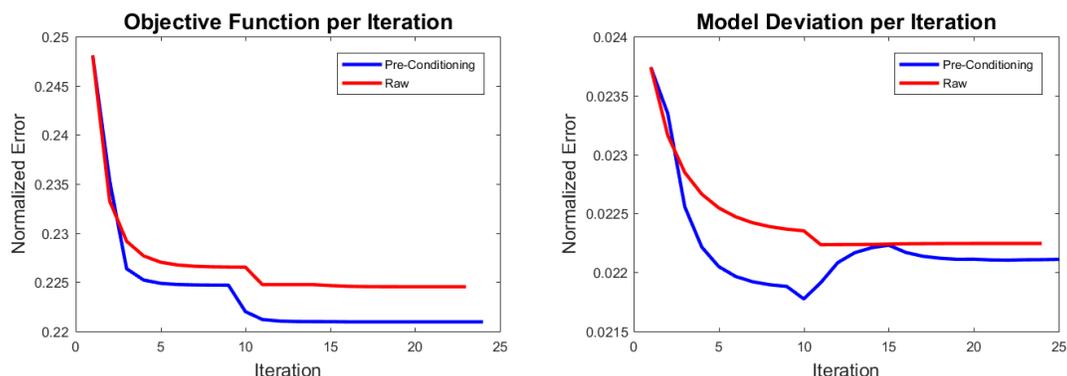


FIG. 10. Plots of the objective function per iteration (left) and model deviation per iteration (right) with the "raw" gradient (red) and pre-conditioned one (blue).

This method showed to be even cheaper than the pre-stack one and the results, even preliminarily, are really promising.

CONCLUSIONS

On the first part of this report, we presented a gradient based on a band-limited impedance inversion of the migrated residuals. This method improved the quality of the update without the need to migrate single frequencies of a selected band for the current iteration, as used on Guarido et al. (2015). Its usage showed to be cheaper in 2 ways: on the gradient computation and on a faster convergence of the objective function. Comparing again with the inversion obtained on Guarido et al. (2015), the BLIMP method resulted on higher resolution models at the deeper areas and more continuity of the geologic features.

Second, we presented a new interpretation of the gradient that is forward modeling and source estimation free. We understand the gradient to be a residual impedance inversion of the acquired data. It is stable, converge fast and is impressively cheap, as the inversion was done on a personal laptop without parallel processing. Two forward modelings are still required on the step length estimation. The resulting model contains all the bigger structures of the Marmousi model, as the inversion worked well on low frequencies up to 6Hz. The model converged early due to a horizontal artifact close to the surface. Immediate future work is to suppress such artifact and allow higher frequencies to be included in the inverted model as well.

Following the forward modeling free gradient and understanding the FWI algorithm as a combination of different seismic processing tools, we inverted the order of the migration and stacking operators and applied a post-stack depth migration (zero offset PSPI) on a stack section. The results are preliminarily but really promising.

In the end we came with a cheap solution to apply an acoustic FWI with promising results and we are confident that the same approximation can be extended to others parameters.

ACKNOWLEDGMENTS

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 thank Soane Mota dos Santos for the suggestions, tips and productive discussions.

REFERENCES

- Bishop, T. N., and Spongberg, M. E., 1984, Seismic tomography: A case study: SEG Technical Program Expanded Abstracts, **313**, 712–713.
- Chipman, J. S., 1999, Linear restrictions, rank reduction, and biased estimation in linear regression: Linear Algebra and its Applications, **289**, No. 1, 55 – 74.
- Claerbout, J. F., 1971, Toward a unifeid theory of reflection mapping: Geophysics, **36**, No. 3, 467–481.
- Cutler, R. T., Bishop, T. N., Wyld, H. W., Shuey, R. T., Kroeger, R. A., Jones, R. C., and Rathbun, M. L., 1984, Seismic tomography: Formulation and methodology: SEG Technical Program Expanded Abstracts, **312**, 711–712.
- Fatti, J. L., Smith, G. C., Vail, P. J., Strauss, P. J., and Levitt, P. R., 1994, Detection of gas in sandstone reservoirs using avo analysis: A 3-d seismic case history using the geostack technique: Geophysics, **59**, No. 9, 1362–1376.
- Ferguson, R., and Margrave, G., 2005, Planned seismic imaging using explicit one-way operators: Geophysics, **70**, No. 5, S101–S109.
- Ferguson, R. J., and Margrave, G., 1996, A simple algorithm for band-limited impedance inversion: CREWES Research Report, **83**, 21.1–21.10.
- Guarido, M., Lines, L., and Ferguson, R., 2014, Full waveform inversion - a synthetic test using the pspi migration: CREWES Research Report, **26**, 26.1–26.23.
- Guarido, M., Lines, L., and Ferguson, R., 2015, Full waveform inversion: a synthetic test using pspi migration: SEG Technical Program Expanded Abstract, **279**, 1456–1460.

- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering, theory and application: Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.
- Langan, R. T., Lerche, I., Cutler, R. T., Bishop, T. N., and Spera, N. J., 1984, Seismic tomography: The accurate and efficient tracing of rays through heterogeneous media: SEG Technical Program Expanded Abstracts, **314**, 713–715.
- Lindseth, R. O., 1979, Synthetic sonic logs—a process for stratigraphic interpretation: *Geophysics*, **44**, No. 1, 3–26.
- Ma, Y., Hale, D., Meng, Z. J., and Gong, B., 2010, Full waveform inversion with image-guided gradient: SEG Technical Program Expanded Abstracts, **198**, 1003–1007.
- Margrave, G., Ferguson, R., and Hogan, C., 2010, Full waveform inversion with wave equation migration and well control: CREWES Research Report, **22**, 63.1–63.20.
- Margrave, G., Yedlin, M., and Innanen, K., 2011, Full waveform inversion and the inverse hessian: CREWES Research Report, **23**, 77.1–77.13.
- Pica, A., Diet, J. P., and Tarantola, A., 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: *Geophysics*, **55**, No. 2, R59–R80.
- Pratt, R. G., Shin, C., and Hick, G. J., 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, No. 2, 341–362.
- Shuey, R. T., 1985, A simplification of the Zoeppritz equations: *Geophysics*, **50**, No. 4, 609–614.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, No. 8, 1259–1266.
- Treitel, S., Lines, L., and Ruckgaber, G., 1995, Seismic impedance estimation: *Geophysical Inversion and Applications - Memorial University of Newfoundland*, **1**, 6–11.
- Vigh, D., and Starr, E. W., 2008, 3d prestack plane-wave, full-waveform inversion: *GEOPHYSICS*, **73**, No. 5, VE135–VE144.
- Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, No. 6, WCC1–WCC26.
- Warner, M., and Guasch, L., 2014, Adaptive waveform inversion: Theory: SEG Technical Program Expanded Abstracts, **207**, 1089–1093.
- Waters, K. H., 1978, *Reflection seismology*: John Wiley and Sons.
- Wenyong, P., Margrave, G., and Innanen, K., 2013, On the role of the deconvolution imaging condition in full waveform inversion: CREWES Research Report, **25**, 72.1–72.19.
- Zhou, C., Cai, W., Luo, Y., Schuster, G. T., and Hassanzadeh, S., 1995, Acoustic wave-equation traveltimes and waveform inversion of crosshole seismic data: *GEOPHYSICS*, **60**, No. 3, 765–773.

APPENDIX

Linearity of the migration operator

Claerbout (1971) suggests to map a reflection point (or layer) by a cross-correlation of the downgoing and upgoing wave fields:

$$Im(x, z) = g(x, z, t) \otimes u(x, z, t) \quad (28)$$

where $u(x, z, t)$ is the upgoing residuals wavefield $d_m - d_0$, d_m is the synthetic data using model m , d_0 is the acquired data and $g(x, z, t)$ is the downgoing source wavefield. From now on we will omit the coordinates x and z on the notation and keep in mind that we are imaging in depth. In the frequency domain (after a Fourier Transform) equation 28 becomes:

$$Im = U(\omega)G^*(\omega) \quad (29)$$

where $*$ denotes the complex conjugate, $U(\omega)$ and $G(\omega)$ are the Fourier Transform of $u(t)$ and $g(t)$, respectively. The linear property of the Fourier Transform leads to $U(\omega) = D_m(\omega) - D_0(\omega)$, where $D_m(\omega)$ and $D_0(\omega)$ are the Fourier Transform of d_m and d_0 , respectively. From equation 29:

$$\begin{aligned} Im &= U(\omega)G^*(\omega) \\ &= (D_m(\omega) - D_0(\omega))G^*(\omega) \\ &= D_m(\omega)G^*(\omega) - D_0(\omega)G^*(\omega) \\ Im(x, z) &= g(x, z, t) \otimes d_m(x, z, t) - g(x, z, t) \otimes d_0(x, z, t) \end{aligned} \quad (30)$$

This means that applying the imaging condition of equation 28 on the residuals u is equivalent to subtract the imaging condition of d_0 to the d_m , proving the linearity property of the migration operator.

Linearity of the stacking operator

For a common imaging point of a seismic survey we have N traces tr_{u_i} of the residuals $tr_{u_i} = tr_{m_i} - tr_{0_i}$ (i is the shot position or number, m relates the trace to the synthetic shot and 0 to the acquired data). Stacking them is:

$$S_u = \sum_{i=1}^N \frac{tr_{u_i}}{N} \quad (31)$$

Now solving the sum of the residual terms of equation 31 we demonstrate the linear property of stacking operator:

$$\begin{aligned}
 S_u &= \sum_{i=1}^N \frac{tr_{u_i}}{N} \\
 &= \sum_{i=1}^N \frac{tr_{m_i} - tr_{0_i}}{N} \\
 &= \frac{1}{N} (tr_{m_1} - tr_{0_1} + tr_{m_2} - tr_{0_2} + \dots + tr_{m_N} - tr_{0_N}) \\
 &= \frac{1}{N} [tr_{m_1} + tr_{m_2} + \dots + tr_{m_N} - (tr_{0_1} + tr_{0_2} + \dots + tr_{0_N})] \\
 &= \frac{1}{N} \left(\sum_{i=1}^N tr_{m_i} - \sum_{i=1}^N tr_{0_i} \right) \\
 &= S_m - S_0
 \end{aligned} \tag{32}$$

Linearity approximation of the impedance inversion operator

Another way to re-write the reflection coefficients R and impedances I of equation 11 is given by Treitel et al. (1995) for the reflection at point k is:

$$I_k = I_0 \prod_{k=1}^N \left(\frac{1 + R_k}{1 - R_k} \right) \tag{33}$$

The next step to to apply a Taylor expansion on the denominator. But first let's understand on the values the reflection coefficient can be. A strong water bottom reflector (water velocity = 1500m/s and first layer velocity = 2100m/s) would have a reflection coefficient of around 0.21 and the squared of it is 0.04. On a "well behaved" Earth, besides the water bottom and high impedance layers as salt or basalt, the reflection coefficients have values of around 0.03 or less, which squared would be 0.0007 and less. We came to this analysis so terms of power 2 and higher on the Taylor expansion and products can be neglected. With this information in mind, equation 33 is:

$$\begin{aligned}
 I_k &= I_0 \prod_{k=1}^N [(1 + R_k)(1 + R_k + R_k^2 + R_k^3 + \dots)] \\
 &\approx I_0 \prod_{k=1}^N [(1 + R_k)(1 + R_k)]
 \end{aligned} \tag{34}$$

Equation 34 leads to an easier solution for the impedance. Now we can open the product and, as previously, neglect powers 2 and higher of the reflection coefficients:

$$\begin{aligned}
I_k &= I_0 \prod_{k=1}^N (1 + 2R_k + R_k^2) \\
&\approx I_0 \prod_{k=1}^N (1 + 2R_k) \\
&= I_0 (1 + 2R_1)(1 + 2R_2)(1 + 2R_3) \cdots (1 + 2R_N) \\
&= I_0 (1 + 2R_1 + 2R_2 + 2R_3 \cdots 2R_N + O(\text{power} \geq 2)) \\
&= I_0 (1 + 2 \sum_{k=1}^N R_k) \tag{35}
\end{aligned}$$

This last result ended up to be very interesting. The sum of the reflection coefficients is a perturbation linear impedance trend with slope equals to $2I_0$ and intercept I_0 . For the FWI problem, if we have an initial model that contains the local Earth's linear trend (low frequency content) we can update only the perturbation, or the cumulative sum of the migrated acquired data with the low frequency removed. The resulting impedance perturbation will have zeros and negative values. As a model update, this means reducing equation 35 to:

$$I_{k_{update}} \approx I_0 \sum_{k=1}^N R_k \tag{36}$$

It is clear equation 36 has the linearity property if we understand that the reflection coefficient of the residuals is the difference of the reflection coefficients of the synthetic and acquired data and the approximation of equation 25 is valid.

In practice, we do not update the model by using equation 36. Instead we compute a more precise impedance inversion by the BLIMP method and then removing the low frequency's linear trend (slope) and the intercept (water velocity in our marine simulation), resulting in the velocity perturbation over the Earth's global trend.