Elastodynamic FWI efficiency study with partial stacking in 2D

Vladimir N. Zubov, CREWES, University of Calgary*
vzubov@ucalgary.ca
and
Gary F. Margrave, CREWES, University of Calgary
margrave@ucalgary.ca
and
Michael P. Lamoureux, CREWES, University of Calgary
mikel@ucalgary.ca

Summary

We study a method of accelerating conventional full waveform inversion using a data sampling approach implemented as partial stacking or fractional shots. This iterative, partial stacking FWI scheme has a simple algorithmic structure and high computational efficiency due to data and process decomposition. The data sampling algorithm provides a stable and noise resistance solution as shown in a numerical comparison with full stack FWI using synthetic data and non-random noise. The numerical results achieved on the Marmousi model, solving independently for either density or bulk modulus parameters, suggests the possibility of further partial stacking improvements using filtering and modelling characteristic estimations in FWI.

Introduction

Full Waveform Inversion, as well as many other inverse problems in physics, is known for its mathematical and computational complexity. This contributes to the reputation within both the mathematical and geophysical communities that FWI as a challenging problem. Better modelling and computer implementation could help to improve its reputation as a viable solution to the seismic imaging problem. Our goal with this work is to help simplify the method and improve numerical efficiency using the technique of partial stacking.

We begin the FWI model with the elastodynamic equations (Virieux 1986) representing the fundamental conservation laws in an ideal elastic 2D medium. Ideal elasticity is a useful simplified model, although it does ignore the real physical effects of viscosity and dissipation. We also assume a simplified horizontal structure in the physical model, allowing us to use a lower spatial grid resolution for the density and bulk modulus, to improve the algorithm's efficiency through shot decomposition (partial stacking). The data sampling method in FWI, together with low frequency filtering, is a natural extension that corresponds to the multigrid diagonal and block-diagonal preconditioning technique (Olshanski 2012). The conventional stacking Full Waveform Inversion iterative scheme used in the present study is simplified both in the block diagram and in the computer algorithm, to keep the computational process simple and understandable while achieving the desired efficiency.

The partial stacking approach is to improve the stability of the FWI model, increasing the independence of neighbouring shots in minimizing the misfit data through an artificial distance increase between shots in the partial stack. As a result, there is less contribution in the misfit from neighbours in the stack during the non-linear minimization. This approach is in contrast to shot gathering, or simultaneous

shooting, where the imaging condition is derived as an averaged imaging conditions of all shots in the gather. The independence of the shots is determined by the uniqueness of wavelets and also with uniqueness of the noise impact into data measurements for each particular shot.

Theory and/or Method

The elastodynamic equations governing the wave propagation in our 2D FWI are as follows:

$$\begin{cases} \frac{\partial}{\partial t}\rho\frac{\partial u_x}{\partial t} = \frac{\partial}{\partial x}(\lambda+2\mu)\frac{\partial u_x}{\partial x} + \frac{\partial}{\partial x}\lambda\frac{\partial u_z}{\partial z} + \frac{\partial}{\partial z}\mu\frac{\partial u_z}{\partial x} + \frac{\partial}{\partial z}\mu\frac{\partial u_x}{\partial z} \\ \frac{\partial}{\partial t}\rho\frac{\partial u_z}{\partial t} = \frac{\partial}{\partial x}\mu\frac{\partial u_z}{\partial x} + \frac{\partial}{\partial x}\mu\frac{\partial u_x}{\partial z} + \frac{\partial}{\partial z}\lambda\frac{\partial u_x}{\partial x} + \frac{\partial}{\partial z}(\lambda+2\mu)\frac{\partial u_z}{\partial z} \end{cases} \quad (1)$$

where (u_x, u_z) is the deformation vector of the medium, λ and μ are bulk moduli and ρ is a density. With these variables, we formulate two independent FWIs as inverse coefficient problems for bulk modulus λ and the density field ρ respectively. Both FWI are posed in here traditional way (Virieux 2009) as a misfit objective function minimization from the impulsive point source response on the number of the geophones at the air-earth boundary layer. As a result, the FWI iterative block diagram can be presented the following way (Figure 1).

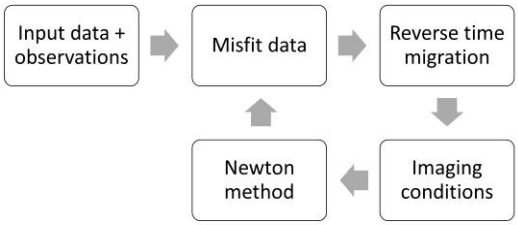

Figure 1. FWI algorithm iterative block diagram

In this FWI algorithm, the input data block includes both observations and the geophysical expert's position on modelling spatial and temporal grids' characteristic sizes as well as the source wavelet signature, assumed known for simplicity. The misfit data block is a forward wave propagation using the estimated values for the physical parameters, while the misfit data derivation (Virieux 2009) uses not the full stack of shots, but only a part of it. We call this the partial stack, a version of data sampling.

The choice of the shots in the partial stack changes as we repeat the FWI iterations and serves the following purposes:

- maximize the minimum distance between each two neighbor shots in the partial stack;
- involve all shots in the FWI equally;
- keep the shot selection as both periodic and piece-wise linear.

In the next section we used the shots switching strategy presented in the Table 1.

Table 1. Shots switching strategy in partial stacking FWI:
blue cells – active shots, white cells – inactive shots.

FWI iteration	Full stack of shots										
	1	2	3	4	5	6	7	8	9	10	11
1	→				→				→		
2		→				→				→	
3			→				→				→
4				→				→			
5	←				←				←		
6				←				←			
7			←				←				←
8		←				←				←	
9	→				→				→		

The next block of reverse time migration includes simultaneous backwards in time wave propagations for the source wavelet with the direct operator (1) and for the misfit data using adjoint to operator (1). We need to mention that the equation (1) is both self-adjoint and time direction invariant, which means that differential evolution operator from (1) is used for RTM with only changes in initial conditions.

The imaging conditions block of the diagram includes a misfit data gradient, derived in the traditional way (Tarantola 1984, Hasanov 2011) using the following formulas for density FWI and bulk modulus λ FWI respectfully:

$$I_\rho(d) = \int_0^T \left(\frac{\partial \Phi_x}{\partial t}\frac{\partial u_x}{\partial t} + \frac{\partial \Phi_z}{\partial t}\frac{\partial u_z}{\partial t} \right) dt \tag{2}$$

$$I_\lambda(d) = \int_0^T \left(\frac{\partial \Phi_x}{\partial x} + \frac{\partial \Phi_z}{\partial z} \right)\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} \right) dt \tag{3}$$

where u is a solution of direct operator source wave propagation, d is a misfit, and Φ is a misfit data migrated back in time. The integrals in (2) and (3) are the standard imaging conditions formulas (37A) and (37B) in Tarantola's paper and are computed explicitly. The simplification made in the gradient derivation loses information about its sign. That is why we use the second order Newton method in the misfit function minimization as it is invariant in the gradient sign direction and also provides second order accuracy in gradient step search.

It is an important to mention that in a partial stacking version of the FWI imaging condition for each shot in the partial stack is derived independently from neighbor shots using the corresponding misfit objective function.

The last block is the usual Newton method, which minimizes the misfit data norm d_2 (4) as follows:

- for each shot in the partial stack, the dependence of d_2 from the gradient step ε in gradient direction $I_{\rho_n}(d)$ (in density FWI case) is approximated by a parabola (Figure 2) with 3 points;
- for each shot in the partial stack, the optimal gradient step ε^* is estimated with a minimum of the corresponding parabola (Figure 2), $\varepsilon^* = b/2a$ and the density field correction for this particular shot is taken as $\Delta\rho_i = \varepsilon^* I_{\rho_n}(d)$;
- for the whole partial stack, the next density field ρ_{n+1} is estimated with a sum (or weighted with σ_i sum which is optional) of all $\Delta\rho_i$ applied to ρ_n, so that $\rho_{n+1} = \rho_n + \Sigma_{\text{partial stack}} \sigma_i \Delta\rho_i$.

$$d_2 = \left\| \left(\frac{\partial}{\partial t} u_{x\,\text{estimated}} - \frac{\partial}{\partial t} u_{x\,\text{observed}} \right)^2 + \left(\frac{\partial}{\partial t} u_{z\,\text{estimated}} - \frac{\partial}{\partial t} u_{z\,\text{observed}} \right)^2 \right\|_{L_2}^{1/2} \tag{4}$$

The objective function (4) represents the vertical component of the misfit, which is a deformation of the speed function. The energy norm in (5) is specified to increase the weight of the first arrivals.

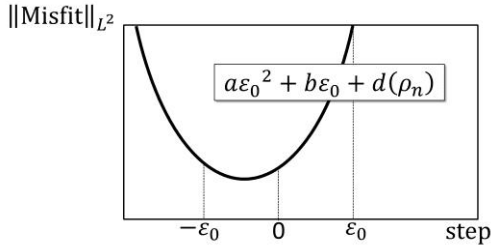

Figure 2. Paralobic approximation of the misfit norm dependance from the gradient step in imaging conditions direction with 3 points.

The second order Newton method implicitly uses domain decomposition with overlaps. Due to these overlaps of fixed size, the partial stacking method will run just a few times slower than the

corresponding shot gathering approach applied to the same FWI algorithm (Figure 1), dependent upon the size of overlaps. On the other hand, the partial stacking treats neighboring shots with different and incomparable wavelet signatures as independent, thus not affecting each other's objective function. This contributes to stability in the FWI algorithm and to predictability of its convergence.

Keeping both simplicity and predictability in the FWI algorithm, we can test this robust and fast algorithm for computational efficiency and the noise resistance.

Examples

We give two examples to show the numerical stability and convergence behaviour of the partial stacking approach in the FWI iterative algorithm. Both experiments run with the same physical parameters in the same modelling conditions, so the two experiments are comparable.

In a spatial area Ω of size 4×1 we introduce a rectangular grid with 321×475 points, and a point size of 0.0125×0.0021. The grid points are stretched by a factor of 6 in the horizontal direction, making the model more horizontal than the standard. The spatial domain Ω consists of 2 layers: air and earth with a horizontal linear interface in between. The density values used in the corresponding FWI is presented in Figure 3 and cover both layers of air and earth. In this case, the other parameters λ and μ in (2) are taken as constant 1 in the earth, while in the air we set $\lambda=\mu=0.01$ and $\rho=0.16$. This creates a realistic difference in the wave propagation speed in air and earth. The propagation speed in air is fixed and assumed known, so that the inversion problem is formulated for the earth only.

All 6 point sources in the full stack model shown in Figure 6 are located at 5 grid points under the surface. The line of geophones situated on the interface between air and earth is limited with the corresponding spatial window in which each source is processed. An example of this window is presented in the Figure 3 as well.

Starting with initial smooth guess of the density parameter, preserving the zero-frequency trend of the exact reflection coefficient field, we apply 128 iterations of the FWI to obtain the following approximate density solution. In these 128 iterations, we changed filtering 32 times starting with low frequencies (Figure 4 b) and continuously involving higher frequency components. Each filtering is used for 8 iterations of the FWI. For each FWI iteration, just one iteration of the corresponding Newton gradient search is implemented.

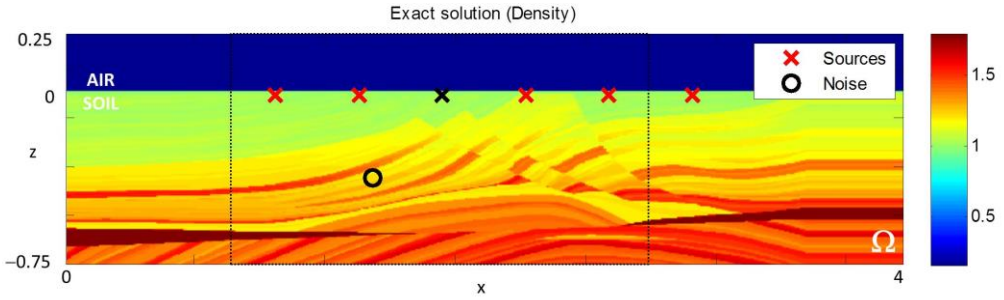

Figure 3. Density field for the synthetic observation's exact solution with a dotted line representing the spatial window for black x source processing and periodical noise affecting only this source.

Each shot wavelet, marketed in the Figure 3 with red and black x, has the same wavelet signature in all examples (Figure 4 a) and, as a result, the same low frequency filtering strategy (Figure 4 b) is applied for all shots in the stack at the same time. For noise testing, we insert a high frequency noise source with an unspecified location affected only one shot. We assume we don't know which shot is affected (Figure 6). We also assume that we don't contribute any other error to the existing finite difference approximation and numerical integration.

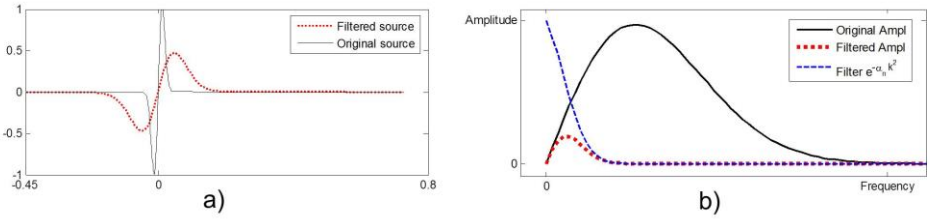

Figure 4. Source function $f(t) = e^{-(\eta \cdot t)^2} \sin(\omega \cdot t)$ filtering result:
a) signal in time domain; b) amplitude spectrum and filter in frequency domain

We compare full stack FWI using 6 shots' stack with the partial stacking FWI using 25% of the 11 shots on each iteration with the corresponding (Table 1) shots' switching. The numerical FWI convergence for the experiment is presented in Figure 5. The partial stacking FWI runs in two times faster than the full stack FWI. In spite of this, the convergence results for both stacking approaches are almost the same. We note in other experiment (not presented here) we have observed faster convergence in the full stack case in deeper layers. As a result, it appears that partial stacking FWI may have more advantages as a dynamic and filter dependent procedure within an iterative FWI scheme and should probably involve more shots in the stack with low frequency filtering than with the high frequency one.

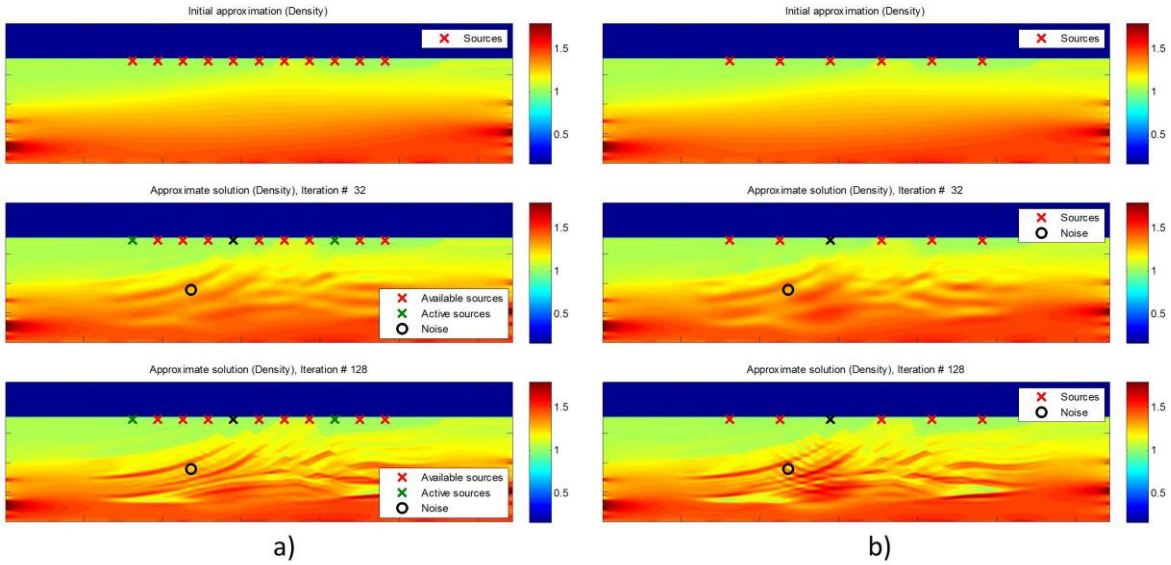

Figure 5. Approximate solution ρ_n convergence behavior with periodical noise affecting black x shot:
a) – partial stack FWI for 11 sources; b) – full stack 6 shots FWI.

Studying the noise impact, the specified high frequency noise has almost no effect on both stacks until the moment when filtering starts increasing the weight of the noise frequency in the misfit data masking. It appears after the 32nd iteration for both stacks (Figure 5) simultaneously. Moreover, the partial stack successfully averaged the defective shot with additional neighbor shots not involved in the full stack, thus keeping a comparable level of computational complexity for both stacks. As a result, the noise significantly distorted the final full stack FWI solution while partial stack shows good accuracy.

In the second experiment we show the bulk modulus partial stacking FWI experiment. The convergence of the iterative FWI in this case appears to behave similarly to the density FWI. We hope to combine both methods in future studies, for a joint FWI. A comparison of the results in the two approaches is presented in Figures 6-7. The exact solution in bulk modulus FWI are considered the following (in the layer of earth only): λ_{exact} (Figure 6 b) is just inverse density in corresponding density FWI (Figure 6 a), $\mu_{exact} \equiv \lambda_{exact}$ and $\rho_{soil} \equiv 1$.

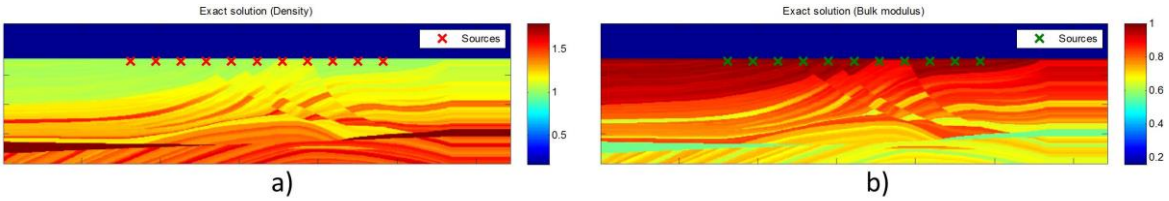

Figure 6. Exact solution for partial stack FWI: a – density field FWI; b – Bulk modulus FWI.

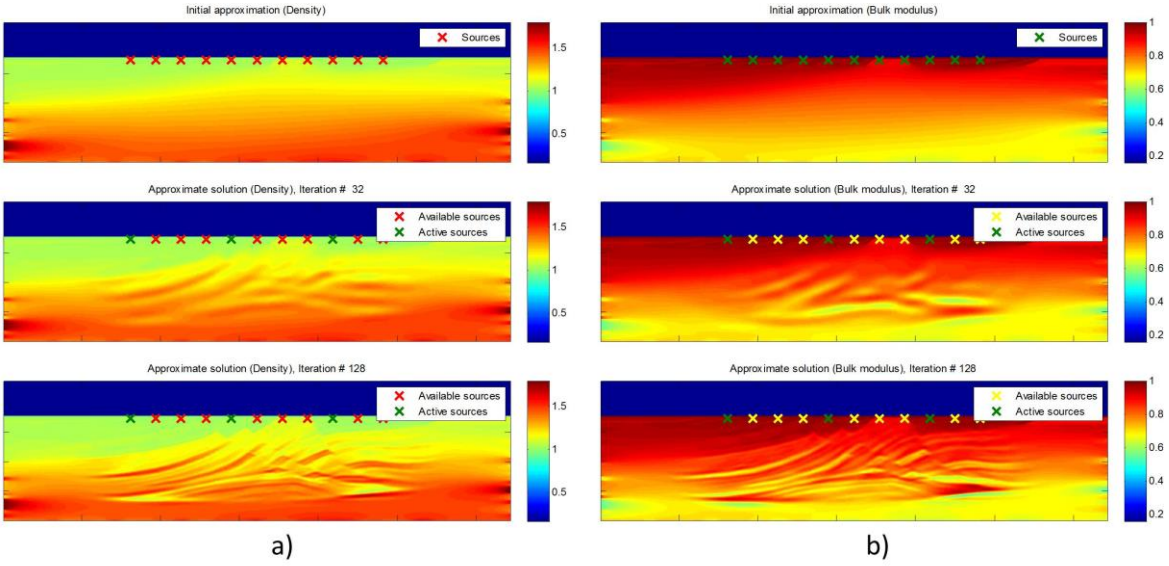

Figure 7. Approximate solution: a – density partial stack FWI; b – bulk modulus partial stack FWI

At the 32nd iteration, the FWI results for bulk modulus are better at deeper layers than in density FWI (Figure 7, iteration 32). At the same time, the high frequency error cancellation is more successful in density FWI (Figure 7, iteration 128). Note that given how bulk modulus λ and μ appear in the PDE, both parameters are required to be differentiable. For this reason, it is hard to expect great success in high frequencies FWI for λ, as compared to ρ.

Conclusions

The techniques presented here for optimizing the FWI algorithm appear to support fast and effective data processing in the case of large stacks of shots. The partial stack method accelerates the stack convergence for both density and bulk modulus FWI.

The results obtained with the Marmousi density model (with preserved vertical resolution of the field) and partial stack 6 shots FWI (Figure 7 a) supports the hypothesis that the FWI, with either partial stack or shot gathering techniques, will be fast and numerically economical.

The numerical study for both density and bulk modulus FWI shows potential to further optimize the stacking algorithm by synchronizing it with frequency filtering. In future work, we intend to use filtering and partial stacking to compare with shot gathering techniques of the same computational difficulty and identical modelling conditions, to get a better understanding of the advantages of partial stacking.

Acknowledgements

We thank the CREWES consortium and its sponsors, the agencies NSERC and PIMS, and Department of Mathematics and Statistics for their support of this work. We would also like to thank Joe Wong for his comments and suggestions.

References

Hasanov A., Pektas B. and Erdem A. 2011, Comparative analysis of inverse coefficient problems for parabolic equations. Part I: adjoint problem approach, Inverse Problems in Science and Engineering, 19:5, 599-615

Hogan C. and Margrave G., 2006, Measurement of convergence in plane-wave migration: CREWES annual report 38, University of Calgary.

Margrave G., Ferguson R. and Hogan C., 2011, Full Waveform Inversion Using One-way Migration and Well Calibration: CSEG Convention Abstracts

Olshanskii M.A. 2012, Lecture notes on multigrid methods: Moscow, Institute of Numerical Mathematics, RAS.

Pratt R.G., Shin C. and Hicks G.J. 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysics 133, 341-362.

Tarantola A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 74, No. 8, 1259-1266.

Virieux J. 1986 P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, 51, No 4, 889-901.

Virieux J. and Operto S. 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics 74, No. 6, WCC1-WCC26.