

Local optimization approaches for simultaneous AVO inversion based on re-parameterized Zoeppritz equations

Mariana Lume and Kristopher Innanen
CREWES, University of Calgary

Summary

Linearized AVO inversion methods, such as the weighted stacking, are based on approximations of the Zoeppritz equations subject to several assumptions, including the limitation of incidence angles to 35-40°. Thus, in long-offset seismic acquisitions these approaches fail. In this project, we focused on developing a nonlinear inversion appropriate for these circumstances, modifying the Zoeppritz equations in terms of the fractional density and compressional and shear impedances and considering synthetic P-P and P-S datasets with different noise and frequency content. Overall, three different local optimization algorithms, under a good initial model, demonstrated a superior performance respect to the simultaneous weighted stacking, producing more accurate results for most elastic parameters. Additionally, the advantages of including both datasets instead of only the P-P reflectivities were illustrated.

Forward modeling

To perform the forward problem, we considered plane wave Zoeppritz equations since mathematically these are not conditioned to any particular range of angles. However, in reality a plane wave approximation would be made, since seismic data is not produced by plane but spherical waves. These equations can be written as:

$$\mathbf{P}\mathbf{u} = \mathbf{b} \quad (1)$$

where:

$$\mathbf{P} = \begin{bmatrix} -X & -\sqrt{1-B^2X^2} & CX & \sqrt{1-D^2X^2} \\ \sqrt{1-X^2} & -BX & \sqrt{1-C^2X^2} & -DX \\ 2B^2X\sqrt{1-X^2} & B(1-2B^2X^2) & 2AD^2X\sqrt{1-C^2X^2} & AD(1-2D^2X^2) \\ -(1-2B^2X^2) & 2B^2X\sqrt{1-B^2X^2} & AC(1-2D^2X^2) & -2AD^2X\sqrt{1-D^2X^2} \end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix} R_{pp} \\ R_{ps} \\ T_{pp} \\ T_{ps} \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} X \\ \sqrt{1-X^2} \\ 2B^2X\sqrt{1-X^2} \\ 1-2B^2X^2 \end{bmatrix}; \quad \text{and} \quad A = \frac{\rho_2}{\rho_1}; \quad B = \frac{\beta_1}{\alpha_1}; \quad C = \frac{I_2 \rho_1}{I_1 \rho_2}; \quad D = \frac{J_2 \rho_1}{J_1 \rho_2}; \quad X = \sin \theta_1$$

here α , β , ρ , I and J are the P-wave velocity, S-wave velocity, density, compressional impedance and shear impedance, respectively. Moreover, equation 1 was re-parameterized in terms of the fractional impedances and fractional density, which is familiar from AVO analysis. To accomplish this, terms A , C , and D were modified using the following equation commonly applied when linearizing problems:

$$\frac{Y_2}{Y_1} = \left(1 + \frac{1}{2} \frac{\Delta Y}{Y}\right) / \left(1 - \frac{1}{2} \frac{\Delta Y}{Y}\right) \quad (2)$$

As a result, each element of \mathbf{P} is a nonlinear function of $\Delta I/I$, $\Delta J/J$, and $\Delta \rho/\rho$. The vector \mathbf{u} is obtained for a chosen P-wave incidence angle θ_1 . To work with simultaneous data, different incidence angles for P-P and P-S reflectivities were computed through raytracing, in order each pair of plane waves reach the same receiver. Hence, the forward problem needs to be performed with each of these angles to later construct the appropriate vector of coefficients.

Iterative nonlinear simultaneous inversion

In an unconstrained optimization, an objective function, without restrictions on the values that the variables can take, is minimized (Nocedal and Wright, 2006). This objective function was constructed measuring the L2 norm between predicted and observed (d_{obs}) data vectors:

$$\Phi(\mathbf{m}) = \frac{1}{2} \sum_{j=1}^N \left(\mathbf{S} \mathbf{u}(\mathbf{m}, \theta_j) - \mathbf{d}_{obs}(\theta_j) \right)^T \mathbf{W}^T \mathbf{W} \left(\mathbf{S} \mathbf{u}(\mathbf{m}, \theta_j) - \mathbf{d}_{obs}(\theta_j) \right) \quad (3)$$

where:

$$\mathbf{m} = \begin{bmatrix} \Delta I \\ I \\ \Delta J \\ J \\ \Delta \rho \\ \rho \end{bmatrix}^T; \quad \mathbf{W} = \begin{bmatrix} 1/\sigma_{pp} & 0 \\ 0 & 1/\sigma_{ps} \end{bmatrix}; \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}; \quad \frac{\Delta q}{q} = \frac{\Delta I}{I} - \frac{\Delta J}{J}$$

Inside \mathbf{W} , σ is the standard deviation of the dataset; \mathbf{S} is a sampling operator used to extract only the P-P and P-S reflectivities from \mathbf{u} . Moreover, with the elastic parameters of \mathbf{m} , values of the fractional Vp/Vs ratio ($\Delta q/q$) can be calculated.

To approach the local minimum, the algorithm selects a direction \mathbf{p} and a step length α to move per iteration, lowering the cost function until either no more progress is made or the solution is approximated with a given accuracy. The local gradient and/or curvature of the objective function is used to compute the model update (Sen and Stoffa, 2020), according to:

$$\Delta \mathbf{m} = \alpha \mathbf{p} = -\alpha \mathbf{B}^{-1} \mathbf{g} \quad (4)$$

where \mathbf{B} depends on the applied method. In this study, it was set equal to the identity matrix for the Steepest Descent method, the Hessian (\mathbf{H}) matrix for the Gauss-Newton method, and $\mathbf{H} + \lambda \text{diag}(\mathbf{H})$ for the Levenberg-Marquardt modification, where λ is a damping parameter adjusted per iteration and equivalent to a step size (Tolle, 2003).

In addition, the gradient (\mathbf{g}) and Hessian (\mathbf{H}) take the form:

$$\mathbf{g} = \sum_{j=1}^N \mathbf{J}_j^T \mathbf{W}^T \mathbf{W} \left(\mathbf{S} \mathbf{u}(\mathbf{m}, \theta_j) - \mathbf{d}_{obs}(\theta_j) \right); \quad (5) \quad \mathbf{H} = \sum_{j=1}^N \mathbf{J}_j^T \mathbf{W}^T \mathbf{W} \mathbf{J}_j \quad (6)$$

Where \mathbf{J} is the Jacobian matrix, formed by derivatives of the predicted data with respect to each model parameter. Each of its elements, per incidence angle, are obtained through:

$$J_{\mu}^i = S_K^i \frac{\partial u^K}{\partial m^{\mu}} = -S_K^i (P^{-1})_L^K \frac{\partial P_M^L}{\partial m^{\mu}} u^M; \quad i = 1,2; \mu = 1,2,3; K = M = L = 1,2,3,4 \quad (7)$$

When computing **J**, the main new task is to determine all 48 elements of $\frac{\partial P_M^L}{\partial m^{\mu}}$ in equation 7. Additionally, the rows of $\frac{\partial u^K}{\partial m^{\mu}}$ are formed by the derivative of a coefficient with respect to each model parameter; then, to consider the raytracing approach, this matrix is computed twice (one time with P-P and another time with the P-S incidence angles), in order the first and third rows from $\frac{\partial u^K}{\partial m^{\mu}_{pp}}$ and the second and fourth rows from $\frac{\partial u^K}{\partial m^{\mu}_{ps}}$ form the final $\frac{\partial u^K}{\partial m^{\mu}}$ matrix.

Later, values of the model vector are modified per iteration (k) according to:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta \mathbf{m}_k \quad (8)$$

Comparison between optimization methods

We designed a two-layer model of solid units in welded contact that does not produce a critical angle and assuming a range of offsets from 0 to 4000m with intervals of 80m. This model meets most of the underlying assumptions of the weighted stacking (Larsen, 1999), i.e., weak contrast of elastic parameters and a value of the Vp/Vs ratio between 1.5 and 2. However, the forward modeling showed that almost half of these incidence angles are higher than 35°, reaching a maximum of 53° for the P-P dataset and 65° for the P-S data. Thus, it is expected that the weighted stacking does not produce accurate estimates under these circumstances.

For the nonlinear inversion, the initial model was constructed from weak perturbations of the true values of α , β , and ρ . Nevertheless, since $\Delta I/I$ and $\Delta J/J$ result from multiplying velocities and densities, their initial perturbations were not that small. In addition, estimations were evaluated through:

$$\%error = \frac{\frac{\Delta(I, J, q \text{ or } \rho)}{(I, J, q \text{ or } \rho)} \Big|_{CALCULATED} - \frac{\Delta(I, J, q \text{ or } \rho)}{(I, J, q \text{ or } \rho)} \Big|_{TRUE}}{\frac{\Delta(I, J, q \text{ or } \rho)}{(I, J, q \text{ or } \rho)} \Big|_{TRUE}} \times 100 \quad (9)$$

Figure 1 illustrates results when inverting broadband and noise free reflectivities. Pure Gauss-Newton had the fastest convergence rate, finding the minimum point at the second iteration. However, Steepest Descent and Levenberg-Marquardt had a slower performance, reaching convergence at the 17th and 10th iteration, respectively. In addition, estimations were very close to the true values with almost zero data residuals. Particularly for Steepest Descent, the RMS error of P-P reflectivities increased in early iterations, indicating more instability. Moreover, results from all optimizations got trapped into the same local minimum, but these methods were significantly superior than the weighted stacking, evidencing error differences of around 2.5% between estimates and allowing the calculation of the fractional density.

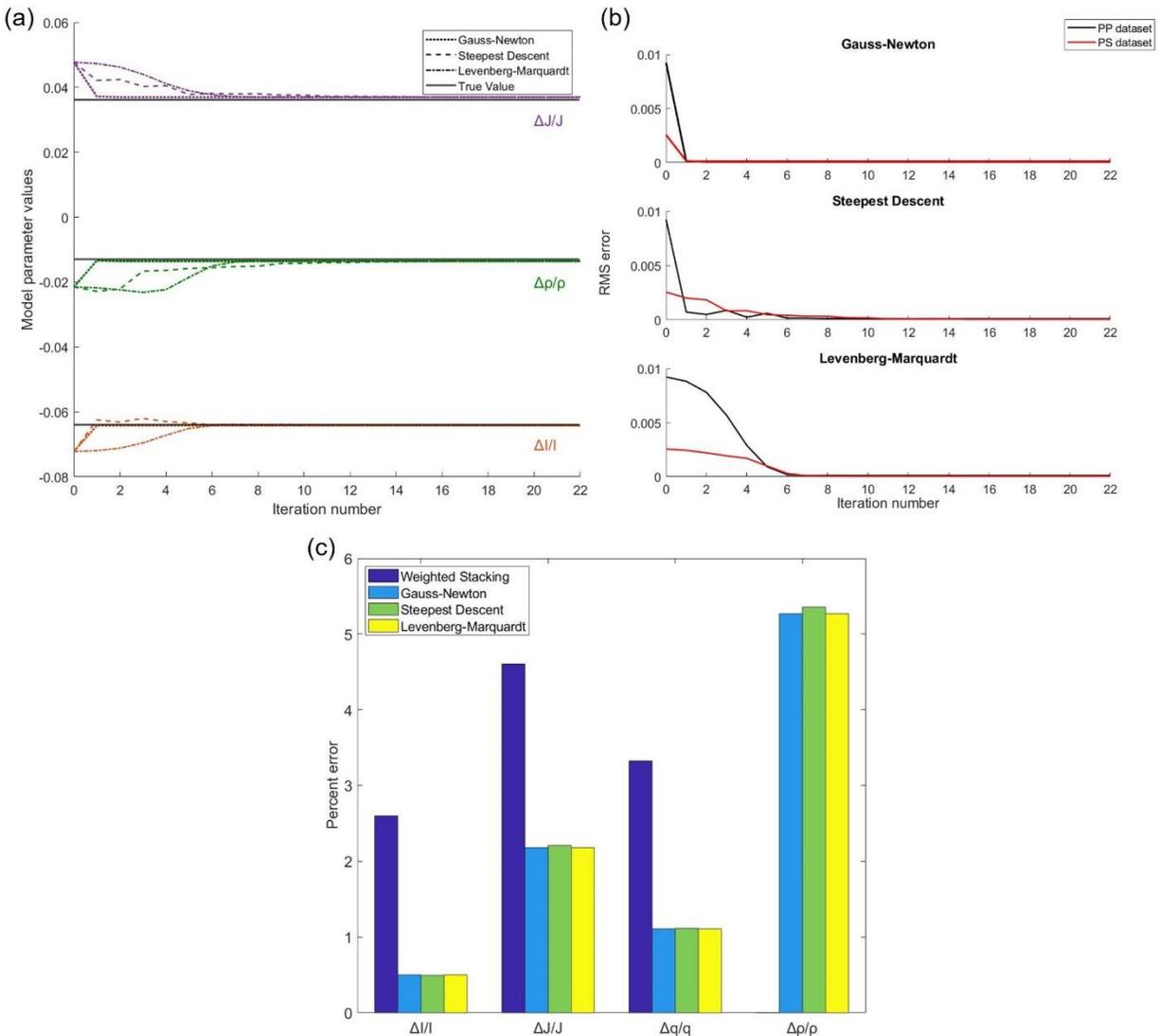


Figure 1. (a) and (b) Convergence of optimization algorithms using broadband and noise free reflectivities. (c) Accuracy test performed with estimates from linearized and nonlinear inversions.

Since the observed data can be treated as a random variable, because measurements always contain random noise (Zhdanov, 2002), broadband and band-limited reflectivities were contaminated with noise and each inversion was repeated 5000 times for statistical analyses, comparing the maximum likelihood solutions. Figure 2a indicates that for noisy broadband data, results from each optimization were very similar, resembling to those of the noise free case. For noisy band-limited data, Figure 2b shows that results were almost identical between optimizations, but with smaller accuracy than those obtained in Figure 2a. However, errors stayed within an acceptable range, except for $\Delta \rho/\rho$ which reached a dramatic value of approximately 66%. Therefore, P-P and P-S datasets have little sensitivity to the fractional density, not doing a proper constraint of its band-limited true value. Overall, although the nature of the stacking

procedure helps to suppress noise (Larsen, 1999), once more the nonlinear approach produced better estimations of most model parameters and for both reflectivity cases.

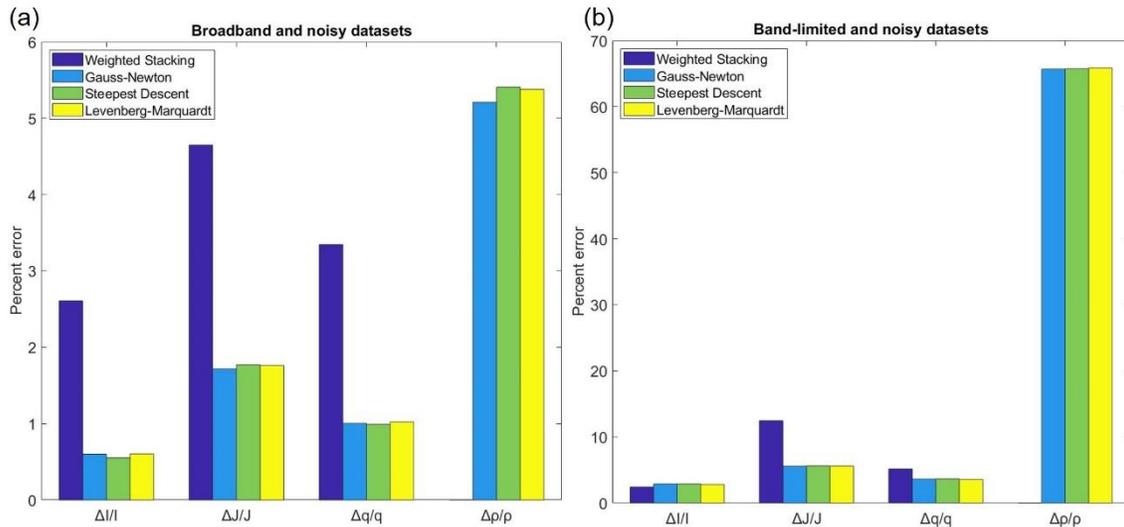


Figure 2. Accuracy tests performed with maximum likelihood solutions from linearized and nonlinear inversions. A signal-to-noise ratio of 8 was assigned to R_{pp} and of 4 to R_{ps} .

Advantages of the simultaneous nonlinear inversion

The benefits of including P-S reflectivities in the inversion were studied with the Pure Gauss-Newton algorithm. For the conventional P-P inversion, only the first row of \mathbf{S} , as well as all θ_{pp} were considered. In this case, the condition number of \mathbf{H} was approximately 300, while in the simultaneous technique it was approximately 24. Thus, similar to the analysis shown by Larsen (1999), incorporating P-S data stabilized the inversion. Figure 3 illustrates the posterior uncertainty distribution of each estimate after statistically inverting noisy band-limited datasets. The conventional inversion generated an almost perfect maximum likelihood solution for $\Delta I/I$, but this method showed lack of accuracy and precision for the remaining parameters. Although the simultaneous inversion sacrificed accuracy for $\Delta I/I$, it produced maximum likelihood solutions of $\Delta J/J$ and $\Delta q/q$ closer to their true values and all the estimations were much more precise, hence there is more probability to compute better estimates with this approach. Unfortunately, the nonlinear inversion of simultaneous or conventional reflectivities could not generate a good maximum likelihood solution for $\Delta \rho/\rho$, but including P-S reflectivities guaranteed more precision.

Conclusions

Synthetic data allowed to confirm the outperformance of the simultaneous nonlinear AVO inversion over the weighted stacking under scenarios of long-offsets. Conditioned to a good initial model, estimations of the fractional compressional and shear impedances, as well as the fractional V_p/V_s ratio, from local optimizations, were very accurate with remarkably smaller percent errors. Additionally, regardless the applied optimization algorithm and depending on the type of reflectivities inverted, convergence was reached to the same minimum point, but with

different convergence features. Moreover, advantages of applying a simultaneous nonlinear inversion over a conventional P-P technique were demonstrated by the high accuracy and precision of the results. Finally, the nonlinear inversion had a negative impact when estimating $\Delta\rho/\rho$ from band-limited reflectivities, indicating the inadequacy of the dataset to produce a proper constraint to the right answer. Hence, in future work, it is important to determine improvements for a better inversion of this elastic parameter.

Acknowledgements

We would like to thank the sponsors of the CREWES project as well NSERC (Natural Science and Engineering Research Council of Canada) under the grant CRDPJ 543578-19 for making this work possible through their financial support.

References

Larsen, J. A., 1999, AVO analysis by simultaneous P-P and P-S weighted stacking: M.Sc. thesis, University of Calgary.

Nocedal, J., and Wright, S. J., 2006, Numerical Optimization: Springer.

Sen, M. K., and Stoffa, P. L., 2020, Inverse theory, global optimization, in Gupta, H. K., Ed., Encyclopedia of Solid Earth Geophysics: Springer Netherlands, 625–632.

Tolle, J. W., 2003, Nonlinear programming, in Meyers, R. A., Ed., Encyclopedia of Physical Science and Technology: Academic Press, 583–596.

Zhdanov, M., 2002, Geophysical Inverse Theory and Regularization Problems: Elsevier Science.

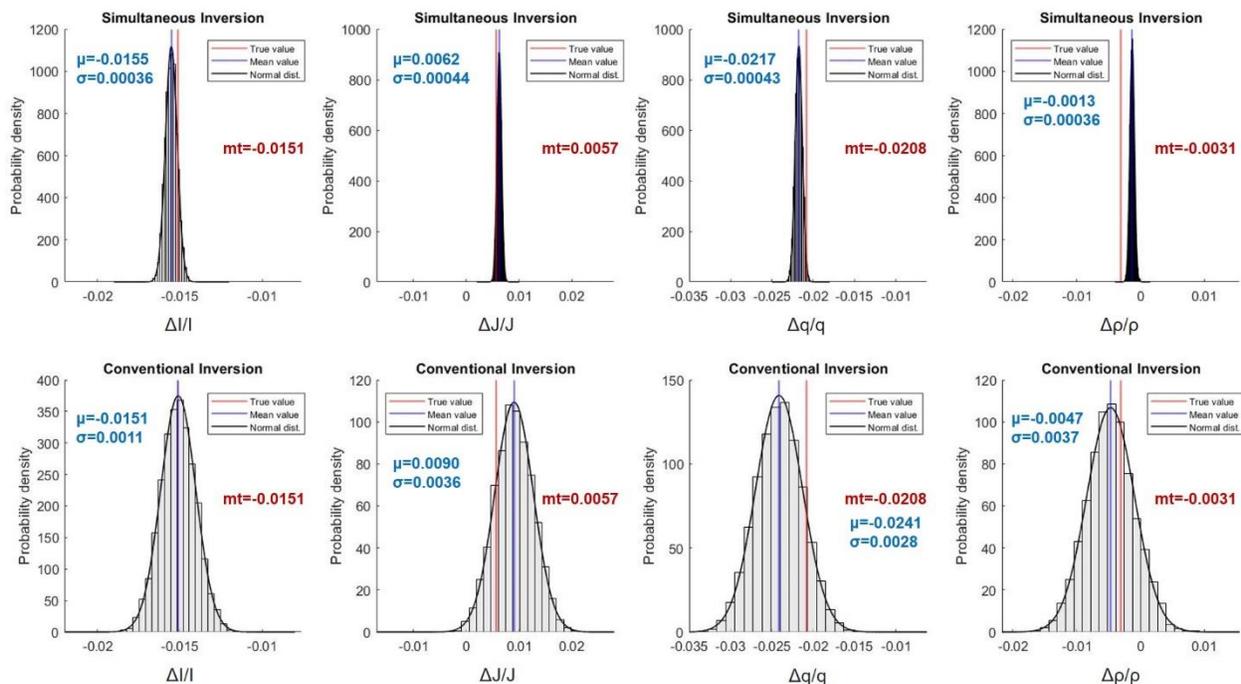


Figure 3. Posterior uncertainty distributions that resulted from noisy band-limited reflectivities; “ mt ” is the band-limited true fractional value. Reflectivities were band-limited as in Figure 2b and contaminated with the same noise characteristics.