A Novel Nonlinear Conjugate-Gradient AVO Waveform Inversion Method with Full Knott-Zoeppritz Forward Modeling and Local Minima Circumvention Techniques

Anthony Barone[1] and Anthony Vassiliou[2]*
[1]University of Texas at Austin, Institute for Geophysics, [2]GeoEnergy Inc., Houston, Texas

Summary

We present a novel nonlinear waveform inversion method for subsurface parameter estimation. This method utilizes the full Knott-Zoeppritz equations for forward modeling seismic data in combination with an effective and efficient gradient-based optimization scheme that is capable of bypassing local minima. The use of the Knott-Zoeppritz equations allows for significantly better recovery of shear wave and density parameters than conventional AVO approximations allow. Our optimization scheme uses a flexible error function that can be modified to produce optimal results for multiple use cases. We apply our method to synthetically generated seismic data and invert for compressional wave velocity (V_p), shear wave velocity (V_s), and bulk density (ρ). Synthetic results show our method to be extremely effective, even when the signal-to-noise ratio (SNR) is poor or the initial model is inaccurate.

Introduction

Over the past several decades, Amplitude-vs-Offset (AVO) information in seismic data has been successfully used to estimate subsurface elastic properties. The physical laws governing the reflection of seismic energy at a layer interface is related to elastic properties of the layers above and below the interface and the incident angle of the wave. This information in turn can be used to formulate in inverse problem to estimate the elastic properties of the layers.

Traditionally, this inverse problem has been linearized using AVO approximations. This linearization simplifies the inversion, but comes at a cost: the approximation error increases significantly at high incident angles, making the forward modeling of far offset seismic data unreliable. In practice this significantly reduces the accuracy of shear wave and density estimates, since seismic reflectivity (R_{PP}) is more sensitive to these parameters at farther offsets / higher incident angles (Asveth et al., 2010). Using the full Knott-Zoeppritz equations in forward modeling avoids this issue and allows for better estimates of V_s and ρ, but requires a nonlinear optimization scheme. In this work, we present an optimization scheme that is designed explicitly for this purpose. Our method is based on conjugate-gradients and utilizes a quadratic stepsize code that is capable of escaping local minima. The optimization is efficient and has the throughput required for inverting large seismic datasets (several thousand iterations per hour using one 4-core/8-thread CPU). We successfully test our method with synthetic seismic data over a wide range of conditions. Synthetic data was modeled from a real well log using Zoeppritz equations, a convolutional model, a Ricker wavelet, and various levels of noise. Additionally, we successfully performed this inversion on real 3D seismic data from an unconventional play. Unfortunately, due to data licensing and confidentiality agreements we are not able to include these results in this paper, but we will briefly discuss the process we used.

Method

The overall workflow of our optimization scheme is summarized in Figure 1, shown below. This inversion uses the full Knott-Zoeppritz equations and convolutional modeling to predict synthetic seismic data. In the convolutional forward modeling we used a Ricker wavelet for the synthetic data inversion and a statistically derived wavelet for the real data inversion. During the real data inversion we converted to the angle domain using the current P-wave velocity model and Equation 1. Note that this requires the forward modeler must be capable of using time-varying incident angles. Iteratively updating the incident angles ensures that the incident angles are closer to their true values when the final model is inverted. The synthetic data, however, was generated in angle domain and as such we did not convert the synthetic data to offset.

$$\sin(\theta) = \frac{V_{int}}{V_{rms}}\left(\frac{x}{\sqrt{x^2+\left(v_{rms}\, t_0\right)^2}}\right) \quad , \quad \begin{matrix} x = offset \\ t_0 = time \end{matrix} \quad . \qquad (1)$$

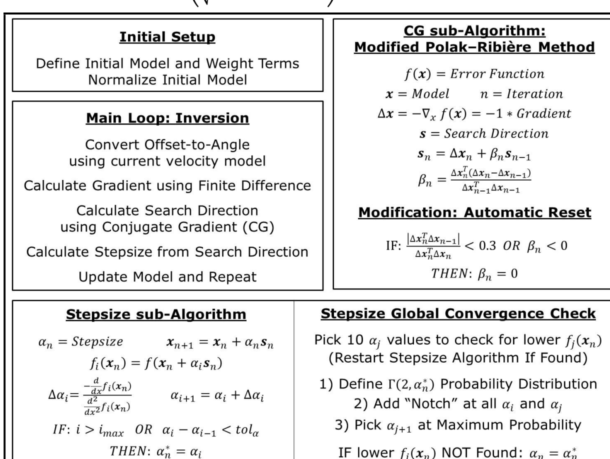

Figure 1: A diagram summarizing the main points of our inversion method. See Equation 2 for the specific formulation of our Error function and Equation 1 for the offset-to-angle conversion formula.

Prior to starting the inversion, the initial model is normalized. This serves to standardize the amplitudes of the parameters and stabilize the inversion. V_p and V_s were both normalized by the mean V_p, and density was normalized by the mean density. This normalization does not affect the layer contrasts used in the Zoeppritz equations, and thus does not affect the forward modeling operation. This normalization is reversed after the final model is found by the inversion algorithm.

In order to increase the range of situations in which our inversion method is successful, we chose to use a very flexible and general error function. Our error function, shown in Equation 1, consists of elements of L2 data norm, L2 model norm, total variation norm, and a constraint on V_p/V_s ratio of inverted results. Including L2 data and L2 model norm is standard practice; however, the total variation norm is included to help promote sparse layer-like models, and the V_p/V_s ratio constraint ensures the inversion results are physically realistic.

$$
\begin{aligned}
E(\boldsymbol{m}) = &\left(\boldsymbol{d} - g(\boldsymbol{m})\right)^T \underline{W_d}\left(\boldsymbol{d} - g(\boldsymbol{m})\right) \\
&+\alpha_1(\boldsymbol{m} - \boldsymbol{m_0})^T \underline{W_m}(\boldsymbol{m} - \boldsymbol{m_0}) \\
&+\alpha_2 \left|\underline{D_1}\ \underline{W_{TV}}\ \boldsymbol{m}\right|_1 + \alpha_3 \left(\frac{Vp}{Vs}\ constraint\right), \qquad (2)
\end{aligned}
$$

where $\boldsymbol{m}$ and $\boldsymbol{m_0}$ are the current and initial model, $\boldsymbol{d}$ is the seismic data, $g(\boldsymbol{m})$ is the forward operator used to make synthetics, α_1 to α_3 are weights, $\underline{D_1}$ is a 1st derivative finite difference matrix, $\underline{W_d}$, $\underline{W_m}$, and $\underline{W_{TV}}$ are weighting matrices, and $|\ldots|_1$ indicates the L1 norm. Specific weights are chosen via trial and error, though we found that having error primarily come from the L2 data norm, followed by the total variation norm, and lastly by the L2 model norm and V_p/V_s constraint, works well in practice. We constructed our V_p/V_s constraint by constructing a probability distribution (P) of V_p/V_s ratios using the well information as a rough guide. The error is then calculated using: $E\left(\frac{V_p}{V_s}\right) = -\log\left[P\left(\frac{V_p}{V_s}\right)\right]$. Refer to Figure 2 for a visual example of the probability distribution (P).

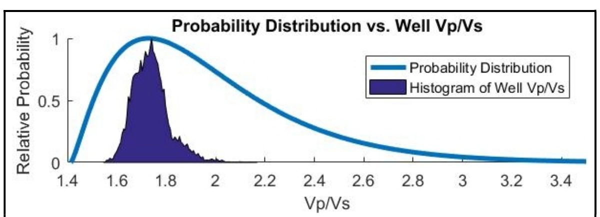

Figure 2: Example of a probabilility distribution used to construct the V_p/V_s constaint in the error function. The specific form of this distribution will vary based on the dataset, but in general should resemble what is shown in this figure.

After all parameters are set the inversion can be started. Once the (optional) offset-to-angle conversion has finished, the gradient of the error function is calculated using a standard finite difference scheme. Care is taken to avoid unnecessarily re-calculating information when possible. The search direction is found using a modified version of the Polak–Ribière conjugate gradient method (Polak and Ribière, 1969). Their method is modified to include automatic directional resets if the β parameter falls below zero (e.g., Hager and Zhang, 2006; Dai and Yuan, 1999) or if successive gradients are no longer sufficiently conjugate. This method is described in Equations 3 – 6.

$$\Delta\boldsymbol{x} = -\nabla_x f(\boldsymbol{x}) = -1 * Gradient\ , \qquad (3)$$

$$\boldsymbol{s}_n = \Delta\boldsymbol{x}_n + \beta_n \boldsymbol{s}_{n-1}\ , \qquad (4)$$

$$\beta_n = \max\left[0, \frac{\Delta x_n^T(\Delta x_n - \Delta x_{n-1})}{\Delta x_{n-1}^T \Delta x_{n-1}}\right], \qquad (5)$$

$$IF\text{:}\ \frac{|\Delta x_n^T \Delta x_{n-1}|}{\Delta x_n^T \Delta x_n} < 0.3 \ \rightarrow\ \beta_n = 0\ , \qquad (6)$$

where $\boldsymbol{x}$ is the model, $f(\boldsymbol{x})$ is the error function, $\boldsymbol{s}_n$ is the search direction, and n is the iteration number.

Once the search direction is found, the stepsize can be calculated using a line-search method. Our method assumes the error to be either quadratic or cubic in nature. Under this assumption, derivatives along the search direction are found using a finite difference technique. These derivatives are related to the polynomial coefficients, and thus can be used in standard root-finding solutions for quadratics and cubics to estimate the minimum value. The solution for this becomes the beginning of a new stepsize sub-iteration. When the stepsize change falls below some tolerance or a set number of sub-iterations have passed, the final stepsize estimate is found. As an example, in the quadratic case:

$$f_i(\boldsymbol{x}_n) = f(\boldsymbol{x}_n + \alpha_i \boldsymbol{s}_n)\ , \qquad (7)$$

$$\Delta\alpha_i = \frac{-\frac{d}{dx} f_i(x_n)}{\frac{d^2}{dx^2} f_i(x_n)} = -r\frac{1/2(e_+ - e_-)}{e_+ + e_- - 2e_0}, \qquad (8)$$

$$\alpha_{i+1} = \alpha_i + \Delta\alpha_i\ , \qquad (9)$$

where e_+, e_-, and e_0 are the error function values for positive, negative, and zero finite difference perturbation (r). At this point we then perform a check for global convergence by testing the error function at 10 additional optimally chosen stepsizes. These points are chosen such that the overall distribution of all attempted stepsizes mimics a $\Gamma(2, \alpha_n^*)$ distribution, where α_n^* is the stepsize estimate being tested. To do this, the initial distribution if multiplied by a notch distribution for all attempted α_i, and the maximum probability is found with a targeted brute-force method. This process ensures the stepsizes used for checking global convergence are not stepsizes with already known error. If a lower error is found, the stepsize algorithm is re-started using that stepsize as a new seed; otherwise, α_n^* is assumed to produce the global minimum.

After the search direction and stepsize have been found, the model can be updated using the standard update rule:

$$x_{n+1} = x_n + \alpha_n s_n \,. \tag{10}$$

This algorithm does not explicitly consider the Wolfe Conditions for convergence (described in several papers, including Hager and Zhang, 2006); however, extensive testing showed that the search direction and stepsize found by our method satisfy these in effectively 100% of scenarios. As such, we do not include it in order to increase the efficiency of the inversion algorithm.

The description above summarized the core functions of our inversion method. Being that it is an iterative method, after the model is updated the inversion repeats itself, using the new model to update the incident angles and gradient, and in turn using this to find new search directions and stepsizes. After a final inverted model has been established, the final step is to undo the data normalization to revert the model parameters to their original data domain.

Inversion Results on Synthetic Data

We ran the inversion described in the previously section on synthetically generated seismic data and real seismic data for an unconventional field. The synthetic data was generated using Ricker wavelet in convolutional modeling a real well log as the model. Various amounts of random white noise were added to the synthetics to test our methods response with different SNR's. In our case study we ran the inversion for 25 iterations to ensure optimal results, although the reduction in error and the change in the inverted model are relatively minor after ~10 iterations. The inversion processed individual iterations at a rate of several thousand per hour on a single 4-core/8-thread CPU, verifying the computational feasibility of using our method for large datasets. Note that each CDP is evaluated independently and thus the inversion can easily be parallelized, making inversions of any size practical.

Figures 3 – 5 overview our inversion of the synthetic dataset. Figure 3 displays V_p, V_s, and ρ as well-log curves and compares the true model, the initial model, and the inverted model. The initial model is a frequency filtered version of the true well log with a 0-0-6-8 Hz frequency filter. Figure 4 shows the data match between the "true" synthetic data and the synthetic data predicted by the inversion. The synthetic data used in Figures 3 and 4 has a SNR of 5. Figure 5 shows how error progresses at every iteration for the inversion shown in Figures 3 and 4. We additionally test how the inversion performs under less ideal scenarios, specifically with noisy data and with poor starting models. Figure 6 shows how the inverted models changes as signal to noise ratio is reduced, as well as how the inversion results when using a linear starting model.

Discussion

Our proposed inversion method is capable of inverting for P-wave, S-wave and density information strictly from the use of P-P reflection data. The use for the full Knott-Zoeppritz equations in forward modeling allows for significantly better utilization of far offset data, which is the domain in which density and (to a lesser extent) S-wave differences have a more significant effect on seismic reflectivity. Standard AVO approximations are inaccurate in these regimes, and thus the estimates of parameters that rely on this information are inherently erred.

Overall, the inverted model found by our method is extremely accurate. Comparing our estimated model to the well log data, as shown in Figure 3, indicates that all major events and a many minor events are being recovered correctly. The predicted data sections in Figures 4 and 5 show that the data is being matched very well, even in the presence of extreme noise. Figure 4 shows synthetics for the extreme case of SNR=0.05, and Figure 5 shows that our method is capable of handling noisy data, and at a signal to noise ratio of 0.2 (i.e., 5x as much noise as data) our results were largely unaffected. Note: SNR is defined by the ratio of signal variance to noise variance. Figure 5 shows that our method is not very sensitive to the initial model except outside of recovering low frequency information. Lastly, Figure 6 shows that our inversion converges quickly and that there is minimal improvement after ~10 iterations. Combined, these factors indicate that our inversion method consistently and reliably produces accurate and believable estimates of subsurface elastic parameters.

Conclusions

We present a novel method for implementing nonlinear AVO waveform inversion. This gradient-based method uses the full Knott-Zoeppritz equations in forward modeling, significantly improving the accuracy of estimates for density and shear wave velocity. This inversion uses an automatically resetting conjugate gradient scheme to ensure quick convergence and implements an effective and efficient method for escaping local minima in the error function. Our method is extremely robust: it flawlessly handles noisy data without issue, and can use a poor starting model so long as it contains some low frequency information. Our approach is computationally efficient and practically feasible: tens of thousands of CDP's can be inverted in a matter of days on a single consumer-grade CPU. Furthermore, the independence between CDP's makes massive parallelization easy to implement for larger scale projects. Overall, we believe that this method provides an excellent improvement over standard AVO inversion techniques and offers an excellent balance between accuracy and efficiency.

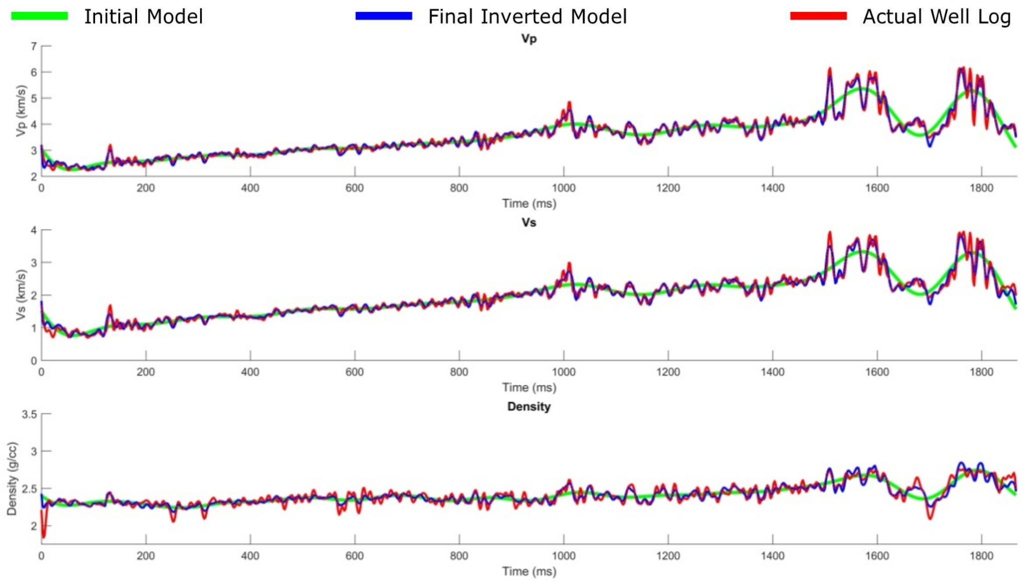

Figure 3: A comparrison between the inversion results and the well log data at the well location. The smooth initial models (Green), final inverted models (Blue), and actual well logs (Red) are shown for P-wave velocity (*Top*), S-wave velocity (*Middle*), and Bulk Density (*Bottom*). Our inversion method is able to recover the vast majority of the variability in the well log, and is a close match both in terms of phase and amplitude. X-axis represents travel time and Y-axis represents the parameter value.

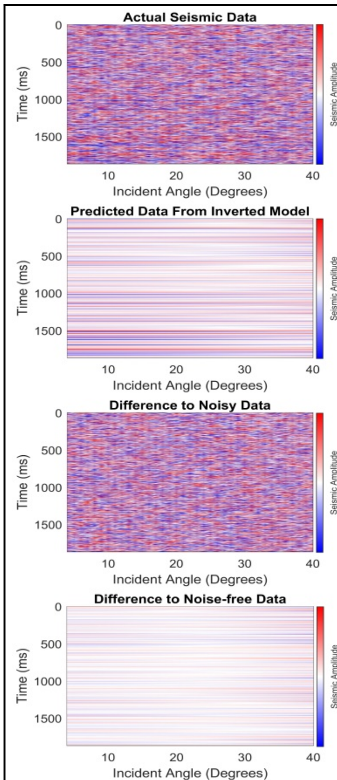

Figure 4: The prestack seismic data for the inversion shown in Figure 3. (*Top*) the actual seismic data, with noise added to make the overall signal to noise ratio equal to 0.05. (*Upper Middle*) the seismic data predicted by the inversion. (*Lower Middle*) the difference between predicted and noisy data, and (*Bottom)* the difference between the predicted and noise-free data.. The inversion reproduces the seismic data almost exactly, resulting in effectively no signal in the difference section. All plots use the same colorscale.

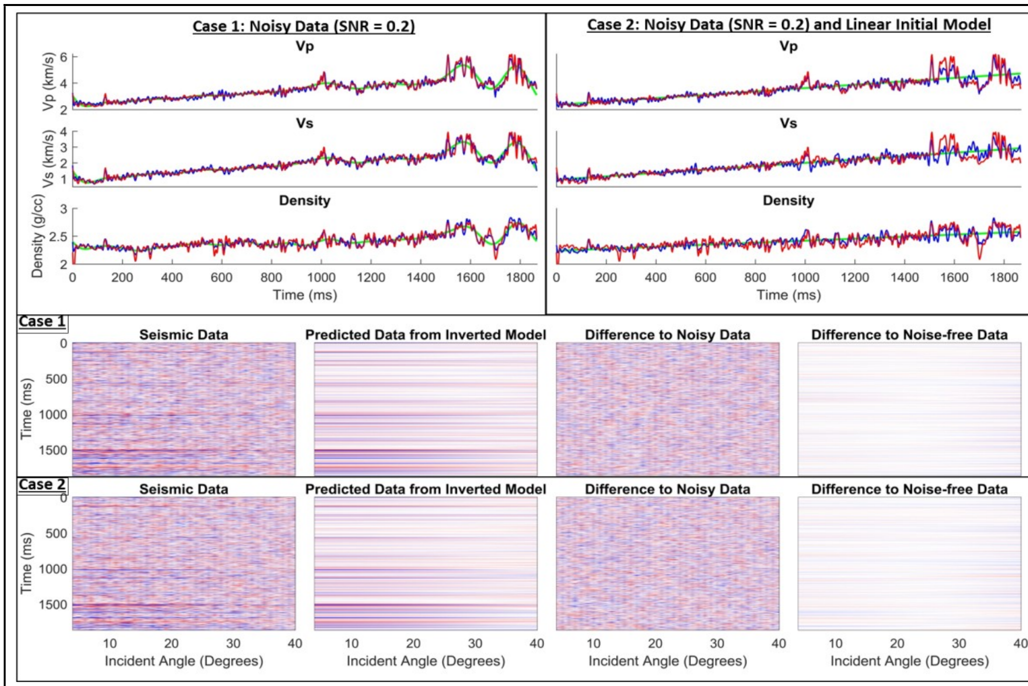

Figure 5: A demonstration of how our inversion method performs in two traditionally difficult scenarios: 1) on very noisy data (SNR = 0.2) and 2) on very noisy data and with a poor initial model (linear initial model). (*Upper Left*) Inverted Vp, Vs and density results for the noisy data. (*Upper Right*) Inverted Vp, Vs and density results for the noisy data using a linear initial model. Our results are not significantly effected by these difficulties, with the exception of missing low frequency info in the linear initial model scenario (e.g., 1500 – 1800 ms). (*Bottom*) A comparrison of the noisy seismic data, seismic data predicted from the inverted model, the difference between the noisy and predicted data, and the difference between the noise-free and predicted data. The Upper Row represents "Case 1" (noisy data), and the Lower Row represents "Case 2" (noisy data and linear initial model). All plots use the same colorscale. Ideally, the noise section difference should contain only noise and the noise-free difference section should be empty.

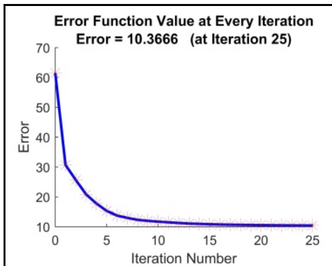

Figure 6: A plot showing how error evolved on every iteration. Error values for iterations 1 to 25 are shown. All other figures show the results from 25 iterations. X axis is iteration number and Y-axis is error value. There is minimal change in error after ~10 iterations.