GeoConvention 2015
GEOSCIENCE
new horizons
Telus Convention Centre • 4-8 May 2015 • Calgary, AB Canada

Total variation tomographic inversion via the Alternating Direction Method of Multipliers

Landon Safron and Mauricio D. Sacchi

Department of Physics, University of Alberta

Summary

Geophysical inverse problems are typically ill-posed and require *a priori* information to estimate a meaningful solution. If the unknown model is assumed to contain edges and sharp features, total variation (TV) regularization can be applied to retrieve a solution that present sharp boundaries. In this paper, we cast the seismic tomography inverse problem as a $\ell_2 - \ell_1$ least squares problem with total variation regularization. We make a substitution to change the problem into a form that can be solved using the alternating direction method of multipliers (ADMM). Testing the method with noisy synthetic VSP traveltime data results in an edge-preserving model that approximately matches the true model. Similar results are obtained when we apply the method to the inversion of the classical Dix equation.

Introduction

Seismic tomography is an imaging technique that can be used to determine the Earth's velocity-structure. First, the image is separated into sections that are assumed to be have the same material properties. Then, using measurements made of seismic waves passing through the subsurface, the properties of these sections are determined. Clearly, seismic tomography is an inverse problem. Like other geophysical inverse problems, the data used in seismic tomography is contaminated with unknown noise. As a result, infinitely many subsurface models satisfy the same set of observations and the inverse problem is said to be ill-posed.

A common approach to solving many geophysical inverse problems is to cast them as regularized least squares problems. One such example is Tikhonov regularization (Tikhonov and Arsenin, 1977), which is often used to choose the smoothest solution that satisfies the data. Smooth solutions are computationally inexpensive and can be appropriately used to solve many inverse problems, such as vertical seismic profiles (Lizarralde and Swift, 1999).

While smooth solutions can be justified in many settings (Constable et al., 1987), they are not always appropriate. This is the case when imaging geological interfaces with strong acoustic velocity contrasts (e.g. salt bodies). In these situations, we expect to see sharp boundaries and interfaces in the model. Therefore, a better choice of regularization is to look for a *blocky* solution, i.e., one that allows for edges and sharp features. This is precisely the goal of the total variation regularization proposed by Rudin et al. (1992).

In this paper, we apply the alternating direction method of multipliers to seismic tomography with total variation regularization. We test the method with noisy synthetic vertical seismic profile (VSP) and common mid-point (CMP) traveltime data.

Total variation regularization

Let d be a N x 1 vector of seismic data, L be a data kernel of size N x M, m be a subsurface model of size M x 1 and n be additive noise of size N x 1, where N and M are the number of observations and model parameters, respectively. Then the seismic data can be found from the forward problem

$$d = Lm + n. \quad (1)$$

In reality, we are more interested in the inverse problem (tomography), i.e., finding m of size M x 1 for a known d and L and an unknown n. Because the noise is unknown, we cannot uniquely solve for m. Regularization is therefore necessary to stabilize the problem. Casting the inverse problem as a least squares problem gives us

$$\min_m R(m) + \mu||Lm - d||_2^2, \quad (2)$$

where $||.||_2$ is the ℓ_2 norm, $R(m)$ is a regularization function and μ is a tradeoff parameter. In the case with total variation regularization, we set

$$R(m) = \sum_{i=1}^{N} ||D_i m||_1^1, \quad (3)$$

where $||.||_1$ is the ℓ_1 norm, D is a N x N first difference operator and D_i is the i^{th} row of D. The solution to the inverse problem is therefore

$$\min_m \sum_{i=1}^{n} ||D_i m||_1^1 + \frac{\mu}{2}||Lm - d||_2^2. \quad (4)$$

Setting $y_i = D_i m$ transforms (1.4) into the constrained problem

$$\min_{m,y} \{\sum_{i=1}^{n} ||y_i||_1^1 + \frac{\mu}{2}||Lm - d||_2^2 : y_i = D_i m\}. \quad (5)$$

The inverse problem is now in a form that can be solved iteratively using the alternating direction method of multipliers.

Alternating direction method of multipliers

The alternating direction method of multipliers was first proposed by Glowinski and Marrocco (1975) and Gabay and Mercier (1976). The method provides a fast iterative scheme for solving optimization problems of the form

$$\min_{x,z} \{P(x) + Q(z) : Ax + Bz = c\}, \quad (6)$$

where P and Q are convex functions. The alternating direction method of multipliers proposes to solve (6) by looking for saddle points of the augmented Lagrangian function

$$\mathscr{L}_{\mathscr{A}}(x,z,\lambda) \equiv P(x) + Q(z) + \lambda^T(Ax + Bz - c) + \frac{\beta}{2}||Ax + Bz - c||_2^2, \quad (7)$$

where λ is an estimate of the Lagrange multiplier, β is the dual update length and the superscript T denotes the transpose. This leads to the iterative scheme

$$\begin{cases} x^{k+1} \leftarrow arg\ \min_x \mathscr{L}_{\mathscr{A}}(x, z^k, \lambda^k), \\ z^{k+1} \leftarrow arg\ \min_z \mathscr{L}_{\mathscr{A}}(x^{k+1}, z, \lambda^k), \\ \lambda^{k+1} \leftarrow \lambda^k - \beta(x^{k+1} - Dz^{k+1}), \end{cases}$$

where k denotes the iteration number.

Applying ADMM to the TV regularized seismic tomography problem (7) results in the augmented Lagrangian function

$$\mathscr{L}_{\mathscr{A}}(m,y,\lambda) \equiv \sum_{i=1}^{N} \left(||y_i||_1^1 - \lambda_i^T (y_i - D_i m) + \frac{\beta}{2} ||y_i - D_i m||_2^2 \right) + \frac{\mu}{2} ||Lm - d||_2^2. \quad (8)$$

Starting with $x = x^k$ and $\lambda = \lambda^k$, we iterate as follows:

$$\begin{cases} y^{k+1} \leftarrow arg\ \min_y \mathscr{L}_{\mathscr{A}}(m^k, y, \lambda^k), \\ m^{k+1} \leftarrow arg\ \min_m \mathscr{L}_{\mathscr{A}}(m, y^{k+1}, \lambda^k), \\ \lambda^{k+1} \leftarrow \lambda^k - \beta(y^{k+1} - Dm^{k+1}). \end{cases}$$

In Tao and Yang (2009), the authors recast the minimizations as closed form solutions:

$$y_i^{k+1} = max\left\{ \left| D_i m^k + \frac{1}{\beta}(\lambda^k)_i \right| - \frac{1}{\beta}, 0 \right\} \circ sgn\left(D_i m^k + \frac{1}{\beta}(\lambda^k)_i \right), \quad (9)$$

$$m^{k+1} = \left(D^T (y^{k+1} - \frac{1}{\beta}\lambda^k) + \frac{\mu}{\beta} L^T d \right) (D^T D + \frac{\mu}{\beta} L^T L)^{-1}, \quad (10)$$

where "$\circ$" and "sgn" are the piece-wise product and and signum function, respectively. After updating y and m, λ is updated as

$$\lambda^{k+1} \leftarrow \lambda^k - \beta(y^{k+1} - Dm^{k+1}). \quad (11)$$

All that remains is to define a convergence criteria. The authors of Tao and Yang (2009) use the relative change in the model as a test for convergence, i.e.,

$$\frac{||m^{k+1} - m^k||_2^2}{max\{||m^k||_2^2, 1\}} < \varepsilon, \quad (12)$$

where ε is the point of convergence. To summarize, the alternating direction method of multipliers is given by the following algorithm:

Algorithm 1 ADMM

Input: d, L, μ, β and λ_0. Initialize m = d and $\lambda = \lambda_0$.

While *not converged*, **Do**
 (1) Update y^{k+1} with (9)
 (2) Update m^{k+1} with (10)
 (3) Update λ^{k+1} with (11)
End While

Model selection

Up to this point, we have established that ADMM can be used to solve TV regularized tomography problems for a given μ and β. The problem then is to find the value of μ and β that produce the model that best fits the observed data. In Constable et al. (1987), the authors choose the best model according to the weighted least squares criterion. The idea comes from the assumption that the data are independent measurements of random variables, meaning the residual error should follow a Gaussian distribution. Consequently, if we define χ^2 as

$$\chi^2 = \frac{||Lm - d||_2^2}{\sigma^2}, \quad (13)$$

where σ is the standard error of the noise, then the best solution is the one that results in a value of χ^2 closest to N, the number of observations. In other words, E[χ^2]=N.

In Tao and Yang (2009), the authors find that the quality of the inversion has little relevance to the value of β. It is therefore reasonable to choose β based entirely on the speed of convergence of the ADMM algorithm. As a result, the problem of choosing the best model is reduced to choosing the best value of the tradeoff parameter, μ. Therefore, a good approach is to iteratively increase (or decrease) μ until χ^2 matches N with an acceptable level of misfit.

We set the point of convergence for the ADMM algorithm to be at $\varepsilon = 10^{-11}$. Increasing the precision beyond this point results in no appreciable increase in the accuracy of the inverse solution. Of course, the selection of ε largely depends on the computational power available and the goal of the inversion.

Example: VSP traveltime inversion

We demonstrate the effectiveness of ADMM for solving the TV regularized inverse problem by using noisy synthetic VSP traveltime data. We begin by defining a 1-dimensional velocity model that comprises 6 horizontal layers. The theoretical survey uses 500 equally spaced receivers lowered into a borehole and a source placed at the Earths surface at 0m offset. Direct-arrival traveltime data is then computed for each receiver. Finally, we contaminate the traveltime data with Gaussian noise with standard error $\sigma = 1ms$.

To set the problem up according to equation (1), we define d as the noisy direct-arrival traveltimes, L as the causal integration operator multiplied by the receiver spacing and m as the unknown acoustic reciprocal velocity model. The inverse problem is clearly ill-posed. We choose to implement total variation regularization to select the best edge-preserving model. Trial and error reveals that setting $\beta = 4500$ allows the ADMM algorithm to converge to a blocky solution in a relatively minimal number of iterations.

After approximately 1000 iterations, the ADMM algorithm converged to the inverse solution shown in Figure 1. Unlike the smooth inversion, the TV regularized inverse model predicts most of the interfaces and interval velocities that are present in the true model. The figure also highlights the fact that the noise level limits the resolution of the model. This is best observed at shallow depths, where the true traveltime signal is small and may be overwhelmed by the random noise. TV regularization can interpret this as having additional non-real layers.

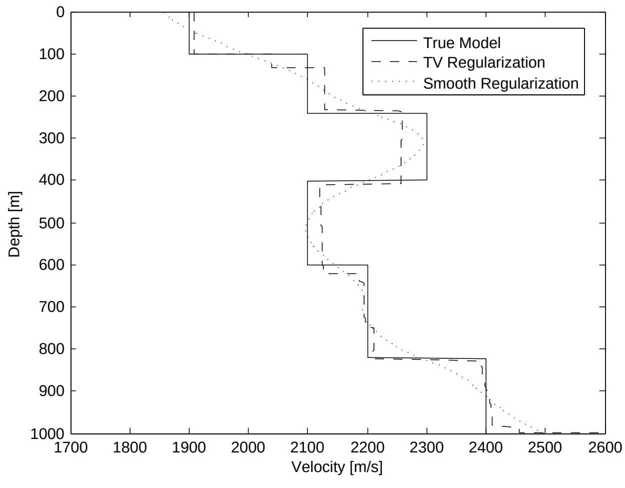
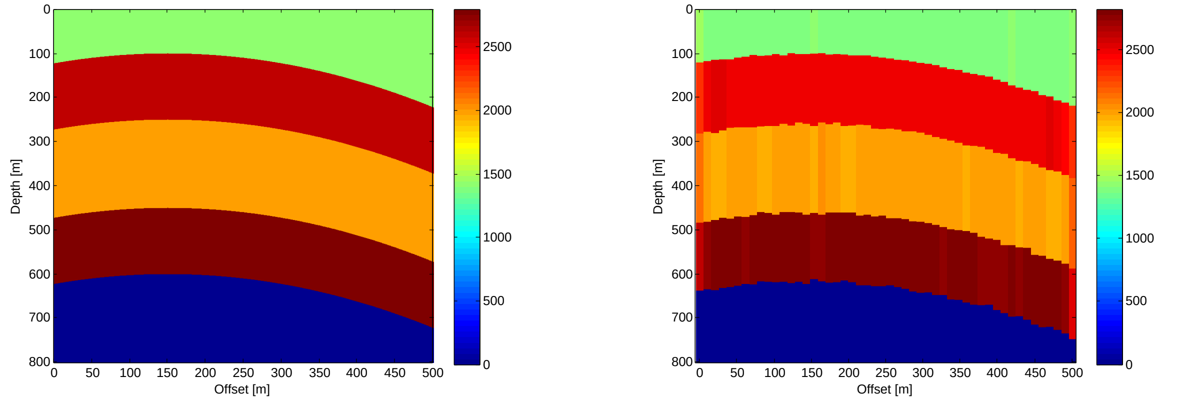


Figure 1: Inverted vertical seismic profile using TV regularization and smooth regularization.



(a) (b)

Figure 2: Dix inversion example. (a) True subsurface model. The deepest layer was assigned an acoustic velocity of zero as a reminder that the properties of this layer cannot be recovered. (b) Inverse model found using TV regularization. The model consists of layers that are nearly uniform.

Example: Dix equation

The classical Dix equation relates root-mean-square (RMS) velocities to interval velocities for a horizontally layered section. The relationship is given by

$$V_{RMS}^2 = \frac{\sum_{n=1}^{p} V_n^2 \Delta t_{0,n}}{t_{0,n}}, \tag{14}$$

where p is the number of layers to the n^{th} interface, V_n is the acoustic velocity of the n^{th} layer, $\Delta t_{0,n}$ is the zero-offset two-way traveltime through the n^{th} layer and $t_{0,n}$ is the zero-offset two-way traveltime for the n^{th} reflection. We can rewrite the equation in the form shown in (1) by setting $d = t_{0,n}V_{RMS}^2$, L as the causal integration operator and $m = V_n^2 \Delta t_{0,n}$.

Dix equation can be used to solve for a 1-dimensional subsurface model based on data from 1 CMP. However, we propose to use Dix equation to image a 2-dimensional subsurface comprised of 5 folded layers. 1 CMP will not be sufficient to resolve the folded strata. Instead, we propose to concatenate the results of 50 equally spaced CMPs. We solve each CMP according to Dix equation, subject to a lateral edge-preserving condition as enforced by total variation regularization.

We begin by computing synthetic data. We calculate true RMS velocities and then contaminate them with random noise with a standard error of 5m/s. Similarly, the zero-offset two-way traveltimes for the n^{th} reflection, $t_{0,n}$, are computed by contaminating the true $t_{0,n}$'s with Gaussian noise with a standard error of 2ms.

Infinitely many models can satisfy the CMP traveltime data. We choose total variation regularization to look for the edge-preserving model that best fits the data. Trial and error reveals that setting $\beta = 80000$ allows the ADMM algorithm to converge to a blocky solution in a relatively minimal number of iterations. Figure 2 shows the results after approximately 700 iterations. The shape and acoustic velocities of the folded layers approximately resemble the true model. The figure also shows that TV regularization was successful at recovering an edge preserving model.

Conclusion

In this paper, we have discussed the application of the alternating direction method of multipliers to seismic tomography with total variation regularization. This ultimately required recasting the TV regularized inverse problem as a constrained $\ell_2 - \ell_1$ least squares problem, which could then be solved iteratively. Experimental results for VSP and CMP traveltime data confirm that the method converges to an accurate solution. The speed of convergence for the method could likely be improved with the addition of an automated selection of the dual update parameter β.

Acknowledgements

We thank the sponsors of the Signal Analysis and Imaging Group (SAIG) for supporting our research program. Landon Safron was also supported by NSERC via the Undergraduate Student Research Awards (USRA) program.

References

Constable, S. C., R. L. Parker, and C. G. Constable, 1987, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, **52**, 289–300.

Gabay, D. and B. Mercier, 1976, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations: Comput. Math. Appl., **2**, 17–40.

Glowinski, R. and A. Marrocco, 1975, Sur lapproximation par elements finis dordre un, et la resolution par penalisation-dualite dune classe de problemes de dirichlet nonlineaires: Rev. Francaise dAut. Inf. Rech. Oper., **2**, 41–76.

Lizarralde, D. and S. Swift, 1999, Smooth inversion of VSP traveltime data: Geophysics, **64**, 659–661.

Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Physica D, **60**, 259–268.

Tao, M. and J. Yang, 2009, Alternating direction algorithms for total variation deconvolution in image reconstruction: TR0918 Department of Mathematics, Nanjing University.

Tikhonov, A. and V. Arsenin, 1977, Solutions of ill-posed problems: Winston, Washington.