Regularization algorithms for seismic inverse problems

Hongbo Zhou, Yu Zhang, Samuel H. Gray, Veritas DGC Inc.
Guanquan Zhang, Chinese Academy of Sciences

Summary

Seismic inverse problems are known to be ill-posed and ill-conditioned because of noisy data and inadequate model parameterizations. To obtain a unique and stable solution, both data and model have to be regularized. How do we do this? Given the wealth of data and model regularization techniques available, are there criteria for choosing an adequate regularization algorithm for a particular problem? In this abstract, we begin to answer this question. First, we discuss the properties of some conventionally used regularization functionals, and we compare them systematically. Then we discuss specific linear/nonlinear functionals, which can be used for certain special purposes. Finally, we show an example, applying these techniques to migration velocity analysis.

Introduction

Solving an inverse problem means making inferences about physical models from the observed data. For seismic inverse problems, the observed data might be reflection traveltimes or residual curvature on depth-migrated gathers (Bube and Langan, 1997; Zhou et al., 2001), and the physical model to be inferred is typically the velocity structure within the Earth. A common way to estimate a model is to seek the model that gives the best fit to the data in the sense that the data residual, measured by some metric (usually the L_2 norm), is made as small as possible.

In real problems, the final estimated model usually differs significantly from the idealized model. One cause of this difference is noisy data, since the observations (picked reflection traveltimes or moveout on gathers) are usually inconsistent, and are contaminated with errors that have some effect on the estimation. These solutions need *data regularization* to reduce the effects of noisy data. Another cause may be the mathematical description of the model. Generally, most seismic inverse problems are ill-posed (under-determined, with a large null space) and ill-conditioned or unstable, so that small errors in the data may lead to large variations in the model. So there may exist many models, usually an infinite number of them, that fit the same observed data equally well. Usually, the size and shape of the model null space depends partly on the choice of model space parameterization. Therefore, besides data regularization, one should also apply a *model regularization* or preconditioning in order to reduce the size of the null space and stabilize the algorithm. This step restricts the set of admissible solutions, and provides *a priori* information (e.g., model smoothness).

In general, given a forward model, i.e., a relationship between the model u and observed data b

$$Au = b,$$

the inverse problem can be formulated as a variational problem to minimize the misfit functional

$$J(u) = F(r(u)) + \epsilon R(u). \tag{1}$$

Here $F(r)$ is called the data term, usually an even, non-negative function, $r(u) = Au - b$ is the data residual, and $R(u)$ is called the regularization term. The goal of Eq. (1) is to estimate the model u from the data residual $r(u)$. The roles of $R(u)$ are to incorporate *a priori* knowledge of the model u and to regularize the solution. In principle, $\epsilon R(u)$ should be as small as possible, allowing the solution method to attack the data residuals, but in practice the ill-posedness of the unconstrained inverse problem often forces this term to be larger than we wish it to be. If $\epsilon R(u)$ is much greater than $F(r)$, our solution is as regular as we've constrained it to be, and it incorporates prior knowledge, but it ignores the data residuals.

Mathematically it is obvious (but geophysically we tend to ignore) that, by choosing particular forms for F, R, and a particular value for ϵ, we have already begun to constrain the solution of equation (1). The constraints can be either mathematical (continuity, smoothness, etc.) or geophysical (geological interfaces, maximal and minimal velocity values, etc.). Therefore, our choice of objective function should depend on our *a priori* knowledge of the solution and on our assumptions about its properties, e.g., its smoothness. A high-fidelity inversion result can be obtained only with good choices for the functionals $F(r)$ and $R(u)$.

In this abstract, we address the problem of appropriate choices for $F(r)$ and $R(u)$ by discussing the properties of some conventionally used regularization functionals and comparing them systematically. We also introduce some specific linear/nonlinear functionals which can be used for certain special purposes. For example, some functionals can be used to detect the sharp model boundaries from the data automatically, and others can be used to preserve model resolution near geological interfaces without compromising their ability to suppress random data noise.

Data regularization

Optimization problems are often formulated as least-squares problems, where the L_2 norm of the data misfit needs to be minimized. However, least-squares solutions tend to be very sensitive to outliers, i.e., individual data points that lie far away from the bulk of the data (Claerbout, 2001). Therefore, to reduce the influence of outliers, we must sometimes use a norm different from L_2. From the enormous selection of choices for the function $F(r)$, we discuss three of the most popular ones.

- L_p norm: $F(r) = \frac{1}{p}|r|^p$, where $1 \leq p \leq 2$. When $p = 2$, we have a least-squares problem; the solution tends to be very sensitive to spiky data residuals. The L_1 norm ($p = 1$) overcomes this problem by overweighting data terms with small residuals and underweighting the large spiky data residuals. Other p-values between 1 and 2 perform between L_1 and L_2. In numerical calculations, solving optimized L_p norm problems often requires the use of IRLS (Iteratively Reweighted Least Squares) algorithm (Bube and Langan, 1997; Claerbout, 2001) in which extremely small residuals are replaced by a threshold value. This causes some artificial errors in the minimization, and a loss of resolution in the solution.

- Hybrid L_1/L_2: $F(r) = \left(1 + r^2/r_0^2\right)^{p/2} - 1$, where $1 \leq p \leq 2$. When r/r_0 is small, $F(r) \approx \frac{p}{2}|r/r_0|^2$, and when r/r_0 is large, $F(r) \approx |r/r_0|^p$. Therefore, $F(r)$ behaves like the L_2 norm for small residuals, and like the L_p norm for large residuals (Bube and Langan, 1997). This hybrid norm underweights large residual errors, but it treats small values as L_2 does, simply by ignoring their influence. Therefore, unlike the L_1 norm, this function $F(r)$ does not emphasize data terms with small residuals.

- Cauchy distribution: $F(r) = \ln(1 + r^2/r_0^2)$. Similar to the L_1 norm, this function is not guaranteed to produce unique solutions. However, it underweights larger data residuals more severely than the hybrid L_1/L_2 norm does.

Model regularization

The data regularization techniques mentioned above result from minimizing different misfit functionals of data residual. However, because of the intrinsic under-determined nature of seismic inverse problems, the inverse matrices might not exist or, if they exist, they might not have any meaning. Even when the inverse matrices exist, they can be ill-conditioned (become nearly singular and extremely sensitive to data errors). When this happens, the solution becomes extremely unstable. To overcome this difficulty, the models also need to be regularized by introducing functional R.

Linear isotropic regularization Linear isotropic regularization uses the L_2 norm to penalize the roughness in the solution and attempts to emphasize the large scale features of the model. Also, according to the variational calculus, such regularization functions often lead to linear differential equations for the inverted model u which are relatively easy to solve.

- H^0 norm:
$$R(u) = \|u\|_{L_2}^2 = \left(\int |u(\vec{x})|^2 d\vec{x}\right)^{1/2}. \qquad (2)$$
This is the Levenberg-Marquardt damping operator. It provides stable, but oscillatory, solutions for a given model parameterization. One serious disadvantage of this model regularization norm is that the solution depends heavily on the model parameterization.

- H^1 semi-norm:
$$R(u) = ||\nabla u||_{L_2}^2, \qquad (3)$$
where ∇ represents the gradient operator. Minimizing the objective function with this regularization term leads to the solving of a elliptic equation
$$-\triangle u = 0,$$
where $\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplace operator.

- H^2 semi-norm:
$$R(u) = ||\triangle u||_{L_2}^2. \qquad (4)$$
Minimizing this functional is equivalent to solving a bi-harmonic equation (Smith and Wessel, 1990)
$$\triangle^2 u = 0.$$

- $H^1 + H^2$ semi-norm: The physical interpretation of this regularizer is splines in tension, with an energy functional $R(u)$ defined as
$$R(u) = (1-t)||\triangle u||^2 + t||\nabla u||^2, \qquad (5)$$
where $0 \leq t \leq 1$ is a tension parameter. The corresponding partial differential equation is
$$(1-t)\triangle^2 u - t\triangle u = 0. \qquad (6)$$
When $t = 0$, this equation becomes bi-harmonic, which produces a smooth solution, but creates artificial oscillations in the unconstrained regions around sharp changes in the gradient; when $t = 1$, it becomes Laplace's equation, whose solution is a linear function with no extraneous inflections in the solution, but with discontinuous derivatives. Intermediate values of t allow us to achieve a compromise: a smooth solution surface with constrained oscillations (Smith and Wessel, 1990).

Nonlinear isotropic regularization The main problem using quadratic regularization functionals in the preceding section is their inability to "respect" the discontinuities of the model. A strategy for overcoming this problem is to use the gradient as an edge detector, and encourage intra-region smoothing over inter-region smoothing. Then, locations where the gradient is large will have a large likelihood of being an edge (Perona and Malik, 1990).

- L^p norm, $1 \le p < 2$: When $p = 1$, the variational problem becomes the total variation (TV) norm, which is proposed as a regularization functionals for the image restoration problem (Rudin et al., 1992):

$$R(u) = TV(u) = \int \sqrt{u_x^2 + u_y^2 + u_z^2}\, dx. \quad (7)$$

Its corresponding differential equation, with homogeneous Neumann boundary conditions for u, is

$$0 = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right). \quad (8)$$

While the H^0 semi-norm honors larger values of the solution, the H^1 semi-norm honors linear functions over others. However, the TV norm allows more solution functions, including discontinuous ones, to approximate a given function.

- Cauchy distribution: Minimizing the energy functional

$$R(u) = \int \ln(1 + |\nabla u|^2) dx, \quad (9)$$

results in its corresponding differential equation

$$0 = -\nabla \cdot \left(\frac{1}{1 + |\nabla u|^2} \nabla u \right). \quad (10)$$

This corresponds to the very famous Perona-Malik (1990) nonlinear diffusion filter. One problem with this functional is that $R(u)$ will not have a well-posed solution unless regularized (Alvarez, et al., 1992).

- Hybrid L_1/L_2: Minimizing the energy functional

$$R(u) = \int \sqrt{1 + |\nabla u|^2}\, dx. \quad (11)$$

gives a solution which satisfies

$$0 = -\nabla \cdot \left(\frac{1}{\sqrt{1 + |\nabla u|^2}} \nabla u \right). \quad (12)$$

This is similar to the Cauchy distribution, and it provides a unique solution.

Linear anisotropic regularization In migration velocity analysis, sometimes we have information about the subsurface dips. When this happens, it is desirable to let our solution (the estimated velocity function) follow the given structures. Suppose the normalized normal direction $\vec{n} = (n_x, n_y, n_z)$ of a reflecting surface is given. Then we can project any vector into the plane orthogonal to $\vec{n}$ by using the projection operator $P_{\vec{n}}$, defined as

$$P_{\vec{n}} \vec{v} = (I - \vec{n} \otimes \vec{n})\, \vec{v} = \vec{v} - (\vec{n} \cdot \vec{v})\vec{n}. \quad (13)$$

To make sure that diffusion takes place only in the plane orthogonal to $\vec{n}$, we minimize the energy functional

$$J(u) = \frac{1}{2} \int |P_{\vec{n}} \nabla u|^2 dx. \quad (14)$$

Its solution satisfies

$$-\nabla \cdot (P_{\vec{n}} \nabla u) = 0, \quad (15)$$

and it can be solved by the following parabolic differential equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (P_{\vec{n}} . \nabla u)\,. \quad (16)$$

Expanding the RHS of this equation produces

$$\begin{aligned} & \nabla \cdot (P_{\vec{n}} \nabla u) \\ = \; & (1 - n_x^2) \frac{\partial^2 u}{\partial x^2} + (1 - n_y^2) \frac{\partial^2 u}{\partial y^2} + (1 - n_z^2) \frac{\partial^2 u}{\partial z^2} \\ & -2 n_x n_y \frac{\partial^2 u}{\partial x \partial y} - 2 n_x n_y \frac{\partial^2 u}{\partial x \partial z} - 2 n_y n_z \frac{\partial^2 u}{\partial y \partial z}. \end{aligned} \quad (17)$$

This is exactly the same result, derived differently, as that of Schwab (1997) who used it to detect plane reflectors.

Numerical tests

In this section, we perform The residual curvature analysis using the technique discussed above to a two-layer model with a horizontal reflector. The model is 100 × 100m wide and 50 × 100m deep and is divided into boxes of a width and height of 100m × 100m. There are 100 × 50 velocity parameters. The true velocity for the first layer is 1000 m/s, and the migration velocity is chosen as 900 m/s. The depth residuals are computed analytically over an offset range of 0m to 3000m. The common image point locations are picked at every other grid point, i.e., the input data grid is coarser than the solution grid.

Figure 1 (top) shows the velocity perturbation from migration velocity analysis without applying any regularization. The solution oscillates because the density of the analysis locations (every second grid location) is too coarse for the inversion. Figure 1 (middle) shows the result of linear isotropic smoothing using H^1 semi-norm (3), a significant improvement over the solution without regularization, but still showing a certain amount of smearing. This smearing is greatly reduced in Figure 1 (bottom), the result from hybrid L_1/L_2 nonlinear isotropic regularization, which shows the sharp boundary at the right reflector location.

Conclusion

Seismic inverse problems are inherently non-unique, and *a priori* information is often needed to reduce the ambiguity in the solution. With so many choices of data and model regularization algorithms, a natural question to ask is which one we should choose. In general, there is no single correct answer to this question. But if we understand the physics behind a particular problem, finding a solution becomes relatively easy. As a general rule, an appropriate choice of regularization depends on our *a priori* knowedge of the solution and an understanding of the logic behind the various candidate regularization algorithms. The solution - the inverted model - will depend critically on our choice of regularization. For example, in velocity analysis, if we know in advance that velocity is piecewise constant, we will use the TV (total variation) norm. If we know the velocity is a bi-cubic spline, we will use H^2 norm. If we know in advance that the solution is spiky, then the L_1 norm, the hybrid L_1/L_2 norm, or the Cauchy distribution can be applied to the model. Of course, the seismic velocity field within the Earth is usually none of these. So our choice of norm is a compromise that attempts to balance the range of possible velocity distributions with the prospect of obtaining a plausible, well-behaved solution in an efficient manner.

Acknowledgement

Part of the fourth author's research was supported by the Major Basic Research Program and National Natural Science Foundation of China.

References

Alvarez, L., Lions, P-L., and Morel, J-M., 1992, Image selective smoothing and edge detection by nonlinear diffusion, II: SIAM J. Numer. Anal., V. 29, 845–866.

Bube, K. P. and Langan, R.T., 1997, Hybrid ℓ^1/ℓ^2 minimization with applications to tomography: Geophysics, v.62, 1183-1195.

Claerbout, J.F., 2001, Geophysical estimation by example: Environmental soundings image construction, Multidimensional autoregression: http://sepwww.stanford.edu/sep/prof/, 2001.

Perona, P. and Malik, J., 1990, Scale-space and edge detection Using Anisotropic diffusion: IEEE Transaction on Pattern Analysis and Machine Intelligence, v. 12, 629-639.

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

Schwab, M., 1997, Cross product operator detects plane reflectors: SEP-94, 181–189.

Smith, W. H. and Wessel, P., 1990, Gridding with continuous curvature splines in tension: Geophysics, v. 55, 293-305.

Zhou, H., Guo, J. and Young, J., 2001, An alternative residual-curvature updating method for prestack depth migration: 71st Ann. Internat. Mtg., SEG, 877–880.

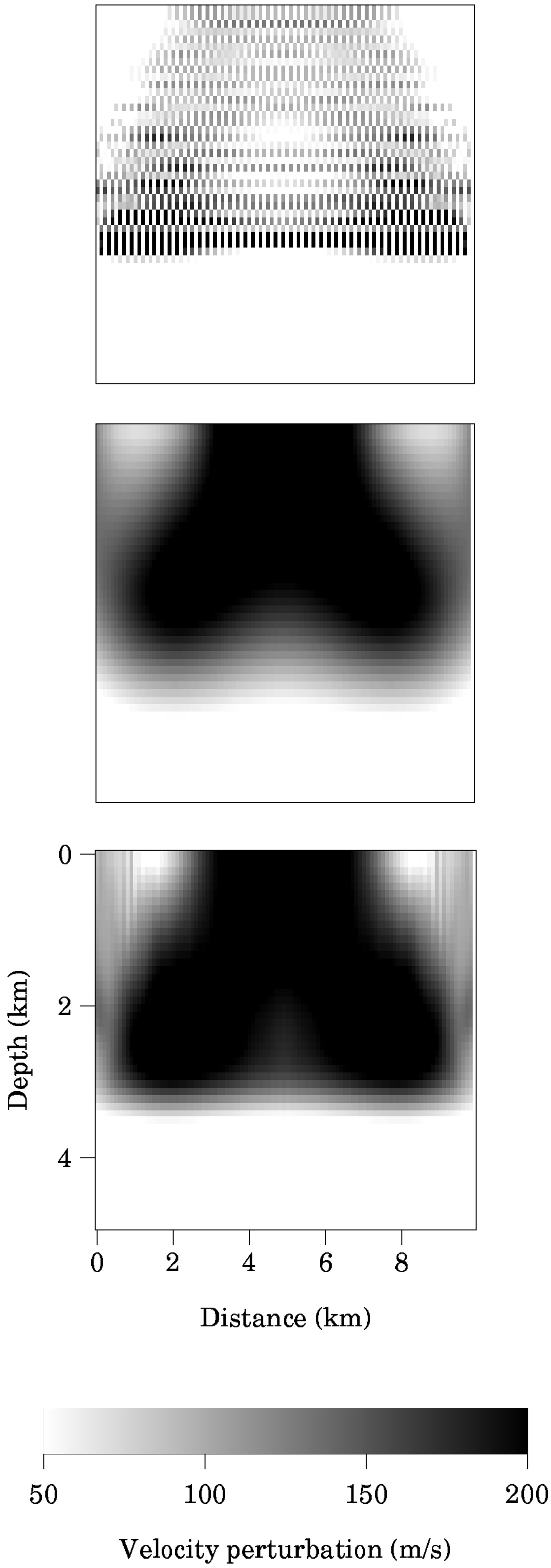

Fig. 1: Top figure shows perturbation of the velocity without any smoothing; the middle figure is the result with linear isotropic smoothing; and the bottom figure is the result of nonlinear isotropic smoothing