W031

Regularized Full Waveform Inversion Including Prior Model Information

A. Asnaashari* (Joseph Fourier University), R. Brossier (Joseph Fourier University), S. Garambois (Joseph Fourier University), F. Audebert (Total E&P), P. Thore (Total E&P) & J. Virieux (Joseph Fourier University)

SUMMARY

In this study, we propose a regularized time-domain Full Waveform Inversion (FWI) scheme based on a two-term model penalty in the misfit definition: the Tikhonov term to ensure smoothness and a prior model term to attract the inversion toward a given direction. We illustrate that the prior model could reduce the sensitivity of inversion with respect to a non-accurate initial model. This procedure delivers a robust model as compared to "classical" FWI and can help inversion to solve the cycle-skipping problem. In the Marmousi dataset application, an investigation was done to identify the respective influences of initial model and prior model in complex media. Finally, we propose a simple dynamic approach based on a first derivative of the objective function curve to take out gradually the effect of the prior model in the inversion scheme, once the cycle-skipping problem has been solved.

Introduction

Preconditioning or regularization techniques may alleviate the non-uniqueness of ill-posed inverse problems. Tikhonov and Arsenin (1977) have proposed a regularization strategy inside the optimization step for finding the smoothest model that explains the data. Preconditioning techniques acting as a smooth operator on the model update (Operto et al., 2006) may add strong prior features of the expected structure through directive Laplacian smoothing, for example (Guitton et al., 2010). Regularization schemes that preserve edges and contrasts have also been developed for specific Full Waveform Inversion (FWI) application through ℓ_1 model penalty (Guitton, 2011). All these regularization techniques allow the inversion scheme to be stabilized by assuming a particular representation or structure of the velocity model (smoothness, sparsity etc.). However, no prior on model values is generally used in most of full waveform inversion implementations as no information is usually available for exploration applications. For monitoring purposes, where many different data types have already been collected, such as sonic logs, well data or stratigraphic recordings, one may want to use such available prior information in the FWI scheme, as can be done for other velocity model building techniques. One major challenge for monitoring remains the necessity to derive a robust high-resolution baseline model, a key for time-lapse parameter variations (Asnaashari et al., 2011). Thus, by taking into account the available prior information in the FWI procedure, a robust baseline model reconstruction could be achieved.
In this study, we propose a regularized time-domain FWI scheme based on a two-term model penalty in the misfit definition: the Tikhonov term to ensure smoothness and a generalized Tikhonov term to attract the inversion towards a prior model. First, we present the theoretical framework of our study. Then, through a Marmousi II synthetic application, we show critical effects of the prior model penalty term on FWI results. Moreover, the fundamental different influence of the prior model and the starting model in the FWI procedure is highlighted.

Theory

The full waveform inversion relies on an iterative local optimization problem that is generally introduced as a linearized least-squares problem. The optimization attempts to minimize the residuals between the observed and the modeled wavefields at receivers. The linearized inverse problem remains ill-posed and, therefore, requires extra information generally introduced through regularization. For time-lapse applications or other specific applications, where other information such as sonic logs, stratigraphic data or geological constraints are available, it is crucial to take these into account in the inversion procedure to ensure robust and consistent results.
Our misfit function is based on a classical Tikhonov regularization function (Tikhonov and Arsenin, 1977), plus a second penalty term for prior model misfit that is often referred as generalized Tikhonov regularization. This last term estimates residuals between the current model at a given iteration and the prior model we consider at the same iteration. The misfit function $\mathscr{C}(\mathbf{m})$ can be written as

$$\mathscr{C}(\mathbf{m}) = \mathscr{C}_d(\mathbf{m}) + \lambda_1 \mathscr{C}_{1_m}(\mathbf{m}) + \lambda_2 \mathscr{C}_{2_m}(\mathbf{m}), \tag{1}$$

where $\mathscr{C}_d(\mathbf{m})$ is the data misfit term, $\mathscr{C}_{1_m}(\mathbf{m})$ is the Tikhonov term and $\mathscr{C}_{2_m}(\mathbf{m})$ is the prior model misfit term. Two regularization hyper-parameters λ_1 and λ_2 are introduced and allow each penalty term to be weighted with respect to each other and to the data term. The misfit function can be explicited using the ℓ_2 norm as

$$\mathscr{C}(\mathbf{m}) = \frac{1}{2}\Big\{ (\mathbf{d}_{obs} - \mathbf{d}(\mathbf{m}))^T \mathbf{W}_d^T \mathbf{W}_d (\mathbf{d}_{obs} - \mathbf{d}(\mathbf{m})) \Big\} + \frac{\lambda_1}{2}\{\mathbf{m}^T \mathbf{B}^T \mathbf{B}\mathbf{m}\} + \frac{\lambda_2}{2}\Big\{ (\mathbf{m} - \mathbf{m}_p)^T \mathbf{W}_m^T \mathbf{W}_m (\mathbf{m} - \mathbf{m}_p) \Big\}, \tag{2}$$

where vectors $\mathbf{d}_{obs}$ and $\mathbf{d}(\mathbf{m})$ are the observed and computed data, respectively. The matrix $\mathbf{W}_d$ is a weighting operator on the data misfit, that is often considered to be diagonal, as generally no correlation is present between measurements. We can even consider a uniform measurement quality on all receivers, leading to a scaled identity matrix as $\mathbf{W}_d = \sigma_d I$. The synthetic data depend non-linearly on model parameters denoted as $\mathbf{m} = \{m_i\}_{i=1,N_m}$, where the number of unknowns is denoted by N_m: these model parameters should be determined through the inverse procedure. The second term of the objective function corresponds to the Tikhonov regularization where we minimize the first spatial derivatives of the model through the first-order derivative

operator matrix $\mathbf{B}$. The third term is related to the prior model $\mathbf{m}_p$ designed from different information and that could be set prior to the inversion. This prior could also be adapted iteratively during the inversion procedure. The matrix $\mathbf{W}_m$ is a weighting operator on the model space that we consider to be diagonal. It includes prior uncertainty of the model parameters. The influence of $\mathbf{W}_m$ plays a critical role in driving the inversion procedure towards a given minimum. Please note that the misfit function mixing data and model quantities is built dimensionless, thanks to the matrices and operators $\mathbf{W}_d$, $\mathbf{W}_m$ and $\mathbf{B}$.
Minimizing the misfit function leads to the normal equation system which can be written as

$$\mathscr{H}_m \Delta\mathbf{m} = -\mathscr{G}_m, \tag{3}$$

where the gradient and the Hessian of the misfit function are denoted respectively $\mathscr{G}_m$ and $\mathscr{H}_m$. The gradient expression can be written with three terms as

$$\mathscr{G}_m = -\left(\frac{\partial \mathbf{d}(\mathbf{m})}{\partial \mathbf{m}}\right)^T \mathbf{W}_d^T \mathbf{W}_d \left(\mathbf{d}_{obs} - \mathbf{d}(\mathbf{m})\right) + \lambda_1 \mathbf{B}^T \mathbf{B}\mathbf{m} + \lambda_2 \mathbf{W}_m^T \mathbf{W}_m \left(\mathbf{m} - \mathbf{m}_p\right). \tag{4}$$

The sensitivity matrix $\mathbf{J} = \partial\mathbf{d}(\mathbf{m})/\partial\mathbf{m}$ is composed by the Fréchet derivatives of the synthetic data with respect to the model parameters. For the data contribution, the gradient is efficiently computed with an adjoint formulation (Plessix, 2006) without an explicit computation of $\mathbf{J}$. We solve the normal equation system by a quasi-Newton method using the L-BFGS-B scheme (Byrd et al., 1995). We have performed the inversion in the time domain although the approach is also valid in the frequency domain.

Marmousi application

A selected target of the Marmousi II P-wave velocity distribution (Martin et al., 2006) and a homogeneous density model are considered. The target exhibits two gas sand traps (Figure 1.a). The acquisition geometry contains 54 explosive sources, located along a horizontal line at 16 m depth, each 50 m. The layout is the same for all the shots, one fixed horizontal receiver line at 15 m depth and two fixed vertical lines of receivers inside two wells at $x = 50$ m and $x = 2700$ m with a 12.5 m interval. We do not consider any sources in wells as it is often more difficult and impractical, especially if one wishes to re-iterate the experiment periodically. On the contrary, we take benefit of sensors located permanently in wells for time-lapse imaging that increase dramatically our illumination for the baseline reconstruction.
The synthetic seismograms are computed using a Ricker wavelet (10 Hz central frequency) in our time-domain finite-difference code. We do not consider free surface in this application. Note that for inversion, the Tikhonov regularization parameter is fixed to $\lambda_1 = 20$ in all the further applications, as we want to show the influence of the prior term only. This value imposes a soft smoothing constraint. A highly smoothed velocity model (Figure 1.c) is first used as initial model for FWI. The pressure recorded data are used as observed data, both at the surface and in wells. A first investigation (Figure 2.a) is performed with a "classical" regularized FWI method without considering any prior model (equivalent to $\lambda_2 = 0$ in eq. (2)). It is clear that the optimization is trapped in a local minimum, due to cycle-skipping problem, especially in depth bellow 700 m and on the left part of the model until the second fault. Obviously, the velocity inside the two reservoir areas is not well recovered for this configuration.
In a second investigation, the inversion is performed using prior information extracted from wells. The true velocity values inside the two wells are considered to be known, as sonic logging is often used in such a case. We built a 2D prior model through a linear interpolation of the well properties (Figure 1.b). We also designed a 2D prior uncertainty model that takes into account the distance to the wells but also an increasing uncertainty in depth (not shown here). Figure 2.b shows the reconstructed velocity model by the FWI starting from the highly smoothed initial model, using the prior model and appropriate prior uncertainty model and $\lambda_2 = 3 \times 10^5$. We can see the significant improvement in the result using the prior information when compared to the "classical approach". In this case, the prior model allows the cycle-skipping problem to be mitigated and helps the inversion to converge to the global minimum. However the footprint of the prior can be noticed in the final model.

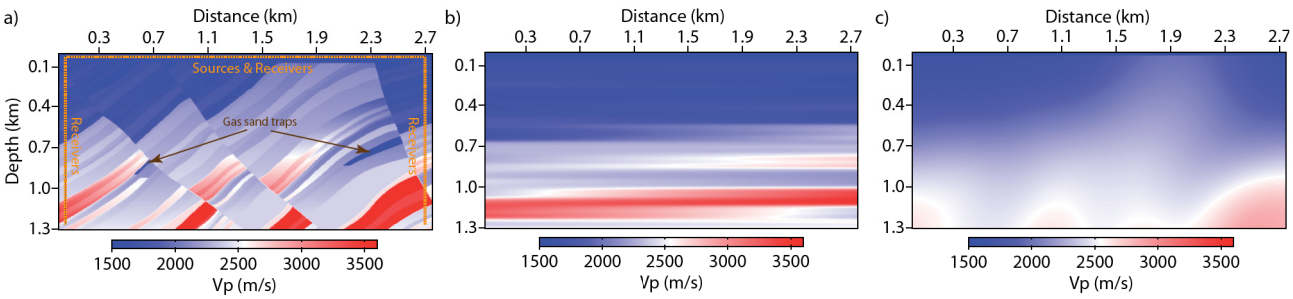

Figure 1 *(a) The target V_p model selected from Marmousi II model and the acquisition geometry; (b) the prior model built by linear interpolation of the two well-velocity profiles; (c) the highly smoothed model.*

Initial versus prior models

In this part, we address the relative role of the prior and initial models in the inversion procedure. A first natural idea could be to consider the prior model (Figure 1.b) as the initial model of FWI: since this model helps the FWI when used as prior, why would it not be a good initial one for inversion? Figure 2.c illustrates the inversion result using the classical regularized FWI (same tuning as Figure 2.a) and the "prior" model (Figure 1.b) as initial model. We can clearly see that inversion converged toward a local minimum, far from being satisfactory. Moreover, the optimization process stopped after few iterations in that case. The shallow part on the right part of model seems satisfactory but the left and the deep parts seems to be badly driven by this initial model built from interpolation in this strongly 2D structures. One interpretation of this failure is related to the major difference between the initial and the prior model: the initial model must be kinematically accurate and should not generate erroneous arrivals in the computed data, as we use it to solve the wave equation to compute synthetic data. It is much more difficult for inversion to suppress or shift a structure than creating a new one. On the contrary, the prior model is never used for solving the wave equation. It can contain any structure that can drive inversion toward the global minimum. In our case, the prior model allows FWI to be driven and partially fills the lack of low wavenumbers that can not be extracted from the data only.

Dynamic prior weighting

For practical applications in complex environments, the prior model derived from extra information on the target zone may be far away from true model. Even if the prior model can significantly improve result by an appropriate driving of the inversion, the final model can keep a significant footprint of the prior model structure. As shown on Figure 2.b, the result exhibits ghost interfaces coming from the interpolated prior. These footprints of prior should not honor the data itself. However, keeping a fixed weight on the prior term of the misfit function prevents the result to be improved as the prior model is intrinsically wrong in such a case. Therefore, one can investigate a dynamic weighting of the prior, in order to decrease the weight of the prior (λ_2) through the iterations of the optimization. The objective is to maximize the prior weight in the early iterations to steer inversion, and progressively decrease this weight to let the data improve the results once inversion as reach the global minimum valley of the misfit function.

In our study, we used an empirical approach based on the derivative of cost function evolution with iterations to iteratively decrease the λ_2 value. Under a fixed value of the derivative, meaning that the cost function evolution is too flat, we relax the prior term weight to give more freedom to the data part. Of course, this dynamic weighting requires an extra tuning parameter, but the result obtained with this strategy (Figure 2.d) shows an impressive improvement as compared to the fixed weight result (Figure 2.b). It shows that combining the prior information and a dynamic weighting of the prior allows (1) to steer the inversion toward the global minimum valley of the misfit function, mitigating cycle-skipping issues and filling the potential lack of low wavenumbers, while (2) progressively letting the data misfit term improve results when the prior model becomes too far from reality at the late iterations of FWI.

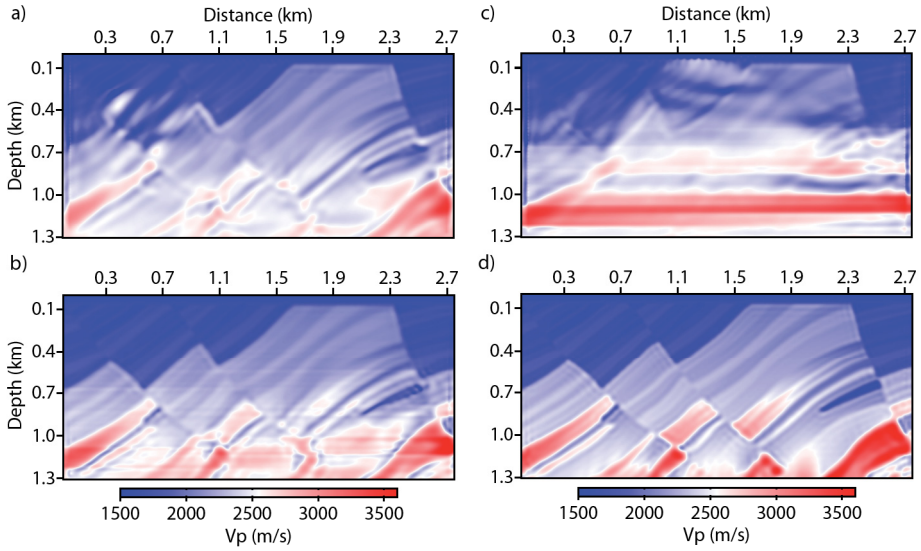

Figure 2 *The reconstructed V_p models by FWI, (a) starting from the smooth initial model with "classical" FWI, (b) starting from the smooth initial model and using the prior model and appropriate uncertainty model, (c) starting from 1D "prior" model with "classical" FWI, (d) as case (b) using a dynamic prior weighting.*

Conclusion

We have proposed a regularized FWI scheme based on a two-term model penalty in the misfit definition: the Tikhonov term to ensure smoothness and a prior model term to attract the inversion in a given direction. Prior model penalty allows the non-uniqueness issue of FWI to be reduced, particularly for time-lapse applications where prior such as sonic log and geological constraints can be available. We have demonstrated the specific influence of the initial and prior models in this framework: we show that the initial model must be kinematically consistent and should not contain erroneous structure, whereas the prior model can contain any structure that can help and drive inversion toward the global minimum valley. In addition, we have suggested a dynamic weighting approach of the prior weight, in order to progressively decrease the influence of prior through the optimization iterations. The combination of prior and dynamic weighting enables the lack of low wavenumbers to be mitigated when appropriate prior is available and helps FWI to converge toward the global minimum.

Acknowledgements

We would like to thank TOTAL Exploration & Production and SEISCOPE consortium for supporting this study. This work was performed by accessing to the high-performance computing facilities of CIMENT (Université de Grenoble, France) and to the HPC resources of GENCI-CINES under Grant 2010-046091.

References

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P. and Virieux, J. [2011] Sensitivity analysis of time-lapse images obtained by differential waveform inversion with respect to reference model. *SEG Technical Program Expanded Abstracts*, **30**(1), 2482–2486, doi:10.1190/1.3627707.

Byrd, R., Lu, P. and Nocedal, J. [1995] A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, **16**, 1190–1208.

Guitton, A. [2011] A blocky regularization scheme for full waveform inversion. *SEG Technical Program Expanded Abstracts*, **30**(1), 2418–2422.

Guitton, A., Ayeni, G. and Gonzales, G. [2010] A preconditioning scheme for full waveform inversion. *SEG Technical Program Expanded Abstracts*, **29**(1), 1008–1012, doi:10.1190/1.3513017.

Martin, G.S., Wiley, R. and Marfurt, K.J. [2006] Marmousi2: An elastic upgrade for marmousi. *The Leading Edge*, **25**(2), 156–166, doi:10.1190/1.2172306.

Operto, S., Virieux, J., Dessa, J.X. and Pascal, G. [2006] Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough. *Journal of Geophysical Research*, **111**(B09306), doi:10.1029/2005JB003835.

Plessix, R.E. [2006] A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**(2), 495–503.

Tikhonov, A. and Arsenin, V. [1977] *Solution of ill-posed problems*. Winston, Washington, DC.