Mismatches between physics and operators for least squares Kirchhoff and reverse time migrations

Daniel Trad

CREWES-University of Calgary

Summary

The standard approach for least-squares migration (LSMIG) is designed to calculate the reflectivity model that can best predict the data under the Born approximation assumption. This is achieved by iteratively modifying the reflectivity model with updates obtained by mapping the prediction errors from data to model space. This approach assumes that all prediction errors are due to incorrect reflectivity. This assumption is no fulfilled in general, except for the very special case of synthetic data and a matching modeling operator in the LSMIG algorithm. Because LSMIG assumes Born modeling, even for synthetic data the prediction error is always due in part to mismatches between "physics" and data. Therefore, each update to the reflectivity contains wrong components. The larger the mismatch, the larger the contribution of these wrong components. Usually, this noise is attenuated by model regularization, but not always effectively because model constraints can't account for all possible operator errors. Often the model regularization can be the dominant effect in the inversion, so that instead of seeing the effect of inverting the Hessian, we may see the effects of filtering the noise. In this work, I compare several situations for LSMIG using both Kirchhoff and RTM operators and discuss many possible ways in which data and model spaces can be affected by wrong mappings. Also, I discuss localized residual adaptive filtering instead of model filtering and data pre-conditioning as a possible way to remove inversion noise.

Introduction

Inversion is a very common technique in seismic processing and other branches of science, consisting of creating a model that can predict acquired data. This approach has three main components, the model m, the data d, and the operator L that connects them. Choosing the kind of model defines the operator. In seismic we use many types of models. For example, in Full Waveform inversion (FWI) we use some function of velocity, like slowness square. In Least squares migration (LSMIG) we use a model that resembles reflectivity (Kirchhoff modeling), or velocity perturbations (scattering or Born modeling). When the goal of our processing is to predict data, for example in noise attenuation or interpolation, we have flexibility in choosing the model and the operator as long as they predict the part of the data we want. In FWI and LSMIG, on the other hand, we need to make sure that the operator resembles the physics that generated the data because the goal is the model itself. In other words, a wrong operator may predict the data correctly and the model incorrectly.

In all these techniques the most common approach is to calculate the model that best predict the data: we minimize the residual square or average difference between data and predictions. This minimization always proceeds in some form of optimization, which depends on whether the operator represents a linear or non-linear transformation. In LSMIG, at difference of FWI, we keep the operator independent of the model and assume that the source wavefield changes only with the background velocity and not the reflectivity (Schuster, 2017). Linearity implies that all errors present in the residuals can be mapped to and corrected by changes in the reflectivity. The consequence of this assumption is far reaching because residuals contain components that are not predicted by the model-operator combination, and components that are predicted with wrong amplitude and phase. Since the least squares mechanism is designed to reduce error in the prediction, it will try to use any component of the model space to predict something similar to the input data. For example, an aliased operator can use low frequencies to produce data with higher frequencies. A mismatch between physics and mathematics will not only prevent the inversion algorithm to produce zero residuals but also will produce models with artifacts or noise. In this work, I will discuss this problem and show some ways of limiting its effects.

Adjoint operators for least squares migration

Least squares migration calculates the model that best predict the data in a least squares sense:

$$J = \|d - Lm\|^2 + \|W_m m\|^2 \tag{1}$$

where d is the acquired data, L the prediction operator, m the reflectivity model and W_m a regularization function to constrain solutions of m. I will assume W_m to be the identity for simplicity, but conclusions are still valid for the general case. The solution to this problem is the least squares reflectivity obtained by

$$m_{lsrtm} = (L^H L + \lambda I)^{-1} L^H d \tag{2}$$

Where L^H is the adjoint operator that performs the mapping from data to reflectivity and satisfies adjoint numerical properties with the forward operator L (Claerbout, 1992). In general, $L^H\ d$ resembles the model obtained by migration and is called the adjoint model ($m_{adjoint}$). The main difference between migration and LSMIG is the elimination of the Hessian, whose diagonal components represent differences in illumination between different parts of the model and off-diagonal components represent the crosstalk between model components. Thus, by eliminating the Hessian (transforming it into an identity operator), we remove differences in illumination and increase sharpness of the image by removing interference (or blurring effect) introduced by the non-orthogonality of the operator (Schuster, 2017):

$$m_{lsrtm} = (L^H L + \lambda I)^{-1} m_{migration} = (L^H L + \lambda I)^{-1} (L^H L) m_{true} \tag{4}$$

In practice these equations are solved by an iterative solver, where each iteration calculates the gradient or direction along which the prediction improves by migrating residuals. The distance to travel along this direction or step size is calculated by a parametric approach that requires the operators L and L^H to be a perfect adjoint pair. This requires to choose one of the operators and calculate the other. For Kirchhoff migration, the forward and adjoint operators can be represented as a exchange in the inner summation between data and model (Nemeth at al, 1999):

$$m(x,y,z,h) += \int w^*(x,y,z,t) d(s,r,t) ds, dr, dt$$
$$d(s,r,t) += \int w(x,y,z,t) m(x,y,z,h) dx, dy, dz, dh \tag{5}$$

where *m(x,y,z,h)* represents the image at each point in space and offset (or angle), and *d(s,r,t,h)* represents each data point in shot, receiver and time coordinates. These are just simplified expressions because Kirchhoff migration contains many filters and weights. The details of the weights *w(x,y,z,t)* in equation (5) depend on whether we start from a Kirchhoff migration with deconvolution imaging condition or a Kirchhoff modelling with cross-correlation imaging condition, or we use antialias filtering, and other preconditioners (Trad, 2017).

For RTM implementation, the forward adjoint pairs are more complex than for Kirchhoff migration. If we start from the adjoint operator to resemble an RTM migration operator, then the forward operator has to be designed as a Born modeling obtained from reflectors as a convolutional model, also known as demigration.

$$r(x,y,z) = \int \ddot{u}_s(x,y,z,t) \otimes u_r(x,y,z,T-t) dt \tag{2}$$

$$u_r(x,y,z,t) = \int \ddot{u}_s(x,y,z,t) \otimes r(x,y,z) dt \tag{3}$$

This may seem a bit strange at first because data are usually predicted by finite difference approximations of the wave equation in the exact velocity model. However, finite difference prediction is not an adjoint operator of imaging. Instead, correlation (RTM) is the adjoint pair of demigration (convolution). In reality there are some subtle differences between the RTM operator used in LSRTM and a typical RTM algorithm

(Ji, 2009), and those details have to taken into account if we want to use a linear conjugate gradient algorithm (Xu and Sacchi, 2016, Chen and Sacchi, 2017).

Mapping between data and model spaces

Independently of the algorithm details used for the inversion, each iteration performs two key tasks:

a) A mapping from the current reflectivity model to the data space (modeling).

b) A mapping from the prediction errors to the model space (migration or gradient estimation).

The main goal in this paper is to discuss the errors involved in these two steps. Figure 1 shows a diagram indicating different possibilities for errors in the case of migration. Some components of the data map to reflectivity with strong connections, but the strength of these connectors varies (different illumination). These differences are eliminated by LSMIG, because they are contained in the diagonal of the Hessian, but also, can be removed by cheaper methods. In general, these mappings are not unique (same data maps to different components of the model) and this cross-talk blurs the image. This blurring is removed by LSMIG because it is contained in the off-diagonal components of the Hessian. This is the sharpening effect of LSMIG that justifies the extra computation time of inverting the Hessian.

Some data components have no mapping to the model, for example random noise, and they do no harm nor help the image. Similarly, some model components do not predict valid data, for example high dips or under-resolution features. Finally, there are data and model components that have wrong mappings, for example if the velocities (and therefore operators) are not exact. Inside an iterative LSMIG, these components create wrong updates that become the origin of the noise we often see with iterations.

In practice, we use a global measure for residuals as an indicator that fitting improves and predictions are closer to the input. However, the decreasing of the residual does not imply that predictions are becoming better everywhere. One possible way to control the noise introduced by wrong operators is to locally detect residual components that are consistently increasing with iterations and turn them down.

Examples

Figure 2a shows a RTM result for the Marmousi model with 25 shots, while Figure 2b shows the equivalent LSRTM after 9 iterations only. Inversion has helped the image by the sharpening effect and the illumination compensation. However, this result depends heavily on the operator and the velocity model to be exact. It takes just a mild distortion on either to decrease or eliminate the improvement we observe. This is, unfortunately, the normal situation when working with real data sets. When applying similar ideas to LSMIG with Kirchhoff, the problem appears even with synthetics because data generated with finite difference algorithms have many features that can not be predicted properly by ray tracing approximations. This effect is more obvious for structured data, which is why LSK has not been too successful in practice for complex structures. Figure 3a shows a Kirchhoff migration, and Figure 3b shows a modified version of LSK (9 iterations). The change consists on removing from residuals components that are not correctly predicted by the Kirchhoff operator. This is done during iterations, instead of the most common approach of filtering data before migration.

Conclusions

LSMIG has the important characteristic of sharpenning reflectivity images and compensating for illumination. This positive features heavily depending on the matching between operators and physics. The only true solution is to design better operators but that is costly and difficult. A compromise solution is to turn off during the inversion residual components that can not properly be predicted. That requires intelligent algorithms that detect those events by properly tracking evolution of residuals with iterations.

Acknowledgements

My gratitude to Sam Gray for many insightful discussions, CREWES sponsors and NSERC (Natural Science and Engineering Research Council of Canada) for support through the grant CRDPJ 461179-13.

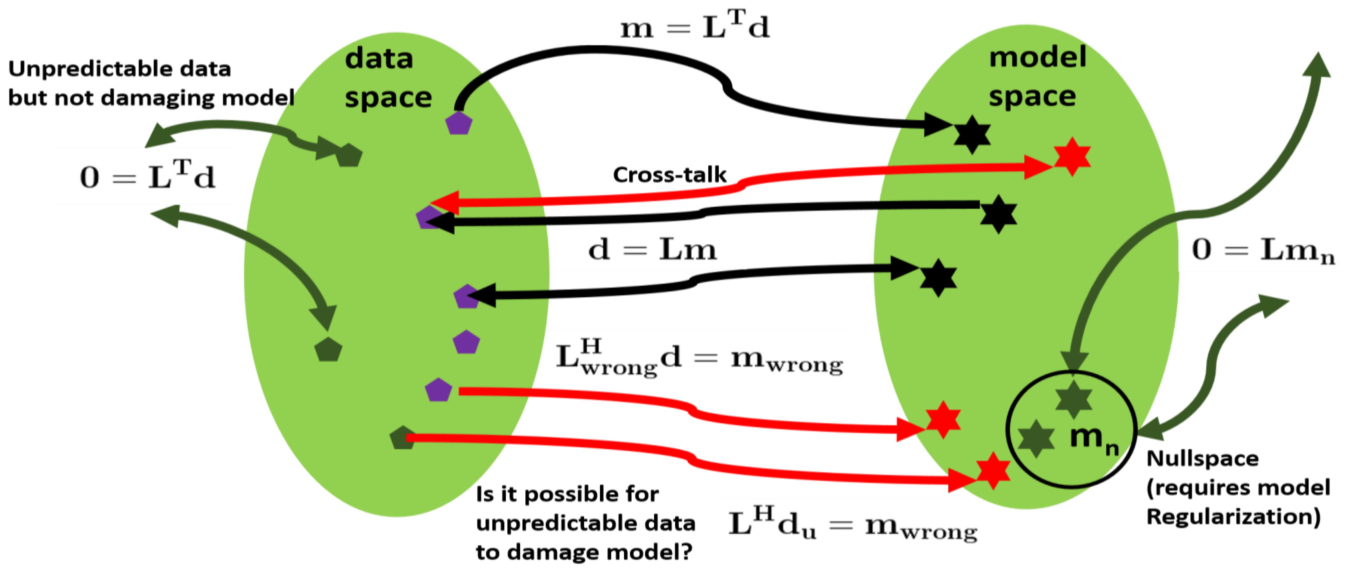

Figure 1: Mapping between data and model space for an inverse problem

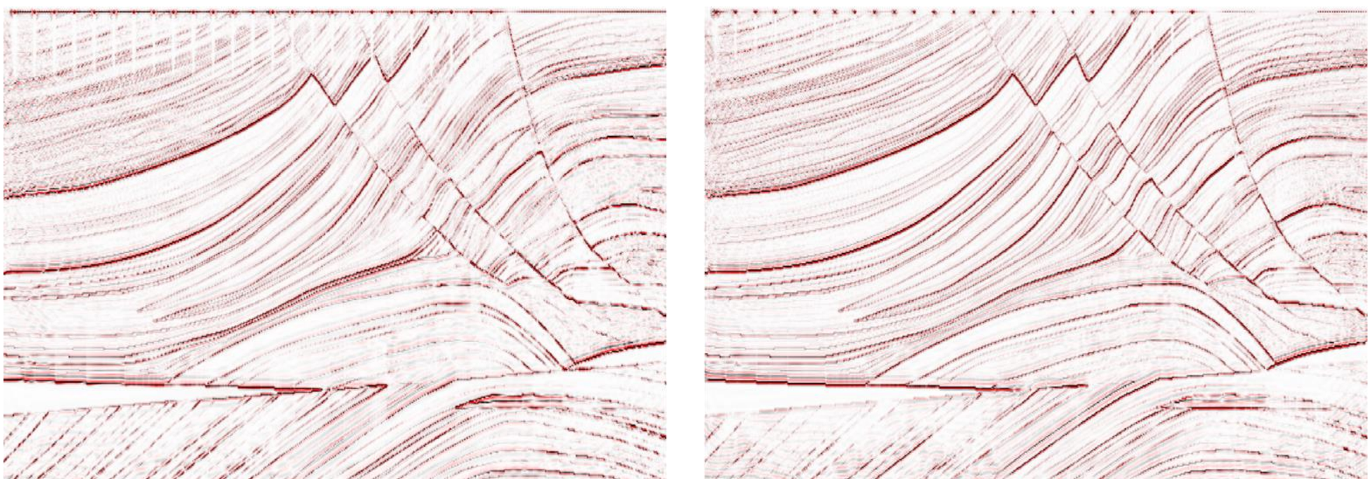

Figure2: Best scenario for least squares migration: correct operator with exact velocity. a) RTM (25 shots), b) LSRTM with 9 iterations. (AGC applied in both figures for display).

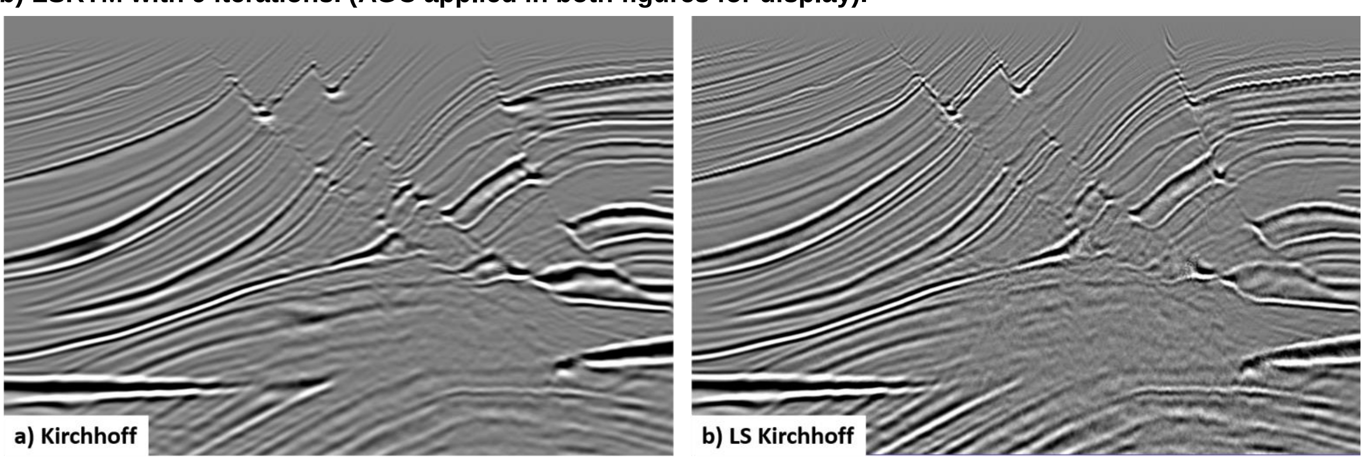

Figure 3: Common situation for least squares migration when operator does not completely match the physics. a) Kirchhoff migration, b) LS Kirchhoff with residual adaptive weighting.

References

Chen, K., and Sacchi, M. D., 2017, Elastic least-squares reverse time migration via linearized elastic full waveform inversion with pseudo-hessian preconditioning: GEOPHYSICS, 82, No. 5, S341–S358.

Claerbout, J., 1992, Earth sounding analysis, Processing versus inversion: Blackwell Scientific Publications, Inc.

Ji, J., 2009, An exact adjoint operation pair in time extrapolation and its application in least-squares reverse time migration: GEOPHYSICS, 74, No. 5, H27–H33.

Nemeth, T., Wu, C., and Schuster, G. T., 1999, Least squares migration of incomplete reflection data: GEOPHYSICS, 64, No. 1, 208–221.

Trad, D., 2017, Challenges for Least Squares Kirchhoff depth migration, SEG Technical Program Expanded Abstracts: 4545-4549.

Schuster, G., 2017, Seismic Inversion: Society of Exploration Geophysicists.

Xu, L., and Sacchi, M., 2016, Least squares reverse time migration with model space preconditioning and exact forward/adjoint pairs, in EAGE Annual Conference Proceedings, (EAGE, Vienna).