Source estimation for frequency-domain FWI with robust penalties

Aleksandr Y. Aravkin[1], Tristan van Leeuwen[1], Henri Calandra[2], and Felix J. Herrmann[1]
[1] Dept. of Earth and Ocean sciences University of British Columbia Vancouver, BC, Canada
[2] Total Exploration and Production, Pau, France

January 17, 2012

Abstract

Source estimation is an essential component of full waveform inversion. In the standard frequency-domain formulation, there is closed form solution for the the optimal source weights, which can thus be cheaply estimated on the fly. A growing body of work underscores the importance of robust modeling for data with large outliers or artifacts that are not captured by the forward model. Effectively, the least-squares penalty on the residual is replaced by a robust penalty, such as Huber, Hybrid ℓ_1-ℓ_2 or Student's t. As we will demonstrate, it is essential to use the same robust penalty for source estimation. In this abstract, we present a general approach to robust waveform inversion with robust source estimation. In this general formulation, there is no closed form solution for the optimal source weights so we need to solve a scalar optimization problem to obtain these weights. We can efficiently solve this optimization problem with a Newton-like method in a few iterations. The computational cost involved is of the same order as the usual least-squares source estimation procedure. We show numerical examples illustrating robust source estimation and robust waveform inversion on synthetic data with outliers.

Introduction

Full waveform inversion (FWI) is an approach to obtain velocity model parameters from seismic data. FWI is naturally cast as a nonlinear least squares optimization problem (Tarantola, 1984; Pratt et al., 1998). A growing body of work is concerned with replacing the least-squares misfit by a robust misfit. A compelling motivation for these approaches is the need for methods which can recover good solutions even when data contain artifacts unexplained by the forward model. Bube and Langan (1997) considered a hybrid misfit to avoid non-uniqueness problems with the Huber norm for linear inverse problems, and Bube and Nemeth (2007) explored computational advantages of line searches for the hybrid misfit in the same context. Guitton and Symes (2003) suggested using the Huber norm for FWI. Shin and Ha (2008) and Brossier et al. (2010) compared various penalties, including least squares, Huber and ℓ_1. Recently, a reasonable inversion result was obtained using a Student's t-based non-convex misfit in the presence of many large unexplained artifacts in the data (Aravkin et al., 2011a,b), and the advantages of using non-convex penalties was explored using statistical considerations.
It is well-known that source-estimation can be easily incorporated into the least-squares formulation by choosing the source weights for a given model to be a minimizer of a least-squares misfit function (Pratt, 1999). In this abstract, we extend this approach to source estimation in the context of robust FWI, using general twice differentiable misfit functions. The approach is similar to variable projection (Golub and Pereyra, 2003) for nonlinear least squares, and extends to a wide class or robust misfits. After deriving the general approach, we show that we recover the standard method in the least squares case, and give details and experiments showing source estimation using the hybrid (Bube and Langan, 1997) and Student's t (Aravkin et al., 2011b) approaches. We conclude with a synthetic example where Student's t FWI with robust source estimation is used to invert highly corrupted data.

Robust FWI with Robust Source Estimation

Let $\rho(\mathbf{r})$ be any differentiable misfit, such as least squares, Huber, hybrid, or Student's t. Suppose that in addition to the velocity model parameters, we would like to estimate complex source weights, one per shot/frequency pair. The (frequency-domain) FWI problem with source estimation is simply formulated as a joint optimization problem:

$$\min_{\mathbf{m},\mathbf{a}} \phi(\mathbf{m},\mathbf{a}) := \sum_{\omega,i} \rho\left(\mathbf{d}_{i,\omega} - a_{i,\omega}\mathbf{f}_{i,\omega}(\mathbf{m})\right), \tag{1}$$

where i indexes shots, ω indexes frequencies, $\mathbf{d}_{i,\omega}$ is the observed shotrecord, $\mathbf{f}_{i,\omega}(\mathbf{m})$ is the modeled shotrecord for model $\mathbf{m}$ and $\mathbf{a}$ is a vector of complex source weights $a_{i,\omega}$. Note that (1) is completely separable in the source terms, and the estimation of the source weights $a_{\omega,i}$ can therefore be done independently. Exploiting this structure we can define

$$\widehat{a}_{\omega,i}(\mathbf{m}) = \operatorname*{argmin}_{a_{i,\omega}} \rho\left(\mathbf{d}_{i,\omega} - a_{i,\omega}\mathbf{f}_{i,\omega}(\mathbf{m})\right) , \tag{2}$$

that is, the vector of source-weights $\widehat{\mathbf{a}}$ is taken to be the minimizer of (1) for a fixed model $\mathbf{m}$. The source weights recovered in this manner satisfiy

$$\nabla_{\mathbf{a}}\phi(\mathbf{m},\widehat{\mathbf{a}}) = 0 \,. \tag{3}$$

Now we define a modified problem, where we substitute the optimal source weights into the misfit

$$\min_{\mathbf{m}} \widehat{\phi}(\mathbf{m}) := \sum_{\omega,i} \rho\left(\mathbf{d}_{i,\omega} - \widehat{a}_{i,\omega}(\mathbf{m})\mathbf{f}_{i,\omega}(\mathbf{m})\right). \tag{4}$$

We can solve this by making a very minor modification to any algorithm for solving the robust FWI problem — namely the re-estimation of the source weights $\widehat{\mathbf{a}}$ any time the model $\mathbf{m}$ is updated. Even though the source weights depend explicitly on the model, this does not add any extra terms to the gradient. Indeed it is easily verified that the gradient of the modified misfit is the same as the gradient of the original objective, evaluated with the optimal source weights, or

$$\nabla \widehat{\phi}(\mathbf{m}) = \nabla_{\mathbf{m}} \phi(\mathbf{m}, \widehat{\mathbf{a}}) \,. \tag{5}$$

Any minimizer $\widehat{\mathbf{m}}$ of (4) satisfies $\nabla \widehat{\phi}(\widehat{\mathbf{m}}) = 0$, and so by (5) we have $\nabla_{\mathbf{m}} \phi\big(\widehat{\mathbf{m}}, \widehat{\mathbf{a}}(\widehat{\mathbf{m}})\big) = 0$. This fact, together with (3) guarantees that $(\widehat{\mathbf{m}}, \widehat{\mathbf{a}})$ is a local minimizer of (1).
In the next section, we show that (2) generalizes the least squares source estimation strategy, and provide implementation details for the Hybrid and Student's t misfits.

Implementation Details for Optimal Source

In this section, we discuss how to solve the scalar optimization problem (2) with a Newton-like algorithm. We drop the source-frequency subscripts for ease of notation. The basic Newton iteration to solve (2) is given by

$$a^{\nu+1} = a^{\nu} - \frac{g(a^{\nu})}{h(a^{\nu})}, \tag{6}$$

where $g(a)$ and $h(a)$ are the first and second derivatives of the objective in (2) with respect to a. We give explicit details for three popular misfits ρ.

Least Squares. In this case, (2) reduces to minimizing $||\mathbf{d} - a\mathbf{f}(\mathbf{m})||^2$, and we get the closed form solution (Pratt, 1999)

$$\widehat{a}_{ls} = \mathbf{f}(\mathbf{m})^T \mathbf{d} / ||\mathbf{f}(\mathbf{m})||_2^2 \,,$$

which is also the first Newton iteration if we start from $a = 0$.

Hybrid. The *hybrid* objective considered by Bube and Nemeth (2007) is given by
$\rho(\mathbf{r}) = \sum_i \sqrt{1 + (r_i)^2/\sigma^2} - 1$. The source estimation problem for a robust FWI approach using this objective boils down to solving

$$\widehat{a}_{hyb} = \operatorname*{argmin}_a \sum_i \sqrt{1 + (r_i(\mathbf{m}, a))^2/\sigma^2} - 1 \,,$$

where $\mathbf{r}(\mathbf{m}, a) = \mathbf{d} - a\mathbf{f}(\mathbf{m})$. The iteration (6) takes the form

$$a^{\nu+1} = a^{\nu} - \frac{\sum_i r_i(a^{\nu}) f_i / \sqrt{1 + r_i(a^{\nu})}}{\sigma \sum_i f_i^2 / \sqrt{1 + r_i(a^{\nu})}} \,, \tag{7}$$

where the dependence on $\mathbf{m}$ has been dropped.

Student's t The robust function is this case is given by

$$\rho(\mathbf{r}) = \frac{1}{2} \sum_i \log(k + (r_i)^2). \tag{8}$$

While this objective is not convex, a strictly positive approximation to the second derivative is easily obtained (see Aravkin et al., 2012, eq. (5.9)). Thus, the iteration (6) takes the form

$$a^{\nu+1} = a^{\nu} - \frac{\sum_i r_i(a^{\nu}) f_i / (k + (r_i(a^{\nu}))^2)}{\sum_i f_i^2 / (k + (r_i(a^{\nu}))^2)} \,. \tag{9}$$

Source estimation example

To show how robust source-estimation works in practice, we conduct the following stylized experiment. We convolve a shot record with a 10 Hz. Ricker wavelet and replace a number of traces with noise, as shown in figure 1. We reconstruct the source wavelet by fitting this shot record to the original using the

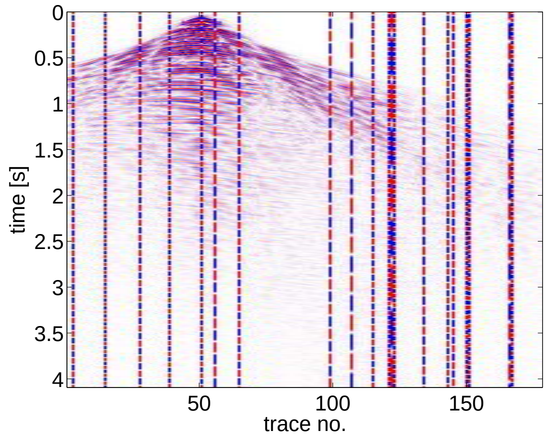

***Figure 1** Data with outliers in the form of bad traces.*

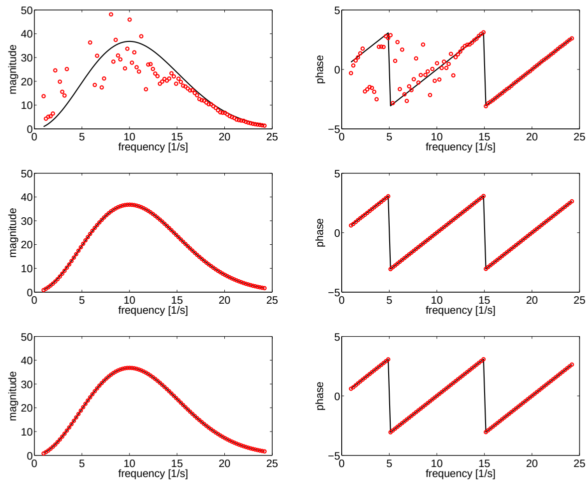

***Figure 2** Estimated source wavelet using Least-Squares (top), Hybrid (middle) and Students t (bottom) approaches.*

least-squares, Hybrid and Student's t approaches, which required 5 or 6 Newton iterations to converge. The results are shown in figure 2. Both the Hybrid and Student's t recover well in this case, while the least-squares reconstruction is quite useless, especially at low frequencies. Note that for each source, a whole vector of data is used to estimate a single complex value; the least-squares estimate in this case is analogous to finding the mean of a set of data, which is not robust to outliers. In contrast, both the Student's t and Hybrid approach return estimates analogous to the median of a large set of data. In this context, similar performance would be expected from Huber and ℓ_1; however, we would not be able to apply a Newton method, since these penalties are not twice differentiable.

Robust FWI example

To illustrate why robust source estimation is important for robust FWI we perform the following three experiments on synthetic data with outliers: *i)* least-squares FWI with least-squares source estimation (LS-LS); *ii)* robust FWI with the Students t penalty and least-squares source estimation (ST-LS) and *iii)* robust FWI with the Students t penalty with corresponding source estimation (ST-ST). We use a frequency-domain modeling operator based on a 9-point discretization of the Helmholtz equation (Jo, 1996). The data are generated for a subset of the Marmousi for 61 equispaced sources, 301 equispaced receivers and 12 frequencies between 3 and 25 Hz. We use an L-BFGS method to fit the model, and source estimation is implemented as described above. The initial model we used was a smoothed version of the original model. To create noise, we replace 20% of the samples in the data with Gaussian noise. A LS-LS reconstruction on data without noise, as well as the reconstructions on data with noise are shown in figure 3. The ST-ST reconstruction is nearly identical to the LS-LS reconstruction without noise, thus demonstrating the ability of this approach to deal with noise. The LS-LS reconstruction with noise is completely meaningless. Although the ST-LS shows some of the underlying structure, this clearly demonstrates the need to use robust source estimation in conjunction with a robust penalty.

Conclusions

We surveyed several *robust* formulations of FWI that use penalty functions that are less sensitive to outliers in the data, including the Huber, hybrid and Students t penalty. In practice, one usually estimates the source wavelet as part of the inversion process. We extended the usual least-squares approach to source estimation to a class of robust formulations with twice-differentiable misfit functions. For every evaluation of the misfit, we solve a scalar optimization problem for each source and frequency to obtain the source weights. This can be done with a Newton method, and our experience is that we need only a

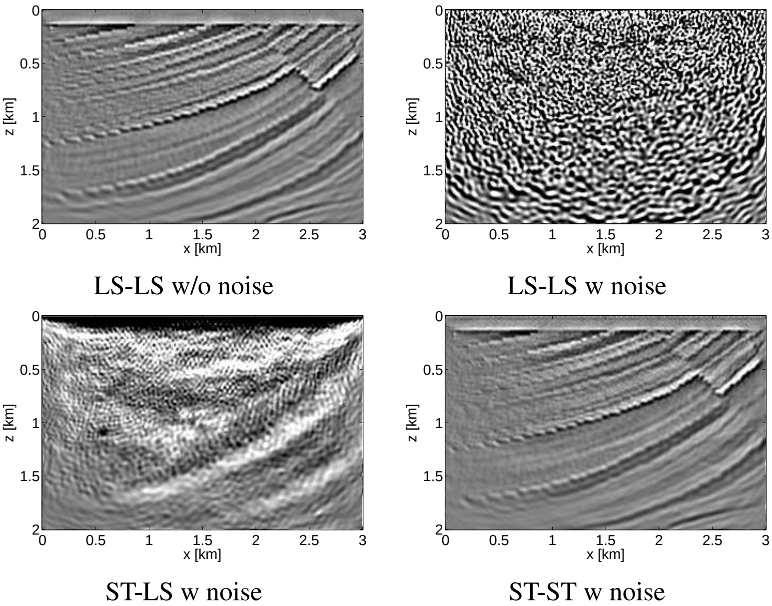

Figure 3 *Reconstructions (difference between initial and final models) for different scenario's.*

few iterations. Therefore, the computational cost incurred is negligible compared to the cost of forward modeling of the wavefield. We demonstrate robust source estimation on a real shot gather and show the uplift of robust source estimation in conjunction with robust FWI.

Acknowledgments

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

References

Aravkin, A., van Leeuwen, T. and Herrmann, F. [2011a] Robust full-waveform inversion using the student's t-distribution. *SEG Technical Program Expanded Abstracts*, **30**(1), 2669–2673, doi:10.1190/1.3627747.

Aravkin, A., Burke, J. and Pillonetto, G. [2012] Robust and trend-following student's t kalman smoothers. *Optimization Online Preprint*.

Aravkin, A., Friedlander, M., Herrmann, F. and van Leeuwen, T. [2011b] Robust inversion, dimensionality reduction, randomized sampling. *Submitted to Math. Prog.*

Brossier, R., Operto, S. and Virieux, J. [2010] Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, **75**(3), R37–R46.

Bube, K.P. and Langan, R.T. [1997] Hybrid ℓ_1/ℓ_2 minimization with applications to tomography. *Geophysics*, **62**(4), 1183–1195.

Bube, K.P. and Nemeth, T. [2007] Fast line searches for the robust solution of linear systems in the hybrid ℓ_1/ℓ_2 and huber norms. *Geophysics*, **72**(2), A13–A17.

Golub, G. and Pereyra, V. [2003] Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems*, **19**(2), R1.

Guitton, A. and Symes, W.W. [2003] Robust inversion of seismic data using the huber norm. *Geophysics*, **68**(4), 1310–1319.

Jo, C.H. [1996] An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator. *Geophysics*, **61**(2), 529, ISSN 1070485X, doi:10.1190/1.1443979.

Pratt, R.G., Shin, C. and Hicks, G. [1998] Gauss-newton and full newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, **133**(2), 341Ð362.

Pratt, R.G. [1999] Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, **64**, 888–901.

Shin, C. and Ha, W. [2008] A comparison between the behavior of objective functions for waveform inversion in the frequency and laplace domains. *Geophysics*, **73**(5), VE119–VE133.

Tarantola, A. [1984] Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, **49**(8), 1259–1266.