# In which domain should we measure the misfit for robust full waveform inversion?

T. van Leeuwen[1], A.Y. Aravkin[1,2], H. Calandra[3] and F.J. Herrmann[1]

[1] Dept. of Earth and Ocean sciences, University of British Columbia, Vancouver, BC, Canada

[2] Dept. of Computer Science, University of British Columbia, Vancouver, BC, Canada

[3] Total Exploration and Production, Pau, France.

February 19, 2013

**Abstract**

Full-waveform inversion relies on minimizing the difference between observed and modeled data, as measured by some penalty function. A popular choice, of course, is the least-squares penalty. However, when outliers are present in the data, the use of robust penalties such as the Huber or Student's t may significantly improve the results since they put relatively less weight on large residuals. In order for robust penalties to be effective, the outliers must be somehow localized and distinguishable from the good data. We propose to first transform the residual into a domain where the outliers are localized before measuring the misfit with a robust penalty. This is exactly how one would normally devise filters to remove the noise before applying conventional FWI. We propose to merge the two steps and let the inversion process implicitly filter out the noise. Results on a synthetic dataset show the effectiveness of the approach.

## Introduction

Full-waveform inversion (FWI) relies on minimizing the difference between observed and modeled data (Tarantola, 1984). The question is, **how should we measure the difference?** There has been a steady flow of papers showing that the least-squares misfit is not the best choice when the data are contaminated with non-Gaussian noise. Instead, so-called *robust* norms, such as the Huber (Guitton and Symes, 2003; Brossier et al., 2010), Hybrid (Bube and Langan, 1997), Cauchy (Crase et al., 1990; Amundsen, 1991) or Student's t (Aravkin et al., 2011) may dramatically improve the results when the noise consists of isolated large outliers that affect only parts of the data. The next question is, **what constitutes an outlier?** Robust penalties work by ignoring a 'few' residuals that are large compared to 'most' residuals. It is easy to imagine that a bad trace would constitute an outlier. Similarly, spurious events may be seen as outliers in the time-offset domain. But, what happens if we do frequency-domain waveform inversion? Noise that was localized in time-offset domain will now be spread out across all frequencies and a robust penalty will not be able to discern the noise from the good data. Figure 1 illustrates this by comparing two different kinds of noise (bad traces and spurious events) in 3 domains ($(t,h)$, $(f,h)$ and $(f,k)$). This observation leads us to pose the question: **In which domain should we measure the misfit?** In this paper we review some recent work on robust FWI and extend it by introducing a transform that will localize the noise. We illustrate the method on a synthetic dataset with artificially added outliers.

## Mathematical formulation

We consider the following frequency-domain formulation of the FWI problem:

$$\min_{\mathbf{m},\mathbf{w}} \phi(\mathbf{m},\mathbf{w}) = \sum_i \rho(B_i(w_i P_i \mathbf{u}_i - \mathbf{d}_i)), \quad \mathbf{u}_i = G_i(\mathbf{m}), \tag{1}$$

where $\mathbf{d}_i$ is the observed data for one frequency and source $i$, $\mathbf{w}$ are source weights, $P_i$ restricts the wavefield to the receiver locations, $G_i(\mathbf{m})$ is the corresponding modelling operator, $B_i$ is a data-processing operator (to be discussed later) and $\rho$ is a robust penalty. Examples are the Student's t penalty:

$$\rho(\mathbf{r}) = \sum_i \log(1 + |r_i|^2/(\sigma^2 k)), \tag{2}$$

and the Huber penalty:

$$\rho(\mathbf{r}) = \sum_{|r_i|\leq\sigma} \frac{1}{2\sigma^2}|r_i|^2 + \sum_{|r_i|>\sigma} \frac{1}{\sigma}|r_i| - \frac{1}{2}. \tag{3}$$

Here, $\sigma$ is a scale parameter and $k$ denotes the degrees of freedom.

**Scale initialization.** The scale parameter $\sigma$ is taken to be a fraction of residual for the initial model $\sigma = \max_{ij} \alpha|(\mathbf{r}_i)_j|$. With this scaling, residuals smaller than $\alpha$ end up in the 'least-squares' part of the penalty, while residuals larger then $\alpha$ end up in the tail of the penalty. This is illustrated in figure 2.

**Source estimation.** For a given model $\mathbf{m}$, we estimate the source wavelet by solving for each source:

$$\overline{w}_i = \underset{w}{\operatorname{argmin}}\, \rho(B_i(wP_i\mathbf{u}_i - \mathbf{d}_i)). \tag{4}$$

This is a scalar optimization problem that can be solved efficiently in a number of ways. The objective function is now effectively a function of the model only: $\overline{\phi}(\mathbf{m}) = \phi(\mathbf{m},\overline{\mathbf{w}})$ and the gradient of the reduced objective is given by $\nabla\overline{\phi}(\mathbf{m}) = \nabla_{\mathbf{m}}\phi(\mathbf{m},\overline{\mathbf{w}})$. For more details on this variable projection approach, we refer to (Aravkin et al., 2012; Aravkin and van Leeuwen, 2012)

**Noise localization.** The processing operator $B_i$ should localize the noise. We can use any transform that would normally be used to filter out the noise (Fourier for periodic noise, Radon for noise with moveout, Curvelets for more complicated coherent events, etc.). The power of this approach is that we are not explicitly filtering out the noise, with the danger of losing information. Instead, we let a robust inversion process decide which parts of the data can be fitted and which should be ignored. Thus, the filtering is done implicitly as part of the inversion process.

## Numerical example

We present results on the Marmousi model. The data are either contaminated by replacing a portion of the traces with large constant values or by adding periodic noise to each shot (basically sin(offset)). The inversion is done for a single frequency of 4 Hz using 20 steepest-descent iterations starting from the model depicted in figure 3. We use the Student's t penalty either in the $(f,h)$ or $(f,k)$ domain. The scaling parameter $\alpha$ is picked by hand, although we could use techniques described by Aravkin and van Leeuwen (2012) to automatically fit it as part of the inversion. Reference results on data without noise are show in figure 4.

In the first experiment, every 5th trace out of every 5th shot is heavily contaminated. The resulting 'noise' is very localized in space and still reasonably localized in the wavenumber domain since 4 out every 5 shots are completely clean. The results are shown in figure 5. The results from the spatial domain formulation look good, but could possibly benefit from a more careful tuning of the parameter $\alpha$. The formulation in the wavenumber domain yields a slightly noisy results but does pretty well because the Student's t formulation allows it to ignore the contaminated shots altogether.

In the second experiment we add a sine wave to every shot with a moderate amplitude (10). This noise is certainly not localized in the spatial domain. The results are shown in figure 6. The formulation in the wavenumber domain does much better, as expected.

## Conclusion and Discussion

The use of robust penalties such as the Huber or Student's t in FWI may significantly improve the results when large outliers are present in the data. Robust penalties are not a magic bullet, though; The outliers have to be localized in some sense so that the robust penalty can distinguish 'good' residuals from bad ones. To increase the applicability of robust penalties we propose to first transform the residuals into a different domain in order to localize the noise. Of course, this is exactly how one would normally devise filters to remove the noise before applying conventional FWI. Our approach allows us to merge the two steps and let the inversion process implicitly filter out the noise. Results on a synthetic dataset show the effectiveness of the approach.

## 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, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.

## References

Amundsen, L. [1991] Comparison of the least-squares criterion and the cauchy criterion in frequency-wavenumber inversion. *Geophysics*, **56**(12), 2027–2035, doi:10.1190/1.1443015.
Aravkin, A., van Leeuwen, T. and Herrmann, F. [2011] 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.Y. and van Leeuwen, T. [2012] Estimating nuisance parameters in inverse problems. *Inverse Problems*, **28**(11), 115016, ISSN 0266-5611.
Aravkin, A.Y., van Leeuwen, T., Calandra, H. and Herrmann, F.J. [2012] Source estimation for frequency-domain fwi with robust penalties. *EAGE technical program*, EAGE, EAGE.
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. and Langan, R. [1997] Hybrid $\ell_1/\ell_2$ minimization with applications to tomography. *Geophysics*, **62**(4), 1183–1195.
Crase, E., Pica, A., Noble, M., McDonald, J. and Tarantola, A. [1990] Robust non-linear waveform inversion: Application to real data. *Geophysics*, **55**, 527–538.
Guitton, A. and Symes, W.W. [2003] Robust inversion of seismic data using the huber norm. *Geophysics*, **68**(4), 1310–1319.
Tarantola, A. [1984] Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, **49**(8), 1259–1266, doi:10.1190/1.1441754.

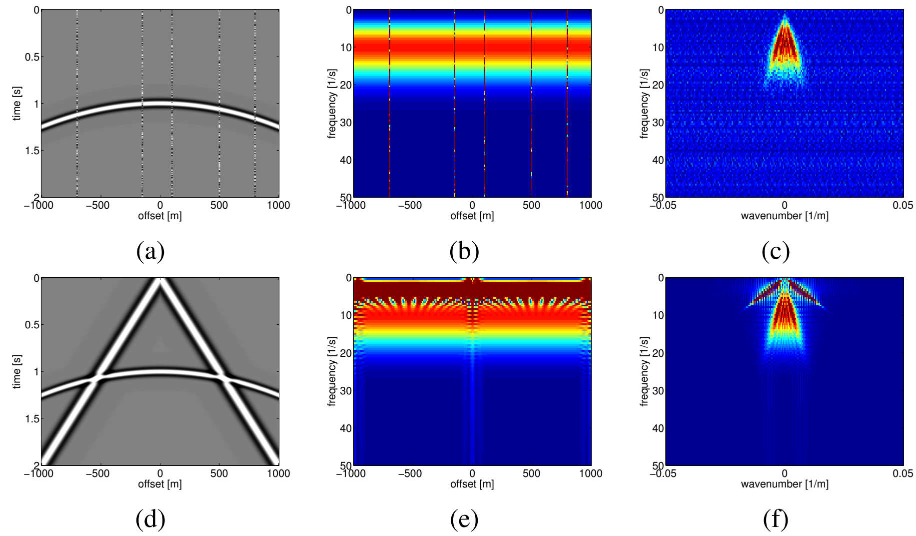



***Figure 1*** *Examples of outliers in different domains. Bad traces (a-c) are localized in both the $(t,h)$ and $(f,h)$, but are spread out in the $(f,k)$ domain. Spurious events (d-f) are localized in the $(t,h)$ and $(f,k)$ domain and are spread out in the $(f,h)$ domain.*

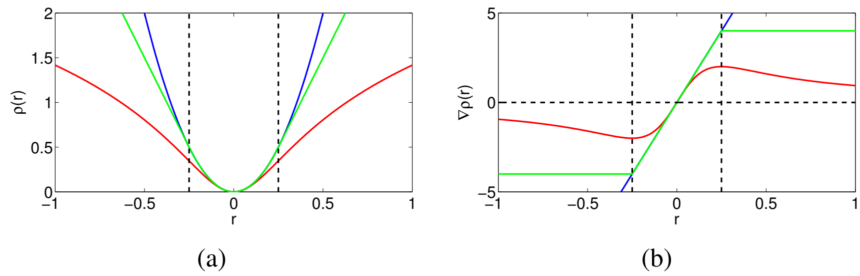



***Figure 2*** *The Student's t (red), Huber (blue) and least-squares penalties for $\alpha = 0.25$ are shown in (a); the corresponding influence functions are shown in (b). The robust penalties behave like the least-squares penalty for residuals $|r| < 0.25$.*

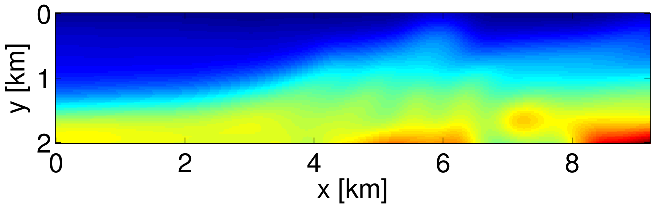
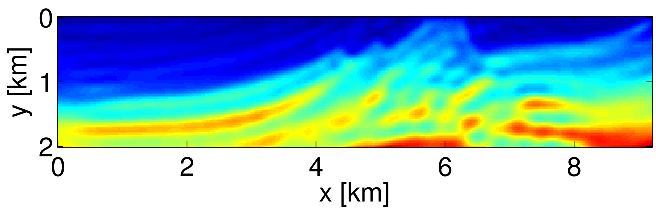
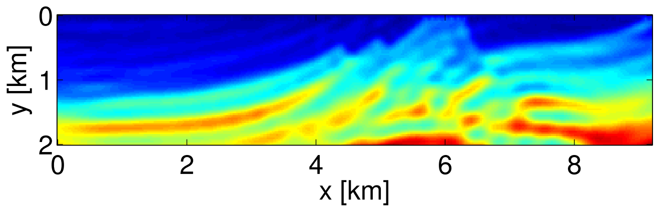



***Figure 3*** *Initial model used for Marmousi test*

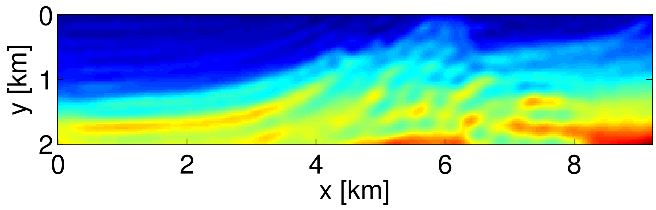



$((f,h),\, \alpha = 1)$

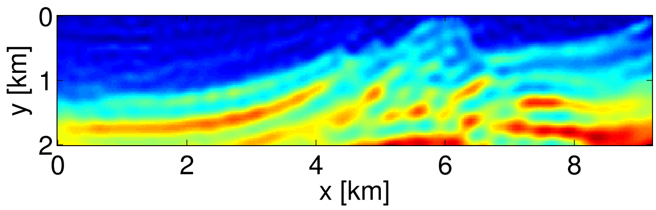



$((f,k),\, \alpha = 0.1)$

***Figure 4*** *Results for robust FWI in spatial and wavenumber domain without noise*

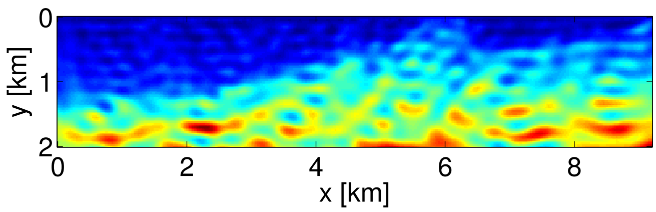



$((f,h),\, \alpha = 1)$

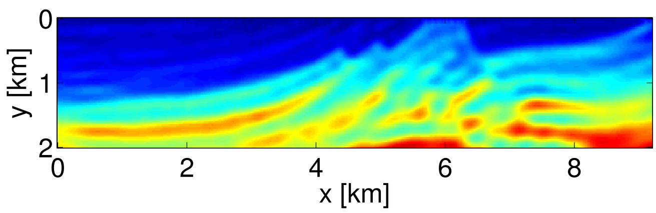



$((f,k),\, \alpha = 1)$

***Figure 5*** *Results for robust FWI in spatial and wavenumber domain with spiky noise in the spatial domain.*




$((f,h),\, \alpha = 1)$




$((f,k),\, \alpha = 0.01)$

***Figure 6*** *Results for robust FWI in spatial and wavenumber domain with a moderate amount of spiky noise in the wavenumber domain.*