

W022

## The Diagonalator, an Alternative Cost Functional for Wave-equation Inversion

P.P. Moghaddam\* (Delft University of Technology) & W.A. Mulder (Shell Global Solutions International BV / Delft University of Technology)

### SUMMARY

---

The classic least-squares cost functional for full waveform inversion suffers from local minima due to loop skipping in the absence of low frequencies in the seismic data. Velocity model building based on subsurface spatial or temporal shifts may break down in the presence of multiples in the data. Cost functionals that translate this idea to the data domain, with offset- or time-shifts, can handle multiples. An earlier data-domain formulation suffered from cross-talk between events. Here, we present a multishot extension that should be less sensitive to cross-talk. It has the property of an annihilator, similar to the functional used for velocity analysis with extended images based on subsurface shifts. However, since it operates in the data domain, it should be able to handle multiples. For 2-D models with line acquisition, the proposed functional applies a singular-value decomposition on the observed data and uses the eigenvectors to build a data panel that should be diagonal in the correct velocity model, but has significant off-diagonal entries in the wrong model. By minimizing these off-diagonal entries or maximizing the main diagonal, the correct model should be found. We therefore named it the diagonalator. We present initial tests on a simple, horizontally layered velocity model.

## Introduction

Full waveform inversion requires a sufficiently accurate starting model for the least-squares misfit functional to converge to the global minimum. The synthetic waveforms should match the observed data to within less than a half-cycle (Bunks et al., 1995; Chauris et al., 2008), otherwise the minimization may encounter a local minimum from which it cannot recover with the usual gradient-based optimization methods. The cause of this problem is the lack of low frequencies in seismic data. Alternatives for the least-squares objective function with the L2-norm (Tarantola and Valette, 1982), such as the Huber norm (Huber, 1973; Guitton and Symes, 2003) or hybrid approaches (Bube and Langan, 1997; Brossier et al., 2010), are more robust in the presence of large isolated and non-Gaussian errors, but still suffer from the same cycle skipping problem as observed with the L2-norm.

Here, we propose a new objective function which should be less sensitive to loop skipping. It is motivated by the work of van Leeuwen and Mulder (2008b), who proposed offset- or time-shifts in the data domain as an alternative to the subsurface space- or time-shifts introduced for wave-equation migration velocity analysis (Faye and Jeannot, 1986; MacKay and Abma, 1992; Rickett and Sava, 2002; Sava and Fomel, 2006; Shen et al., 2003). The latter may fail in the presence of multiples (Mulder and van Leeuwen, 2008). In a NMO setting, the data-domain correlation functional can indeed handle multiples (van Leeuwen and Mulder, 2008a). However, the approach suffers from cross-talk between events (van Leeuwen and Mulder, 2010), requiring the use of a Gabor window. The question remained if the equivalent of an *annihilator* (Symes, 2008) can be constructed for the data-domain velocity-analysis method. The proposed objective function represents a step in that direction. We analyse its convexity and compare it to the classic least-squares formulation with the L2-norm by carrying out a one-parameter scan of the cost functions between two models. We then consider a simple 2-layer model and study the convergence for data with no frequencies below 8Hz. Finally, we compare errors in the reconstructed model for both functionals after a fixed number of iterations for various frequency subsets of the data .

## Method

Given a seismic survey with  $n_s$  shots and  $n_r$  receivers, we build a 2-D matrix per frequency containing normalized data per shot. If all the receivers are co-located for each shot, the  $n_r \times n_s$  data matrix  $\mathbf{A}(\omega)$  has the form

$$\mathbf{A}(\omega) = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n_s} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n_s} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n_r,1} & r_{n_r,2} & \cdots & r_{n_r,n_s} \end{pmatrix} \quad (1)$$

with  $r_{i,j}(\omega)$  is the  $i^{th}$  receiver signal corresponding to the  $j^{th}$  source at angular frequency  $\omega$ . If receivers are not co-located for each source, as in a marine acquisition, some elements of the matrix  $\mathbf{A}$  are zero (Berkhout, 2006). The singular-value decomposition (SVD) of the data matrix is  $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ , where the unitary matrices  $\mathbf{U}$  of size  $n_r \times n_m$  with  $n_m = \min(n_s, n_r)$  and  $\mathbf{V}$  of size  $n_m \times n_s$  contain the left and right eigenvectors, respectively. The diagonal matrix  $\mathbf{S}$  of size  $n_m \times n_m$  contains the singular values of  $\mathbf{A}$ .

Given the observed data, we can order them per frequency into matrices  $\mathbf{A}_{\text{obs}}(\omega)$  and perform an SVD on each, providing  $\mathbf{U}_{\text{obs}}$ ,  $\mathbf{V}_{\text{obs}}$  and  $\mathbf{S}_{\text{obs}}$ . The matrix  $\mathbf{S} = \mathbf{U}_{\text{obs}}^H \mathbf{A} \mathbf{V}_{\text{obs}}$  for data  $\mathbf{A}$  equals  $\mathbf{S}_{\text{obs}}$  in the correct velocity model, so it is diagonal. Otherwise, there may be significant contributions on the off-diagonals. We therefore define a cost function that penalizes off-diagonal entries and the more so if they are further away from the diagonal. Alternatively, one might maximize the energy on and close to the main diagonal with a gaussian weighting function. An example of a cost function for minimization is

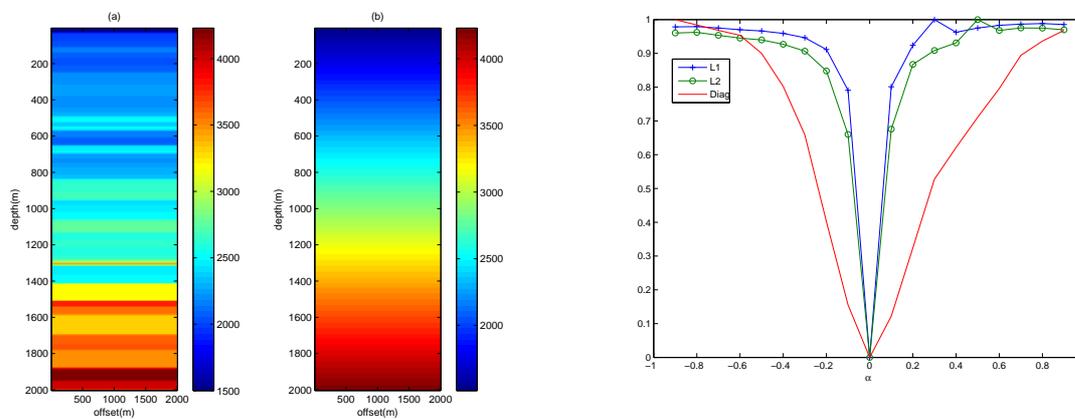
$$J^{\text{Diag}} = \sum_{i=1}^{n_m} \sum_{j=1}^{n_m} w_{ij} |S_{ij}|^2, \quad (2)$$

with  $w_{ij}$  an increasing function of  $|i - j|$ , the distance to the main diagonal, for instance,  $w_{ij} = |i - j|^p$

with  $p > 0$ , or, more severely,  $w_{ij} = \exp(|i - j|)$ . This approach generalizes the offset-shift approach for a single shot in the data domain introduced by van Leeuwen and Mulder (2008b). We call it the diagonalator.

### Comparison to the least-squares cost function

We consider a series of numerical examples to compare the behaviour of the new functional to that of the least-squares cost functional with L1 or L2-norm. We start with a simple convexity test. Given the two horizontally invariant models in Figure 1 with velocities  $v_a(z)$  and  $v_b(z)$ , we consider a range of models  $v(\alpha) = (1 - \alpha)v_a + \alpha v_b$ . The “observed” data are computed in model  $a$ , whereas model  $b$  increased linear in depth. We then computed synthetic seismograms for 20 values of  $\alpha$  ranging from  $-1$  to  $1$  and determined the values of the diagonalator as well as the least-squares functional, using either the L1 or the L2 norm for the latter. All computations were carried out with a 2-D frequency-domain finite-difference code written in Matlab<sup>®</sup>, using frequencies between 8 and 30 Hz. Figure 2 displays the values of the cost functionals as a function of  $\alpha$ . Clearly, the new functional is more convex and has a wider basin of attraction than the least-squares functional, thereby imposing less strict demands on the accuracy of the starting model.



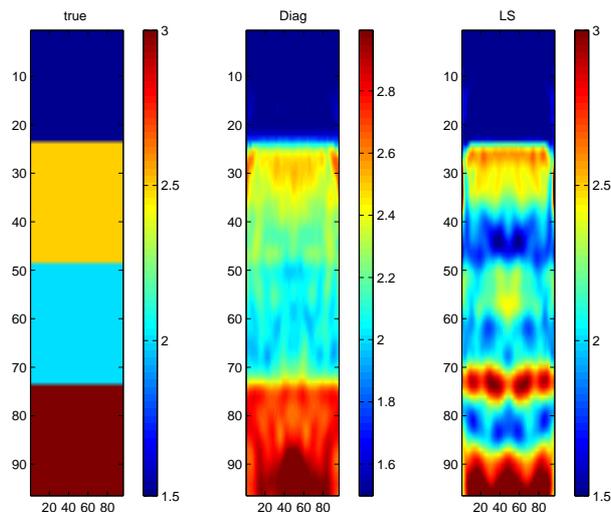
**Figure 1** Test models for evaluation of the new objective function with (a) the true velocity model and (b) a model linear in depth.

**Figure 2** Comparison between the diagonalator and the L1 and L2 least-squares objective functions as a function of the perturbation parameter  $\alpha$ .

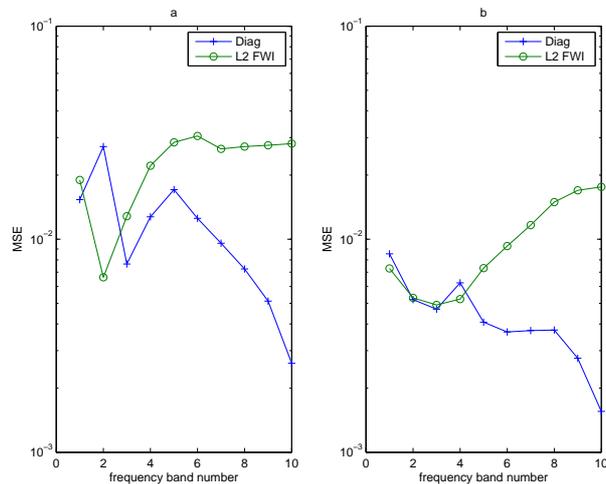
As a second test, we studied the convergence behaviour as a function of frequency band for the simple layered model shown in Figure 3(a). We divided the frequency range between 0 and 20 Hz into bands of 2 Hz wide, containing 4 frequencies each, resulting in 10 different, equally spaced bands. We then performed gradient-based inversion, starting from the same smooth version of the true model for each band and used the 4 frequencies simultaneously. In this way, we obtained 10 different velocity models. To compare the quality of the results, we calculated the average relative model error,  $MSE = \text{avg}((\hat{v} - v)^2 / v^2)$ , with avg the pointwise average,  $\hat{v}$  the inversion result after 30 iterations of L-BFGS minimization (Nocedal and Wright, 2006) and  $v$  the true model. Figure 4 displays the average relative model error for each of the 10 frequency bands, starting from a smooth (a) or from a sharper (b) initial model. We observe that the diagonalator and the least-squares functional produce errors of the same size in the low-frequency bands, where reliable seismic data are more difficult to obtain in practice, and that the diagonalator performs much better in the higher-frequency bands, by an order of magnitude.

As a third test, we started the inversion from a velocity model that was linear in depth and used a multi-scale approach (Ravaut et al., 2004) with data between 8 and 20 Hz. The result for the diagonalator in Figure 3(b) is clearly better than then the least-squares result in Figure 3(c). No regularization was applied, nor did we impose lateral invariance, to emphasize the behaviour of the underlying cost functional. The least-squares inversion result resembles reverse-time migration, the diagonalator recovers

the larger-scale trends.



**Figure 3** (a) true model, (b) diagonalator inversion result, (c) least-squares reconstruction.



**Figure 4** Average relative model error for 10 frequency bands with a 2-Hz bandwidth each, from 0 to 20 Hz, for smooth (a) and sharp (b) initial models.

## Conclusions

We have proposed a new objective function for full waveform inversion, which we call the diagonalator. It is based on the singular-value decomposition of the observed data, ordered in a specific way. If the corresponding matrices with eigenvectors are applied to the modelled data, the resulting matrix should be diagonal in the correct velocity model. The cost functional penalizes off-diagonal entries. We have demonstrated that the objective function is superior to the classic least-squares cost functional, either with the L1 or L2 norm, in terms of convexity and accuracy of the reconstructed model.

## Acknowledgements

Shell GameChanger provided financial support for this project.

## References

Berkhout, A.J. [2006] Seismic processing in the inverse data space. *Geophysics*, **71**(4), A29–A33, doi: 10.1190/1.2217727.

- 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, doi:10.1190/1.3379323.
- Bube, K.P. and Langan, R.T. [1997] Hybrid  $\ell_1/\ell_2$  minimization with applications to tomography. *Geophysics*, **62**(4), 1183–1195, doi:10.1190/1.1444219.
- Bunks, C., Saleck, F.M., Zaleski, S. and Chavent, G. [1995] Multiscale seismic waveform inversion. *Geophysics*, **60**(5), 1457–1473, doi:10.1190/1.1443880.
- Chauris, H., Noble, M. and Taillandier, C. [2008] What initial velocity model do we need for full waveform inversion? *70th EAGE Conference & Exhibition*, Expanded Abstracts.
- Faye, J. and Jeannot, J. [1986] Prestack migration velocities from focusing depth analysis. *SEG Expanded Abstracts*, **5**, 438–440, doi:10.1190/1.1893053.
- Guitton, A. and Symes, W.W. [2003] Robust inversion of seismic data using the Huber norm. *Geophysics*, **68**(4), 1310–1319, doi:10.1190/1.1598124.
- Huber, P.J. [1973] Robust regression: Asymptotics, conjectures, and Monte Carlo. *The Annals of Statistics*, **1**(5), 799–821.
- MacKay, S. and Abma, R. [1992] Imaging and velocity analysis with depth-focusing analysis. *Geophysics*, **57**(12), 1608–1622, doi:10.1190/1.1443228.
- Mulder, W.A. and van Leeuwen, T. [2008] Automatic migration velocity analysis and multiples. *SEG Technical Program Expanded Abstracts*, **27**(1), 3128–3132, doi:10.1190/1.3063996.
- Nocedal, J. and Wright, S. [2006] *Numerical Optimization, 2nd edition*. Springer Series in Operations Research and Financial Engineering, Springer-Verlag, Berlin, New-York, ISBN 978-0-387-30303-1, doi:10.1007/978-0-387-40065-5.
- Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A. and Dell'Aversana, P. [2004] Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International*, **159**(3), 1032–1056, doi:10.1111/j.1365-246X.2004.02442.x.
- Rickett, J.E. and Sava, P.C. [2002] Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics*, **67**(3), 883–889.
- Sava, P. and Fomel, S. [2006] Time-shift imaging condition in seismic migration. *Geophysics*, **71**(6), S200–S217.
- Shen, P., Symes, W.W. and Stolk, C.C. [2003] Differential semblance velocity analysis by wave-equation migration. *SEG Expanded Abstracts*, **22**, 2132–2135, doi:10.1190/1.1817759.
- Symes, W.W. [2008] Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, **56**(6), 765–790, doi:10.1111/j.1365-2478.2008.00698.x.
- Tarantola, A. and Valette, B. [1982] Generalized nonlinear inverse problems solved using the least square criterion. *Reviews of Geophysics*, **20**(2), 219–232, doi:10.1029/RG020i002p00219.
- van Leeuwen, T. and Mulder, W.A. [2008a] Velocity analysis with multiples – NMO modeling for layered velocity structures. *SEG Technical Program Expanded Abstracts*, **27**(1), 1925–1929, doi:10.1190/1.3059275.
- van Leeuwen, T. and Mulder, W. [2008b] Velocity analysis based on data correlation. *Geophysical Prospecting*, **56**(6), 791–803, doi:10.1111/j.1365-2478.2008.00704.x.
- van Leeuwen, T. and Mulder, W. [2010] A comparison of seismic velocity inversion methods for layered acoustics. *Inverse Problems*, **26**(1), 015008, doi:10.1088/0266-5611/26/1/015008.