# Robust elastic frequency-domain full-waveform inversion using the $L_1$ norm

R. Brossier,[1] S. Operto,[2] and J. Virieux[3]

Received 12 June 2009; accepted 21 September 2009; published 23 October 2009.

[1] Elastic frequency-domain full-waveform inversion is here applied to a realistic synthetic offshore model to study the behavior of the least-squares and least-absolute-value norms, in the presence of noisy multi-component seismic data. As expected, the inversion results show that the $L_2$ norm is highly sensitive to non-Gaussian errors in the data and gives rise to high amplitude artifacts in the reconstructed models. The $L_1$ norm shows more robust behavior, whatever the noise characteristics, and allows convergence towards acceptable models, for both the compressional-wave and shear-wave velocity models. This study highlights the sensitivity of the full-waveform inversion results for the chosen norm in the case of noisy data. The marginally used $L_1$ norm reveals a robust alternative to the $L_2$ norm for frequency-domain full-waveform inversion of weakly redundant data. **Citation:** Brossier, R., S. Operto, and J. Virieux (2009), Robust elastic frequency-domain full-waveform inversion using the $L_1$ norm, *Geophys. Res. Lett.*, *36*, L20310, doi:10.1029/2009GL039458.

## 1. Introduction

[2] The physical properties of the Earth are often estimated through seismic-wave analysis. Wave arrival-time information is generally used to determine the tomography of the Earth. In the 1980s, the full-waveform inversion (FWI) method was introduced by *Tarantola* [1984] to exploit the full information contained in the seismograms, in terms of the full waveform, and to infer high-resolution models of the subsurface.

[3] Originally developed in the time domain [*Tarantola*, 1984, 1987], FWI has become tractable and popular since the pioneering work of *Pratt and Worthington* [1990] and *Pratt* [1990] in the frequency domain, and is now investigated more, particularly for active seismic surveys at various scales. Computationally efficient frequency-domain FWI was designed by limiting the inversion to a few discrete frequencies, taking advantage of the redundant wavenumber coverage provided by wide-aperture surveys [*Sirgue and Pratt*, 2004; *Brenders and Pratt*, 2007]. This wavenumber redundancy can be partly sacrificed by using a judicious subset of frequencies in the inversion, at the expense of the signal-to-noise ratio of the reconstructed models. FWI potentially provides high-resolution quantitative images of physical parameters, but suffers from two main difficulties. The first is related to the computational cost of the forward problem; namely, the numerical resolution of the wave equation in heterogeneous media for multiple sources. The second difficulty is related to the ill-posedness and the non-linearity of the inverse problem, which is generally formulated as a least-squares local optimization [*Tarantola and Valette*, 1982], to manage the numerical cost of the forward problem. The ill-posedness and the non-linearity of the FWI arise mainly from the limited accuracy of the starting model in the framework of local optimization approaches, the incomplete illumination of the subsurface provided by conventional seismic surveys, the limited bandwidth sources and presence of noise. Several hierarchical multiscale strategies that proceed from low frequencies to higher frequencies [*Pratt and Worthington*, 1990; *Bunks et al.*, 1995; *Sirgue and Pratt*, 2004; *Brossier et al.*, 2009a], have been proposed to mitigate the non-linearity of the inverse problem.

[4] The noise footprint in seismic imaging is conventionally mitigated by stacking highly redundant multifold data. Mitigating the sensitivity of the inversion to noise is however a key issue in efficient frequency-domain FWI, where the data redundancy is partly sacrificed for computational efficiency. The least-squares objective function remains the most commonly used, although it suffers from poor robustness when dealing with large isolated and non-Gaussian errors. Other norms should therefore be considered. The $L_1$ norm was introduced in time-domain FWI by *Tarantola* [1987] and *Crase et al.* [1990], and has been shown to be weakly sensitive to noise. *Djikpéssé and Tarantola* [1999] used this norm successfully to invert field data from the Gulf of Mexico. Alternatives like the Huber criterion [*Huber*, 1973; *Guitton and Symes*, 2003; *Ha et al.*, 2009], the *sech* criterion [*Crase et al.*, 1990; *Monteiller et al.*, 2005] and the hybrid $L_1/L_2$ criterion [*Bube and Langan*, 1997] can also be considered. All these criteria behave as the $L_2$ norm for small residuals and as the $L_1$ norm for large ones. A threshold that needs to be provided defines where the transition between these two different behaviors takes place. These hybrid criteria are efficient for dealing with outliers in data. However, they assume Gaussian statistics as soon as the $L_2$ norm is used, leading to the same limitations as for the pure $L_2$ norm in the presence of ambient noise if the threshold is not chosen correctly.

[5] In this study, we assess two-dimensional (2D) elastic, frequency-domain FWI from noisy, multi-component seismic data of the oil & gas field of Valhall computed in synthetic velocity models. The results clearly demonstrate the lack of robustness of the classic least-squares norm, whereas the $L_1$ norm that has seen marginal use in the FWI

[1] Géoazur, Observatoire de la Côte d'Azur, Université de Nice Sophia Antipolis, CNRS, Valbonne, France.
[2] Géoazur, Observatoire de la Côte d'Azur, Université de Nice Sophia Antipolis, CNRS, Villefranche-sur-Mer, France.
[3] LGIT, Université Joseph Fourier, Grenoble, France.


0094-8276/09/2009GL039458$05.00

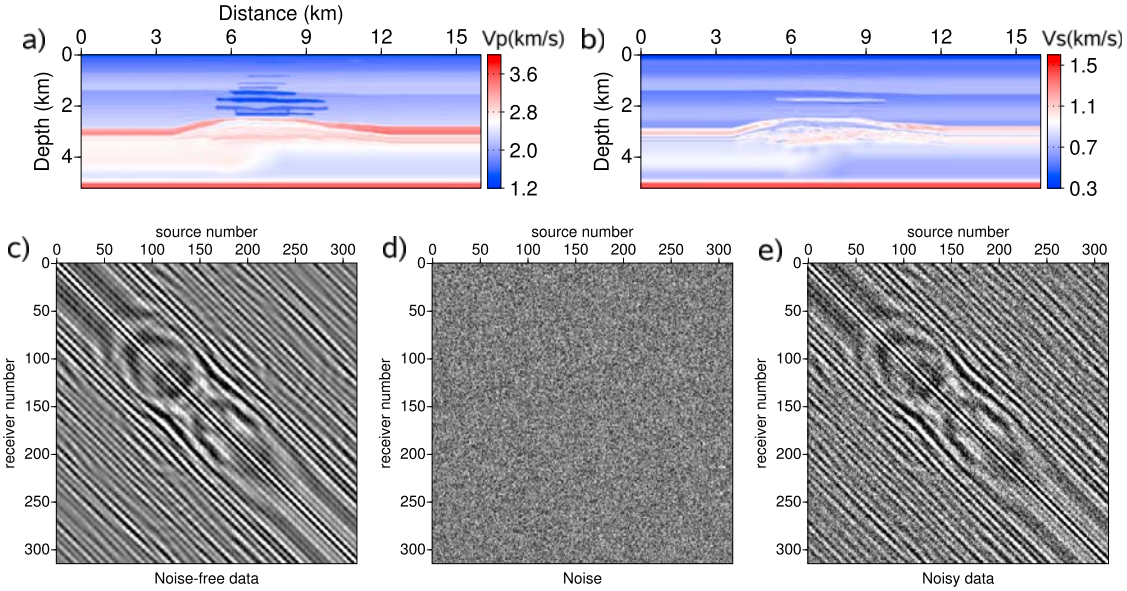


**Figure 1.** Synthetic true Valhall models for (a) $V_P$ and (b) $V_S$. Real part of the hydrophone data for frequency (4 + 10 $i$) Hz. (c) Noise-free data, (d) added noise and (e) noisy data.

community, is shown to be very robust, even in the case of highly decimated data.

## 2. Theory and Algorithm

### 2.1. Least-Squares Norm

[6] The least-squares criterion provides the most usual framework for the development of frequency-domain FWI [*Pratt and Worthington*, 1990; *Pratt*, 1990]. The $L_2$ functional is usually written in the following form:

$$\mathcal{C}_{L_2}^{(k)} = \frac{1}{2}\Delta\mathbf{d}^t\mathbf{S}_d^t\mathbf{S}_d^*\Delta\mathbf{d}^*, \tag{1}$$

where $\Delta\mathbf{d} = \mathbf{d_{obs}} - \mathbf{d_{calc}^{(k)}}$ is the data misfit vector, the difference between the observed data $\mathbf{d_{obs}}$ and the modeled data $\mathbf{d_{calc}^{(k)}}$ computed in the model $\mathbf{m}^{(k)}$. $k$ is the iteration number of the non-linear iterative inversion. Superscript * indicates the conjugate and $\mathbf{S}_d$ is a diagonal weighting matrix applied to the misfit vector to scale the relative contributions of each of its components.

[7] Differentiation of $\mathcal{C}_{L_2}^{(k)}$ with respect to the model parameters gives the following expression of the gradient:

$$\mathcal{G}_{L_2}^{(k)} = \mathcal{R}\left[\mathbf{J}^t\mathbf{S}_d^t\mathbf{S}_d^*\Delta\mathbf{d}^*\right], \tag{2}$$

where $\mathbf{J}$ is the Fréchet derivative matrix. The gradient can be derived from the adjoint-state method that allows the computing of $\mathcal{G}_{L_2}^{(k)}$ without explicitly forming $\mathbf{J}$ [*Plessix*, 2006], giving the expression:

$$\mathcal{G}_{m_i\,L_2}^{(k)} = \mathcal{R}\left[\mathbf{v}^t\left[\frac{\partial\mathbf{A}^t}{\partial m_i}\right]\mathbf{A}^{-1}\mathbf{S}_d^t\mathbf{S}_d^*\Delta\mathbf{d}^*\right], \tag{3}$$

where the incident wavefield $\mathbf{v}$ is linearly related to the source $\mathbf{s}$ through the forward problem operator $\mathbf{A}$ : $\mathbf{Av} = \mathbf{s}$.

[8] The gradient can be seen as a weighted zero-lag convolution between the incident wavefield $\mathbf{v}$ emitted by the seismic source and the residual wavefield back-propagated from the receiver positions $\mathbf{A}^{-1}\,\mathbf{S}_d^t\mathbf{S}_d^*\;\Delta\mathbf{d}^*$.

### 2.2. Least-Absolute-Value Norm

[9] For complex arithmetic numbers, such as frequency-domain data, we can extend the $L_1$ norm developed by *Tarantola* [1987] and *Crase et al.* [1990], written as:

$$\mathcal{C}_{L_1}^{(k)} = \left(\Delta\mathbf{d}^t\mathbf{S}_d^t\mathbf{S}_d^*\Delta\mathbf{d}^*\right)^{1/2}. \tag{4}$$

[10] The gradient $\mathcal{G}_{L_1}^{(k)}$ is given by:

$$\mathcal{G}_{L_1}^{(k)} = \mathcal{R}\left[\mathbf{J}^t\mathbf{S}_d\frac{\Delta\mathbf{d}^*}{|\Delta\mathbf{d}|}\right], \tag{5}$$

assuming that $|\Delta\mathbf{d}| > 0$, considering the machine precision used. For all of the tests that we performed, we never met any case where $|\Delta\mathbf{d}| = 0$.

[11] The $L_1$ and $L_2$ gradients have similar forms, as seen by equations (2) and (5), except that the residual term differs. For the $L_1$ norm, the data residuals are normalized according to their amplitudes, which gives clear insights why the $L_1$ norm is expected to be less sensitive to large residuals.

[12] In the case of real arithmetic numbers, the term $\frac{\Delta\mathbf{d}^*}{|\Delta\mathbf{d}|}$ of expression (5) corresponds to the function *sign* [*Tarantola*, 1987; *Crase et al.*, 1990].

### 2.3. Algorithm

[13] The 2D elastic, frequency-domain FWI algorithm used in this study is described by *Brossier et al.* [2009a]. The reader is thus referred to *Brossier et al.* [2009a] for a complete description of the algorithm. The algorithm embeds three main loops: the outer one is over frequency groups; namely, a set of frequencies simultaneously inverted. The second loop is over time-damping factors that control the amount of information over time that is preserved for inversion in the seismograms. Time damping is

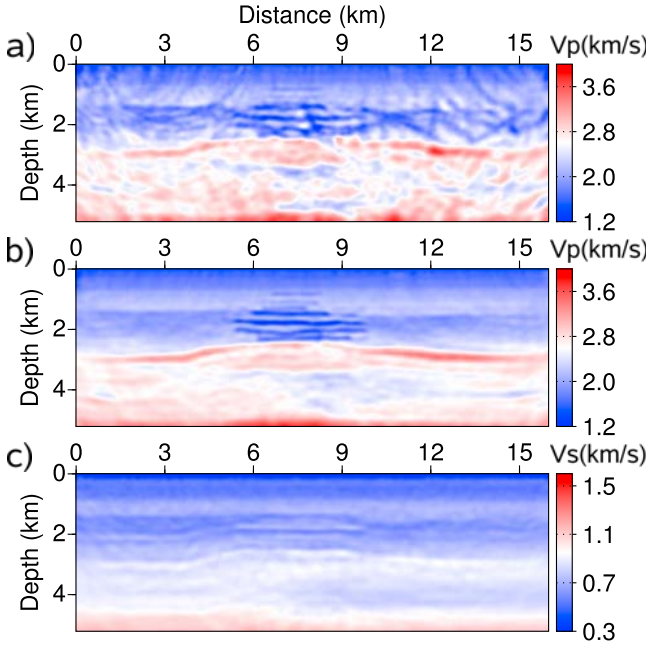


**Figure 2.** Reconstructed models for the test 1 with outliers, for (a) the $V_P$ parameter with $L_2$ and the (b) $V_P$ and (c) $V_S$ parameters with $L_1$ norm.

applied in the frequency-domain modeling by using complex-valued frequencies which is equivalent to damp seismograms in time by $\exp^{-(t-t_0)\gamma}$ where $t_0$ is the first-arrival traveltime. The third loop is over iterations of one frequency group inversion. The two outer nested loops define two hierarchical multiscale levels in the inversion that are helpful to mitigate the non-linearity of the inversion. The forward problem is performed with a discontinuous Galerkin method for solving the elastodynamic equations in the frequency domain [*Brossier et al.*, 2008]. The linear system that results is solved in parallel using the MUMPS LU solver [*Amestoy et al.*, 2006]. The optimization is solved with the quasi-Newton L-BFGS algorithm [*Nocedal*, 1980]. The diagonal of the pseudo-Hessian matrix [*Shin et al.*, 2001] is used as an initial guess for the L-BFGS algorithm for both the $L_1$ and the $L_2$ criteria.

## 3. Numerical Example: Offshore Valhall Model

[14] The numerical example is based on the synthetic Valhall model (Figure 1), which is representative of oil and gas fields in shallow water environments of the North Sea [*Munns*, 1985]. The main targets are the gas cloud in the large sediment layer, and the trapped oil underneath the cap rock, which is composed of chalk, in a deeper part of the model. Gas clouds are easily identified by the low P-wave velocities, whereas their signature is much weaker in the $V_S$ model. The selected acquisition mimics a three-component ocean bottom cable survey [*Kommedal et al.*, 2004]. This environment shows a particularly ill-posed problem for S-wave velocity reconstruction, due to the relatively small shear-wave velocity contrast at the sea bed that prevents recording of significant P-to-S converted waves. A successful inversion requires a multi-step hierarchical strategy in the manner of *Sears et al.* [2008], as developed by *Brossier et al.* [2009b] for noise-free data. We follow the same approach for noisy data:

[15] 1. In the first step, the $V_P$ velocity is reconstructed from the hydrophone data. The forward problem is performed with the elastic discontinuous Galerkin method, but the $V_S$ model is left unchanged during the FWI. The aim of this first stage is to improve the $V_P$ model so as to significantly decrease the P-wave residuals. During this first step, a coarse mesh that is adapted to the $V_P$ wavelength is designed for computational efficiency. In this case, S-wave modeling is affected by numerical dispersion that, however, does not significantly impact on the $V_P$-model reconstruction.

[16] 2. In the second step, the $V_P$ and $V_S$ models are reconstructed simultaneously from the horizontal and vertical components of the geophones. An amplification, with a gain given by the power of 2 of the source-receiver offset, is applied to the data through the matrix $\mathbf{S}_d$. This weighting increases the weight of the intermediate-to-long-offset data at which the converted P-to-S arrivals are recorded.

[17] Five frequencies were inverted successively (2, 3, 4, 5 and 6 Hz) with 3 regularization damping factors applied in cascade for each frequency inversion ($\gamma$ = 2, 0.33, 0.1 $s^{-1}$). Starting models were built by smoothing the true models with a Gaussian filter, the vertical correlation length of which increased linearly from 25 m to 1000 m with depth, and the horizontal one was fixed at 500 m. Ten iterations were performed per damping factor, leading to 30 iterations per frequency inversion. Density is constant and assumed to be known in the inversion. The source is estimated during the inversion by linear inversion [*Pratt*, 1999].

[18] Two tests were performed. For both tests, uniform white noise was introduced into the observed data, computed using the forward-problem engine implemented in the inversion code (the so-called inverse crime). The signal-to-noise ratio was set to 10. Figures 1c–1e show the 4-Hz noise-free and noisy data in the source-receiver domain.

[19] During the first test, we introduced outliers into the data: large errors (noise had been multiplied by 20) were introduced randomly in one trace out of one hundred, to simulate a poorly preprocessed data set. The resulting noise is consequently no more uniform for this test. The $V_P$ models obtained after the first inversion step with the $L_2$ and $L_1$ norms are shown in Figure 2. The $L_1$ norm provides an accurate $V_P$ model, whereas the inversion rapidly converged towards a local minimum when the $L_2$ norm was used. For the $L_2$ norm, the inversion was stopped close to the first step because of the failure. However, the second step was performed for the $L_1$ norm and shows good quality of reconstruction for both $V_P$ and $V_S$ models.

[20] A second test was performed without any outliers, and considering only the ambient noise. The $V_P$ and $V_S$ models obtained with the $L_2$ and $L_1$ norms after the second step are shown in Figure 3 and reveal again the improved robustness of the $L_1$ norm.

## 4. Discussion

[21] The use of the $L_2$ and $L_1$ norms in elastic, frequency-domain FWI highlights the sensitivity of the optimization convergence to the noise.

[22] The first test with outliers illustrates the known behavior of the $L_2$ norm in the presence of high-amplitude

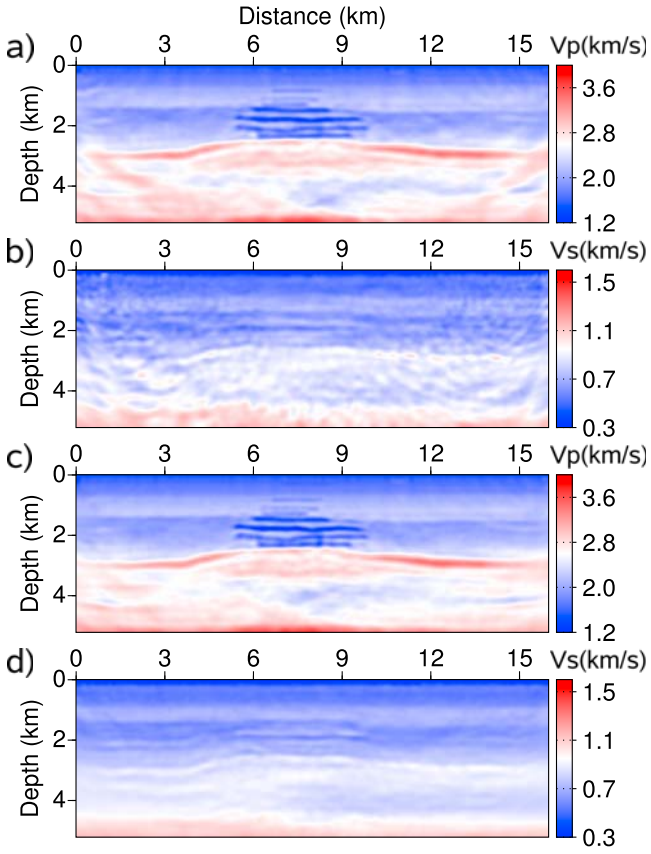


**Figure 3.** Reconstructed $V_P$ and $V_S$ models for the test 2 without outliers, for (a and b) the $L_2$ and (c and d) the $L_1$ norm.

isolated noise. The $L_2$ criterion intrinsically amplifies the weight of the high-amplitude residuals during inversion, hence causing divergence of the optimization if the residuals do not correspond to coherent signals. The $L_1$ norm shows stable behavior in this unfavorable test, because the isolated, high-amplitude outliers have negligible contributions in the final images. This confirms the robustness of this criterion applied to FWI, even when a limited amount of data is considered.

[23] The second test without outliers shows that the $V_P$ reconstruction is robust for both norms, whereas only the $L_1$ norm allowed the successful reconstruct of the $V_S$ model. In this shallow-water environment with low velocity contrasts at the sea bed, the $V_P$ imaging is more linear than the $V_S$ imaging for two main reasons. First, the larger P-wavelengths are less resolving than the S counterparts, and are therefore less sensitive to the inaccuracy of the starting model in the framework of a multiscale reconstruction [*Brossier et al.*, 2009a]. Second, the P-waves dominate the seismic wavefield, whereas the P-to-S waves have a weaker footprint in the data. The limited signature of the S-waves in the data makes the inversion poorly conditioned for the S-wave parameter class, even with noise-free data. *Brossier et al.* [2009b] showed how the hierarchical two-step strategy allow the strengthening of the sensitivity of the inversion to the $V_S$ parameter during the second step, and hence the successful reconstruction of the $V_S$ model with the $L_2$ norm in the case of noise-free data. However, adding noise to the data still contributes to the degradation of the sensitivity of FWI to the P-to-S arrivals. In this case, the two-step strategy implemented with the $L_2$ norm failed to reconstruct the $V_S$ model, and the $L_1$ norm revealed that it was necessary to converge towards acceptable $V_S$ models by mitigating the contribution of the amplitudes in the reconstruction.

## 5. Conclusion

[24] This study presents an application of elastic, frequency-domain FWI of controlled-source synthetic seismic data, to compare the behaviors of the classic $L_2$ and the robust $L_1$ functionals in the presence of noisy data. Our results show first the robust behavior of the $L_1$ norm in the presence of high-amplitude non-Gaussian noise, such as for outliers, whereas the $L_2$ criterion failed to produce acceptable models as expected. Second, the $L_1$ norm allows the successful exploitation of the P-to-S arrivals, which have a relatively weak footprint in the data, to image the $V_S$ velocity model in the presence of white noise. In this case too, the $L_2$ norm failed to reconstruct the $V_S$ model. The robustness of the $L_1$ norm with respect to noise is attributed to the fact that the footprints of the amplitudes are mitigated in the imaging. The $L_1$ norm should be an alternative to the classic $L_2$ norm, particularly when a limited amount of data are to be inverted. This is the case in efficient frequency-domain FWI, where only a limited number of discrete frequencies are inverted independently. The $L_1$ norm can also be useful to invert low-fold passive data, such as teleseismic data lor lithospheric imaging.

[25] **Acknowledgments.** This study is funded by the SEISCOPE consortium (http://seiscope.oca.eu/), which is sponsored by BP, CGG-VERITAS, EXXON-MOBIL, SHELL and TOTAL, and by Agence Nationale de la Recherche (ANR) under project ANR-05-NT05-2-42427. The LU factorization of the impedance matrix was performed with MUMPS (http://graal.ens-lyon.fr/MUMPS/index.html). The mesh generation was performed with the help of TRIANGLE (http://www.cs.cmu.edu/~quake/triangle.html). Access to the high-performance computing facilities of the MESOCENTRE SIGAMM computer centre and IDRIS national (project 082280) centers provided the required computer resources, and we gratefully acknowledge both of these facilities and the support of their staff. Many thanks go to J. Kommendal and L. Sirgue from BP, for providing the elastic synthetic models of Valhall.

## References

Amestoy, P. R., A. Guermouche, J. Y. L'Excellent, and S. Pralet (2006), Hybrid scheduling for the parallel solution of linear systems., *Parallel Comput.*, *32*, 136–156.

Brenders, A. J., and R. G. Pratt (2007), Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data, *Geophys. J. Int.*, *168*, 152–170.

Brossier, R., J. Virieux, and S. Operto (2008), Parsimonious finite-volume frequency-domain method for 2-D P-SV-wave modelling, *Geophys. J. Int.*, *175*, 541–559.

Brossier, R., S. Operto, and J. Virieux (2009a), Seismic imaging of complex onshore structures by two-dimensional elastic frequency-domain full-waveform inversion, *Geophysics*, in press.

Brossier, R., S. Operto, and J. Virieux (2009b), Two-dimensional seismic imaging of the valhall model from synthetic OBC data by frequency-domain elastic full-waveform inversion, *SEG Tech. Program Expanded Abstr.*, in press.

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

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent (1995), Multiscale seismic waveform inversion, *Geophysics*, *60*, 1457–1473.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola (1990), Robust elastic non-linear waveform inversion: Application to real data, *Geophysics*, *55*, 527–538.

Djikpéssé, H. A., and A. Tarantola (1999), Multiparameter $l_1$ norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms, *Geophysics*, *64*, 1023–1035.

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

Ha, T., W. Chung, and C. Shin (2009), Waveform inversion using a back-propagation algorithm and a Huber function norm, *Geophysics*, *74*, R15–R24.

Huber, P. J. (1973), Robust regression: Asymptotics, conjectures, and Monte Carlo, *Ann. Stat.*, *1*(5), 799–821.

Kommedal, J. H., O. I. Barkved, and D. J. Howe (2004), Initial experience operating a permanent 4C seabed array for reservoir monitoring at Valhall, *SEG Tech. Program Expanded Abstr.*, *23*, 2239–2242.

Monteiller, V., J.-L. Got, J. Virieux, and P. Okubo (2005), An efficient algorithm for double-difference tomography and location in heterogeneous media, with an application to the Kilauea volcano, *J. Geophys. Res.*, *110*, B12306, doi:10.1029/2004JB003466.

Munns, J. W. (1985), The Valhall field: A geological overview, *Mar. Pet. Geol.*, *2*, 23–43.

Nocedal, J. (1980), Updating quasi-Newton matrices with limited storage, *Math. Comput.*, *35*(151), 773–782.

Plessix, R.-E. (2006), A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, *167*, 495–503.

Pratt, R. G. (1990), Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method, *Geophys. Prospect.*, *38*, 311–330.

Pratt, R. G. (1999), Seismic waveform inversion in the frequency domain, part I: Theory and verification in a physical scale model, *Geophys. J. Int.*, *64*, 888–901.

Pratt, R. G., and M. H. Worthington (1990), Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method, *Geophys. Prospect.*, *38*, 287–310.

Sears, T., S. Singh, and P. Barton (2008), Elastic full waveform inversion of multi-component OBC seismic data, *Geophys. Prospect.*, *56*, 843–862.

Shin, C., S. Jang, and D. J. Min (2001), Improved amplitude preservation for prestack depth migration by inverse scattering theory, *Geophys. Prospect.*, *49*, 592–606.

Sirgue, L., and R. G. Pratt (2004), Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics*, *69*, 231–248.

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

Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, New York.

Tarantola, A., and B. Valette (1982), Generalized nonlinear inverse problems solved using the least square criterion, *Rev. Geophys. Space Phys.*, *20*, 219–232.

R. Brossier, Géoazur, Observatoire de la Côte d'Azur, Université de Nice Sophia Antipolis, CNRS, 250 rue Albert Einstein, F-06560 Valbonne, France. (brossier@geoazur.unice.fr)

S. Operto, Géoazur, Observatoire de la Côte d'Azur, Université de Nice Sophia Antipolis, CNRS, La Darse, B.P. 48, F-06235 Villefranche-sur-Mer CEDEX, France. (operto@geoazur.obs-vlfr.fr)

J. Virieux, LGIT, Université Joseph Fourier, B.P. 53, F-38041 Grenoble CEDEX 9, France. (jean.virieux@obs.ujf-grenoble.fr)