We 11 10

Fast Full Waveform Inversion with Source Encoding and Second Order Optimization Methods

C. Castellanos Lopez* (Geoazur), L. Métivier (LJK), S. Operto (Geoazur) & R. Brossier (Institut des Sciences de la Terre)

SUMMARY

In the context of full waveform inversion (FWI), second-order optimization methods, which take into account more precisely the effect of the Hessian such as the quasi-Newton l-BFGS method, have shown superior convergence properties than first-order methods.
When using source encoding techniques, the regeneration of the random variables to assemble the sources at each FWI iteration plays a crucial role since it helps to reduce the so-called cross talk noise produced by the encodings. However, it is not clear how to combine the l-BFGS algorithm and encoding methods because, strictly speaking, l-BFGS needs previous iteration estimations, thus prohibiting the regeneration of the code at each iteration. We study how to combine second-order optimization methods with encoding techniques, considering two truncated matrix-free Newton algorithms (Gauss Newton and full Newton) and the l-BFGS algorithm with periodic restarts and we apply our method on the 2004 BP salt model.

Introduction

Full waveform inversion (FWI) is a non-linear ill-posed inverse problem aimed to reconstruct with high resolution the physical parameters of the earth's subsurface, from the restricted knowledge of some seismic observations at receiver's positions on the earth's surface (Tarantola, 1984). The forward problem models the wavefield $u(x)$ for a choice of physical parameters $m(x)$ through the wave equation operator A. In the frequency domain this leads to a linear system $A(m)u = s$, where $s(x)$ is a source, and $x \in \Omega$, where $\Omega \subset \mathbb{R}^2$ is the domain where the solution of the wave equation is found. The associated discrete inverse problem is the minimization of the ℓ_2 norm of the residuals,

$$\min_m \phi \quad = \quad \min_m \sum_{\omega=1}^{N_f} \sum_{i=1}^{N_s} \sum_{r=1}^{N_r} \|P_{r,i} u(m)^i_{\omega,r} - d^i_{\omega,r}\|_2^2 + \lambda \|\nabla m\|_2^2, \tag{1}$$

where the operator P restricts the solution of the forward problem to the space of the measured data d. We consider N_s sources, N_f frequencies, N_r receivers, and a Tikhonov regularization term whose weight is given by λ. Because of the high numerical cost of the forward modelling, one considers simultaneous sources with arbitrary random encodings which allows to reduce the cross-talks coming from connections between one source and acquisition systems of the other sources.

Not much emphasis has been done on the impact of the optimization algorithm when using source encoding techniques. In the context of FWI, second-order optimization methods, which take into account the Hessian, have shown superior convergence properties than first-order methods, such as steepest descent. Alternatively, the quasi-Newton l-BFGS method where the influence of the Hessian is estimated through previous models and gradients has been shown to be effective. When using encoding techniques, the regeneration of the random variables to assemble the sources plays a crucial role since it helps to reduce the so-called cross talk noise produced by the encodings. However, it is not clear how to combine the l-BFGS algorithm and encoding methods because, strictly speaking, l-BFGS needs previous iteration estimations, thus prohibiting the regeneration of the code at each iteration. Therefore, we are interested in studying how to combine second-order optimization methods with encoding techniques. Two matrix-free truncated Newton algorithms (Gauss Newton (GN) and full Newton (FN), (Métivier et al., 2012)) and the l-BFGS algorithm with periodic restarts are chosen for our study.

Method

Popular is the gathering of super sources by summing individual sources encoded in such a way as to make the sum incoherent (Krebs et al., 2009; Neelamani et al., 2008; **?**; Ben Hadj Ali et al., 2011; van Leeuwen et al., 2011). Encoding methods include phase shifts, time shits, convolution with random variables amongst others. For most of the encoding methods, the problem lies on a slow convergence because a large number of incoherent sums must be performed to mitigate the cross talk noise. Other speed up techniques, that do not rely on random encodings have also been proposed (Habashy et al., 2011; van Leeuwen and Herrmann, 2012). Despite the intensive work done, there are still issues that remain challenging. One of the main difficulties in all encoding techniques arises when considering noisy data (Krebs et al., 2009; van Leeuwen et al., 2011).

We will only consider the approach in which a limited number of super sources $\tilde{s}$ are constructed by summing individual encoded sources s^i (Haber et al., 2010), $\tilde{s}^q_{\omega,r} = \sum_{i=1}^{N_s} s^i_{\omega,r} \cdot \gamma^q_i$, where γ^q is an encoding random vector satisfying $Cov(\gamma^q, \gamma^p) = \mathbb{I}$, for $q \in [1, \ldots, K]$ where K is the number of super sources. In a similar fashion, the observed data are also encoded, and the same holds for the computed wavefield since it is linear with respect to the sources. The misfit function can then be rewritten as

$$\tilde{\phi} \quad = \quad \sum_{\omega=1}^{N_f} \sum_{q=1}^{K} \sum_{r=1}^{N_r} \|P_{r,q} \tilde{u}^q_{\omega,r} - \tilde{d}^q_{\omega,r}\|_2^2 + \lambda \|\nabla m\|_2^2. \tag{2}$$

Whenever $K < N_s$ the number of direct problems solved is reduced. We will measure the speed-up S as a percentage $S = \left(1 - \frac{N_e}{N_o}\right) \cdot 100$, where N_e is the total number of direct problems solved when using the encoding technique and N_o using all the individual sources. When $K << N_s$, the descent direction found in each iteration will be noisy because $\tilde{\phi}$ will differ greatly from ϕ, but each iteration will be performed at a very low cost allowing for many steps to be taken which, in average, tend towards an accurate direction (van Leeuwen et al., 2011).

Leaving source encoding techniques aside, the most suitable second-order optimization method for a given problem in FWI is case dependent, in which factors such as the frequencies to be inverted, the initial model and the model properties play a key role. GN, FN and l-BFGS methods differ in the way the Hessian is approximated (Métivier et al., 2012). While l-BFGS method performs a finite difference approximation with M gradients and models, truncated GN and FN methods approximately solve the linear system $H\Delta m = -G$, where H is the Hessian, G the gradient and Δm the model perturbation. Gauss-Newton method is an approximation of the FN method, where the term containing the second order derivatives of the wavefield accounting for double scattering in FWI, are ignored. Numerically, the linear system is solved with an iterative solver where the product of the Hessian with a vector is efficiently computed with a matrix-free algorithm based on a second-order adjoint development (Métivier et al., 2012). Even though truncated Newton methods require solving a greater number of direct problems per iteration in comparison to l-BFGS because of the additional effort invested in solving the linear system, the compromise is beneficial if the convergence speed is increased. Another major difference between the two optimization algorithms lies in the fact that truncated Newton methods construct the second derivative of the misfit function using only the information of the current iteration, contrary to the l-BFGS method. When including source encoding, this dissimilarity has an impact since the encoding random variables γ can be changed in each iteration step of the minimization process in truncation Newton methods, as opposed to l-BFGS method where, to be consistent, the random variables should be kept the same throughout the previous M iterations. Since changing the encoding aids in making the cross-talk less coherent, we have performed periodic restarts of $l-BFGS$, every n iterations ($n = M$). At each restart, a new encoding vector is drawn and the Hessian estimation is restarted, allowing us to change the encoding periodically, while still being able to use the l-BFGS method. Given these differences, one may wonder which is the best optimization method to minimize the encoded misfit (2).

Numerical Tests

For our numerical tests, we use the 2D BP 2004 velocity model (Figure 1) which contains high contrasts in the wave velocities between salt bodies and the underlying sedimentary layers, generating high amplitude multiple reflections. As shown in Métivier et al. (2012), the presence of high-amplitude multiple reflections emphasizes the importance of an accurate estimation of the full Hessian. We solve the acoustic wave equation in the frequency domain for the reconstruction of the P wavespeed. There are 62 sources and 248 receivers per source, located on the surface. We simultaneously invert 9 frequencies from 1 Hz to 9 Hz, with a 1 Hz frequency interval. This single frequency group simplifies the speed up analysis, although imaging below the salt may be hampered by cycle skipping. The starting velocity model (Figure 1) is a smoothed version of the true velocity model. We keep the number of super sources low ($K = 3$), and, when the sources are encoded, l-BFGS is restarted every $M = 5$ iterations.

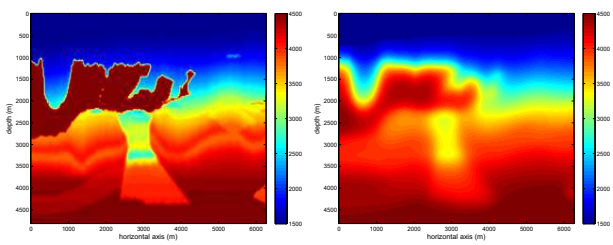

Figure 1 *True Vp velocity model (left) and initial velocity model (right).*

For a fair comparison between models obtained with and without encoding, we show the misfit function reduction attained with both sets of models when the full set of individual sources is used (Figures 2 and 3). We also monitor the evolution of the relative model error $e = \frac{1}{N}\sum_{i=1}^{N}(|m_i^{true} - m_i|/m_i^{true})$, where N is the number of points in the grid. However both measures must be interpreted with care, specially under the influence of noise or data encoding, because a noisy model with high resolution may have the same error as a smoother less noisy one. We use the diagonal terms of the pseudo-Hessian (Shin et al., 2001) as a preconditioner for the truncated Newton methods and as an initial guess of the Hessian in l-BFGS. Three main parameters have to be set before inversion, namely, the regularization parameter λ, a damping factor β in the diagonal pseudo-Hessian approximation to prevent division by very small values, and the maximum number of iterations in the truncated Newton methods to solve the linear system (Métivier et al., 2012), or the restart interval for l-BFGS. We chose the parameters as uniform as possible, for the different methods. The same damping factor β is used for all the tests. For the tests without encoding the λ value used with l-BFGS is one order of magnitude higher than that of the Newton methods, because limiting the internal iterations of the GN and FN has a regularizing effect (Kaltenbacher et al., 2008). For the tests with encoding, all optimization methods share the same λ because restarting l-BFGS has a regularization effect by periodically correcting the directions that may have accumulated errors due to noise in the gradient (Schittkowski, 2011). Using all the sources sequentially for data without noise we found that all optimization methods have similar convergence rates (Figure 2). When the sources are encoded, l-BFGS method has the fastest convergence rate ($S = 89\%$), followed by GN ($S = 67\%$). When 10% of gaussian noise is added to the data (the power of noise is 10% of the power of the data for each frequency), we increase the regularization factor with respect to the previous experiment, decreasing the convergence rate of all methods. Without source encoding, both truncated Newton algorithms seem to get stuck in a local minimum, while L-BFGS stagnates for some iterations but eventually succeeds to find an appropriate descent direction (Figure 3). When the sources are encoded, the misfit reduction for all methods is similar. The model error shows that the Newton methods have improved convergence when source encoding is used, suggesting that the randomness in the search direction may avoid getting stuck in a premature local minimum. The truncated Newton methods have a larger speed-up ($\approx 74\%$) because the inversion had stagnated without the random encoding. The speed up of the l-BFGS is 68%. However, note that the speed up is not a constant value, since the distance between the convergence curves changes as a function of the nonlinear iterations as the misfits is reduced.

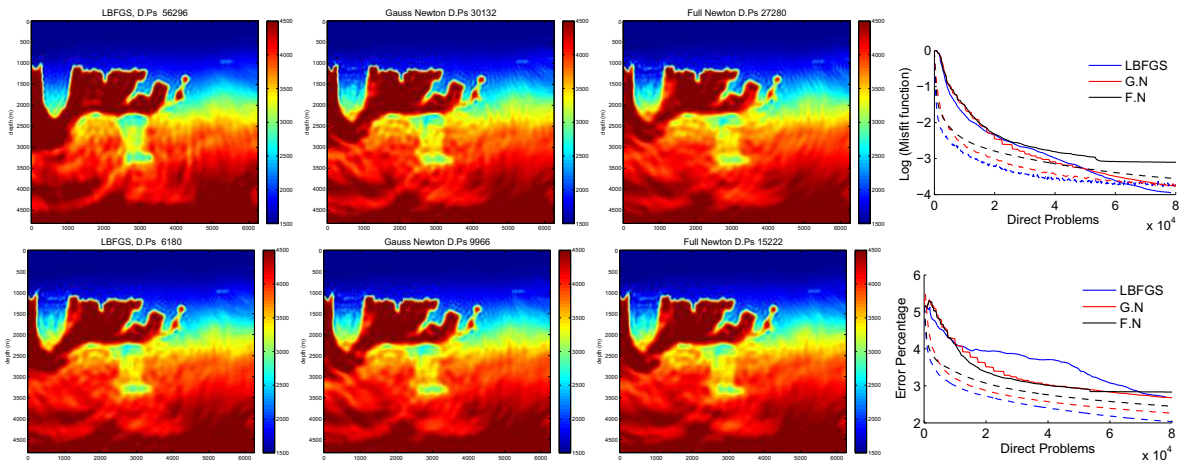

Figure 2 *Data with no noise, Velocity models for the three optimization methods (l-BFGS, GN, and FN from left to right) for a similar model error of $e \approx 3.0$. Top row: no source encoding. Bottom row: using source encoding. For the convergence curves the solid lines represent the inversions using all the individual sources, and dotted lines the inversion with source encoding. $S_{l-BFGS} \approx 89\%$. , $S_{GN} \approx 67\%$, $S_{FN} \approx 44\%$*

Conclusions

We have found that, using source encoding with periodic restarts to regenerate the random vector for the l-BFGS algorithm, allows to converge to an adequate final model. Despite the differences in the

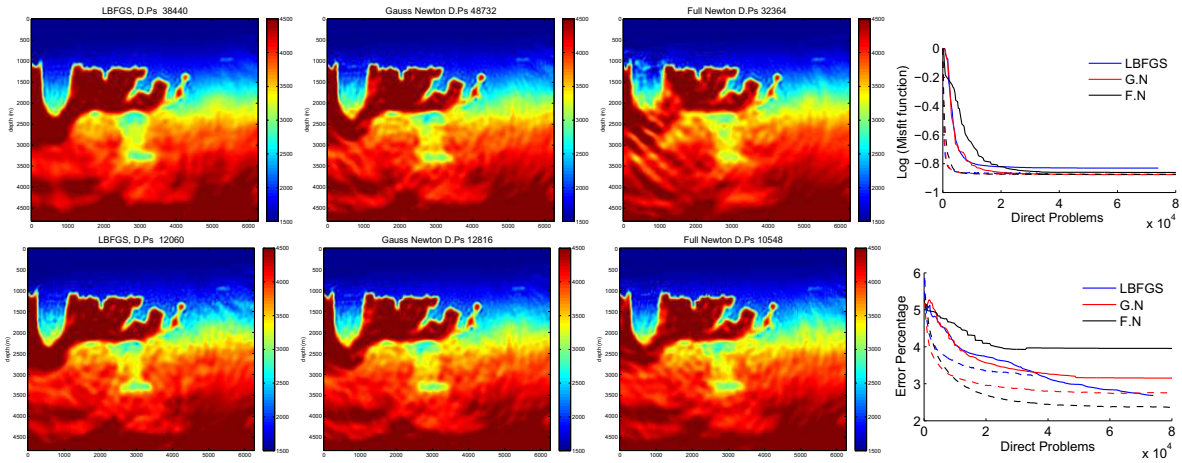

Figure 3 *Data with ten percent of noise. Velocity models for the three optimization methods (l-BFGS, G.N, and F.N from left to right) for a similar model error of $e \approx 3.2$. Top row: no source encoding. Bottom row: using source encoding. For the convergence curves the solid lines represent the inversions using all the individual sources, and dotted lines the inversion with source encoding. $S_{l-BFGS} \approx 68\%$, $S_{GN} \approx 74\%$, $S_{FN} \approx 74\%$.*

second-order optimization methods, when they are combined together with encoding techniques, they behave in a similar fashion in terms of convergence and final quality of the models. The speed-up values for each method show that there is a considerable gain both in the computational cost and, even more, in the convergence as the encoding seems to avoid to be trapped into local minima. We may see a similar behaviour as for semi-global techniques as the simulated annealing algorithm.

Acknowledgements

The authors acknowledge the University of Nice-Sophia Antipolis and the SEISCOPE consortium for the financial support. Computations have been done on the Mesocentre SIGAMM machine, hosted by Observatoire de la Cote d'Azur.

References

Ben Hadj Ali, H., Operto, S. and Virieux, J. [2011] An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics*, **76**(4), R109.

Habashy, T.M., Abubakar, A., Pan, G. and Belani, A. [2011] Source-receiver compression scheme for full-waveform seismic inversion. *Geophysics*, **76**(4), R95–R108.

Haber, E., Chung, M. and Hermann, F. [2010] An effective method for parameter estimation with P.D.E constraints with multiple right hand sides. *Tech. Rep. TR-2010-4*.

Kaltenbacher, B., Neubauer, A. and Scherzer, O. [2008] *Iterative Regularization Methods for Nonlinear Problems*. de Gruyter, Berlin, New York.

Krebs, J. et al. [2009] Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, **74(6)**, WCC105–WCC116.

Métivier, L., Brossier, R., Virieux, J. and Operto, S. [2012] Full waveform inversion and the truncated newton method. *SIAM Journal On Scientific Computing*, **accepted for publication**.

Neelamani, R., Krohn, C.E., Krebs, J.R., Deffenbaugh, M., Anderson, J.E. and Romberg, J.K. [2008] Efficient seismic forward modeling usin simultaneous random sources and sparsity. *78th Annual SEG Conference & Exhibition, Las Vegas*, Society of Exploration Geophysicists.

Schittkowski, K. [2011] A robust implementation of a sequential quadratic programming algorithm with successive error restoration. *Optimization Letters*, **5**, 283–296, ISSN 1862-4472, 10.1007/s11590-010-0207-9.

Shin, C., Jang, S. and Min, D.J. [2001] Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, **49**, 592–606.

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

van Leeuwen, T., Aravkin, A.Y. and Herrmann, F.J. [2011] Seismic waveform inversion by stochastic optimization. *International Journal of Geophysics*, **Volume 2011, ID 689041**, 18 pages.

van Leeuwen, T. and Herrmann, F. [2012] Fast waveform inversion without source-encoding. *Geophysical Prospecting*, **doi: 10.1111/j.1365-2478.2012.01096.x**.