P016

Toward Gauss-Newton and Exact Newton Optimization for Full Waveform Inversion

L. Métivier* (ISTerre), R. Brossier (ISTerre), J. Virieux (ISTerre) & S. Operto (Géoazur)

SUMMARY

Full Waveform Inversion (FWI) applications classically rely on efficient first-order optimization schemes, as the steepest descent or the nonlinear conjugate gradient optimization. However, second-order information provided by the Hessian matrix is proven to give a useful help in the scaling of the FWI problem and in the speed-up of the optimization. In this study, we propose an efficient matrix-free Hessian-vector formalism, that should allow to tackle Gauss-Newton (GN) and Exact-Newton (EN) optimization for large and realistic FWI targets. Our method relies on general second order adjoint formulas, based on a Lagrangian formalism. These formulas yield the possibility of computing Hessian-vector products at the cost of 2 forward simulations per shot. In this context, the computational cost (per shot) of one GN or one EN nonlinear iteration amounts to the resolution of 2 forward simulations for the computation of the gradient plus 2 forward simulations per inner linear conjugate gradient iteration. A numerical test is provided, emphasizing the possible improvement of the resolution when accounting for the exact Hessian in the inversion algorithm.

Introduction

Full Waveform Inversion (FWI) is becoming an efficient tool to derive high resolution quantitative models of the subsurface parameters. The method relies on the minimization, through an iterative procedure, of the residual between recorded data and synthetic data computed by solving the two-way wave equation in a subsurface model. The growth of available computational resources and recent developments of the method makes now possible applications to 2D and 3D data in the acoustic approximation (see for example Prieux et al., 2011; Plessix et al., 2010) and even in the elastic approximation (Brossier et al., 2009).

Most of the FWI applications rely on fast optimization schemes as preconditioned steepest descent or preconditioned conjugate-gradient methods (PCG). Second-order information provided by the Hessian is often neglected in FWI, due to the high computational cost to build this matrix and solve the normal equation system. However, a significant improvement of the results can be obtained using this information: Pratt et al. (1998) have shown the improved resolution of Gauss-Newton method compared to the steepest descent one in a canonical application. Hu et al. (2011) have shown results improvement provided by a non-diagonal truncated Hessian in PCG. Brossier et al. (2009) have shown the estimated Hessian's impact of a quasi-Newton l-BFGS (Nocedal, 1980), on image resolution and convergence speed compared to PCG. Epanomeritakis et al. (2008); Fichtner and Trampert (2011) have also discussed the interest of Hessian for inversion and uncertainty estimation, and also the prohibitive cost of Hessian computation and storage.

In this study, we develop the mathematical framework to propose an efficient matrix-free Hessian-vector product algorithm for FWI, and give an illustration of the interest of accounting for the exact Hessian in the inversion process. The final aim is to tackle Gauss-Newton and Full-Newton method for large scale applications. Our development relies on a general second-order adjoint-state formula, valid either in the time or in the frequency domain.

Problem statement

We consider the forward problem equation

$$S(p)u = \varphi, \tag{1}$$

where p denotes the subsurface parameters (model space), $S(p)$ denotes the linear forward problem operator corresponding to the two-way wave equation discretization[1], φ is the source vector, and u is the wavefield vector. In the following, we consider that p and u have $m \in \mathbb{N}$ components[2]. These notations are general and can be applied either in the time domain or in the frequency domain. The FWI problem is expressed as the least-square problem

$$\min_p f(p) = \frac{1}{2}\sum_{s=1}^{N_s} \|R_s u_s(p) - d_s\|^2, \tag{2}$$

where d_s and $u_s(p)$ are respectively the recorded dataset and the solution of the forward problem associated with the source φ_s, N_s is the total number of sources in the dataset, and R_s is a restriction operator that maps the wavefield u_s to the receivers locations.

The misfit gradient is

$$\nabla f(p) = \mathscr{R}\left(\sum_{s=1}^{N_s} J_s^\dagger R_s^\dagger (R_s u_s(p) - d_s)\right) \tag{3}$$

[1]Note that the operator $S(p)$ depends non-linearly on p

[2]This simplification is satisfied whenever u is discretized on the same grid than p. However, the formulas can be straightforwardly extended to the situation where p and u do not have the same number of components.

where $J_s(p)$ denotes the Jacobian matrix $\partial_p u_s(p)$. The complex conjugate transpose operator is denoted by the symbol † and the real part application by $\mathscr{R}$. The Hessian operator is

$$H(p) = \mathscr{R}\left(\sum_{s=1}^{N_s}\left(J_s^\dagger R_s^\dagger R_s J_s + \sum_{j=1}^{m}[R_s^\dagger(R_s u_s(p) - d_s)]_j H_{sj}\right)\right), \quad H_{sj} = \partial^2_{pp}(u_s)_j(p), \tag{4}$$

and its Gauss-Newton approximation is

$$B(p) = \mathscr{R}\left(\sum_{s=1}^{N_s} J_s^\dagger R_s^\dagger R_s J_s\right) \tag{5}$$

In our study, the problem (2) is solved using a Newton or a Gauss-Newton algorithm, a local iterative optimization method that computes a sequence p_k from an initial guess p_0 using the update formula

$$p_{k+1} = p_k + \alpha_k d_k, \tag{6}$$

where d_k is the solution of

$$H(p_k)d_k = -\nabla f(p_k), \text{ or } B(p_k)d_k = -\nabla f(p_k) \text{ in the Gauss-Newton approximation,} \tag{7}$$

and α_k is computed through a globalization procedure (linesearch or trust region). Explicit computation and storage of matrices $J(p), H(p)$ and $B(p)$ is prohibitive for large scale problems and realistic FWI applications. Hence, the equation (7) should be solved using matrix-free linear iterative solvers, as proposed in Epanomeritakis et al. (2008); Fichtner and Trampert (2011), such as the conjugate gradient (CG) method. This requires to compute efficiently both the gradient $\nabla f(p)$ and Hessian-vector products $H(p)v$ or $B(p)v$ for arbitrary vectors v . While the computation of $\nabla f(p)$ can be performed efficiently through the classical adjoint-state formula (Plessix, 2006)

$$(\nabla f(p))_i = \mathscr{R}\left(\partial_{p_i} S(p)u(p), \lambda(p)\right), \quad i = 1, \ldots, m \tag{8}$$

where $\lambda(p)$ (the adjoint state) is the solution of

$$S(p)^\dagger \lambda = R^\dagger(d - Ru), \tag{9}$$

we propose general second-order adjoint-state formula for the computation of $H(p)v$ and $B(p)v$ related to the Lagrangian formalism and the nonlinear constrained optimization theory.

Computing Hessian vector products

For the sake of clarity, we consider in the following that $N_s = 1$ and we drop index s. Formulas for $N_s > 1$ are directly obtained by summation. First, we define the function $g_v(p)$ such that

$$g_v(p) = (\nabla f(p), v) = \mathscr{R}\left(\left(J^\dagger R^\dagger(Ru(p) - d), v\right)\right), \tag{10}$$

where $u(p)$ is the solution of (1). We have $\nabla g_v(p) = H(p)v$. We define the Lagrangian function associated with the functional $g_v(p)$

$$L_v(p, u, \alpha, \lambda, \mu) = \mathscr{R}\left(R^\dagger(Ru - d), \alpha\right) + \mathscr{R}\left(S(p)u - \varphi, \mu\right) + \mathscr{R}\left(S(p)\alpha + \sum_{j=1}^{m} v_j \partial_{p_j} S(p)u, \lambda\right) \tag{11}$$

The Lagrangian L_v is composed of three terms: the first one accounts for the function g_v, the second one accounts for the constraints on the wavefield u, solution of the forward problem, the third one accounts for the constraints on the first-order derivatives of the wavefield u with respect to p. For $\widetilde{u}$ and $\widetilde{\alpha}$ such that

$$S(p)\widetilde{u} = \varphi, \;\; S(p)\widetilde{\alpha} = -\sum_{j=1}^{m} v_j \partial_{p_j} S(p)\widetilde{u}, \tag{12}$$

we have

$$\nabla g_v(p) = \partial_p L_v\left(p,\widetilde{u},\widetilde{\alpha},\lambda,\mu\right) + \partial_u L_v\left(p,\widetilde{u},\widetilde{\alpha},\lambda,\mu\right)\partial_p\widetilde{u}(p) + \partial_\alpha L_v\left(p,\widetilde{u},\widetilde{\alpha},\lambda,\mu\right)\partial_p\widetilde{\alpha}(p). \tag{13}$$

We define $\widetilde{\lambda}$ and $\widetilde{\mu}$ such that

$$\partial_u L_v\left(p,\widetilde{u},\widetilde{\alpha},\widetilde{\lambda},\widetilde{\mu}\right) = 0, \quad \partial_\alpha L_v\left(p,\widetilde{u},\widetilde{\alpha},\widetilde{\lambda},\widetilde{\mu}\right) = 0. \tag{14}$$

We have

$$S(p)^\dagger\widetilde{\mu} = -R^\dagger R\widetilde{\alpha} - \sum_{j=1}^{m} v_j\left(\partial_{p_j}S(p)\right)^\dagger\widetilde{\lambda}, \quad S(p)^\dagger\widetilde{\lambda} = -R^\dagger(R\widetilde{u}-d) \tag{15}$$

and

$$(H(p)v)_i = \mathscr{R}\left(\left(\left(\partial_{p_i}S(p)\right)\widetilde{u},\widetilde{\mu}\right) + \left(\left(\partial_{p_i}S(p)\right)\widetilde{\alpha},\widetilde{\lambda}\right) + \sum_{j=1}^{m} v_j\left(\left(\partial_{p_j}\partial_{p_i}S(p)\,\widetilde{u},\widetilde{\lambda}\right)\right)\right), i = 1,\ldots,m. \tag{16}$$

Note that $\widetilde{\lambda}$ corresponds to the adjoint state defined for the computation of ∇f. In addition, it can be proved that the computation of $B(p)v$ amounts to setting λ to 0 in equations (15) and (16).

$$(B(p)v)_i = \mathscr{R}\left(\left(\left(\partial_{p_i}S(p)\right)\widetilde{u},\widetilde{\mu}\right)\right), i = 1,\ldots,m \quad \text{with} \quad S(p)^\dagger\widetilde{\mu} = -R^\dagger R\widetilde{\alpha}, \tag{17}$$

The computation of one matrix vector product $H(p)v$ (or $B(p)v$) thus requires to solve one additional forward problem for α and one additional adjoint problem for μ, as reported in Epanomeritakis et al. (2008); Fichtner and Trampert (2011). For $N_s > 1$, the overall computation cost is multiplied by N_s.

Numerical results

We consider the estimation of the pressure wave velocity v_p using a 2D acoustic FWI algorithm in the frequency domain. The exact model v_p^* is composed of a homogeneous background $v_p^0 = 1500$ m.s^{-1} defined on a 2000 m length square, and two inclusions where $v_p^* = 4500$ m.s^{-1}. PML are all around. The two inclusions are distant only from 40 m (see fig.1). The high velocity contrast generates high amplitude multi-scattered waves. We use a full acquisition with four lines of 29 sources placed each 50 m on each side of the domain. Each source is associated with four lines of 29 receivers placed each 50 m all around the domain. The model is discretized over a 101×101 grid with a spatial step of 20 m. The initial guess is the background v_p^0. We use one dataset corresponding to the frequency of 5 Hz. The associated average wavelength is $\lambda \simeq 300$ m. At this frequency, the distance between the two inclusions is smaller than the expected resolution of classical FWI algorithms based on l-BFGS or Gauss-Newton approximation. We compare the results obtained performing 50 iterations of the l-BFGS algorithm, 20 iterations of Gauss-Newton inversion and 20 iterations of Exact-Newton inversion[3]. The corresponding estimated models are plot in Figure 1.

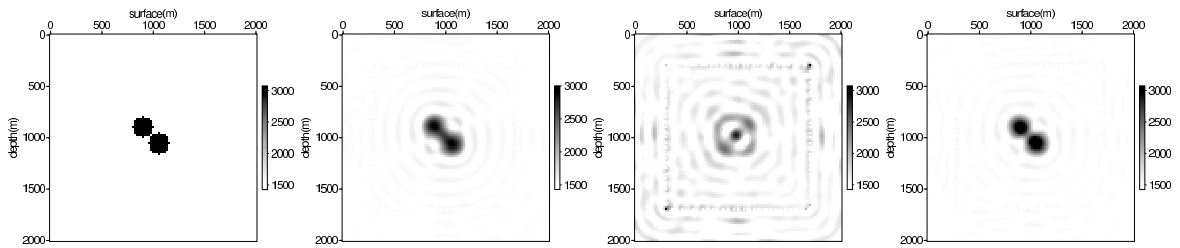

Figure 1 *Acoustic FWI for pressure wave velocity. From left to right, exact model, l-BFGS result, Gauss-Newton result, Exact-Newton result.*

[3]This corresponds approximately to the same computation cost for 2.5 conjugate gradient inner iterations per Newton nonlinear iteration

These results emphasize the role of the second-order part of the Hessian, which is neglected in the Gauss-Newton approximation, and hardly estimated in the l-BFGS approximation. As mentioned by different authors, this part of the Hessian allows to account for double scattering during the inversion (Pratt et al., 1998; Virieux and Operto, 2009). An enhancement of the resolution is obtained: only when using the exact Hessian, the two inclusions are identified. In this case, the normalized misfit is 7×10^{-4}. In the case of the Gauss-Newton inversion, the final normalized misfit is around 0.25: neglecting the double scattered waves prevent the optimizer to converge. In the case of the l-BFGS inversion, the final normalized misfit is around 10^{-3}: the l-BFGS Hessian approximation allows to account for a part of the scattered wavefield.

Conclusions and perspectives

The preliminary results we have obtained illustrate how accounting for the Hessian effect can improve the FWI results. The second-order adjoint formula is an efficient tool to implement Gaus-Newton or Exact-Newton algorithms in a matrix-free fashion. The next step consists now in developing globalization methods either based on linesearch or trust region that render the overall method competitive with the l-BFGS method. Two main problems have to be tackled: the first one is the definition of a proper criterion to stop the inner conjugate gradient iterations, in order to minimize the computation cost. The second one is the development of efficient preconditioners to speed-up the convergence of the inner conjugate gradient loop. Among several possibilities, we are interested in the Newton Steihaug algorithm (Steihaug, 1983), based on the trust-region method, that provides naturally a stopping criterion for the inner iterations. In addition, for each resolution of the inner linear systems, a l-BFGS preconditioner can be computed simultaneously, which is applied to the inner linear system raised by the next outer nonlinear iteration.

Acknowledgements

This research was funded by the SEISCOPE consortium sponsored by BP, CGG-VERITAS, ENI, EXXON MOBIL, PETROBRAS, SAUDI ARAMCO, SHELL, STATOIL and TOTAL. The linear systems were solved with the MUMPS package. This work was performed by accessing to the high-performance computing facilities of CIMENT (Université de Grenoble, France) and to the HPC resources of GENCI-CINES under Grant 2011-046091.

References

Brossier, R., Operto, S. and Virieux, J. [2009] Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, **74**(6), WCC63–WCC76, doi:10.1190/1.3215771.

Epanomeritakis, I., Akçelik, V., Ghattas, O. and Bielak, J. [2008] A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. *Inverse Problems*, **24**, 1–26.

Fichtner, A. and Trampert, J. [2011] Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, **185**(2), 775–798, ISSN 1365-246X, doi:10.1111/j.1365-246X.2011.04966.x.

Hu, W., Abubakar, A., Habashy, T.M. and Liu, J. [2011] Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion. *Geophysical Prospecting*, **59**(3), 477–491, ISSN 1365-2478, doi:10.1111/j.1365-2478.2010.00938.x.

Nocedal, J. [1980] Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, **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. *Geophysical Journal International*, **167**(2), 495–503.

Plessix, R.E., Baeten, G., de Maag, J.W., Klaassen, M., Rujie, Z. and Zhifei, T. [2010] Application of acoustic full waveform inversion to a low-frequency large-offset land data set. *SEG Technical Program Expanded Abstracts*, **29**(1), 930–934, doi:10.1190/1.3513930.

Pratt, R.G., Shin, C. and Hicks, G.J. [1998] Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, **133**, 341–362.

Prieux, V. et al. [2011] On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, **187**, 1495–1515, doi:DOI: 10.1111/j.1365-246X.2011.05209.x, 2011.

Steihaug, T. [1983] The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis*, **20**, 626–637.

Virieux, J. and Operto, S. [2009] An overview of full waveform inversion in exploration geophysics. *Geophysics*, **74**(6), WCC127–WCC152.