

# A sparse reduced Hessian approximation for multi-parameter Wavefield Reconstruction Inversion

Bas Peters\*, and Felix J. Herrmann\*

\*Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

## SUMMARY

Multi-Parameter full-waveform inversion is a challenging problem, because the unknown parameters appear in the same wave equation and the magnitude of the parameters can vary many orders of magnitude. This makes accurate estimation of multiple-parameters very difficult. To mitigate the problems, sequential strategies, regularization methods and scalings of gradients and quasi-Newton Hessians have been proposed. All of these require design, fine-tuning and adaptation to different waveform inversion problems. We propose to use a sparse approximation to the Hessian derived from a penalty-formulation of the objective function. Sparseness allows to have the Hessian in memory and compute update directions at very low cost. This results in decent reconstruction of the multiple parameters at very low additional memory and computational expense.

## INTRODUCTION

Multi-parameter waveform inversion is not a new topic, (Tarantola, 1986) proposes a strategy of inverting for different parameters sequentially. Other research in multi-parameter inversion relates to multiple parameters in multiple PDE's, such as joint inversion of electromagnetic and seismic data. In this area of research a regularization approach was used to promote structurally similar models (Gallardo & Meju, 2004). Some more recent examples in multi-parameter full waveform inversion are Lavoué, Brossier, Métivier, Garambois, & Virieux (2014) and Prioux, Brossier, Operto, & Virieux (2013).

Because regularization which promotes structural similarity relies on a priori information on the geology and sequential inversion strategies may have to be redesigned and fine-tuned for different parameters or different datasets. We will not use any of those ideas in this abstract. Although these sequential inversion and regularization concepts can definitely be very successful, we focus our efforts on fully automatic inversion with minimal use of a priori information and a minimum amount of user intervention to assess if it is a viable alternative to the methods mentioned above.

In this abstract we seek improvement on the gradient based methods (including quasi-Newton methods) by using Hessian information. Our formulation is based on the penalty formulation of the waveform inversion problem as introduced for waveform inversion by Leeuwen & Herrmann (2013). We extend their formulation to the multi-parameter case and derive a sparse Gauss-Newton type approximation to the Hessian. This type of approximation will not be possible when using the adjoint-state method to solve the Lagrangian formulation of a PDE-constrained optimization problem.

Using simple numerical experiments we illustrate the effect of

our Hessian approximation and apply it to a multi-parameter waveform inversion problem.

## WAVEFIELD RECONSTRUCTING INVERSION FORMULATION

We start by defining the equations we wish to use to invert the measured waveforms. The second order partial differential equation (PDE) that describes the acoustic wave motion is

$$\partial_r b \partial_k p + \omega^2 \kappa p = q. \quad (1)$$

This is the Helmholtz equation for a lossless and isotropic medium where there are only volume-injection (monopole) sources. In this abstract we only consider measurements of the pressure field  $p$ . Our unknown parameters which we want to invert are the buoyancy  $b$  and the compressibility  $\kappa$ . The compressibility, buoyancy and velocity ( $c$ ) in an acoustic medium are related as  $\kappa = \frac{b}{c^2}$ .

The general waveform inversion problem is following PDE-constrained optimization problem (in discrete form):

$$\min_{\mathbf{b}, \kappa, \mathbf{u}} \frac{1}{2} \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 \quad \text{s.t.} \quad \mathbf{A}(\mathbf{b}, \kappa)\mathbf{u} = \mathbf{q}. \quad (2)$$

$\mathbf{P}$  is a linear projection operator onto the receiver locations,  $\mathbf{A}$  is the discretized two-parameter Helmholtz matrix,  $\mathbf{u}$  is the pressure wavefield,  $\mathbf{q}$  is the source term and  $\mathbf{d}$  is the measured data. The two parameters to be estimated are the buoyancy  $\mathbf{b}$  and the compressibility  $\kappa$ . Following Leeuwen & Herrmann (2013), we use the Wavefield Reconstructing Inversion (WRI) methodology by using the penalty objective

$$\min_{\mathbf{b}, \kappa, \mathbf{u}} \phi_\lambda(\mathbf{b}, \kappa, \mathbf{u}) = \frac{1}{2} \|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2} \|\mathbf{A}(\mathbf{b}, \kappa)\mathbf{u} - \mathbf{q}\|_2^2. \quad (3)$$

Eq.3 is a combination of data-misfit and PDE-misfit, where the scalar  $\lambda$  determines the importance of each. This is now an unconstrained problem, generated by removing the PDE-constraint and adding it as a quadratic penalty term. We proceed by applying a variational projection as in Aravkin & Leeuwen (2012) to eliminate the field variable by solving

$$\bar{\mathbf{u}} = \arg \min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda \mathbf{A}(\mathbf{b}, \kappa) \\ \mathbf{P} \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2 \quad (4)$$

to obtain the reduced form

$$\min_{\mathbf{b}, \kappa} \bar{\phi}_\lambda(\mathbf{b}, \kappa) = \frac{1}{2} \|\mathbf{P}\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2} \|\mathbf{A}(\mathbf{b}, \kappa)\bar{\mathbf{u}} - \mathbf{q}\|_2^2. \quad (5)$$

Using this formulation, we update the fields  $\bar{\mathbf{u}}$  first, and then update the medium parameters. In other words, we first reconstruct the wavefield, given the PDE and observed data, then we use that field to update the model. So at each iteration of an optimization algorithm, we solve the least-squares problem

and find an update direction using an optimization algorithm of choice. The misfit and gradients can be accumulated as a running sum over frequencies and sources. Therefore, only one field has to be in memory at a time.

To compute gradients and Hessians we need the partial derivatives of the PDE w.r.t the two parameters which are

$$\begin{aligned} \mathbf{G}_b &= \partial \mathbf{A}(\mathbf{b}, \kappa) \bar{\mathbf{u}} / \partial \mathbf{b} \\ \mathbf{G}_\kappa &= \partial \mathbf{A}(\mathbf{b}, \kappa) \bar{\mathbf{u}} / \partial \kappa. \end{aligned} \quad (6)$$

Because of the parameterization choice, the partial derivative w.r.t one parameter does not depend on the other. These partial derivatives can be used to write down the gradient of the reduced objective function (eq.5):

$$\begin{aligned} \nabla_b \bar{\phi}_\lambda &= \lambda^2 \mathbf{G}(\mathbf{b}, \bar{\mathbf{u}})^* (\mathbf{A}(\mathbf{b}, \kappa) \bar{\mathbf{u}} - \mathbf{q}) \\ \nabla_\kappa \bar{\phi}_\lambda &= \lambda^2 \mathbf{G}(\kappa, \bar{\mathbf{u}})^* (\mathbf{A}(\mathbf{b}, \kappa) \bar{\mathbf{u}} - \mathbf{q}). \end{aligned} \quad (7)$$

The structure of these gradients reveal that a perturbation in the compressibility will induce a nonzero gradient for the buoyancy and vice versa. This is the result of the occurrence of both parameters the Helmholtz equation and is one of the reasons multi-parameter inversion of parameters occurring in the same equation is challenging.

## OBTAINING A SPARSE APPROXIMATION TO THE REDUCED HESSIAN WITHOUT COMPUTATIONAL COST

The gradients (eq.7) of the reduced objective (eq.5) can be used for gradient descent-type methods. We will not study these, because the linear convergence rate will make this method infeasible for difficult problems with an ill-conditioned Hessian. Quasi-Newton methods can provide a faster superlinear rate of convergence. The Hessian, of which L-BFGS tries to iteratively obtain an approximation with low-rank updates, has a block structure, where the blocks corresponding to different unknowns can have widely varying Frobenius norms. This results in strong ill-conditioning (can be many orders of magnitude above machine precision). It is therefore questionable how much sense the L-BFGS approximation will make. It is possible to initialize the L-BFGS Hessian with a diagonal to correct for scale differences between Hessian blocks corresponding to the different parameters. This initialization remains a far from trivial choice that has to be made and although sensible and physically inspired choices can be made, the information is not provided by the problem at hand directly. Diagonal scaling can improve compared to the case where just the gradient is used however. See Lavoué et al. (2014) for a nice discussion about the difficulties of multi-parameter inversion and the L-BFGS Hessian. These disappointing results are the motivation to use Hessian information. We begin by deriving the Hessian matrix corresponding to the reduced objective (eq.5). Note that the Hessian and Newton system for the full objective (eq.3) are given by

$$\begin{pmatrix} \nabla_{\bar{\mathbf{u}}, \bar{\mathbf{u}}}^2 \phi_\lambda & \nabla_{\bar{\mathbf{u}}, \kappa}^2 \phi_\lambda & \nabla_{\bar{\mathbf{u}}, \mathbf{b}}^2 \phi_\lambda \\ \nabla_{\kappa, \bar{\mathbf{u}}}^2 \phi_\lambda & \nabla_{\kappa, \kappa}^2 \phi_\lambda & \nabla_{\kappa, \mathbf{b}}^2 \phi_\lambda \\ \nabla_{\mathbf{b}, \bar{\mathbf{u}}}^2 \phi_\lambda & \nabla_{\mathbf{b}, \kappa}^2 \phi_\lambda & \nabla_{\mathbf{b}, \mathbf{b}}^2 \phi_\lambda \end{pmatrix} \begin{pmatrix} \delta_{\bar{\mathbf{u}}} \\ \delta_\kappa \\ \delta_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \nabla_{\bar{\mathbf{u}}} \phi_\lambda \\ \nabla_\kappa \phi_\lambda \\ \nabla_{\mathbf{b}} \phi_\lambda \end{pmatrix}. \quad (8)$$

Solving the least-squares problem in eq.4 is equivalent to setting the first order optimality condition of eq.3 w.r.t  $\mathbf{u}$  to zero:  $\nabla_{\bar{\mathbf{u}}} \phi_\lambda(\bar{\mathbf{u}}, \kappa, \mathbf{b}) = 0$  and then applying block-Gaussian elimination for the  $\delta_{\bar{\mathbf{u}}}$  variable from the full Newton system (eq.8). Another point of view is that the resulting reduced Hessian is the Schur-complement of the full Hessian. To write this compactly, partition the Hessian in eq.8 as

$$\begin{pmatrix} \mathbf{E} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}. \quad (9)$$

The Schur complement follows as

$$(\mathbf{D} - \mathbf{C}\mathbf{E}^{-1}\mathbf{B}) \begin{pmatrix} \delta_\kappa \\ \delta_{\bar{\mathbf{u}}} \end{pmatrix} = \begin{pmatrix} \nabla_\kappa \phi_\lambda \\ \nabla_{\bar{\mathbf{u}}} \phi_\lambda \end{pmatrix} - \mathbf{C}\mathbf{E}^{-1} \nabla_{\bar{\mathbf{u}}} \phi_\lambda(\bar{\mathbf{u}}, \kappa, \mathbf{b}) \quad (10)$$

which instantly simplifies to

$$(\mathbf{D} - \mathbf{C}\mathbf{E}^{-1}\mathbf{B}) \begin{pmatrix} \delta_\kappa \\ \delta_{\bar{\mathbf{u}}} \end{pmatrix} = \begin{pmatrix} \nabla_\kappa \phi_\lambda \\ \nabla_{\bar{\mathbf{u}}} \phi_\lambda \end{pmatrix}, \quad (11)$$

because we set  $\nabla_{\bar{\mathbf{u}}} \phi_\lambda(\bar{\mathbf{u}}, \kappa, \mathbf{b}) = 0$  by solving eq.4. One of the main disadvantages of the reduced Hessians, for both penalty and Lagrangian objective formulations, is that they are dense. This can be readily observed by noting the appearance of the  $\mathbf{A}^{-1}$  block in the system (see Haber, Ascher, & Oldenburg, 2000 for the reduced Hessian based on a Lagrangian objective). This results in enormous memory requirements making it usually impossible to store it in memory for medium sized 2D and larger seismic applications. The reduced dense Hessian can still be used by computing matrix-free matrix-vector products to compute the Newton direction iteratively, see Métivier, Brossier, Virieux, & Operto (2013) for this method in the waveform inversion context. Matrix-free evaluations cost expensive extra PDE or least-squares system solves. Moreover, these matrix-free matrix-vector products use a chain of linear-system solves, inducing complex and amplified error propagation. Therefore this strategy is less suitable for inexact iterative linear-system solution strategies.

The key contribution of this abstract follows here. We will make two approximations to the reduced Hessian. This Hessian originating from the penalty objective formulation can be split in a dense ( $\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ ) and a sparse part ( $\mathbf{D}$ ). As the first approximation, we choose to approximate the reduced Hessian by only its sparse part,  $\mathbf{D}$ . The accuracy of this approximation depends on the choice of the scalar  $\lambda$ , it is more accurate for a small  $\lambda$ . The penalty objective is equivalent to the constrained Lagrangian formulation for  $\lambda \uparrow \infty$ . The use of the sparse part enables us to have it in memory and compute exact Newton-type directions at very low cost, as explained next. Thus, after the first approximation we have the following Hessian:

$$\begin{pmatrix} \nabla_{\kappa, \kappa}^2 \phi_\lambda & \nabla_{\kappa, \mathbf{b}}^2 \phi_\lambda \\ \nabla_{\mathbf{b}, \kappa}^2 \phi_\lambda & \nabla_{\mathbf{b}, \mathbf{b}}^2 \phi_\lambda \end{pmatrix}. \quad (12)$$

The second approximation is to drop the off-diagonal blocks from this sparse part. These blocks can be used to obtain an algorithm which has potentially a better rate of convergence, but this topic is not explored in this abstract. Our final approximation to the reduced Hessian which we will use in the numerical

examples below is given by

$$\tilde{\mathbf{H}} = \begin{pmatrix} \nabla_{\kappa, \kappa}^2 \phi_\lambda & 0 \\ 0 & \nabla_{\mathbf{b}, \mathbf{b}}^2 \phi_\lambda \end{pmatrix} = \begin{pmatrix} \mathbf{G}_\kappa^* \mathbf{G}_\kappa & 0 \\ 0 & \mathbf{G}_\mathbf{b}^* \mathbf{G}_\mathbf{b} \end{pmatrix}. \quad (13)$$

We summarize the important properties of this Hessian approximation which we will refer to as a sparse Gauss-Newton approximation: 1) positive-definiteness, 2) Hermitian and 3) sparse. These properties make the Gauss-Newton step

$$\mathbf{p}_{gn} = \begin{pmatrix} \mathbf{G}_\kappa^* \mathbf{G}_\kappa & 0 \\ 0 & \mathbf{G}_\mathbf{b}^* \mathbf{G}_\mathbf{b} \end{pmatrix}^{-1} \begin{pmatrix} \nabla_{\mathbf{b}} \bar{\phi}_\lambda \\ \nabla_{\kappa} \bar{\phi}_\lambda \end{pmatrix} \quad (14)$$

cheap to compute. The block structure enables parallel solution of both nonzero blocks. The content of these blocks depends of the specific parameterization and discretization used. Here we follow the discretize-then-optimize approach and discretize such that  $\mathbf{G}_\kappa$  is a diagonal matrix. Inverting  $\mathbf{G}_\kappa^* \mathbf{G}_\kappa$  is then trivial. The buoyancy in the Helmholtz equation (eq.1) is connected to differential operators and that is why the partial derivative w.r.t the buoyancy ( $\mathbf{G}_\mathbf{b}$ ) will contain a differential operator. The block  $\mathbf{G}_\mathbf{b}^* \mathbf{G}_\mathbf{b}$  cannot be trivially inverted, but it is just one extra PDE-solve on top of all the linear systems that have to be solved each iteration (eq. 4) to obtain the fields. Note that this block is also symmetric positive definite and suitable for efficient iterative solvers and preconditioners. Therefore this sparse approximate reduced Hessian approach does not consume a lot of storage, does not take any significant computations to form and the solution of the Gauss-Newton system (eq.14) gives rise to a minor contribution to the total computational cost of the Gauss-Newton waveform inversion algorithm. Below is the final algorithm in a nutshell.

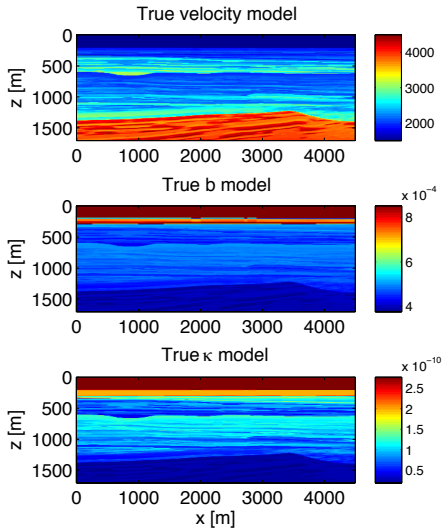


Figure 1: True models for example 1.

## NUMERICAL EXPERIMENTS

### Example 1: A motivational example

We can naively use just the gradient in the L-BFGS method. The result is shown in figure 2 where the true models are shown

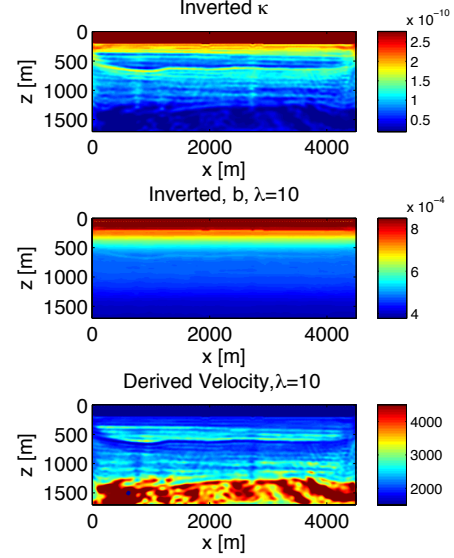


Figure 2: Inverted compressibility and buoyancy for example 1. The velocity computed from the inverted parameters is also shown. The optimization algorithm was a naive L-BFGS implementation.

---

**Algorithm 1** Waveform inversion with a sparse Hessian approximation.

---

**while** Not converged **do**

1.  $\bar{\mathbf{u}} = \arg \min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda \mathbf{A}(\mathbf{b}, \kappa) \\ \mathbf{P} \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$  // Solve
2.  $\mathbf{G}_\kappa, \mathbf{G}_\mathbf{b}, \nabla_{\mathbf{b}} \bar{\phi}_\lambda, \nabla_{\kappa} \bar{\phi}_\lambda$  // Form
3.  $\mathbf{p}_{gn} = \tilde{\mathbf{H}}^{-1} \mathbf{g}$  // Solve
4. find steplength  $\alpha$  // Linesearch
5.  $\mathbf{m} = \mathbf{m} + \alpha \mathbf{p}_{gn}$  // update model

**end**

---

in figure 1. The initial guess was a smoothed version of the true models. The results show that the buoyancy is not updated very much. Nearly all of the updates is in the compressibility model. The velocity derived from the compressibility and buoyancy is actually reasonable.

### Example 2: Heuristics about the effect of the sparse Hessian approximation

To analyse the effect of our proposed Hessian approximation, we construct a model with a point scatterer. It consists of both a compressibility and buoyancy perturbation. The background model is homogeneous. We look at the gradient of the objective w.r.t both the buoyancy and compressibility and compare it to the corresponding Gauss-Newton directions to visually illustrate the effect of the Hessian approximation. The sources and receivers are located at the top of the domain. We see that the Gauss-Newton direction improves the focussing of the buoyancy direction, but it does not do much for the compressibility.

### Example 3: Application to waveform inversion

In this waveform inversion example we try to invert just pres-

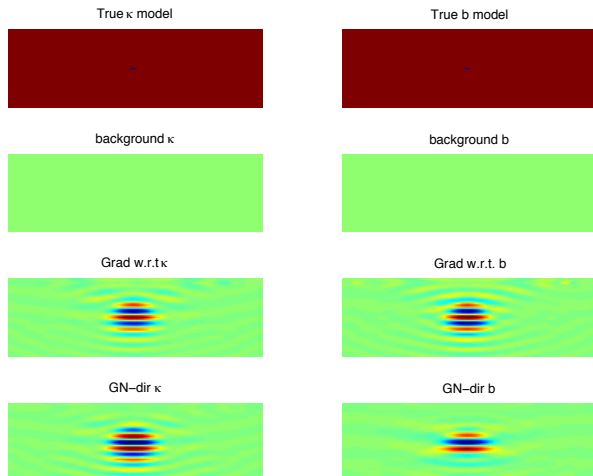


Figure 3: True and initial medium. The gradients and Gauss-Newton directions are shown as well. It can be seen that the Gauss-Newton direction w.r.t. the buoyancy is more focussed than the corresponding gradient direction. The two directions do not show much difference for the compressibility.

sure wavefield data recorded near the surface of the model. There are 77 sources and receivers and the data contains frequencies from 2 to 17 Hertz. We use a frequency continuation approach to invert the data. We invert one batch of just 2 frequencies at a time and use the result as an initial guess for the next batch. The batches are  $\{2, 3\}, \{3, 4\}, \dots, \{16, 17\}$  Hertz. Algorithm 1 was used to solve this inverse problem. The results are shown in figure 4. The reconstructed buoyancy and compressibility show the main structures of the true models (figure 1). The derived velocity model is also quite decent. This synthetic example indicates that our approximate reduced Hessian allows simultaneous inversion of multiple-parameters without the introduction of explicit scalings of gradients or quasi-Newton Hessians. In a way, these type of scalings are implicit in the true Hessian which naturally provides this information based on the actual problem.

## COMPARISON TO ADJOINT-STATE METHOD

The adjoint-state method is based on the Lagrangian formulation to solve eq.2. One proceeds by solving for the field variable and using the result to solve for the Lagrangian multiplier (adjoint-wavefield). This can be used to compute the gradient of the resulting reduced Lagrangian objective w.r.t the medium parameters. This effectively eliminated the PDE-constraint and reduced the constrained problem to an unconstrained one. The resulting reduced Hessian (see for example Haber et al., 2000) contains only one sparse part which is not Hermitian positive definite for all parameters. This makes a similar approach to the one proposed in this abstract impossible. To use Hessian information in this case, the dense Hessian needs to be stored or matrix-vector products have to be computed though, accurate and expensive, extra PDE-solves.

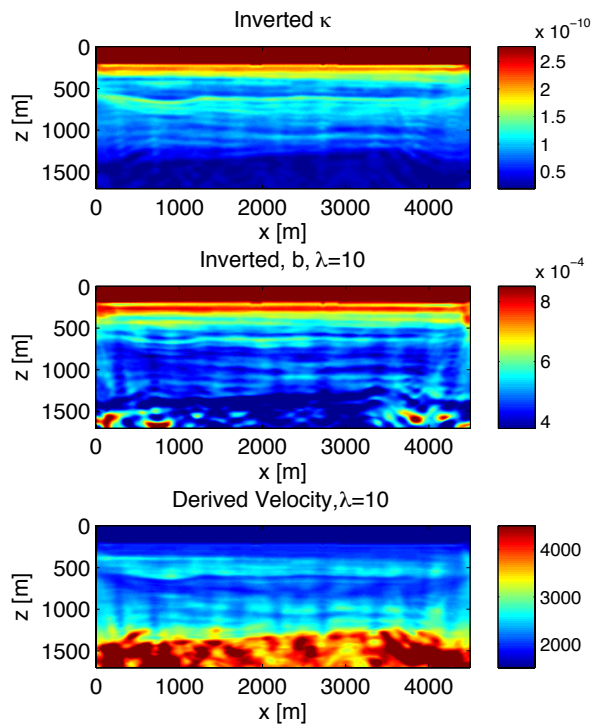


Figure 4: Inverted compressibility and buoyancy for example 3. The velocity computed from the inverted parameters is also shown. Inversion by algorithm 1. Both compressibility and buoyancy results show the main structures of the true model. The derived velocity is also decent.

## CONCLUSIONS AND DISCUSSION

We derived a sparse approximation to the reduced Hessian based on a penalty formulation of the waveform inversion problem. This approximate Hessian enables multi-parameter waveform inversion without the need for incorporating a priori information about the coupling between the parameters by regularization, scaling of gradients or quasi-Newton Hessians or inverting for different parameters sequentially. The derived approximation to the reduced Hessian is Hermitian positive definite by construction and easy to invert when computing the update direction. Further research will focus on including more sparse parts into the approximation as well as the analysis of the practically achievable rate of convergence.

## ACKNOWLEDGEMENTS

This work was financially supported in part by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN 261641-06) and the Collaborative Research and Development Grant DNOISE II (CDRP J 375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, CGG, ION GXT, Petrobras, PGS, Statoil, Total SA, WesternGeco, Woodside.

Aravkin, A. Y., & Leeuwen, T. van. (2012). Estimating nuisance parameters in inverse problems. *Inverse Problems*, 28(11), 115016.

Gallardo, L. A., & Meju, M. A. (2004). Joint two-dimensional dC resistivity and seismic travel time inversion with cross-gradients constraints. *Journal of Geophysical Research: Solid Earth*, 109(B3), n/a–n/a. doi:[10.1029/2003JB002716](https://doi.org/10.1029/2003JB002716)

Haber, E., Ascher, U. M., & Oldenburg, D. (2000). On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16(5), 1263–1280. doi:[10.1088/0266-5611/16/5/309](https://doi.org/10.1088/0266-5611/16/5/309)

Lavoué, F., Brossier, R., Métivier, L., Garambois, S., & Virieux, J. (2014). Two-dimensional permittivity and conductivity imaging by full waveform inversion of multioffset gPR data: a frequency-domain quasi-newton approach. *Geophysical Journal International*, 197(1), 248–268. doi:[10.1093/gji/ggt528](https://doi.org/10.1093/gji/ggt528)

Leeuwen, T. van, & Herrmann, F. J. (2013). Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195, 661–667. doi:[10.1093/gji/ggt258](https://doi.org/10.1093/gji/ggt258)

Métivier, L., Brossier, R., Virieux, J., & Operto, S. (2013). Full waveform inversion and the truncated newton method. *SIAM Journal on Scientific Computing*, 35(2), B401–B437. doi:[10.1137/120877854](https://doi.org/10.1137/120877854)

Prieux, V., Brossier, R., Operto, S., & Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 1: imaging compressional wave speed, density and attenuation. *Geophysical Journal International*, 194(3), 1640–1664. doi:[10.1093/gji/ggt177](https://doi.org/10.1093/gji/ggt177)

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *GEOPHYSICS*, 51(10), 1893–1903. doi:[10.1190/1.1442046](https://doi.org/10.1190/1.1442046)