GEOPHYSICS, VOL. 79, NO. 3 (MAY-JUNE 2014); P. X11–X15, 9 FIGS., 1 TABLE.
10.1190/GEO2013-0466.1

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

Discussion

Comment on: "Application of the variable projection scheme for frequency-domain full-waveform inversion" (M. Li, J. Rickett, and A. Abubakar, Geophysics, 78, no. 6, R249–R257)

Discussion by Tristan van Leeuwen[1], Aleksandr Y. Aravkin[2], and Felix J. Herrmann[3]

In the paper "Application of the variable projection scheme for frequency-domain full-waveform inversion," Li et al. (2013) discuss a method for source estimation in the context of frequency-domain full-waveform inversion (FWI). This method is an extension of earlier work by Aravkin et al. (2012), and Li et al. (2013) suggest similar improvements as concurrent work by Aravkin and van Leeuwen (2012). We have read the work by Li et al. (2013) with great pleasure but feel that their claims are not fully supported by numerical experiments. The goal of this discussion paper is twofold; we would like to

1) clarify some technical details of the extension as presented by Aravkin and van Leeuwen (2012) that are not discussed by Li et al. (2013) and
2) suggest additional numerical experiments to support the claims made by Li et al. (2013).

The remainder of this section contains the main discussion and refers to the appendices for more (technical) details and numerical results.

Li et al. (2013) discuss an automated procedure for estimating the source weights in frequency-domain FWI. The main idea is that these source weights can be estimated on-the-fly during each iteration of FWI using a technique known as "variable projection." A brief mathematical description of the problem is given in Appendix A (under "Problem statement"). The use of the variable-projection method for source estimation in FWI has a long history, as is noted by Li et al. (2013). As far as we know, it was first recognized by Aravkin et al. (2012) that source estimation is indeed an instance of variable projection, and this insight was used to generalize the concept of source estimation to generic misfit functions.

Li et al. (2013) state that when using the variable projection technique to compute the source weights, the Jacobian (which Li et al. (2013) refer to as "the gradient" in the introduction) and Hessian contain correction terms, which were not computed previously. They correctly point out that ignoring these correction terms does not affect the calculation of the gradient (i.e., multiplication of the adjoint of the Jacobian with the residual) so that gradient-descent methods can be applied without computing this correction term. This is exactly the regime discussed by Aravkin et al. (2012), and therefore there are no gradient approximations used in this paper, contrary to what is suggested by Li et al. (2013). For the sake of completeness, we have included a derivation in Appendix A (under "Computing the gradient").

The Hessian does contain a correction term however, and explicit expressions of these terms are presented by Aravkin and van Leeuwen (2012). The variable projection method was also included in a frequency-domain waveform inversion method by van Leeuwen and Herrmann (2013), where a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) approximation of the *full* Hessian was used, which automatically includes the correction terms.

The expressions presented by Li et al. (2013) amount to a Gauss-Newton (GN) approximation of these correction terms. We include a derivation in Appendix A (under "Computing the Hessian"), along with an explicit Schur-complement expression for the reduced Hessian that was not presented in Aravkin and van Leeuwen (2012) nor by Li et al. (2013). From this expression, it is evident that ignoring correction terms amounts to *adding* a positive semidefinite correction term to the true Hessian, which is a common optimization technique (recall, e.g., the Levenberg-Marquardt method). Hence, properly designed optimization strategies can ignore the correction terms and still satisfy standard convergence guarantees (i.e., guarantees for convergence to stationary points). The effect of including correction terms (that can possibly make the Hessian indefinite) is problem-specific, and in some cases, ignoring negative definite corrections in the context of GN methods can improve algorithm performance (Aravkin et al., 2012).

The main contribution of Li et al. (2013) is the computation of a GN approximation of the correction terms in the Hessian. However, Li et al. (2013) do not compare the performance of their algorithm with a "standard" approach that does not include the correction terms. Thus, the actual benefit of including the correction terms

Manuscript received by the Editor 16 December 2013; revised manuscript received 6 February 2014; published online 20 May 2014.
[1]Centrum Wiskunde & Informatica, Amsterdam, The Netherlands. E-mail: t.van.leeuwen@cwi.nl.
[2]IBM T.J. Watson Research Center, Yorktown Heights, New York, USA. E-mail: saravkin@us.ibm.com.
[3]University of British Columbia, Department of Earth, Ocean and Atmospheric Sciences, Vancouver, British Columbia, Canada. E-mail: fherrmann@eos.ubc.ca.
© 2014 Society of Exploration Geophysicists. All rights reserved.

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

is yet to be investigated. To shed some light on this issue, we present some numerical experiments on a toy model depicted in Figure 1, using both a reflection and a transmission setup. More details can be found in Appendix B. First, we compute the effect of the correction term by plotting the actual misfit as well a quadratic model based on the Hessian in Figure 2. Second, we compare the use of an L-BFGS method to a GN method with and without correction terms. These experiments suggest that the correction terms can indeed affect the rate of convergence and the final results (see Figures 3 and 4). However, the differences observed on two toy problems are small and only occur when reducing the misfit by three orders of magnitude, a situation that rarely occurs in practice. The MATLAB code used to conduct the numerical experiment can be found here: https://github.com/tleeuwen/Variable-Projection-for-FWI.git.

In conclusion, the correction terms suggested by Li et al. (2013) are a particular approximation of the full Hessian that follows from the variable projection approach (Aravkin and van Leeuwen, 2012). Numerical experiments on a toy problem suggest that including these terms may improve convergence but might not influence the final result significantly. We expect the influence of the correction terms to be problem dependent and we would like to encourage the authors to repeat this experiment with their code on a more realistic problem and include the results in a reply.

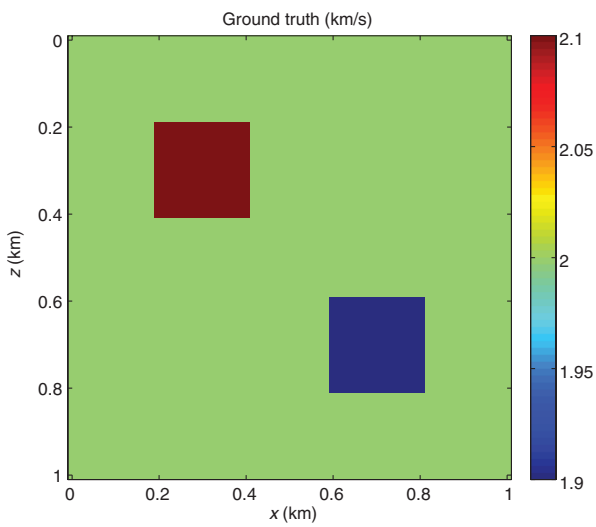

Figure 1. Velocity model used for experiments.

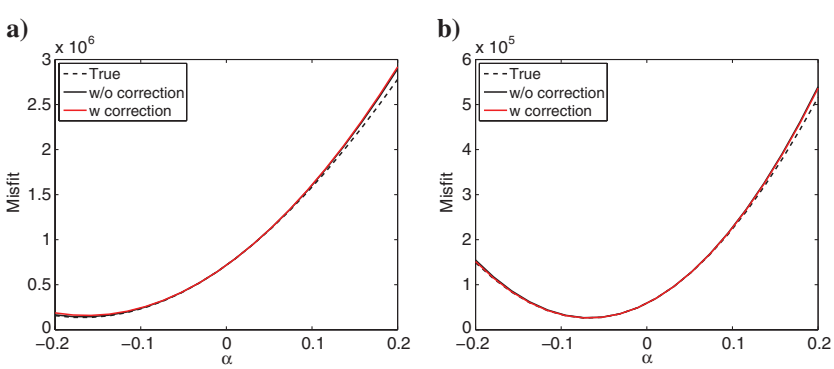

Figure 2. Actual misfit and quadratic model with and without correction terms for transmission (a) and reflection (b) configuration.

ACKNOWLEDGMENTS

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Total SA, Statoil, WesternGeco, and Woodside.

APPENDIX A

THEORY

Problem statement

Li et al. (2013) consider the following optimization problem (we ignore the regularization and weighting introduced in Li et al. [2013] because this does not have an impact on the source estimation):

$$\min_{\mathbf{m},\mathbf{c}} \Phi(\mathbf{m},\mathbf{c}) = \sum_{k,j} \|\mathbf{d}_{k,j} - c_{k,j}\mathbf{s}_{k,j}(\mathbf{m})\|_2^2, \tag{A-1}$$

where $\mathbf{d}_{k,j}$ is the observed data (organized in a vector) for the kth frequency and jth source; $\mathbf{s}_{k,j}(\mathbf{m})$ is the corresponding synthetic data (organized in a vector) for model $\mathbf{m}$; and $c_{k,j}$ are the (possibly complex) source weights, which are assembled in the vector $\mathbf{c}$. These source weights need to be estimated to calibrate the modeling code. Using the variable-projection method entails projecting out the source weights by solving

$$\min_{\mathbf{c}} \Phi(\mathbf{m},\mathbf{c}), \tag{A-2}$$

the solution of which we denote by $\bar{\mathbf{c}}(\mathbf{m})$. Note that the optimal source weight is implicitly defined by

$$\nabla_{\mathbf{c}}\Phi(\mathbf{m},\bar{\mathbf{c}}) = 0. \tag{A-3}$$

Substituting the optimal source weight back into the objective we obtain a *reduced* objective in $\mathbf{m}$ alone, written

$$\bar{\Phi}(\mathbf{m}) = \Phi(\mathbf{m},\bar{\mathbf{c}}(\mathbf{m})). \tag{A-4}$$

Next, we discuss the following three main algorithmic ingredients of the variable projection approach:

1) a method for obtaining $\bar{\mathbf{c}}(\mathbf{m})$
2) an explicit expression for the gradient of $\bar{\Phi}(\mathbf{m})$
3) an explicit expression for the Hessian of $\bar{\Phi}(\mathbf{m})$.

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

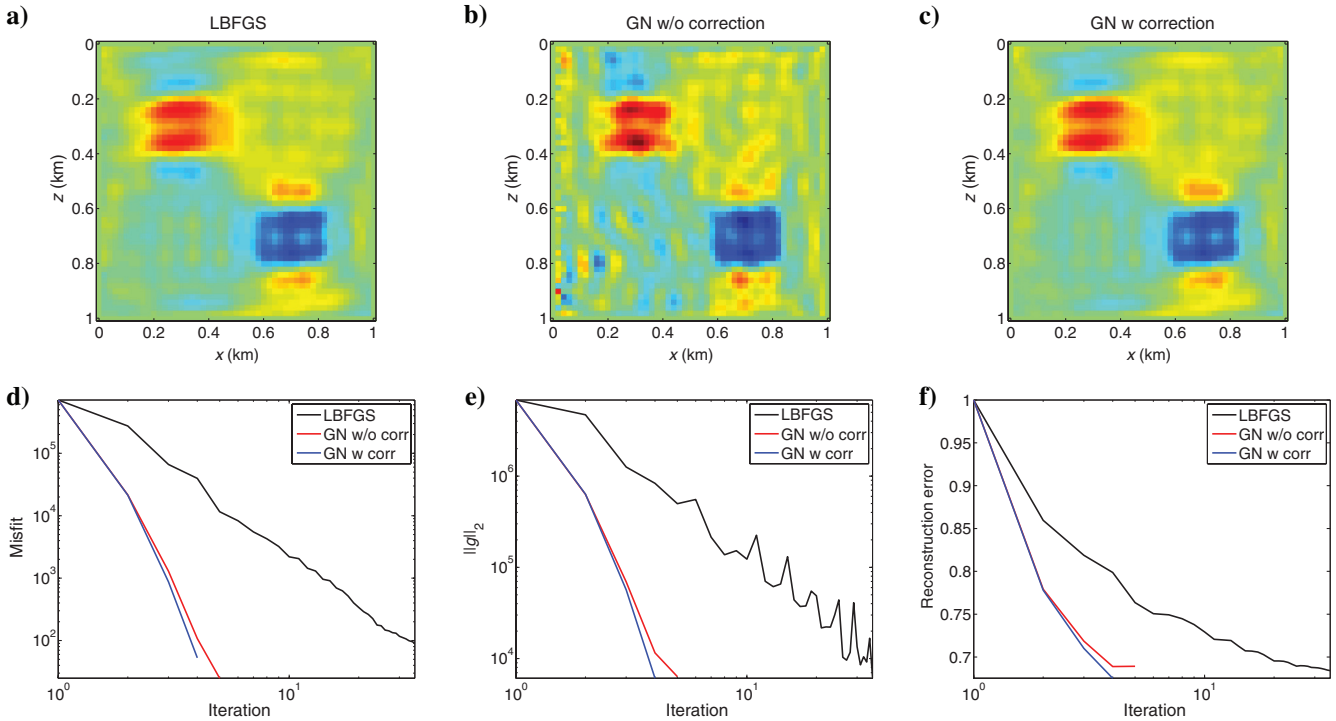

Figure 3. Inversion result for transmission configuration. The reconstructions with L-BFGS and GN with and without correction terms are shown in (a-c). The convergence histories in terms of the misfit, norm of the gradient, and reconstruction error ($\|\mathbf{m}_k - \mathbf{m}_{\text{true}}\|_2/\|\mathbf{m}_0 - \mathbf{m}_{\text{true}}\|$) are shown in (d-f).

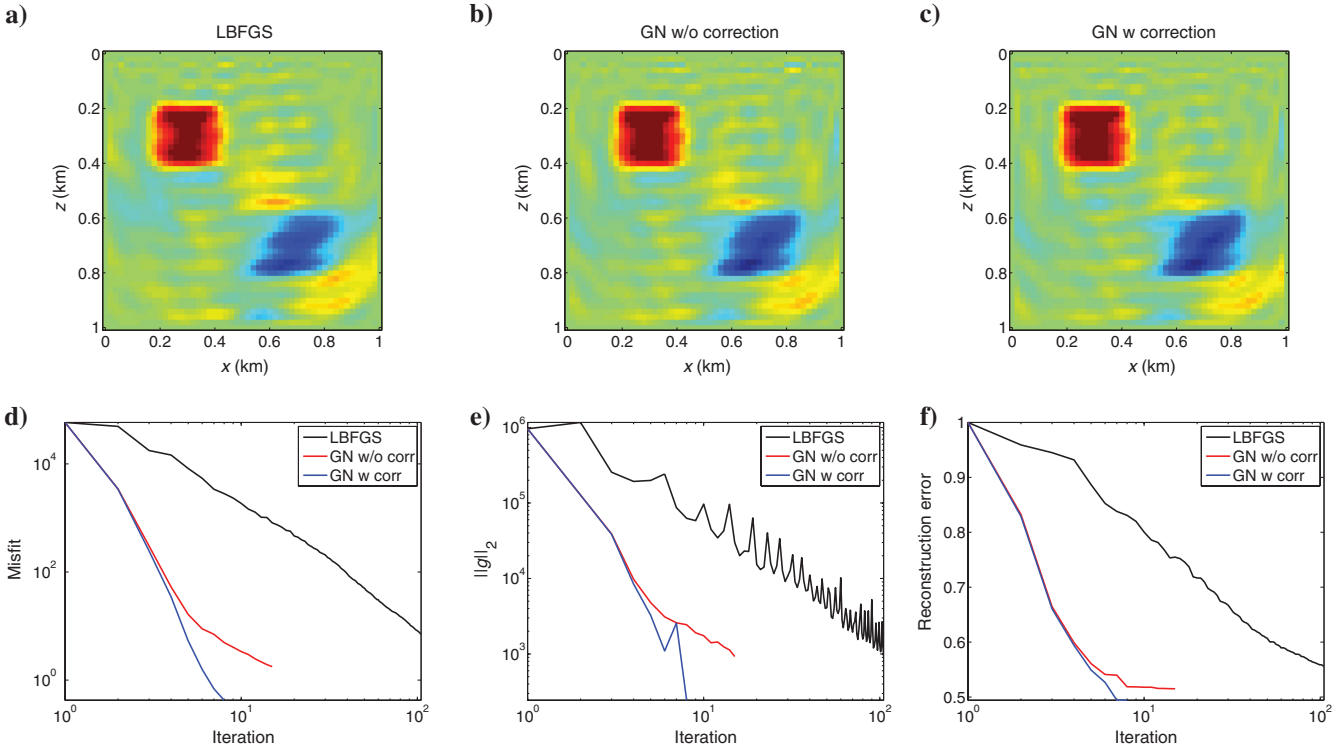

Figure 4. Inversion result for reflection configuration. The reconstructions with L-BFGS and GN with and without correction terms are shown in (a-c). The convergence histories in terms of the misfit, norm of the gradient, and reconstruction error ($\|\mathbf{m}_k - \mathbf{m}_{\text{true}}\|_2/\|\mathbf{m}_0 - \mathbf{m}_{\text{true}}\|$) are shown in (d-f).

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

To optimize the objective, one needs only the first two items because this enables methods such as gradient-descent or L-BFGS. The last item (or one of several possible approximations) lets us use different (Gauss-) Newton variants.

Obtaining the source weights

The main driver behind the variable-projection technique is that the inner optimization problem (A-2) is easy to solve. In this case, there is of course a closed-form solution:

$$c_{k,j} = \mathbf{s}_{k,j}^{*}\mathbf{d}_{k,j}/\|\mathbf{s}_{k,j}\|_2^2, \tag{A-5}$$

where $*$ denotes the complex-conjugate transpose. See also equation A-11 in Li et al. (2013). However, even if the problem does not have a closed-form solution (when using other misfit penalties, for example), one can devise an efficient procedure (e.g., by solving several independent scalar optimization problems in parallel) to obtain $c_{k,j}$. Some examples are discussed by Aravkin and van Leeuwen (2012) and Aravkin et al. (2012).

Computing the gradient

Computation of the gradient can be done via the chain rule, yielding

$$\nabla_{\mathbf{m}}\bar{\Phi}(\mathbf{m}) = \nabla_{\mathbf{m}}\Phi(\mathbf{m},\bar{\mathbf{c}}) + \nabla_{\mathbf{c}}\Phi(\mathbf{m},\bar{\mathbf{c}})\nabla_{\mathbf{m}}\bar{\mathbf{c}}, \tag{A-6}$$

from which it is immediately clear that the second term vanishes if we have really solved $\nabla_{\mathbf{c}}\Phi = 0$ to obtain $\bar{\mathbf{c}}$. In the quadratic case (i.e., with closed form solution given by equation A-5), this is certainly true, and we have

$$\nabla_{\mathbf{m}}\bar{\Phi}(\mathbf{m}) = \nabla_{\mathbf{m}}\Phi(\mathbf{m},\bar{\mathbf{c}}). \tag{A-7}$$

Even if there is no closed-form solution for $\bar{\mathbf{c}}$, we can solve $\nabla_{\mathbf{c}}\Phi = 0$ up to arbitrary precision and safely ignore the correction term.

The same argument can be made by computing the Jacobian of the modeled data $\bar{c}_{k,j}\mathbf{s}_{k,j}$, which consists of two parts:

$$\frac{\partial \bar{c}_{k,j}(\mathbf{m})\mathbf{s}_{k,j}(\mathbf{m})}{\partial \mathbf{m}} = \bar{c}_{k,j}(\mathbf{m})\frac{\partial \mathbf{s}_{k,j}(\mathbf{m})}{\partial \mathbf{m}} + \mathbf{s}_{k,j}\frac{\partial \bar{c}_{k,j}}{\mathbf{m}}. \tag{A-8}$$

Compare this expression with Li et al. (2013), equation A-13. The main point of the original paper on variable projection by Golub and Pereyra (1973) is that the residual $\mathbf{d}_{k,j} - \bar{c}_{k,j}\mathbf{s}_{k,j}(\mathbf{m})$ is in the null space of the adjoint of the second term in the Jacobian, so that this second term vanishes when computing the gradient. Indeed,

$$\left(\frac{\partial \bar{c}_{k,j}}{\partial \mathbf{m}}\right)^{*}\mathbf{s}_{k,j}^{*}(\mathbf{d}_{k,j} - \bar{c}_{k,j}\mathbf{s}_{k,j}) = 0, \tag{A-9}$$

as can be verified by substituting the explicit expression for the optimal source weight presented earlier. This was also noted by Rickett (2013).

Computing the Hessian

The Hessian can be computed by applying the chain rule to the gradient:

$$\nabla_{\mathbf{m}}^2\bar{\Phi}(\mathbf{m}) = \nabla_{\mathbf{m}}^2\Phi(\mathbf{m},\bar{\mathbf{c}}) + \nabla_{\mathbf{m},\mathbf{c}}^2\Phi(\mathbf{m},\bar{\mathbf{c}})\nabla_{\mathbf{m}}\bar{\mathbf{c}}, \tag{A-10}$$

where $\nabla_{\mathbf{m},\mathbf{c}}^2$ denotes the mixed second derivative with regards to $\mathbf{m}$ and $\mathbf{c}$. In this case none of the terms vanish, and we have to compute $\nabla_{\mathbf{m}}\bar{\mathbf{c}}$ to obtain the exact Hessian. These expressions were presented in Aravkin and van Leeuwen (2012), and we suggested using only the first term as an approximation.

To better understand this approximation from an optimization point of view, we compute this term by considering the function $F(\mathbf{m}) = \nabla_{\mathbf{c}}\Phi(\mathbf{m},\bar{\mathbf{c}})$ and computing its gradient:

$$\nabla_{\mathbf{m}}F(\mathbf{m}) = \nabla_{\mathbf{m},\mathbf{c}}^2\Phi(\mathbf{m},\bar{\mathbf{c}}) + \nabla_{\mathbf{c}}^2\Phi(\mathbf{m},\bar{\mathbf{c}})\nabla_{\mathbf{m}}\bar{\mathbf{c}}. \tag{A-11}$$

Using the stationarity argument, we find that $F(\mathbf{m}) = 0$, and hence $\nabla_{\mathbf{m}}F(\mathbf{m}) = 0$, yielding

$$\nabla_{\mathbf{m},\mathbf{c}}^2\Phi(\mathbf{m},\bar{\mathbf{c}}) = -\nabla_{\mathbf{c}}^2\Phi(\mathbf{m},\bar{\mathbf{c}})\nabla_{\mathbf{m}}\bar{\mathbf{c}}, \tag{A-12}$$

which gives us

$$\nabla_{\mathbf{m}}\bar{\mathbf{c}} = -(\nabla_{\mathbf{c}}^2\Phi)^{-1}\nabla_{\mathbf{m},\mathbf{c}}^2\Phi. \tag{A-13}$$

Finally, the Hessian of the reduced objective can be written as

$$\nabla_{\mathbf{m}}^2\bar{\Phi}(\mathbf{m}) = \nabla_{\mathbf{m}}^2\Phi - \nabla_{\mathbf{m},\mathbf{c}}^2\Phi(\nabla_{\mathbf{c}}^2\Phi)^{-1}\nabla_{\mathbf{m},\mathbf{c}}^2\Phi. \tag{A-14}$$

A similar expression is also presented by Ruhe and Wedin (1980). Note also that this is precisely the Schur complement of $\bar{\mathbf{c}}$ in the full Hessian of $\Phi(\mathbf{m},\mathbf{c})$, and equation A-14 makes it very clear that ignoring the correction term amounts to a positive definite modification to $\nabla_{\mathbf{m}}^2\bar{\Phi}(\mathbf{m})$, or to its GN approximation, depending on the algorithm. Such a modification may have a positive rather than a detrimental effect on the algorithm because the correction term may actually render the Hessian indefinite, possibly causing the GN method to diverge. We therefore encourage its further investigation both from a theoretical and numerical perspective.

APPENDIX B

NUMERICAL EXPERIMENTS

To shed some preliminary light on the effect of the correction term, we conducted a few numerical experiments on a toy problem. The MATLAB code used to perform the experiments can be found at https://github.com/tleeuwen/Variable-Projection-for-FWI.git.

The simulated data are given by

$$\mathbf{s}_{k,j}(\mathbf{m}) = PA_k(\mathbf{m})^{-1}\mathbf{q}_j, \tag{B-1}$$

where $A_k = \omega_k^2\mathrm{diag}(\mathbf{m}) + \nabla^2$ is a five-point finite-difference discretization of the Helmholtz operator (with absorbing boundary conditions) for the kth frequency, $\mathbf{q}_j$ is the jth source function, and P is the detection operator that samples the wavefield at the receiver locations. The Jacobian is given by

$$J_{k,j} = \frac{\partial \mathbf{s}_{k,j}(\mathbf{m})}{\partial \mathbf{m}} = PA_k(\mathbf{m})^{-1}G_k(\mathbf{m})\mathrm{diag}(\mathbf{u}_{k,j}), \tag{B-2}$$

where $\mathbf{u}_{k,j} = A_k(\mathbf{m})^{-1}\mathbf{q}_j$ and $G_k(\mathbf{m})$ contains derivatives of A_k with respect to $\mathbf{m}$. For the experiments, we ignore the contributions of the

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

boundaries to the gradient, in which case $G_k(\mathbf{m}) = \omega_k^2 \mathrm{diag}(\mathbf{w})$, where $\mathbf{w} = 1$ on the internal grid points and $\mathbf{w} = 0$ on the boundary.

The gradient of the source weight is given by

$$\frac{\partial \bar{c}_{k,j}}{\partial \mathbf{m}} = \frac{1}{\|\mathbf{s}_{k,j}\|_2^2} J_{k,j}^*(\mathbf{d}_{k,j} - 2\bar{c}_{k,j}\mathbf{s}_{k,j}). \tag{B-3}$$

Then, the Jacobian of the reduced objective is given by (see equation A-8)

$$\bar{J}_{k,j} = \bar{c}_{k,j} J_{k,j} + \mathbf{s}_{k,j} \frac{\partial \bar{c}_{k,j}}{\partial \mathbf{m}}. \tag{B-4}$$

The gradient and Hessian are now given by

$$\mathbf{g} = \sum_{k,j} \bar{J}_{k,j}^*(\mathbf{s}_{k,j} - \bar{c}_{k,j}\mathbf{s}_{k,j}), \tag{B-5}$$

$$H = \sum_{k,j} \bar{J}_{k,j}^* \bar{J}_{k,j}. \tag{B-6}$$

Toy problem

For the following experiments, we use the velocity model depicted in Figure 1. We use either a reflection or a transmission (crosswell) setup. For all the experiments, we use a grid spacing of 20 m, three frequencies: {3,5,10} Hz and 49 equispaced sources and receivers. The observed data are weighted with a random amplitude factor for each source and frequency. For all the experiments, the initial model $\mathbf{m}_0$ is constant with a velocity of 2000 m/s.

Experiment 1

To asses how much the correction terms contribute to the Hessian, we consider the quadratic model:

$$q(\mathbf{m}_0, \Delta\mathbf{m}) = \bar{\Phi}(\mathbf{m}_0) + \Delta\mathbf{m}^T \mathbf{g}(\mathbf{m}_0) + \frac{1}{2}\Delta\mathbf{m}^T H(\mathbf{m}_0)\Delta\mathbf{m} \tag{B-7}$$

and compare this to the actual misfit $\bar{\Phi}(\mathbf{m}_0 + \Delta\mathbf{m})$. Figure 2 shows the quadratic models $q(\mathbf{m}_0, \alpha\Delta\mathbf{m})$ with and without correction terms as well as the actual misfit $\bar{\Phi}(\mathbf{m}_0 + \alpha\Delta\mathbf{m})$, with $\Delta\mathbf{m} = \nabla\bar{\Phi}(\mathbf{m}_0)/\|\nabla\bar{\Phi}(\mathbf{m}_0)\|_2$ as a function of α. For this particular setup, the correction terms do not contribute significantly. In the next experiment, we investigate how the correction term affects the convergence when using a GN method.

Experiment 2

We solve the optimization problem by iteratively updating the model

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \alpha_i \Delta\mathbf{m}_i. \tag{B-8}$$

The search direction $\Delta\mathbf{m}_i$ is either determined by applying the L-BFGS inverse Hessian to $-\mathbf{g}_i$ using a two-loop recursion (see Nocedal and Wright [2000], Algorithm 7.4), or by solving $H_i\Delta\mathbf{m}_i = -\mathbf{g}_i$ up to a prescribed relative tolerance (10^{-1} in this case) using the conjugate gradient (CG), were, H_i can either be the conventional GN Hessian or contain the correction terms. We use a weak Wolfe line search to determine α_i and stop the iterations when the norm of the gradient drops below a preset relative tolerance (10^{-3} in this case). We compute the cost of the inversion in terms of the number of PDE solves; two PDE solves for a gradient computation and three for a Hessian-vector product.

Table 1. Overview of iteration counts and computational cost in terms of the number of PDE solves for the transmission (a) and reflection (b) experiments.

Transmission	L-BFGS	GN w/o corr.	GN w corr.
Iterations	34	5	3
PDE solves	76	477	104
Reflection	**L-BFGS**	**GN w/o corr.**	**GN w corr.**
Iterations	104	15	7
PDE solves	222	1235	889

Figure 3 shows the results for the transmission experiment. Here, the correction terms hardly affect the rate of convergence, although the results look slightly better when including the correction terms. For the reflection experiment, shown in Figure 4, the correction terms significantly affect the rate of convergence but have little effect on the final reconstruction. In both cases, the L-BFGS results are nearly identical to the results obtained with the GN method with correction term; however, it is significantly cheaper (in terms of the number of PDE solves) as can be seen from Table 1.

REFERENCES

Aravkin, A. Y., J. V. Burke, and G. Pillonetto, 2012, Robust and trend-following Kalman smoothers using Student's T, *in* M. Kinnaert, ed., SYSID 2012: 16th IFAC Symposium on System Identification: IFAC, 1215–1220, doi: 10.3182/20120711-3-BE-2027.00283.

Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: Inverse Problems, **28**, 115016, doi: 10.1088/0266-5611/28/11/115016.

Aravkin, A. Y., T. van Leeuwen, H. Calandra, and F. J. Hermann, 2012, Source estimation for frequency-domain FWI with robust penalties: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, P018.

Golub, G., and V. Pereyra, 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM Journal on Numerical Analysis, **10**, 413–432, doi: 10.1137/0710036.

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme for frequency-domain full-waveform inversion: Geophysics, **78**, no. 6, R249–R257, doi: 10.1190/geo2012-0351.1.

Nocedal, J., and S. Wright, 2000, Numerical optimization: Springer, Springer Series in Operations Research.

Rickett, J., 2013, The variable projection method for waveform inversion with an unknown source function: Geophysical Prospecting, **61**, 874–881, doi: 10.1111/1365-2478.12008.

Ruhe, A., and P. A. Wedin, 1980, Algorithms for separable nonlinear least squares problems: SIAM Review, **22**, 318–337, doi: 10.1137/1022057.

van Leeuwen, T., and F. J. Herrmann, 2013, Fast waveform inversion without source-encoding: Geophysical Prospecting, **61**, 10–19, doi: 10.1111/j.1365-2478.2012.01096.x.

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

Reply to the discussion

Maokun Li[4], James Rickett[5], and Aria Abubakar[4]

First of all, we would like to thank the authors for their interest in our paper. The detailed discussions and numerical tests help us to further understand the variable projection method and its application to seismic full-waveform inversion (FWI). We also thank the authors for adding references on this topic (Aravkin and van Leeuwen, 2012).

As we described in our paper (Li et al., 2013), the residual vector lies in the null space of the Jacobian matrix of the calibration coefficient. Therefore, we can still maintain the accuracy of the inversion algorithms, such as the nonlinear conjugate gradient method, in which we only need gradient information (product of the Jacobian matrix with the residual vector). We agree with the authors in the discussion that the gradient of the data misfit function in Aravkin et al. (2012) is accurate. We apologize for the potentially misleading information to the readers in this regard.

Moreover, both Aravkin and van Leeuwen (2012) and we agree that the approximated Jacobian without the correction term will affect the accuracy of other algorithms requiring more than the gradient information. The effect needs further investigation. On the other hand, computing the full Jacobian matrix needs few additional resources in the frequency-domain FWI algorithm. Therefore, we choose to use the full Jacobian matrix in our computation. As a comparison, we rerun the Marmousi data inversion using Jacobian with and without this approximation. We use the same settings as in our paper (Li et al., 2013). Note that we added 10% of error to the data. First, we inverted the data using the preconditioned nonlinear conjugate gradient (P-NLCG) method. The results are shown in Figure 5. We observe that the inverted model using the approximated Jacobian is close to the one using the full Jacobian. This verifies our guess on the accuracy of gradient-based inversion algorithms. Compared with the true model, the inverted model misfits are 0.12584 using the approximated Jacobian and 0.12282 using the full Jacobian. The inverted model using the full Jacobian matches the true model slightly better than the one using the approximated Jacobian. This is because the P-NLCG method requires an approximated Hessian matrix to construct a preconditioner. Using the full Jacobian matrix can construct a better preconditioner to accelerate the convergence of inversion. We also plot the data misfit at each iteration in the inversion of the first frequency in Figure 6 and observe a faster convergence by using the full Jacobian to construct the Hessian.

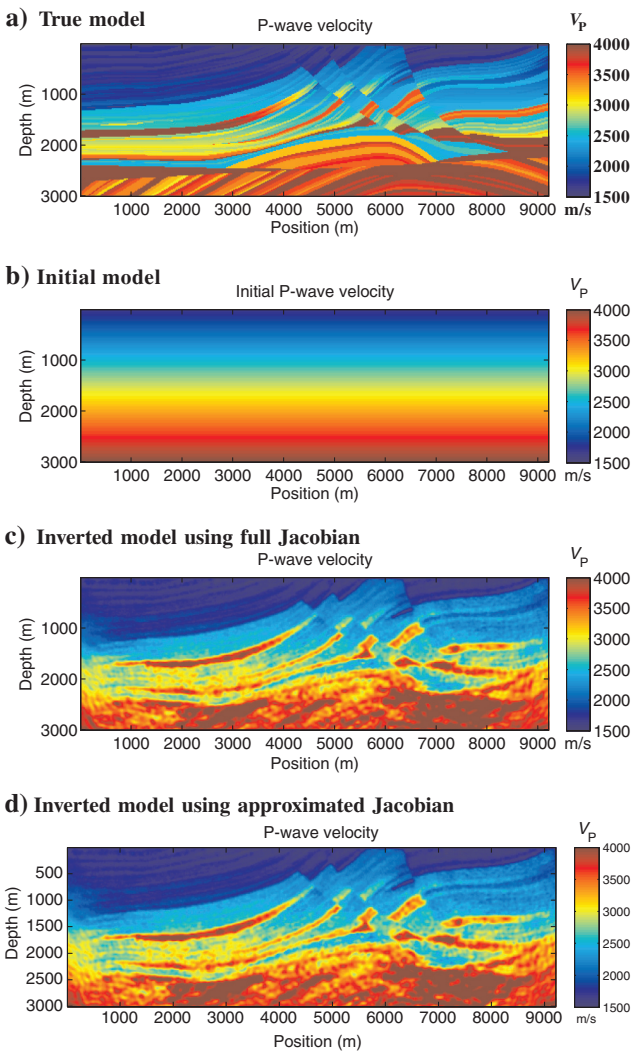

Figure 5. Marmousi data inversion using the P-NLCG method.

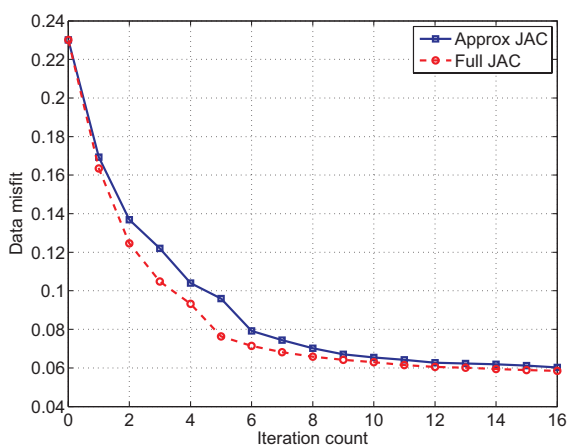

Figure 6. Comparison of data misfit at each iteration between using the approximated and full Jacobian.

[4]Schlumberger-Doll Research, Mathematics and Modeling Department, Cambridge, Massachusetts, USA. E-mail: mli7@slb.com; aria.abubakar@gmail.com.
[5]Schlumberger-Gould Research, Cambridge, UK. E-mail: jrickett@slb.com.

Downloaded 01/07/15 to 131.211.104.187. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

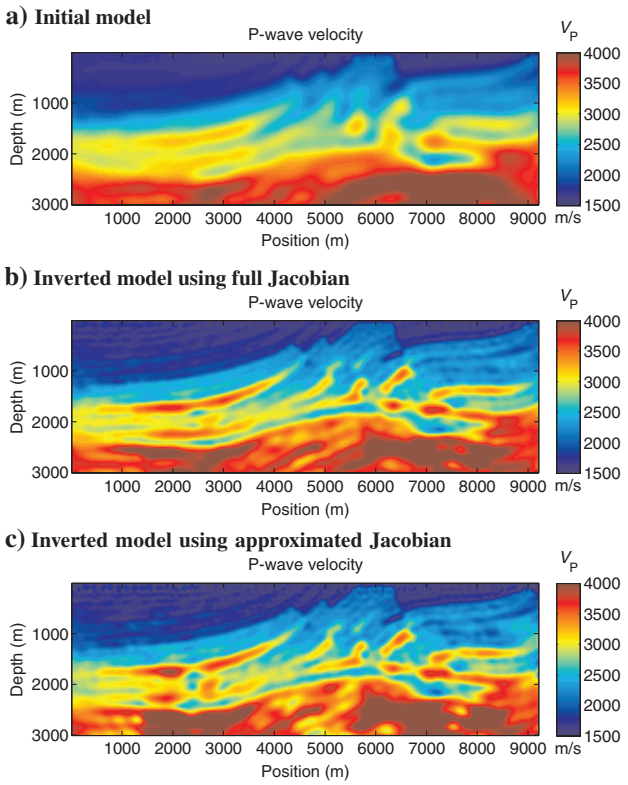

Figure 7. Marmousi data inversion at 6 Hz using the GN method.

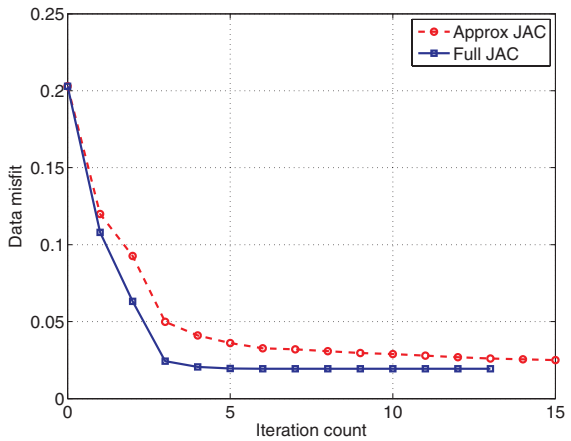

Figure 8. Comparison of data misfit at each iteration between using the approximated and full Jacobian in the GN method.

We also run the Marmousi data inversion using the Gauss-Newton (GN) inversion algorithm. First, we run a single-frequency data inversion at 6 Hz. The initial model is shown in Figure 7a. We add 5% noise to the data. The other settings are the same as in Li et al. (2013). The inverted models using the full and approximate Jacobian are shown in Figure 7b and 7c with model misfit 0.11486 and 0.13284, respectively. The inverted model using the full Jacobian is slightly better than the one using the approximate Jacobian. Figure 8 shows the data misfit at every iteration. We also observe a fast convergence using the full Jacobian.

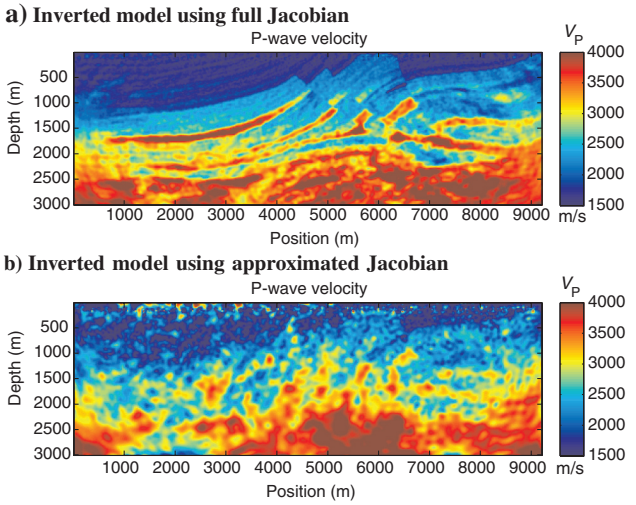

Figure 9. Marmousi data inversion using the GN method.

We then run the Marmousi data inversion using the GN method with all the data as in Li et al. (2013). Note that here we added 10% noise to the data. We set the maximum number of inversion iterations to five for every frequency. The inverted model using the full Jacobian is shown in Figure 9a. The inversion process reconstructs the Marmousi model very well with a model misfit of 0.11942. The inverted model using the approximated Jacobian is shown in Figure 9b with a larger model misfit of 0.18245. From this comparison, we think that increasing the accuracy in the Jacobian could increase the robustness of the GN inversion because it leads to a better approximation of the Hessian.

In conclusion, we observe that the gradient vector is still exact with the approximated Jacobian in the variable projection method. Therefore, the nonlinear conjugate gradient method will maintain its accuracy. For other inversion methods requiring more than gradient information (P-NLCG or GN method), the convergence of the inversion could be changed by the approximation. The model inverted using the full Jacobian can have a better accuracy than the model inverted using the approximate Jacobian. The data misfit between the measured and simulated data is also lower. Moreover, when the data noise level increases, having an accurate Jacobian can increase the robustness of the inversion. Therefore, we think using a full Jacobian is helpful in the variable projection method if the computational cost is affordable. In the end, we would like to thank the authors in the discussion again. We look forward to more studies to improve the accuracy and efficiency of this scheme.

REFERENCES

Aravkin, A., T. van Leeuwen, H. Calandra, and F. Herrmann, 2012, Source estimation for frequency-domain FWI with robust penalties: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, P018.

Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: Inverse Problems, **28**, 115016, doi: 10.1088/0266-5611/28/11/115016.

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme for frequency-domain full-waveform inversion: Geophysics, **78**, no. 6, R249–R257, doi: 10.1190/geo2012-0351.1.