Seismic full waveform inversion with nonlinear sensitivities

Kris Innanen

ABSTRACT

The convergence properties and physical interpretability of full waveform inversion updates are key issues as we contemplate practical FWI. One extension of standard FWI updates that has been only superficially broached by the seismic community is the idea that the sensitivity, or Fréchet kernel, used in the gradient calculation could be improved by accommodating higher order model/wavefield relations. We present and analyze an approach to constructing a second order sensitivity, and demonstrate its natural accommodation of nonlinear reflection amplitudes of the type encountered when contrasts causing reflections are large. In a companion paper we have proposed a way for second order data-model interactions to be properly incorporated in the inverse Hessian. Those are not incorporated in the updates we study here, and evidently do not adversely affect the inclusion in FWI of second-order reflectivity.

INTRODUCTION

Seismic full waveform inversion (Lailly, 1983; Tarantola, 1984), which is being pursued in broadband land multicomponent settings by CREWES (Margrave et al., 2013; Innanen, 2014a), involves defining an objective function (usually based on a least-squares norm) and a starting model, and iteratively updating towards the minimum, assuming no local minima obstruct the path. Descent-based methods like Newton, quasi-Newton, Gauss-Newton, and gradient-based (Virieux and Operto, 2009; Operto et al., 2013; Pratt, 1999; Shin et al., 2001; Plessix et al., 2013) all make use of the local *gradient* of the objective function. The idea is that in order to find a valley bottom, going downhill, or something like downhill, is your best option.

The gradient is the multidimensional derivative of the objective function with respect to the model parameters. Because the objective function is itself a function of the forward model, the gradient calculation ends up involving the derivative of the forward modelling functional with respect to the model parameters. This internal derivative can be studied in the framework of the Fréchet derivative (McGillivray and Oldenburg, 1990) and/or Gâteaux derivative, and specific forms can be arrived at via the Born approximation (e.g., Innanen, 2014b), or rejection of nonlinearity in the relationship between the perturbed field and the perturbed medium (Tarantola, 1984), or via adjoint state techniques (Plessix, 2006; Liu and Trompe, 2006; Yedlin and Van Vorst, 2010).

Wu and Zheng (2014); Wu et al. (2014) point out that within the Fréchet derivative the linearization of the medium perturbation and field perturbation relationship has accuracy and convergence consequences for full waveform inversion, and formally invoke a nonlinear renormalized solution of the scattering problem with which to compute the sensitivity. Such extensions have also been introduced in optical tomography (Kwon and Yazici, 2010).

We will do something similar but with a specific focus on reflected seismic amplitudes.

We point out that at any one step in iterated inversion, the objective function itself is approximated quadratically — i.e., as a quantity with at most second-order variability with respect to the model parameters. What we attempt is to calculate the sensitivities accurate to the same order to which we have approximated the objective function – namely, second order. This leads to updates which in any one step precisely account for first and second order data-model behaviour. This may have significant applicability in the incorporation of AVO in FWI. Standard FWI updates with the correct inverse Hessian additions have been shown to be, iteration by iteration, equivalent to applying linearized AVO inversion to reflection data (Innanen, 2014b). If our second order sensitivities have been constructed properly, we should likewise be able to see second order AVO (Stovas and Ursin, 2001; Innanen, 2013) being correctly accounted for in each FWI iteration, when the FWI equations are applied in circumstances matching those in which AVO / AVA is typically carried out.

This paper proceeds in two steps. After introducing the terms and equations, the inverse scattering series, nominally itself a framework for inversion (Weglein et al., 2003), is used to arrive at nonlinear sensitivity calculations. Essentially, the denominator of the ratio $\delta G/\delta s$, prior to taking the limit $\delta s \to 0$, is built up from the sum of the first and second order inverse scattering solutions for the perturbation. The general scalar formula for 2nd order sensitivities as included in a Gauss-Newton update is the result of this first step. The second step is to demonstrate analytically that the first iteration of a reflection full waveform inversion scheme using these sensitivities reconstructs a single interface correctly to second order in one step. A rough illustration of the subsequent convergence over 10 iterations is given, suggesting that the uptick in convergence is significant.

Terms and equations

Equations

The wave fields P giving rise to our seismic data will be assumed to satisfy the scalar equation

$$\left[\nabla^2 + \frac{\omega^2}{c^2(\mathbf{r})}\right] P(\mathbf{r}, \mathbf{r}_s) = \delta(\mathbf{r} - \mathbf{r}_s) \tag{1}$$

in the space frequency domain $P = P(\mathbf{r}, \mathbf{r}_s, \omega)$ for a receiver at $\mathbf{r}$ and a source at $\mathbf{r}_s$. For the purposes of inversion we express the velocity in terms of the squared slowness parameter s:

$$\left[\nabla^2 + \omega^2 s(\mathbf{r})\right] P(\mathbf{r}, \mathbf{r}_s) = \delta(\mathbf{r} - \mathbf{r}_s). \tag{2}$$

Here $s(\mathbf{r})$ is the actual distribution of wave velocities in the subsurface. We suppose we have access to a reference medium $s_0(\mathbf{r})$, different from $s(\mathbf{r})$, and also the wave field G propagating through it, which satisfies

$$\left[\nabla^2 + \omega^2 s_0(\mathbf{r})\right] G(\mathbf{r}, \mathbf{r}_s) = \delta(\mathbf{r} - \mathbf{r}_s). \tag{3}$$

A Lippmann-Schwinger or scattering equation can be developed relating P and G:

$$P(\mathbf{r}, \mathbf{r}_s) = G(\mathbf{r}, \mathbf{r}_s) - \omega^2 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}')\delta s(\mathbf{r}')P(\mathbf{r}', \mathbf{r}_s). \tag{4}$$

The second term on the right-hand side can therefore be equated to the difference between P and G,

$$\delta G(\mathbf{r}_g, \mathbf{r}_s) = P(\mathbf{r}_g, \mathbf{r}_s) - G(\mathbf{r}_g, \mathbf{r}_s) = \delta P(\mathbf{r}_g, \mathbf{r}_s), \tag{5}$$

which we can variously associate with a small change in the field at any notional point in the medium (which we will refer to as δG, as in the left side of equation 5), or the residuals, the difference between a modelled and a measured field on some well defined measurement surface (which we will refer to as δP, as in the right side of equation 5).

Full waveform inversion quantities

Full waveform inversion is the search for the distribution of medium properties which minimizes the sum of the squares of the differences between measured and modelled data. We seek, in other words, the medium for which

$$\phi(s) = \frac{1}{2} \sum_{s,g} \int d\omega |\delta P(\mathbf{r}_g, \mathbf{r}_s)|^2 \tag{6}$$

is at its smallest value. This is typically done iteratively, that is we have in hand a current model iterate $s_{n-1}(\mathbf{r})$ and we work to calculate an update $\delta s_{n-1}(\mathbf{r})$ in order to determine the next model iterate:

$$s_n(\mathbf{r}) = s_{n-1}(\mathbf{r}) + \delta s_{n-1}(\mathbf{r}). \tag{7}$$

In a Newton method the update has the basic form

$$\delta s_n(\mathbf{r}) = \int d\mathbf{r}' H^{-1}(\mathbf{r}, \mathbf{r}') g(\mathbf{r}'), \tag{8}$$

where g is referred to as the gradient and H^{-1} is referred to as the inverse Hessian. These are functional derivatives of the objective function ϕ:

$$g(\mathbf{r}) = \frac{\partial \phi(s)}{\partial s(\mathbf{r})}, \quad H(\mathbf{r}, \mathbf{r}') = \frac{\partial^2 \phi(s)}{\partial s(\mathbf{r}) \partial s(\mathbf{r}')}, \tag{9}$$

and H^1 is the functional inverse of H. It can be shown (Margrave et al., 2011) that the gradient is given by

$$g(\mathbf{r}) = -\sum_{g,s} \int d\omega \frac{\partial G(\mathbf{r}_g, \mathbf{r}_s)}{\partial s(\mathbf{r})} \delta P^*(\mathbf{r}_g, \mathbf{r}_s), \tag{10}$$

where δP^* is the complex conjugate of the residuals (the difference between the data and the current modelled field), in a product with the sensitivities or Jacobian $\partial G/\partial s$. In this paper we will not be concerned with the full form of the Hessian, but rather an approximation of it which when used forms a Gauss-Newton update:

$$H(\mathbf{r}, \mathbf{r}') \approx H_{\text{GN}}(\mathbf{r}, \mathbf{r}') = \sum_{g,s} \int d\omega \frac{\partial G(\mathbf{r}_g, \mathbf{r}_s)}{\partial s(\mathbf{r})} \frac{\partial G(\mathbf{r}_g, \mathbf{r}_s)}{\partial s(\mathbf{r}')} \tag{11}$$

Our focus in this paper will be on the sensitivities.

LINEAR AND NONLINEAR SENSITIVITIES

In a companion paper, we point out that the full Newton update is inappropriate when the sensitivities are approximated linearly (which we will define presently). Here we will devise a response to this issue, and extend the sensitivity calculation to include low order nonlinear corrections.

General sensitivities

Let us begin with a general discussion on forms for sensitivities. This requires us to analyze a ratio of the change in the wave (i.e., δG) produced by a change in the medium (i.e., δs) as the latter tends to zero. A common means to arrive at this relationship is through the Born series:

$$\begin{aligned}\delta G(\mathbf{r}_g, \mathbf{r}_s) =& - \omega^2 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}')\delta s(\mathbf{r}')G(\mathbf{r}', \mathbf{r}_s) \\ &+ \omega^4 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}')\delta s(\mathbf{r}') \int d\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'')\delta s(\mathbf{r}'')G(\mathbf{r}'', \mathbf{r}_s) \\ &- ...\end{aligned} \tag{12}$$

which is subject to some manipulations we will turn to presently. First, however, we will consider different degrees of accuracy with which the denominator of the ratio $\delta G/\delta s$ can be included in the sensitivity. Equation (12) can be formally inverted (e.g., Weglein et al., 2003) through an expansion of $\delta s(\mathbf{r})$ in series. We write

$$\delta s(\mathbf{r}) = \delta s_1(\mathbf{r}) + \delta s_2(\mathbf{r}) + ..., \tag{13}$$

where $\delta s_n(\mathbf{r})$ is defined to be nth order in $\delta G(\mathbf{r}_g, \mathbf{r}_s)$, then substitute this series into equation (12) and equate like orders. At first order we obtain

$$\delta G(\mathbf{r}_g, \mathbf{r}_s) = -\omega^2 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}')\delta s_1(\mathbf{r}')G(\mathbf{r}', \mathbf{r}_s), \tag{14}$$

whereas at second order we have

$$\begin{aligned}&\int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}')\delta s_2(\mathbf{r}')G(\mathbf{r}', \mathbf{r}_s) \\ &\qquad = \omega^2 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}')\delta s_1(\mathbf{r}') \int d\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'')\delta s_1(\mathbf{r}'')G(\mathbf{r}'', \mathbf{r}_s),\end{aligned} \tag{15}$$

etc. In a *direct inversion* formulation (e.g., Zhang and Weglein, 2009a,b) the right-hand side of equation (13) is built up order by order, with equation (14) first being solved for $\delta s_1(\mathbf{r})$, and then that result being used to solve for $\delta s_2(\mathbf{r})$ in equation (16), and so on. Here we will consider the order-by-order construction of $\delta s(\mathbf{r})$ not as an end in itself, but rather as a means to calculate approximations for the denominator of the sensitivities $\delta G/\delta s$. We define indexed sensitivities of the following form. At order one we have:

$$\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s)}{\partial s(\mathbf{r})}\right)_1 = \lim_{\delta s \to 0} \frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\delta s(\mathbf{r})}, \quad \delta s(\mathbf{r}) \approx \delta s_1(\mathbf{r}); \tag{16}$$

whereas at order two we have

$$\left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})}\right)_2 = \lim_{\delta s\to 0}\frac{\delta G(\mathbf{r}_g,\mathbf{r}_s)}{\delta s(\mathbf{r})},\quad \delta s(\mathbf{r}) \approx \delta s_1(\mathbf{r}) + \delta s_2(\mathbf{r}), \tag{17}$$

or, generally, at order N we have

$$\left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})}\right)_N = \lim_{\delta s\to 0}\frac{\delta G(\mathbf{r}_g,\mathbf{r}_s)}{\delta s(\mathbf{r})},\quad \delta s(\mathbf{r}) \approx \sum_{n=1}^{N}\delta s_n(\mathbf{r}). \tag{18}$$

Linearized sensitivities

The general expression for sensitivities in equation (18) reduces to the familiar form used in seismic FWI as follows. We begin with equation (14):

$$\delta G(\mathbf{r}_g,\mathbf{r}_s) = -\omega^2\int d\mathbf{r}' G(\mathbf{r}_g,\mathbf{r}')\delta s_1(\mathbf{r}')G(\mathbf{r}',\mathbf{r}_s). \tag{19}$$

From this we must determine the ratio of δG and δs (in this case to be approximated by δs_1, the latter at a particular point in space $\mathbf{r}$. To do this we replace $\delta s_1(\mathbf{r}')$ with

$$\delta s_1(\mathbf{r}') = \delta s_1(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}'), \tag{20}$$

from which we obtain

$$\delta G(\mathbf{r}_g,\mathbf{r}_s) = -\omega^2\delta s_1(\mathbf{r})G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}_s). \tag{21}$$

We then solve this for $\delta s_1(\mathbf{r})$ and make the first inverse Born approximation $\delta s(\mathbf{r}) \approx \delta s_1(\mathbf{r})$:

$$\delta s(\mathbf{r}) \approx \delta s_1(\mathbf{r}) = -\delta G(\mathbf{r}_g,\mathbf{r}_s)[\omega^2 G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}_s)]^{-1}. \tag{22}$$

This quantity* is now ready to be used in the denominator of the first order sensitivity expression in equation (16):

$$\begin{aligned}\left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})}\right)_1 &= \lim_{\delta s\to 0}\frac{\delta G(\mathbf{r}_g,\mathbf{r}_s)}{\delta s(\mathbf{r})}\\ &= \lim_{\delta s\to 0}\left(\frac{\delta G(\mathbf{r}_g,\mathbf{r}_s)}{\delta G(\mathbf{r}_g,\mathbf{r}_s)\left[-\omega^2 G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}_s)\right]^{-1}}\right)\\ &= -\omega^2 G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}_s),\end{aligned} \tag{23}$$

*Seeing an equation like (22) could make a practical observer question this whole process. First of all, having that equation in hand, it would appear our job is done – we wanted to know $\delta s(\mathbf{r})$, and there it is: a formula for $\delta s(\mathbf{r})$. Second of all, everything was prescriptive right up until equation (20), but then we made a rather presumptuous looking replacement $\delta s(\mathbf{r}') \to \delta s(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')$. Do we really think the Earth model is of this form? By what right do we do this? The answer to the first question is, yes, this is, in a sense, a formula for inversion for $\delta s(\mathbf{r})$. You could actually use it, too, *if* your initial medium model and the actual model were identical except at an infinitely local point $\mathbf{r}$. If you could say with certainty that it was, then from a single measurement of δG (at frequency ω) you could determine what that change was, to first order. Our interest in more complicated differences between the initial velocity model and the actual Earth makes this impractical. The short answer to the second question is that we did this not because the model $\delta s(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')$ was a realistic model of the Earth, but, rather, because the particular δG associated with the model $\delta s(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')$, not the δG caused by the full perturbation, is the δG called for by the definition of the gradient formula.

from which we recover the standard form for the full waveform inversion sensitivities as it is currently used in the geophysics community.

Sensitivities associated with 2nd order collocated scattering

The usefulness of the general expression in equation (18) is that it permits us to extend beyond equation (23) to nonlinear forms if desired. Here we will advance one such form and subsequently analyze it. We will now assume a second order medium, meaning we accept the approximation $\delta s(\mathbf{r}) \approx \delta s_1(\mathbf{r}) + \delta s_2(\mathbf{r})$. Having determined $\delta s_1(\mathbf{r})$ in the previous section:

$$\delta s_1(\mathbf{r}) = -\frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s)}, \tag{24}$$

it remains to use equation (16) and determine $\delta s_2(\mathbf{r})$. Inclusion of a second order term invokes "double scattering", i.e., contains a picture of wave data which admits propagation from the source to a scatter point, propagation from that scatter point to a second scatter point, followed by propagation to the receiver. Here we will restrict ourselves to a subset of these contributions to the wave field, in which the two scattering points are collocated in space. The data-medium nonlinearity we are including is, in other words, only that which appears from local interactions of the wave with the medium. This restriction is put in place by assuming the forms

$$\begin{aligned} \delta s_1(\mathbf{r}') &= \delta s_1(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'), \\ \delta s_1(\mathbf{r}'') &= \delta s_1(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}''), \\ \delta s_2(\mathbf{r}') &= \delta s_2(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'). \end{aligned} \tag{25}$$

These forms when substituted into equation (16) lead to

$$\delta s_2(\mathbf{r}) = \omega^2 \delta s_1^2(\mathbf{r}) G(\mathbf{r}, \mathbf{r}). \tag{26}$$

The field G grows without bound as the source and receiver points converge, bringing the meaning of the quantity $G(\mathbf{r}, \mathbf{r})$ into question. However, such issues are seen in the Fourier domain to be easily managed by taking principle values of inverse Fourier transform integrals, so we will not dwell on the value of $G(\mathbf{r}, \mathbf{r})$ at the moment but assume it can be assigned a finite value. Substituting equation (24) into equation (26) we have

$$\begin{aligned} \delta s_2(\mathbf{r}) &= \omega^2 \delta s_1^2(\mathbf{r}) G(\mathbf{r}, \mathbf{r}) \\ &= \omega^2 \left[-\frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s)} \right]^2 G(\mathbf{r}, \mathbf{r}) \\ &= \frac{\delta G(\mathbf{r}_g, \mathbf{r}_s) \delta G^*(\mathbf{r}_g, \mathbf{r}_s)}{\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s) G^*(\mathbf{r}_g, \mathbf{r}) G^*(\mathbf{r}, \mathbf{r}_s)} G(\mathbf{r}, \mathbf{r}), \end{aligned} \tag{27}$$

and thus the sum $\delta s_1 + \delta s_2$ is

$$\begin{aligned} \delta s(\mathbf{r}) &\approx \delta s_1(\mathbf{r}) + \delta s_2(\mathbf{r}) \\ &= -\frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s)} \left[1 - \frac{\delta G^*(\mathbf{r}_g, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r})}{G^*(\mathbf{r}_g, \mathbf{r}) G^*(\mathbf{r}, \mathbf{r}_s)} \right], \end{aligned} \tag{28}$$

which permits us to form the denominator of the 2nd order sensitivity formula in equation (17). Because δs_2 is second order in δG, by definition, we notice that in forming the ratio $\delta G/\delta s$ there will be a "left over" δG not folded into the sensitivity expression. This free variation in the field we replace with the residuals $\delta G(\mathbf{r}_g, \mathbf{r}_s) \to \delta P(\mathbf{r}_g, \mathbf{r}_s)$:

$$\delta s(\mathbf{r}) = -\frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s)} \left[1 - \frac{\delta P^*(\mathbf{r}_g, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r})}{G^*(\mathbf{r}_g, \mathbf{r}) G^*(\mathbf{r}, \mathbf{r}_s)}\right], \tag{29}$$

so that we have, finally,

$$\begin{aligned}
\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s)}{\partial s(\mathbf{r})}\right)_2 &= \lim_{\delta s \to 0} \frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\delta s(\mathbf{r})} \\
&= \lim_{\delta s \to 0} \delta G(\mathbf{r}_g, \mathbf{r}_s) \left[-\frac{\delta G(\mathbf{r}_g, \mathbf{r}_s)}{\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s)} \left(1 - \frac{\delta P^*(\mathbf{r}_g, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r})}{G^*(\mathbf{r}_g, \mathbf{r}) G^*(\mathbf{r}, \mathbf{r}_s)}\right)\right]^{-1} \\
&= -\omega^2 G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s) \left(1 + \frac{\delta P^*(\mathbf{r}_g, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r})}{G^*(\mathbf{r}_g, \mathbf{r}) G^*(\mathbf{r}, \mathbf{r}_s)}\right),
\end{aligned} \tag{30}$$

as the form for the second order sensitivity matrix.

GAUSS-NEWTON UPDATES WITH LINEAR/NONLINEAR SENSITIVITIES

We have pointed out that the nonlinearity included in the sensitivities is associated with collocated scattering, and thus (we expect) it will be most appropriate for dealing with "local" nonlinear wave/medium relations (e.g., nonlinear AVO). Because the nonlinear components of the inverse Hessian contribute off-diagonally (Innanen, 2014b), and thus are not involved with collocated scattering points, our expectation is that the inverse Hessian will not contribute significantly to the nonlinear updates we are currently interested in. We will thus focus on the nonlinear sensitivities, and de-emphasize nonlinear inverse Hessian contributions, forming Gauss-Newton updates of the following kind. Given a gradient of order n,

$$g_n(\mathbf{r}) = -\sum_{g,s} \int d\omega \left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s)}{\partial s(\mathbf{r})}\right)_n \delta P^*(\mathbf{r}_g, \mathbf{r}_s), \tag{31}$$

the associated Gauss-Newton update is

$$\delta s_{\mathrm{GN}_n}(\mathbf{r}) = -\int d\mathbf{r}' H_{\mathrm{GN}}^{-1}(\mathbf{r}, \mathbf{r}') g_n(\mathbf{r}'), \tag{32}$$

where

$$H_{\mathrm{GN}}(\mathbf{r}, \mathbf{r}') = \sum_{g,s} \int d\omega \omega^4 G(\mathbf{r}_g, \mathbf{r}') G(\mathbf{r}', \mathbf{r}_s) G^*(\mathbf{r}_g, \mathbf{r}) G^*(\mathbf{r}, \mathbf{r}_s). \tag{33}$$

1st order Gauss-Newton step

The general Gauss-Newton update with nth order sensitivities in equation (32) reduces to the standard Gauss-Newton update

$$\begin{aligned}
&\delta s_{\mathrm{GN}_1}(\mathbf{r}) \\
&\quad = -\int d\mathbf{r}' H_{\mathrm{GN}}^{-1}(\mathbf{r},\mathbf{r}') g_1(\mathbf{r}') \\
&\quad = -\int d\mathbf{r}' H_{\mathrm{GN}}^{-1}(\mathbf{r},\mathbf{r}') \left[\sum_{s,g} \int d\omega \omega^2 G(\mathbf{r}_g,\mathbf{r}') G(\mathbf{r}',\mathbf{r}_s) \delta P^*(\mathbf{r}_g,\mathbf{r}_s) \right],
\end{aligned} \tag{34}$$

used in seismic FWI given linearized ($n = 1$) sensitivities.

2nd order / collocated scattering Gauss-Newton step

Using the second order sensitivities arrived at in equation (30), we have instead:

$$\begin{aligned}
&\delta s_{\mathrm{GN}_2}(\mathbf{r}) \\
&\quad = -\int d\mathbf{r}' H_{\mathrm{GN}}^{-1}(\mathbf{r},\mathbf{r}') g_2(\mathbf{r}') \\
&\quad = -\int d\mathbf{r}' H_{\mathrm{GN}}^{-1}(\mathbf{r},\mathbf{r}') \left[\sum_{s,g} \int d\omega \omega^2 G(\mathbf{r}_g,\mathbf{r}') G(\mathbf{r}',\mathbf{r}_s) \left(1 + \frac{\delta P^*(\mathbf{r}_g,\mathbf{r}_s) G(\mathbf{r}',\mathbf{r}')}{G^*(\mathbf{r}_g,\mathbf{r}') G^*(\mathbf{r}',\mathbf{r}_s)} \right) \right].
\end{aligned} \tag{35}$$

We will investigate this update formula with an analytic and numerical example in the next section.

AN EXAMPLE OF THE 2ND ORDER GAUSS-NEWTON STEP

The character of a second order Gauss-Newton update can be investigated both analytically and numerically. To focus on its behaviour in relation to precritical reflection amplitudes, we will consider the reconstruction of a single scalar boundary using data at normal incidence. Primarily our interest is in the accuracy of a single update, but we will make some rough remarks about the convergence over several iterations.

One-interface model

The model we study is illustrated in Figure 1. We will assume a homogeneous background medium (Figure 1a) characterized by velocity c_0. The true model to be determined is illustrated in Figure 1b. It gives rise to a single reflected primary with reflection coefficient $R = (c_1 - c_0)/(c_1 + c_0)$. As we shall see, with perfectly sampled, full bandwidth data, and assuming that we know the correct velocity c_0 in the uppermost medium, at any given FWI iteration the reconstruction will have the basic form illustrated in Figure 1c. The interface is correctly located and its shape is correctly reconstructed, but the "height" of the step is not the same as that of the true model (c_1' rather than c_1). The aim here will be to quantify the rate at which c_1' converges to c_1.

Qualitative insight can be arrived at by viewing this as a full 1D experiment in itself, or a "zoom in" on a region of an Earth model where a jump in P-wave velocity overlies a background whose variation is smooth enough to locally appear homogeneous.

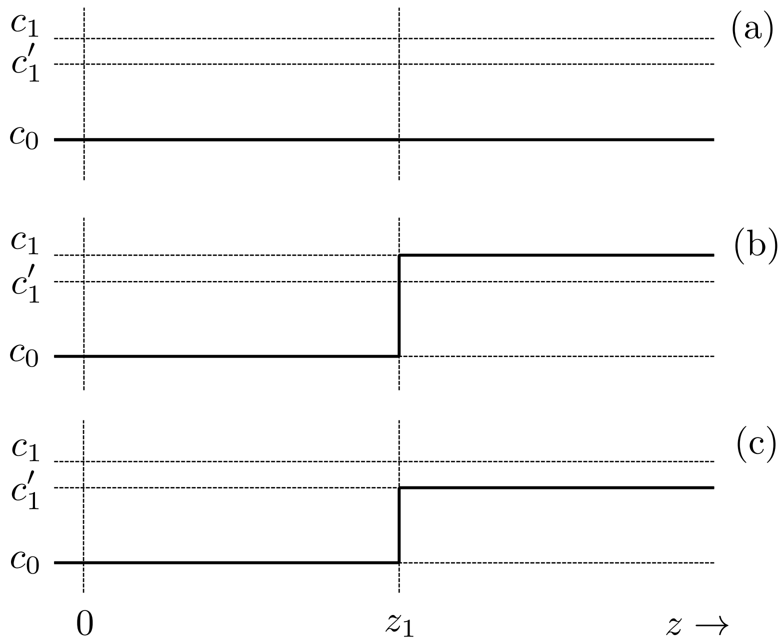

FIG. 1. Analytic model used to investigate the character of a second order Gauss-Newton update. (a) Homogeneous initial model characterized by c_0. (b) True model consisting of a single interface at depth z_1. (c) Any instance of a FWI update will tend to be in error by the velocity value estimated below the interface (i.e., c_1' rather than c_1.

Analytic evaluation of sensitivities and Gauss-Newton update

The wave field P measured above (i.e., to the left in Figure 1) the interface at $z_g = z_s = 0$ is given analytically by

$$P(z_g = 0, z_s = 0) = \frac{1}{i2k} + R\frac{e^{i2kz_1}}{i2k}, \tag{36}$$

where $k = \omega/c_0$. The first term is the direct arrival, and the second term is the reflection from z_1. The Green's function for the first iteration is

$$G(z, z') = \frac{e^{ik|z-z'|}}{i2k}. \tag{37}$$

Thus, the complex conjugate of the residuals for the first iteration, $P^*(0,0) - G^*(0,0)$, is

$$\delta P^*(0,0) = -R\frac{e^{-i2kz_1}}{i2k}. \tag{38}$$

Putting these into the linearized sensitivity formula, we obtain

$$\left(\frac{\partial G(0,0)}{\partial s(z)}\right)_1 = -\omega^2 G(0,z)G(z,0) = -\omega^2\frac{e^{i2kz}}{(i2k)^2} = \frac{c_0^2}{4}e^{i2kz}; \tag{39}$$

substituting them into the second order sensitivity formula produces instead

$$\begin{aligned}\left(\frac{\partial G(0,0)}{\partial s(z)}\right)_2 &= -\omega^2 G(0,z)G(z,0)\left[1-\frac{\delta P^*(0,0)G(z,z)}{G^*(0,z)G^*(z,0)}\right]\\ &= -\omega^2\frac{e^{i2kz}}{(i2k)^2}\left[1-\left(-R\frac{e^{-i2kz_1}}{i2k}\right)\left(\frac{1}{i2k}\right)\left(\frac{e^{-i2kz}}{(-i2k)^2}\right)^{-1}\right]\\ &= \frac{c_0^2}{4}e^{i2kz}\left[1+Re^{i2k(z-z_1)}\right].\end{aligned} \tag{40}$$

These each lead to a different gradient, g_1 and g_2 respectively:

$$\begin{aligned} g_1(z) &= -\int d\omega\left(\frac{\partial G(0,0)}{\partial s(z)}\right)_1 \delta P^*(0,0)\\ &= -\int d\omega\frac{c_0^2}{4}e^{i2kz}\left(-R\frac{e^{-i2kz_1}}{i2k}\right)\\ &= \frac{c_0^2 R}{4}\int d\omega\frac{e^{i2k(z-z_1)}}{i2k}\\ &= \frac{c_0^3 R}{8}\int d2k\frac{e^{i2k(z-z_1)}}{i2k}\\ &= \frac{\pi c_0^3 R}{4}S(z-z_1),\end{aligned} \tag{41}$$

and

$$\begin{aligned} g_2(z) &= -\int d\omega\left(\frac{\partial G(0,0)}{\partial s(z)}\right)_2 \delta P^*(0,0)\\ &= -\int d\omega\frac{c_0^2}{4}e^{i2kz}\left[1+Re^{i2k(z-z_1)}\right]\left(-R\frac{e^{-i2kz_1}}{i2k}\right)\\ &= \frac{c_0^3 R}{8}\int d2k\left[\frac{e^{i2k(z-z_1)}}{i2k}-R^2\frac{e^{i2k(2z-2z_1)}}{i2k}\right]\\ &= \frac{\pi c_0^3}{4}[R-2R^2]S(z-z_1).\end{aligned} \tag{42}$$

The last line above is a consequence of the property of the Heaviside function $S(ax+b) = S(x-b/a)$ if $a>0$ (Bracewell, 1978). To complete a single first- or second-order Gauss-Newton update we form the approximate Hessian operator

$$\begin{aligned} H_{\text{GN}}(z,z') &= \int d\omega\omega^4 G(0,z')G(z',0)G^*(0,z)G^*(z,0)\\ &= \int d\omega\omega^4\frac{e^{i2k(z'-z)}}{(i2k)^4}\\ &= \frac{c_0^5}{32}\int d2k e^{i2k(z'-z)}\\ &= \frac{c_0^5\pi}{16}\delta(z-z'),\end{aligned} \tag{43}$$

whose inverse is

$$H_{\text{GN}}^{-1}(z,z') = \frac{16}{c_0^5\pi}\delta(z-z'). \tag{44}$$

Combining the gradients and Hessians, we finally produce the two updates:

$$\begin{aligned}\delta s_{\text{GN}_1}(z) &= -\int dz' H_{\text{GN}}^{-1}(z,z')g_1(z') \\ &= -\left(\frac{4R}{c_0^2}\right)S(z-z_1),\end{aligned} \tag{45}$$

generated with linearized sensitivities, and

$$\begin{aligned}\delta s_{\text{GN}_2}(z) &= -\int dz' H_{\text{GN}}^{-1}(z,z')g_2(z') \\ &= -\left(\frac{4R-8R^2}{c_0^2}\right)S(z-z_1),\end{aligned} \tag{46}$$

with second order sensitivities.

Convergence of Gauss-Newton updates formed with 1st *vs* 2nd order sensitivities

There are two ways to discuss and compare equations (45) and (46) to gain insight into their convergence properties when iterated. First, we can ask *how close is each first iteration to the true model*? And second, we can actually iterate the formula. The former is a more precisely defined, analytic question (though it requires us to assume the convergence rate is determined by the accuracy of the first iteration), and the second is numerical in nature.

The reflection coefficient, which provides our update with all of its useful amplitude information in this example, is

$$R = \frac{c_1-c_0}{c_1+c_0} = \frac{1-c_0/c_1}{1+c_0/c_1}. \tag{47}$$

The "perfect" update δs, taking us from a homogeneous background to the precise correct interface size c_1, is

$$\delta s = s_1 - s_0 = -s_0\left(1-\frac{s_1}{s_0}\right) = -s_0\left(1-\frac{c_0^2}{c_1^2}\right). \tag{48}$$

The fraction in the reflection coefficient can be written

$$\frac{c_0}{c_1} = \left(1+\frac{\delta s}{s_0}\right)^{1/2} = 1 + \frac{1}{2}\left(\frac{\delta s}{s_0}\right) - \frac{1}{8}\left(\frac{\delta s}{s_0}\right)^2 + ..., \tag{49}$$

which means R can be expanded in orders of the perfect FWI update:

$$R = \frac{1}{4}\left(\frac{\delta s}{s_0}\right) + \frac{1}{8}\left(\frac{\delta s}{s_0}\right)^2 + ... \quad . \tag{50}$$

Supposing the second order term in equation (50) to be large enough to signify, let us now substitute this form for R into the result in equation (46):

$$\begin{aligned}\delta s_{\mathrm{GN}_2}(z) &= \frac{1}{c_0^2}\left[\left(\frac{\delta s}{s_0}\right) + \frac{1}{2}\left(\frac{\delta s}{s_0}\right)^2 - \frac{1}{2}\left(\frac{\delta s}{s_0}\right)^2\right] S(z-z_1) \\ &= \frac{1}{c_0^2}\left(\frac{\delta s}{s_0}\right) S(z-z_1) \\ &= \delta s\; S(z-z_1).\end{aligned} \tag{51}$$

This means the candidate update $\delta s_{\mathrm{GN}_2}(z)$ is exact (i.e., indistinguishable from $\delta s(z) = \delta s S(z-z_1)$) to second order. The difference between first and second order sensitivities is the third term in the square brackets in equation (51). Notice that if it had been absent, the nonlinear character of R (the second term in square brackets) would have produced a second order discrepancy between the GN update and the exact update; the second order sensitivity term suppresses this.

Figures 2–4 illustrate the numerical convergence boost we achieve by having each iteration correct to second order rather than first order. The effect is naturally more noticeable as contrasts increase. The convergences are approximate, as the Green's functions are approximated to zeroth order at every iteration.

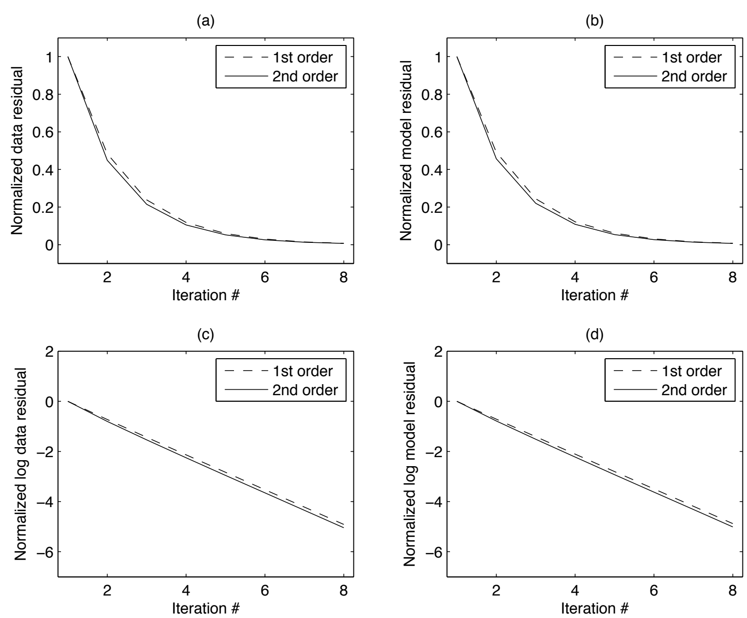

FIG. 2. Convergence over 8 iterations of the reconstruction of the single scalar interface. Normalized data; model; log data; and log model errors are considered. In this case the actual medium properties were $c_0 = 1500$m/s; $c_1 = 1600$m/s.

CONCLUSIONS

FWI is being pursued by CREWES in application to broadband land multicomponent. The standard descent-based methods like Newton, quasi-Newton, Gauss-Newton, and gradient-based are likely means by which iterative updating will take place, but given

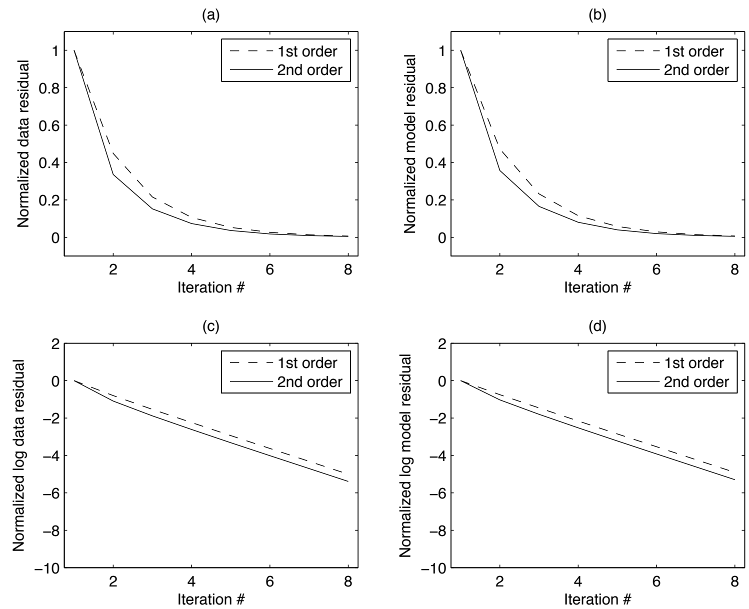

FIG. 3. Convergence over 8 iterations of the reconstruction of the single scalar interface. Normalized data; model; log data; and log model errors are considered. In this case the actual medium properties were $c_0 = 1500$m/s; $c_1 = 1600$m/s.

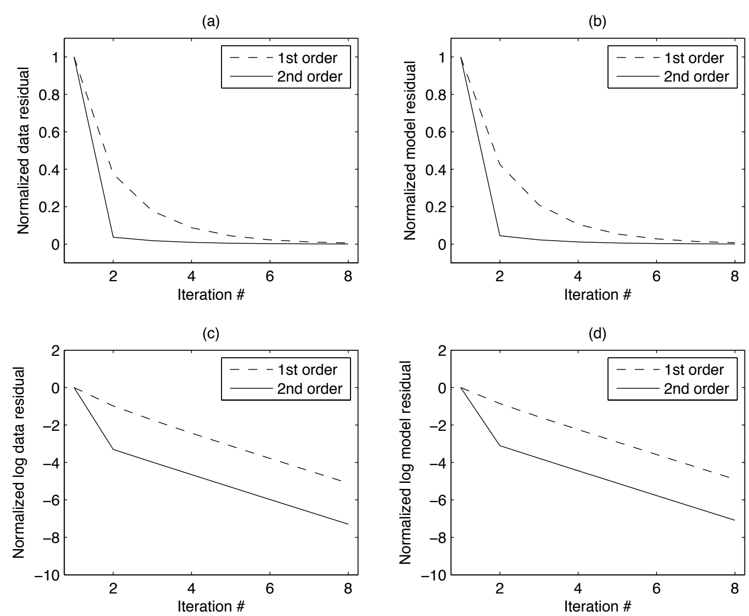

FIG. 4. Convergence over 8 iterations of the reconstruction of the single scalar interface. Normalized data; model; log data; and log model errors are considered. In this case the actual medium properties were $c_0 = 1500$m/s; $c_1 = 1600$m/s.

the occasional importance of nonlinearity in reflection amplitudes, we may be able to do better by conceiving of sensitivities which go beyond first order.

We have calculated the sensitivities accurate to the same order — 2nd — to which we have approximated the objective function – namely, second order. Consequent rough analysis of convergence indicates an exceedingly rapid convergence of parts of the model determined by reflection amplitudes (phase has not as yet been tested). Furthermore, we see analytically that the first update in the construction of a single boundary gives the correct answer to second order, in contrast to the standard sensitivities-based gradient, which produces updates in a similar circumstance accurate to leading order.

This research is an offshoot of our general FWI / IMMI research. In particular it is a forward-looking addendum to the multicomponent elastic updates we contemplate for IMMI on broadband land data (Margrave et al., 2013); assuming the (at least) partial use of reflection amplitudes, their intrinsically nonlinear character (Stovas and Ursin, 2001) can be leveraged according to the approach we have discussed here. Any means by which we can increase convergence rates must be vigorously sought.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for continued support. This work was funded by CREWES and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 379744-08.

REFERENCES

Bracewell, R. M., 1978, The Fourier Transform and Its Applications: McGraw-Hill, Inc.

Innanen, K. A., 2013, Coupling in amplitude variation with offset and the Wiggins approximation: Geophysics, **78**, No. 4, N21–N33.

Innanen, K. A., 2014a, Parameterization of multicomponent elastic reflection full waveform inversion: CREWES Annual Reports, **26**.

Innanen, K. A., 2014b, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, 717–734.

Kwon, K., and Yazici, B., 2010, Born expansion and Fréchet derivatives in nonlinear diffuse optical tomography: Computers and Mathematics with Applications, **59**, No. 11, 3377–3397.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Liu, Q., and Trompe, J., 2006, Finite-frequency kernels based on adjoint methods: Bulletin of the Seimological Society of America, **96**, No. 6, 2383–2397.

Margrave, G. F., Innanen, K. A., Lawton, D. C., Bancroft, J. B., Lamoureux, M., and Lines, L. R., 2013, CREWES 5-year research plan: Towards broadband multicomponent seismology and practical iterated inversion: CREWES Annual Report, **25**.

Margrave, G. F., Yedlin, M., and Innanen, K. A., 2011, Full waveform inversion and the inverse Hessian: CREWES Annual Research Report, **23**.

McGillivray, P. R., and Oldenburg, D. W., 1990, Methods for calculating fréchet derivatives and sensitivities for the non-linear inverse problem: a comparative study: Geophysical Prospecting, **38**, No. 5, 499–524.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice: The Leading Edge, **Sep**, 1040–1054.

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**, No. 2, 495–503.

Plessix, R.-E., Milcik, P., Rynia, H., Stopin, A., Matson, K., and Abri, S., 2013, Multiparameter full-waveform inversion: marine and land examples: The Leading Edge, **Sep**, 1030–1038.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model: Geophysics, **64**, No. 3, 888–901.

Shin, C., Yoon, K., Marfurt, K. J., Park, K., Yang, D., Lim, H. Y., Chung, S., and Shin, S., 2001, Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion: Geophysics, **66**, No. 6, 1856–1863.

Stovas, A., and Ursin, B., 2001, Second-order approximations of the reflection and transmission coefficients between two viscoelastic isotropic media: Journal of Seismic Exploration, **9**, No. 3, 223–233.

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

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1.

Weglein, A. B., Araújo, F. V., Carvalho, P. M., Stolt, R. H., Matson, K. H., Coates, R. T., Corrigan, D., Foster, D. J., Shaw, S. A., and Zhang, H., 2003, Inverse scattering series and seismic exploration: Inverse Problems, **19**, R27–R83.

Wu, R. S., Hu, C., and Wang, B., 2014, Nonlinear sensitivity operator and inverse thin-slab propagator for tomographic waveform inversion: SEG Technical Program Expanded Abstracts.

Wu, R. S., and Zheng, Y., 2014, Non-linear partial derivative and its De Wolf approximation for non-linear seismic inversion: Geophysical Journal International, **196**, No. 3, 1827–1843.

Yedlin, M. J., and Van Vorst, D. G., 2010, Tutorial on the continuous and discrete adjoint state method and basic implementation: CREWES Annual Reports, **22**.

Zhang, H., and Weglein, A. B., 2009a, Direct nonlinear inversion of multiparameter 1D acoustic media using inverse scattering subseries: Geophysics, **74**, No. 6, WCD29–WCD39.

Zhang, H., and Weglein, A. B., 2009b, Direct nonlinear inversion of multiparameter 1D elastic media using the inverse scattering series: Geophysics, **74**, No. 6, WCD15–WCD27.