A second-order preconditioner for Newton updates in seismic full waveform inversion

Kris Innanen

ABSTRACT

Newton updates (i.e., the updates invoked in the most general forms of seismic full waveform inversion) are intrinsically nonlinear, in the sense that they engage the data twice, once in the gradient and once in the inverse Hessian. However, a simple univariate example demonstrates that the nonlinear nature of the Newton step is not always used advantageously in inversion. This can be fixed to low order by introducing a formal parameter λ to the nonlinear part of the inverse Hessian, and determining a value for it which correctly implements second-order nonlinearity. The approach extends without requiring any additional conceptual leaps to a multidimensional scalar full waveform inversion problem, provided that development makes use of a nonlinear sensitivity expression such as that developed in a companion report in this volume.

INTRODUCTION

Seismic inversion via a local descent-based optimization algorithm (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) is sometimes referred to as being "iterative linear" (Weglein et al., 2003), or "quasi-linear" (Wu and Zheng, 2014; Wu et al., 2014), that is, an approach which solves a fundamentally nonlinear problem by solving a linearized version of the problem, updating, then repeating. This name originates from analysis of gradient-based full waveform inversion algorithms, wherein the data are engaged once in a gradient derivable through Born-approximate methods. Characterizing the full iterated inverse problem, even when it is gradient-based, as being linear at every step, should be done with some caution, however, since the forward modelled wave field begins to propagate with significant nonlinearity (including, for instance, multiples) as the model converges.

In the case of full Newton updates, the iterative linear label is not applicable. The inverse Hessian operator, which alters the direction and length of the gradient vector in a Newton update, depends on the data through the residuals. Although from a computational point of view incorporation of the full nonlinear Hessian is not currently possible in seismic FWI, updates which incorporate approximate versions are relatively common (e.g., Shin et al., 2001; Pan et al., 2013). An update formed by a composition of (1) the gradient, which depends on the data, and (2) the inverse Hessian, which also depends on the data, is nonlinear. A Newton update, therefore, although it involves linearized aspects (i.e., through its reliance on the first order Fréchet kernel, as discussed by, e.g., Tarantola, 1984; Dahlen et al., 2000), is not linear in the data, even during a given iteration.

However, and this is a big however, whether a Newton update is *meaningfully or usefully nonlinear* is another question entirely. In this paper we will pursue the consequences of an observation, developed in a simple univariate environment, that Newton updates solving a simple minimization problem with a nonlinear forward model, do not, in general, make optimal use of the nonlinear part of the inverse Hessian to speed up convergence.

The observation is not solely negative, as it also is suggestive of a simple preconditioner, in the form of a factor λ in front of the residual dependent part of the inverse Hessian, which enforces the update to solve in one step for the minimum of a problem in which the forward operator is exactly second order in character*. We can arrive at the requisite value for the preconditioning factor ($\lambda = -1/2$) in several ways; one of which carries over to the infinite dimensional forms needed to analyze seismic full waveform inversion updates.

The approach is then applied to a scalar 3D seismic full waveform inversion formulation. The only significant alteration of standard FWI derivation necessary to realize this is that the sensitivity or Jacobian matrix has to be extended to second order also, there being no point in retaining some nonlinear terms if others are neglected.

UNIVARIATE PRIMER

The issues at hand, pertaining to the accuracy and convergence of Newton steps when the forward model is nonlinear and/or when sensitivities are approximated, can be demonstrated by considering a univariate minimization problem.

Newton updates with a nonlinear forward modelling operator

A univariate Newton update

We will adopt a data (d) and model (x) relationship of the form

$$d = F(x). \tag{1}$$

We then seek the model value x^* which minimizes the objective function

$$\phi(x) = \frac{1}{2}\left[F(x) - d\right]^2. \tag{2}$$

Given a starting point x_0, one Newton step towards x^* from x_0 is calculated from the ratio of derivatives of ϕ:

$$\Delta x_{\mathrm{N}} = -\frac{\phi'(x_0)}{\phi''(x_0)}. \tag{3}$$

Substituting the particular form for ϕ in equation (2) we obtain

$$\Delta x_{\mathrm{N}} = -\frac{\left[F(x_0) - d\right]F'(x_0)}{\left[F'(x_0)\right]^2 + \left[F(x_0) - d\right]F''(x_0)}, \tag{4}$$

which can be organized as follows:

$$\Delta x_{\mathrm{N}} = -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0) + H_{\mathrm{NL}}(x_0)}, \tag{5}$$

*This is as compared to the ability of a standard Newton update to solve exactly in one step for the minimum of a problem in which the forward operator is exactly linear.

where

$$\begin{aligned}
&r(x_0) = F(x_0) - d \ \text{ are the residuals,}\\
&J(x_0) = F'(x_0) \ \text{ are the univariate sensitivities, and}\\
&H_{\mathrm{GN}}(x_0) = \left[F'(x_0)\right]^2 \quad \text{and}\\
&H_{\mathrm{NL}}(x_0) = r(x_0)F''(x_0)
\end{aligned} \tag{6}$$

are the residual independent and residual dependent parts of the Hessian matrix respectively. A Gauss-Newton update neglects $H_{\mathrm{NL}}(x_0)$ on the assumption of small residuals:

$$\Delta x_{\mathrm{GN}} = -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)}. \tag{7}$$

One way of categorizing Newton and Gauss-Newton updates is by introducing the parameter λ:

$$\Delta x_{\lambda} = -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0) + \lambda H_{\mathrm{NL}}(x_0)}, \tag{8}$$

in which case:

$$\Delta x_{\lambda}\Big|_{\lambda=1} = \Delta x_{\mathrm{N}}, \quad \text{and} \quad \Delta x_{\lambda}\Big|_{\lambda=0} = \Delta x_{\mathrm{GN}}. \tag{9}$$

Newton updates given a nonlinear forward modelling operator

When the forward modelling operator $F(x)$ is nonlinear, we might intuitively expect the Newton update to have better convergence properties than the Gauss-Newton type, since $H_{\mathrm{NL}}(x_0)$ engages the residuals nonlinearly. This is not the case, however — we can easily find examples where convergency is not improved at all. To do so, we will restrict ourselves to a certain class of nonlinear forward modelling operators F. We will ask that F be *second order*, in the following sense. A perfect datum satisfies

$$d = F(x^*), \tag{10}$$

where x^* is the true model. Let Δx be defined as the difference between any initial x_0 and the exact solution x^*:

$$x^* = x_0 + \Delta x. \tag{11}$$

The forward model will be taken to be *second order* if for any x^*, x_0 pair we can with negligible error write

$$d = F(x^*) = F(x_0 + \Delta x) = F(x_0) + F'(x_0)\Delta x + \frac{1}{2}F''(x_0)\Delta x^2. \tag{12}$$

That is, the forward problem is nonlinear but not so nonlinear that terms in Δx^3 and higher are needed. Next, we re-write the Newton update and expand in binomial series in the residuals:

$$
\begin{aligned}
\Delta x_{\mathrm{N}} &= -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)+H_{\mathrm{NL}}(x_0)} \\
&= -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)}\left[1+\frac{H_{\mathrm{NL}}(x_0)}{H_{\mathrm{GN}}(x_0)}\right]^{-1} \\
&= -\frac{F'(x_0)}{[F'(x_0)]^2}r(x_0)+\frac{F'(x_0)F''(x_0)}{[F'(x_0)]^4}r^2(x_0)+\dots\ .
\end{aligned}
\tag{13}
$$

The residuals can be expressed using equation (12) as

$$
r(x_0)=F(x_0)-d=-F'(x_0)\Delta x-\frac{1}{2}F''(x_0)\Delta x^2. \tag{14}
$$

Substituting equation (14) into equation (13), we arrive at an expression which relates the Newton update Δx_{N} to the ideal update Δx (the one that takes us directly to the correct answer):

$$
\begin{aligned}
\Delta x_{\mathrm{N}} &= \frac{F'(x_0)}{[F'(x_0)]^2}\left[F'(x_0)\Delta x+\frac{1}{2}F''(x_0)\Delta x^2\right]+\frac{F'(x_0)F''(x_0)}{[F'(x_0)]^2}\Delta x^2+\dots \\
&= \Delta x+\frac{1}{2}\frac{F'(x_0)F''(x_0)}{[F'(x_0)]^2}\Delta x^2+\frac{F'(x_0)F''(x_0)}{[F'(x_0)]^2}\Delta x^2+\dots \\
&= \Delta x+\frac{3}{2}\frac{F''(x_0)}{F'(x_0)}\Delta x^2+\dots\ .
\end{aligned}
\tag{15}
$$

This relation is consistent with general wisdom about the Newton update: namely, to first order, it is equivalent to the ideal update, $\Delta x_{\mathrm{N}}\approx\Delta x$. However, in spite of the fact that it engages the data at second order via $r^2(x_0)$ etc., the Newton update is *not* equivalent to the ideal update to second order – there is a discrepancy of $(3/2)(F''/F')\Delta x^2$.

A second order preconditioner

We can force the update to do what we might have believed the Newton update should do – namely, to correctly engage the data nonlinearly in order to speed up convergence. We do so as follows. Take the particular re-writing of the update in equation (13), but re-introduce the parameter λ in front of H_{NL}, now considering it to be a free and choosable preconditioning quantity:

$$
\begin{aligned}
\Delta x_{\lambda} &= -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)}\left[1+\lambda\frac{H_{\mathrm{NL}}(x_0)}{H_{\mathrm{GN}}(x_0)}\right]^{-1} \\
&= -\frac{F'(x_0)}{[F'(x_0)]^2}r(x_0)+\lambda\frac{F'(x_0)F''(x_0)}{[F'(x_0)]^4}r^2(x_0)+\dots\ .
\end{aligned}
\tag{16}
$$

Once again substituting the second order forward model in equation (14), we this time arrive at the more general λ update

$$
\Delta x_{\lambda}=\Delta x+\left(\frac{1}{2}+\lambda\right)\frac{F''(x_0)}{F'(x_0)}\Delta x^2+\dots\ . \tag{17}
$$

This in turn is suggestive that selecting the special number $\lambda = -1/2$ produces an update which matches the exact update up to second order. In other words, for univariate optimization the update

$$\Delta x_{-1/2} = -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)}\left[1 + \frac{1}{2}\frac{H_{\mathrm{NL}}(x_0)}{H_{\mathrm{GN}}(x_0)}\right] \tag{18}$$

appears to be optimal. More generally, this mode of analysis (comparing Δx_λ with the exact Δx order by order in the presence of a preconditioning parameter) appears to be useful for tuning candidate updates.

Deriving the preconditioner by explicit comparison of orders

The main goal in this paper is to derive a more complex, seismic/wave inversion realization of the second order preconditioner λ. The derivation in the previous section is too simple to lend itself well to that effort. Here we present a slightly less intuitive derivation whose main positive feature is that a version of it can be employed in the more complex FWI situation. We re-write the basic update in equation (13) as

$$-\left[H_{\mathrm{GN}}(x_0) + \lambda H_{\mathrm{NL}}(x_0)\right]\Delta x_\lambda = J(x_0)r(x_0), \tag{19}$$

wherein both $H_{\mathrm{NL}}(x_0)$ and the right-hand side are functions of the residuals r. Substituting the expansion in equation (14) in the left-hand side above, we obtain

$$\begin{aligned}
\mathrm{LHS} &= -\left[H_{\mathrm{GN}}(x_0) + \lambda H_{\mathrm{NL}}(x_0)\right]\Delta x_\lambda \\
&= -\left[F'(x_0)\right]^2\Delta x_\lambda - \lambda F''(x_0)\left[-F'(x_0)\Delta x - \frac{1}{2}F''(x_0)\Delta x^2 - ...\right]\Delta x_\lambda \\
&= \lambda F''(x_0)F'(x_0)\Delta x\Delta x_\lambda - \left[F'(x_0)\right]^2\Delta x_\lambda,
\end{aligned} \tag{20}$$

and doing likewise to the right-hand side we obtain

$$\begin{aligned}
\mathrm{RHS} &= F'(x_0)\left[-F'(x_0)\Delta x - \frac{1}{2}F''(x_0)\Delta x^2 - ...\right] \\
&= -\frac{1}{2}F''(x_0)F'(x_0)\Delta x^2 - \left[F'(x_0)\right]^2\Delta x.
\end{aligned} \tag{21}$$

Defining Δx^2 and $\Delta x\Delta x_\lambda$ to be of identical (second) order, we recover by comparing coefficients in equations (20) and (21) the result that if Δx_λ is to be equivalent to Δx to second order, it must be that $\lambda = -1/2$.

Convergence

We now have three variations on the Newton step to consider and compare. In this section we will do so with a toy forward model designed to exhibit a flexible degree of nonlinearity. The three update types are

$$\begin{aligned}\Delta x_{\mathrm{N}} &= -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)+H_{\mathrm{NL}}(x_0)},\\ \Delta x_{\mathrm{GN}} &= -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)},\\ \Delta x_{-1/2} &= -\frac{J(x_0)r(x_0)}{H_{\mathrm{GN}}(x_0)}\left[1+\frac{1}{2}\frac{H_{\mathrm{NL}}(x_0)}{H_{\mathrm{GN}}(x_0)}\right],\end{aligned} \tag{22}$$

where to reiterate

$$\begin{aligned} r(x_0) &= F(x_0) - d,\\ J(x_0) &= F'(x_0),\\ H_{\mathrm{GN}}(x_0) &= \left[F'(x_0)\right]^2,\\ H_{\mathrm{NL}}(x_0) &= r(x_0)F''(x_0),\end{aligned} \tag{23}$$

and $d = F(x^*)$ is the single exact datum. We will experiment with the following toy forward model:

$$F(x) = c_0 + c_1 x + c_2 x^2 \tag{24}$$

which is exactly "second order", and whose degree of nonlinearity can be varied by increasing or decreasing c_2 relative to c_0 and c_1. We will examine three cases, "low" nonlinearity ($c_0 = 1$, $c_1 = -2$, $c_2 = 0.3$), "moderate" nonlinearity ($c_0 = 1$, $c_1 = -2$, $c_2 = 0.6$), and "high" nonlinearity ($c_0 = 1$, $c_1 = -2$, $c_2 = 1.0$). None of the cases exhibit a degree of nonlinearity which introduces local minima; our interest is to pursue the results of non-quadratic curvature in ϕ. See Figure 1.

We select the exact model $x^* = 0.29$, the starting point $x_0 = 0.45$, then calculate the datum $d = F(0.29)$, and iterate:

$$\begin{aligned} x_{\mathrm{N}}^{(i)} &= x_{\mathrm{N}}^{(i-1)} + \Delta x_{\mathrm{N}}^{(i-1)}\\ x_{\mathrm{GN}}^{(i)} &= x_{\mathrm{GN}}^{(i-1)} + \Delta x_{\mathrm{GN}}^{(i-1)}\\ x_{\text{-1/2}}^{(i)} &= x_{\text{-1/2}}^{(i-1)} + \Delta x_{\text{-1/2}}^{(i-1)}.\end{aligned} \tag{25}$$

In Figure 2 we plot various error quantities over the course of 6 iterations for the "low nonlinearity" case. The left column plots the absolute data residuals normalized to the starting residual value, with the top (Figure 2a) being the straight residuals and the bottom (Figure 2b) being the log residuals. The right column is identical but the model residuals are plotted instead. The red curve is the Newton update, the blue curve is the Gauss-Newton update, and the black curve is the second-order preconditioned update. The three are comparable, especially at and beyond the third iteration, but with the second order slightly outcompeting the Gauss-Newton update, and the Gauss-Newton slightly outcompeting the Newton update.

In Figures 3–4 we repeat the iterations for the "moderate" and "high" nonlinearity cases. Here the discrepancies between the step types become more vivid, with the second order update outpacing the Gauss-Newton update even more dramatically, and likewise the Gauss-Newton outpacing the Newton.

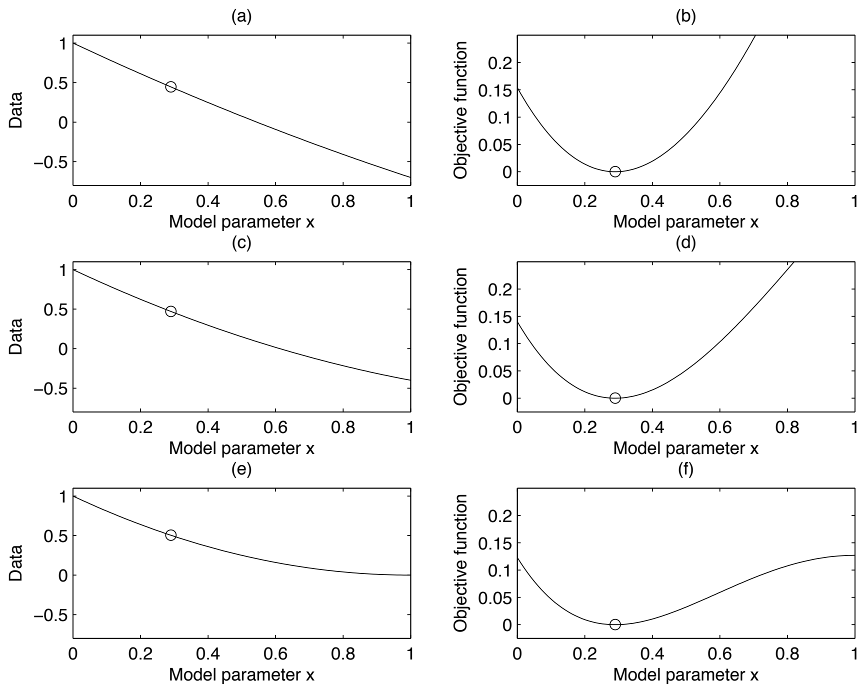

FIG. 1. Three univariate updating problems with varying degrees of forward model nonlinearity. Left column: the forward model $F(x)$ plotted vs x, using equation (24), with $c_0 = 1$, $c_1 = -2$, and c_2 varying from (a) 0.3, (c) 0.6, and (e) 1.0. The exact model $x^* = 0.29$ is plotted as a circle. Right column: (b)-(f) the objective functions $\phi(x)$ for each of the forward model examples.

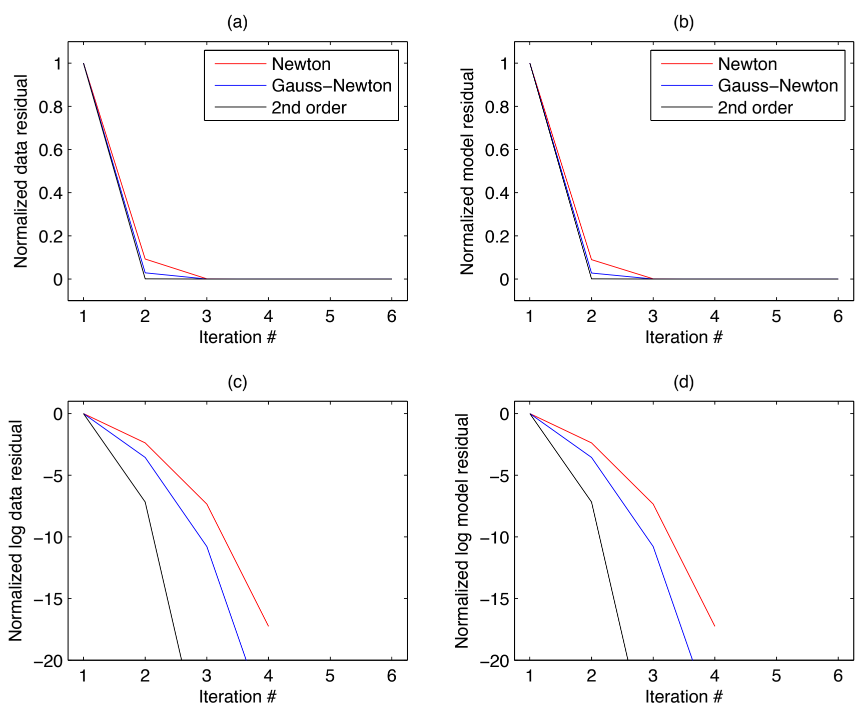

FIG. 2. "Low nonlinearity" case. (a) Normalized absolute data residuals; (b) normalized absolute model residuals; (c) normalized log absolute data residuals; (d) normalized log absolute model residuals.

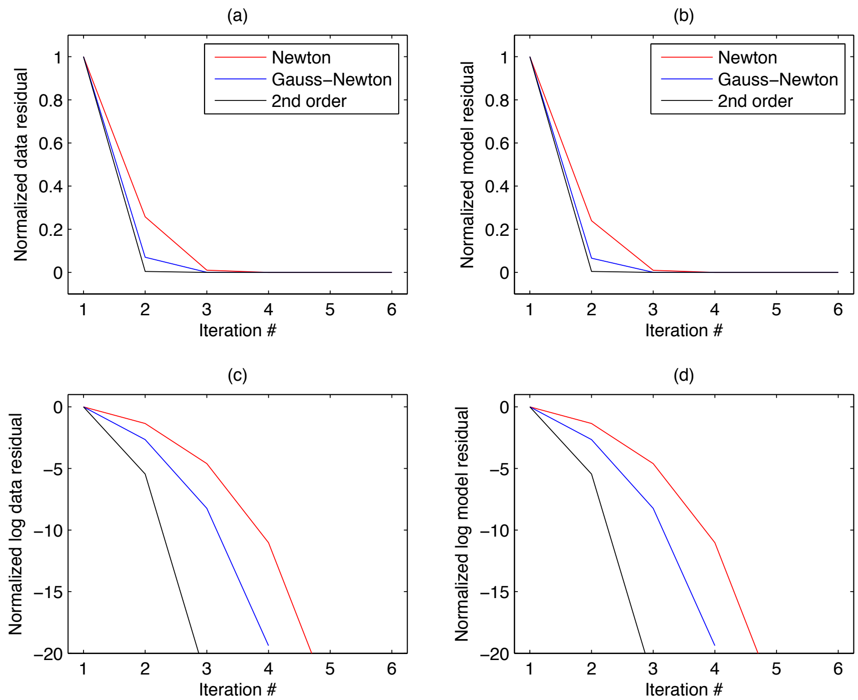

FIG. 3. "Moderate nonlinearity" case. (a) Normalized absolute data residuals; (b) normalized absolute model residuals; (c) normalized log absolute data residuals; (d) normalized log absolute model residuals.

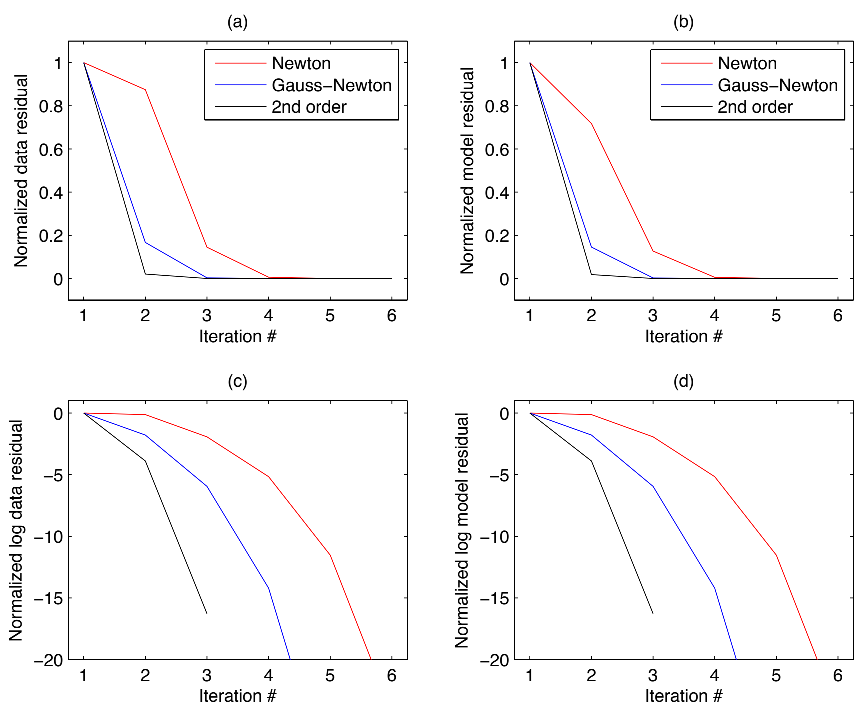

FIG. 4. "High nonlinearity" case. (a) Normalized absolute data residuals; (b) normalized absolute model residuals; (c) normalized log absolute data residuals; (d) normalized log absolute model residuals.

The influence of linearized sensitivities

One aspect of FWI that will impact the decision whether or not to try to engage the nonlinear part of the inverse Hessian is its standard use of linearized sensitivities[†]. In a companion paper (Innanen, 2014a) we extend some way beyond linearity in the seismic FWI sensitivities. The ability to do this will be important when considering our current suggestion, that nonlinear parts of the Hessian are useful — that is likely true only if we simultaneously increase the order of the Fréchet derivative. It is chancy at best to incorporate nonlinearity in one part of a mathematical quantity if it is being ignored in another.

In all of the univariate development so far, we have been assuming access to an exact sensitivity $F'(x_0)$. Supposing our gradient were constructed not with the exact sensitivity $F'(x_0)$ but with an approximated version of it $F'_{\rm lin}(x_0)$, we would expect the Gauss-Newton inverse Hessian approximation to be affected, as well as any instance of the sensitivity being used. *However*, generally an exact (or, at least, nonlinear) forward modelling is assumed in the construction of the residuals. This leads to a form for the general update of:

$$\begin{aligned}\Delta x_\lambda &= \frac{F'_{\rm lin}(x_0)}{[F'_{\rm lin}(x_0)]^2}\left[F'(x_0)\Delta x + \frac{1}{2}F''(x_0)\Delta x^2\right] + \lambda\frac{F'_{\rm lin}(x_0)F''(x_0)}{[F'_{\rm lin}(x_0)]^2}\Delta x^2 + ...\\ &= \left(\frac{F'(x_0)}{F'_{\rm lin}(x_0)}\right)\Delta x + \left(\frac{1}{2}+\lambda\right)\frac{F''(x_0)}{F'(x_0)}\Delta x^2 + ...\end{aligned} \tag{26}$$

In contrast to our earlier results, here there is *no* value of λ which brings Δx into agreement with Δx_λ. Proper incorporation of 2nd order behaviour in a modified Newton step requires the sensitivities/Jacobian matrix to be accurate to at least that order also.

INFERRING A SECOND-ORDER PRECONDITIONER FOR SEISMIC FWI

The main result of this paper involves adapting the simple λ update approach presented in univariate form in the previous section to the seismic FWI case. We will assume a 3D scalar (P-wave velocity) problem. See Innanen (2014b) for a review of the terms and basic equations of functional versions of multidimensional FWI.

The λ update and associated gradient

Given a 3D scalar gradient $g(\mathbf{r})$, the associated Hessian can be calculated as

$$H(\mathbf{r},\mathbf{r}') = \frac{\partial}{\partial s(\mathbf{r})}g(\mathbf{r}'). \tag{27}$$

The gradient $g(\mathbf{r})$ is a sum over experimental variables (in this case source and receiver locations and temporal frequency) of the product of the sensitivities and the residuals. Thus,

[†] See the course notes of Schuster (*http://utam.gg.utah.edu/stanford/*), under the subtitle *waveform inversion algorithm*, for a clear exposition of the linearization as it appears through a scattering derivation; to see where linearization occurs in adjoint methods, see for instance McGillivray and Oldenburg (1990).

its derivative with respect to the model gives a Hessian function of the form

$$H(\mathbf{r},\mathbf{r}') = \sum_{g,s} \int d\omega \left[\frac{\partial G^*(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})} \frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r}')} - \frac{\partial}{\partial s(\mathbf{r})} \left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r}')} \right) \delta P^* \right]. \tag{28}$$

We will substitute into this expression the sensitivities constructed to second order in the residuals as discussed by Innanen (2014a):

$$\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})} = -\omega^2 G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}_s) - \omega^2 G^\dagger(\mathbf{r}_g,\mathbf{r},\mathbf{r}_s)G(\mathbf{r},\mathbf{r})\delta P^*, \tag{29}$$

where we have defined

$$G^\dagger(\mathbf{r}_g,\mathbf{r},\mathbf{r}_s) = \frac{G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}_s)}{G^*(\mathbf{r}_g,\mathbf{r})G^*(\mathbf{r},\mathbf{r}_s)}. \tag{30}$$

Making use of the result

$$\frac{\partial}{\partial s(\mathbf{r})} \delta P^*(\mathbf{r}_g,\mathbf{r}_s) = -\frac{\partial G^*(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})}, \tag{31}$$

we obtain, to lowest order in δP, for the two terms in equation (28),

$$\begin{aligned} \frac{\partial}{\partial s(\mathbf{r})} \left(\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r}')} \right) = \omega^4 [& G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}')G(\mathbf{r}',\mathbf{r}_s) + G(\mathbf{r}_g,\mathbf{r}')G(\mathbf{r}',\mathbf{r})G(\mathbf{r},\mathbf{r}_s) \\ & - G^\dagger(\mathbf{r}_g,\mathbf{r},\mathbf{r}_s)G^*(\mathbf{r}_g,\mathbf{r})G^*(\mathbf{r},\mathbf{r}_s)G(\mathbf{r},\mathbf{r})], \end{aligned} \tag{32}$$

and

$$\begin{aligned} \frac{\partial G^*(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r})} \frac{\partial G(\mathbf{r}_g,\mathbf{r}_s)}{\partial s(\mathbf{r}')} = \omega^4 [& G(\mathbf{r}_g,\mathbf{r}')G(\mathbf{r}',\mathbf{r}_s)G^*(\mathbf{r}_g,\mathbf{r})G^*(\mathbf{r},\mathbf{r}_s) \\ & + G(\mathbf{r}_g,\mathbf{r}')G(\mathbf{r}',\mathbf{r}_s)G^{\dagger *}(\mathbf{r}_g,\mathbf{r},\mathbf{r}_s)G^*(\mathbf{r},\mathbf{r})\delta P \\ & + G^*(\mathbf{r}_g,\mathbf{r})G^*(\mathbf{r},\mathbf{r}_s)G^\dagger(\mathbf{r}_g,\mathbf{r}',\mathbf{r}_s)G(\mathbf{r}',\mathbf{r}')\delta P^*]. \end{aligned} \tag{33}$$

After substitution of these results into equation (28), the Hessian takes the form

$$H(\mathbf{r},\mathbf{r}') = \sum_{g,s} \int d\omega \omega^4 H_{\text{GN}}(\mathbf{r},\mathbf{r}') \left[1 + \frac{H_{\text{NL}}^{\text{c}}(\mathbf{r},\mathbf{r}')}{H_{\text{GN}}(\mathbf{r},\mathbf{r}')} \delta P^* + \frac{H_{\text{NL}}^{\text{nc}}(\mathbf{r},\mathbf{r}')}{H_{\text{GN}}(\mathbf{r},\mathbf{r}')} \delta P \right], \tag{34}$$

where

$$\begin{aligned} H_{\text{NL}}^{\text{c}}(\mathbf{r},\mathbf{r}') = & G(\mathbf{r}_g,\mathbf{r})G(\mathbf{r},\mathbf{r}')G(\mathbf{r}',\mathbf{r}_s) + G(\mathbf{r}_g,\mathbf{r}')G(\mathbf{r}',\mathbf{r})G(\mathbf{r},\mathbf{r}_s) \\ & - G^*(\mathbf{r}_g,\mathbf{r})G^*(\mathbf{r},\mathbf{r}_s)G^\dagger(\mathbf{r}_g,\mathbf{r},\mathbf{r}_s)G(\mathbf{r},\mathbf{r}) \\ & - G^*(\mathbf{r}_g,\mathbf{r}')G^*(\mathbf{r}',\mathbf{r}_s)G^\dagger(\mathbf{r}_g,\mathbf{r}',\mathbf{r}_s)G(\mathbf{r}',\mathbf{r}'), \end{aligned} \tag{35}$$

and

$$H_{\text{NL}}^{\text{nc}}(\mathbf{r},\mathbf{r}') = G(\mathbf{r}_g,\mathbf{r}')G(\mathbf{r}',\mathbf{r}_s)G^{\dagger *}(\mathbf{r}_g,\mathbf{r},\mathbf{r}_s)G(\mathbf{r},\mathbf{r}). \tag{36}$$

This is now in a form that allows us to incorporate the preconditioner λ, in analogy to the λ update we discussed in the univariate section. The gradient, λ-update and λ-preconditioned Hessian are related by

$$g(\mathbf{r}) = \int d\mathbf{r}' H_\lambda(\mathbf{r}, \mathbf{r}') \delta s_\lambda(\mathbf{r}'), \tag{37}$$

where

$$H_\lambda(\mathbf{r}, \mathbf{r}') = \sum_{g,s} \int d\omega \omega^4 H_{\mathrm{GN}}(\mathbf{r}, \mathbf{r}') \left[1 + \lambda(\mathbf{r}, \mathbf{r}') \left(\frac{H_{\mathrm{NL}}^{\mathrm{c}}(\mathbf{r}, \mathbf{r}')}{H_{\mathrm{GN}}(\mathbf{r}, \mathbf{r}')} \delta P^* + \frac{H_{\mathrm{NL}}^{\mathrm{nc}}(\mathbf{r}, \mathbf{r}')}{H_{\mathrm{GN}}(\mathbf{r}, \mathbf{r}')} \delta P \right) \right].$$

We will now construct series expansions of both sides of equation (37), with the left-hand side entirely expressed in terms of the exact perturbation $\delta s(\mathbf{r})$ and the right-hand side expressed in combinations of $\delta s(\mathbf{r})$ and the update form $\delta s_\lambda(\mathbf{r})$. The two sides of the equation will be compared in a manner similar to the comparison between equations (20) and (21).

The gradient in terms of the exact step $\delta s(\mathbf{r})$

We now reconstruct the 3D wave version of the univariate equation (19). The gradient to second order is

$$g(\mathbf{r}) = -\sum_{g,s} \int d\omega \omega^2 \left[G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s) \delta P^* + G^\dagger(\mathbf{r}_g, \mathbf{r}, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r}) \delta P^* \delta P^* \right]. \tag{38}$$

If our scalar seismic forward problem is assumed to be intrinsically "second order", we may write it using a Born series expansion

$$\begin{aligned} \delta P^* \approx & -\omega^2 \int d\mathbf{r}' G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) \delta s(\mathbf{r}') \\ & + \omega^4 \int d\mathbf{r}' G^*(\mathbf{r}_g, \mathbf{r}') \delta s(\mathbf{r}') \int d\mathbf{r}'' G^*(\mathbf{r}', \mathbf{r}'') G^*(\mathbf{r}'', \mathbf{r}_s) \delta s(\mathbf{r}''), \end{aligned} \tag{39}$$

truncated beyond the second term. Making the same assumption of collocated scattering we made in constructing the sensitivities (Innanen, 2014a), this becomes

$$\begin{aligned} \delta P^* \approx & -\omega^2 \int d\mathbf{r}' \delta s(\mathbf{r}') G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) \\ & + \omega^4 \int d\mathbf{r}' \delta s^2(\mathbf{r}') G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) G^*(\mathbf{r}', \mathbf{r}'), \end{aligned} \tag{40}$$

or, when needed,

$$\delta P^* \times \delta P^* \approx \omega^4 \int d\mathbf{r}' \delta s^2(\mathbf{r}') \left[G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) \right]^2. \tag{41}$$

These we may now substitute into the gradient, forming a second-order accurate relationship between $g(\mathbf{r})$ and the exact update $\delta s(\mathbf{r})$:

$$
\begin{aligned}
g(\mathbf{r}) = \sum_{g,s} \int d\omega\omega^4 \int d\mathbf{r}' \delta s(\mathbf{r}') G(\mathbf{r}_g, \mathbf{r}) G(\mathbf{r}, \mathbf{r}_s) G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) \\
- \sum_{g,s} \int d\omega\omega^6 \int d\mathbf{r}' \delta s^2(\mathbf{r}') \left[G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) G^*(\mathbf{r}', \mathbf{r}') \right. \\
\left. + [G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s)]^2 G^\dagger(\mathbf{r}_g, \mathbf{r}, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r}) \right] .
\end{aligned}
\tag{42}
$$

Because the Gauss-Newton Hessian approximation appears in the first order term, for convenience we instead write the gradient expression as

$$
\begin{aligned}
g(\mathbf{r}) = \sum_{g,s} \int d\omega\omega^4 \int d\mathbf{r}' \delta s(\mathbf{r}') H_{\mathrm{GN}}(\mathbf{r}, \mathbf{r}') \\
- \sum_{g,s} \int d\omega\omega^6 \int d\mathbf{r}' \delta s^2(\mathbf{r}') \left[G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) G^*(\mathbf{r}', \mathbf{r}') \right. \\
\left. + [G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s)]^2 G^\dagger(\mathbf{r}_g, \mathbf{r}, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r}) \right] .
\end{aligned}
\tag{43}
$$

This is analogous to the right-hand side of equation (19).

The λ gradient in terms of the exact step $\delta s(\mathbf{r})$

In order to construct the 3D scalar wave version of the left-hand side of equation (19), we substitute appropriate collocated scattering integrals for the residuals appearing in equation (37), finding

$$
\begin{aligned}
g(\mathbf{r}) = \sum_{g,s} \int d\omega\omega^4 \int d\mathbf{r}' \delta s_\lambda(\mathbf{r}') H_{\mathrm{GN}}(\mathbf{r}, \mathbf{r}') \\
- \sum_{g,s} \int d\omega\omega^6 \int d\mathbf{r}' \delta s(\mathbf{r}') \delta s_\lambda(\mathbf{r}') \lambda(\mathbf{r}, \mathbf{r}') \left[H_{\mathrm{NL}}^{\mathrm{nc}}(\mathbf{r}, \mathbf{r}') G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) \right. \\
\left. + H_{\mathrm{NL}}^{\mathrm{nc}}(\mathbf{r}, \mathbf{r}') G(\mathbf{r}_g, \mathbf{r}') G(\mathbf{r}', \mathbf{r}_s) \right] .
\end{aligned}
\tag{44}
$$

Inferring λ through explicit comparison of orders

Equations (43) and (44) can now be compared explicitly at first and second orders. It follows that for the λ update $\delta s_\lambda(\mathbf{r})$ to be equivalent to second order to the exact update $\delta s(\mathbf{r})$ the preconditioner must be of the form

$$
\lambda(\mathbf{r}, \mathbf{r}') = \frac{G^*(\mathbf{r}', \mathbf{r}') + G^*(\mathbf{r}_g, \mathbf{r}') G^*(\mathbf{r}', \mathbf{r}_s) G^\dagger(\mathbf{r}_g, \mathbf{r}, \mathbf{r}_s) G(\mathbf{r}, \mathbf{r})}{H_{\mathrm{NL}}^{\mathrm{c}}(\mathbf{r}, \mathbf{r}') + G^\dagger(\mathbf{r}_g, \mathbf{r}', \mathbf{r}_s) H_{\mathrm{NL}}^{\mathrm{nc}}(\mathbf{r}, \mathbf{r}')} . \tag{45}
$$

CONCLUSIONS

Newton updates (i.e., the type invoked in the most general forms of seismic full waveform inversion) are intrinsically nonlinear, since they invoke the data twice, once in the gradient and once in the inverse Hessian. But, this nonlinearity does not accomplish what we might expect it to, i.e., to correctly engage the residuals such that a single update exactly solves for the minimum when the forward modelling operator is exactly second order. An order by order argument allows us to infer a preconditioning factor (or operator) which corrects this failing in a scalar full waveform inversion theory appropriate for multidimensional and (for the moment) scalar media.

In the companion paper to this report (Innanen, 2014a), we show that it is the nonlinear sensitivity or Fréchet kernel that is responsible for the corrections required because of the nonlinearity of the reflection coefficient - velocity contrast relationship. Using the extended sensitivity, the correct second order update is achieved using only the Gauss-Newton approximate form of the Hessian (which is independent of the residuals). So, what then is the role of the nonlinear, preconditioned inverse Hessian we have just derived in FWI? What job does it do? One possibility stems from the (longstanding Shin et al., 2001) observation that the inverse Hessian corrects for otherwise neglected propagation phenomena. It may be that the data which characterize the medium overlying a particular interface are engaged to correct for its mispositioning and amplitude. This would be similar to the direct mechanisms observed in the mathematics of nonlinear inverse scattering (discussed by, e.g., Weglein et al., 2003; Shaw et al., 2004; Innanen, 2008; Zhang and Weglein, 2009a,b). This is a matter of ongoing investigation.

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

Dahlen, F. A., Hung, S. H., and Nolet, G., 2000, Fréchet kernels for finite-frequency traveltimes – i. Theory: Geophysical Journal International, **141**, 157–174.

Innanen, K. A., 2008, A direct non-linear inversion of primary wave data reflecting from extended, heterogeneous media: Inverse Problems, , No. 24, 035,021.

Innanen, K. A., 2014a, Gauss-Newton updates with nonlinear sensitivities: 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.

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.

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.

Pan, W., Innanen, K. A., and Margrave, G. F., 2013, Full waveform inversion with phase-encoded pseudo-Hessian: CREWES Annual Reports, **25**.

Shaw, S. A., Weglein, A. B., Foster, D. J., Matson, K. H., and Keys, R. G., 2004, Isolation of a leading order depth imaging series and analysis of its convergence properties: Journal of Seismic Exploration, **2**, 157–195.

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.

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 Wold approximation for non-linear seismic inversion: Geophysical Journal International, **196**, No. 3, 1827–1843.

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.