Full waveform inversion and the inverse Hessian

Gary Margrave, Matt Yedlin and Kris Innanen

ABSTRACT

Full waveform inversion involves defining an objective function, and then moving in steps from some starting point to the minimum of that objective function. Gradient based steps have long been shown to involve seismic migrations, particularly, migrations which make us of a correlation-based imaging condition. More sophisticated steps, like Gauss-Newton and quasi-Newton, alter the step by involving the inverse Hessian or approximations thereof. Our interest is in the geophysical, and practical, influence of the Hessian. We derive a wave physics interpretation of the Hessian, use it to flesh out a published statement of Virieux, namely that performing a quasi-Newton step amounts to applying a gain correction for amplitude losses in wave propagation, and finally show that in doing so the quasi-Newton step is equivalent to migration with a deconvolution imaging condition rather than a correlation imaging condition.

INTRODUCTION

This paper is about full waveform inversion: how to do it, and also how to think about it. The details are technical, but the point is entirely practical: to attach geophysical significance to the mathematical operations of waveform inversion, and to recognize that it can often be implemented with items that are already in our seismic imaging toolboxes.

Full waveform inversion (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) is an application of the methods of multivariate optimization to the seismic inverse problem, in which the parameters of the Earth's subsurface are estimated from measurements of seismic wave fields. An optimization problem is solved when an objective function is minimized. This can happen in a number of ways. For instance, in order of "completeness", taking

1. Gauss-Newton (or just Newton) steps,
2. Quasi-Newton steps, or
3. Gradient-based steps

towards the minimum. The full Gauss-Newton step exactly minimizes a local quadratic approximation of the objective function, with a step direction and length that is a composition of the inverse Hessian and gradient. The quasi-Newton step does so with an approximate (usually less complex) inverse Hessian. Finally, the gradient-based step uses the gradient of the objective function followed by a line search to take its step towards the minimum.

In seismic inversion, these steps are, thankfully, not simply huge empty numerical exercises: real geophysics enters these minimizations. It turns out that the gradient is actually a *migration* of seismic data, calculated using a correlation-based imaging condition. This interpretation allows us to think about and analyze each step in waveform inversion in terms of physical concepts, and also turns one's seismic migration toolbox into a [gradient-based] full waveform inversion toolbox.

However, of late full waveform inversion researchers have begun to look seriously at taking quasi-Newton and even Gauss-Newton steps. Numerically this is no mean endeavour, and though mathematically taking a sophisticated Gauss-Newton step rather than a plain old gradient based step may be a move forward, we may worry about introducing big numerical machinery without concomitant geophysical meaning.

But geophysical interpretation, again in terms of physical concepts and known and available seismic migration tools, is available for quasi-Newton, and likely full Gauss-Newton steps also. The purpose of this paper is to develop these interpretations, which in our opinion are critical to successful seismic inversion. In particular:

1. Using 3D scalar full waveform inversion as a framework, we illustrate via Gâteaux derivatives the role of the gradient and inverse Hessian in taking a single Gauss-Newton step towards the inverse solution;
2. We re-derive using a nonlinear scattering formulation the interpretation of a gradient-based inversion step as being equivalent to *migration of data residuals using a correlation based imaging condition*;
3. We extend this wave-based interpretation to include the action of the Hessian and approximate Hessian;
4. With physical arguments and dimensional analysis we flesh out the statement of Virieux and Operto (2009), that premultiplying the gradient by the inverse approximate Hessian (i.e., performing a quasi-Newton step rather than a gradient-based step) amounts to applying a gain correction to correct for amplitude losses in modelled wave propagation;
5. We identify this quasi-Newton step as being equivalent to applying a *deconvolution* imaging condition, rather than a correlation imaging condition, in the migration interpretation.

1. ONE GAUSS-NEWTON STEP IN FULL WAVEFORM INVERSION

Equations of motion

We will consider two equations, one of which is satisfied by our modeled field G, and one which is satisfied by the "actual" field P, projections of which onto the measurement surface will be the measured data. They are assumed to satisfy

$$\begin{aligned} \left[\nabla^2 + \omega^2 s_0^{(n)}(\mathbf{r})\right] G(\mathbf{r}, \mathbf{r}_s, \omega) &= \delta(\mathbf{r} - \mathbf{r}_s) \\ \left[\nabla^2 + \omega^2 s(\mathbf{r})\right] P(\mathbf{r}, \mathbf{r}_s, \omega) &= \delta(\mathbf{r} - \mathbf{r}_s) \end{aligned} \tag{1}$$

where $s_0(\mathbf{r})$ is the known reference/background medium at the nth iteration in the inverse process, and $s(\mathbf{r})$ is the unknown actual medium. The latter is the "right answer", towards which the former is expected to converge. The quantities s are related to the scalar wave velocities of the two media by

$$s(\mathbf{r}) \equiv \frac{1}{c^2(\mathbf{r})}, \quad s_0^{(n)}(\mathbf{r}) \equiv \frac{1}{c_{0,n}^2(\mathbf{r})}. \tag{2}$$

Reiterating,

$$\begin{aligned} &P(\mathbf{r}_g, \mathbf{r}_s, \omega): \quad && \text{Field in actual medium,} \quad \text{DATA} = P|_{\text{ms}} \\ &G(\mathbf{r}_g, \mathbf{r}_s, \omega): \quad && \text{Modeled field in current medium model iteration.} \end{aligned} \tag{3}$$

The modelled field G depends on $s_0^{(n)}(\mathbf{r})$, but the field P, from a projection of which onto the measurement surface we derive our data, does not. We will keep track of things that do so depend via their arguments, e.g., by expressing $G = G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)})$.

Objective function and its expansion

We will evaluate the proximity of $s_0^{(n)}(\mathbf{r})$ to the actual distribution of medium velocities by considering the quantity δP:

$$\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}) \equiv P(\mathbf{r}_g, \mathbf{r}_s, \omega) - G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}). \tag{4}$$

We will examine the problem of updating s_0, from $s_0^{(n)}$ to $s_0^{(n+1)}$ so that the magnitude of δP is maximally reduced. This update will coincide with a minimization of the objective function

$$\Phi\left(s_0^{(n)}\right) \equiv \frac{1}{2} \int d\omega \left(\sum_{s,g} |\delta P|^2 \right). \tag{5}$$

In Appendix A we establish via Gâteaux theory the particulars allowing $\Phi\left(s_0^{(n)} + \delta s_0^{(n)}\right)$ to be be approximated by:

$$\Phi\left(s_0^{(n)} + \delta s_0^{(n)}\right) \approx \Phi\left(s_0^{(n)}\right) + \int d\mathbf{r}' \frac{\partial \Phi}{\partial s_0^{(n)}(\mathbf{r}')} \delta s_0^{(n)}(\mathbf{r}'). \tag{6}$$

Since extrema of Φ are identified through analysis of its gradients, we compute

$$\begin{aligned} \frac{\partial \Phi\left(s_0^{(n)} + \delta s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}')} &\approx \frac{\partial \Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}')} + \frac{\partial}{\partial s_0^{(n)}(\mathbf{r}')} \int d\mathbf{r} \left[\frac{\partial \Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r})} \right] \delta s_0^{(n)}(\mathbf{r}) \\ &= \frac{\partial \Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}')} + \int d\mathbf{r} \left[\frac{\partial^2 \Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}') \partial s_0^{(n)}(\mathbf{r})} \right] \delta s_0^{(n)}(\mathbf{r}) \\ &= g^{(n)}(\mathbf{r}') + \int d\mathbf{r} H^{(n)}(\mathbf{r}', \mathbf{r}) \delta s_0^{(n)}(\mathbf{r}), \end{aligned} \tag{7}$$

where

$$g^{(n)}(\mathbf{r}') = \frac{\partial \Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}')} \tag{8}$$

is the gradient and

$$H^{(n)}(\mathbf{r}', \mathbf{r}) = \frac{\partial^2 \Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}')\partial s_0^{(n)}(\mathbf{r})} \tag{9}$$

is the Hessian. We point out that the Hessian is symmetric under an exchange of $\mathbf{r}$ and $\mathbf{r}'$.

One Gauss-Newton step

Equations (7)–(9) express the objective function in one place in terms of its value and its derivatives at another. It is the functional equivalent of the univariate Taylor's series approximation $f'(x + \Delta x) \approx f'(x) + f''(x)\Delta x$. It is the beginning point for taking a step towards the nearest minimum of Φ. If the step $\delta s_0^{(n)}(\mathbf{r})$ is to take us to this point on Φ, i.e., where the left hand side of equation (7) is zero, it evidently must be true that

$$\int d\mathbf{r} H^{(n)}(\mathbf{r}', \mathbf{r})\delta s_0^{(n)}(\mathbf{r}) = -g^{(n)}(\mathbf{r}'). \tag{10}$$

We assume an inverse Hessian $H^{(n)-}(\mathbf{r}, \mathbf{r}')$ defined such that

$$\int d\mathbf{r}' H^{(n)}(\mathbf{r}'', \mathbf{r}')H^{(n)-}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r}'' - \mathbf{r}), \tag{11}$$

by multiplying equation (10) through by $H^{(n)-}$ and integrating we have

$$\begin{aligned}
&\int d\mathbf{r}' H^{(n)-}(\mathbf{r}'', \mathbf{r}') \left[\int d\mathbf{r} H^{(n)}(\mathbf{r}, \mathbf{r}')\delta s_0^{(n)}(\mathbf{r})\right] = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r}'', \mathbf{r}')g^{(n)}(\mathbf{r}') \\
&\int d\mathbf{r}\ \delta s_0^{(n)}(\mathbf{r}) \int d\mathbf{r}' H^{(n)}(\mathbf{r}, \mathbf{r}')H^{(n)-}(\mathbf{r}'', \mathbf{r}') = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r}'', \mathbf{r}')g^{(n)}(\mathbf{r}') \\
&\int d\mathbf{r}\ \delta s_0^{(n)}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'') = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r}'', \mathbf{r}')g^{(n)}(\mathbf{r}')
\end{aligned} \tag{12}$$

or

$$\delta s_0^{(n)}(\mathbf{r}'') = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r}'', \mathbf{r}')g^{(n)}(\mathbf{r}'). \tag{13}$$

So, a Newton step towards minimizing Φ is devised by deriving appropriate forms for the gradient $g^{(n)}(\mathbf{r})$ and the inverse Hessian $H^{(n)-}(\mathbf{r}, \mathbf{r}')$. Both of these quantities have wave physical interpretations.

2. A NONLINEAR PERTURBATIVE DERIVATION OF THE GRADIENT

We begin with a full perturbative derivation of the form of the gradient. Substituting equation (5) into equation (8) we have for the gradient

$$\begin{aligned} g^{(n)}(\mathbf{r}) &= \frac{\partial\Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r})} \\ &= \frac{1}{2}\sum_{s,g}\int d\omega\left[-P\frac{\partial G^*(s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r})} - P^*\frac{\partial G(s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r})} + \frac{\partial G^*(s_0^{(n)})G(s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r})}\right] \\ &= \frac{1}{2}\sum_{s,g}\int d\omega[(G-P)^*G' + (G-P)G^{*\prime})], \end{aligned} \tag{14}$$

where the prime indicates the derivative with respect to $s_0^{(n)}$. Since

$$Z + Z^* = \mathrm{Re}Z + i\mathrm{Im}Z + \mathrm{Re}Z - i\mathrm{Im}Z = 2\mathrm{Re}Z, \tag{15}$$

we have

$$g^{(n)}(\mathbf{r}) = -\sum_{s,g}\int d\omega\mathrm{Re}\left\{\frac{\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r})}\delta P^*(\mathbf{r}_g,\mathbf{r}_s,\omega|s_0^{(n)})\right\}. \tag{16}$$

The quantity $\partial G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_0^{(n)})/\partial s_0^{(n)}(\mathbf{r})$ is the Fréchet derivative. Let us determine it by developing a relationship between small changes the field, $\delta G(\mathbf{r}_g,\mathbf{r}_s,\omega|s_0^{(n)})$, and small changes in the medium, $\delta s_0^{(n)}(\mathbf{r})$. Let $G_U(\mathbf{r}_g,\mathbf{r}_s,\omega)$ be the field in an unperturbed state, and let $G_P(\mathbf{r}_g,\mathbf{r}_s,\omega)$ be the field in a perturbed state. Let the former satisfy

$$\left[\nabla_g^2 + \omega^2 s_U^{(n)}(\mathbf{r}_g)\right]G_U(\mathbf{r}_g,\mathbf{r}_s,\omega) = \delta(\mathbf{r}_g - \mathbf{r}_s), \tag{17}$$

and the latter satisfy

$$\left[\nabla_g^2 + \omega^2 s_P^{(n)}(\mathbf{r}_g)\right]G_P(\mathbf{r}_g,\mathbf{r}_s,\omega) = \delta(\mathbf{r}_g - \mathbf{r}_s). \tag{18}$$

In the perturbative view, we express $s_P^{(n)}$ in terms of $s_U^{(n)}$ via

$$s_P^{(n)}(\mathbf{r}) = s_U^{(n)}(\mathbf{r}) + \delta s^{(n)}(\mathbf{r}) \tag{19}$$

so that equation (18) can be written

$$\left[\nabla_g^2 + \omega^2 s_U^{(n)}(\mathbf{r}_g)\right]G_P(\mathbf{r}_g,\mathbf{r}_s,\omega) = \delta(\mathbf{r}_g - \mathbf{r}_s) - \omega^2\delta s(\mathbf{r}_g)G_P(\mathbf{r}_g,\mathbf{r}_s,\omega) \tag{20}$$

and the operator in brackets $[\cdot]$ can be inverted to obtain

$$G_P(\mathbf{r}_g,\mathbf{r}_s,\omega) = G_U(\mathbf{r}_g,\mathbf{r}_s,\omega) - \omega^2\int d\mathbf{r}'G_U(\mathbf{r}_g,\mathbf{r}',\omega)\delta s_P(\mathbf{r}')G_P(\mathbf{r}',\mathbf{r}_s,\omega). \tag{21}$$

Now, the difference $G_P - G_U$ is one of our desired quantities, namely $\delta G(\mathbf{r}_g, \mathbf{r}_s, \omega)$. We form this difference on the left hand side and eliminate G_P from the right hand side by expanding in series

$$\begin{aligned}\delta G(\mathbf{r}_g, \mathbf{r}_s, \omega) &= -\omega^2 \int d\mathbf{r}' G_U(\mathbf{r}_g, \mathbf{r}', \omega)\delta s(\mathbf{r}')G_U(\mathbf{r}', \mathbf{r}_s, \omega) \\ &+ \omega^4 \int d\mathbf{r}' G_U(\mathbf{r}_g, \mathbf{r}', \omega)\delta s(\mathbf{r}') \int d\mathbf{r}'' G_U(\mathbf{r}', \mathbf{r}'', \omega)\delta s(\mathbf{r}'')G_U(\mathbf{r}'', \mathbf{r}_s, \omega) \\ &+ \ldots \ .\end{aligned} \tag{22}$$

We next further specify the perturbation, such that it characterizes a local change at an arbitrary $\mathbf{r}$. That is, we set

$$\delta s(\mathbf{r}') = \delta s \times \delta(\mathbf{r}' - \mathbf{r}), \tag{23}$$

where δs is now a scalar quantity localized at the point $\mathbf{r}$. Substituting equation (23) into equation (22), the integrals are easily evaluated and we obtain

$$\begin{aligned}&\delta G(\mathbf{r}_g, \mathbf{r}_s, \omega) \\ &= -\omega^2 \delta s G(\mathbf{r}_g, \mathbf{r}, \omega)G(\mathbf{r}, \mathbf{r}_s, \omega)\left[1 - \omega^2 \delta s G_{\mathrm{R}} + \omega^4 \delta s^2 G_{\mathrm{R}}^2 - \ldots\right] \\ &= -\frac{\omega^2 \delta s G(\mathbf{r}_g, \mathbf{r}, \omega)G(\mathbf{r}, \mathbf{r}_s, \omega)}{1 + \omega^2 G_{\mathrm{R}} \delta s},\end{aligned} \tag{24}$$

where in the second step we have recognized and collapsed the series $(1+x)^{-1} = 1 - x + x^2 - \ldots$. The quantity $G_{\mathrm{R}} \equiv G(\mathbf{r}, \mathbf{r}, \omega | s_0^{(n)})$ is the Green's function evaluated with source and receiver coincident at $\mathbf{r}$, which is singular in multiple dimensions. Our sense is that this is primarily a mathematical issue, since the variation in any real wave field that comes with a small change in the medium must surely not be infinite. We take this as an indication that some finite principle value is probably available to assign to G_{R}, whose exact value is not relevant now. Continuing, we are in a position to compute the Fréchet derivative. If

$$\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r})} = \lim_{\delta s \to 0} \frac{\delta G}{\delta s}, \tag{25}$$

then forming this ratio in equation (24) and taking the limit the denominator goes to unity and we confirm that

$$\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)})}{\partial s_0^{(n)}(\mathbf{r})} = -\omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_0^{(n)})G(\mathbf{r}, \mathbf{r}_s, \omega | s_0^{(n)}), \tag{26}$$

and hence, from (16), that

$$g^{(n)}(\mathbf{r}) = \sum_{s,g} \int d\omega\ \omega^2 [G(\mathbf{r}, \mathbf{r}_s, \omega | s_0^{(n)})] \times [G(\mathbf{r}_g, \mathbf{r}, \omega | s_0^{(n)})\delta P^*(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)})]. \tag{27}$$

This gradient is the source of the "migration" interpretation of full waveform inversion, namely, that the time reversed residuals, δP^*, are propagated into the medium via $G(\mathbf{r}_g, \mathbf{r}, \omega | s_0^{(n)})$, the source field is propagated into the medium via $G(\mathbf{r}, \mathbf{r}_s, \omega | s_0^{(n)})$, and the gradient is formed by correlating the two.

3. WAVE INTERPRETATION OF THE HESSIAN / APPROXIMATE HESSIAN

There is a similar wave/geophysical interpretation available to the Hessian. We may leverage some of the effort of computing $g^{(n)}(\mathbf{r})$ to compute $H^{(n)}(\mathbf{r},\mathbf{r}')$. Since

$$H^{(n)}(\mathbf{r},\mathbf{r}') = \frac{\partial^2\Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r})\partial s_0^{(n)}(\mathbf{r}')} = \frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}\left[\frac{\partial\Phi\left(s_0^{(n)}\right)}{\partial s_0^{(n)}(\mathbf{r}')}\right] = \frac{\partial}{\partial s_0^{(n)}(\mathbf{r})} g^{(n)}(\mathbf{r}'), \tag{28}$$

we may divide it up into two parts:

$$\begin{aligned} H^{(n)}(\mathbf{r},\mathbf{r}') &= \frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}\sum_{s,g}\int d\omega\; \omega^2 G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega)\delta P^*(\mathbf{r}_g,\mathbf{r}_s,\omega) \\ &= H_1^{(n)}(\mathbf{r},\mathbf{r}') + H_2^{(n)}(\mathbf{r},\mathbf{r}'), \end{aligned} \tag{29}$$

where

$$H_1^{(n)}(\mathbf{r},\mathbf{r}') = \sum_{s,g}\int d\omega\; \omega^2\left[\frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega)\right]\delta P^*(\mathbf{r}_g,\mathbf{r}_s,\omega), \tag{30}$$

and

$$H_2^{(n)}(\mathbf{r},\mathbf{r}') = \sum_{s,g}\int d\omega\; \omega^2 G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega)\left[\frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}\delta P^*(\mathbf{r}_g,\mathbf{r}_s,\omega)\right]. \tag{31}$$

(We have omitted the dependence of the fields and residuals on $s_0^{(n)}$ to save space.) Consider $H_1^{(n)}(\mathbf{r},\mathbf{r}')$ first. Making use of equation (26), we have that

$$\begin{aligned} &\frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega) \\ &= \frac{\partial G(\mathbf{r}_g,\mathbf{r}',\omega)}{\partial s_0^{(n)}(\mathbf{r})}G(\mathbf{r}',\mathbf{r}_s,\omega) + G(\mathbf{r}_g,\mathbf{r}',\omega)\frac{\partial G(\mathbf{r}',\mathbf{r}_s,\omega)}{\partial s_0^{(n)}(\mathbf{r})} \\ &= -\omega^2\left[G(\mathbf{r}_g,\mathbf{r},\omega)G(\mathbf{r},\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega) + G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r},\omega)G(\mathbf{r},\mathbf{r}_s,\omega)\right] \end{aligned} \tag{32}$$

Hence

$$\begin{aligned} H_1^{(n)}(\mathbf{r},\mathbf{r}') = -\sum_{s,g}\int d\omega\; \omega^4[&G(\mathbf{r}_g,\mathbf{r},\omega)G(\mathbf{r},\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega) \\ &+ G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r},\omega)G(\mathbf{r},\mathbf{r}_s,\omega)]\delta P^*(\mathbf{r}_g,\mathbf{r}_s,\omega). \end{aligned} \tag{33}$$

Next $H_2^{(n)}(\mathbf{r},\mathbf{r}')$. Since

$$\begin{aligned} \frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}\delta P^* &= \frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}\left[P^*(\mathbf{r}_g,\mathbf{r}_s,\omega) - G^*(\mathbf{r}_g,\mathbf{r}_s,\omega)\right] \\ &= -\frac{\partial}{\partial s_0^{(n)}(\mathbf{r})}G^*(\mathbf{r}_g,\mathbf{r}_s,\omega) \\ &= \omega^2 G^*(\mathbf{r}_g,\mathbf{r},\omega)G^*(\mathbf{r},\mathbf{r}_s,\omega), \end{aligned} \tag{34}$$

we have

$$H_2^{(n)}(\mathbf{r},\mathbf{r}') = \sum_{s,g} \int d\omega\ \omega^4 G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega)G^*(\mathbf{r}_g,\mathbf{r},\omega)G^*(\mathbf{r},\mathbf{r}_s,\omega). \tag{35}$$

The full Hessian, therefore, is expressible in terms of Green's functions propagating in the background medium.

The approximate Hessian

Note that if the residuals δP are small, $H_1^{(n)}$ may be neglected, in which case the Hessian is approximately

$$H^{(n)}(\mathbf{r},\mathbf{r}') \approx \sum_{s,g} \int d\omega\ \omega^4 G(\mathbf{r}_g,\mathbf{r}',\omega)G(\mathbf{r}',\mathbf{r}_s,\omega)G^*(\mathbf{r}_g,\mathbf{r},\omega)G^*(\mathbf{r},\mathbf{r}_s,\omega). \tag{36}$$

This approximate wave form for the Hessian is consistent with the approximate Hessian used in a quasi-Newton step of full waveform inversion, a point we will further clarify in the next section.

4. GAIN CORRECTION IN A QUASI-NEWTON STEP

We may also discuss the Hessian in seismic migration terms. Since the original remarks directing our thinking on this issue come from Virieux and Operto (2009) we will use the language of that paper, which is matrix-vector rather than functional. The interpretation of the Gauss-Newton result for the parameter update vector $\Delta\mathbf{p}$, as given by those authors, yields to us a further interpretation of a gain correction that is consistent with the deconvolution imaging condition. The relevant equation in that paper states that the update for this vector is the gradient that is premultiplied by the inverse of the approximate and stabilized Hessian, i.e.,

$$\begin{aligned}\Delta\mathbf{p} &= -\mathrm{Re}\left[\left(\mathbf{J}^\dagger\mathbf{W}_d\mathbf{J}\right) + \epsilon\mathbf{W}_m\right]^{-1}\mathrm{Re}\left[\mathbf{J}^\dagger\mathbf{W}_d\Delta\mathbf{d}\right]\\ &= -\mathrm{Re}\left[\left(\mathbf{J}^\dagger\mathbf{W}_d\mathbf{J}\right) + \epsilon\mathbf{W}_m\right]^{-1}\mathrm{Re}\left[\mathbf{J}^T\mathbf{W}_d\Delta\mathbf{d}^*\right].\end{aligned} \tag{37}$$

In equation (37), $\mathbf{J}$ is the Jacobian matrix, $\mathbf{W}_d$ is a data-weighting matrix, $\mathbf{W}_m$ is a regularization matrix and $\Delta\mathbf{d}$ is the data residual at the recievers.

First a point about units. Since the Jacobian, the derivative of $\mathbf{u}$ (the forward-propagated predicted field) with respect to $\mathbf{p}$ has dimensions of data divided by parameters, the inverse Hessian provides us with the necessary gain correction so that the gradient term is multiplied by the proper units. This is seen by doing a unit analysis of equation (37). Denoting the units operator by $[\cdot]$, we have that

$$[\Delta\mathbf{p}] = \left(\frac{\text{data}}{\text{parameters}}\right)^{-2} \times \frac{\text{data}}{\text{parameters}} \times \text{data} = \text{ parameters }. \tag{38}$$

What is still left is explicitly relating $\mathbf{J}$ to the wavefields and scattering effects.

We can examine the gradient term in the numertor of equation (37) and the approximate stabilized inverse Hessian by first noting that

$$\mathbf{J} = \frac{\partial \mathbf{u}}{\partial \mathbf{p}} = \mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial \mathbf{p}}\mathbf{u}. \tag{39}$$

In equation (39), $\mathbf{B}$ is the forward modelling operator, $\mathbf{B}^{-1}$ is the Green's operator, and the derivative of $\mathbf{B}$ with respect to a particular member of $\mathbf{p}$, p_i, represents the scattering effect of a spatial Dirac impulse at the appropriate point. We now look at the gradient, the "numerator" in equation (37). Substituting equation (39) into $\mathrm{Re}(\mathbf{J}^T\mathbf{W}_d\Delta\mathbf{d}^*)$, and for simplicity, setting $\mathbf{W}_d = \mathbf{I}$, we obtain

$$\text{gradient} = \mathrm{Re}\left(\mathbf{J}^T\Delta\mathbf{d}^*\right) = \mathrm{Re}\left[\left(\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial \mathbf{p}}\mathbf{u}\right)^T \Delta\mathbf{d}^*\right] \tag{40}$$

Expansion of the transpose in equation (40), results in the final expression for the gradient as

$$\text{gradient} = \mathrm{Re}\left[\mathbf{u}^T \times \frac{\partial \mathbf{B}}{\partial \mathbf{p}}^T \times \underbrace{\mathbf{B}^{-1^T}\Delta\mathbf{d}^*}_{\substack{\text{back-propagated, time}\\ \text{reversed residual}}}\right]. \tag{41}$$

The gradient computed in equation (41) corresponds to the gradient computed in equation (27). At first sight there appear to be some differences, as there are two Green's operators in equation (27), which are not explict in (41). However, in equation (41) the field $\mathbf{u}$ is $\mathbf{B}^{-1}$ multiplying the source term in frequency, which we have taken to be unity. Furthermore, due to the choice of the Helmholtz equation parameter s, the term $(\partial\mathbf{B}/\partial\mathbf{p})^T$ is simply ω^2. Thus, utilizing the foregoing, equation (41) corresponds precisely to equation (27).

The final result for the gradient, as shown in equation (41), represents conventional reverse time migration, as a cross-correlation of the forward propagated modeled field with the backpropagated data. However, there is no gain correction.

Next, we consider the inverse approximate Hessian, which is like a "denominator" in equation (37). If we substitute equation (39) into the approximate Hessian as given by equation (37),and setting $W_d = I$, we have that

$$\mathrm{Re}\left[\left(\mathbf{J}^\dagger\mathbf{W}_d\mathbf{J}\right) + \epsilon\mathbf{W}_m\right] = \mathrm{Re}\left[\underbrace{\left(\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial \mathbf{p}} \times \mathbf{u}\right)^\dagger \left(\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial \mathbf{p}} \times \mathbf{u}\right)}_{\text{KEY TERM}} + \epsilon\mathbf{W}_m\right]. \tag{42}$$

Expanding the KEY TERM, and setting ϵ to zero for simplicity, we obtain

$$\text{KEY TERM} = \mathrm{Re}\left[\mathbf{u}^\dagger \overbrace{\left(\frac{\partial \mathbf{B}}{\partial \mathbf{p}}\right)^\dagger \underbrace{\mathbf{B}^{-1^\dagger}\mathbf{B}^{-1}}_{\text{geometrical spreading}} \frac{\partial \mathbf{B}}{\partial \mathbf{p}}}^{\text{scatterer weighting}} \times \mathbf{u}\right] \tag{43}$$

The expression (43) corresponds precisely to exactly to equation (36), using the same correspondences as for the gradient.

The KEY TERM is the autocorrelation of the modeled forward propagated wavefield u with the gain correction for geometrical spreading and a scatterer weighting operator. If we consider the symbolic division of the gradient by the inverse approximate Hessian, we recover the deconvolution imaging condition.

5. QUASI-NEWTON STEPS AND DECONVOLUTION IMAGING CONDITIONS

In the previous section, we showed how to derive the approximate discrete Hessian using the discrete Jacobian. It is now a straightforward matter to show directly how this amounts to migration with a deconvolution imaging condition. We begin with equation (37).

We substitute equations (41) and (43) into equation (37), and use the correspondences described in the previous section. Then we obtain

$$\Delta\mathbf{p} = \frac{\text{gradient}}{\text{KEY TERM}} = \frac{\text{Re}\left[\mathbf{u}^T \times \omega^2 \times \left(\mathbf{B}^{-1}\right)^T \times \Delta\mathbf{d}^*\right]}{\text{Re}\left[\mathbf{u}^\dagger \times \omega^4 \left(\mathbf{B}^{-1}\right)^\dagger \left(\mathbf{B}^{-1}\right) \times \mathbf{u}\right]} \tag{44}$$

Now the term $\left(\mathbf{B}^{-1}\right)^\dagger \left(\mathbf{B}^{-1}\right)$ represents geometrical spreading, which for a homogeneous medium is r^{-2}; the term $\left(\mathbf{B}^{-1}\right)^T \times \Delta\mathbf{d}^*$ is the backpropagated time-reversed data residual, which in reverse-time migration is simply the back-propagated time-reversed data. With these simplifications in mind, we have

$$\Delta\mathbf{p} = \frac{\text{gradient}}{\text{KEY TERM}} = \frac{r^2}{\omega^2} \underbrace{\frac{\text{Re}\left(\mathbf{u}^T \times \text{BPTR}\right)}{\text{Re}\left(\mathbf{u}^\dagger\mathbf{u}\right)}}_{\substack{\text{deconvolution}\\ \text{imaging condition}}}. \tag{45}$$

We recognize that equation (45) in the time-domain is the gain-corrected zero lag cross-correlation between the downward propagated field and the time-reversed recorded data field divided by the autocorrelation of the downward propagated field. This is equivalent, after appropriate gain, to deconvolving the back-propagated data by the downward propagated data at the image point. We have demonstrated that the first quasi-Newton step in the least squares inversion algorithm has the exact analog to common processing work flow in industrial reverse-time migration.

Let us re-express our arguments less formally. We argue that the simplest form of full waveform inversion, gradient-based stepping, uses a correlation imaging condition that lacks gain correction. It follows that a line search to scale the gradient is a crude average gain correction, but probably not as good as standard practice. We have argued that the approximate Hessian, used in the quasi-Newton approach, is as a gain correction and has a direct interpretation as applying a deconvolution imaging condition. As used in industry practice, the deconvolution imaging condition is a direct estimate of a reflection coefficient. Since we are seeking an update to an impedance model, which is a velocity model in the constant density acoustic approximation, it is desirable to convert the reflection coefficient into an impedance update. In Margrave et al. (2010) this was done entirely by matching to well control; however, we could also use the approximation

$$R = \frac{\Delta I}{2I} \rightarrow R_k = \frac{\Delta I_k}{2I_{k-1}} \rightarrow \Delta I_k = 2I_{k-1}R_k. \tag{46}$$

In the last expression, the impedance model at iteration $k-1$ is used to scale the reflection coefficient estimated at iteration k to obtain an impedance update for iteration k. The estimate for R_k might come from a deconvolution imaging condition applied in an industry standard migration, or from a correlation imaging condition if the data are gained before migration. The estimate of ΔI_k is presumably what is obtained from a quasi-Newton implementation of full waveform inversion.

DISCUSSION AND CONCLUSIONS

Seismic imaging, which we take as generally as possible to mean the creation of a subsurface image from seismic data by any means, has advanced a great deal in the roughly 40 years since the advent of digital data processing (in the early 1970s). Most of the imaging algorithms developed in the first 30 years, or up until about 2000, have been called migration algorithms. In the last ten years, inversion methods such as FWI have been steadily gaining prominence and we are gradually learning what novelty they have to offer. It should come as no surprise, that migration and FWI are closely related since both are derived from the same wave theory and since the developers of FWI initially thought of it as iterative migration (e.g., Lailly, 1983).

In last year's CREWES report, Margrave et al. (2010) presented an iterative migration scheme using well control and simple impedance inversion as a practical approach to FWI. They advocated using standard industry migration algorithms, such as the depth-stepping PSPI, as alternatives to reverse-time migration. Here we have extended that perspective by examining the role of Claerbout's two imaging conditions (correlation and deconvolution Claerbout, 1971) in full waveform inversion. Standard migration practice often uses the correlation imaging condition, because of its inherent stability, but always the data is gain corrected first. Less common, but still frequent, is the use of a stabilized deconvolution imaging condition. Simple arguments suffice to show that the latter does not require a prior gain correction and can directly estimate reflection coefficients (provided that source and receiver consistent amplitude scalars have been applied).

ACKNOWLEDGMENTS

The authors are grateful for the support of the CREWES project and its sponsors. The authors would also like to acknowledge Daryl Van Vorst (EECE – UBC) whose discussions with Matt Yedlin helped shed light on aspects of the interpretation of the Jacobian.

APPENDIX A: GÂTEAUX DERIVATIVES - A PIECE OF CAKE!

Gradients of the type we take in equation (6) derive from the theory of Gâteaux derivatives. The Gâteaux differential or Gâteaux derivative is an essential mathematical tool that is used to compute the variation of a functional. In our case the functional is the misfit between the predicted data and measured data and is calculated using Euclidean distance squared (the L_2 norm). The variation is taken with respect to the velocity parameters describing our functional. Usually, this variation is computed using the Fréchet derivative, which is a stronger form of the derivative. With new misfit functionals, we may have to appeal to the Gâteaux derivative to define gradients properly.

Simply put the Gâteaux derivative is a generalization to function spaces of the idea of the directional derivative, usually taught in a second year calculus course. Thus, we begin, via a basic example, with a simple review of the directional derivative. Let's assume that we have a function that is smooth and we can take its derivatives. Denote this scalar function, in two dimensions, by $F(\mathbf{u})$, with $\mathbf{u}$ being the point (x, y). We would like to compute its derivative in the direction $\mathbf{v}$, where $\mathbf{v} = (a, b)$. A reasonable way to do this would be to write it as

$$dF(\mathbf{u};\mathbf{v}) = \lim_{t\to 0}\frac{F(\mathbf{u}+t\mathbf{v}) - F(\mathbf{u})}{t} = \frac{d}{dt}F(\mathbf{u}+t\mathbf{v})\bigg|_{t=0} \tag{47}$$

Let's now do this explicitly using the chain rule for derivatives:

$$\begin{aligned}
\frac{d}{dt}F(\mathbf{u}+t\mathbf{v})\bigg|_{t=0} &= \frac{d}{dt}F(x+ta, y+tb)\bigg|_{t=0} \\
&= \frac{\partial F(x+ta, y+tb)}{\partial(x+ta)}\frac{d\,(x+ta)}{dt} + \frac{\partial F(x+ta, y+tb)}{\partial(y+tb)}\frac{d\,(y+tb)}{dt}\bigg|_{t=0} \\
&= \frac{\partial F(x,y)}{\partial x}a + \frac{\partial F(x,y)}{\partial y}b \\
&= \nabla F(x,y)\cdot\mathbf{v}
\end{aligned} \tag{48}$$

We can see from (48) that the directional derivative is a scalar product of the gradient of $F(x, y)$ and a displacement vector $\mathbf{v}$ and as a result, we also see the motivation for using the expression Gâteaux differential instead of Gâteaux derivative.

Now with the motivation behind us, we use exactly the same definition for the Gâteaux differential as in (47). The only difference is that we generalize all the items in (47). The point $\mathbf{u} = (x, y)$ becomes a member of a topological vector space, denoted by X. On this space an inner product is defined. If our space is a Hilbert space then the inner product of two functions f and g is given by

$$< f, g >= \int f^*(x)g(x)\,dx, \tag{49}$$

where the integral is performed over the domain of the function and $*$ refers to complex conjugate. The vector $\mathbf{v}$, is now also a member of the Hilbert space, X. We can call it ϕ. Our scalar function $F(x, y)$ becomes our functional, which is a mapping from X to the real numbers. Two points are worth noting. First, there is no single Gâteaux differential at $\mathbf{u}$. The differential depends on ϕ. Secondly, the Gâteaux differential may be nonlinear in ϕ, which is not the case for the Fréchet derivative. Thus a functional which is Fréchet differentiable is Gâteaux differentiable, while the converse is not true. We can now write our Gâteaux differential as

$$dF(\mathbf{u};\phi) = \lim_{t\to 0} \frac{F(\mathbf{u}+t\phi) - F(\mathbf{u})}{t} = \frac{d}{dt} F(\mathbf{u}+t\phi)\bigg|_{t=0}. \tag{50}$$

A simple example is now in order. Consider the functional

$$F(\mathbf{u}) = \int_a^b \left[\left(\frac{\partial \mathbf{u}}{\partial x}\right)^2 + \mathbf{u}^2 \right] dx. \tag{51}$$

So, now

$$dF(\mathbf{u};\phi) = \frac{d}{dt} \left\{ \int_a^b \left[\left(\frac{\partial \mathbf{u}}{\partial x} + t\frac{\partial \phi}{\partial x} \right)^2 + (\mathbf{u}+t\phi)^2 \right] dx \right\}\bigg|_{t=0}. \tag{52}$$

After interchanging differentiation with respect to t and integration with respect to x, under the usual smoothness assumptions, and then setting $t = 0$, we finally obtain

$$dF(\mathbf{u};\phi) = \int_a^b \left(\mathbf{u}\phi + \frac{\partial \mathbf{u}}{\partial x}\frac{\partial \phi}{\partial x} \right) dx. \tag{53}$$

We can now integrate by (53) by parts and assume that $\frac{\partial \mathbf{u}}{\partial x}$ vanishes at the endpoints a and b and finally obtain

$$dF(\mathbf{u};\phi) = \int_a^b \left(\mathbf{u} - \frac{\partial^2 \mathbf{u}}{\partial x^2} \right) \phi dx, \tag{54}$$

which is clearly an inner product of ϕ with

$$\left(\mathbf{u} - \frac{\partial^2 \mathbf{u}}{\partial x^2} \right).$$

REFERENCES

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, No. 3, 467–481.

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.

Margrave, G. F., Ferguson, R. J., and Hogan, C. M., 2010, Full waveform inversion with wave equation migration and well control: CREWES Annual Research Report, **22**.

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.