Introduction

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 use 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.

Summary of results

Full waveform inversion (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) is solved when an objective function is minimized. This happens by taking

(i) Gauss-Newton (or just Newton) steps,
(ii) Quasi-Newton steps, or
(iii) Gradient-based steps
towards the minimum. Gradient based steps are the most common, but researchers have begun to consider quasi-Newton steps. Our purpose is to interpret such steps. In particular:

1. We illustrate the role of the gradient and inverse Hessian in taking a single Gauss-Newton step towards the solution;
2. We re-derive using a scattering formulation the migration/correlation interpretation of gradient based stepping;
3. We extend this wave-based interpretation to include the Hessian;
4. We flesh out the statement of Virieux (2009), that using the inverse approximate Hessian applies a gain correction;
5. We identify (4) as equivalent to use of a deconvolution imaging condition in the migration interpretation.

Gain correction in a Quasi-Newton step

We may also discuss the Hessian in seismic migration terms, using matrix-vector rather than functional notation. The interpretation of the Newton result for the parameter update vector \( \Delta p \) yields a further interpretation of a gain correction consistent with the deconvolution imaging condition. The update is:

\[
\Delta p = - \text{Re} \left[ \frac{\partial^2 \Phi}{\partial p^2} \right]^{-1} \text{Re} \left[ J^T W d \Delta d^T \right].
\]  

(1)

\( J \) is the Jacobian matrix, \( W_d \) is a data-weighting matrix, \( W_m \) is a regularization matrix and \( \Delta d \) is the data residual. Since the Jacobian has dimensions of data / parameters, the inverse Hessian provides the necessary gain so that the gradient is multiplied by the proper units. Denoting the units operator by \( \left[ \right] \), we have

\[
\left[ \Delta p \right] = \left( \frac{\text{data}}{\text{parameters}} \right)^2 \times \frac{\text{data}}{\text{parameters}} \times \text{data} = \text{parameters}\n
\]  

(2)

What is still left explicitly relates \( J \) to the wavefields and scattering effects. We can examine the gradient term and the approximate stabilized inverse Hessian by first noting that

\[
J = \frac{i u}{i p} = B^{-1} \frac{i u}{i p}.
\]  

(3)

In equation (3), B is the forward modelling operator, \( B^{-1} \) is the Green's operator, and the derivative of B with respect to a particular member of \( p, p \), represents the scattering effect of a spatial Dirac impulse at the appropriate point. We now look at the gradient, the "numerator" in equation (1). Substituting equation (3) into \( \text{Re}(J^T W_d \Delta d^T) \), and for simplicity, setting \( W_d = I \), we obtain

\[
\text{gradient} = \text{Re} \left[ J^T \Delta d^T \right] = \text{Re} \left[ B^{-1} \frac{i u}{i p} \right] \Delta d^T
\]  

(4)

Expansion of the transpose in equation (4), results in the final expression for the gradient as (real part implied):

\[
\text{gradient} = \frac{i u}{i p} \times \frac{B^{-1}}{B^{-1}} \times \frac{i u}{i p} \times \frac{B^{-1} - B^{-1}}{B^{-1} - B^{-1} \Delta d^T}.
\]  

(5)

The gradient in (5), which corresponds to \( g \) in the previous section, represents reverse time migration: a cross-correlation of the modeled field with the backpropagated data. However, there is no gain correction. The inverse approximate Hessian is incorporated by substituting (3) into (1) with \( W_d = I \). The real parts become

\[
\left( J^T W_d + \varepsilon W_m \right) = \left( B^{-1} \frac{i u}{i p} \right) \times \left( B^{-1} \frac{i u}{i p} \right) = \text{KEY TERM}.
\]  

(6)

Expanding the real part of the KEY TERM, and setting \( \varepsilon \to 0 \), we have

\[
\text{KEY TERM} = \frac{i u}{i p} \times \frac{B^{-1}}{B^{-1}} \times \frac{i u}{i p} \times \frac{B^{-1} - B^{-1}}{B^{-1} - B^{-1} \Delta d^T}.
\]  

(7)

The KEY TERM is the autocorrelation of the modeled wavefield \( u \) gain-corrected for geometrical spreading with a scatterer weighting operator. Hence we recover a slight modification of the deconvolution imaging condition.

Quasi-Newton and deconvolution imaging conditions

It is now possible to show directly how quasi-Newton amounts to migration with a deconvolution imaging condition. We begin with equation (1). We substitute (5) and (7) into (1), obtaining

\[
\Delta p = \text{gradient KEY TERM} = \frac{i u}{i p} \times \frac{B^{-1}}{B^{-1}} \times \frac{i u}{i p} \times \frac{B^{-1} - B^{-1}}{B^{-1} - B^{-1} \Delta d^T}.
\]  

(8)

The term \( B^{-1} \Delta d \) represents geometrical spreading, which for a homogeneous medium is \( r^{-2} \), the term \( B^{-1} - B^{-1} \Delta d^T \) is the backpropagated time-reversed (BPTR) data residual. With these simplifications in mind, we have

\[
\Delta p = \text{gradient KEY TERM} = \frac{i u}{i p} \times \frac{B^{-1}}{B^{-1}} \times \frac{i u}{i p} \times \frac{B^{-1} - B^{-1} \Delta d^T}{B^{-1} - B^{-1} \Delta d^T}.
\]  

(9)

This in the time-domain is the gain-corrected zero lag cross-correlation between the downward propagated field and the time-reversed data, divided by the autocorrelation of the downward propagated field. This is equivalent to deconvolving the back-propagated data by the downward propagated data at the image point.

Discussion

The simplest form of full waveform inversion, gradient-based stepping, uses a correlation imaging condition that lacks gain correction. The approximate Hessian, used in the quasi-Newton approach, has as a gain correction and has a direct interpretation as applying a deconvolution imaging condition. 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, one converts the \( R \) into an impedance update. In Margrave et al. (2010), this was done by matching to well control, however, we could also use the approximation

\[
R = \frac{\Delta f}{2 \Delta f} \to R_k = \frac{\Delta f}{2 \Delta f} \to \Delta f_k = 2 \Delta f - R_k,
\]  

(10)

in which the impedance model at iteration \( k - 1 \) scales the reflection coefficient at iteration \( k \), to updating the impedance for iteration \( k \). \( R_k \) might come from a deconvolution imaging condition, or from a correlation imaging condition if the data are gained before migration. The estimate of \( \Delta f_k \) is presumably what is obtained from a quasi-Newton implementation of full waveform inversion.

Bibliography