Full waveform inversion and the inverse Hessian

Gary Margrave[1], Matt Yedlin[2] and Kris Innanen[1]

[1] Dept. Geoscience, University of Calgary,
[2] Dept. Electrical and Computer Engineering, UBC

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 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.

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.

Gradient and Hessian: continuous forms

In the corresponding CREWES report, we derive the gradient and Hessian functions assuming an objective function of the form

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

where δP are the data residuals, finding, respectively

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

and

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

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 $\Delta\mathbf{p}$ yields a further interpretation of a gain correction consistent with the deconvolution imaging condition. The update is:

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

$\mathbf{J}$ 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. 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 $[\cdot]$, we have

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

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

$$\mathbf{J} = \frac{\partial\mathbf{u}}{\partial\mathbf{p}} = \mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\mathbf{p}}\mathbf{u}. \tag{3}$$

In equation (3), $\mathbf{B}$ 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 (1). Substituting equation (3) into $\mathrm{Re}(\mathbf{J}^T\mathbf{W}_d\Delta\mathbf{d}^*)$, and for simplicity, setting $\mathbf{W}_d = \mathbf{I}$, we obtain

$$\text{gradient} = \mathrm{Re}\left(\mathbf{J}^T\Delta\mathbf{d}^*\right) = \mathrm{Re}\left[\left(\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\mathbf{p}}\mathbf{u}\right)^T\Delta\mathbf{d}^*\right] \tag{4}$$

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

$$\text{gradient} = \mathbf{u}^T \times \frac{\partial\mathbf{B}^T}{\partial\mathbf{p}} \times \underbrace{\mathbf{B}^{-1^T}\Delta\mathbf{d}^*}_{\text{back-propagated, time reversed residual}}. \tag{5}$$

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

$$(\mathbf{J}^\dagger\mathbf{W}_d\mathbf{J}) + \epsilon\mathbf{W}_m = \underbrace{\left(\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\mathbf{p}}\times\mathbf{u}\right)^\dagger\left(\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\mathbf{p}}\times\mathbf{u}\right)}_{\text{KEY TERM}} + \epsilon\mathbf{W}_m. \tag{6}$$

Expanding the real part of the KEY TERM, and setting $\epsilon \to 0$, we have

$$\text{KEY TERM} = \mathbf{u}^\dagger \overbrace{\left(\frac{\partial\mathbf{B}}{\partial\mathbf{p}}\right)^\dagger \underbrace{\mathbf{B}^{-1^\dagger}\mathbf{B}^{-1}}_{\text{geometrical spreading}} \frac{\partial\mathbf{B}}{\partial\mathbf{p}}}^{\text{scatterer weighting}} \times\mathbf{u} \tag{7}$$

The KEY TERM is the autocorrelation of the modeled wavefield $\mathbf{u}$, 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

$$\Delta\mathbf{p} = \frac{\text{gradient}}{\text{KEY TERM}} = \frac{\mathrm{Re}\left[\mathbf{u}^T \times \omega^2 \times \left(\mathbf{B}^{-1}\right)^T \times \Delta\mathbf{d}^*\right]}{\mathrm{Re}\left[\mathbf{u}^\dagger \times \omega^4 \left(\mathbf{B}^{-1}\right)^\dagger\left(\mathbf{B}^{-1}\right) \times \mathbf{u}\right]} \tag{8}$$

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 (BPTR) data residual. With these simplifications in mind, we have

$$\Delta\mathbf{p} = \frac{\text{gradient}}{\text{KEY TERM}} = \frac{r^2}{\omega^2}\underbrace{\frac{\mathrm{Re}\left(\mathbf{u}^T \times \text{BPTR}\right)}{\mathrm{Re}\left(\mathbf{u}^\dagger\mathbf{u}\right)}}_{\text{deconvolution imaging condition}}. \tag{9}$$

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, is 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

$$R = \frac{\Delta I}{2I} \to R_k = \frac{\Delta I_k}{2I_{k-1}} \to \Delta I_k = 2I_{k-1}R_k, \tag{10}$$

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 ΔI_k is presumably what is obtained from a quasi-Newton implementation of full waveform inversion.

Bibliography

- 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.