A framework for multiparameter full waveform inversion of precritical reflection seismic data

Kris Innanen

ABSTRACT

Formulas for computing full and quasi-Newton steps in seismic full waveform inversion, specifically designed for pre-critical reflection experiments, are derived. The formulas are partly continuous and partly discrete. The discrete aspect of the problem is connected to the multiplicity of parameters, whereas the continuous aspect is connected to the distribution in space of the unknowns. We analyze the opportunities this formulation provides for forming quasi-Newton steps. There are two different kinds we can invoke, which we refer to as parameter-type and space-type. The parameter-type approximation appears to retain the ability of reflection FWI to correctly update one parameter when several are responsible for the amplitude content of the data. A third approximation can be created by invoking both simultaneously. All three are simple to implement, since they each amount to the setting of different, and well-defined, off-diagonal Hessian operator elements to zero.

INTRODUCTION

In full waveform inversion, or FWI (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009), we iteratively solve for the properties of the subsurface using the techniques of multivariate optimization. Tomographic modes of FWI, for instance in crosswell applications (Pratt and Shipp, 1999), have achieved significant success, but less progress has been reported for reflection modes. A great deal of the information in surface seismic data comes from pre-critical reflections, and if FWI is to become a fully realized technology, the challenges associated with this data type must be addressed.

Progress has been slow in this area with some exceptions (Hicks and Pratt, 2001), but there has been noticeable community movement in this direction recently. A significant fraction of the talks at the 2013 SEG FWI sessions involved inverting reflection data (Brossier et al., 2013; Wang et al., 2013). Some have even treated precritical reflection data, though given the issues associated with missing wavelengths, Kelly et al. (2013) classifies positive outcomes in this regime “rare”. We are heartened, however, by evidence that the incorporation of well-control in iterative inversion (Margrave et al., 2010) could be a practical source for missing wavelengths. In this paper we will consider practical pre-critical reflection FWI to be a meaningful possibility.

A pressing theoretical issue is that reflection FWI efforts have been primarily concerned with the scalar problem (i.e., inversion for P-wave velocity). Several decades of industry experience with AVO (e.g., Foster et al., 2010) has clarified that any discussion of “full waveforms” in reflection seismology requires at a minimum an isotropic elastic model to proceed. Less than this, and the measured seismic amplitudes are inadequately accounted for. A framework for reflection FWI must be flexible enough to incorporate multiple parameters.

Fortunately, multiparameter FWI has moved forward significantly outside of the pre-critical reflection regime. The September 2013 issue of the Leading Edge contains several papers on the subject (e.g., Plessix et al., 2013; Operto et al., 2013). Although the focus in many of these studies has been on anisotropic models and their effect on travel times rather than amplitudes, many of the mathematical points likely carry over. For instance, purely gradient-based methods suffer because the update in any one parameter proceeds absent accounting for the fact that data phases and amplitudes are co-determined by more than that one parameter. Errors caused by this problem are referred to as "cross-talk". Operto et al. (2013) show that, given N discrete points in the Earth at which we assign M independent parameters, the Hessian appears as an $M \times M$ block matrix, in which each element is itself an $N \times N$ matrix. The submatrices account for coupling between parameters, and the authors infer that the incorporation of the inverse Hessian in an FWI step should reduce cross-talk. Only the off-diagonal components of the Hessian are sensitive to coupling between different parameters, and so it is in these components the mechanism for cross-talk reduction must reside.

Knowing no reason why these albeit qualitative ideas should not also hold for pre-critical reflection mode data, our analysis will focus on the character of the Hessian and various approximations thereof. We seek a formulation of FWI which easily allows us to analyze and manipulate the Hessian. The formulation should facilitate algorithm development, but also naturally provide insights into the character of the multiparameter inverse problem. We do this by taking a mixed continuous-discrete approach, in which the distribution of properties in space is treated continuously but the individual parameters remain discrete. In this formulation, approximate Hessians can be devised which involve rejection of various types of off-diagonal matrix/functional elements.

In this paper the goal, in short, is to analyze FWI in terms of the parameters which specifically determine precritical reflection amplitudes. We treat the 2-parameter case and the 3-parameter case individually. In a companion paper in this report (Innanen, 2013), we will use this FWI framework to comment directly on the role of AVO information in determining the direction of gradient-based, full Newton and quasi-Newton steps in reflection mode FWI.

SUMMARY OF RESULTS

Because of its rather involved logical pathways, here we provide a short summary of the key results of this paper. The remainder of the paper is involved with deriving these results and providing details concerning their ingredients and how to determine them. We consider 1, 2 and 3 parameter cases. We label the unknowns in each of these cases s, (s_κ, s_ρ), and (s_P, s_S, s_ρ), to correspond with the scalar case (involving wave velocity), the acoustic case (involving bulk modulus and density), and the elastic-isotropic case (involving P-wave velocity, S-wave velocity, and density) respectively. This labelling is strictly unnecessary, since the results in this paper hold for *any* 1-, 2- and 3-parameter problems, but to accord to the equations a sense of their likely eventual use, we have done so.

Full Newton updates

At the nth FWI iteration we have in hand the distribution in space $(\mathbf{r})$ of one model parameter, say $s^{(n)}(\mathbf{r})$, and we wish to update it, adding a step $\delta s^{(n)}(\mathbf{r})$ to attain the $n+1$th iterate:

$$s^{(n+1)}(\mathbf{r}) = s^{(n)}(\mathbf{r}) + \delta s^{(n)}(\mathbf{r}). \tag{1}$$

It is not difficult to show (e.g., Margrave et al., 2011) that in the continuous, full-Newton case, that update is given by

$$\delta s(\mathbf{r}) = -\int d\mathbf{r}' H^{-1}(\mathbf{r}, \mathbf{r}') g(\mathbf{r}'), \tag{2}$$

where g is the gradient of a suitably chosen objective function and H^{-1} is the functional inverse of the Hessian function. Both of these quantities are formed by combining the observed data with the medium properties of the nth iteration and the modeled wave field propagating through this nth medium. The two-parameter (e.g., acoustic) problem is similar, but with two continuous functions of space rather than one, organized into a vector:

$$\begin{bmatrix} s_\kappa^{(n+1)}(\mathbf{r}) \\ s_\rho^{(n+1)}(\mathbf{r}) \end{bmatrix} = \begin{bmatrix} s_\kappa^{(n)}(\mathbf{r}) \\ s_\rho^{(n)}(\mathbf{r}) \end{bmatrix} + \begin{bmatrix} \delta s_\kappa^{(n)}(\mathbf{r}) \\ \delta s_\rho^{(n)}(\mathbf{r}) \end{bmatrix}. \tag{3}$$

The update steps $\delta s_\kappa^{(n)}(\mathbf{r})$ and $\delta s_\rho^{(n)}(\mathbf{r})$ are calculated through a formula generalizing equation (2):

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} = -\int d\mathbf{r}' \mathcal{H}_2^{-1}(\mathbf{r}, \mathbf{r}') \int d\mathbf{r}'' \begin{bmatrix} H_{\rho\rho}(\mathbf{r}', \mathbf{r}'') & -H_{\rho\kappa}(\mathbf{r}', \mathbf{r}'') \\ -H_{\kappa\rho}(\mathbf{r}', \mathbf{r}'') & H_{\kappa\kappa}(\mathbf{r}', \mathbf{r}'') \end{bmatrix} \begin{bmatrix} g_\kappa(\mathbf{r}'') \\ g_\rho(\mathbf{r}'') \end{bmatrix},$$

where $g_\kappa(\mathbf{r})$ and $g_\rho(\mathbf{r})$ are the gradients for each of the two parameters (e.g., bulk modulus and density), the four functions $H_{\kappa\kappa}$, $H_{\kappa\rho}$, $H_{\rho\kappa}$ and $H_{\rho\rho}$ are generalizations of the function H in the scalar case ($\mathcal{H}_2^{-1}$ is the inverse of $\mathcal{H}_2$, which itself is a combination of the above four Hessian functions). As in the scalar case, these six functions are completely determined by the measured data and the nth iteration of the medium and the wave field. The three-parameter (e.g., elastic-isotropic) case is a straightforward extension of the two-parameter case. Three model updates are now considered:

$$\begin{bmatrix} s_P^{(n+1)}(\mathbf{r}) \\ s_S^{(n+1)}(\mathbf{r}) \\ s_\rho^{(n+1)}(\mathbf{r}) \end{bmatrix} = \begin{bmatrix} s_P^{(n)}(\mathbf{r}) \\ s_S^{(n)}(\mathbf{r}) \\ s_\rho^{(n)}(\mathbf{r}) \end{bmatrix} + \begin{bmatrix} \delta s_P^{(n)}(\mathbf{r}) \\ \delta s_S^{(n)}(\mathbf{r}) \\ \delta s_\rho^{(n)}(\mathbf{r}) \end{bmatrix}. \tag{4}$$

On the left hand side, the current model iterates, $s_P^{(n)}(\mathbf{r})$, $s_S^{(n)}(\mathbf{r})$ and $s_\rho^{(n)}(\mathbf{r})$, are transformed into the next model iterates by adding the steps, which are determined through the formula

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} = -\int d\mathbf{r}' \mathcal{H}_3^{-1}(\mathbf{r}, \mathbf{r}') \int d\mathbf{r}'' \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{bmatrix} \begin{bmatrix} g_P(\mathbf{r}'') \\ g_S(\mathbf{r}'') \\ g_\rho(\mathbf{r}'') \end{bmatrix}.$$

As before, the functions $g_P(\mathbf{r})$, $g_S(\mathbf{r})$ and $g_\rho(\mathbf{r})$ represent the gradients for each of the three parameters (e.g., P-wave velocity, S-wave velocity and density), the $\Lambda_{ij} = \Lambda_{ij}(\mathbf{r}', \mathbf{r}'')$ are combinations of Hessian functions similar to those of the two-parameter case, and $\mathcal{H}_3^{-1}$ is the inverse of a function $\mathcal{H}_3$ which is itself a combination of the functions Λ_{ij}. The twelve functions are again all determined by the measured data and the models/wavefields of the current iterate.

Quasi-Newton updates

Quasi-Newton steps involve approximate versions of the formulas in the previous section, in which only parts of the functions H_{ij} and $\mathcal{H}_k$ and the matrices containing them are invoked. Usually the Hessian is approximated by neglecting its off-diagonal elements. A hybrid continuous-discrete formulation involves two different ways of neglecting off-diagonal elements, each with different consequences for the direction of the step.

Neglecting diagonal elements in the scalar Hessian in equation (2) takes the form of disallowing contributions to the integral from $H^{-1}(\mathbf{r}, \mathbf{r}')$ when $\mathbf{r} \neq \mathbf{r}'$, that is, setting $H(\mathbf{r}, \mathbf{r}') = \Gamma(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')$ such that

$$\delta s(\mathbf{r}) \approx -\frac{1}{\Gamma(\mathbf{r})} g(\mathbf{r}). \tag{5}$$

We will refer to this as the *parameter-type Hessian approximation*, since what is neglected in the Hessian accounts for correlation between distinct points in space $\mathbf{r}$ and $\mathbf{r}'$ where $\mathbf{r} \neq \mathbf{r}'$, while leaving the influence of parameters amongst themselves alone. We can see this more clearly by applying the same approximation to the two parameter case:

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\frac{1}{\Gamma_{\kappa\kappa}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{\rho\kappa}(\mathbf{r})\Gamma_{\kappa\rho}(\mathbf{r})} \begin{bmatrix} \Gamma_{\rho\rho}(\mathbf{r}) & -\Gamma_{\rho\kappa}(\mathbf{r}) \\ -\Gamma_{\kappa\rho}(\mathbf{r}) & \Gamma_{\kappa\kappa}(\mathbf{r}) \end{bmatrix} \begin{bmatrix} g_\kappa(\mathbf{r}) \\ g_\rho(\mathbf{r}) \end{bmatrix}. \tag{6}$$

The 2×2 matrix representing the interaction of modulus with density and vice versa is retained, and thus, evidently, the ability for the modulus to be updated in the presence of variations in both modulus and density. The functions Γ_{ij} are completely determined by the modelled waves and the data at the nth step; these must be found through analysis if we wish to say something mathematical about the update, but they can be found by simply setting off-diagonal Hessian matrix (and/or submatrix) elements to zero in implementations.

The opposite approximation is available to us also, in which we maintain the possibility of correlation from one spatial position to another, but neglect off-diagonal terms corresponding to inter-parameter coupling. This leads to a different two-parameter update formula:

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\int d\mathbf{r}' \begin{bmatrix} H_{\kappa\kappa}^{-1}(\mathbf{r}, \mathbf{r}') g_\kappa(\mathbf{r}') \\ H_{\rho\rho}^{-1}(\mathbf{r}, \mathbf{r}') g_\rho(\mathbf{r}') \end{bmatrix}. \tag{7}$$

We will refer to this as the *space-type Hessian approximation*. Finally, imposing both parameter and space type restrictions on the Hessian, we have a third formula:

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\begin{bmatrix} \Gamma_{\kappa\kappa}^{-1}(\mathbf{r}) g_\kappa(\mathbf{r}) \\ \Gamma_{\rho\rho}^{-1}(\mathbf{r}) g_\rho(\mathbf{r}) \end{bmatrix}. \tag{8}$$

The three parameter case is a straightforward extension of the two parameter case, with a parameter-type approximation:

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} = -\frac{1}{\mathcal{H}_0(\mathbf{r})} \begin{bmatrix} H'_{11}(\mathbf{r}) & H'_{12}(\mathbf{r}) & H'_{13}(\mathbf{r}) \\ H'_{21}(\mathbf{r}) & H'_{22}(\mathbf{r}) & H'_{23}(\mathbf{r}) \\ H'_{31}(\mathbf{r}) & H'_{32}(\mathbf{r}) & H'_{33}(\mathbf{r}) \end{bmatrix} \begin{bmatrix} g_P(\mathbf{r}) \\ g_S(\mathbf{r}) \\ g_\rho(\mathbf{r}) \end{bmatrix}, \tag{9}$$

and space-type approximation:

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\int d\mathbf{r}' \begin{bmatrix} H^{-1}_{PP}(\mathbf{r},\mathbf{r}')g_P(\mathbf{r}') \\ H^{-1}_{SS}(\mathbf{r},\mathbf{r}')g_S(\mathbf{r}') \\ H^{-1}_{\rho\rho}(\mathbf{r},\mathbf{r}')g_\rho(\mathbf{r}') \end{bmatrix}, \tag{10}$$

and a mixed approximation:

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\begin{bmatrix} \Gamma^{-1}_{PP}(\mathbf{r})g_P(\mathbf{r}) \\ \Gamma^{-1}_{SS}(\mathbf{r})g_S(\mathbf{r}) \\ \Gamma^{-1}_{\rho\rho}(\mathbf{r})g_\rho(\mathbf{r}) \end{bmatrix}, \tag{11}$$

all available.

MINIMIZATION SCHEME

Univariate, bivariate and trivariate templates

The seismic inverse problem is multivariate. The scalar problem, in which one value of a discrete or continuous variable must be determined for each point in space, exposes one aspect of this multiplicity. Using Gâteaux methods (Margrave et al., 2011) this scalar minimization problem is straightforward to derive. Interestingly, it can also be written down with very little formal mathematics by first writing down a *univariate* minimization problem as a template, and then replacing the simple operations it involves (e.g., derivatives) with the correct continuous operator. The two-parameter FWI problem can likewise be derived using a bivariate template, and the three-parameter problem using a trivariate template.

Let us first review the simple minimization schemes which will act as templates. The univariate, bivariate and trivariate minimization problems involve the objective functions

$$\begin{aligned} \phi_1 &= \phi_1(x), \\ \phi_2 &= \phi_2(x,y), \quad \text{and} \\ \phi_3 &= \phi_3(x,y,z). \end{aligned} \tag{12}$$

Starting somewhere on these surfaces ϕ_i, we seek to take roughly "downhill" steps which take us to the minimum. In the univariate case, one Newton step is found by recognizing that when this step, say δx, is taken, it must be true that for a quadratic ϕ_1:

$$g_1 = -H_1\delta x, \quad \text{where} \quad g_1 = \frac{d\phi_1}{dx} \quad \text{and} \quad H_1 = \frac{d^2\phi_1}{dx^2}. \tag{13}$$

The use of g_1 and H_1 seems a little overblown for such a simple case, but it helps in interpreting the slightly increased complexity of the bivariate version of this relationship:

$$\mathbf{g}_2 = -\mathbf{H}_2 \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}, \quad \text{where} \quad \mathbf{g}_2 = \begin{bmatrix} \partial_x \phi_2 \\ \partial_y \phi_2 \end{bmatrix} \quad \text{and} \quad \mathbf{H}_2 = \begin{bmatrix} \partial_{xx}\phi_2 & \partial_{xy}\phi_2 \\ \partial_{yx}\phi_2 & \partial_{yy}\phi_2 \end{bmatrix},$$

and the trivariate version:

$$\mathbf{g}_3 = -\mathbf{H}_3 \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix}, \quad \text{where} \quad \mathbf{g}_3 = \begin{bmatrix} \partial_x \phi_3 \\ \partial_y \phi_3 \\ \partial_z \phi_3 \end{bmatrix} \quad \text{and} \quad \mathbf{H}_3 = \begin{bmatrix} \partial_{xx}\phi_3 & \partial_{xy}\phi_3 & \partial_{xz}\phi_3 \\ \partial_{yx}\phi_3 & \partial_{yy}\phi_3 & \partial_{yz}\phi_3 \\ \partial_{zx}\phi_3 & \partial_{zy}\phi_3 & \partial_{zz}\phi_3 \end{bmatrix}.$$

With these relationships in hand, we may solve for either one, two, or three step lengths:

$$\delta x = -H_1^{-1} g_1, \quad \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = -\mathbf{H}_2^{-1}\mathbf{g}_2, \quad \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix} = -\mathbf{H}_3^{-1}\mathbf{g}_3. \tag{14}$$

In the univariate case, the solution requires a multiplicative inverse, or reciprocal only, but in the bivariate and trivariate cases we must invert small matrices. The formulas for these inverses are:

$$\mathbf{H}_2^{-1} = \frac{1}{\det \mathbf{H}_2} \begin{bmatrix} \partial_{yy}\phi_2 & -\partial_{xy}\phi_2 \\ -\partial_{yx}\phi_2 & \partial_{xx}\phi_2 \end{bmatrix},$$

where

$$\det \mathbf{H}_2 = \partial_{xx}\phi_2 \partial_{yy}\phi_2 - \partial_{yx}\phi_2 \partial_{xy}\phi_2, \tag{15}$$

and

$$\mathbf{H}_3^{-1} = \frac{1}{\det \mathbf{H}_3} \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix},$$

where

$$\begin{aligned}
&K_{11} = \partial_{yy}\phi_3\partial_{zz}\phi_3 - \partial_{zy}\phi_3\partial_{yz}\phi_3, \quad K_{12} = \partial_{zy}\phi_3\partial_{xz}\phi_3 - \partial_{xy}\phi_3\partial_{zz}\phi_3 \\
&K_{13} = \partial_{xy}\phi_3\partial_{zy}\phi_3 - \partial_{yy}\phi_3\partial_{xz}\phi_3, \quad K_{21} = \partial_{zx}\phi_3\partial_{yz}\phi_3 - \partial_{yx}\phi_3\partial_{zz}\phi_3 \\
&K_{22} = \partial_{xx}\phi_3\partial_{zz}\phi_3 - \partial_{xz}\phi_3\partial_{zx}\phi_3, \quad K_{23} = \partial_{yx}\phi_3\partial_{xz}\phi_3 - \partial_{xx}\phi_3\partial_{yz}\phi_3 \\
&K_{31} = \partial_{yx}\phi_3\partial_{zy}\phi_3 - \partial_{zx}\phi_3\partial_{yy}\phi_3, \quad K_{32} = \partial_{zx}\phi_3\partial_{xy}\phi_3 - \partial_{xx}\phi_3\partial_{zy}\phi_3 \\
&K_{33} = \partial_{xx}\phi_3\partial_{yy}\phi_3 - \partial_{xy}\phi_3\partial_{yx}\phi_3
\end{aligned}$$

and where

$$\begin{aligned}
\det \mathbf{H}_3 = &\ \partial_{xx}\phi_3 \left(\partial_{yy}\phi_3\partial_{zz}\phi_3 - \partial_{yz}\phi_3\partial_{zy}\phi_3\right) \\
&- \partial_{xy}\phi_3 \left(\partial_{yx}\phi_3\partial_{zz}\phi_3 - \partial_{yz}\phi_3\partial_{zx}\phi_3\right) \\
&+ \partial_{xz}\phi_3 \left(\partial_{yx}\phi_3\partial_{zy}\phi_3 - \partial_{yy}\phi_3\partial_{zx}\phi_3\right).
\end{aligned} \tag{16}$$

We will refer to these formulas in the following sections, where we develop their use as templates for multiparameter FWI.

The univariate template and the minimization of a scalar function of one function

In Margrave et al. (2011) we spent some time developing the scalar FWI problem, using methods based on Gâteaux theory. Here we will review the minimization, focusing on the fact that the scalar problem can be reproduced by starting from the univariate minimization formula in equation (14), and simply replacing arithmetic operations with the right integral operations. In scalar minimization, rather than an objective function of one variable, we invoke a scalar function of one function of space:

$$\Phi_1\left(s(\mathbf{r})\right). \tag{17}$$

A generalized Taylor's series expansion of the objective function about s is truncated as follows:

$$\Phi_1\left(s+\delta s\right) \approx \Phi_1\left(s\right) + \int d\mathbf{r}' \frac{\partial \Phi_1\left(s\right)}{\partial s(\mathbf{r}')}\delta s(\mathbf{r}'). \tag{18}$$

From this model we could find the generalized roots of Φ_1, but we may transform it to a minimization problem by replacing Φ_1 with its derivative with respect to the model:

$$\frac{\partial \Phi_1\left(s+\delta s\right)}{\partial s(\mathbf{r})} \approx \frac{\partial \phi_1\left(s\right)}{\partial s(\mathbf{r})} + \int d\mathbf{r}' \frac{\partial^2 \Phi_1\left(s\right)}{\partial s(\mathbf{r})\partial s(\mathbf{r}')}\delta s(\mathbf{r}'), \tag{19}$$

or, using the g, H notation of equation (13),

$$\frac{\partial \Phi_1\left(s+\delta s\right)}{\partial s(\mathbf{r})} \approx g_1(\mathbf{r}) + \int d\mathbf{r}' H_1(\mathbf{r},\mathbf{r}')\delta s(\mathbf{r}'), \tag{20}$$

where

$$g_1(\mathbf{r}) = \frac{\partial \phi_1\left(s\right)}{\partial s(\mathbf{r})}, \quad H_1(\mathbf{r},\mathbf{r}') = \frac{\partial^2 \Phi_1\left(s\right)}{\partial s(\mathbf{r})\partial s(\mathbf{r}')}. \tag{21}$$

If, within this approximation, the step δs carries us to the nearest local minimum of Φ_1, that step must set the left hand side of equation (20) to zero. Solving for δs under these circumstances we arrive at the formula

$$\delta s(\mathbf{r}) = -\int d\mathbf{r}' H_1^{-1}(\mathbf{r},\mathbf{r}') g_1(\mathbf{r}'), \tag{22}$$

where H_1^{-1} is the inverse of H_1 in the sense that

$$\int d\mathbf{r}'' H_1^{-1}(\mathbf{r},\mathbf{r}'') H_1(\mathbf{r}'',\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}'). \tag{23}$$

With this in hand, and assuming that this δs has been constructed to perform the nth update, we may then calculate the $n+1$th iterate:

$$s^{(n+1)}(\mathbf{r}) = s^{(n)}(\mathbf{r}) + \delta s^{(n)}(\mathbf{r}). \tag{24}$$

Equation (22) is the key equation here, and it is the analogue of the univariate equation, i.e.,

$$\delta s(\mathbf{r}) = -\int d\mathbf{r}' H_1^{-1}(\mathbf{r},\mathbf{r}')g_1(\mathbf{r}') \quad \leftrightarrow \quad \delta x = -H_1^{-1}g_1. \tag{25}$$

Thus in moving from a univariate to a unifunctional minimization problem, the product of the inverse Hessian and the gradient is transformed to an inner product involving functional extensions of derivatives of the kind in equation (21), and reciprocals are replaced with inner products with inverse functions defined as per equation (23).

Quasi-Newton step

In functional minimization, a Hessian element is considered to be off-diagonal if $\mathbf{r} \neq \mathbf{r}'$, and on-diagonal if $\mathbf{r} = \mathbf{r}'$. Thus, we create an approximate Hessian involving its diagonal components only with the replacement

$$H(\mathbf{r},\mathbf{r}') \approx \Gamma(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}'), \tag{26}$$

where the weights $\Gamma(\mathbf{r})$, while left unspecified here, are fully determined by the problem at hand. When substituted into equation (25), this leads the quasi-Newton step

$$\delta s(\mathbf{r}) \approx -\frac{1}{\Gamma(\mathbf{r})} g(\mathbf{r}). \tag{27}$$

The bivariate template and acoustic minimization

Let us next treat the acoustic minimization problem, arriving quickly at the desired results by using the bivariate template and the relations implied by (25). We now conceive of a scalar objective function which depends on two scalar functions, $s_\kappa(\mathbf{r})$ and $s_\rho(\mathbf{r})$:

$$\Phi_2\left(s_\kappa(\mathbf{r}), s_\rho(\mathbf{r})\right). \tag{28}$$

The bivariate template, a two-variable minimization problem has the form

$$\mathbf{g}_2 = -\mathbf{H}_2 \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} \tag{29}$$

or

$$\begin{bmatrix} \partial_x \phi_2 \\ \partial_y \phi_2 \end{bmatrix} = \begin{bmatrix} \partial_{xx}\phi_2 & \partial_{xy}\phi_2 \\ -\partial_{yx}\phi_2 & \partial_{yy}\phi_2 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}, \tag{30}$$

and the formula for the steps δx and δy is

$$\begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = -\frac{1}{\partial_{xx}\phi_2\partial_{yy}\phi_2 - \partial_{yx}\phi_2\partial_{xy}\phi_2} \begin{bmatrix} \partial_{yy}\phi_2 & -\partial_{yx}\phi_2 \\ -\partial_{xy}\phi_2 & \partial_{xx}\phi_2 \end{bmatrix} \begin{bmatrix} \partial_x \phi_2 \\ \partial_y \phi_2 \end{bmatrix}. \tag{31}$$

In order to produce the continuous version of this problem, we recall that arithmetic products are replaced with inner products, and multiplication with reciprocal quantities is replaced with functional inverses defined as in equation (23). Thus, the determinant $\partial_{xx}\phi_2\partial_{yy}\phi_2 - \partial_{yx}\phi_2\partial_{xy}\phi_2$ is replaced with the quantity

$$\mathcal{H}_2(\mathbf{r},\mathbf{r}') = \int d\mathbf{r}'' \left[H_{\kappa\kappa}(\mathbf{r},\mathbf{r}'')H_{\rho\rho}(\mathbf{r}'',\mathbf{r}') - H_{\rho\kappa}(\mathbf{r},\mathbf{r}'')H_{\kappa\rho}(\mathbf{r}'',\mathbf{r}')\right], \tag{32}$$

where the functions H_{ij} take the place of the second derivatives:

$$H_{\kappa\kappa}(\mathbf{r},\mathbf{r}') = \frac{\partial^2\Phi_2\left(s_\kappa, s_\rho\right)}{\partial s_\kappa(\mathbf{r})\partial s_\kappa(\mathbf{r}')}, \quad H_{\kappa\rho}(\mathbf{r},\mathbf{r}') = \frac{\partial^2\Phi_2\left(s_\kappa, s_\rho\right)}{\partial s_\kappa(\mathbf{r})\partial s_\rho(\mathbf{r}')}, \tag{33}$$

and

$$H_{\rho\kappa}(\mathbf{r},\mathbf{r}') = \frac{\partial^2\Phi_2\left(s_\kappa, s_\rho\right)}{\partial s_\rho(\mathbf{r})\partial s_\kappa(\mathbf{r}')}, \quad H_{\rho\rho}(\mathbf{r},\mathbf{r}') = \frac{\partial^2\Phi_2\left(s_\kappa, s_\rho\right)}{\partial s_\rho(\mathbf{r})\partial s_\rho(\mathbf{r}')}. \tag{34}$$

The two parameter acoustic Newton step is then written

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} = \int d\mathbf{r}'\mathcal{H}_2^{-1}(\mathbf{r},\mathbf{r}') \int d\mathbf{r}'' \begin{bmatrix} -H_{\rho\rho}(\mathbf{r}',\mathbf{r}'') & H_{\rho\kappa}(\mathbf{r}',\mathbf{r}'') \\ H_{\kappa\rho}(\mathbf{r}',\mathbf{r}'') & -H_{\kappa\kappa}(\mathbf{r}',\mathbf{r}'') \end{bmatrix} \begin{bmatrix} g_\kappa(\mathbf{r}'') \\ g_\rho(\mathbf{r}'') \end{bmatrix}, \tag{35}$$

where the gradients are, explicitly,

$$g_\kappa(\mathbf{r}) = \frac{\partial\Phi_2\left(s_\kappa, s_\rho\right)}{\partial s_\kappa(\mathbf{r})}, \quad g_\rho(\mathbf{r}) = \frac{\partial\Phi_2\left(s_\kappa, s_\rho\right)}{\partial s_\rho(\mathbf{r})} \tag{36}$$

and where $\mathcal{H}_2^{-1}$ is the inverse of $\mathcal{H}_2$ in the sense that

$$\int d\mathbf{r}''\mathcal{H}_2^{-1}(\mathbf{r}'',\mathbf{r}')\mathcal{H}_2(\mathbf{r}'',\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}'). \tag{37}$$

After the calculation of this set of two step lengths, we may add them to the current (say, nth) model iterates to complete the update:

$$\begin{bmatrix} s_\kappa^{(n+1)}(\mathbf{r}) \\ s_\rho^{(n+1)}(\mathbf{r}) \end{bmatrix} = \begin{bmatrix} s_\kappa^{(n)}(\mathbf{r}) \\ s_\rho^{(n)}(\mathbf{r}) \end{bmatrix} + \begin{bmatrix} \delta s_\kappa^{(n)}(\mathbf{r}) \\ \delta s_\rho^{(n)}(\mathbf{r}) \end{bmatrix}. \tag{38}$$

The parameter-type approximate Hessian

In large scale implementations of FWI, full Newton steps are impractical, since inverting the Hessian matrix is computationally intensive. In gradient-based methods this is avoided by setting the Hessian equal to the identity matrix, or a delta function in the continuous case. A middle ground is a quasi-Newton step, in which the Hessian is approximated by some quantity which is constructed and inverted relatively easily and yet provides some of the what the full Hessian accomplished.

We permit two different types of Hessian approximation natural to the hybrid continuous-discrete framework to be made. We will refer to them as *parameter-type* approximations and *space-type* approximations, to distinguish the sense in which the approximation is made. Both approximations involve neglecting non-diagonal parts of the Hessian. Parameter-type approximations arise from making choices similar to that in equation (26), in which parts of the function $H(\mathbf{r}, \mathbf{r}')$ where $\mathbf{r} = \mathbf{r}'$ were retained, and all other parts were neglected. As in the scalar case, in the two parameter acoustic case this is brought about by setting

$$\begin{aligned} H_{\kappa\kappa}(\mathbf{r}, \mathbf{r}') &\approx \Gamma_{\kappa\kappa}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'), \quad H_{\kappa\rho}(\mathbf{r}, \mathbf{r}') \approx \Gamma_{\kappa\rho}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'), \\ H_{\rho\kappa}(\mathbf{r}, \mathbf{r}') &\approx \Gamma_{\rho\kappa}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'), \quad H_{\rho\rho}(\mathbf{r}, \mathbf{r}') \approx \Gamma_{\rho\rho}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'), \end{aligned} \tag{39}$$

that is, by insisting that the Hessian functions be expressible as delta functions with the weights $\Gamma_{\kappa\kappa}(\mathbf{r})$, $\Gamma_{\kappa\rho}(\mathbf{r})$, $\Gamma_{\rho\kappa}(\mathbf{r})$ and $\Gamma_{\rho\rho}(\mathbf{r})$ (whose precise analytic forms are dictated by the particular dimensionality, data set, and iteration number). Substituting these forms into equation (32), we firstly find that

$$\begin{aligned} \mathcal{H}_2(\mathbf{r}, \mathbf{r}') &\approx \int d\mathbf{r}'' \left[\Gamma_{\kappa\kappa}\delta(\mathbf{r} - \mathbf{r}'')\Gamma_{\rho\rho}\delta(\mathbf{r}'' - \mathbf{r}') - \Gamma_{\rho\kappa}\delta(\mathbf{r} - \mathbf{r}'')\Gamma_{\kappa\rho}\delta(\mathbf{r}'' - \mathbf{r}')\right] \\ &\approx \left[\Gamma_{\kappa\kappa}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{\rho\kappa}(\mathbf{r})\Gamma_{\kappa\rho}(\mathbf{r})\right]\delta(\mathbf{r} - \mathbf{r}'), \end{aligned} \tag{40}$$

which has the straightforward inverse

$$\mathcal{H}_2^{-1}(\mathbf{r}, \mathbf{r}') \approx \left[\Gamma_{\kappa\kappa}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{\rho\kappa}(\mathbf{r})\Gamma_{\kappa\rho}(\mathbf{r})\right]^{-1}\delta(\mathbf{r} - \mathbf{r}'). \tag{41}$$

Lastly substituting the approximations in equations (39) and (41) into equation (35), and evaluating the remaining integrals using the sifting property of the delta function, we arrive at the formula for the parameter-type quasi-Newton step:

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\frac{1}{\Gamma_{\kappa\kappa}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{\rho\kappa}(\mathbf{r})\Gamma_{\kappa\rho}(\mathbf{r})} \begin{bmatrix} \Gamma_{\rho\rho}(\mathbf{r}) & -\Gamma_{\rho\kappa}(\mathbf{r}) \\ -\Gamma_{\kappa\rho}(\mathbf{r}) & \Gamma_{\kappa\kappa}(\mathbf{r}) \end{bmatrix} \begin{bmatrix} g_\kappa(\mathbf{r}) \\ g_\rho(\mathbf{r}) \end{bmatrix}. \tag{42}$$

Notice that in the 2×2 matrix, off-diagonal elements are still present—those corresponding to coupling between the two parameters. Only the inner products over space are suppressed.

The space-type approximate Hessian

The opposite, space-type approximation, is found by retaining the spatial off-diagonal components of the Hessian functions, but suppressing the off-diagonal elements of the 2×2 system in equation (35). That is, we let

$$\begin{bmatrix} H_{\kappa\kappa}(\mathbf{r}, \mathbf{r}') & H_{\kappa\rho}(\mathbf{r}, \mathbf{r}') \\ H_{\rho\kappa}(\mathbf{r}, \mathbf{r}') & H_{\rho\rho}(\mathbf{r}, \mathbf{r}') \end{bmatrix} \approx \begin{bmatrix} H_{\kappa\kappa}(\mathbf{r}, \mathbf{r}') & 0 \\ 0 & H_{\rho\rho}(\mathbf{r}, \mathbf{r}') \end{bmatrix}, \tag{43}$$

in which case

$$\mathcal{H}_2(\mathbf{r}, \mathbf{r}') \approx \int d\mathbf{r}'' H_{\kappa\kappa}(\mathbf{r}, \mathbf{r}'') H_{\rho\rho}(\mathbf{r}'', \mathbf{r}'). \tag{44}$$

It is not difficult to show that if $H_{\rho\rho}^{-1}$ and $H_{\kappa\kappa}^{-1}$ are the inverses of $H_{\rho\rho}$ and $H_{\kappa\kappa}$ in the usual sense (e.g., equation 23), then $\mathcal{H}_2$ in equation (44) has the inverse

$$\mathcal{H}_2^{-1}(\mathbf{r},\mathbf{r}') \approx \int d\mathbf{r}'' H_{\rho\rho}^{-1}(\mathbf{r},\mathbf{r}'')H_{\kappa\kappa}^{-1}(\mathbf{r}'',\mathbf{r}'). \tag{45}$$

Assuming sufficient symmetry within the Hessian functions that

$$\int d\mathbf{r}'' H_{\rho\rho}^{-1}(\mathbf{r},\mathbf{r}'')H_{\kappa\kappa}^{-1}(\mathbf{r}'',\mathbf{r}') = \int d\mathbf{r}'' H_{\kappa\kappa}^{-1}(\mathbf{r},\mathbf{r}'')H_{\rho\rho}^{-1}(\mathbf{r}'',\mathbf{r}'), \tag{46}$$

holds, the substitution of equations (43) and (45) into equation (35), we find the formula for the space type quasi-Newton step:

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\int d\mathbf{r}' \begin{bmatrix} H_{\kappa\kappa}^{-1}(\mathbf{r},\mathbf{r}')g_\kappa(\mathbf{r}') \\ H_{\rho\rho}^{-1}(\mathbf{r},\mathbf{r}')g_\rho(\mathbf{r}') \end{bmatrix}. \tag{47}$$

Here the Hessian retains the ability to influence the step length, at some particular point in space, using values of the gradient at *other* points in space. However, it has lost the ability to affect the bulk modulus update with density gradient values, and vice versa.

Combined approximate Hessian

The two types of off-diagonal suppression can both be incorporated simultaneously, leading to a fully diagonal Hessian approximation. The result is

$$\begin{bmatrix} \delta s_\kappa(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx - \begin{bmatrix} \Gamma_{\kappa\kappa}^{-1}(\mathbf{r})g_\kappa(\mathbf{r}) \\ \Gamma_{\rho\rho}^{-1}(\mathbf{r})g_\rho(\mathbf{r}) \end{bmatrix}. \tag{48}$$

The trivariate template and the minimization of a scalar function of three functions

The extension of the hybrid continuous-discrete minimization approach to three parameters, appropriate for, e.g., the elastic-isotropic problem, is straightforward, though with each added parameter the formulas become significantly more complicated. The objective function to be minimized is a scalar function of three functions $s_P(\mathbf{r})$, $s_S(\mathbf{r})$, and $s_\rho(\mathbf{r})$:

$$\Phi_3\left(s_P(\mathbf{r}), s_S(\mathbf{r}), s_\rho(\mathbf{r})\right). \tag{49}$$

Applying the trivariate template to this three-parameter update problem in the same manner as in the two parameter case, we arrive at the following formula for a full Newton step:

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} = -\int d\mathbf{r}'\mathcal{H}_3^{-1}(\mathbf{r},\mathbf{r}') \int d\mathbf{r}'' \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{bmatrix} \begin{bmatrix} g_P(\mathbf{r}'') \\ g_S(\mathbf{r}'') \\ g_\rho(\mathbf{r}'') \end{bmatrix}, \tag{50}$$

where the space dependence of the functions $\Lambda_{ij} = \Lambda_{ij}(\mathbf{r}',\mathbf{r}'')$ is implied. These Λ_{ij} are themselves mixtures of Hessian functions similar to those used in the two-parameter case:

$$\begin{aligned}
&\Lambda_{11} = H_{SS}H_{\rho\rho} - H_{\rho S}H_{S\rho},\ \Lambda_{12} = H_{\rho S}H_{P\rho} - H_{PS}H_{\rho\rho}, \\
&\Lambda_{13} = H_{PS}H_{\rho S} - H_{SS}H_{P\rho},\ \Lambda_{21} = H_{\rho P}H_{S\rho} - H_{SP}H_{\rho\rho}, \\
&\Lambda_{22} = H_{PP}H_{\rho\rho} - H_{P\rho}H_{\rho P},\ \Lambda_{23} = H_{SP}H_{P\rho} - H_{PP}H_{S\rho}, \\
&\Lambda_{31} = H_{SP}H_{\rho S} - H_{\rho P}H_{SS},\ \Lambda_{32} = H_{\rho P}H_{PS} - H_{PP}H_{\rho S}, \\
&\Lambda_{33} = H_{PP}H_{SS} - H_{PS}H_{SP},
\end{aligned}$$

where here too the space dependence of each function $H_{AB} = H_{AB}(\mathbf{r}', \mathbf{r}'')$ is implied. These H_{AB} elements have the expected second-derivative Hessian forms

$$H_{AB}(\mathbf{r}, \mathbf{r}') = \frac{\partial^2 \Phi_3 (s_P, s_S, s_\rho)}{\partial s_A(\mathbf{r}) \partial s_B(\mathbf{r}')}. \tag{51}$$

Meanwhile the generalized determinant is expressible in three terms

$$\mathcal{H}_3(\mathbf{r}, \mathbf{r}') = \mathcal{H}_P(\mathbf{r}, \mathbf{r}') + \mathcal{H}_S(\mathbf{r}, \mathbf{r}') + \mathcal{H}_\rho(\mathbf{r}, \mathbf{r}') \tag{52}$$

where

$$\mathcal{H}_P(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' H_{PP}(\mathbf{r}, \mathbf{r}'') \int d\mathbf{r}''' \left[H_{SS}(\mathbf{r}'', \mathbf{r}''') H_{\rho\rho}(\mathbf{r}''', \mathbf{r}') - H_{S\rho}(\mathbf{r}'', \mathbf{r}''') H_{\rho S}(\mathbf{r}''', \mathbf{r}') \right]$$
$$\mathcal{H}_S(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' H_{PS}(\mathbf{r}, \mathbf{r}'') \int d\mathbf{r}''' \left[H_{SP}(\mathbf{r}'', \mathbf{r}''') H_{\rho\rho}(\mathbf{r}''', \mathbf{r}') - H_{S\rho}(\mathbf{r}'', \mathbf{r}''') H_{\rho P}(\mathbf{r}''', \mathbf{r}') \right]$$
$$\mathcal{H}_\rho(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' H_{P\rho}(\mathbf{r}, \mathbf{r}'') \int d\mathbf{r}''' \left[H_{SP}(\mathbf{r}'', \mathbf{r}''') H_{\rho S}(\mathbf{r}''', \mathbf{r}') - H_{SS}(\mathbf{r}'', \mathbf{r}''') H_{\rho P}(\mathbf{r}''', \mathbf{r}') \right],$$

and the inverse of $\mathcal{H}_3$, namely $\mathcal{H}_3^{-1}$, is defined such that

$$\int d\mathbf{r}'' \mathcal{H}_3^{-1}(\mathbf{r}'', \mathbf{r}') \mathcal{H}_3(\mathbf{r}'', \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \tag{53}$$

With these quantities in hand at the nth step in a full Newton inversion, we may update to the $n+1$th iterate via

$$\begin{bmatrix} s_P^{(n+1)}(\mathbf{r}) \\ s_S^{(n+1)}(\mathbf{r}) \\ s_\rho^{(n+1)}(\mathbf{r}) \end{bmatrix} = \begin{bmatrix} s_P^{(n)}(\mathbf{r}) \\ s_S^{(n)}(\mathbf{r}) \\ s_\rho^{(n)}(\mathbf{r}) \end{bmatrix} + \begin{bmatrix} \delta s_P^{(n)}(\mathbf{r}) \\ \delta s_S^{(n)}(\mathbf{r}) \\ \delta s_\rho^{(n)}(\mathbf{r}) \end{bmatrix}. \tag{54}$$

Parameter-type approximate Hessian

In the parameter type approximate Hessian, we neglect all terms in all functions $H_{AB}(\mathbf{r}, \mathbf{r}')$ for which $\mathbf{r} \neq \mathbf{r}'$, that is we again assume a form

$$H_{AB}(\mathbf{r}, \mathbf{r}') \approx \Gamma_{AB}(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'), \tag{55}$$

for all combinations of $\{A, B\} = \{P, S, \rho\}$. This transforms the generalized determinant into

$$\mathcal{H}_3(\mathbf{r}, \mathbf{r}') \approx \mathcal{H}_3^0(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'), \tag{56}$$

where

$$\begin{aligned} \mathcal{H}_3^0(\mathbf{r}) =& \Gamma_{PP}(\mathbf{r}) \left[\Gamma_{SS}(\mathbf{r}) \Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{S\rho}(\mathbf{r}) \Gamma_{\rho S}(\mathbf{r}) \right] \\ &+ \Gamma_{PS}(\mathbf{r}) \left[\Gamma_{SP}(\mathbf{r}) \Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{S\rho}(\mathbf{r}) \Gamma_{\rho P}(\mathbf{r}) \right] \\ &+ \Gamma_{P\rho}(\mathbf{r}) \left[\Gamma_{SP}(\mathbf{r}) \Gamma_{\rho S}(\mathbf{r}) - \Gamma_{SS}(\mathbf{r}) \Gamma_{\rho P}(\mathbf{r}) \right], \end{aligned} \tag{57}$$

with the weights Γ determined by the details of the specific inversion problem at hand. The inverse is thus

$$\mathcal{H}_3^{-1}(\mathbf{r},\mathbf{r}') \approx \frac{1}{\mathcal{H}_3^0(\mathbf{r})}\delta(\mathbf{r}-\mathbf{r}'), \tag{58}$$

and when this is substituted into equation (50) we arrive at the parameter type quasi-Newton step formula

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} = \frac{1}{\mathcal{H}_3^0(\mathbf{r})} \begin{bmatrix} \Lambda'_{11}(\mathbf{r}) & \Lambda'_{12}(\mathbf{r}) & \Lambda'_{13}(\mathbf{r}) \\ \Lambda'_{21}(\mathbf{r}) & \Lambda'_{22}(\mathbf{r}) & \Lambda'_{23}(\mathbf{r}) \\ \Lambda'_{31}(\mathbf{r}) & \Lambda'_{32}(\mathbf{r}) & \Lambda'_{33}(\mathbf{r}) \end{bmatrix} \begin{bmatrix} g_P(\mathbf{r}) \\ g_S(\mathbf{r}) \\ g_\rho(\mathbf{r}) \end{bmatrix}, \tag{59}$$

where

$$\begin{aligned}
\Lambda'_{11}(\mathbf{r}) &= \Gamma_{SS}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{\rho S}(\mathbf{r})\Gamma_{S\rho}(\mathbf{r}),\\
\Lambda'_{12}(\mathbf{r}) &= \Gamma_{\rho S}(\mathbf{r})\Gamma_{P\rho}(\mathbf{r}) - \Gamma_{PS}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}),\\
\Lambda'_{13}(\mathbf{r}) &= \Gamma_{PS}(\mathbf{r})\Gamma_{\rho S}(\mathbf{r}) - \Gamma_{SS}(\mathbf{r})\Gamma_{P\rho}(\mathbf{r}),\\
\Lambda'_{21}(\mathbf{r}) &= \Gamma_{\rho P}(\mathbf{r})\Gamma_{S\rho}(\mathbf{r}) - \Gamma_{SP}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}),\\
\Lambda'_{22}(\mathbf{r}) &= \Gamma_{PP}(\mathbf{r})\Gamma_{\rho\rho}(\mathbf{r}) - \Gamma_{P\rho}(\mathbf{r})\Gamma_{\rho P}(\mathbf{r}),\\
\Lambda'_{23}(\mathbf{r}) &= \Gamma_{SP}(\mathbf{r})\Gamma_{P\rho}(\mathbf{r}) - \Gamma_{PP}(\mathbf{r})\Gamma_{S\rho}(\mathbf{r}),\\
\Lambda'_{31}(\mathbf{r}) &= \Gamma_{SP}(\mathbf{r})\Gamma_{\rho S}(\mathbf{r}) - \Gamma_{\rho P}(\mathbf{r})\Gamma_{SS}(\mathbf{r}),\\
\Lambda'_{32}(\mathbf{r}) &= \Gamma_{\rho P}(\mathbf{r})\Gamma_{PS}(\mathbf{r}) - \Gamma_{PP}(\mathbf{r})\Gamma_{\rho S}(\mathbf{r}), \quad \text{and}\\
\Lambda'_{33}(\mathbf{r}) &= \Gamma_{PP}(\mathbf{r})\Gamma_{SS}(\mathbf{r}) - \Gamma_{PS}(\mathbf{r})\Gamma_{SP}(\mathbf{r}).
\end{aligned}$$

Space-type approximate Hessian

The space-type approximate Hessian and the associated quasi-Newton step can again be brought about by suppressing of off-diagonal elements of the 3×3 matrix in the Newton step in equation (50), as well as the correct parts of the generalized determinant. That is, we set

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{bmatrix} \approx \begin{bmatrix} H_{SS}H_{\rho\rho} & 0 & 0 \\ 0 & H_{PP}H_{\rho\rho} & 0 \\ 0 & 0 & H_{PP}H_{SS} \end{bmatrix}, \tag{60}$$

and further

$$\mathcal{H}_3(\mathbf{r},\mathbf{r}') \approx \int d\mathbf{r}'' \int d\mathbf{r}''' H_{PP}(\mathbf{r},\mathbf{r}'')H_{SS}(\mathbf{r}'',\mathbf{r}''')H_{\rho\rho}(\mathbf{r}''',\mathbf{r}'). \tag{61}$$

This leads through arguments identical to those in the two-parameter case to the space-type quasi-Newton step formula

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx -\int d\mathbf{r}' \begin{bmatrix} H_{PP}^{-1}(\mathbf{r},\mathbf{r}')g_P(\mathbf{r}') \\ H_{SS}^{-1}(\mathbf{r},\mathbf{r}')g_S(\mathbf{r}') \\ H_{\rho\rho}^{-1}(\mathbf{r},\mathbf{r}')g_\rho(\mathbf{r}') \end{bmatrix}. \tag{62}$$

Combined approximate Hessian

Finally, again both types of off-diagonal element suppression can be applied simultaneously, leading to a combined approximate Hessian and quasi-Newton step of

$$\begin{bmatrix} \delta s_P(\mathbf{r}) \\ \delta s_S(\mathbf{r}) \\ \delta s_\rho(\mathbf{r}) \end{bmatrix} \approx - \begin{bmatrix} \Gamma_{PP}^{-1}(\mathbf{r}) g_P(\mathbf{r}) \\ \Gamma_{SS}^{-1}(\mathbf{r}) g_S(\mathbf{r}) \\ \Gamma_{\rho\rho}^{-1}(\mathbf{r}) g_\rho(\mathbf{r}) \end{bmatrix}. \tag{63}$$

CONCLUSIONS

In this paper we have presented some formulas for computing full and quasi-Newton steps in seismic full waveform inversion. The framework for doing so which is partly continuous and partly discrete. The discrete aspect of the problem is directly connected to the multiplicity of parameters — e.g., bulk modulus and density in the two parameter acoustic case, and P-wave velocity, S-wave velocity and density in the three parameter elastic case. The continuous aspect of the problem is connected to the distribution in space of the unknowns. The full Newton step direction and length is then a matrix-integral operation involving quantities which essentiall reproduce the multivariate gradient vector and Hessian matrix in standard optimization theory, but in a way which intuitively exposes their behaviour.

We have analyzed the opportunities a hybrid continuous-discrete formulation provides for quasi-Newton steps, in which efficiently calculable approximations of the Hessian operator are used, rather than the computationally burdensome full Hessian. A common approach in multivariate optimization is to utilize only the diagonal elements of the Hessian matrix. Because our Newton step involves continuous and discrete inner products, there are two different *kinds* of off-diagonal suppression we can engage in. We refer to these as parameter-type approximations and space-type approximations. Qualitatively, the parameter-type approximation would appear to retain the ability of reflection FWI to correctly update one parameter when several are responsible for the amplitude content of the data. We leave a quantitative demonstration of this fact for the next paper in this two-paper set.

ACKNOWLEDGMENTS

This work was funded by the CREWES project. CREWES sponsors and research personnel are thanked.

REFERENCES

Brossier, R., Operto, S., and Virieux, J., 2013, Toward data-domain waveform inversion of reflected waves: SEG Expanded Abstracts, 892–897.

Foster, D. J., Keys, R. G., and Lane, F. D., 2010, Interpretation of AVO anomalies: Geophysics, **75**, 75A3–75A13.

Hicks, G., and Pratt, R. G., 2001, Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit: Geophysics, **66**, 598–612.

Innanen, K. A., 2013, A framework for iterative multiparameter seismic inversion: CREWES Annual Research Report, **25**.

Kelly, S., Ramos-Martinez, J., Zou, K., and Tsimelzon, B., 2013, Inversion of refractions and reflections by full-waveform inversion for marine streamer data: classification of problem types and solution methods: The Leading Edge, **Sep**, 1130–1138.

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**.

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

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., 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., and Shipp, R. M., 1999, Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data: Geophysics, **64**, No. 3, 902–914.

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.

Wang, S., Chen, F., Zhang, H., and Shen, Y., 2013, Reflection-based full waveform inversion (RFWI) in the frequency domain: SEG Expanded Abstracts, 877–881.