SCATTERING-BASED DECOMPOSITION OF SENSITIVITY KERNELS FOR FULL WAVEFORM INVERSION

D. L. Macedo, I. Vasconcelos, and J. Schleicher

email: *dlmacbr@gmail.com*

keywords: *Full waveform inversion, Sensitivity kernels, Scattering theory*

ABSTRACT

With the increasing demand in complexity for subsurface models in environments such as subsalt, sub-basalt and pre-salt, full-waveform inversion (FWI) is quickly becoming one of the model-building methods of choice. While in principle capable of handling all of the nonlinearity in the data, in practice nonlinear gradient-based FWI is limited due to its notorious sensitivity to the choice of starting models. To help addressing model convergence issues in FWI, in this paper we analyse the role of nonlinearity in the so-called sensitivity kernels, which are the centrepiece of gradient-based FWI algorithms. Using a scattering-based approach and assuming acoustic-only data, we start by reparameterizing the subsurface model in terms of smooth and singular components, which contain low and high frequency information respectively, for both compressibility and density. This leads to a decomposition of the data into a reference field that is sensitive only to the smooth model, and a scattered field sensitive to both smooth and sharp model components. Focussing on the model backprojections from the scattered data only, we provide expressions for the Fréchet-derivative sensitivity kernels of all four model parameters. Our results provide for the decomposition of current FWI kernels into no less than thirteen sub-kernels which have explicitly different levels of nonlinearity with respect to the data and the background model. This capability to discern levels of nonlinearity within FWI kernels is key to understanding model convergence in gradient-based, iterative FWI. We illustrate this by analysing some of the sub-kernel terms in detail. The scattering-based FWI kernel decomposition we provide could have broad potential applications, such as devising multiscale FWI algorithms, and improving velocity model building in the image domain using extended image gathers.

INTRODUCTION

For many years, the most common imaging techniques were based on ray theory, such as Kirchhoff migration. But lately, as the industry has been facing geologically more complex areas where these techniques were not successful, new methods based on wave-equation migration came into play – in the beginning, one-way wave-equation based, and more recently, two-way wave-equation. All this became possible due to new acquisition techniques which give better illumination of the subsurface, and more powerful computational capacities.

Those new methods require more and more refined earth models. However, even though migration has advanced quickly with the raise of available computer power, constructing these models is still mostly ray-based. Recently, one tool, based on the two-way wave-equation, has been studied and developed for Earth modelling: full waveform inversion (FWI) (see, e.g. Vigh et al., 2009).

The basic idea behind FWI is the minimization of an objective function that measures the difference between the observed seismic data and synthetic data, modelled using a numerical earth model. In the last decades, many studies on FWI were carried out. For the state of the art, the reader is referred to the works of Virieux and Operto (2009) and Vigh et al. (2009).

A series of paper published in the eighties (Lailly, 1983; Tarantola, 1984, 1986) introduced the gradient-based FWI method in applied geophysics. In essence, these methods propose that a model can be updated iteratively with the help of so-called sensitivity kernels (SK) (Tromp et al., 2005). The SKs are operators that describe the change in the wavefield due to changes in the model parameters. With the adjoint of the SKs, one can evaluate the change in the earth model associated with wavefield residuals.

While in principle capable of handling all aspects of wave propagation contained in the data, including full nonlinearity, in practice nonlinear gradient-based FWI is limited due to its notorious sensitivity to the choice of starting models. To help addressing model convergence issues in FWI, in this paper we analyse the role of nonlinearity in the sensitivity kernels.

To do so, we use a scattering-based approach (Vasconcelos, 2008). Assuming an acoustic medium, we reparameterize the subsurface model in terms of smooth and singular components for both compressibility and density. This leads to a decomposition of the data into a reference field that is sensitive only to the smooth model, and a scattered field sensitive to both smooth and sharp model components. Focussing on the model backprojections from the scattered data only, we provide expressions for the Fréchet-derivative sensitivity kernels for all four model parameters (perturbations of both smooth and singular parts of compressibility and density).

The scattering-based FWI kernel decomposition we provide could have broad potential applications, such as devising multiscale FWI algorithms, and improving velocity model building in the image domain using extended image (EI) gathers (Rickett and Sava, 2002; Sava and Fomel, 2003; Symes, 2008; Sava and Vasconcelos, 2009, 2010). As shown by Vasconcelos et al. (2009, 2010), there is a connection between the extended image conditions and the interferometry formalism: the EIs behave like locally scattered wavefields in the image domain.

FULL-WAVEFORM INVERSION AND SENSITIVITY KERNELS

The acoustic wave equation states a non-linear relation between the wavefield and the model parameters, which can be written as

$$p = f(\boldsymbol{m}), \tag{1}$$

where p is the acoustic wavefield, $\boldsymbol{m}$ are the model parameters, and f denotes the functional relating the two, i.e., the wave equation.

When a perturbation $\boldsymbol{\delta m}$ is introduced in the model parameters, the perturbed wavefield may be expressed as

$$p + \delta p^{true} = f(\boldsymbol{m} + \boldsymbol{\delta m}) = f(\boldsymbol{m}) + \boldsymbol{\Phi}_f \boldsymbol{\delta m} + O(\boldsymbol{\delta m}^2) \,, \tag{2}$$

where δp^{true} is the wavefield perturbation (or residual) and $\boldsymbol{\Phi}_f$ are the first derivatives of the wavefield with respect to the model parameters. The subscript f stands for "full wavefield". The latter are also know as Frèchet derivatives. If only first-order terms in the model perturbations are taken into account, we obtain the linearized wavefield perturbation,

$$\delta p = \boldsymbol{\Phi}_f \boldsymbol{\delta m} \,. \tag{3}$$

The expression for $\boldsymbol{\Phi}_f$ is obtained with the help of *secondary or adjoint sources* (Tarantola, 1984). These can be understood as pseudo-sources that give rise to the wavefield residuals due to perturbations in the model parameters.

The acoustic wave-equation used to describe wavefield propagation can be represented as

$$\mathcal{L}\left[p(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] = \delta(\boldsymbol{x} - \boldsymbol{x}_s) S(t) \,, \tag{4}$$

with the acoustic full wave differential operator, also simply referred to as full wave operator, given by

$$\mathcal{L} = \left\{ \frac{1}{K(\boldsymbol{x})} \frac{\partial^2}{\partial t^2} - \nabla \cdot \left(\frac{1}{\rho(\boldsymbol{x})} \nabla \right) \right\} \,, \tag{5}$$

where K is the bulk modulus and ρ is the density.

The secondary sources are obtained when model parameter perturbations δK and $\delta \rho$ are introduced in the operator $\mathcal{L}$. Retaining only the first-order contribution yields (Tarantola, 1984, see appendix)

$$\mathcal{L}\left[\delta p(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] = -\delta \mathcal{L}\left[p(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] , \tag{6}$$

where the full secondary potential, $\delta\mathcal{L}$, is defined as

$$\delta\mathcal{L} = -\left\{\frac{\delta K(\boldsymbol{x})}{K^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{\delta\rho(\boldsymbol{x})}{\rho^2(\boldsymbol{x})}\nabla\right)\right\}. \tag{7}$$

The right-hand side of equation (6) is the secondary source, which can be forward propagated with the help of the Green's function of (4), $G(\boldsymbol{x},t;\boldsymbol{x}_s)$, to yield the wavefield perturbation

$$\delta p(\boldsymbol{x},t;\boldsymbol{x}_s) = -\int_{\mathbb{V}} d^3\boldsymbol{x}'\, G(\boldsymbol{x},t;\boldsymbol{x}') * \delta\mathcal{L}\left[p(\boldsymbol{x}',t;\boldsymbol{x}_s)\right]. \tag{8}$$

Here, the symbol $*$ denotes time-convolution.

The discretized version of equation (8) can be written in the frequency domain as

$$\widehat{\delta p} = \boldsymbol{\Phi}_f\,\boldsymbol{\delta m} = \left[\begin{array}{cc} \boldsymbol{U}_f & \boldsymbol{V}_f \end{array}\right]\left[\begin{array}{c} \boldsymbol{\delta K} \\ \boldsymbol{\delta\rho} \end{array}\right]. \tag{9}$$

which defines the bulk modulus and density Frèchet derivatives, $\boldsymbol{U}_f$ and $\boldsymbol{V}_f$, respectively.

For simplicity, let us assume constant density, i.e., $\delta\rho = 0\,\forall\boldsymbol{x}$. Using the reciprocity propriety of the Green's function, equation (8) becomes

$$\delta p(\boldsymbol{x}_g,t;\boldsymbol{x}_s) = \int_{\mathbb{V}} d^3\boldsymbol{x}'\left[\frac{1}{K^2(\boldsymbol{x}')}\,G(\boldsymbol{x}',t;\boldsymbol{x}_g) * \frac{\partial^2 p}{\partial t^2}(\boldsymbol{x}',t;\boldsymbol{x}_s)\right]\delta K(\boldsymbol{x}'). \tag{10}$$

Here, we recognize the expression in brackets as the bulk modulus Frèchet derivative for the source-receiver pair $(\boldsymbol{x}_s,\boldsymbol{x}_g)$. A similar procedure provides the corresponding expression for the density Frèchet derivative.

From equation (9), the estimate for the bulk modulus and density perturbations $\boldsymbol{\delta K}^{\text{est}}$ and $\boldsymbol{\delta\rho}^{\text{est}}$ are obtained from the wavefield residual $\widehat{\delta p}$ by

$$\left[\begin{array}{c} \boldsymbol{\delta K}^{\text{est}} \\ \boldsymbol{\delta\rho}^{\text{est}} \end{array}\right] = \boldsymbol{\delta m}^{\boldsymbol{k}} = \boldsymbol{\Phi}_f^{\dagger}\,\widehat{\delta p} = \left[\begin{array}{c} \boldsymbol{U}_f^{\dagger} \\ \boldsymbol{V}_f^{\dagger} \end{array}\right]\widehat{\delta p}\,, \tag{11}$$

where the superscript † denotes the adjoint matrix.

WAVEFIELD AND SENSITIVITY KERNEL DECOMPOSITIONS

In scattering theory, the full wavefield p is interpreted as being composed by two components: a reference wavefield, p_0, and a scattered wavefield, p_s, i.e., $p = p_0 + p_s$. The former is generated by the source that gives rise to the full wavefield and propagates in a smooth background medium that contains only low frequency information, i.e., no sharp contrasts, reflectors and/or diffractors. The latter is generated when the reference wavefield hits the singular part of the medium model which contains the high frequency information. It propagates in the complete medium containing all inhomogeneities. Of course, other interpretations of p_0 and p_s are also possible without changing the theory below.

Mathematically, the medium decomposition and wavefield separation lead to a set of differential equations,

$$\mathcal{L}^B\left[p_0(\boldsymbol{x},t;\boldsymbol{x}_s)\right] = \delta(\boldsymbol{x}-\boldsymbol{x}_s)S(t), \tag{12}$$

$$\mathcal{L}\left[p_s(\boldsymbol{x},t;\boldsymbol{x}_s)\right] = -\mathcal{V}\left[p_0(\boldsymbol{x},t;\boldsymbol{x}_s)\right], \tag{13}$$

with $\mathcal{V} = \mathcal{L} - \mathcal{L}^B$ and

$$\mathcal{L}^B = \left\{\frac{1}{K_B(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho_B(\boldsymbol{x})}\nabla\right)\right\}, \tag{14}$$

$$\mathcal{L} = \left\{\frac{1}{K_B(\boldsymbol{x})+K_S(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho_B(\boldsymbol{x})+\rho_S(\boldsymbol{x})}\nabla\right)\right\}, \tag{15}$$

where K_B and ρ_B denote the parameters of the background model, K_S and ρ_S represent the singular part, i.e., $K = K_B + K_S$ and $\rho = \rho_B + \rho_S$. Also, operator $\mathcal{V}$ is known as the scattering potential, and $\mathcal{L}^B$ is called the smooth wave operator.

Classically, it is assumed that the singular parts of the medium can be considered part of small perturbations to the smooth background (Tarantola, 1984). Then, the problem can be linearized as discussed above. However, sometimes first approximations to both parts are already available. Then, a more appropriate procedure needs to invert for perturbations of both the smooth *and* the singular part of the model.

This begs the question: Once perturbations are introduced either in the smooth and/or in the singular part of the model, what will be the reference and scattered wavefield residuals? In other words, what are the sensitivity kernels for these wavefields? To answer this question, we must look at the expression for secondary sources, to be decomposed into one for each wavefield.

Secondary sources for the reference wavefield residual

The differential equation that describes the reference wavefield propagation has the same functional form of that for the full wavefield, equation (4). Therefore, the equation for the corresponding residual has the form of equation (6), i.e.,

$$\mathcal{L}^B \left[\delta p_0(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] = -\delta\mathcal{L}^B \left[p_0(\boldsymbol{x}, t; \boldsymbol{x}_s)\right], \tag{16}$$

where the smooth secondary potential, $\delta\mathcal{L}^B$, is defined as

$$\delta\mathcal{L}^B = -\left\{\frac{\delta K_B(\boldsymbol{x})}{K_B^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{\delta\rho_B(\boldsymbol{x})}{\rho_B^2(\boldsymbol{x})}\nabla\right)\right\}. \tag{17}$$

Here, δK_B and $\delta\rho_B$ are the perturbations of the medium parameters for the smooth background medium.

With the help of the Green's function G_0 of equation (12), we can evaluate the reference wavefield residual by

$$\delta p_0(\boldsymbol{x}, t; \boldsymbol{x}_s) = -\int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_0(\boldsymbol{x}, t; \boldsymbol{x}') * \delta\mathcal{L}^B \left[p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right]. \tag{18}$$

Correspondingly to equation (8), also expression (18) can be discretized in the frequency domain as

$$\widehat{\delta p}_0 = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{V} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta K}_B \\ \boldsymbol{\delta\rho}_B \\ \boldsymbol{\delta K}_S \\ \boldsymbol{\delta\rho}_S \end{bmatrix}. \tag{19}$$

Here $\boldsymbol{U}$ e $\boldsymbol{V}$ are the sensitivity kernels of the reference wavefield respect to the smooth parts of the bulk modulus and density. In equation (19), we have already made explicit that $\widehat{\delta p}_0$ does not depend on the perturbations of the singular part of the model, δK_S and $\delta\rho_S$.

For an explicit expression of the sensitivity kernel, we again assume for simplicity that $\delta\rho_B = 0\ \forall\boldsymbol{x}$. Again using the Green's function reciprocity, equation (18) becomes

$$\delta p_0(\boldsymbol{x}_g, t; \boldsymbol{x}_s) = \int_{\mathbb{V}} d^3\boldsymbol{x}' \left[\frac{1}{K_B^2(\boldsymbol{x}')}\, G_0(\boldsymbol{x}', t; \boldsymbol{x}_g) * \frac{\partial^2 p_0}{\partial t^2}(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \delta K_B(\boldsymbol{x}'), \tag{20}$$

where the expression in brackets is the sensitivity kernel with respect to the smooth part of the bulk modulus for the source-receiver pair $(\boldsymbol{x}_s, \boldsymbol{x}_g)$. In other words, it corresponds to $\boldsymbol{U}$ in equation (19).

Secondary sources for the scattered wavefield residual

To obtain the sensitivity kernels of the scattered wavefield, we introduce the model perturbations δK_B, δK_S, $\delta\rho_B$, and $\delta\rho_S$ in equation (13). As shown in Appendix A, this leads to

$$\begin{aligned} \mathcal{L}\left[\delta p_s(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] &= \Delta s(\boldsymbol{x}, t; \boldsymbol{x}_s) \\ &= \Delta s_1(\boldsymbol{x}, t; \boldsymbol{x}_s) + \Delta s_2(\boldsymbol{x}, t; \boldsymbol{x}_s) + \Delta s_3(\boldsymbol{x}, t; \boldsymbol{x}_s) + \Delta s_4(\boldsymbol{x}, t; \boldsymbol{x}_s) \\ &= -\mathcal{V}\left[\delta p_0(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] - \delta\mathcal{L}\left[p_0(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] \\ &\quad -\delta\mathcal{L}\left[p_s(\boldsymbol{x}, t; \boldsymbol{x}_s)\right] + \delta\mathcal{L}^B\left[p_0(\boldsymbol{x}, t; \boldsymbol{x}_s)\right]. \end{aligned} \tag{21}$$

This equation allows us to evaluate the scattered wavefield residual, δp_s, resulting in

$$\delta p_s(\boldsymbol{x}, t; \boldsymbol{x}_s) = \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G(\boldsymbol{x}, t; \boldsymbol{x}') * \Delta s(\boldsymbol{x}', t; \boldsymbol{x}_s)\ , \tag{22}$$

where $G(\boldsymbol{x}, t; \boldsymbol{x}')$ is the Green's function of the full wave operator $\mathcal{L}$.

If we consider the Green's function G as composed by reference and scattered wavefields, i.e. $G = G_0 + G_s$, both obeying the set of equations (12) and (13) with an impulsive point source, then expression (22) can be decomposed into 8 terms (numbered from left to right, top to bottom):

$$\begin{aligned}
\delta p_s(\boldsymbol{x}_g, t; \boldsymbol{x}_s) = \sum_{i=1}^{n=8} \delta p_{s,i}(\boldsymbol{x}_g, t; \boldsymbol{x}_s) = \\
- \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \mathcal{V}\left[\delta p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \ - \ \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_0(\boldsymbol{x}', t; \boldsymbol{x}_g) * \mathcal{V}\left[\delta p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \\
- \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \delta\mathcal{L}\left[p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \ - \ \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_0(\boldsymbol{x}', t; \boldsymbol{x}_g) * \delta\mathcal{L}\left[p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \\
- \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \delta\mathcal{L}\left[p_s(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \ - \ \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_0(\boldsymbol{x}', t; \boldsymbol{x}_g) * \delta\mathcal{L}\left[p_s(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \\
+ \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \delta\mathcal{L}^B\left[p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right] \ + \ \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_0(\boldsymbol{x}', t; \boldsymbol{x}_g) * \delta\mathcal{L}^B\left[p_0(\boldsymbol{x}', t; \boldsymbol{x}_s)\right]\ ,
\end{aligned} \tag{23}$$

where Green's function reciprocity was used once again.

In discretized form, equation (23) reads

$$\widehat{\delta p}_s = \left[\ \boldsymbol{U}_B \quad \boldsymbol{V}_B \quad \boldsymbol{U}_S \quad \boldsymbol{V}_S\ \right] \begin{bmatrix} \boldsymbol{\delta K}_B \\ \boldsymbol{\delta \rho}_B \\ \boldsymbol{\delta K}_S \\ \boldsymbol{\delta \rho}_S \end{bmatrix}, \tag{24}$$

where each of the four terms has two contributions, one from G_0 and one from G_s. Here, $\boldsymbol{U}_B$ e $\boldsymbol{V}_B$ are the sensitivity kernels of the scattered wavefield respect to the smooth parts of the bulk modulus and density, and $\boldsymbol{U}_S$ e $\boldsymbol{V}_S$ are the sensitivity kernels of the scattered wavefield respect to the singular parts of the bulk modulus and density, respectively.

In order to look through the terms of (23) and understand them, it is important to notice that all of them have the same functional form

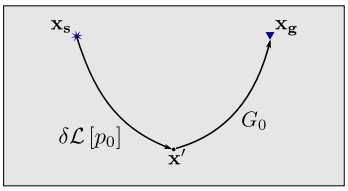
$$\delta p_{s,i}(\boldsymbol{x}_g, t; \boldsymbol{x}_s) = \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_r(\boldsymbol{x}', t; \boldsymbol{x}_g) * \mathcal{P}\left[P_s(\boldsymbol{x}', t; \boldsymbol{x}_s)\right], \tag{25}$$

where P_{sou} represents a source-related wavefield, G_r a receiver-related Green's function that acts like a propagator, and $\mathcal{P}$ a scattering potential. At each point $\boldsymbol{x}'$, one of the four secondary sources is activated through the application of $\mathcal{P}$ to P_s. Each secondary source term is propagated either by G_s – odd terms in equation (23) – or G_0 – even terms.

First of all it is important to note that the eighth term of equation (23) is the negative of the reference wavefield residual, equation (18). This leads to

$$\begin{aligned}
\delta p_s(\boldsymbol{x}_g, t; \boldsymbol{x}_s) - \delta p_{s,8}(\boldsymbol{x}_g, t; \boldsymbol{x}_s) &= \delta p_s(\boldsymbol{x}_g, t; \boldsymbol{x}_s) + \delta p_0(\boldsymbol{x}_g, t; \boldsymbol{x}_s) \\
&= \delta p(\boldsymbol{x}_g, t; \boldsymbol{x}_s) = \sum_{i=1}^{n=7} \delta p_{s,i}(\boldsymbol{x}_g, t; \boldsymbol{x}_s).
\end{aligned} \tag{26}$$

This shows us that this split into separate terms does not add new wavefield contributions as compared to the classical form for the full wavefield residual given by expression (8). In other words, the sum of the terms above yields the same total Frèchet derivative as the conventional formulation. However, by using this approach we have successfully decomposed the sensitivity kernels for the full wavefield into a set of

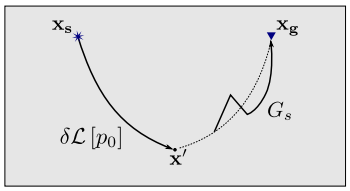


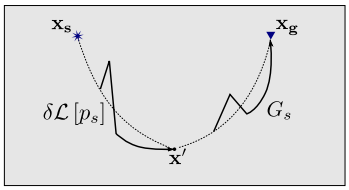
Single scattering

$$\delta p_{s,4}(\mathbf{x_g}, t; \mathbf{x_s}) = -\int_V d^3\mathbf{x}'\, G_0(\mathbf{x}', t; \mathbf{x_g}) * \delta\mathcal{L}\,[p_0(\mathbf{x}', t; \mathbf{x_s})]$$

Multiple scattering

$$\delta p_{s,3}(\mathbf{x_g}, t; \mathbf{x_s}) = -\int_V d^3\mathbf{x}'\, G_s(\mathbf{x}', t; \mathbf{x_g}) * \delta\mathcal{L}\,[p_0(\mathbf{x}', t; \mathbf{x_s})]$$





Strong multiple scattering

$$\delta p_{s,5}(\mathbf{x_g}, t; \mathbf{x_s}) = -\int_V d^3\mathbf{x}'\, G_s(\mathbf{x}', t; \mathbf{x_g}) * \delta\mathcal{L}\,[p_s(\mathbf{x}', t; \mathbf{x_s})]$$

Figure 1: Three different levels of data nonlinearity made explicit by our decomposition (23).

subkernels. With this decomposition we uncover the different levels of nonlinearity that else are hidden in the sensitivity kernel for the full wavefield (see Figure 1).

Each term from $\delta p_{s,3}$ to $\delta p_{s,6}$ can be further decomposed, because the secondary potential $\delta\mathcal{L}$ depends on perturbations of both the smooth and singular parts of the model. For example, $\delta p_{s,3}$ can be written as

$$\begin{aligned}\delta p_{s,3}(\boldsymbol{x}_g, t; \boldsymbol{x}_s) \;=\; & \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \left\{ \frac{\delta K_B(\boldsymbol{x}')}{K(\boldsymbol{x}')^2} \frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{\delta\rho_B(\boldsymbol{x}')}{\rho(\boldsymbol{x}')^2}\nabla \right) \right\} p_0(\boldsymbol{x}', t; \boldsymbol{x}_s) \\ & - \int_{\mathbb{V}} d^3\boldsymbol{x}'\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \left\{ \frac{\delta K_S(\boldsymbol{x}')}{K(\boldsymbol{x}')^2} \frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{\delta\rho_S(\boldsymbol{x}')}{\rho(\boldsymbol{x}')^2}\nabla \right) \right\} p_0(\boldsymbol{x}', t; \boldsymbol{x}_s), \quad (27)\end{aligned}$$

where, as before, $K = K_B + K_S$ and $\rho = \rho_B + \rho_S$. For constant density, i.e., $\delta\rho_B = 0$ and $\delta\rho_S = 0\,\forall\boldsymbol{x}$, this yields explicitly

$$\begin{aligned}\delta p_{s,3}(\boldsymbol{x}_g, t; \boldsymbol{x}_s) \;=\; & -\left\{ \int_{\mathbb{V}} d^3\boldsymbol{x}' \left[\frac{1}{(K_B(\boldsymbol{x}') + K_S(\boldsymbol{x}'))^2}\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \frac{\partial^2 p_0}{\partial t^2}(\boldsymbol{x}', t; \boldsymbol{x}_s) \right] \delta K_B(\boldsymbol{x}') \right. \\ & \left. + \int_{\mathbb{V}} d^3\boldsymbol{x}' \left[\frac{1}{(K_B(\boldsymbol{x}') + K_S(\boldsymbol{x}'))^2}\, G_s(\boldsymbol{x}', t; \boldsymbol{x}_g) * \frac{\partial^2 p_0}{\partial t^2}(\boldsymbol{x}', t; \boldsymbol{x}_s) \right] \delta K_S(\boldsymbol{x}') \right\}. \quad (28)\end{aligned}$$

The expression between the first brackets is the contribution $\boldsymbol{U}_B$ to the sensitivity kernel of the scattered wavefield residual with respect to the smooth background of the bulk modulus. The one between the second brackets is the corresponding contribution $\boldsymbol{U}_S$ with respect to the singular part.

The same decomposition can be done for the fourth, fifth and sixth terms of equation (23). Therefore, for each parameter of the full model – bulk modulus and density – we end up with thirteen subkernels composing the full wavefield sensitivity kernel:

(a) one subkernel to evaluate the reference wavefield residual which depends only on perturbations of the smooth part of the model;

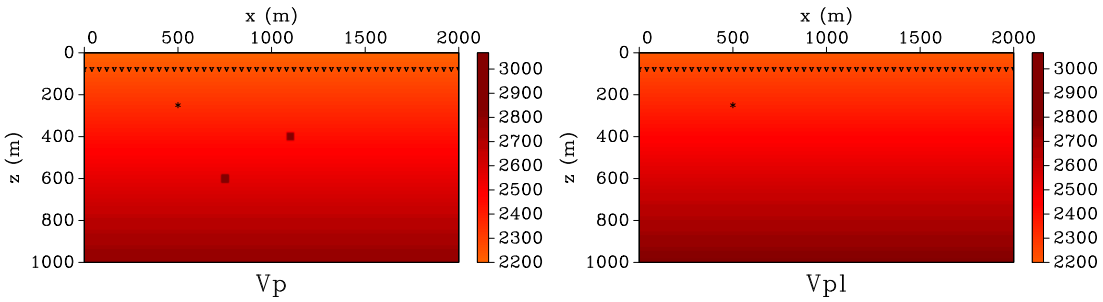
(b) twelve subkernels to evaluate the scattered wavefield residual, among which

  i. eight depend on perturbations of the smooth part of the model;

  ii. four depend on perturbations of the singular part.

Mathematically, we have from equations (19) and (24), under consideration of $\delta\rho_B = 0$ and $\delta\rho_S = 0\,\forall\boldsymbol{x}$,

$$\delta p_0 = \boldsymbol{U}\,\boldsymbol{\delta K}_B, \quad \text{and} \quad \delta p_s = \boldsymbol{U}_B\,\boldsymbol{\delta K}_B + \boldsymbol{U}_S\,\boldsymbol{\delta K}_S = \left(\sum_{i=1}^{n=8} \boldsymbol{U}_{B,i} \right) \boldsymbol{\delta K}_B + \left(\sum_{i=3}^{n=6} \boldsymbol{U}_{S,i} \right) \boldsymbol{\delta K}_S\,. \quad (29)$$



(a) Full velocity model **(b)** Smooth velocity model

Figure 2: Unperturbed P velocity model used for all numerical examples. Also shown are the source position (black dot), receiver positions (black triangles) and the two scatterers embedded in the model (dark red squares).

NUMERICAL EXAMPLES

In order to validate our results we performed numerical experiments using the open-source software package *Madagascar*. All numerical experiments followed these basic steps:

1. An unperturbed model was defined, and the full and reference wavefields were modelled (the scattered wavefield was evaluated by its definition, $p_s = p - p_0$).
2. A perturbation was introduced in the model and the perturbed wavefields were modelled and the true residuals evaluated.
3. From the model perturbation, the linearized residuals of the full, reference and scattered were evaluated and compared to the true ones.

In all models, we introduced no perturbation in the density so we can isolate the influence of the bulk modulus. Because the P velocity is more sensitive to perturbations in bulk modulus than in density, we perturbed the P velocity model and obtain the bulk modulus perturbation from the relation $v_p = \sqrt{K/\rho}$.

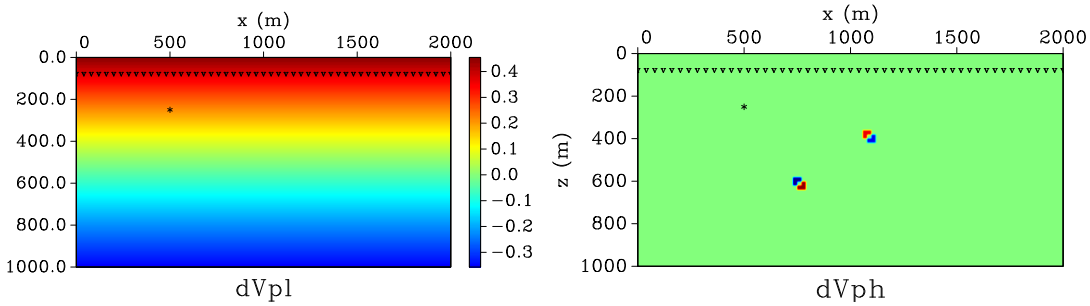
Figure 2 shows the unperturbed P-velocity model used in all numerical experiments. We chose a very simple model to facilitate a more intuitive interpretation of the results. The smooth part of the model, seen in Figure 2(b) has a near-surface velocity of 2200 m/s and a vertical gradient of 0.6/s. The singular part consists of two 36 m wide square scatterers, (see Figure 2(a)). Finite-difference modelling was used to simulated the wavefields from a 30 Hz Ricker point pressure source placed at (500 m, 250 m). A horizontal receiver array was placed at depth $z = 80$ m with a spacing of 4 m, the same as the horizontal grid spacing. The vertical grid spacing is 2 m and the time spacing used in the modelling process was 0.4 ms. The data were then resampled to an interval of 2 ms.

Numerical experiment #1

In the first experiment we introduced perturbations in both the smooth and singular parts of the model. We slightly modified the smooth background model (see Figure 3(a), scale in %) and displaced the diffraction points (see Figure 3(b)). The true wavefield residual due to this model perturbations, recorded at a control receiver at $x = 1008$ m we refer to as 'crN', is shown in Figure 4. It is calculated as the difference between the modelled wavefields in the perturbed and unperturbed models.

Figure 5 compares the true residuals of both reference and scattered wavefields (red lines) to the ones obtained from synthesising equation (23) (blue lines). Phase and waveform are well matchd. Due to a problem with the modeller, the synthesised residual differs in amplitude (see scale to the right) from the true residual (scale to the left).

Another way to verify the decomposition validity is to compare the sum of the reference and scattered residuals with the full wavefield residual evaluated by Tarantola's expression (10). This is done in Figure 6. The sum of the reference and scattered residuals (blue and red lines) completely covers the full solution (invisible black line). The magenta line shows the difference between the residuals. This proves that the scaling factor between the true and evaluated residuals is not a consequence of the decomposition.



(a) Velocity model perturbation (background) **(b)** Velocity model perturbation (singular part)

Figure 3: Perturbed P velocity model used in numerical experiment #1.

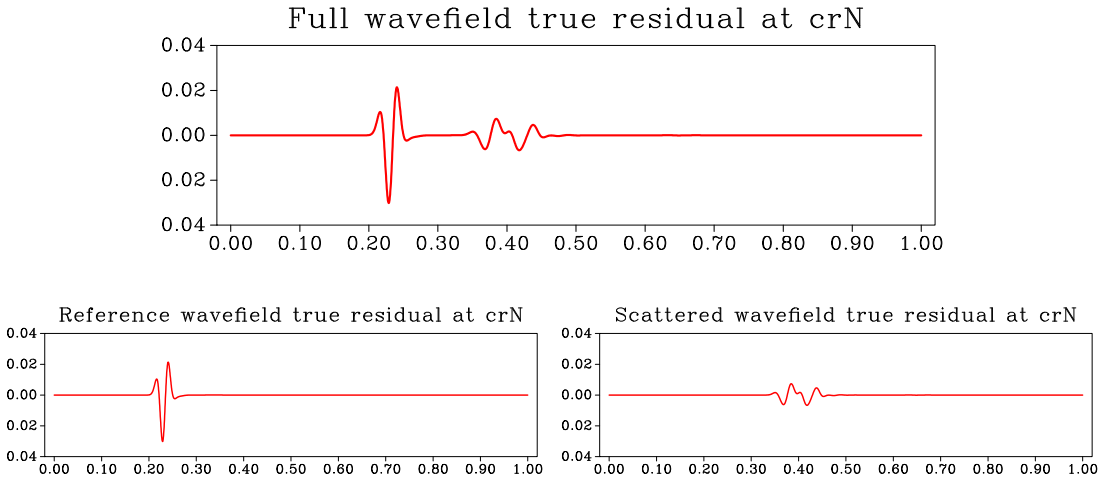


Figure 4: EXPERIMENT #1. True wavefield residual recorded at the control receiver crN ($x = 1008$ m) and its decomposition into its reference and scattered parts.

Finally, in Figure 7 we can see the individual contributions of each one of the terms of equation (23). First of all, we can see that, as expected, some terms are much more important than others. In red we highlight the most important contributions (4th and 8th term). It is worthwhile to notice that these correspond to the single scattering terms according to the classification in Figure 1. The contributions depicted in blue (3rd and 6th term) are the next smaller ones. They correspond to the multiple scattering terms. At last, the contributions depicted in magenta (1st, 2nd, 5th, and 7th) show almost no contributions. They correspond to the strong multiple scattering terms.

These results must not encourage the idea that the terms related to multiple and/or strong multiple scattering are not important. They only show that in this model, these phenomena are not relevant. In other numerical tests not included here, their relative importance was stronger.

Numerical experiment #2

In the second experiment, we perturbed only the singular part of the model (see Figure 8). In this experiment, the residuals were recorded at a different control receiver at $x = 1440$ m, which we refer to as 'crT'. The true residual for this model perturbation is depicted in Figure 9. As expected, only the scattered wavefield residual is nonzero.

Figure 10 shows the comparison between the true and synthesised residuals of both reference and scattered wavefields. Again, the wavefield match is very good except for a constant factor (see scales at left and right). In Figure 11 we compare the reference and scattered residuals with the full wavefield residual

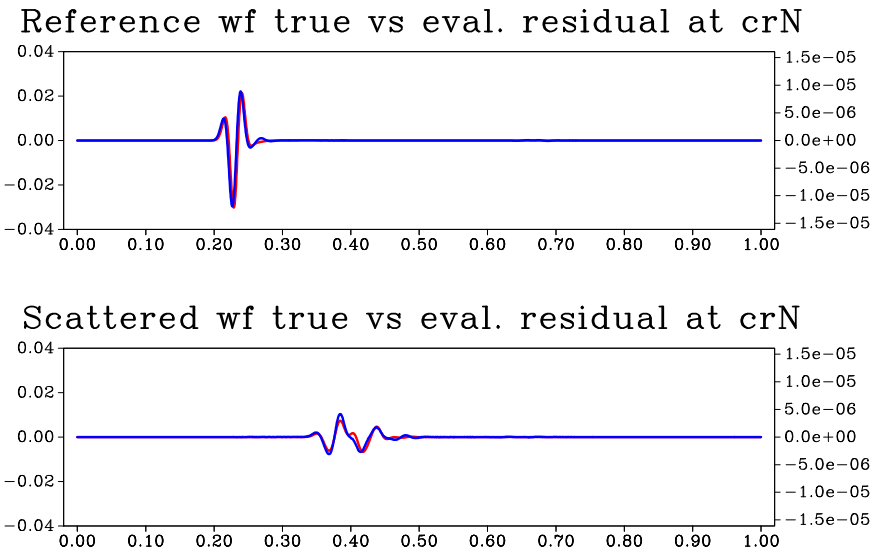


Figure 5: EXPERIMENT #1. Comparison between true (red line) and evaluated (blue line) residuals of the reference and scattered wavefields at the control receiver crN ($x = 1008$ m).

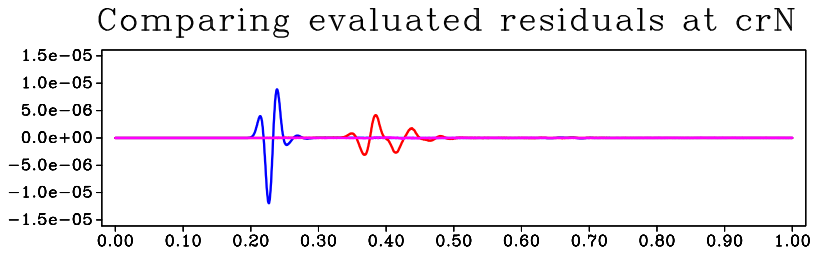


Figure 6: EXPERIMENT #1. The blue and red lines show the reference and scattered residuals, respectively, evaluated at crN. The magenta line shows the difference between the full wavefield residual and the sum of the reference and scattered residuals.

evaluated by Tarantola's expression (10). Again the magenta line shows that they match perfectly within the range of some numerical fluctuations. Finally, Figure 12, colorcoded as before, compares the importance of the individual contributions of equation (23). The conclusions are the same as for experiment #1.

BACKPROJECTING THE RESIDUALS INTO MODEL SPACE

Let us now briefly discuss how the wavefield residuals are used for model updating. This is achieved by backprojecting the residuals into the model space. For this purpose, we need the adjoint sensitivity kernels.

For simplicity, we consider only perturbations in the singular part of the bulk modulus. This corresponds to the situation in experiment #2. The other terms can be calculated analogously. From equations (29), we obtain

$$\delta p_0 = 0 \quad \text{and} \quad \delta p_s = \boldsymbol{U}_S\, \boldsymbol{\delta K}_S = \left(\sum_{i=3}^{n=6} \boldsymbol{U}_{S,i} \right) \boldsymbol{\delta K}_S. \tag{30}$$

Approximately inverting equation (30) we find the estimate of the model perturbations by backprojecting the (scattered) wavefield residual with the help of the adjoint operator $\boldsymbol{U}_S^\dagger$:

$$\boldsymbol{\delta K}_S^{\text{est}} = \boldsymbol{U}_S^\dagger\, \delta p_s = \left(\sum_{i=3}^{n=6} \boldsymbol{U}_{S,i}^\dagger \right) \delta p_s = \sum_{i=3}^{n=6} \boldsymbol{\delta K}_S^{\text{est},i}. \tag{31}$$

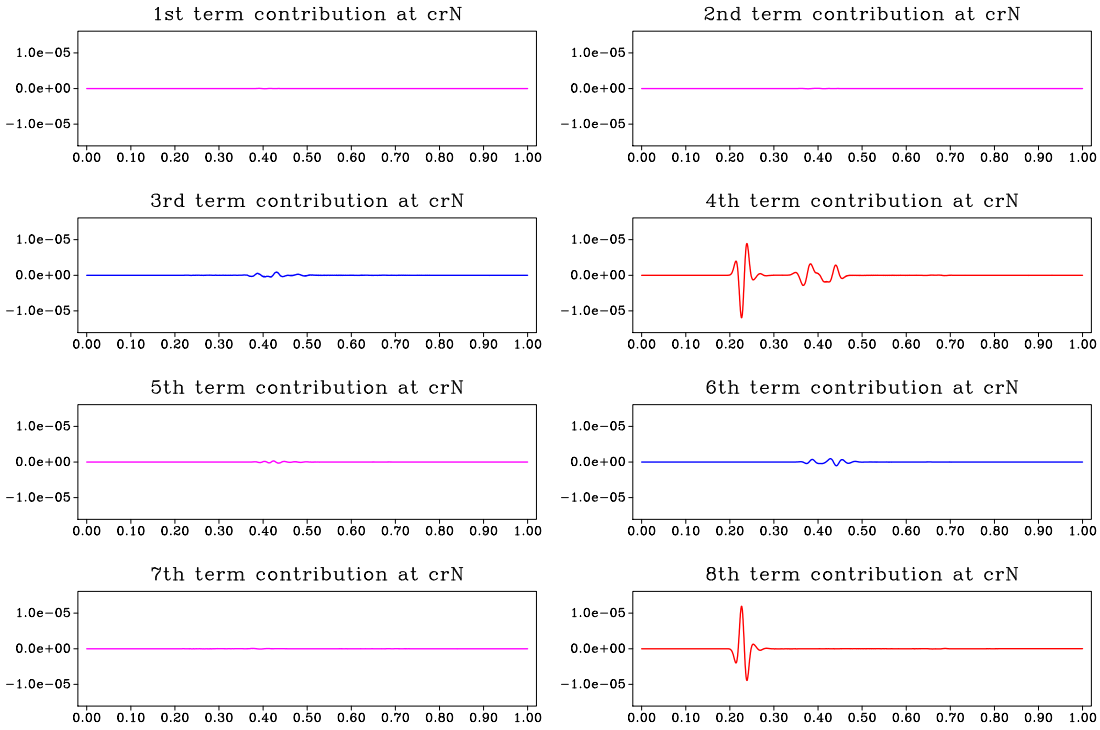


Figure 7: EXPERIMENT #1. Individual contributions of each one of the terms of (23).

How do these adjoint operators look like? From equation (23), we get for the third term

$$\delta K_S^{\mathrm{est},3}(\boldsymbol{x}_i') = -\int d\omega\, \frac{\omega^2}{K_0^2(\boldsymbol{x}_i')} \overbrace{\widehat{p}_0^{\dagger}(\boldsymbol{x}_i', \omega; \boldsymbol{x}_s)}^{\text{direct wavefield}} \underbrace{\widehat{G}_0^{\dagger}(\boldsymbol{x}_i', \omega; \boldsymbol{x}_g)\widehat{\delta p}_s(\boldsymbol{x}_g, \omega; \boldsymbol{x}_s)}_{\text{back-propagation of } \widehat{\delta p}_0}. \tag{32}$$

The other terms of (31) are obtained analogously using appropriate wavefield substitutions.

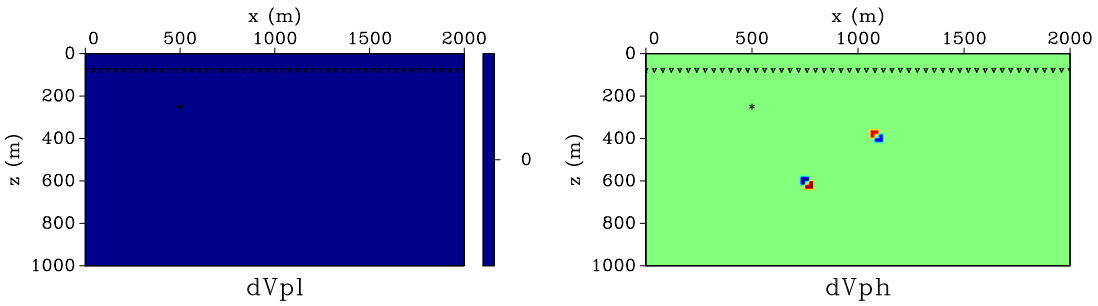
It is interesting to note that this third term is exactly the expression traditionally used in the model updating process of FWI, where only wavefields propagating in a smooth velocity model are used in both forward and back propagation steps. With the analysis above, we have shown that this actually means leaving behind three other first-order terms with respect to the model parameters.

CONCLUSIONS

In this paper we have presented a successful decomposition of standard full-waveform inversion (FWI) sensitivity kernels into several sub-kernels. For this purpose, we have developed a scattering formulation that decomposes the model parameters into a smooth background and a singular component, both of which are then perturbed. As a consequence, the single sensitivity kernel of the standard formulation of FWI is decomposed into a sum of sub-kernels.

The superposition of all our sub-kernels yields the full Frèchet gradient as in the standard formulations of FWI. The asset of the decomposition is that it allows to recognize in each individual sub-kernel the different contributions in terms of nonlinearity with respect to data components. We have demonstrated that the contributions of each sub-kernel can be physically interpreted in terms of orders of scattering with respect to the unknown model parameters. Thus, our scattering parameterization and sub-kernel approach will allow for better control of nonlinear effects at each iteration of FWI model optimization.

In addition to its potential benefits in controlling convergence of FWI routines, we point out that the underlying scattering formulation of our method allows for a direct connection to migration-type imaging and model building. Classically, it is assumed that the singular parts of the medium can be considered part



(a) Velocity model perturbation (background) **(b)** Velocity model perturbation (singular part)

Figure 8: Perturbed P velocity model used in the numerical experiment #2.

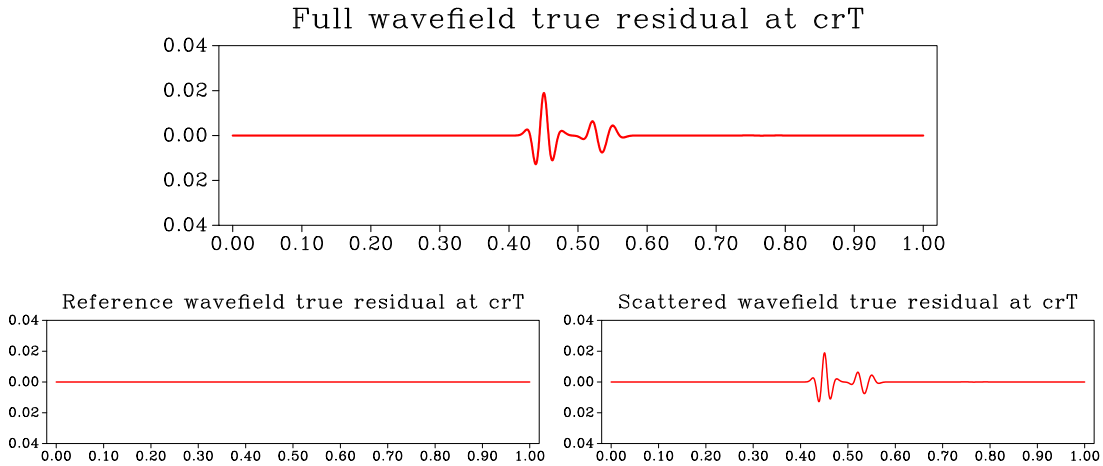


Figure 9: EXPERIMENT #2. Wavefield true residual recorded at the control receiver crT ($x = 1440$ m).

of small perturbations to the smooth background. That assumption does not hold if the initial estimate for the smooth model is not good enough or if strong contrasts are present in the model. However, a migrated seismic image can in principle be taken as an estimate for the sharp/singular part of the subsurface model, while models from velocity analysis are a proxy for the smooth model component. Under that framework, our formalism provides for an explicit method for jointly using velocity models and migrated seismic images in FWI as well as for understanding their interplay in the model-building process. Likewise, the scattering kernels presented here can be directly used in the context of extended images (EIs) in devising nonlinear wave-equation migration velocity model building techniques.

On the downside, we expect practical challenges to arise when defining how to separate/represent the smooth and sharp model components, and particularly when decomposing the corresponding data components. Possible ideas include high and lowpass filtering. So while there are potential benefits in describing and including higher-order nonlinear terms in FWI using our formulation, achieving them in practice with both synthetic and field data is the subject of ongoing research.

The next step in this research will consist of analysing the contributions of the sub-kernels in more complicated models, aiming at an evaluation of the importance of the higher-order terms as compared to the dominant (single scattering) term. Moreover, we will also analyse and exemplify the terms for background and singular density perturbations so as to study their superposition in the model updating procedure.

ACKNOWLEDGMENTS

We thank the Brazilian National Research Council CNPq as well as Petrobras, Schlumberger, and the sponsors of the *Wave Inversion Technology (WIT) Consortium* for kindly supporting this research.

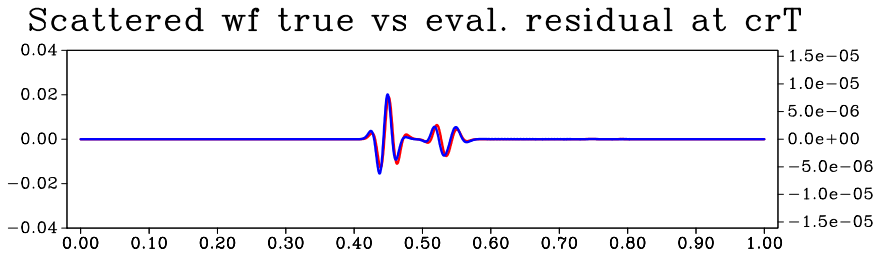


Figure 10: EXPERIMENT #2. Comparison between true (red line) and evaluated (blue line) residuals of the reference and scattered wavefields at the control receiver crT ($x = 1440$ m).

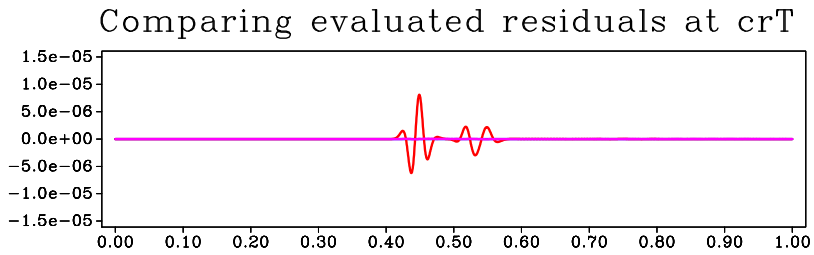


Figure 11: EXPERIMENT #2. The blue and red lines show the reference and scattered residuals evaluated at crN respectively. The magenta line shows the difference between the full wavefield residual and the sum of the reference and scattered residuals.

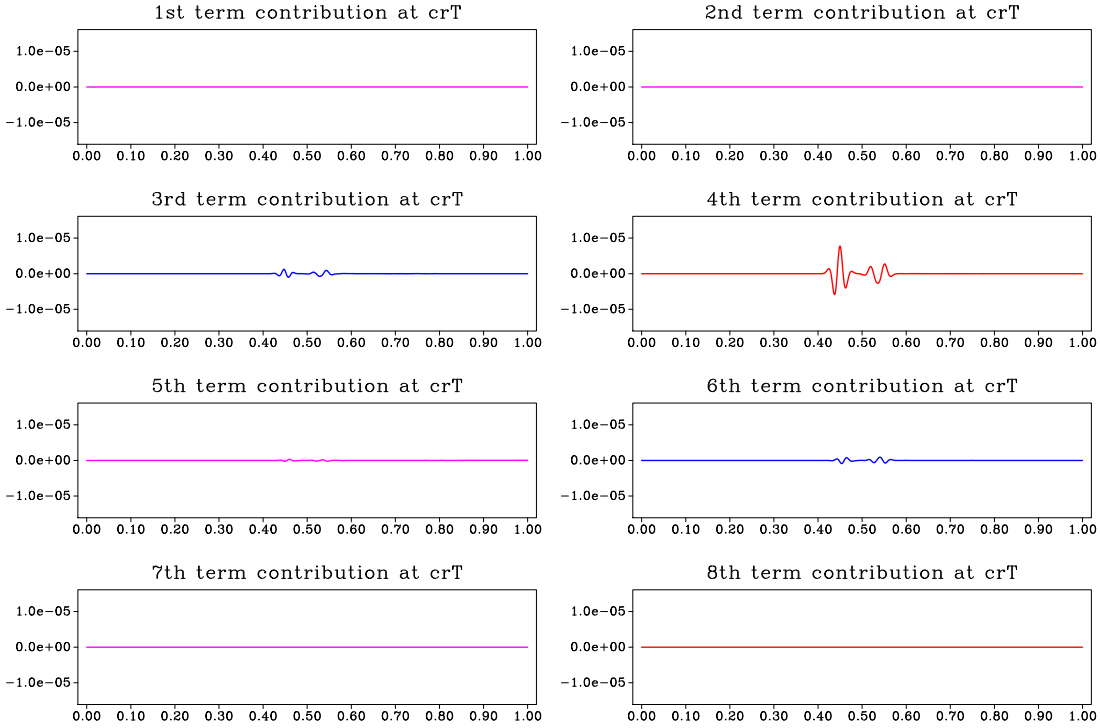


Figure 12: EXPERIMENT #2. Individual contributions of each one of the terms of (23).

REFERENCES

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, J., Robinson, E., and A.Weglein, editors, *Inverse scattering theory and application*, pages 206–220. Society for Industrial and Applied Mathematics (SIAM).

Rickett, J. E. and Sava, P. C. (2002). Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics*, 67(3):883–889.

Sava, P. and Vasconcelos, I. (2009). Extended common-image-point gathers for wave-equation migration. *EAGE Exp. Abs.*

Sava, P. and Vasconcelos, I. (2010). Extended imaging conditions for wave-equation migration. Geophys. Prosp, accepted.

Sava, P. C. and Fomel, S. (2003). Angle-domain common-image gathers by wavefield continuation methods. *Geophysics*, 68(3):1065–1074.

Symes, W. W. (2008). Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56:765–790.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51(10):1893–1903.

Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal an banana-doughnut kernels. *Geophys. J. Int.*, 160:295–216.

Vasconcelos, I. (2008). Generalized representations of perturbed fields – applications in seismic interferometry and migration. *SEG Exp. Abs.*

Vasconcelos, I., Sava, P., and Douma, H. (2009). Wave-equation extended images via image-domain interferometry. *EAGE Exp. Abs.*

Vasconcelos, I., Sava, P., and Douma, H. (2010). Nonlinear extended images via image-domain interferometry. *Geophysics*, 75(6):SA105–SA115.

Vigh, D., Starr, E. W., and Kapoor, J. (2009). Developing earth models with full waveform inversion. *The Leading Edge*, 28(4):432–435.

Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

APPENDIX A

PERTURBATION OF BACKGROUND AND SINGULAR PART

On perturbation of the model parameters in the wave equation (13), we obtain

$$\tilde{\mathcal{L}}\ \left[p_s(\boldsymbol{x},t;\boldsymbol{x}_s)+\delta p_s(\boldsymbol{x},t;\boldsymbol{x}_s)\right] = -\tilde{\mathcal{V}}\left[p_0(\boldsymbol{x},t;\boldsymbol{x}_s)+\delta p_0(\boldsymbol{x},t;\boldsymbol{x}_s)\right], \tag{33}$$

with $\tilde{\mathcal{V}} = \tilde{\mathcal{L}} - \tilde{\mathcal{L}}^B$ and

$$\tilde{\mathcal{L}}^B = \left\{\frac{1}{K_B(\boldsymbol{x})+\delta K_B(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho_B(\boldsymbol{x})+\delta\rho_B(\boldsymbol{x})}\nabla\right)\right\}; \tag{34}$$

$$\tilde{\mathcal{L}} = \left\{\frac{1}{[K_B(\boldsymbol{x})+\delta K_B(\boldsymbol{x})]+[K_S(\boldsymbol{x})+\delta K_S(\boldsymbol{x})]}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{[\rho_B(\boldsymbol{x})+\delta\rho_B(\boldsymbol{x})]+[\rho_S(\boldsymbol{x})+\delta\rho_S(\boldsymbol{x})]}\nabla\right)\right\}; \tag{35}$$

where the tilde denotes a perturbed potential and/or operator. Use of

$$\frac{1}{(h+\delta h)+(g+\delta g)} = \frac{1}{h+g} - \frac{\delta h}{(h+g)^2} - \frac{\delta g}{(h+g)^2} + O(\delta h \delta g, \delta h^2, \delta g^2), \tag{36}$$

in equation (35) yields

$$\tilde{\mathcal{L}} \approx \bar{\mathcal{L}} = \left\{ \left(\frac{1}{K_B+K_S} - \frac{\delta K_B}{(K_B+K_S)^2} - \frac{\delta K_S}{(K_B+K_S)^2} \right) \frac{\partial^2}{\partial t^2} - \nabla \cdot \left[\left(\frac{1}{\rho_B+\rho_S} - \frac{\delta \rho_B}{(\rho_B+\rho_S)^2} - \frac{\delta \rho_S}{(\rho_B+\rho_S)^2} \right) \nabla \right] \right\}, \tag{37}$$

if the cross and higher-order terms are neglected. Analogously, using

$$\frac{1}{h+\delta h} = \frac{1}{h} - \frac{\delta h}{h^2} + O(\delta h^2), \tag{38}$$

in equation (34) we get

$$\tilde{\mathcal{L}}^B \approx \bar{\mathcal{L}}^B = \left\{ \left(\frac{1}{K_B} - \frac{\delta K_B}{K_B^2} \right) \frac{\partial^2}{\partial t^2} - \nabla \cdot \left[\left(\frac{1}{\rho_B} - \frac{\delta \rho_B}{\rho_B^2} \right) \nabla \right] \right\}. \tag{39}$$

The full secondary potential is defined in equation (7). In terms of the individual perturbations of the background and singular parts, it reads

$$\delta\mathcal{L} = -\left\{ \left(\frac{\delta K_B + \delta K_S}{(K_B+K_S)^2} \right) \frac{\partial^2}{\partial t^2} - \nabla \cdot \left[\left(\frac{\delta \rho_B + \delta \rho_S}{(\rho_B+\rho_S)^2} \right) \nabla \right] \right\} . \tag{40}$$

Using this expression and equation (17), operators $\tilde{\mathcal{L}}$, $\tilde{\mathcal{L}}^B$ e $\tilde{\mathcal{V}}$ can be expressed as

$$\tilde{\mathcal{L}}^B \approx \bar{\mathcal{L}}^B = \mathcal{L}^B + \delta\mathcal{L}^B; \tag{41}$$

$$\tilde{\mathcal{L}} \approx \bar{\mathcal{L}} = \mathcal{L} + \delta\mathcal{L}; \tag{42}$$

$$\tilde{\mathcal{V}} \approx \bar{\mathcal{V}} = \mathcal{V} + (\delta\mathcal{L} - \delta\mathcal{L}^B). \tag{43}$$

With these approximations, equation (33) can be recast into the form

$$(\mathcal{L} + \delta\mathcal{L}) \left[(p_s + \delta p_s) \right] = - \left(\mathcal{V} + \delta\mathcal{L} - \delta\mathcal{L}^B \right) \left[(p_0 + \delta p_0) \right] . \tag{44}$$

Next we distribute terms in equation (44) and neglect all cross-terms between wavefield and medium perturbations. This results in

$$\mathcal{L}\left[p_s\right] + \mathcal{L}\left[\delta p_s\right] + \delta\mathcal{L}\left[p_s\right] = -\mathcal{V}\left[p_0\right] - \mathcal{V}\left[\delta p_0\right] - \delta\mathcal{L}\left[p_0\right] + \delta\mathcal{L}^B\left[p_0\right] . \tag{45}$$

Finally, taking the difference between equations (45) and (13), we obtain

$$\mathcal{L}\left[\delta p_s\right] = -\mathcal{V}\left[\delta p_0\right] - \delta\mathcal{L}\left[p_0\right] - \delta\mathcal{L}\left[p_s\right] + \delta\mathcal{L}^B\left[p_0\right] . \tag{46}$$