Two-dimensional elastic full waveform inversion using Born and Rytov formulations in the frequency domain

Céline Gélis, Jean Virieux, Gilles Grandjean

Geophys. J. Int. (2007) **168,** 605–633 doi: 10.1111/j.1365-246X.2006.03135.x

Two-dimensional elastic full waveform inversion using Born and Rytov formulations in the frequency domain

C. Gélis,[1,2] J. Virieux[1] and G. Grandjean[2]

[1] *Géosciences Azur, Nice, France*
[2] *BRGM, Orléans, France. E-mail: virieux@geoazur.unice.fr*

Accepted 2006 July 10. Received 2006 July 8; in original form 2005 August 2

SUMMARY

We perform the full elastic waveform inversion in the frequency domain in a 2-D geometry. This method allows imaging of two physical seismic parameters, using vertical and horizontal field components. The forward problem is discretized using finite difference, allowing to simulate the full elastic wavefield propagation. Moreover, it is solved in the frequency domain, a fast approach for multisource and multireceiver acquisition. The non-linear inversion is based on a pre-conditioned gradient method, where Born and Rytov formulations are used to compute Fréchet derivatives. Parameter perturbations linearly depend on fields perturbations in the Born kernel, and on the generalized complex phases of fields in the Rytov kernel, giving different Fréchet derivatives. The gradient is pre-conditioned with the diagonal part of the inverse Hessian matrix, allowing to better estimate the stepping in the optimization direction. Non-linearity is taken into account by updating parameters at each iteration and proceeding from low to high frequencies. The latter allows as well to progressively introduce smaller wavelengths in parameter images. On a very simple synthetic example, we examine the way the inversion determines the V_p (P-wave velocity) and V_s (S-wave velocity) images. We highlight that, with a transmission acquisition, final parameter images weakly depend on the chosen formulation to compute Fréchet derivatives and on the inverted parameters choice. Of course, convergence strongly depends on the medium wavenumber illumination which is related somehow to the acquisition geometry. With a reflection acquisition, the Born formulation allows to better recover scatterers. Moreover, the medium anomalies are not well reconstructed when surface waves propagate in the medium. This may be due to the evanescent nature of surface waves. By selecting first body waves and then surface waves, we improve the convergence and properly reconstruct anomalies. This shows us that preparation of the seismic data before the inversion is as critical as the initial model selection.

Key words: Born and Rytov formulations, diffraction tomography, finite difference methods, medium wavenumber illumination, seismic imaging, waveform inversion.

1 INTRODUCTION

Quantitative imaging using full wave equation has been achieved through the use of the adjoint formulation problem for seismic data in the last 20 yr. Both formulations in time domain (Lailly 1984; Tarantola 1984; Gauthier *et al.* 1986) and in frequency domain (Pratt *et al.* 1996; Pratt 1999; Ravaut *et al.* 2004) have been implemented and applied to various synthetic and real data examples with specific advantages on both sides. Easier seismic traces processing in time domain will allow progressive introduction of phases by increasing the time domain window in both observed and synthetic data (Kolb *et al.* 1986; Shipp & Singh 2002; Sheng 2004). Efficient ways of solving the forward problem in the frequency domain make the frequency formulation appealing (Stekl & Pratt 1998). Moreover, the progressive introduction of higher frequencies allows both to introduce and mitigate the non-linearity and recover shorter and shorter heterogeneities (Pratt 1999; Sirgue 2003). Furthermore, for wide-angle data acquisitions, this frequency approach efficiently takes benefit of the wavenumber redundancy by limiting the number of inverted frequencies (Pratt 1990; Sirgue & Pratt 2004). The attenuation may be introduced, which has been applied to real data examples (Hicks & Pratt 2001).

All kinds of waves are considered in the forward modelling with our approach, and it will avoid any ghost images arising in the imaging procedure from incomplete wave reconstitution as it may happen when considering asymptotic ray seismograms (Lambaré *et al.* 1992; Jin 1992;

Forgues 1996). In this asymptotic approximation, the forward problem is based on a linearization between parameters and data perturbations based on ray theory (Beylkin 1985). While retaining the same adjoint formulation for fitting waveforms, the so-called Ray+Born formulation provides a better control of the diffracting point illumination, leading to rather efficient algorithms by adapting locally the weighting used in the optimization scheme for both the acoustic case (Lambaré *et al.* 1992) and the elastic one (Jin 1992; Jin *et al.* 1992; Forgues 1996). This high-frequency approach does not require to go from low to high frequencies. Still, the capacity of Ray+Born formulation of considering high frequencies should be acknowledged, especially in 3-D geometry (Lambaré *et al.* 2003). Moreover, attenuation reconstruction has been considered by Tarantola (1988) and Ribodetti & Virieux (1998).

While time domain formulation has been applied for recovering elastic parameters using multicomponent data (Tarantola 1987; Mora 1987, 1988), frequency formulation for elastic parameters reconstruction has only been applied by Pratt (1990) for cross-hole data to our knowledge. Moreover, investigations on the misfit function definition have not been performed as for the time domain formulation where Crase *et al.* (1990) has introduced a norm which mimics a L1 norm, allowing new Fréchet derivatives to be estimated. Discussions about data and parameter sets we should consider have not yet been addressed for the frequency approach in the elastic case.

The Fréchet derivatives calculation depends on the chosen approximation to linearly link data and parameters perturbations. Born (Clayton & Stolt 1981; Beylkin 1985; Beydoun & Tarantola 1988; Beydoun & Mendes 1989; Lambaré *et al.* 1992) and Rytov (Bleistein 1987; Beydoun & Tarantola 1988; Beylkin & Burridge 1990) approximations are currently used, although they proceed differently since the data space is not the same : while the Born formulation focuses on the velocity field, the Rytov formulation is based on the velocity field complex phase. Some authors (Beydoun & Tarantola 1988) argue that the Born formulation is well suited for reflection acquisitions, whereas the Rytov formulation is more efficient with transmission acquisitions. Moreover, Woodward (1992) explained that the Rytov formulation is better adapted to the frequency domain, whereas the Born formulation has a more direct interpretation in the time domain. However, Pratt (1990), Pratt *et al.* (1998) used the Born formulation in the frequency domain while studying the full waveform inversion with acoustic waves. Pratt & Worthington (1988) applied the acoustic full waveform inversion on real data using Born and Rytov formulations. They showed that the Born formulation better reconstructs the edges of a discrete object. In order to understand and clarify the influence of the chosen formulation in the calculation of Fréchet derivatives for the elastic case, we have considered Born and Rytov formulations to calculate Fréchet derivatives.

We shall concentrate our attention on the reconstruction of elastic parameters in a 2-D medium using either single-component data or multicomponent data. The attenuation parameter will not be considered in this study although such extension is possible. The minimization of particle velocity trace residuals or complex phase trace residuals using iterative linearized algorithm will be outlined. Then, we shall consider how to solve efficiently in the frequency domain full wave equation by a finite difference scheme different from the one already used by Stekl & Pratt (1998) or by Virieux (1986). While the forward problem is fully solved, we use Born and Rytov formulations to compute Fréchet derivatives, respectively, for amplitude fitting and (complex) phase fitting. On simple synthetic examples, we shall analyse the influence of these two different misfit functions as well as the effect of different sets of elastic parameters. For each example, the complete velocity models have to be reconstructed (background and anomalies), even if we choose an initial homogeneous medium. The influence of the data acquisition geometry will be discussed for the reconstruction of different types of anomalies.

Finally, we introduce the free surface, allowing us to model the surface waves propagation in the forward problem. Using surface waves to determine the surface properties such as seismic velocities or anisotropy has already been done by several authors (Snieder 1986; Nolet 1987; Debayle & Kennett 2000a,b). The importance of the free surface will be analysed and compared with previous reconstructions, for addressing the specific contribution of both body and surface waves in the resulting images. We shall highlight the difficulty to deal with surface waves, whose amplitude exponentially decays with depth, in our inversion engine and show one possibility to efficiently take them into account.

This study will allow us to draw conclusions regarding the potential of the method for addressing challenging problems as crustal imaging, reservoir characterization, subsurface reconstruction, cavity detection and so on. We shall concentrate in this article on issues of the proposed method when elastic waves are considered. More realistic examples will be considered in future works.

2 INVERSION FORMULATION

The reconstruction of medium parameters from seismograms is a non-unique inverse problem because non-linearity is present in the relation between model and data spaces. Phase wavelets should be roughly in phase before adjustment to avoid cycle-skipping (Sirgue 2003). The general relation g between the model $\mathbf{m}$ and the data $\mathbf{d}$ can be expressed by

$$\mathbf{d} = g(\mathbf{m}). \tag{1}$$

We choose to perform a non-linear inversion (Tarantola 1984, 1987) in the least-square sense because Fréchet derivatives can be efficiently computed. We emphasize that $g(\mathbf{m})$ represents the fully solved forward problem. The misfit function E is equal to

$$E(\mathbf{m}) = \frac{1}{2}[g(\mathbf{m}) - \mathbf{d}_{\mathbf{obs}}]^{\dagger}[g(\mathbf{m}) - \mathbf{d}_{\mathbf{obs}}], \tag{2}$$

where $\mathbf{d}_{\mathbf{obs}}$ represents observed data whereas $\mathbf{d} = g(\mathbf{m})$ are the calculated data with the $\mathbf{m}$ model. The symbol † stands for the adjoint operator (complex conjugate). In this formulation, each source–receiver couple contribution is implicitely summed. Weighted least squares can be introduced to increase the importance of a specific subset as, for example, far-offset traces (Operto *et al.* 2004). The misfit function contains the sum of squares of differences between the observed and calculated data, that is, data differences or residuals $\Delta\mathbf{d}$.

First, we shall consider a data space of particle velocities $V^{\rm calc}(\mathbf{r}, \omega, \mathbf{s})$ recorded at receivers $\mathbf{r}$ for different sources $\mathbf{s}$ and for different frequencies ω. Therefore, in this case, synthetic data $g(\mathbf{m}) = [V_x^{\rm calc}(\mathbf{r}, \omega, \mathbf{s}), V_z^{\rm calc}(\mathbf{r}, \omega, \mathbf{s})]$ and observed data $\mathbf{d}_{\rm obs} = [V_x^{\rm obs}(\mathbf{r}, \omega, \mathbf{s}), V_z^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})]$ are particle velocity amplitudes. The misfit function E_b at a given frequency ω will be equal to

$$E_b(\mathbf{m}) = \frac{1}{2} \sum_{\mathbf{s}} \sum_{\mathbf{r}} [V^{\rm calc}(\mathbf{r}, \omega, \mathbf{s}) - V^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})]^{\dagger} [V^{\rm calc}(\mathbf{r}, \omega, \mathbf{s}) - V^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})]. \tag{3}$$

This misfit function will be considered as the Born misfit function because the Born formulation (Clayton & Stolt 1981; Beylkin 1985; Beydoun & Tarantola 1988; Beydoun & Mendes 1989; Lambaré *et al.* 1992) is used to compute associated Fréchet derivatives as we shall see later on.

As an alternative, we may consider the data $\mathbf{d}$ as the generalized complex phase of the particle velocity field which includes the Neperian complex logarithm *Ln* of the velocity amplitude as the real part and the phase of the velocity field as the imaginary part. In this case, particle velocity phases are used for both synthetic data $g(\mathbf{m}) = [Ln[V_x^{\rm calc}(\mathbf{r}, \omega, \mathbf{s})], Ln[V_z^{\rm calc}(\mathbf{r}, \omega, \mathbf{s})]]$ and observed data $\mathbf{d}_{\rm obs} = [Ln[V_x^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})], Ln[V_z^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})]]$. The misfit function $E_r(\mathbf{m})$ is expressed for each frequency ω as

$$E_r(\mathbf{m}) = \frac{1}{2} \sum_{\mathbf{s}} \sum_{\mathbf{r}} [Ln[V^{\rm calc}(\mathbf{r}, \omega, \mathbf{s})] - Ln[V^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})]]^{\dagger} [Ln[V^{\rm calc}(\mathbf{r}, \omega, \mathbf{s})] - Ln[V^{\rm obs}(\mathbf{r}, \omega, \mathbf{s})]]. \tag{4}$$

Fréchet derivatives will be obtained through the Rytov formulation (Bleistein 1987; Beydoun & Tarantola 1988; Beylkin & Burridge 1990). We call this misfit function the Rytov misfit function. Unwrapping phases is necessary for avoiding jumps in the phase perturbation estimation.

Because the data space is different, we may expect different behaviours of the minimization procedure for Born and Rytov formulations. Let us remind that the forward problem is solved by a purely numerical technique and that we have selected a finite difference method.

The misfit function has a rather complex shape with many minima. Once an initial model is chosen, we perform iterations to reach the neighbouring minimum and take the best model calculated at one iteration as the new initial model for the next iteration. We perform a second-order Taylor–Lagrange development around a given initial model $\mathbf{m_0}$ to get

$$E(\mathbf{m_0} + \delta\mathbf{m}) = E(\mathbf{m_0}) + \nabla_{\mathbf{m}} E(\mathbf{m_0}) \delta\mathbf{m} + 1/2 \delta\mathbf{m}^{\dagger} \mathcal{H}(\mathbf{m_0}) \delta\mathbf{m} + \mathcal{O}(\delta\mathbf{m}^2), \tag{5}$$

where the gradient is denoted $\nabla_{\boldsymbol{m}} E$, the Hessian $\mathcal{H}$, $\delta\mathbf{m}$ is a small parameters perturbation and $\mathcal{O}(\delta\mathbf{m}^2)$ is a quantity that we neglect. A local minimum of E is reached when the increment in the model $\delta\mathbf{m}$ verifies the following equation,

$$\mathcal{H}(\mathbf{m_0}) \delta\mathbf{m} = -\nabla_{\mathbf{m}} E(\mathbf{m_0}). \tag{6}$$

Let us introduce now the Fréchet derivative denoted as $\mathcal{B}_0 = (\partial g / \partial \mathbf{m})(\mathbf{m_0})$. How to estimate this matrix and the way it is involved in the gradient estimation will be discussed in a later paragraph. Taking the derivative of eq. (2) with respect to the model $\mathbf{m}$ gives the following expression

$$\nabla_{\mathbf{m}} E(\mathbf{m}) = \Re \left\{ \left(\frac{\partial g}{\partial \mathbf{m}} \right)^{\dagger} [g(\mathbf{m}) - \mathbf{d}_{\rm obs}] \right\}, \tag{7}$$

where $\Re$ indicates the real part. When $\mathbf{m} = \mathbf{m}_0$, the first term of the right-hand side is the adjoint operator of the Fréchet derivative and the second one is the data residuals $\Delta\mathbf{d}$. Therefore, the gradient vector may be expressed by the adjoint operator $\mathcal{B}_0^{\dagger}$ applied to data residuals $\Delta\mathbf{d}$ following the linearized inversion formalism (Tarantola 1987). After another derivative, the misfit function with respect to the model $\mathbf{m}$ gives the Hessian operator expression

$$\mathcal{H}(\mathbf{m}) = \left(\frac{\partial^2 g}{\partial \mathbf{m}^2} \right)^{\dagger} [g(\mathbf{m}) - \mathbf{d}_{\rm obs}] + \left(\frac{\partial g}{\partial \mathbf{m}} \right)^{\dagger} \frac{\partial g}{\partial \mathbf{m}}. \tag{8}$$

The first term is generally dropped off in non-linear problems (Pratt *et al.* 1998). We only keep the second term, which turns out to be the matrix $\mathcal{B}_0^{\dagger} \mathcal{B}_0$ in the model space (Tarantola 1987). The following normal equation,

$$\Re[\mathcal{B}_0^{\dagger} \mathcal{B}_0] \delta\mathbf{m} = \Re\left[\mathcal{B}_0^{\dagger} \Delta\mathbf{d}\right], \tag{9}$$

coming from the transformation of eq. (6), should be verified by the model increment for reducing the misfit function. The Hessian matrix is difficult to invert especially when the model space has a high number of degrees of freedom. We shall assume that this matrix is diagonally behaving for solving the system through a gradient method. Since we take its real part, the right-hand side of eq. (9) may be written as

$$\Re\left(\mathcal{B}_0^{\dagger} \Delta\mathbf{d}\right) = \Re\left(\mathcal{B}_0^{t} \Delta\mathbf{d}^{*}\right), \tag{10}$$

where $\mathcal{B}_0^{t}$ stands for the $\mathcal{B}_0$ transpose and complex conjugate residuals, corresponding to time-reversed residuals in the time domain, are denoted by $\Delta\mathbf{d}^{*}$.

Any pre-conditioning transforming the system (9) into an improved diagonal construction of the Hessian matrix will speed up the convergence towards the minimum of the misfit function (Shin *et al.* 2001b). This has been achieved, for example, in the Ray+Born method because the diagonal structure of the Hessian matrix, depending strongly on the data acquisition geometry, could be related to local ray properties (Jin *et al.* 1992).

We shall add a constant coefficient ϵ to avoid divisions by too small numbers when inverting the diagonal approximation of the Hessian matrix. When this regularization coefficient increases, the inverse diagonal Hessian provides a weaker influence, decreasing the jump between old and new models. The new model is estimated through

$$\mathbf{m}^{l+1} = \mathbf{m}^{l} - \alpha_l \left\{ \mathrm{diag}\left[\Re\left(\mathcal{B}_0^{\dagger}\mathcal{B}_0\right)\right]^{(l)} + \epsilon^l I \right\}^{-1} \Re\left[\left(\mathcal{B}_0^{t}\right)^{(l)} \Delta \mathbf{d}_l^{*}\right], \tag{11}$$

where the model $\mathbf{m}^l$ corresponds to the model $\mathbf{m}$ at the lth iteration, The number α_l is the stepping at the lth iteration. The stepping α_l is obtained by sampling a parabolic shape along the gradient direction. Because we use a line search approach as Pratt *et al.* (1996), this requires two additional forward modelling at each iteration. Please note that the matrix $\mathcal{B}_0$ is computed for each iteration l. To simplify the notations, we will not write the iteration number l in all following equations.

With Gauss–Newton or full Newton methods (Pratt *et al.* 1998), the complete Hessian matrix given by the eq. (8) has to be inverted. It contains the first term of the Hessian operator in the full Newton method, which deals with multiple diffractions. The second term of the Hessian operator allows to take into account the source band-limited frequency content and the incomplete medium illumination by the acquisition system. Nevertheless, inverting the complete Hessian matrix is time-consuming and requires significant numerical resources although not unrealistic in 2-D geometry. We have not tried this important feature investigated by Plessix & Mulder (2004). Moreover, when the Hessian matrix is not taken into account and the gradient is used alone, the gradient is usually calculated with the adjoint operator.

3 THE FORWARD FORMULATION

The forward model will be solved for each new medium in the frequency domain because it allows to efficiently consider multisources acquisitions. In this study, we consider an isotropic 2-D linear elastic medium. The propagation of 2-D P-SV waves in elastic media can be expressed thanks to the matrix formalism of Pratt *et al.* (1998) and Stekl & Pratt (1998). Using first an operator notation, the wave equation may be compactly written as

$$\mathcal{W}\mathcal{X} = \mathcal{S}, \tag{12}$$

where $\mathcal{X}^t$ is the $[V_x(\mathbf{x}, \omega)\, V_z(\mathbf{x}, \omega)]^t$ velocity field vector, $\mathcal{S}^t$ is the $[S_x(\mathbf{x}, \omega)\, S_z(\mathbf{x}, \omega)]^t$ source vector and $\mathcal{W}(\mathbf{x}, \omega)$ is the wave operator defined as

$$\mathcal{W}(\mathbf{x}, \omega) = \begin{bmatrix} \mathcal{W}_{xx}(\mathbf{x}, \omega) & \mathcal{W}_{xz}(\mathbf{x}, \omega) \\ \mathcal{W}_{zx}(\mathbf{x}, \omega) & \mathcal{W}_{zz}(\mathbf{x}, \omega) \end{bmatrix}, \tag{13}$$

where

$$\begin{aligned}
\mathcal{W}_{xx}(\mathbf{x}, \omega) &= -\omega^2 \rho(\mathbf{x}) - \frac{\partial}{\partial x}[\lambda(\mathbf{x}) + 2\mu(\mathbf{x})]\frac{\partial}{\partial x} - \frac{\partial}{\partial z}\mu(\mathbf{x})\frac{\partial}{\partial z} \\
\mathcal{W}_{xz}(\mathbf{x}, \omega) &= -\frac{\partial}{\partial x}\lambda(\mathbf{x})\frac{\partial}{\partial z} - \frac{\partial}{\partial z}\mu(\mathbf{x})\frac{\partial}{\partial x} \\
\mathcal{W}_{zx}(\mathbf{x}, \omega) &= -\frac{\partial}{\partial z}\lambda(\mathbf{x})\frac{\partial}{\partial x} - \frac{\partial}{\partial x}\mu(\mathbf{x})\frac{\partial}{\partial z} \\
\mathcal{W}_{zz}(\mathbf{x}, \omega) &= -\omega^2 \rho(\mathbf{x}) - \frac{\partial}{\partial z}[\lambda(\mathbf{x}) + 2\mu(\mathbf{x})]\frac{\partial}{\partial z} - \frac{\partial}{\partial x}\mu(\mathbf{x})\frac{\partial}{\partial x},
\end{aligned} \tag{14}$$

where ω is the frequency, $\rho(\mathbf{x})$ is the density, $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ are Lamé parameters. These parameters describe spatially variable properties of the medium and are related to P ans S seismic velocities through

$$\lambda(\mathbf{x}) + 2\mu(\mathbf{x}) = \rho(\mathbf{x}) V_p(\mathbf{x})^2 \quad \text{and} \quad \mu(\mathbf{x}) = \rho(\mathbf{x}) V_s(\mathbf{x})^2. \tag{15}$$

Let us emphasize that V_p and V_s are model parameters (mean seismic wave velocities) whereas previously defined V_x and V_z are data parameters (mean horizontal and vertical particle velocities).

If the source is a dirac in the time domain, elementary solutions correspond to the Green functions of the medium: $G_{ip}(\mathbf{x}, \omega, \mathbf{s})$ is the velocity field recorded at $\mathbf{x}$ in the ith direction and emitted by a point force source located at $\mathbf{s}$ and acting in the pth direction at the frequency ω. Therefore, the field component $\mathcal{X}_i$ of eq. (12) can be written as the product of the Green function $G_{ip}(\mathbf{r}, \omega, \mathbf{s})$ calculated at the position $\mathbf{r}$ with the source frequency content $S_p^f(\omega, \mathbf{s})$. This Green function depends on the medium velocity structure.

Please note that these equations are not linear with respect to parameters ρ, λ and μ. We perform the non-linear forward modelling, allowing to take multiple scattering into account. In order to fully solve the forward problem in the frequency domain (eq. 12), we resort to finite difference techniques because of their simplicity. The spatial stencil we use is a rotated one, in which derivatives are calculated along 45° rotated axes with respect to the reference Cartesian axes. This stencil has been shown to have enough accuracy in a velocity–stress staggered grid by Saenger *et al.* (2000). We apply twice this stencil of first-order derivatives to get second-order derivatives, following the parsimonious approach of Luo & Schuster (1990) and Hustedt *et al.* (2004). At the end, this stencil is equivalent to the popular P1 finite element stencil (Abramowitz & Stegun 1965). A rather simple way to introduce the free surface boundary condition is zeroing elastic coefficients above the free surface (Hayashi *et al.* 2001; Graves 1996; Ohminato & Chouet 1997), making this stencil a very attractive one (Saenger & Bohlen 2004; Gélis *et al.* 2005).

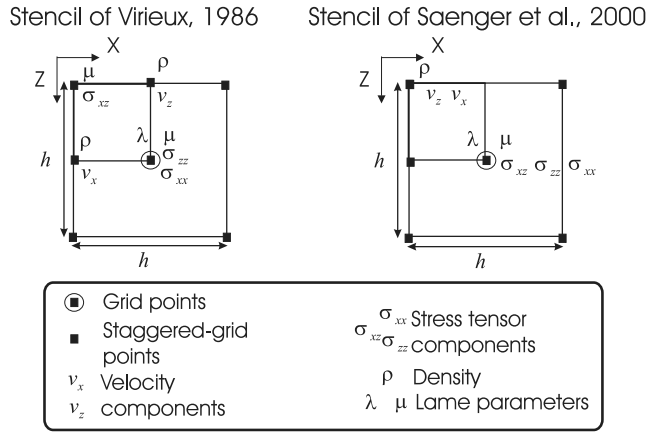

Figure 1. Comparison between the second-order finite difference stencils of Virieux (1986) (left) and Saenger *et al.* (2000) (right).

This stencil defines elastic properties at a single node of the finite difference grid and density at another single node diagonally shifted, making a partially staggered grid different from the one of Virieux (1986) (Fig. 1). It prevents any spurious spatial interpolation problem between neighbouring medium parameters during the inversion procedure and the medium description for the forward calculation.

Because the numerical grid is always finite, we introduce absorbing boundary conditions using Perfectly Matched Layers procedure (Bérenger 1994; Hastings *et al.* 1996; Hustedt *et al.* 2004). This efficient technique for avoiding numerical reflections from grid edges is a key element for the frequency formulation. It prevents any frequency resonance from the grid itself which would have complicated the frequency sampling strategy in our inversion procedure.

Using this finite difference stencil, one can express eq. (12) through a matrix notation (Pratt *et al.* 1996; Stekl & Pratt 1998) which leads to the following rather sparse discrete system

$$W \quad X = S, \tag{16}$$

where the vector X contains the horizontal particle velocity denoted as V_x and the vertical particle velocity denoted as V_z for each point of the medium. The so-called impedance matrix W is a square sparse matrix of dimension $(2 * nx * nz)^2$ where the number of nodes in the x and z directions are, respectively, denoted nx and nz. Only 18 coefficients are non-zero for each line of the matrix. This block-diagonal matrix shape depends on the medium points ordering. The source function $S(\omega)$ is applied at prescribed discrete positions.

In practice, we solve this system by decomposing W in a LU form, where L is a lower triangular matrix and U is an upper one. By performing direct and back substitutions, the forward problem is efficiently solved for each source. One must be aware that this decomposition fills up L and U matrices, which do not have a sparse structure any more. We have selected the MUMPS software (Amestoy *et al.* 2000, 2001) which helps handling this fill-in of the matrix through parallel computing using MPI library. Once this decomposition is performed, which is the most CPU-demanding part (around 70 per cent of the total forward computational effort on average for examples presented in this paper), computing the solution for different sources is simply a matter of substitution, and this can be done efficiently with an incremental approach as the number of sources increases.

The impedance matrix W is a symmetric matrix except in PML zones, thanks to the reciprocity principle (Aki & Richards 2002). Each column of the matrix W^{-1} corresponds to the Green function value at each point in the medium. More precisely, the Green function $G_{xz}(\mathbf{x}, \omega, \mathbf{s})$, which is the horizontal particle velocity recorded at position $\mathbf{x}$ for a vertical point source emitted in $\mathbf{s}$, is equal to the Green function $G_{zx}(\mathbf{s}, \omega, \mathbf{x})$, which is the vertical particle velocity recorded in $\mathbf{s}$ for a horizontal point source emitted in $\mathbf{x}$. As the matrix W depends on the frequency, considering constant values of the point source at each frequency leads to a dirac function in the time domain if all frequencies are considered. However, we use a limited number of frequency components in the inversion and do not reconstruct seismograms in the time domain.

A 2-D-gridded wavefield, denoted as a 'frequency map' (Hustedt *et al.* 2004), is any field component computed at a given frequency for a given source to each spatial point of the model. At the selected frequency, with the help of our finite difference technique, we simulate all kinds of waves (propagating, evanescent, diffracted ones) existing in the medium. Vertical and horizontal velocity components frequency maps are shown in the top right of Fig. 2 for a vertical force source in an infinite heterogeneous medium.

When applying inversion procedure in the next section, the particle velocity will be required at each point of the grid for each source. Moreover, we will need to compute Green functions for impulsive sources located at receiver locations. This is related to the computation of the diagonal part of the Hessian, as we will see in the next section.

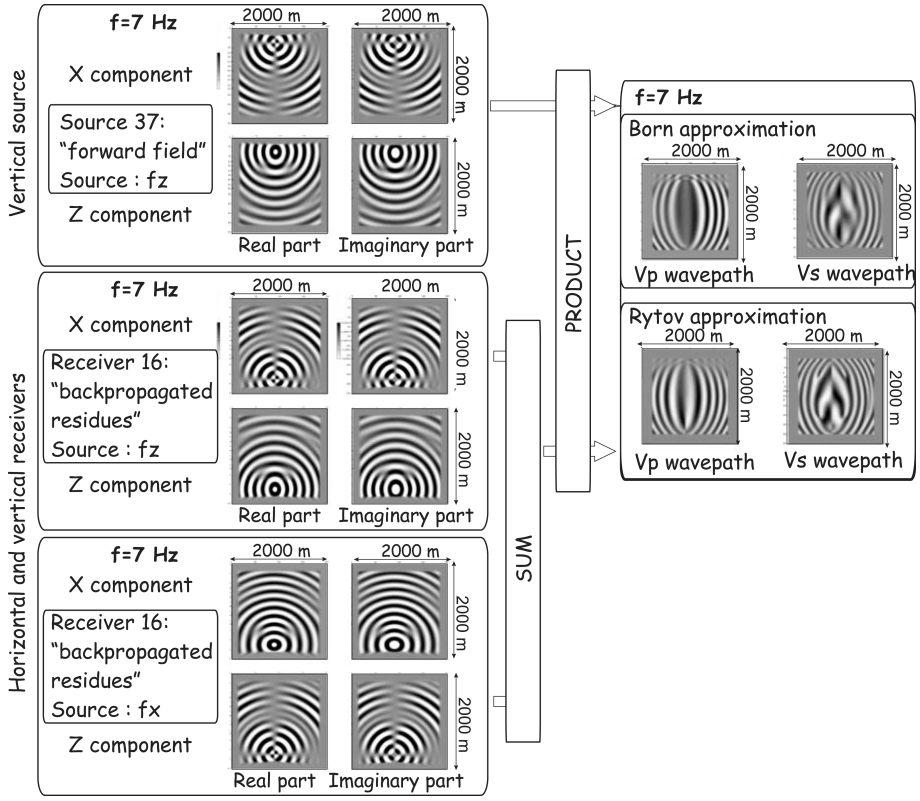

Figure 2. Gradient construction with the Born and Rytov formulations. Frequency maps show the wave propagation at a given frequency on the left for a real vertical source (top) and the backpropagation of residuals located at receiver place for vertical residuals (middle) and horizontal residuals (bottom). To compute the gradient, the forward field is multiplied in the frequency domain with the sum of the fields due to backpropagated residuals. On the right the contributions of one source–receiver couple to the gradient are displayed for V_p and V_s anomalies and for Born and Rytov formulations. Please note the difference between the V_p and V_s contributions to the gradient for one source–receiver couple. Please remark as well the differences between wavepaths obtained with Born and Rytov formulations

4 THE GRADIENT ESTIMATION FOR V_p AND V_s VELOCITY PARAMETERS FOR BORN MISFIT FUNCTION

Fréchet derivatives may be deduced from a local relation between medium parameters perturbations and particle velocity amplitudes perturbations. This relation requires only the first term of the Born series for the first derivative estimation. Because we fully solve the forward problem at each iteration, we do include multiple diffractions in our perturbation estimation.

Spatial derivatives of data perturbations could be obtained through the discrete propagator W previously defined in the eq. (16) (Pratt *et al.* 1996, 1998). The elastic adjoint operator (Pratt *et al.* 1998) could be written as

$$\Re\left[\mathcal{B}_0^t \Delta \mathbf{d}^*\right] = -\mathcal{S}^t (\mathcal{W}^{-1})^t \, \frac{\partial \mathcal{W}}{\partial m} \, \mathcal{W}^{-1} \Delta V^*. \tag{17}$$

In this equation, we recognize the Born Kernel equal to $(\mathcal{W}^{-1})^t \; \partial\mathcal{W}/\partial m \; \mathcal{W}^{-1}$ and surrounded by two excitation terms: the real source $\mathcal{S}^t$ and the non-real source ΔV^*, composed of conjugate residuals located at receivers places. With another reading of this equation, we can distinguish three different terms, corresponding to physical phenomena: the forward propagation of the sources, the backpropagation of the residuals and the derivatives of the differencing operator. We now detail them for the discrete case.

The discrete expression $S^t(W^{-1})^t$ of the forward propagation operator $\mathcal{S}^t(\mathcal{W}^{-1})^t$ is the forward propagating field (Pratt *et al.* 1996, 1998). We may consider a dirac impulse source, discretized by a unitary vector in the model grid for a given frequency. Since $WW^{-1} = I$ where I is the identity matrix, the matrix W^{-1} describes discrete Green functions split into

$$\begin{bmatrix} G^0_{xx}(ix, iz, \omega, \mathbf{s}) & G^0_{xz}(ix, iz, \omega, \mathbf{s}) \\ G^0_{zx}(ix, iz, \omega, \mathbf{s}) & G^0_{zz}(ix, iz, \omega, \mathbf{s}) \end{bmatrix} \tag{18}$$

at the node (ix, iz) of the model grid for a spatial point source $\mathbf{s}$ (ixs, izs) of the model grid.

The discrete term $W^{-1}\Delta V^*$ represents backpropagated residuals (Pratt *et al.* 1996, 1998). Thanks to the reciprocity principle, $(W^{-1})^t = W^{-1}$, and therefore, W^{-1} can be expressed as Green functions emitted from receivers,

$$\begin{bmatrix} G^0_{xx}(ix,iz,\omega,\mathbf{r}) & G^0_{zx}(ix,iz,\omega,\mathbf{r}) \\ G^0_{xz}(ix,iz,\omega,\mathbf{r}) & G^0_{zz}(ix,iz,\omega,\mathbf{r}) \end{bmatrix}, \tag{19}$$

where the receiver is at the discrete position $\mathbf{r}$ (ixr, izr).

Conjugate data residuals ΔV^* correspond to time-reversed residuals in the time domain. They may be backpropagated at once, by using as a non-real source all the residuals located at corresponding receivers for one source. This allows to efficiently compute the misfit function gradient but this prevents from calculating the discrete Hessian matrix diagonal part. Indeed, in order to obtain the discrete Hessian matrix diagonal coefficients, we have to evaluate the discrete expression $B_0^\dagger B_0$ (eq. 8) where B_0 is the discrete Born Kernel equal to $(W^{-1})^t\, \partial W/\partial m\, W^{-1}$. We can equivalently use $B_0^t B_0^*$ since we take the real part. Therefore, in order to get the discrete Hessian matrix diagonal coefficients, we need to compute Green functions for sources located at each receiver. One Green function is then multiplied with the conjugate data residual at this receiver for the selected frequency in order to get the backpropagated field from this residual. We follow the second possibility to get the discrete Hessian matrix diagonal coefficients.

Moreover, as we consider two oriented data fields (the vertical and the horizontal ones), two Green functions have to be calculated for each receiver. This allows to backpropagate horizontal and vertical residuals.

Finally, we compute the discrete expression of operator $\partial\mathcal{W}/\partial V_p$ thanks to eqs (14) and (15) and the finite difference stencil we use. The discrete partial operator $\Delta W/\Delta V_p$ is equal to

$$\frac{\Delta W}{\Delta V_p}(ix,iz,\omega) = \begin{bmatrix} \Delta W_{xx}(ix,iz,\omega)/\Delta V_p & \Delta W_{xz}(ix,iz,\omega)/\Delta V_p \\ \Delta W_{zx}(ix,iz,\omega)/\Delta V_p & \Delta W_{zz}(ix,iz,\omega)/\Delta V_p \end{bmatrix}, \tag{20}$$

where partial derivatives are

$$\begin{aligned}
\Delta W_{xx}(ix,iz,\omega)/\Delta V_p &= -\frac{\Delta}{\Delta x}2V_p(ix,iz)\frac{\Delta}{\Delta x} \\
\Delta W_{xz}(ix,iz,\omega)/\Delta V_p &= -\frac{\Delta}{\Delta x}2V_p(ix,iz)\frac{\Delta}{\Delta z} \\
\Delta W_{zx}(ix,iz,\omega)/\Delta V_p &= -\frac{\Delta}{\Delta z}2V_p(ix,iz)\frac{\Delta}{\Delta x} \\
\Delta W_{zz}(ix,iz,\omega)/\Delta V_p &= -\frac{\Delta}{\Delta z}2V_p(ix,iz)\frac{\Delta}{\Delta z},
\end{aligned} \tag{21}$$

where formal notations $\Delta/\Delta x$ and $\Delta/\Delta z$ represent finite difference operators deduced from the discrete forward formulation.

Similarly, the discrete Fréchet derivative $\Delta W/\Delta V_s$ is equal to

$$\frac{\Delta W}{\Delta V_s}(ix,iz,\omega) = \begin{bmatrix} \Delta W_{xx}(ix,iz,\omega)/\Delta V_s & \Delta W_{xz}(ix,iz,\omega)/\Delta V_s \\ \Delta W_{zx}(ix,iz,\omega)/\Delta V_s & \Delta W_{zz}(ix,iz,\omega)/\Delta V_s \end{bmatrix}, \tag{22}$$

with following expressions of partial derivatives,

$$\begin{aligned}
\Delta W_{xx}(ix,iz,\omega)/\Delta V_s &= \frac{\Delta}{\Delta x}2V_s(ix,iz)\frac{\Delta}{\Delta x} - \frac{\Delta}{\Delta z}2V_s(ix,iz)\frac{\Delta}{\Delta z} \\
\Delta W_{xz}(ix,iz,\omega)/\Delta V_s &= -\frac{\Delta}{\Delta x}4V_s(ix,iz)\frac{\Delta}{\Delta z} - \frac{\Delta}{\Delta z}2V_s(ix,iz)\frac{\Delta}{\Delta x} \\
\Delta W_{zx}(ix,iz,\omega)/\Delta V_s &= -\frac{\Delta}{\Delta z}4V_s(ix,iz)\frac{\Delta}{\Delta x} - \frac{\Delta}{\Delta x}2V_s(ix,iz)\frac{\Delta}{\Delta z} \\
\Delta W_{zz}(ix,iz,\omega)/\Delta V_s &= \frac{\Delta}{\Delta z}2V_s(ix,iz)\frac{\Delta}{\Delta z} - \frac{\Delta}{\Delta x}2V_s(ix,iz)\frac{\Delta}{\Delta x}.
\end{aligned} \tag{23}$$

Discrete operators $\Delta W/\Delta V_p(ix,iz,\omega)$ and $\Delta W/\Delta V_s(ix,iz,\omega)$ are finite difference diffracting terms, expressing the interaction of propagating and backpropagating fields with medium parameters (Pratt *et al.* 1998). They are sparse matrices thanks to the finite difference construction of the matrix W and can be readily estimated from it. Eight non-zero coefficients per line appear in the finite differences matrix nearby the diagonal, depending on the diffracting point position. These coefficients are equal to $\pm 1/2V_p/h^2$ for the V_p parameter (they are equal to $\pm 1/2V_s/h^2$ or $\pm 3/2V_s/h^2$ when considering V_s as parameter), where the finite difference grid step is denoted h.

Injecting discrete estimations of the forward propagation of the sources (eq. 18), the backpropagation of the residuals (eq. 19) and the derivative of the diffracting operators (eqs 20–23) in the eq. (17) will provide the final discrete formulae we shall use in our approach of optimization. Let us write down explicitly the Born misfit function gradient for an arbitrary seismic source S and two-components receivers (therefore, horizontal and vertical residuals). Velocity field components appear for the forward field term through

$$\begin{aligned}
V^0_x(ix,iz,\omega,\mathbf{s}) &= G^0_{xx}(ix,iz,\omega,\mathbf{s})S^f_x(\omega,\mathbf{s}) + G^0_{xz}(ix,iz,\omega,\mathbf{s})S^f_z(\omega,\mathbf{s}) \\
V^0_z(ix,iz,\omega,\mathbf{s}) &= G^0_{zx}(ix,iz,\omega,\mathbf{s})S^f_x(\omega,\mathbf{s}) + G^0_{zz}(ix,iz,\omega,\mathbf{s})S^f_z(\omega,\mathbf{s}),
\end{aligned} \tag{24}$$

for an arbitrary oriented source. Considering as well vertical and horizontal residuals $\Delta V_x^*(\mathbf{r}, \omega, \mathbf{s})$ and $\Delta V_z^*(\mathbf{r}, \omega, \mathbf{s})$, the discrete Born misfit function gradient for the V_p parameter is equal to

$$
\begin{aligned}
\Re[(B_0)^t \Delta \mathbf{d}_b^*]^{V_p} = & -V_p^0(ix, iz)/2h^2 \sum_{\mathbf{s}} \Bigg\{ \sum_{\mathbf{r}} \{ \\
& + \big\{ - G_{xz}^0(ix-1, iz-1, \omega, \mathbf{r}) - G_{zz}^0(ix-1, iz-1, \omega, \mathbf{r}) \\
& - G_{xz}^0(ix-1, iz, \omega, \mathbf{r}) + G_{zz}^0(ix-1, iz, \omega, \mathbf{r}) \\
& + G_{xz}^0(ix, iz-1, \omega, \mathbf{r}) - G_{zz}^0(ix, iz-1, \omega, \mathbf{r}) \\
& + G_{xz}^0(ix, iz, \omega, \mathbf{r}) + G_{zz}^0(ix, iz, \omega, \mathbf{r}) \big\} \Delta V_z^*(\mathbf{r}, \omega, \mathbf{s}) \\
& + \big\{ - G_{xx}^0(ix-1, iz-1, \omega, \mathbf{r}) - G_{zx}^0(ix-1, iz-1, \omega, \mathbf{r}) \\
& - G_{xx}^0(ix-1, iz, \omega, \mathbf{r}) + G_{zx}^0(ix-1, iz, \omega, \mathbf{r}) \\
& + G_{xx}^0(ix, iz-1, \omega, \mathbf{r}) - G_{zx}^0(ix, iz-1, \omega, \mathbf{r}) \\
& + G_{xx}^0(ix, iz, \omega, \mathbf{r}) + G_{zx}^0(ix, iz, \omega, \mathbf{r}) \big\} \Delta V_x^*(\mathbf{r}, \omega, \mathbf{s}) \big\} \\
& \big\{ - V_x^0(ix-1, iz-1, \omega, \mathbf{s}) - V_z^0(ix-1, iz-1, \omega, \mathbf{s}) \\
& - V_x^0(ix-1, iz, \omega, \mathbf{s}) + V_z^0(ix-1, iz, \omega, \mathbf{s}) \\
& + V_x^0(ix, iz-1, \omega, \mathbf{s}) - V_z^0(ix, iz-1, \omega, \mathbf{s}) \\
& + V_x^0(ix, iz, \omega, \mathbf{s}) + V_z^0(ix, iz, \omega, \mathbf{s}) \big\} \Bigg\}.
\end{aligned}
\tag{25}
$$

In this lengthy expression, the forward field spatial derivatives are expressed by a linear combination of terms such as $V_x^0(ix, iz, \omega, \mathbf{s})$, propagating from the real seismic source. Conjugate data residuals, corresponding to time-reversed residuals in the time domain, are considered as individual sources [terms as $\Delta V_x^*(\mathbf{r}, \omega, \mathbf{s})$] backpropagated to the diffracting point (ix, iz) through discrete Green function components combination [terms such as $G_{zx}^0(ix, iz, \omega, \mathbf{r})$]. The interaction of these two field values with medium scatterers appears in the $-V_p^0(ix, iz)/2h^2$ term. Our finite difference stencil collects wave propagation contributions from four different nodes of the mesh.

Derivatives with respect to the parameter V_s could be obtained using the same procedure and are not given explicitly here. Let us just specify that the discrete Born misfit function gradient has a more complicated expression for the V_s parameter.

In eq. (25), we recognize the discrete Born Kernel B_0^t coefficients, corresponding to the whole expression in which conjugate data residuals have to be omitted. To compute the discrete Hessian matrix diagonal terms, the discrete Born Kernel coefficients must be multiplied with their conjugate. This explains why we have to compute Green functions for sources located at receivers location, and not directly fields due to the backpropagation of residuals.

We have started from the discrete wave operator and have deduced the discrete expression of Fréchet derivatives. We may take another road and proceed from the continuum formulation of the wave operator through Green functions and thus express the Fréchet derivatives as proposed by Tarantola (1987). By the discretization of these derivatives using the selected finite difference stencil, we obtain term to term discrete expressions as the eq. (25) as we show in Appendix A.

The physical meaning and the contribution of different factors are shown on Fig. 2 for one source–receiver couple: fields coming from the backpropagation of horizontal and vertical residuals are summed. The resulting field is then multiplied with the field coming from the real source. Both fields meet together at the scatterer location, at the time corresponding to the time propagation from the source to the scatterer. Fig. 2 shows as well that the contribution of one source–receiver couple is different for V_p and V_s parameters reconstruction. The V_p parameter image for one source–receiver couple has a shape similar to the acoustic velocity parameter image in the acoustic configuration (Pratt *et al.* 1998; Ravaut *et al.* 2004), whereas the V_s parameter image is strongly different.

Let us now consider the complex phase data related to the Rytov misfit function.

5 THE GRADIENT ESTIMATION FOR V_p AND V_s VELOCITY PARAMETERS FOR RYTOV MISFIT FUNCTION

Fitting observed and calculated field complex phases leads to another data space. The non-linear relation g relates complex phase perturbations to parameters perturbations. The Fréchet derivative is defined by an operator, denoted $\mathcal{B}_0^r$. In order to estimate its discrete expression B_0^r, discrete parameters perturbations have to be linearly related to discrete field complex phase perturbations. This constitutes the Rytov formulation for medium perturbation.

The unknown medium is the superposition of a reference background and small perturbations. The ith field velocity component $V_i(\mathbf{x})$ that propagates in the medium is linked to the ith velocity field component $V_i^0(\mathbf{x})$ that propagates in the reference background and to the

complex generalized phase shift $\delta\psi_i(\mathbf{x})$ due to the presence of small perturbations in the medium through

$$V_i(\mathbf{x}) = V_i^0(\mathbf{x})e^{\delta\psi_i(\mathbf{x})}. \tag{26}$$

Since the misfit function contains field complex phases, we focus on the complex Neperian logarithm of the velocity field $V_i(\mathbf{x})$.

Moreover, as already pointed out in the acoustic case by Devaney (1981), Woodward (1992), Wu (2003), a relation between Green functions perturbations $\delta G_{pq}(\mathbf{r}, \omega, \mathbf{s})$ and perturbations of their complex phase $\delta\psi_{pq}(\mathbf{r}, \omega, \mathbf{s})$ exists and is expressed through

$$\delta G_{pq}(\mathbf{r}, \omega, \mathbf{s}) = G_{pq}^0(\mathbf{r}, \omega, \mathbf{s})\delta\psi_{pq}(\mathbf{r}, \omega, \mathbf{s}). \tag{27}$$

It means that, in the weak-scattering limit, Rytov data perturbations $\delta\psi_{pq}(\mathbf{r}, \omega, \mathbf{s})$ can be related to Born data pertubations $\delta G_{pq}(\mathbf{r}, \omega, \mathbf{s})$.

Similarly to the Born approach, the perturbation of complex phases can be deduced from model parameter perturbations as suggested by Pratt *et al.* (1998) through a discrete formulation based on finite difference stencil. Before doing so, let us write down operator expressions. The complex source phase $\mathcal{S}_r$ and receiver phase $\mathcal{X}_r$ are related to the field source $\mathcal{S}$ and synthetic receiver seismograms $\mathcal{X}$ through $\mathcal{X}_r = Ln(\mathcal{X})$ and $\mathcal{S}_r = Ln(\mathcal{S})$. The Rytov elastic adjoint operator is equal to

$$\Re\left[\left(\mathcal{B}_0^r\right)^t \Delta d_r^* \right] = \left(\frac{\partial \mathcal{X}_r}{\partial m} \right)^t \Delta d_r^*, \tag{28}$$

where d_r^* are data considered in the Rytov case and $\mathcal{B}_0^r$ is the Fréchet derivatives matrix for a forward problem with a $\mathcal{S}_r$ source. As already pointed out, the data we shall consider in the Rytov formulation are complex phases and may be written as $\Delta[Ln(V)^*]$ where V represent real seismograms which are the data considered in the Born approach. Furthermore, partial derivatives of seismograms with respect to model parameters may be expressed as

$$\frac{\partial \mathcal{X}_r}{\partial m} = \frac{\partial Ln(\mathcal{X})}{\partial m} = \mathcal{X}^{-1}\frac{\partial \mathcal{X}}{\partial m}. \tag{29}$$

The notation $\mathcal{X}^{-1}$ means that we take the inverse of each term composing the vector $\mathcal{X}$.

Therefore, the eq. (28) becomes

$$\Re\left[\left(\mathcal{B}_0^r\right)^t \Delta d_r^* \right] = \frac{\partial \mathcal{X}^t}{\partial m}(\mathcal{X}^{-1})^t \Delta[Ln(V)^*]. \tag{30}$$

Because $\partial\mathcal{X}^t/\partial m = \mathcal{X}^t\ \partial\mathcal{W}/\partial m\ \mathcal{W}^{-1}$ (eqs 15 and 16 of Pratt *et al.* 1998), we deduce sequentially

$$\begin{aligned} \Re\left[\left(\mathcal{B}_0^r\right)^t \Delta d_r^* \right] &= -\mathcal{S}^t(\mathcal{W}^{-1})^t \frac{\partial \mathcal{W}}{\partial m}\mathcal{W}^{-1}(\mathcal{X}^{-1})^t \Delta[Ln(V)^*] \\ \Re\left[\left(\mathcal{B}_0^r\right)^t \Delta d_r^* \right] &= -\mathcal{S}^t(\mathcal{W}^{-1})^t \frac{\partial \mathcal{W}}{\partial m}\mathcal{W}^{-1}(\mathcal{S}^{-1})^t\mathcal{W}^t \Delta[Ln(V)^*]. \end{aligned} \tag{31}$$

The term $(\mathcal{W}^{-1})^t\partial\mathcal{W}/\partial m\mathcal{W}^{-1}(\mathcal{X}^{-1})^t$ corresponds to the Rytov Kernel. The term $(\mathcal{X}^{-1})^t = (\mathcal{S}^{-1})^t\mathcal{W}^t$, appearing at the end of this expression corresponds to the forward field, propagating from the source to the receiver (eqs 27 and 29). Since the forward field term $(\mathcal{X}^{-1})^t$ does not depend on the diffracting point, it may be moved in the discrete Rytov Kernel expression, so that we may consider the expression $(\mathcal{W}^{-1})^t(\mathcal{X}^{-1})^t\partial\mathcal{W}/\partial m\ \mathcal{W}^{-1}$ as the new Rytov Kernel.

Using this new expression, we can distinguish and physically interpret three terms of the misfit function gradient, which are different from these obtained with the Born formulation but are still similar. The forward propagating field $\mathcal{W}^{-1}S$ from the source to the diffracting point is multiplied with $(\mathcal{X}^{-1})^t$, where $\mathcal{X}$ is the reference forward propagating field propagating from the source to the receiver. Therefore, we obtain the propagation of the forward field complex phase from the source to the diffracting point. This represents a fundamental difference with respect to the Born approach. Backpropagated complex phase residuals appear in $\mathcal{W}^{-1}\Delta[Ln(d^*)]$, which is also specific to the Rytov approach. These two 'complex phases fields' interact with local medium parameters variations $\partial\mathcal{W}/\partial m$.

Let us now consider discrete formulation. The continuous operator $\partial\mathcal{W}/\partial m$ is replaced by the discrete one $\Delta W/\Delta m$, defined by eqs (20)–(23). The matrix W in $(X^{-1})^t = W^t(S^{-1})^t$, corresponding to the forward field propagating from the source to the receiver and independent of the diffracting point (ix, iz) (eqs 27 and 29), can be expressed with Green functions between sources and receiver points. By combining previously detailed discrete terms, we deduce the global discrete expression of the Rytov misfit function gradient, similar to the Born one (eq. 25). For an arbitrary source and vertical and horizontal receivers, the Rytov misfit function gradient for the V_p parameter equals to

$$
\begin{aligned}
\Re\left[(B_0^r)^t \Delta \mathbf{d}_r^*\right]^{V_p} = & -V_p^0(ix,iz)/2h^2 \sum_{\mathbf{s}} \left\{ \sum_{\mathbf{r}} \{ \right. \\
& + \{ -G_{xz}^0(ix-1,iz-1,\omega,\mathbf{r}) - G_{zz}^0(ix-1,iz-1,\omega,\mathbf{r}) \\
& - G_{xz}^0(ix-1,iz,\omega,\mathbf{r}) + G_{zz}^0(ix-1,iz,\omega,\mathbf{r}) \\
& + G_{xz}^0(ix,iz-1,\omega,\mathbf{r}) - G_{zz}^0(ix,iz-1,\omega,\mathbf{r}) \\
& + G_{xz}^0(ix,iz,\omega,\mathbf{r}) + G_{zz}^0(ix,iz,\omega,\mathbf{r}) \} \Delta Ln(V_z(\mathbf{r},\omega,\mathbf{s})^*) \\
& + \{ -G_{xx}^0(ix-1,iz-1,\omega,\mathbf{r}) - G_{zx}^0(ix-1,iz-1,\omega,\mathbf{r}) \\
& - G_{xx}^0(ix-1,iz,\omega,\mathbf{r}) + G_{zx}^0(ix-1,iz,\omega,\mathbf{r}) \\
& + G_{xx}^0(ix,iz-1,\omega,\mathbf{r}) - G_{zx}^0(ix,iz-1,\omega,\mathbf{r}) \\
& \left. + G_{xx}^0(ix,iz,\omega,\mathbf{r}) + G_{zx}^0(ix,iz,\omega,\mathbf{r}) \} \Delta Ln(V_x(\mathbf{r},\omega,\mathbf{s})^*) \right\} \\
& \left\{ \frac{1}{V_x^0(\mathbf{r},\omega,\mathbf{s})} \{ -V_x^0(ix-1,iz-1,\omega,\mathbf{s}) - V_x^0(ix-1,iz,\omega,\mathbf{s}) \right. \\
& + V_x^0(ix,iz-1,\omega,\mathbf{s}) + V_x^0(ix,iz,\omega,\mathbf{s}) \} \\
& + \frac{1}{V_z^0(\mathbf{r},\omega,\mathbf{s})} \{ -V_z^0(ix-1,iz-1,\omega,\mathbf{s}) + V_z^0(ix-1,iz,\omega,\mathbf{s}) \\
& \left. \left. - V_z^0(ix,iz-1,\omega,\mathbf{s}) + V_z^0(ix,iz,\omega,\mathbf{s}) \} \right\} . \right\}
\end{aligned}
\tag{32}
$$

In this expression, we clearly distinguish the forward field spatial terms (propagation from the real seismic source divided by the source/receiver forward field) from the backpropagated residuals, coming from the propagation (via Green functions derivatives) of conjugate complex phase residuals considered as sources. In both cases, propagating fields are velocity complex phases fields. The interaction of these two fields with the medium scatterers appears in the $-V_p^0(ix,iz)/2h^2$ term and is equal to the one appearing in the eq. (25) for the Born formulation.

Similarly to the Born formulation, this expression may be deduced from the discretization of equations coming from the Tarantola (1987) approach and detailed in Appendix B. Moreover, contributions of different factors are shown in the Fig. 2 for one source–receiver couple. Please note again that V_p and V_s parameter images differ. We will analyse differences between the Born and Rytov approaches in the next section.

6 CONVERGENCE BEHAVIOUR FOR BORN AND RYTOV FORMULATIONS

Let us consider a very simple example in order to analyse how non-linear optimization performs when Fréchet derivatives are calculated both for Born (amplitude fitting) and Rytov (phase fitting) formulations. Comparative quantification is difficult because one method may be more efficient for a specific data acquisition geometry and for different types of heterogeneities. By considering rather simple diffracting elements, convergence is analysed regardless deeper efficiency investigation of Born and Rytov methods.

Let us consider an infinite medium with a *P*-wave velocity of 1500 m s^{-1} and an *S*-wave velocity of 1200 m s^{-1}. Two finite-sized anomalous disks of radius 100 m are inserted with *P*-wave velocity of 1800 m s^{-1} and an *S*-wave velocity of 1440 m s^{-1}. The numerical grid size is 201 by 201 points with a spatial step of 10 m. The vertical point force source is a Ricker signal centred on 5 Hz which represents a rather impulsive signal. Fifteen shots are defined per edge separated by 100 m (Fig. 3). Each source is recorded by 36 geophones located on the opposite side along the bold line with a stepping of 40 m. This transmission acquisition geometry allows to densely sample the medium and illuminates anomalies with different diffracting angles. We choose to image two anomalies because multiple scattering occurs. The forward problem should take care of that and ghost images, if any, should vanish.

The initial medium is the constant velocity medium without anomalies and we proceed from low frequencies towards high frequencies. This allows us to reconstruct long wavelength anomalies before going into details. This careful introduction of the frequency content will also prevent fast focusing of anomalies that may lead to some local minimum in the model space. One may hope we avoid somehow cycle-skipping which is often more dramatic at high frequency than at low frequency. For each frequency, the initial model comes from the inversion of the previously inverted frequency. Twenty iterations are performed at each frequency. We progressively introduce details in the reconstructed images. The inverted frequencies are 1.75, 3, 4.25 and 10.25 Hz. Their choice depends on the acquisition geometry and on the anomaly position to be reconstructed as showed Sirgue & Pratt (2004) for a reflection acquisition geometry. We show in Fig. 4 and Appendix C that the relation proposed by Sirgue & Pratt (2004) may be used for a transmission acquisition geometry. Thus, this relation can be used to select frequencies, such as to cover a continuous wavenumber spectrum in the target zone.

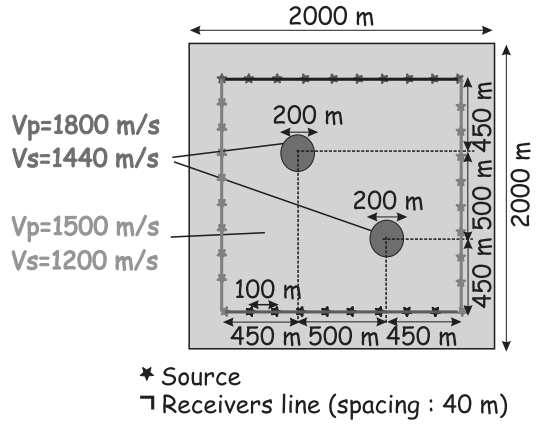

Figure 3. Background and transmission acquisition geometry synthetic example for illustration of Born and Rytov formulations performances. Receivers recording waves coming from the corresponding sources are the same colour. Background medium is homogeneous. Perturbations are 20 per cent high. Acquisition geometry provides transmission effects.

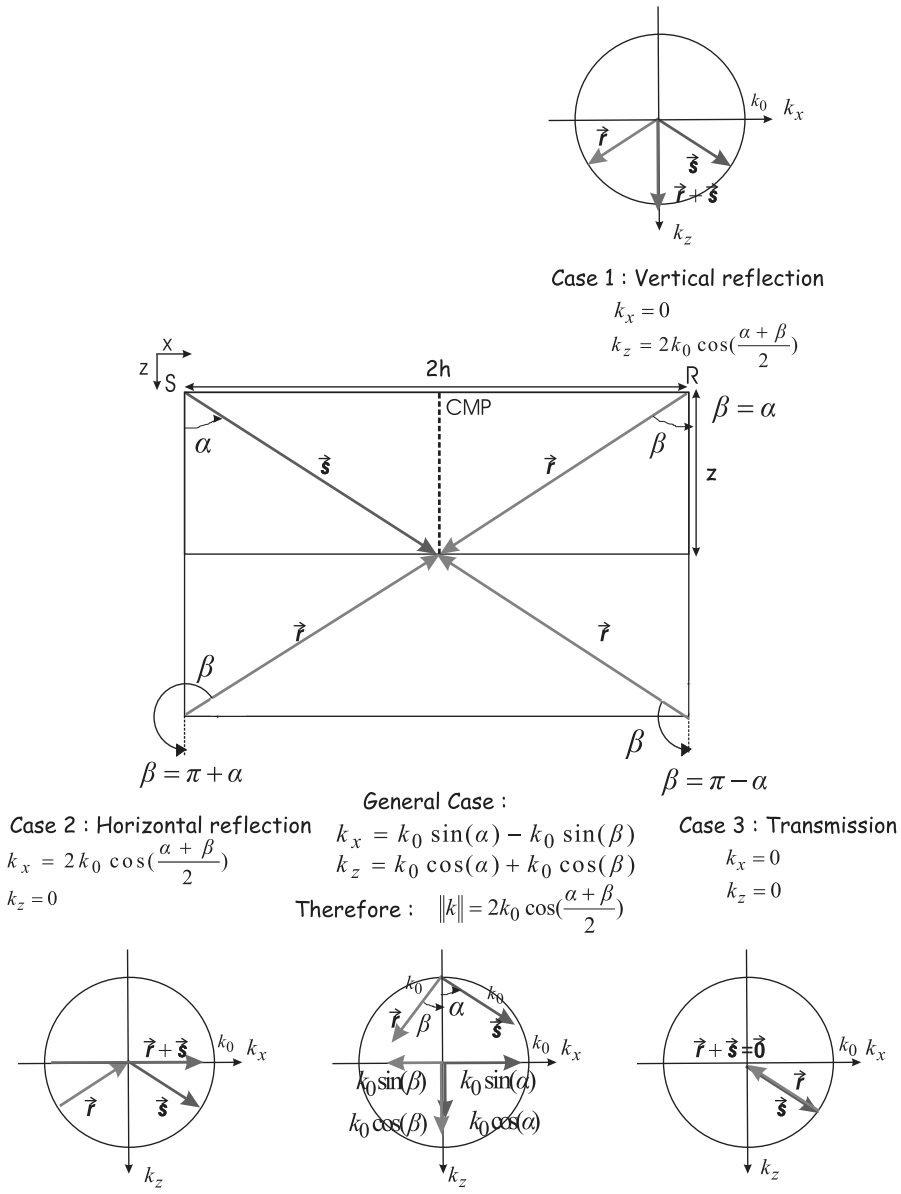

Figure 4. Wavenumber imaging, depending on the source and receiver locations. The wavenumber norm remains the same whatever the source and receivers positions are.

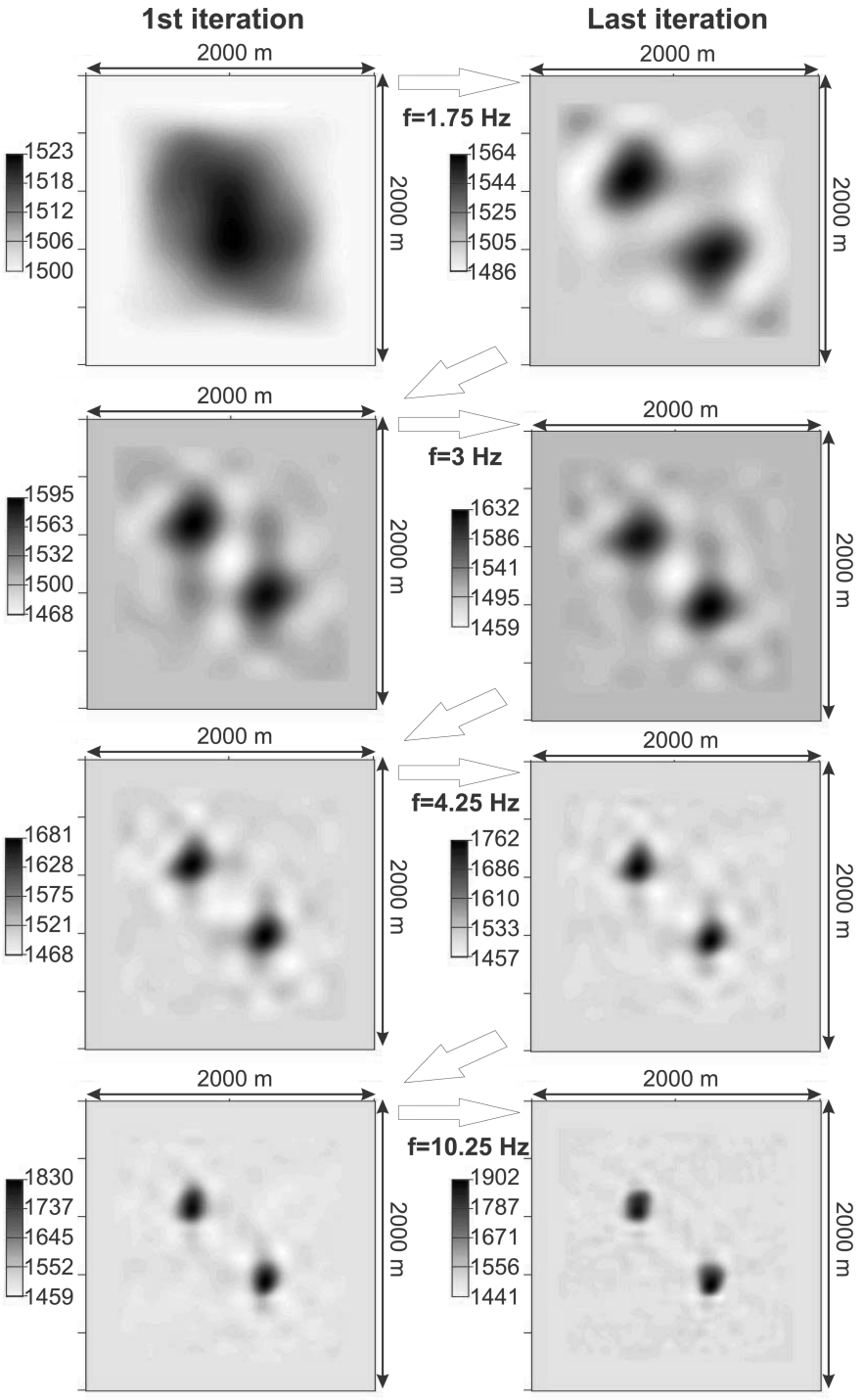

Figure 5. V_p progressive imaging with the frequency increase in the Born formulation. Please note the importance of frequency in focusing anomalies.

Figs 5 and 6 show the progressive reconstruction of anomalies both in shape and in velocity amplitude for V_p and V_s parameters when considering the Born formulation. V_p and V_s parameters are simultaneously inverted and differently reconstructed. Rytov formulation, although equivalent for weak anomalies, behaves differently when considering this example where anomalies have amplitude perturbation as high as 20 per cent (see Figs 7 and 8).

Although the Born and Rytov formulations are strictly restricted to small perturbations, in this experiment the acquisition redundancy allows the use of a homogeneous starting model without preventing the accurate reconstruction of the anomalies in shape, depth and amplitude. Moreover, final parameter images obtained with the Born and Rytov formulations are almost the same, showing the good convergence of the inversion in both cases, although patterns to search the minima are not identical.

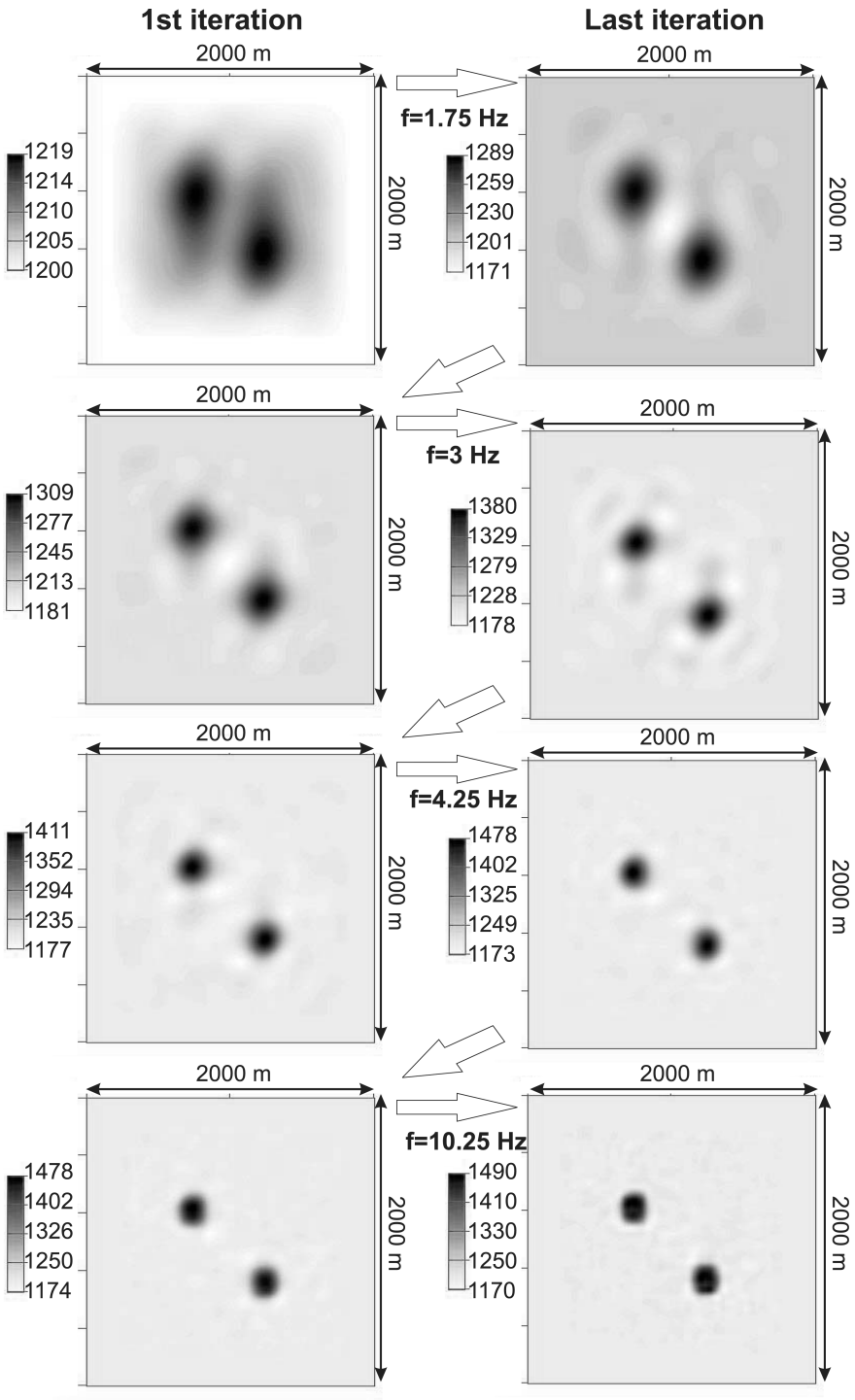

Figure 6. V_s progressive imaging with the frequency increase in the Born formulation. Please note the importance of frequency in focusing anomalies. V_p (Fig. 5) and V_s parameters are differently reconstructed with the Born formulation.

The size of reconstructed scatterers depends on the selected wavelength L. Details progressively appear when the inverted frequency increases, allowing to characterize scatterers more finely. The way anomalies are recovered is different when inverted parameters change, as already noticed by Mora (1987).

The choice of the first frequency is not critical in this very specific example of very well-localized anomalies but one may see that broad reconstruction at low frequencies does not prevent further refinements in this case. The frequency content will be essentially limited by the source spectrum in real applications (Sirgue 2003).

To better understand differences between Born and Rytov formulations, we focus on the gradient of the misfit function expression. Fig. 2 shows Born and Rytov wavepaths (Woodward 1992) that represent the kernels of the misfit function gradient for the V_p and V_s parameters for

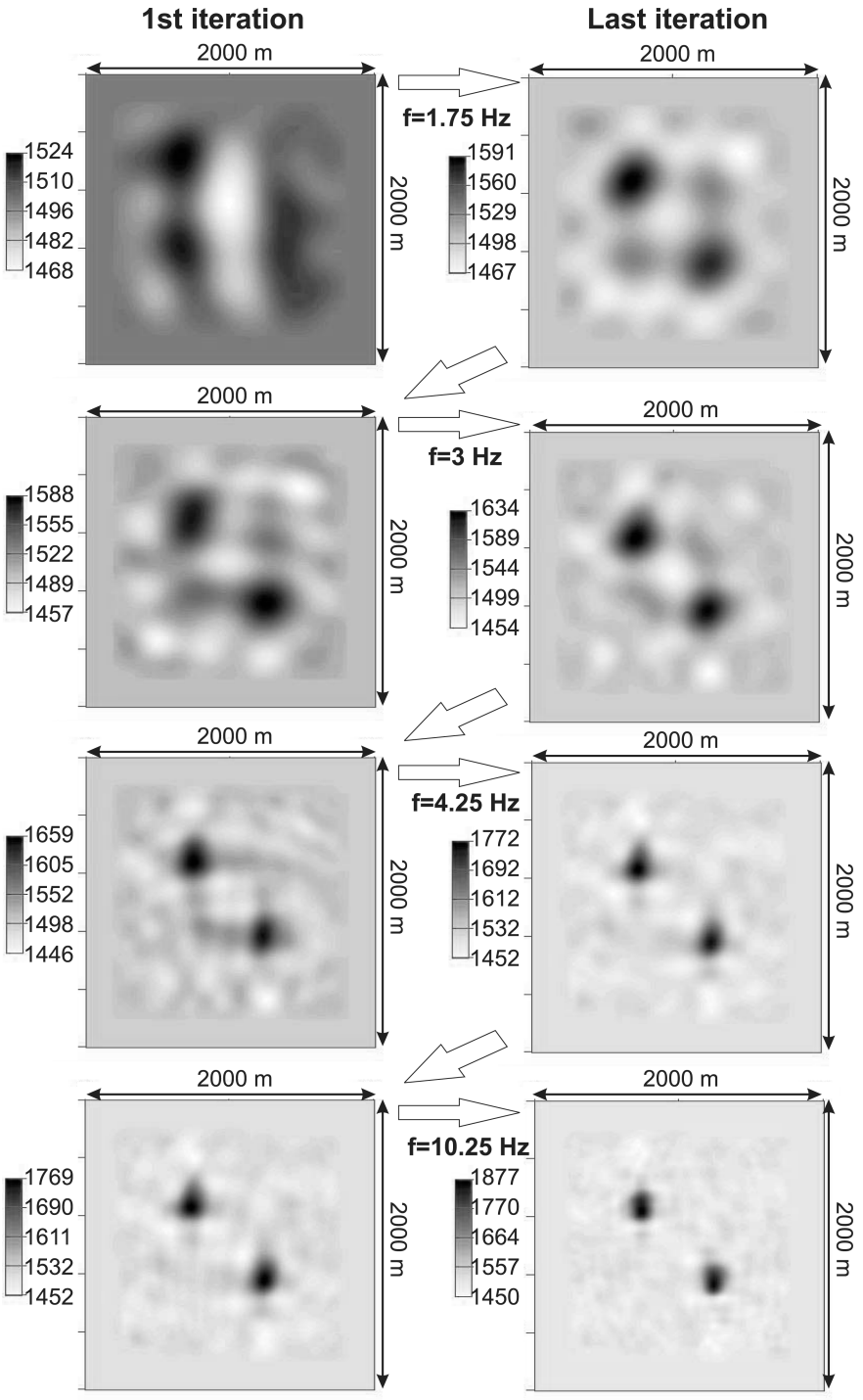

Figure 7. V_p progressive imaging with the frequency increase in the Rytov formulation. Please note the importance of frequency in focusing anomalies. The V_p parameter is differently reconstructed with the Born (Fig. 5) and the Rytov formulations. The final image is similar to the one obtained with the Born formulation (Fig. 5).

one source–receiver couple. They correspond to eqs (25) and (32) when only one source–receiver couple is considered in the summation. They are obtained here by simultaneously inverting the V_p and V_s parameters. This wavepath describes the path followed by a scattered wavefield from the source to the scatterer and from the scatterer to the receiver. Once more, this highlights that one simple diffraction is considered with the Born and Rytov formulations. The central area is equivalent to the first Fresnel zone (Woodward 1992; Pratt *et al.* 1996). The width of the first Fresnel zone for one wavepath depends on the inverted frequency through the considered wavelength L and the distance between the source and the receiver O since the first Fresnel zone width equals $\sqrt{L * O}$ (Pratt *et al.* 1998). For higher frequencies, the wavelength is smaller and positive and negative variations zones are closer. By summing up such wavepaths for all source–receiver couples, Fresnel zones

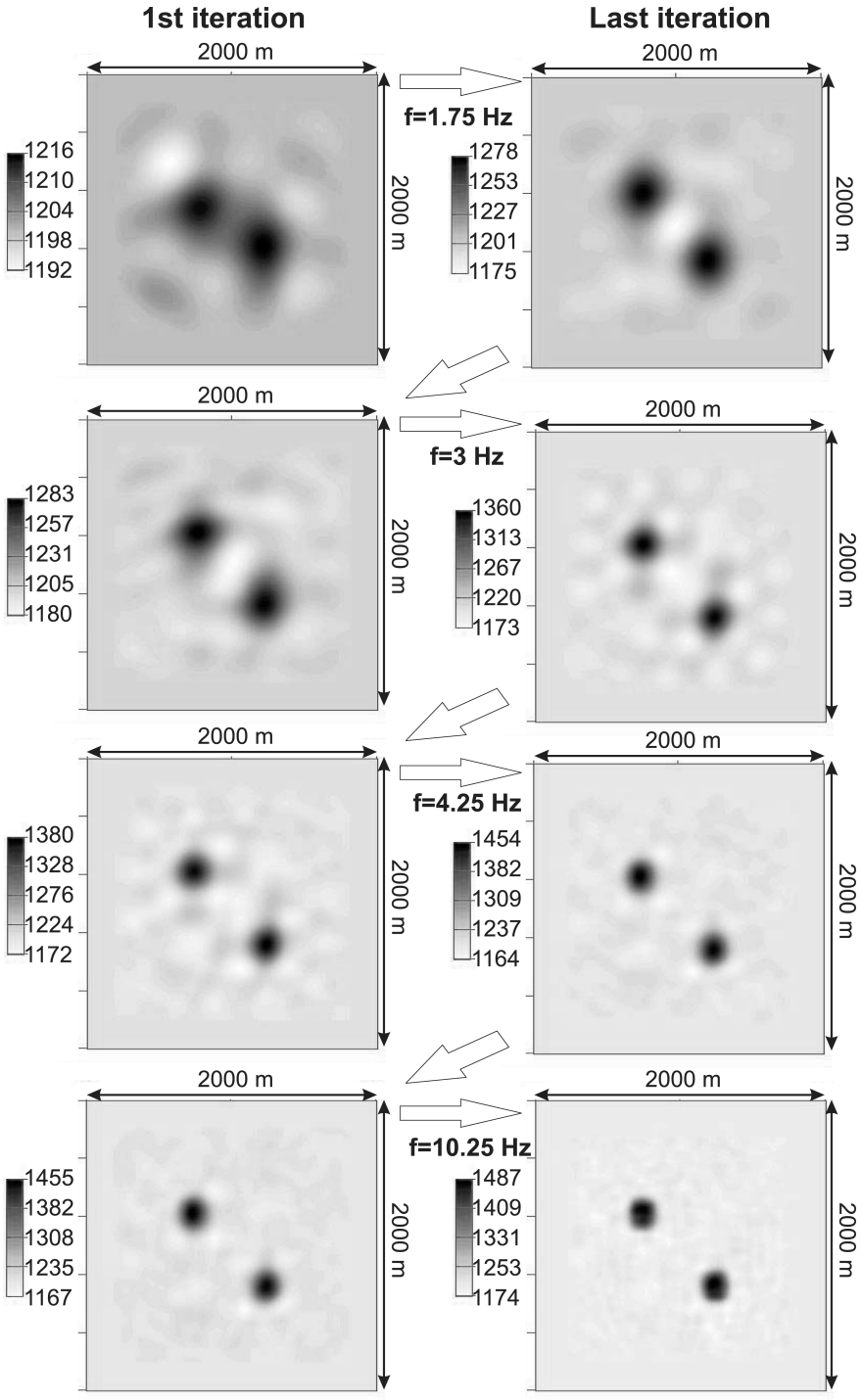

Figure 8. V_s progressive imaging with the frequency increase in the Rytov formulation. Please note the importance of frequency in focusing anomalies. V_p (Fig. 7) and V_s parameters are differently reconstructed with the Rytov formulation. The V_s parameter is differently reconstructed with the Born (Fig. 6) and the Rytov formulations. The final image is similar to the one obtained with the Born formulation (Fig. 6).

containing significant energy (due to scatterers) are progressively surimposed and finally allow to locate the scatterers. As each wave going from one source to one receiver differently illuminates anomalies, denser and wider acquisition geometries provide better images in practice (Pratt *et al.* 1996).

Fig. 2 shows as well that the inversion differently reconstructs V_p and V_s anomalies for the Born and Rytov formulations. As Figs 5–8 show, images obtained from Born and Rytov formulations differ. Differences are weak and mainly located in the first Fresnel zone. This was already noticed by Woodward (1992) who shows Rytov and Born wavepaths for the acoustic case. When considering the time domain and,

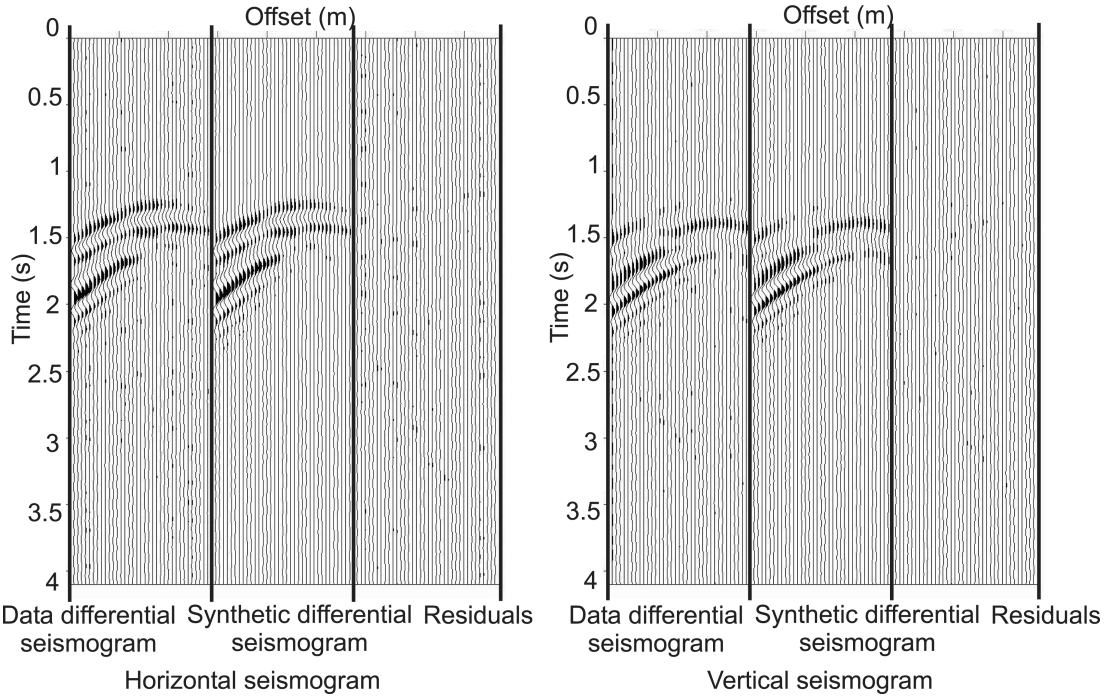

Figure 9. Comparison of data differential seismograms (difference between data to be inverted and data calculated in the initial model), synthetic differential seismograms (difference between data calculated in the final model and data calculated in the initial model) and residuals (difference between data differential seismograms and synthetic differential seismograms) with the Born formulation and a transmission acquisition geometry. Please note that residuals are very weak, which indicate the very good fit of data.

therefore, the contribution of all frequencies, Woodward (1992), Dahlen *et al.* (2000), Dahlen & Baig (2002), Spetzler & Snieder (2004) noticed that heterogeneities located on the ray path do not affect traveltimes but only wave amplitudes.

We now focus on the practical way Born and Rytov formulations allow to reconstruct anomalies for the V_p and V_s parameters. The Born formulation images heterogeneities in the medium as a series of independent simple scatterers. The diffracted field is, therefore, considered as a series of neighbouring hyperbolae due to the presence of simple scatterers. These hyperbolae result from the convolution in the time domain of the forward field and the backpropagated field residuals. These hyperbolae constructively and successively are added in the time domain, thus reproducing the desired diffracting wavefield. The Fig. 9 shows three differential seismograms for the vertical component, a source located on the bottom left of the model and the corresponding receivers. The first differential seismogram represents the difference between the initial field and the data field, the second one shows the difference between the initial field and the field calculated in the final inverted model. The third seismogram, called residual seismogram, contains the difference between the first and second seismograms. The inversion explains 92 per cent of the vertical seismograms energy and 87 per cent of the horizontal seismograms energy. Unexplained energy mainly comes from small phases shifts, indicating that the final inverted model corresponds to a minimum model.

The Rytov formulation proceeds in a different manner. The medium perturbations are reconstructed owing to the phase shifts and logarithmic amplitude variations they induce on receivers. The diffracted field is considered as a series of waves of different phase shifts, coming from different scatterers. Residuals (Fig. 10) calculated in the Rytov formulation contain almost the same level of energy as for the one deduced by the Born formulation: 93 per cent of the vertical seismograms energy and 88 per cent of the horizontal seismograms energy are explained.

As a partial conclusion, scatterers are well located spatially. They have the correct size and the velocity amplitude is fairly well estimated. Optimization is quite robust leaving unexplained energy in residuals. The final images very weakly depend on the chosen Born or Rytov formulation, whereas the road taken to reach the global minimum of the misfit function is quite different (see the first parameter images for the first frequency in Figs 5–8). We may expect more complex convergence when other data acquisition geometries will be considered.

7 CHOICE OF INVERTED PARAMETERS

Selection of the parameter space is expected to be important and non-linear relations may lead to different results. Based on radiation pattern diagrams (Forgues 1996), we may consider four other parameters couples, denoted (V_ϕ, V_s), (λ, μ), (κ, μ) and [$Ln(V_p)$, $Ln(V_s)$]. V_ϕ and κ

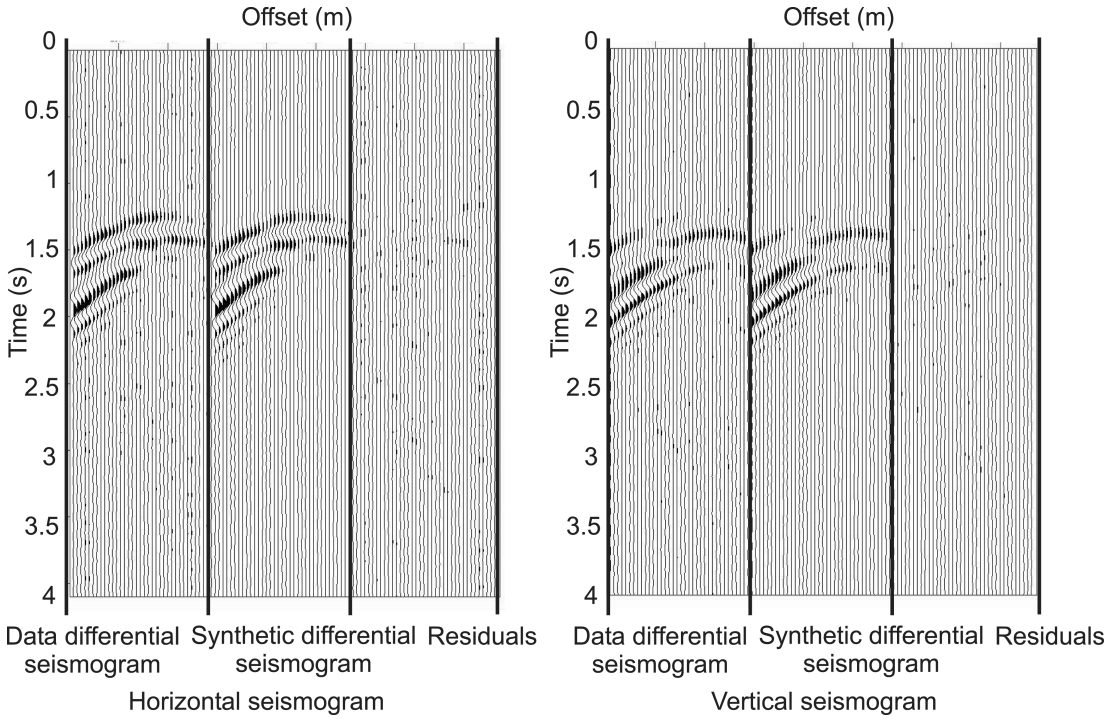

Figure 10. Comparison of data differential seismograms (difference between data to be inverted and data calculated in the initial model), synthetic differential seismograms (difference between data calculated in the final model and data calculated in the initial model) and residuals (difference between data differential seismograms and synthetic differential seismograms) with the Rytov formulation and a transmission acquisition geometry. Please note that residuals are very weak, which indicate the very good fit of data. Final results are equivalent to the ones obtained with the Born formulation (Fig. 9).

are defined by relation (15) and following expressions

$$V_\phi(\mathbf{x})^2 = \rho(\mathbf{x})\left[V_p(\mathbf{x})^2 - \frac{4}{3}V_s(\mathbf{x})^2\right]$$
$$\kappa(\mathbf{x}) = \rho(\mathbf{x})\left[V_p(\mathbf{x})^2 - \frac{4}{3}V_s(\mathbf{x})^2\right]. \tag{33}$$

The bulk velocity is denoted by V_ϕ (Gorbatov & Kennett 2003) and is currently inverted in the acoustic case (Ravaut *et al.* 2004; Operto *et al.* 2004; Dessa *et al.* 2004). By choosing V_ϕ instead of V_p, we separate strict acoustic phenomena to shear ones in model reconstruction. We consider as well Lamé parameters couples: (λ, μ) and also (κ, μ). The parameter set [$Ln(V_p)$, $Ln(V_s)$] may be of interest because derivatives are related to relative variations of seismic velocities V_p and V_s.

The geometry and amplitudes of anomalous disks to be recovered are identical to the previous example as well as the acquisition geometry. Selected parameters are simultaneously inverted for each couple. After the inversion, we reconstruct the V_p and V_s parameters for analysis and horizontal profiles crossing the lowest anomaly (Fig. 3) are shown in Figs 11 and 12.

Whatever the inverted couple we select, final images show Gaussian-shape amplitude variations due to the finite frequency effects that do not allow to recover sharp velocity contrasts. The Born and Rytov formulations similarly reconstruct scatterers in amplitude and in shape. With the Rytov formulation, we have not tried to unwrap the phase and have adopted to ignore any data with a phase shift higher than 2π.

As a partial conclusion, results obtained with any parameters couple are equivalent for the data acquisition geometry and the frequency content we have considered.

8 THE INFLUENCE OF ACQUISITION GEOMETRY

The acquisition geometry plays a key role in the image reconstruction (Mora 1988; Pratt *et al.* 1998). We now illustrate its influence in a synthetic example based on the geometry of a real experiment which has taken place in the North of France near a railway (Leparoux *et al.* 2002). Detection of identified near-surface cavities, embedded at a maximal depth of 5 m, was the objective. We consider a zone of 45 m width and 20 m depth (Fig. 13a). Two disk anomalies are inserted with roofs located at 2 and 4 m depth, respectively, and their radius is equal to 1.2 m. The background medium has following properties: V_p equals 888 m s^{-1}, V_s equals 431 m s^{-1} and ρ equals 1600 kg m^{-3}. The model is discretized with a 0.1 m numerical spatial step and contains 451 points horizontally and 201 points vertically.

In this synthetic simulation, we do not introduce yet the free surface for better analysis. Nevertheless, the medium has been discretized following the rule of 30 points per wavelength (Saenger & Bohlen 2004) used when the free surface is introduced, in order to compare these results with the ones obtained with the free surface.

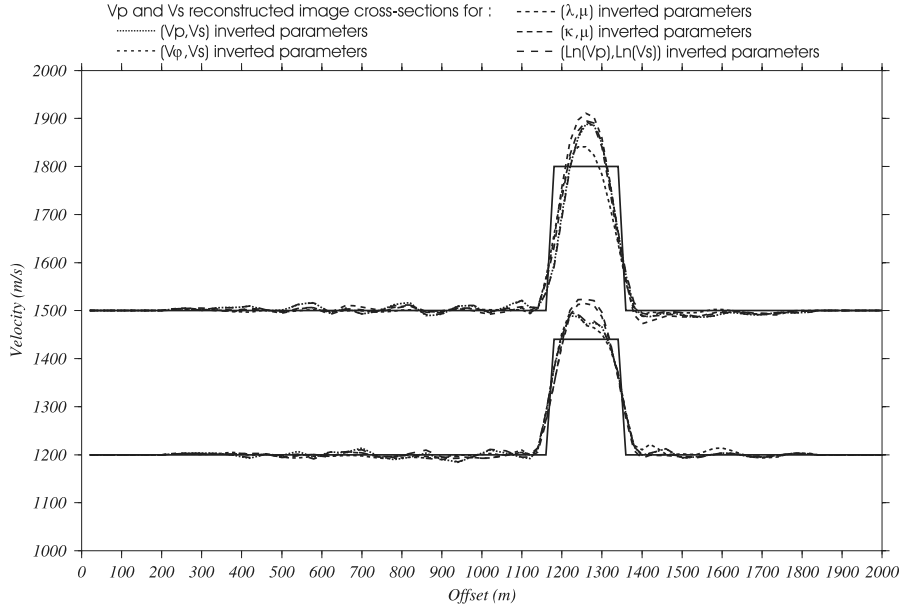

Figure 11. V_p and V_s parameters cross-sections extracted along a horizontal line crossing the lowest anomaly (see Fig. 3) and calculated with the Born formulation for different parameters couples. Anomalies size, location, shape and amplitudes are well reconstructed. The inverted parameters couple choice has a weak influence with this transmission acquisition geometry.

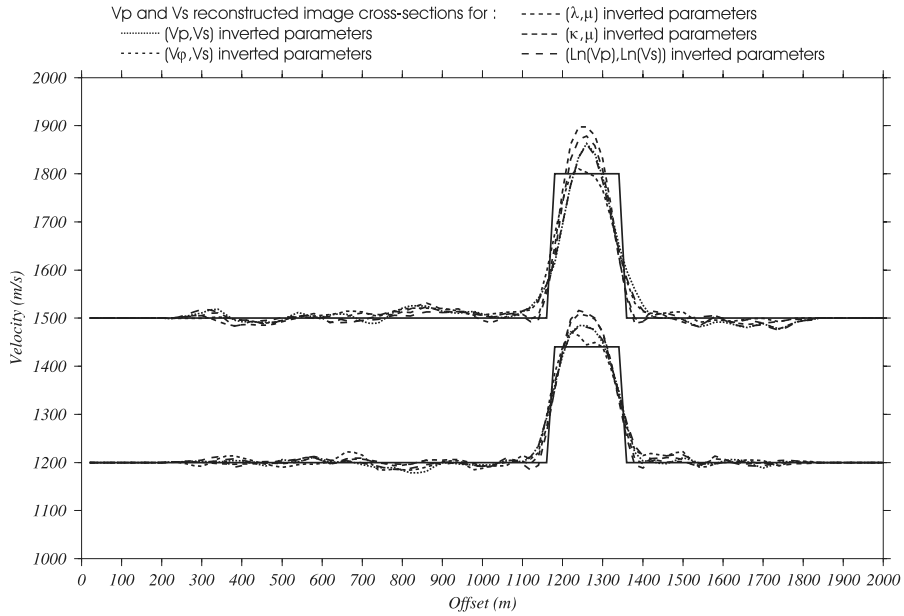

Figure 12. V_p and V_s parameters cross-sections extracted along a horizonatal line crossing the lowest anomaly (see Fig. 3) and calculated with the Rytov formulation for different parameters couples. Anomalies size, location, shape and amplitudes are well reconstructed. The inverted parameters couple choice has a weak influence with this transmission acquisition geometry. Final results are equivalent to the ones obtained with the Born formulation (Fig. 11).

Sources and receivers are located above anomalies. The acquisition configuration is expected to influence the parameters final images since anomalies are differently illuminated with respect to the previous transmission acquisition configuration example.

If sources and receivers are located on opposite sides with respect to scatterers, the latter will be contained in the first Fresnel zone for some source–receiver couples and, therefore, the full waveform inversion performs more or less like a traveltime tomography (Pratt *et al.* 1996). On the contrary, if sources and receivers are located on the same side with respect to scatterers, the latter are located on external Fresnel zones corresponding to high phase shifts, and therefore, will influence later times in seismograms. One may think that the full waveform inversion provides migration-like tomography (Pratt *et al.* 1996).

Fourty-three receivers record horizontal and vertical velocity fields coming from thirty-seven vertical force sources (Fig. 13a). The spacing between sources is 1 m as well as for receivers. Each source is recorded by all receivers. The source is a Ricker wavelet. The frequency content is centred around 88 Hz. We select four inverted frequencies at 10, 20, 50 and 130 Hz.

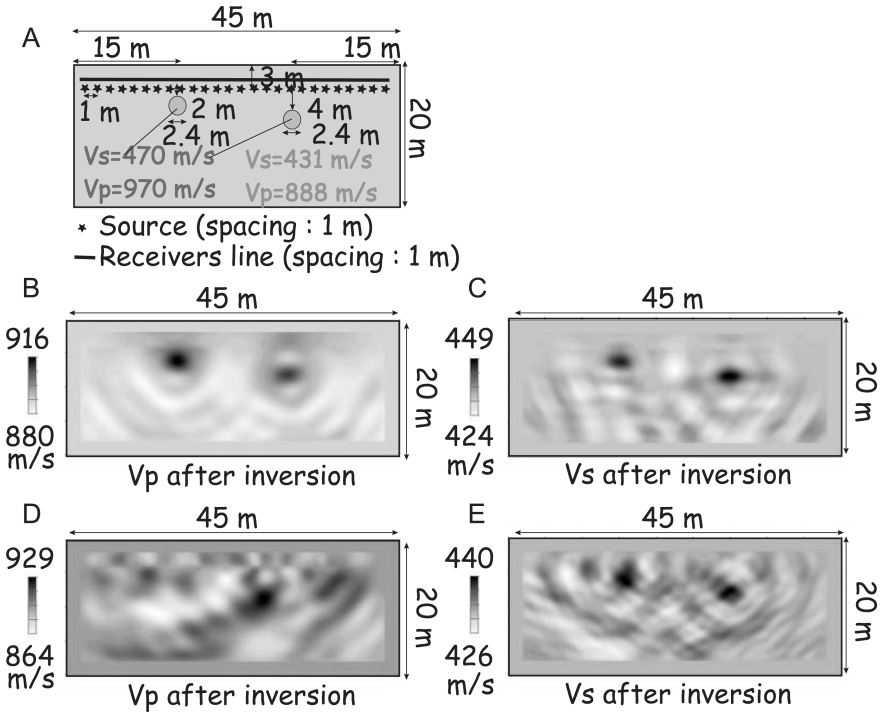

Figure 13. (a) Background and reflection acquisition geometry. (b) V_p parameter calculated with the Born formulation. (c) V_s (right) parameter calculated with the Born formulation. (d) V_p parameter calculated with the Rytov formulation. (e) V_s parameter calculated with the Rytov formulation. Please note the better anomalies reconstruction with the Born formulation for V_p and V_s parameters.

We choose to perform the elastic full waveform inversion with the Born and the Rytov formulations. Moreover, as our main goal is locating anomalies and determining perturbation amplitudes in real cases, we select the (V_p, V_s) seismic velocities as inverted parameters.

The background structure is known as for the previous example. We focus our attention on anomalies reconstruction. We assume that long wavelengths of the model have already been determined with another method such as the first arrival traveltime tomography (Improta *et al.* 2002) or the Spectral Analysis of Surface Waves (Nazarian & Stokoe 1984).

Results obtained from the inversion are presented on the Figs 13(b)–(e). Anomalies are better reconstructed with the Born formulation than with the Rytov formulation, due to the reflection acquisition (Beydoun & Tarantola 1988). These results are in good agreement with the ones obtained by Pratt & Worthington (1988). They show in the acoustic case that the Born formulation allows to better reconstruct the edges of discrete objects than the Rytov approximation. In the V_p and V_s parameter images calculated with the Born formulation, we clearly distinguish the two anomalies although small ghosts appear beneath them. They are due to the limited aperture of the acquisition configuration. Let us point out that we perform an image of each point of the medium, therefore, background parameters are imaged as well, even if we take as their initial values the true ones. Moreover, as the background is homogeneous, no diving wave propagates, thus preventing us from getting some knowledge about the medium parameters from grazing angles. We explain 60 per cent of energy in vertical seismograms and 45 per cent of energy in horizontal seismograms.

9 THE INFLUENCE OF THE FREE SURFACE

Let us now introduce the free surface in our previous example. We choose the Born formulation since it is better suited for reflection acquisition (Beydoun & Tarantola 1988) as the previous example showed. Moreover, Keller (1969) and Woodward (1992) explained the Rytov formulation should be used only with one diffracted wave, in order to model phase shift due to one wave. When the free surface is present, surface and body diffracted waves contribute to data residuals.

The effect of the free surface is taken into account by the finite difference modelling of the vacuum zeroing elastic coefficients above the free surface while keeping a small density (Saenger *et al.* 2000). Stresses are zero on the free surface. The free surface has a stair-case geometry and we must include thirty points per wavelength in our simulations for accurate modelling (Bohlen & Saenger 2003; Gélis *et al.* 2005).

The reflection acquisition geometry remains the same as for the previous realistic example. Receivers are located at the free surface whereas sources are embedded at 1 m depth in order to properly emit downwards the energy inside the solid medium (Fig. 14a).

The presence of strong surface waves in seismograms may introduce a supplementary difficulty for the optimization scheme: surface waves result from constructive interferences of reflected/transmitted body waves at the free surface and decay exponentially with depth. One

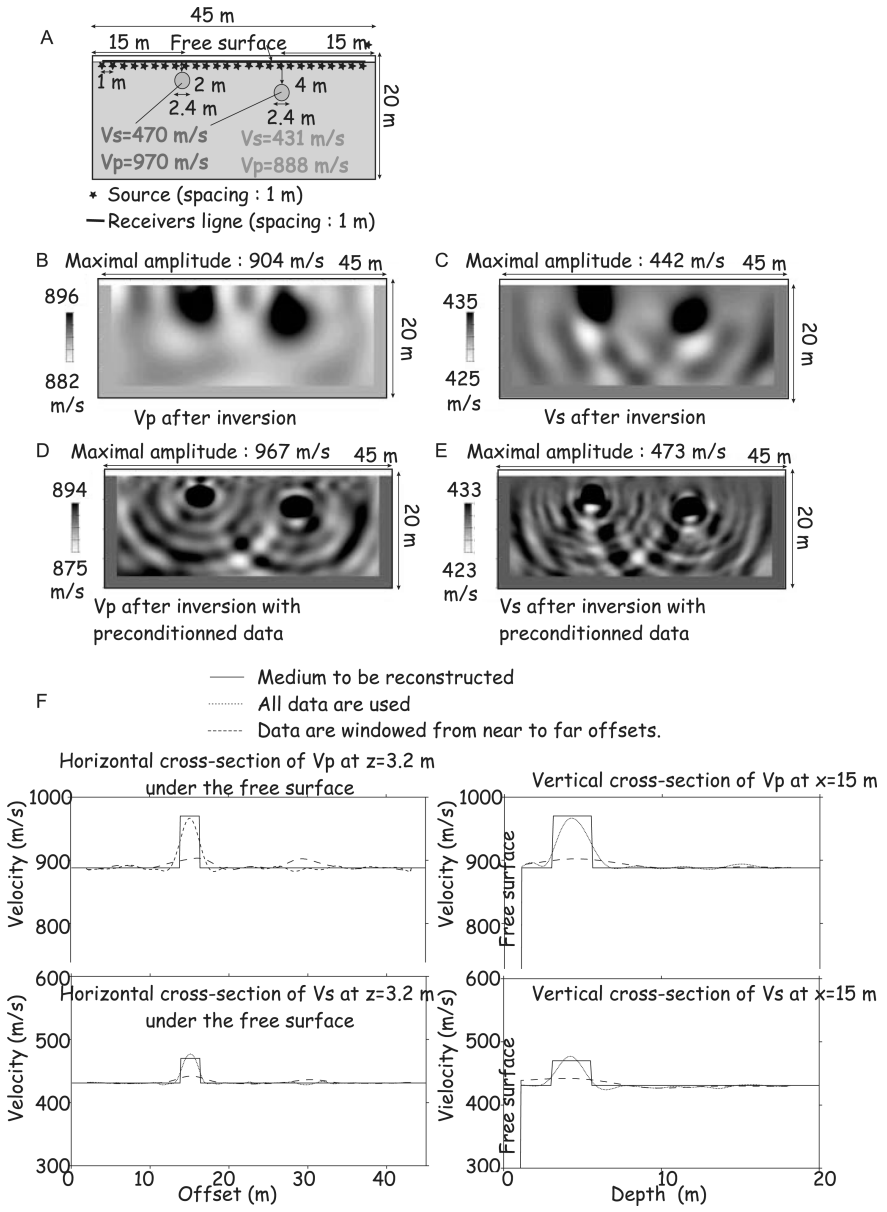

Figure 14. (a) Background, free surface and reflection acquisition geometry. (b) V_p parameter calculated with the Born formulation and the full data set. (c) V_s parameter calculated with the Born formulation and the full data set. (d) V_p parameter calculated with the Born formulation and progressively introduced data set. (e) V_s parameter calculated with the Born formulation and progressively introduced data set. Parameters maps colours are saturated to highlight reconstructed anomalies. (f) Curves show horizontal and vertical cross-sections of V_p and V_s parameters models. Velocity anomalies are better characterized when data are progressively introduced.

may wonder how velocity anomalies are extracted from surface waves because of their evanescent nature. We present here a synthetic example to understand how the inversion works with surface waves and how we could deal with them.

The frequency sampling remains the same as in the previous example, since its choice is based on the continuous wavenumber coverage. Let us point out that this frequency sampling is well suited for surface waves. The exponential decay of surface waves amplitude with depth makes the frequency choice for surface waves less dense, as we explain in Appendix C, for a reflection acquisition geometry.

By performing the inversion with the whole data set, V_p and V_s anomalies are correctly detected and located, but spurious features appear (Figs 14b and c). The comparison with the previous example without free surface indicates that this convergence towards a local minimum can be attributed to surface waves and to the difficulty to fit both body and surface waves. We may consider first body waves and introduce surface waves later on. This data pre-conditioning can be performed by successively performing inversion with different data sets. We first consider short-offset data containing mainly diffracted body waves and then, we use the final results as initial models to perform an inversion with higher offset data. Therefore, we progressively introduce higher-offset data during the inversion with more and more surface wave content. The entire inversion is performed through the frequency range with a pre-conditioned data before moving to the other selected data set. For

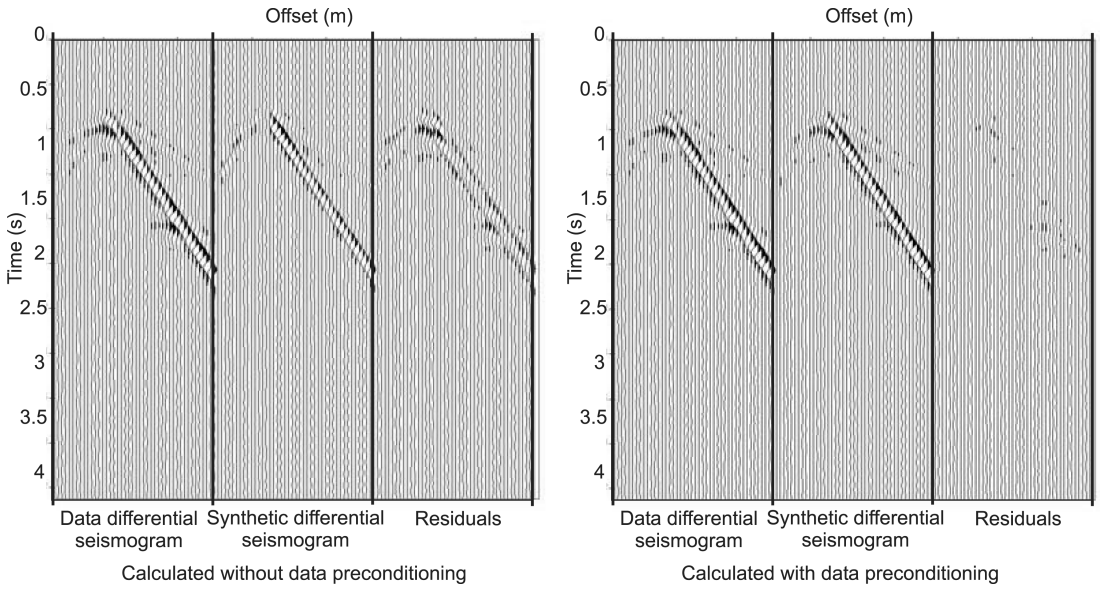

Figure 15. Comparison of data differential seismograms (difference between data to be inverted and data calculated in the initial model), synthetic differential seismograms (difference between data calculated in the final model and data calculated in the initial model) and residuals (difference between data differential seismograms and synthetic differential seismograms) calculated without and with data pre-conditioning for the horizontal component with the Born formulation. Data residuals are weaker when data are progressively introduced.

each inversion, the frequency choice has to be adapted to ensure the coverage of a continuous wavenumber spectrum. Therefore, the number of imaged frequencies is higher than with the whole data set and they are equal to 10, 15, 20, 35, 50, 90 and 130 Hz.

This near-to-far offset data selection is possible because the background model is known in this synthetic example and anomalies will focus at the right place. More realistic strategies could be envisioned although it is more complex in the frequency domain than in the time domain. An (ω, k) filtering will reduce effects of surface waves but it must be handled with care for erasing initially surface waves. Let us remind that the frequency increase during the inversion is also a data pre-conditioning.

The V_p and V_s parameter images show better reconstructed anomalies when surface waves are progressively introduced (Figs 14d and e) with respect to the case where body and surface waves are inverted at once (Figs 14b and c). Anomalie amplitudes have been increased. We explain 59 per cent of the vertical component when inverting directly the whole data set and 97 per cent with offset pre-conditioning. Similarly, we improve fitted amplitude from 48 per cent for the horizontal component to 91 per cent. When data are pre-conditioned, the energy left in residual seismograms is considerably reduced (Fig. 15). Because surface waves contain most of energy, they control the way optimization is performed: incorrect interpretation of these surface wave in diffracted body waves will drive the optimization to a local minimum.

10 DISCUSSION AND CONCLUSION

The optimization method

In this paper, we perform the elastic full waveform inversion following an iterative gradient optimization scheme. Other approaches may be applied to find the global minimum. First, conjugate gradient methods allow to reach the global minimum more quickly (Polak 1971; Luo & Schuster 1991). The Hessian matrix is as well considered as diagonal behaving (Zadler *et al.* 2004). The main advantage of this method is its faster convergence but final models remain in the same misfit function area (Molinari *et al.* 2002). In the Ray-Born approach (Jin *et al.* 1992), the Hessian diagonal approximation is counter-balanced by a local illumination coefficient depending on the considered point. Pratt *et al.* (1998) show that iterations in the gradient method allows to progressively introduce Hessian matrix out-of-diagonal terms, related to the data acquisition aperture and the limited frequency bandwidth. Shin *et al.* (2001a) use the virtual source concept developed by Pratt *et al.* (1998) and modify it to render the Hessian matrix more diagonal, considering that virtual sources are more uncorrelated than when using backpropagated residuals. However, this approach is more time-consuming since each virtual source must be calculated to construct the Hessian matrix. Plessix & Mulder (2004) show that, in the acoustic case, the diagonal approximation of the Hessian misleads the inversion engine. Moreover, they show that the non-diagonal elements size depends on the background velocity and on the depth. By including some non-diagonal terms in the diagonal Hessian matrix, medium imaging is improved. This could be a further improvement of our optimization scheme in the elastic case.

The main difficulty with gradient or Newton methods (Pratt *et al.* 1998) remains the error estimation and uncertainty assessment to define the quality of final parameter images.

The data and model spaces choice

In this study, we highlight the data and model spaces choice influence by considering Born and Rytov formulations and several inverted parameters couples. Model space explorations performed with Born and Rytov formulations differently converge to the global minimum. When the problem is ill-posed, for example with limited aperture acquisition geometries, different model exploration may lead to different local minima. On the contrary, similar results coming from different formulations indicate the good quality of the final parameter images. Moreover, the inverted parameters couple choice will influence the final parameter images by better recovering anomaly amplitudes or edges. This effect becomes more important when the number of inverted parameters increases. For example, if one aims at recovering seismic velocities and density, several couples of physical parameters may be chosen as Forgues (1996) shows and final images may differ. In order to choose parameters couples, diffraction patterns must be considered. Kormendi & Dietrich (1991) find that seismic impedance inversion gives better results than seismic velocity inversion.

Data pre-conditioning

Another way to modify the misfit function or better control its gradient direction concerns the data pre-conditioning. In this study, we perform two kinds of data selection by isolating only near-offset data or by ignoring out-of-phase data with the Rytov formulation. This data discrimination is closely related to the L2 norm choice, very sensitive to few outliers. Operto *et al.* (2004) strongly weights far-offset data in their inversion scheme to first reconstruct background parameters. Our data pre-conditioning with offset represents an opposite case of this offset weighting when the long wavelenghts of the background are known. In the elastic case, it could also be possible to weight horizontal versus vertical data. This should be studied in details in another paper. Tuning the inversion with selected data is a key point to avoid local minima and better converge towards the global one. The weighting choice depends on the background knowledge and the type of information we first aim to recover.

The surface waves modelling

In this approach, we do not discriminate surface waves from body waves in the inversion. The frequency modelling allows to take into account all waves propagating at a given frequency. In practice, this may give rise to some difficulties since surface waves are stronger than body waves in recorded seismograms and contain lower frequencies. As the inversion begins with low frequencies and progressively considers higher frequencies, surface waves normally first control the inversion. Methods using surface waves have been developed by Campman *et al.* (2004) and Herman *et al.* (2000) to eliminate scattered surface waves masking waves coming from deeper areas. They construct an effective medium containing anomalies explaining observed scattered waves but they do not search to finely image these anomalies. This method allows to roughly detect scatterers and may be used in our approach as first indicator of medium anomalies. This method allows to construct a less smooth background than the SASW (Nazarian & Stokoe 1984) and/or the traveltime tomography (Improta *et al.* 2002).

Combining surface waves and body waves is still a challenge because of their different physical behaviour and nature. We present here an efficient way to discriminate between body and surface waves and to take information from body waves first. Other data pre-conditioning may be used to reach this purpose. A more detailed study of the surface waves influence on the inversion scheme must now be done in another forthcoming paper, to introduce surface waves in the optimal way.

Conclusion

In this paper, we have performed the elastic full waveform inversion following a gradient optimization method. The forward problem allows to model all waves propagation in complex media and to properly simulate the free surface. We highlight that the reconstruction of several parameters is possible as long as data allow to extract information. This means that we need to know the background medium and that this inverse scheme works for backscattering as long as macromodel is available. Moreover, we highlight problems that may be encountered with surface waves. Their evanescent nature may introduce instabilities when depth increases but body waves help collaborative recontruction as soon as we discriminate data. The next step of this study will be to precisely understand how to introduce them in the inversion in the optimal way.

ACKNOWLEDGMENTS

We thank P. R. Amestoy and I. S. Duff and J.-Y. L'Excellent for allowing us to use the MUMPS free software (web page: http://www.enseeiht.fr/lima/apo/MUMPS/). We greatly thank S. Operto for his very useful comments and precise review. Moreover, we greatly thank René-Edouard Plessix and Gerhard Pratt for their detailed and precise helpful comments and their relevant and constructive review. This work has been supported by BRGM through a contract SG/DJVB02-188.

REFERENCES

Abramowitz, M. & Stegun, I., 1965. *Handbook of Mathematical Functions,* Dover, New York.

Aki, K. & Richards, P., 2002. *Quantitative Seismology, Theory and Methods,* 2nd edn, University Science Books, Sausalito, California.

Amestoy, P.R., Duff, I.S. & L'Excellent, J.-Y., 2000. Multifrontal parallel distributed symmetric and unsymmetric solvers, *Comput. Methods Appl. Mech. Eng.,* **184,** 501–520.

Amestoy, P.R., Duff, I.S., Koster, J. & L'Excellent, J.-Y., 2001. A fully asynchronous multifrontal solver using distributed dynamic scheduling, *SIAM J. Matrix Anal. Appl.,* **23**(1), 15–41.

Bérenger, J., 1994. A perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Geophys.,* **114,** 185–200.

Beydoun, W. & Mendes, M., 1989. Elastic Ray-Born l^2 migration-inversion, *Geophys. J.,* **97,** 151–160.

Beydoun, W. & Tarantola, A., 1988. First Born and Rytov approximations: modelling and inversion conditions in a canonical example, *J. acoust. Soc. Am.,* **83,** 1045–1055.

Beylkin, G., 1985. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized randon transform, *J. Math. Phys.,* **26**(1), 99–108.

Beylkin, G. & Burridge, R., 1990. Linearized inverse scattering problems in acoustics and elasticity, *Wave Motion,* **12,** 15–52.

Bleistein, N., 1987. On the imaging of reflectors in the earth, *Geophysics,* **52,** 931–942.

Bohlen, T. & Saenger, E., 2003. 3D visco elastic finite difference modeling using the rotated staggered grid—tests of numerical accuracy, in *EAGE 65th Conference and Exhibition,* Stavanger, Norway.

Campman, X., van Wijk, K., Riyanti, C.D., Scales, J.A. & Herman, G., 2004. Imaging scattered seismic surface waves, *Near Surf. Geophys.,* **2**(4), 223–230.

Chapman, C., 2004. *Fundamentals of Seismic Waves Propagation,* Cambridge Press, Cambridge, England.

Clayton, R. & Stolt, R., 1981. A Born-WKBJ inversion method for acoustic reflexion data, *Geophysics,* **46,** 1559–1565.

Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic non-linear waveform inversion: application to real data, *Geophysics,* **55,** 527–538.

Dahlen, F. & Baig, A., 2002. Fréchet kernels for body waves amplitudes, *Geophys. J. Int.,* **150,** 440–466.

Dahlen, F., Hung, S. & Nolet, G., 2000. Fréchet kernels for finite-difference travaltimes—I. Theory, *Geophys. J. Int.,* **141,** 157–174.

Debayle, E. & Kennett, B., 2000a. The australian continental upper mantle: structure and deformation inferred from surface waves, *J. geophys. Res.,* **105**(11), 25 423–25 450.

Debayle, E. & Kennett, B., 2000b. Anisotropy in the australian upper mantle from love and rayleigh waveform inversion, *Earth planet. Sci. Lett.,* **184,** 339–351.

Dessa, J., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J. & Kaneda, Y., 2004. Multiscale seismic imaging of the eastern nankai trough by full waveform inversion, *Geophys. Res. Lett.,* **31,** L18606.

Devaney, A., 1981. Inverse scattering theory within the rytov approximation, *Opt. Lett.,* **6,** 374–376.

Forgues, E., 1996. Inversion linéraisée multiparamètres via la théorie des rais, *PhD thesis*, Université Paris VII.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional non-linear inversion of seismic waveforms: numerical results, *Geophysics,* **51,** 1387–1403.

Gélis, C., Leparoux, D., Virieux, J., Bitri, A., Operto, S. & Grandjean, G., 2005. Numerical modeling of surface waves over shallow cavities, *J. Environ. Eng. Geophys.,* **10,** 49–59.

Gorbatov, A. & Kennett, B., 2003. Joint bulk-sound and shear tomography for western pacific subduction zones, *Earth planet. Sci. Lett.,* **210,** 527–543.

Graves, R., 1996. Simulating seismic waves propagation in 3D elastic media using staggered-grid finite differences, *J. acoust. Soc. Am.,* **100,** 3061–3069.

Hastings, F., Scheider, J. & Broschat, S., 1996. Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation, *J. acoust. Soc. Am.,* **100,** 3061–3069.

Hayashi, K., Burns, D. & Toksz, M., 2001. Discontinuous-grid finite-difference seismic modelling including surface topography, *Bull. seism. Soc. Am.,* **91**(6), 1750–1764.

Herman, G., Milligan, P., Huggins, R. & Rector, J., 2000. Imaging shallow objects and heterogeneities with scattered guided waves, *Geophysics,* **1,** 247–252.

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

Hustedt, B., Operto, S. & Virieux, J., 2004. Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling, *Geophys. J. Int.,* **157,** 1269–1296.

Improta, L., Zollo, A., Herrero, A., Frattini, R., Virieux, J. & Dell'Aversana, P., 2002. Seismic imaging of complex structures by non-linear traveltimes inversion of dense wide-angle data: application to a thrust belt, *Geophys. J. Int.,* **151,** 264–278.

Jin, S., 1992. Inversion de données de sismique pétrolière : séparation des paramètres élastiques et détermination des vitesses de référence, *PhD thesis*, Institut de Géophysique du Globe de Paris.

Jin, S., Madariaga, R., Virieux, J. & Lambaré, G., 1992. Two-dimensional asymptotic iterative elastic inversion, *Geophys. J. Int.,* **108,** 575–588.

Keller, J., 1969. Acuracy and validity of the Born and Rytov approximations, *J. Opt. Soc. Am.,* **59,** 1003–1004.

Kolb, P., Collino, F. & Lailly, P., 1986. Prestack inversion of 1D medium, *Proc. IEEE,* **74,** 498–508.

Kormendi, F. & Dietrich, M., 1991. Non-linear waveform inversion of plane-wave seismograms in stratified elastic media, *Geophysics,* **56**(5), 664–674.

Lailly, P., 1984. *The seismic inverse problem as a sequence of before stack migrations, in: Inverse Scattering Theory and Applications,* SIAM, Philadelphia.

Lambaré, G., Virieux, J., Madariaga, R. & Jin, S., 1992. Iterative asymptotic inversion in the acoustic approximation, *Geophysics,* **57,** 1138–1154.

Lambaré, G., Operto, S., Podvin, P. & Thierry, P., 2003. 3-D ray-Born migration/inversion: Theory, *Geophysics,* **68**(4), 1348–1356.

Leparoux, D., Bitri, A. & Grandjean, G., 2002. Détection de cavités souterraines par ondes sismiques de surface le long de lignes TGV. expérimentation sur un site test, in: *Journées AGAP,* LCPC, Nantes.

Luo, Y. & Schuster, G., 1990. Parsimonious staggered grid finite-differencing of the wave equation, *Geophys. Res. Lett.,* **17**(2), 155–158.

Luo, Y. & Schuster, G., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), 645–653.

Molinari, M., Cox, S., Blott, B. & Daniell, G., 2002. Comparison of algorithms for non-linear inverse 3D electrical tomography reconstruction, *Physiol. Meas.,* **23,** 95–104.

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics,* **52,** 1211–1228.

Mora, P., 1988. Elastic wavefield inversion of reflection and transmission data, *Geophysics,* **53,** 750–759.

Nazarian, S. & Stokoe, K., 1984. In situ shear wave velocities from spectral analysis of surface waves, *Proc. 8th Conf. on Earthquake Eng., San Francisco,* **3,** 31–38.

Nolet, G., 1987. *Seismic tomography, Seismic wave Propagation and Seismic Tomography,* pp. 1–23, ed. Nolet, G., Reidel, Dordrecht.

Ohminato, T. & Chouet, B., 1997. A free-surface boundary condition for including 3D topography in the finite-difference method, *Bull. seism. Soc. Am.,* **87**(2), 494–515.

Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A. & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from multi-fold wide aperture seismic data, *Geophys. Prospect.,* **52,** 625–651.

Plessix, R. & Mulder, W., 2004. Frequency domain finite difference amplitude preserving migration, *Geophys. J. Int.,* **157,** 975–987.

Polak, E., 1971. *Computational Methods in Optimization,* Academic Press, New York.

Pratt, R., 1990. Inverse theory applied to multi-source cross-hole tomography. Part II: elastic wave-equation method, *Geophys. Prospect.,* **38,** 311–330.

Pratt, R., 1999. Seismic waveform inversion in the frequency domain, Part I: theory and verifi- cation in a physic scale model, *Geophys. J. Int.,* **64,** 888–901.

Pratt, R. & Worthington, M., 1988. The application of diffraction tomography to cross-hole seismic data, *Geophysics,* **53**(10), 1284–1294.

Pratt, R., Song, Z., Williamson, P. & Warner, M., 1996. Two-dimensional velocity model from wide-angle seismic data by wavefield inversion, *Geophys. J. Int.,* **124,** 323–340.

Pratt, R., Shin, C. & Hicks, G., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A. & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from multifold wide angle seismic data by frequency domain full waveform inversion: application to a thrust belt, *Geophys. J. Int.,* **159,** 1032–1056.

Ribodetti, A. & Virieux, J., 1998. Asymptotic theory for imaging the attenuation factor Q, *Geophysics,* **31**(5), 1767–1778.

Saenger, E. & Bohlen, T., 2004. Finite difference modeling of viscoelastic and anisotropic wave propagation using the rotated staggered grid, *Geophysics,* **69**(2), 583–591.

Saenger, E., Gold, N. & Shapiro, S., 2000. Modeling the propagation of elastic waves using a modified finite-difference grid, *Wave Motion,* **31,** 77–92.

Sheng, J., 2004. High Resolution Seismic Tomography With the Generalized Radon Transform and Early Arrival Waveform Inversion, *PhD thesis*, University of Utah.

Shin, C., Jang, S. & Min, D., 2001a. Improved amplitude preservation for prestack depth migration by inverse scattering theory, *Geophys. Prospect.,* **49,** 592–606.

Shin, C., Yoon, K., Marfurt, K., Park, K., Yang, D., Lim, H., Chung, S. & Shin, S., 2001b. Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion, *Geophysics,* **66**(6), 1856–1863.

Shipp, R. & Singh, S., 2002. Two-dimensional full waveform inversion of wide-aperture marine streamer data, *Geophys. J. Int.,* **151,** 325–344.

Sirgue, L., 2003. Inversion de la forme d'onde dans le domaine fréquentiel de données sismiques grands déports. Thèse de doctorat. Sciences de la Terre, *PhD thesis*, Laboratoire de géologie de l'Ecole Normale Supéerieure. UniversitéParis XI.

Sirgue, L. & Pratt, R., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics,* **69**(1), 231–248.

Snieder, R., 1986. 3D linearized scattering of surface waves and a formalism for surface waves holography, *Geophys. J. R. astr. Soc.,* **84,** 581–601.

Spetzler, J. & Snieder, R., 2004. The fresnel volume and transmitted waves, *Geophysics,* **69**(3), 653–663.

Stekl, I. & Pratt, R., 1998. Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators, *Geophysics,* **63,** 1779–1794.

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

Tarantola, A., 1987. *Inverse Problem Theory,* Elsevier Sci, New York.

Tarantola, A., 1988. Theoritical background for the inversion of seismic waveforms including elasticity and attenuation, *Pageoph,* **128,** 365–399.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity stress finitedifference method, *Geophysics,* **51,** 889–901.

Woodward, M., 1992. Wave-equation tomography, *Geophysics,* **57,** 15–26.

Wu, R., 2003. Wave propagation, scattering and imaging using dual-domain one-way and onereturn propagation, *Pure appl. Geophys.,* **160,** 509–539.

Zadler, B., Le Rousseau, J., Scales, J. & Smith, M., 2004. Resonant ultrasound spectroscopy: theory and application, *Geophys. J. Int.,* **156,** 154–169.

APPENDIX A: BORN KERNEL COMPUTATION

We follow the approach described in many textbooks (Tarantola 1987; Chapman 2004) and linearly link the Green function perturbations to the parameters perturbations, allowing to define the so-called Born kernel **K**. For any 2-D medium, the perturbation of the Green function component along the direction q at the receiver position $\mathbf{r}$ for a point source $\mathbf{s}$ applied along the direction p is given for the first-order Born formulation, also denoted Born formulation, by the expression

$$\begin{aligned}\delta G_{pq}(\mathbf{r},\omega,\mathbf{s}) &= \int_M \left\{ \omega^2 \delta\rho(\mathbf{x}) G^0_{ip}(\mathbf{x},\omega,\mathbf{r}) G^0_{iq}(\mathbf{x},\omega,\mathbf{s}) \right. \\ &\quad \left. - \delta C_{ijkl}(\mathbf{x}) \frac{\partial G^0_{ip}(\mathbf{x},\omega,\mathbf{r})}{\partial x_j} \frac{\partial G^0_{kq}(\mathbf{x},\omega,\mathbf{s})}{\partial x_l} \right\} d\mathbf{x} \\ &= -\int_M \mathbf{K_{pq}}(\mathbf{r},\mathbf{x},\mathbf{s},\omega)\, d\mathbf{x}, \end{aligned} \tag{A1}$$

in which the implicit Einstein convention for summation is used, where M is the set of all points in the medium that may be considered as scatterers and where

$$\begin{aligned}\rho(\mathbf{x}) &= \rho^0(\mathbf{x}) + \delta\rho(\mathbf{x}) \\ C_{ijkl}(\mathbf{x}) &= C^0_{ijkl}(\mathbf{x}) + \delta C_{ijkl}(\mathbf{x}). \end{aligned} \tag{A2}$$

For an isotropic linear elastic medium, the fourth-order elastic tensor $C_{ijkl}(\mathbf{x})$ becomes

$$C_{ijkl}(\mathbf{x}) = \lambda(\mathbf{x})\delta_{ij}\delta_{kl} + \mu(\mathbf{x})(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{A3}$$

where λ and μ are Lamé parameters.

This expression, which has been obtained using the reciprocity theorem, allows estimation of Fréchet derivatives with respect to velocities. We neglect here derivatives with respect to the ρ parameter which is kept constant throughout this study.

As an explicit example, the Green function $\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})$ can be expressed with respect to V_p and V_s physical parameters perturbations δV_p and δV_s. One may write

$$\begin{aligned}\delta G_{zz}(\mathbf{r},\omega,\mathbf{s}) = \int_M \Big\{ & \left[2\rho^0(\mathbf{x})V_p^0(\mathbf{x})\delta V_p^0(\mathbf{x}) - 4\rho^0(\mathbf{x})V_s^0(\mathbf{x})\delta V_s^0(\mathbf{x})\right]\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ & + \left[2\rho^0(\mathbf{x})V_p^0(\mathbf{x})\delta V_p^0(\mathbf{x}) - 4\rho^0(\mathbf{x})V_s^0(\mathbf{x})\delta V_s^0(\mathbf{x})\right]\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} \\ & + 2\rho^0(\mathbf{x})V_s^0(\mathbf{x})\delta V_s^0(\mathbf{x})\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ & + 2\rho^0(\mathbf{x})V_s^0(\mathbf{x})\delta V_s^0(\mathbf{x})\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} \\ & + 2\rho^0(\mathbf{x})V_s^0(\mathbf{x})\delta V_s^0(\mathbf{x})\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} \\ & + 2\rho^0(\mathbf{x})V_s^0(\mathbf{x})\delta V_s^0(\mathbf{x})\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ & + 2\rho^0(\mathbf{x})V_p^0(\mathbf{x})\delta V_p^0(\mathbf{x})\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} \\ & + 2\rho^0(\mathbf{x})V_p^0(\mathbf{x})\delta V_p^0(\mathbf{x})\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}\Big\} d\mathbf{x}. \end{aligned} \tag{A4}$$

The Born Kernel defined in eq. (A1) comes from the contribution of the different spatial Green functions perturbations with respect to the perturbations of one physical parameter, such as V_p ou V_s. These Green functions perturbations are then multiplied with conjugate velocity residuals in order to get the misfit function gradient (eq. 11). Moreover, in order to take the real source term into account, one may write

$$\begin{aligned}\Re\left[\mathcal{B}_0'\Delta\mathbf{d}^*\right]^{V_p} = & S_z^f\frac{\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})}\Delta V_z^*(\omega,\mathbf{r}) + S_x^f\frac{\delta G_{zx}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})}\Delta V_z^*(\omega,\mathbf{r}) \\ & + S_z^f\frac{\delta G_{xz}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})}\Delta V_x^*(\omega,\mathbf{r}) + S_x^f\frac{\delta G_{xx}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})}\Delta V_z^*(\omega,\mathbf{r}), \end{aligned} \tag{A5}$$

which is physically similar to eq. (17) expressed with discrete fields.

To go on with the $\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})$ example, we link the Green function perturbations $\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})$ to δV_p and δV_s physical parameters perturbations. Thanks to eq. (A4), the expression $\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})/\delta V_p(\mathbf{x})$ is equal to:

$$\begin{aligned}\frac{\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})} = & -2V_p^0(\mathbf{x})\left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x} + \frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\right\} \\ & \left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} + \frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}\right\} \end{aligned} \tag{A6}$$

Similarly,

$$\begin{aligned}\frac{\delta G_{zz}(\mathbf{r},\omega,\mathbf{s})}{\delta V_s(\mathbf{x})} = & -2V_s^0(\mathbf{x})\left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z} + \frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\right\} \\ & \left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} + \frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}\right\} \\ & - 2\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ & - 2\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}. \end{aligned} \tag{A7}$$

Other components of Born gradient function could be similarly estimated for V_p and V_s giving following expressions

$$\begin{aligned}\frac{\delta G_{xx}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})} = & -2V_p^0(\mathbf{x})\left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x} + \frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\right\} \\ & \left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} + \frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}\right\}, \end{aligned} \tag{A8}$$

$$\begin{aligned}\frac{\delta G_{xz}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})} = & -2V_p^0(\mathbf{x})\left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x} + \frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\right\} \\ & \left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x} + \frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}\right\}, \end{aligned} \tag{A9}$$

$$\frac{\delta G_{zx}(\mathbf{r},\omega,\mathbf{s})}{\delta V_p(\mathbf{x})} = -2V_p^0(\mathbf{x})\left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}+\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\right\} \left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}+\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}\right\}, \tag{A10}$$

$$\begin{aligned}\frac{\delta G_{xx}(\mathbf{r},\omega,\mathbf{s})}{\delta V_s(\mathbf{x})} = &-2V_s^0(\mathbf{x})\left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}+\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\right\} \left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}+\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}\right\} \\ &-2\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ &-2\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x},\end{aligned} \tag{A11}$$

$$\begin{aligned}\frac{\delta G_{xz}(\mathbf{r},\omega,\mathbf{s})}{\delta V_s(\mathbf{x})} = &-2V_s^0(\mathbf{x})\left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}+\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\right\} \left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}+\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}\right\} \\ &-2\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ &-2\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x},\end{aligned} \tag{A12}$$

$$\begin{aligned}\frac{\delta G_{zx}(\mathbf{r},\omega,\mathbf{s})}{\delta V_s(\mathbf{x})} = &-2V_s^0(\mathbf{x})\left\{\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}+\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\right\} \left\{\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z}+\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}\right\} \\ &-2\frac{\partial G_{xz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial x}\frac{\partial G_{zx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial z} \\ &-2\frac{\partial G_{zz}^0(\mathbf{x},\omega,\mathbf{r})}{\partial z}\frac{\partial G_{xx}^0(\mathbf{x},\omega,\mathbf{s})}{\partial x}.\end{aligned} \tag{A13}$$

By inserting eqs (A6), (A8), (A9) and (A10) into eq. (A5), we obtain the continuous expression of the gradient misfit function for the V_p parameter. When this equation is discretized thanks to the finite difference stencil used throughout this study, we obtain term by term the discrete eq. (25).

When computing the gradient misfit function for the V_s parameter, the same final expression is obtained with the discretization of analytical expressions such as (A5) and with the use of the Pratt *et al.* (1998) formalism where discrete fields are always considered (equations such as 25).

APPENDIX B: ELASTIC RYTOV ANALYTICAL FORMULATION

We first recall that equations governing the waves propagation in a homogeneous elastic medium defined by the homogeneous density $\rho^0(\mathbf{x})$ and the fourth-order elastic tensor $C_{ijkl}^0(\mathbf{x})$ are (Tarantola 1987; Chapman 2004):

$$-\omega^2\rho^0(\mathbf{x})V_i^0(\mathbf{x},\omega)-\frac{\partial}{\partial x_j}\left[C_{ijkl}^0(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}=S_i(\mathbf{x},\omega)\right]. \tag{B1}$$

We then consider weak heterogeneities (scatters) in the medium that perturb waves through eq. (A5). Therefore, eq. (B1) can be written as

$$\begin{aligned}&-\omega^2\rho^0(\mathbf{x})V_i(\mathbf{x},\omega)-\frac{\partial}{\partial x_j}\left[C_{ijkl}^0(\mathbf{x})\frac{\partial V_k(\mathbf{x},\omega)}{\partial x_l}\right] \\ &\quad = S_i(\mathbf{x},\omega)+\omega^2\delta\rho(\mathbf{x})V_i(\mathbf{x},\omega)+\frac{\partial}{\partial x_j}\left[\delta C_{ijkl}(\mathbf{x})\frac{\partial V_k(\mathbf{x},\omega)}{\partial x_l}\right].\end{aligned} \tag{B2}$$

Because of the scatters presence, propagating fields $V_i(\mathbf{x},\omega)$ are perturbed and are shifted with respect to referenced propagating fields $V_i^0(\mathbf{x},\omega)$ in the homogeneous medium through

$$\begin{aligned}V_i(\mathbf{x},\omega) &= V_i^0(\mathbf{x},\omega)e^{\delta\psi_i(\mathbf{x},\omega)} \\ V_k(\mathbf{x},\omega) &= V_k^0(\mathbf{x},\omega)e^{\delta\psi_k(\mathbf{x},\omega)}.\end{aligned} \tag{B3}$$

Therefore, the left hand side of eq. (B2) equals to

$$
\begin{aligned}
&-\omega^2\rho^0(\mathbf{x})V_i(\mathbf{x},\omega)-\frac{\partial}{\partial x_j}\left[C^0_{ijkl}(\mathbf{x})\frac{\partial V_k(\mathbf{x},\omega)}{\partial x_l}\right]\\
&\quad=\Bigg[-\omega^2\rho^0(\mathbf{x})V_i^0(\mathbf{x},\omega)e^{\delta\psi_i(\mathbf{x},\omega)-\delta\psi_k(\mathbf{x},\omega)}-\frac{\partial C^0_{ijkl}(\mathbf{x})}{\partial x_j}\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}-C^0_{ijkl}(\mathbf{x})\frac{\partial^2 V_k^0(\mathbf{x},\omega)}{\partial x_k\partial x_l}-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_j}\\
&\quad-\frac{\partial C^0_{ijkl}(\mathbf{x})}{\partial x_j}V_k^0(\mathbf{x},\omega)\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_l}-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_j}\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_l}-C^0_{ijkl}(\mathbf{x})V_k^0(\mathbf{x},\omega)\frac{\partial^2\delta\psi_k(\mathbf{x},\omega)}{\partial x_k\partial x_l}\Bigg]e^{\delta\psi_k(\mathbf{x},\omega)}.
\end{aligned}
\tag{B4}
$$

We perform a first-order Taylor–Lagrange development of $e^{\delta\psi_i(\mathbf{x},\omega)}$ and $e^{\delta\psi_i(\mathbf{x},\omega)-\delta\psi_k(\mathbf{x},\omega)}$ since $\delta\psi_i(\mathbf{x},\omega)$ and $\delta\psi_i(\mathbf{x},\omega)-\delta\psi_k(\mathbf{x},\omega)$ are considered as small perturbations. Therefore, we approximate these expressions using $e^x=1+\mathbf{x}$. We develop eq. (B1) and elimine common terms which gives us the following expression

$$
\begin{aligned}
&-\omega^2\rho^0(\mathbf{x})V_i(\mathbf{x},\omega)-\frac{\partial}{\partial x_j}\left[C^0_{ijkl}(\mathbf{x})\frac{\partial V_k(\mathbf{x},\omega)}{\partial x_l}\right]=-\omega^2\rho^0(\mathbf{x})V_i^0(\mathbf{x},\omega)[\delta\psi_i(\mathbf{x},\omega)-\delta\psi_k(\mathbf{x},\omega)]\\
&\quad-\frac{\partial C^0_{ijkl}(\mathbf{x})}{\partial x_j}\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}\delta\psi_k(\mathbf{x},\omega)-C^0_{ijkl}(\mathbf{x})\frac{\partial^2 V_k^0(\mathbf{x},\omega)}{\partial x_l\partial x_j}\delta\psi_k(\mathbf{x},\omega)\\
&\quad-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}\frac{\partial[\delta\psi_k(\mathbf{x},\omega)]}{\partial x_j}[1+\delta\psi_k(\mathbf{x},\omega)]-\frac{\partial C^0_{ijkl}(\mathbf{x})}{\partial x_j}V_k^0(\mathbf{x},\omega)\frac{\partial[\delta\psi_k(\mathbf{x},\omega)]}{\partial x_l}[1+\delta\psi_k(\mathbf{x},\omega)]\\
&\quad-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_j}\frac{\partial[\delta\psi_k(\mathbf{x},\omega)]}{\partial x_l}[1+\delta\psi_k(\mathbf{x},\omega)]-C^0_{ijkl}(\mathbf{x})V_k^0(\mathbf{x},\omega)\frac{\partial^2\delta\psi_k(\mathbf{x},\omega)}{\partial x_l\partial x_j}[1+\delta\psi_k(\mathbf{x},\omega)]
\end{aligned}
\tag{B5}
$$

The last expression is equal to $\delta\rho(\mathbf{x})V_i(\mathbf{x},\omega)+\frac{\partial}{\partial x_j}(\delta C_{ijkl}(\mathbf{x})\frac{\partial V_k(\mathbf{x},\omega)}{\partial x_l})+S_i(\mathbf{x},\omega)$ since it is the right hand side of eq. (B2). Since

$$
\begin{aligned}
&\delta\rho(\mathbf{x})V_i(\mathbf{x},\omega)+\frac{\partial}{\partial x_j}\left(\delta C_{ijkl}(\mathbf{x})\frac{\partial V_k(\mathbf{x},\omega)}{\partial x_l}\right)+S_i(\mathbf{x},\omega)\\
&\quad=\delta\rho(\mathbf{x})V_i^0(\mathbf{x},\omega)[1+\delta\psi_i(\mathbf{x},\omega)]+\frac{\partial}{\partial x_j}\left(\delta C_{ijkl}(\mathbf{x})\frac{\partial\{V_k^0(\mathbf{x},\omega)[1+\delta\psi_k(\mathbf{x},\omega)]\}}{\partial x_l}\right)+S_i(\mathbf{x},\omega),
\end{aligned}
\tag{B6}
$$

we get the expression

$$
\begin{aligned}
&-\omega^2\rho^0(\mathbf{x})V_i^0(\mathbf{x},\omega)\delta\psi_i(\mathbf{x},\omega)-\frac{\partial}{\partial x_j}\left(C^0_{ijkl}(\mathbf{x})\frac{\partial[V_k^0(\mathbf{x},\omega)\delta\psi_k(\mathbf{x},\omega)]}{\partial x_l}\right)=\omega^2\rho^0(\mathbf{x})V_i^0(\mathbf{x},\omega)\delta\psi_k(\mathbf{x},\omega)\\
&\quad-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_j}\delta\psi_k(\mathbf{x},\omega)-\frac{\partial C^0_{ijkl}(\mathbf{x})}{\partial x_j}V_k^0(\mathbf{x},\omega)\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_l}\delta\psi_k(\mathbf{x},\omega)\\
&\quad-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_j}\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_l}\delta\psi_k(\mathbf{x},\omega)-C^0_{ijkl}(\mathbf{x})V_k^0(\mathbf{x},\omega)\frac{\partial^2\delta\psi_k(\mathbf{x},\omega)}{\partial x_l\partial x_j}\delta\psi_k(\mathbf{x},\omega)\\
&\quad+\delta\rho(\mathbf{x})V_i^0(\mathbf{x},\omega)[1+\delta\psi_i(\mathbf{x},\omega)]+\frac{\partial}{\partial x_j}\left(\delta C_{ijkl}(\mathbf{x})\frac{\partial\{V_k^0(\mathbf{x},\omega)[1+\delta\psi_k(\mathbf{x},\omega)]\}}{\partial x_l}\right).
\end{aligned}
\tag{B7}
$$

The right hand side of the equation is the source term $S_c^i(\mathbf{x},\omega)$ corresponding to scatters effects. The propagating field is now $V_i^0(\mathbf{x},\omega)\,\delta\,\psi_i(\mathbf{x},\omega)$ instead of $V_i^0(\mathbf{x},\omega)\,e^{\delta\psi_i(\mathbf{x},\omega)}$ as in eq. (B1). Following the classical scattering field formulation, we write

$$
V_n^0(\mathbf{r},\omega)\delta\psi_n(\mathbf{r},\omega)=\int_M G^0_{ni}(\mathbf{r},\mathbf{x})S_c^i(\mathbf{x},\omega)\,d\mathbf{x}.
\tag{B8}
$$

By replacing $S_c^i(\mathbf{x},\omega)$ with its value, we finally obtain

$$
\begin{aligned}
\delta\psi_n(\mathbf{r})V_n^0(\mathbf{r},\omega)=\int_M\Bigg(&\omega^2\rho^0(\mathbf{x})V_i^0(\mathbf{x},\omega)\delta\psi_k(\mathbf{x},\omega)-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_j}\delta\psi_k(\mathbf{x},\omega)-\frac{\partial C^0_{ijkl}(\mathbf{x})}{\partial x_j}V_k^0(\mathbf{x},\omega)\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_l}\delta\psi_k(\mathbf{x},\omega)\\
&-C^0_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_j}\frac{\partial\delta\psi_k(\mathbf{x},\omega)}{\partial x_l}\delta\psi_k(\mathbf{x},\omega)-C^0_{ijkl}(\mathbf{x})V_k^0(\mathbf{x},\omega)\frac{\partial^2\delta\psi_k(\mathbf{x})}{\partial x_l\partial x_j}\delta\psi_k(\mathbf{x},\omega)\Bigg)+\delta\rho(\mathbf{x})V_i^0(\mathbf{x},\omega)[1+\delta\psi_i(\mathbf{x},\omega)]\\
&+\frac{\partial}{\partial x_j}\left[\delta C_{ijkl}(\mathbf{x})\frac{\partial\left\{V_k^0(\mathbf{x},\omega)[1+\delta\psi_k(\mathbf{x},\omega)]\right\}}{\partial x_l}\right]G^0_{ni}(\mathbf{r},\mathbf{x})\,d\mathbf{x}.
\end{aligned}
\tag{B9}
$$

By neglecting multiple scattering terms as for the Born formulation, performing integration by part for the second remaining term of the right-hand side and considering no perturbation on the model edges, we finally get the perturbed phase

$$
\delta\psi_n(\mathbf{r})V_n^0(\mathbf{r},\omega)=\int_M\left[\omega^2\delta\rho(\mathbf{x})V_i^0(\mathbf{x},\omega)G^0_{ni}(\mathbf{r},\omega,\mathbf{x})-\frac{\partial G^0_{ni}(\mathbf{r},\omega,\mathbf{x})}{\partial x_j}\delta C_{ijkl}(\mathbf{x})\frac{\partial V_k^0(\mathbf{x},\omega)}{\partial x_l}\right]d\mathbf{x}.
\tag{B10}
$$

Following the Born formulation approach, we factorize the source term $S_q^f(\mathbf{r},\omega)$ in the previous equation and, therefore, we obtain

$$
\delta\psi_{nq}(\mathbf{x})G^0_{nq}(\mathbf{r},\omega,\mathbf{s})=\int_M\left[\omega^2\delta\rho(\mathbf{x})G^0_{iq}(\mathbf{x},\omega,\mathbf{s})G^0_{ni}(\mathbf{r},\mathbf{x})-\frac{\partial G^0_{ni}(\mathbf{r},\omega,\mathbf{x})}{\partial x_j}\delta C_{ijkl}(\mathbf{x})\frac{\partial G^0_{kq}(\mathbf{x},\omega,\mathbf{s})}{\partial x_l}\right]d\mathbf{x}.
\tag{B11}
$$

Dividing by the reference Green function $G^0_{nq}(\mathbf{r}, \omega, \mathbf{s})$, we obtain at last

$$\delta\psi_{nq}(\mathbf{x}) = \frac{1}{G^0_{nq}(\mathbf{r}, \omega, \mathbf{s})} \int_M \left[\omega^2 \delta\rho(\mathbf{x}) G^0_{iq}(\mathbf{x}, \omega, \mathbf{s}) G^0_{ni}(\mathbf{r}, \mathbf{x}) - \frac{\partial G^0_{ni}(\mathbf{r}, \omega, \mathbf{x})}{\partial x_j} \delta C_{ijkl}(\mathbf{x}) \frac{\partial G^0_{kq}(\mathbf{x}, \omega, \mathbf{s})}{\partial x_l} \right] d\mathbf{x}. \tag{B12}$$

This equation linearly connects model parameters and complex phases through Rytov formulation. The elastic adjoint operator will be deduced and discretized (see eq. 32 for the general expression) as for the Born formulation.

APPENDIX C: CHOICE OF INVERTED FREQUENCIES

A1 For a transmission acquisition geometry and only body waves

The Fig. 4 explains how we obtain the expressions for k_x and k_z in any acquisition geometry. They are equal to

$$\begin{aligned} k_x &= k_0 sin(\alpha) - k_0 sin(\beta) \\ k_z &= k_0 cos(\alpha) + k_0 cos(\beta) \end{aligned} \tag{C1}$$

where k_0 is the vertical wavenumber (Sirgue 2003; Sirgue & Pratt 2004) and α and β are defined on Fig. 4 as angles between the source or receiver vector and the vertical direction.

Using the trigonometric formulae $\cos(\alpha) + \cos(\beta) = 2\cos(\frac{\alpha+\beta}{2})\cos(\frac{\alpha-\beta}{2})$ and $\sin(\alpha) - \sin(\beta) = 2\cos(\frac{\alpha+\beta}{2})\sin(\frac{\alpha-\beta}{2})$ and $\cos^2(\frac{\alpha-\beta}{2}) + \sin^2(\frac{\alpha-\beta}{2}) = 1$, we get

$$\begin{aligned} \|k\| &= k_x^2 + k_z^2 \\ &= 2k_0 \cos\left(\frac{\alpha+\beta}{2}\right) \end{aligned} \tag{C2}$$

and

$$\begin{aligned} k_x &= \|k\| \sin\left(\frac{\alpha-\beta}{2}\right) \\ k_z &= \|k\| \cos\left(\frac{\alpha-\beta}{2}\right) \end{aligned} \tag{C3}$$

In the pure transmission case (Fig. 4 case 3), the imaged wavenumber at a given frequency is zero (continuous component of the medium). Therefore, the frequency choice is determined by souce-receiver couples corresponding to horizontal or vertical geomoetries (case 1 and 2 in Fig. 4). Thus, the inverted frequencies may be chosen as equal to the ones calculated for the reflection acquisition geometry determined by Sirgue (2003), Sirgue & Pratt (2004).

A2 For a reflection acquisition geometry and surface waves

With only body waves, frequencies are chosen such as (Sirgue 2003; Sirgue & Pratt 2004):

$$\begin{aligned} k_{z\ \min} &= 2k_0/\sqrt{1 + h^2/z^2} \\ k_{z\ \max} &= 2k_0 \end{aligned} \tag{C4}$$

where h is the half-offset, z is the target depth, k_0 is the vertical wavenumber, $k_{z\min}$ and $k_{z\max}$ are minimal and maximal imaged wavenumbers for one frequency. When surface waves are present, their exponential decay with depth must be introduced, leading to

$$\begin{aligned} k_{z\ \min} &= 2k_0/\sqrt{1 + h^2/z^2}\exp(-2k_0 z) \\ k_{z\ \max} &= 2k_0 \exp(-2k_0 z) \end{aligned} \tag{C5}$$

Sirgue (2003), Sirgue & Pratt (2004) explained that the wavenumber spectrum continuous imaging condition is

$$k_{z\ \min}(f_{n+1}) = k_{z\ \max}(f_n) \tag{C6}$$

for two consecutive frequencies f_n and f_{n+1}. This gives

$$f_{n+1} = f_n/\sqrt{1 + h^2/z^2} \tag{C7}$$

for body waves, thanks to eq. (C4). For surface waves, thanks to eq. (C5), this equation becomes

$$\begin{aligned} F_{n+1} &= F_n/\sqrt{1 + h^2/z^2}\ \exp(2z/c[F_{n+1} - F_n)] \\ F_{n+1} &= f_{n+1}\ \exp(2z/c[F_{n+1} - F_n)], \end{aligned} \tag{C8}$$

where f (small letters) are frequencies chosen for body waves whereas F (block letters) are frequencies chosen for surface waves. As $\exp(2z/c\ [F_{n+1} - F_n)] > 1$ since $2z/c\ (F_{n+1} - F_n) > 0$, the frequency choice is less constraining for surface waves than for body waves. Therefore, we choose frequencies corresponding to the body waves imaging conditions.

Let us point out that the frequency choice proposed by Sirgue & Pratt (2004) is based on the wave propagation in a homogeneous medium. In more realistic media, inverted frequencies distribution may be denser to take benefit of the wavenumber spectrum overlapping and redundancy.