An introduction to full waveform inversion

J. Virieux, ISTerre, jean.virieux@univ-grenoble-alpes.fr.
A. Asnaashari, PGS, amir.asnaashari@pgs.com.
R. Brossier, ISTerre, romain.brossier@univ-grenoble-alpes.fr.
L. Métivier, LJK/ISTerre, ludovic.metivier@univ-grenoble-alpes.fr.
A. Ribodetti, Geoazur, CNRS/IRS/UNS/OCA, alessandra.ribodetti@geoazur.unice.fr.
W. Zhou, ISTerre, wei.zhou@univ-grenoble-alpes.fr.

Published January 2017. **http://dx.doi.org/10.1190/1.9781560803027.entry6.**

Abstract

Full waveform inversion (FWI) is a high-resolution seismic imaging technique that is based on using the entire content of seismic traces for extracting physical parameters of the medium sampled by seismic waves. The widespread strategy of seismic imaging, the single scattering formulation, at the core of FWI, assumes no prior scale in the model description. Each unexplained residual data sample at receivers for one source is assumed to come from any point of the medium, and only the summation over sources and receivers helps in locating medium property anomalies, regardless of what type of phase is involved. This pixel-oriented perturbation leads to the local optimization approach, which is a linearized differential approach based on the Newton equation. For a least-squares misfit function, there are both the gradient vector and Hessian matrix, in addition to approximations that can be considered for the related Newton equation. The forward problem of the wave propagation, used thousands of times during optimization, should be efficient, and these equations are expressed either as a first-order hyperbolic system of velocity-stress or as a second-order hyperbolic system of displacement (or velocity) only, by using a self-adjoint formulation in both cases. Gradient vectors are built as a zero-lag crosscorrelation in time between incident and adjoint wavefields with forward and backward patterns and also could be used for obtaining Hessian-matrix approximations. Resolution and uncertainties are relevant, although the actual state of the art does not provide meaningful estimation of these quantities: the FWI remains a deterministic approach at this time. An examplary North Sea data set from the Valhall reservoir illustrates the successful story for high-resolution imaging based on data-driven components, with paleorivers stored in sediments, imprints of glaciers in the bedrock, and gas clouds at different scales in the image. Alternative sources of information on the medium, such as sonic logs and geologic interpretation, are illustrated through a model-driven component of the misfit function. Although methods can be used to increase the speed of the workflow, they are quite costly. The multiparameter reconstruction, which is mandatory for elastic FWI, starts to be feasible if one improves the Hessian-matrix influence. Thus, FWI is becoming a mature strategy for high-resolution seismic imaging.

Introduction

Remote-sensing methods are essential for characterizing the interior of the Earth at different scales, from the shallow subsurface to the inner core. Interaction between seismic waves and matter is expected to provide the highest resolution of these indirect osculating physical methods, compared with the results from gravimetry, geomagnetism, or electromagnetism.

Natural events, such as earthquakes, make the problem quite challenging because one also needs to characterize the sources, whereas controlled-source seismology focuses mainly on reconstruction of the structure.

During the rapid growth of seismic imaging methods over the last 50 years, two types of interactions have been identified. The first is a smooth interaction between waves and the medium that occurs where wave directions are slightly deviated by variations of the medium properties. One can assume that small variations in the wavefront result from this type of interaction, thereby allowing waves to continue to propagate forward after the interaction, in a forward scattering that is related to a transmission regime. The second is a rough interaction that occurs where wave directions are significantly modified by medium properties, resulting in a backward scattering that is related to a reflection regime; after that interaction the waves move backward. For example, reflection waves have a completely different direction of propagation, compared with that of the incident wave. Waves are bounced back toward the receiver position. One phase may mix both interactions during its propagation inside the medium. Other types of interactions could contribute to the model updating, in the form of multiple scattering effects, and could lead to a diffusive regime that requires other strategies for model building (Wu, 1985; Wu and Aki, 1988). Let us point out that forward modeling will introduce these multiple scattering and diffusive effects for the current model if there are corresponding heterogeneities.

These two types of interaction can be observed on the hierarchical organization of seismograms or traces (Figure 1). For a smooth interaction, traveltimes of a given phase could vary smoothly, whereas different bursts of energy occuring along the time axis characterize the rough interaction. These two probing interactions of medium properties have led naturally to two strategies for imaging the medium. The first type of reconstruction is based on tomographic approaches in which one tries to reconstruct the smoothly varying properties of the medium by tracking essentially slow variations of phases (or traveltimes). Thus, for these analyses, we have many approaches depending on the phases we consider as first-arrival phases, refraction phases, reflection phases, or surface-wave phases, among the various possible phases (combining these approaches improves the reconstruction) (Thurber and Ritsema, 2007). For example, first-arrival phases have inflections, as do the reflection phases, leading to traveltime tomographies. Reflection phases exhibit this smooth interaction on their downward and upward paths, along with the rough interaction at the interface or at diffracting points.

The second type of reconstruction is based on the so-called migration, either in time or in space, connected with Huygens' principle in which rapid variations of medium properties are tentatively reconstructed. Recent formulations have improved such imaging toward more quantitative estimations of the medium properties, such as least-squares migration (Nemeth et al., 1999). A related tool is amplitude-variation-with-offset (AVO) or amplitude-variation-with-angle (AVA) analysis, which characterizes a target zone from a single phase (Demirbag et al., 1993; Ursin and Ekren, 1995; Gray et al., 1998).

Considering the wavenumber domain, deduced by applying a Fourier transformation from the space domain, the low-wavenumber content is investigated by traveltime tomographic methods, whereas the high-wavenumber content is built by migration methods. Claerbout (1985) gives a simple presentation of such reconstructions, essentially along the depth direction, showing a poor reconstruction of intermediate wavenumber ranges (Figure 2). Are seismic waves inadequately sensitive to intermediate wavenumber ranges or are we designing standard imaging approaches that are not sensitive to intermediate wavenumbers? What seismic-trace information should we consider for building the low-wavenumber content? What information do we need for the high-wavenumber content? Do we need additional information, especially when we try to extend the two zones at the limit of the expected resolution of seismic data?

Tarantola (1984) has investigated an approach for avoiding this distinction between the expected smooth and rough interactions. On the basis of the single scattering formalism, all phases are considered under the same type of interaction (Devaney, 1984), regardless of the modification of the wave propagation induced by the properties of the medium. This formalism for updating the medium properties is embedded into the linearized approach called full waveform inversion (FWI); under the same concept of wave/medium interaction (each piece of information in the data

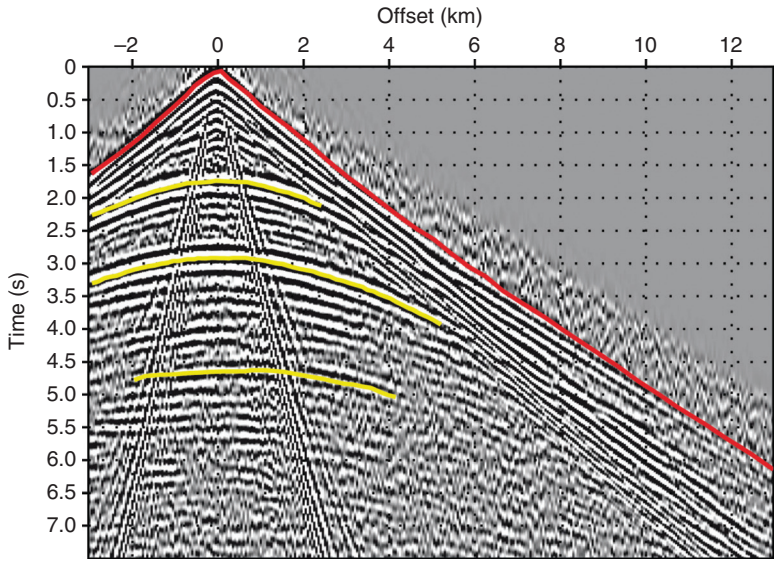

Figure 1. Common shot gather with different organized hierarchies in the time recording. Among these hierarchies, bursts of energy with reflection phases (yellow curves) express the reflection regime with time delays, whereas undulations visible both on direct waves (red curves) and on hyperbolic branches of reflected phases (yellow curves) identify the transmission regime.

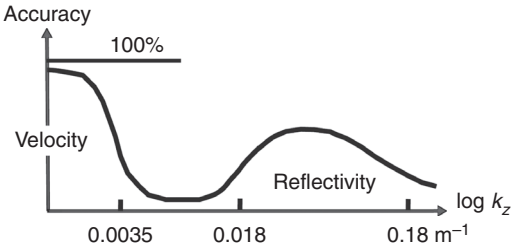

Figure 2. Extraction percentage of vertical model variation when considering the vertical component k_z of the wavenumber vector. Tomographic tools reconstruct the smooth part of the velocity structure, whereas migration tools reconstruct the contrasted part of the velocity structure. There is an apparent insensitivity at intermediate values of wavenumbers. After Figure 1.4-3 of Claerbout (1985). Used by courtesy of J. Claerbout.

domain the same way), one should consider all phases contained in the traces for imaging. Any type of propagation should be taken into consideration in the same way for model updating. The multiple interactions of waves with the medium come only when the forward modeling is performed.

The key point of such a formulation derives from a unique way of considering the interaction between waves and matter. No prior scale separation is considered in this imaging strategy. Still, FWI is a linearized formulation in which the nonlinearity of waves with respect to the medium properties is introduced by the wave-propagation engine. In our approach we do not use any linearization of the forward modeling — such as the linear Born approach — which is considered in migration approaches. When we reconstruct our model by beginning with an initial model and revising it iteratively, nonlinearity is implicitly introduced by repeated exact forward modeling at each update stage. We also point out that the misfit between observed data and synthetic data is a continuous function that will behave well locally, thereby allowing the introduction of the stochastic optimization approach under the Bayesian approximation, as Tarantola (1987) promoted.

We shall first consider the prior assumption of the single scattering approach, in which no scale separation is used. We then discuss what is the associated forward problem performed in the current model (not the initial model), so that we can provide practical measures for updating the model. We highlight the components necessary for seismic imaging, leaving the technical description of the adjoint formulation to the Appendix. We next discuss the resolution and uncertainty estimation in this context of a deterministic approach. We have all of the required elements for an illustration of full waveform inversion, and we highlight FWI results obtained for the Valhall reservoir in the North Sea by different groups.

We need to start from an initial model, but the challenging issue of defining that model has required many tentative strategies that we shall recall for possible future investigation by the reader. We prefer to focus on a specific one that is driven by the usual concepts of seismic imaging. This strategy may give the reader a better, more practical understanding of how full waveform inversion works. Because full waveform inversion is a heavily computational task, we briefly present strategies for speeding up the algorithmic aspect of the total workflow, regardless of the possible hardware solutions to be considered. Requirements for multiparameter reconstruction are analyzed and should lead scientists to concentrate future research efforts on the Hessian matrix. Different imprints of the parameters inside the data, and the possible tradeoffs among the parameters, will be near-future challenges of full waveform inversion. With that, we conclude our examination of this important technique, which has found its place in tools deployed by the seismic community for improving seismic imaging by increasing the extraction of information from more data samples than was previously possible.

Single scattering formulation

Extracting information contained in records of particle motion can be quite challenging because the seismic waves being recorded are traveling over large distances, thereby making ambiguous the position in the medium where property changes are inducing modifications in wave travel. To overcome such ambiguity, we proceed through a perturbation strategy in which observed seismic traces are compared with synthetic ones computed in a given medium. From residuals between observed and synthetic traces, we use perturbation iteratively to update the medium properties by employing full wave modeling to recompute synthetic traces in each updated medium. Designing the initial medium is, therefore, a critical step of such a local procedure. The estimation of property perturbation, starting from a given model with specific properties, is based on a single scattering formulation in which we attempt to reconstruct perturbation values at a single point. The perturbations are added to those of the current medium. Therefore, when we want to update the model, we consider the medium to be a sum of rather independent diffraction points (or pixels) where properties should be updated without any hierarchical strategy in the analysis of the types of waves (related to medium/wave interaction) we are examining in the data domain. Other strategies may represent the model with surfaces — such as in the common-reflection-surface (CRS) method, which links many points of the medium in a collaborative manner (Hubral et al., 1998; Heilmann et al., 2006; Koglin et al., 2006). Recall that the forward modeling will link pixels through the full resolution of the wave equation; multiple diffractions and reflections are included in the propagation. During updating, pixel interaction is more complex to assess and will depend on the misfit definition and on the related descent direction. It is different from the linking of the forward modeling.

Geophysical diffraction tomography (Devaney, 1984; Wu and Toksöz, 1987; Mora, 1989; Huang and Schuster, 2014) based on the single scattering formulation will provide a quantified estimation of wavenumber sampling from seismic data. This theory includes all scattering angles related to forward and backward scatterings, which are not distinguished. Their contribution will be quite different in the process of model building because backward scattering (small angles) has a tendency to locate the possible model-property variation, whereas the forward scattering (flat angles) has difficulties doing the same thing when considering all of the contributions coming from the acquisition. The record contribution at the receiver coming from a scattering point illuminated by the incident wave from the source depends on local orientations of the incident and scattered waves. Waves hitting the scattering point are leaving the source with an angle ϕ_s and will illuminate the scattering point with an angle Φ_s. Scattered waves leave the scattering point with an angle Φ_r and reach the receiver with an angle ϕ_r. The forward slowness vector $\mathbf{p}_s$ and the backward slowness vector $\mathbf{p}_r$ are constructed with an illumination angle θ for obtaining the illumination vector $\mathbf{q}$, which is linked to the wavenumber vector $\mathbf{k}$ through the expression

$$\mathbf{k} = 2\omega\mathbf{q} = \frac{4\pi f}{c}\cos\left(\frac{\theta}{2}\right)\mathbf{n} = \frac{4\pi}{\lambda}\cos\left(\frac{\theta}{2}\right)\mathbf{n}, \quad (1)$$

where the angular frequency is denoted by $\omega = 2\pi f$ (frequency is denoted by f), the local wavelength is represented by λ (related to the

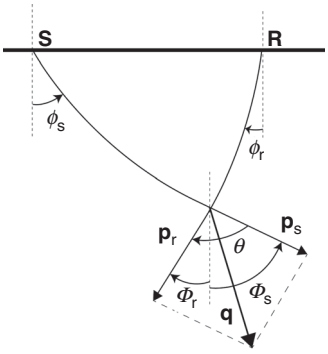

Figure 3. Single scattering geometric configuration. At the point of diffraction, where model properties have to be reconstructed, the incident wave arrives with an angle Φ_s, providing a wavenumber vector $\mathbf{p}_s$. Corresponding receiver data or residuals are back-propagated in the medium and will arrive with an angle Φ_r at the diffracting point, providing a wavenumber vector $\mathbf{p}_r$. The angle θ between these two vectors is the illumination angle, and the vectorial composition of these vectors provides the illumination vector q, a key ingredient for parameter imaging at the diffracting point. Waves leave the source with an angle ϕ_s and arrive with an angle ϕ_r at the receiver.

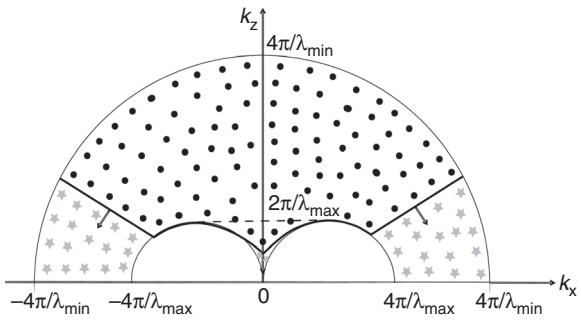

Figure 4. Continuous sampling of wavenumber space based on the single scattering formulation, regardless of the types of waves. The zone of sampling is the black-dotted zone. As offset range increases, the wavenumber sampling can conquer the grey-star zone.

local velocity c), and the vector $\mathbf{n}$ is the unitary vector colinear to both illumination and wavenumber vectors (Figure 3). One can see that we may obtain small wavenumber values by decreasing the frequency or by having angle θ close to π, which is the transmission regime we considered in the introduction. In other words, even high-frequency contents in phases (short wavelengths) may provide low-wavenumber information (small wavenumbers) of the model if we are able to identify the so-called transmission regime of these phases. Of course, this analysis and the related updating are performed in the current model and will be modified at each iteration.

Considering a source frequency band between f_{min} and f_{max} that is related to a minimum wavelength λ_{min} and a maximum wavelength λ_{max}, the sampling of the wavenumber space is shown in Figure 4 for a homogeneous medium with a flat interface and is inspired from figures by Wu and Toksöz (1987) and Mora (1989) when they were investigating a broad range of frequencies. One can see that low-vertical-wavenumber values are difficult to sample and depend strongly on the minimum frequency of the source as well as on the offset range. With increases in the offset range come commensurate improvements in transmission configuration and, therefore, the contribution of diving waves, thereby leading to a better sampling of low vertical wavenumbers. Under the single scattering formulation, in which no phase interpretation is performed, we have difficulty filling some zones of the wavenumber spectrum when using the data. If necessary for the wave-propagation modeling (especially the low-wavenumber components), these zones should be present in the initial model. We shall see that going back to phase identification without precise picking will allow a compatible fill-in of these zones for better full waveform reconstruction.

Local optimization strategy

Let us assume that the model space is described over a diffraction domain $\mathcal{D}$ where property perturbations $\Delta\mathbf{m}$ should be updated at the scattering point $\mathbf{x}$. Because we have at our disposal a model inside which wave solutions can be provided between source positions $\mathbf{x}_s$ spanning a domain $\partial\Omega_s$ or receiver positions $\mathbf{x}_r$ spanning a domain $\partial\Omega_r$ and any scattering position $\mathbf{x}$, we may consider the imaging operator $\mathcal{O}$ to be designed from current model $\mathbf{m}$ and available data $\mathbf{d}$ as

$$\Delta\mathbf{m}(\mathbf{x}) = \mathcal{O}^{\mathbf{m}}_{\text{data}}\Big[\Delta^{\mathbf{m}}\mathbf{d}(\mathbf{x}_s,\mathbf{x}_r,t),\mathbf{x}_s,\mathbf{x},\mathbf{x}_r,\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x},t),\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_r,\mathbf{x},t)\Big], \quad (2)$$

where the model perturbations $\Delta\mathbf{m}$ to be defined are linked to the residuals $\Delta^{\mathbf{m}}\mathrm{d}(\mathbf{x}_s,\mathbf{x}_r,t)$ between observed traces $\mathbf{d}_{\text{obs}}(\mathbf{x}_s,\mathbf{x}_r,t)$ and synthetic traces $\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_r,t)$. Synthetic wavefields between

sources and scatterers $\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x},t)$ on one side and between receivers and scatterers $\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_r,\mathbf{x},t)$ on the other side are necessary ingredients, as are the geometric positions of sources, receivers, and scatterers.

The imaging operator could be designed through different approaches, such as the direct inversion approach based on an integral formulation (Bleistein, 1987; Miller et al., 1987; Beylkin and Burridge, 1990; Weglein et al., 2009) or through an optimization inversion such as the least-squares minimization (Tarantola, 1987). We shall consider the second approach and emphasize that a possible bridge could be built with the direct inversion (Jin et al., 1992). Therefore, the FWI problem is defined as a nonlinear least-squares minimization problem of the misfit function $\mathcal{C}$ through the rather standard least-squares misfit function

$$\min_{\mathbf{m}} \mathcal{C}(\mathbf{m}) = \frac{1}{2}\int_0^{T_w}\int_{\partial\Omega_s}\int_{\partial\Omega_r} |\, \mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_r,\mathbf{x}_s,t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_r,\mathbf{x}_s,t)\,|^2\, d\mathbf{x}_r d\mathbf{x}_s dt, \quad (3)$$

where the modulus operation is denoted by the symbol $|\cdot|$. Other misfit functions could be devised as long as we can perform the derivation with respect to model parameters. The duration of observation, because the origin time is set to zero, is given by T_w. This problem is solved through local nonlinear optimization algorithms. In the framework of these methods, an iterative sequence $\mathbf{m}_k(\mathbf{x})$ is built from an initial guess $\mathbf{m}_{\text{ini}}(\mathbf{x})$, such that

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k, \quad (4)$$

where the scalar parameter $\alpha_k \in \mathbb{R}$ is computed through a linesearch or a trust-region globalization process (Nocedal and Wright, 2006; Bonnans et al., 2006). In the particular case of the Newton method for linear problems, the parameter α_k is set to 1 and the model update $\Delta\mathbf{m}_k$ is deduced from the Newton equation

$$\mathcal{H}(\mathbf{m}_k)\Delta\mathbf{m}_k = -\nabla\mathcal{C}(\mathbf{m}_k). \quad (5)$$

In this expression, the operators $\mathcal{H}(\mathbf{m}_k)$ and $\nabla\mathcal{C}(\mathbf{m}_k)$ are, respectively, the Hessian matrix and the gradient vector of the functional $\mathcal{C}(\mathbf{m})$ at the iteration k. These two key components of the FWI method can be expressed in terms of the Jacobian operator $\mathcal{J}(\mathbf{m}) = \partial\mathbf{d}_{\text{syn}}/\partial\mathbf{m}$ as

$$\nabla\mathcal{C}(\mathbf{m}) = \mathcal{J}^{\mathrm{T}}(\mathbf{m})(\mathbf{d}_{\text{syn}}(\mathbf{m}) - \mathbf{d}_{\text{obs}}),$$
$$\mathcal{H}(\mathbf{m}) = \mathcal{J}^{\mathrm{T}}(\mathbf{m})\mathcal{J}(\mathbf{m}) + \frac{\partial\mathcal{J}}{\partial\mathbf{m}}(\mathbf{d}_{\text{syn}}(\mathbf{m}) - \mathbf{d}_{\text{obs}}), \quad (6)$$

where the transpose operator is denoted by the superscript T. A common approximation of the Newton method consists of considering only the first term of the right side in the expression of the Hessian operator (equation 6) and is referred to as the Gauss-Newton (GN) approximation. In this context, the model update at iteration k is the solution of the linear system

$$\mathcal{H}_{\text{GN}}(\mathbf{m}_k)\Delta\mathbf{m}_k = -\nabla\mathcal{C}(\mathbf{m}_k), \text{ with}$$
$$\mathcal{H}_{\text{GN}}(\mathbf{m}) = \mathcal{J}^{\mathrm{T}}(\mathbf{m})\mathcal{J}(\mathbf{m}). \quad (7)$$

This equation can be rewritten in terms of the Jacobian operator and the model update only through the equation

$$\mathcal{J}^{\mathrm{T}}(\mathbf{m}_k)\mathcal{J}(\mathbf{m}_k)\Delta\mathbf{m}_k = -\mathcal{J}^{\mathrm{T}}(\mathbf{m}_k)(\mathbf{d}_{\text{syn}}(\mathbf{m}_k) - \mathbf{d}_{\text{obs}}). \quad (8)$$

This Gauss-Newton approximation (equation 8) is important because it reveals that the model update at iteration k is computed as the least-squares solution of the simple linear system

$$\mathcal{J}(\mathbf{m}_k)\Delta\mathbf{m}_k = -(\mathbf{d}_{\text{syn}}(\mathbf{m}_k) - \mathbf{d}_{\text{obs}}), \quad (9)$$

which is a locally linearized expression around the current model $\mathbf{m}_k$ and provides a local linear link between data perturbation and model perturbation. Let us reiterate the alternative: finding an approximation of the inverse of the Jacobian operator, sometimes named the direct inversion formulation, will provide a different strategy from that of the least-squares operator $-[\mathcal{J}^{\mathrm{T}}(\mathbf{m}_k)\mathcal{J}(\mathbf{m}_k)]^{-1}\mathcal{J}^{\mathrm{T}}(\mathbf{m}_k)$. One can cite the inverse Fourier transform or the inverse of the generalized Radon transform as examples of direct inversions (Beylkin, 1985).

One can see that as soon as we have discrete data and model spaces, the single scattering formulation for updating the current model relies essentially on tools designed from linear algebra. Acquisition configuration and initial model

design are the two other types of complementary information that we need for performing the full waveform inversion as a data-driven strategy of imaging. In order to obtain a practical evaluation of the gradient vector and the Hessian matrix of the misfit function $\mathcal{C}$, we must consider the hyperbolic partial differential equation for constructing the wave solution inside a known medium. This is the forward problem.

Wave propagation

Wave-propagation modeling is an internal process of the seismic imaging workflow and should be designed with a consideration of the optimization we have selected. Here we will examine forward modeling by itself. We formulate it as a time-implicit differential system suitable for optimization but also as a time-explicit differential system suitable for computational purposes. The Earth is a heterogeneous medium with attenuation and anisotropic properties, making the modeling of wave propagation a challenging problem. We will not discuss attenuation in this introductory presentation of FWI because additional attention is required (see the review by Yang et al. (2016) for including attenuation in the full waveform context). A consideration of anisotropy does not introduce new difficulties and, therefore, can be undertaken here. Following continuum mechanics, the motion at a point of the medium $\mathbf{x} = (x, y, z)$ should follow the partial differential equation (PDE)

$$\begin{aligned} \rho \partial_t v_i &= \partial_j \sigma_{ij} + f_i^{\mathrm{e}} \\ \partial_t \sigma_{ij} &= c_{ijkl} \partial_t \varepsilon_{kl} + \partial_t \sigma_{ij}^{\mathrm{e}}, \end{aligned} \tag{10}$$

where the density is denoted by the symbol ρ, the i component of the particle velocity by v_i, the ij component of the second-order stress tensor by σ_{ij}, the $ijkl$ component of the fourth-order stiffness tensor by c_{ijkl} (sometimes called the elastic tensor), and the second-order strain tensor by ε_{ij}. Indices i, j, k, l span the coordinates $[x, y, z]$ (or [1, 2, 3], as is often used). Recall that the strain is defined by the expression $\varepsilon_{ij} = (\partial_i u_j + \partial_j u_i)/2$ where the i component of the particle displacement is denoted by u_i. Recall also that all mentioned tensors are symmetrical; the system is conservative. The first equation is Newton's law for dynamics, and the second equation is the time derivative of the linear Hooke's law characterizing the elastic rheology of the material defined as $\sigma_{ij} = c_{ijkl}\varepsilon_{kl}$. The medium is excited by external forces f^{e} or by internal stress σ^{e} coming from a local failure of the rheology law (explosion or fracture).

We use Voigt indexing, $11 \to 1, 22 \to 2, 33 \to 3$, 23 *or* $32 \to 4$, 13 *or* $31 \to 5$, 12 or $21 \to 6$, and we define the nine-component vector $\mathbf{w}^{\mathrm{t}} = (v_x\ v_y\ v_z\ \sigma_{xx}\ \sigma_{yy}\ \sigma_{zz}\ \sigma_{yz}\ \sigma_{xz}\ \sigma_{xy})^{\mathrm{T}}$ with $v_x \equiv w_1$, $v_y \equiv w_2$, $v_z \equiv w_3$, $\sigma_{xx} \equiv \sigma_1 = w_4$, $\sigma_{yy} \equiv \sigma_2 = w_5$, $\sigma_{zz} \equiv \sigma_3 = w_6$, $\sigma_{yz} \equiv \sigma_4 = w_7$, $\sigma_{xz} \equiv \sigma_5 = w_8$, $\sigma_{xy} \equiv \sigma_6 = w_9$. Substituting strain by particle velocities (the time derivative of displacement), the elastodynamic system can be written through scalar notation for each component w_I as

$$\partial_t w_I(\mathbf{x}_i, t) = (\mathbf{A}\boldsymbol{\mathcal{D}})_{IJ}\, w_J(\mathbf{x}_i, t) + f_I^w, \tag{11}$$

with implicit summation on repeated capital indices from values 1 to 9. The source term $\mathbf{f}^w$ could act on the nine different components. The vector $\mathbf{w}$ is evaluated at the position $\mathbf{x}_i$ in the discrete mesh used for numerical integration. The extended stiffness matrix $\mathbf{A}$ has the following structure

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_{3\times3} & 0_{3\times6} \\ 0_{6\times3} & \mathbf{C}_{6\times6} \end{pmatrix}, \tag{12}$$

where the diagonal matrix

$$\mathbf{B} = \begin{pmatrix} b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{pmatrix} \tag{13}$$

introduces buoyancy b, which is the inverse of density ρ, and where the stiffness matrix

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ \cdot & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ \cdot & \cdot & c_{33} & c_{34} & c_{35} & c_{36} \\ \cdot & \cdot & \cdot & c_{44} & c_{45} & c_{46} \\ \cdot & \cdot & \cdot & \cdot & c_{55} & c_{56} \\ \cdot & \cdot & \cdot & \cdot & \cdot & c_{66} \end{pmatrix} \tag{14}$$

is symmetrical. The symmetrical matrix $\boldsymbol{\mathcal{D}}$ contains differential operators with respect to space coordinates:

$$\boldsymbol{\mathcal{D}} = \begin{pmatrix} 0_{3\times3} & \mathbf{D}_{3\times6} \\ \mathbf{D}^t_{6\times3} & 0_{6\times6} \end{pmatrix} \tag{15}$$

with the rectangular submatrix

$$\mathbf{D}_{3\times 6} = \begin{pmatrix} \partial_x & 0 & 0 & 0 & \partial_z & \partial_y \\ 0 & \partial_y & 0 & \partial_z & 0 & \partial_x \\ 0 & 0 & \partial_z & \partial_y & \partial_x & 0 \end{pmatrix}. \qquad (16)$$

In matrix form, we have the first-order hyperbolic system

$$\partial_t \mathbf{w} = \mathbf{B}_w \mathbf{w} + \mathbf{f}^w \qquad (17)$$

with an impedance matrix $\mathbf{B}_w$, which is now not symmetrical (because matrices $\mathbf{A}$ and $\boldsymbol{\mathcal{D}}$ do not commute), as shown by boxes when considering its complete expression for the particular case of isotropic media (with Lamé coefficients λ and μ)

$$\mathbf{B}_w = \begin{pmatrix} 0 & 0 & 0 & \boxed{b}\partial_x \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \boxed{(\lambda+2\mu)}\partial_x & \lambda\partial_y & \lambda\partial_z & 0 \\ \lambda\partial_x & (\lambda+2\mu)\partial_y & \lambda\partial_z & 0 \\ \lambda\partial_x & \lambda\partial_y & (\lambda+2\mu)\partial_z & 0 \\ 0 & \mu\partial_z & \mu\partial_y & 0 \\ \mu\partial_z & 0 & \mu\partial_y & 0 \\ \mu\partial_y & \mu\partial_x & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & b\partial_z & b\partial_y \\ b\partial_y & 0 & b\partial_z & 0 & b\partial_x \\ 0 & b\partial_z & b\partial_y & b\partial_x & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad (18)$$

The excitation $\mathbf{f}^w$ is composed from forces and stresses of the system (equation 10). We shall see that we need to solve partial differential equations defined by the transpose matrix $\mathbf{B}_w^{\mathrm{T}}$, which is real. System 10 governing the wave propagation is not a self-adjoint differential system. Because the wave propagation preserves the total energy in nonattenuating media, we design a self-adjoint system, providing the same solution as in system 10.

If we introduce the compliance tensor $\mathbf{S}$, for which we shall use Voigt compact notation, we consider the following PDE

$$\begin{aligned} \rho\partial_t v_i - f_i^{\mathrm{e}} &= \partial_j \sigma_{ij} \\ s_{ijkl}\partial_t \sigma_{ij} - s_{ijkl}\partial_t \sigma_{ij}^{\mathrm{e}} &= (\partial_k v_l + \partial_l v_k)/2. \end{aligned} \qquad (19)$$

All spatial partial derivatives have been purposely written on the right side and, therefore, are not intermingled with the properties of the medium. All matrices are now symmetrical and, therefore, the energy will be preserved during propagation — an important property of wave propagation in nonattenuating media. Now we understand why system 10 also preserves the energy, whereas it was not obvious initially. Moreover, this separation will be quite helpful when we take the derivative of these partial differential equations with respect to the medium parameters, which are expressed only on the left side of these equations. We introduce an extended mass/compliance matrix $\mathbf{M}$ and we write in a compact form the symmetrical first-order hyperbolic PDE, still using the scalar notation for each component

$$M_{IJ}\,\partial_t w_J(\mathbf{x}_i,t) - M_{IJ}\,f_J^w = \boldsymbol{\mathcal{D}}_{IK}\,w_K(\mathbf{x}_i,t), \qquad (20)$$

where a product of matrices is avoided (Burridge, 1996). As we shall see, this structure is suitable for elaborating on the imaging condition, but because it is an implicit system it is not suitable for computing wavefields. We prefer to solve the first-order hyperbolic system (equation 11) by evaluating the time evolution of the wavefield $\mathbf{w}$ from initial values. Let us repeat that the system of equation 20 will be the one used formally for seismic image updating, regardless of the system selected for computing wavefields. These two systems (equations 11 and 20) provide the same continuum solution. Numerical solutions may differ only by numerical errors coming from the discretization we assume in the numerical algorithm we select; such errors will be neglected.

Alternatively, we may eliminate the stress components in the first-order hyperbolic system (equation 11), leading to a second-order hyperbolic system in which the wavefield $\mathbf{w}$ is only the particle velocity $\mathbf{v}$. We drastically reduce the memory requirements of wave-propagation simulation because the wavefield now only has three

components. This system is self-adjoint because we have second-order derivatives in space and in time. The introduction of the compliance matrix illustrates this convenient common property shared by first-order and second-order hyperbolic systems. We obtain the system

$$\rho\partial_{tt}v_i = \partial_j(c_{ijkl}\partial_l v_k) + f_i^{\mathrm{e}} + \partial_j\sigma_{ij}^{\mathrm{e}}. \tag{21}$$

This system must be complemented by the estimation of the strain ε or equivalently by the stress σ because we will need them when considering derivatives of the PDE with respect to medium parameters for the imaging condition. This requirement mitigates partially the memory reduction we have identified. We must consider an extended system

$$\begin{aligned} \rho\partial_{tt}v_i &= \partial_j(c_{ijkl}\partial_l v_k) + f_i^{\mathrm{e}} + \partial_j\sigma_{ij}^{\mathrm{e}}, \\ \partial_t\varepsilon_{ij} &= (\partial_i v_j + \partial_j v_i)/2. \end{aligned} \tag{22}$$

The first line is self-consistent and provides the wavefield at each time step. The second line of system 22 is required for some of the imaging conditions we shall consider in the next section. In other words, these additional fields $\partial_t\varepsilon$ (or equivalently $\partial_t\sigma$) are only composed when needed for computing seismic imaging ingredients. Their numerical estimation should be integrated into the numerical tool of the forward problem, whereas that estimation is not included when only the wavefield is needed. This alternative formulation is attractive for different applications because it mitigates the memory requirement. Of course, with whatever system we select, the imaging condition expressed by the gradient will be the same quantity.

We can move from time integration to frequency sampling, and in so doing a linear system has to be solved with complex arithmetic operations. Often we prefer to convert the second-order hyperbolic system in time into a linear discrete system that will be smaller than the one deduced from the first-order hyperbolic system. Computer resource needs are reduced for solving the linear system either through a direct solver technique based on sparse matrix-vector manipulation (MUMPS team, 2011) or through an iterative solver technique (Plessix, 2007). The direct solver technique is quite appealing when considering many sources, but the memory requirement is high. Recently improved approaches have broadened the domain of applications because memory requirements have been reduced and scalability has been dramatically improved (Amestoy et al., 2015). Iterative solver techniques have been found to be efficient strategies for the acoustic case (Plessix, 2007; Riyanti et al., 2007; Erlangga and Nabben, 2009; Neklyudov et al., 2014) but the elastic case is still an open question (Gosselin-Cliche and Giroux, 2014; Li et al., 2015).

Currently, the optimal strategy is unclear — whether we should consider frequency formulation or time formulation for the forward problem. The frequency formulation could include attenuation without any extra cost, whereas incorporating attenuation is more tedious in the time formulation. On the other hand, selecting data is easier in the time domain than in the frequency domain. We must wait for numerical investigations by other groups to identify cases when one option is superior to the other one; we expect that the choice will be case-dependent and could change during the imaging process. Until now, from a practical point of view, most implementations of FWI have been performed in the time domain because multiple frequency contents could be considered simultaneously, which is important when dealing with reflection phases.

Seismic imaging ingredients

This section is technical, and a demonstration of the results are presented in the Appendix. The most obvious strategy for minimizing the misfit function uses the estimation of Fréchet derivatives; that is, it uses the data derivative with respect to the model parameter, also called the sensitivity matrix, which has a dimensional complexity related to the dimension of the data space multiplied by the dimension of the model space. The related system of equation 9 is solved in the least-squares sense, leading to the system of equation 5, which is solved by different techniques, such as the conjugate gradient promoted, for example, by the least-squares (LSQR) algorithm (Paige and Saunders, 1982). When looking at the system of equation 5, one can see that the local minimization of the misfit function will need only the estimation of its gradient vector with a model complexity (i.e., the number of degrees of freedom of the model), and sometimes an estimate of its Hessian matrix with a model-square complexity (the square of the number of the degrees of freedom), independent of the dimension

of the data space in equation 5. The estimate of the sensitivity matrix is not strictly required, even though it will yield significant information for fine-tuning the optimization provided we can afford the expense of estimation and storage. We do not need to evaluate the Fréchet derivatives in order to update the model; this is the adjoint formulation (see Chavent, 1974; Tromp et al., 2005; Plessix, 2006b; Chavent, 2009, for reviews), which is described in more detail in the Appendix. Here we only describe the final expressions. Let us write this misfit function again,

$$\mathcal{C}(\mathbf{m})=\frac{1}{2}\int_0^{T_w}\int_{\partial\Omega_s}\int_{\partial\Omega_r}\left[\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_r,t)-\mathbf{d}_{\text{obs}}(\mathbf{x}_s,\mathbf{x}_r,t)\right]^2 \times d\mathbf{x}_r d\mathbf{x}_s dt, \tag{23}$$

with the relation between synthetic data and computed wavefields given by the expression

$$\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_r,t)=R_r\mathbf{w}^{\mathbf{m}}(\mathbf{x}_s,\mathbf{x},t), \tag{24}$$

where we define the operator R_r as the projection of the wavefield vector $\mathbf{w}^m$ onto data components at receiver positions. In a discrete world, this operator is a rectangular matrix. The vector $\mathbf{w}^m$ must satisfy the system of equation 20, which is expressed in time. We, therefore, have two equality constraints, one derived from the projection of the wavefield onto the recorded data at receivers and the other one from the wave equation verified by the wavefield. In mathematical optimization, the method of Lagrange multipliers is the mathematical tool for performing such a minimization under constraints. This leads us to introduce two new vectors. The first, λ^m, has the same number of components as the wavefield vector $\mathbf{w}^m$, whereas the second, μ^m, has the same number of components as the data vector $\mathbf{d}^{\mathbf{m}}_{\text{syn}}$ and is simply connected to the projector R_r. The first vector is called the adjoint wavefield associated with one source $\mathbf{x}_s$ and should verify the partial differential system

$$\begin{aligned}&M_{IJ}\partial_t\lambda_J(\mathbf{x}_s,\mathbf{x},t)\\&\quad-\sum_r R_r^T\left[d^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_r,t)-d_{\text{obs}}(\mathbf{x}_s,\mathbf{x}_r,t)\right]_I\\&=\mathcal{D}_{IK}(\nabla)\lambda_K(\mathbf{x}_s,\mathbf{x},t),\end{aligned} \tag{25}$$

which turns out to be identical to the implicit system of the wave propagation (equation 20). The sum is over receivers related to the selected source. The source term of this adjoint system defines this adjoint wavefield and comes from inserting the projection R_r^T of data residuals back into the modeling discrete grid. We send back into the medium unexplained features of the data in order to locate possible zones of contribution inside the medium. The conditions on this adjoint wavefield will be final conditions, because $\lambda(\mathbf{x}, T_w) = 0$. We deduce an explicit system

$$\begin{aligned}\partial_t\lambda_I(\mathbf{x}_s,\mathbf{x},t)&=(\mathbf{A}\mathcal{D})_{IJ}\lambda_J(\mathbf{x}_s,\mathbf{x},t)\\&\quad+\sum_r A_{IK}R_r^T\left(d^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_r,t)-d_{\text{obs}}(\mathbf{x}_s,\mathbf{x}_r,t)\right)_K,\end{aligned} \tag{26}$$

identical to equation 11, in which only the source term has been modified, thereby introducing medium properties at receivers. Note that the adjoint source will simultaneously return all of the residuals for one source. This system can be integrated back in time to the time of origin. Incident and adjoint wavefields are not obtained simultaneously. Although we do not consider attenuation in this presentation, let us note, to avoid confusion, that this adjoint system is still stable when attenuation is included. The attenuation effect is an amplitude decrease during the backward integration of the adjoint wavefield and does not provide any amplitude increase (Tarantola, 1988). Having the incident wavefield be at the same time as the adjoint wavefield is required for crosscorrelation, in order to build up the gradient. Many methods exist for making the incident wavefield available for crosscorrelation, such as saving the incident field during the forward computation with efficient compression and disk storage through efficient input/output strategies (Sun and Fu, 2013; Prabhat and Koziol, 2014). Alternatively, during the integration of the adjoint wavefield backward in time, one may recompute the incident wavefield either by using time reversal integration (Clapp, 2008; Dussaud et al., 2008; Brossier et al., 2014), or by using optimal checkpoint strategies based on repeated time forward integration (Griewank, 1992; Griewank and Walther, 2000; Symes, 2007; Anderson et al., 2012). We may deduce the gradient of the misfit function as

$$\frac{\partial\mathcal{C}}{\partial\mathbf{m}}(\mathbf{x})=\sum_s\int_0^{T_w}dt\,\lambda^T(\mathbf{x},t)\frac{\partial\mathbf{M}}{\partial\mathbf{m}}(\mathbf{x}) \times[\partial_t\mathbf{w}(\mathbf{x},t)-\mathrm{Mf}^W(\mathbf{x},t)]. \tag{27}$$

The sum is over sources.

In the frequency domain, monochromatic wavefields are stationary fields distributed inside the medium; they have amplitude and phase because they are complex. We consider one source from which we have a monochromatic concentric wavefield because we are working with a homogeneous medium. We have the same concentric shape of the wavefield from the receiver corresponding to the adjoint field. Figure 5 is the product of these two wavefields, which is the interference at a given frequency in a homogeneous medium. This interference is the gradient contribution for one pair (source, receiver), and it displays where we should update the model in order to explain what is observed and not explained at the receiver. This is sometimes called the sensitivity kernel. In the first Fresnel zone, primarily sampled by direct waves, waves provide a potential contribution for updating the medium with an expected smooth resolution (Woodward, 1992), whereas outer interference fringes, mainly sampled by reflected waves, yield a zone with an expected high resolution. The location of this zone strongly depends on the current model we are considering. The first Fresnel zone depends on the acquisition geometry and more weakly on the model, whereas the outer fringe is highly dependent on the model and does not depend on the acquisition. Estimating model perturbation inside the first Fresnel zone will fill in the low-wavenumber part of the medium because we are not able to discriminate between contributions of the different points of this zone. On the contrary, when we start the optimization, the initial model should be accurate enough to locate the outer fringe at nearly the right place, as we shall discuss in the section concerning the initial model definition. If we misposition this outer fringe, we shall construct a high-wavenumber content at an incorrect place and it will be very difficult to remove it through later iterations. This is a complementary interpretation of the cycle-skipping issue expressed in the oscillatory time signals. Counting fringes from the first Fresnel zone without missing one could be performed without an initial model, a counting nearly impossible to achieve with seismic acquisition. In seismics and in seismology, we rely on the initial model design and we avoid such counting. We put the fringe at nearly the right place with the help of a good initial model. We say that the initial model is kinematically compatible.

Away from sources, the gradient of the misfit function is the zero-lag crosscorrelation between the adjoint wavefield and the time derivative of the incident wavefield, which is multiplied by a local scattering operator. Note that there are no explicit spatial derivatives in this estimation of the gradient based on the first-order hyperbolic system (velocity-stress):

$$\frac{\partial \mathcal{C}}{\partial \mathbf{m}}(\mathbf{x}) = \sum_{s} \int_0^{T_w} dt\, \lambda^{\mathrm{T}}(\mathbf{x},t) \frac{\partial \mathbf{M}}{\partial \mathbf{m}}(\mathbf{x}) \partial_t \mathbf{w}(\mathbf{x},t), \qquad (28)$$

where we have omitted the dependence of the incident and adjoint wavefields with respect to

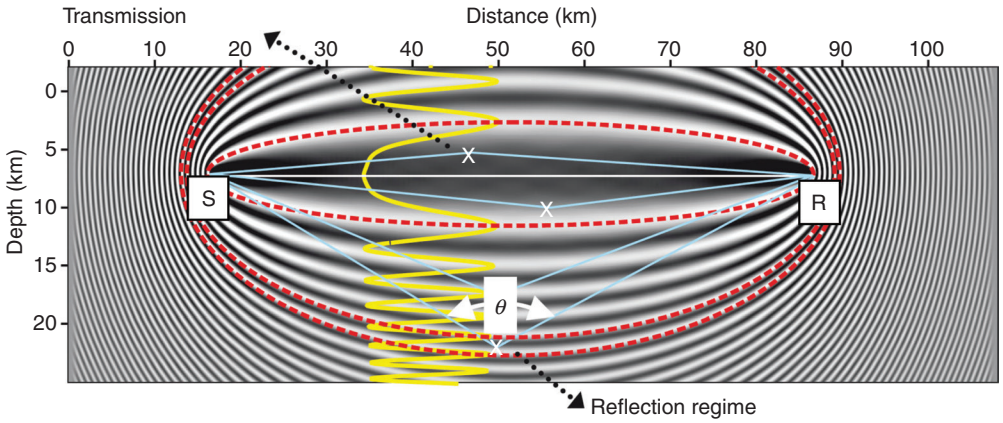

Figure 5. Misfit function gradient when considering a homogeneous background medium and monochromatic incident and adjoint fields; the zero-lag crosscorrelation between these two fields highlights two different contributions. The central ellipse represents the first Fresnel zone, where phase time does not differ by more than half a period with respect to the direct arrival. The outer fringes represent migration isochrones whose width decreases with depth according to the scattering angle.

the source. Using the density and the compliance components, we can write

$$\frac{\partial \mathcal{C}}{\partial \rho}(\mathbf{x}_i) = \sum_{\mathrm{s}} \int_0^{T_w} dt \Big[\partial_t v_I(\mathbf{x}_i) \lambda_I^v(\mathbf{x}_i) \Big]$$
$$\frac{\partial \mathcal{C}}{\partial S_{kl}}(\mathbf{x}_i) = \sum_{\mathrm{s}} \int_0^{T_w} dt \left[\partial_t \sigma_I(\mathbf{x}_i) \left(\frac{\partial \mathbf{S}}{\partial S_{kl}} \right)_{IJ} \lambda_J^\sigma(\mathbf{x}_i) \right], \tag{29}$$

with implicit summation over capital indices. The 6×6 matrix $\partial \mathbf{S}/\partial S_{kl}$ for each component S_{kl} is composed with elements of values zero or one. We have split the nine-component adjoint wavefield λ into the three-component adjoint wavefield λ^v related to particle velocity and into the six-component adjoint wavefield λ^σ related to stress. The physical interpretation of the gradient contribution based on displacement and stress components should rely on expressions 29, in which the scattering matrix is explicitly identified. An equivalent formulation could be written using stiffness components and is widely used in seismology, where wave-propagation modeling is performed with a second-order hyperbolic system (displacement) (Liu and Tromp, 2006).

The gradient estimation is required for updating the model, whereas we can avoid estimation of the inverse of the Hessian matrix when solving the Newton equation (equation 5). Omitting this estimation will decrease drastically the problem complexity. Most algorithms of FWI are essentially based on gradient estimations, such as the steepest-descent method or the conjugate-gradient method, although in recent investigations the impact of the Hessian matrix is increasingly being taken into consideration. An efficient way of evaluating its influence is through the quasi-Newton method, for which limited storage has been promoted by Nocedal and Wright (2006). This approach only requires a small number (roughly 10 or 20) of stored gradients and models for approximate estimations of the impact of the Hessian matrix.

In the future, we may need to go one step farther and consider the full Newton method, which will require estimation of the product of the Hessian matrix with any model vector — a key ingredient of a conjugate-gradient method for solving this Newton equation. An efficient way to estimate that product is through the so-called second-order adjoint formulation, which will not be described here. Instead, the reader is referred to the article by Métivier et al. (2014) in which a matrix-free approach is presented. This approach will allow efficient estimation of the influence of the Hessian matrix, which will be important for mitigating any acquisition deficit, for improving the convergence rate, and for reducing the tradeoff between parameters. Because we never estimate the inverse of the Hessian matrix, we cannot provide any information on resolution and uncertainty of the seismic image — an analysis that we consider in the next section.

An extensive description of the impact of the Hessian matrix will be provided when multiple parameter imaging is considered.

In the frequency domain, the expression of the misfit gradient vector is composed of three terms: the incident and adjoint wavefields and the scattering matrix. At the point i in the mesh for the model parameter m_l, the gradient is

$$\left(\frac{\partial \mathcal{C}(\mathbf{m})}{\partial m_l} \right)_i = \imath \omega \sum_{j=1}^{N_s} \lambda^t(\mathbf{x}, \omega)_I^j \left(\frac{\partial M}{\partial m_l}(\mathbf{x}_i) \right)_{IJ} \mathbf{w}^*(\mathbf{x}, \omega)_J^j, \tag{30}$$

where $\mathbf{w}(\mathbf{x}, \omega)$ are conjugate residual vectors because we are now working in a complex space (Virieux and Operto, 2009) and the circular frequency is expressed by ω. The imaginary number is denoted by $\imath = \sqrt{-1}$. Implicit summation is considered on repeated capital indices.

Finally, let us mention that we also may need to estimate the source time signal. The deconvolution of the source wavelet is a simple operation in the frequency domain (Pratt, 1999) and a slightly more complex one in the time domain. The main point is the selection of the data time window for performing the deconvolution. We may select near-field data, far-field data, or ocean-bottom reflection data. Once this preliminary source wavelet is built, we will use anticausal mute and/or band-pass filtering for finalizing our estimation of the source wavelet. Readers interested in source extraction also can consult the article by Plessix and Cao (2011), who discuss using an adjoint formulation. That could be done for each source and at each iteration, if needed. We may assume that it should be done once at the beginning and could be the same for all sources. These different strategies are application-dependent.

Resolution analysis and formal uncertainty estimation

Although the relation between seismic data and medium properties is essentially nonlinear, we can expand the misfit function around the expected global-minimum model $\tilde{\mathbf{m}}$. The misfit function has a Taylor expansion

$$\mathcal{C}(\mathbf{m}) \approx \mathcal{C}(\tilde{\mathbf{m}}) + \nabla\mathcal{C}(\tilde{\mathbf{m}})(\mathbf{m}-\tilde{\mathbf{m}}) + \frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^t\mathcal{H}(\tilde{\mathbf{m}})(\mathbf{m}-\tilde{\mathbf{m}}) \quad (31)$$

reducing to a quadratic expansion form

$$\mathcal{C}(\mathbf{m}) \sim \mathcal{C}(\tilde{\mathbf{m}}) + \frac{1}{2}\big(\mathbf{m}-\tilde{\mathbf{m}}\big)^t\mathcal{H}(\tilde{\mathbf{m}})\big(\mathbf{m}-\tilde{\mathbf{m}}\big) \quad (32)$$

at the minimum we have reached by zeroing the gradient vector $\nabla\mathcal{C}(\tilde{\mathbf{m}})$. There is no guarantee that it will be a minimum and more specifically the global minimum we are seeking. At the current local minimum, the Hessian matrix should have real positive eigenvalues with some of them near or equal to zero. We always try to mitigate the influences of these small eigenvalues during the optimization procedure. We shall assume Gaussian statistics, a basic hypothesis in the framework of the Bayesian formulation (Tarantola, 1987). From the Taylor expansion (equation 32), the probability density

$$pd(\mathbf{m}) \propto e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^t\mathcal{H}(\tilde{\mathbf{m}})(\mathbf{m}-\tilde{\mathbf{m}})} \quad (33)$$

expresses the statistical distribution of possible models. Let us repeat that the gradient vector is zero at the minimum and, therefore, is not in the probability density function when expressed at the minimum. We assume that seismic traces are not correlated and are of the same quality. Moreover, let us assume that the statistics on the data are Gaussian and the units are normalized, giving us a data variance of one for each data value. Assuming that we have reached the global minimum, the Hessian matrix is the inverse of the posterior covariance matrix evaluated at the minimum $\tilde{\mathbf{m}}$ and, therefore, it will include the information for formal resolution and uncertainties as well as tradeoffs (Fichtner and Trampert, 2011). Using a singular value decomposition (SVD) of the Hessian matrix, we may deduce that the probability density will be the product of the probability densities for each model parameter: a rather simple Bayesian structure that is of limited validity in practical applications for a reliable uncertainty estimation.

Probing the structure of the inverse of the Hessian matrix will provide a resolution analysis as well as an uncertainty estimation, but that is not an easy task because the Hessian matrix has a complexity that increases as the square of the model complexity. The recursive expression of the inverse of the Hessian for the limited-memory Broyden-Fletcher-Godlfarb-Shanno (BFGS) algorithm will allow a discrete estimation of resolution and uncertainty under the quadratic approximation. Of course, we have to store or compress the Hessian matrix, which has mainly a block-band-diagonal structure, because neighboring points should interact whereas far-apart points are less expected to interact. Another approach will be through the second-order adjoint formulation with matrix-free operations. At the minimum, we perturb the gradient at the position in which we are interested, and we solve the restricted Newton equation with the gradient restriction at the selected position. The resulting model vector will provide the resolution and uncertainty around the selected position.

Another numerical alternative comes from the low-rank approximation of the Hessian matrix, which is supposed to have a pseudo-differential structure inducing a sparse shape of this matrix. Therefore, through randomized sampling, one can probe the most dominant eigenvalue components of the Hessian matrix itself (Candes and Demanet, 2005; Demanet et al., 2012), leading to fair approximations of the Hessian matrix (Herrmann, 2010; Fang et al., 2014). We are still operating under the quadratic assumption.

Because the true misfit function differs from a quadratic shape, we can introduce the resolution matrix

$$\tilde{\mathbf{m}}(\mathbf{x}) = \mathcal{R}(\mathbf{x},\mathbf{x}')\mathbf{m}(\mathbf{x}'). \quad (34)$$

Ideally, the desired resolution matrix $\mathcal{R}$ should be the identity. We expect not to be too far away from a diagonal structure and, therefore, we expect interaction only between neighboring elements. This will lead to a low-rank approximation, in which the complexity of the matrix $\mathcal{R}$ will be $k\mathcal{O}(n)$, where the integer k defines the halo around the current parameter when the model

is characterized by n degrees of freedom. Numerical probing comes with a finite small perturbation δm^j of one parameter j away from the minimum $\tilde{\mathbf{m}}$. Synthetic data at the new model will differ from synthetic data at the minimum and, therefore, could be used for a model reconstruction that provides an estimated perturbation $\delta\tilde{\mathbf{m}}_i$ of parameters with the same expected resolution matrix defined at the minimum. We build numerically one line of the resolution matrix

$$\mathcal{R}_{ij} = \frac{\delta \tilde{m}_i}{\delta m^j} \tag{35}$$

and we may proceed with another perturbation j over positions and parameter classes, provided we can store the matrix. We have the capacity of computing the spread function $SPREAD(\mathcal{R}) = ||\mathcal{R} - I||^2$ over the medium. The *SPREAD* function is a compact quantity for quality control of probed zones of the medium. Moreover, we can estimate the probability density as well as tradeoffs between parameters. We may quantify somehow the trustworthiness of the formal uncertainty and resolution by finding the Gaussian distribution of $\delta\tilde{\mathbf{m}}$ for the specified initial perturbation δm^j. Of course, we may sample this resolution matrix coarsely, especially on regions of interest.

The difficulties encountered with sampling the Hessian matrix accurately, and the question of how accurate the quadratic approximation is when considering high-dimensional spaces, lead to cautiousness in promoting these formal quantities. The FWI is a rather deterministic approach providing the "best" model at a given level without providing insight on quantified resolution and uncertainty. In the future, a more sophisticated sampling strategy based on ensemble methods (Evensen, 2009) could provide more-quantitative assessments of resolution and uncertainty. Such a strategy would require significant computer resources.

An example of a real application

Real applications of FWI have been successfully achieved on various seismic data sets. Currently, FWI is implemented mostly in the time domain because a wide range of applications with various survey sizes and different acquisition geometries can be investigated, in both exploration seismics and seismology (e.g., Komatitsch et al., 2002; Vigh and Starr, 2008; Mulder and Plessix, 2008; Krebs et al., 2009; Plessix and Perkins, 2010; Sirgue et al., 2010; Routh et al., 2011; Peter et al., 2011; Plessix et al., 2012; Bansal et al., 2013; Vigh et al., 2013; Fichtner et al., 2013; Schiemenz and Igel, 2013; Warner et al., 2013a; Zhu et al., 2015). Such applications are generally quite demanding on computational resources. We would like to focus on a specific environment where the FWI has been demonstrated to perform quite well both in the frequency domain (Operto et al., 2015) and in the time domain (Warner et al., 2013a).

A marine environment for characterizing reservoirs below sedimentary layers using permanent ocean bottom cables (OBC) or nodes provides a good configuration for FWI applications. There is a high density of sources and the recording at the sea bottom provides four-component data. We shall illustrate FWI high-resolution imaging on a real OBC data set from the North Sea (Operto et al., 2015). This acquisition is a fixed-spread with wide-azimuth configuration. For this illustration, the forward modeling is a 3D viscoacoustic vertical-axis transverse isotropy (VTI) finite-difference simulation in the frequency domain (Operto et al., 2014). Only the vertical velocity is reconstructed. Four-component OBC acquisition comprises ~2300 receivers and ~50,000 shots in the Valhall oil field (Barkved and Heavey, 2003). Only the hydrophone component is taken into account here, because we consider just the acoustic approximation. The acquisition layout covers an area of 145 km^2 and the maximum depth of the subsurface model is 4.5 km. Data are inverted in the 3.5- to 10-Hz frequency band.

Warner et al. (2013a) processed successfully a similar data set in the same kind of environment using a time approach. In addition to other substantiating evidence, high-quality images have also been obtained and give us further confidence in the usefulness of FWI. Readers interested in an extensive discussion of the FWI application in the Valhall oil field should consult the paper by Operto et al. (2015). Here we provide a brief discussion of the image improvements achieved by fitting the whole content of seismograms.

We need to start from an initial model. A vertical-velocity model and Thomsen's parameter models were built by reflection traveltime tomography and were provided to us (courtesy of

BP). A density model is built from the initial vertical-velocity model by using Gardner's law, and it is kept fixed over at least ten iterations. Thomsen's parameters are also kept fixed. A homogeneous model of the quality factor is used below the sea bottom, with a value of $Q = 200$. Starting from this initial model, we sequentially invert the vertical velocity for 11 frequencies, beginning with 3.5 Hz and ending with 10 Hz. We have adapted the finite-difference grid to each frequency. Note that the final grid size of 35 m accurately matches the sea bottom at a depth of 70 m. We compare horizontal slices at three different depths: 175 m (Figure 6a and 6d), 500 m (Figure 6b and 6e), and 1000 m (Figure 6c and 6f) in the initial tomographic velocity model and in the final FWI model.

In Figure 6d, one can see an improved reconstruction of paleochannels stored in the superficial sedimentary layer, as well as the imprints of glaciers sliding over the bedrock in Figure 6e. Both of these features are difficult to see in the tomographic images. The gas cloud is identified in the tomographic image (Figure 6c), but the reconstruction in Figure 6f by the FWI gives an unprecedented sharpness of contours, because of the broadband resolution we achieved. We conclude that the resolution has been significantly improved by FWI compared with that from traveltime tomography.

Comparing seismograms of real and synthetic data also illustrates the capacity of FWI to extract more information (Figure 7). Synthetic seismograms in time, computed by a time forward

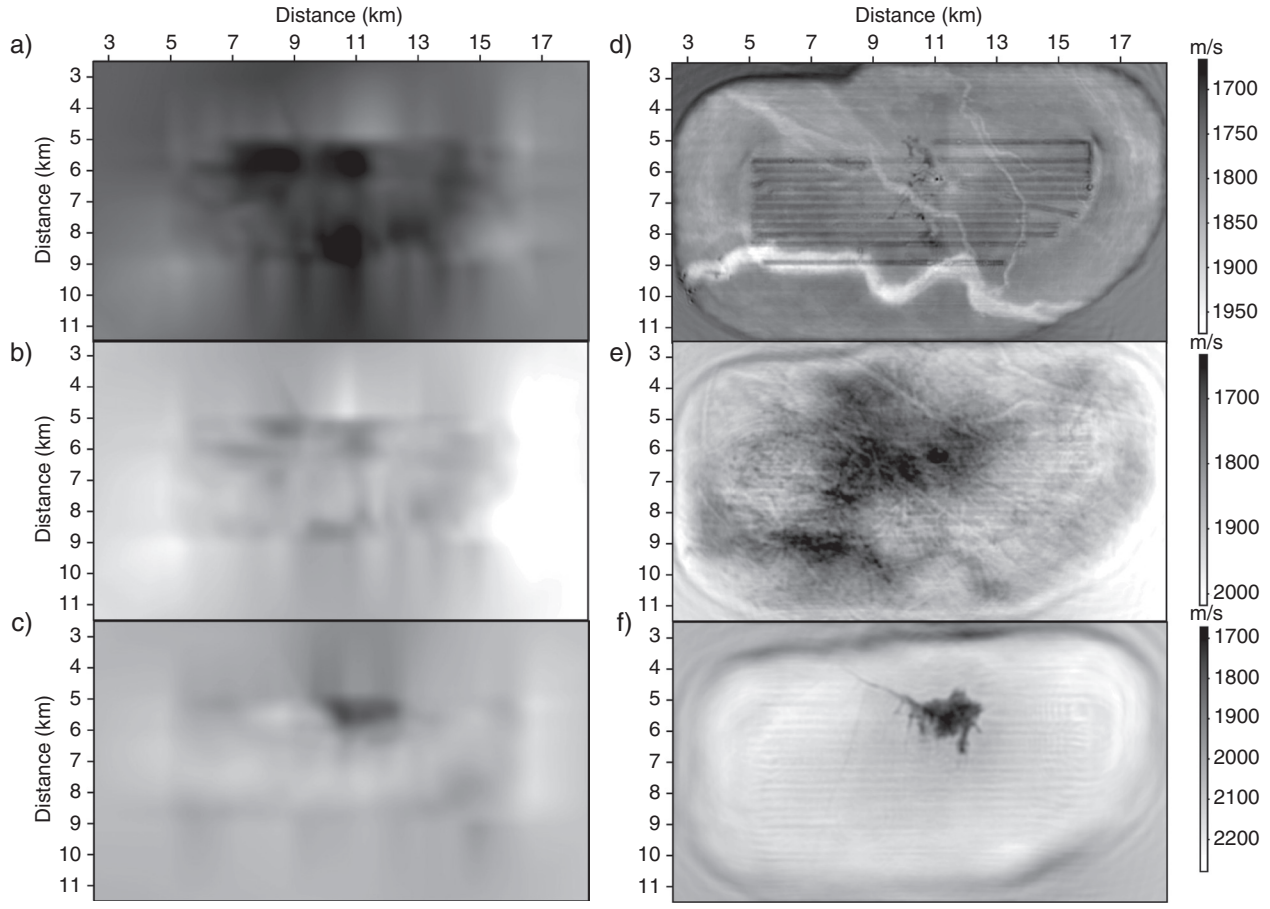

Figure 6. Initial model in the left panels and final model in the right panels. Horizontal sections are shown at depths of 175 m, 500 m, and 1000 m. Note the significant increase of resolution from left to right. The paleoriver features are nicely visible in (d) but cannot be seen in (a), at a depth of 175 m. In (d), one can identify also the strong imprint of the acquisition. At a depth of 500 m, imprints on bedrock of former glaciers are focused in (e) but are not visible in (b). Finally, at a depth of 1000 m, very precise contours of the gas cloud are clear in (f) but are completely blurred in (c). Acquisition imprints decrease with depth but are still present.

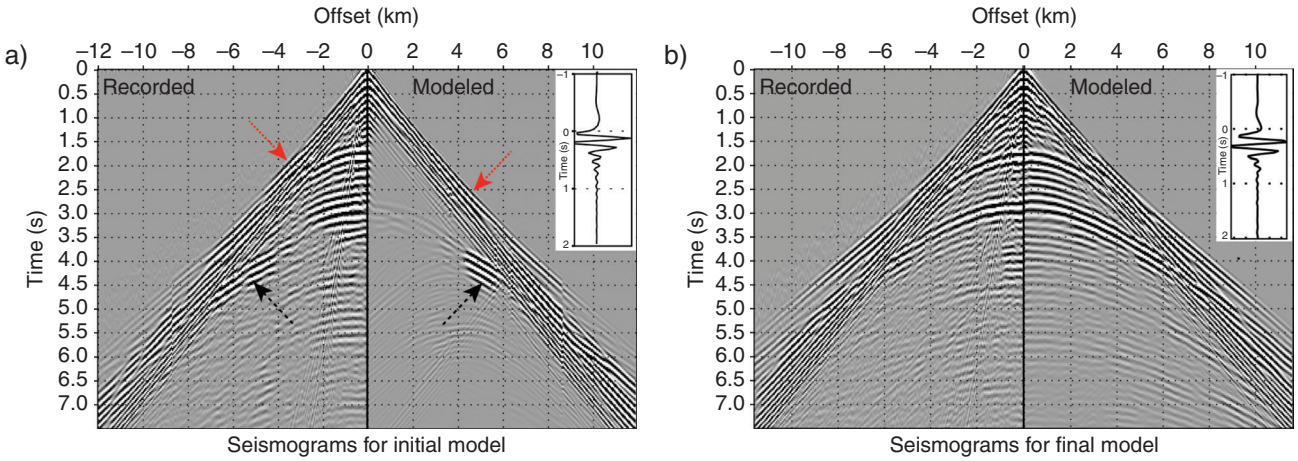

Figure 7. A common receiver gather in (a) the initial model and (b) the final model. Each panel has recorded traces and mirror-modeled traces. Red arrows identify first-arrival phases, and black arrows identify post-critical phases. Note the increasing extraction of information achieved by the FWI because of the better match between synthetic and real traces. The source signal is shown as an insert and differs between the initial and final models.

modeling, show improved agreement with real seismograms, highlighting the efficient extraction of information that FWI has achieved. Let us point out the slightly different amplitudes for reflection phases in this figure, which might suggest that density or impedance should be introduced and should direct us toward multiparameter inversion.

The high-resolution reconstruction with FWI is linked to its significant sensitivity to the initial model, which should be carefully designed. We discuss the cycle-skipping issue in the next section and explain that we cannot escape the fundamental ambiguity related to local minima when considering local minimization strategies. We must introduce some prior information that could be pertinent for the seismic data we consider, but we cannot guarantee that it will work as desired for any specific data sets.

Cycle skipping and initial model design

Cycle skipping in FWI originates from the oscillatory nature of seismic data. Counting wriggles from the source is impossible, taking into account the seismic data acquisition protocol. We may rely on velocity structures for predicting, with a time accuracy below half of the signal period, where the synthetic wriggle should be relative to the data wriggle in the time recording. If the initial velocity is not accurate enough, we face a cycle-skipping ambiguity and we may be trapped into a local minimum during the optimization procedure by attempting to fit the wrong wriggle. This is especially true when the misfit function is the ℓ_2 norm and we are comparing a data sample and a synthetic sample of a given record.

Therefore, we should rely on robust methods such as traveltime tomography for constructing initial models. Because traveltimes and phases for dispersive waves are robust observables, we should rely on them as long as we are able to extract them from seismograms because they are separable. Phase picking, the most challenging step of initial velocity reconstruction, can be performed in the data domain or in the image domain (Lambaré, 2008). The resulting pickings are interpreted during the related inversion procedure, which is far less difficult and intensive than the one for FWI. Strong and expert human inputs are required for driving the volumetric picking (threshold, cleaning, and peeling off the observables) and for specifying the identification and the interpretation of the analyzed phase. Even if the inversion is less non-linear, we need various penalty strategies in this tomographic inversion step of the velocity model construction. Note the significant humantime consumption in this workflow.

Two other strategies, based on more autonomous workflows, have emerged for overcoming the issue of cycle skipping. They are based primarily on misfit-function designs and/or model-space definitions.

The data domain strategy is mainly driven by the design of more robust objective functions that operate essentially by mitigating the oscillations of the seismic signal. Phase-only approaches have been investigated by Bednar et al. (2007), whereas analysis of the envelope has been suggested by Bŏzdag et al. (2011), and analysis of the instantaneous phase has been proposed by Maggi et al. (2009) and Lee and Chen (2013). Unwrapping or dynamic warping (Choi and Alkhalifah, 2011; Alkhalifah and Choi, 2012; Ma and Hale, 2013; Perrone et al., 2015) is another way of comparing signals. Crosscorrelation (Luo and Schuster, 1991; Tromp et al., 2005; van Leeuwen and Mulder, 2010) is natural, and zero-lag crosscorrelations (Routh et al., 2011) are also time series operations that mitigate amplitude-variations. Finally, deconvolution also reduces oscillations and is promoted by Luo and Sava (2011) and Warner and Guasch (2014), whereas integration, as another oscillation reduction tool, has been proposed by Donno et al. (2013). Finally, let us mention optimal transport as a new way of comparing seismic signals or gathers (Engquist and Froese, 2014; Métivier et al., 2016a, b). These different techniques modify the misfit function and require further investigation, especially when considering complex media; some limitations are expected and determining where these techniques perform efficiently will require further work.

The other strategy for the misfit-function design is an unphysical extension of the model domain in order to mitigate any errors or biases resulting from the model description. Introducing subsurface offsets and time-shifts is a method for relaxing imaging conditions because the model is not yet built correctly and prevents focusing of the incident and adjoint fields at the point of interest (Symes, 2008; Rickett and Sava, 2002; Sava and Vasconcelos, 2009; Biondi and Almomin, 2012; Almomin and Biondi, 2012; Biondi and Almomin, 2013, 2014). Alternatively, constrained optimization based on a penalty strategy will extend the model domain by integrating wavefields as components of the model space. Consequently, the wave equation is not required to be verified exactly by wavefields at each iteration. We update iteratively both the wavefields (for verifying the wave equation) and model parameters (for fitting the data) (van Leeuwen and Herrmann, 2013).

Whatever the misfit-function strategies are, they will introduce prior knowledge regarding how to interpret different phases and, therefore, a specific investigation should be performed to better understand that prior information. Often, these guided strategies will not identify clearly what are the assumptions enabling phase separability, phase identification, and phase interpretation.

In this article on FWI, we do not want to delve into strategies that would involve new concepts. But, one strategy complements the description of FWI we have presented. We will describe it and detail its relation to previous strategies (Chavent and Jacewitz, 1995; Plessix et al., 1999; Clément et al., 2001), sometimes in a broader context such as that of differential waveforms (Symes and Kern, 1994; Chauris and Plessix, 2012). We are convinced that this strategy, by mitigating inherent difficulties of FWI, will improve the reader's understanding of how FWI really works.

The primary contribution to the FWI gradient vector should come from the first Fresnel zone contribution (the so-called forward scattering) — essentially, from direct/diving waves. We should avoid the contribution from backward scattering, which requires a quite accurate velocity medium. During the hierarchical analysis of the seismic data, we have shown that other phases, such as reflections, provide forward-scattering information that is similar to information from direct phases. We should be able to distinguish this forward scattering from the backward scattering contribution of these reflected phases. Separation of these contributions also has been promoted by Snieder et al. (1989), Staal and Verschuur (2013), and Berkhout (2014), by using specific forward modeling. A new model description should be introduced by considering sharp contrasts of model parameters inside the volume to be imaged: we return to prior scale information assumed in basic methods of seismic imaging. We lose the unique concept of the single scattering formulation (the central hypothesis of FWI) for the interpretation of all phases, whatever they are. In this new formulation, we are back to smooth probing of the medium by waves, as assumed by traveltime tomography. The fundamental difference here is that picking is not required and the phase analysis is progressively updated.

To illustrate this strategy, we consider only acoustic propagation and leave the extension to the elastic case for future work. Readers interested in details can find them in Zhou et al. (2015). The sharp-constrast description $\delta\mathbf{m}$ performed through the P-wave impedance parameter, denoted by $\mathbf{IH}_p$, will absorb backward scattering, whereas the pixel velocity description $\mathbf{m}$, denoted by $\mathbf{VL}_p$, will take into account only the information brought by the forward scattering. Note that the impedance parameter is still pixelized along a line. In other words, this new model description is linked to a separation between forward and backward scatterings — a dramatically different strategy compared with the one driving the standard FWI. Splitting the forward and backward scatterings in a reflection phase comes with a new description of the medium, in which we consider the forward scattering as a reflection on its travel down to the interface and back up to the receiver. This is true also for diffractions and more-complex phases. This contribution is added to the one resulting from direct/diving waves. The gradient related to this smooth updating of the model $\mathbf{VL}_p$ is shown in Figure 8. It will contain essentially first Fresnel zones and, therefore, a low-wavenumber description of the velocity corresponding to smooth information. High-wavenumber content has mostly disappeared and is mitigated by selecting the impedance as the complementary parameter for absorbing this high-wavenumber content. This approach is the so-called joint full waveform inversion (JFWI) (Zhou et al., 2015).

We shall proceed in two steps, following alternative strategies promoted recently by Berkhout (2014). We begin with the buildup of the impedance $\mathbf{IH}_p$, starting from a rather simple initial model $\mathbf{VL}_p^0$ using the misfit function

$$\mathcal{C}(\mathbf{IH}_p) = \frac{1}{2}\left\| W^r\left(d_{obs}^{refl} - d_{pred}^{refl}(\mathbf{VL}_p^0, \mathbf{IH}_p)\right)\right\|_2^2, \quad (36)$$

in which we have selected just near-zero-offset data d_{obs}^{refl} that provide only migration isochrones to be compared with synthetic data d_{pred}^{refl}. Typically, the range of offsets from the source will be several hundred meters. We conduct the inversion through a small number of iterations because we are essentially interested in the impedance contrast. Once the impedance has been evaluated, we proceed in minimization

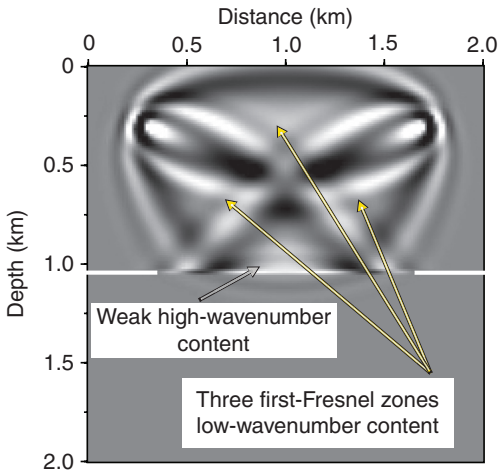

Figure 8. Gradient of the joint FWI, which is mainly composed of the first Fresnel zone between a source and a receiver and the two tilted, partially cut first Fresnel zones between a virtual source and a real receiver and a real source and a virtual receiver. The virtual positions are images of real positions with respect to the model interface, which is supposed to be known. A weak high-wavenumber content can be identified near the horizontal interface (white horizontal line) and will be mitigated by summing over sources and receivers during construction of the total gradient.

of the velocity $\mathbf{VL}_p$ by using the new misfit function

$$\mathcal{C}_{JFWI}(\mathbf{VL}_p) = \frac{1}{2}\left\| W^d\left(d_{obs}^{div} - d_{pred}^{div}(\mathbf{VL}_p)\right)\right\|_2^2 + \frac{1}{2}\left\| W^r\left(d_{obs}^{refl} - d_{pred}^{refl}(\mathbf{VL}_p, \mathbf{IH}_p)\right)\right\|_2^2, \quad (37)$$

in which we have made a rather rough separation between diving waves and reflected waves. The previously recovered impedance $\mathbf{IH}_p$ is kept fixed during this inversion step. Now we consider a broad range of offsets for the $\mathbf{VL}_p$ reconstruction and keep the impedance description fixed. This reconstruction provides a new smooth $\mathbf{VL}_p$ model. Therefore, we can update the impedance in this new $\mathbf{VL}_p$ model, and so on.

Figure 9 shows the true model, composed of a velocity structure $\mathbf{V}_p$ and an impedance structure $\mathbf{I}_p$ that we want to reconstruct, as well as the rather simple 1D medium we use for starting the inversion cycle in order to obtain the high-wavenumber impedance $\mathbf{IH}_p$ and the low-wavenumber velocity $\mathbf{VL}_p$. After this cycle is

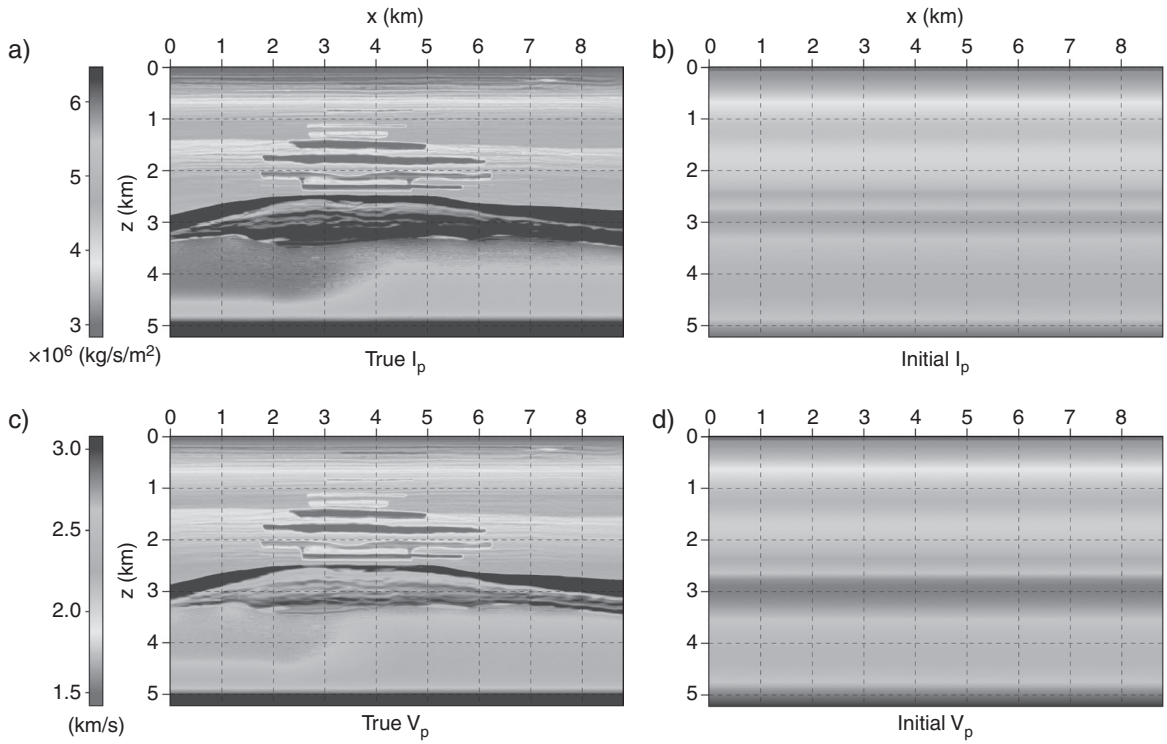

Figure 9. On the left, a synthetic Valhall example with a sedimentary layer, a stratified gas cloud above the reservoir, and a flat interface below the reservoir: (a) true impedance $\mathbf{I}_p$ and (c) true velocity $\mathbf{V}_p$.On the right, the 1D simple initial model: (b) initial impedance and (d) initial velocity.

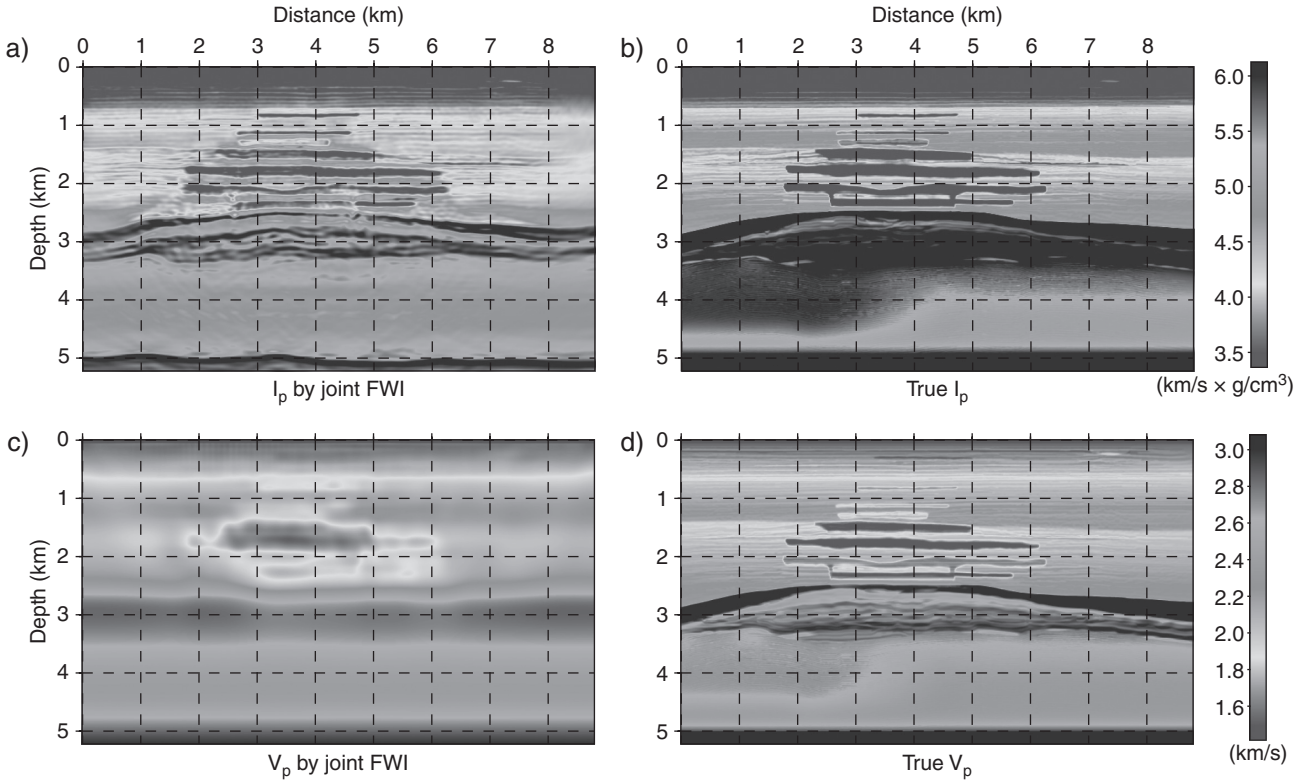

Figure 10. (a) The impedance $\mathbf{IH}_p$ after the cycling procedure; (b) the true impedance for comparison; (c) the velocity $\mathbf{VL}_p$, which has a mostly smooth content; and (d) the true $\mathbf{V}_p$ for comparison.

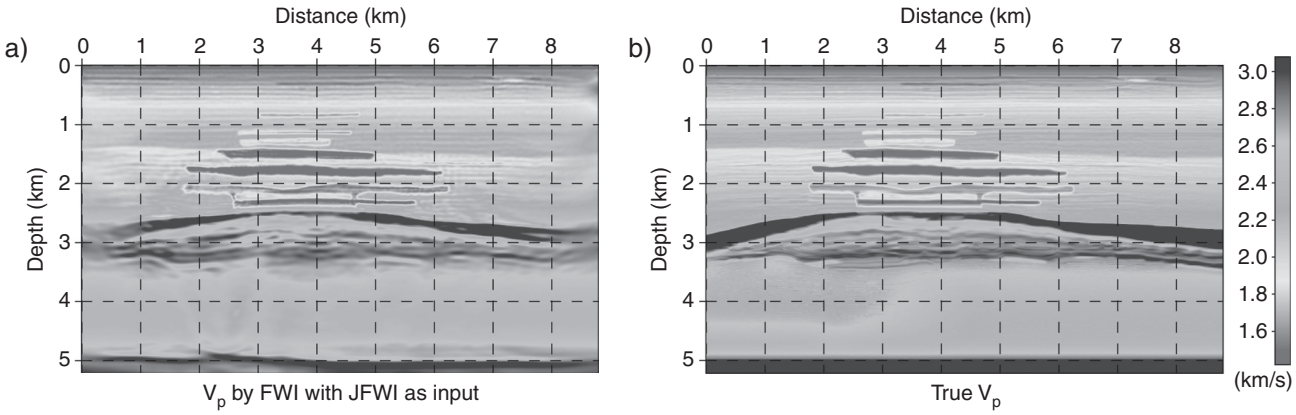

Figure 11. (a) The FWI velocity V_p and (b) the true velocity model. Observe the increase in the high-wavenumber content, but be aware that we have assumed very simple variations of density, as is often done for standard FWI.

repeated approximately 20 times, as shown in Figure 10, the result is a reconstruction of the impedance $\mathbf{IH}_p$ that has a strong high-wavenumber content. The reconstruction of the velocity $\mathbf{VL}_p$ has a low-wavenumber content that benefits from diving and reflected phases. This smooth velocity structure could be used as the initial velocity model in the standard full waveform inversion for an improved velocity reconstruction, assuming a simple density model as shown in Figure 11.

This workflow relies on a simple modification of the misfit function and assumes a rough phase separation between mostly horizontal propagation of diving waves and mostly vertical propagation of reflected waves. In the JFWI workflow, we have purposely mitigated any high-wavenumber content in the velocity reconstruction, because we chose the impedance parameter, which can absorb the high-wavenumber component of the reconstruction. By doing so, we have avoided the strong requirement of an accurate initial model when we are considering reflections in their reflection regime. Updating the impedance by considering near-zero offsets affects the high-wavenumber content again and puts it at the right place in the new smooth velocity model. For the computational aspect of such an approach, the cost is simply for the JFWI part of the whole cycle to be twice the standard FWI cost. The impedance portion should be considered as well and turns out to be fast because few iterations are required for constructing the new impedance model through a kind of least-squares migration.

Regularization and prior constraints

Sometimes we have a good understanding of what we expect as reconstructed models, either because we have other information such as sonic logs, stratigraphic data, or geologic constraints or because we have other remote-sensing approaches such as gravimetry or electromagnetic prospecting. Taking these other approaches into account in the inversion process will ensure robust and consistent results (Tikhonov and Arsenin, 1977). We may consider integrating into the misfit function such prior information either by preconditioning the data gradient that reduces the model description (Fomel and Claerbout, 2003; Guitton et al., 2012) or by adding a model term in the misfit function (Asnaashari et al., 2013). Both approaches are pertinent and complement nicely the purely data-driven strategy often applied for the FWI.

We shall now describe the model-driven aspect of the FWI. (Asnaashari et al., 2013), because recent investigations of the FWI workflow have focused more on the data-driven aspect and have neglected the model strategy. The new misfit function will be defined by the expression

$$\mathcal{C}(\mathbf{m}) = \mathcal{C}_d(\mathbf{m}) + \lambda_1 \mathcal{C}_{1_m}(\mathbf{m}) + \lambda_2 \mathcal{C}_{2_m}(\mathbf{m}), \qquad (38)$$

where the Tikhonov term is denoted by $\mathcal{C}_{1_m}(\mathbf{m})$ for inversion robustness and the prior model misfit term is denoted by $\mathcal{C}_{2_m}(\mathbf{m})$. The previous data misfit function is $\mathcal{C}_d$. Two regularization hyperparameters λ_1 and λ_2 are introduced, to allow weighting of the penalty terms with

respect to each other and to the data term. They have to be estimated either theoretically or numerically. Let us express these three terms in a more explicit way for the particular case of ℓ_2 norms. The data term is the one we have considered previously. The second term of the misfit function is the Tikhonov term and can be written as

$$\begin{aligned}\mathcal{C}_{1_m}(\mathbf{m}) &= \|\mathbf{B}_x\mathbf{m}\|^2 + \|\mathbf{B}_z\mathbf{m}\|^2 \\ &= \frac{1}{2}\{\mathbf{m}^T\mathbf{B}_x^T\,\mathbf{B}_x\mathbf{m} + \mathbf{m}^T\mathbf{B}_z^T\,\mathbf{B}_z\mathbf{m}\} \\ &= \frac{1}{2}\{\mathbf{m}^T\mathbf{D}\mathbf{m}\},\end{aligned} \tag{39}$$

where $\mathbf{B}_x$ and $\mathbf{B}_z$ are the first-order spatial derivative operator matrices with respect to x and z, respectively. In practice, they can be reduced to the second-order Laplacian operator $\mathbf{D}$. We use a classical five-point finite-difference stencil to implement the operator $\mathbf{D}$, which provides a connection between neighboring points. The third term of the misfit function is related to the prior model $\mathbf{m}_{\text{prior}}$, which can be designed from different information and could be set prior to the seismic inversion, but which also could be adapted iteratively during the inversion procedure. This so-called prior model norm term is computed using the expression

$$\begin{aligned}\mathcal{C}_{2_m}(\mathbf{m}) &= \|\mathbf{W}_m(\mathbf{m}-\mathbf{m}_{\text{prior}})\|^2 \\ &= \frac{1}{2}\left\{(\mathbf{m}-\mathbf{m}_{\text{prior}})^T\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m}-\mathbf{m}_{\text{prior}})\right\},\end{aligned} \tag{40}$$

where the matrix $\mathbf{W}_m$ is a weighting operator on the model space. This weighting operator has a dimensional complexity that is the square of the model complexity, as does the one for the Hessian matrix. This matrix also can be seen to be the inverse of the square root of the covariance matrix of the model and contains prior uncertainty information on the model parameters. Only diagonal structures of the $\mathbf{W}_m$ matrix, $diag(\mathbf{W}_m^T\mathbf{W}_m) = 1/\sigma^2(\mathbf{m})$, are considered here, although off-diagonal terms should play an important role. The prior weighting model $\sigma^2(\mathbf{m})$ contains both the prior model uncertainty (the variance) and an acquisition-correction function, which is to be designed.

Note that the model term of the misfit function is dimensionless, because of the introduction of the matrix $\mathbf{W}_m$. Hyperparameters λ_1 and λ_2 absorb data and Tikhonov dimension issues. The total gradient has a data component, coming from the data misfit $\mathcal{C}_d$ and evaluated by the adjoint method, and a model component, coming from the model misfit $\mathcal{C}_{1_m} + \mathcal{C}_{2_m}$, which yields

$$\gamma_{\text{model}} = \lambda_1\mathbf{D}\mathbf{m} + \lambda_2\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m}-\mathbf{m}_p), \tag{41}$$

an expression that is numerically straightforward to evaluate. When we are considering a model term of the misfit function, three difficulties arise: the definition of the prior model, the construction of the weighting function, and the computation of its inverse, which is the covariance. Because we consider only a diagonal matrix in this introduction, the inverse will be efficiently computed. To a lesser extent, estimation of hyperparameters adds an additional complexity, although numerical experiences provide reasonable values quite rapidly for most common configurations.

Let us consider a synthetic example with two gas-sand traps in a model for which the acquisition is deployed at the free surface as well as in wells (Figure 12). As shown in the figure, the initial model is a smooth version of the true model such that first-arrival waveforms are well fitted. We have assumed perfect knowledge of the source wavelet in order to emphasize the prior model component in this illustration. For real applications, many effects will contribute and will require skillful strategies. The prior model is a simple linear interpolation between the two sonic logs recorded in wells (Figure 12). The data in such a model will deviate significantly from the data to be fitted.

Designing the model weighting matrix is of crucial importance. We assume it to be diagonal in this article, for simplicity, although off-diagonal information is expected to be quite important. The variance along this diagonal will be the product of two components (Figure 13). One will come from the expected knowledge of velocity in nearby wells. This prior variance will increase as we move away from the wells. Moreover, we have to handle the acquisition configuration in our expected prior variance. This prior variance should be larger at depth than at the free surface, where the sources and some of the receivers are located. Let us repeat that the total variance is simply the product of these two designed prior variances, and the results are much more

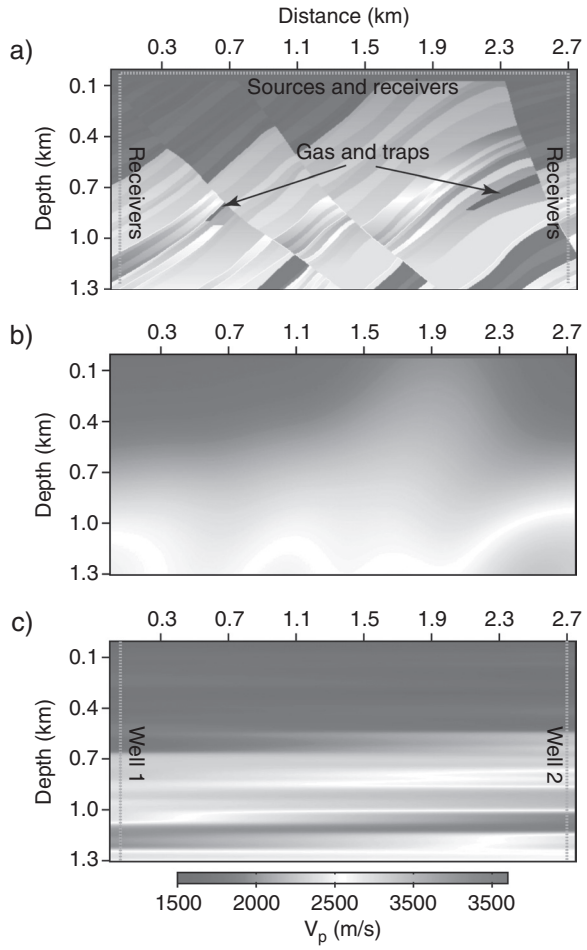

Figure 12. (a) The true model inspired by a Marmousi configuration, with shots at the free surface and receivers in two wells as well as at the free surface, (b) the smooth initial model, which allows us to have a rather good description of first-arrival waveforms, and (c) the prior model, which is a linear interpolation between sonic logs in the two wells. Data in this model are very different from real synthetic data.

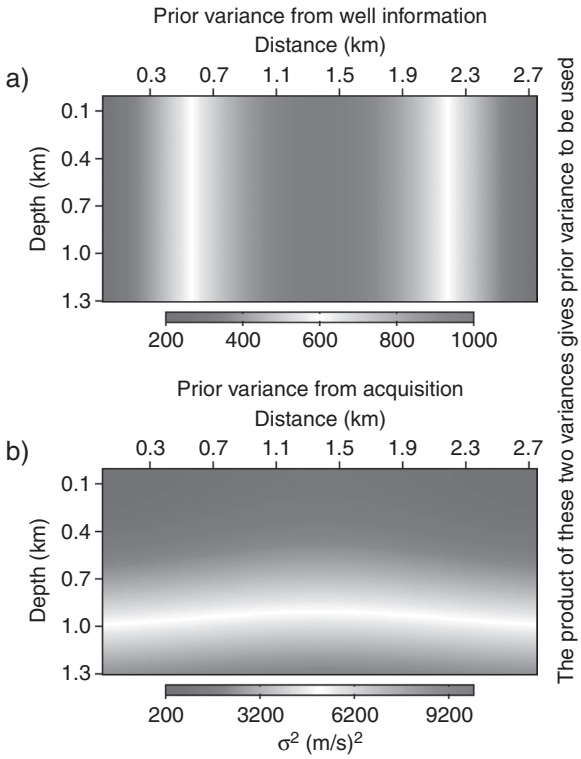

Figure 13. A prior weighting diagonal matrix as the product of the diagonal matrices shown in panels (a) and (b): the matrix in (a) is related to well information, with a small expected variance around wells, whereas the one in (b) is designed to take into account the acquisition geometry, with an increase of the variance with depth.

sensitive to the shape of this total variance than to its absolute values.

The hyperparameters λ_1 and λ_2 are adjusted empirically through a few trials. Moreover, we design an automatic decrease of the hyperparameter λ_2 through iterations based on the decreasing rate of the misfit function, in such a way that the final image is only related to the data misfit. We have used this dynamic tuning to steadily modify the misfit function during the inversion process. The initial data as well as the final data are shown in Figure 14. The final data are almost identical to the true data, which are not shown because it is not possible to see the differences in the data from this synthetic test.

Reconstructed models are displayed in Figure 15, where we can see the influence of the prior information helping the optimization to converge to the correct minimum by modifying the search path. It does not mean that the optimization has allowed an increase of the total misfit function; it has simply taken another path for converging to another local minimum. It may have induced locally an implicit jump in the data component of the misfit function.

Speeding up FWI

The full waveform inversion is a rather intensive task, and its application has been popularized primarily in the context of acoustic propagation. Even for that case, we may need to reduce the cost of the modeling or increase the convergence of the optimization at the expense of the model reconstruction.

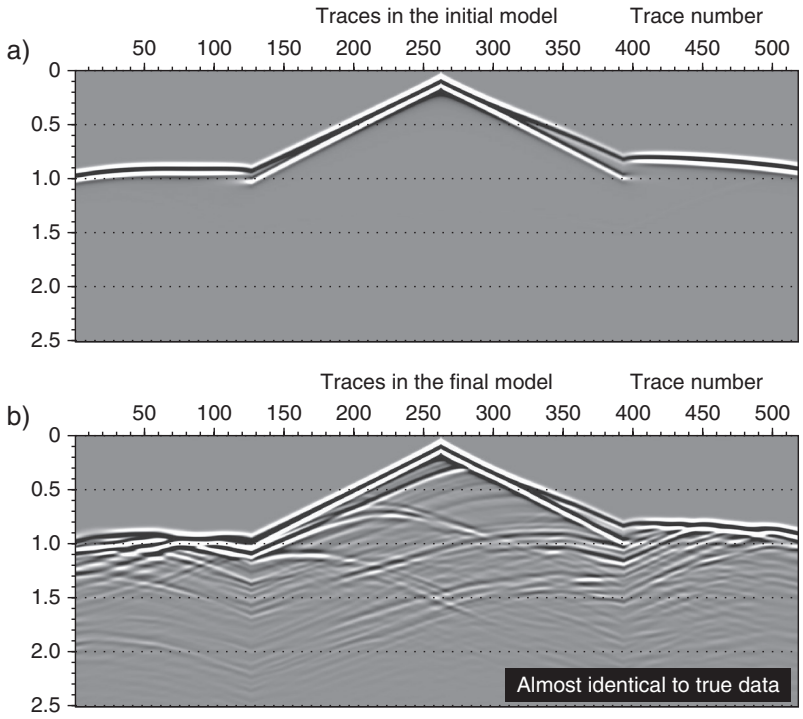

Figure 14. (a) The initial data and (b) the final model data. For this synthetic example, the model data are nearly identical to the true data, which are not shown here.

When FWI is performed in the frequency domain, sparse frequency sampling is possible, and criteria have been designed by Sirgue and Pratt (2004) for such selective sampling. Warner et al. (2013b) discuss in detail how to subsample sources when performing FWI in time and, benefiting from this investigation, Operto et al. (2015) discuss the pros and cons of the frequency and time approaches. Herrmann (2010) promotes randomized sampling to increase efficiency. These strategies organize in different ways the feeding of the optimization kernel by selected seismic data. They may reduce computer resource needs, with the goal of obtaining the same quality universally in the image reconstruction.

For a 3D target-oriented investigation under the acoustic approximation, a frequency approach with a direct solver is quite appealing and competes well with time approaches. Because the wave equation is a linear equation, we can simulate waves for more than one source at the same time, as has been proposed for migration (Romero et al., 2000). For simplicity, let us consider the frequency formulation of FWI for the acoustic pressure p, as described by Ben Hadj Ali et al. (2011), whereas Krebs et al. (2009) consider time approaches. The computational burden in FWI can be mitigated by summing encoded sources into supershots $\overline{\mathbf{s}} = \sum_{j=1}^{N_s} a_j \mathbf{s}_j$, where a_j is a complex number. The encoded signature is defined by the phase $a_j = \exp(\imath\phi_j)$ of modulus one. The adjoint source can be deduced with a similar summation. Because the conjugate of the coefficient a_j for each term of the adjoint source is considered, thanks to the linearity of the partial differential equations, the sum will be constructive for the source and its related adjoint source. The sum will be destructive for any unrelated source and adjoint source and will be considered as crosstalk. At position i in the mesh, the gradient of the parameter m_l can be expressed as

$$\left(\frac{\partial \mathcal{C}(\mathbf{m})}{\partial m_l}\right)_i = \sum_{j=1}^{N_s} \mathbf{p}_{ij}^{\mathrm{T}} \frac{\partial \mathbf{M}}{\partial \mathbf{m}_{li}} \mathbf{r}_{ij}^{*} + \sum_{j=1}^{N_s} \sum_{\substack{k=1 \\ k\neq j}}^{N_s} a_{ij} a_{ik}^{*} \mathbf{p}_{ij}^{\mathrm{T}} \frac{\partial \mathbf{M}}{\partial \mathbf{m}_{li}} \mathbf{r}_{ik}^{*}, \tag{42}$$

where the adjoint wavefield $\mathbf{r}$ is composed of complex-conjugate quantities.

The first term in the right side of equation 42 corresponds to the standard gradient formed by stacking the contribution of each individual shot, whereas the second term corresponds to crosstalk

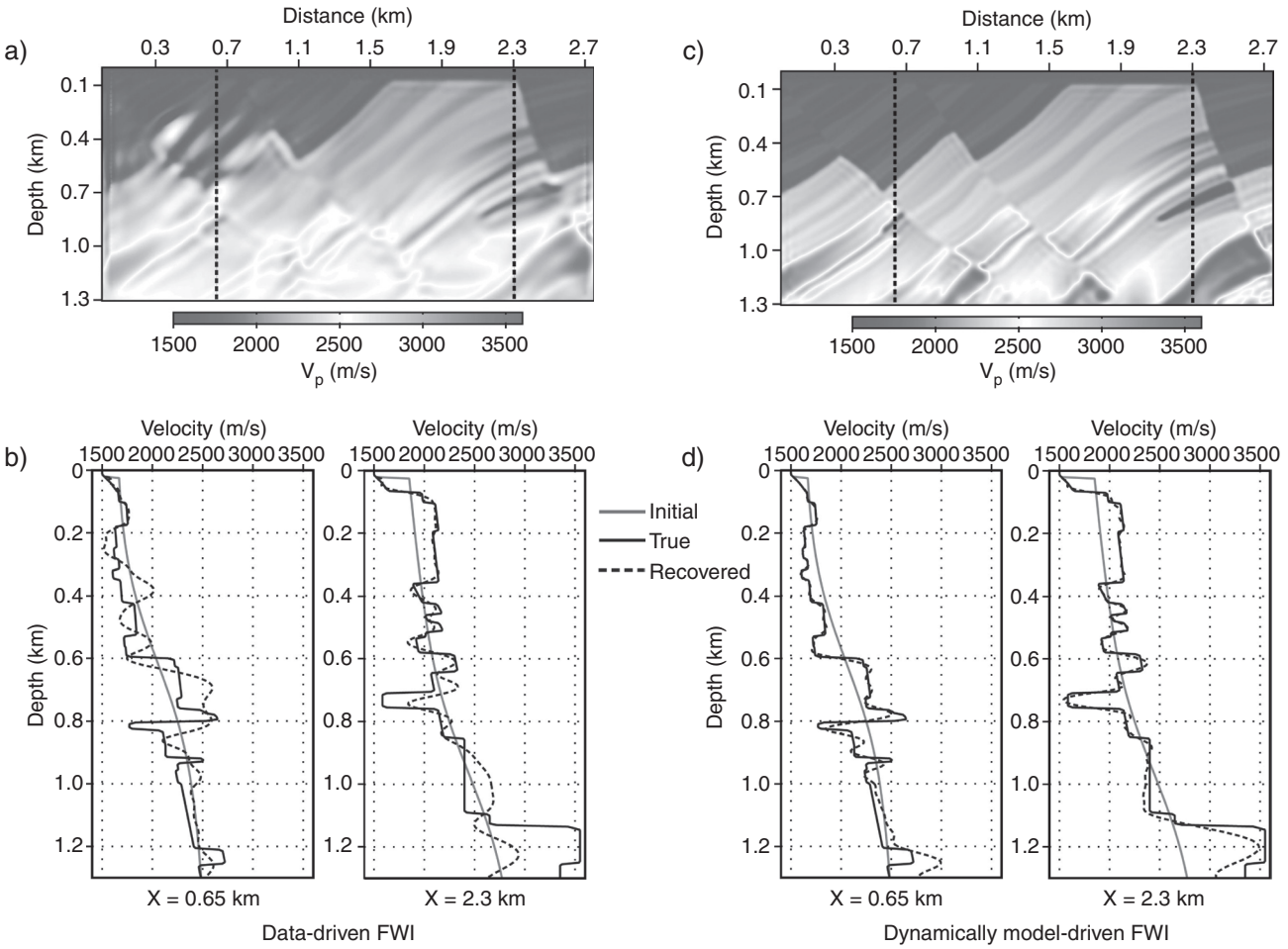

Figure 15. On the left, results without prior information, using only a data misfit term. As shown in panel (a), the left part of the medium image is locked into a secondary minimum. The gas-sand trap on the right side is recovered, whereas the smaller one on the left side is not resolved. Panel (b) shows sonic logs on two vertical dotted profiles of panel (a) and illustrates the difficulty of reconstruction, especially at depth. On the right, results are shown when dynamically prior information was used. As is shown in panel (c), the medium image is well reconstructed with the two gas-sand traps. (d) Sonic logs on two vertical profiles show the reconstruction improvement quantitatively.

interferences between sources j and k. This extra term, referred to as crosstalk noise in the following, alters the imaging result and, therefore, can be considered as noise. Therefore, the goal of the code phases ϕ_j is to minimize the crosstalk noise. Random codes seem to be quite efficient in reducing the noise over iterations (Schuster et al., 2011; van Leeuwen et al., 2011; Godwin and Sava, 2013). Of course, efficiency is difficult to analyze without considering the optimization engine. Castellanos et al. (2015) perform an investigation and conclude that a similar load for computer resources will be needed regardless of the technique used for the optimization. A slight advantage is assigned to the truncated Newton method, which seems to provide the final solution with the smallest variance. As noticed by Castellanos et al. (2015), convergence is improved because small local minima do not prevent the updating from proceeding (Martin et al., 2012). Coarsening the seismic data in a more or less random way also will mitigate the local-minima influence. For nonlinear problems, there is no formal proof of convergence of such random strategy. Finally, some authors suggest combining stochastic approaches when far from the minimum and moving to a deterministic approach when nearby (Friedlander and Schmidt, 2012; van Leeuwen and Herrmann, 2012).

Many strategies have been elaborated over the last 10 years to mitigate the computational burden of FWI. We shall see in the near future

different proposals for making FWI more affordable, especially for the elastic case.

Multiparameter FWI

Because seismic acquisition in the future will include increasing numbers of offsets, partition of energy between P-waves and S-waves should be considered in seismic imaging. Therefore, elastic wave propagation should be examined, although it is a challenging problem that dramatically increases computer demand. Moreover, we shall face the problem of inverting more than one parameter. Even in the case of acoustic propagation, we also need to consider other parameters such as density, attenuation, and anisotropy (Tarantola, 1986; Crase et al., 1990, 1992; Plessix, 2006a; Epanomeritakis et al., 2008; Plessix and Cao, 2011; Prieux et al., 2013a; Gholami et al., 2013; Plessix et al., 2014, among others). For elastic cases, numerical illustrations are still quite simple and real applications remain rare (Prieux et al., 2013b; Schäfer et al., 2013; Vigh et al., 2014; Stopin et al., 2014; Borisov et al., 2015).

Operto et al. (2013) discuss the challenges of multiple parameters. It is crucially important to identify strategies that allow us to extract information on multiple parameters that have different imprints in the recorded data. Alternative inversions are an often-used technique that may provide adequate convergence when we are able to identify a related hierarchy in the data with respect to the different parameters. If not, we need to invert these different parameters simultaneously and, therefore, proper balancing among the parameters should be performed even at the initial stage of the inversion. Any incorrect projection onto a parameter related to a tradeoff between parameters will introduce features that will be very difficult to remove at later iterations in the update model. Another pertinent strategy for mitigating such a tradeoff is based on the subspace method, which takes into account local projection onto a subspace of model parameters (Kennett et al., 1988; Kennett and Sambridge, 1998; Baumstein, 2014, 2015). In addition to the subspace method, we need to consider constraints on this multiparameter reconstruction as, for example, projections onto convex sets (POCS) (Baumstein, 2013).

In a multiparameter framework, we are interested in the reconstruction of several classes of parameters $\mathbf{m}_j$, with $j = 1, M$, where M is the number of parameter classes to be reconstructed. Using these notations and following equation 6, the gradient of the misfit function with respect to the parameter class j is given by

$$\frac{\partial \mathcal{C}}{\partial \mathbf{m}_j}(\mathbf{x}) = \int_0^{T_w}\int_{\partial\Omega_s}\int_{\partial\Omega_r} \frac{\partial \mathbf{d}^{\mathbf{m}}_{\text{syn}}}{\partial \mathbf{m}_j}(\mathbf{x}_s,\mathbf{x}_r,\mathbf{x},t) \times (\mathbf{d}^{\mathbf{m}}_{\text{syn}}(\mathbf{x}_s,\mathbf{x}_r,t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_s,\mathbf{x}_r,t))d\mathbf{x}_s d\mathbf{x}_r dt. \tag{43}$$

Equation 43 is particularly interesting because it reveals that the gradient of the misfit function with respect to parameter class j is computed as the sum over the sources and receivers of the zero-lag crosscorrelation in time of the residuals with the partial derivatives of the synthetic data with respect to parameter class j.

The latter quantity can be expressed as

$$\frac{\partial \mathbf{d}^{\mathbf{m}}_{\text{syn}}}{\partial \mathbf{m}_j} = d_{\text{syn}}(\mathbf{m} + \mathbf{dm}_j) - d_{\text{syn}}(\mathbf{m}) + o(\|\mathbf{dm}_j\|^2), \tag{44}$$

in which $\mathbf{dm}_j$ is a perturbation of the model parameter j. The partial derivatives of the synthetic data with respect to parameter class j correspond to the perturbation of the signal recorded at the receivers, which would be caused by the introduction of a perturbation $\mathbf{dm}_j$ in the first-order approximation. This is the single scattering approximation — the perturbed signal can be interpreted as the one generated by the scatterer $\mathbf{dm}_j$ acting as a secondary source.

From equations 43 and 44, we see that the only differences in the updates brought by the gradient to the different parameter classes come from the variation of the perturbed signal with the parameter class. If two parameter classes have similar scattering responses, the gradient of the misfit function does not allow us to distinguish between these two parameter classes. This is what is often referred to as tradeoff or crosstalk between parameter classes.

Of course, the structure of the Hessian operator will depend on the parameter selection, which can mitigate the tradeoff interaction among parameters (Innanen, 2014). We may also look at the expression of the Hessian operator. For the sake of simplicity, we will only consider the Gauss-Newton approximation. In a multiparameter context, this matrix is composed

of M^2 blocks with a symmetric structure. The ijth block is given by

$$(H_{\mathrm{GN}})_{ij}(\mathbf{x},\mathbf{x}') = \int_0^{T_w}\int_{\partial\Omega_s}\int_{\partial\Omega_r} \frac{\partial \mathbf{d}^{\mathbf{m}}_{\mathrm{syn}}}{\partial \mathbf{m}_j}(\mathbf{x}_s,\mathbf{x}_r,\mathbf{x},t) \times \frac{\partial \mathbf{d}^{\mathbf{m}}_{\mathrm{syn}}}{\partial \mathbf{m}_i}(\mathbf{x}_s,\mathbf{x}_r,\mathbf{x}',t)\,d\mathbf{x}_s\,d\mathbf{x}_r\,dt. \quad (45)$$

Now we see that block ij of the Gauss-Newton operator is computed as the sum over all sources and receivers of the zero-lag crosscorrelation in time of the signals scattered by perturbations of the parameter classes i and j. As such, it is a measure of the correlation between the scattering response of parameter class i and that of parameter class j. This reveals that the inverse Hessian operator can act as a decoupling operator on the gradient and should help remove crosstalk artifacts.

In order to illustrate these basic concepts of multiparameter inversion, we consider the simple example of 2D acoustic-frequency-domain FWI for simultaneous reconstruction of the P-wave velocity and density. The scattering response (or radiation pattern) of these two parameters for a given point of the medium in the acoustic approximation is illustrated in Figure 16. As can be seen at the scattering point, the radiation pattern of the P-wave velocity is isotropic. The radiated energy has the same amplitude in all directions. Conversely, the radiation pattern of the density at the same scattering point is directional because the energy of the scattered signal is mostly concentrated at short-angle illumination. From these two diagrams, one can thus expect that the coupling between the two parameters will be stronger if the seismic acquisition concentrates on small illumination angles, whereas it should be less severe if the acquisition spans a broader range of illumination angles.

Now consider the following tutorial experiment. A square homogeneous background is perturbed in its center with two nonoverlapping inclusions of P-wave velocity and density, respectively (Figure 17). Seismic data are acquired on this perturbed medium. The initial model is set to the background homogeneous model, and we attempt to simultaneously recover the two perturbations from the seismic data by using multiparameter FWI.

Two acquisition configurations are considered. The first is a fixed-spread surface acquisition with sources and receivers on top of the domain only (with just small illumination angles). The second follows a fixed-spread full-illumination design with sources and receivers all around the

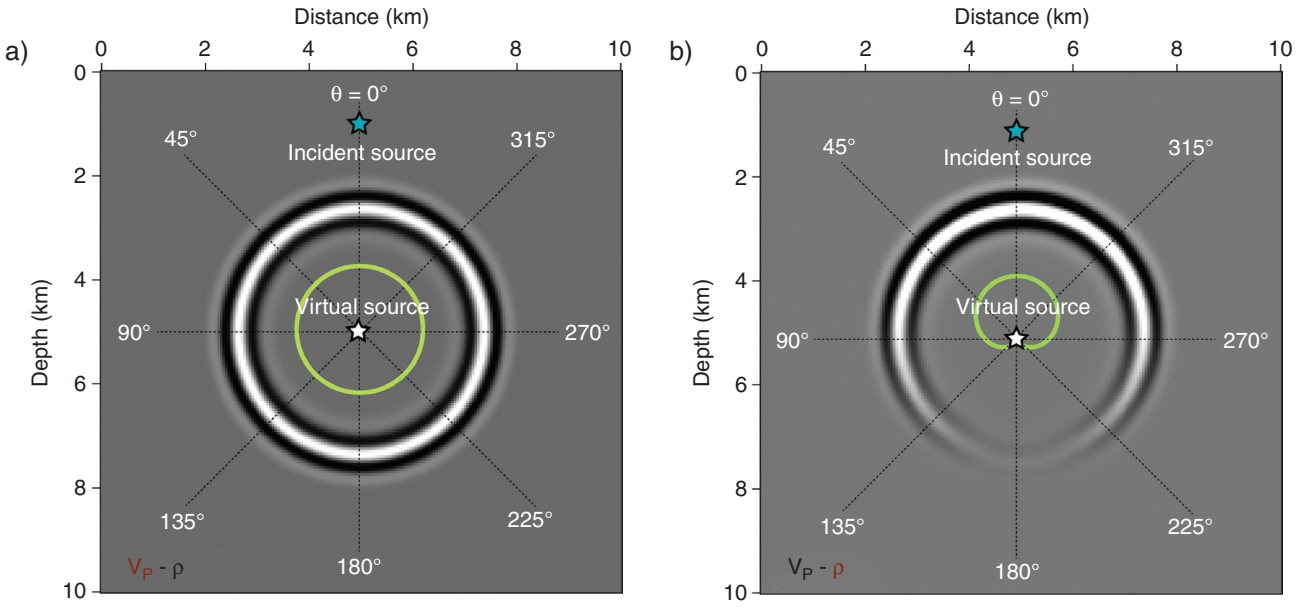

Figure 16. Radiation pattern at the scattering point for (a) P-wave velocity and (b) density, in the acoustic approximation. Whereas the radiation pattern of the P-wave velocity is isotropic (the energy is scattered with the same intensity in all directions), the radiation of the density is directional. The energy is mostly scattered for short illumination angles.

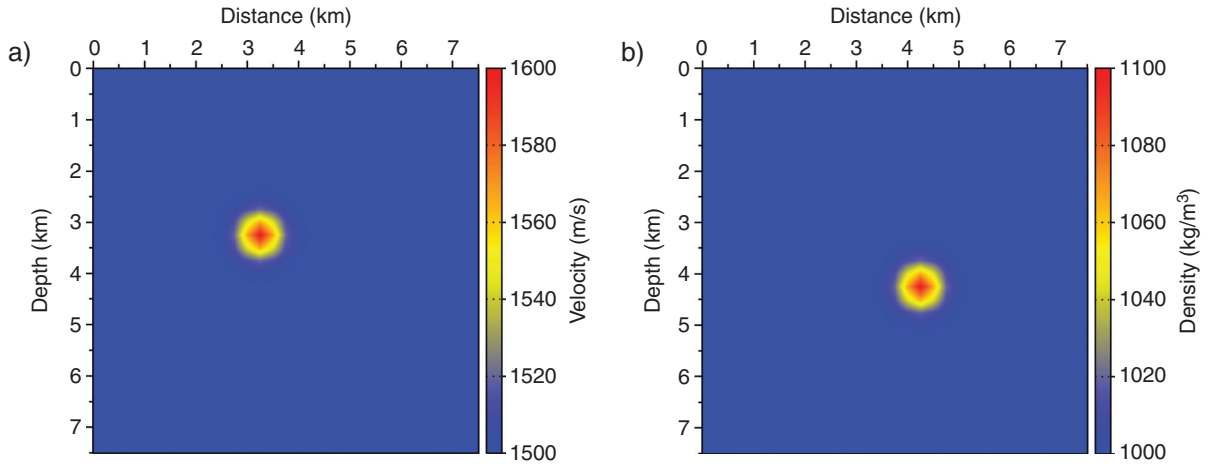

Figure 17. Perturbed (a) P-wave velocity and (b) density models used for the multiparameter tutorial test.

target zone (with a broad range of illumination angles). In Figure 18, the model perturbation deduced from gradients with respect to the P-wave velocity parameter and the density parameter in the surface-acquisition case are presented. As could be expected, the similarity of the radiation patterns of these two parameters for small illumination angles makes their gradients almost identical.

In Figure 19, the model perturbation deduced from gradients with respect to the P-wave velocity parameter and the density parameter are presented in the full-acquisition case. Despite a broader range of illumination angles, the P-wave velocity gradient contains a strong imprint of the density perturbation. The density gradient contains two perturbations of equal intensity: one at the position of the P-wave velocity and one at the position of the density. Compared with the surface illumination case, the geometric shape of the full-acquisition perturbation is better recovered. However, the tradeoff between the two perturbations is still strong.

The Gauss-Newton operator is computed for the two acquisition setups and is presented in Figure 20. In both cases, the upper diagonal block corresponds to the second-order derivatives with respect to the P-wave velocity, whereas the lower diagonal block corresponds to the second-order derivatives with respect to the density. The off-diagonal blocks are identical, as a result of the symmetry of the Hessian operator. They correspond to the cross derivatives with respect to P-wave velocity and density (see equation 45). As can be seen, the amplitude of the P-wave velocity block is larger in the full-illumination setup. This is due to the similar sensitivity of this parameter to all illumination angles, whereas the density is only sensitive to short illumination angles. Because the Gauss-Newton operator is built as a sum over source and receiver pairs, the contribution of large-offset pairs accumulates for the P-wave velocity, whereas the corresponding contribution for the density is almost negligible. Conversely, for surface acquisition, the amplitudes of the blocks are comparable, because only short illumination angles contribute to these blocks.

To further analyze the differences between the two configurations, we perform a singular value decomposition (SVD) of the two operators (full illumination versus surface illumination). The singular value distributions are presented in Figure 21. One can see that the decrease of the singular values is faster for surface illumination, which indicates a poorer conditioning of the operator in this case. On the basis of this SVD decomposition, it is possible to derive an approximate inverse of the Gauss-Newton operator by truncating the singular values below a certain level, represented by dotted lines in Figure 21. Neglecting the smallest singular values defines a regularized version of the inverse Gauss-Newton operator. We apply this approximate inverse to the gradient in the surface-acquisition case (Figure 18) and the results are presented in Figure 22. The perturbations brought to the parameter by this preconditioned gradient are decoupled correctly compared with the one brought by the gradient alone. The shape of the perturbations is recovered. That illustrates the potential for the Hessian operator to mitigate cross-talk between parameters.

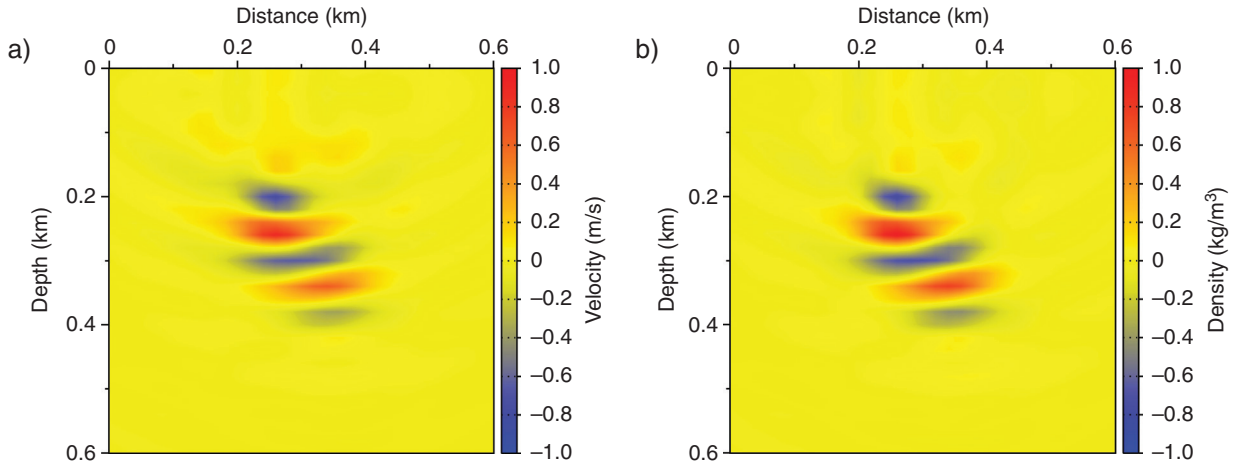

Figure 18. Model perturbation deduced from gradients obtained for (a) the P-wave velocity and (b) the density in the surface-acquisition case.

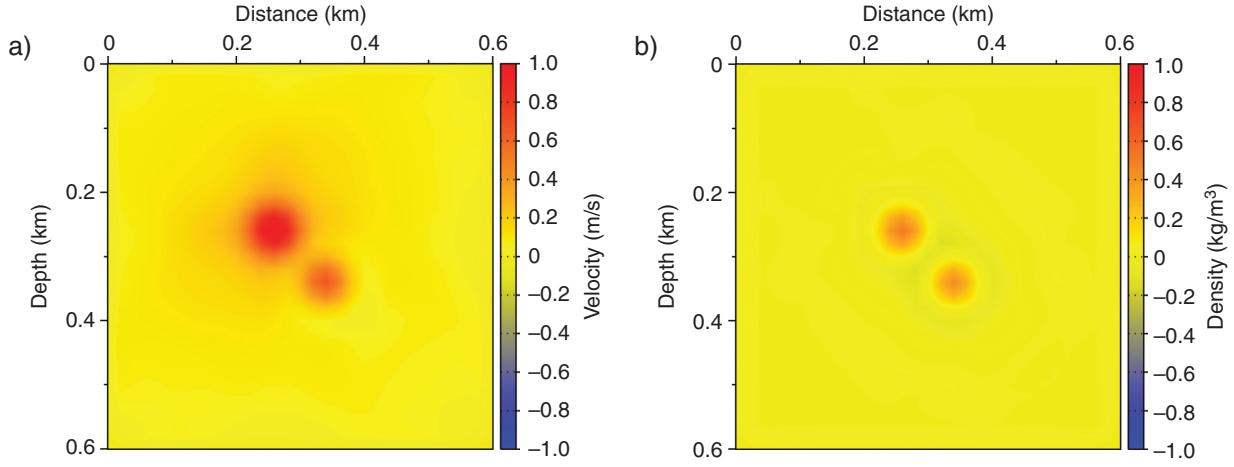

Figure 19. Model perturbation deduced from gradients obtained for (a) the P-wave velocity and (b) the density in the full-acquisition case.

However, computing an approximate inverse of the Gauss-Newton operator through a truncated SVD method is not affordable for large-scale optimization problems. Approximations of the inverse Hessian operator adapted to large-scale problems have to be considered — namely *l*-BFGS (Nocedal, 1980; Nocedal and Wright, 2006) or the truncated Newton method (Nash, 2000).

The *l*-BFGS strategy consists of rank 2 updates of an initial approximation taken as the identity if no prior information is known. These rank 2 updates are based on the *l*-previous values of the gradient, so the *l*-BFGS strategy does not require computation of quantities other than the gradient in order to build an approximation of the inverse Hessian operator. However, the quality of this approximation in the early iterations strongly depends on the prior knowledge of the inverse Hessian operator. Starting from the identity, the initial model updates are along the direction of the model gradient. In a multiparameter context, as illustrated by the tutorial case study presented here, this can be harmful for a decoupled reconstruction of the parameter classes. Tradeoffs are introduced in the early stage of the inversion, making it extremely difficult to decouple the parameters in the later iterations.

On the other hand, the truncated Newton method is based on an inexact solution of the Newton equation (equation 5) through a matrix-free conjugate-gradient method. This requires the ability to compute Hessian-vector products efficiently. As emphasized in Epanomeritakis

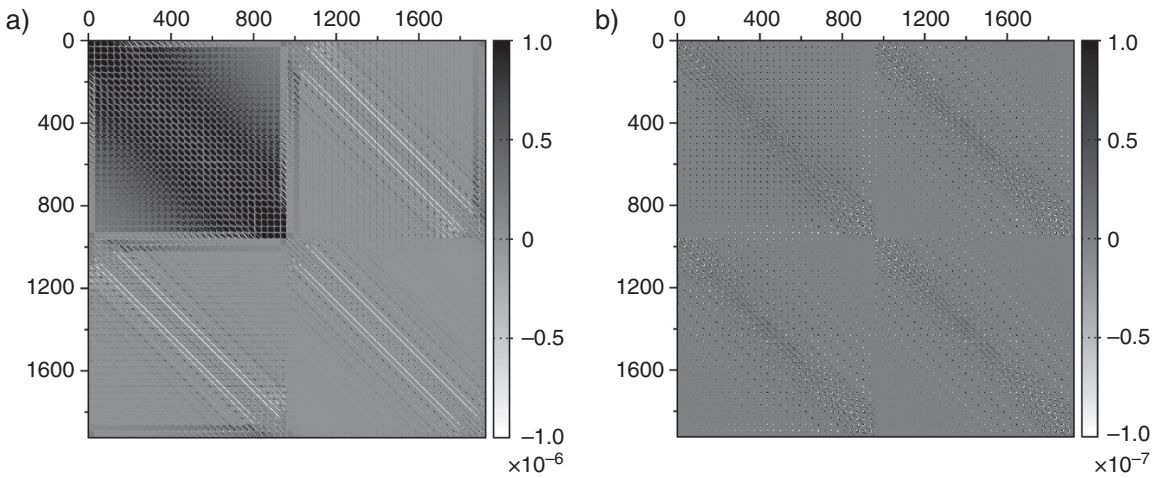

Figure 20. Gauss-Newton operator for (a) full acquisition and (b) surface acquisition. Upper diagonal blocks correspond to second-order derivatives with respect to the P-wave velocity. Lower diagonal blocks correspond to second-order derivatives with respect to the density. Off diagonal blocks correspond to crosscorrelation of the derivatives with respect to P-wave velocity and density and express the crosstalk between the two parameters. The isotropic radiation pattern of the P-wave velocity makes the misfit function sensitive to this parameter for all illumination angles. This explains the stronger amplitude of the upper diagonal block for full illumination, whereas the amplitudes of the different blocks are comparable for surface acquisition.

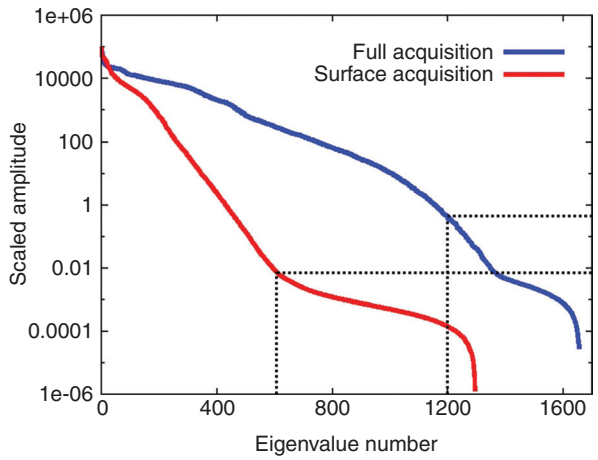

Figure 21. Normalized singular values of the Gauss-Newton operator for full illumination (blue) and surface illumination (red). The decrease of the singular values is faster for surface illumination, which indicates a poorer conditioning of the operator in this case. Dotted lines specify threshold values used for building preconditioned gradients.

et al. (2008), Fichtner and Trampert (2011), and Métivier et al. (2013, 2014), this can be performed through a generalization of the adjoint-state strategy usually implemented for computation of the gradient. The computation of Hessian-vector products thus only requires additional wave-propagation problems to be solved. The advantage of the truncated Newton method in a multiparameter FWI context is that the accuracy of the inverse Hessian approximation should be the same from one iteration to the other. This is especially important for the first iterations: the truncated Newton method should enable us to better decouple the parameters in the early stages of the inversion.

The optimization strategies used in the context of realistic-scale multiparameter FWI applications should be completed with efficient preconditioning strategies. In this context, a preconditioner is an approximation of the inverse Hessian operator, which can be computed at low cost and inserted in the optimization framework. Both the *l*-BFGS and truncated Newton methods can be used with a preconditioner (Métivier and Brossier, 2016). A generalization of the diagonal preconditioner proposed by Shin et al. (2001) and improved by Choi and Shin (2008) has been recently promoted by Innanen (2014) and Métivier et al. (2015). This strategy amounts to approximating each block of the Hessian by a diagonal matrix, following the pseudo-Hessian approximation of Choi and Shin (2008). The preconditioner is computed as an inverse of this block approximation, which can be computed straightforwardly through local inversion of these corresponding submatrices. The leading

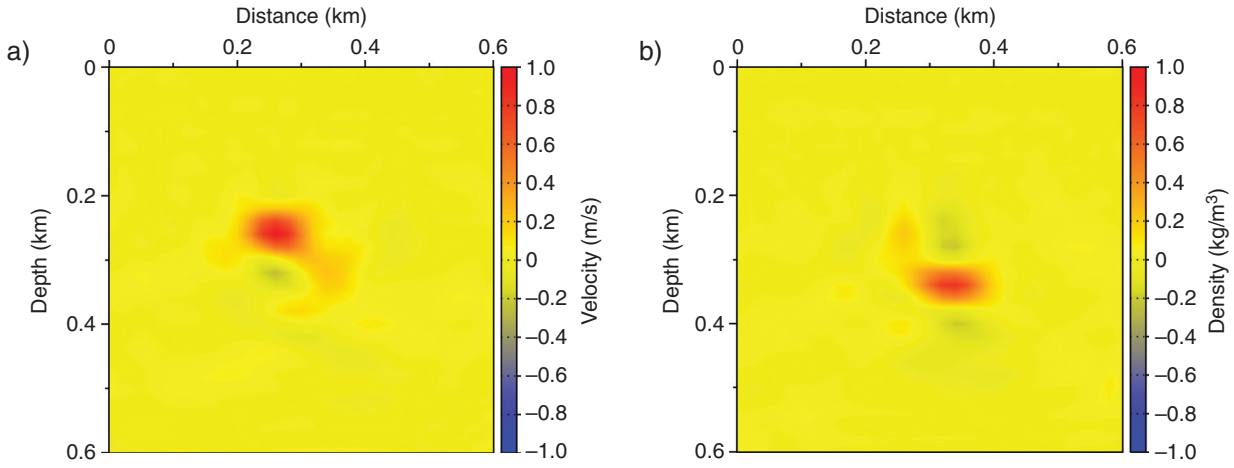

Figure 22. Model updates obtained after applying the approximate inverse Gauss-Newton operator to the gradient in the surface-acquisition case. The two perturbations in (a) P-wave velocity and (b) density are correctly decoupled. The shape of the perturbation is recovered.

idea is to account in this preconditioner for the local coupling between parameters, neglecting the spatial interaction that may exist between parameters of different classes. Preliminary results seem to indicate improved reliability of multiparameter inversion results by using this preconditioner in conjunction with the truncated Newton method (Métivier et al., 2015).

Summary and the future of FWI

Full waveform inversion has been found to be a useful tool for extracting pieces of information from seismograms/traces. Of course, this potentially high-resolution performance comes with strict requirements related to the cycle-skipping problem, because we are concerned about seismic oscillatory signals. The workflow related to the model updating has been mitigated through an adjoint formulation in which the gradient vector of the misfit function with respect to model parameters is evaluated directly without dealing with Fréchet derivatives. An extention for handling the Hessian matrix is still a challenging subject, although a second adjoint formulation could provide an efficient strategy.

The fundamental feature of FWI is the absence of prior scale separation for signal interpretation; the single scattering formulation is at the core of FWI and each seismic phase is interpreted the same way by using this scattering assumption. That is the physical intuition on which FWI is based. FWI may benefit from the different hierarchies found in the seismic data. Various formulations have been developed in recent years that are based on such hierarchies. Basic concepts related to linear partial differential equations (forward modeling) as well as linear algebra (linearized optimization) can be handled quite naturally when starting an FWI investigation. The gradient can be computed efficiently by using either a second-order hyperbolic wave equation or a first-order hyperbolic wave equation. The choice is really application-dependent.

Uncertainty quantification and model-driven ingredients are important new trends in FWI. The construction of the initial model to make FWI accurate and efficient is a very active research area because it is a critical issue for this technique. A simple and systematic strategy is based on the introduction of scale separation tied to FWI concepts. No phase picking is performed in this strategy. Other strategies may perform as well for high-resolution seismic imaging, and considerable progress will be made. Of course, an extension to 3D will be the selective criterion for those different methods.

FWI is evolving toward elastic probing of the medium because the focus is increasingly on amplitude information. This probing will require multiparameter reconstruction in which the Hessian matrix, the second derivative of the misfit function, will play a crucial role.

For people involved in high-resolution seismic imaging, the trend is the steady movement from impressive real-world examples that are based on acoustic propagation to well-constrained multiparameter reconstructions that increasingly are based on elastic propagation. At the same

time, other challenges, such as microscale characterization (Dupuy et al., 2016), may need alternative optimization strategies such as global optimization.

Appendix A: Lagrange multipliers

Optimization under constraints can be achieved using Lagrange multipliers. We present a systematic if not automatic workflow for building the misfit-function gradient. Let us reiterate that the expression of the misfit function is

$$\mathcal{C}(\mathbf{m}) = \frac{1}{2}\int_0^{T_w}\int_{\partial\Omega_s}\int_{\partial\Omega_r}\left(\mathbf{d}_{\text{syn}}(\mathbf{x}_r,\mathbf{x}_s,t)-\mathbf{d}_{\text{obs}}(\mathbf{x}_r,\mathbf{x}_s,t)\right)^2 \times d\mathbf{x}_r d\mathbf{x}_s dt, \tag{A-1}$$

and the relation between synthetic data and computed wavefields is given by the expression

$$\mathbf{d}_{\text{syn}}(\mathbf{x}_r,t) = R_r\mathbf{w}(\mathbf{x},t). \tag{A-2}$$

The first adjoint wavefield λ has the same number of components as does the wavefield vector $\mathbf{w}$, whereas the second field μ has the same number of components as does the data vector $\mathbf{d}_{\text{syn}}$. They are related to a new auxiliary expression $\mathcal{L}$, called the Lagrangian, which depends now on independent quantities ($\mathbf{m}$, $\mathbf{w}$, λ, $\mathbf{d}_{\text{syn}}$, μ). At sampled model points, often called realizations, these quantities $\bar{\mu},\bar{\mathbf{w}},\bar{\lambda}$ are linked through the wave/adjoint equations and the projection relation to the model vector $\mathbf{m}$. Such conditions are defined as constraints. Over the entire domain Ω, the Lagrangian can be written as

$$\begin{aligned}\mathcal{L}(\mathbf{m},\mathbf{d},\mu,\mathbf{w},\lambda) &= \mathcal{C}(\mathbf{m},\mathbf{w}) + \sum_s\sum_r\int_0^{T_w} dt\,\mu(\mathbf{x}_r,t)^{\mathrm{T}} \\ &\times\left(\mathbf{d}_{\text{syn}}(\mathbf{x}_r,t) - R_r\mathbf{w}(\mathbf{x},t)\right) + \sum_s\sum_r\int_0^{T_w}\int_\Omega dt d\mathbf{x}\,\lambda(\mathbf{x},t)^{\mathrm{T}} \\ &\times\left[\mathbf{M}\partial_t\mathbf{w}(\mathbf{x},t) - \mathbf{M}\mathbf{f}^{\mathrm{w}}(\mathbf{x},t) - \mathcal{D}(\nabla)\mathbf{w}(\mathbf{x},t)\right].\end{aligned} \tag{A-3}$$

A necessary requirement is that the differential form $\delta\mathcal{L}$ should be zero at the minimum, which is identical for Lagrangian and misfit functions. Therefore, the five gradients of $\mathcal{L}$ with respect to arguments (which are considered as independent) should be equal to zero through the list:

(1) Zeroing the gradient with respect to the vector μ comes from the projection relation (equation 24).
(2) Zeroing the gradient with respect to the vector $\mathbf{d}_{\text{syn}}$ leads to the equation

$$\mu(\mathbf{x}_r,t) = (\mathbf{d}_{\text{syn}}(\mathbf{x}_r,t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_r,t)), \tag{A-4}$$

at each receiver, taking into account that $\partial\mathcal{C}/\partial\mathbf{d}_{\text{syn}} = (\mathbf{d}_{\text{syn}}(\mathbf{x}_r,t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_r,t))$.
(3) Zeroing the gradient with respect to the vector λ gives the wave equation.
(4) Zeroing the gradient with respect to the vector $\mathbf{w}$ provides the partial differential equation verified by the vector λ, which is called the adjoint wavefield.
(5) Zeroing the gradient of the Lagrangian function with respect to the parameter vector $\mathbf{m}$ will provide a new perturbation solution for updating the model.

Let us work out the fourth condition. Away from sources, we find the expression

$$\begin{aligned}\frac{\partial\mathcal{L}^{\mathrm{T}}}{\partial\mathbf{w}}\cdot\delta\mathbf{w} &= \frac{\partial\mathcal{C}^{\mathrm{T}}}{\partial\mathbf{w}}\cdot\delta\mathbf{w} + \sum_s\sum_r\int_0^{T_w}\int_\Omega dt d\mathbf{x}\lambda(\mathbf{x},t)^{\mathrm{T}} \\ &\times\frac{\partial}{\partial\mathbf{w}}[\mathbf{M}\,\partial_t\mathbf{w}(\mathbf{x},t) - \mathcal{D}(\nabla)\mathbf{w}(\mathbf{x},t)]\cdot\delta\mathbf{w},\end{aligned} \tag{A-5}$$

and, with the help of the symmetry of matrices $\mathbf{S}$ and $\mathcal{D}$ as well as an integration by parts, we find

$$\begin{aligned}\frac{\partial\mathcal{L}^{\mathrm{T}}}{\partial\mathbf{w}}\cdot\delta\mathbf{w} &= \frac{\partial\mathcal{C}^{\mathrm{T}}}{\partial\mathbf{w}}\cdot\delta\mathbf{w} - \sum_s\sum_r\int_0^{T_w}\int_\Omega dt d\mathbf{x} \\ &\times\left[\mathbf{M}\partial_t\lambda(\mathbf{x},t) - \mathcal{D}(\nabla)\lambda(\mathbf{x},t)\right]^{\mathrm{T}}\cdot\delta\mathbf{w},\end{aligned} \tag{A-6}$$

which should be equal to zero at the minimum. For the integration by parts, we have considered that boundary conditions are such that direct $\mathbf{w}$ and adjoint λ wavefields are zero outside the domain Ω, that initial conditions for the direct wavefield are zero, and that final conditions for the adjoint wavefield are zero. The gradient of the misfit function $\mathcal{C}$ with respect to the wavefield $\mathbf{w}$ can be written as

$$\frac{\partial\mathcal{C}^{\mathrm{T}}}{\partial\mathbf{w}} = \sum_r R_r^{\mathrm{T}}\frac{\partial\mathcal{C}^{\mathrm{T}}}{\partial\mathbf{d}_{\text{syn}}} = \sum_r R_r^{\mathrm{T}}[R_r\mathbf{w}(\mathbf{x},t) - \mathbf{d}_{\text{obs}}(\mathbf{x}_r,t)]. \tag{A-7}$$

We deduce that the adjoint wavefield should verify the following partial differential system

$$\mathbf{M}\,\partial_t\lambda(\mathbf{x},t) - \boldsymbol{\mathcal{D}}(\nabla)\lambda(\mathbf{x},t) = \sum_{\mathrm{r}} R_{\mathrm{r}}^{\mathrm{T}}[\mathbf{d}_{\mathrm{syn}}(\mathbf{x}_{\mathrm{r}},t) - \mathbf{d}_{\mathrm{obs}}(\mathbf{x}_{\mathrm{r}},t)] \quad \text{(A-8)}$$

with conditions $\lambda(\mathbf{x}, T_w) = 0$. The source terms of this equation are the residuals at the receivers. Note that we sum only over receivers associated with the same source. Receivers emit unexplained residuals inside the medium, and we integrate in time from the final recording time. Equivalently, we can revert time in the residual source and use the final residuals as initial conditions for a backward time integration using a negative time step. We consequently propagate the adjoint wavefield in the forward way as the direct wavefield using the same code. Of course, we shall not use the implicit system (equation 25) for the adjoint field estimation but instead will use the equivalent explicit system

$$\partial_t\lambda(\mathbf{x},t) - \mathbf{A}\boldsymbol{\mathcal{D}}(\nabla)\lambda(\mathbf{x},t) = \sum_{\mathrm{r}} \mathbf{A}(\mathbf{x})R_{\mathrm{r}}^{\mathrm{T}}(\mathbf{d}_{\mathrm{syn}}(\mathbf{x}_{\mathrm{r}},t) - \mathbf{d}_{\mathrm{obs}}(\mathbf{x},t)), \quad \text{(A-9)}$$

where the adjoint source term related to residuals has been modified by medium properties at the receiver positions, because we want to solve the adjoint problem using the explicit time integration. This source term is highly dependent on the local properties at receivers and, therefore, on the recording quantities at these positions.

Finally, let us consider the fifth condition. Zeroing the gradient of the Lagrangian $\mathcal{L}$ with respect to the expected new model solution $\mathbf{m} + \Delta\mathbf{m}$ will give us the expression

$$\frac{\partial\mathcal{L}^{\mathrm{T}}}{\partial\mathbf{m}}(\mathbf{m} + \Delta\mathbf{m}) = 0 \quad \text{(A-10)}$$

to be solved through the Newton method. For that, we will need the gradient $(\partial\mathcal{L}/\partial\mathbf{m})^{\mathrm{T}}\times(\mathbf{m},\mathbf{d},\bar{\mu},\bar{\mathbf{w}},\lambda)$ at the current solution (a model realization in which fields, identified by the upper bar sign, verify constraints). This gradient turns out to be equal to the misfit gradient $(\partial\mathcal{C}/\partial\mathbf{m})^{\mathrm{T}}(\mathbf{m})$ at these realizations. When considering the Lagrangian function with independent variables, the misfit function does not depend explicitly on the model vector. Consequently, we have

$$\frac{\partial\mathcal{L}^{\mathrm{T}}}{\partial\mathbf{m}}\cdot\delta\mathbf{m} = \sum_{\mathrm{s}}\int_0^{T_{\mathrm{w}}} dt\left[\lambda^{\mathrm{T}}(\mathbf{x},t)\frac{\partial\mathbf{M}}{\partial\mathbf{m}}(\mathbf{x})(\partial_t\mathbf{w}(\mathbf{x},t) - \mathbf{s}'(\mathbf{x},t))\right]^{\mathrm{T}}\cdot\delta\mathbf{m}. \quad \text{(A-11)}$$

At the realization point, we have

$$\frac{\partial\mathcal{L}(\mathbf{m},\mathbf{d},\bar{\mu}(\mathbf{m}),\bar{\mathbf{w}}(\mathbf{m}),\bar{\lambda}(\mathbf{m}))}{\partial\mathbf{m}} = \frac{\partial\mathcal{C}}{\partial\mathbf{m}}(\mathbf{m}), \quad \text{(A-12)}$$

and, therefore, one can deduce the misfit function gradient we are seeking,

$$\frac{\partial\mathcal{C}}{\partial\mathbf{m}}(\mathbf{x}) = \sum_{\mathrm{s}}\int_0^{T_{\mathrm{w}}} dt\lambda^{\mathrm{T}}(\mathbf{x},t)\frac{\partial\mathbf{M}}{\partial\mathbf{m}}(\mathbf{x})[\partial_t\mathbf{w}(\mathbf{x},t) - \mathbf{s}'(\mathbf{x},t)]. \quad \text{(A-13)}$$

Note that there are no explicit spatial derivatives in this estimation of the gradient, which is based on the first-order hyperbolic system (velocity-stress):

$$\frac{\partial\mathcal{C}}{\partial\mathbf{m}}(\mathbf{x}) = \sum_{\mathrm{s}}\int_0^{T_{\mathrm{w}}} dt\lambda^{\mathrm{T}}(\mathbf{x},t)\frac{\partial\mathbf{M}}{\partial\mathbf{m}}(\mathbf{x})\partial_t\mathbf{w}(\mathbf{x},t). \quad \text{(A-14)}$$

Using the density and the compliance components, we find again the expressions

$$\frac{\partial\mathcal{C}}{\partial\rho}(\mathbf{x}_i) = \sum_{\mathrm{s}}\int_0^{T_{\mathrm{w}}} dt[\partial_t v_I(\mathbf{x}_i)\lambda_I^v(\mathbf{x}_i)]$$
$$\frac{\partial\mathcal{C}}{\partial S_{kl}}(\mathbf{x}_i) = \sum_{\mathrm{s}}\int_0^{T_{\mathrm{w}}} dt\left[\partial_t\sigma_I(\mathbf{x}_i)\left(\frac{\partial\mathbf{S}}{\partial S_{kl}}\right)_{IJ}\lambda_I^{\sigma}(\mathbf{x}_i)\right], \quad \text{(A-15)}$$

with implicit summation over the capital indices. The matrix $\partial\mathbf{S}/\partial S_{IJ}$ is composed of elements of values zero or one. From the identity $\mathbf{SC} = \mathbf{CS} = \mathbf{I}$, let us remark that the gradient of the compliance matrix with respect to model parameters is

$$\frac{\partial\mathbf{S}}{\partial\mathbf{m}}(\mathbf{x}_i) = -\mathbf{S}(\mathbf{x}_i)\frac{\partial\mathbf{C}}{\partial\mathbf{m}}(\mathbf{x}_i)\mathbf{S}(\mathbf{x}_i) = -\mathbf{C}^{-1}(\mathbf{x}_i)\frac{\partial\mathbf{C}}{\partial\mathbf{m}}(\mathbf{x}_i)\mathbf{C}^{-1}(\mathbf{x}_i). \quad \text{(A-16)}$$

Away from sources, the gradient with respect to model parameters can be decomposed into

$$\frac{\partial \mathcal{C}}{\partial \rho}(\mathbf{x}_i) = \sum_s \int_0^{T_w} dt[\partial_t \nu_I(\mathbf{x}_i)\lambda_I^{\nu}(\mathbf{x}_i)]$$
$$\frac{\partial \mathcal{C}}{\partial C_{kl}}(\mathbf{x}_i) = -\sum_s \int_0^{T_w} dt[C_{MI}^{-1}(\mathbf{x}_i)\partial_t \sigma_I(\mathbf{x}_i,t)]^t \times \left(\frac{\partial \mathbf{C}}{\partial C_{kl}}(x_i)\right)_{MN} [C_{NJ}^{-1}(\mathbf{x}_i)\lambda_J^{\sigma}(\mathbf{x}_i,t)], \tag{A-17}$$

which can be used as long as particle velocity and stresses are available with implicit summations over capital indices. The matrix $\partial \mathbf{M}/\partial M_{IJ}$ is composed of elements of values zero or one. We may consider the equivalent system

$$\frac{\partial \mathcal{C}}{\partial \rho}(\mathbf{x}_i) = \sum_s \int_0^{T_w} dt[\partial_t \nu_I(\mathbf{x}_i)\lambda_I^{\nu}(\mathbf{x}_i)]$$
$$\frac{\partial \mathcal{C}}{\partial C_{IJ}}(\mathbf{x}_i) = -\sum_s \int_0^{T_w} dt\, \partial_t \varepsilon_I(\mathbf{x}_i,t)\frac{\partial C}{\partial C_{IJ}}(x_i)\lambda_J^{\varepsilon}(\mathbf{x}_i,t), \tag{A-18}$$

as shown in the main text, which can be obtained with the extended second-order hyperbolic system.

Acknowledgments

Many thanks to the editors, Kees Wapenaar and Vladimir Grechka, for their careful reading and comments, and to René-Édouard Plessix for his remarks and comments; their input improved the content of this article. Many thanks to Anne Thomas and Susan Stamm for their careful editorial work. Continuous discussions with Stéphane Operto have made this work possible. We are strongly indebted to him for providing his results on the Valhall data set. Authors obtained support from the sponsors of the SEISCOPE consortium http://seiscope2.osug.fr, sponsored by BP, CGG, Chevron, Exxon-Mobil, JGI, Petrobras, Saudi Aramco, Schlumberger, Shell, Sinopec, Statoil, Total, and Woodside. The MUMPS package, available on http://graal.ens-lyon.fr/MUMPS/index.html and http://mumps.enseeiht.fr, has been used for solving linear systems. We thank the developers from the MUMPS research team (P. Amestoy, A. Buttari, T. Mary, J. -Y. L'Excellent, and C. Weisbecker) for useful discussions and recommendations concerning the interfacing of MUMPS in our imaging code. This study was granted access to the HPC resources of SIGAMM (http://crimson.oca.eu) and CIMENT computer centres and CINES/IDRIS under the allocation 046091 made by GENCI.

We also thank BP Norge AS and their Valhall partner Hess Norge AS for allowing access to data from the Valhall data set, as well as to the well-log velocities and initial FWI models.

References

Alkhalifah, T., and Y. Choi, 2012, Taming waveform inversion non-linearity through phase unwrapping of the model and objective functions: Geophysical Journal International, **191**, no. 3, 1171–1178, http://dx.doi.org/10.1111/j.1365-246X.2012.05699.x.

Almomin, A., and B. Biondi, 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5, http://dx.doi.org/10.1190/segam2012-0976.1.

Amestoy, P. R., C. Ashcraft, O. Boiteau, A. Buttari, J.-Y. L'Excellent, and C. Weisbecker, 2015, Improving multifrontal methods by means of block low-rank representations: SIAM Journal on Scientific Computing, **37**, no. 3, A1451–A1474, http://dx.doi.org/10.1137/120903476.

Anderson, J. E., L. Tan, and D. Wang, 2012, Time-reversal checkpointing methods for RTM and FWI: Geophysics, **77**, no. 4, S93–S103, http://dx.doi.org/10.1190/geo2011-0114.1.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2013, Regularized seismic full waveform inversion with prior model information: Geophysics, **78**, no. 2, R25–R36, http://dx.doi.org/10.1190/geo2012-0104.1.

Bansal, R., P. Routh, J. Krebs, S. Lee, A. Baumstein, J. Anderson, N. Downey, S. Lazaratos, R. Lu, and S. Saldarriaga, 2013, Full wavefield inversion of ocean bottom node data: 75th Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2013, Extended Abstracts, **2013**, We1104, 2506, http://dx.doi.org/10.3997/2214-4609.20130828.

Barkved, O., P. Heavey, R. Kjelstadii, T.Kleppan, and T. G. Kristiansen, 2003, Valhall field — still on plateau after 20 years of production: SPE 83957, https://www.earth.ox.ac.uk/~michaelw/teaching_yr3/Valhall.pdf, accessed 13 September 2016.

Baumstein, A., 2013, POCS-based geophysical constraints in multi-parameter full wavefield inversion: 75th Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2013, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20130409.

Baumstein, A., 2014, Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1121–1125, http://dx.doi.org/10.1190/segam2014-0546.1.

Baumstein, A., 2015, Borehole-constrained multi-parameter full waveform inversion: 77th Conference and Exhibition, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.201412904.

Bednar, J. B., C. Shin, and S. Pyun, 2007, Comparison of waveform inversion, part 2: Phase approach: Geophysical Prospecting, **55**, no. 4, 465–475, http://dx.doi.org/10.1111/j.1365-2478.2007.00618.x.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, **76**, no. 4, R109–R124, http://dx.doi.org/10.1190/1.3581357.

Berkhout, A., 2014, An outlook on the future of seismic imaging, part II: Full-wavefield migration: Geophysical Prospecting, **62**, no. 5, 931–949, http://dx.doi.org/10.1111/1365-2478.12154.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, no. 1, 99–108, http://dx.doi.org/10.1063/1.526755.

Beylkin, G., and R. Burridge, 1990, Linearized inverse scattering problems in acoustics and elasticity: Wave Motion, **12**, no. 1, 15–52, http://dx.doi.org/10.1016/0165-2125(90)90017-X.

Biondi, B., and A. Almomin, 2012, Tomographic full waveform inversion (TWFI) by combining full waveform inversion with wave-equation migration velocity analysis: 82nd Annual International Meeting, SEG, Expanded abstracts, 1–5, http://dx.doi.org/10.1190/segam2012-0275.1.

Biondi, B., and A. Almomin, 2013, Tomographic full waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis: The Leading Edge, **32**, no. 9, 1074–1080, http://dx.doi.org/10.1190/tle32091074.1.

Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics **79**, no. 3, WA129–WA140, http://dx.doi.org/10.1190/geo2013-0340.1.

Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, no. 7, 931–942, http://dx.doi.org/10.1190/1.1442363.

Bonnans, J.-F., J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, 2006, Numerical optimization, theoretical and practical aspects: Universitext Series, Springer Science & Business Media.

Borisov, D., and S. C. Singh, 2015, Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: A synthetic study: Geophysical Journal International, **201**, no. 3, 1215–1234, http://dx.doi.org/10.1093/gji/ggv048.

Brossier, R., B. Pajot, L. Combe, S. Operto, L. Métivier, and J. Virieux, 2014, Time and frequency- domain FWI implementations based on time solver-analysis of computational complexities: 76th Conference and Exhibition, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20141139.

Bozdag, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, no. 2, 845–870, http://dx.doi.org/10.1111/j.1365-246X.2011.04970.x.

Burridge, R., 1996, Elastic waves in anisotropic media: Schlumberger-Doll Research, ftp://math.stanford.edu/pub/mgss/burridge2.ps.gz, accessed 13 September 2016.

Candès, E. J., and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: Communications on Pure and Applied Mathematics, **58**, no. 11, 1472–1528, http://dx.doi.org/10.1002/cpa.20078.

Castellanos, C., L. Métivier, S. Operto, R. Brossier, and J. Virieux, 2015, Fast full waveform inversion with source encoding and second-order optimization methods: Geophysical Journal International, **200**, no. 2, 718–742, http://dx.doi.org/10.1093/gji/ggu427.

Chauris, H., and R.-É. Plessix, 2012, Investigating the differential waveform inversion: 74th Conference and Exhibition, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20149790.

Chavent, G., 1974, Identification of functional parameters in partial differential equations, *in* R. E. Goodson and M. Polis, eds., Identification of parameters in distributed systems: American Society of Mechanical Engineers, 31–48.

Chavent, G., 2009, Nonlinear least squares for inverse problems: Theoretical foundations and step-by-step guide for applications: Springer, available online ©2010, http://dx.doi.org/10.1007/978-90-481-2785-6.

Chavent, G., and C. A. Jacewitz, 1995, Determination of background velocities by multiple migration fitting: Geophysics, **60**, no. 2, 476–490, http://dx.doi.org/10.1190/1.1443785.

Choi, Y., and T. Alkhalifah, 2011, Frequency-domain waveform inversion using the unwrapped phase:

81st Annual International Meeting, SEG, Expanded Abstracts, **30**, 2576–2580, http://dx.doi.org/10.1190/1.3627727.

Choi, Y., D.-J. Min, and C. Shin, 2008, Frequency-domain elastic full waveform inversion using the new pseudo-Hessian matrix: Experience of elastic Marmousi-2 synthetic data: Bulletin of the Seismological Society of America, **98**, no. 5, 2402–2415, http://dx.doi.org/10.1785/0120070179.

Claerbout, J., 1985, Imaging the earth's interior: Blackwell Scientific Publications (original publisher), now available online in its entirety courtesy of J. Claerbout, http://sep.standford.edu/sep/prof/eie2/, accessed 10 November 2016.

Clapp, R., 2008, Reverse time migration: Saving the boundaries: Technical Report SEP-136, Stanford Exploration Project, http://sepwww.stanford.edu/data/media/public/docs/sep138/bob1/Fig/paper.pdf, accessed 4 November 2016.

Clément, F., G. Chavent, and S. Gómez, 2001, Migration-based traveltime waveform inversion of 2-D simple structures: A synthetic example: Geophysics, **66**, no. 3, 845–860, http://dx.doi.org/10.1190/1.1444974.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, no. 5, 527–538, http://dx.doi.org/10.1190/1.1442864.

Crase, E., C. Wideman, M. Noble, and A. Tarantola, 1992, Nonlinear elastic waveform inversion of land seismic reflection data: Journal of Geophysical Research, **97**, no. B4, 4685–4703, http://dx.doi.org/10.1029/90JB00832.

Demanet, L., P.-D. Létourneau, N. Boumal, H. Calandra, J. Chiu, and S. Snelson, 2012, Matrix probing: A randomized preconditioner for the wave-equation Hessian: Applied and Computational Harmonic Analysis, **32**, no. 2, 155–168, http://dx.doi.org/10.1016/j.acha.2011.03.006.

Demirbag, E., C. Çoruh, and J. Costain, 1993, Inversion of P-wave AVO, *in* J. P. Castagna and M. M. Backus, eds., Offset-dependent reflectivity — Theory and practice of AVO analysis: SEG Investigations in Geophysics, no. 8, 287–302.

Devaney, A., 1984, Geophysical diffraction tomography: IEEE Transactions on Geoscience and Remote Sensing, **GE-22**, no. 1, 3–13, http://dx.doi.org/10.1109/TGRS.1984.350573.

Donno, D., H. Chauris, and H. Calandra, 2013, Estimating the background velocity model with the normalized integration method: 75th Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2013, Extended Abstracts, Tu0704, http://dx.doi.org/10.3997/2214-4609.20130411.

Dupuy, B., A. Asnaashari, R. Brossier, S. Garambois, L. Métivier, A. Ribodetti, and J. Virieux, 2016, A downscaling strategy from FWI to microscale reservoir properties from high-resolution images: The Leading Edge, **35**, no. 2, 146–150, http://dx.doi.org/10.1190/tle35020146.1.

Dussaud, E., W. W. Symes, P. Williamson, L. Lemaistre, P. Singer, B. Denel, and A. Cherrett, 2008, Computational strategies for reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271, http://dx.doi.org/10.1190/1.3059336.

Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences, **12**, no. 5, 979–988, http://dx.doi.org/10.4310/CMS.2014.v12.n5.a7.

Epanomeritakis, I., V. Akçelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: Inverse Problems, **24**, no. 3, http://dx.doi.org/10.1088/0266-5611/24/3/034015.

Erlangga, Y. A., and R. Nabben, 2009, Algebraic multilevel Krylov methods: SIAM Journal on Scientific Computing, **31**, no. 5, 3417–3437, http://dx.doi.org/10.1137/080731657.

Evensen, G., 2009, Data assimilation: The ensemble Kalman filter, 2nd ed.: Springer.

Fang, Z., F. J. Herrmann, and C. D. Silva, 2014, Fast uncertainty quantification for 2D full-waveform inversion with randomized source subsampling: 76th Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2014, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20140715.

Fichtner, A., and J. Trampert, 2011, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, no. 3, 1604–1624, http://dx.doi.org/10.1111/j.1365-246X.2011.05218.x.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villaseñor, 2013, Multiscale full waveform inversion: Geophysical Journal International, **194**, no. 1, 534–556, http://dx.doi.org/10.1093/gji/ggt118.

Fomel, S., and J. F. Claerbout, 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: Geophysics, 68, no. 2, 1–12, http://dx.doi.org/10.1190/1.1567228.

Friedlander, M. P., and M. W. Schmidt, 2012, Hybrid deterministic-stochastic methods for data fitting: SIAM Journal on Scientific Computing, **34**,

no. 3, A1380–A1405, http://dx.doi.org/10.1137/110830629.

Gholami, Y., R. Brossier, S. Operto, V. Prieux, A. Ribodetti, and J. Virieux, 2013, Which parametrization is suitable for acoustic vertical transverse isotropic full waveform inversion? Part 2: Synthetic and real data case studies from Valhall: Geophysics, **78**, no. 2, R107–R124, http://dx.doi.org/10.1190/geo2012-0203.1.

Godwin, J., and P. Sava, 2013, A comparison of shot-encoding schemes for wave-equation migration: Geophysical Prospecting, **61**, no. S1, 391–408, http://dx.doi.org/10.1111/j.1365-2478.2013.01125.x.

Gosselin-Cliche, B., and B. Giroux, 2014, 3D frequency-domain finite-difference viscoelastic-wave modeling using weighted average 27-point operators with optimal coefficients: Geophysics, **79**, no. 3, T169–T188, http://dx.doi.org/10.1190/geo2013-0368.1.

Gray, D., T. Chen, and B. Goodway, 1998, 3-D AVO and migration: 68th Annual International Meeting, SEG, Expanded Abstracts, 234–235, http://dx.doi.org/10.1190/1.1820386.

Griewank, A., 1992, Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation: Optimization Methods and Software, **1**, no. 1, 35–54, http://dx.doi.org/10.1080/10556789208805505.

Griewank, A., and A. Walther, 2000, Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation: ACM Transactions on Mathematical Software (TOMS), **26**, no. 1, 19–45, http://dx.doi.org/10.1145/347837.347846.

Guitton, A., G. Ayeni, and E. Díaz, 2012, Constrained full-waveform inversion by model reparameterization: Geophysics, **77**, no. 2, R117–R127, http://dx.doi.org/10.1190/geo2011-0196.1.

Heilmann, Z., J. Mann, and I. Koglin, 2006, CRS-stack-based seismic imaging considering top-surface topography: Geophysical Prospecting, **54**, no. 6, 681–695, http://dx.doi.org/10.1111/j.1365-2478.2005.00565.x.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, no. 6, WB173–WB187, http://dx.doi.org/10.1190/1.3506147.

Huang, Y., and G. T. Schuster, 2014, Resolution limits for wave equation imaging: Journal of Applied Geophysics, **107**, 137–148, http://dx.doi.org/10.1016/j.jappgeo.2014.05.018.

Hubral, P., G. Höcht, and R. Jäger, 1998, An introduction to the common reflection surface stack: 60th Conference and Exhibition, EAGE, Extended Abstracts, 1–19, http://dx.doi.org/10.3997/2214-4609.201408165.

Innanen, K. A., 2014, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, no. 2, 717–734, http://dx.doi.org/10.1093/gji/ggu291.

Jin, S., R. Madariaga, J. Virieux, and G. Lambaré, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, no. 2, 575–588, http://dx.doi.org/10.1111/j.1365-246X.1992.tb04637.x.

Kennett, B. L. N., and M. Sambridge, 1998, Inversion for multiple parameter classes: Geophysical Journal International, **135**, no. 1, 304–306, http://dx.doi.org/10.1046/j.1365-246X.1998.00657.x.

Kennett, B. L. N., M. S. Sambridge, and P. R. Williamson, 1988, Subspace methods for large inverse problems with multiple parameter classes: Geophysical Journal International, **94**, no. 2, 237–247, http://dx.doi.org/10.1111/j.1365-246X.1988.tb05898.x.

Koglin, I., J. Mann, and Z. Heilmann, 2006, CRS-stack-based residual static correction: Geophysica Prospecting, **54**, no. 6, 697–707, http://dx.doi.org/10.1111/j.1365-2478.2005.00562.x.

Komatitsch, D., J. Ritsema, and J. Tromp, 2002, The spectral-element method, Beowulf computing, and global seismology: Science, **298**, no. 5599, 1737–1742, http://dx.doi.org/10.1126/science.1076024.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, no. 6, WCC177–WCC188, http://dx.doi.org/10.1190/1.3230502.

Lambaré, G., 2008, Stereotomography: Geophysics, **73**, no. 5, VE25–VE34, http://dx.doi.org/10.1190/1.2952039.

Lee, E.-J., and P. Chen, 2013, Automating seismic waveform analysis for full 3-D waveform inversions: Geophysical Journal International, 194, no. 1, 572–589, http://dx.doi.org/10.1093/gji/ggt124.

Li, Y., L. Métivier, R. Brossier, B. Han, and J. Virieux, 2015, 2D and 3D frequency-domain elastic wave modeling in complex media with a parallel iterative solver: Geophysics, **80**, no. 3, T101–T118, http://dx.doi.org/10.1190/geo2014-0480.1.

Liu, Q., and J. Tromp, 2006, Finite-frequency kernels based on adjoint methods: Bulletin of the Seismological Society of America, **96**, no. 6, 2383–2397, http://dx.doi.org/10.1785/0120060041.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion:

81st Annual International Meeting, SEG, Expanded Abstracts, **30**, 2788–2792, http://dx.doi.org/10.1190/1.3627773.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, no. 5, 645–653, http://dx.doi.org/10.1190/1.1443081.

Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: Geophysics, **78**, no. 6, R223–R233, http://dx.doi.org/10.1190/geo2013-0004.1.

Maggi, A., C. Tape, M. Chen, D. Chao, and J. Tromp, 2009, An automated time-window selection algorithm for seismic tomography: Geophysical Journal International, **178**, no. 1, 257–281, http://dx.doi.org/10.1111/j.1365-246X.2009.04099.x.

Martin, J., L. C. Wilcox, C. Burstedde, and O. Ghattas, 2012, A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion: SIAM Journal on Scientific Computing, **34**, no. 3, A1460–A1487, http://dx.doi.org/10.1137/110845598.

Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014, Full waveform inversion and the truncated Newton method: Quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, no. 6, 1353–1375, http://dx.doi.org/10.1111/1365-2478.12136.

Métivier, L., and R. Brossier, 2016, The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: Geophysics, **81**, no. 2, F1–F15, http://dx.doi.org/10.1190/geo2015-0031.1.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016a, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: Geophysical Journal International, **205**, no. 1, 345–377, http://dx.doi.org/10.1093/gji/ggw014.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016b, Optimal transport approach for seismic tomography: Application to 3D full waveform inversion: Inverse Problems, **32**, no. 11, paper ID 115008, 1–32, http://dx.doi.org/10.1088/0266-5611/32/11/115008.

Métivier, L., R. Brossier, S. Operto, and J. Virieux, 2015, Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: Preconditioned truncated Newton approach: 85th Annual international Meeting, SEG, Expanded Abstracts, 1198–1203, http://dx.doi.org/10.1190/segam2015-5875643.1.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, no. 2, B401–B437, http://dx.doi.org/10.1137/120877854.

Miller, D., M. Oristaglio, and G. Beylkin, 1987, A new slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, no. 7, 943–964, http://dx.doi.org/10.1190/1.1442364.

Mora, P. R., 1989, Inversion = migration + tomography: Geophysics, **54**, no. 12, 1575–1586, http://dx.doi.org/10.1190/1.1442625.

Mulder, W., and R.-É. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, no. 6, 827–841, http://dx.doi.org/10.1111/j.1365-2478.2008.00708.x.

MUMPS team, 2011, MUltifrontal Massively Parallel Solver (4.10.0) users' guide, May 10, 2011. https://openmodelica.org/svn/OpenModelicaExternal/trunk/3rdParty/Ipopt-3.11.6/ThirdParty/Mumps/MUMPS/doc/userguide_4.10.0.ps, accessed 15 September 2016.

Nash, S. G., 2000, A survey of truncated-Newton methods: Journal of Computational and Applied Mathematics, **124**, nos. 1–2, 45–59, http://dx.doi.org/10.1016/S0377-0427(00)00426-X.

Neklyudov, D., M. Dmitriev, M. Belonosov, and V. Tcheverda, 2014, Frequency-domain iterative solver for 3D acoustic wave equation with two-stage semi-analytical preconditioner: 76th Conference and Exhibition, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20140811.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, no. 1, 208–221, http://dx.doi.org/10.1190/1.1444517.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782, http://dx.doi.org/10.1090/S0025-5718-1980-0572855-7.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer, available as a PDF at http://home.agh.edu.pl/~pba/pdfdoc/Numerical_Optimization.pdf, accessed 15 September 2016.

Operto, S., R. Brossier, L. Combe, L. Métivier, A. Ribodetti, and J. Virieux, 2014, Computationally efficient three-dimensional acoustic finite-difference frequency-domain seismic modeling in vertical transversely isotropic media with sparse direct solver: Geophysics, **79**, no. 5, T257–T275, http://dx.doi.org/10.1190/geo2013-0478.1.

Operto, S., R. Brossier, Y. Gholami, V. Prieux, A. Ribodetti, L. Métivier, and J. Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory

to practice: The Leading Edge, **32**, no. 9, 1040–1054, http://dx.doi.org/10.1190/tle32091040.1.

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: Application to Valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, **202**, no. 2, 1362–1391, http://dx.doi.org/10.1093/gji/ggv226.

Paige, C. C., and M. A. Saunders, 1982, Algorithm 583: LSQR: Sparse linear equations and least squares problems: ACM Transactions on Mathematical Software (TOMS), **8**, no. 2, 195–209, http://dx.doi.org/10.1145/355993.356000.

Perrone, F., P. Sava, and J. Panizzardi, 2015, Wavefield tomography based on local image correlations: Geophysical Prospecting, **63**, no. 1, 35–54, http://dx.doi.org/10.1111/1365-2478.12163.

Peter, D., D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini, and J. Tromp, 2011, Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes: Geophysical Journal International, **186**, no. 2, 721–739, http://dx.doi.org/10.1111/j.1365-246X.2011.05044.x.

Plessix, R.-É., 2006a, Estimation of velocity and attenuation coefficient maps from crosswell seismic data: Geophysics, **71**, no. 6, S235–S240, http://dx.doi.org/10.1190/1.2345051.

Plessix, R.-É., 2006b, Review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, no. 2, 495–503, http://dx.doi.org/10.1111/j.1365-246X.2006.02978.x.

Plessix, R.-É., 2007, A Helmholtz iterative solver for 3D seismic-imaging problems: Geophysics, **72**, no. 5, SM185–SM194, http://dx.doi.org/10.1190/1.2738849.

Plessix, R.-É., G. Baeten, J. W. de Maag, F. ten Kroode, and Z. Rujie, 2012, Full waveform inversion and distance separated simultaneous sweeping: A study with a land seismic data set: Geophysical Prospecting, **60**, no. 4, 733–747, http://dx.doi.org/10.1111/j.1365-2478.2011.01036.x.

Plessix, R.-É., and Q. Cao, 2011, A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium: Geophysical Journal International, **185**, no. 1, 539–556, http://dx.doi.org/10.1111/j.1365-246X.2011.04957.x.

Plessix, R.-É., Y.-H. De Roeck, and G. Chavent, 1999, Waveform inversion of reflection seismic data for kinematic parameters by local inversion: SIAM Journal on Scientific Computing, **20**, no. 3, 1033–1052, http://dx.doi.org/10.1137/S1064827596311980.

Plessix, R.-É., and C. Perkins, 2010, Thematic set: Full waveform inversion of a deep water ocean bottom seismometer dataset: First Break, **28**, no. 4, 71–78, http://dx.doi.org/10.3997/1365-2397.2010013.

Plessix, R.-É., A. Stopin, P. Milcik, and K. Matson, 2014, Acoustic and anisotropic multi-parameter seismic full waveform inversion case studies: 84th Annual International Meeting, SEG, Expanded Abstracts, 1056–1060, http://dx.doi.org/10.1190/segam2014-0646.1.

Prabhat, and Q. Koziol, eds., 2014, High performance parallel I/O: Chapman and Hall/CRC Inc.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, **64**, no. 3, 888–901, http://dx.doi.org/10.1190/1.1444597.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013a, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field, Part 1: Imaging compressional wavespeed, density and attenuation: Geophysical Journal International, **194**, no. 3, 1640–1664, http://dx.doi.org/10.1093/gji/ggt177.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013b, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field, Part 2: Imaging compressive-wave and shear-wave velocities: Geophysical Journal International, **194**, no. 3, 1665–1681, http://dx.doi.org/10.1093/gji/ggt178.

Rickett, J. E., and P. C. Sava, 2002, Offset and angle-domain common image-point gathers for shot-profile migration: Geophysics, **67**, no. 3, 883–889, http://dx.doi.org/10.1190/1.1484531.

Riyanti, C. D., A. Kononov, Y. A. Erlangga, C. Vuik, C. W. Oosterlee, R.-É. Plessix, and W. A. Mulder, 2007, A parallel multigrid-based preconditioner for the 3D heterogeneous high-frequency Helmholtz equation: Journal of Computational Physics, **224**, no. 1, 431–448, http://dx.doi.org/10.1016/j.jcp.2007.03.033.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, no. 2, 426–436, http://dx.doi.org/10.1190/1.1444737.

Routh, P. S., J. R. Krebs, S. Lazaratos, A. I. Baumstein, I. Chikichev, S. Lee, N. Downey, D. Hinkley, and J. E. Anderson, 2011, Full-wavefield inversion of marine streamer data with the encoded

simultaneous source method: 73rd Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2011, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20149730.

Sava, P., and I. Vasconcelos, 2009, Extended common-image-point gathers for wave-equation migration: 71st Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2009, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.201400361.

Schäfer, M., L. Groos, T. Forbriger, and T. Bohlen, 2013, 2D full waveform inversion of recorded shallow seismic Rayleigh waves on a significantly 2D structure: 19th European Meeting of Environmental and Engineering Geophysics, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20131338, available as a PDF from https://www.researchgate.net/publication/266633089_2D_Full_Waveform_Inversion_of_Recorded_Shallow_Seismic_Rayleigh_Waves_on_a_Significantly_2D_Structure, accessed 15 September 2016.

Schiemenz, A., and H. Igel, 2013, Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: Application to Valhall ocean-bottom cable data: Geophysical Journal International, **195**, no. 3, 1970–1988, http://dx.doi.org/10.1093/gji/ggt362.

Schuster, G. T., X. Wang, Y. Huang, W. Dai, and C. Boonyasiriwat, 2011, Theory of multisource crosstalk reduction by phase-encoded statics: Geophysical Journal International, **184**, no. 3, 1289–1303, http://dx.doi.org/10.1111/j.1365-246X.2010.04906.x.

Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, no. 5, 592–606, http://dx.doi.org/10.1046/j.1365-2478.2001.00279.x.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Thematic set: Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, no. 4, 65–70, http://dx.doi.org/10.3997/1365-2397.2010012.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, no. 1, 231–248, http://dx.doi.org/10.1190/1.1649391.

Snieder, R., M. Y. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, **54**, no. 8, 991–1000, http://dx.doi.org/10.1190/1.1442742.

Staal, X. R., and D. J. Verschuur, 2013, Joint migration inversion, imaging including all multiples with automatic velocity update: 75th Conference and Exhibition, EAGE, incorporating SPE EUROPEC 2013, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20130375.

Stopin, A., R.-É. Plessix, and S. Al Abri, 2014, Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman: Geophysics, **79**, no. 3, WA69–WA77, http://dx.doi.org/10.1190/geo2013-0323.1.

Sun, W., and L.-Y. Fu, 2013, Two effective approaches to reduce data storage in reverse time migration: Computers & Geosciences, **56**, 69–75, http://dx.doi.org/10.1016/j.cageo.2013.03.013.

Symes, W. W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, no. 5, SM213–SM221,http://dx.doi.org/10.1190/1.2742686.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, no. 6, 765–790, http://dx.doi.org/10.1111/j.1365-2478.2008.00698.x.

Symes, W. W., and M. Kern, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **42**, no. 6, 565–614, http://dx.doi.org/10.1111/j.1365-2478.1994.tb00231.x.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, no. 8, 1259–1266, http://dx.doi.org/10.1190/1.1441754.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, no. 10, 1893–1903, http://dx.doi.org/10.1190/1.1442046.

Tarantola, A., 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier, available s a PDF from http://www.ipgp.fr/~tarantola/Files/Professional/Papers_PDF/InverseProblemTheory.pdf, accessed 15 September 2016.

Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms including elasticity and attenuation: Pure and Applied Geophysics, **128**, no. 1, 365–399, http://dx.doi.org/10.1007/BF01772605.

Thurber, C., and J. Ritsema, 2007, Theory and observations — Seismic tomography and inverse methods, *in* G. Schubert, ed., Treatise on Geophysics, volume 1: Seismology and structure of the Earth: Elsevier, 323–360.

Tikhonov, A. N., and V. J. Arsenin, 1977, Solutions of ill-posed problems: Wiley.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and

banana-doughnut kernels: Geophysical Journal International, **160**, no. 1, 195–216, http://dx.doi.org/10.1111/j.1365-246X.2004.02453.x.

Ursin, B., and B. O. Ekren, 1995, Robust AVO analysis: Geophysics, **60**, no. 2, 317–326, http://dx.doi.org/10.1190/1.1443768.

van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: International Journal of Geophysics, **2011**, http://dx.doi.org/10.1155/2011/689041, available as a PDF at https://www.hindawi.com/journals/ijge/2011/689041/, accessed 15 September 2016.

van Leeuwen, T., and F. J. Herrmann, 2012, Fast waveform inversion without source-encoding: Geophysical Prospecting, **61**, 10–19, version of record online July 2012, special online issue 2013, http://dx.doi.org/10.1111/j.1365-2478.2012.01096.x.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, no. 1, 661–667, http://dx.doi.org/10.1093/gji/ggt258.

van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, no. 3, 1383–1394, http://dx.doi.org/10.1111/j.1365-246X.2010.04681.x.

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, no. 2, R63–R77, http://dx.doi.org/10.1190/geo2013-0055.1.

Vigh, D., N. Moldoveanu, K. Jiao, W. Huang, and J. Kapoor, 2013, Ultralong-offset data acquisition can complement full-waveform inversion and lead to improved subsalt imaging: The Leading Edge, **32**, no. 9, 1116–1122, http://dx.doi.org/10.1190/tle32091116.1.

Vigh, D., and E. W. Starr, 2008, 3D prestack plane-wave, full-waveform inversion: Geophysics, **73**, no. 5, VE135–VE144, http://dx.doi.org/10.1190/1.2952623.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, http://dx.doi.org/10.1190/1.3238367.

Warner, M., and L. Guasch, 2014, Adaptative waveform inversion — FWI without cycle skipping — Theory: 76th Conference and Exhibition, EAGE, Extended Abstracts, We E106 13, http://dx.doi.org/10.3997/2214-4609.20141092.

Warner, M., T. Nangoo, N. Shah, A. Umpleby, and J. Morgan, 2013a, Full-waveform inversion of cycle-skipped seismic data by frequency down-shifting: 83rd Annual International Meeting, SEG, Expanded Abstracts, 903–907, http://dx.doi.org/10.1190/segam2013-1067.1.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013b, Anisotropic 3D full-waveform inversion: Geophysics, **78**, no. 2, R59–R80, http://dx.doi.org/10.1190/geo2012-0338.1.

Weglein, A., H. Zhang, A. C. Ramírez, F. Liu, and J. E. Lira, 2009, Clarifying the underlying and fundamental meaning of the approximate linear inversion of seismic data: Geophysics, **74**, no. 6, WCD1–WCD13, http://dx.doi.org/10.1190/1.3256286.

Woodward, M. J., 1992, Wave-equation tomography: Geophysics, **57**, no. 1, 15–26, http://dx.doi.org/10.1190/1.1443179.

Wu, R. S., 1985, Multiple scattering and energy transfer of seismic waves — Separation of scattering effect from intrinsic attenuation, I: Theoretical modelling: Geophysical Journal International, **82**, no. 1, 57–80, http://dx.doi.org/10.1111/j.1365-246X.1985.tb05128.x.

Wu, R. S., and K. Aki, 1988, Multiple scattering and energy transfer of seismic waves — Separation of scattering effect from intrinsic attenuation, II: Application of the theory to Hindu Kush region: Pure and Applied Geophysics, **128**, no. 1, 49–80, http://dx.doi.org/10.1007/BF01772590.

Wu, R. S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, no. 1, 11–25, http://dx.doi.org/10.1190/1.1442237.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016, A review on the systematic formulation of 3-D multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal International, **207**, no. 1, 129–149, http://dx.doi.org/10.1093/gji/ggw262.

Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, no. 3, 1535–1554, http://dx.doi.org/10.1093/gji/ggv228.

Zhu, H., E. Bozdağ, and J. Tromp, 2015, Seismic structure of the European upper mantle based on adjoint tomography: Geophysical Journal International, **201**, no. 1, 18–52, http://dx.doi.org/10.1093/gji/ggu492.