An overview of full-waveform inversion in exploration geophysics

J. Virieux[1] and S. Operto[2]

ABSTRACT

Full-waveform inversion (FWI) is a challenging data-fitting procedure based on full-wavefield modeling to extract quantitative information from seismograms. High-resolution imaging at half the propagated wavelength is expected. Recent advances in high-performance computing and multifold/multicomponent wide-aperture and wide-azimuth acquisitions make 3D acoustic FWI feasible today. Key ingredients of FWI are an efficient forward-modeling engine and a local differential approach, in which the gradient and the Hessian operators are efficiently estimated. Local optimization does not, however, prevent convergence of the misfit function toward local minima because of the limited accuracy of the starting model, the lack of low frequencies, the presence of noise, and the approximate modeling of the wave-physics complexity. Different hierarchical multiscale strategies are designed to mitigate the nonlinearity and ill-posedness of FWI by incorporating progressively shorter wavelengths in the parameter space. Synthetic and real-data case studies address reconstructing various parameters, from V_P and V_S velocities to density, anisotropy, and attenuation. This review attempts to illuminate the state of the art of FWI. Crucial jumps, however, remain necessary to make it as popular as migration techniques. The challenges can be categorized as (1) building accurate starting models with automatic procedures and/or recording low frequencies, (2) defining new minimization criteria to mitigate the sensitivity of FWI to amplitude errors and increasing the robustness of FWI when multiple parameter classes are estimated, and (3) improving computational efficiency by data-compression techniques to make 3D elastic FWI feasible.

INTRODUCTION

Seismic waves bring to the surface information gathered on the physical properties of the earth. Since the discovery of modern seismology at the end of the 19th century, the main discoveries have arisen from using traveltime information (Oldham, 1906; Gutenberg, 1914; Lehmann, 1936). Then there was a hiatus until the 1980s for amplitude interpretation, when global seismic networks could provide enough calibrated seismograms to compute accurate synthetic seismograms using normal-mode summation. Differential seismograms estimated through the Born approximation have been used as perturbations for matching long-period seismograms, which can provide high-resolution upper-mantle tomography (Gilbert and Dziewonski, 1975; Woodhouse and Dziewonski, 1984). The sensitivity or Fréchet derivative matrix, i.e., the partial derivative of seismic data with respect to the model parameters, is explicitly estimated before proceeding to inversion of the linearized system. The normal-mode description allows a limited number of parameters to be inverted (a few hundred parameters), which makes the optimization procedure feasible through explicit sensitivity matrix estimation in spite of the high number of seismograms.

Meanwhile, exploration seismology has taken up the challenge of high-resolution imaging of the subsurface by designing dense, multifold acquisition systems. Construction of the sensitivity matrix is too prohibitive because the number of parameters exceed 10,000. Instead, another road has been taken to perform high-resolution imaging. Using the exploding-reflector concept, and after some kinematic corrections, amplitude summation has provided detailed images of the subsurface for reservoir determination and characterization (Claerbout, 1971, 1976). The sum of the traveltimes from a specific point of the interface toward the source and the receiver should coincide with the time of large amplitudes in the seismogram. The reflectivity as an amplitude attribute of related seismic traces at the selected point of the reflector provides the migrated image needed for seismic stratigraphic interpretation. Although migration is more a concept for converting seismic data recorded in the time-space domain

into images of physical properties, we often refer to it as the geometric description of the short wavelengths of the subsurface. A velocity macromodel or background model provides the kinematic information required to focus waves inside the medium.

The limited offsets recorded by seismic reflection surveys and the limited-frequency bandwidth of seismic sources make seismic imaging poorly sensitive to intermediate wavelengths (Jannane et al., 1989). This is the motivation behind a two-step workflow: construct the macromodel using kinematic information, and then the amplitude projection using different types of migrations (Claerbout and Doherty, 1972; Gazdag, 1978; Stolt, 1978; Baysal et al., 1983; Yilmaz, 2001; Biondi and Symes, 2004). This procedure is efficient for relatively simple geologic targets in shallow-water environments, although more limited performances have been achieved for imaging complex structures such as salt domes, subbasalt targets, thrust belts, and foothills. In complex geologic environments, building an accurate velocity background model for migration is challenging. Various approaches for iterative updating of the macromodel reconstruction have been proposed (Snieder et al., 1989; Docherty et al., 2003), but they remain limited by the poor sensitivity of the reflection seismic data to the large and intermediate wavelengths of the subsurface.

Simultaneous with the global seismology inversion scheme, Lailly (1983) and Tarantola (1984) recast the migration imaging principle of Claerbout (1971, 1976) as a local optimization problem, the aim of which is least-squares minimization of the misfit between recorded and modeled data. They show that the gradient of the misfit function along which the perturbation model is searched can be built by crosscorrelating the incident wavefield emitted from the source and the back-propagated residual wavefields. The perturbation model obtained after the first iteration of the local optimization looks like a migrated image obtained by reverse-time migration. One difference is that the seismic wavefield recorded at the receiver is back propagated in reverse time migration, whereas the data misfit is back propagated in the waveform inversion of Lailly (1983) and Tarantola (1984). When added to the initial velocity, the velocity perturbations lead to an updated velocity model, which is used as a starting model for the next iteration of minimizing the misfit function. The impressive amount of data included in seismograms (each sample of a time series must be considered) is involved in gradient estimation. This estimation is performed by summation over sources, receivers, and time.

Waveform-fitting imaging was quite computer demanding at that time, even for 2D geometries (Gauthier et al., 1986). However, it has been applied successfully in various studies using forward-modeling techniques such as reflectivity techniques in layered media (Kormendi and Dietrich, 1991), finite-difference techniques (Kolb et al., 1986; Ikelle et al., 1988; Crase et al., 1990; Pica et al., 1990; Djikpéssé and Tarantola, 1999), finite-element methods (Choi et al., 2008), and extended ray theory (Cary and Chapman, 1988; Koren et al., 1991; Sambridge and Drijkoningen, 1992). A less computationally intensive approach is achieved by Jin et al. (1992) and Lambaré et al. (1992), who establish the theoretical connection between ray-based generalized Radon reconstruction techniques (Beylkin, 1985; Bleistein, 1987; Beylkin and Burridge, 1990) and least-squares optimization (Tarantola, 1987). By defining a specific norm in the data space, which varies from one focusing point to the next, they were able to recast the asymptotic Radon transform as an iterative least-squares optimization after diagonalizing the Hessian operator. Applications on 2D synthetic data and real data are provided (Thierry et

al., 1999b; Operto et al., 2000) and 3D extension is possible (Thierry et al., 1999a; Operto et al., 2003) because of efficient asymptotic forward modeling (Lucio et al., 1996). Because the Green's functions are computed in smoothed media with the ray theory, the forward problem is linearized with the Born approximation, and the optimization is iterated linearly, which means the background model remains the same over the iterations. These imaging methods are generally called migration/inversion or true-amplitude prestack depth migration (PSDM). The main difference with the waveform-inversion methods we describe is that the smooth background model does not change over iterations and only the single scattered wavefield is modeled by linearizing the forward problem.

Alternatively, the full information content in the seismogram can be considered in the optimization. This leads us to full-waveform inversion (FWI), where full-wave equation modeling is performed at each iteration of the optimization in the final model of the previous iteration. All types of waves are involved in the optimization, including diving waves, supercritical reflections, and multiscattered waves such as multiples. The techniques used for the forward modeling vary and include volumetric methods such as finite-element methods (Marfurt, 1984; Min et al., 2003), finite-difference methods (Virieux, 1986), finite-volume methods (Brossier et al., 2008), and pseudospectral methods (Danecek and Seriani, 2008); boundary integral methods such as reflectivity methods (Kennett, 1983); generalized screen methods (Wu, 2003); discrete wavenumber methods (Bouchon et al., 1989); generalized ray methods such as WKBJ and Maslov seismograms (Chapman, 1985); full-wave theory (de Hoop, 1960); and diffraction theory (Klem-Musatov and Aizenberg, 1985).

FWI has not been recognized as an efficient seismic imaging technique because pioneering applications were restricted to seismic reflection data. For short-offset acquisition, the seismic wavefield is rather insensitive to intermediate wavelengths; therefore, the optimization cannot adequately reconstruct the true velocity structure through iterative updates. Only when a sufficiently accurate initial model is provided can waveform-fitting converge to the velocity structure through such updates. For sampling the initial model, sophisticated investigations with global and semiglobal techniques (Koren et al., 1991; Jin and Madariaga, 1993, 1994; Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002) have been performed. The rather poor performance of these investigations that arises from insensitivity to intermediate wavelengths has led many researchers to believe that this optimization technique is not particularly efficient.

Only with the benefit of long-offset and transmission data to reconstruct the large and intermediate wavelengths of the structure has FWI reached its maturity as highlighted by Mora (1987, 1988), Pratt and Worthington (1990), Pratt et al. (1996), and Pratt (1999). FWI attempts to characterize a broad and continuous wavenumber spectrum at each point of the model, reunifying macromodel building and migration tasks into a single procedure. Historical crosshole and wide-angle surface data examples illustrate the capacity of simultaneous reconstruction of the entire spatial spectrum (e.g., Pratt, 1999; Ravaut et al., 2004; Brenders and Pratt, 2007a). However, robust application of FWI to long-offset data remains challenging because of increasing nonlinearities introduced by wavefields propagated over several tens of wavelengths and various incidence angles (Sirgue, 2006).

Here, we consider the main aspects of FWI. First, we review the forward-modeling problem that underlies FWI. Efficient numerical

modeling of the full seismic wavefield is a central issue in FWI, especially for 3D problems.

In the second part, we review the main theoretical aspects of FWI based on a least-squares local optimization approach. We follow the compact matrix formalism for its simplicity (Pratt et al., 1998; Pratt, 1999), which leads to a clear interpretation of the gradient and the Hessian of the objective function. Once the gradient is estimated, we review different optimization algorithms used to compute the perturbation model. We conclude the methodology section by the source estimation problem in FWI.

In the third part, we review some key features of FWI. First, we highlight the relationships between the experimental setup (source bandwidth, acquisition geometry) and the spatial resolution of FWI. The resolution analysis provides the necessary guidelines to design the multiscale FWI algorithms required to mitigate the nonlinearity of FWI. We discuss the pros and cons of the time and frequency domains for efficient multiscale algorithms. We provide a few words concerning the parallel implementation of FWI techniques because these are computer demanding. Then we review some alternatives to the least-squares criterion and the Born linearization. A key issue of FWI is the initial model from which the local optimization is started. We also discuss several tomographic approaches to building a starting model.

In the fourth part, we review the main case studies of FWI subdivided into three categories of case studies: acoustic, multiparameter, and three dimensional. Finally, we discuss the future challenges raised by the revival of interest in FWI that has been shown by the exploration and the earthquake-seismology communities.

THE FORWARD PROBLEM

Let us first introduce the notations for the forward problem, namely, modeling the full seismic wavefield. The reader is referred to Robertsson et al. (2007) for an up-to-date series of publications on modern seismic-modeling methods.

We use matrix notations to denote the partial-differential operators of the wave equation (Marfurt, 1984; Carcione et al., 2002). The most popular direct method to discretize the wave equation in the time and frequency domains is the finite-difference method (Virieux, 1986; Levander, 1988; Graves, 1996; Operto et al., 2007), although more sophisticated finite-element or finite-volume approaches can be considered. This is especially true when accurate boundary conditions through unstructured meshes must be implemented (e.g., Komatitsch and Vilotte, 1998; Dumbser and Kaser, 2006).

In the time domain, we have

$$\mathbf{M}(\mathbf{x})\frac{d^2\mathbf{u}(\mathbf{x},t)}{dt^2} = \mathbf{A}(\mathbf{x})\mathbf{u}(\mathbf{x},t) + \mathbf{s}(\mathbf{x},t), \qquad (1)$$

where $\mathbf{M}$ and $\mathbf{A}$ are the mass and the stiffness matrices, respectively (Marfurt, 1984). The source term is denoted by $\mathbf{s}$ and the seismic wavefield by $\mathbf{u}$. In the acoustic approximation, $\mathbf{u}$ generally represents pressure, although in the elastic case $\mathbf{u}$ generally represents horizontal and vertical particle velocities. The time is denoted by t and the spatial coordinates by $\mathbf{x}$. Equation 1 generally is solved with an explicit time-marching algorithm: The value of the wavefield at a time step $(n + 1)$ at a spatial position is inferred from the value of the wavefields at previous time steps. Implicit time-marching algorithms are avoided because they require solving a linear system (Marfurt, 1984). If both velocity and stress wavefields are helpful,

the system of second-order equations can be recast as a first-order hyperbolic velocity-stress system by incorporating the necessary auxiliary variables (Virieux, 1986).

In the frequency domain, the wave equation reduces to a system of linear equations; the right-hand side is the source and the solution is the seismic wavefield. This system can be written compactly as

$$\mathbf{B}(\mathbf{x},\omega)\mathbf{u}(\mathbf{x},\omega) = \mathbf{s}(\mathbf{x},\omega), \qquad (2)$$

where $\mathbf{B}$ is the impedance matrix (Marfurt, 1984). The sparse complex-valued matrix $\mathbf{B}$ has a symmetric pattern, although it is not symmetric because of absorbing boundary conditions (Hustedt et al., 2004; Operto et al., 2007).

Equation 2 can be solved by a decomposition of $\mathbf{B}$ such as lower and upper (LU) triangular decomposition, leading to direct-solver techniques. The advantage of the direct-solver approach is that once the decomposition is performed, equation 2 is efficiently solved for multiple sources using forward and backward substitutions (Marfurt, 1984). The direct-solver approach is efficient for 2D forward problems (Jo et al., 1996; Stekl and Pratt, 1998; Hustedt et al., 2004). However, the time and memory complexities of LU factorization and its limited scalability on large-scale distributed memory platforms prevent use of the approach for large-scale 3D problems (i.e., problems involving more than 10 million unknowns; Operto et al., 2007).

Iterative solvers provide an alternative approach for solving the time-harmonic wave equation (Riyanti et al., 2006, 2007; Plessix, 2007; Erlangga and Herrmann, 2008). Iterative solvers currently are implemented with Krylov subspace methods (Saad, 2003) that are preconditioned by solving the dampened time-harmonic wave equation. The solution of the dampened wave equation is computed with one cycle of a multigrid. The main advantage of the iterative approach is the low memory requirement, although the main drawback results from a difficulty to design an efficient preconditioner because the impedance matrix is indefinite. To our knowledge, the extension to elastic wave equations still needs to be investigated. As for the time-domain approach, the time complexity of the iterative approach increases linearly with the number of sources in contrast to the direct-solver approach.

An intermediate approach between the direct and iterative methods consists of a hybrid direct-iterative approach based on a domain decomposition method and the Schur complement system (Saad, 2003; Sourbier et al., 2008). The iterative solver is used to solve the reduced Schur complement system, the solution of which is the wavefield at interface nodes between subdomains. The direct solver is used to factorize local impedance matrices that are assembled on each subdomain. Briefly, the hybrid approach provides a compromise in terms of memory savings and multisource-simulation efficiency between the direct and the iterative approaches.

The last possible approach to compute monochromatic wavefields is to perform the modeling in the time domain and extract the frequency-domain solution either by discrete Fourier transform in the loop over the time steps (Sirgue et al., 2008) or by phase-sensitivity detection once the steady-state regime is reached (Nihei and Li, 2007). One advantage of the approach based on the discrete Fourier transform is that an arbitrary number of frequencies can be extracted within the loop over time steps at minimal extra cost. Second, time windowing can be easily applied, which is not the case when the modeling is performed in the frequency domain. Time windowing allows the extraction of specific arrivals for FWI (early arrivals, reflections, PS converted waves), which is often useful to mitigate the

nonlinearity of the inversion by judicious data preconditioning (Sears et al., 2008; Brossier et al., 2009a).

Among all of these possible approaches, the iterative-solver approach theoretically has the best time complexity (here, "complexity" denotes how the computational cost of an algorithm grows with the size of the computational domain) if the number of iterations is independent of the frequency (Erlangga and Herrmann, 2008). In practice, the number of iterations generally increases linearly with frequency. In this case, the time complexities of the time-domain and the iterative-solver approach are equivalent (Plessix, 2007).

The reader is referred to Plessix (2007, 2009) and Virieux et al. (2009) for more detailed complexity analyses of seismic modeling based on different numerical approaches. A discussion on the pros and cons of time-domain versus frequency-domain seismic modeling with application to FWI is also provided in Vigh and Starr (2008b) and Warner et al. (2008).

Source implementation is an important issue in FWI. The spatial reciprocity of Green's functions can be exploited in FWI to mitigate the number of forward problems if the number of receivers is significantly smaller than the number of sources (Aki and Richards, 1980). The reciprocity of Green's functions also allows matching data emitted by explosions and recorded by directional sensors, with pressure synthetics computed for directional forces (Operto et al., 2006). Of note, the spatial reciprocity is satisfied theoretically for the unidirectional sensor and the unidirectional impulse source. However, the spatial reciprocity of the Green's functions can also be used for explosive sources by virtue of the superposition principle. Indeed, explosions can be represented by double dipoles or, in other words, by four unidirectional impulse sources.

A final comment concerns the relationship between the discretization required to solve the forward problem and that required to reconstruct the physical parameters. Often during FWI, these two discretizations are identical, although it is recommended that the fingerprint of the forward problem be kept minimal in FWI.

The properties of the subsurface that we want to quantify are embedded in the coefficients of matrices $\mathbf{M}$, $\mathbf{A}$, or $\mathbf{B}$ of equations 1 and 2. The relationship between the seismic wavefield and the parameters is nonlinear and can be written compactly through the operator $\mathbf{G}$, defined as

$$\mathbf{u} = \mathbf{G}(\mathbf{m}) \qquad (3)$$

in the time domain or in the frequency domain.

FWI AS A LEAST-SQUARES LOCAL OPTIMIZATION

We follow the simplest view of FWI based on the so-called length method (Menke, 1984). For information on probabilistic maximum likelihood or generalized inverse formulations, the reader is referred to Menke (1984), Tarantola (1987), Scales and Smith (1994), and Sen and Stoffa (1995).

We define the misfit vector $\Delta\mathbf{d} = \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}(\mathbf{m})$ of dimension N by the differences at the receiver positions between the recorded seismic data $\mathbf{d}_{\text{obs}}$ and the modeled seismic data $\mathbf{d}_{\text{cal}}(\mathbf{m})$ for each source-receiver pair of the seismic survey. Here, $\mathbf{d}_{\text{cal}}$ can be related to the modeled seismic wavefield $\mathbf{u}$ by a detection operator $\mathcal{R}$, which extracts the values of the wavefields computed in the full computational domain at the receiver positions for each source: $\mathbf{d}_{\text{cal}} = \mathcal{R}\mathbf{u}$. The model $\mathbf{m}$ represents some physical parameters of the subsurface discretized over the computational domain.

In the simplest case corresponding to the monoparameter acoustic approximation, the model parameters are the P-wave velocities defined at each node of the numerical mesh used to discretize the inverse problem. In the extreme case, the model parameters correspond to the 21 elastic moduli that characterize linear triclinic elastic media, the density, and some memory variables that characterize the anelastic behavior of the subsurface (Toksöz and Johnston, 1981). The most common discretization consists of projection of the continuous model of the subsurface on a multidimensional Dirac comb, although a more complex basis can be considered (see Appendix A in Pratt et al. [1998] for a discussion on alternative parameterizations). We define a norm $\mathcal{C}(\mathbf{m})$ of this misfit vector $\Delta\mathbf{d}$, which is referred to as the misfit function or the objective function. We focus below on the least-squares norm, which is easier to manipulate mathematically (Tarantola, 1987). Other norms are discussed later.

The Born approximation and the linearization of the inverse problem

The least-squares norm is given by

$$\mathcal{C}(\mathbf{m}) = \frac{1}{2}\Delta\mathbf{d}^{\dagger}\Delta\mathbf{d}, \qquad (4)$$

where $\dagger$ denotes the adjoint operator (transpose conjugate).

In the time domain, the implicit summation in equation 4 is performed over the number of source-channel pairs and the number of time samples in the seismograms, where a channel is one component of a multicomponent sensor. In the frequency domain, the summation over frequencies replaces that over time. In the time domain, the misfit vector is real valued; in the frequency domain, it is complex valued.

The minimum of the misfit function $\mathcal{C}(\mathbf{m})$ is sought in the vicinity of the starting model $\mathbf{m}_0$. The FWI is essentially a local optimization. In the framework of the Born approximation, we assume that the updated model $\mathbf{m}$ of dimension M can be written as the sum of the starting model $\mathbf{m}_0$ plus a perturbation model $\Delta\mathbf{m}$: $\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}$. In the following, we assume that $\mathbf{m}$ is real valued.

A second-order Taylor-Lagrange development of the misfit function in the vicinity of $\mathbf{m}_0$ gives the expression

$$C(\mathbf{m}_0 + \Delta\mathbf{m}) = C(\mathbf{m}_0) + \sum_{j=1}^{M} \frac{\partial C(\mathbf{m}_0)}{\partial m_j} \Delta m_j$$
$$+ \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M} \frac{\partial^2 C(\mathbf{m}_0)}{\partial m_j \partial m_k} \Delta m_j \Delta m_k + \mathcal{O}(\mathbf{m}^3). \qquad (5)$$

Taking the derivative with respect to the model parameter m_l results in

$$\frac{\partial C(\mathbf{m})}{\partial m_l} = \frac{\partial C(\mathbf{m}_0)}{\partial m_l} + \sum_{j=1}^{M} \frac{\partial^2 C(\mathbf{m}_0)}{\partial m_j \partial m_l} \Delta m_j. \qquad (6)$$

The minimum of the misfit function in the vicinity of point $\mathbf{m}_0$ is reached when the first derivative of the misfit function vanishes. This gives the perturbation model vector:

$$\Delta\mathbf{m} = -\left[\frac{\partial^2 C(\mathbf{m}_0)}{\partial \mathbf{m}^2}\right]^{-1} \frac{\partial C(\mathbf{m}_0)}{\partial \mathbf{m}}. \qquad (7)$$

The perturbation model is searched in the opposite direction of the steepest ascent (i.e., the gradient) of the misfit function at point $\mathbf{m}_0$. The second derivative of the misfit function is the Hessian; it defines the curvature of the misfit function at $\mathbf{m}_0$. Of note, the error term $\mathcal{O}(\mathbf{m}^3)$ in equation 5 is zero when the misfit function is a quadratic function of $\mathbf{m}$. This is the case for linear forward problems such as $\mathbf{u} = \mathbf{G}.\mathbf{m}$. In this case, the expression of the perturbation model of equation 7 gives the minimum of the misfit function in one iteration. In FWI, the relationship between the data and the model is nonlinear and the inversion needs to be iterated several times to converge toward the minimum of the misfit function.

Normal equations: The Newton, Gauss-Newton, and steepest-descent methods

Basic equations

The derivative of $C(\mathbf{m})$ with respect to the model parameter m_l gives

$$\frac{\partial C(\mathbf{m})}{\partial m_l} = -\frac{1}{2}\sum_{i=1}^{N}\left[\left(\frac{\partial d_{\mathrm{cal}_i}}{\partial m_l}\right)(d_{\mathrm{obs}_i} - d_{\mathrm{cal}_i})^*\right.$$

$$\left.+ (d_{\mathrm{obs}_i} - d_{\mathrm{cal}_i})\frac{\partial d_{\mathrm{cal}_i}^*}{\partial m_l}\right]$$

$$= -\sum_{i=1}^{N}\Re\left[\left(\frac{\partial d_{\mathrm{cal}_i}}{\partial m_l}\right)^*(d_{\mathrm{obs}_i} - d_{\mathrm{cal}_i})\right], \quad (8)$$

where the real part and the conjugate of a complex number are denoted by $\Re$ and $*$, respectively. In matrix form, equation 8 translates to

$$\nabla C_{\mathbf{m}} = \frac{\partial C(\mathbf{m})}{\partial \mathbf{m}} = -\Re\left[\left(\frac{\partial \mathbf{d}_{\mathrm{cal}}(\mathbf{m})}{\partial \mathbf{m}}\right)^{\dagger}(\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{\mathrm{cal}}(\mathbf{m}))\right]$$

$$= -\Re[\mathbf{J}^{\dagger}\Delta\mathbf{d}], \quad (9)$$

where $\mathbf{J}$ is the sensitivity or the Fréchet derivative matrix. In equation 9, $\nabla C_{\mathbf{m}}$ is a vector of dimension M. Taking $\mathbf{m} = \mathbf{m}_0$ in equation 9 provides the descent direction along which the perturbation model is searched in equation 7.

Differentiation of the gradient expression 8, with respect to the model parameters gives the following expression in matrix form for the Hessian (see Pratt et al. [1998] for details):

$$\frac{\partial^2 C(\mathbf{m}_0)}{\partial \mathbf{m}^2} = \Re[\mathbf{J}_0^{\dagger}\mathbf{J}_0] + \Re\left[\frac{\partial \mathbf{J}_0^t}{\partial \mathbf{m}^t}(\Delta\mathbf{d}_0^* \ldots \Delta\mathbf{d}_0^*)\right]. \quad (10)$$

Inserting the expression of the gradient (equation 9) and the Hessian (equation 10) into equation 7 gives the following for the perturbation model:

$$\Delta\mathbf{m} = -\left\{\Re\left[\mathbf{J}_0^{\dagger}\mathbf{J}_0 + \frac{\partial \mathbf{J}_0^t}{\partial \mathbf{m}^t}(\Delta\mathbf{d}_0^* \ldots \Delta\mathbf{d}_0^*)\right]\right\}^{-1}\Re[\mathbf{J}_0^{\dagger}\Delta\mathbf{d}_0]. \quad (11)$$

The method solving the normal equations, e.g., equation 11, generally is referred to as the Newton method, which is locally quadratically convergent.

For linear problems ($\mathbf{d}=\mathbf{G}.\mathbf{m}$), the second term in the Hessian is zero because the second-order derivative of the data with respect to

model parameters is zero. Most of the time, this second-order term is neglected for nonlinear inverse problems. In the following, the remaining term in the Hessian, i.e., $\mathbf{H}_a = \mathbf{J}_0^{\dagger}\mathbf{J}_0$, is referred to as the approximate Hessian. The method which solves equation 11 when only $\mathbf{H}_a$ is estimated is referred to as the Gauss-Newton method.

Alternatively, the inverse of the Hessian in equation 11 can be replaced by a scalar α, the so-called step length, leading to the gradient or steepest-descent method. The step length can be estimated by a line-search method, for which a linear approximation of the forward problem is used (Gauthier et al., 1986; Tarantola, 1987). In the linear approximation framework, the second-order Taylor-Lagrange development of the misfit function gives

$$C(\mathbf{m} - \alpha\nabla C(\mathbf{m}_0)) = C(\mathbf{m}) - \alpha\langle\nabla C(\mathbf{m})|\nabla C(\mathbf{m}_0)\rangle$$

$$+ \frac{1}{2}\alpha^2\mathbf{H}_a(\mathbf{m})\langle\nabla C(\mathbf{m}_0)|\nabla C(\mathbf{m}_0)\rangle, \quad (12)$$

where we assume a model perturbation of the form $\Delta\mathbf{m} = \alpha\nabla C(\mathbf{m}_0)$. In equation 12, we replace the second-order derivative of the misfit function by the approximate Hessian in the second term on the right-hand side. Inserting the expression of the approximate Hessian $\mathbf{H}_a$ into the previous expression, zeroing the partial derivative of the misfit function with respect to α, and using $\mathbf{m} = \mathbf{m}_0$ gives

$$\alpha = \frac{\langle\nabla C(\mathbf{m}_0)|\nabla C(\mathbf{m}_0)\rangle}{\langle\,\mathbf{J}^t(\mathbf{m}_0)\nabla C(\mathbf{m}_0)|\mathbf{J}^t(\mathbf{m}_0)\nabla C(\mathbf{m}_0))\rangle}. \quad (13)$$

The term $\mathbf{J}^t(\mathbf{m}_0)\nabla C(\mathbf{m}_0)$ is computed conventionally using a first-order-accurate finite-difference approximation of the partial derivative of $\mathbf{G}$,

$$\frac{\partial \mathbf{G}(\mathbf{m}_0)}{\partial \mathbf{m}}\nabla C(\mathbf{m}_0) = \frac{1}{\varepsilon}(\mathbf{G}(\mathbf{m}_0 + \varepsilon\nabla C(\mathbf{m}_0)) - \mathbf{G}(\mathbf{m}_0)), \quad (14)$$

with a small parameter ε. Estimation of α requires solving an extra forward problem per shot for the perturbed model $\mathbf{m}_0 + \varepsilon\nabla C(\mathbf{m}_0)$. This line-search technique is extended to multiple-parameter classes by Sambridge et al. (1991) using a subspace approach. In this case, one forward problem must be solved per parameter class, which can be computationally expensive. Alternatively, the step length can be estimated by parabolic interpolation through three points, ($\alpha, C(\mathbf{m}_0 + \alpha\nabla C(\mathbf{m}_0))$). The minimum of the parabola provides the desired α. In this case, two extra forward problems per shot must be solved because we already have a third point corresponding to $(0, C(\mathbf{m}_0))$ (see Figure 1 in Vigh et al. [2009] for an illustration).

Pratt et al. (1998) illustrate how quality and rate of convergence of the inversion depend significantly on the Newton, Gauss-Newton, or gradient method used. Importantly, they show how the gradient method can fail to converge toward an acceptable model, however many iterations, unlike the Newton and Gauss-Newton methods. They interpret this failure as the result of the difficulty of estimating a reliable step length. However, gradient methods can be significantly improved by scaling (i.e., dividing) the gradient by the diagonal terms of $\mathbf{H}_a$ or of the pseudo-Hessian (Shin et al., 2001a).

Numerical algorithms: The conjugate-gradient method

Over the last decade, the most popular local optimization algorithm for solving FWI problems was based on the conjugate-gradient method (Mora, 1987; Tarantola, 1987; Crase et al., 1990). Here, the model is updated at the iteration n in the direction of $\mathbf{p}^{(n)}$, which is a linear combination of the gradient at iteration n, $\nabla C^{(n)}$, and the direction $\mathbf{p}^{(n-1)}$:

$$\mathbf{p}^{(n)} = \nabla C^n + \beta^{(n)} \mathbf{p}^{(n-1)}. \tag{15}$$

The scalar $\beta^{(n)}$ is designed to guarantee that $\mathbf{p}^{(n)}$ and $\mathbf{p}^{(n-1)}$ are conjugate. Among the different variants of the conjugate-gradient method to derive the expression of $\beta^{(n)}$, the Polak-Ribière formula (Polak and Ribière, 1969) is generally used for FWI:

$$\beta^{(n)} = \frac{\langle \nabla C^{(n)} - \nabla C^{(n-1)} | \nabla C^{(n)} \rangle}{\| \nabla C^{(n)} \|^2}. \tag{16}$$

In FWI, the preconditioned gradient $W_m^{-1} \nabla C^{(n)}$ is used for $\mathbf{p}^{(n)}$, where W_m is a weighting operator that is introduced in the next section (Mora, 1987). Only three vectors of dimension M, i.e., $\nabla C^{(n)}$, $\nabla C^{(n-1)}$, and $\mathbf{p^{(n-1)}}$, are required to implement the conjugate-gradient method.

Numerical algorithms: Quasi-Newton algorithms

Finite approximations of the Hessian and its inverse can be computed using quasi-Newton methods such as the BFGS algorithm (named after its discoverers Broyden, Fletcher, Goldfarb, and Shanno; see Nocedal [1980] for a review). The main idea is to update the approximation of the Hessian or its inverse $\mathcal{H}^{(n)}$ at each iteration of the inversion, taking into account the additional knowledge provided by $\nabla C^{(n)}$ at iteration n. In these approaches, the approximation of the Hessian or its inverse is explicitly formed.

For large-scale problems such as FWI in which the cost of storing and working with the approximation of the Hessian matrix is prohibitive, a limited-memory variant of the quasi-Newton BFGS method known as the L-BFGS algorithm allows computing in a recursive manner $\mathcal{H}^{(n)} \nabla C^{(n)}$ without explicitly forming $\mathcal{H}^{(n)}$. Only a few gradients of the previous nonlinear iterations (typically, 3–20 iterations) need to be stored in L-BFGS, which represents a negligible storage and computational cost compared to the conjugate-gradient algorithm (see Nocedal, 1980; p. 177–180). The L-BFGS algorithm requires an initial guess $\mathcal{H}^{(0)}$, which can be provided by the inverse of the diagonal Hessian (Brossier et al., 2009a). For multiparameter FWI, the L-BFGS algorithm provides a suitable scaling of the gradients computed for each parameter class and hence provides a computationally efficient alternative to the subspace method of Sambridge et al. (1991). A comparison between the conjugate-gradient method and the L-BFGS method for a realistic onshore application of multiparameter elastic FWI is shown in Brossier et al. (2009a).

Newton and Gauss-Newton algorithms

The more accurate, although more computationally intensive, Gauss-Newton and Newton algorithms are described in Akcelik (2002), Askan et al. (2007), Askan and Bielak (2008), and Epanomeritakis et al. (2008), with an application to a 2D synthetic model of the San Fernando Valley using the SH-wave equation. At each nonlinear FWI iteration, a matrix-free conjugate-gradient method is used to solve the reduced Karush-Kuhn-Tucker (KKT) optimal system, which turns out to be similar to the normal equation system

(equation 11). Neither the full Hessian nor the sensitivity matrix is formed explicitly; only the application of the Hessian to a vector needs to be performed at each iteration of the conjugate-gradient algorithm.

Application of the Hessian to a vector requires performing two forward problems per shot for the incident and the adjoint wavefields (Akcelik, 2002). Because these two simulations are performed at each iteration of the conjugate-gradient algorithm, an efficient preconditioner must be used to mitigate the number of iterations of the conjugate-gradient algorithm. Epanomeritakis et al. (2008) use a variant of the L-BFGS method for the preconditioner of the conjugate gradient, in which the curvature of the objective function is updated at each iteration of the conjugate gradient using the Hessian-vector products collected over the iterations.

Regularization and preconditioning of inversion

As widely stressed, FWI is an ill-posed problem, meaning that an infinite number of models matches the data. Some regularizations are conventionally applied to the inversion to make it better posed (Menke, 1984; Tarantola, 1987; Scales et al., 1990). The misfit function can be augmented as follows:

$$\mathcal{C}(\mathbf{m}) = \frac{1}{2} \Delta \mathbf{d}^\dagger \mathbf{W}_d \Delta \mathbf{d} + \frac{1}{2} \varepsilon (\mathbf{m} - \mathbf{m}_{\text{prior}})^\dagger \mathbf{W}_m (\mathbf{m} - \mathbf{m}_{\text{prior}}), \tag{17}$$

where $\mathbf{W}_d = \mathbf{S}_d^t \mathbf{S}_d$ and $\mathbf{W}_m = \mathbf{S}_m^t \mathbf{S}_m$. Weighting operators are $\mathbf{W}_d$ and $\mathbf{W}_m$, the inverse of the data and model covariance operators in the frame of the Bayesian formulation of FWI (Tarantola, 1987). The operator $\mathbf{S}_d$ can be implemented as a diagonal weighting operator that controls the respective weight of each element of the data-misfit vector. For example, Operto et al. (2006) use $\mathbf{S}_d$ as a power of the source-receiver offset to strengthen the contribution of large-offset data for crustal-scale imaging. In geophysical applications where the smoothest model that fits the data is often sought, the aim of the least-squares regularization term in the augmented misfit function (equation 17) is to penalize the roughness of the model $\mathbf{m}$, hence defining the so-called Tikhonov regularization (see Hansen [1998] for a review on regularization methods). The operator $\mathbf{S}_m$ is generally a roughness operator, such as the first- or second-difference matrices (Press et al., 1986, 1007).

For linear problems (assuming the second term of the Hessian is neglected), the minimization of the weighted misfit function gives the perturbation model:

$$\Delta \mathbf{m} = - \{\Re(\mathbf{J}_0^\dagger \mathbf{W}_d \mathbf{J}_0) + \varepsilon \mathbf{W}_m\}^{-1} \Re[\mathbf{J}_0^\dagger \mathbf{W}_d \Delta \mathbf{d}_0], \tag{18}$$

where we use $\mathbf{m}_{\text{prior}} = \mathbf{m}_0$. Of note, equation 18 is equivalent to Tarantola (1987, p. 70) and Menke (1984, p. 55):

$$\Delta \mathbf{m} = -\mathbf{W}_m^{-1} \{\Re(\mathbf{J}_0 \mathbf{W}_m^{-1} \mathbf{J}_0^\dagger) + \varepsilon \mathbf{W}_d^{-1}\}^{-1} \Re[\mathbf{J}_0^\dagger \Delta \mathbf{d}_0]. \tag{19}$$

Equation 19 can be more tractable from a computational viewpoint when $N < M$. Because $\mathbf{W}_m$ is a roughness operator, $\mathbf{W}_m^{-1}$ is a smoothing operator. It can be implemented, for example, with a multidimensional adaptive Gaussian smoother (Ravaut et al., 2004; Operto et al., 2006) or with a low-pass filter in the wavenumber domain (Sirgue, 2003).

For the steepest-descent algorithm, the regularized solution for the perturbation model is given by

$$\mathbf{\Delta m} = -\alpha \mathbf{W}_m^{-1} \Re[\mathbf{J}_0^\dagger \mathbf{W}_d \mathbf{\Delta d}_0], \tag{20}$$

where the scaling performed by the diagonal terms of the approximate Hessian can be embedded in the operator $\mathbf{W}_m^{-1}$ in addition to the smoothing operator.

A more complete and rigorous mathematical derivation of these equations is presented in Tarantola (1987).

Alternative regularizations based on minimizing the total variation of the model have been developed mainly by the image-processing and electromagnetic communities. The aim of the total variation or edge-preserving regularization is to preserve both the blocky and the smooth characteristics of the model (Vogel and Oman, 1996; Vogel, 2002). Total variation (TV) regularization is conventionally implemented by minimizing the $\mathcal{L}_1$-norm of the model-misfit function $R_{\mathrm{TV}} = (\mathbf{\Delta m} \mathbf{W}_m \mathbf{\Delta m})^{1/2}$. Alternatively, van den Berg and Abubakar (2001) implement TV regularization as a multiplicative constraint in the original misfit function. In this framework, the original misfit function can be seen as the weighting factor of the regularization term, which is automatically updated by the optimization process without the need for heuristic tuning. TV regularization is applied to FWI in Askan and Bielak (2008). The weighted $\mathcal{L}_2$-norm regularization applied to frequency-domain FWI is shown in Hu et al. (2009) and Abubakar et al. (2009).

The gradient and Hessian in FWI: Interpretation and computation

A clear interpretation of the gradient and Hessian is given by Pratt et al. (1998) using the compact matrix formalism of frequency-domain FWI. A review is given here. Let us consider the forward-problem equation given by equation 2 for one source and one frequency. In the following, we assume that the model is discretized in a finite-difference sense using a uniform grid of nodes.

Differentiation of equation 2 with respect to the model parameter m_l gives the expression of the partial derivative wavefield $\partial\mathbf{u}/\partial m_\ell$ by solving the following system:

$$\mathbf{B}\frac{\partial\mathbf{u}}{\partial m_\ell} = \mathbf{f}^{(\ell)}, \tag{21}$$

where

$$\mathbf{f}^{(\ell)} = -\frac{\partial\mathbf{B}}{\partial m_\ell}\mathbf{u}. \tag{22}$$

An analogy between the forward-problem equation 2 and equation 21 shows that the partial-derivative wavefield can be computed by solving one forward problem, the source of which is given by $\mathbf{f}^{(\ell)}$. This so-called virtual secondary source is formed by the product of $\partial\mathbf{B}/\partial m_\ell$ and the incident wavefield $\mathbf{u}$. The matrix $\partial\mathbf{B}/\partial m_\ell$ is built by differentiating each coefficient of the forward-problem operator $\mathbf{B}$ with respect to m_ℓ. Because the discretized differential operators in $\mathbf{B}$ generally have local support, the matrix $\partial\mathbf{B}/\partial m_\ell$ is extremely sparse.

The spatial support of the virtual secondary source is centered on the position of m_ℓ, whereas the temporal support of $\mathbf{f}^{(\ell)}$ is centered around the arrival time of the incident wavefield at the position of m_ℓ. Therefore, the partial-derivative wavefield with respect to the model parameter m_ℓ can be interpreted as the wavefield emitted by the seismic source $\mathbf{s}$ and scattered by a point diffractor located at m_ℓ. The radiation pattern of the virtual secondary source is controlled by the operator $\partial\mathbf{B}/\partial m_\ell$. Analysis of this radiation pattern for different pa-

rameter classes allows us to assess to what extent parameters of different natures are uncoupled in the tomographic reconstruction as a function of the diffraction angle and to what extent they can be reliably reconstructed during FWI. Radiation patterns for the isotropic acoustic, elastic, and viscoelastic wave equations are shown in Wu and Aki (1985), Tarantola (1986), Ribodetti and Virieux (1996), and Forgues and Lambaré (1997).

Because the gradient is formed by the zero-lag correlation between the partial-derivative wavefield and the data residual, these have the same meaning: They represent perturbation wavefields scattered by the missing heterogeneities in the starting model $\mathbf{m}_0$ (Tarantola, 1984; Pratt et al., 1998). This interpretation draws some connections between FWI and diffraction tomography; the perturbation model can be represented by a series of closely spaced diffractors. By virtue of Huygens' principle, the image of the model perturbations is built by the superposition of the elementary image of each diffractor, and the seismic wavefield perturbation is built by superposition of the wavefields scattered by each point diffractor (Mc-Mechan and Fuis, 1987).

The approximate Hessian is formed by the zero-lag correlation between the partial-derivative wavefields, e.g., equation 10. The diagonal terms of the approximate Hessian contain the zero-lag autocorrelation and therefore represent the square of the amplitude of the partial-derivative wavefield. Scaling the gradient by these diagonal terms removes from the gradient the geometric amplitude of the partial-derivative wavefields and the residuals. In the framework of surface seismic experiments, the effects of the scaling performed by the diagonal Hessian provide a good balance between shallow and deep perturbations. A diagonal Hessian is shown in Ravaut et al. (2004, their Figure 12). The off-diagonal terms of the Hessian are computed by correlation between partial-derivative wavefields associated with different model parameters. For 1D media, the approximate Hessian is a band-diagonal matrix, and the numerical bandwidth decreases as the frequency increases. The off-diagonal elements of the approximate Hessian account for the limited-bandwidth effects that result from the experimental setup. Applying its inverse to the gradient can be interpreted as a deconvolution of the gradient from these limited-bandwidth effects.

An illustration of the scaling and deconvolution effects performed by the diagonal Hessian on one hand and the approximate Hessian on the other hand is provided in Figure 1. A single inclusion in a homogeneous background model (Figure 1a) is reconstructed by one iteration of FWI using a gradient method preconditioned by the diagonal terms of the approximate Hessian (Figure 1b) and by a Gauss-Newton method (Figure 1c). The image of the inclusion is sharper when the Gauss-Newton algorithm is used. The corresponding approximate Hessian and its diagonal elements are shown in Figure 2. An interpretation of the second term of the Hessian (equation 10) is given in Pratt et al. (1998). This term accounts for multiscattering events such as multiples in the reconstruction procedure. Through iterations, we might correct effects caused by this missing term as long as convergence is achieved.

Although equation 21 gives some clear insight into the physical sense of the gradient of the misfit function, it is impractical from a computer-implementation point of view; with the computer explicitly forming the sensitivity matrix with equation 21, it would require performing as many forward problems as the number of model parameters $m_\ell(\ell = 1, M)$ for each source of the survey. To mitigate this computational burden, the spatial reciprocity of Green's functions can be exploited as shown below.

Inserting the expression of the partial derivative of the wavefield (equation 21) in the expression of the gradient of equation 9 gives the following expression of the gradient:

$$\nabla C_\ell = \Re \left[\mathbf{u}^t \left[\frac{\partial \mathbf{B}}{\partial m_\ell} \right]^t \mathbf{B}^{-1^t} (\mathcal{P} \Delta \mathbf{d})^* \right], \qquad (23)$$

where $\mathcal{P}$ denotes an operator that augments the residual data vector with zeroes in the full computational domain so that the dimension of the augmented vector matches the dimension of the matrix $\mathbf{B}^{-1^t}$ (Pratt et al., 1998). The column of $\mathbf{B}^{-1}$ corresponds to the Green's functions for unit impulse sources located at each node of the model. By virtue of the spatial reciprocity of the Green's functions, $\mathbf{B}^{-1}$ is

symmetric. Therefore, $\mathbf{B}^{-1^t}$ can be substituted by $\mathbf{B}^{-1}$ in equation 23, which gives

$$\nabla C_\ell = \Re \left[\mathbf{u}^t \left[\frac{\partial \mathbf{B}}{\partial m_\ell} \right]^t \mathbf{B}^{-1} (\mathcal{P} \Delta \mathbf{d}^*) \right] = \Re \left[\mathbf{u}^t \left[\frac{\partial \mathbf{B}}{\partial m_\ell} \right]^t \mathbf{r}_b \right]. \qquad (24)$$

The wavefield $\mathbf{r}_b$ corresponds to the back-propagated residual wavefield. All of the residuals associated with one seismic source are assembled to form one residual source. The back propagation in time is indicated by the conjugate operator in the frequency domain. The number of forward seismic problems for computing the gradient is reduced to two: one to compute the incident wavefield $\mathbf{u}$ and one to back propagate the corresponding residuals. The underlying imaging principle is reverse-time migration, which relies on the correspondence of the arrival times of the incident wavefield and the
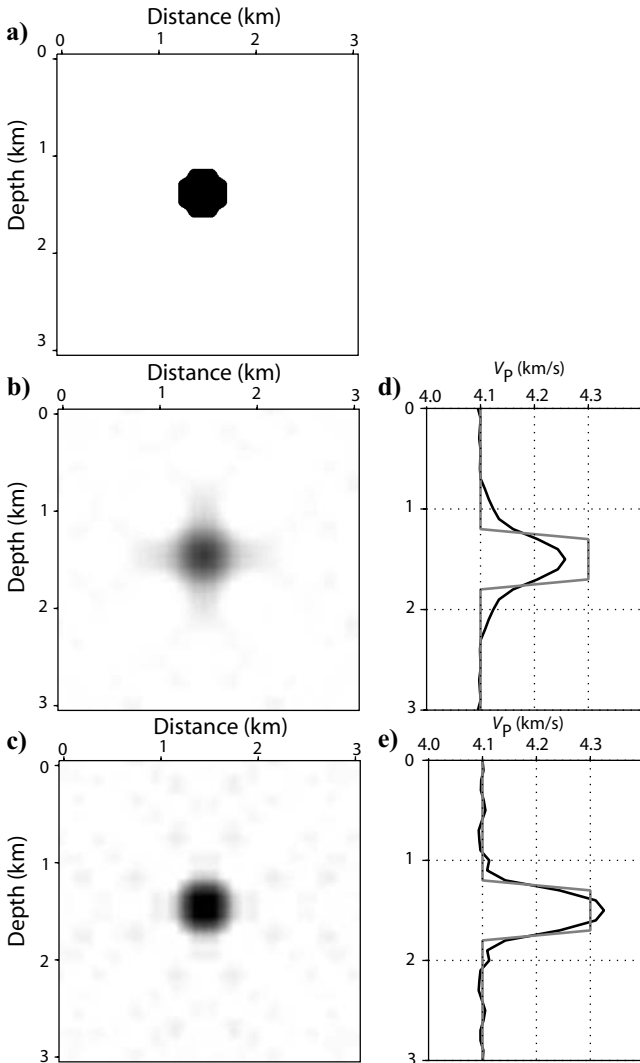

Figure 1. Reconstruction of an inclusion by frequency-domain FWI. (a) True model and FWI models built (b) by a preconditioned gradient method and (c) by a Gauss-Newton method. Four frequencies (4, 5, 7, and 10 Hz) were inverted. One iteration per frequency was computed. Fourteen shots were deployed along the top and left edges of the model. Shots along the top edge were recorded by 14 receivers along the bottom edge; shots along the left edges were recorded by 14 receivers along the right edge. The P-wave velocities in the background and in the inclusion are 4.0 and 4.2 km/s, respectively. Vertical velocity profiles are extracted from the true model (gray line) and the FWI models (black line) for (d) the gradient and (e) the Gauss-Newton inversions.
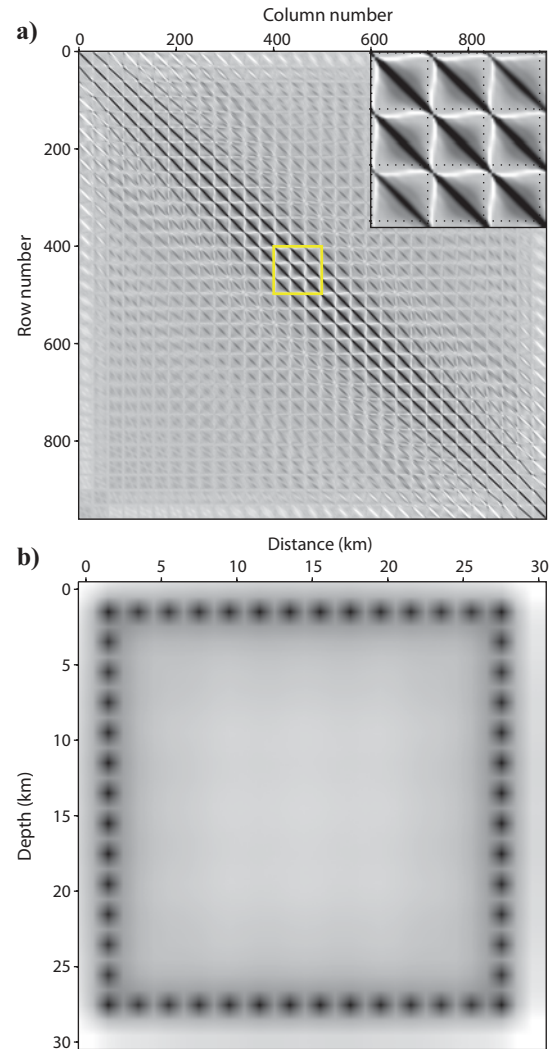

Figure 2. Hessian operator. (a) Approximate Hessian corresponding to the 31 × 31 model of Figure 1 for a frequency of 4 Hz. A close-up of the area delineated by the yellow square highlights the band-diagonal structure of the Hessian. (b) Corresponding diagonal terms of the Hessian plotted in the distance-depth domain. The high-amplitude coefficients indicate source and receiver positions. Scaling the gradient by this map removes the geometric amplitude effects from the wavefields.

back-propagated wavefield at the position of heterogeneity (Claerbout, 1971; Lailly, 1983; Tarantola, 1984).

The approach that consists of computing the gradient of the misfit function without explicitly building the sensitivity matrix is often referred to as the adjoint-wavefield approach by the geophysical community. The underlying mathematical theory is the adjoint-state method of the optimization theory (Lions, 1972; Chavent, 1974). Interesting links exist between optimization techniques used in FWI and assimilation methods, widely used in fluid mechanics (Talagrand and Courtier, 1987). A detailed review of the adjoint-state method with illustrations from several seismic problems is given in Tromp et al. (2005), Askan (2006), Plessix (2006), and Epanomeritakis et al. (2008). The expression of the gradient of the frequency-domain FWI misfit function (equation 24) is derived from the adjoint-state method and the method of the Lagrange multiplier in Appendix A.

For multiple sources and multiple frequencies, the gradient is formed by the summation over these sources and frequencies:

$$\nabla C_\ell = \sum_{i=1}^{N_\omega} \sum_{s=1}^{N_s} \Re \left[[\mathbf{B}_i^{-1}\mathbf{s}_s]^t \left[\frac{\partial \mathbf{B}_i}{\partial m_\ell} \right]^t [\mathbf{B}_i^{-1}(\mathcal{P}\boldsymbol{\Delta}\mathbf{d}_{i,s}^*)] \right].$$

(25)

We also need to note that matrices $\mathbf{B}_i^{-1}(i = 1, N_\omega)$ do not depend on shots; therefore, any speedup toward resolving systems that involve these matrices with multiple sources should be considered (Marfurt, 1984; Jo et al., 1996; Stekl and Pratt, 1998).

By comparing the expressions of the gradient in equations 9 and 24, we can conclude that one element of the sensitivity matrix is given by

$$\mathbf{J}_{k(s,r),\ell} = \mathbf{u}_s^t \left[\frac{\partial \mathbf{B}^t}{\partial m_\ell} \right] \mathbf{B}^{-1} \delta_r,$$

(26)

where $k(s,r)$ denotes a source-receiver couple of the acquisition geometry, with s and r denoting a shot and a receiver position, respectively. An impulse source δ_r is located at receiver position r. If the sensitivity matrix must be built, one forward problem for the incident wavefield and one forward problem per receiver position must be computed. Therefore, the number of simulations to build the sensitivity matrix can be higher than that required by gradient estimation if the number of nonredundant receiver positions significantly exceeds the number of nonredundant shots, or vice versa. Computing each term of the sensitivity matrix is also required to compute the diagonal terms of the approximate Hessian $\mathbf{H_a}$ (Shin et al., 2001b). To mitigate the resulting computational burden for coarse OBS surveys, Operto et al. (2006) suggest computing the diagonal terms of $\mathbf{H_a}$ for a decimated shot acquisition. Alternatively, Shin et al. (2001a) propose using an approximation of the diagonal Hessian, which can be computed at the same cost as the gradient.

Although the matrix-free adjoint approach is widely used in exploration seismology, the earthquake-seismology community tends to favor the scattering-integral method, which is based on the explicit building of the sensitivity matrix (Chen et al., 2007). The linear system relating the model perturbation to the data perturbation is formed and solved with a conjugate-gradient algorithm such as LSQR (Paige and Saunders, 1982a). A comparative complexity analysis of the adjoint approach and the scattering-integral approach is presented in Chen et al. (2007), who conclude that the scattering-integral approach outperforms the adjoint approach for a regional to-

mographic problem. Indeed, the superiority of one approach over the other is highly dependent on the acquisition geometry (the relative number of sources and receivers) and the number of model parameters.

The formalism in equation 25 has been kept as general as possible and can relate to the acoustic or the elastic wave equation. In the acoustic case, the wavefield is the pressure scalar wavefield; in the elastic case, the wavefield ideally is formed by the components of the particle velocity and the pressure if the sensors have four components. Equation 25 can be translated in the time domain using Parseval's relation. The expression of the gradient in equation 25 can be developed equivalently using a functional analysis (Tarantola, 1984). The partial derivatives of the wavefield with respect to the model parameters are provided by the kernel of the Born integral that relates the model perturbations to the wavefield perturbations. Multiplying the transpose of the resulting operator by the conjugate of the data residuals provides the expression of the gradient. The two formalisms (matrix and functional) give the same expression, provided the discretization of the partial differential operators are performed consistently in the two approaches. The derivation in the frequency domain of the gradient of the misfit function using the two formalisms is explicitly illustrated by Gelis et al. (2007).

Source estimation

Source excitation is generally unknown and must be considered as an unknown of the problem (Pratt, 1999). The source wavelet can be estimated by solving a linear inverse problem because the relationship between the seismic wavefield and the source is linear (equation 2). The solution for the source is given by the expression

$$\mathbf{s} = \frac{\langle \mathbf{g}_{cal}(\mathbf{m}_0) | \mathbf{d}_{obs} \rangle}{\langle \mathbf{g}_{cal}(\mathbf{m}_0) | \mathbf{g}_{cal}(\mathbf{m}_0) \rangle},$$

(27)

where $\mathbf{g}_{cal}(\mathbf{m}_0)$ denotes the Green's functions computed in the starting model $\mathbf{m}_0$. The source function can be estimated directly in the FWI algorithm once the incident wavefields have been modeled. The source and the medium are updated alternatively over iterations of the FWI. Note that it is possible to take advantage of source estimation to design alternative misfit functions based on the differential semblance optimization (Pratt and Symes, 2002) or to define more heuristic criteria to stop the iteration of the inversion (Jaiswal et al., 2009).

Alternatively, new misfit functions have been designed so the inversion becomes independent of the source function (Lee and Kim, 2003; Zhou and Greenhalgh, 2003). The governing idea of the method is to normalize each seismogram of a shot gather by the sum of all of the seismograms. This removes the dependency of the normalized data with respect to the source and modifies the misfit function. The drawback is that this approach requires an explicit estimate of the sensitivity matrix; the normalized residuals cannot be back propagated because they do not satisfy the wave equation.

SOME KEY FEATURES OF FWI

Resolution power of FWI and relationship to the experimental setup

The interpretation of the partial-derivative wavefield as the wavefield scattered by the missing heterogeneities provides some connections between FWI and generalized diffraction tomography (Devaney and Zhang, 1991; Gelius et al., 1991). Diffraction tomography

recasts the imaging as an inverse Fourier transform (Devaney, 1982; Wu and Toksoz, 1987; Sirgue and Pratt, 2004; Lecomte et al., 2005). Let us consider a homogeneous background model of velocity c_0, an incident monochromatic plane wave propagating in the direction $\hat{\mathbf{s}}$ and a scattered monochromatic plane wave in the far-field approximation propagating in the direction $\hat{\mathbf{r}}$ (Figure 3). If amplitude effects are not taken into account, the incident and scattered Green's functions can be written compactly as

$$G_0(\mathbf{x},\mathbf{s}) = \exp(ik_0\hat{\mathbf{s}}.\mathbf{x}),$$
$$G_0(\mathbf{x},\mathbf{r}) = \exp(ik_0\hat{\mathbf{r}}.\mathbf{x}), \qquad (28)$$

with the relation $k_0 = \omega/c_0$. Inserting the expression of the incident and scattered plane waves into the gradient of the misfit function of equation 24 gives the expression (Sirgue and Pratt, 2004)

$$\nabla C(\mathbf{m}) = -\omega^2 \sum_\omega \sum_s \sum_r \mathcal{R}\{\exp(-ik_0(\hat{\mathbf{s}}+\hat{\mathbf{r}}).\mathbf{x})\mathbf{\Delta d}\}.$$
$$(29)$$

Equation 29 has the form of a truncated Fourier series where the integration variable is the scattering wavenumber vector given by $\mathbf{k} = k_0(\hat{\mathbf{s}}+\hat{\mathbf{r}})$. The coefficients of the series are the data residuals. The summation is performed over sources, receivers, and frequencies that control the truncation and sampling of the Fourier series.

We can express the scattering wavenumber vector $k_0(\hat{\mathbf{s}}+\hat{\mathbf{r}})$ as a function of frequency, diffraction angle, or aperture to highlight the relationship between the experimental setup and the spatial resolution of the reconstruction (Figure 4):

$$\mathbf{k} = \frac{2f}{c_0}\cos\left(\frac{\theta}{2}\right)\mathbf{n}, \qquad (30)$$

where $\mathbf{n}$ is a unit vector in the direction of the slowness vector ($\hat{\mathbf{s}} + \hat{\mathbf{r}}$). Equation 30 was also derived in the framework of the ray + Born migration/inversion, recast as the inverse of a generalized Radon transform (Miller et al., 1987) or as a least-squares inverse problem (Lambaré et al., 2003).

Several key conclusions can be derived from equation 30. First, one frequency and one aperture in the data space map one wavenumber in the model space. Therefore, frequency and aperture have redundant control of the wavenumber coverage. This redundancy increases with aperture bandwidth. Pratt and Worthington (1990), Sirgue and Pratt (2004), and Brenders and Pratt (2007a) propose decimating this wavenumber-coverage redundancy in frequency-domain FWI by limiting the inversion to a few discrete frequencies. This data reduction leads to computationally efficient frequency-do-

main FWI and allows managing a compact volume of data, two clear advantages with respect to time-domain FWI. The guideline for selecting the frequencies to be used in the FWI is that the maximum wavenumber imaged by a frequency matches the minimum vertical wavenumber imaged by the next frequency (Sirgue and Pratt, 2004, their Figure 3). According to this guideline, the frequency interval increases with the frequency.

Second, the low frequencies of the data and the wide apertures help resolve the intermediate and large wavelengths of the medium. At the other end of the spectrum, the maximum wavenumber, constrained by $\theta = 0$ and the highest frequency, leads to a maximum resolution of half a wavelength if normal-incidence reflections are recorded. Third, for surface acquisitions, long offsets are helpful for sampling the small horizontal wavenumbers of dipping structures such as flanks of salt domes.

A frequency-domain sensitivity kernel for point sources, referred to as the wavepath by Woodward (1992), is shown in Figure 5. The interference picture shows zones of equiphase over which the residuals are back projected during FWI. The central zone of elliptical shape is the first Fresnel zone of width $\sqrt{\lambda o_{sr}}$, where o_{sr} is the source-receiver offset. Residuals that match the first arrival with an error lower than half a period will be back projected constructively over the first Fresnel zone, updating the large wavelengths of the structure. The outer fringes are isochrones on which residuals associated with later-arriving reflection phases will be back projected, providing an update of the shorter wavelengths of the medium, just like PSDM (Lecomte, 2008). The width of the isochrones, which gives some insight into the vertical resolution in Figure 4, is given by the modulus of the wavenumber of equation 30.

To illustrate the relationship between FWI resolution and the experimental setup, we show the FWI reconstruction of an inclusion in a homogeneous background for three acquisition geometries (Figure 6). In the crosshole experiment (Figure 6a), FWI has reconstructed a low-pass-filtered (smoothed) version of the vertical section of the inclusion and a band-pass-filtered version of the horizontal section of the inclusion. This anisotropy of the imaging results from the transmission-like reconstruction of the vertical wavenumbers and the reflection-like reconstruction of the horizontal wavenumbers of the inclusion. In the case of the double crosshole experiment (Figure 6b), the vertical and horizontal wavenumber spectra of the inclusion have

Figure 4. Illustration of the main parameters in diffraction tomography and their relationships. Key: λ, wavelength; θ, diffraction or aperture angle; c, P-wave velocity; f, frequency; $\mathbf{p_s}$, $\mathbf{p_R}$, $\mathbf{q}$, slowness vectors; $\mathbf{k}$, wavenumber vector; $\mathbf{x}$, diffractor point; S and R, source and receiver positions.
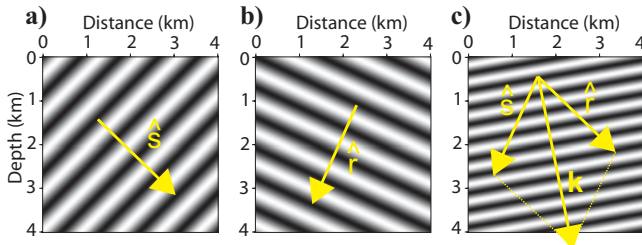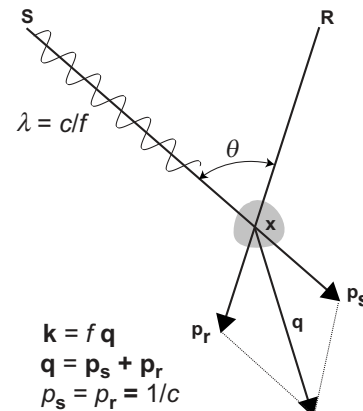
Figure 3. Resolution analysis of FWI. (a) Incident monochromatic plane wave (real part). (b) Scattered monochromatic plane wave (real part). (c) Gradient of FWI describing one wavenumber component (real part) built from the plane waves shown in (a) and (b).

been partly filled because of the combined use of transmission and reflection wavepaths. Of note, the vertical section shows a lack of low wavenumbers, whereas the horizontal section exhibits a deficit of low wavenumbers because the maximum horizontal source-receiver offset is two times higher than the vertical one (Figure 6b). Therefore, the aperture illuminations of the horizontal and vertical wavenumbers differ. For the surface acquisition (Figure 6c), the vertical section exhibits a strong deficit of low wavenumbers because of the lack of large-aperture illumination. Of note, the pick-to-pick amplitude of the perturbation is fully recovered in Figure 6c. The horizontal section of the inclusion is poorly recovered because of the poor illumination of the horizontal wavenumbers from the surface.

The ability of the wide apertures to resolve the large wavelengths of the medium has prompted some studies to consider long-offset acquisitions as a promising approach to design well-posed FWI problems (Pratt et al., 1996; Ravaut et al., 2004). For example, equation 30 can suggest that the long wavelengths of the medium can be resolved whatever the source bandwidth, provided that wide-aperture data are recorded by the acquisition geometry. However, all of the conclusions derived so far rely on the Born approximation. The Born approximation requires that the starting model allows matching the observed traveltimes with an error less than half the period (Beydoun and Tarantola, 1988). If not, the so-called cycle-skipping artifacts will lead to convergence toward a local minimum (Figure 7).

Pratt et al. (2008) translates this condition in terms of relative time error $\Delta t / T_L$ as a function of the number of propagated wavelengths N_λ, expressed as

$$\frac{\Delta t}{T_L} < \frac{1}{N_\lambda}, \tag{31}$$

where T_L denotes the duration of the simulation. Condition 31 shows that the traveltime error must be less than 1% for an offset involving 50 propagated wavelengths, a condition unlikely to be satisfied if FWI is applied without data preconditioning. Therefore, some studies consider that recording low frequencies (<1 Hz) is the best strategy to design well-posed FWI (Sirgue, 2006). Unfortunately, such

low frequencies cannot be recorded during controlled-source experiments. As an alternative to low frequencies, multiscale layer-stripping approaches where longer offsets, shorter apertures, and longer recording times are progressively introduced in the inversion, have been designed to mitigate the nonlinearity of the inversion.

Multiscale FWI: Time domain versus frequency domain

FWI can be implemented in the time domain or in the frequency domain. FWI was originally developed in the time domain (Taran-
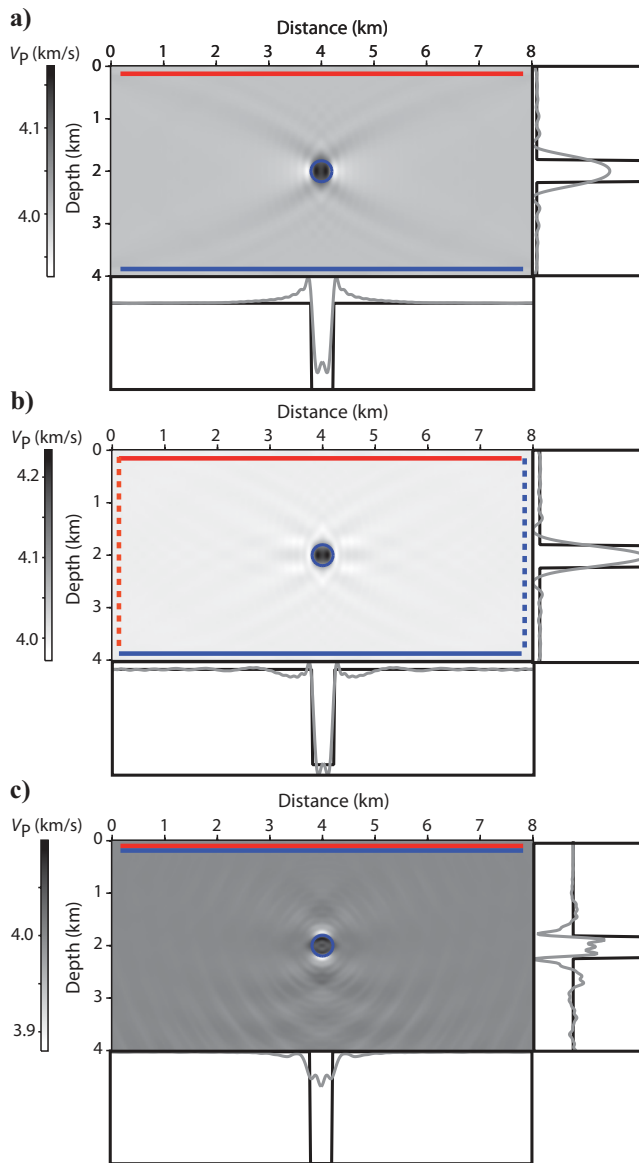
Figure 6. Imaging an inclusion by FWI. (a) Crosshole experiment. Source and receiver lines are in red and blue, respectively. The contour of the inclusion with a diameter of 400 m is delineated by the blue circle. The true velocity in the inclusion is 4.2 km/s, whereas the velocity in the background is 4 km/s. Six frequencies (4, 7, 9, 11, 12, and 15 Hz) were inverted successively, and 20 iterations per frequency were computed. The black and gray curves along the right and bottom sides of the model are velocity profiles across the center of the inclusion extracted from the exact model and the reconstructed model, respectively. (b) Same as (a) for a vertical and horizontal crosshole experiment (the shots along the red dashed line are recorded by only the receivers along the vertical blue dashed line). (c) Same as (a) for a surface experiment.
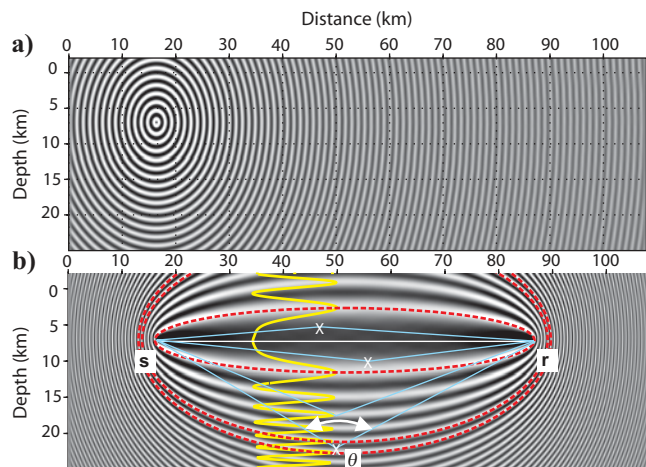

Figure 5. Wavepath. (a) Monochromatic Green's function for a point source. (b) Wavepath for a receiver located at a horizontal distance of 70 km from the source. The frequency is 5 Hz and the velocity in the homogeneous background is 6 km/s. The dashed red lines delineate the first Fresnel zone and an isochrone surface. The yellow line is a vertical section across the wavepath. The blue lines represent diffraction paths within the first Fresnel zone and from the isochrone.

tola, 1984; Gauthier et al., 1986; Mora, 1987; Crase et al., 1990) whereas the frequency-domain approach was proposed mainly in the 1990s by Pratt and collaborators (Pratt, 1990; Pratt and Worthington, 1990; Pratt and Goulty, 1991), first with application to crosshole data and later with application to wide-aperture surface seismic data (Pratt et al., 1996).

The nonlinearity of FWI has prompted many studies to develop some hierarchical multiscale strategies to mitigate this nonlinearity. Apart from computational efficiency, the flexibility offered by the time domain or the frequency domain to implement efficient multiscale strategies is one of the main criteria that favors one domain rather than the other. The multiscale strategy successively processes data subsets of increasing resolution power to incorporate smaller wavenumbers in the tomographic models. In the time domain, Bunks et al. (1995) propose successive inversion of subdata sets of increasing high-frequency content because low frequencies are less sensitive to cycle-skipping artifacts. The frequency domain provides a more natural framework for this multiscale approach by performing successive inversions of increasing frequencies. In the frequency domain, single or multiple frequencies (i.e., frequency group) can be inverted at a time.

Although a few discrete frequencies theoretically are sufficient to fill the wavenumber spectrum for wide-aperture acquisitions, simultaneous inversion of multiple frequencies improves the signal-to-noise ratio and the robustness of FWI when complex wave phenomena are observed (i.e., guide waves, surface waves, dispersive waves). Therefore, a trade-off between computational efficiency and quality of imaging must be found. When simultaneous multifrequency inversion is performed, the bandwidth of the frequency group ideally must be as large as possible to mitigate the nonlinearity of FWI in terms of the nonunicity of the solution, whereas the maximum frequency of the group must be sufficiently low to prevent cycle-skipping artifacts. An illustration of this tuning of FWI is given in Brossier et al. (2009a) in the framework of elastic seismic imaging of complex onshore models from the joint inversion of surface waves and body waves.
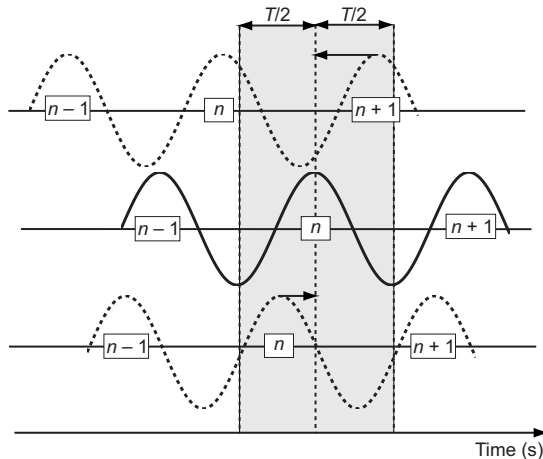

Figure 7. Schematic of cycle-skipping artifacts in FWI. The solid black line represents a monochromatic seismogram of period T as a function of time. The upper dashed line represents the modeled monochromatic seismograms with a time delay greater than $T/2$. In this case, FWI will update the model such that the $n+1$th cycle of the modeled seismograms will match the nth cycle of the observed seismogram, leading to an erroneous model. In the bottom example, FWI will update the model such that the modeled and recorded nth cycle are in-phase because the time delay is less than $T/2$.

The regularization effects introduced by hierarchical inversion of increasing frequencies might not be sufficient to provide reliable FWI results for realistic frequencies and realistic starting models in the case of complex structures. This has prompted some studies to design additional regularization levels in FWI. One of these is to select a subset of arrivals as a function of time. An aim of this time windowing is to remove arrivals that are not predicted by the physics of the wave equation implemented in FWI (for example, PS-converted waves in the frame of acoustic FWI). A second aim is to perform a heuristic selection of aperture angles in the data. Considering a narrow time window centered on the first arrival leads to so-called early-arrival waveform tomography (Sheng et al., 2006). Time windowing the data around the first arrivals is equivalent to selecting the large-aperture components of the data to image the large and intermediate wavelengths of the medium. Alternatively, time windowing can be applied to isolate later-arriving reflections or PS-converted phases to focus on imaging a specific reflector or a specific parameter class, such as the S-wave velocity (Shipp and Singh, 2002; Sears et al., 2008; Brossier et al., 2009a).

The frequency domain is the most appropriate to select one or a few frequencies for FWI, but the time domain is the most appropriate to select one type of arrival for FWI. Indeed, time windowing cannot be applied in frequency-domain modeling, in which only one or few frequencies are modeled at a time. A last resort is the use of complex-valued frequencies, which is equivalent to the exponential damping of a signal $p(t)$ in time from an arbitrary traveltime t_0 (Sirgue, 2003; Brenders and Pratt, 2007b):

$$P\left(\omega + \frac{i}{\tau}\right)e^{t_0/\tau} = \int_{-\infty}^{+\infty} p(t)e^{-(t-t_0)/\tau}e^{i\omega t}dt, \qquad (32)$$

where $P(\omega)$ denotes the Fourier transform of $p(t)$ and τ is the damping factor.

A last regularization level can be implemented by layer stripping, in which the imaging proceeds hierarchically from the shallow part to the deep part. Layer stripping in FWI can be applied by combined offset and temporal windowing (Shipp and Singh, 2002; Wang and Rao, 2009).

These three levels of regularization — frequency dependent, time dependent, and offset dependent — can be combined in one integrated multiloop FWI workflow. An example is provided in Shin and Ha (2009) and Brossier et al. (2009a), in which the frequency- and time-dependent regularizations are implemented into two nested loops over frequency groups and time-damping factors. In this approach, the frequencies increase in the outer loop and the damping factors decrease in the inner loop. In Figure 8, the V_P and V_S models of the overthrust model are inferred from the successive inversion of two groups of five frequencies (Brossier et al., 2009a). The frequencies of the first group range from 1.7 to 3.5 Hz, whereas those of the second group range from 3.5 to 7.2 Hz. Five damping factors of τ between 0.67 and 30.0 s were applied hierarchically for data preconditioning during the inversion of each frequency group. Without these two regularization levels associated with frequency and aperture selections, FWI fails to converge toward acceptable models.

In summary, the implementation of FWI in the frequency domain allows the easy implementation of multiscale FWI based on the hierarchical inversion of groups of frequencies of arbitrary bandwidth and sampling intervals. Time-domain modeling provides the most flexible framework to apply time windowing of arbitrary geometry.

This makes frequency-domain FWI based on time-domain modeling an attractive strategy to design robust FWI algorithms. This is especially true for 3D problems, for which time-domain modeling has several advantages with respect to frequency-domain modeling (Sirgue et al., 2008).

On the parallel implementation of FWI

FWI algorithms must be implemented in parallel to address large-scale 3D problems. Depending on the numerical technique for solving the forward problem, different parallel strategies can be considered for FWI. If the forward problem is based on numerical methods such as time-domain modeling or iterative solvers, which are not demanding in terms of memory, a coarse-grained parallelism that consists of distributing sources over processors is generally used and the forward problem is performed sequentially on each processor for each source (Plessix, 2009). If the number of processors significantly exceeds the number of shots, which can be the case if source encoding techniques are used (Krebs et al., 2009), a second level of parallelism can be viewed by domain decomposition of the physical computational domain. A comprehensive review of different algorithms to efficiently compute the forward and the adjoint wavefields in time-domain FWI is presented by Akcelik (2002).

In contract, if the forward problem is based on a method that embeds a memory-expensive preprocessing step, such as LU factorization in the frequency-domain direct-solver approach, parallelism must be based on a domain decomposition of the computational domain. Each processor computes a subdomain of the wavefields for all sources. Examples of such algorithms are described in Ben Hadj Ali et al. (2008a), Brossier et al. (2009a), and Sourbier et al. (2009a, 2009b). A contrast source inversion (CSI) method is described by Abubakar et al. (2009), which allows a decrease in the number of LU factorization in frequency-domain FWI at the expense of the number of iterations.

Variants of classic least-squares and Born-approximation FWI

Although the most popular approach of FWI is based on minimizing the least-squares norm of the data misfit on the one hand and on the Born approximation for estimating partial-derivative wavefields on the other, several variants of FWI have been proposed over the last decade. These variants relate to the definition of the minimization criteria, the representation of the data (amplitude, phase, logarithm of the complex-valued data, envelope) in the misfit, or the linearization procedure of the inverse problem.

The choice of the minimization criterion

The least-squares norm approach assumes a Gaussian distribution of the misfit (Tarantola, 1987). Poor results can be obtained when this assumption is violated, for example, when large-amplitude outliers affect the data. Therefore, careful quality control of the data must be carried out before least-squares inversion. Crase et al. (1990) investigate several norms such as the least-squares norm $\mathcal{L}_2$, the least-absolute-values norm $\mathcal{L}_1$, the Cauchy criterion norm, and the hyperbolic secant (sech) criterion in FWI (Figure 9). The $\mathcal{L}_1$-norm specifically ignores the amplitude of the residuals during back propagation of the residuals when gradient building, making this criterion less sensitive to large errors in the data. The Cauchy and sech criteria can be considered a combination of the $\mathcal{L}_1$- and the

$\mathcal{L}_2$-norms with different transitions between the norms. Crase et al. (1990) conclude that the most reliable FWI results have been obtained with the Cauchy and the sech criteria.

The $\mathcal{L}_2$ and Cauchy criteria are also compared by Amundsen (1991) in the framework of frequency-wavenumber-domain FWI for stratified media described by velocity, density, and layer thicknesses (Amundsen and Ursin, 1991). They consider random noise and weather noise and conclude in both cases that the Cauchy criterion leads to the more robust results.

The Huber norm also combines the $\mathcal{L}_1$- and the $\mathcal{L}_2$-norms; it is combined with quasi-Newton L-BFGS by Guitton and Symes (2003) and Bube and Nemeth (2007). The Huber norm is also used in the framework of frequency-domain FWI by Ha et al. (2009) and shows an overall more robust behavior than the $\mathcal{L}_2$-norm.

The choice of the linearization method

The sensitivity matrix is generally computed with the Born approximation, which assumes a linear tangent relationship between the model and wavefield perturbations (Woodward, 1992). This linear relationship between the perturbations can be inferred from the assumption that the wavefield computed in the updated model is the wavefield computed in the starting model plus the perturbation wavefield.
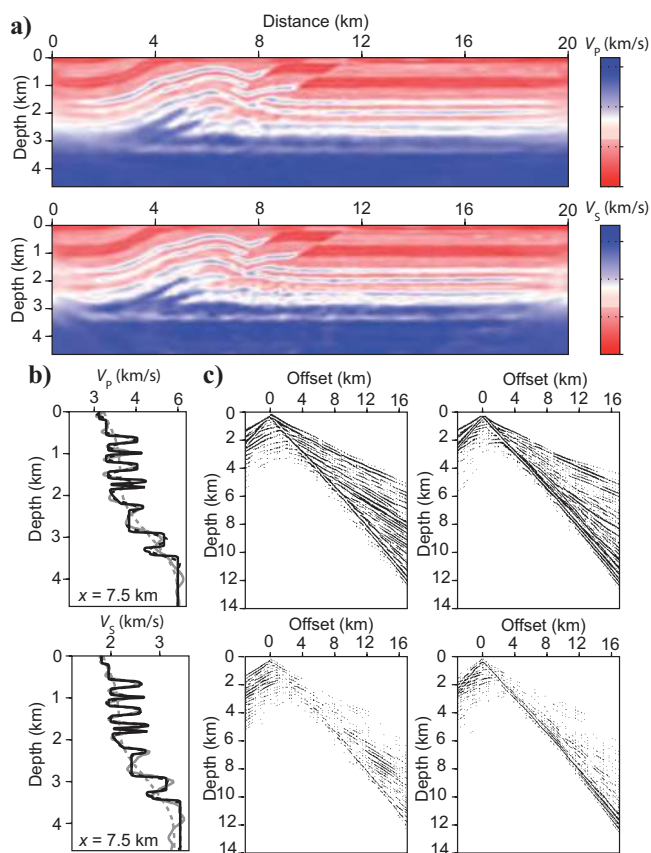

Figure 8. Multiscale strategy for elastic FWI with application to the overthrust model. (a) FWI velocity models V_P (top) and V_S (bottom). (b) Comparison between the logs from the true model (black), the starting model (dashed gray), and the final FWI model (solid gray). (c) Synthetic seismograms computed in the final FWI models for the horizontal (left) and vertical (right) components of particle velocity. The bottom panels are the final residuals between seismograms computed in the true and in the final FWI models (image courtesy R. Brossier).

The Rytov approach considers the generalized phase as the wavefield (Woodward, 1992). The Rytov approximation provides a linear relationship between the complex-phase perturbation and the model perturbation by assuming that the wavefield computed in the updated model is related to the wavefield computed in the starting model through $\mathbf{u}(\mathbf{x},\omega) = \mathbf{u}_0(\mathbf{x},\omega)\exp(\Delta\phi(\mathbf{x},\omega))$, where $\Delta\phi(\mathbf{x},\omega)$ denotes the complex-phase perturbation. The sensitivity of the Rytov kernel is zero on the Fermat raypath because the traveltime is stationary along this path. A linear relationship between the model perturbations and the logarithm of the amplitude ratio $\mathrm{Ln}[A(\omega)/A_0(\omega)]$ is also provided by the Rytov approximation by taking the real part of the sensitivity kernel of the Rytov integral instead of the imaginary part that provides the phase perturbation.

The Born approximation is valid in the case of weak and small perturbations, but the Rytov approximation is supposed to be valid for large-aperture angles and a small amount of scattering per wavelength, i.e., smooth perturbations or smooth variation in the phase-perturbation gradient (Beydoun and Tarantola, 1988). Although several analyses of the Rytov approximation have been carried out, it remains unclear to what extent its domain of validity differs significantly from that of the Born approximation. A comparison between the Born approximation and the Rytov approximation in the framework of elastic frequency-domain FWI is presented in Gelis et al. (2007). The main advantage of the Rytov approximation might be to provide a natural separation between phase and amplitude (e.g., Woodward, 1992). This separation allows the implementation of phase and amplitude inversions (Bednar et al., 2007; Pyun et al.,

Figure 9. Different functionals for FWI. (a) Value of four functionals as a function of real-valued data residual. The $\mathcal{L}_1$, $\mathcal{L}_2$, Huber, and mixed $\mathcal{L}_1$–$\mathcal{L}_2$ functionals are plotted as indicated. (b) Amplitude of the residual source used to compute the back-propagated adjoint wavefield for the different functionals shown in (a). Note that the back-propagated source in the $\mathcal{L}_1$-norm is not sensitive to the residual amplitude and therefore is less sensitive to large-amplitude residuals than the $\mathcal{L}_2$-norm (image courtesy R. Brossier).

2007) from a frequency-domain FWI code using a logarithmic norm (Shin and Min, 2006; Shin et al., 2007), the use of which leads to the Rytov approximation in the framework of local optimization problems.

Another approach in the time-frequency domain is developed by Fichtner et al. (2008) for continental- and global-scale FWI, in which the misfit of the phase and the misfit of the envelopes are minimized in a least-squares sense. The expected benefit from this approach is to mitigate the nonlinearity of FWI by separating the phase and amplitude in the inversion and by inverting the envelope instead of the amplitudes, the former being more linearly related to the data.

Building starting models for FWI

The ultimate goal in seismic imaging is to be able to apply FWI reliably from scratch, i.e., without the need for sophisticated a priori information. Unfortunately, because multidimensional FWI at present can only be attacked through local optimization approaches, building an accurate starting model for FWI remains one of the most topical issues because very low frequencies (<1 Hz) still cannot be recorded in the framework of controlled-source experiments.

A starting model for FWI can be built by reflection tomography and migration-based velocity analysis such as those used in oil and gas exploration. A review of the tomographic workflow is given in Woodward et al. (2008). Other possible approaches for building accurate starting models, which should tend toward a more automatic procedure and might be more closely related to FWI, are first-arrival traveltime tomography (FATT), stereotomography, and Laplace-domain inversion.

FATT performs nonlinear inversions of first-arrival traveltimes to produce smooth models of the subsurface (e.g., Nolet, 1987; Hole, 1992; Zelt and Barton, 1998). Traveltime residuals are back projected along the raypaths to compute the sensitivity matrix. The tomographic system, augmented with smoothing regularization, generally is solved with a conjugate-gradient algorithm such as LSQR (Paige and Saunders, 1982b). Alternatively, the adjoint-state method can be applied to FATT, which avoids the explicit building of the sensitivity matrix, just as in FWI (Taillandier et al., 2009). The spatial resolution of FATT is estimated to be the width of the first Fresnel zone (Williamson, 1991; Figure 5).

Examples of applications of FWI to real data using a starting model built by FATT are shown, for example, in Ravaut et al. (2004), Operto et al. (2006), Jaiswal et al. (2008, 2009), and Malinowsky and Operto (2008) for surface acquisitions; in Dessa and Pascal (2003) in the framework of ultrasonic experimental data; in Pratt and Goulty (1991) for crosshole data; and in Gao et al. (2006b) for VSP data.

Several blind tests that correspond to surface acquisitions have been tackled by joint FATT and FWI. Results at the oil-exploration scale and at the lithospheric scale are presented in Brenders and Pratt (2007a, 2007b, 2007c) and suggest that very low frequencies and very large offsets are required to obtain reliable FWI results when the starting model is built by FATT. For example, only the upper part of the BP benchmark model was imaged successfully by Brenders and Pratt (2007c) using a starting frequency as small as 0.5 Hz and a maximum offset of 16 km. Another drawback of FATT is that the method is not suitable when low-velocity zones exist because these low-velocity zones create shadow zones.

Reliable picking of first-arrival times is also a difficult issue when low-velocity zones exist. Fitting first-arrival traveltimes does not guarantee that computed traveltimes of later-arriving phases, such as
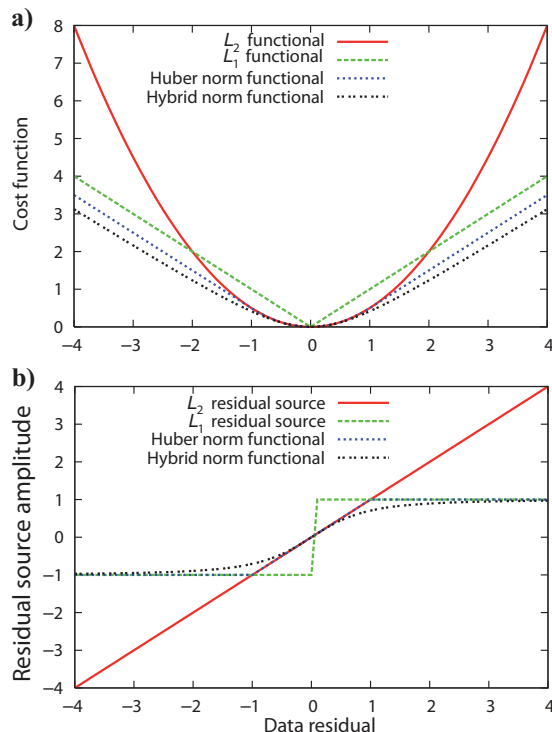
reflections, will match the true reflection traveltimes with an error that does not exceed half a period, especially if anisotropy affects the wavefield. We should stress that FATT can be recast as a phase inversion of the first arrival using a frequency-domain waveform-inversion algorithm within which complex-valued frequencies are implemented (Min and Shin, 2006; Ellefsen, 2009). Compared to FATT based on the high-frequency approximation, this approach helps account for the finite-frequency effect of the data in the sensitivity kernel of the tomography. Judicious selection of the real and imaginary parts of the frequency allows extraction of the phase of the first arrival. The principles and some applications of the method are presented in Min and Shin (2006) and Ellefsen (2009) for near-surface applications. This is strongly related to finite-frequency tomography (Montelli et al., 2004).

Traveltime tomography methods that can manage refraction and reflection traveltimes should provide more consistent starting models for FWI. Among these methods, stereotomography is probably one of the most promising approaches because it exploits the slope of locally coherent events and a reliable semiautomatic picking procedure has been developed (Lambaré, 2008). Applications of stereotomography to synthetic and real data sets are presented in Billette and Lambaré (1998), Alerini et al. (2002), Billette et al. (2003), Lambaré and Alérini (2005), and Dummong et al. (2008).

To illustrate the sensitivity of FWI to the accuracy of different starting models, Figure 10 shows FWI reconstructions of the synthetic Valhall model when the starting model is built by FATT and reflection stereotomography (Prieux et al., 2009). In the case of stereotomography, the maximum offset is 9 km and only the reflection traveltimes are used (Lambaré and Alérini, 2005), whereas the maximum offset is 32 km for FATT (Prieux et al., 2009). Stereotomography successfully reconstructs the large wavelength within the gas cloud down to a maximum depth of 2.5 km; FATT fails to reconstruct the large wavelengths of the low-velocity zone associated with the gas cloud. However, the FWI model inferred from the FATT starting model shows an accurate reconstruction of the shallow part of the model. These results suggest that joint inversion of refraction and reflection traveltimes by stereotomography can provide a promising framework to build starting models for FWI.

A third approach to building a starting model for FWI can be provided by Laplace-domain and Laplace-Fourier-domain inversions (Shin and Cha, 2008, 2009; Shin and Ha, 2008). The Laplace-domain inversion can be viewed as a frequency-domain waveform inversion using complex-valued frequencies (see equation 32), the real part of which is zero and the imaginary part of which controls the time damping of the seismic wavefield. In other words, the principle is the inversion of the DC component of damped seismograms where the Laplace variable s corresponds to $1/\tau$ in equation 32. The DC of the undamped data is zero, but the DC of the damped data is not and is exploited in Laplace-domain waveform inversion. The information contained in the data can be similar to the amplitude of the wavefield (Shin and Cha, 2009). Laplace-domain waveform inversion provides a smooth reconstruction of the velocity model, which can be used as a starting model for Laplace-Fourier and classical frequency-domain waveform inversions.

The Laplace-Fourier domain is equivalent to performing an inversion of seismograms damped in time. The results shown in Shin and Cha (2009) suggest that this method can be applied to frequencies smaller than the minimum frequency of the source bandwidth. The ability of the method to use frequencies smaller than the frequencies effectively propagated by the seismic source is called a mirage resur-

rection of the low frequencies by Shin and Cha (2009). An application to real data from the Gulf of Mexico is shown in Shin and Cha (2009). For the real-data application, frequencies between 0 and 2 Hz in combination with nine Laplace damping constants are used for the Laplace-Fourier-domain inversion, the final model of which is used as the starting model for standard frequency-domain FWI.

Joint application of Laplace-domain, Laplace-Fourier-domain and Fourier-domain FWI to the BP benchmark model is illustrated in Figure 11 (Shin and Cha, 2009). The starting model is a simple velocity-gradient model (Figure 11b). A first velocity model of the large wavelengths is obtained by Laplace-domain inversion (Figure 11c), which is subsequently used as a starting model for inversion in the Laplace-Fourier-domain inversion, the final model of which is shown in Figure 11d. During this stage, the starting frequency used in the inversion of the damped data is as low as 0.01 Hz. The final model obtained after frequency-domain FWI is shown in Figure 11e. All of the structures were successfully imaged, beginning with a very crude starting model.
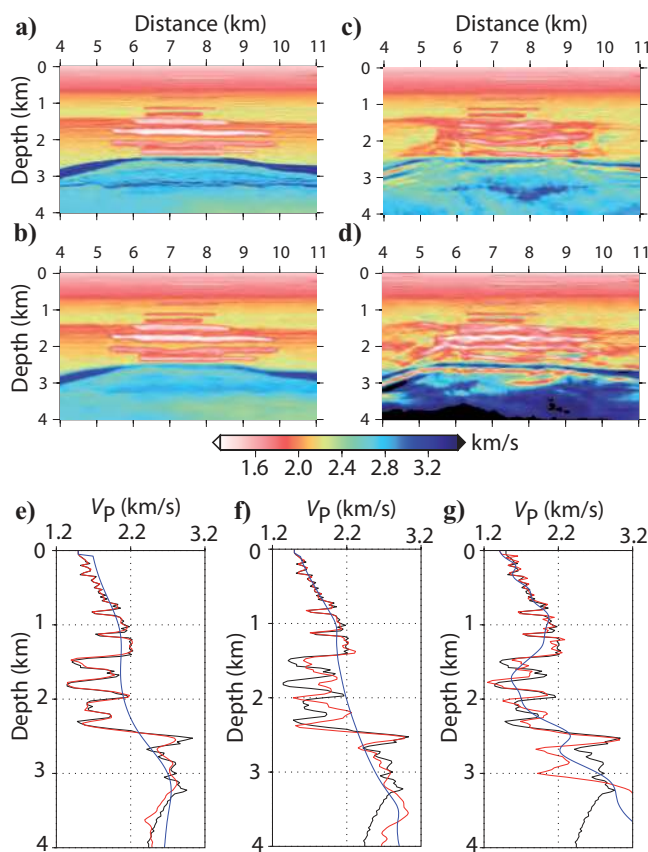

Figure 10. (a) Close-up of the synthetic Valhall velocity model centered on the gas layer. (b) FWI model built from a starting model obtained by smoothing the true model with a Gaussian filter with horizontal and vertical correlation lengths of 500 m. (c) FWI model from a starting model built by FATT (Prieux et al., 2009). (d) FWI model from a starting model built by stereotomography (Lambaré and Alérini, 2005). (e) Velocity profiles at a distance of 7.5 km extracted from the true model (black line), from the starting model built by smoothing the true model (blue line), and from the FWI model of (b) (red line). (f) Same as (e) for the starting model built by FATT and (c) the corresponding FWI model. (g) Same as (e) for the starting model built by stereotomography and (d) the corresponding FWI model. The frequencies used in the inversion are between 4 and 15 Hz.

CASE STUDIES

Applications of FWI have been applied essentially to synthetic examples for which high-resolution images have been constructed. Because FWI is an attractive approach, the number of real-data applications has increased quite rapidly, from monoparameter reconstructions of the V_P parameter to multiparameter ones.

Monoparameter acoustic FWI

Most of the recent real-data case studies of FWI at different scales and for different acquisition geometries have been performed in the
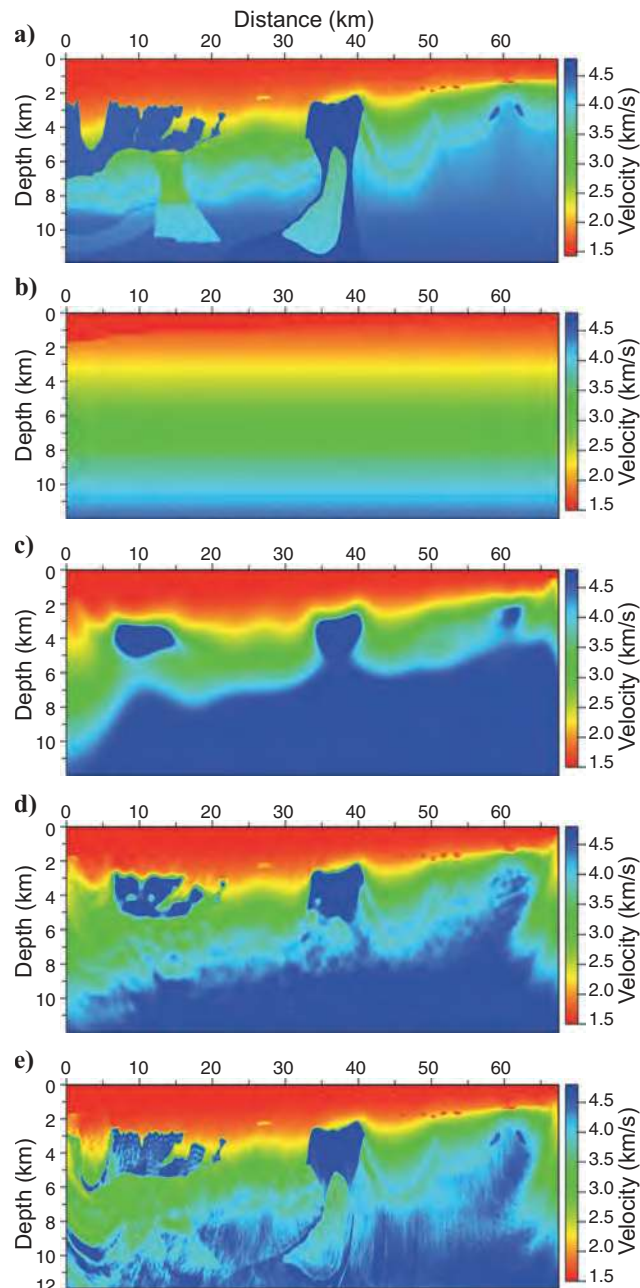

Figure 11. Laplace-Fourier-domain waveform inversion. (a) BP benchmark model. (b) Starting model. (c) Velocity model after Laplace-domain inversion. (d) Velocity model after Laplace-Fourier-domain inversion. (e) Velocity model after frequency-domain FWI (image courtesy C. Shin and Y. H. Cha).

acoustic isotropic approximation, considering only V_P as the model parameter (Dessa and Pascal, 2003; Ravaut et al., 2004; Chironi et al., 2006; Gao et al., 2006a, 2006b; Operto et al., 2006; Bleibinhaus et al., 2007; Ernst et al., 2007; Jaiswal et al., 2008, 2009; Shin and Cha, 2009). Although the acoustic approximation can be questioned in the framework of FWI because of unreliable amplitudes, one advantage of acoustic FWI is dealing with less computationally expensive forward modeling than in the elastic case. Moreover, acoustic FWI is better posed than elastic FWI because only the dominant parameter V_P can be involved in the inversion.

Specific waveform-inversion data processing generally is designed to account for the amplitude errors introduced by acoustic modeling (Pratt, 1999; Ravaut et al., 2004; Brenders and Pratt, 2007b). The amplitude discrepancies in the P-wavefield result from incorrectly modeling the amplitude-variation-with-offset (AVO) effects and incorrectly modeling the directivity of the source and receiver (Brenders and Pratt, 2007b). Acoustic-wave modeling generally is based on resolving the acoustic-wave equation in pressure; therefore, the particle-velocity synthetic wavefields might not be computed (Hustedt et al., 2004). If the receivers are geophones, the physical measurements collected in the field (particle velocities) are not the same as those computed by the seismic modeling engine (pressure). A match between the vertical geophone data and the pressure synthetics can, however, be performed by exploiting the reciprocity of the Green's functions if the sources are explosions (Operto et al., 2006).

In contrast, if the sources and receivers are directional, the pressure wavefield cannot account for the directivity of the sources and receivers, and heuristic amplitude corrections must be applied before inversion. Brenders and Pratt (2007b) propose optimizing an empirical correction law for the decay of the rms amplitudes with offset. Applying this correction law to the modeled data matches the main AVO trend of the observed data before FWI. Using this data preprocessing, Brenders and Pratt (2007b) successfully image the onshore lithospheric model of the CCSS blind test (Zelt et al., 2005) by acoustic FWI of synthetic elastic data. This strategy is also used successfully by Jaiswal et al. (2008, 2009) in the framework of acoustic FWI of real onshore data in the Naga thrust and fold belt in India.

Successful application of acoustic FWI to synthetic elastic data computed in the marine Valhall model from an OBC acquisition is presented by Brossier et al. (2009b). An application of acoustic FWI to real onshore long-offset data recorded in the southern Apennines thrust belt is illustrated in Figure 12 (Ravaut et al., 2004). The velocity model is validated locally by comparison with a VSP log. Application of PSDM using the final FWI model as a starting model contributes to the validation of the relevance of the velocity structure reconstructed by FWI (Operto et al., 2004). Some guidelines based on numerical examples of the domain of validity of acoustic FWI applied to elastic data are also provided in Barnes and Charara (2008).

Multiparameter FWI

Because FWI accounts for the full wavefield, the seismic modeling embedded in the FWI algorithm theoretically should honor as far as possible all of the physics of wave propagation. This is especially required by FWI of wide-aperture data, in which significant AVO and azimuthal anisotropic effects should be observed in the data. The requirement of realistic seismic modeling has prompted some studies to extend monoparameter acoustic FWI to account for parameter classes other than the P-wave velocity, such as density, attenuation,

shear-wave velocity or related parameters, and anisotropy. The fact that additional parameter classes are taken into account in FWI increases in the ill-posedness of the inverse problem because more degrees of freedom are considered in the parameterization and because the sensitivity of the inversion can change significantly from one parameter class to the next.

Different parameter classes can be more or less coupled as a function of the aperture angle. This coupling can be assessed by plotting the radiation pattern of each parameter class using some asymptotic analyses. Diffraction patterns of the different combination of parameters for the acoustic, elastic, and viscoelastic wave equation are shown in Wu and Aki (1985), Tarantola (1986), Ribodetti and Virieux (1996), and Forgues and Lambaré (1997). An alternative is to plot the sensitivity kernel, i.e., that obtained by summing all of the rows of the sensitivity matrix for the full acquisition and for different combinations of parameters and to qualitatively assess which combination provides the best image (Luo et al., 2009). Hierarchical strategies that successively operate on different parameter classes should be designed to mitigate the ill-posedness of FWI (Tarantola, 1986; Kamei and Pratt, 2008; Sears et al., 2008; Brossier et al., 2009b).

Density

Density is difficult to reconstruct (Forgues and Lambaré, 1997). As an illustration, acoustic radiation patterns are shown in Figure 13 for different combinations of parameters (I_P, ρ), (I_P, V_P), and (V_P, ρ), where I_P denotes P-wave impedance. The radiation pattern of V_P is isotropic because the operator $\partial \mathbf{B}/\partial m_l$ reduces to a scalar for V_P and therefore represents an explosion. On the other hand, the density has
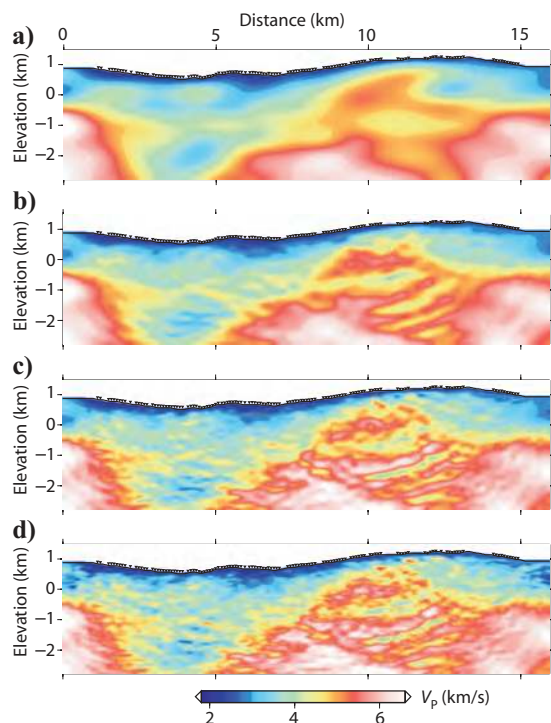

Figure 12. Two-dimensional seismic imaging of a thrust belt in the southern Apennines, Italy, from long-offset data by frequency-domain FWI. Thirteen frequencies from 6 to 20 Hz were inverted successively. (a) Starting model for FWI developed by FATT (Improta et al., 2002). (b–d) FWI model after inversion of (b) the starting 6-Hz, (c) 10-Hz, and (d) 20-Hz frequencies (Ravaut et al., 2004).
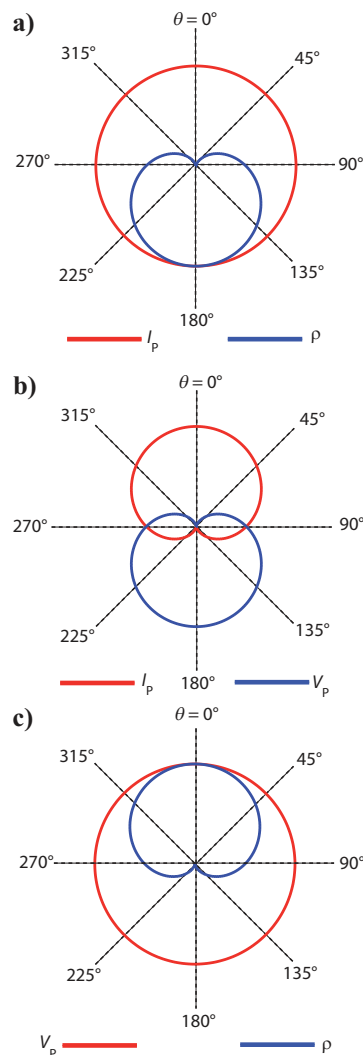
the same radiation pattern as V_P at short apertures but does not scatter energy at wide apertures because the secondary source $\mathbf{f}^\ell$ corresponds to a vertical force for the density. Because V_P and ρ have the same radiation pattern at short apertures, these two parameters are difficult to reconstruct from short-offset data. For such data, the P-wave impedance can be considered a reliable parameter for FWI. If wide-aperture data are available, V_P and I_P might provide the most judicious parameterization because they scatter energy for different aperture bands (wide and short apertures, respectively; Figure 13b).

A successful reconstruction of the density parameter in the case of the Marmousi case study is presented by Choi et al. (2008). However, the use of an unrealistically low frequency (0.125 Hz) brings into question the practical implication of these results.

Attenuation

The attenuation reconstruction can be implemented in frequency-domain seismic-wave modeling using complex velocities (Toksöz and Johnston, 1981). The most commonly applied attenuation/dispersion model is referred to as the Kolsky-Futterman model (Kolsky, 1956; Futterman, 1962). This model has linear frequency dependence of the attenuation coefficient, whereas the deviation from

Figure 13. Radiation pattern of different parameter classes in acoustic FWI. (a) $I_{P\text{-}\rho}$; (b) $I_P\text{-}V_P$; (c) $V_{P\text{-}\rho}$.

constant-phase velocity is accounted for through a term that varies as the logarithm of the frequency. This model is obtained by imposing causality on the wave pulse and assuming the absorption coefficient is strictly proportional to the frequency over a restricted range of frequencies.

Using the Kolsky-Futterman model, the complex velocity $\overline{c}$ is given by

$$\overline{c} = c\left[\left(1 + \frac{1}{\pi Q}|\log(\omega/\omega_r)|\right) + i\frac{\text{sgn}(\omega)}{2Q}\right]^{-1}, \quad (33)$$

where Q is the attenuation factor and ω_r is a reference frequency.

In frequency-domain FWI, the real and imaginary parts of the velocity can be processed as two independent real-model parameters without any particular implementation difficulty. However, the reconstruction of attenuation is an ill-posed problem. Ribodetti et al. (2000) show that in the framework of the ray + Born inversion, P-wave velocity and Q are uncoupled (i.e., the Hessian is nonsingular) only if a reflector is illuminated from both sides (with circular acquisition as in medical imaging; see Figure 5b of Ribodetti et al., 2000). In contrast, the Hessian becomes singular for a surface acquisition. Mulder and Hak (2009) also conclude that V_P and Q cannot be imaged simultaneously from short-aperture data because the Hilbert transform with respect to depth of the complex-valued perturbation model for V_P and Q produces the same wavefield perturbation in the framework of the Born approximation as the original perturbation model. Despite this theoretical limitation, preconditioning of the Hessian is investigated by Hak and Mulder (2008) to improve the convergence of the joint inversion for V_P and Q.

Assessment and application of viscoacoustic frequency-domain FWI is presented at various scales by Liao and McMechan (1995, 1996), Song et al. (1995), Hicks and Pratt (2001), Pratt et al. (2005), Malinowsky et al. (2007), and Smithyman et al. (2008). Kamei and Pratt (2008) recommend inversion for V_P only in a first step and then joint inversion for V_P and Q in a second step because the reliability of the attenuation reconstruction strongly depends on the accuracy of the starting V_P model. Indeed, accurate V_P models are required before reconstructing Q so that the inversion can discriminate the intrinsic attenuation from the extrinsic attenuation.

Elastic parameters

A limited number of applications of elastic FWI have been proposed. Because V_P is the dominant parameter in elastic FWI, Tarantola (1986) recommends inversion first for V_P and I_P, second for V_S and I_S, and finally for density. This strategy might be suitable if the footprint of the S-wave velocity structure on the seismic wavefield is sufficiently small. This hierarchical strategy over parameter classes is illustrated by Sears et al. (2008), who assess time-domain FWI of multicomponent OBC data with synthetic examples. They highlight how the behavior of FWI becomes ill-posed for S-wave velocity reconstruction when the S-wave velocity contrast at the sea bottom is small. In this case, the S-wave velocity structure has a minor footprint on the seismic wavefield because the amount of PS conversion is small at the sea bottom. In this configuration, they recommend inversion first for V_P, using only the vertical component; second for V_P and V_S from the vertical component; and finally for V_S, using both components. The aim of the second stage is to reconstruct the intermediate wavelengths of the S-wave velocity structure by exploiting the AVO behavior of the P-waves.

In contrast, Brossier et al. (2009a) conclude that joint inversion for V_P and V_S with judicious hierarchical data preconditioning by time damping is necessary for inversion of land data involving both body waves and surface waves. The strong sensitivity of the high-amplitude surface waves to the near-surface S-wave velocity structure requires inversion for V_S during the early stages of the inversion. This makes the elastic inversion of onshore data highly nonlinear when the surface waves are preserved in the data.

A recent application of elastic FWI to a gas field in China is presented by Shi et al. (2007). They invert for the Lamé parameters and unambiguously image Poisson's ratio anomalies associated with the presence of gas. They accelerate the convergence of the inversion by computing an efficient step length using an adaptive controller based on the theory of model-reference nonlinear control. Several logs available along the profile confirm the reliability of this gas-layer reconstruction.

Anisotropy

Reconstruction of anisotropic parameters by FWI is probably one of the most undeveloped and challenging fields of investigation. Vertically transverse isotropy (VTI) or tilted transversely isotropic (TTI) media are generally considered a realistic representation of geologic media in oil and gas exploration, although fractured media require an orthorhombic description (Tsvankin, 2001). The normal-moveout (NMO) P-wave velocity in VTI media depends on only two parameters: the NMO velocity for a horizontal reflector $V_{\text{NMO}}(0)$ $= V_{P_0}\sqrt{1 + 2\delta}$ and the $\eta = (\varepsilon - \delta)/(1 + 2\delta)$ parameter (Alkhalifah and Tsvankin, 1995), which is a combination of Thomsen's parameters ε and δ (Thomsen, 1986). The dependency of NMO velocity in VTI media on a limited subset of anisotropic parameters suggests that defining the parameter classes to be involved in FWI will be a key task. Another issue will be to assess to what extent FWI can be performed in the acoustic approximation knowing that acoustic media are by definition isotropic (Grechka et al., 2004). The kinematic and dynamic accuracy of an acoustic TTI wave equation for FWI is discussed in Operto et al. (2009).

A feasibility study of FWI in VTI media for crosswell acquisitions is presented in Barnes et al. (2008). They invert for five parameter classes — V_P, V_S, density, δ, and ε — and show reliable reconstruction of the classes, even with noisy data. Pratt et al. (2001, 2008) apply anisotropic FWI to crosshole real data. The results of Pratt et al. (2001) highlight the difficulty in discriminating layer-induced anisotropy from intrinsic anisotropy in FWI.

Further investigations of anisotropic FWI in the case of surface seismic data are required. In particular, the benefit of wide apertures in resolving as many anisotropic parameters as possible needs to be investigated (Jones et al., 1999).

Three-dimensional FWI

Because of the continuous increase in computational power and the evolution of acquisition systems toward wide-aperture and wide-azimuth acquisition, 3D acoustic FWI is feasible today. In three dimensions, the computational burden of multisource seismic modeling is one of the main issues. The pros and cons of time-domain versus frequency-domain seismic modeling for FWI have been discussed. Another issue is assessing the impact of azimuth coverage on FWI. Sirgue et al. (2007) show the footprint of the azimuth coverage in 3D surveys on the velocity model reconstructed by FWI. Their imaging confirms the importance of wide-azimuth surveys for FWI of coarse acquisitions such as node surveys.

Most 3D FWI applications have been limited to low frequencies (<7 Hz). At these frequencies, FWI can be seen as a tool for velocity-model-building rather than a self-contained seismic-imaging tool that continuously proceeds from the velocity-model-building task to the migration task through the continuous sampling of wavenumbers (Ben Hadj Ali et al., 2008b).

Ben Hadj Ali et al. (2008a) apply a frequency-domain FWI algorithm to a series of synthetic data sets. The forward problem is solved in the frequency domain with a massively parallel direct solver. Plessix (2009) presents an application to real ocean-bottom-seismometer (OBS) data. Seismic modeling is performed in the frequency domain with an iterative solver. The inverted frequencies range between 2 and 5 Hz. The inversion converges to a similar velocity model down to the top of a salt body using two different starting velocity models, with one a simple velocity-gradient model. An application to a real 3D streamer data set is presented by Warner et al. (2008) for the imaging of a shallow channel. They perform seismic modeling in the frequency domain using an iterative solver (Warner et al., 2007).

Three-dimensional time-domain FWI is developed by Vigh and Starr (2008a), in which the input data in FWI are plane-wave gathers rather than shot gathers. The main motivation behind the use of plane-wave shot gathers is to mitigate the computational burden by decimating the volume of data. The computational cost is reduced by one order of magnitude for 2D applications and by a factor 3 for 3D applications when the plane-wave-based approach is used instead of the shot-based approach.

Sirgue et al. (2009) apply 3D frequency-domain FWI to the hydrophone component of OBC data from the shallow-water Valhall field. Frequencies between 3.5 and 7 Hz are inverted successively, using a starting model built by ray-based reflection tomography. They successfully image a complex network of shallow channels at 150 m depth and a gas cloud between 1000 and 2500 m depth. Although the spacing between the cables is as high as 300 m, a limited footprint of the acquisition is visible in the reconstructed models. Comparisons of depth-migrated sections computed from the reflection tomography model and the FWI velocity model show the improvements provided by FWI, both in the shallow structure and at the reservoir level below the gas cloud. The step-change improvement in the quality of the depth-migrated image results from the high-resolution nature of the velocity model from FWI and the accounting of the intrabed multiples by the two-way wave-equation modeling engine. Comparisons between depth slices across the channels and the gas cloud extracted from the reflection tomography and the FWI models highlight the significant resolution improvement provided by FWI (Figure 14).

Solving large-scale 3D elastic problems is probably beyond our current tools because of the computational burden of 3D elastic modeling for many sources. This has prompted several studies to develop strategies to mitigate the number of forward simulations required during migration or FWI of large data sets. One of these approaches stacks the seismic sources before modeling (Capdeville et al., 2005). Because the relationship between the seismic wavefield and the source is linear, stacking sources is equivalent to emitting each source simultaneously instead of independently. This assem-
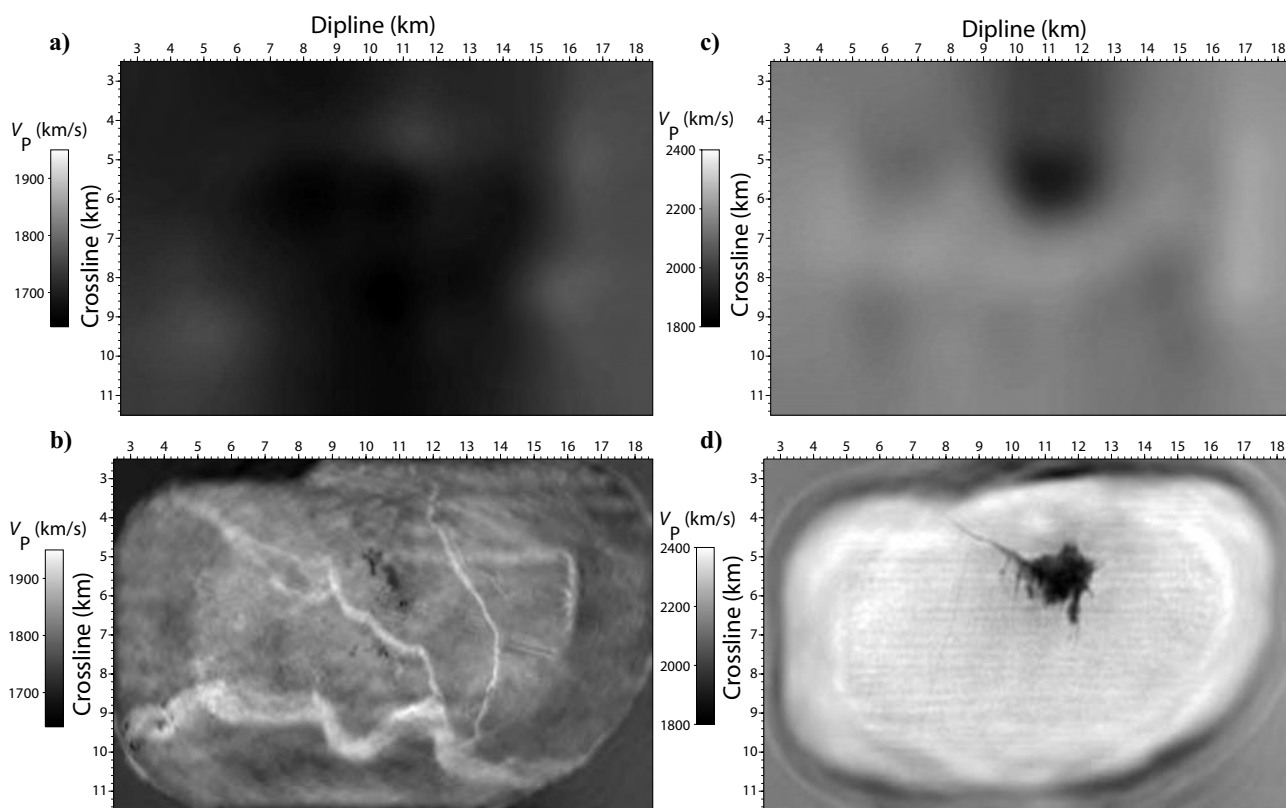

Figure 14. Imaging of the Valhall field by 3D FWI. Depth slices extracted from the velocity models. (a,c) Built by ray-based reflection tomography. (b,d) Built by 3D FWI. (c,d) The shallow slice at 150 m depth shows a complex network of channels in the FWI model, although the deeper slice at 1050 m depth shows a much higher resolution of the top of the gas cloud in the FWI model (image courtesy L. Sirgue and O. I. Barkved, BP).

blage generates artifacts in the imaging that arise from the undesired correlation between each independent source wavefield with the back-propagated residual wavefields associated with the other sources of the stack. Applying some random-phase encoding to each source before assemblage mitigates these artifacts. The method originally was applied to migration algorithms (Romero et al., 2000). Promising applications to time- and frequency-domain FWI are presented in Krebs et al. (2009) and Ben Hadj Ali et al. (2009a, 2009b). Alternatively, Herrmann et al. (2009) propose to recover the source-separated wavefields from the simultaneous simulation before FWI.

An illustration of the source-encoding technique is provided in Figure 15, in which a dip section of the overthrust model is built three ways: by conventional frequency-domain FWI (i.e., without source assemblage; Figure 15b), by FWI with source assemblage but without phase encoding (Figure 15c), and by FWI with source assemblage and phase encoding (Figure 15d). The models were obtained by successive inversions of four groups of two frequencies ranging between 3.5 and 20 Hz. The number of shots was 200, and no noise was added to the data. The number of iterations per frequency group to obtain the final FWI models without and with the source assemblage was 15 and 200, respectively. The time to compute the model of Figure 15d was seven times less than the time to build the model of Figure 15b. More details can be found in Ben Hadj Ali et al. (2009). If the phase-encoding technique is seen as sufficiently robust, especially in the presence of noise, it is likely that 3D elastic FWI can be viewed in the near future using sophisticated modeling engines.
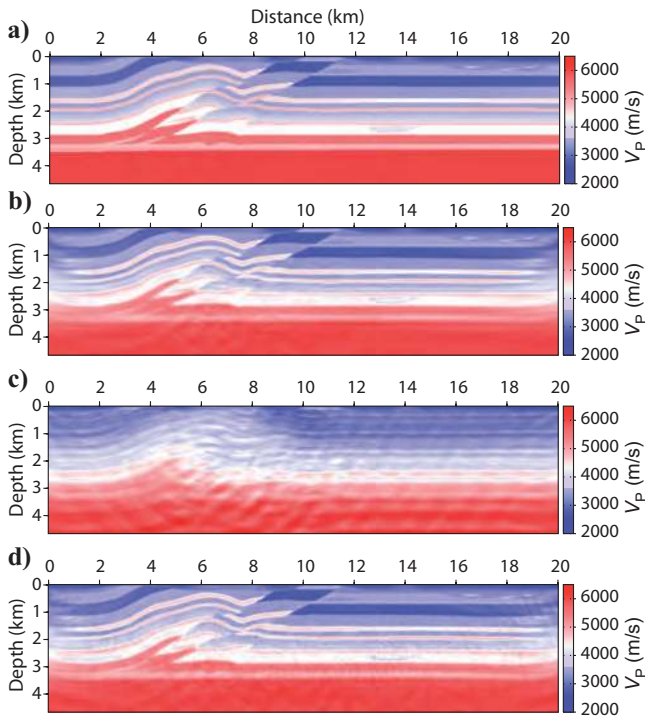

Figure 15. Application of source encoding in FWI; 200 sources and receivers were on the surface. (a) Dip section of the overthrust model. (b) FWI model obtained without source assemblage and phase encoding; 15 iterations per frequency were computed. (c) FWI model after assemblage of all the sources in one super shot. No phase encoding was applied. (d) FWI model obtained with source assemblage and phase encoding; 200 iterations per frequency were computed (image courtesy of H. Ben Hadj Ali).

DISCUSSION

Most of the FWI methods presented and assessed in the literature are based on the local least-squares optimization formulation, in which the misfit between the observed seismograms and the modeled ones are minimized in the time domain or in the frequency domain. Without very low frequencies (<1 Hz), it remains very difficult to obtain reliable results from these approaches when considering real data, especially at high frequencies. Clearly, new formulations of FWI are needed to proceed further.

Recent improvements relate to Hessian estimation, which has been shown to be quite important for better convergence toward the solution. A systematic strategy since the beginning of FWI investigations has been the progressive introduction of the entire content of seismograms through multiscale approaches to partially prevent the convergence toward secondary minima. Transforming the data at the

Table 1. Nomenclature, listed as introduced in the text.

Symbol	Description
$\mathbf{x}$	Spatial coordinates (m)
$t, f, \omega = 2\pi/f$	Time (s), frequency (Hz), circular frequency (rad/s)
$\mathbf{k}, k = c/f$	Wavenumber vector, wavenumber (1/m), where c is wavespeed
$\lambda = 1/k$	Wavelength (m)
$\mathbf{M}, \mathbf{A}, \mathbf{B}$	Mass, stiffness, impedance matrices in the wave equation
$\mathbf{u}(\mathbf{x},t/\omega), \mathbf{s}(\mathbf{x},t/\omega)$	Wavefield solution of the wave equation and seismic source
V_P, V_S, ρ	P-wave and S-wave velocities (m/s), density (kg/m^3)
Q	Attenuation factor for P-waves
δ, ε	Anisotropic Thomsen's parameters
$\mathbf{m}, \mathbf{m}_0, \Delta\mathbf{m}$	Updated and starting FWI models, perturbation model
$\mathbf{d}_{obs}, \mathbf{d}_{cal}(\mathbf{m})$	Recorded and modeled data
$\Delta\mathbf{d} = \mathbf{d}_{obs} - \mathbf{d}_{cal}(\mathbf{m})$	Data-misfit vector
$\mathcal{C}(\mathbf{m}), \nabla C_{\mathbf{m}}$	Misfit function, gradient of the misfit function
$\mathbf{J} = [\partial\mathbf{d}/\partial\mathbf{m}]$	Sensitivity or Fréchet derivative matrix
α	Step length in gradient methods
n	Iteration number in FWI
$\mathbf{H}_a = \mathbf{J}^\dagger\mathbf{J}$	Approximate Hessian
$\mathbf{p}^{(n)}, \beta^{(n)}$	Descent direction and Polak and Ribière coefficient in conjugate gradients
$\mathbf{W}_d = \mathbf{S}_d^t\mathbf{S}_d$	Weighting operators in the data space
$\mathbf{W}_m = \mathbf{S}_m^t\mathbf{S}_m$	Weighting operators in the model space
ε	Damping in damped least-squares FWI
$\mathbf{f}^{(\ell)}$	Virtual source in FWI for diffractor m_ℓ
s, r	Source and receiver indices
θ	Diffraction or aperture angle
τ	Damping coefficient for time damping of seismograms

different stages of the local inversion procedure (particle displacement, particle velocity, logarithms of these quantities, divergence) can also provide benefits for global convergence of the optimization procedure.

Do we need to mitigate the nonlinearity of the inverse problem through more sophisticated search strategies such as semiglobal approaches or even global sampling of the model space? Because this search is quite computationally intensive, should we reduce the dimensionality of the parameter models? To perform such global searches, should we speed up the forward model at the expense of its precision? If so, can we consider these procedures as intermediate strategies for approaching the global minimum?

A pragmatic workflow, from low to high spatial-frequency content, should be expected to move hierarchically from imaging the large medium wavelengths to the short medium wavelengths. A step-by-step procedure for extracting the information, starting from traveltimes and proceeding to true amplitudes, might include many intermediate steps for interpreting new observables (polarization, apparent velocity, envelope).

Another aspect is the quality control of a reconstruction. An objective analysis for uncertainty estimation is important and might rely on semiglobal investigations once the global minimum has been found. Various strategies can be considered that would be manageable as statistical procedures (bootstrapping or jackknifing techniques) or as local nondifferential approaches (simplex, simulation annealing, genetic algorithms).

Finally, solving large-scale 3D elastic problems remains beyond our present tools, although one must be aware that these aids are right around the corner. Because massive data acquisition for 3D reconstruction has been achieved, we indeed expect an improvement in our data-crunching for high-resolution imaging.

An appealing reconstruction is 4D imaging, which is based on time-lapse evolution of targets inside the earth. Differential data are available, providing us with new information for tracking the evolution of the subsurface parameters. Thus, fluid tracking and variations in solid-rock matrices are possible challenges for FWI in the near future.

CONCLUSION

FWI is the last-course procedure for extracting information from seismograms. We have shown the conceptual efforts that have been carried out over the last 30 years to provide FWI as a possible tool for high-resolution imaging. These efforts have been focused on development of large-scale numerical optimization techniques, efficient resolution of the two-way wave equation, judicious model parameterization for multiparameter reconstruction, multiscale strategies to mitigate the ill-posedness of FWI, and specific waveform-inversion data preprocessing.

FWI is mature enough today for prototype application to 3D real data sets. Although applications to 3D real data have shown promising results at low frequencies (<7 Hz), it is still unclear to what extent FWI can be applied efficiently at higher frequencies. To answer this question, a more quantitative understanding of FWI sensitivity to the accuracy of the starting model, to the noise, and to the amplitude accuracies is probably required.

If FWI remains limited to low frequencies, it will remain a tool to build background models for migration. In the opposite case, FWI will tend toward a self-contained processing workflow that can reunify macromodel building and migration tasks.

The present is exciting because realistic applications are becoming possible right now. However, new strategies must be found to make this technique as attractive as the scientific issues require. Fields of investigation should address the need to speed up the forward problem by means of providing new hardware (GPUs) and software (compressive sensing), defining new minimization criteria in the model and data spaces, and incorporating more sophisticated wave phenomena (attenuation, elasticity, anisotropy) in modeling and inversion.

APPENDIX A

APPLICATION OF THE ADJOINT-STATE METHOD TO FWI

In this appendix, we provide some guidelines for the derivation of the gradient of the misfit function (equation 24) with the adjoint-state method and Lagrange multipliers. The reader is referred to Nocedal and Wright (1999) for a review of constrained optimization and to Plessix (2006) for a review of the adjoint-state method and its application to FWI.

First, we introduce the Lagrangian function $\mathcal{L}$ corresponding to the misfit function C augmented with equality constraints:

$$\mathcal{L}(\bar{\mathbf{u}},\mathbf{m},\lambda) = \frac{1}{2}\langle\mathcal{P}(\mathbf{d}_{\text{obs}} - \mathcal{R}\bar{\mathbf{u}})|\mathcal{P}(\mathbf{d}_{\text{obs}} - \mathcal{R}\bar{\mathbf{u}})\rangle$$

$$- \langle\lambda|\mathbf{B}(\mathbf{m})\bar{\mathbf{u}} - \mathbf{s}\rangle \qquad \text{(A-1)}$$

The equality constraints correspond to the forward-problem equation, namely, the state equation, which must be satisfied by the seismic wavefield. A realization $\mathbf{u}$ of the state equation is the so-called state variable. In equation A-1, we introduce the variable $\bar{\mathbf{u}}$ to distinguish any element of the state variable space from a realization of the state equation (Plessix, 2006).

The vector λ, the dimension of which is that of the wavefield $\mathbf{u}$, is the Lagrange multiplier; it corresponds to the adjoint-state variables.

In the framework of the theory of constrained optimization, the first-order necessary optimality conditions known as the Karush-Kuhn-Tucker (KKT) conditions state that the solution of the optimization problem is obtained at the stationary points of $\mathcal{L}$.

The first condition, $[\partial \mathcal{L}/\partial \lambda]_{\bar{\mathbf{u}} = cste, \mathbf{m} = cste} = \mathbf{0}$, leads to the forward-problem equation: $\mathbf{B}(\mathbf{m})\mathbf{u} = \mathbf{s}$. Resolving the state equation for $\mathbf{m} = \mathbf{m}_0$ provides the incident wavefield for FWI.

The second condition, $[\partial \mathcal{L}/\partial \bar{\mathbf{u}}]_{\lambda = cste, \mathbf{m} = cste} = 0$, leads to the so-called adjoint-state equation, expressed as

$$\frac{\partial \mathcal{L}(\bar{\mathbf{u}}, \mathbf{m}, \lambda)}{\partial \bar{\mathbf{u}}} = \mathcal{P}(\mathbf{d}_{obs} - \mathcal{R}\bar{\mathbf{u}}) - \mathbf{B}^{\dagger}(\mathbf{m})\lambda = \mathbf{0}. \quad \text{(A-2)}$$

For the derivation of equation A-2, we use the fact that $\langle \lambda, \mathbf{B}(\mathbf{m})\bar{\mathbf{u}} \rangle = \langle \mathbf{B}^{\dagger}(\mathbf{m})\lambda, \bar{\mathbf{u}} \rangle$ and that the source does not depend on $\bar{\mathbf{u}}$. Choosing $\mathbf{m} = \mathbf{m}_0$ and $\bar{\mathbf{u}} = \mathbf{u}(\mathbf{m}_0)$ in equation A-2 leads to

$$\mathbf{B}^{\dagger}(\mathbf{m}_0)\lambda = \mathcal{P}(\Delta \mathbf{d}_0), \quad \text{(A-3)}$$

which can be rewritten equivalently as

$$\lambda^* = \mathbf{B}^{-1}(\mathbf{m}_0)\mathcal{P}(\Delta \mathbf{d}_0^*), \quad \text{(A-4)}$$

where we exploit the fact that $\mathbf{B}^{-1}$ is symmetric by virtue of the spatial reciprocity of Green's functions. The adjoint-state variables are computed by solving a forward problem for a composite source formed by the conjugate of the residuals, which is equivalent to back propagation of the residuals in the model.

The third condition, $[\partial \mathcal{L}/\partial \mathbf{m}]_{\bar{\mathbf{u}} = cste, \lambda = cste} = 0$, defines the minimum of $\mathcal{L}$ in a comparable way as for the unconstrained minimization of the misfit function C. We have

$$\frac{\partial \mathcal{L}(\bar{\mathbf{u}}, \mathbf{m}, \lambda)}{\partial \mathbf{m}} = \left\langle \lambda \left| \frac{\partial \mathbf{B}(\mathbf{m})}{\partial \mathbf{m}} \bar{\mathbf{u}} \right. \right\rangle. \quad \text{(A-5)}$$

For any realization of the forward problem $\mathbf{u}$, $\mathcal{L}(\mathbf{u}, \mathbf{m}, \lambda) = C(\mathbf{m})$. Therefore, equation A-5 gives the expression of the desired gradient of C as a function of the state variable and adjoint-state variable when $\bar{\mathbf{u}} = \mathbf{u}$:

$$\frac{\partial C}{\partial \mathbf{m}} = \left\langle \lambda \left| \frac{\partial \mathbf{B}(\mathbf{m})}{\partial \mathbf{m}} \mathbf{u} \right. \right\rangle. \quad \text{(A-6)}$$

Inserting the expression of λ (equation A-4) into equation A-3 and choosing $\mathbf{m} = \mathbf{m}_0$ gives the expression of the gradient of C at the point $\mathbf{m}_0$ in the opposite direction of which a minimum of C is sought for in FWI:

$$\frac{\partial C(\mathbf{m}_0)}{\partial \mathbf{m}} = \mathbf{u}^t(\mathbf{m}_0) \frac{\partial \mathbf{B}^t(\mathbf{m}_0)}{\partial \mathbf{m}} \mathbf{B}^{-1}(\mathbf{m}_0)\mathcal{P}(\Delta \mathbf{d}_0^*). \quad \text{(A-7)}$$

Equation A-7 is equivalent to equation 24 in the main text.

REFERENCES

Abubakar, A., W. Hu, T. Habashy, and P. van den Berg, 2009, A finite-difference contrast source inversion algorithm for full-waveform seismic applications: Geophysics, this issue.

Akcelik, V., 2002, Multiscale Newton-Krylov methods for inverse acoustic wave propagation: Ph.D. thesis, Carnegie Mellon University.

Aki, K., and P. Richards, 1980, Quantitative seismology: Theory and methods: W.H. Freeman & Co.

Alerini, M., S. Le Bégat, G. Lambaré, and R. Baina, 2002, 2D PP- and PS-stereotomography for a multicomponent datset: 72nd Annual International Meeting, SEG, Expanded Abstracts, 838–841.

Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: Geophysics, **56**, 2027–2035.

Amundsen, L., and B. Ursin, 1991, Frequency-wavenumber inversion of acoustic data: Geophysics, **56**, 1027–1039.

Askan, A., 2006, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Ph.D. thesis, Carnegie Mellon University.

Askan, A., V. Akcelik, J. Bielak, and O. Ghattas, 2007, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Bulletin of the Seismological Society of America, **97**, 1990–2008.

Askan, A., and J. Bielak, 2008, Full anelastic waveform tomography including model uncertainty: Bulletin of the Seismological Society of America, **98**, 2975–2989.

Barnes, C., and M. Charara, 2008, Full-waveform inversion results when using acoustic approximation instead of elastic medium: 78th Annual International Meeting, SEG, Expanded Abstracts, 1895–1899.

Barnes, C., M. Charara, and T. Tsuchiya, 2008, Feasibility study for an anisotropic full waveform inversion of cross-well data: Geophysical Prospecting, **56**, 897–906.

Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

Bednar, J. B., C. Shin, and S. Pyun, 2007, Comparison of waveform inversion, Part 3: Amplitude approach: Geophysical Prospecting, **55**, 477–485.

Ben Hadj Ali, H., S. Operto, and J. Virieux, 2008a, Velocity model building by 3D frequency-domain full-waveform inversion of wide-aperture seismic data: Geophysics, **73**, no. 5, VE101–VE117.

Ben Hadj Ali, H., S. Operto, J. Virieux, and F. Sourbier, 2008b, 3D frequency-domain full-waveform tomography based on a domain decomposition forward problem: 78th Annual International Meeting, SEG, Expanded Abstracts, 1945–1949.

———, 2009a, Efficient 3D frequency-domain full waveform inversion with phase encoding: 71st Conference & Technical Exhibition, EAGE, Extended Abstracts, 5812.

———, 2009b, Three-dimensional frequency-domain full-waveform inversion with phase encoding: 79th Annual International Meeting, SEG, Expanded Abstracts, 2288–2292.

Beydoun, W. B., and A. Tarantola, 1988, First Born and Rytov approximation: Modeling and inversion conditions in a canonical example: Journal of the Acoustical Society of America, **83**, 1045–1055.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.

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

Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophysical Journal International, **135**, 671–680.

Billette, F., S. Le Bégat, P. Podvin, and G. Lambaré, 2003, Practical aspects and applications of 2D stereotomography: Geophysics, **68**, 1008–1021.

Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis, 2007, Structure of the California Coast Ranges and San Andreas fault at SAFOD from seismic waveform inversion and reflection imaging: Journal of Geophysical Research, doi: 10.1029/2006JB004611.

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

Bouchon, M., M. Campillo, and S. Gaffet, 1989, A boundary integral equation — Discrete wavenumber representation method to study wave propagation in multilayered media having irregular interfaces: Geophysics, **54**, 1134–1140.

Brenders, A. J., and R. G. Pratt, 2007a, Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152–170.

———, 2007b, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.

———, 2007c, Waveform tomography of marine seismic data: What can limited offset offer?: 75th Annual International Meeting, SEG, Expanded Abstracts, 3024–3029.

Brossier, R., S. Operto, and J. Virieux, 2009a, Seismic imaging of complex structures by 2D elastic frequency-domain full-waveform inversion: Geophysics, this issue.

———, 2009b, Two-dimensional seismic imaging of the Valhall model from synthetic OBC data by frequency-domain elastic full-waveform inversion: 79th Annual International Meeting, SEG, Expanded Abstracts, 2293–2297.

Brossier, R., J. Virieux, and S. Operto, 2008, Parsimonious finite-volume fre-

quency-domain method for 2D P-SV wave modelling: Geophysical Journal International, **175**, 541–559.

Bube, K. P., and T. Nemeth, 2007, Fast line searches for the robust solution of linear systems in the hybrid ℓ_1/ℓ_2 and Huber norms: Geophysics, **72**, no. 2, A13–A17.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Capdeville, Y., Y. Gung, and B. Romanowicz, 2005, Towards global earth tomography using the spectral element method: A technique based on source stacking: Geophysical Journal International, **162**, 541–554.

Carcione, J. M., G. C. Herman, and A. P. E. ten Kroode, 2002, Seismic modeling: Geophysics, **67**, 1304–1325.

Cary, P., and C. Chapman, 1988, Automatic 1-D waveform inversion of marine seismic refraction data: Geophysical Journal of the Royal Astronomical Society, **93**, 527–546.

Chapman, C., 1985, Ray theory and its extension: WKBJ and Maslov seismograms: Journal of Geophysics, **58**, 27–43.

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

Chen, P., T. Jordan, and L. Zhao, 2007, Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods: Geophysical Journal International, **170**, 175–181.

Chironi, C., J. V. Morgan, and M. R. Warner, 2006, Imaging of intrabasalt and subbasalt structure with full wavefield seismic tomography: Journal of Geophysical Research, **11**, B05313, doi: 10.1029/2004JB003595.

Choi, Y., D.-J. Min, and C. Shin, 2008, Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupled media: Geophysical Prospecting, **56**, 863–881.

Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

———, 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co., Inc.

Claerbout, J., and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, **37**, 741–768.

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

Danecek, P., and G. Seriani, 2008, An efficient parallel Chebyshev pseudospectral method for large-scale 3D seismic forward modelling: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, P046.

de Hoop, A., 1960, A modification of Cagniard's method for solving seismic pulse problems: Applied Scientific Research, **8**, 349–356.

Dessa, J. X., and G. Pascal, 2003, Combined traveltime and frequency-domain seismic waveform inversion: A case study on multi-offset ultrasonic data: Geophysical Journal International, **154**, 117–133.

Devaney, A. J., 1982, A filtered backprojection algorithm for diffraction tomography: Ultrasonic Imaging, **4**, 336–350.

Devaney, A. J., and D. Zhang, 1991, Geophysical diffraction tomography in a layered background: Wave Motion, **14**, 243–265.

Djikpéssé, H., and A. Tarantola, 1999, Multiparameter ℓ_1-norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1023–1035.

Docherty, P., R. Silva, S. Singh, Z.-M. Song, and M. Wood, 2003, Migration velocity analysis using a genetic algorithm: Geophysical Prospecting, **45**, 865–878.

Dumbser, M., and M. Kaser, 2006, An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes — II. The three-dimensional isotropic case: Geophysical Journal International, **167**, 319–336.

Dummong, S., K. Meier, D. Gajewski, and C. Hubscher, 2008, Comparison of prestack stereotomography and NIP wave tomography for velocity model building: Instances from the Messinian evaporites: Geophysics, **73**, no. 5, VE291–VE302.

Ellefsen, K., 2009, A comparison of phase inversion and traveltime tomography for processing of near-surface refraction traveltimes: Geophysics, this issue.

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**, 1–26.

Erlangga, Y. A., and F. J. Herrmann, 2008, An iterative multilevel method for computing wavefields in frequency-domain seismic inversion: 78th Annual International Meeting, SEG, Expanded Abstracts, 1956–1960.

Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, no. 5, J53–J64.

Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, **175**, 665–685.

Forgues, E., and G. Lambaré, 1997, Parameterization study for acoustic and

elastic ray + Born inversion: Journal of Seismic Exploration, **6**, 253–278.

Futterman, W., 1962, Dispersive body waves: Journal of Geophysics Research, **67**, 5279–5291.

Gao, F., A. R. Levander, R. G. Pratt, C. A. Zelt, and G. L. Fradelizio, 2006a, Waveform tomography at a groundwater contamination site: Surface reflection data: Geophysics, **72**, no. 5, G45–G55.

———, 2006b, Waveform tomography at a groundwater contamination site: VSP-surface data set: Geophysics, **71**, no. 1, H1–H11.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.

Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, **43**, 1342–1351.

Gelis, C., J. Virieux, and G. Grandjean, 2007, 2D elastic waveform inversion using Born and Rytov approximations in the frequency domain: Geophysical Journal International, **168**, 605–633.

Gelius, L. J., I. Johansen, N. Sponheim, and J. J. Stamnes, 1991, A generalized diffraction tomography algorithm: Journal of the Acoustical Society of America, **89**, 523–528.

Gilbert, F., and A. Dziewonski, 1975, An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra: Philosophical Transactions of the Royal Society of London, **278**, 187–269.

Graves, R., 1996, Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences: Bulletin of the Seismological Society of America, **86**, 1091–1106.

Grechka, V., L. Zhang, and J. W. Rector III, 2004, Shear waves in acoustic anisotropic media: Geophysics, **69**, 576–582.

Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, 1310–1319.

Gutenberg, B., 1914, Über erdbenwellen viia. beobachtungen an registrierungen von fernbeben in göttingen und folgerungen über die konstitution des erdkörpers: Nachrichten von der Könglichen Gesellschaft der Wissenschaften zu Göttinge, Mathematisch-Physikalische Klasse, 125–176.

Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: Geophysics, **74**, no. 3, R15–R24.

Hak, B., and W. Mulder, 2008, Preconditioning for linearized inversion of attenuation and velocity perturbations: 78th Annual International Meeting, SEG, Expanded Abstracts, 2036–2040.

Hansen, C., ed., 1998, Rank-deficient and discrete ill-posed problems — Numerical aspects of linear inversion: Society for Industrial and Applied Mathematics.

Herrmann, F. J., Y. Erlangga, and T. T. Y. Lin, 2009, Compressive sensing applied to full-wave form inversion: 71st Conference & Technical Exhibition, EAGE, Extended Abstracts, S016.

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

Hole, J. A., 1992, Nonlinear high-resolution three-dimensional seismic travel time tomography: Journal of Geophysical Research, **97**, 6553–6562.

Hu, W., A. Abubakar, and T. M. Habashy, 2009, Simultaneous multifrequency inversion of full-waveform seismic data: Geophysics, **74**, no. 2, R1–R14.

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

Ikelle, L., J. P. Diet, and A. Tarantola, 1988, Linearized inversion of multioffset seismic reflection data in the ω-k domain: Depth-dependent reference medium: Geophysics, **53**, 50–64.

Improta, L., A. Zollo, A. Herrero, R. Frattini, J. Virieux, and P. DellÁversana, 2002, Seismic imaging of complex structures by non-linear traveltimes inversion of dense wide-angle data: Application to a thrust belt: Geophysical Journal International, **151**, 264–278.

Jaiswal, P., C. Zelt, A. W. Bally, and R. Dasgupta, 2008, 2-D traveltime and waveform inversion for improved seismic imaging: Naga thrust and fold belt, India: Geophysical Journal International, doi: 10.1111/j.1365-246X.2007.03691.x.

Jaiswal, P., C. Zelt, R. Dasgupta, and K. Nath, 2009, Seismic imaging of the Naga thrust using multiscale waveform inversion: Geophysics, this issue.

Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, and D. Trézéguet, 1989, Wavelengths of earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910.

Jin, S., and R. Madariaga, 1993, Background velocity inversion by a genetic algorithm: Geophysical Research Letters, **20**, 93–96.

———, 1994, Non-linear velocity inversion by a two-step Monte Carlo method: Geophysics, **59**, 577–590.

Jin, S., R. Madariaga, J. Virieux, and G. Lambaré, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, 575–588.

Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference,

frequency-space 2D scalar extrapolator: Geophysics, **61**, 529–537.

Jones, K., M. Warner, and J. Brittan, 1999, Anisotropy in multi-offset deep-crustal seismic experiments: Geophysical Journal International, **138**, 300–318.

Kamei, R., and R. G. Pratt, 2008, Waveform tomography strategies for imaging attenuation structure for cross-hole data: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F019.

Kennett, B. L. N., 1983, Seismic wave propagation in stratified media: Cambridge University Press.

Klem-Musatov, K. D., and A. M. Aizenberg, 1985, Seismic modelling by methods of the theory of edge waves: Journal of Geophysics, **57**, 90–105.

Kolb P., F. Collino, and P. Lailly, 1986, Prestack inversion of 1-D medium: Proceedings of the IEEE, 498–508.

Kolsky, H., 1956, The propagation of stress pulses in viscoelastic solids: Philosophical Magazine, **1**, 693–710.

Komatitsch, D., and J. P. Vilotte, 1998, The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of the Seismological Society of America, **88**, 368–392.

Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, 20289–20299.

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

Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, A. Baumstein, M. D. Lacasse, and S. Lee, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, this issue.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Lambaré, G., 2008, Stereotomography: Geophysics, **73**, no. 5, VE25–VE34.

Lambaré, G., and M. Alérini, 2005, Semi-automatic picking PP-PS stereotomography: Application to the synthetic Valhall data set: 75th Annual International Meeting, SEG, Expanded Abstracts, 943–946.

Lambaré, G., S. Operto, P. Podvin, P. Thierry, and M. Noble, 2003, 3-D ray + Born migration/inversion — Part 1: Theory: Geophysics, **68**, 1348–1356.

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

Lecomte, I., 2008, Resolution and illumination analyses in PSDM: A ray-based approach: The Leading Edge, **27**, 650–663.

Lecomte, I., S.-E. Hamran, and L.-J. Gelius, 2005, Improving Kirchhoff migration with repeated local plane-wave imaging? A SAR-inspired signal-processing approach in prestack depth imaging: Geophysical Prospecting, **53**, 767–785.

Lee, K. H., and H. J. Kim, 2003, Source-independent full-waveform inversion of seismic data: Geophysics, **68**, 2010–2015.

Lehmann, I., 1936, P′: Publications du Bureau Central Séismologique International, A14, 87–115.

Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436.

Liao, Q., and G. A. McMechan, 1995, 2.5D full-wavefield viscoacoustic inversion: Geophysical Prospecting, **43**, 1043–1059.

———, 1996, Multifrequency viscoacoustic modeling and inversion: Geophysics, **61**, 1371–1378.

Lions, J., 1972, Nonhomogeneous boundary value problems and applications: Springer-Verlag, Berlin.

Lucio, P. S., G. Lambaré, and A. Hanyga, 1996, 3D multivalued travel time and amplitude maps: Pure and Applied Geophysics, **148**, 113–136.

Luo, Y., T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Seismic modeling and imaging based upon spectral-element and adjoint methods: The Leading Edge, **28**, 568–574.

Malinowsky, M., and S. Operto, 2008, Quantitative imaging of the permo-mesozoic complex and its basement by frequency domain waveform tomography of wide-aperture seismic data from the polish basin: Geophysical Prospecting, **56**, 805–825.

Malinowsky, M., A. Ribodetti, and S. Operto, 2007, Multiparameter full-waveform inversion for velocity and attenuation — Refining the imaging of a sedimentary basin: 69th Conference & Technical Exhibition, EAGE, Extended Abstracts, P276.

Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.

McMechan, G. A., and G. S. Fuis, 1987, Ray equation migration of wide-angle reflections from southern Alaska: Journal of Geophysical Research, **92**, 407–420.

Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

Miller, D., M. Oristaglio, and G. Beylkin, 1987, A new slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, 943–964.

Min, D.-J., and C. Shin, 2006, Refraction tomography using a waveform-inversion back-propagation technique: Geophysics, **71**, no. 3, R21–R30.

Min, D.-J., C. Shin, R. G. Pratt, and H. S. Yoo, 2003, Weighted-averaging fi-

nite-element method for 2D elastic wave equations in the frequency domain: Bulletin of the Seismological Society of America, **93**, 904–921.

Montelli, R., G. Nolet, F. Dahlen, G. Masters, E. Engdahl, and S. Hung, 2004, Finite-frequency tomography reveals a variety of plumes in the mantle: Science, **303**, 338–343.

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

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

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447.

Mulder, W. A., and B. Hak, 2009, Simultaneous imaging of velocity and attenuation perturbations from seismic data is nearly impossible: 71st Conference & Technical Exhibition, EAGE, Extended Abstracts, S043.

Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD): Geophysical Journal International, **169**, 1069–1078.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.

Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer.

Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics: D. Reidel Publ. Co.

Oldham, R., 1906, The constitution of the earth: Quarterly Journal of the Geological Society of London, **62**, 456–475.

Operto, S., G. Lambaré, P. Podvin, and P. Thierry, 2003, 3-D ray-Born migration/inversion, Part 2: Application to the SEG/EAGE overthrust experiment: Geophysics, **68**, 1357–1370.

Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Quantitative imaging of complex structures from multifold wide aperture seismic data: A case study: Geophysical Prospecting, **52**, 625–651.

Operto, S., J. Virieux, P. Amestoy, J. Y. L'Excellent, L. Giraud, and H. Ben-Hadj-Ali, 2007, 3D finite-difference frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, no. 5, SM195–SM211.

Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research, doi: 10.1029/2005JB003835.

Operto, S., J. Virieux, A. Ribodetti, and J. E. Anderson, 2009, Finite-difference frequency-domain modeling of viscoacoustic wave propagation in two-dimensional tilted transversely isotropic media: Geophysics, **74**, no. 5, T75–T95, doi: 10.1190/1.3157243.

Operto, S., S. Xu, and G. Lambaré, 2000, Can we image quantitatively complex models with rays?: Geophysics, **65**, 1223–1238.

Paige, C. C., and M. A. Saunders, 1982a, Algorithm 583 LSQR: Sparse linear equations and least squares problems: Association for Computing Machinery Transactions on Mathematical Software, **8**, 195–209.

———, 1982b, LSQR: An algorithm for sparse linear equations and sparse least squares: Association for Computing Machinery Transactions on Mathematical Software, **8**, 43–71.

Pica, A., J. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in laterally invariant medium: Geophysics, **55**, 284–292.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

———, 2007, A Helmholtz iterative solver for 3D seismic-imaging problems: Geophysics, **72**, no. 5, SM185–SM194.

———, 2009, 3D frequency-domain full-waveform inversion with an iterative solver: Geophysics, this issue.

Polak, E., and G. Ribière, 1969, Note sur la convergence de méthodes de directions conjuguées: Revue Française d'Informatique et de Recherche Opérationnelle, **16**, 35–43.

Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, **38**, 311–330.

———, 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R. G., and N. R. Goulty, 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: Geophysics, **56**, 204–224.

Pratt, R. G., F. Hou, K. Bauer, and M. Weber, 2005, Waveform tomography images of velocity and inelastic attenuation from the Mallik 2002 crosshole seismic surveys, *in* S. R. Dallimore, and T. S. Collett, eds., Scientific results from the Mallik 2002 gas hydrate production research well program, Mackenzie Delta, Northwest Territories, Canada: Geological Survey of Canada.

Pratt, R. G., R. E. Plessix, and W. A. Mulder, 2001, Seismic waveform tomography: The effect of layering and anisotropy: 63rd Conference & Technical Exhibition, EAGE, Extended Abstracts, P092.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton

methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

Pratt, R. G., L. Sirgue, B. Hornby, and J. Wolfe, 2008, Cross-well waveform tomography in fine-layered sediments — Meeting the challenges of anisotropy: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F020.

Pratt, R. G., Z. M. Song, P. R. Williamson, and M. Warner, 1996, Two-dimensional velocity model from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340.

Pratt, R. G., and W. Symes, 2002, Semblance and differential semblance optimisation for waveform tomography: A frequency domain implementation: Sub-basalt imaging conference, Expanded Abstracts, 183–184.

Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography, Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Veltering, 1986, Numerical recipes: The art of scientific computing: Cambridge University Press.

Prieux, V., S. Operto, R. Brossier, and J. Virieux, 2009, Application of acoustic full waveform inversion to the synthetic Valhall model: 77th Annual International Meeting, SEG, Expanded Abstracts, 2268–2272.

Pyun, S., C. Shin, and J. B. Bednar, 2007, Comparison of waveform inversion, Part 3: Phase approach: Geophysical Prospecting, **55**, 465–475.

Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, **159**, 1032–1056.

Ribodetti, A., S. Operto, J. Virieux, G. Lambaré, H.-P. Valéro, and D. Gibert, 2000, Asymptotic viscoacoustic diffraction tomography of ultrasonic laboratory data: A tool for rock properties analysis: Geophysical Journal International, **140**, 324–340.

Ribodetti, A., and J. Virieux, 1996, Asymptotic theory for imaging the attenuation factors Q_p and Q_s: Conference of Aixe-les-Bains, INRIA — French National Institute for Research in Computer Science and Control, Expanded Abstracts, 334–353.

Riyanti, C. D., Y. A. Erlangga, R. E. Plessix, W. A. Mulder, C. Vuik, and C. Oosterlee, 2006, A new iterative solver for the time-harmonic wave equation: Geophysics, **71**, no. 5, E57–E63.

Riyanti, C. D., A. Kononov, Y. A. Erlangga, C. Vuik, C. W. Oosterlee, R. E. Plessix, and W. A. Mulder, 2007, A parallel multigrid-based preconditioner for the 3D heterogeneous high-frequency Helmholtz equation: Journal of Computational Physics, **224**, 431–448.

Robertsson, J. O., B. Bednar, J. Blanch, C. Kostov, and D. J. van Manen, 2007, Introduction to the supplement on seismic modeling with applications to acquisition, processing and interpretation: Geophysics, **72**, no. 5, SM1–SM4.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Saad, Y., 2003, Iterative methods for sparse linear systems, 2nd ed.: Society for Industrial and Applied Mathematics.

Sambridge, M., and G. Drijkoningen, 1992, Genetic algorithms in seismic waveform inversion: Geophysical Journal International, **109**, 323–342.

Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1–29.

Sambridge, M. S., A. Tarantola, and B. L. N. Kennett, 1991, An alternative strategy for non linear inversion of seismic waveforms: Geophysical Prospecting, **39**, 457–491.

Scales, J. A., P. Docherty, and A. Gersztenkorn, 1990, Regularization of nonlinear inverse problems: Imaging the near-surface weathering layer: Inverse Problems, **6**, 115–131.

Scales, J. A., and M. L. Smith, 1994, Introductory geophysical inverse theory: Samizdat Press.

Sears, T. J., S. C. Singh, and P. J. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.

Sen, M. K., and P. L. Stoffa, 1995, Global optimization methods in geophysical inversion: Elsevier Science Publ. Co., Inc.

Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**, no. 4, U47–U57.

Shi, Y., W. Zhao, and H. Cao, 2007, Nonlinear process control of wave-equation inversion and its application in the detection of gas: Geophysics, **72**, no. 1, R9–R18.

Shin, C., and W. Ha, 2008, A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains: Geophysics, **73**, no. 5, VE119–VE133.

Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931.

——, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, **177**, 1067–1079.

Shin, C., S. Jang, and D. J. Min, 2001a, Improved amplitude preservation for

prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Shin, C., and D.-J. Min, 2006, Waveform inversion using a logarithmic wavefield: Geophysics, **71**, no. 3, R31–R42.

Shin, C., S. Pyun, and J. B. Bednar, 2007, Comparison of waveform inversion, Part 1: Conventional wavefield vs. logarithmic wavefield: Geophysical Prospecting, **55**, 449–464.

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

Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: Geophysical Journal International, **151**, 325–344.

Sirgue, L., 2003, Inversion de la forme d'onde dans le domaine fréquentiel de données sismiques grand offset: Ph.D. thesis, Université Paris and Queen's University.

——, 2006, The importance of low frequency and large offset in waveform inversion: 68th Conference & Technical Exhibition, EAGE, Extended Abstracts, A037.

Sirgue, L., O. I. Barkved, J. P. V. Gestel, O. J. Askim, and J. H. Kommedal, 2009, 2D waveform inversion on Valhall wide-azimuth OBS: 71st Conference & Technical Exhibition, EAGE, Extended Abstracts, U038.

Sirgue, L., J. Etgen, and U. Albertin, 2007, 3D full waveform inversion: Wide versus narrow azimuth acquisitions: 77th Annual International Meeting, SEG, Expanded Abstracts, 1760–1764.

——, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F022.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Smithyman, B. R., R. G. Pratt, J. G. Hayles, and R. J. Wittebolle, 2008, Near surface void detection using seismic Q-factor waveform tomography: 70th Annual International Meeting, EAGE, Expanded Abstracts, F018.

Snieder, R., M. 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**, 991–1000.

Song, Z., P. Williamson, and G. Pratt, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data, Part 2: Inversion method, synthetic experiments and real-data results: Geophysics, **60**, 786–809.

Sourbier, F., A. Haiddar, L. Giraud, S. Operto, and J. Virieux, 2008, Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method: A tool for 3D full-waveform inversion?: 78th Annual International Meeting, SEG, Expanded Abstracts, 2147–2151.

Sourbier, F., S. Operto, J. Virieux, P. Amestoy, and J. Y. L'Excellent, 2009a, FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 1: Algorithm: Computers and Geosciences, **35**, 487–495.

——, 2009b, FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 2: Numerical examples and scalability analysis: Computers & Geosciences, **35**, 496–514.

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

Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, 23–48.

Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009, First-arrival travel time tomography based on the adjoint state method: Geophysics, this issue.

Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation, I: Theory: Quarterly Journal of the Royal Meteorological Society, **113**, 1311–1328.

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

——, 1986, A strategy for nonlinear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

——, 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Science Publ. Co., Inc.

Thierry, P., G. Lambaré, P. Podvin, and M. Noble, 1999a, 3-D preserved amplitude prestack depth migration on a workstation: Geophysics, **64**, 222–229.

Thierry, P., S. Operto, and G. Lambaré, 1999b, Fast 2D ray-Born inversion/migration in complex media: Geophysics, **64**, 162–181.

Thomsen, L. A., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

Toksöz, M. N., and D. H. Johnston, 1981, Seismic wave attenuation: SEG.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

Tsvankin, I., 2001, Seismic signature and analysis of reflection data in anisotropic media: Elsevier Scientific Publ. Co., Inc.

van den Berg, P. M., and A. Abubakar, 2001, Contrast source inversion meth-

od: State of the art: Progress in Electromagnetics Research, **34**, 189–218.

Vigh, D., and E. W. Starr, 2008a, 3D plane-wave full-waveform inversion: Geophysics, **73**, no. 5, VE135–VE144.

——, 2008b, Comparisons for waveform inversion, time domain or frequency domain?: 78th Annual International Meeting, SEG, Expanded Abstracts, 1890–1894.

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

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

Virieux, J., S. Operto, H. B. H. Ali, R. Brossier, V. Etienne, F. Sourbier, L. Giraud, and A. Haidar, 2009, Seismic wave modeling for seismic imaging: The Leading Edge, **28**, 538–544.

Vogel, C., 2002, Computational methods for inverse problems: Society of Industrial and Applied Mathematics.

Vogel, C. R., and M. E. Oman, 1996, Iterative methods for total variation denoising: SIAM Journal on Scientific Computing, **17**, 227–238.

Wang, Y., and Y. Rao, 2009, Reflection seismic waveform tomography: Journal of Geophysical Research, doi: 10.1029/2008JB005916.

Warner, M., I. Stekl, and A. Umpleby, 2007, Full wavefield seismic tomography — Iterative forward modeling in 3D: 69th Conference & Technical Exhibition, EAGE, Extended Abstracts, C025.

——, 2008, 3D wavefield tomography: synthetic and field data examples: 78th Annual International Meeting, SEG, Expanded Abstracts, 3330–3334.

Williamson, P., 1991, A guide to the limits of resolution imposed by scatter-ing in ray tomography: Geophysics, **56**, 202–207.

Woodhouse, J., and A. Dziewonski, 1984, Mapping the upper mantle: Three dimensional modelling of earth structure by inversion of seismic waveforms: Journal of Geophysical Research, **89**, 5953–5986.

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

Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: Geophysics, **73**, no. 5, VE5–VE11.

Wu, R. S., 2003, Wave propagation, scattering and imaging using dual-domain one-way and one-return propagation: Pure and Applied Geophysics, **160**, 509–539.

Wu, R. S., and K. Aki, 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595.

Wu, R.-S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25.

Yilmaz, O., 2001, Seismic data analysis: processing, inversion and interpretation of seismic data: SEG.

Zelt, C., and P. J. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe basin: Journal of Geophysical Research, **103**, 7187–7210.

Zelt, C. A., R. G. Pratt, A. J. Brenders, S. Hanson-Hedgecock, and J. A. Hole, 2005, Advancements in long-offset seismic imaging: A blind test of travel-time and waveform tomography: Eos, Transactions of the American Geophysical Union, **86**, Abstract S52A-04.

Zhou, B., and S. A. Greenhalgh, 2003, Crosshole seismic inversion with normalized full-waveform amplitude data: Geophysics, **68**, 1320–1330.