Wavefield tomography based on local image correlations

Francesco Perrone[1] and Paul Sava[1]
[1] *Center for Wave Phenomena, Colorado School of Mines*

ABSTRACT
The estimation of a velocity model from seismic data is a crucial step for obtaining a high-quality image of the subsurface. Velocity estimation is usually formulated as an optimization problem where an objective function measures the mismatch between synthetic and recorded wavefields and its gradient is used to update the model. The objective function can be defined in the data-space (as in full-waveform inversion) or in the image-space (as in migration velocity analysis). In general, the latter leads to smooth objective functions, which are monomodal in a wider basin about the global minimum compared to the objective functions defined in the data space. Nonetheless, migration velocity analysis requires construction of common-image gathers at fixed spatial locations and subsampling of the image in order to assess the consistency between the trial velocity model and the observed data. We present an objective function that extracts the velocity error information directly in the image domain without computing common-image gathers. Because of the dimensionality of the problem, gradient-based methods (such as the conjugate-gradient algorithm) are used in the optimization procedure. In order to include the full complexity of the wavefield in the velocity estimation algorithm, we consider a two-way (as opposed to one-way) wave operator, we do not linearize the imaging operator with respect to the model parameters (as in linearized wave-equation migration velocity analysis), and compute the gradient of the objective function using the adjoint-state method. We illustrate our velocity estimation methodology with a few synthetic examples.

Key words: tomography, migration velocity analysis, imaging

1 INTRODUCTION

Seismic imaging involves the estimation of wave propagation velocities in the subsurface from seismic data recorded at the surface. Seismic velocities are related to other physical parameters (for example, density and compressibility, which characterize the lithology of the Earth), and rock-mechanics parameters (for example, porosity and fluid overpressure, which are crucial in reservoir engineering) (Carcione, 2007).

Seismic imaging includes the estimation of both the *position* of the structures that generate the data recorded at the surface and of a *model* that describes wave propagation in the subsurface. The two problems are closely related since a model is necessary to infer the position of the reflectors. Waves recorded at the surface are extrapolated in a model of the subsurface (by solving a wave equation) and crosscorrelated with a synthetic source wavefield simulated in the same model (Claerbout, 1985). Under the single scattering approximation, reflectors are located where the source and receiver wavefields match in time and space.

Wave-equation tomography (Tarantola, 1984; Woodward, 1992; Biondi and Sava, 1999) is a family of techniques that estimate the velocity model parameters from finite bandwidth signals recorded at the surface. The inversion is usually formulated as an optimization problem where the correct velocity model minimizes an objective function that measures the inconsistency between a trial model and the observations. The objective function can be defined either in the data-space (full-waveform inversion) or in the image-space (migration velocity analysis).

Full-waveform inversion (FWI) (Tarantola, 1984; Pratt, 1999; Sirgue and Pratt, 2004) addresses the estimation problem in the data-space and measures the mismatch between the observations and simulated data. Full-waveform inversion aims to reconstruct the exact model that generates the recorded data. By matching both traveltimes and amplitudes, full-waveform inversion allows one to achieve high-resolution (Sirgue et al., 2010). Nonetheless, a source estimate is needed, the physics of wave propagation (for example, isotropic vs. anisotropic, acoustic vs. elastic, etc.) must be correctly modelled, and a good parametrization (for example, impedance vs. velocity contrasts) is crucial (Kelly et al., 2010). Moreover, an accurate initial model is key to avoid cycle skipping and con-

verge to the global minimum (instead of a local minimum) of the objective function.

Because of the nonlinearity of the wavefields with respect to the velocity model, the objective function in the data domain is highly multimodal (Santosa and Symes, 1989), and local optimization methods can easily get trapped into local minima and fail to converge to the correct model. This is particularly true for reflection full-waveform inversion. Refraction full-waveform inversion focuses on diving waves and mutes the data, retaining only the diving energy (Pratt, 1999). This leads to a better-behaved objective function but requires very long offsets in order to record the refracted energy. Moreover, this approach limits the depth at which a robust inversion result can be expected.

Migration velocity analysis (MVA) (Fowler, 1985; Faye and Jeannot, 1986; Al-Yahya, 1989; Chavent and Jacewitz, 1995; Biondi and Sava, 1999; Sava et al., 2005; Albertin et al., 2006) defines the objective function in the image space and is based on the semblance principle (Al-Yahya, 1989). If the velocity model is correct, images from different experiments must be consistent with each other because a single Earth model generates the recorded data. A measure of consistency is usually computed through conventional semblance (Taner and Koehler, 1969) or differential semblance (Symes and Carazzone, 1991). These two functionals analyze a set of migrated images at fixed locations in space; they consider all the shots that illuminate the points under investigation. Migration velocity analysis leads to smooth objective functions and well-behaved optimization problems (Symes, 1991; Symes and Carazzone, 1991) and is less sensitive than full-waveform inversion to the initial model. On the other hand, because we do not use amplitudes in the imaging step, the estimated model has lower resolution than the ideal full-waveform inversion result (Uwe Albertin, personal communication).

Migration velocity analysis measures either the invariance of the migrated images in an auxiliary dimension (reflection angle, shot, etc.) (Al-Yahya, 1989; Rickett and Sava, 2002; Sava and Fomel, 2003; Xie and Yang, 2008) or focusing in an extended space (Rickett and Sava, 2002; Symes, 2008; Sava and Vasconcelos, 2009; Yang and Sava, 2011). All these approaches require the migration of the entire survey in order to analyze the moveout curve in common-image gathers or measure focusing at a specific spatial location. The dimensionality of the (extended) image space and computational complexity of the velocity analysis step rapidly explodes for realistic case scenarios. Moreover, because of the high memory requirement for storing the partial information from each experiment, only a subset of the image points can be considered in the evaluation of the objective function. Illumination holes and/or irregular acquisition geometries can also impact the quality of the common-image gathers but no systematic study of this problem is reported in the literature to our knowledge.

Reverse-time migration (Baysal et al., 1983; McMechan, 1983) is routinely used in exploration geophysics because of its ability to correctly handle the full complexity of the wave propagation phenomena (under the assuption that the physics of wave propagation in the subsurface is correctly modeled). Its computational cost is nevertheless still prohibitive and limits its integration into a migration velocity analysis loop that requires the extraction of common-image gathers for moveout analysis. A migration velocity analysis procedure based on a full-wave propagation engine allows exploitation of more complex wave phenomena (e.g. overturning reflection, prismatic waves, and multiples) and increases the amount of information in the data that can be used (Farmer et al., 2006). Nonetheless, because of the intrinsic cost of wave extrapolation in the time domain, a different approach must be considered for an effective and efficient implementation of velocity model building.

Up to now, the powerful tool offered by reverse-time migration has been used only for reconstructing an image of the impedance contrasts in the subsurface. The velocity model used for generating a migrated image is obtained with other techniques, for example through traveltime tomography (Bishop et al., 1985) or through wave-equation migration velocity analysis (Biondi and Sava, 1999; Sava and Biondi, 2004; Shen and Symes, 2008), which are based on asymptotic approximations and/or linearization of the wave equation operator. These techniques do not exploit all the information encoded in the recorded wavefields because they are either inaccurate in complex velocity models (traveltime tomography) or unable to properly model part of the propagation phenomena (one-way methods). Designing a velocity model-building procedure based on a two-way wave propagation engine will allow us to exploit the full complexity of the wavefields and will make the velocity analysis step conceptually consistent with the imaging algorithm used.

Because reverse-time migration is so demanding from a computational and storage point of view, we propose to analyze the semblance of small groups of migrated images from nearby experiments, i.e. neighboring shots or plane-waves with similar ray-parameter or take-off angle, at every image point illuminated by the experiments. Reverse-time migration works only in common-shot and common-receiver configurations, so the choice of the shot-domain for performing migration velocity analysis with a two-way engine may be preferable. Here, we use the word "shot" in its broader sense to include synthetic shot-gathers like plane-wave sources (Whitmore, 1995; Liu et al., 2006; Stoffa et al., 2006; Zhang et al., 2005), random shot-encoded sources (Morton and Ober, 1998; Romero et al., 2000) or any other phase-/amplitude-encoded source (Soubaras, 2006; Perrone and Sava, 2011) in the range of shot-profile migration. Other techniques for velocity model building in the shot-domain use different strategies. Nonetheless, all of them consider the entire survey and migrate all the data before starting the migration velocity analysis loop (de Vries and Berkhout, 1984; Yilmaz and Chambers, 1984; Al-Yahya, 1989; Chavent and Jacewitz, 1995; Sava and Biondi, 2004; Shen and Symes, 2008; Symes, 2008; Xie and Yang, 2008). We propose an objective function that evaluates the degree of semblance between images through local correlations in the image space and does not need common-image gathers (CIGs). We use the morphologic relationship between

images from nearby experiments to define an objective function that measures shifts in the image space. The methodology described in this work follows from the linearized wave-equation MVA operator proposed by (Perrone et al., 2012) but removes the linearization of the wave-operator with respect to the model parameters and thus accounts for the full two-way nature of the wavefields. We compute the gradient of the objective function with the adjoint-state method (Plessix, 2006). Our approach reduces the memory requirements and avoids the need to pick moveout on gathers, and allows us to include all points illuminated by the seismic experiments in the velocity model building.

2 THEORY

In order to set up an optimization problem, we need to define a residual in either the data space, $r\left(t\right)$, or the image space, $r\left(\mathbf{x}\right)$, that represents an indirect measure of the error in the model parameters. An objective function can then be defined as the energy (l_2 norm) of the residual in the appropriate domain. Other norms (for example the l_1 norm) have been used to exploit the concept of sparsity and gain robustness against the unavoidable noise in the data (Claerbout and Muir, 1973; Tarantola, 2005). In the simple case of linear problems, the l_2 norm leads to quadratic optimization problems for which we can easily compute the gradient of the objective function; in contrast, the l_1 norm leads to nonquadratic objective function which are much more difficult to differentiate. This fundamental difference makes the l_2 norm a more popular choice for setting up an inverse problem in the optimization framework. In the following, we briefly review how the residual is defined; then, we introduce our measure of velocity error in the image space and the resulting objective function. The adjoint-state calculation of the gradient concludes the section.

2.1 Data space residuals

The residual $r\left(t\right)$ can be defined in different ways: in full-waveform inversion, the residual is a function of time, and it is defined as the difference between the synthetic and observed data (Tarantola, 1984):

$$r(t) = d^{syn}\left(t\right) - d^{obs}\left(t\right).$$

This residual captures information about the traveltime and amplitude error due to an incorrect velocity model. The observed data set a time reference and the mismatch in traveltime can be directly backprojected in the velocity model: a positive time lag corresponds to a velocity model that is too fast; a negative time lag corresponds to a velocity model that is too slow. A different residual function can be obtained computing the crosscorrelation of the synthetic and observed data and penalizing the nonzero lags of the crosscorrelation (van Leeuwen and Mulder, 2008):

$$r(t) = \sum_{\tau} P(\tau)\left(d^{syn}\left(t\right) *_{\tau} d^{obs}\left(t\right)\right).$$

The asterisk denotes the crosscorrelation operation, the variable τ indicates the crosscorrelation lag in the time domain, the penalty operator $P(\tau)$ can be any monotonic function of τ that is zero at $\tau = 0$. In this case, the focus is only on the traveltime information and not on the amplitude error in the data. If the velocity model is incorrect, the crosscorrelation of the synthetic and observed data shifts from the zero lag and indicates the sign of the velocity error. The penalty operator annihilates the signal around zero lag and highlights the traveltime mismatch.

2.2 Image space residuals

In migration velocity analysis, the residual $r(\mathbf{x})$ is computed using the source and receiver wavefields extrapolated in the subsurface model, and thus it is defined in the image space. Various measure of velocity error are reported in the literature, most of them are based on the curvature of the moveout observed in common-image gathers in various domains (offset, reflection angle, etc.). Recently, the focusing criterion (Shen and Symes, 2008; Symes, 2008; Sava and Vasconcelos, 2009) have been used for defining a residual in the image domain that represent a proxy for the velocity error. Both the construction of common-image gathers and the measure of focusing require the migration of all the data that illuminate the portion of the model under investigation.

We present an objective function that does not require the explicit construction of common-image gathers and operates directly on the migrated images. The design of this objective function follows from previous work on linearized wave-equation MVA (Perrone et al., 2012), which used the warping relationship between images to construct an image perturbation that measures the model error and drives the model update. Here, we estimate the relative shifts between images using penalized local correlations (a bilinear operator) and are able to integrate this measure into a full-wavefield inversion algorithm using the adjoint-state method. Our goal is to assess the correctness of the velocity model from a small subset of images from nearby experiments but including all the points in the image into the objective function and without increasing the computational cost. We want to trade complete illumination (necessary to measure focusing) at a single spatial location for subsurface aperture (i.e., all the points in the image illuminated by the shots under analysis). We review the measure of consistency based on the semblance principle and introduced in Perrone et al. (2012), define a residual $r\left(\mathbf{x}\right)$ in the image space, and then formulate an objective function as the l_2 norm of the residual.

There are at least two options for measuring similarity between two images. The first one is to compute the pointwise difference, in which case the residual function is

$$r\left(\mathbf{x}\right) = R_{i+1}\left(\mathbf{x}\right) - R_i\left(\mathbf{x}\right), \tag{1}$$

where $R_i\left(\mathbf{x}\right)$ and $R_{i+1}\left(\mathbf{x}\right)$ are the migrated images for ith and $(i+1)$th shot, respectively. The two shots are assumed to be close and to image the same area in the subsurface. Plessix (2006) uses the energy of the image difference as a regulariza-

tion term for full-waveform inversion and not as a stand-alone objective function. Notice that if the velocity model used in migration is severely inaccurate, we can have cycle skipping in the image space or, in other words, the difference between the two images can produce two events if the position of the same reflector in the two images changes more than a quarter of a wavelength. This problem is analogous to what happens in full-waveform inversion in the data space.

An alternative solution is to compute local correlations of the two images at each point $\mathbf{x}$ and penalize according to the local dip estimated from one of the images. This approach measures the relative shift of one image with respect to the other directly in the image space and can be used for velocity analysis. Our methodology is the dual in the image domain of the data correlation strategy developed by van Leeuwen and Mulder (2008) in the frameowrk of full-waveform inversion. We express the residual as

$$r(\mathbf{x}) = \int P(\mathbf{x}, \boldsymbol{\lambda})\, c_i(\mathbf{x}, \boldsymbol{\lambda})\, d\boldsymbol{\lambda}, \tag{2}$$

where

$$c_i(\mathbf{x}, \boldsymbol{\lambda}) = \int_{w(\mathbf{x})} R_i\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) d\boldsymbol{\xi} \tag{3}$$

is the local correlation of the images $R_i(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$ (Hale, 2007b) and $P(\mathbf{x}, \boldsymbol{\lambda})$ is a penalty operator that highlights the shift of the local correlation along the dip direction. The vector $\boldsymbol{\lambda}$ denotes the correlation lag in the image space and the index i scans the shot position. Local Gaussian windows $w(\mathbf{x})$ centered about the image points $\mathbf{x}$ weight the input images and allow computation of local crosscorrelations.

2.3 The imaging condition

The residual $r(\mathbf{x})$ in equation 2 is computed from the migrated images $R_i(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$. The dependency on the source and receiver wavefields is hidden in the imaging condition we use to construct the migrated image. An image of the subsurface $R(\mathbf{x})$ is conventionally computed as the zero-lag time correlation of the source and receiver wavefields:

$$R(\mathbf{x}) = \int_T u_s(\mathbf{x}, t)\, u_r(\mathbf{x}, t)\, dt, \tag{4}$$

where T is the recording time interval in the data, or the zero-lag frequency correlation of the source and receiver wavefields:

$$R(\mathbf{x}) = \int_\Omega \overline{u_s(\mathbf{x}, \omega)} u_r(\mathbf{x}, \omega)\, d\omega, \tag{5}$$

where the overbar denotes the complex conjugation operator and Ω denotes the frequency bandwidth of the signals.

2.4 Image correlation objective function

We restate the semblance principle as follows: if the velocity model is correct, two images from nearby experiments construct the final image along the direction of the reflectors (Perrone et al., 2012). This is equivalent to the standard assumption that, if the model is correct, the prestack image-cube is invariant with respect to the shot position, i.e., no moveout in the shot-domain common-image gathers (Xie and Yang, 2008).

The structural dip is an attribute that is commonly extracted in seimic processing (van Vliet and Verbeek, 1995; Fomel, 2002; Hale, 2007a). The similarity of two images along the structural dip can be measured by means of local correlations (Hale, 2007b). We illustrate this idea on a simple model with a single horizontal density interface (Figure 1); the correct velocity model is constant and equal to 2.0 km/s. Figure 2 shows the dip and apparent displacement vector field (white and black arrows, respectively) at particular spatial locations. Observe that the two vectors are orthogonal when the velocity model is correct.

The computation of the displacement field requires a highly nonlinear procedure, which extracts the *maximum* of the local correlations at every image point (Hale, 2007b). We reformulate the problem of evaluating the orthogonality between dip and displacement using local correlations (as defined in equation 3) penalized along the dip direction (Figure 3). If the velocity model is correct, the *maximum* of the correlation lies along the reflector slope (Figure 3(a)); otherwise, we observe a deviation that represents the relative shift of one image with respect to the other (Figure 3(b)). To measure the shift between the two images, we consider the penalty operator

$$P(\mathbf{x}, \boldsymbol{\lambda}) = \boldsymbol{\lambda} \cdot \boldsymbol{\nu}(\mathbf{x}),$$

where $\boldsymbol{\lambda}$ is the correlation lag vector and the dip vector $\boldsymbol{\nu}(\mathbf{x})$ is the normal to the reflector at $\mathbf{x}$. The penalty operator $P(\mathbf{x}, \boldsymbol{\lambda})$ is a linear function in the dip direction (normal to the reflector) and the isopenalty lines are parallel to the reflector. The penalty operator is identically zero in the direction tangential to the reflector. The local correlation of the two images is an even function with respect to the direction of the reflector: it is so because the two images illuminate the same area of the model and are therefore similar. The penalty operator changes the symmetry of the local correlation: if the velocity model is correct, the product between $P(\mathbf{x}, \boldsymbol{\lambda})$ and $c(\mathbf{x}, \boldsymbol{\lambda})$ is an odd function in the dip direction and by stacking over the correlation lags we obtain zero (Figure 3(a)); on the other hand, if the velocity model is incorrect, the maximum of the local correlation deviates from the direction of the reflector (Figure 3(b)) and is enhanced by the penalty operator. The stack over the correlation lags does not sum to zero because the function is not odd in the dip direction and we obtain a measure of relative shift between the two images. Since the maximum of the local correlation is positive, the sign of the penalty operator in the dip direction indicates the direction of the relative shift of one image with respect to the other. Figure 4 shows the local correlations, penalty operators, and penalized local correlations for a particular location on the reflector in the three cases in Figure 2. The rows of Figure 4 refer to different models (from top to bottom: a too high, the correct, and a too low velocity model). The point on the image has lateral position $x = 3.5$ km and vertical that depends on the imaged reflector.

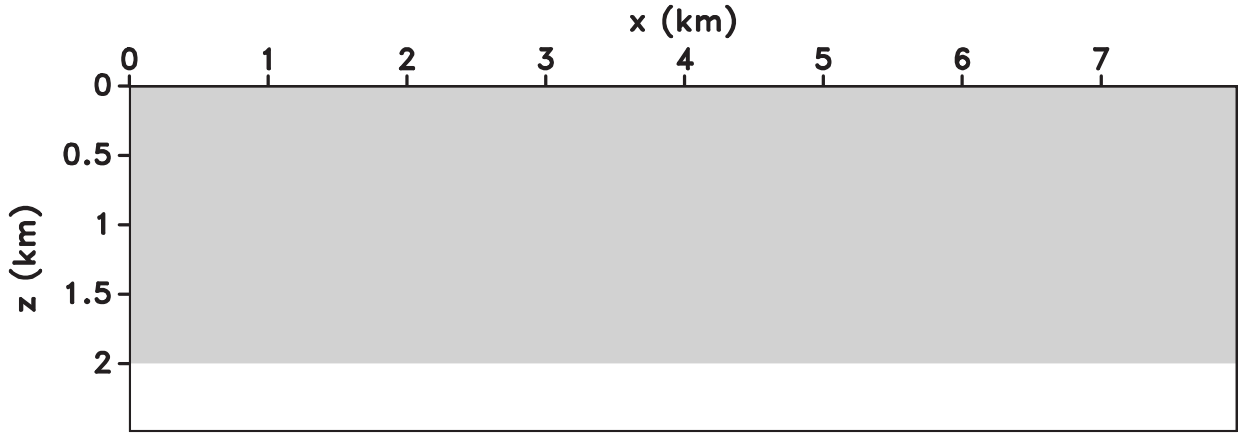


Figure 1. Test model: single horizontal density interface in a constant 0.5 s/km slowness medium. The lighter shades of grey correspond to higher values of density.

Figure 4(a), 4(d), and 4(g) show the local correlations computed from two nearby shots; observe the shift in the peak of the local correlation and the orientation of the correlation panels. Figures 4(b), 4(e), and 4(h) show the penalty operators that we use to highlight the inconsistency between the dip measured on the image and the local correlations. The orientation of the penalty operators does indeed depend on the measured dip. The applied tapering reduces truncation artifacts from the edges of the local correlation panels. Figures 4(c), 4(f), and 4(i) show the penalized correlations. The apparent shift is now clear as a sign unbalance in the panels. When the velocity model is correct, the penalized correlation is an odd function (i.e.,zero mean value) (Figure 4(f)); on the other hand, the panels lose their symmetry when the model is incorrect (Figures 4(c) and 4(i)). The sign of the mean value of the penalized panels measures the apparent shift between the two images.

We define the objective function

$$\begin{aligned} \mathcal{J}(m) &= \frac{1}{2}\sum_i \|r_i(\mathbf{x})\|^2_{\mathbf{x}} \\ &= \frac{1}{2}\sum_i \left\| \int P(\mathbf{x},\boldsymbol{\lambda})\, c_i(\mathbf{x},\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right\|^2_{\mathbf{x}}, \end{aligned}$$

where m represents the model parameters. Figure 5 shows the values of the objective function for different constant perturbations of the model for the simple example shown in Figure 1. The objective function is smooth and convex in the range of errors considered.

2.5 Computation of the gradient with the adjoint-state method

Because of the dimensionality of the model, the optimization problem is addressed using a local gradient-based method (steepest descent, conjugate gradient, etc.) (Vogel, 2002). The computation of the derivative of the state variables with respect to the model parameters (Fréchet derivatives) is not practical (or even possible) because of the large number of dimensions of the model space. The adjoint-state method (Lions, 1972; Plessix, 2006) is an efficient algorithm that computes the gradient of an objective function, which depends on some variables that describe the state of the physical system under analysis (state variables), without computing the Fréchet derivatives of these variables with respect to the model parameters.

The adjoint-state method consists of four steps: the computation of the state variables, the computation of the adjoint sources, the computation of the adjoint-state variables, and finally the computation of the gradient of the objective function. The first and third step (the evaluation of the state and adjoint-state variables) require the solution of the wave-equation that governs the physics of the problem. The adjoint-state method is computationally efficient because only two wavefield simultations are required at each iteration for computing the gradient of the objective function instead of a simulation for each parameter in the model. The Fréchet derivatives of the state variables with respect to the model parameters are extremely big objects: they are defined in the space given by the cartesian product between the space of the state variables and the space of the model. They are not practical to compute and impossible to store in memory for realistic case scenarios in exploration geophysics. Nonetheless, they describe the sensitivity of each state variable to changes in each model parameter and are a powerful tool for resolution analysis.

In the following, we refer to the generic state variable using the letter u, the adjoint sources are indicated as g, and the adjoint-state variables are designated by the letter a. For wavefield tomography, there are two adjoint sources per experiment i: the source wavefield $u_{s,i}(\mathbf{x},t)$ and receiver wavefield $u_{r,i}(\mathbf{x},t)$, where i represents the shot index.

The state variables u are obtained from the solution to the forward problem

$$\mathcal{L}(m)\, u = f, \tag{6}$$

where f is the the source term vector, $\mathcal{L}(m)$ is the forward

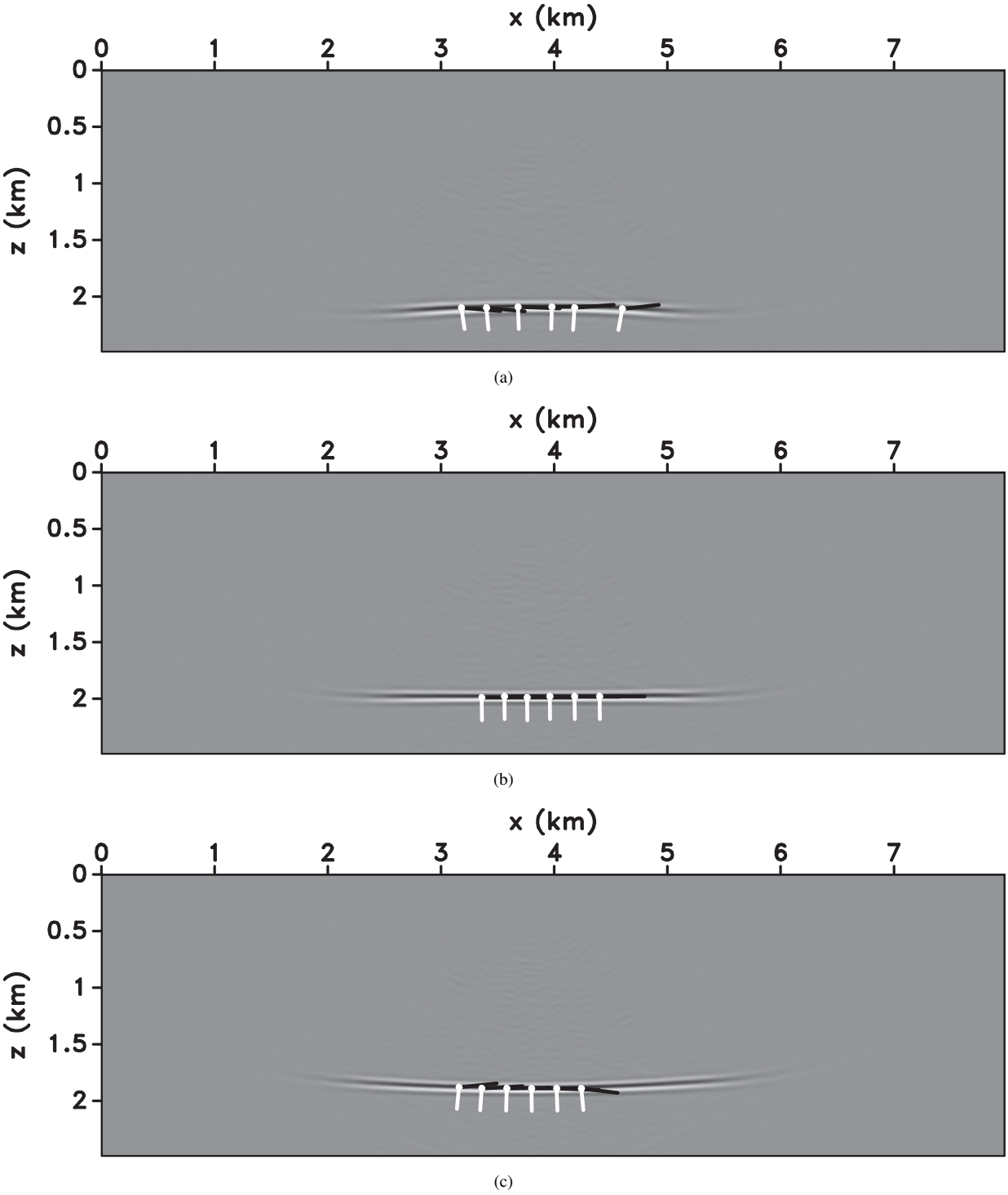


Figure 2. Relationship between the dip field and the apparent displacement: dip (white) and displacement (black) vector fields for different values of slowness error for a single horizontal reflectors: (a) $\frac{\Delta m}{m} = 10\%$, $\frac{\Delta m}{m} = 0$, and (c) $\frac{\Delta m}{m} = 10\%$. The correct slowness value is 0.5 s/km.

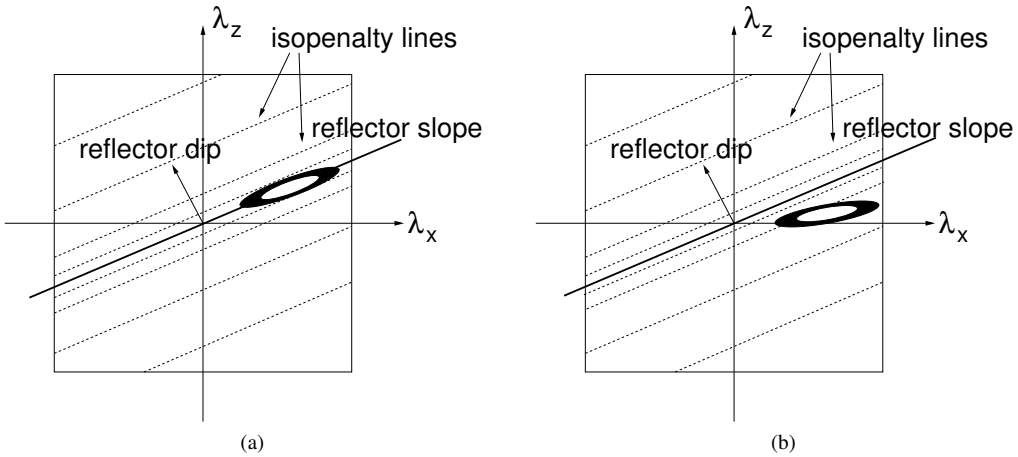


Figure 3. The cartoon describes the behavior of local correlations for the correct and for a incorrect model. When the model is correct, the the maximum of the correlation of two images is aligned with the slope of the reflector (a), otherwise, the maximum of the correlation deviates (b).

modeling operator, and m indicates the model parameters. $\mathcal{L}$ can be a one-way or two-way wave operator; thus we can use either downward continuation (Stolt, 1978; Gazdag and Sguazzero, 1984; Stoffa et al., 1990) or reverse-time migration (Baysal et al., 1983; McMechan, 1983) in the wavefield tomography procedure. Here, we consider a two-way wave operator (d'Alambert or Helmholtz), and m represents the squared slowness. As shown in the following, the choice of squared slowness is convenient because it simplifies the final calculation of the gradient and allows us to obtain an expression that is independent on the model parameters at the current iteration.

The adjoint sources g are the partial derivatives of the objective function with respect to the state variables:

$$g = \frac{\partial \mathcal{J}}{\partial u}. \tag{7}$$

The actual expression of the adjoint sources depends on the objective function. The particular design of the objective function impacts the adjoint sources and characterizes the particular wavefield tomography strategy proposed. In the following sections, we derive the expressions for the adjoint sources for an objective function based on local image correlations (equation 6).

We solve the adjoint problem and compute the adjoint variables a using the adjoint sources g as the force term:

$$\mathcal{L}^{*}(m)\, a = g, \tag{8}$$

where $\mathcal{L}^{*}(m)$ is the operator adjoint to $\mathcal{L}(m)$. Finally, the gradient is given by the inner product

$$\frac{\partial \mathcal{J}}{\partial m} = -\left\langle a, \frac{\partial \mathcal{L}}{\partial m} \right\rangle. \tag{9}$$

For example, for the Helmholz operator parametrized in terms of slowness squared, $\mathcal{L} = -\nabla^2 - m\omega^2$, the partial derivative $\frac{\partial \mathcal{L}}{\partial m}$ returns a simple scaling factor $-\omega^2$, and the inner product $\langle a, \frac{\partial \mathcal{L}}{\partial m} \rangle = -\omega^2 \langle a, u \rangle$ can also be seen as the zero-lag correlation of u and a, similar to the procedure employed in conventional FWI.

2.6 Adjoint sources

As described in the previous section, the adjoint sources $g = \frac{\partial \mathcal{J}}{\partial u}$ are the derivatives of the objective function with respect to the state variables. In our objective function, we consider the dip field a slowly varying function of state variables and neglect the derivative of the penalty operator with respect to u. The details of the derivation of the expressions for the adjoint sources are in appendix A; here, we report the final result:

$$\begin{aligned} g_{s,i} &= u_{r,i}(\boldsymbol{\eta}, t) \left[r_{i-1} \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\eta} + \frac{\boldsymbol{\lambda}}{2}\right) R_{i-1}(\boldsymbol{\eta} - \boldsymbol{\lambda})\; d\boldsymbol{\lambda} \right. \\ &+ \left. r_i \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\eta} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}(\boldsymbol{\eta} + \boldsymbol{\lambda})\; d\boldsymbol{\lambda} \right], \end{aligned} \tag{10}$$

and

$$\begin{aligned} g_{r,i} &= u_{s,i}(\boldsymbol{\eta}, t) \left[r_{i-1} \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\eta} + \frac{\boldsymbol{\lambda}}{2}\right) R_{i-1}(\boldsymbol{\eta} - \boldsymbol{\lambda})\; d\boldsymbol{\lambda} \right. \\ &+ \left. r_i \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\eta} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}(\boldsymbol{\eta} + \boldsymbol{\lambda})\; d\boldsymbol{\lambda} \right], \end{aligned} \tag{11}$$

where $g_{s,i} = g_{s,i}(\mathbf{x}, \boldsymbol{\eta}, t)$ and $g_{r,i} = g_{r,i}(\mathbf{x}, \boldsymbol{\eta}, t)$ indicate the adjoint sources for the source and receiver wavefield of the ith shot, respectively, and $r_i = r_i(\mathbf{x})$ is defined as in equation 2.

The expressions of the adjoint sources in equation 10 and 11 are quite complex and need to be explained in some detail. Notice that the adjoint sources depend on 3 variables $(\mathbf{x}, \boldsymbol{\eta}, t)$. The vector $\mathbf{x}$ represents the physical space, where model and image are defined; the variable t identifies the time axis; $\boldsymbol{\eta}$ is an auxiliary vector defined in the physical space that spans the local window around every image point. The adjoint source is obtained by spreading the value of the residual at each point in a window around the image point and by weighting it by the

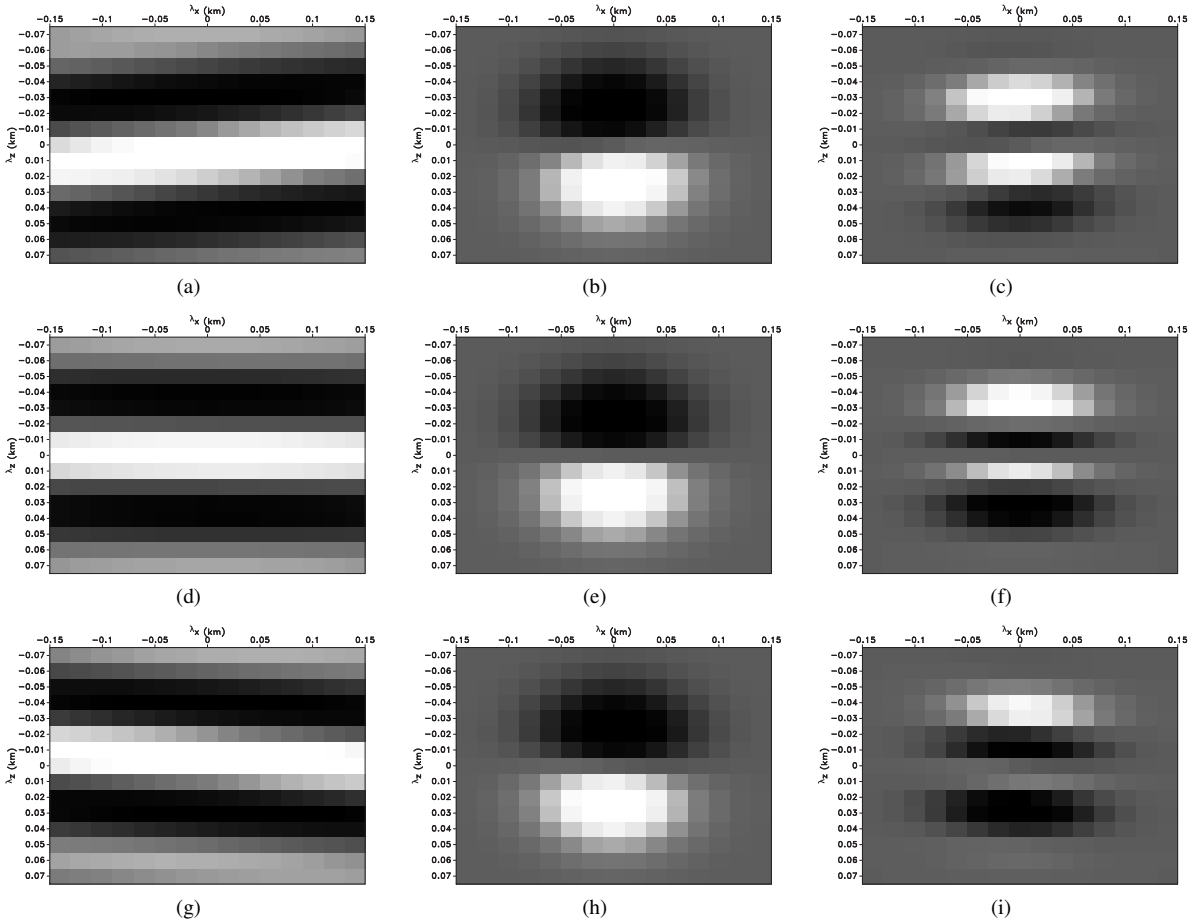


Figure 4. Measure of the relative displacement between two images using penalized local correlations. We consider 2 images from nearby experiments of the reflector in Figure 2; each row (from top to bottom) represents a different model: too high, correct, and too low; the columns show (from left to right) the local correlation panel, the penalty operator used, and the penalized local correlation. For each model, we pick a point on the reflector with lateral coordinate $x = 3.5$ km. The vertical coordinate changes as a function of model parameters because the depth of the reflector changes. Observe the asymmetry of the penalized local correlations for the wrong models: the mean value (i.e., the stack over the correlation lags) of the penalized correlations gives us a measure of the relative shift between two images.

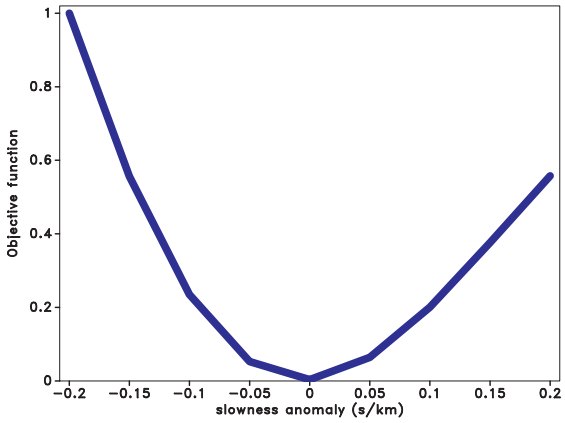


Figure 5. Values of the objective function for different errors in the slowness model. We consider constant perturbations ranging from -10% to 10% of the exact value (0.5 s/km).

the integral over $\boldsymbol{\lambda}$, which represents a local convolution of the image and the penalty operator. The residual in equation 2 measures the relative displacement of one image with respect to the other; the adjoint source thus estimates the curvature of the moveout in the shot-domain common-image gathers at each image point and scales the background wavefields by this value.

Equations 10 and 11 depend on two images, $R_{i-1}(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$, and may suffer from cycle skipping if the distance between the two shots and the model complexity cause the reconstucted reflectors to be more distant than a quarter of a wavelegth. In order to overcome this problem, we use the central image $R_i(\mathbf{x})$ to compute the adjoint sources.

The dependence of the penalty operator $P(\mathbf{x}, \boldsymbol{\lambda})$ on the state variables u passes through the definition of the dip vector $\boldsymbol{\nu}(\mathbf{x})$, which is normal to the reflectors at every point in the image. The dip field defines the wavefront set of the function that represents the reflection events in the image (Chang et al., 1987); the link between the image (and the wavefields) and the dip field can be written exploiting the concept of gradient square tensor (van Vliet and Verbeek, 1995): the dip vector is the eigenvector associated with the largest eigenvalues. There is no simple linear relationship between the dip vector and the wavefields, i.e. the state variables. Neglecting this term in the computation of the adjoint sources may introduce an error of the same order of magnitude of the term considered. A thorough study of $\partial P/\partial u$ and its impact on the computation of both the adjoint sources and the gradient of the objective function is subject for future research.

In the following test, we consider the simple density interface in Figure 1 and three shots 60 m apart starting from $x = 4$ km. The data are computed in a constant 2.0 km/s velocity medium using a two-way finite-difference scheme. We migrate the data with three different slowness models and perform sensitivity analysis for our methodology. We show the residuals, the objective function as a function of the model perturbation, and the gradient with respect to the model paramenters.

2.7 Flat reflector in a constant velocity medium

Figure 6 shows the migrated images of a shot located at $x = 4$ km using three different velocity models. Observe that with a single shot, the image of a horizontal reflector curves up or down depending on the sign of the velocity error. The relative displacement between different images is measured by the residual defined in equation 2. Figure 7 shows the residuals calculated using two images from adjacent shots. Observe the change in sign across the zero-offset reflection point and how the sign changes with the velocity error. The shift between the images tells us that the velocity model contains errors but it does not unequivocally relate with the sign of the error. We can combine the residuals and obtain an estimate of the curvature of the shot-domain common-image gathers (Figure 8) that is more directly connected to the sign of the velocity error. The sign flip at the edges of the reflector are due to the limited aperture of the experiment and the fact that the three images do not overlap over the entire extent of the reflector.

In Figure 9, we show the gradient computed from the correlation objective function. We consider three shots to compute the residuals and use only the wavefields of the central experiment to compute the gradient. Observe the correlation between the sign and pattern of the curvature estimate in Figure 8 and the computed gradient of the objective function. The gradient of the objective function represents the direction of maximum *increase* of the function itself and its sign depends on the chosen parametrization. We compute the gradient with respect to slowness squared; this means that a positive gradient points toward slower models whereas a negative gradient points toward faster models.

3 EXAMPLES

We compute the gradients for the objective function based on local image correlations using the adjoint-state method and show a few inversion tests. We discuss 3 cases of increasing complexity. First, we consider a horizontal density interface (Figure 1) in a homogeneous slowness model. We use 30 sources evenly spaced 200 m apart at the surface and receivers located at every grid point at $z = 0$. We demonstrate how the model is updated through inversion by means of a gradient-based method and the adjoint-state method. Second, we run an inversion test on a simple heterogeneous model with different layers using 40 shots evenly spaced 200 m apart and receivers at every grid point on the surface. We use shot-domain common-image gathers to assess the quality of the result and to show that our methodology is able to obtain a set of images that is invariant with respect to the experiment position. The final test uses the Marmousi model. The data are generated with a single scattering code and migrated with a downward continuation scheme.

3.1 Homogeneous slowness model

We consider again the simple model in Figure 1 with a single horizontal density interface. The data are generated using a time-domain finite difference code with absorbing boundary conditions implemented on all sides of the model. Random Gaussian noise is added to the data and the signal-to-noise ratio is equal to 10. The correct model (Figure 10) is a homogeneous slowness layer and the initial model is homogeneous but with an incorrect value of slowness (Figure 11(a)). We model 30 shots at the surface starting from $x = 1.01$ km; the spacing between the shots is 200 m and the 400 receivers are located at every grid point on the surface.

In the image obtained with the inital model (Figure 11(b)), the reflector is mispositioned because of the erroneous value of the model parameters; the shot-domain common-image gathers computed for illustration purposes show that different shots image the reflector at different depths (Figure 11(c)). We run five iterations of steepest descent (Vogel, 2002) and reconstruct the slowness model above the re-

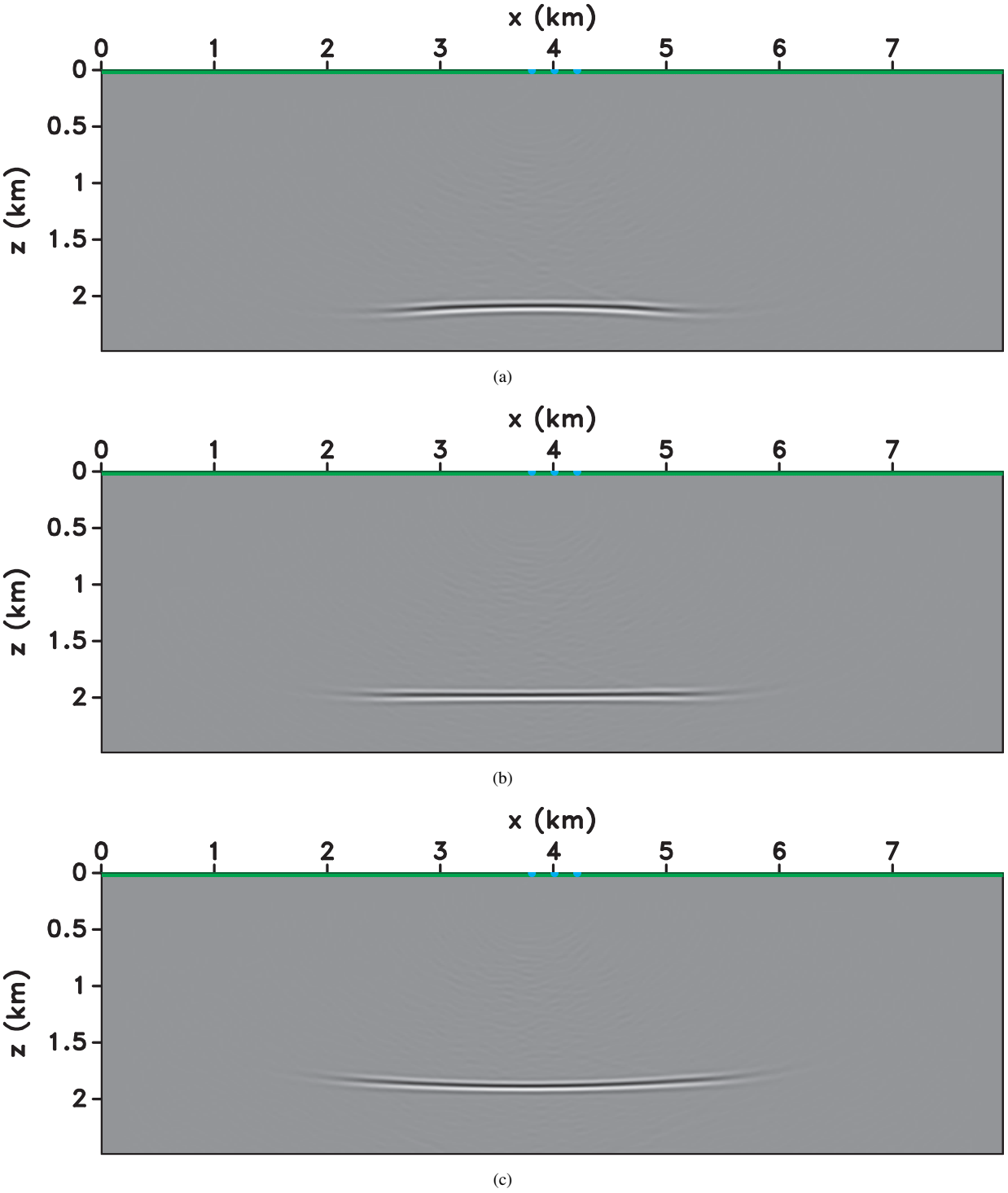


Figure 6. We consider the simple model in Figure 1. We migrate the data with a slowness model that is too low (a), correct (b), and too high (c). The slowness squared error is constant across the entire model and equal to 10% of the background. The correct slowness value is 0.5 s/km.

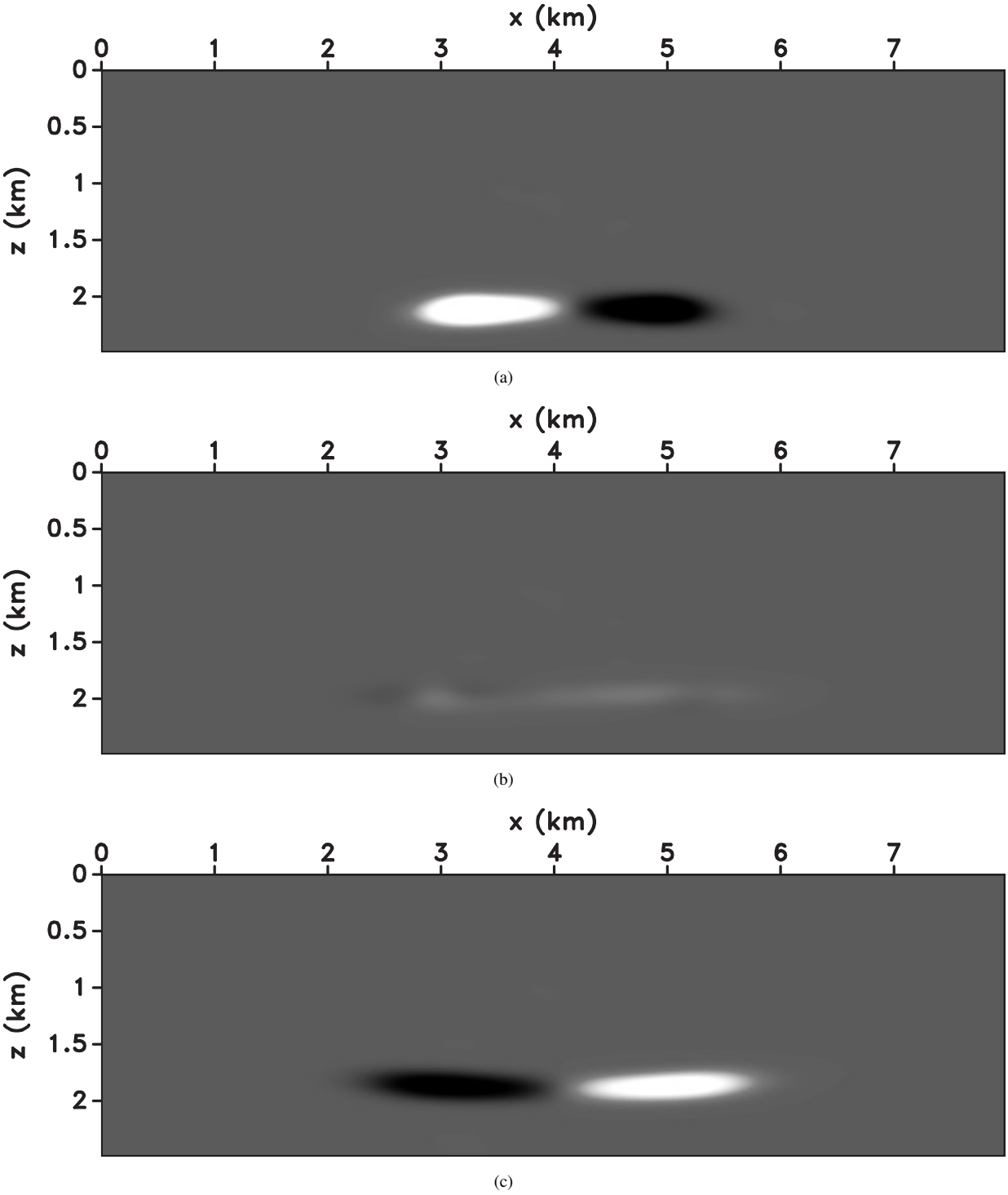


Figure 7. For the three cases in Figure 6, we compute the residuals according to Equation 2. The sign of the residual indicates the relative displacement of one image with respect to the other.

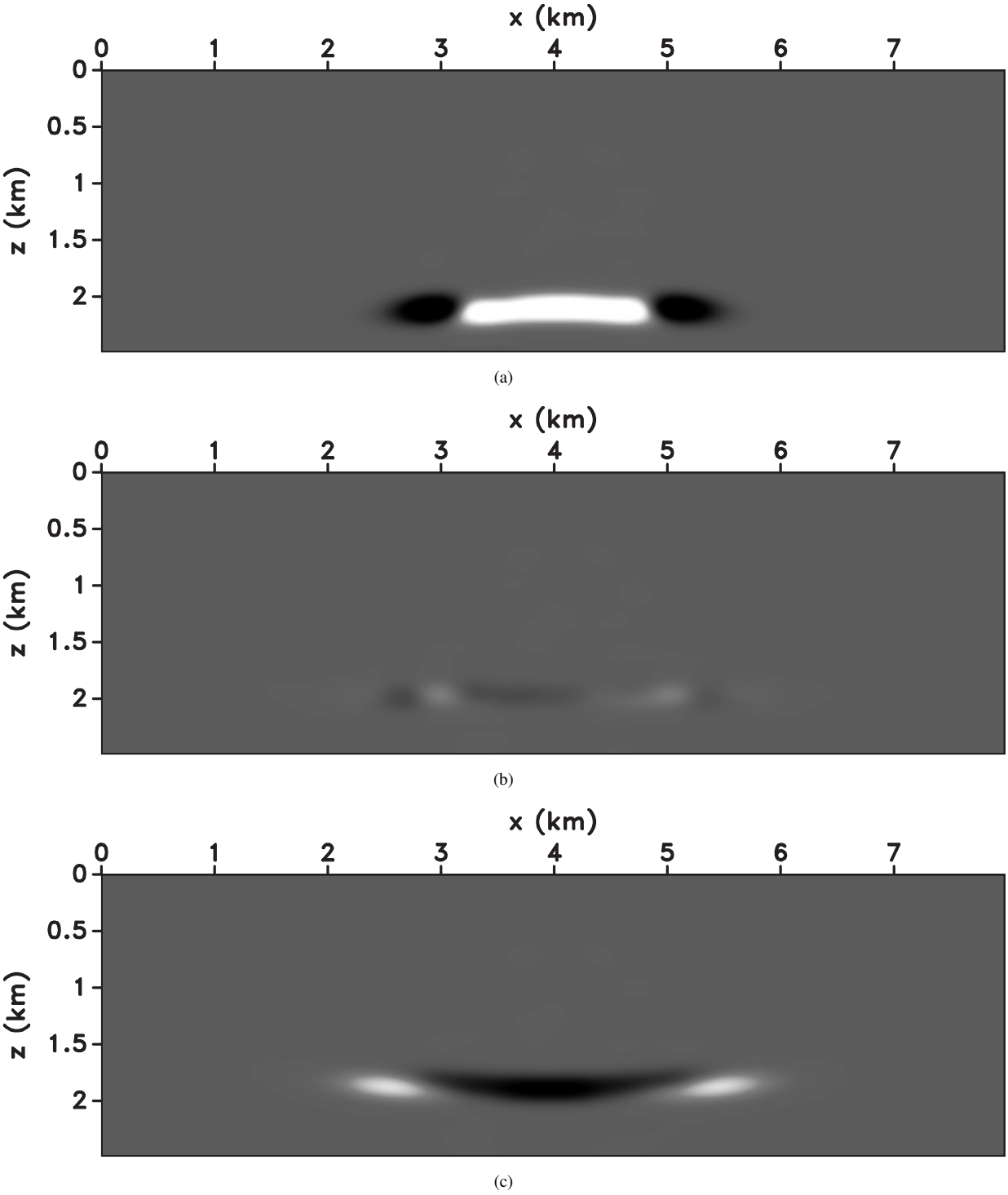


Figure 8. The difference between the residuals from adjacent experiments give us an estimate of the curvature of the shot-domain common-image gathers. This information correlates well with the velocity error in the model.

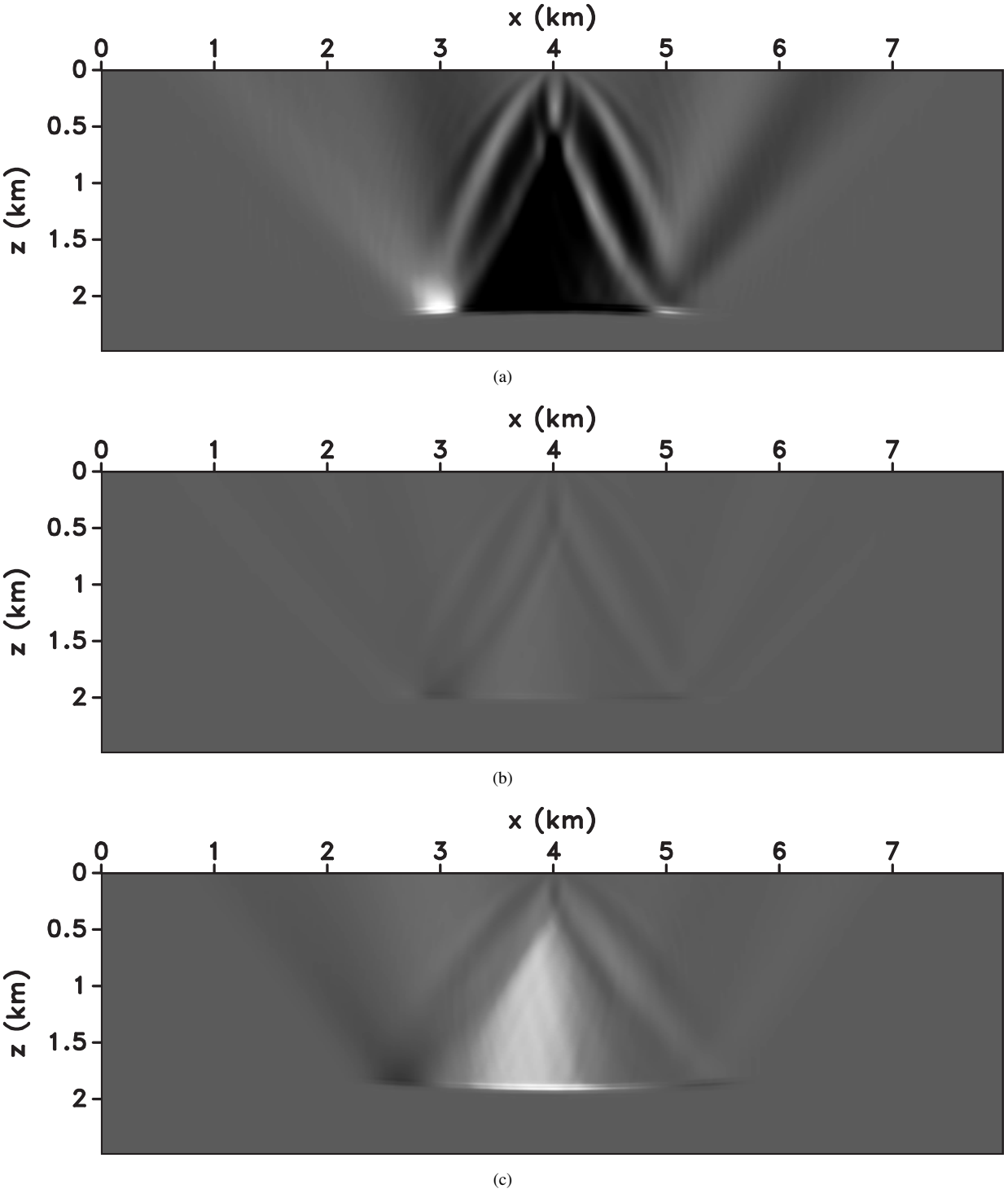


Figure 9. Gradient of the correlation objective function for a slowness model that is too low (a), correct (b), and too high (c). The slowness squared error is constant everywhere and equal to 10% of the background. The distance between the two shots is 60 m. We consider three shots and computed the gradient using the wavefields of the central experiment.

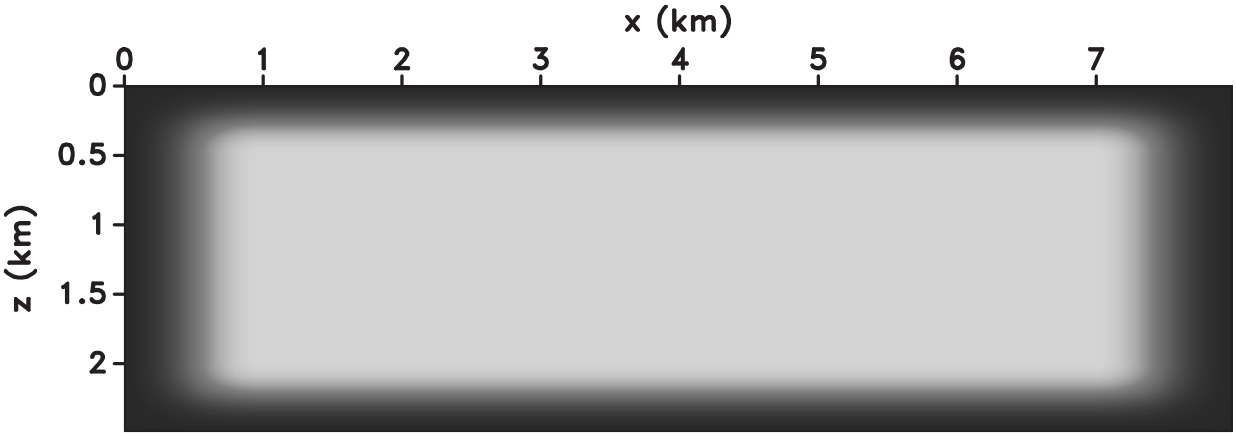


Figure 10. Slowness model used to generate the full-acoustic data. Absorbing boundary conditions have been applied at all sides of the model. The lighter shades of grey indicate higher slowness.

flector (Figure 12(a)). The migrated image obtained with the final model (Figure 12(b)) is shifted upward as a result of the higher slowness value. The common-image gathers in Figure 12(c) show the focusing obtained with the reconstructed model. The far left and far right edges of the reflector show residual curvature, which is due to the lack of illumination and of constraints from the data. The lack of illumination leads to a higher value of slowness in the central part of the model: the inversion compensates the lower slowness on the sides with a higher slowness in the middle. Observe the rapid decrease of the objective function as the gathers are flattened (Figure 13).

3.2 Synthetic laterally heterogeneous model

We use the synthetic heterogeneous model in Figure 14 to test our inversion algorithm. The model has a few layers with different dips and a sincline structure with an inversion in the otherwise decreasing slowness trend. We generate full-acoustic (i.e., non-Born) data with absorbing boundary conditions (no free-surface multiples) and random Gaussian noise with a signal-to-noise ratio equal to 10. We model 40 shots evenly spaced 200 m apart and receivers at every grid point on the surface. We heavily smooth the initial model in Figure 14 and obtain the initial slowness model for migration (Figure 15(a)). Figure 15(b) shows the stack of the migrated images obtained from the initial model; if we compare the correct model in Figure 14 and the migrated image we can observe the mispositioning of the reflecting interfaces. We construct 10 shot-domain common image gathers to assess the overall focusing of the image. These gathers are simply the juxtaposition of the migrated images at fixed lateral positions (from $x = 0$ km to $x = 9$ km every 1 km). Figure 15(c) shows such gathers for the initial model: the position of the reflectors changes as a function of experiment (the images do not satisfy the semblance principle) and the model is thus not accurate. After 15 waveform tomography steps we recover the model in Figure 16(a). The model is still quite smooth but the migrated image shows a noticeable improvement in the position of the reflectors (Figure 16(b)), and the shot-domain common-image gathers in Figure 16(c) clearly indicate better focusing. Moreover, the sign of the reflectors is now consistent with the actual slowness contrasts in the exact model. We impose conformity between the model update and the layers in the image by means of a structure-oriented smoothing operator, which steers the gradient of the objective function by smoothing along the reflector slope. This approach is conceptually similar to the sparse inversion proposed by Ma et al. (2010). Figure 17 shows the evolution of the objective function with iterations. Notice the monotonic decrease in the value of the residual and the flattening when the algorithm converges. The inversion also shows a rapid convergence in the very first iterations of steepest-descent: the structure-oriented smoothing forces the gradient to conform to the geometry of the layers and speeds up convergence to an acceptable model. Despite the simplicity of the model, the inaccuracy of the wave kinematics and the absence of overturning energy (diving waves) make the data-domain techniques (such as full waveform inversion) ineffective in this particular case.

3.3 Marmousi model

We set up a more complex inversion test using the Marmousi model. We model the data using a finite-difference single-scattering (i.e., Born) modeling code and absorbing boundary conditions on the 4 sides of the model. We simulate 78 shots, the spacing between the shots is 0.08 km, and the first source is at $x = 0.96$ km; receivers are placed at every grid point at $z = 0$ km. Conflicting dips are not yet handled correctly by our methodology because they make the definition of the penalty operators ambiguous. For this reason, we restrict the inversion to the first 1.5 km in depth, where the reflectors are more coherent.

The initial model is a heavily smoothed version of the correct Marmousi slowness model. Figure 18 shows the ini-

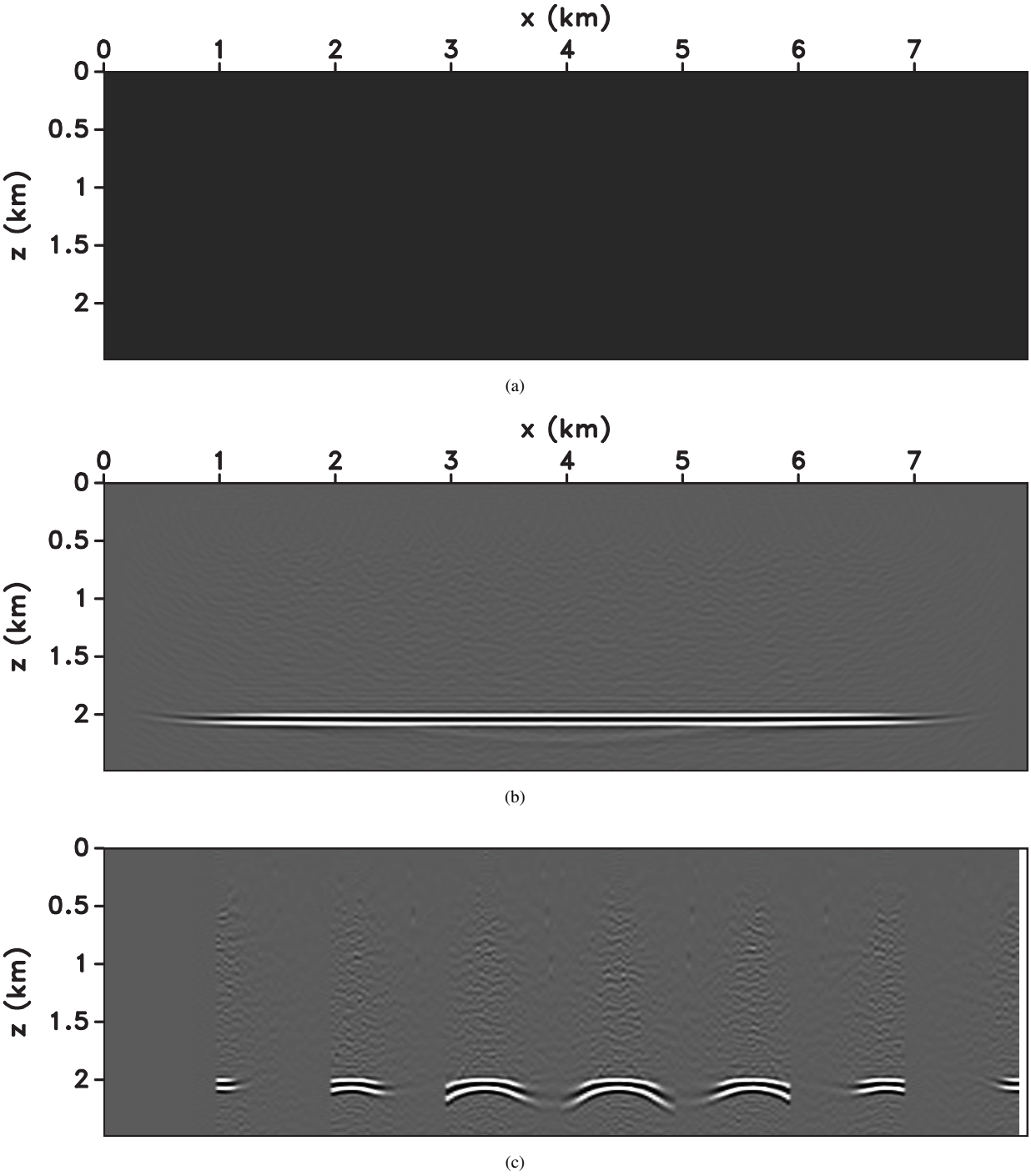


Figure 11. Initial slowness model (a), associated migrated image (b), and shot-domain common-image gathers (c). Observe the curvature of the gathers: different experiments image the reflector at different depths.

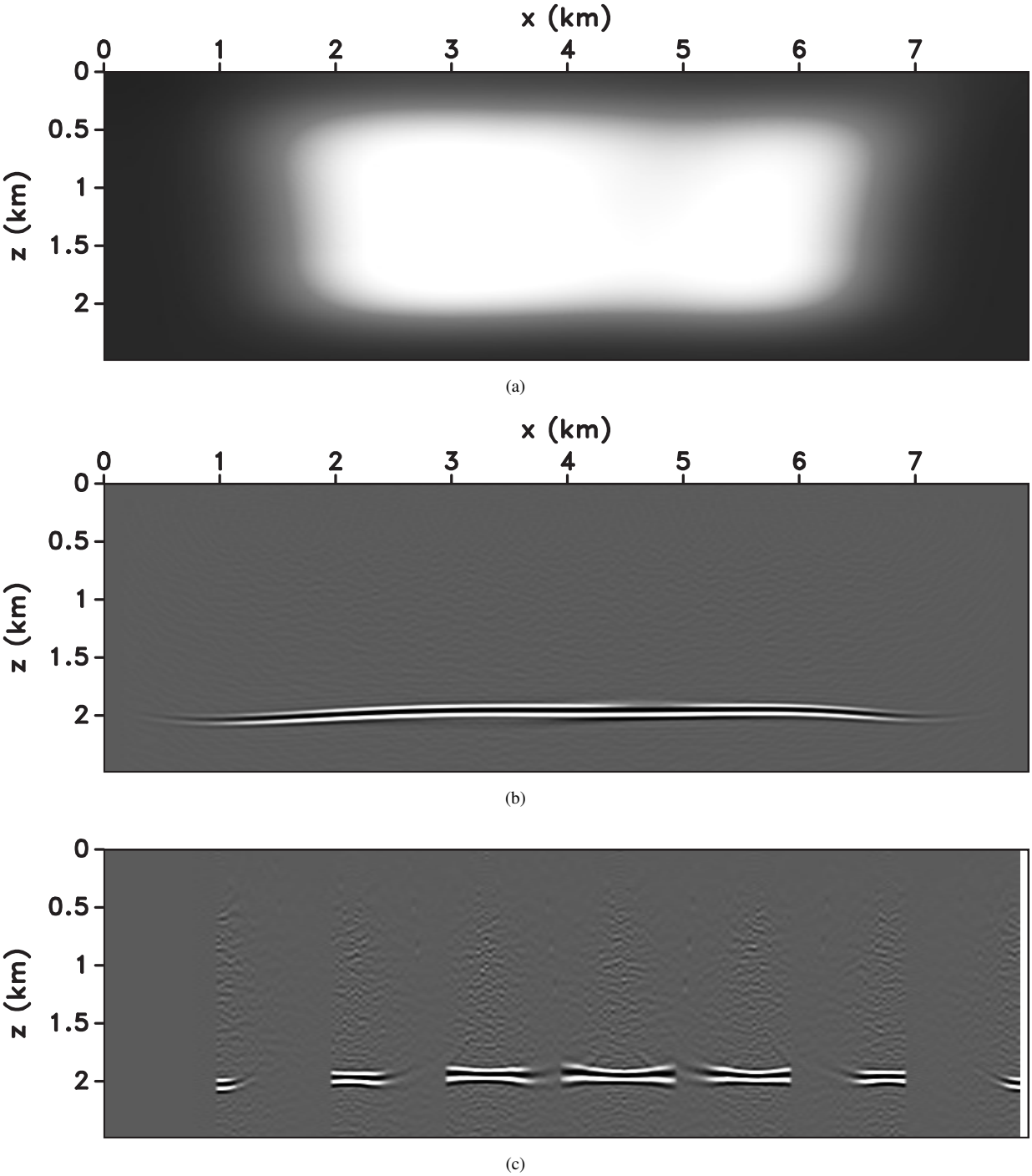


Figure 12. Reconstructed slowness model (a), final migrated image (b), and shot-domain common-image gathers (c). The inversion involves 5 iterations of steepest descent. Observe the horizontal gathers that indicate the invariance of the image with respect to the shot position. The area of the model that is well illuminated is smaller than the actual extent of the perturbation and the inverted model has a higher value of slowness than the correct medium.

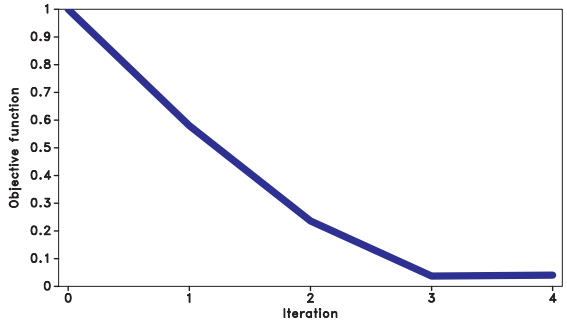


Figure 13. The objective function decreases as a function of iteration and flattens out when the common-image gathers are horizontal.

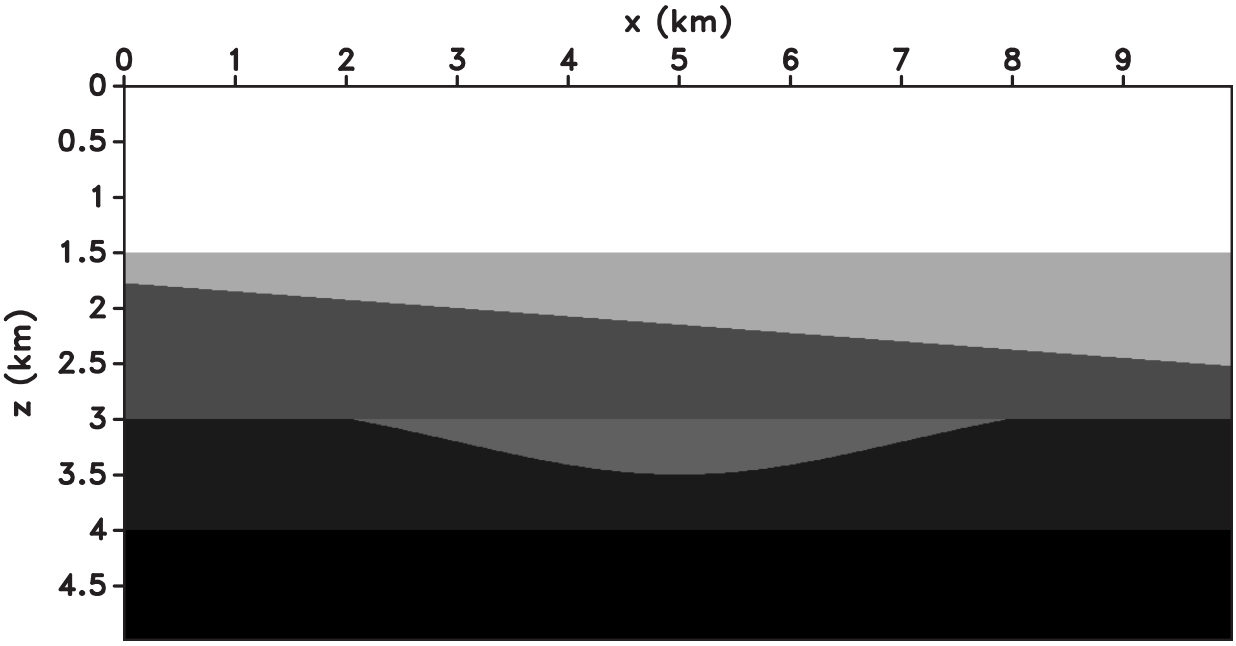


Figure 14. Slowness model used to generate the full-acoustic data. Absorbing boundary conditions have been applied at all sides of the model. The lighter shades of grey indicate faster layers. The top layer has slowness of water (0.66 s/km).

tial slowness model, the associated migrated image, and shot-domain common-image gathers extracted every 1 km from $x = 0$ km. Because of the regularity of the geologic structures in the shallow part of the model, the migrated image is not severely distorted, nonetheless the reflectors are mispositioned and the faults are not correctly focused. The shot-domain common-image gathers (Figure 18(c)) show variation of the depth of imaged reflectors as function of experiment, which indicates model inaccuracy.

We run 14 iterations of steepest-descent and obtain the results in Figure 19. The slowness model is corrected by the inversion procedure and the imaged interfaces are moved toward the correct position. The fault planes are better focused and we reduce the interface crossings at the unconformities ($x = 3$ km, $x = 4$ km and $x = 5$ km). The shot-domain common-image gather in Figure **??** show flatter horizontal events that indicate a more accurate kinematic model, and additional events are focused at the bottom of the model (at about $z = 3$ km). Observe also that the common-image gathers in poorly or non illuminated parts of the model contribute mainly migration noise and thus cannot be used to assess the correctness of the velocity model.

In Figure 20, the objective function shows a constant decrease. The residuals are computed over the entire image but we use only the points in the first 1.5 km in depth to update the velocity model. For complex areas with conflicting dips and complicated geologic features (where there is no clear definition of the dip field), a more sophisticated design of the penalty operator is key to obtaining meaningful residuals and a reliable gradient. The eigenvalues and eigenvectors of the gradient square tensors (van Vliet and Verbeek, 1995) can be used

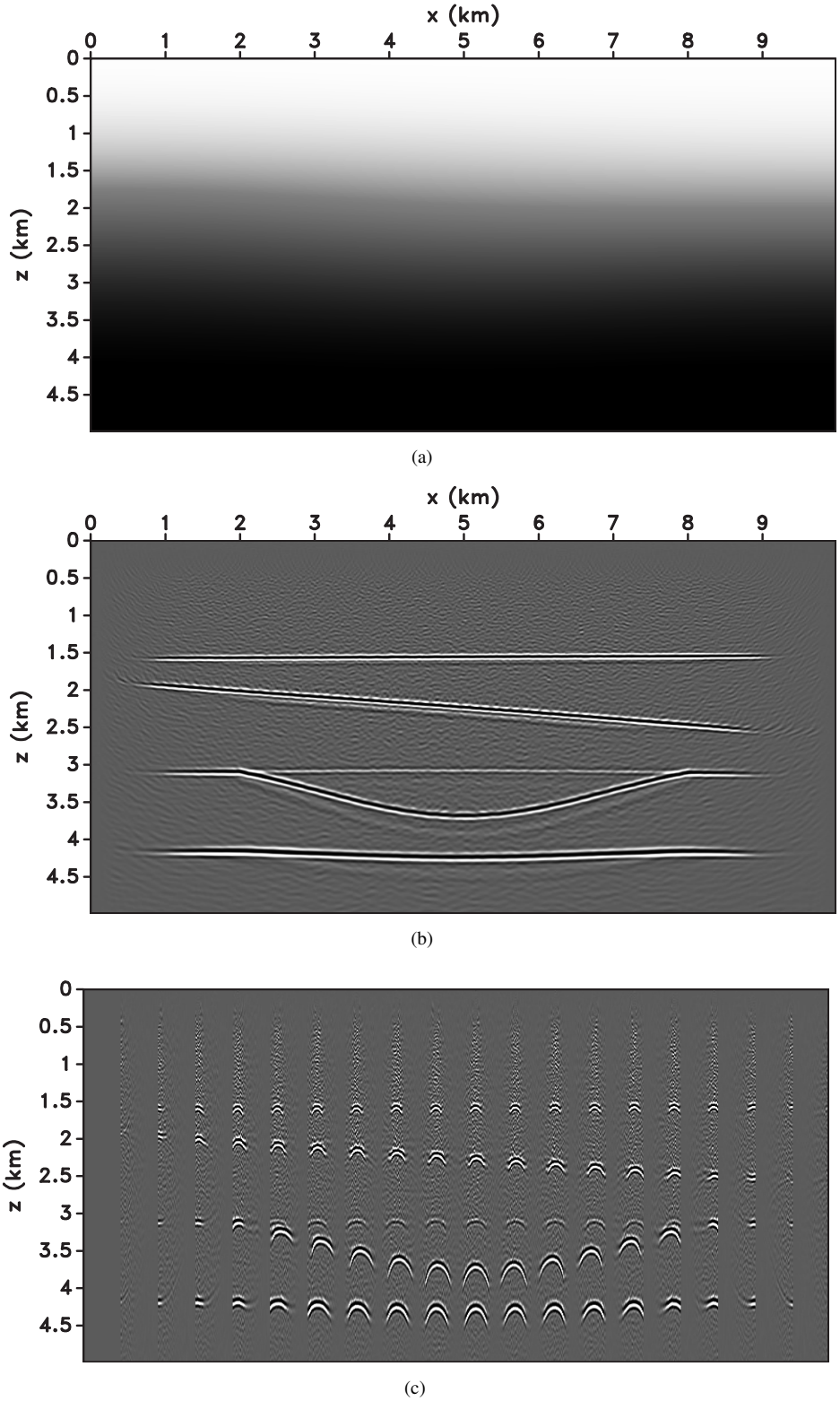
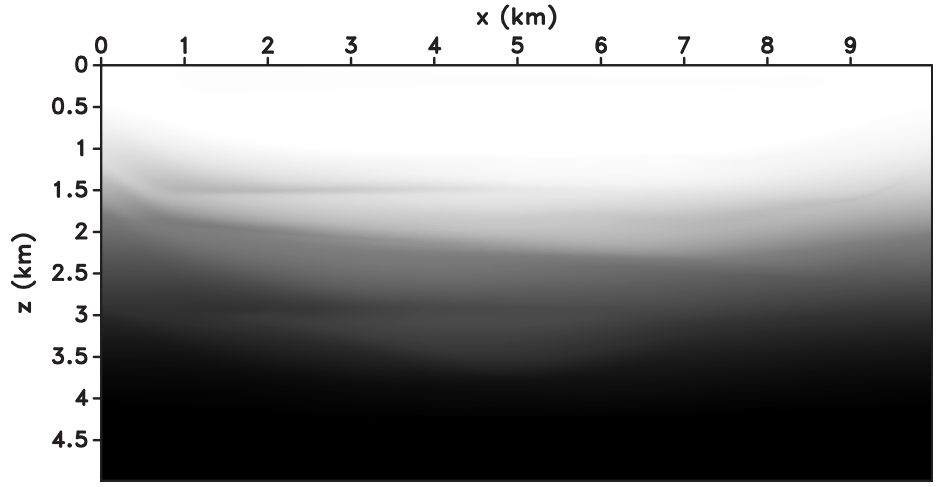
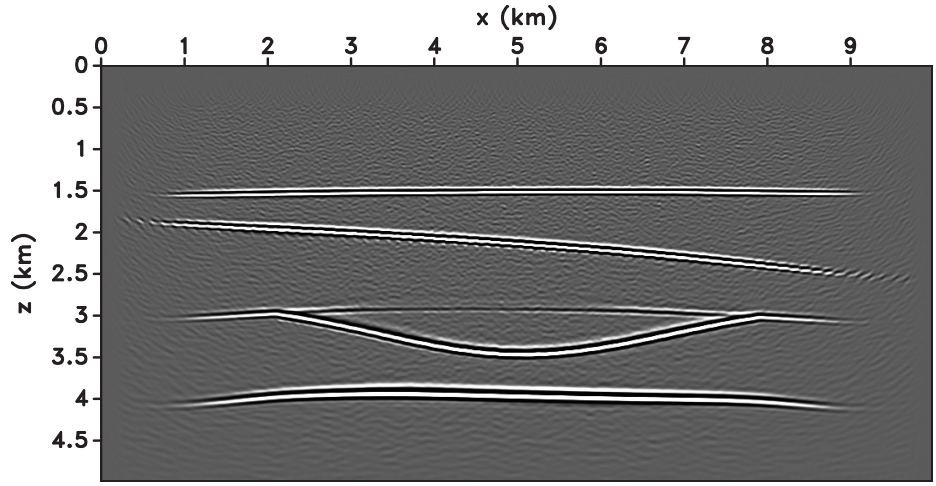


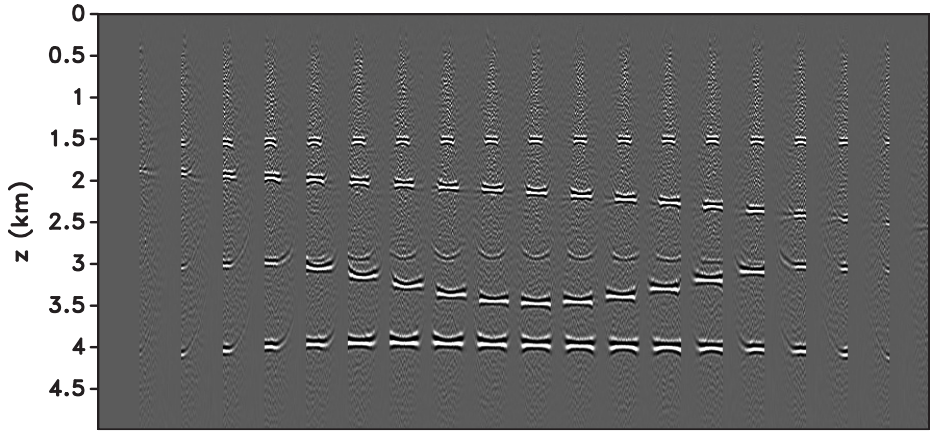
Figure 15. Initial slowness model (a), associated migrated image (b), and shot-domain common-image gathers. The reflectors are severely defocused and mispositioned. Observe the curvature of the common-image gathers that shows how different experiments image the reflectors at different depths.



(a)



(b)



(c)

Figure 16. Slowness model after 15 iterations of waveform inversion (a), migrated image (b), and shot-domain common-image gathers (c). Observe the improved focusing and positioning of the reflectors. The common-image gathers show that the images are invariant in the shot direction and the model is thus kinematically correct. Observe the profile of the various layers in the reconstructed model.

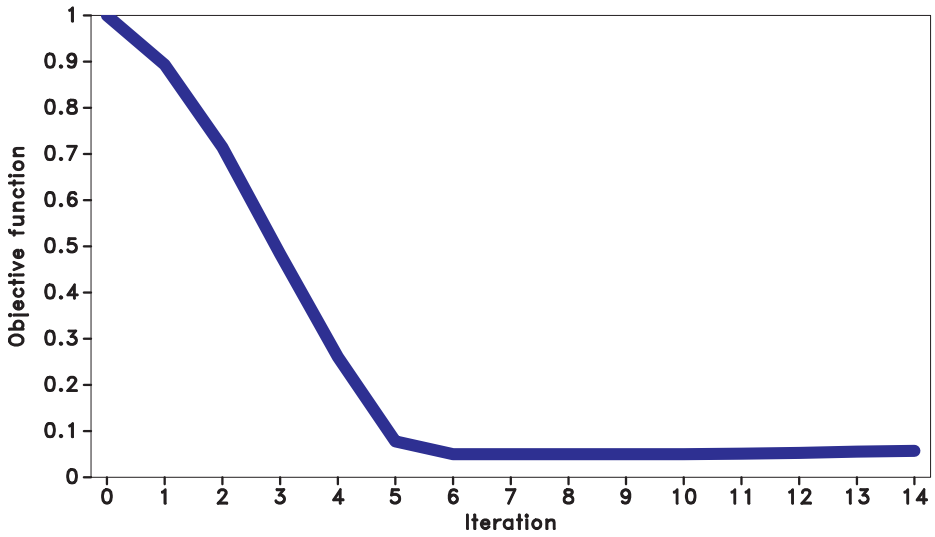


Figure 17. Evolution of the objective function with iterations. Observe the sharp decrease of the objective function and the rapid convergence to a kinematically accurate model.

to define ellipses that, in turn, may offer a structure oriented criterion for the definition of the penalty operators.

4 DISCUSSION

A complete image of the subsurface is the superposition of partial images from individual experiments. The semblance principle (Al-Yahya, 1989) is one common criterion for assessing the correctness of the velocity model used for imaging the survey: when the velocity model is correct, all shots locate reflectors at the same position, i.e., the image is invariant along the experiment axis. Several shots are needed to evaluate a velocity error at a single point in space.

We can evaluate the invariance along the experiment axis by computing the energy of the first derivative in that dimension. The first derivative acts as a penalty operator by highlighting and enhancing deviations from the horizontal direction along the shot axis. The stack of the energy of the first derivative is the differential semblance operator applied directly in the shot domain (Symes, 1991; Plessix, 2006).

Here, we explore an alternative statement of the semblance principle: when the velocity model is correct, images from different shots constructively interfere and build up the image perpendicular to the structural dip or, equivalently, parallel to the reflector slope, at every point in the image space. The structural dip is a commonly extracted attribute and can be linked to the image itself by means of the gradient square tensors (van Vliet and Verbeek, 1995). Unfortunately, the relationship between wavefields and the dip field is nonlinear: the dip vector represents the eigenvector associated with the largest eigenvalue of the gradient square tensor. The integration of information about the dip variation with respect to model perturbation is not straightforward, and further research is needed to develop an efficient method to exploit this information for velocity analysis purposes. Nonetheless, we are able to measure the semblance of two images through appropriately penalized local correlations of pairs of images. If the velocity model is correct, the maximum of the local correlation is along the reflector at every point in the image; if the model is incorrect, the maximum deviates from the reflector slope. The penalty operator is space-dependent and annihilates the correlation panels orthogonal to the reflector dip. Because of the dependency on the velocity model, we measure the dip field at each tomographic iteration; the estimation can be carried out efficiently using gradient square tensors (van Vliet and Verbeek, 1995; Hale, 2007a).

The correlation objective function is superior to the image difference in many respects. First, the difference between two images depends on image amplitudes that change as a function of the shot position and cannot be matched point-wise as in the standard implementation of full-waveform inversion in the data space. The amplitude patterns affect the residual in the image space and effectively contribute to the adjoint source calculation, even if the velocity model is correct and the gradient of the objective function is zero (Mulder and Kroode, 2002). By penalizing local correlations, we reduce the dependence of the objective function on amplitudes, thus increasing the robustness and reducing the systematic bias caused by the amplitude differences between the images. A downside of the correlation operator is the loss of spatial resolution. Correlation in the spatial domain is equivalent to multiplication in the dual frequency/wavenumber domain; for signals with both finite spatial support and bandwidth, the multiplication of the spectra decreases the bandwidth, i.e., increases the width of the signal in the spatial domain. To increase the resolution and accuracy of the evaluation of the relative shifts between images, deconvolution is a viable improvement over correlation. If we assume that two images from nearby experiments are

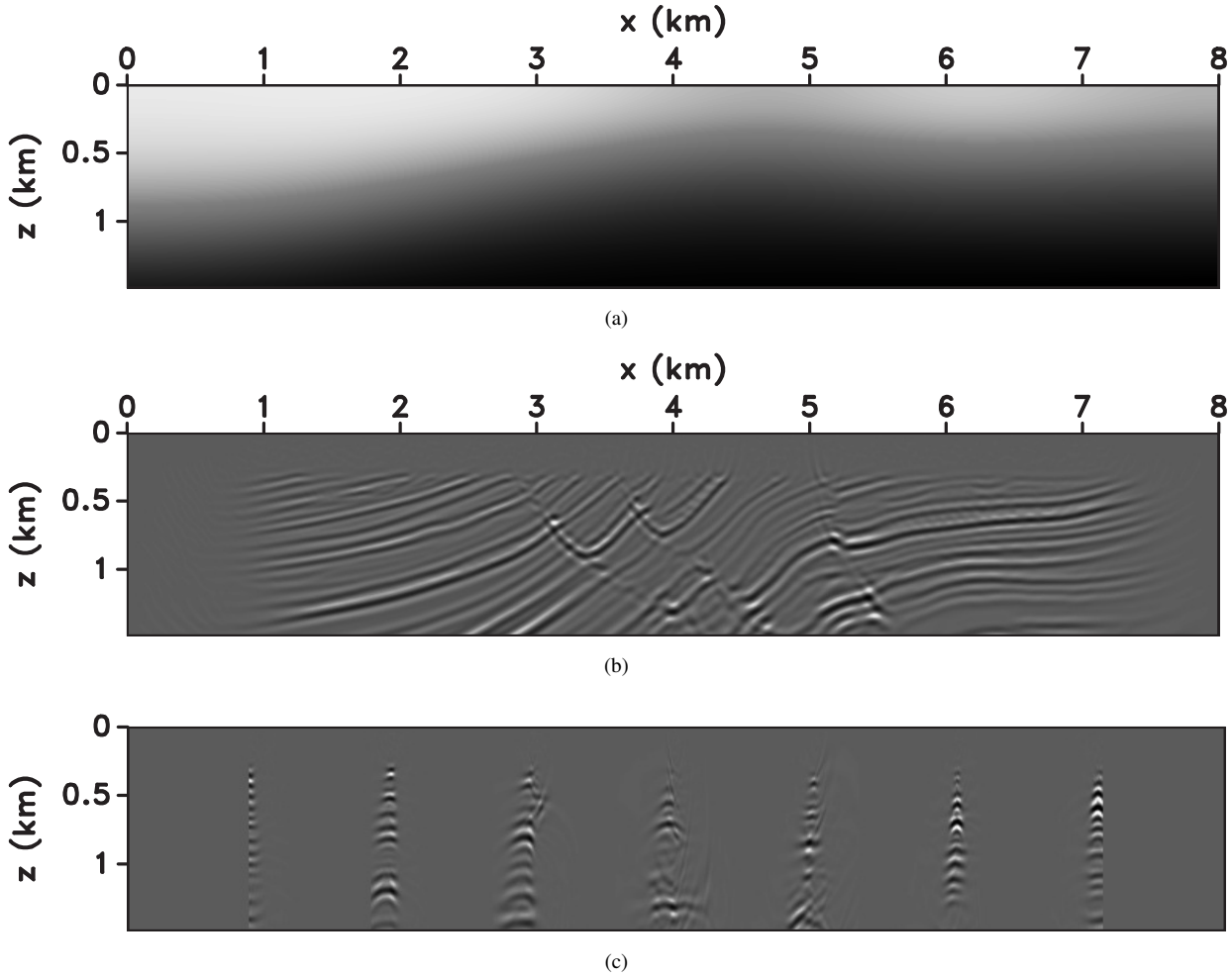


Figure 18. Initial slowness model (a), initial migrated images (b), and shot-domain common-image gathers (c). The initial model is obtained by heavily smoothing and scaling by 0.9 the correct model.

linked by a simple spatial shift, a local deconvolution approach would ideally produce a bandlimited spatial delta function indicating the direction of apparent displacement. On the other hand, since deconvolution amplifies the noise in the data (because of the division in the frequency domain), additional care is necessary to stabilize the result.

A thorough study of the effect of the shot distance on the computed gradient would be beneficial in relation to phenomena such as cycle skipping, which hinders many velocity analysis strategies (especially in the data domain but not exclusively). The image domain is intrinsically less prone to cycle skipping problems; nevertheless the question remains about what happens when two images illuminate rather different areas of the subsurface and the local correlations cannot be used to estimate a reliable and meaningful shift between the two images. From our current knowledge and understanding, the shots must be close enough to provide comparable images of the subsurface and avoid cycle skipping in the image domain. The sampling of the shot position and the illumination pattern in the subsurface can also create a scenario in which the two images do not overlap in certain local windows.

Our method is based on locally coherent events, such as locally smooth reflectors. Conflicting dips, fault planes, and areas where the definition of a reflection plane is ambiguous represent open problems because in these areas we cannot define penalty operators. In this respect, plane-wave migration may represent a valid solution because of the implicit spatial filtering of the image: each plane-wave reconstructs a particular subset of the dips in the model, thus reducing the ambiguity in defining the penalty operators.

In contrast with other techniques in the image space (Bishop et al., 1985; Biondi and Symes, 2004; Lambaré et al., 2004; Xie and Yang, 2008; Yang and Sava, 2010, 2011), we do not need to compute image gathers (or extended images). A further cost-saving factor is given by the selection of points on the imaged reflectors (Yang and Sava, 2010) and away from

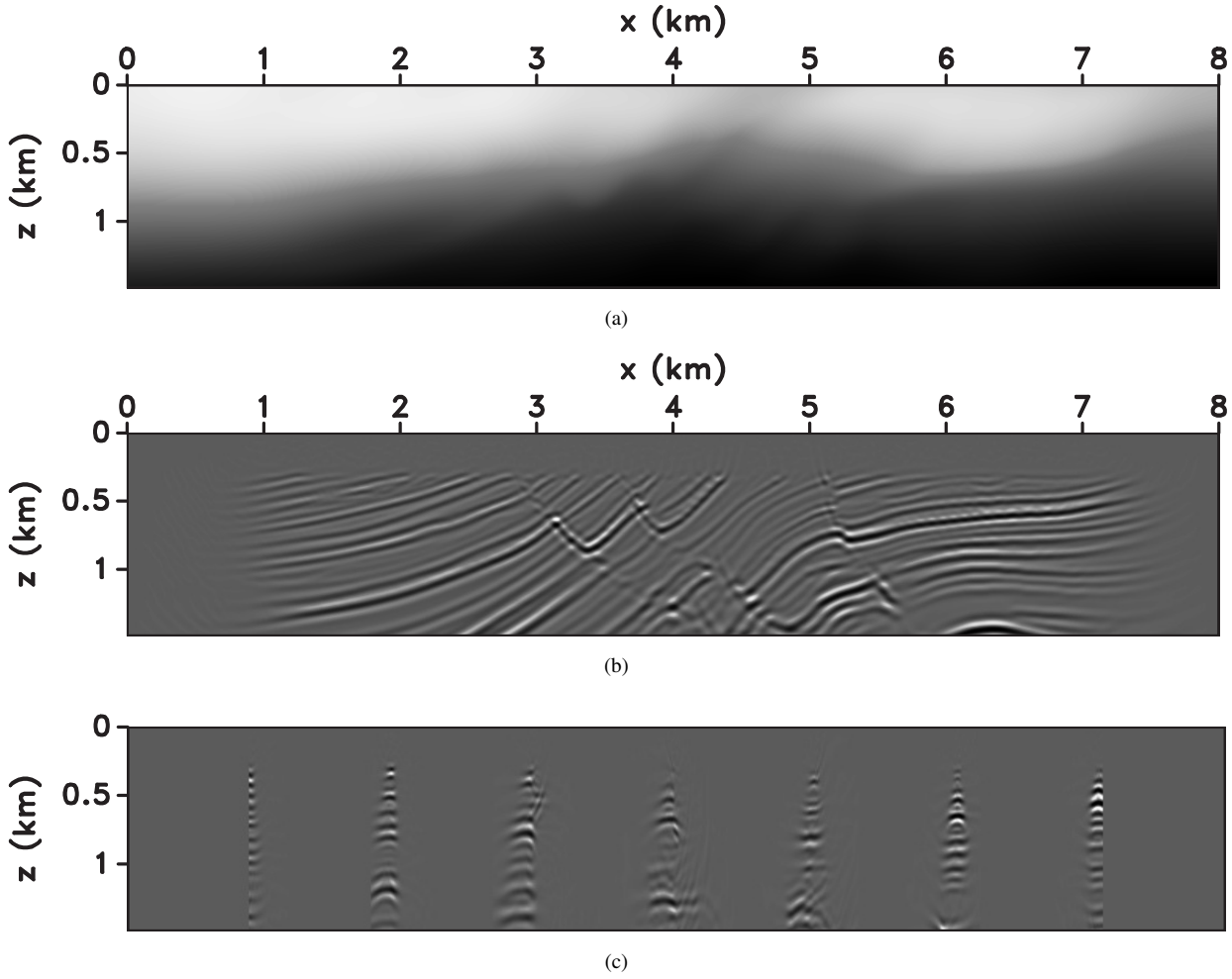


Figure 19. Updated model (a), migrated image (b), and shot-domain common-image gathers (c) after 14 iterations of steepest descent.

complex areas (pinch-outs or areas with conflicting dips). The computation of the local correlation is carried out at every image point using the efficient method developed by Hale (2006).

The ability to extract velocity information from pairs of experiments adds a new degree of freedom for implementing a model update procedure. Here, we indicate two possible strategies. We can simultaneously include all the shots in the survey in the definition of the objective function or we can proceed iteratively and update the velocity model using the information obtained from a single pair of experiments before moving to the next pair. Nearby shots probe similar portions of the model and provide comparable images; we can use the information extracted from an initial group to update the model used for imaging a second group of experiments. Since in any migration velocity analysis scheme we have to image the entire survey at least once, the iterative update of the model over shots becomes cost-effective if we actually reduce the number of global migrations of the entire dataset.

In 2D, we can easily define an order for the experiments; in the general 3D scenario, we have one extra degree of freedom. The definition of the objective function remains the same but a given shot can have more than 2 neighboring experiments (the acquisition is defined on a bidimensional grid). We can generalize the concept of correlation using the semblance functional (Taner and Koehler, 1969), which is nonlinear in the input signals and makes the computation of the gradient more involved. We can also separately analyze each pair of experiments given a reference model. Further research and numerical tests are needed to assess efficiency and robustness of one strategy over the other.

5 CONCLUSION

We develop an objective function for migration velocity analysis in the shot-image domain. Our methodology is based on local image correlations and a reformulation of the semblance principle involving only small groups of migrated images. The optimization criterion is the minimization of the apparent shift

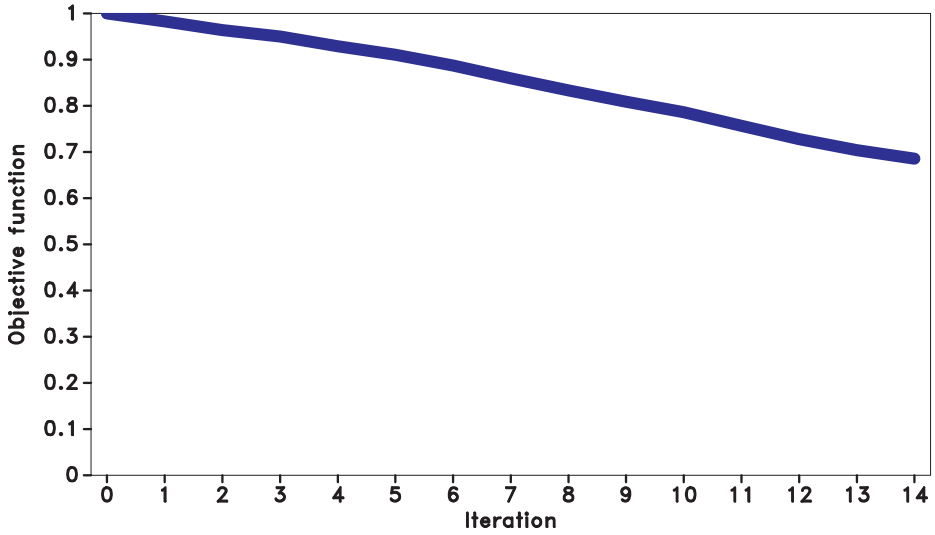


Figure 20. Evolution of the objective function versus the number of iterations

between the migrated images and naturally leads to a differential semblance optimization problem, which is iteratively solved using a gradient-based method. The linearity of the operators in the objective function makes the computation of the gradient practical using the adjoint-state method. Our approach is full-wave because we do not rely on the linearization of the wave extrapolation procedure with respect to the model parameters to construct a migration velocity analysis operator. We iteratively solve a nonlinear problem with a gradient-based algorithm and the simulation of the wavefields is fully nonlinear with respect to the model parameters. We test our method using a few synthetic examples of increasing complexity.

The structural features in the migrated image and the variation of these features as function of experiment supply valuable information about the model parameters. Here, we show that it is possible, without constructing common-image gathers , to extract information about the model from a very limited number of experiments using a warping relationship between migrated images. Despite the fact that we are neglecting the variation of the image dip field with respect to the model parameters, we are nevertheless able to correct anomalies in the model.

6 ACKNOWLEDGMENTS

We would like to thank eni E&P for the permission to publish this work and Nicola Bienati for the valuable conversations on tomography and inversion. The reproducible numerical examples in this paper use the Madagascar open-source software package freely available from http://www.reproducibility.org. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory.

7 A. COMPUTATION OF THE SOURCE AND RECEIVER ADJOINT SOURCE

Our objective function is defined as

$$\begin{aligned} \mathcal{J}(m) &= \frac{1}{2}\sum_i \|r_i(\mathbf{x})\|^2_{\mathbf{x}} \\ &= \frac{1}{2}\sum_i \left\| \int P(\mathbf{x},\boldsymbol{\lambda})\, c_i(\mathbf{x},\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right\|^2_{\mathbf{x}}, \end{aligned} \tag{A-1}$$

where $c_i(\mathbf{x},\boldsymbol{\lambda}) = \int_{w(\mathbf{x})} R_{i+1}\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_i\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) d\boldsymbol{\xi}$ is the local correlation of two images from the nearby shots i and $(i+1)$, and $P(\mathbf{x},\boldsymbol{\lambda})$ is an operator that annihilates the local correlation panel along the direction of the structure at every point in the image.

The adjoint-state method computes the gradient of the objective function with respect to the model parameters by solving two forward problems for each state variable of the system. This procedure turns out to be extremely computationally efficient since it avoids perturbing each model parameter and computing the resulting state variable perturbation (Lions, 1972; Plessix, 2006). The solution of the auxiliary forward problem requires different sources (adjoint sources) that are computed by differentiating the objective function with respect to the state variables.

We now derive the mathematical expression for the source and receiver adjoint sources. The computation of the derivative of a functional with respect to a function requires the definition of an inner product in the appropriate functional space; moreover, it requires the concept of the Fréchet differential (Vogel, 2002).

Let us denote the state variables of the problem, i.e., the source and receiver wavefields for the ith shot, by $u_{s,i}$ and $u_{r,i}$, respectively. We derive the expression of the adjoint source for the source wavefield. The derivation and expression

of the adjoint source for the receiver wavefield are analogous. First, we make explicit the dependency of the residual $r_i(\mathbf{x})$ (and thus of the objective function $\mathcal{J}$) on the state variable:

$$r_i(\mathbf{x}) = \int P(\mathbf{x},\boldsymbol{\lambda})\, w(\mathbf{x}-\boldsymbol{\xi})\, u_{s,i}\left(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2},t\right)\cdot \cdot\; u_{r,i}\left(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2},t\right) R_{i+1}\left(\boldsymbol{\xi}+\frac{\boldsymbol{\lambda}}{2}\right) dV, \quad \text{(A-2)}$$

where the integral is a triple integral over the volume element $dV = dtd\boldsymbol{\xi}d\boldsymbol{\lambda}$. Second, we compute the Fréchet differential of the objective function $\mathcal{J}(m)$ with respect to an arbitrary perturbation of the state variable $\delta u_{s,i}$. The objective function in equation A-1 is the sum over shots of energy of the residues in equation A-2 over the spatial domain. In order to simplify the derivation, we first write the differential of the objective function as a function of the differential of the residual using perturbation theory:

$$\mathcal{J}[u_{s,i}]+\delta\mathcal{J}[u_{s,i},\delta u_{s,i}] = \frac{1}{2}\int \left(r_{i-1}+\delta r_{i-1}[u_{s,i},\delta u_{s,i}]\right)^2 + \left(r_i+\delta r_i[u_{s,i},\delta u_{s,i}]\right)^2 d\mathbf{x}. \quad \text{(A-3)}$$

Expanding the squares in equation A-3 and considering only the terms that are linear in the wavefield perturbation, we obtain an expression for the Fréchet differential of the objective function:

$$\delta\mathcal{J}[u_{s,i},\delta u_{s,i}] = \int r_{i-1}\delta r_{i-1}[u_{s,i},\delta u_{s,i}]+r_i\delta r_i[u_{s,i},\delta u_{s,i}]d\mathbf{x}. \quad \text{(A-4)}$$

The square brackets in equation A-4 indicate the double dependency of the Fréchet differential on the state variable and its arbitrary perturbation.

Assuming $r_i(\mathbf{x})$ is Fréchet differentiable, we compute the Fréchet differential using the limit of the incremental ratio:

$$\delta r_i[u_{s,i},\delta u_{s,i}] = \lim_{h\to 0}\frac{r_i[u_{s,i}+h\delta u_{s,i}]-r_i[u_{s,i}]}{h}, \quad \text{(A-5)}$$

where h is a scalar, and $\delta u_{s,i}$ is an arbitrary perturbation of the state variable $u_{s,i}$ and belongs to the same functional space. The residual r_i is a linear functional of $u_{s,i}$, and computing the limit in equation A-5 we obtain

$$\delta r_i = \int P(\mathbf{x},\boldsymbol{\lambda})\, w(\mathbf{x}-\boldsymbol{\xi})\, \delta u_{s,i}\left(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2},t\right)\cdot \cdot\; u_{r,i}\left(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2},t\right) R_{i+1}\left(\boldsymbol{\xi}+\frac{\boldsymbol{\lambda}}{2}\right) dV. \quad \text{(A-6)}$$

We want to express equation A-6 as an inner product between an operator (the Fréchet derivative) and the state variable perturbation $\delta u_{s,i}$; we perform a change of variables to make the perturbation ofthe state variable $\delta u_{s,i}$ depend only on a single independent variable. We define two new variables $\boldsymbol{\eta} = \boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}$ and $\boldsymbol{\chi} = \boldsymbol{\lambda}$; the Jacobian of this transformation is trivial and equal to 1; then, in the new coordinate system, we have

$$\delta r_i = \int P(\mathbf{x},\boldsymbol{\chi})\, w\left(\mathbf{x}-\boldsymbol{\eta}-\frac{\boldsymbol{\chi}}{2}\right)\delta u_{s,i}(\boldsymbol{\eta},t)\cdot \cdot\; u_{r,i}(\boldsymbol{\eta},t)\, R_j(\boldsymbol{\eta}+\boldsymbol{\chi})\, dV', \quad \text{(A-7)}$$

where the triple integral is over the new volume element $dV' = dtd\boldsymbol{\eta}d\boldsymbol{\chi}$. From equation A-7, we find the expression of the gradient of the residual r_i with respect to the state variable $u_{s,i}$:

$$\nabla_{u_{s,i}} r_i = \int P(\mathbf{x},\boldsymbol{\chi})\, w\left(\mathbf{x}-\boldsymbol{\eta}-\frac{\boldsymbol{\chi}}{2}\right) u_{r,i}(\boldsymbol{\eta},t)\cdot \cdot\; R_{i+1}(\boldsymbol{\eta}+\boldsymbol{\chi})\, d\boldsymbol{\chi}. \quad \text{(A-8)}$$

Combining equation A-8 into A-4, we can finally write the equation for the gradient of the objective function $\mathcal{J}$ with respect to the state variable $u_{s,i}$ as

$$\nabla_{u_{s,i}}\mathcal{J} = u_{r,i}(\boldsymbol{\eta},t)\left[r_{i-1}P(\mathbf{x},\boldsymbol{\lambda}) * R_{i-1}(\mathbf{x}) + r_i P(\mathbf{x},\boldsymbol{\lambda}) \star R_{i+1}(\mathbf{x})\right], \quad \text{(A-9)}$$

where

$$P(\mathbf{x},\boldsymbol{\lambda}) * R_{i-1}(\mathbf{x}) = \int w\left(\mathbf{x}-\boldsymbol{\eta}+\frac{\boldsymbol{\chi}}{2}\right)\cdot \cdot\; P(\mathbf{x},\boldsymbol{\chi})\, R_{i-1}(\boldsymbol{\eta}-\boldsymbol{\chi})\, d\boldsymbol{\chi}$$

and

$$P(\mathbf{x},\boldsymbol{\lambda}) \star R_{i+1}(\mathbf{x}) = \int w\left(\mathbf{x}-\boldsymbol{\eta}-\frac{\boldsymbol{\chi}}{2}\right)\cdot \cdot\; P(\mathbf{x},\boldsymbol{\chi})\, R_{i+1}(\boldsymbol{\eta}+\boldsymbol{\chi})\, d\boldsymbol{\chi}$$

are the local convolution and local correlation of the penalty operator with the image over the local window w, respectively. The derivation for the receiver-side adjoint source is totally analogous and leads to the formula

$$\nabla_{u_{r,i}}\mathcal{J} = u_{s,i}(\boldsymbol{\eta},t)\left[r_{i-1}P(\mathbf{x},\boldsymbol{\lambda}) * R_{i-1}(\mathbf{x}) + r_i P(\mathbf{x},\boldsymbol{\lambda}) \star R_{i+1}(\mathbf{x})\right]. \quad \text{(A-10)}$$

Equations A-9 and A-10 are the expressions for the adjoint sources in equations 10 and 11 in the paper.

REFERENCES

Al-Yahya, K., 1989, Velocity analysis by iterative profile migration: Geophysics, **54**, P. 718–729.

Albertin, U., P. Sava, J. Etgen, and M. Maharramov, 2006, Adjoint wave-equation velocity analysis: Presented at the 74th Ann. Internat. Mtg., SEG.

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

Biondi, B., and P. Sava, 1999, Wave-equation migration velocity analysis: Presented at the 69th Ann. Internat. Mtg., SEG.

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

Bishop, T. N., K. P. Bube, R. T. Cutler, R. T. Langan, P. L. Love, J. R. Resnick, R. T. Shuey, D. A. Spindler, and H. W. Wyld, 1985, Tomographic determination of velocity and depth in laterally varying media: Geophysics, **50**, 903–923.

Carcione, J., 2007, Wave fields in real media wave propagation in anisotropic, anelastic, porous and electromagnetic media: Elsevier. Handbook of Geophysical Exploration.

Chang, W., P. Carrion, and G. Beylkin, 1987, Wavefront sets of solutions to linearised inverse scattering problems: Inverse Problems, **3**, 683 – 690.

Chavent, G., and C. A. Jacewitz, 1995, Determination of background velocities by multiple migration fitting: Geophysics, **60**, 476–490.

Claerbout, J. F., 1985, Imaging the earth's interior: Blackwell Publishing.

Claerbout, J. F., and F. Muir, 1973, Robust modeling with erratic data: Geophysics, **38**, 826–844.

de Vries, D., and A. J. Berkhout, 1984, Velocity analysis based on minimum entropy: Geophysics, **49**, 2132–2142.

Farmer, P. A., I. F. Jones, H. Zhou, R. I. Bloor, and M. C. Goodwin, 2006, Application of reverse time migration to complex imaging problems: First Break, **24**, 65–73.

Faye, J.-P., and J.-P. Jeannot, 1986, Prestack migration velocity analysis from focusing depth imaging: Presented at the 56th Ann. Internat. Mtg., Soc. of Expl. Geophys.

Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

Fowler, P., 1985, Migration velocity analysis by optimization: Technical Report 44, Stanford Exploration Project.

Gazdag, J., and P. Sguazzero, 1984, Migration of seismic data by phase shift plus interpolation: Geophysics, **49**.

Hale, D., 2006, An efficient method for computing local cross-correlations of multi-dimensional signals: Technical Report CWP-544, Center for Wave Phenomena, Colorado School of Mines.

———, 2007a, Local dip filtering with directional laplacian: Technical Report CWP-567, Center for Wave Phenomena, Colorado School of Mines.

———, 2007b, A method for estimating apparent displacement vectors from time-lapse seismic data: Technical Report CWP-566, Center for Wave Phenomena, Colorado School of Mines.

Kelly, S., J. R. Martinez, B. tsimelzon, and S. Crawley, 2010, A comparison of inversion results for two full-waveform methods that utilize the lowest frequencies in dual-sensor recordings: Presented at the 80th Ann. Internat. Mtg., Soc. of Expl. Geophys.

Lambaré, G., M. Alerini, R. Baina, and P. Podvin, 2004, Stereotomography: a semi-automatic approach for velocity macromodel estimation: Geophysical Prospecting, **52**, 671–681.

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

Liu, F., D. W. Hanson, N. D. Whitmore, R. S. Day, and R. H. Stolt, 2006, Toward a unified analysis for source plane wave migration: Geophysics, **71**, S129–S139.

Ma, Y., D. Hale, Z. Meng, and B. Gong, 2010, Full waveform inversion with image-guided gradient: Presented at the 80th Ann. Internat. Mtg., Soc. of Expl. Geophys.

McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: Geophysical Prospecting, **31**, 413–420.

Morton, S. A., and C. C. Ober, 1998, Faster shot-record migration using phase encoding: Presented at the 68th Ann. Internat. Mtg., SEG.

Mulder, W. A., and A. P. E. Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184 – 1191.

Perrone, F., and P. Sava, 2011, Wave-equation migration with dithered plane waves: Geophysical Prospecting.

Perrone, F., P. Sava, C. Andreoletti, and N. Bienati, 2012, Wave-equation migration velocity analysis by image warping: Technical Report CWP-708, Center for Wave Phenomena, Colorado School of Mines.

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

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

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

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.

Santosa, F., and W. W. Symes, 1989, An analysis of least-squares velocity inversion, *in* Geophysical Monograph no. 4: Soc. of Expl. Geophys.

Sava, P., and B. Biondi, 2004, Wave-equation migration velocity analysis. i. theory: Geophysical Prospecting, **52**, 593–606.

Sava, P., B. Biondi, and J. Etgen, 2005, Wave-equation migration velocity analysis by focusing diffractions and reflections: Geophysics, **70**, U19–U27.

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

Sava, P., and I. Vasconcelos, 2009, Extended common-image-point gathers for wave-equation migration: Presented at the 71th Ann. Internat. Mtg., European Association of Geoscientists and Engineers.

Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–VE59.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: the

next leap forward in imaging at valhall: First Break, **28**, 65–70.

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

Soubaras, R., 2006, Modulated-shot migration: 76th Ann. Internat. Mtg., Soc. of Expl. Geophys., 2430–2433.

Stoffa, P. L., J. T. Fokkema, R. M. de Luna Freire, and W. P. Kessiger, 1990, Split-step fourier migration: Geophysics, **55**.

Stoffa, P. L., M. K. Sen, R. K. Seifoullaev, R. C. Pestana, and J. T. Fokkema, 2006, Plane-wave depth migration: Geophysics, **71**, S261–S272.

Stolt, R. H., 1978, Migration by fourier transform: Geophysics, **43**.

Symes, W. W., 1991, Layered velocity inversion: A model problem from reflection seismology problem from reflection seismology: SIAM J. on Math. Analysis, **22**, pp. 680–716.

———, 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Taner, M. T., and F. Koehler, 1969, Velocity spectra - digital computer derivation and applications of velocity functions: Geophysics, **34**, 859 – 881.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 71–92.

———, 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

van Leeuwen, T., and W. A. Mulder, 2008, Velocity analysis based on data correlation: Geophysical Prospecting, **56**, 791 – 803.

van Vliet, L. J., and P. W. Verbeek, 1995, Estimators for orientation and anisotropy in digitalized images: Proceedings of the first annual conference of the Advanced School for Computing and Imaging, ASCI'95, 442–450.

Vogel, C., 2002, Computational methods for inverse problems: SIAM.

Whitmore, N., 1995, An imaging hierarchy for common-angle plane wave seismograms: PhD thesis, University of Tulsa.

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

Xie, X. B., and H. Yang, 2008, The finite-frequency sensitivity kernel for migration residual moveout and its applications in migration velocity analysis: Geophysics, **73**, S241–S249.

Yang, T., and P. Sava, 2010, Wave-equation migration velocity analysis with extended common-image-point gathers: Presented at the 80th Ann. Internat. Mtg., Soc. of Expl. Geophys.

———, 2011, Wave-equation migration velocity analysis with time-shift imaging: Geophysical Prospecting, **59**, 635–650.

Yilmaz, O., and R. Chambers, 1984, Migration velocity analysis by wave-field extrapolation: Geophysics, **49**, P. 1664–1674.

Zhang, Y., J. Sun, C.Notfors, S. H. Gray, L. Chernis, and J. Young, 2005, Delayed-shot 3d depth migration: Geophysics, **70**, E21–E28.