Image-guided full waveform inversion

Yong Ma[1], Dave Hale[1], Bin Gong[2] & Zhaobo (Joe) Meng[3]
[1] *Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA*
[2] *Seismic Technology, ConocoPhillips Company, Houston, TX 77252, USA*
[3] *Formerly Seismic Technology, ConocoPhillips Company, Houston, TX 77252, USA; presently In-Depth Geophysical, Houston, TX 77074, USA*

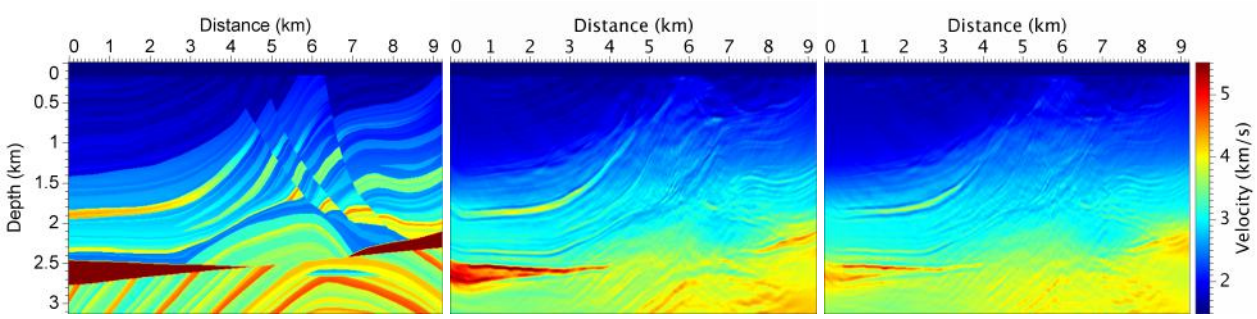


Figure 1. The true Marmousi II velocity model (a) and inverted models with (b) and without (c) the image-guided technique. Both inversions use highly smoothed Marmousi II velocity as the initial model; a 15Hz Ricker wavelet is used as the source for the 11 shots in the inversion.

ABSTRACT
Multiple problems, including high computational cost, spurious local minima, and solutions with no geologic sense, have prevented widespread application of full waveform inversion (FWI), especially FWI of seismic reflections. These problems are fundamentally related to a large number of model parameters and to the absence of low frequencies in recorded seismograms.
Instead of inverting for all the parameters in a dense model, image-guided full waveform inversion inverts for a sparse model space that contains far fewer parameters. We represent a model with a sparse set of values, and from these values, we use image-guided interpolation (IGI) and its adjoint operator to compute finely- and uniformly-sampled models that can fit recorded data in FWI. Because of this sparse representation, image-guided FWI updates more blocky models, and this blockiness in the model space mitigates the absence of low frequencies in recorded data. Moreover, IGI honors imaged structures, so image-guided FWI built in this way yields models that are geologically sensible.

Key words: image-guided, full waveform inversion, reduced model space

1 INTRODUCTION

Full waveform inversion (FWI) (Tarantola, 1984; Pratt, 1999) uses recorded seismic data $\mathbf{d}$ to estimate parameters of a subsurface model $\mathbf{m}$, by minimizing the difference between recorded data $\mathbf{d}$ and synthetic data $\mathbf{F}(\mathbf{m})$, where $\mathbf{F}$ is a forward operator that synthesizes data. All information in recorded seismic waveforms should, in principle, be taken into account in minimizing this difference. In other words, FWI comprehensively minimizes differences in traveltimes, amplitudes, converted waves, multiples, etc. between recorded and synthetic data. This all-or-nothing approach distinguishes FWI from other methods, such as traveltime tomography

(Stork, 1992; Woodward, 1992; Vasco and Majer, 1993; Zelt and Barton, 1998), which focuses on only traveltime differences.

FWI is attractive in its capability to estimate a subsurface model with generally higher resolution (Operto et al., 2004) than traveltime tomography and migration velocity analysis (MVA) (Yilmaz and Chambers, 1984; Sava and Biondi, 2004). Another advantage of FWI over traveltime tomography or MVA is that FWI can estimate multiple parameters: density (Forgues and Lambaré, 1997), attenuation (Liao and McMechan, 1996), elasticity (Shi et al., 2007), anisotropy (Barnes et al., 2008; Pratt et al., 2008), etc. Although FWI has a long history and promising benefits, two major obstacles – high computational cost and nonunique solutions – have prevented its widespread application in exploration seismology.

FWI requires large numbers of seismic wavefield simulations and reconstructions, and computational cost is proportional to the number of sources. FWI also requires multiple iterations to minimize data misfit, and computational cost is proportional to the number of required iterations. Therefore, various methods have been applied to reduce computational cost. One such method is to apply phase-encoding techniques (Krebs et al., 2009) that combine all shots together to form a simultaneous source. The cost of FWI using encoding techniques is thereby reduced by a factor ideally equal to the number of encoded shots divided by the number of recorded shots. To reduce the number of required iterations, one may use a sparse representation of a model space and reduce the number of model parameters. The wavelet transform is a representative technique used in inverse problems to reduce the number of parameters (Meng and Scales, 1996).

Because FWI is a typical underdetermined problem, many different models may yield synthetic data that match recorded data within a reasonable tolerance. This nonuniqueness problem is caused mainly by local minima in a data misfit function, and the presence of local minima is due to the fact that the forward operator $\mathbf{F}$ is generally a nonlinear function of the model $\mathbf{m}$. Strong nonlinearity in reflection FWI makes this local-minima problem more severe (Snieder et al., 1989). Cycle-skipping also causes nonunique solutions in FWI. Cycle-skipping occurs if the phase difference (time delay) between synthetic and recorded data is larger than half a period of the dominant wavelet. In practice, the cycle-skipping problem typically appears because it can be difficult to obtain an adequate initial model that is consistent with unrecorded low frequencies.

Both local-minima and cycle-skipping problems lead to models that poorly approximate the subsurface. To mitigate such problems, multiscale approaches (Bunks, 1995; Sirgue and Pratt, 2004; Boonyasiriwat et al., 2009) have been proposed. These methods recursively add higher-frequency details to models first computed from lower-frequency data. The fidelity of multiscale techniques depends fundamentally on the fidelity of low-frequency content in recorded data. In practice, the low frequencies required to bootstrap the multiscale approach may be unavailable. Other methods for addressing these problems have been proposed as well, e.g., inverting high-wavenumber and low-wavenumber components separately (Snieder et al., 1989; Hicks and Pratt, 2001).

We solve these problems in a different way. Following Meng (2009), who proposes to use subsurface dips to constrain the inversion, we investigate the image-guided gradient (Ma et al., 2010) to complement low frequencies that are usually unavailable in recorded data. In this paper, we propose image-guided sparse FWI, which aims to make FWI more efficient and more stable and to generate geologically sensible results. In image-guided FWI, we use image-guided interpolation (IGI) (Hale, 2009a) and its adjoint operator in order to apply structural constrains derived from migrated images. We first reformulate FWI in a sparse model space, by efficiently choosing sample points. We then solve the sparse FWI by using a modified image-guided conjugate-gradient method. This image-guided sparse FWI is tested on the Marmousi II model and with realistically high-frequency data.

2 SPARSE-MODEL FULL WAVEFORM INVERSION

Because the forward operator $\mathbf{F}$ has no inverse $\mathbf{F}^{-1}$ for almost any geophysical inverse problem, we cannot simply invert the model from the data using $\mathbf{m} = \mathbf{F}^{-1}(\mathbf{d})$. Instead, FWI is usually formulated as a least-squares optimization problem, in which we compute a model $\mathbf{m}$ that minimizes the data misfit function

$$E(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|^2 , \tag{1}$$

where $\|.\|$ denotes an L2 norm.

We begin with an initial model $\mathbf{m}_0$, which can be found using other inversion methods (e.g., traveltime tomography or migration velocity analysis); then we iteratively reduce the data misfit $E(\mathbf{m})$ by applying Newton-like methods. In the i^{th} iteration, the Taylor series expansion of equation 1 about the model $\mathbf{m}_i$ is

$$E(\mathbf{m}_i + \delta\mathbf{m}_i) = E(\mathbf{m}_i) + \delta\mathbf{m}_i^T \mathbf{g}_i + \frac{1}{2}\delta\mathbf{m}_i^T \mathbf{H}_i \delta\mathbf{m}_i + \dots , \tag{2}$$

where $\mathbf{g}_i \equiv \mathbf{g}(\mathbf{m}_i) = \dfrac{\partial E}{\partial \mathbf{m}_i}$ denotes the gradient of the data misfit E evaluated at $\mathbf{m}_i$ and $\mathbf{H}_i$ denotes the Hessian matrix comprised of the 2nd partial derivatives of $E(\mathbf{m})$, again evaluated at $\mathbf{m}_i$. If we ignore any term higher than 2nd order in equation 2, this Taylor series approximation is quadratic in the model perturbation

$\delta\mathbf{m}_i$, and we can minimize the data misfit $E(\mathbf{m}_i)$ by solving a set of linear equations:

$$\mathbf{H}_i\delta\mathbf{m}_i = -\mathbf{g}_i \tag{3}$$

with a Newton method (Pratt et al., 1998) solution

$$\delta\mathbf{m}_i = -\mathbf{H}_i^{-1}\mathbf{g}_i \ . \tag{4}$$

Unfortunately, the large size of the Hessian matrix $\mathbf{H}_i$, which is directly determined by the number of parameters, prevents the application of Newton or Newton-like methods in realistic cases. Moreover, FWI is usually ill-posed in practice due to a typically large condition number of the Hessian matrix (Tarantola, 2005). A large condition number tends to appear especially when an inverse problem has a large number of model parameters in $\mathbf{m}$. If the change of a model parameter in $\mathbf{m}$ does not cause significant change in the data misfit function $E(\mathbf{m})$, the Hessian matrix $\mathbf{H}$ will have a small (or nearly zero) eigenvalue. As a consequence, the condition number of the Hessian matrix will be large enough that the gradient-descent method (Vigh and Starr, 2008) used to solve an FWI problem converges slowly.

2.1 Inverse problem in sparse model space

Inspired by the intuitive relationship between the convergence rate and the number of model parameters, we pose an FWI problem that inverts for only a few model parameters, to which the data misfit function is sensitive. We then reduce the condition number of the Hessian matrix and thereby the number of required iterations.

Following an approach similar to subspace methods (Kennett et al., 1988; Oldenburg et al., 1993) and the point collocation scheme (Pratt et al., 1998, Appendix A), we reconstruct a finely- and uniformly-sampled (dense) model $\mathbf{m}$ from a sparse model $\mathbf{s}$ that contains a much smaller number of model parameters than does the dense model $\mathbf{m}$:

$$\mathbf{m} = \mathbf{R}\mathbf{s} \ , \tag{5}$$

where $\mathbf{R}$ denotes a linear operator that interpolates model parameters from the sparse model to the dense model.

Differentiating both sides of equation 5, we have

$$\delta\mathbf{m} = \mathbf{R}\delta\mathbf{s} \ . \tag{6}$$

Then, substituting equation 6 into equation 4, we can reformulate the inverse problem posed in equation 4, with respect to a smaller number of model parameters in the sparse model $\mathbf{s}$, as

$$\mathbf{H}_i\mathbf{R}\delta\mathbf{s}_i = -\mathbf{g}_i \ . \tag{7}$$

Because $\mathbf{R}$ is not a square matrix, equation 7 is different from conventional preconditioning (Benzi, 2002).

2.2 Solution in sparse model

We cannot solve equation 7 with a solution like $\delta\mathbf{s}_i = -\left(\mathbf{H}_i\mathbf{R}\right)^{-1}\mathbf{g}_i$ in the sparse domain $\mathbf{s}$ because equation 7 is overdetermined; i.e., there are more equations than parameters. Therefore, we modify equation 7 to be

$$\mathbf{R}^T\mathbf{H}_i\mathbf{R}\delta\mathbf{s}_i = -\mathbf{R}^T\mathbf{g}_i \ , \tag{8}$$

and thereby obtain a least-squares solution for equation 7 in the sparse domain $\mathbf{s}$:

$$\delta\mathbf{s}_i = -\left(\mathbf{R}^T\mathbf{H}_i\mathbf{R}\right)^{-1}\mathbf{R}^T\mathbf{g}_i \ , \tag{9}$$

where $\mathbf{R}^T$ is the adjoint operator of $\mathbf{R}$. This adjoint operator projects model parameters from the dense model $\mathbf{m}$ to the sparse model $\mathbf{s}$. Unfortunately, equation 9 is usually hard to implement because in practice the Hessian matrix is extremely expensive to compute and store.

Alternatively, the model update $\delta\mathbf{s}$ can be iteratively approximated by replacing the inverse of the projected Hessian matrix $\left(\mathbf{R}^T\mathbf{H}_i\mathbf{R}\right)$ with a scalar step length α_i:

$$\mathbf{s}_{i+1} = \mathbf{s}_i - \alpha_i\mathbf{h}_i^{\mathbf{s}} \ , \tag{10}$$

where the conjugate direction $\mathbf{h}_i^{\mathbf{s}}$ (Ma et al., 2010) is determined by

$$\begin{aligned}
\mathbf{h}_0^{\mathbf{s}} &= \mathbf{R}^T\mathbf{g}_0 \ , \\
\beta_i &= \frac{\left(\mathbf{R}^T\mathbf{g}_i\right)^T\left(\mathbf{R}^T\mathbf{g}_i - \mathbf{R}^T\mathbf{g}_{i-1}\right)}{\left(\mathbf{R}^T\mathbf{g}_{i-1}\right)^T\mathbf{R}^T\mathbf{g}_{i-1}} \ , \\
\mathbf{h}_i^{\mathbf{s}} &= \mathbf{R}^T\mathbf{g}_i + \beta_i\mathbf{h}_{i-1}^{\mathbf{s}} \ .
\end{aligned} \tag{11}$$

In equation 10, the step length α_i can be found with a line-search method (Nocedal and Wright, 2000). Equation 11 employs $\mathbf{R}^T\mathbf{g}_i$ instead of $\mathbf{g}_i$, implying that equation 10 provides a solution for the FWI problem in the sparse domain $\mathbf{s}$. Because of fewer model parameters involved, the projected Hessian matrix $\left(\mathbf{R}^T\mathbf{H}_i\mathbf{R}\right)$ can become better-conditioned and thus equation 10 can require fewer iterations to converge to a solution model $\mathbf{s}$.

In reality, we need a dense update $\delta\mathbf{m}$ to compute synthetic data $\mathbf{F}(\mathbf{m})$ and to fit recorded data $\mathbf{d}$. For this reason, we apply the linear operator $\mathbf{R}$ to both sides of equation 10 and thereby interpolate the sparse model update $\delta\mathbf{s}_i$ to obtain the dense model update $\delta\mathbf{m}_i$:

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \alpha_i\mathbf{h}_i^{\mathbf{m}} \ , \tag{12}$$

where we compute the search direction $\mathbf{h}_i^{\mathbf{m}}$ by projecting the sparse conjugate direction $\mathbf{h}_i^{\mathbf{s}}$ to the dense domain:

$$\begin{aligned}
\mathbf{h}_0^{\mathbf{m}} &= \mathbf{R}\mathbf{h}_0^{\mathbf{s}} = \mathbf{R}\mathbf{R}^T\mathbf{g}_0 \ , \\
\beta_i &= \frac{\left(\mathbf{R}^T\mathbf{g}_i\right)^T\left(\mathbf{R}^T\mathbf{g}_i - \mathbf{R}^T\mathbf{g}_{i-1}\right)}{\left(\mathbf{R}^T\mathbf{g}_{i-1}\right)^T\mathbf{R}^T\mathbf{g}_{i-1}} \ , \\
\mathbf{h}_i^{\mathbf{m}} &= \mathbf{R}\mathbf{R}^T\mathbf{g}_i + \beta_i\mathbf{h}_{i-1}^{\mathbf{m}} \ .
\end{aligned} \tag{13}$$

Equation 13 provides a solution $\mathbf{m}$ for FWI in the dense model space with the advantages derived from solving for $\mathbf{s}$ in the sparse model space.

2.3 Implementation of sparse-model FWI

An implementation of sparse FWI based on conjugate gradients consists of four steps performed iteratively, beginning with an initial model $\mathbf{m}_0$:

(i) compute the data difference $\mathbf{d} - \mathbf{F}(\mathbf{m}_i)$;
(ii) compute the gradient $\mathbf{g}_i$, $\mathbf{R}^T\mathbf{g}_i$, $\mathbf{R}\mathbf{R}^T\mathbf{g}_i$, and the update direction $\mathbf{h}_i^{\mathbf{m}}$;
(iii) search for a step length α_i;
(iv) update the model with $\mathbf{m}_{i+1} = \mathbf{m}_i + \delta\mathbf{m}_i$.

3 CHOICE OF R

The operator $\mathbf{R}$ can take different forms, including Fourier transform, wavelet transform, cubic splines, etc. In this paper, we implement $\mathbf{R}$ with image-guided interpolation (IGI) (Hale, 2009a) specifically because IGI accounts for imaged subsurface structure. IGI uses structure tensors (van Vliet and Verbeek, 1995; Fehmers and Höcker, 2003) to guide interpolation of a few spatially scattered values, thereby making the interpolant conform to structure in images.

3.1 Image-guided interpolation

The input of IGI is a set of scattered data, a set

$$\mathcal{F} = \{f_1, f_2, ..., f_K\}$$

of K known sample values $f_k \in \mathbb{R}$ that correspond to a set

$$\chi = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K\}$$

of K known sample points $\mathbf{x}_k \in \mathbb{R}^n$. Combining these two sets forms a space (e.g., our sparse model $\mathbf{s}$), in which $\mathcal{F}$ and χ denote sample values and coordinates, respectively. The result of the interpolation is a function $q(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, such that $q(\mathbf{x}_k) = f_k$. Here, the dense model $\mathbf{m}$ consists of all interpolation points $\mathbf{x}$ and values $q(\mathbf{x})$.

Image-guided interpolation is a two-step process (Hale, 2009a):

$$\mathbf{R} = \mathbf{Q}\mathbf{P}\ , \tag{14}$$

where $\mathbf{P}$ and $\mathbf{Q}$ denote nearest neighbor interpolation and blending of nearest neighbors, respectively. Examples of applying IGI can be found in Hale (2009a). Appendix A describes in more detail the operators $\mathbf{P}$ and $\mathbf{Q}$. Intuitively, we can describe these two operators as:

> (i) $\mathbf{P}$: *scatters* values f_k from nearest sample points $\mathbf{x}_k$ to obtain uniformly sampled interpolated values;
>
> (ii) $\mathbf{Q}$: *smooths* the uniformly sampled nearest neighbor interpolant.

3.2 Adjoint image-guided interpolation

Because $\mathbf{Q}^T = \mathbf{Q}$, we can configure the adjoint of image-guided interpolation as

$$\mathbf{R}^T = \mathbf{P}^T\mathbf{Q}^T = \mathbf{P}^T\mathbf{Q}\ . \tag{15}$$

The adjoint operator $\mathbf{R}^T$ is again a two-step process:

> (i) $\mathbf{Q}^T = \mathbf{Q}$: *smooths* uniformly sampled values;
>
> (ii) $\mathbf{P}^T$: *gathers* uniformly sampled values from nearest neighbors $\mathbf{x}$ to the scattered sample points $\mathbf{x}_k$.

In equation 13, we sequentially apply the IGI operator $\mathbf{R}$ and its adjoint operator $\mathbf{R}^T$ to produce the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_i$. The adjoint operator $\mathbf{R}^T$ first gathers information to sample points from nearest neighborhoods that conform to structural features in the gradient $\mathbf{g}_i$. The IGI operator $\mathbf{R}$ then scatters the gathered information back to the same neighborhoods. Through this *gather-scatter* process, the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_i$ generates low wavenumbers in models $\mathbf{m}$.

4 STRUCTURALLY CONSTRAINED SAMPLE SELECTION

A set of properly chosen locations for scattered samples is essential for implementing image-guided sparse FWI. The samples should be representative, such that image-guided interpolation can reconstruct an accurate dense model $\mathbf{m}$ from a sparse model $\mathbf{s}$. In general, we must:

- *locate samples between reflectors*, so that the gather-scatter process ($\mathbf{R}\mathbf{R}^T$) can produce low wavenumbers between reflectors. We should especially avoid putting samples on reflectors.
- *locate samples along geological features*. To reduce redundancy, we should place fewer samples along structural features than across features. Moreover, to better honor structural features, we should put more samples in structurally complex areas than in simple areas.

Figures 2a and 2b show examples of uniform sample and pseudo-random sample selections, respectively. The uniform sample selection and the pseudo-random selection are easy to implement, however, neither of

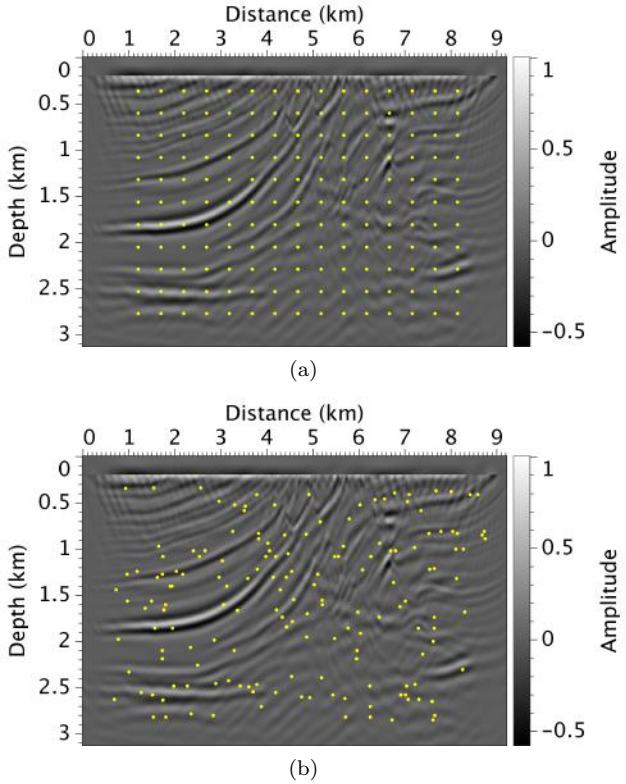


Figure 2. A uniform sample selection (a) and a pseudo-random sample selection (b). A total of 165 samples are chosen from the densely sampled model space.

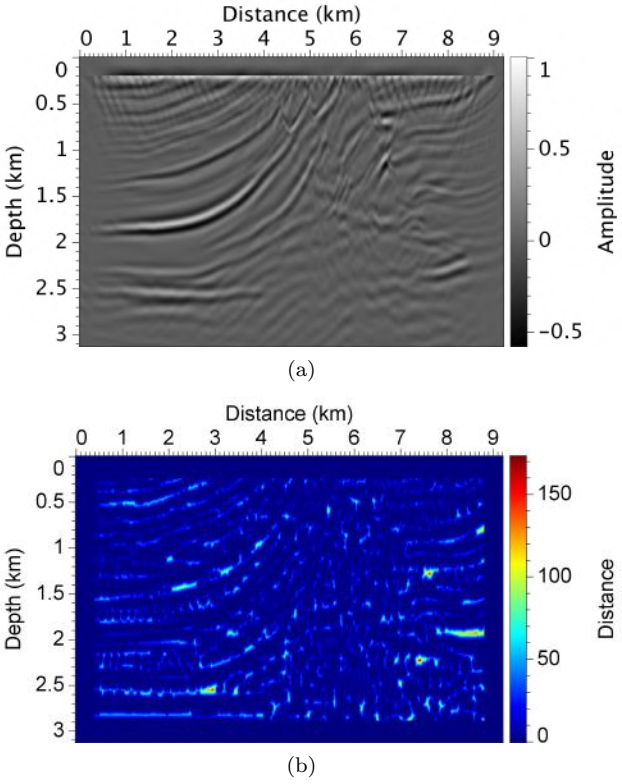


Figure 3. (a) A migrated image $I(\mathbf{x})$ and (b) a distance map $d(\mathbf{x})$ of the migrated image. Zero distance indicates reflectors, and nonzero values indicate areas between reflectors.

them can satisfy both of the above criteria. Given a fixed number of samples, the uniform samples fail to follow structural structures; many samples lie on reflectors, as shown in Figure 2a, and those samples are undesirable. The pseudo-random selection has the same problem and creates samples that are too close, as shown in Figure 2b. In this paper, we investigate and then employ a structurally constrained sample selection scheme, which satisfies both criteria.

4.1 Distance transform

A migrated image $I(\mathbf{x})$ (Figure 3a) can be considered as a combination of two parts: reflectors and areas delimited by reflectors. To choose samples between reflectors, we must first distinguish reflectors from the areas between them. For this purpose, we use a distance transform (DT) (Fabbri et al., 2008), which was first introduced in computer vision and image processing.

A distance transform computes for each pixel of an image the smallest distance to a given subset pixels. This given subset is a region of interest in the DT. Appendix B describes the distance transform in detail. For our sample selection problem, we treat reflectors as regions of interest and compute the distance from each sample in the migrated image to the nearest reflectors.

Figure 3b shows a distance map $d(\mathbf{x})$, which illustrates, for a migrated image, how far each point is from the nearest reflector. As we can see from Figure 3b, a large part of the distance map is nearly zero, which indicates reflectors. The remainder of the distance map shows areas between the reflectors, and a larger value implies a larger distance from reflectors.

If we choose only samples that have nonzero distances in Figure 3b, the selection result will satisfy our first criteria. Normally, we choose the first sample at the location with the largest distance. To keep samples sparse, we avoid placing another sample in a nearby area surrounding the first sample. We refer to this area as an exclusion region, where no sample can be chosen.

4.2 Structurally constrained selection

The rejection region can take different shapes. Figure 4 compares three types of rejection regions: a rectangle, a circle and a structure-constrained ellipse. Neither the rectangle or the circle follows the structural features in the geological layers. These two types of rejection regions cross the reflectors, and as a consequence it risks

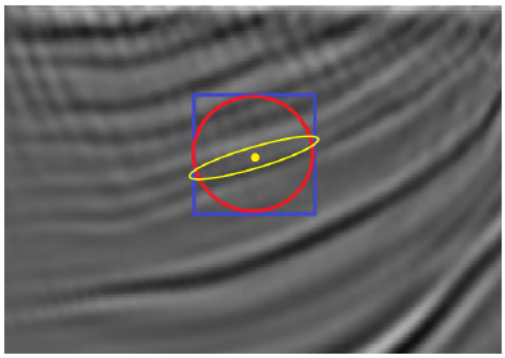


Figure 4. A chosen sample (a yellow dot) in geological layers and three different rejection regions: a rectangular (blue), a circle (red), and a structurally constrained ellipse (yellow). An exclusion region is where no a second sample appears.

missing possible samples. To make the sample selection follow the second criteria, we use image structure to construct a structurally constrained exclusion region, which is shown as the yellow ellipse in Figure 4. This region does not cross imaged reflectors.

To construct a structurally constrained exclusion region, we use a tensor field $\mathbf{D}(\mathbf{x})$, as we do in the image-guided interpolation and its adjoint operator. Pseudo code for implementing the structurally constrained sample selection:

while $d(\mathbf{x}) > 0$ for some $\mathbf{x}$ **do**
  pick the sample $\mathbf{x}_k$ with the largest $d(\mathbf{x})$
  solve equation A1 for $t(\mathbf{x})$ and $t(\mathbf{x}_k) = 0$
  find the structurally constrained region
  where $t(\mathbf{x}) \le t_0$
  exclude every other sample in that region by
  setting $d(\mathbf{x}) = 0$, if $t(\mathbf{x}) \le t_0$;
end while

Figure 5a shows tensor fields $\mathbf{D}(\mathbf{x})$ computed for the migrated image, and Figure 5b shows a total of 165 samples picked with this structurally constrained selection procedure. According to our previously mentioned two criteria, Figure 5b shows a better distribution of samples than do the uniform and pseudo-random selections shown in Figure 2.

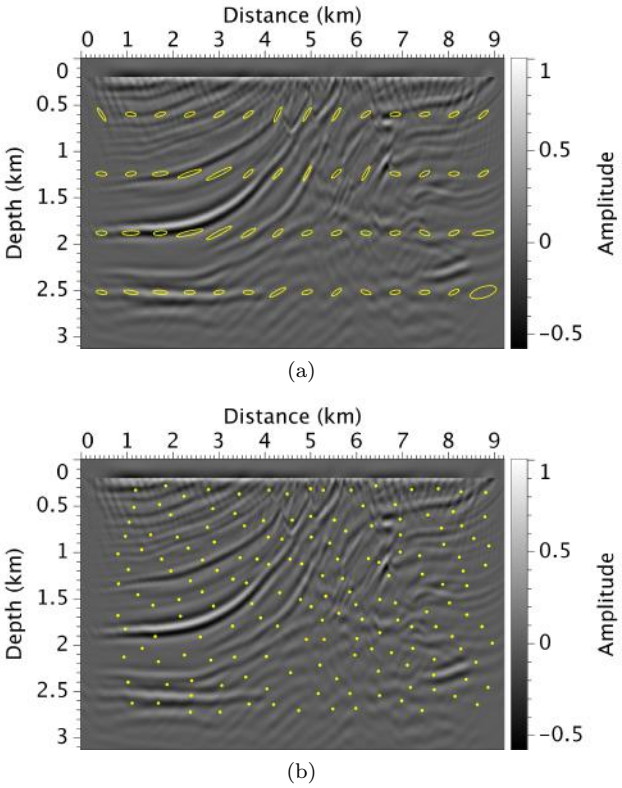


Figure 5. A metric tensor field $\mathbf{D}(\mathbf{x})$ illustrated by ellipses (a) and structurally constrained sample selection (b). A total of 165 samples are chosen for sparse-model FWI.

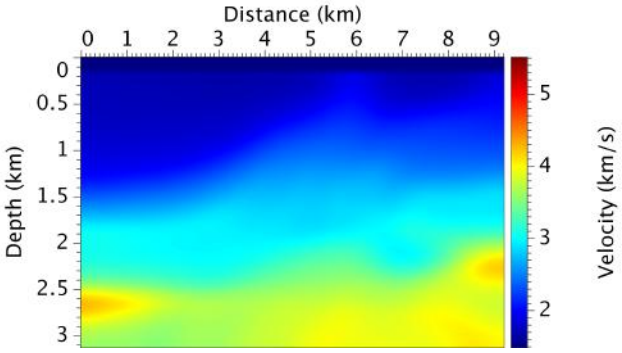


Figure 6. Initial velocity for FWI. It is a highly smoothed version of the true Marmousi II velocity shown in Figure 1a.

5 MARMOUSI EXAMPLE

To illustrate the feasibility of image-guided FWI, we test the algorithm using the Marmousi II model and compare image-guided FWI results with conventional FWI results. In this example, we employ 11 evenly distributed shots on the surface, and a 15Hz Ricker wavelet is used as the source for simulating wavefields. The source and receiver intervals are 0.76 km and 0.024 km, respectively. We refer to the model in Figure 1a as the true model $\mathbf{m}$. Figure 6 displays the initial model $\mathbf{m}_0$ that we used for inversion; it is a highly smoothed version of the true model $\mathbf{m}$.

We first create data $\mathbf{d} = \mathbf{F}(\mathbf{m})$ using the true model $\mathbf{m}$. Henceforth, for consistency with the previous discussion, we refer to these data as the recorded data, even though we compute these noise-free data using a finite-difference solution to a 2D acoustic constant-density wave equation. For example, Figure 7a shows a

common-shot gather for shot number 1 of the recorded data $\mathbf{d}$. Figure 7b shows the corresponding synthetic data $\mathbf{F}(\mathbf{m}_0)$ computed for the initial model $\mathbf{m}_0$. Figure 7c displays the data residual $\mathbf{d} - \mathbf{F}(\mathbf{m}_0)$, which is a part of the data that cannot be explained by the initial model. In the four steps of image-guided FWI, computation of this data residual is step (i).

In step (ii), we compute the gradient of the data misfit. As discussed in Tarantola and Valette (1982), this gradient is equal to the output of reverse time migration (RTM) applied to the data residual, using the current model $\mathbf{m}_0$. This implementation of the gradient is also referred to as the adjoint-state method (Tromp et al., 2005). Figure 8a shows the gradient $\mathbf{g}_0$ computed in this way for the first iteration.

Also in step (ii), we compute the image-guided gradient. To obtain this gradient, one must know the tensor field $\mathbf{D}(\mathbf{x})$ that corresponds to the structure of the subsurface. We obtain this metric tensor field $\mathbf{D}(\mathbf{x})$ from the migrated image. Figure 5a displays ellipses that illustrate tensors for a migrated image, which is a RTM result of recorded data $\mathbf{d}$ with the initial model $\mathbf{m}_0$. We must also choose a set of sparse sample points. In this example, we employ the structurally constrained sample selection scheme and automatically pick a total of 165 samples, as depicted by the dots in Figure 5b. Figure 8b shows the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_0$ computed in this way for the 1st iteration of image-guided FWI.

In step (iii), we use a quadratic line-search algorithm to compute a step length α_0. The search direction $\mathbf{h}_0^{\mathbf{m}}$ is computed using conjugate gradients in equation 13, but for this first iteration is simply the image-guided gradient.

Finally, in step (iv), we update the current velocity model according to equation 12. Figure 1b shows the updated velocity after 25 iterations of image-guided FWI. For comparison, Figure 1c shows the updated velocity after 25 iterations of conventional FWI without the application of the image-guided technique.

6 DISCUSSIONS

Due to the gather-scatter operation $\mathbf{R}\mathbf{R}^T$, the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_0$ shown in Figure 8b contains significantly more low-wavenumber content than the original gradient $\mathbf{g}_0$ in Figure 8a. In addition to low-wavenumber components, the image-guided gradient preserves the structural features in the model.

However, both the original gradient and the image-guided gradient are highest in the shallow part. This uneven distribution will cause FWI to mainly update the shallow part of the model. To deal with this problem, we use the inverse of a seismic illumination factor (Hubral et al., 1999; Xie et al., 2006) as a preconditioner to scale the directions in Figure 8a and 8b before our line search. We also use layer stripping to gradually update the model from the shallow part to deep, but this increases the computational cost. In the Marmousi example, we use layer stripping in groups of 4 iterations. We update the upper half of the model in the first 4 iterations and the lower half of the model in the next 4 iterations. Figure 8 shows only the 1st iteration.

The velocity estimated by image-guided FWI shows coherent structures and makes better geological sense, as shown in Figure 1b. Image-guided FWI, with the application of the image-guided gradient, also converges faster than does conventional FWI. After 25 iterations, the data residual for image-guided FWI is smaller than the residual for conventional FWI, as shown in Figure 9. Moreover, Figure 10a shows the data misfit functions for image-guided FWI and conventional FWI.

Although we never know the true model in practice, for the synthetic study in this paper, it is worthwhile to compare the model misfit function (an L2 norm of the difference between the true model and the estimated model) as well. Figure 10b shows the model misfit as a function of iteration number. Interestingly, but unsurprisingly, the model misfit function of image-guided FWI shows a quite different trend from conventional FWI. Due to the lack of low frequencies in the recorded data, conventional FWI in this Marmousi example never reduce the model misfit, even though the data misfit decreases monotonically. In contrast, the model misfit of image-guided FWI decreases significantly.

7 CONCLUSIONS

In this paper, we have proposed image-guided sparse FWI, which inverts for subsurface parameters in a sparse model space. Image-guided sparse FWI is implemented with a modified conjugate-gradient method that employs an image-guided gradient. We also proposed an efficient way to select sparse model sample locations in a structurally constrained fashion.

We test our method on the Marmousi II model . By using image-guided interpolation and its adjoint operator, we construct an image-guided gradient that mitigates the lack of low frequencies in the recorded data, and thereby improve both inversion speed and quality. Because structural features in images are considered, models updated by image-guided FWI are more sensible geologically.

8 ACKNOWLEDGMENT

This work is sponsored by a research agreement between ConocoPhillips and Colorado School of Mines (SST-20090254-SRA). Yong Ma thanks Thomas Cullison (CWP) for discussions on sample selection and Diane Witters for help in learning how to polish this manuscript.

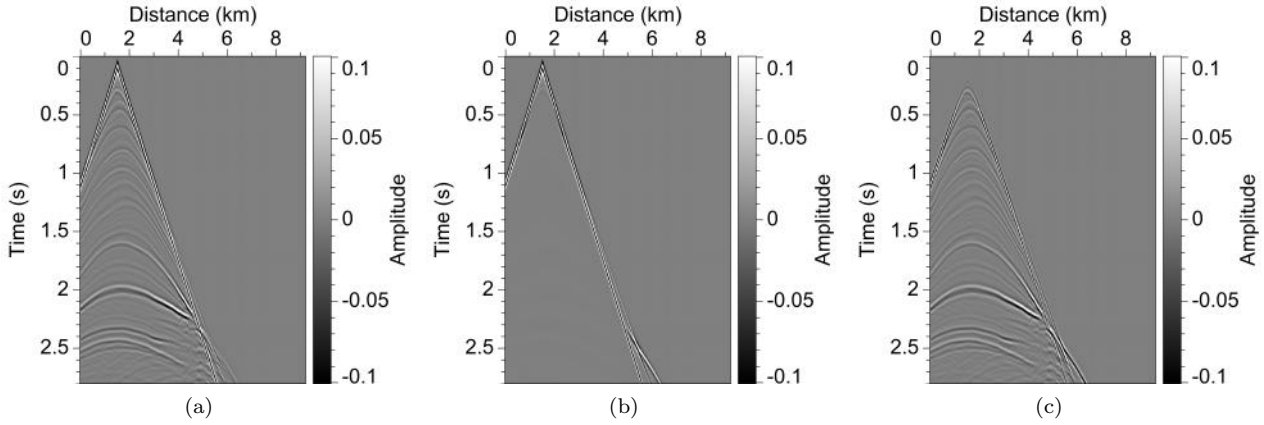


Figure 7. A shot gather from the recorded data $\mathbf{d}$ (a), a shot gather from the synthetic data $\mathbf{F}(\mathbf{m}_0)$ (b), and the initial data residual $\mathbf{d} - \mathbf{F}(\mathbf{m}_0)$ (c).

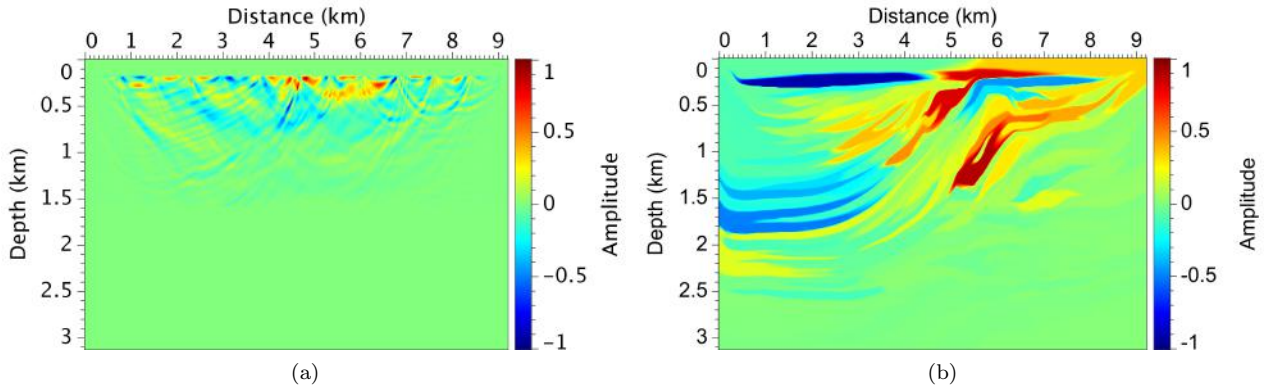


Figure 8. The gradient $\mathbf{g}_0$ (a) and the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_0$ (b), both in the 1st iteration.

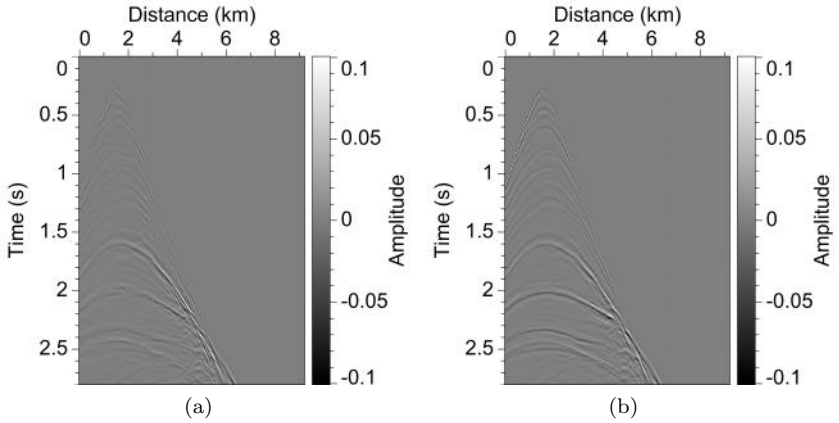
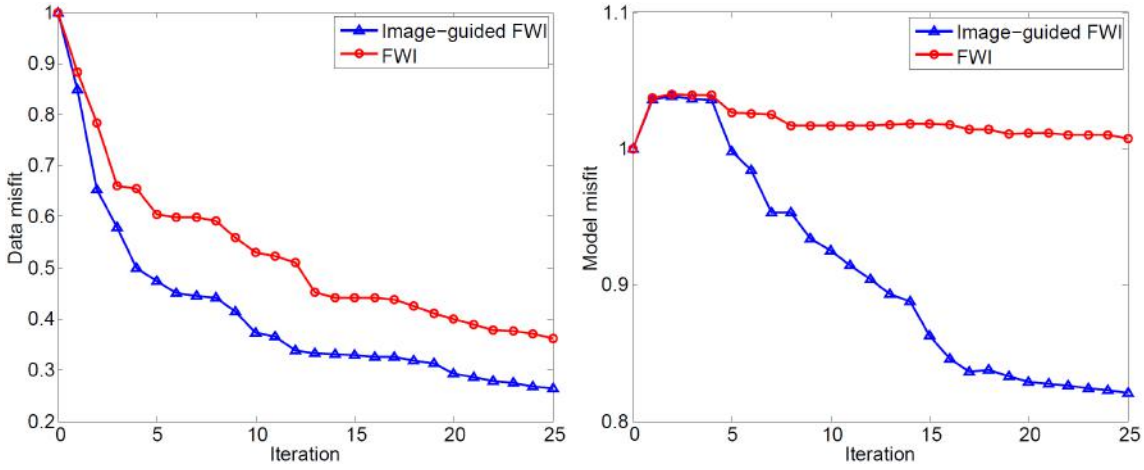


Figure 9. The data residual for one shot after 25 iterations of image-guided FWI (a) and conventional FWI (b).



(a) (b)

Figure 10. Convergence of between image-guided FWI and conventional FWI: the data misfit function (a) and the model misfit function (b).

APPENDIX A: IMAGE-GUIDED INTERPOLATION AND ITS ADJOINT OPERATORS

A1 Image-guided interpolation

We follow the steps in Hale (2009a) to describe the details of nearest neighbor interpolation $\mathbf{P}$ and blended neighbor interpolation $\mathbf{Q}$:

(i) $\mathbf{P}$: solve

$$\nabla t(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla t(\mathbf{x}) = 1, \mathbf{x} \notin \chi \; ; \quad t(\mathbf{x}) = 0, \mathbf{x} \in \chi \tag{A1}$$

for
$t(\mathbf{x})$: the minimum time from $\mathbf{x}$ to the nearest known sample point $\mathbf{x}_k$, and
$p(\mathbf{x})$: the nearest neighbor interpolant corresponding to f_k, the value of the sample point $\mathbf{x}_k$ nearest to the point $\mathbf{x}$.

(ii) $\mathbf{Q}$: for a specified constant $e \geq 2$ ($e = 4$ in this paper), solve

$$q(\mathbf{x}) - \frac{1}{e} \nabla \cdot t^2(\mathbf{x}) \mathbf{D}(\mathbf{x}) \cdot \nabla q(\mathbf{x}) = p(\mathbf{x}) \tag{A2}$$

for the blended neighbor interpolant $q(\mathbf{x})$.

In equation A1, the metric tensor field $\mathbf{D}(\mathbf{x})$ (van Vliet and Verbeek, 1995; Fehmers and Höcker, 2003) represents structural features of the subsurface, such as structural orientation, coherence, and dimensionality. In n dimensions, each metric tensor field $\mathbf{D}$ is a symmetric positive-definite $n \times n$ matrix (Hale, 2009a). Here, the minimum time $t(\mathbf{x})$ is a non-Euclidean distance between a sample point $\mathbf{x}_k$ and an interpolation point $\mathbf{x}$. By this measure of distance, we say that a sample point $\mathbf{x}_k$ is nearest to a point $\mathbf{x}$ if the time $t(\mathbf{x})$ to $\mathbf{x}_k$ is less than $t(\mathbf{x})$ to any other sample point.

A2 Adjoint image-guided interpolation

Letting $\mathbf{p}$ and $\mathbf{q}$ denote vectors that contain all values in $p(\mathbf{x})$ and $q(\mathbf{x})$, respectively, we can rewrite equation A2 in a matrix-vector form:

$$\left(\mathbf{I} + \mathbf{B}^T \mathbf{D} \mathbf{B}\right) \mathbf{q} = \mathbf{p} \; , \tag{A3}$$

where $\mathbf{B}$ corresponds to a finite-difference approximation of the gradient operator (Hale, 2009b). Therefore, $\mathbf{q} = \mathbf{Q}\mathbf{p}$, where

$$\mathbf{Q} = \left(\mathbf{I} + \mathbf{B}^T \mathbf{D} \mathbf{B}\right)^{-1} \; , \tag{A4}$$

and this inverse can be efficiently approximated by conjugate-gradient iterations because $\mathbf{I} + \mathbf{B}^T \mathbf{D} \mathbf{B}$ is symmetric and positive-definite (SPD).

Note that $\mathbf{Q}^T = \mathbf{Q}$, so we can write the adjoint image-guided interpolation as

$$\mathbf{R}^T = \mathbf{P}^T \mathbf{Q}^T = \mathbf{P}^T \mathbf{Q} \; . \tag{A5}$$

APPENDIX B: EUCLIDEAN DISTANCE TRANSFORM

The distance transform (DT) (Fabbri et al., 2008) computes the distance of each pixel of an image to a given subset of pixels. Let $I : \Omega \subset \mathbb{Z}^2 \rightarrow \{0, 1\}$ represent a binary image (e.g., Figure B1a) where the domain $\Omega = \{1, ..., n_1\} \times \{1, ..., n_2\}$. In image processing, 1 is associated with white, and 0 with black. Hence, two sets can be defined in the following way:

$$\mathcal{O} = \{\mathbf{x} \in \Omega \mid I(\mathbf{x}) = 1\} \; , \tag{B1}$$

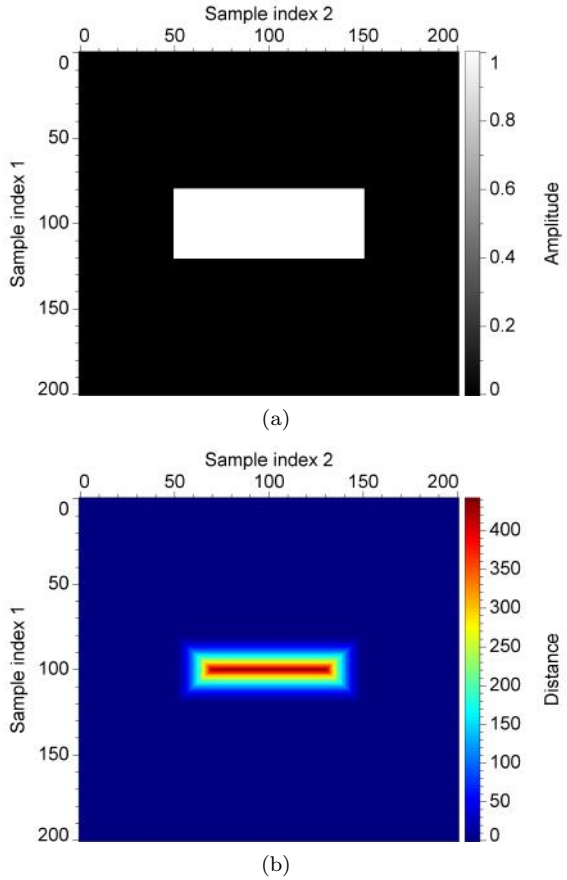


Figure B1. (a) A binary image and (b) its distance transform.

and its complement

$$\mathcal{O}^c = \{\mathbf{x} \in \Omega \mid I(\mathbf{x}) = 0\} . \tag{B2}$$

In image processing literature, the set $\mathcal{O}$ is referred to as object or foreground and can consist of any subset of the image domain Ω. The set of black pixels $\mathcal{O}^c$ is called background. In the DT, $\mathcal{O}^c$ is the given subset.

DT is, thereby, the transformation that produces a map $d(\mathbf{x})$, which shows the smallest distance from this pixel $\mathbf{x}$ to $\mathcal{O}^c$:

$$d(\mathbf{x}) := \min\{l(\mathbf{x},\mathbf{y}) \mid \mathbf{y} \in \mathcal{O}^c\} , \tag{B3}$$

where the DT kernel $l(\mathbf{x},\mathbf{y})$ can take different forms, but $l(\mathbf{x},\mathbf{y})$ is usually the Euclidean distance, defined as $l(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2$. In this article, for simplicity we employ $l(\mathbf{x},\mathbf{y}) = (\|\mathbf{x}-\mathbf{y}\|_2)^2$. Figure B1b shows the distance map computed for the binary image in Figure B1a.

REFERENCES

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.

Benzi, M., 2002, Preconditioning techniques for large linear systems: A survey: Journal of Computational Physics, **182**, 418–477.

Boonyasiriwat, C., P. Valasek, P. Routh, W. Cao, G. T. Schuster, and B. Macy, 2009, An efficient multiscale method for time-domain waveform tomography: Geophysics, **74**, WCC59–WCC68.

Bunks, C., 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457.

Fabbri, R., L. da F. Costa, J. C. Torelli, and O. M. Bruno, 2008, 2D Euclidean distance transform algorithms: A comparative survey: ACM Computing Surveys, **40**, 2:1–2:44.

Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293.

Forgues, E., and G. Lambaré, 1997, Parameterization study for acoustic and elastic ray + born inversion: Journal of Seismic Exploration, **6**, 253–278.

Hale, D., 2009a, Image-guided blended neighbor interpolation of scattered data: SEG Technical Program Expanded Abstracts, **28**, 1127–1131.

———, 2009b, Structure-oriented smoothing and semblance: CWP Report, **635**.

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.

Hubral, P., G. Hoecht, and R. Jaeger, 1999, Seismic illumination: The Leading Edge, **18**, 1268–1271.

Kennett, B., M. Sambridge, and P. Williamson, 1988, Subspace methods for large inverse problems with multiple parameter classes: Geophysical Journal International, **94**, 237–247.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.

Liao, O., and G. A. McMechan, 1996, Multifrequency viscoacoustic modeling and inversion: Geophysics, **61**, 1371–1378.

Ma, Y., D. Hale, Z. Meng, and B. Gong, 2010, Full waveform inversion with image-guided gradient: SEG Technical Program Expanded Abstracts, **29**, 1003–1007.

Meng, Z., 2009, Dip guided full waveform inversion: Patent, 41279–USPRO.

Meng, Z., and J. A. Scales, 1996, 2D tomography in multi-resolution analysis model space: SEG Technical Program Expanded Abstracts, **15**, 1126–1129.

Nocedal, J., and S. J. Wright, 2000, Numerical Optimization: Springer.

Oldenburg, D., P. McGillvray, and R. Ellis, 1993, Generalized subspace methods for large-scale inverse problems: Geophysical Journal International, **114**,

12–20.

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

Pratt, R., C. Shin, and G. Hicks, 1998, Gauss-Newton and full newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

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

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

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

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**, R9–R18.

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

Snieder, R., M. Y. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, **54**, 991–1000.

Stork, C., 1992, Reflection tomography in the postmigrated domain: Geophysics, **57**, 680–692.

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

———, 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics.

Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least-squares criterion: Reviews of Geophysics, **20**, 219–232.

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

van Vliet, L. J., and P. W. Verbeek, 1995, Estimators for orientation and anisotropy in digitized images: Proceeding of the First Annual Conference of the Advanced School for Computing and Imaging, 442–450.

Vasco, D., and E. Majer, 1993, Wavepath travel-time tomography: Geophysical Journal International, **115**, 1055–1069.

Vigh, D., and E. Starr, 2008, 3D prestack plane-wave, full-waveform inversion: Geophysics, **73**, VE135–VE144.

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

Xie, X., S. Jin, and R. Wu, 2006, Wave-equation-based seismic illumination analysis: Geophysics, **71**, S169–S177.

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

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