

A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration

Hejun Zhu¹, Siwei Li², Sergey Fomel³, Georg Stadler⁴, and Omar Ghattas⁵

ABSTRACT

Full-waveform inversion (FWI) enables us to obtain high-resolution subsurface images; however, estimating model uncertainties associated with this technique is still a challenging problem. We have used a Bayesian inference framework to estimate model uncertainties associated with FWI. The uncertainties were assessed based on an a posteriori covariance operator, evaluated at the maximum a posteriori model. For the prior distribution, we have used a spatially nonstationary covariance operator based on a plane-wave construction with local dips measured from migrated images. Preconditioned frequency-domain FWI was used to estimate the maximum a posteriori model. Efficient manipulation of the posterior covariance was based on a low-rank approximation of the data misfit Hessian

preconditioned by the prior covariance operator. The strong decay of the singular values indicated that data were mostly informative about a low-dimensional subspace of model parameters. To reduce computational cost of the randomized singular value decomposition, we have used a Hessian approximation based on point-spread functions. The 2D numerical examples with synthetic data confirmed that the method can effectively estimate uncertainties for FWI. Visual comparisons of random samples drawn from the prior and posterior distributions have allowed us to evaluate model uncertainties. Furthermore, we found out how statistical quantities, such as means and pointwise standard deviation fields, can be efficiently extracted from the prior and posterior distributions. These fields helped us to objectively assess subsurface images provided by FWI.

INTRODUCTION

Full-waveform inversion (FWI) was proposed to estimate physical parameters of subsurface by fully exploiting information embedded in seismic waveforms (Lailly, 1983; Tarantola, 1984). FWI has been used to infer 2D (Pratt et al., 1998; Pratt, 1999; Brossier et al., 2009) and 3D (Ben-Hadj-Ali et al., 2008; Epanomeritakis et al., 2008; Warner et al., 2013) variations of different seismic parameters, such as velocity, density, anisotropy, and attenuation. In recent years, FWI has become an important imaging technique in exploration and global seismology (Fichtner et al., 2009; Tape et al., 2009; Virieux and Operto, 2009; Zhu et al., 2012, 2013; Zhu and Tromp, 2013).

In FWI, gradient-based optimization methods, such as nonlinear conjugate gradients (Fletcher and Reeves, 1964) or limited-memory BFGS methods (Nocedal, 1980), are used to search for the best-fitting models by minimizing discrepancies between observed and predicted seismograms. The adjoint method is usually applied to numerically calculate misfit gradients by correlating forward and adjoint wavefields (Tarantola, 1984; Tromp et al., 2005; Plessix, 2006), which are the numerical solutions of wave equations.

Most current FWI studies have focused on how to efficiently and accurately search for the best-fitting models. Only a few studies discussed issues related to uncertainty and resolution. However, without the crucial information about uncertainty, it is difficult to objectively assess and interpret inverted subsurface models, espe-

Manuscript received by the Editor 18 November 2015; revised manuscript received 14 June 2016; published online 11 August 2016.

¹The University of Texas at Dallas, Department of Geosciences, Richardson, Texas, USA. E-mail: hejun.zhu@utdallas.edu.

²Chevron, Houston, Texas, USA. E-mail: llisiw@gmail.com.

³The University of Texas at Austin, Bureau of Economic Geology, Jackson School of Geosciences, Austin, Texas, USA. E-mail: sergey.fomel@beg.utexas.edu.

⁴New York University, Courant Institute of Mathematical Sciences, New York City, New York, USA. E-mail: stadler@cims.nyu.edu.

⁵The University of Texas at Austin, The Institute for Computational Engineering and Sciences, Jackson School of Geosciences, Department of Mechanical Engineering, Austin, Texas, USA. E-mail: omar@ices.utexas.edu.

© 2016 Society of Exploration Geophysicists. All rights reserved.

cially for small-scale details resolved by FWI. Fichtner and Trampert (2011, 2012) discuss the resolution of FWI by analyzing the Fourier components of the Hessian kernels. Martin et al. (2012) propose a Hessian-accelerated Markov-chain Monte Carlo (MCMC) method for sampling the posterior of Bayesian seismic inversion problems. Zhu et al. (2015) use the point-spread functions (PSFs) of the Hessian to evaluate resolution at selected locations with interesting geologic features. Fichtner and van Leeuwen (2015) analyze direction- and position-dependent resolution lengths for waveform tomography by randomly sampling the Hessian and computing the autocorrelations of the sampled results.

Resolution and uncertainty are long lasting topics in inverse theory. Basic tools for analyzing this information include resolution and posterior covariance matrices (Backus and Gilbert, 1968, 1970). For large-scale inverse problems with millions of model parameters, it is infeasible to assemble and analyze these matrices. Some recovery tests, such as checkerboard or spike tests, are routinely used to estimate resolution for large-scale tomographic inversions. However, such recovery experiments might produce misleading conclusions in terms of resolution capability (Lévéque et al., 1993). Other studies estimated the resolution of large-scale inverse problems by modifying classic inversion algorithms, such as LSQR (Paige and Saunders, 1982; Zhang and McMechan, 1995; Yao et al., 1999; Zhang and Thurber, 2007). Recently, several stochastic sampling approaches have been proposed to extract resolution information for large-scale inverse problems. For instance, MacCarthy et al. (2011) use random samples to extract the diagonal elements of the resolution matrix. An (2012) calculates spatially varying resolution lengths based on the Gaussian approximation of the resolution operators. Trampert et al. (2013) use random samples to obtain resolution lengths for tomographic models, such as S40RTS (Ritsema et al., 2011). A recent detailed review about uncertainty assessments for seismic tomography is provided by Rawlinson et al. (2014).

Bayesian inference provides a systematic way to quantify uncertainties for geophysical inverse problems. Bui-Thanh et al. (2013) formulate seismic tomography in an infinite-dimensional Bayesian inference framework by taking data uncertainties, a priori model information, and likelihood functions into account. By linearizing the forward modeling operator around the maximum a posteriori model, they derive an approximation for the posterior covariance. Because of high dimensionality of the posterior covariances, it is difficult to directly analyze their properties and extract valuable information. Bui-Thanh et al. (2013) use a randomized singular-value decomposition (SVD) (Liberty et al., 2007; Halko et al., 2011) to analyze the spectrum of the prior-preconditioned data misfit Hessian and to construct an efficient-to-manipulate representation. This procedure allowed them to characterize and analyze the posterior covariance of 3D synthetic examples for global seismic tomography (Bui-Thanh et al., 2013). Gouveia and Scales (1998) and Osypov et al. (2013) use a similar approach to quantify model uncertainties by analyzing the posterior covariance in the vicinity of the maximum a posteriori model. In these studies, a priori information was inferred from other independent assessments, such as well-log samples and rock-physics measurements. Similar approaches of using Bayesian inference to estimate model uncertainties have also been discussed for other geophysical inverse problems (Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002; Tarantola, 2005).

In the Bayesian inference framework, the a priori distribution is an important component. The a priori covariance operator used by Bui-Thanh et al. (2013) was approximated as the inverse of the square of an anisotropic elliptic operator, with the leading term corresponding to a biharmonic operator. This imposes a certain correlation structure on the velocity models. Clapp et al. (2004) propose the application of steering filters in reflection tomography. Steering filters incorporate geologic slope information to describe the nonstationary correlation of model parameters at different locations. An analogous idea has been adopted to build appropriate model preconditioners in FWI to achieve faster convergence (Guitton et al., 2012). Hale (2014) generalizes the concept of Matérn covariance by incorporating local geologic information derived from seismic images. By solving partial differential equations with spatially varying coefficients, he was able to build anisotropic and nonstationary model covariances. Fomel and Guitton (2006) show that the model covariance in least-squares inversion can be constructed based on a plane-wave construction (PWC) with local dip information (Fomel, 2002). These methods allowed the construction of nonstationary a priori covariance operators without storing large covariance matrices in memory. In this paper, we use PWC to build a priori covariance operators. The local dips are derived from migrated images, assuming that subsurface velocity variations follow geologic structures (Clapp et al., 2004).

One difficulty for the randomized SVD in Bui-Thanh et al. (2013) is high computational costs associated with sampling the Hessian operator. In this paper, we explore the possibilities of reducing computational costs by using the PSFs of the Hessian. The paper is organized as follows. First, we discuss the Bayesian inference framework for FWI. Then, we introduce a randomized SVD approach to analyze the singular values and vectors of the prior-preconditioned data misfit Hessian. We discuss choices of a priori covariance operators in the following section. Next, we compare random samples from the prior and posterior distributions. Finally, we introduce the idea of using the PSFs of the Hessian to reduce computational costs in the randomized SVD. We use several numerical examples to illustrate this procedure.

THEORY

Bayesian inference

The Bayesian inference framework (Tarantola, 2005) allows us to incorporate in inverse problems prior information, uncertainties in measurements, model parameterization, and theories. The solution to a Bayesian inverse problem is described as an a posteriori probability density function (PDF), $\rho(\mathbf{m}|\mathbf{d})$, which represents the conditional probability distribution of a model \mathbf{m} given observations \mathbf{d} . Based on the Bayes' theorem, it can be expressed as

$$\rho(\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{m})\rho(\mathbf{d}|\mathbf{m}), \quad (1)$$

where $\rho(\mathbf{m})$ is the a priori distribution that is used to describe a priori information on model parameters. It is usually inferred from independent measurements, such as well logs, rock-physics studies, or geologic surveys. Here, $\rho(\mathbf{d}|\mathbf{m})$ denotes the likelihood function that quantifies the fits between observations and predictions given a model \mathbf{m} .

In this study, we assume that a priori information and the likelihood function are described as multidimensional Gaussian distributions

$$\rho(\mathbf{m}) \propto \exp\left\{-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T \boldsymbol{\Gamma}_{\text{prior}}^{-1} (\mathbf{m} - \mathbf{m}_0)\right\}, \quad (2)$$

$$\rho(\mathbf{d}|\mathbf{m}) \propto \exp\left\{-\frac{1}{2}[\mathbf{d} - \mathbf{F}(\mathbf{m})]^T \boldsymbol{\Gamma}_d^{-1} [\mathbf{d} - \mathbf{F}(\mathbf{m})]\right\}, \quad (3)$$

where \mathbf{m}_0 is the a priori mean model and $\boldsymbol{\Gamma}_{\text{prior}}$ is the a priori covariance. Here, $\mathbf{F}(\mathbf{m})$ denotes a general forward-modeling operator that quantifies a linear or nonlinear relationship between models and predictions. In FWI, it represents the numerical solution of the wave equations restricted at observation locations. The $\boldsymbol{\Gamma}_d$ represents data uncertainties, such as errors in measurements or theories.

Combining equations 1–3, we have

$$\rho(\mathbf{m}|\mathbf{d}) \propto \exp\left[-\frac{1}{2}\chi(\mathbf{m})\right], \quad (4)$$

with misfit function $\chi(\mathbf{m})$ defined as (Tarantola, 2005)

$$\begin{aligned} \chi(\mathbf{m}) = & [\mathbf{d} - \mathbf{F}(\mathbf{m})]^T \boldsymbol{\Gamma}_d^{-1} [\mathbf{d} - \mathbf{F}(\mathbf{m})] \\ & + (\mathbf{m} - \mathbf{m}_0)^T \boldsymbol{\Gamma}_{\text{prior}}^{-1} (\mathbf{m} - \mathbf{m}_0). \end{aligned} \quad (5)$$

Although a priori and the likelihood function are Gaussian distributions, the posterior PDF is not necessarily a Gaussian distribution because of the nonlinearity of the forward modeling operator $\mathbf{F}(\mathbf{m})$. Following Gouveia and Scales (1998) and Bui-Thanh et al. (2013), we linearize the forward modeling operator around the maximum a posteriori model \mathbf{m}_{MAP} . Then, the posterior PDF can be approximated as a multidimensional Gaussian distribution,

$$\rho(\mathbf{m}|\mathbf{d}) \propto \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_{\text{MAP}})^T \boldsymbol{\Gamma}_{\text{post}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{MAP}})\right]. \quad (6)$$

The maximum a posteriori model \mathbf{m}_{MAP} can be determined based on least-squares inversion. Here, $\boldsymbol{\Gamma}_{\text{post}}$ is the posterior covariance that is given by

$$\boldsymbol{\Gamma}_{\text{post}} = (\mathbf{H} + \boldsymbol{\Gamma}_{\text{prior}}^{-1})^{-1} = \boldsymbol{\Gamma}_{\text{prior}}^{1/2} (\boldsymbol{\Gamma}_{\text{prior}}^{1/2} \mathbf{H} \boldsymbol{\Gamma}_{\text{prior}}^{1/2} + \mathbf{I})^{-1} \boldsymbol{\Gamma}_{\text{prior}}^{1/2}, \quad (7)$$

where \mathbf{H} is known as the Gauss-Newton approximation of the data misfit Hessian (Pratt et al., 1998), which can be computed by $\mathbf{J}^T \boldsymbol{\Gamma}_d^{-1} \mathbf{J}$, where \mathbf{J} is the Fréchet derivative $\partial \mathbf{F}/\partial \mathbf{m}$. To quantify model uncertainties, we need to approximate and be able to efficiently manipulate the posterior covariance $\boldsymbol{\Gamma}_{\text{post}}$.

Randomized SVD

For large-scale geophysical inverse problems, such as 3D FWI, the size of the posterior covariance $\boldsymbol{\Gamma}_{\text{post}}$ is typically extremely large, with dimension equals to $n \times n$, where n is the total number of discretized model parameters. Therefore, in most cases, it is impossible to directly manipulate and analyze the posterior covariance. Randomized SVD (Liberty et al., 2007; Halko et al., 2011; Bui-Thanh et al., 2013) is an efficient way to approximate large matrices and analyze their spectra, especially for matrices with low-rank property. Bui-Thanh et al. (2013) use this method to decompose and analyze the prior-preconditioned Hessian ($\boldsymbol{\Gamma}_{\text{prior}}^{1/2} \mathbf{H} \boldsymbol{\Gamma}_{\text{prior}}^{1/2}$) as

$$\boldsymbol{\Gamma}_{\text{prior}}^{1/2} \mathbf{H} \boldsymbol{\Gamma}_{\text{prior}}^{1/2} \approx \mathbf{V}_r \boldsymbol{\Lambda}_r \mathbf{V}_r^T, \quad (8)$$

where $\boldsymbol{\Lambda}_r$ is a matrix with diagonal elements equal to the dominant r th singular values of the prior-preconditioned Hessian and \mathbf{V}_r are the associated singular vectors. In this study, we use the single-pass version of the randomized SVD algorithm (Halko et al., 2011) to extract the singular values and vectors of the prior-preconditioned Hessian. Details of this algorithm can be found in Appendix A. The recipe for computing Hessian-vector products, which is an important step in the randomized SVD, is illustrated in Appendix B.

Sampling the prior and posterior distributions

Once we have computed a low-rank approximation of the prior-preconditioned Hessian, we can manipulate the prior and posterior covariances, for instance, by taking their square roots. Then, random samples from the Gaussian prior and posterior distributions can be drawn as follows:

$$\begin{aligned} \mathbf{m}_{\text{prior}} &= \mathbf{m}_0 + \boldsymbol{\Gamma}_{\text{prior}}^{1/2} \mathbf{R}, \\ \mathbf{m}_{\text{post}} &= \mathbf{m}_{\text{MAP}} + \boldsymbol{\Gamma}_{\text{post}}^{1/2} \mathbf{R}, \end{aligned} \quad (9)$$

where \mathbf{R} is a random vector drawn from a normal distribution with zero mean and identity covariance. The $\boldsymbol{\Gamma}_{\text{post}}^{1/2}$ is calculated as

$$\boldsymbol{\Gamma}_{\text{post}}^{1/2} = \boldsymbol{\Gamma}_{\text{prior}}^{1/2} (\mathbf{V}_r \mathbf{P}_r \mathbf{V}_r^T + \mathbf{I}), \quad (10)$$

where $\mathbf{P}_r = \text{diag}(1/\sqrt{\lambda_1 + 1} - 1, \dots, 1/\sqrt{\lambda_r + 1} - 1)$ and λ_i are the singular values of the prior-preconditioned data misfit Hessian. To make equations 8 and 10 good approximations, we truncate the singular values approximately 0.1. Visual comparisons of random samples from the prior and posterior distributions allow us to assess model uncertainties associated with inversion.

Choices of a priori covariance

To define Bayesian inverse problems, we have to properly choose a priori covariance operators that describe correlations among different model parameters, i.e., parameters at different locations or parameters belonging to different classes. The a priori covariance operator has the same size as the a posteriori covariance. For small-scale inverse problems, it is possible to explicitly build a priori covariance matrices. For instance, Gouveia and Scales

(1998) use the fluctuation of well-log measurements to build a priori covariance matrices for 1D nonlinear waveform inversion.

However, for large-scale inverse problems, it is usually impossible to explicitly construct and store covariance matrices in memory. Therefore, in most cases, we have to build operators to represent covariance information, so that covariance operators can be applied to vectors in model spaces. In Bui-Thanh et al. (2013), the square of an elliptic operator was chosen as the inverse of the prior covariance to solve global tomographic problems, which can be generalized as an anisotropic operator.

Because most geophysical signals are nonstationary, the prior covariance operators should capture this nonstationary behavior. In addition, the a priori covariance operators should describe some independent measurements that geophysical models should follow. Steering filters (Clapp et al., 2004) and Matérn covariance (Hale, 2014) are examples of covariance operators, which take these properties into account.

For locally planar structures, the outcome of plane-wave destruction (PWD) (Fomel, 2002) is nearly white, suggesting that PWD involves the inverse spectra of models. Therefore, the inverse of PWD, i.e., PWC, can be a good candidate for a priori covariance operators. Fomel and Claerbout (2003) show that model covariance in the context of least-squares inversion can be represented as

$$\boldsymbol{\Gamma}_{\text{prior}} = \mathbf{P}\mathbf{P}^T = (\mathbf{D}^T\mathbf{D})^{-1}, \quad (11)$$

where \mathbf{P} and \mathbf{D} are the preconditioning and regularization operators, respectively. Here, \mathbf{P}^T and \mathbf{D}^T are their adjoint. Fomel and Guitton (2006) chose PWC as the preconditioning operator \mathbf{P} . Local dips used in PWC are measured based on PWD to describe linear trends in geophysical models. In this study, we construct a nonstationary a priori covariance operator based on PWC with local dips measured

from migrated images, assuming that velocity variations follow impedance contrasts.

Preconditioning and regularization

To find the maximum a posteriori model \mathbf{m}_{MAP} , we need to solve a regularized least-squares inverse problem (equation 5). Fomel and Claerbout (2003) reformulate a regularized inverse problem to a precondition problem to improve the convergence of the inversion

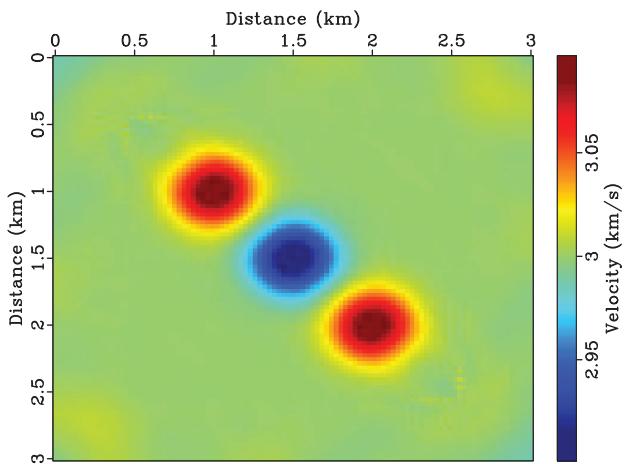


Figure 2. The first experiment with Gaussian anomalies. The maximum a posteriori model is obtained after 40 iterations. Shots and receivers are equally distributed around the model to achieve good data coverages. A homogeneous starting model is used in this experiment.

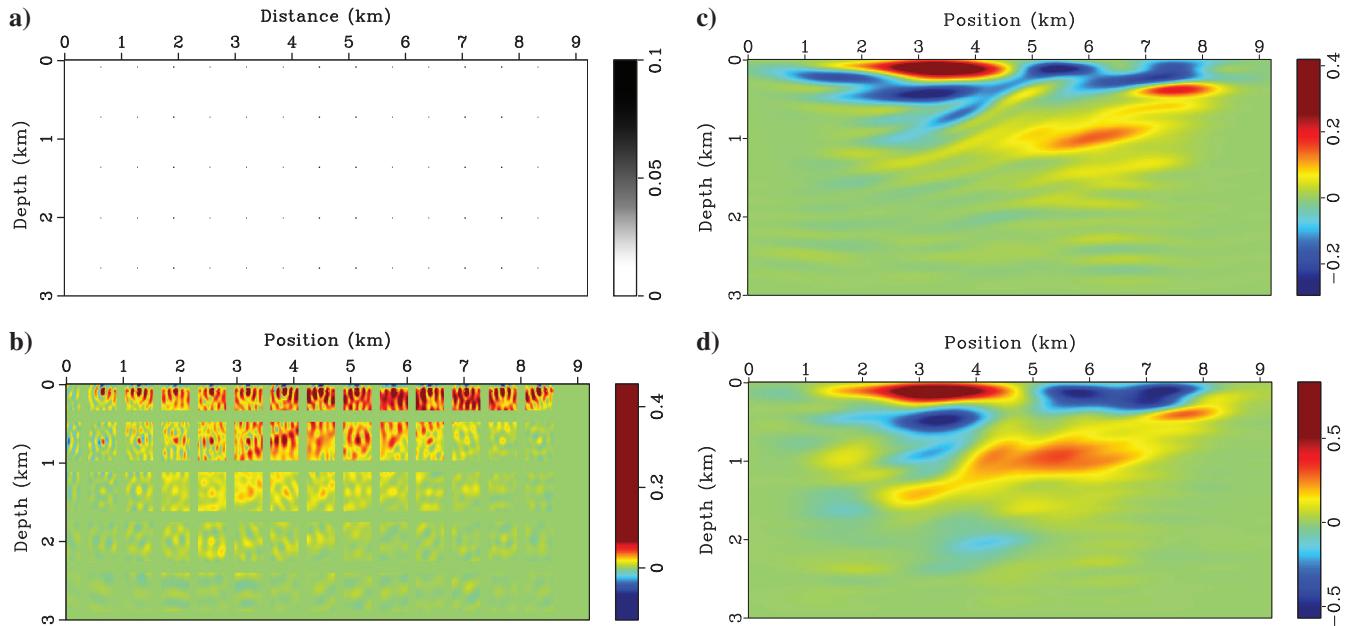


Figure 1. Hessian-vector products based on the PSFs of the Hessian. The 2D Marmousi model is used as background velocity model. (a) Spikes used to calculate the PSFs in panel (b). (c and d) The products of a random vector with the true Hessian and its PSF approximation, respectively.

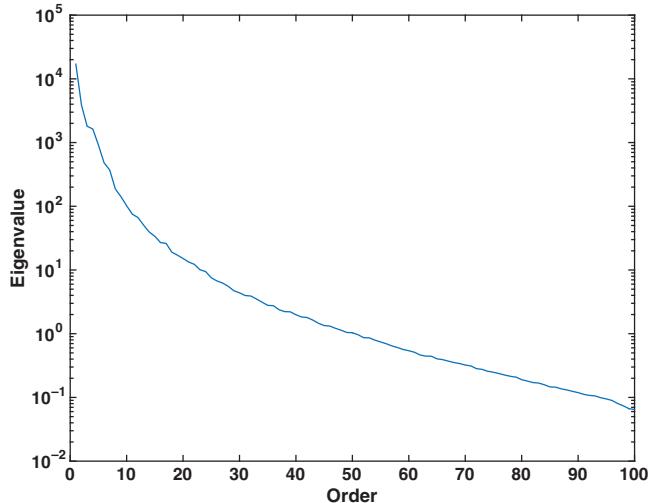


Figure 3. The spectrum of the prior-preconditioned data misfit Hessian for the first experiment with good data coverage.

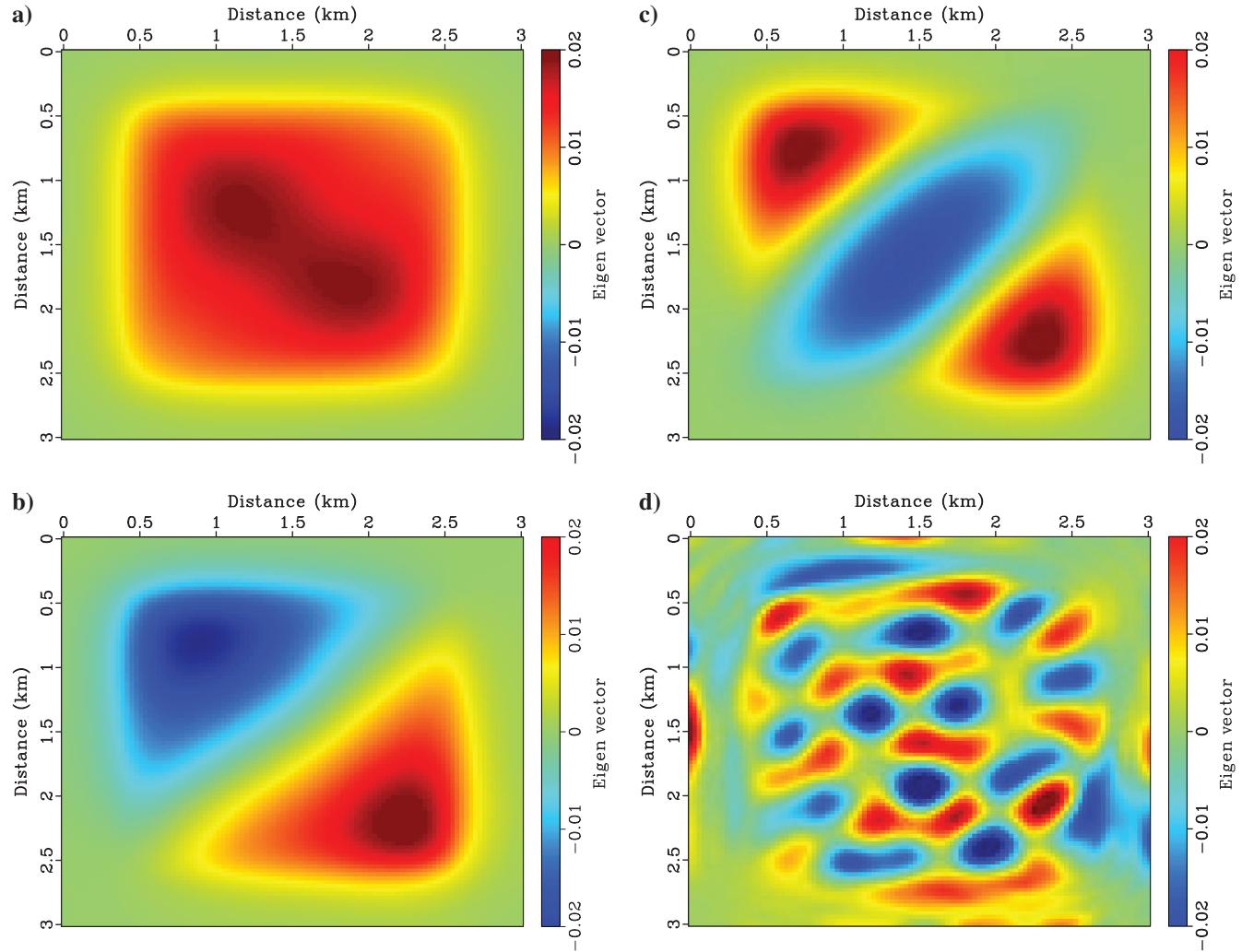


Figure 4. The (a) first, (b) second, (c) third, and (d) 50th singular vectors of the prior-preconditioned data misfit Hessian for the first experiment with good data coverage. Singular vectors corresponding to small singular values show small-scale patterns.

algorithm. This strategy has been applied in image processing (Fomel and Claerbout, 2003; Fomel and Guitton, 2006), reflection tomography (Clapp et al., 2004), and FWI (Guitton et al., 2012). Because of the nonstationary preconditioner we choose (PWC operator), we can train the misfit gradients to follow local geologic trends, which allows us to build more geologically meaningful velocity models. In FWI iterations, the misfit gradient direction \mathbf{g} is preconditioned with the PWC operator \mathbf{P} as

$$\bar{\mathbf{g}} = \mathbf{P}\mathbf{P}^T\mathbf{g}. \quad (12)$$

Then, the new model is calculated based on the model from the previous iteration with the new misfit gradient $\bar{\mathbf{g}}$ given in equation 12.

Reducing computational costs with the PSFs of the Hessian

To extract the singular values and vectors of the prior-preconditioned misfit Hessian, we have to perform many Hessian-vector

products in the randomized SVD algorithm (Appendix A). The number of the Hessian-vector products equals the estimated rank of the prior-preconditioned misfit Hessian, which is typically several hundred. To reduce computational costs, we use the PSFs of the Hessian to approximate the products of the Hessian with random vectors. The PSF represents one column/row of the Hessian operator (Chen and Xie, 2015; Tang and Lee, 2015). With the assumption that the PSFs concentrate locally, i.e., not change dramatically from location to location within certain distances, we use samples and interpolation to approximate the Hessian-vector products. Here, we use the 2D Marmousi model to illustrate this idea. First, we compute the products of the Hessian with a group of spikes (Figure 1a), which are separated several wavelengths to reduce crosstalk between each individual PSF. Window functions are used to extract these PSFs (Figure 1b), which are saved on disk. To compute the products of the Hessian with a random vector, we read in these PSFs and compute their products with the random vectors. Interpolation is applied for locations not sampled by the PSFs. Linear or local dip honored interpolation (Fomel, 2002) can be used in this calculation. The PSFs approximated Hessian-vector products

should work better for models with simple structures than complicated models due to the assumptions described before. To reduce crosstalk between different PSFs and achieve accurate interpolation results, we use several groups of spikes to sample the Hessian. Therefore, the total number of the Hessian-vector products equal the number of spike groups instead of the estimated rank of the prior-preconditioned Hessian. Figure 1c and 1d compares the products of the true Hessian (Figure 1c) and its PSFs (Figure 1d) with a random vector.

NUMERICAL EXAMPLES

Gaussian model

In this section, we use several 2D numerical examples to demonstrate the procedure described in the previous section. The first example is a simple 2D model with Gaussian anomalies. The true velocity model involves three Gaussian anomalies with alternative signs along its diagonal. The starting model is homogeneous with a velocity of 3 km/s. We perform two experiments, one with good data coverage and one with relatively poor data coverage.

For the first experiment, we have full coverage for the model. The 70 shots are uniformly distributed along the left and bottom sides of the model, whereas 70 receivers are uniformly distributed along the right and top sides. The 2D frequency-domain FWI (Pratt et al., 1998) is used to recover the velocity perturbations. A sparse solver (Davis, 2004) is applied to numerically solve the 2D acoustic, constant density Helmholtz equation. Within a multiscale inversion framework (Bunks et al., 1995; Sirgue and Pratt, 2004), we start with 1 Hz data and gradually increase the frequency content to 10 Hz. After 40 iterations, we obtain the maximum a posteriori model shown in Figure 2. Because of excellent data coverage, we are able to recover the three Gaussian anomalies.

Then, we use a randomized SVD method (Liberty et al., 2007; Halko et al., 2011; Bui-Thanh et al., 2013) to extract the singular values and vectors of the prior-preconditioned data misfit Hessian. Figure 3 shows the spectrum of the prior-preconditioned misfit Hessian. One hundred random vectors are used to sample the prior-preconditioned Hessian. We observe that the magnitudes of the singular values decay rapidly, suggesting that data are mostly informative about a low-dimensional subspace of model parameters.

Figure 4 shows the 1st, 2nd, 3rd, and 50th singular vectors of the prior-preconditioned misfit Hessian. For small singular values, more oscillatory patterns are observed in the singular vectors. The first three low-order singular vectors involve low-wavenumber features, suggesting that the misfit function is very sensitive to the changes of model parameters in these directions. On the other hand, the 50th singular vector only involves high-wavenumber components, indicating that the misfit function is almost insensitive to the modifications of model parameters in this mode.

Using the singular values and vectors of the prior-preconditioned misfit Hessian (Figures 3 and 4), we are able to manipulate the posterior covariance and generate random samples from the prior and posterior distributions based on equation 9. Then, we can calculate statistical quantities, such as the pointwise standard deviation. Figure 5 compares the pointwise standard deviations of the prior and posterior distributions. We observe that there are significant reductions in the pointwise standard deviations for the center part of the model, which is well-constrained by data. For regions with poor

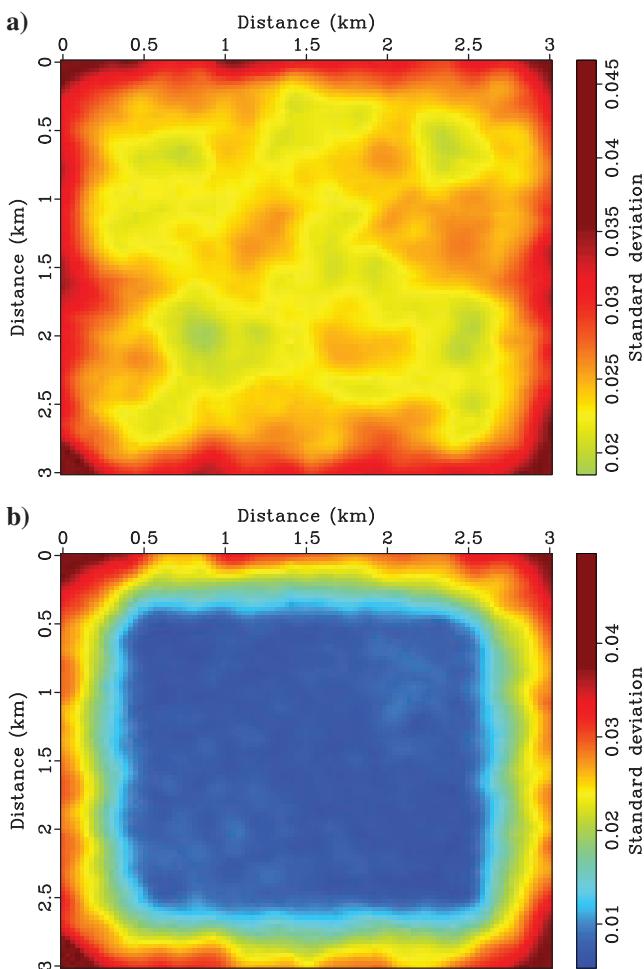


Figure 5. Comparisons of the pointwise standard deviations of the (a) prior and (b) posterior PDFs. Significant reductions in the standard deviations are observed in the center part of the model with good data coverage.

data coverage, the standard deviations are barely changed from their prior values.

To compute singular values and vectors as shown in Figures 3 and 4, we perform 100 Hessian-vector products. Here, we use the PSFs of the Hessian to approximate these products and reduce computational costs. Figure 6a and 6b shows the spikes used to sample the Hessian and their corresponding PSFs. Figure 6c compares the singular values of the prior-preconditioned Hessian with the true Hessian and its PSFs. For the case with PSFs, we only use 16 Hessian-vector products. Figure 6d shows the pointwise standard deviation of the posterior PDFs based on PSFs, which is similar to Figure 5b but with lower computational costs.

The second example has the same setting as the first experiment, except shots and receivers are distributed only along the left and top sides of the model, respectively. Therefore, we only have good data constraints for the model on the top left sides. After 40 iterations, we obtain the maximum a posteriori model as illustrated in Figure 7. Here, we are only able to recover anomalies in areas with good data constraints.

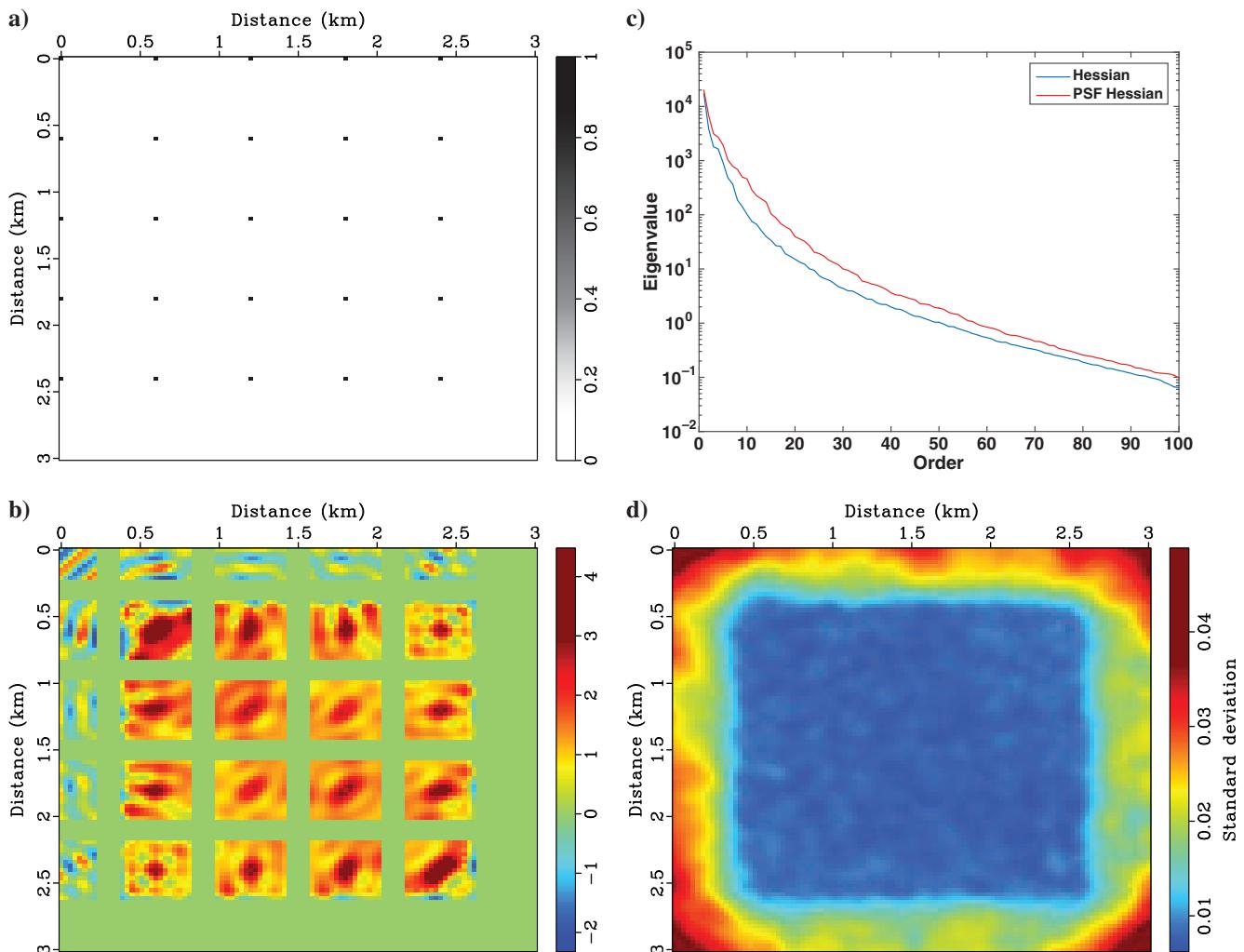


Figure 6. Solutions based on the PSFs of the Hessian. (a) Spikes used to calculate the PSFs in panel (b). (c) Compares the singular values of true Hessian and the PSFs. (d) Illustrates the pointwise standard deviations of the posterior PDF based on the PSFs of the Hessian. Only 16 Hessian-vector products are used in this example.

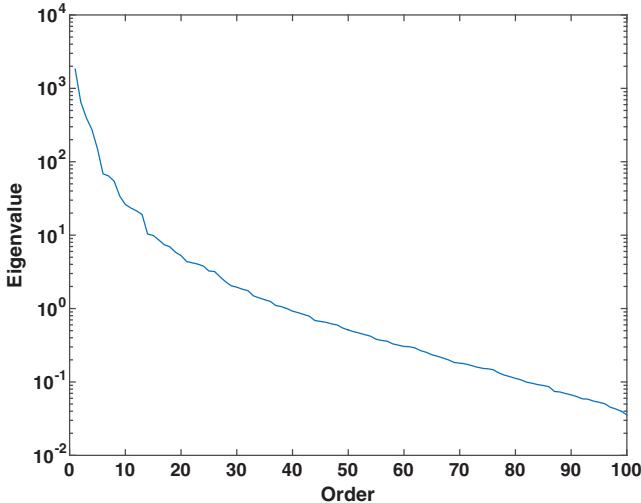


Figure 8. The spectrum of the prior-preconditioned data misfit Hessian for the second experiment with poor data coverage.

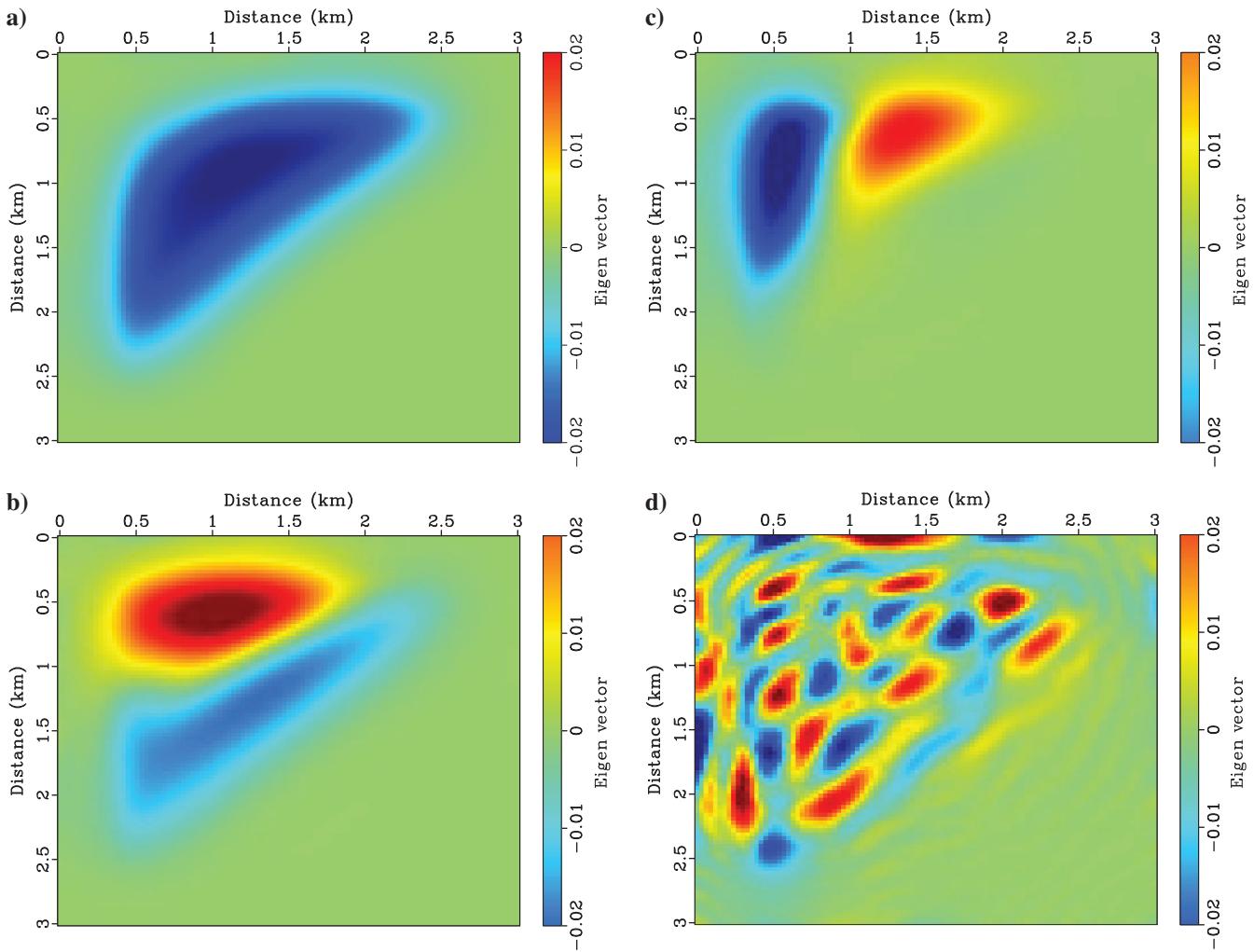


Figure 9. The singular vectors of the prior-preconditioned data misfit Hessian for the second experiment. The same setting as Figure 4.

Similarly, we can extract singular values and vectors of the data misfit Hessian as illustrated in Figures 8 and 9. The pattern of the singular values is similar to the first experiment (Figure 3), suggesting the low-rank property of the prior-preconditioned misfit Hessian in this experiment. The singular vectors are dominant in the regions with good data constraints, whereas they have small sensitivities for areas with poor data constraints. Again, the modes corresponding to large singular values contain low-wavenumber information, whereas those corresponding to small singular values contain high-wavenumber patterns.

Then, we calculate the pointwise standard deviations of the prior and posterior distributions as illustrated in Figure 10. At this time, only the top left portion of the model is well-constrained, i.e., with large reductions in standard deviations.

Comparing Figures 5 and 10, it is clear that the pointwise standard deviations are significantly reduced from the prior to the posterior distributions when we have good data constraints. For the areas with poor illumination, the reductions of the standard deviations are relatively small, suggesting poor resolution and large uncertainties.

In Figure 11, we show the solutions based on the PSFs of the Hessian. The singular values and the standard deviation of the posterior PDF based on PSFs are similar to the true Hessian operator except only 16 Hessian-vector products are used. Results in Figures 6 and 11 suggest that the PSFs' approximation works well for models with simple structures.

Marmousi model

In this section, we use the 2D Marmousi model to further test the proposed framework. One hundred sixty-five shots and 185 receivers are used to illuminate the model. Each shot is recorded by all receivers. The finite-difference method is used to solve the acoustic wave equation. The grid size is 188×576 , and grid space is 16×16 m. Following the multiscale inversion strategy (Bunks et al., 1995; Sirgue and Pratt, 2004), we gradually increase frequency content from 1 to 8 Hz. For each frequency group, we perform 10 preconditioned conjugate gradient iterations.

Figure 12 is the true velocity model used to generate synthetic data. The a priori mean model \mathbf{m}_0 in Figure 13a is obtained by smoothing the true model with a Gaussian function. This mean model is also used as the starting model in the least-squares inversion. We choose a relatively good a priori mean model in this study because in the calculations of the maximum a posterior model, we have to use a regularized FWI, which is a highly nonlinear inverse problem. If the starting model is relatively poor, then the inversion cannot guarantee convergence toward the correct maximum a posterior model.

After 80 nonlinear, preconditioned conjugate gradient iterations, we obtain the maximum a posterior model \mathbf{m}_{MAP} as shown in Figure 13b. We observe that the shallow parts of the Marmousi model are better recovered in comparison with the deeper parts. However, without quantitative uncertainty information, it is difficult to objectively assess the reliability of the best-fitting model.

We build the a priori covariance operator based on PWC with local dips measured from migrated images. This construction is based on the assumption that velocity variations follow high-wavenumber reflector variations. In addition, these migration images are derived from high-frequency data compared with the regularized FWI, which mainly rely on the low-frequency components of records. In this example, we perform frequency-domain reverse time migration (RTM; Mulder and Plessix, 2004) with frequencies ranging from 5 to 15 Hz. The a priori mean model (Figure 13a) is used as background velocity model in RTM. The migrated image is shown in Figure 14a, where a Laplace filter is applied to remove low-wavenumber artifacts in the shallow parts of the image. The local dips (Figure 14b) are measured based on PWD (Fomel, 2002).

Next, we build the PWC operator based on the local dips shown in Figure 14b. Figure 15a shows a “texture” image produced by the action of PWC on a random vector. We observe that PWC aligns the random signals along predefined nonstationary dip directions. In contrast, Figure 15b shows a “texture” image based on a stationary covariance operator on the same random vector. It is clear that PWC allows us to better represent nonstationary correlations between model parameters at different locations, as long as we are able to extract local dips from other measurements, such as migrated images in this case.

We use 400 random vectors with each component drawn from a normal distribution with zero mean and unit variance to sample the prior-preconditioned data misfit Hessian. Figure 16 shows the

spectrum of the preconditioned misfit Hessian. The strong decay of its singular values indicates that the data are only informative to a low-dimensional subspace of the model parameters, consistent with the previous experiments with Gaussian anomalies. Figure 17 illustrates the 1st, 5th, 15th, and 200th singular vectors of the preconditioned Hessian. We observe that the singular vectors corresponding to small singular values are more oscillatory. This pattern is consistent with our previous experiments, suggesting that the rate of change in misfits is large only when we change the model along the singular vector directions corresponding to dominant singular values.

After these computations, we can draw random samples from the prior and posterior distributions. The left plots in Figure 18 show three random samples drawn from the a priori PDF. Due to the specific a priori covariance operator chosen in this study, the prior samples involve the patterns of predefined slope directions (Figure 14b). The variations among different a priori samples reflect the assumed a priori distribution. The right plots in Figure 18 illustrate three corresponding random samples drawn from the posterior distribution. These posterior samples are dominated by the patterns of the

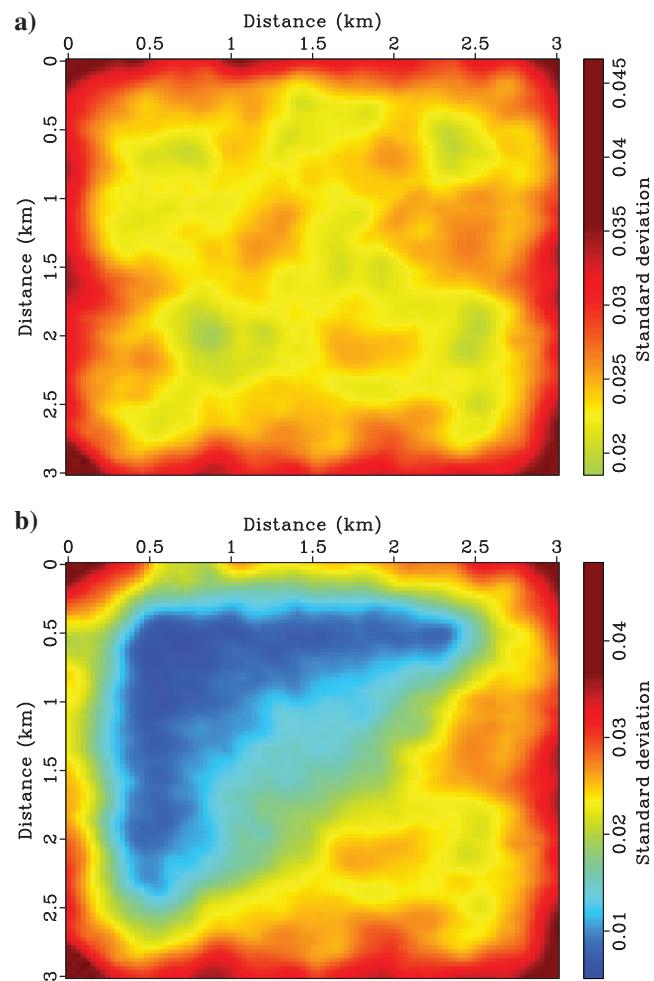


Figure 10. Comparisons of the pointwise standard deviations of the (a) prior and (b) posterior distributions for the second experiment with a poor data constraint. Only the areas with good data coverage involve significant reductions in the standard deviations.

maximum a posteriori model (Figure 13b). The variations among different posterior samples allow us to analyze uncertainties for the posterior distribution.

Furthermore, we can compare the standard deviations of the prior and posterior distributions (Figure 19). The standard deviations are significantly reduced for the shallow part of the model. In the center part of the model with better data coverage, the reductions are greater than those near the left and right boundaries because of the configuration of survey.

Some crossvalidation tools as suggested by Rimstad and Omre (2014) and Rimstad et al. (2012) can help us to quantitatively assess the uncertainties associated with inversion, especially for synthetic experiments. To better compare the prior and posterior samples, we follow some of their approaches, such as analyzing depth profiles for the prior and posterior samples. Figure 20 shows depth profiles for random samples drawn from the prior and posterior distributions at $X = 4.48$ (Figure 20a) and $X = 0.8$ km (Figure 20b). Two hundred random samples are plotted in Figure 20. In the middle of the model (Figure 20a), model uncertainties are significantly reduced at shallow depths (< 2 km). In contrast, at greater depths (> 2 km), the uncertainties are only slightly reduced. On the left side of the

model (Figure 20b), the uncertainties are only reduced at depths shallower than 0.25 km. At greater depths, the uncertainties of the posterior samples are almost unchanged compared with the prior samples.

Figure 21 shows the pointwise standard deviation of the posterior PDF based on the PSFs of the Hessian. Only 32 Hessian-vector products are used compared with 400 with the randomized SVD

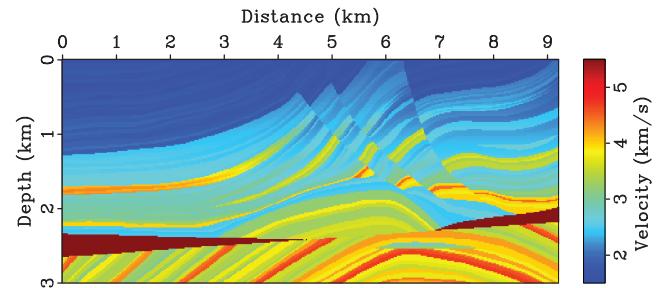


Figure 12. True Marmousi model used to generate “data.”

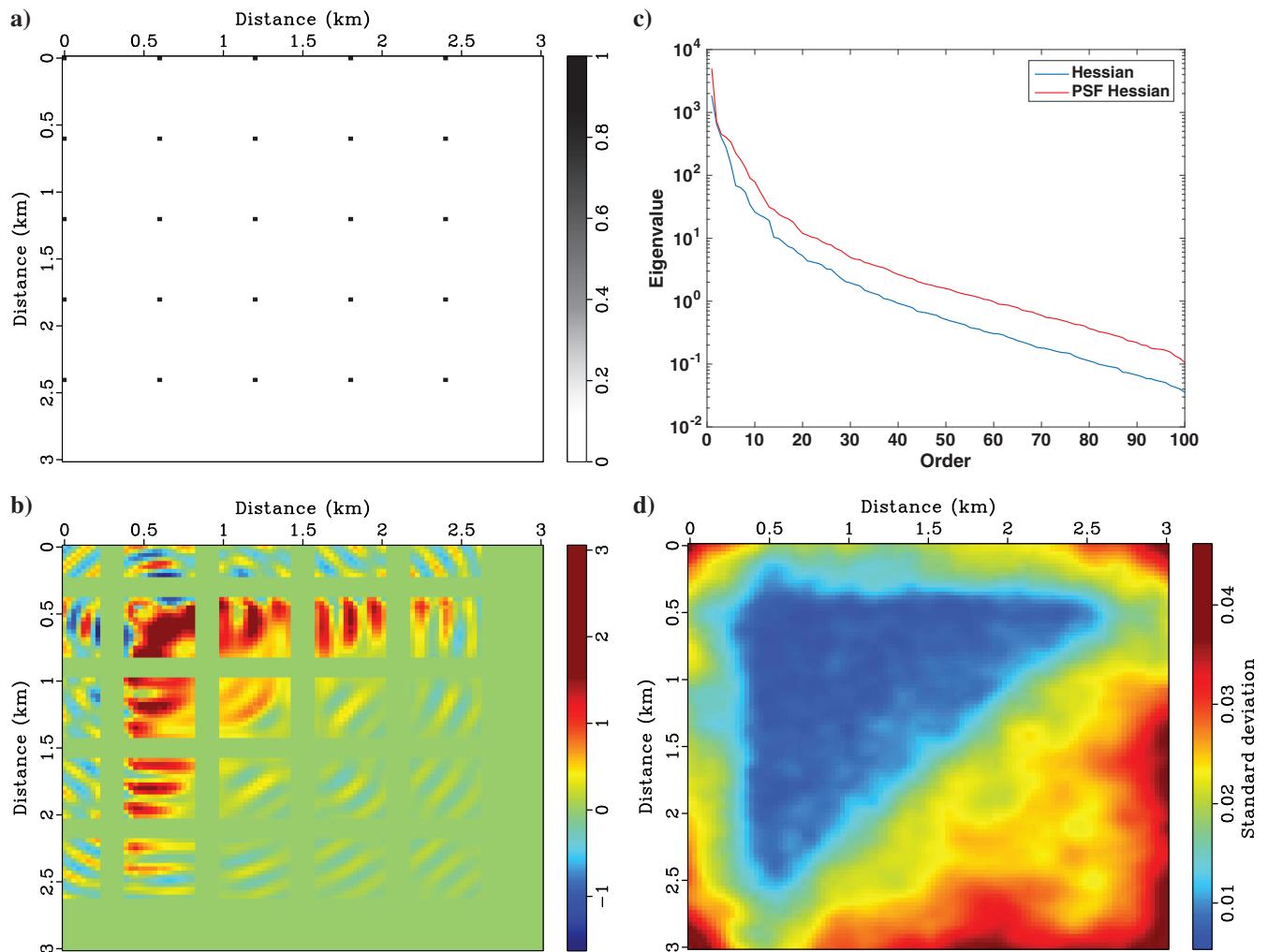


Figure 11. The same setting as Figure 6 except for the second experiment with a poor data constraint.

approach. Although the pointwise standard deviations at shallow depths are overestimated, the overall patterns are similar to Figure 19b. These overestimations might come from crosstalk and the nonlocal behavior of the PSFs.

DISCUSSION

Due to various uncertainties in measurements, theories, and model parameterizations, the solution to an inverse problem is

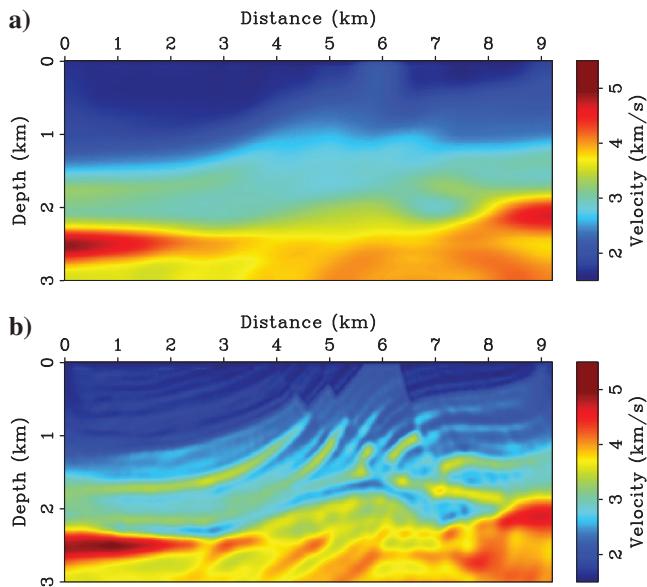


Figure 13. (a) The a priori mean model \mathbf{m}_0 is obtained by smoothing the true model in Figure 12, which is also chosen as the starting model for FWI. (b) The maximum a posteriori model \mathbf{m}_{MAP} is obtained after 80 iterations.

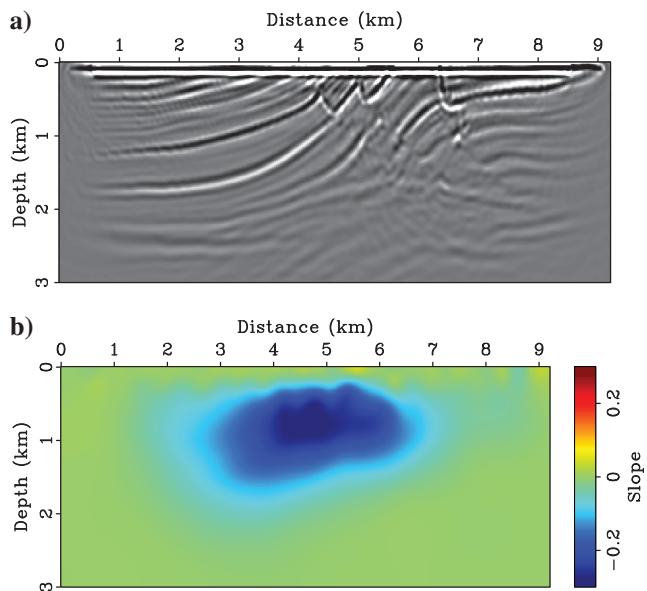


Figure 14. Migrated image based on RTM (a) and local dips measured from PWD (b). The a priori mean model (Figure 13a) is used as a background velocity model in RTM.

inherently probabilistic (Tarantola, 2005). It is not enough to present the maximum a posteriori model with minimum discrepancies between observations and predictions. To provide a full characterization of the a posteriori PDF, we can rely on statistical sampling methods, such as MCMC (Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002), to directly sample the prior and posterior distributions. However, for large-scale geophysical inverse problems, the high dimensions of the model spaces and the expensive forward calculations make this a very challenging problem. Therefore, currently it is still prohibitive to fully rely on statistical sampling methods to solve large-scale geophysical inverse problems. Most current studies focus on finding the best-

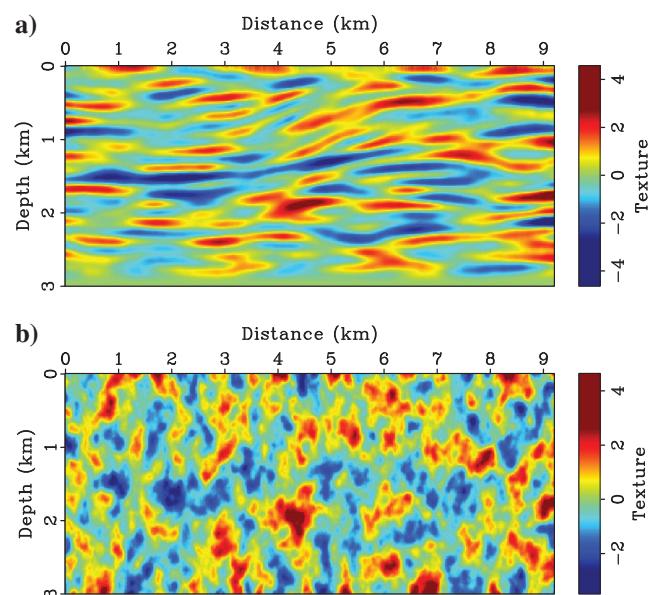


Figure 15. “Texture” images with a random vector applied to a non-stationary covariance operator based on (a) PWC and (b) a stationary covariance operator.

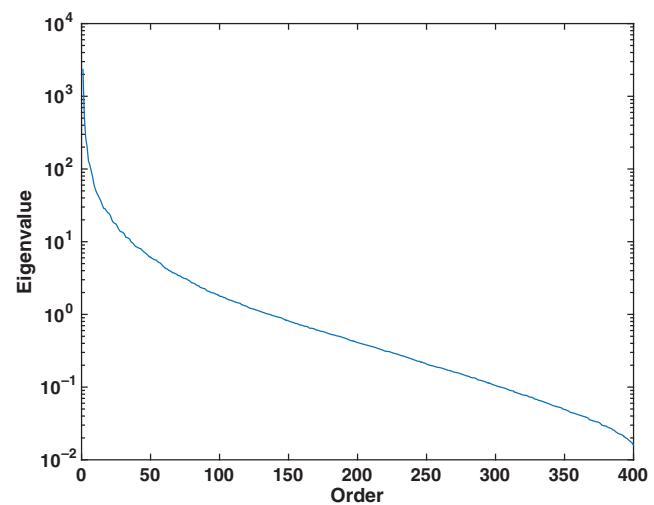


Figure 16. The spectrum of the prior-preconditioned data misfit Hessian for the Marmousi experiment.

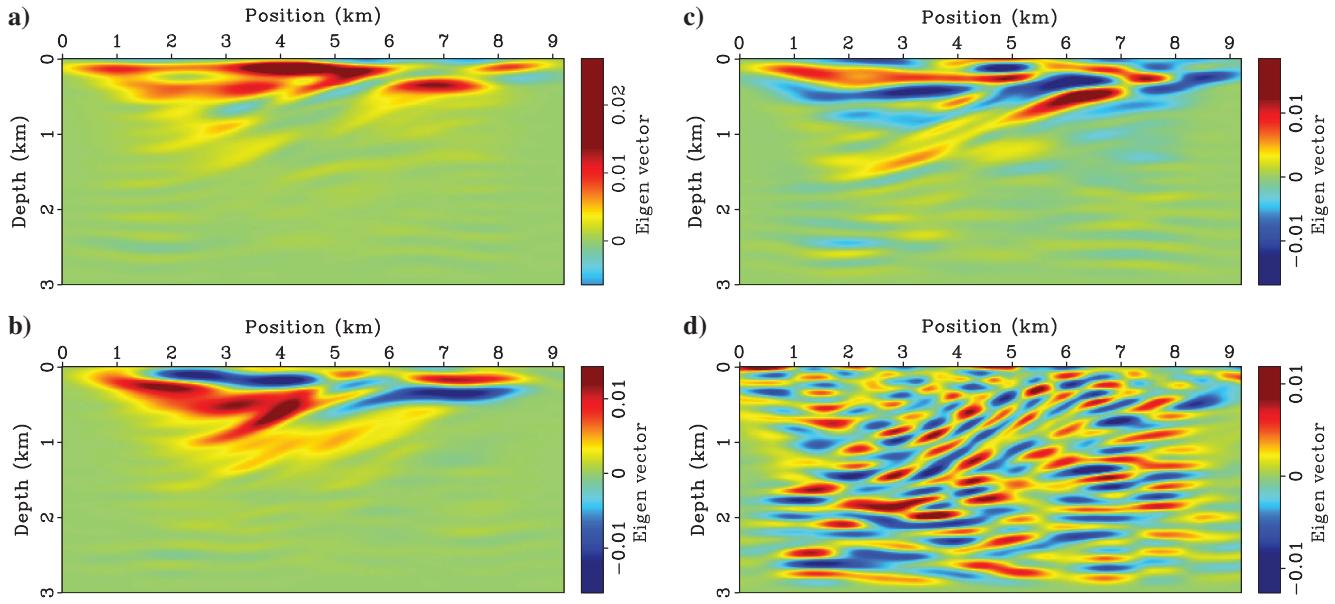


Figure 17. The (a) first, (b) fifth, (c) 15th, and (d) 200th singular vectors of the prior-preconditioned data misfit Hessian for the Marmousi experiment.

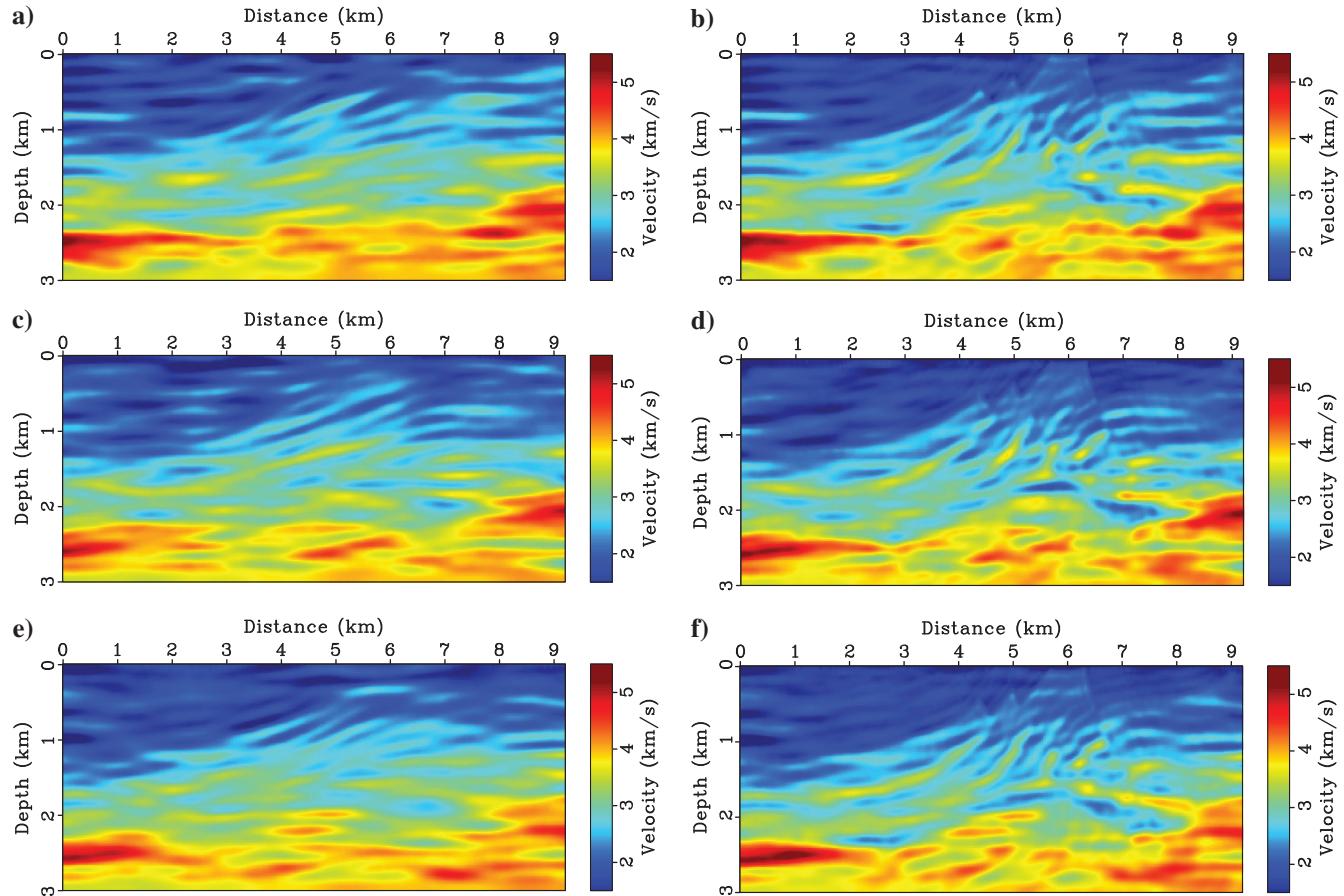


Figure 18. (a, c, and e) Random samples drawn from the prior and (b, d, and f) posterior PDFs for the Marmousi experiment.

fitting models, i.e., the maximum a posteriori model, based on optimization methods. This framework usually has difficulties in evaluating resolution and uncertainty. For classical linear inverse problems, there are tools, such as checkerboard tests and resolution matrices, to assess resolution of the solutions. However, for large-scale, nonlinear inverse problems, such as FWI, it is expensive to perform such tests.

In this study, within the Bayesian inference framework, we are able to obtain and analyze the maximum a posteriori model and posterior covariances by linearizing the forward-modeling operator in the vicinity of the maximum a posteriori model. Once we can analyze and manipulate the posterior covariances, we have access to a variety of tools for analyzing uncertainties, such as visual comparisons of random samples, pointwise standard deviations, and depth profiles. Using these tools, we are able to provide statistical quantities for assessing the resolution and uncertainties of our inversion results. This information can be important in practice, especially when we want to interpret high-resolution, small-scale features produced by FWI. Compared with previous works, such as MacCarthy et al. (2011) and An (2012), we are able to extract more information (singular values and vectors) about the covariance or resolution matrices by low-rank approximation.

Our analysis currently restricted to the Gaussian approximation of the posterior PDF by linearizing the forward modeling operator around the maximum a posteriori model. Therefore, it is still a local analysis and cannot guarantee to provide uncertainty estimation for the global minimum if the problem is highly nonlinear, such as FWI. How to estimate uncertainty without invoking the Gaussian approximation is an important research problem. High-dimensional model spaces and expensive forward computations are the two most challenging issues in this direction. Moreover, extensions to multi-parameter cases also need further exploration.

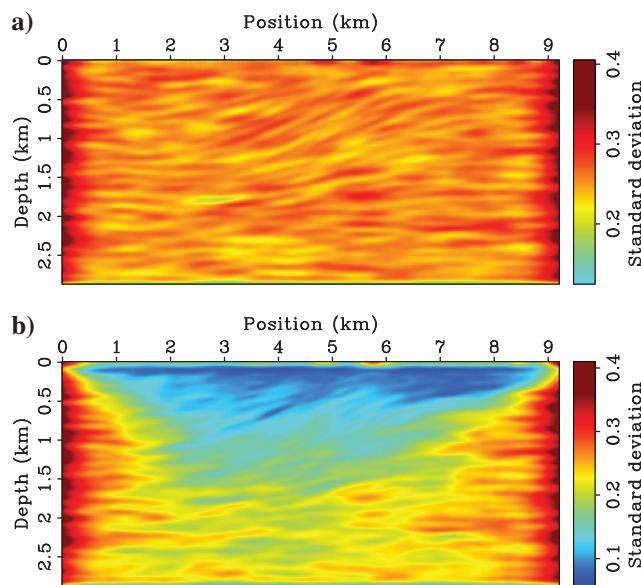


Figure 19. Comparisons of the pointwise standard deviations of the (a) prior and (b) posterior distributions for the Marmousi experiment.

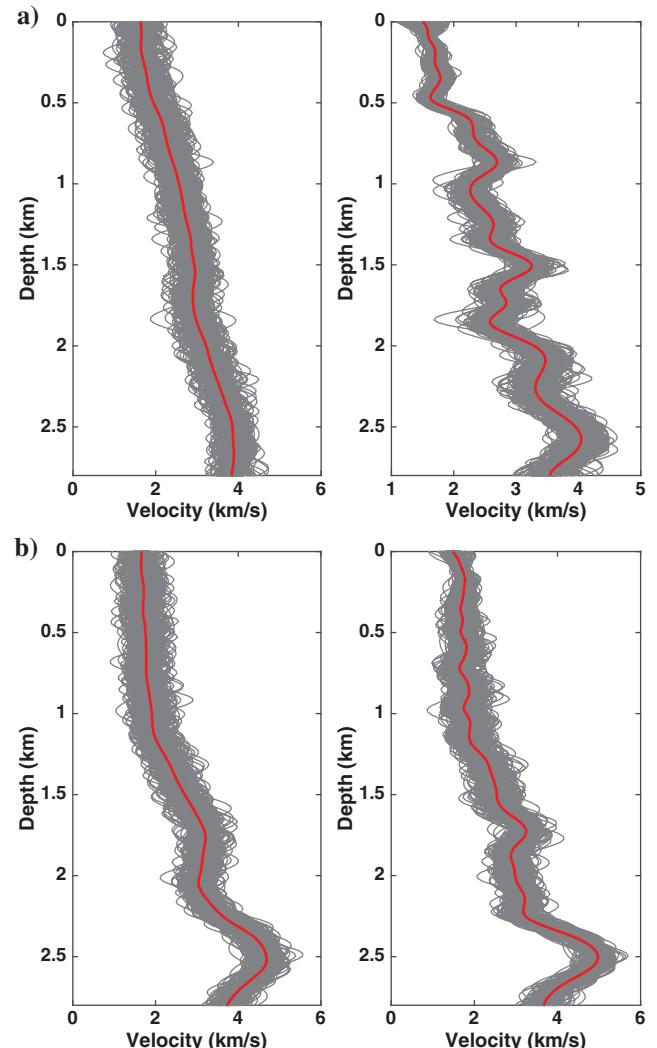


Figure 20. Comparisons of depth profiles between the (left) prior and (right) posterior distributions. (a) Depth profiles in the middle part of the model ($X = 4.48$ km). (b) Depth profiles in the left side of the model ($X = 0.8$ km). Red profiles are mean velocities, and gray profiles are random samples drawn from the prior and posterior distributions.

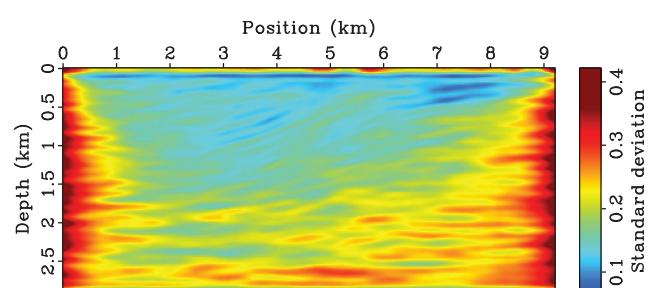


Figure 21. The standard deviation of the posterior PDF based on the PSFs of the Hessian. Only 32 Hessian-vector products are used here compared with 400 used for Figure 19b.

CONCLUSION

In this study, we have adopted a Bayesian inference framework to estimate model uncertainties in FWI. To generate geologically meaningful results, we use a nonstationary a priori covariance based on PWC with local dips measured from migrated images. By linearizing the forward-modeling operators around the maximum a posteriori model, the a posteriori PDF is approximated as a Gaussian distribution. The key step to estimate uncertainties is to efficiently manipulate the posterior covariance. A randomized SVD approach is invoked to approximate and decompose the prior-preconditioned misfit Hessian, which is the key component of the posterior covariance. The spectra of the preconditioned misfit Hessian indicate that data are mostly informative about a low-dimensional subspace of the model parameters. Once we estimate these singular values and vectors, we can draw random samples, as well as compute statistical quantities, such as pointwise standard deviations, from the prior and posterior distributions. To reduce computational costs in the randomized SVD, we use the PSFs of the Hessian to approximate the Hessian-vector products. We demonstrate this procedure in several 2D synthetic experiments. As expected, areas with good data coverages are generally better constrained, i.e., with smaller uncertainties, than the areas with poor data coverages.

ACKNOWLEDGMENTS

This work was supported by the Jackson School Distinguished Postdoctoral Fellowship and the Texas Consortium for Computational Seismology. We thank K. Osypov and B. Crestel for their helpful discussions. We thank K. Van Wijk, A. Fichtner, M. Sambridge, and one anonymous reviewer for their comments and suggestions that helped to improve the manuscript. We thank the Texas Advanced Computing Center for providing computational resources.

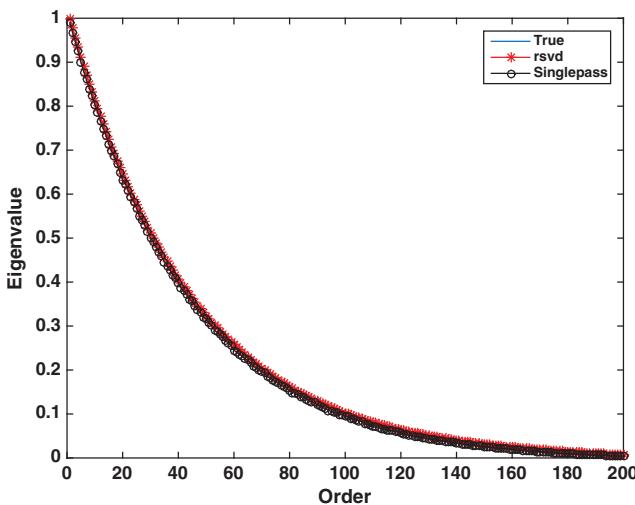


Figure A-1. Comparisons of true (blue) and estimated singular values from classic (red stars) and single-pass (black circles) randomized SVD. 200 random vectors are used to sample a symmetric 3000×3000 matrix.

APPENDIX A

SINGLE-PASS RANDOMIZED SVD

In this study, we use a single-pass randomized SVD (Halko et al., 2011) to estimate the singular values and vectors of the prior-preconditioned Hessian. Compared with classic randomized SVD (Liberty et al., 2007), it only requires the products of a matrix with one set of vectors. The algorithm is illustrated as follows to process a symmetric $m \times m$ matrix \mathbf{H} :

- 1) Generate a random $m \times r$ matrix \mathbf{X} , where r has the same order as the rank of the matrix \mathbf{H} .
- 2) Compute the sample matrix \mathbf{Y} : $\mathbf{Y} = \mathbf{HX}$.
- 3) Perform QR decomposition for the matrix \mathbf{Y} : $\mathbf{Y} = \mathbf{QR}$.
- 4) Solve a linear system to obtain the matrix \mathbf{B} : $\mathbf{Q}^T \mathbf{Y} = \mathbf{B}(\mathbf{Q}^T \mathbf{X})$.
- 5) Perform a SVD for the matrix \mathbf{B} : $\mathbf{B} = \mathbf{U}\Lambda\mathbf{U}^T$.
- 6) Compute the singular vector \mathbf{V} : $\mathbf{V} = \mathbf{QU}$.
- 7) Then, $\mathbf{H} \approx \mathbf{V}\Lambda\mathbf{V}^T$.

We use a simple symmetric 3000×3000 matrix to compare results from classic and single-pass randomized SVD. Two hundred random vectors are used to sample the original matrix. The percentage error between the original and estimated matrices is computed by

$$\text{error} = 100 \times \frac{\sqrt{\sum_{i,j} |\mathbf{H}_{i,j} - \mathbf{H}_{i,j}^{\text{est}}|^2}}{\sqrt{\sum_{i,j} |\mathbf{H}_{i,j}|^2}}, \quad (\text{A-1})$$

where $\mathbf{H}^{\text{est}} = \mathbf{V}\Lambda\mathbf{V}^T$.

Comparisons between the true and estimated singular values from classic and single-pass randomized SVD are illustrated in Figure A-1. The singular values from both algorithms fit the true values well. The percentage errors between the true and estimated matrices are 1.418% and 5.8598% for the classic and single-pass randomized SVD, suggesting that the single-pass algorithm can provide reasonable estimations for the singular values and vectors of large matrices with the low-rank property.

APPENDIX B

HESSIAN-VECTOR PRODUCTS

In the single-pass randomized SVD (Appendix A), computing Hessian-vector products is an important step for estimating the singular values and vectors of the Hessian matrix. Several algorithms for computing Hessian-vector products have been explored in the literature before (Santosa and Symes, 1988; Fichtner and Trampert, 2011; Métivier et al., 2014). In this appendix, we illustrate how to compute Hessian-vector products in the frequency domain and use a simple example to compare results with different approximations.

The first way to compute Hessian-vector products is based on a finite-difference approximation (Zhu et al., 2015),

$$\mathbf{Hx} \approx \mathbf{g}(\mathbf{m} + \mathbf{x}) - \mathbf{g}(\mathbf{m}), \quad (\text{B-1})$$

where \mathbf{H} is the Hessian, the second derivative of the misfit; \mathbf{g} is the gradient, the first derivative of the misfit; \mathbf{m} denotes the background velocity model; and \mathbf{x} is the vector used to sample the Hessian. Zhu et al. (2015) use this approximation to estimate resolution at specific locations for a continental-scale tomographic experiment.

The second way to compute Hessian-vector products is based on adjoint methods (Santosa and Symes, 1988; Fichtner and Trampert, 2011; Métivier et al., 2014). First, we define a least-squares misfit J as follows:

$$J = \frac{1}{2} (\mathbf{d} - \mathbf{Ru})^\dagger (\mathbf{d} - \mathbf{Ru}), \quad (\text{B-2})$$

where \mathbf{u} and \mathbf{d} are the simulated wavefields and observed data. Here, \mathbf{R} is the restriction operator to restrict predictions at receiver locations. The gradient \mathbf{g} is calculated as

$$\mathbf{g} = \frac{\partial J}{\partial \mathbf{m}} = - \left(\frac{\partial \mathbf{u}}{\partial \mathbf{m}} \right)^\dagger \mathbf{R}^\dagger (\mathbf{d} - \mathbf{Ru}). \quad (\text{B-3})$$

The Hessian \mathbf{H} is calculated as

$$\mathbf{H} = \frac{\partial^2 J}{\partial \mathbf{m}^2} = \left(\frac{\partial \mathbf{u}}{\partial \mathbf{m}} \right)^\dagger \mathbf{R}^\dagger \mathbf{R} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{m}} \right) - \left(\frac{\partial^2 \mathbf{u}}{\partial \mathbf{m}^2} \right)^\dagger \mathbf{R}^\dagger (\mathbf{d} - \mathbf{Ru}). \quad (\text{B-4})$$

The Hessian has two contributions (Pratt et al., 1998),

$$\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2, \quad (\text{B-5})$$

where

$$\begin{aligned} \mathbf{H}_1 &= \left(\frac{\partial \mathbf{u}}{\partial \mathbf{m}} \right)^\dagger \mathbf{R}^\dagger \mathbf{R} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{m}} \right), \\ \mathbf{H}_2 &= - \left(\frac{\partial^2 \mathbf{u}}{\partial \mathbf{m}^2} \right)^\dagger \mathbf{R}^\dagger (\mathbf{d} - \mathbf{Ru}). \end{aligned} \quad (\text{B-6})$$

The first contribution \mathbf{H}_1 is the Gauss-Newton approximation Hessian and the second contribution \mathbf{H}_2 comes from multiscattering effects. To compute \mathbf{g} , \mathbf{H}_1 , and \mathbf{H}_2 , we have to know how to evaluate $\partial \mathbf{u} / \partial \mathbf{m}$ and $\partial^2 \mathbf{u} / \partial \mathbf{m}^2$.

The discretized Helmholtz equation with finite-difference or finite-element methods can be expressed as (Marfurt, 1984)

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{f}. \quad (\text{B-7})$$

Then, the first-order partial derivative wavefield is

$$\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u}. \quad (\text{B-8})$$

The second-order partial derivative wavefield is

$$\frac{\partial^2 \mathbf{u}}{\partial \mathbf{m}^2} = -\mathbf{A}^{-1} \left(2 \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \frac{\partial \mathbf{u}}{\partial \mathbf{m}} + \frac{\partial^2 \mathbf{A}}{\partial \mathbf{m}^2} \mathbf{u} \right). \quad (\text{B-9})$$

With the first- and second-order partial derivative wavefields, the gradient is computed as

$$\mathbf{g} = \mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger (\mathbf{A}^\dagger)^{-1} \mathbf{R}^\dagger (\mathbf{d} - \mathbf{Ru}), \quad (\text{B-10})$$

which can be computed by solving the following forward and adjoint problems:

$$\begin{aligned} \mathbf{Au} &= \mathbf{f}, \\ \mathbf{A}^\dagger \mathbf{v} &= \mathbf{R}^\dagger (\mathbf{d} - \mathbf{Ru}), \\ \mathbf{g} &= \text{Re} \left(\mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger \mathbf{v} \right). \end{aligned} \quad (\text{B-11})$$

For the Gauss-Newton Hessian \mathbf{H}_1 , we have

$$\mathbf{H}_1 = \mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger (\mathbf{A}^\dagger)^{-1} \mathbf{R}^\dagger \mathbf{RA}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u}. \quad (\text{B-12})$$

For the second part of the Hessian \mathbf{H}_2 , we have

$$\begin{aligned} \mathbf{H}_2 &= \left[2 \left(\frac{\partial \mathbf{u}}{\partial \mathbf{m}} \right)^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger + \mathbf{u}^\dagger \left(\frac{\partial^2 \mathbf{A}}{\partial \mathbf{m}^2} \right)^\dagger \right] (\mathbf{A}^\dagger)^{-1} \mathbf{R}^\dagger (\mathbf{d} - \mathbf{Ru}), \\ &= -2 \mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger (\mathbf{A}^\dagger)^{-1} \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger \mathbf{v} + \mathbf{u}^\dagger \left(\frac{\partial^2 \mathbf{A}}{\partial \mathbf{m}^2} \right)^\dagger \mathbf{v}, \end{aligned} \quad (\text{B-13})$$

where \mathbf{v} is the adjoint wavefield from equation B-11.

Then, the full Hessian is expressed as

$$\begin{aligned} \mathbf{H} &= \mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger (\mathbf{A}^\dagger)^{-1} \left[\mathbf{R}^\dagger \mathbf{RA}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u} - 2 \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger \mathbf{v} \right] \\ &\quad + \mathbf{u}^\dagger \left(\frac{\partial^2 \mathbf{A}}{\partial \mathbf{m}^2} \right)^\dagger \mathbf{v}. \end{aligned} \quad (\text{B-14})$$

The Hessian-vector product \mathbf{Hx} can be calculated by solving one additional pair of forward and adjoint problems as follows:

$$\begin{aligned} \mathbf{Au}_1 &= \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{ux}, \\ \mathbf{A}^\dagger \mathbf{v}_1 &= \mathbf{R}^\dagger \mathbf{Ru}_1 - 2 \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger \mathbf{vx}, \\ \mathbf{Hx} &= \text{Re} \left(\mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger \mathbf{v}_1 + \mathbf{u}^\dagger \left(\frac{\partial^2 \mathbf{A}}{\partial \mathbf{m}^2} \right)^\dagger \mathbf{vx} \right). \end{aligned} \quad (\text{B-15})$$

Furthermore, the product of the Gauss-Newton approximation Hessian \mathbf{H}_1 with a vector \mathbf{x} can be computed by the following forward and adjoint equations:

$$\begin{aligned} \mathbf{Au}_1 &= \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{ux}, \\ \mathbf{A}^\dagger \mathbf{v}_1 &= \mathbf{R}^\dagger \mathbf{Ru}_1, \\ \mathbf{H}_1 \mathbf{x} &= \text{Re} \left(\mathbf{u}^\dagger \left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^\dagger \mathbf{v}_1 \right). \end{aligned} \quad (\text{B-16})$$

We use a simple example with the background velocity model shown in Figure 2 to compare these three ways for computing the Hessian-vector products. A random vector (Figure B-1a) is used to sample the Hessian. The Hessian-vector products based on equations B-1, B-15, and B-16 are shown in Figure B-1b–B-1d. They are nearly identical to each other, suggesting that (1) the finite-difference approximation is a good approximation for this simple

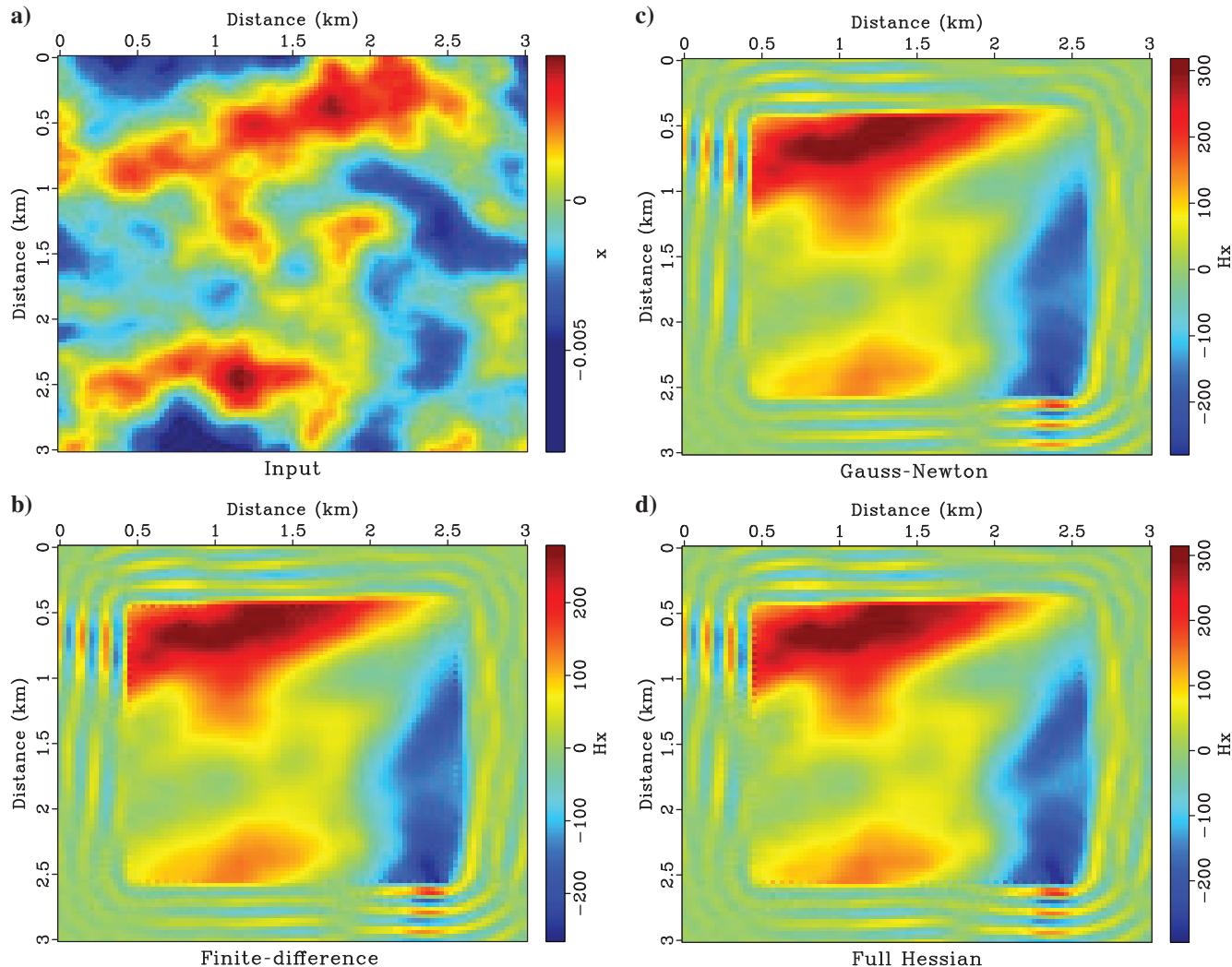


Figure B-1. Comparisons of Hessian-vector products with different approximations. (a) An input random vector. (b-d) Hessian-vector products computed with finite-difference approximation, Gauss-Newton approximation and full Hessian, respectively.

example and (2) the Gauss-Newton Hessian \mathbf{H}_1 makes a major contribution to the full Hessian.

REFERENCES

- An, M., 2012, A simple method for determining the spatial resolution of a general inverse problem: *Geophysical Journal International*, **191**, 849–864, doi: [10.1111/j.1365-246X.2012.05661.x](https://doi.org/10.1111/j.1365-246X.2012.05661.x).
- Backus, G., and F. Gilbert, 1968, The resolving power of gross earth data: *Geophysical Journal of the Royal Astronomical Society*, **16**, 169–205, doi: [10.1111/j.1365-246X.1968.tb00216.x](https://doi.org/10.1111/j.1365-246X.1968.tb00216.x).
- Backus, G., and F. Gilbert, 1970, Uniqueness in the inversion of inaccurate gross earth data: *Philosophical Transactions of the Royal Society A*, **266**, 123–192, doi: [10.1098/rsta.1970.0005](https://doi.org/10.1098/rsta.1970.0005).
- Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2008, Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data: *Geophysics*, **73**, no. 5, VE101–VE117, doi: [10.1190/1.2957948](https://doi.org/10.1190/1.2957948).
- Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: *Geophysics*, **74**, no. 6, WCC105–WCC118, doi: [10.1190/1.3215771](https://doi.org/10.1190/1.3215771).
- Bui-Thanh, T., O. Ghattas, J. Martin, and G. Stadler, 2013, A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion: SIAM Journal on Scientific Computing, **35**, A2494–A2523, doi: [10.1137/12089586X](https://doi.org/10.1137/12089586X).
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473, doi: [10.1190/1.1443880](https://doi.org/10.1190/1.1443880).
- Chen, B., and X. Xie, 2015, An efficient method for broadband seismic illumination and resolution analyses: 85th Annual International Meeting, SEG, Expanded Abstracts, 4227–4231.
- Clapp, R., B. Biondi, and J. Claerbout, 2004, Incorporating geologic information into reflection tomography: *Geophysics*, **69**, 533–546, doi: [10.1190/1.1707073](https://doi.org/10.1190/1.1707073).
- Davis, T., 2004, A column pre-ordering strategy for the unsymmetric-pattern multifrontal method: ACM Transactions on Mathematical Software, **30**, 165–195, doi: [10.1145/992200](https://doi.org/10.1145/992200).
- Epanomeritakis, I., V. Akcelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: *Inverse Problems*, **24**, 34015–34041, doi: [10.1088/0266-5611/24/3/034015](https://doi.org/10.1088/0266-5611/24/3/034015).
- Fichtner, A., B. Kennett, H. Igel, and H. Bunge, 2009, Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods: *Geophysical Journal International*, **179**, 1703–1725, doi: [10.1111/j.1365-246X.2009.04368.x](https://doi.org/10.1111/j.1365-246X.2009.04368.x).
- Fichtner, A., and J. Trampert, 2011, Hessian kernels of seismic data functionals base upon adjoint techniques: *Geophysical Journal International*, **185**, 775–798, doi: [10.1111/j.1365-246X.2011.04966.x](https://doi.org/10.1111/j.1365-246X.2011.04966.x).

- Fichtner, A., and J. Trampert, 2012, Resolution analysis in full waveform inversion: *Geophysical Journal International*, **187**, 1604–1624, doi: [10.1111/j.1365-246X.2011.05218.x](https://doi.org/10.1111/j.1365-246X.2011.05218.x).
- Fichtner, A., and T. van Leeuwen, 2015, Resolution analysis by random probing: *Journal of Geophysical Research*, **120**, 5549–5573, doi: [10.1002/2015JB012106](https://doi.org/10.1002/2015JB012106).
- Fletcher, R., and C. Reeves, 1964, Function minimization by conjugate gradients: *Computer Journal*, **7**, 149–154, doi: [10.1093/comjnl/7.2.149](https://doi.org/10.1093/comjnl/7.2.149).
- Fomel, S., 2002, Applications of plane-wave destruction filters: *Geophysics*, **67**, 1946–1960, doi: [10.1190/1.1527095](https://doi.org/10.1190/1.1527095).
- Fomel, S., and J. Claerbout, 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: *Geophysics*, **68**, 577–588, doi: [10.1190/1.1567228](https://doi.org/10.1190/1.1567228).
- Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model reparameterization using plane-wave construction: *Geophysics*, **71**, no. 5, A43–A47, doi: [10.1190/1.2335609](https://doi.org/10.1190/1.2335609).
- Gouveia, W., and J. Scales, 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: *Journal of Geophysical Research*, **103**, 2759–2779, doi: [10.1029/97JB02933](https://doi.org/10.1029/97JB02933).
- Guitton, A., G. Ayeni, and E. Diaz, 2012, Constrained full-waveform inversion by model reparameterization: *Geophysics*, **77**, no. 2, R117–R127, doi: [10.1190/geo2011-0196.1](https://doi.org/10.1190/geo2011-0196.1).
- Hale, D., 2014, Implementing an anisotropic and spatially varying Matérn model covariance with smoothing filters: *CWP Report*, **815**, 281–290.
- Halko, N., P. Martinsson, and J. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: *SIAM Review*, **53**, 217–288, doi: [10.1137/090771806](https://doi.org/10.1137/090771806).
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration, in J. Bednar, ed., Conference on inverse scattering: Theory and application: SIAM, 206–220.
- Lévéque, J., L. Rivera, and G. Wittlinger, 1993, On the use of the checkerboard test to assess the resolution of tomographic inversions: *Geophysical Journal International*, **115**, 313–318, doi: [10.1111/j.1365-246X.1993.tb05605.x](https://doi.org/10.1111/j.1365-246X.1993.tb05605.x).
- Liberty, E., F. Woolfe, P. Martinsson, V. Rokhlin, and M. Tygert, 2007, Randomized algorithms for the low-rank approximation of matrices: *Proceedings of the National Academy of Science*, **104**, 20167.
- MacCarthy, J., B. Borchers, and R. Aster, 2011, Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross-validation for large geophysical inverse problems: *Journal of Geophysical Research*, **116**, B10304, doi: [10.1029/2011JB008234](https://doi.org/10.1029/2011JB008234).
- Marfurt, K., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave-equations: *Geophysics*, **49**, 533–549, doi: [10.1190/1.1441689](https://doi.org/10.1190/1.1441689).
- Martin, J., L. Wilcox, C. Burstedde, and O. Ghattas, 2012, A stochastic Newton MCMC method for large-scale statistical inverse problems with applications to seismic inversion: *SIAM Journal on Scientific Computing*, **34**, A1460–A1487, doi: [10.1137/110845598](https://doi.org/10.1137/110845598).
- Metivier, L., R. Brossier, J. Virieux, and S. Operto, 2014, Full waveform inversion and the truncated Newton method: Quantitative imaging of complex subsurface structures: *Geophysical Prospecting*, **62**, 1353–1375, doi: [10.1111/1365-2478.12136](https://doi.org/10.1111/1365-2478.12136).
- Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: *Journal of Geophysical Research*, **100**, 12431–12447, doi: [10.1029/94JB03097](https://doi.org/10.1029/94JB03097).
- Mulder, W., and R. Plessix, 2004, How to choose a subset of frequencies in frequency-domain finite-difference migration: *Geophysical Journal International*, **158**, 801–812, doi: [10.1111/j.1365-246X.2004.02336.x](https://doi.org/10.1111/j.1365-246X.2004.02336.x).
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: *Mathematics of Computation*, **35**, 773–773, doi: [10.1090/S0025-5718-1980-0572855-7](https://doi.org/10.1090/S0025-5718-1980-0572855-7).
- Osypov, K., Y. Yang, A. Fournier, N. Ivanova, R. Bachrach, C. Yarman, Y. You, D. Nichols, and M. Woodward, 2013, Model-uncertainty quantification in seismic tomography: Method and applications: *Geophysical Prospecting*, **61**, 1114–1134, doi: [10.1111/1365-2478.12058](https://doi.org/10.1111/1365-2478.12058).
- Paige, C., and M. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: *ACM Transactions on Mathematical Software*, **8**, 43–71, doi: [10.1145/355984.355989](https://doi.org/10.1145/355984.355989).
- Plessix, R., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503, doi: [10.1111/j.1365-246X.2006.02978.x](https://doi.org/10.1111/j.1365-246X.2006.02978.x).
- Pratt, R., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi: [10.1046/j.1365-246X.1998.00498.x](https://doi.org/10.1046/j.1365-246X.1998.00498.x).
- 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, doi: [10.1190/1.1444597](https://doi.org/10.1190/1.1444597).
- Rawlinson, N., A. Fichtner, M. Sambridge, and M. Young, 2014, Seismic tomography and the assessment of uncertainty: *Advances in Geophysics*, **55**, 1–76, doi: [10.1016/bs.agph.2014.08.001](https://doi.org/10.1016/bs.agph.2014.08.001).
- Rimstad, K., P. Avseth, and H. Omre, 2012, Hierarchical Bayesian lithology/fluid prediction: A North Sea case study: *Geophysics*, **77**, no. 2, B69–B85, doi: [10.1190/geo2011-0202.1](https://doi.org/10.1190/geo2011-0202.1).
- Rimstad, K., and H. Omre, 2014, Skew-Gaussian random fields: *Spatial Statistics*, **10**, 43–62, doi: [10.1016/j.spasta.2014.08.001](https://doi.org/10.1016/j.spasta.2014.08.001).
- Ritsema, J., A. Deuss, H. Heijst, and J. Woodhouse, 2011, S40RTS: A degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements: *Geophysical Journal International*, **184**, 1223–1236, doi: [10.1111/j.1365-246X.2010.04884.x](https://doi.org/10.1111/j.1365-246X.2010.04884.x).
- Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: *Reviews of Geophysics*, **40**, 3–1–3–29, doi: [10.1029/2000RG000089](https://doi.org/10.1029/2000RG000089).
- Santosa, F., and W. Symes, 1988, Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology: *Inverse Problems*, **4**, 211–233, doi: [10.1088/0266-5611/4/1/017](https://doi.org/10.1088/0266-5611/4/1/017).
- Sirgue, L., and G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248, doi: [10.1190/1.1649391](https://doi.org/10.1190/1.1649391).
- Tang, Y., and S. Lee, 2015, Multi-parameter full wavefield inversion using non-stationary point-spread functions: 85th Annual International Meeting, SEG, Expanded Abstracts, 1111–1115.
- Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2009, Adjoint tomography of the southern California crust: *Science*, **325**, 988–992, doi: [10.1126/science.1175298](https://doi.org/10.1126/science.1175298).
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: [10.1190/1.1441754](https://doi.org/10.1190/1.1441754).
- Tarantola, A., 2005, Inverse problem theory and method for model parameter estimation: *SIAM*.
- Trampert, J., A. Fichtner, and J. Ritsema, 2013, Resolution tests revisited: The power of random numbers: *Geophysical Journal International*, **192**, 676–680, doi: [10.1093/gji/ggs057](https://doi.org/10.1093/gji/ggs057).
- Tromp, J., C. Tape, and Q. Y. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: *Geophysical Journal International*, **160**, 195–216, doi: [10.1111/j.1365-246X.2004.02453.x](https://doi.org/10.1111/j.1365-246X.2004.02453.x).
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Warner, M., A. Ratcliffe, and T. Nangoo, 2013, Anisotropic 3D full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80, doi: [10.1190/geo2012-0338.1](https://doi.org/10.1190/geo2012-0338.1).
- Yao, Z., R. Roberts, and A. Tryggvason, 1999, Calculating resolution and covariance matrices for seismic tomography with the LSQR method: *Geophysical Journal International*, **138**, 886–894, doi: [10.1046/j.1365-246x.1999.00925.x](https://doi.org/10.1046/j.1365-246x.1999.00925.x).
- Zhang, H., and C. Thurber, 2007, Estimating the model resolution matrix for large seismic tomography problems based on Lanczos bidiagonalization with partial reorthogonalization: *Geophysical Journal International*, **170**, 337–345, doi: [10.1111/j.1365-246X.2007.03418.x](https://doi.org/10.1111/j.1365-246X.2007.03418.x).
- Zhang, J., and G. McMechan, 1995, Estimation of resolution and covariance for large matrix inversions: *Geophysical Journal International*, **121**, 409–426, doi: [10.1111/j.1365-246X.1995.tb05722.x](https://doi.org/10.1111/j.1365-246X.1995.tb05722.x).
- Zhu, H., E. Bozdag, T. Duffy, and J. Tromp, 2013, Seismic attenuation beneath Europe and the North Atlantic: Implications for water in the mantle: *Earth and Planetary Science Letters*, **381**, 1–11, doi: [10.1016/j.epsl.2013.08.030](https://doi.org/10.1016/j.epsl.2013.08.030).
- Zhu, H., E. Bozdag, D. Peter, and J. Tromp, 2012, Structure of the European upper mantle revealed by adjoint tomography: *Nature Geoscience*, **5**, 493–498, doi: [10.1038/ngeo1501](https://doi.org/10.1038/ngeo1501).
- Zhu, H., E. Bozdag, and J. Tromp, 2015, Seismic structure of the European upper mantle based on adjoint tomography: *Geophysical Journal International*, **201**, 18–52, doi: [10.1093/gji/ggu492](https://doi.org/10.1093/gji/ggu492).
- Zhu, H., and J. Tromp, 2013, Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean: *Science*, **341**, 871–875, doi: [10.1126/science.1241335](https://doi.org/10.1126/science.1241335).