Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion

Changsoo Shin*, Kwangjin Yoon‡, Kurt J. Marfurt**, Keunpil Park§, Dongwoo Yang‡, Harry Y. Lim‡, Seunghwan Chung§, and Sungryul Shin§§

ABSTRACT

Linearized inversion of surface seismic data for a model of the earth's subsurface requires estimating the sensitivity of the seismic response to perturbations in the earth's subsurface. This sensitivity, or Jacobian, matrix is usually quite expensive to estimate for all but the simplest model parameterizations. We exploit the numerical structure of the finite-element method, modern sparse matrix technology, and source–receiver reciprocity to develop an algorithm that explicitly calculates the Jacobian matrix at only the cost of a forward model solution. Furthermore, we show that we can achieve improved subsurface images using only one inversion iteration through proper scaling of the image by a diagonal approximation of the Hessian matrix, as predicted by the classical Gauss-Newton method. Our method is applicable to the full suite of wave scattering problems amenable to finite-element forward modeling. We demonstrate our method through some simple 2-D synthetic examples.

INTRODUCTION

The primary goal of reflection seismology is to map structural and stratigraphic features in the earth's subsurface. Geophysicists have used a variety of mathematical tools to map such features, from efficient, fast, and kinematically accurate Kirchhoff schemes based on asymptotic ray theory (Bleistein, 1987) to more amplitude-friendly but computationally expensive wave equation extrapolation-based techniques (e.g., Berkhout and Wapenaar, 1993; Gazdag and Squazzero, 1984).

Whitmore (1983), Baysal et al. (1983), and McMechan (1983) use the hyperbolic wave equation to image the subsurface. In this reverse time migration technique, the subsurface is imaged by a zero-lag convolution between the forward-modeled wavefields and the backward-propagated residuals. Rearranging the sequence of mathematical operations allows us to perform common offset wave equation imaging (Ehinger et al., 1996) but at two to four times the cost.

Lailly (1984) was perhaps the first person to show that the imaging of seismic reflectors is the first approximation to the inversion of seismic data for obtaining the unknown parameters of impedance and velocity at every point in the earth's subsurface. In the seismic inverse problem, geophysicists use iterative nonlinear optimization techniques such as the gradient, Gauss-Newton, and full Newton methods to minimize the least-squares residual between the measured seismic data and the forward-modeled responses (Pratt et al., 1998). The key to the solution of these inverse problems is the Jacobian, the matrix that represents the change in the data to small perturbations of velocity and impedance in the earth's subsurface. In the inversion based on asymptotic ray theory, the Jacobian can be computed efficiently (Lambaré et al., 1992). However, in the waveform-based inverse problem, the explicit calculation of the Jacobian $\mathbf{J}$ can be extremely expensive (Woodhouse and Dziewonski, 1984). Lailly (1984), Tarantola (1984), Geller and Hara (1993), Mora (1987), and Pratt et al. (1998) avoid explicitly calculating the Jacobian and instead calculate the steepest descent direction of the objective function $\mathbf{J}^T\mathbf{d}$, where $\mathbf{d}$ is the observed data, by using an adjoint state technique. To

Table of symbols

σ_i = acoustic impedance at element i
v_i = approximate velocity at element i
ω = frequency, rad/s
ρ_i = density at node i
$\mathbf{M}$ = $n \times n$ finite-element mass matrix
$\mathbf{C}$ = $n \times n$ finite-element damping matrix
$\mathbf{K}$ = $n \times n$ finite-element stiffness matrix
$\mathbf{S}$ = $-\omega^2\mathbf{M} + i\omega\mathbf{C} + \mathbf{K}$ = $n \times n$ complex impedance matrix
$\mathbf{f}_j$ = jth $n \times 1$ source distribution along the surface of the earth
$\mathbf{d}_j$ = $n \times 1$ data vector generated by $\mathbf{f}_j$
$\mathbf{F}$ = $n \times ns$ matrix composed of ns source vectors $\mathbf{f}_j$
$\mathbf{D}$ = $n \times ns$ matrix composed of the ns source vectors $\mathbf{d}_j$ from source vectors $\mathbf{F}$
$\mathbf{m}$ = model vector composed of elements σ_i
$\bar{\mathbf{f}}_{i,j} = \frac{\partial \mathbf{S}}{\partial m_i}\mathbf{d}_j$ = virtual source vector from a perturbation of model parameter m_i for the jth true source
$\mathbf{J}_j = \frac{\partial \mathbf{d}_j}{\partial \mathbf{m}} = -\mathbf{S}^{-1}\bar{\mathbf{f}}_j$ = Jacobian matrix for the jth virtual source vector and frequency ω

calculate the gradient of the l_2 norm of the residuals, they back-propagated the residuals and then correlated them with the forward-propagated wavefields. Both reverse-time migration and waveform inversion were based on multichannel common shot experiments. It is computationally inefficient to use the adjoint state technique to process data that have been sorted to common midpoint (CMP) or common-offset gathers.

In this paper, we compute the Jacobian explicitly, allowing us to generate reverse-time migration and waveform inversion algorithms that can handle data of arbitrary source–receiver configurations, including both CMP and common offset sorted data sets. We begin by exploiting the efficient forward and backward substitution properties of sparse matrices associated with frequency-domain, finite-element solutions outlined by Marfurt and Shin (1989) and exploited by Shin (1988) and Pratt et al. (1998) to calculate the Jacobian matrix needed for seismic imaging and impedance inversion. We then invoke the principle of source–receiver reciprocity to reduce the required number of computations in the Jacobian calculation. We conclude by showing how this approach can be applied to acoustic imaging using the full-wave equation, including multiples, head waves, and primaries.

PARAMETERIZATION AND JACOBIAN CALCULATION

Following Chavent and Jacewitz (1995), we parameterize our subsurface by an $N_x \times N_z$ rectangular grid of nodes. At each node i, we identify a velocity v_i and an impedance $\sigma_i = \rho_i v_i$, where ρ_i is the density. For seismic imaging and impedance inversion, we assume that we have estimated smoothed, reasonably accurate values of v_i by independent means, such as traveltime tomography (e.g., Stork, 1992). In this manner, we define our unknown model parameter vector $\mathbf{m}$ to be

$$\mathbf{m} \equiv \left(\sigma_1, \sigma_2, \ldots, \sigma_{N_xN_z}\right)^T \qquad (1)$$

We now write the discretized equation of motion (Marfurt and Shin, 1989) in the frequency domain as

$$[-\omega^2\mathbf{M}(\mathbf{m}) + i\omega\mathbf{C}(\mathbf{m}) + \mathbf{K}(\mathbf{m})]\,\mathbf{d}(\omega)$$
$$= \mathbf{S}(\mathbf{m}, \omega)\,\mathbf{d}(\omega) = \mathbf{f}(\omega), \qquad (2)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ denote the mass, damping, and stiffness matrices (e.g., Zienkiewicz, 1977), respectively; $\mathbf{d}(\omega)$ is the modeled data at angular frequency ω; $\mathbf{f}(\omega)$ the source vector; $\mathbf{S} \equiv -\omega^2\mathbf{M} + i\omega\mathbf{C} + \mathbf{K}$ is the complex impedance matrix; and $i = \sqrt{-1}$. We solve the modeled data vector $\mathbf{d}$ by inverting the matrix $\mathbf{S}$ in equation (2):

$$\mathbf{d} = \mathbf{S}^{-1}\mathbf{f}. \qquad (3)$$

By including intrinsic attenuation and absorbing boundary conditions, we avoid zero eigenvalues in $\mathbf{S}$. Further stability is obtained by setting $\omega = \omega + i\alpha$, which, by use of the shifting theorem, allows us to suppress wraparound when transforming our solution back to the time domain (Marfurt and Shin, 1989). The amplitude and phase of the modeled data $\mathbf{d}$ depend on the frequency content of the source vector $\mathbf{f}$. By solving for the band-limited approximation to the temporal impulse response, we can easily obtain seismograms from different seismic sources (dynamite, air gun, etc.) by simple multiplication. If we set the value of the source vector $\mathbf{f}$ at the ith node to be 1 and the value at all other nodes to be 0, we recognize that $\mathbf{S}^{-1}$ in equation (3) is a finite-element approximation to the Green's function. If we denote a sequence of source vectors by $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{ns})$, where ns is the number of sources or shots we wish to simulate, and the corresponding solution by $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{ns})$, then

$$\mathbf{D} = \mathbf{S}^{-1}\mathbf{F} \qquad (4)$$

for each frequency ω. For the acoustic and elastic wave equations, the complex impedance matrix $\mathbf{S}$ and its inverse $\mathbf{S}^{-1}$ are symmetric. In the acoustic or elastic problem, because $\mathbf{S}^{-1}$ is symmetric, d_j^i, the data modeled at node i from a unit source at node j, is equal to d_i^j, the data modeled at node j from a unit source at node i. In the elastic problem, i and j correspond to the degrees of freedom. Using this source and receiver reciprocity, we can efficiently implement the explicit calculation of the Jacobian $\mathbf{J}$ (see the Appendix).

SEISMIC IMAGING

Hagedoorn (1954) discusses two approaches for migrating seismic data. In his method of maximum convexity, the amplitude of each data point for a given source–receiver pair is broadcast over an ellipse in depth. This is the approach used in most implementations of wavefield extrapolation migration algorithms, including all implementations of reverse time migration. In Hagedoorn's method of maximum convexity, the reflectivity at each depth point is estimated by summing along the hyperbolic diffraction curve for a common shot gather. This approach is commonly used in many prestack Kirchhoff migration algorithms. Our estimate of the depth model for a given common shot gather,

$$\mathbf{m} = \int_{-\omega_{max}}^{\omega_{max}} \mathbf{J}^T(\mathbf{m}, \omega)\,\tilde{\mathbf{d}}(\omega)\,d\omega, \qquad (5)$$

is the frequency-domain implementation of Hagedoorn's second approach, where ω is the angular frequency, ω_{max} is the maximum frequency of the source wavelet, and $\tilde{\mathbf{d}}$ is the complex conjugate of $\mathbf{d}$. Our current implementation differs in that equation (5) includes all events—primaries, multiples, and headwaves—in the calculation.

SEISMIC INVERSION

If we use the Gauss–Newton method to estimate our unknown model parameters (impedance) $\delta\mathbf{m}$, we obtain

$$\delta\mathbf{m} = -\mathbf{H}^{-1}\nabla_{\mathbf{m}}E \qquad (6)$$

(Pratt et al., 1998), where $\delta\mathbf{m}$ is the change in the model parameter vector, $\mathbf{H}$ is the Hessian matrix, and $\nabla_{\mathbf{m}}E$ is the gradient vector of the objective function E, which is given by $E \doteq (1/2)\,\delta\mathbf{d}^T\delta\mathbf{d}$, where $\delta\mathbf{d} = \mathbf{d}(\mathbf{m}) - \mathbf{d}_{obs}$ and where $\mathbf{d}_{obs}$ is the observed data measured along the earth's surface. Near the global minimum of the objective function, we can approximate $\mathbf{H}$ by $\mathbf{H}_a$,

$$\mathbf{H}_a = \mathbf{J}^T\mathbf{J}, \qquad (7)$$

and the gradient $\nabla_{\mathbf{m}}E$ by

$$\nabla_{\mathbf{m}}E = \frac{\partial}{\partial\mathbf{m}}\left(\frac{1}{2}\,\delta\mathbf{d}^T\delta\mathbf{d}\right) = \mathbf{J}^T\delta\mathbf{d}. \qquad (8)$$

If we have not modeled any data in equation (8), then $\mathbf{d}(\mathbf{m}) = 0$ and $\delta\mathbf{d} = -\mathbf{d}_{obs}$, such that equation (6) degenerates into what we call least-squares migration (Gray, 1997; Lines, 1999, personal communication). In this case, the normal equations can be written as (Pratt et al., 1998)

$$\begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,N_e} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,N_e} \\ \cdots & \cdots & \cdots & \cdots \\ \phi_{N_e,1} & \phi_{N_e,2} & \cdots & \phi_{N_e,N_e} \end{bmatrix} \begin{bmatrix} \delta m_1 \\ \delta m_2 \\ \vdots \\ \delta m_{N_e} \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_{N_e} \end{bmatrix}, \qquad (9)$$

where $N_e = N_x N_z$ and where

$$\phi_{ij} = \int_{-\omega_{max}}^{\omega_{max}} \int_0^{x_{max}} \frac{\partial d(x,\mathbf{m},\omega)}{\partial m_i} \frac{\partial \tilde{d}(x,\mathbf{m},\omega)}{\partial m_j}\, dx\, d\omega,$$

$$R_i = -\int_{-\omega_{max}}^{\omega_{max}} \int_0^{x_{max}} \frac{\partial d(x,\mathbf{m},\omega)}{\partial m_i} \tilde{d}_{obs}(x,\omega)\, dx\, d\omega,$$

$i = 1,\ldots,N_e$, and $j = 1,\ldots,N_e$. Here, N_e is the number of the elements in the model, x_{max} is the largest offset distance, $\partial d(x,\mathbf{m},\omega)/\partial m_i$ is the Fourier transform of the partial derivative wavefield, and $d_{obs}(x,\omega)$ is the Fourier transform of the observed wavefield.

Before discussing damped least-squares migration, we need to examine the approximate Hessian matrix $\mathbf{H}_a$. The main diagonal element of the approximate Hessian matrix is the zero-lag value of the autocorrelation of the partial derivative wavefield, which always has a positive value. The off-diagonal elements of the approximate Hessian matrix are the zero-lag values of the crosscorrelation of the partial derivative wavefields. In the high-frequency limit and for models having relatively smooth impedance variation, the approximate Hessian matrix is diagonally dominant (Pratt et al., 1998; Gray, 1997; Lambaré et al., 1992; Jin et al., 1992). Lambaré et al. (1992) show that for uniformly sampled data around a scattering point, the Hessian matrix can be approximated as a delta function (diagonal element of the Hessian matrix). However, for band-limited seismic sources, the structure of the approximate Hessian matrix becomes banded (Pratt et al., 1998). We assume that the diagonal elements of the approximate Hessian matrix represent the approximate Hessian matrix near the solution, giving us a very

good preconditioner for an iterative solution. Under this assumption, the problem of large-scale migration and inversion of seismic data becomes more tractable.

The amplitude of R_i in equation (9) decreases with depth almost exponentially. Gray (1997) comments that this is a poorly scaled migration. After dividing the right-hand vector in equation (9) by the diagonal elements of the approximate Hessian matrix, the change in our model vector ($\delta\mathbf{m}$) in the least-squares migration is the scaled or normalized migrated depth image. We can use $\delta\mathbf{m}$ in the waveform inversion to obtain an updated initial impedance model, $\mathbf{m} + \delta\mathbf{m}$. In practice, $\mathbf{H}_a$ may be ill conditioned or singular in zones having poor seismic illumination, such that we need to stabilize our inversion by adding a damping factor λ to the main diagonal elements of $\mathbf{H}_a$:

$$\delta\mathbf{m} = (\mathrm{diag}\,\mathbf{H}_a + \lambda\mathbf{I})^{-1}\mathbf{J}^T\delta\mathbf{d}. \qquad (10)$$

To illustrate how much the approximate Hessian matrix normalizes the poorly scaled migrated image, we examine the 2-D geological model shown in Figure 1. We generate 519 common shot gather seismograms using the frequency-domain, finite-element method described by Marfurt and Shin (1989). The source–receiver interval is 5 m. We model 225 discrete frequencies ranging from 0 to 75 Hz at a frequency interval of 0.333 Hz. In imaging the synthetic data, we consider only the Jacobian because of the perturbation of density. In Figure 2, we display the unscaled image $\mathbf{J}^T\mathbf{d}$, showing an unbalanced image having a high-amplitude discontinuity at the first reflector, high-amplitude artifacts near the shot points, and a
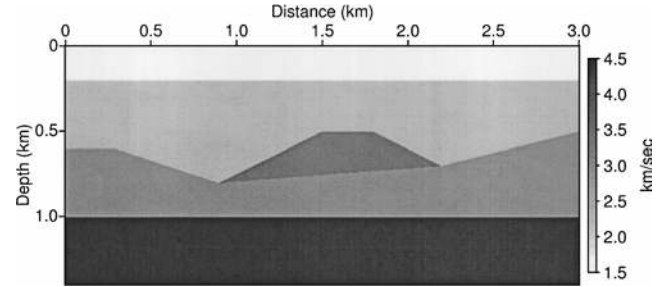

FIG. 1. A 2-D geological model constructed to test the least-squares migration using the Gauss-Newton method. The horizontal distance is discretized into 601 columns, and the vertical distance is discretized into 281 rows.
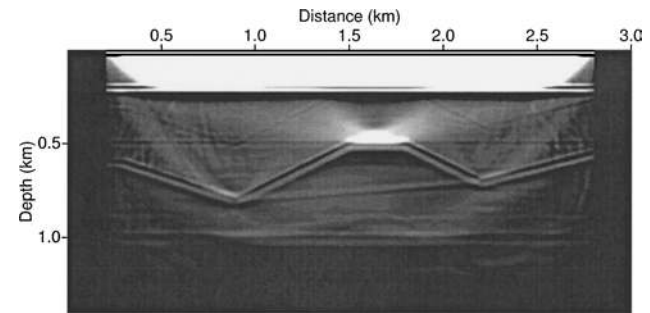

FIG. 2. A depth image obtained by using only the right-hand side vector of equation (9) without any normalization or scaling. As noted by Gray (1997), this image is poorly scaled. Similar scaling can be obtained by a prestack reverse time migration technique.

barely discernible image of the deepest reflector. In Figure 3, we display our preconditioner (diag $\mathbf{H}_a + \lambda\mathbf{I}$)$^{-1}$. After collecting diagonal elements of the approximate Hessian matrix at grid points below the surface, their values are plotted against the depth axis. As the depth increases, the reciprocal value increases such that multiplication of the unscaled migrated image (Figure 2) by (diag $\mathbf{H}_a + \lambda\mathbf{I}$)$^{-1}$ looks like an automatic gain control (AGC) of the subsurface image. In Figure 4a, we show the scaled subsurface image calculated by equation (10) with the damping term $\lambda = 0$, while in Figure 4b we show the scaled subsurface image obtained by setting $\lambda = 0.01$ (by adding 1% of the maximum value of diag $\mathbf{H}_a$). Without proper damping, the depth image at 1.2 km suffers from overamplication, as shown in Figure 4a. We note simple multiples, m_1 and m_2, in both Figures 4a and 4b. We also note a backward-traveling multiple, m_{-1}. Since we exploit the finite-element implementation of the full wave equation, we are able not only to pre-

dict all causal multiples but also to generate additional false images.

To evaluate the sensitivity to an incorrect velocity model, we examine the three velocity models shown in Figure 5: a flat-layered model, an incorrect estimate of the true model, and a smoother version of the true model. The depth images displayed in Figures 6a–c correspond to the models shown in Figures 5a–c. We note that Figure 6b provides a better image than Figure 6a. We also note an excellent false image m_{-1} of the structure above its true location, where the multiple has been migrated with the correct velocity model. The image displayed in Figure 6c is the migrated result when we used a smoothed version of the correct velocity model. The simple multiples, m_1 and m_2 still exist. However, when using a smooth velocity model (Figure 5c), we could eliminate the false image m_{-1}.

EXPERIENCE OF MARMOUSI DATA

We now apply the least-squares migration to the Marmousi data set (Bourgeois et al., 1991). The Marmousi velocity model is shown in Figure 7. We use two sets of Marmousi data. One is the original data set downloaded from an Internet site, and the other is our own data set generated using frequency-domain,

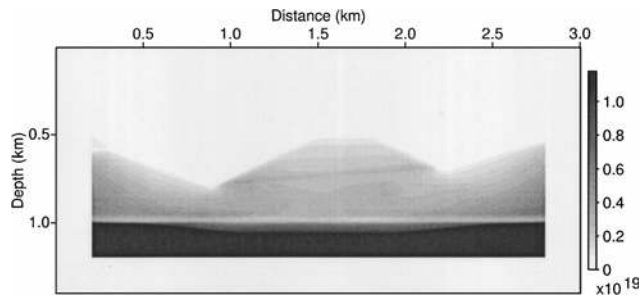
FIG. 3. Reciprocals of the diagonal elements of the approximate Hessian matrix $\mathbf{H}_a = \mathbf{J}^T\mathbf{J}$, which is inversely proportional to the subsurface illumination. Note that (diag $\mathbf{H}_a$)$^{-1}$ increases rapidly with depth.
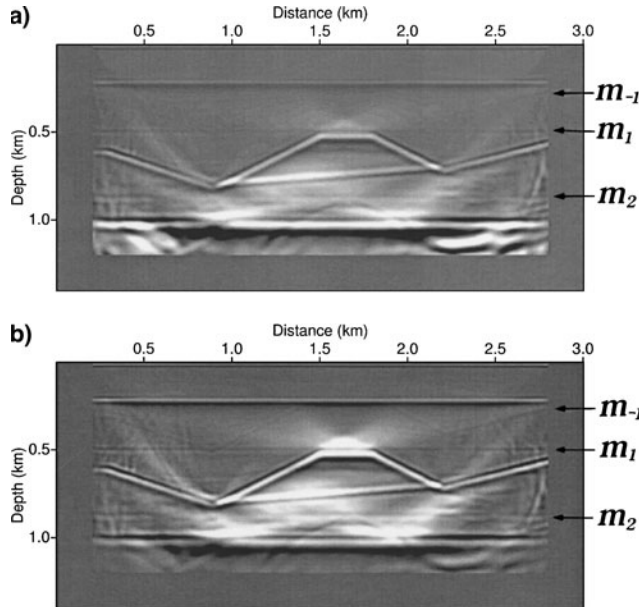

FIG. 4. Depth images obtained by using equation (10) with the damping terms (a) $\lambda = 0$ and (b) $\lambda = 0.01$. Compared with the image shown in Figure 3, these depth images are properly scaled or normalized. Multiple images of the first reflector can be seen at the 0.5-km depth point. Small multiple images of the caplike structure are seen at depths ranging from 0.9 to 1.2 km.
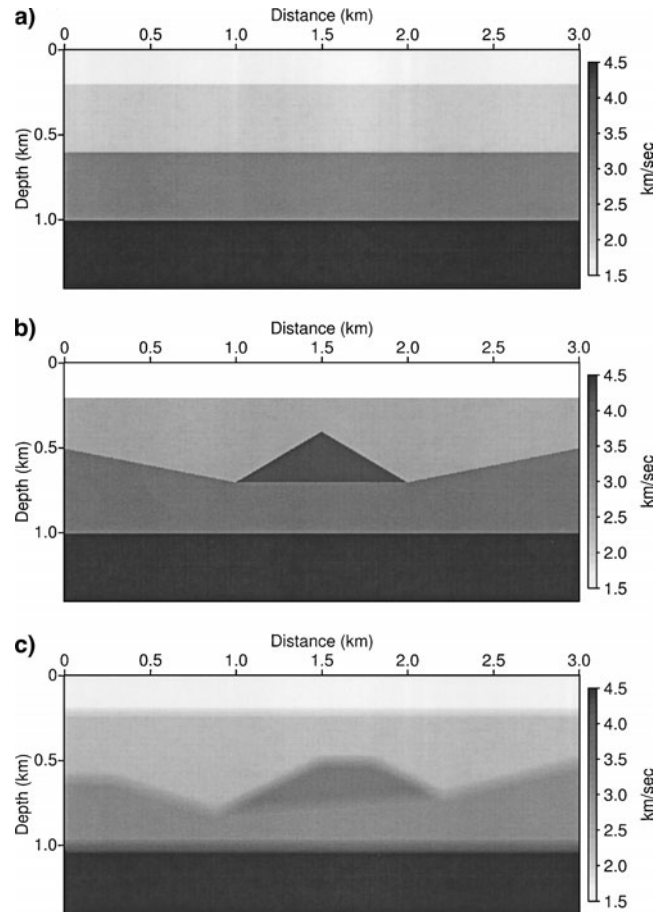

FIG. 5. Three initial velocity models chosen to test the sensitivity of our inversion to errors in velocity. (a) A horizontally layered velocity model. (b) An initial model whose velocity structure is different from the true model shown in Figure 1. (c) A velocity model obtained by smoothing the true model shown in Figure 1.

finite-element modeling. Figure 8 shows shot-gathered seismograms of each data set. The original data were synthesized using a signature of a water gun array obtained by digitizing a real near-field signal and receiver groups consisting of five hydrophones. The depths of shots and receivers were 8 and 12 m, respectively. In our second data set, the sources and the receivers were fixed at 4 m depth. In contrast to the original data, we did not use group arrays to mimic recent advances in acquisition hardware. We used a source wavelet of a Gaussian first derivative function, $-2at \exp^{-at^2}$, with major frequency at 15.6 Hz.
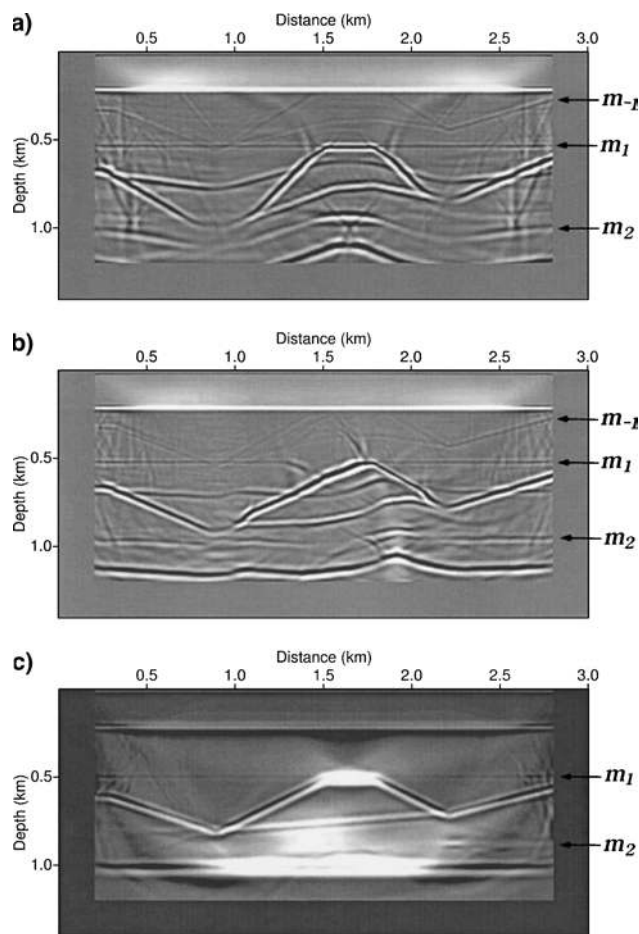

FIG. 6. The final depth images corresponding to three different initial velocity models shown in Figure 5.
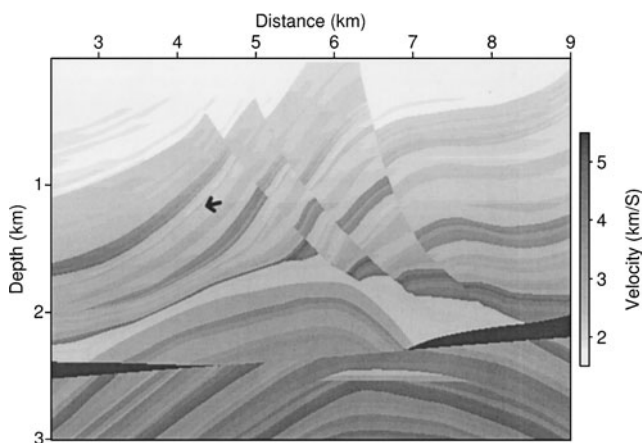

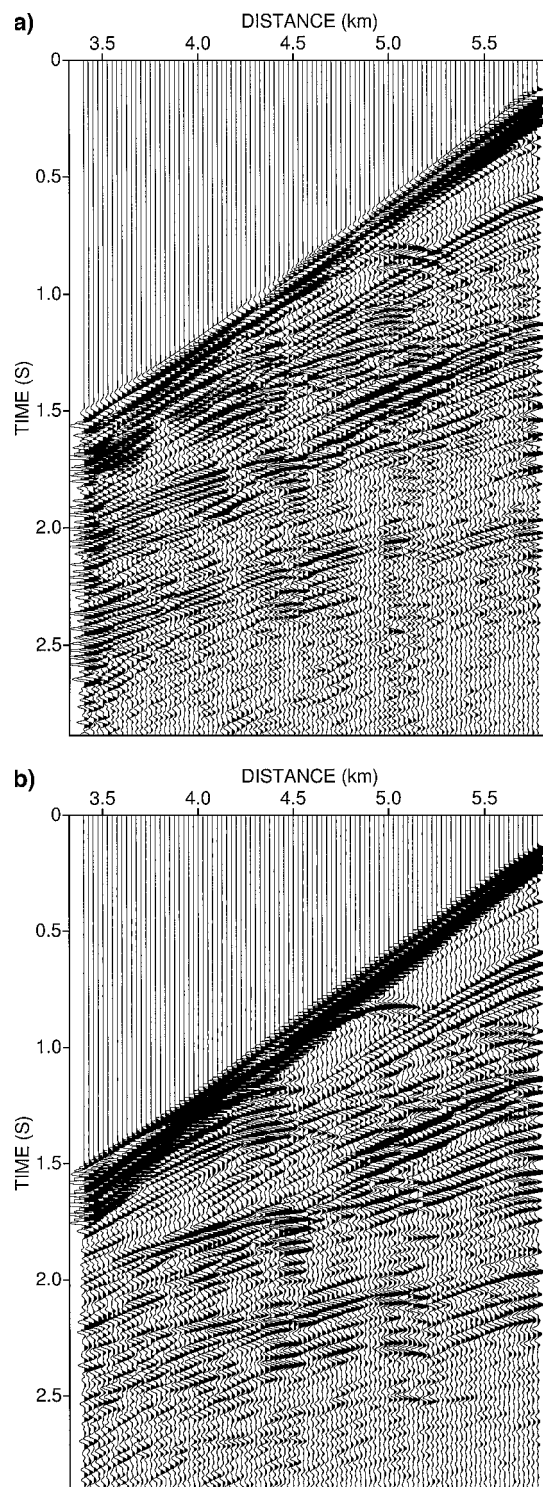FIG. 7. The Marmousi velocity model (after Bourgeois et al., 1991).

FIG. 8. Shot-gathered seismograms of (a) the original data set and (b) our data set generated by frequency-domain, finite-element modeling.

Figure 9 shows two elements of the huge Jacobian matrix $\mathbf{J}^T$, obtained when we locate a source at the distance of 5975 m and perturb the density and the bulk modulus of an element whose location is indicated by an arrow in Figure 7. Figures 10a
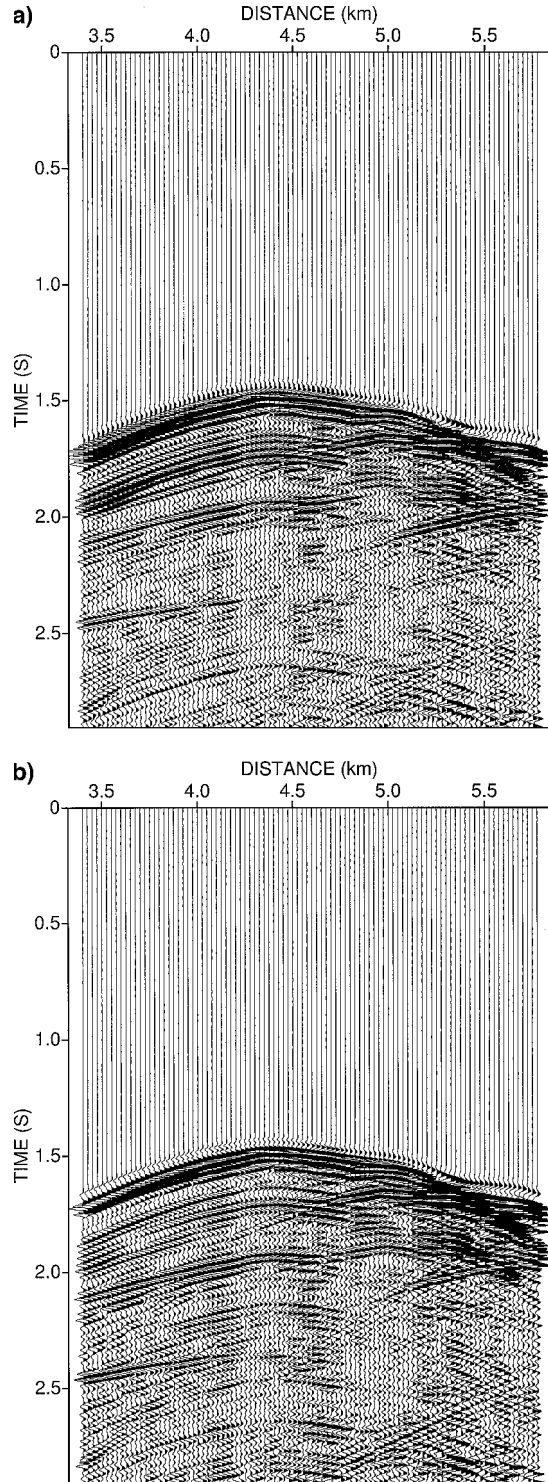
and 10b are final results of least-squares migration using the original data set and our data set, respectively. In Figure 11, we display images obtained by applying AGC to the unscaled images, $\mathbf{J}^T\mathbf{d}$. As Shin and Chung (1999) claim, the AGC to the unscaled image $\mathbf{J}^T\mathbf{d}$ is qualitatively equivalent to dividing it by the diagonal elements of the approximated Hessian matrix.

In the images shown in Figures 10 and 11, we can see well-defined reservoir zones and discover our data set gives a better image than the downloaded data set. We speculate that our data set provides a superior image over the downloaded data because of the intrinsic difference between the finite-difference model and the frequency-domain, finite-element model; the different alignments of sources and receivers between the two data sets; and the use of a correct source wavelet and the single phone recording in modern acquisition systems.

CONCLUSION

By exploiting sparse matrix technology and source–receiver reciprocity of the numerical Green's function of the acoustic wave equation, we have developed a reasonably efficient and very general means of computing the partial derivative

FIG. 9. Partial derivative seismograms, two elements of the Jacobian ($\mathbf{J}^T$), generated by perturbing (a) the density and (b) the bulk modulus of the element whose location is indicated by an arrow in Figure 7 when a source is given at the distance 5975 m.
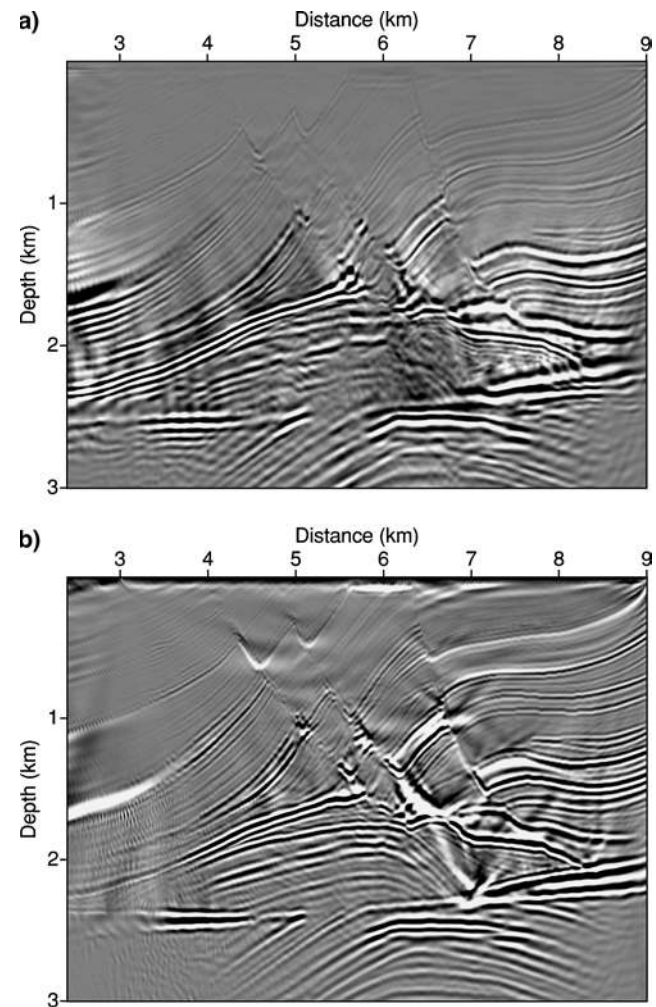

FIG. 10. Results obtained by applying least-squares migration to (a) the original data set and (b) our higher resolution recomputed data set.
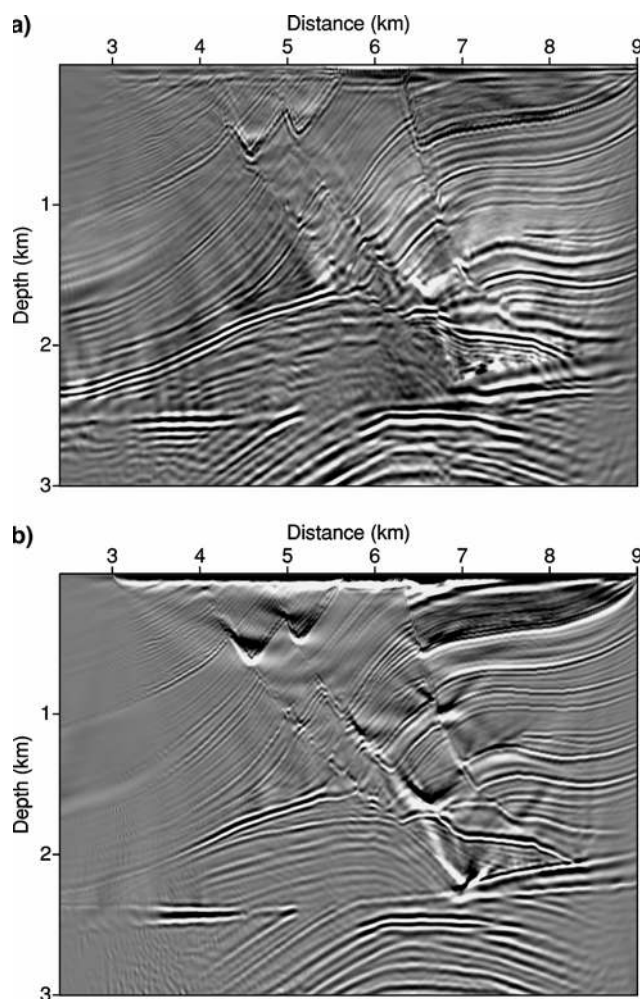
a)

b)

FIG. 11. Results obtained by applying AGC to the unscaled images $\mathbf{J}^T\mathbf{d}$, obtained from (a) the original data set and (b) our higher resolution recomputed data set.

seismograms (Jacobian matrix) necessary for some seismic inversion and imaging algorithms. A direct computation of the partial derivative seismograms allows our least-squares migration and inversion to handle any source–receiver configuration, including CMP and common offset gathers. By normalizing the poorly scaled image using the diagonal approximation to the Hessian, we can obtain high-fidelity subsurface images whose amplitude is controlled by the geological model and acquisition geometry rather than by some statistical method. While our method may be expensive, we strongly feel that the efficient calculation of the partial derivative wavefields will allow us to perform full waveform inversion for a large class of problems

amenable to finite-element solution, including the inversion for shear impedance, anisotropy, and attenuation.

REFERENCES

Baysal, E., Kosloff, D. D., and Sherwood, J. W. C., 1983, Reverse time migration: Geophysics, 48, 1514–1524.
Berkhout, A. J., and Wapenaar, C. P. A., 1993, A unified approach to acoustical reflection imaging, part II: The inverse problem: J. Acoust. Soc. Am., **93**, 2017–2023.
Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, 931–942.
Bourgeois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P., and Versteeg, R., 1991, Marmousi, model and data, in Versteeg, R., and Grau, G., Eds., The Marmousi experience, Proceedings of the 1990 EAEG workshop on Practical Aspects of Seismic Data Inversion: EAEG, 5-16.
Chavent, G., and Jacewitz, C. A., 1995, Determination of background velocities by multiple migration fitting: Geophysics, **60**, 476–490.
Ehinger, A., Lailly, P., and Marfurt, K. J., 1996, Green's function implementation of common-offset, wave-equation migration: Geophysics, **61**, 1813–1821.
Gazdag, J., and Squazzero, P., 1984, Migration of seismic data by phase-shift plus interpolation: Geophysics, **49**, 124–131.
Geller, R. J., and Hara, T., 1993, Two efficient algorithms for iterative linearized inversion of seismic waveform data: Geophys. J. Internat., **115**, 699–710.
Gray, S. H., 1997, True-amplitude seismic migration: A comparison of three approaches: Geophysics, **62**, 929–936.
Hagedoorn, J. G., 1954, A process of seismic reflection interpretation: Geophys. Prosp., **2**, 85–127.
Jin, S., Madariaga, R., Virieux, J., and Lambaré, G., 1992, Two dimensional asymptotic iterative elastic inversion: Geophys. J. Internat., **108**, 575–588.
Lailly, P., 1984, The seismic inverse problem as a sequence of before stack migration, in Bednar, J. B., Redner, R., Robinson, E., and Weglein, A., Eds., Conference on inverse scattering: Theory and application: Soc. Ind. Appl. Math., 206–220.
Lambaré, G., Virieux, J., Madariaga, R., and Jin, S., 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, **57**, 1138–1154.
Marfurt, K. J., and Shin, C. S., 1989, The future of iterative modeling of geophysical exploration, in Eisner, E., Ed., Supercomputers in seismic exploration: Elsevier, 203–228.
McMechan, G. A., 1983, Migration by exploration of time dependent boundary values: Geophys. J. Roy. Astr. Soc., **71**, 613–628.
Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, **52**, 1211–1228.
Pratt, R. G., Shin, C., and Hicks, G. J., 1998, Gauss-Newton and full Newton methods in frequency–space seismic waveform inversion: Geophys. J. Internat., **133**, 341–362.
Shin, C., 1988, Nonlinear elastic wave inversion by blocky parameterization: Ph.D. thesis, Univ. of Tulsa.
Shin, C., and Chung, S., 1999, Understanding CMP stacking hyperbola in terms of partial derivative wavefield: Geophysics, **64**, 1774–1782.
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.
Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: 53rd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 827–830.
Woodhouse, J. H., and Dziewonski, A. M., 1984, Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms: J. Geophys. Res., **89**, 5953–5986.
Zienkiewicz, O. C., 1977, The finite element method, 3rd ed.: McGrawHill.

APPENDIX
EFFICIENT CALCULATION OF THE JACOBIAN MATRIX

The frequency-domain, finite-element, or finite-difference formulation for scalar and elastic wave equations can be written as

$$\mathbf{K(m)d(m, \omega)} + i\omega\mathbf{C(m)d(m, \omega)}$$
$$- \omega^2\mathbf{M(m)d(m, \omega)} = \mathbf{f}(\omega), \qquad (A\text{-}1)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the square $n \times n$ (n is the number of nodal points) mass, damping, and stiffness matrices, respectively; $\mathbf{f}$ and $\mathbf{d}$ are the $n \times 1$ source and data vectors, respectively; $\mathbf{m}$ is the model vector comprised of the impedance and velocity at each nodal coordinate; ω is the angular frequency; and $i = \sqrt{-1}$. For simplicity, we express equation (A-1) as

$$\mathbf{S(m, \omega)d(m, \omega)} = \mathbf{f}(\omega), \qquad (A\text{-}2)$$

where the complex impedance matrix $\mathbf{S}$ is given by

$$\mathbf{S} = \mathbf{K} + i\omega\mathbf{C} - \omega^2\mathbf{M}. \qquad (A\text{-}3)$$

In principle, we could calculate the inverse of $\mathbf{S}$ to obtain $\mathbf{d} = \mathbf{S}^{-1}\mathbf{f} \equiv \mathbf{Gf}$, where we recognize $\mathbf{G}$ to be the Green's function. In practice, we do not explicitly calculate $\mathbf{S}^{-1}$ but rather decompose $\mathbf{S}$ into the product of a lower triangular matrix $\mathbf{L}$ and an upper triangular matrix $\mathbf{U}$. We then calculate the wavefield $\mathbf{d}$ by forward and backward substitution. Matrices associated with 2-D finite-element and finite-difference meshes are amenable to modern sparse matrix techniques such as the nested dissection method used by Marfurt and Shin (1989) for large-scale seismic modeling. Taking the partial derivative of equation (A-2) with respect to the parameter m_i yields

$$\mathbf{S(m, \omega)}\frac{\partial\mathbf{d(m, \omega)}}{\partial m_i} + \frac{\partial\mathbf{S(m, \omega)}}{\partial m_i}\mathbf{d(m, \omega)} = 0. \qquad (A\text{-}4)$$

After arranging equation (A-4), we obtain

$$\mathbf{S}\frac{\partial\mathbf{d}}{\partial m_i} = \bar{\mathbf{f}}_i, \qquad (A\text{-}5)$$

where the virtual source vector $\bar{\mathbf{f}}_i$ associated with a perturbation of the ith model parameter is given by

$$\bar{\mathbf{f}}_i = -\frac{\partial\mathbf{S}}{\partial m_i}\mathbf{d}. \qquad (A\text{-}6)$$

We recognize $\partial d_j/\partial m_i$ as the (i, j)th element of the data sensitivity, or Jacobian, matrix $\mathbf{J}$. Since $\mathbf{d}$ is a function of the true source $\mathbf{f}$, we obtain a different Jacobian for each of the ns sources. Solving equation (A-5) for $\partial\mathbf{d}/\partial m_i$ and denoting $\bar{\mathbf{F}}$ as a matrix composed of all virtual source vectors $\bar{\mathbf{f}}_i$, we obtain

$$\mathbf{J} = \frac{\partial\mathbf{d}}{\partial\mathbf{m}} = \mathbf{S}^{-1}\bar{\mathbf{F}}. \qquad (A\text{-}7)$$

We further limit our computational effort by noting that our sources and receivers typically lie along or slightly below the earth's surface (Figure A-1). If we number our mesh with the large numbers encompassing the source and receivers, we can reduce our forward and backward substitution effort by $>95\%$. Next, we exploit the principle of reciprocity between any subsurface image position i and any surface source s or receiver r. Instead of calculating the Green's function (impulse response) measured at the N_x nodes along the earth's surface by exciting an impulsive source at each of the $N_x \times N_z$ subsurface image points, we calculate and store the impulse response measured in the subsurface corresponding to all source and receiver locations along the earth's surface. We can now efficiently calculate the Jacobian by multiplying the virtual source given by equation (A-6) with a previously factored Green's function $\mathbf{S}^{-1}$. For rectangular finite elements, virtual sources have local support such that they are defined only at each of the four nodal points of the element (Shin, 1988). Therefore, at the lth nodal point of the surface, the partial derivative wavefield resulting from a perturbation of a parameter in the ith element and a true source at the jth nodal point of the surface is given as

$$\left.\frac{\partial d_l}{\partial m_i}\right|_j = \sum_{k=1}^{4} \bar{f}_{i,j}^k\, d_l^k, \qquad (A\text{-}8)$$

where k is one of the nodal points of the ith element, $\bar{f}_{i,j}^k$ is the virtual source at the kth nodal point of the ith element from the jth true source, and d_l^k is the wavefield generated by the lth true source of the surface and measured at the kth nodal point.
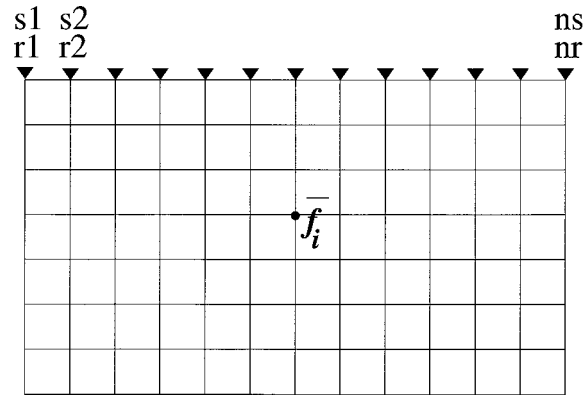

FIG. A-1. An $N_x \times N_z$ finite-element mesh with nr receivers and ns sources distributed along the surface; $\bar{f}_i$ indicates a virtual source from a perturbation of the ith model parameter. The principle of reciprocity states that the wavefield $d_{r,i}$ measured at the rth receiver location from a unit source f_i is identical to the wavefield $d_{i,r}$ measured at the ith subsurface location from a unit source at the rth receiver location along the surface.