GEOPHYSICS, VOL. 60, NO. 5 (SEPTEMBER-OCTOBER 1995), P. 1457–1473, 20 FIGS., 1 TABLE.

Multiscale seismic waveform inversion

Carey Bunks*, Fatimetou M. Saleck‡, S. Zaleski**, and G. Chavent§

ABSTRACT

Iterative inversion methods have been unsuccessful at inverting seismic data obtained from complicated earth models (e.g. the Marmousi model), the primary difficulty being the presence of numerous local minima in the objective function. The presence of local minima at all scales in the seismic inversion problem prevent iterative methods of inversion from attaining a reasonable degree of convergence to the neighborhood of the global minimum. The multigrid method is a technique that improves the performance of iterative inversion by decomposing the problem by scale. At long scales there are fewer local minima and those that remain are further apart from each other. Thus, at long scales iterative methods can get closer to the neighborhood of the global minimum. We apply the multigrid method to a subsampled, low-frequency version of the Marmousi data set. Although issues of source estimation, source bandwidth, and noise are not treated, results show that iterative inversion methods perform much better when employed with a decomposition by scale. Furthermore, the method greatly reduces the computational burden of the inversion that will be of importance for 3-D extensions to the method.

INTRODUCTION

Some of the earliest approaches to seismic inversion avoided the problem of local minima by linearizing the relation between observed seismic data and the velocity model. Linearized inversion is justified when the initial velocity model is in the neighborhood of the global minimum of the objective function. This is a reasonable assumption for seismic inversion problems where the geology of the zone is not too complex or where good a priori knowledge of the background velocity field is available. Numerous papers on linearized waveform inversion have been published (Berkhout, 1984; Devaney, 1984; Esmersoy, 1986; Levy and Esmersoy, 1988; and Tarantola, 1984). This approach produced many interesting results on the properties of linearized inversion (conditions for invertibility, limits of resolution of seismic data) and gave rise to inversion algorithms that have physically intuitive interpretations (for example the back projection and generalized radon transform algorithms). However, these methods did not seem to be able to recover long wavelength components of the velocity model, and consequently gave rise to questions about the observability of the complex wave vector field in the seismic data. In the linearized analysis (Devaney, 1984), it is seen that certain components of the complex wavefield vector are not observable. However, the paper by Mora (1989) shows that under normal operating conditions for seismic exploration (existence of reflectors in the velocity model) that all the wave vector components should be observable when using a nonlinear approach to the seismic waveform inversion problem. Specifically, Mora shows that the observable wave vector components of the velocity field are bounded above by the highest frequency in the source and below by the effective opening of the seismic array. This is important because it means that long wavelength components of the velocity field are observable even when the source is band-limited. The conditions for observability are that the effective receiver array opening is sufficiently large or that the reflector structure is sufficiently complex.

Full nonlinear waveform velocity inversion as described in Mora (1987), Pica et al. (1990), and Tarantola (1986, 1988) seeks to determine the optimum velocity model by globally minimizing the objective function using an iterative descent method. The direct application of this inversion technique to real and synthetic data has been disappointing because numerous local minima in the objective function prevent iterative optimization techniques from working effectively. As illustrated in Figure 1, these local minima are largely

Manuscript received by the Editor January 20, 1994; revised manuscript received August 16, 1994.
*Total, Cedex 47, 92069 Paris La Défense, France.
‡Formerly Total, Cedex 47, 92069 Paris La Défense, France; presently Geomath International, BP 213, 232 Avenue Napoleon Bonaparte, 92502 Rueil Malmaison Cedex, France.
**Laboratoire de Modélisation en Mécanique, URA CNRS no. 229, Université VI Jussieu, Tour 66 Case 162, 75252 Paris Cedex 05, France.
§Inria, Domaine du Voluceau, 78153 La Chesnay Cedex, France.

caused by two effects. First, when an initial velocity model gives rise to large kinematic errors between observed and modeled data (as illustrated in Figure 1a), a perturbation of the velocity model has no effect on the square integral of the residuals, and consequently the gradient is zero without being at the global minimum. Second, the high-frequency components of the wavelet give rise to a multimodal correlation function (illustrated in Figure 1b), which explains many of the local minima in the objective function.

Thus, the main theoretical difficulty for nonlinear seismic inversion is the presence of numerous local minima in the objective function. These local minima impede iterative techniques from finding the global minimum unless the initial model for the velocity field is already in the neighborhood of the global solution. For example, the Marmousi experience (Versteeg and Grau, 1991; Versteeg, 1991) showed that standard as well as modern estimation techniques failed to obtain velocity models sufficiently accurate to correctly image the earth under a complex overburden. This difficulty has given rise to many inversion schemes that try to eliminate or avoid the problem of local minima either by the use of simplifying approximations to the wave equation or by the construction of constrained objective functions.

One alternative approach to full waveform velocity inversion is the well known reflection tomography method (Bishop et al., 1985; Bording et al., 1987). Reflection tomography is based on a reformulation of the seismic inversion problem using the eikonal equation as an approximation to the acoustic wave equation. The modeled data are represented by the traveltimes of first-arrival reflections and these are compared to the traveltimes of reflected waves in the observed data. This formulation of the seismic inversion problem avoids many of the problems of the waveform inversion problem as illustrated in Figure 1. Since the objective function is based directly on kinematic errors of observed and modeled data there are fewer local minima in the objective function. Furthermore, the use of observed and modeled traveltimes reduces the number of local minima introduced by a multimodal correlation function (wavelets are converted to impulses by picking). The reflection tomography method works well for seismic data of moderate complexity but still cannot properly invert complicated data such as that coming from the Marmousi model. This is partially because of the fact that first arrivals are not always the most energetic (Geoltrain and Brac, 1993), partially because of the poorly posed nature of the tomography approach (more unknowns than data), and partially because of the failure of ray theoretic methods to correctly model long wavelength wave phenomena (Biondi, 1992).

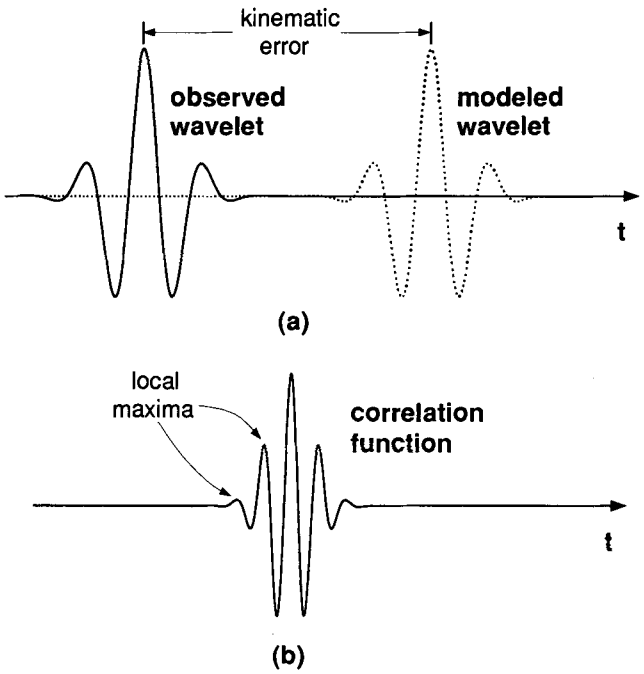

FIG. 1. Source of local minima in objective function for seismic inversion. (a) Large kinematic errors give rise to gradients that are zero, and (b) high frequencies in the wavelet give rise to local maxima in the correlation function.

A number of approaches have been developed recently that attempt to reduce the number of local minima in the objective function by introducing geometrically coherent constraints on the observed seismic data after depth migration. Examples of this approach can be found in Al-Yahya (1987); Bunks (1991, 1992); Chavent and Jacewitz (1990); Symes and Carazzone (1991); and van Trier (1990). The main idea of these approaches is that seismic data must be geometrically coherent after migration. In principle, the migration reduces large kinematic errors and thus reduces the number of local minima in the objective function. These methods are still under examination and no conclusive results are available at this time.

The above discussion implies that the primary problem of seismic inversion continues to be the presence of local minima in the objective function. This has led to a number of papers that have explored the possibility of diminishing the problem of local minima by a decomposition of the seismic inversion problem by scale. Once the problem has been decomposed by scale the long scale components are solved first with the idea being that at long scales the number of local minima is greatly reduced and those that remain are further apart from each other. The application of an iterative descent method at a long scale of the problem is thus more likely (than at a short scale) to find the global minimum or at least a local minimum that is in the neighborhood of the global minimum. The solution found at the long scale of the problem is then recursively refined by using it as an initial solution at increasingly shorter scales.

For the 1-D inversion problem, scale decomposition has been shown to be very effective, and striking results have been obtained by Kolb et al. (1986) and then Hadjee (1988). More recently, several ideas related to scale decomposition for the 2-D problem have been published (Lindgren et al., 1989; Lindgren, 1992), however, the formulation of the problem did not lend itself to conclusive results. Finally, a scale decomposition method has been demonstrated on a modified version of the 2-D Marmousi data set in Saleck et al. (1993). The objective of this paper is to describe in fuller detail with more conclusive examples these recent results for 2-D prestack seismic inversion.

The organization of this paper is as follows. First, a review of the classic gradient calculation for the continuous and discrete formulations of generalized waveform inversion is

given. This is followed by a general description of the multigrid method and our application of this method to seismic inversion. Several examples based on the Marmousi data set are presented. These examples illustrate the effectiveness of the multigrid method for circumventing the problems of local minima in the objective function and show that the scale decomposition of the multigrid method can be used to greatly accelerate the computational aspect of the inversion problem. Finally, we discuss the results, give our perspectives for future work, and present our conclusions.

CLASSIC SEISMIC WAVEFORM INVERSION

The results presented in this paper depend on the accurate calculation of the gradient. The development of the continuous gradient calculation for seismic inversion is available in the literature, but as far as we know its practical implementation as a numerical scheme has never been published. Consequently, this section first develops the continuous gradient calculation for the classic seismic waveform inversion problem and then explicitly specifies the finite-difference schemes used to numerically resolve the problem. This is followed by a description of the full discrete problem formulation and resulting gradient calculation. As will be seen, the solution to the discrete problem formulation has some important details that cannot be directly inferred from the solution to the continuous problem formulation.

CONTINUOUS PROBLEM FORMULATION

Seismic inversion attempts to find the velocity model that minimizes the difference between observed and modeled seismic data. The modeled data are generated by numerically integrating a wave equation such as the constant density acoustic wave equation

$$\frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} = \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} + s\right), \tag{1}$$

where, in general, the initial conditions for the wavefield are

$$\begin{aligned} p(x, z, 0) &= 0 \\ \frac{d}{dt}p(x, z, 0) &= 0. \end{aligned} \tag{2}$$

Here the wavefield p and the source signal s are functions of the space variables x and z and the time variable t, and the velocity v is a function of x and z only.

Implementation of the seismic inversion problem requires the calculation of the gradient of an objective function of the observed and modeled data with respect to the velocity field. In general the observed data are available for some time interval $t \in [0, T]$ and over some spatial domain $(x, z) \in H$. A measure of the difference between observed and modeled data can be obtained using the norm

$$J(v) = \frac{1}{2}\int_{(x,z)\in H} dx\, dz \times \int_0^T dt\, \{p(x, z, t) - \tilde{p}(x, z, t)\}^2, \tag{3}$$

where $\tilde{p}$ and p are the observed and modeled data, respectively. Given a velocity v, the modeled data are calculated from equation (1). The velocity field which minimizes equation (3) is the desired solution to the inversion problem.

Gradient methods are commonly used to solve inversion problems. These methods are iterative and begin by choosing an initial velocity model and then constructing a sequence v_n such that

$$v_{n+1} = v_n + \alpha_n d_n, \tag{4}$$

where α_n is a small positive scalar obtained by a linear search (Luenberger, 1969) and d_n is a descent direction such as the negative of the gradient, $-dJ/dv$. The calculation of the gradient can be obtained by direct differentiation of equation (3) and substitution using equation (1), however, a very compact and elegant derivation can be had using the Lagrange multiplier method of constrained optimization.

The objective of seismic inversion is to find the minimum of the objective function in equation (3) subject to the constraint given by equation (1). Rewriting equation (1) in the form $G(p, v) = 0$ yields the equation

$$G(p, v) \triangleq \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} + s\right) - \frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} = 0. \tag{5}$$

The gradient of the objective function in equation (3) constrained by the wave equation (5) can be obtained by the Lagrange multiplier technique (Alexéev et al., 1982; Hildebrand, 1965; Lanczos, 1962; Luenberger, 1969). The method constructs the Lagrangian S, which is given by the sum of the objective function and the linear span of the constraints

$$S = J + \int_{(x,z)\in H}\int_0^T \lambda G(p, v)\, dt\, dx\, dz. \tag{6}$$

Here the span of the constraints is obtained by calculating the usual scalar product of the wavefield constraint relation with the unknown (Lagrange) multiplier function, $\lambda(x, z, t)$. The solution to the inversion problem is found at the stationary points of the Lagrangian with respect to p and v. The stationary points of S with respect to p generate what is known as the adjoint or dual wave equation. To calculate the stationary points of S with respect to p we begin by integrating equation (6) twice by parts

$$S = J + \int_{(x,z)\in H}\int_0^T p\left[\frac{\partial^2 \lambda}{\partial x^2} + \frac{\partial^2 \lambda}{\partial z^2} + s - \frac{1}{v^2}\frac{\partial^2 \lambda}{\partial t^2}\right] dt\, dx\, dz, \tag{7}$$

and then calculating $\partial S/\partial p = 0$. This yields the adjoint wave equation

$$\frac{1}{v^2}\frac{\partial^2 \lambda}{\partial t^2} = \left(\frac{\partial^2 \lambda}{\partial x^2} + \frac{\partial^2 \lambda}{\partial z^2}\right) + (p - \tilde{p}), \tag{8}$$

with associated final conditions

$$\lambda(x, z, T) = 0$$
$$\frac{d}{dt}\lambda(x, z, T) = 0. \tag{9}$$

It can be seen that equation (8) is dynamically equivalent to equation (1) with final conditions instead of initial conditions. Thus the adjoint equation is a wave equation that evolves backward in time and has the residual function $(p - \tilde{p})$ as a source.

Finally, the gradient is obtained by calculating $\partial S/\partial v$

$$\frac{\partial S}{\partial v} = \frac{2}{v^3}\int_0^T \lambda \frac{\partial^2 p}{\partial t^2}\, dt, \tag{10}$$

where p is obtained by integrating equation (1), and λ is obtained by integrating equation (8). We note that the gradient calculation just described has an interpretation as a migration followed by the application of an imaging condition (first noted in Lailly, 1984).

THE DISCRETIZED PROBLEM FORMULATION

In this paper the numerical integration of equation (1) is obtained by using the classic second-order, finite-difference scheme (Bamberger et al., 1980; Kelly et al., 1976) to approximate the derivatives. Using the notation $p^{\ell}_{n,m} = p(n\Delta x, m\Delta z, \ell\Delta t)$ where $n = 0, 1, \ldots, N$; $m = 0, 1, \ldots, M$; and $\ell = 0, 1, \ldots, L$ the partial derivatives in equation (1) can be approximated by

$$\frac{\partial^2 p}{\partial t^2} \approx \frac{p^{\ell+1}_{n,m} - 2p^{\ell}_{n,m} + p^{\ell-1}_{n,m}}{\Delta t^2}$$
$$\frac{\partial^2 p}{\partial x^2} \approx \frac{p^{\ell}_{n+1,m} - 2p^{\ell}_{n,m} + p^{\ell}_{n-1,m}}{\Delta x^2} \tag{11}$$
$$\frac{\partial^2 p}{\partial z^2} \approx \frac{p^{\ell}_{n,m+1} - 2p^{\ell}_{n,m} + p^{\ell}_{n,m-1}}{\Delta z^2}.$$

Applying these approximations to equation (5) and performing the necessary algebra yields the discretized wave equation (in constraint form)

$$G(p^{\ell}_{n,m}, v_{n,m}) = s^{\ell}_{n,m} + \left[\frac{p^{\ell}_{n+1,m} - 2p^{\ell}_{n,m} + p^{\ell}_{n-1,m}}{\Delta x^2}\right] + \left[\frac{p^{\ell}_{n,m+1} - 2p^{\ell}_{n,m} + p^{\ell}_{n,m-1}}{\Delta z^2}\right] - \frac{1}{v^2_{n,m}}\frac{p^{\ell+1}_{n,m} - 2p^{\ell}_{n,m} + p^{\ell-1}_{n,m}}{\Delta t^2} = 0, \tag{12}$$

where the initial conditions are

$$p^{-1}_{n,m} = 0$$
$$p^{0}_{n,m} = 0. \tag{13}$$

Special conditions are necessary at the extremities of the spatial domain if absorbing boundaries are desired. The conditions used in this paper are those of Reynolds (1978). The condition we use at the surface $z = 0$ is a free boundary condition. Thus, $p^{\ell}_{n,-1} = 0$.

The implementation of the finite-difference scheme for the constant density acoustic wave equation gives rise to problems of numerical instability and numerical dispersion for some choices of the sampling steps Δx, Δz, and Δt (Bamberger et al., 1980). For the second-order scheme described in the previous section, numerical stability is ensured when

$$\min(\Delta x, \Delta z) > \sqrt{2}\Delta t \max(v), \tag{14}$$

and numerical dispersion is limited when

$$\max(\Delta x, \Delta z) < \frac{\min(v)}{10 f_{\max}}, \tag{15}$$

where $f_{\max}$ is the maximum frequency contained in the source s.

Analogous to the objective function for the continuous formulation of the seismic inversion problem, the objective function for the discrete problem formulation is

$$J(v) = \frac{1}{2}\sum_{(n,m)\in H}\sum_{\ell=1}^{L}[p^{\ell}_{n,m} - \tilde{p}^{\ell}_{n,m}]^2, \tag{16}$$

where the sum for $(n, m) \in H$ is over a discrete set of points.

The discrete Lagrangian is obtained by constructing the sum of the objective function in equation (16) and the linear span of the discrete constraints given in equation (12)

$$S = J + \sum_{\ell=0}^{L}\sum_{n=0}^{N}\sum_{m=0}^{M}\lambda^{\ell}_{n,m}G(p^{\ell}_{n,m}, v_{n,m}). \tag{17}$$

Here the unknown (Lagrange) multiplier constants are represented by $\lambda^{\ell}_{n,m}$. As for the continuous formulation, the solution to the discretized inversion problem is obtained at the stationary points of the Lagrangian with respect to $p^{\ell}_{n,m}$ and $v_{n,m}$.

The stationary points of equation (17) with respect to $p^{\ell}_{n,m}$ yield the discrete adjoint equation. Thus, calculating $\partial S/\partial p^{\ell}_{n,m} = 0$ yields

$$\lambda^{\ell-1}_{n,m} = 2\lambda^{\ell}_{n,m} - \lambda^{\ell+1}_{n,m} + v^2_{n,m}\Delta t^2\left[\frac{\lambda^{\ell}_{n+1,m} - 2\lambda^{\ell}_{n,m} + \lambda^{\ell}_{n-1,m}}{\Delta x^2}\right] + v^2_{n,m}\Delta t^2\left[\frac{\lambda^{\ell}_{n,m+1} - 2\lambda^{\ell}_{n,m} + \lambda^{\ell}_{n,m-1}}{\Delta z^2}\right] + (p^{\ell}_{n,m} - \tilde{p}^{\ell}_{n,m}), \tag{18}$$

where $(p^{\ell}_{n,m} - \tilde{p}^{\ell}_{n,m})$ is the discrete residual function and where the final conditions are

$$\lambda^{L}_{n,m} = 0$$
$$\lambda^{L-1}_{n,m} = p^{L}_{n,m} - \tilde{p}^{L}_{n,m}. \tag{19}$$

Note that the final conditions in equation (19) are not symmetric with the initial conditions in equation (13) and do not follow directly from the final conditions determined for the continuous problem formulation described in equation (9). This shows that the correct numerical calculation of the gradient depends on the reformulation of the inversion problem using the exact finite-difference operator used to discretize the problem.

The discrete gradient is obtained by calculating $\partial S/\partial v_{n,m}$ which yields

$$\frac{\partial S}{\partial v_{n,m}} = \frac{2}{v_{n,m}^3} \sum_{\ell=0}^{L} \lambda_{n,m}^{\ell} \left[\frac{p_{n,m}^{\ell+1} - 2p_{n,m}^{\ell} + p_{n,m}^{\ell-1}}{\Delta t^2} \right], \quad (20)$$

where $p_{n,m}^{\ell}$ is obtained by integrating equation (12), and $\lambda_{n,m}^{\ell}$ is obtained by integrating equation (18).

Since the objective functions in equation (3) and (16) are known to have numerous local minima, the success of iterative descent methods depends on the initial guess, $v_0(n,m)$. A poor initial guess leads to a solution that corresponds to the closest local minimum of the objective function. Another important point is that the forward modeling required for the evaluation of equation (3) is obtained by the numerical integration of equation (1). To avoid instability and numerical dispersion, the sampling in time and space of this equation must be inferior to limits imposed by equations (14) and (15). These limits lead to intense computational burdens and memory requirements. In this paper we propose an optimization procedure based on gradient calculations applied to the seismic inversion problem after a decomposition by scale. This technique of decomposing an inversion problem by scale is called the multigrid method (the details of which are discussed in the next section). The multigrid method is more successful than standard gradient methods at finding a global minimum of the seismic inversion objective function, and the method has the potential to diminish the computational burden associated with the modeling part of the inversion problem.

THE MULTIGRID METHOD

The multigrid method (Brandt, 1977; Briggs, 1987; Press and Teukolsky, 1991) is a family of techniques that are useful for solving large-scale linear and nonlinear problems. The method is based on decomposing a problem by scale, followed by the resolution of each scale component by a suitable relaxation operator. This approach has two advantages. First, for each scale of the problem Nyquist's sampling theorem can be applied to appropriately resample the problem onto a smaller grid. This accelerates the effectiveness and reduces the computational cost of the relaxation operator at that scale. Second, for the nonlinear aspects of the problem the solution of the long scale component is followed by a fining-up procedure at successively shorter scales. This approach is more effective at finding the global minimum of the problem as is heuristically illustrated in Figure 2.

Figure 2 shows a multimodal objective function and a decomposition of this function onto four scales. The longest scale component of the problem, shown in Figure 2e, contains slowly varying features and a single minimum. Consequently, iterative relaxation methods easily find the global minimum of this component of the problem. The global minimum of the long scale component is in the neighborhood of the global minimum for the problem at the second longest scale, shown in Figure 2d. Thus, an application of a relaxation method initialized with the solution obtained from the long scale problem has a good chance of finding the global minimum of the second longest scale problem. Successively fining upward in this way succeeds in finding the global minimum of the original fine scale problem illustrated in Figure 2a.

Note that most of the work of the method is required at the longest scale of the problem illustrated in Figure 2e. The relaxation method used at this scale has to move from the initial guess to the global minimum, two points that may be quite distant from each other. Afterwards, however, at each successive scale only a few iterations of our relaxation method should be necessary to refine the solution at that scale since the solution obtained from the previous scale is in the neighborhood of the new global minimum.

It is important to note that Figure 2 illustrates the most optimistic application of the multigrid method and the method may not always work as smoothly as suggested in the figure. In particular, there may not be sufficiently low-frequency content in the problem to obtain a long scale

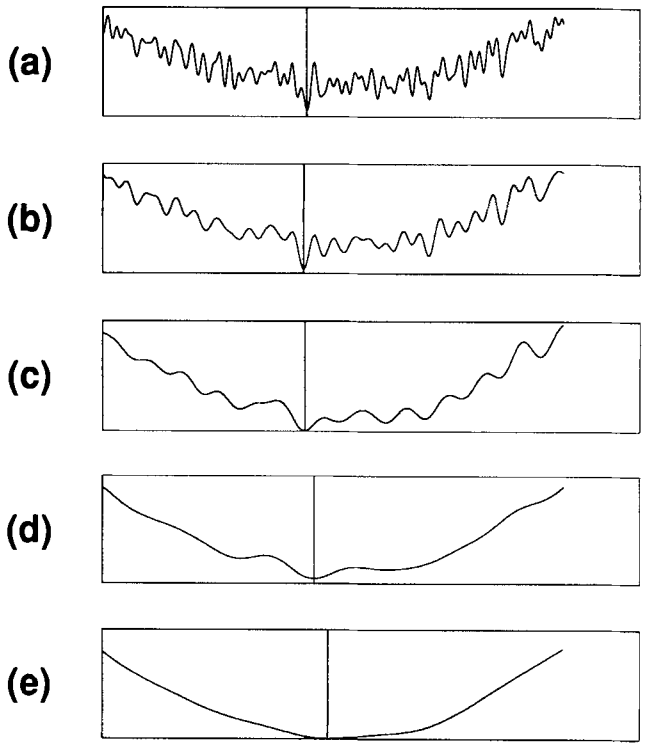

FIG. 2. Heuristic illustration of the multigrid method applied to a nonlinear problem. Frame (a) illustrates a 1-D objective function and frames (b)–(e) illustrate the same objective function at ever increasing scale lengths. A gradient method applied to the shortest scale succeeds in finding the local minimum closest to the starting point, whereas the gradient method applied to the longest scale finds the global minimum (at that scale) regardless of the starting point. A fining up procedure helps find the estimate of the position of the global minimum for the fine scale problem.

decomposition that is convex. In this case, Figure 2d is a more representative starting point for the long scale component of the problem than is Figure 2e. Under these circumstances local minima may impede finding the global minimum for every initial starting point of an iterative descent method. However, a solution obtained at this scale is certainly closer to the global minimum than one found at a shorter scale since there are fewer local minima and those that remain are further apart from each other.

Another important point is that the transitions from scale length to scale length illustrated in Figure 2 have the property that the global solution at one scale is always in the convex neighborhood of the global minimum at the next shortest scale. There is no theoretical guarantee that this property exists for all applications of the multigrid method and substantial effort may be necessary to find scale transitions that reasonably maintain this property. In the rest of this section, a more rigorous description of the multigrid method is given and some of the implementation issues for seismic inversion are discussed.

ELEMENTS OF THE MULTIGRID METHOD

The heuristic description of the multigrid method illustrated in Figure 2 defines a multiscale algorithm for solving nonlinear problems. The implementation of this algorithm requires three elements. The first element is an operator that restricts the original problem to longer and longer scales. The second element is an operator that performs relaxation on each scale. The third component is an operator that injects the solution available at a long scale into the problem setting at a shorter scale. To simplify the presentation, these three elements are defined with respect to a canonical two-level problem that is then generalized to N levels.

The two-level algorithm is illustrated in Figure 3. The problem to be resolved is presumed to be well defined on a space that has a shortest scale length of h. A natural way to define h is as being proportional to the inverse of the highest frequency of the problem; however, other definitions of scale length are possible. A decomposition by scale means a redefinition of the original problem at longer scales. Sequences of scale lengths for multigrid problems commonly evolve by octaves so that a decomposition onto N scales yields the scale length sequence $h, 2h, 4h, \ldots, 2^N h$. (We point out that the decomposition of scale by octaves is common in theoretical discussions of multigrid method but that practical implementations may require finer decompositions to be successful).

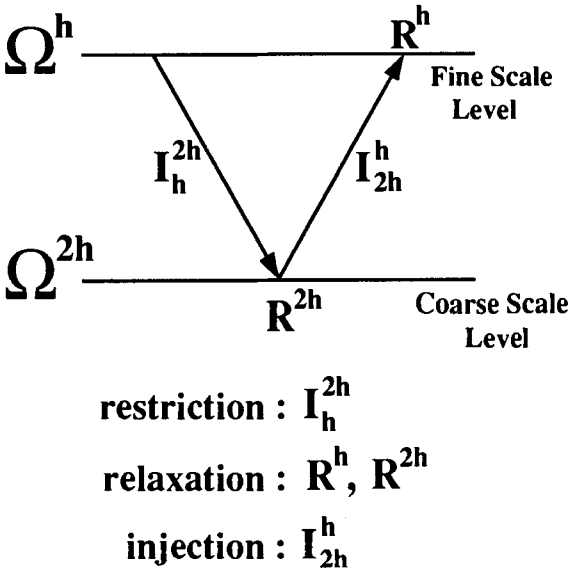

FIG. 3. Description of the two-level paradigm. Ω^h and Ω^{2h} indicate the formulation of the optimization problem at scale lengths h and $2h$, respectively. The restriction, relaxation, and injection operations are represented by I_h^{2h}, R, and I_{2h}^h, respectively.

Let Ω^h represent the space at scale h on which the original problem is defined and Ω^{2h} the space on which we wish to define the scale decomposition of the problem at length $2h$. The operator I_h^{2h} which maps the original problem from Ω^h to Ω^{2h} is called the restriction operator, and its action is represented by

$$I_h^{2h}:\Omega^h \rightarrow \Omega^{2h}. \tag{21}$$

Schematically, the action of I_h^{2h} is as illustrated in Figure 3. Another common term for this operator is the decimation operator and the most common implementations of this operator are based on leaky low-pass filtering followed by subsampling.

Once the problem has been restricted to the long-scale length $2h$, the solution at this scale length is obtained by relaxation. The relaxation operator at scale length $2h$ is represented by the notation R^{2h}. Schematically, the action of R^{2h} is illustrated in Figure 3. Commonly used relaxation methods are the Jacobi and the Gauss-Seidel operators (Briggs, 1987).

Once the relaxation operator has been used to solve the component of the problem at scale length $2h$, solution is then injected (i.e., interpolated) back up to the scale level h with the injection operator I_{2h}^h. The action of this operator is defined by

$$I_{2h}^h:\Omega^{2h} \rightarrow \Omega^h. \tag{22}$$

The injected solution from scale length $2h$ is used to initialize the search for the short scale components of the problem at scale length h.

As indicated in Figure 3, the parts of the problem that were not resolved at scale level $2h$ are solved by the relaxation operator R^h. This completes the description of the two-scale level multigrid solution for nonlinear problems. The generalization to N levels is straightforward and is illustrated in Figure 4. This approach to solving nonlinear problems is known as nested iteration or as the down-to-up scheme.

APPLICATION TO SEISMIC INVERSION

The implementation of the multigrid method requires appropriate choices for the restriction, relaxation, and injection operators. These choices are not unique and not all choices will yield successful multigrid solutions. Each problem has its own physical constraints and the multigrid method must be employed in a way that conforms to these constraints. In this section, we discuss the choices made for

the application of the multigrid method to the seismic inversion problem.

To begin, the notion of scale needs to be defined. Since the seismic inversion problem is multidimensional in time and space, it is necessary to define a scale decomposition that is coherent with the physics of the problem. In fact, it is the highest frequency of the source wavelet, f_{max}, and the smallest value of the velocity model, v_{min}, which completely define the shortest scale length of the seismic inversion problem since the shortest spatial wavelength λ_{min} is equal to the ratio v_{min}/f_{max}.

The seismic inversion problem has several components that may be decomposed by scale. The observed seismic data have both temporal and spatial features, the source has temporal features, and the velocity model has spatial features. Thus, the decomposition by scale can be applied to the source, to the data, to the velocity model, or to any combination of the three. In this paper, two types of decompositions are presented. The first is a temporal frequency decomposition of the source and observed seismic data, and the second is a temporal frequency decomposition plus a spatial decomposition of the observed seismic data and the velocity model.

The restriction operator chosen for the seismic inversion problem is implemented as a leaky (large transition band) low-pass filter. The restriction may be applied only to the source wavelet and the observed seismic data since the velocity model does not need to be restricted (the initial estimate is obtained at the longest scale length). However, the restriction of the velocity model is interesting since the reduction of the size of the finite-difference grid gives rise to important reductions in the CPU time necessary to calculate relaxations. Restriction to longer scales diminishes the frequency of the temporal data as well as that of spatial data. Thus, each restriction that reduces the spectral range by an octave potentially reduces the overall sampling scheme by eight (twice for each dimension x, z, and t). This has very important consequences for seismic inversion which requires intensive computing to perform the forward and backward modeling steps necessary to calculate the gradient. Reductions in CPU time are practically an order of magnitude for each octave of scale reduction, and given that the main effort on the inversion problem should be expended at the longest scale length, means that important time savings should be realizable.

The injection operator chosen for the seismic inversion problem is implemented as the adjoint of a nine-point, nearest-neighbor smoothing filter (illustrated in Figure 5). The injection is applied only to the velocity model. The source wavelet and observed seismic data are already available at each level from the restriction phase of the algorithm.

The relaxation is performed by a quasi-Newton algorithm (Luenberger, 1989) driven by the gradient calculation described in the section on classic seismic waveform inversion. This is an important point since the standard Jacobi and Gauss-Seidel methods are not stable for hyperbolic equations that have all their poles on the imaginary-axis in the Laplace transform domain. The quasi-Newton method retains previously calculated values of the gradient to estimate the inverse of the Hessian (Luenberger, 1989). The convergence of the quasi-Newton method is greatly superior to that of the gradient or conjugate gradient methods.

RELATION TO OTHER MULTISCALE APPROACHES

Simulated annealing (Geman and Geman, 1984; Kirkpatrick et al., 1983; Marroquin, 1985; Metropolis et al., 1953) is a Monte Carlo optimization technique similar to the multigrid method. Simulated annealing, like multigriding, performs optimization at different scales; however, simulated annealing controls the scale of the search operator (as would a stochastic gradient method) and multigrid controls the scale of the objective function. This has important consequences for the speed of resolution and for the computational feasibility of the seismic inversion problem.

When the multigrid method is applied to a long scale component of the inversion problem, the smoothness of the objective function at this scale allows the relaxation operator to rapidly find the neighborhood of the global minimum (see

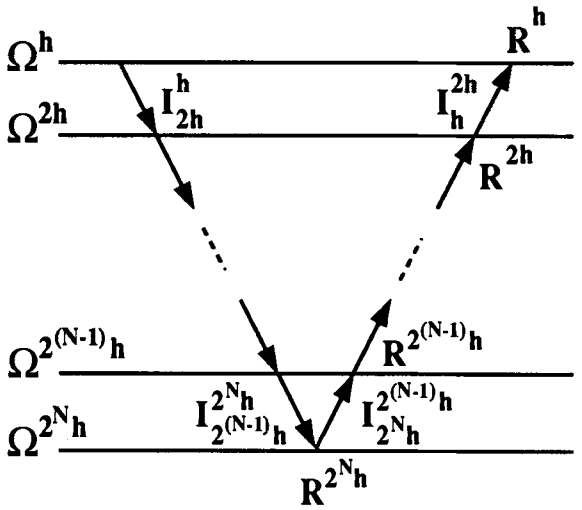

FIG. 4. Down-to-up scheme.

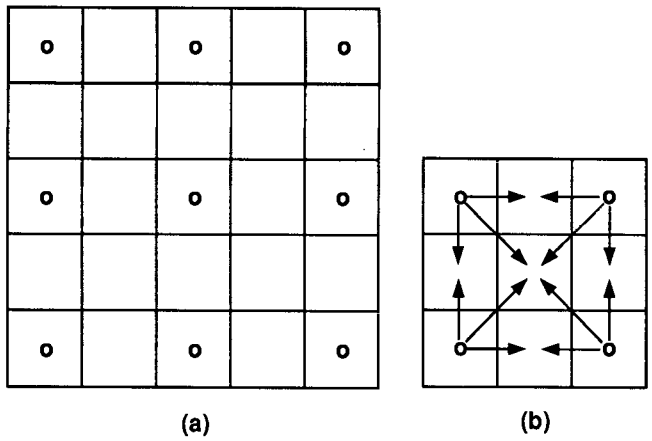

FIG. 5. Restriction and injection schemes. (a) Velocity model from scale Ω^{2h} (pixels containing circles) imbedded in the velocity grid at scale Ω^h. (b) Arrows indicate the elements from the scale Ω^{2h} used to perform the averaging to obtain the elements at scale Ω^h.

Figure 2e). When simulated annealing is employed at a high temperature the search operator allows a sampling of the objective function at long scales, however, the objective function continues to be full of local minima (such as in Figure 2a), and the local information at the long scale is not useful in orienting the search operator towards the neighborhood of the global minimum. Furthermore, the fact that the multigrid technique changes the scale of the objective function means that coarser sampling schemes can be used for the forward and backward modeling steps needed to compute the gradient. This fact greatly accelerates the relaxation of the inversion problem, especially at long scales where it is the most needed. Consequently, the multigrid method is much better adapted to the seismic inversion problem than is the simulated annealing approach.

Another domain of scientific research that decomposes problems by scale is that of wavelets (Daubechies, 1988; Mallat, 1989). The method of scale decomposition used in this paper is based on a frequency decomposition; however, wavelet decompositions might be more efficient and further reduce the computational burden of the seismic inversion problem. The exploration of this question could be an interesting direction for future work.

EXAMPLES

The objective of this section is to show that full nonlinear waveform inversion can be successful for complicated data when a rational multigrid strategy is employed. The examples in this section come from a part of the Marmousi model, the details of which are described in the following. It is important to note that the data used in the examples are not the original Marmousi data but data regenerated from the original velocity model using the finite-difference operator described in the section on classic waveform inversion and a known Ricker wavelet. Our motivation for this is to demonstrate that the multigrid method can effectively eliminate the problem of local minima in the inversion problem. With the remodeled data we avoid problems related to modeling accuracy and source wavelet determination, thus placing the emphasis squarely on the optimization part of the problem.

Another very important point is that we subsampled the original Marmousi velocity model and changed its physical dimensions before performing the remodeling of the seismic data. This was necessary to ensure that the inversion problem was in the scope of our computational resources. An important consequence of the change of the model's physical dimensions is that the resulting remodeled seismic data have a much lower frequency content than do the original band-limited Marmousi data.

The parts of the Marmousi model used for the examples in this section are illustrated in Figure 6. The figure illustrates the entire Marmousi velocity model and indicates the regions that were used for the two examples. The first example corresponds to the larger region delineated by a solid black line in Figure 6. The example based on this portion of the Marmousi model illustrates that the multiscale approach is effective at resolving the problem of local minima in the inversion problem. The second example, which corresponds to the smaller region delineated by a solid black line in Figure 6, illustrates that the calculations of the multiscale inversion can be greatly accelerated when subsampling is performed with scale decomposition.

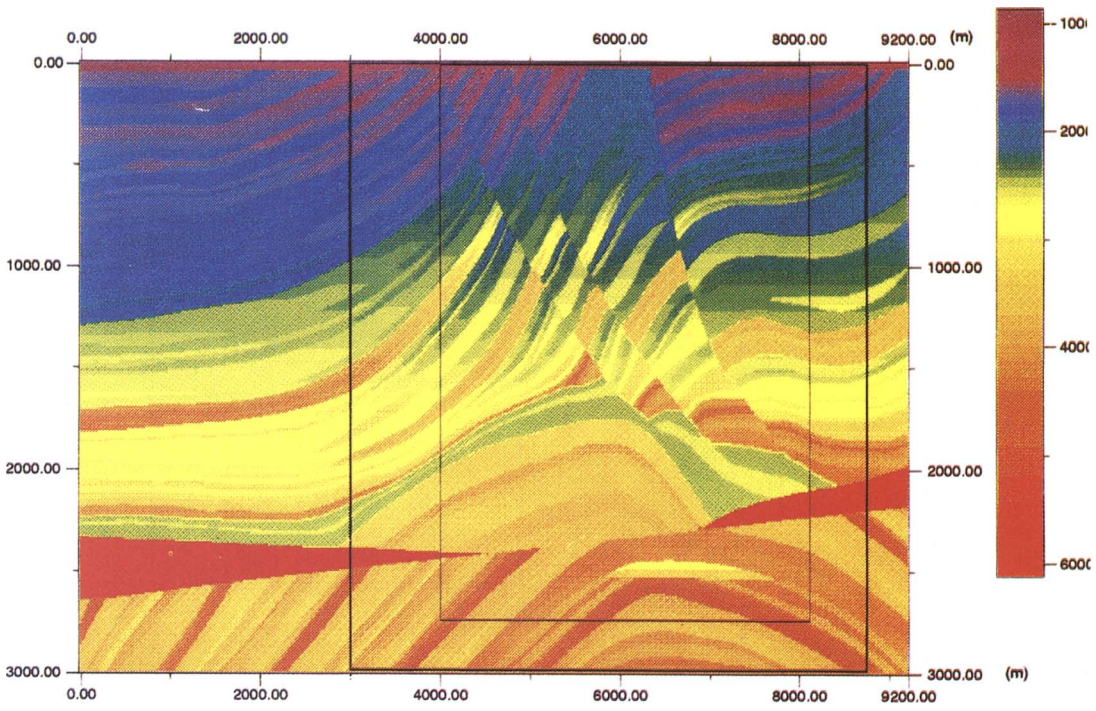

FIG. 6. Full Marmousi velocity model illustrating regions to be used for the two examples in this paper. The velocity color bar has units of m/s.

Example 1

The first example, which corresponds to the larger region delineated by a solid black line in Figure 6, is illustrated schematically in Figure 7. As can be seen, the part of the Marmousi velocity model used for the example corresponds to approximately 60% of the central portion of the model. Two important modifications were made to the extracted part of the original velocity model.

The first modification is a rescaling of the sampling step sizes Δx and Δz. The original Marmousi model has a grid of dimensions 751 × 2301 with a sampling step size of 4 m for Δx and Δz. As is illustrated in Figure 7, the original model is subsampled every 1 in 8 points, giving sampling step sizes of 32 m for Δx and Δz. For this part of the velocity model, this creates a grid of 94 × 175 points which is the biggest model that our computational resources (because of limitations in memory) could handle given the number of times steps necessary to generate the data. Unfortunately, given the source spectrum and the minimum of the velocity model (1500 m/s), the restrictions on sampling imposed by equation (15) require a sampling step size of 6 m for Δx and Δz to avoid numerical dispersion. Consequently, to maintain the complexity of the geology of the Marmousi velocity model without increasing the number of samples, the sampling step sizes were changed from 32 m to 6 m. This change effectively compresses the physical dimensions of the original extracted model from 3000 m in depth and 5600 m in breadth to 558 m in depth and 1044 m in breadth. This scale change is illustrated in Figure 7.

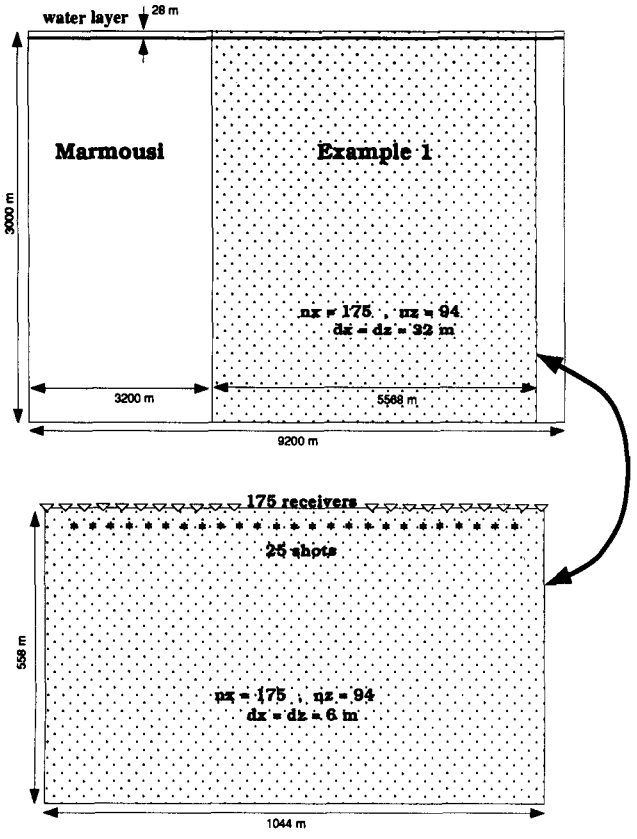

FIG. 7. Geometry of modified Marmousi velocity model and acquisition geometry for first example.

The second modification is that the acquisition geometry was limited to 25 shot points (once again because of problems of memory capacity) equispaced along the surface with 175 receivers per shot point (one receiver for each grid point in x). These implementation details are recapitulated in Table 1 and Figure 7.

The regenerated Marmousi data obtained from the extracted and modified velocity model presented in Figure 7 and Table 1 were used to perform the seismic inversion for the first example. The multigrid strategy is a down-to-up scheme where the scale decomposition was performed uniquely on the time component of the observed seismic data and the source function. The restriction was performed by leaky low-pass filtering where the different scales were obtained by filtering the data and source with filters having cutoff frequencies at 7 Hz, 10 Hz, and 15 Hz, respectively. Figure 8 illustrates the effect of the filtering operation on the source wavelet. Figure 8a shows the original, unfiltered source wavelet and Figures 8b–d illustrate the source wavelet filtered from 0–15 Hz, 0–10 Hz, and 0–7 Hz, respectively. Each filter is a FIR Hamming windowed low-pass filter (Oppenheim and Schafer, 1975; Rabiner and Gold, 1975) and does not distort the phase of the filtered signal since the filter phase is linear. To see the effect that filtering has on the observed seismic data, the first unfiltered shot point of the regenerated Marmousi data as well as these data filtered with the three low-pass filters are illustrated in Figures 9–12.

The longest scale component of the inversion problem is solved for a velocity model restricted to the class of those models that are laterally homogeneous and linearly increasing with depth. We have further constrained the velocity model to have a value of 1500 m/s at the surface, and the only unknown for the optimal longest scale velocity model is the value at 3000 m (i.e., 558 m in the modified velocity model). The optimal value at 3000 m was found by calculating the objective function for the 0–7 Hz data over the range of values $(1500 + 100k)$ m/s for $k = 0, 1, \ldots, 40$. This

Table 1. Parameters for examples 1 and 2.

	Example 1	Example 2
Water bottom	6 m	78 m
Depth of source	6 m	96 m
Distance between receivers	6 m	6 m
Number of receivers	175	129
Central frequency of source	30 Hz	30 Hz
Distance between shot points	30 m	30 m
Number of shots	25	16
Width of velocity model	1044 m	768 m
Depth of velocity model	558 m	768 m
Recording time	1.05 s	0.9 s
Finite-difference parameters	$nx = 175$, $nz = 94$	$nx = nz = 129$
	$dx = dz = 6$ m	$dx = dz = 6$ m
	$dt = 0.75$ ms	$dt = 0.75$ ms

constitutes an exhaustive search for the long scale component of the velocity model and is performed in exactly 41 forward modeling steps. The best linearly increasing with depth velocity model found by this search has a value of 1500 m/s at the surface and 4300 m/s at a depth of 3000 m.

The solution to the exhaustive search for the data filtered to 0–7 Hz is used to initiate the inversion problem with the data filtered to 0–10 Hz. Here the velocity model is free to vary in x and z and the standard gradient is calculated for these data. The gradient calculations are used by a quasi-Newton optimization program (Luenberger, 1989). For this example, 50 iterations of the quasi-Newton program were calculated. The resulting velocity model at this scale was then used to initialize the inversion problem with the data filtered to 0–15 Hz for which 50 iterations of quasi-Newton optimization were calculated. The resulting velocity model was then used to initialize the inversion problem with the unfiltered data where 25 iterations of quasi-Newton were calculated. The CPU time per iteration was 46 minutes on our Convex C210 computer (40 Mflop/s sustained), and 270 minutes were needed to perform the exhaustive search for the vertical gradient. The total CPU time used for the 125 multiscale iterations plus the exhaustive vertical gradient search was 6020 minutes or 100.33 hours.

The resulting velocity models for the 0–10 Hz, 0–15 Hz, and unfiltered data are illustrated in Figures 13–15, respectively. As can be seen, the inversion of the 0–10 Hz data yields a good estimate of the velocity model, and the sequence of inversions for 0–15 Hz and in addition the unfiltered data adds high-frequency details of the model. The overall inversion successfully recovers many of the features of the original velocity model (compare with the real Marmousi model illustrated in Figure 16). In particular much fine detail in the upper faulted region is imaged as well as the boundaries and interior of the tortoise shell feature in the center of the model. The salt on the right side of the tortoise shell is well defined but does not have the correct velocity value. Finally, a good portion of the anticline under the unconformity can be easily inferred from the image of the inverted model in Figure 15.

Using the inverted model in Figure 15, we performed a prestack finite-difference migration of the original Marmousi data (as opposed to our regenerated data). The result is shown in Figure 17. As can be seen, almost all of the Marmousi model is correctly imaged. The faulted region is well defined with many fine bed details visible; the boundary of the tortoise shell feature as well as internal bedding is clear; the unconformity is well defined across the entire image; and the anticline and the petroleum objective (the flat spot found in the middle of the feature) are clearly imaged.

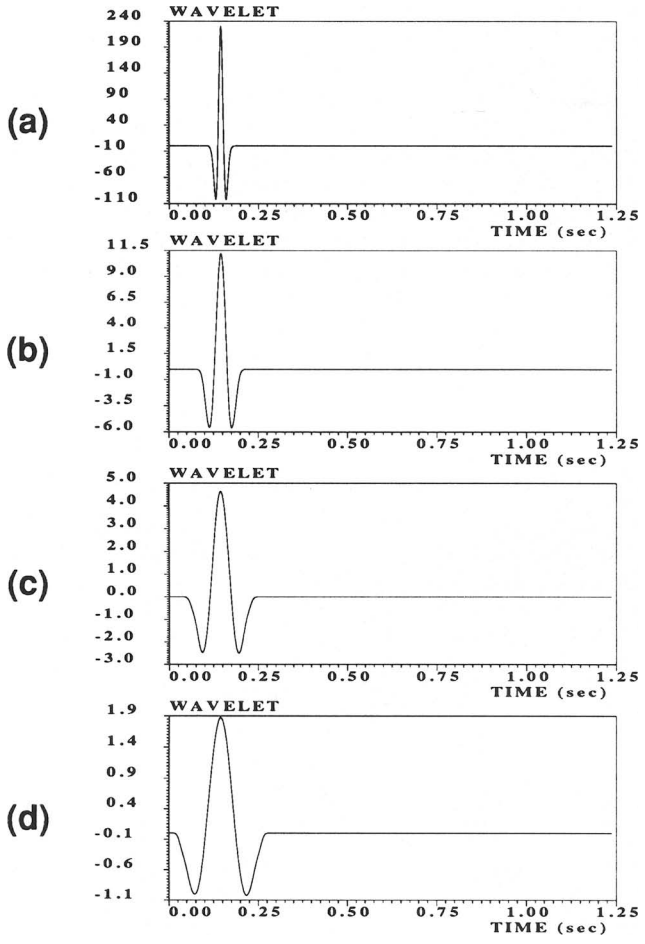

FIG. 8. Filtered source wavelets. (a) The original source wavelet, (b) source wavelet filtered 0–15 Hz, (c) source wavelet filtered 0–10 Hz, (d) source wavelet filtered 0–7 Hz.

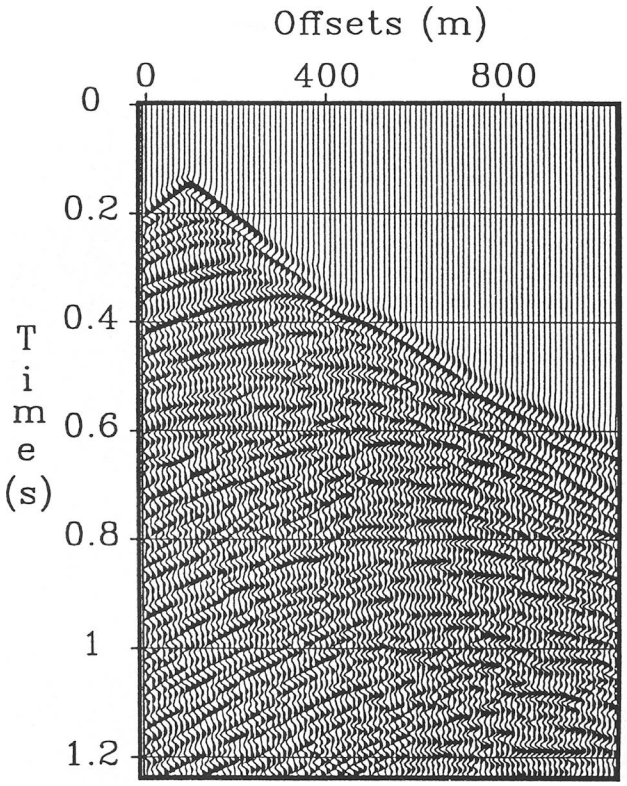

FIG. 9. Unfiltered first shotpoint.

Example 2

The second example, which corresponds to the smaller region delineated by a solid black line in Figure 6, is illustrated schematically in Figure 18. The extracted part of the velocity model is slightly smaller than for the first example. The objective of the second example is to illustrate the effect of scale decomposition when used in conjunction with subsampling of the temporal and spatial data. The idea is simple: as the scale of the inversion problem is increased by low-pass filtering, subsampling of the finite-difference grid is applied [within the limits of stability and numerical dispersion specified in equations (14) and (15)]. As will be shown, the subsampling greatly accelerates the computation of the inversion; however, it also gives rise to some implementation problems.

For the second example the inversion is implemented using the two modifications to the extracted part of the velocity model described in the first example as well as with a third modification. This modification is the addition of a supplementary water layer (see Figure 18). This additional layer avoids a problem with the source geometry when the subsampling is applied to the finite-difference grid of the velocity model. The explanation is as follows. As the scale of the second example is lengthened, the number of points in the finite-difference model for the velocity diminishes because of subsampling. This means that the positions of the sources get closer to the surface (in number of grid points) and eventually the sources are located on the surface boundary. Since the finite-difference modeling scheme employs a free boundary condition at the surface, a source located on

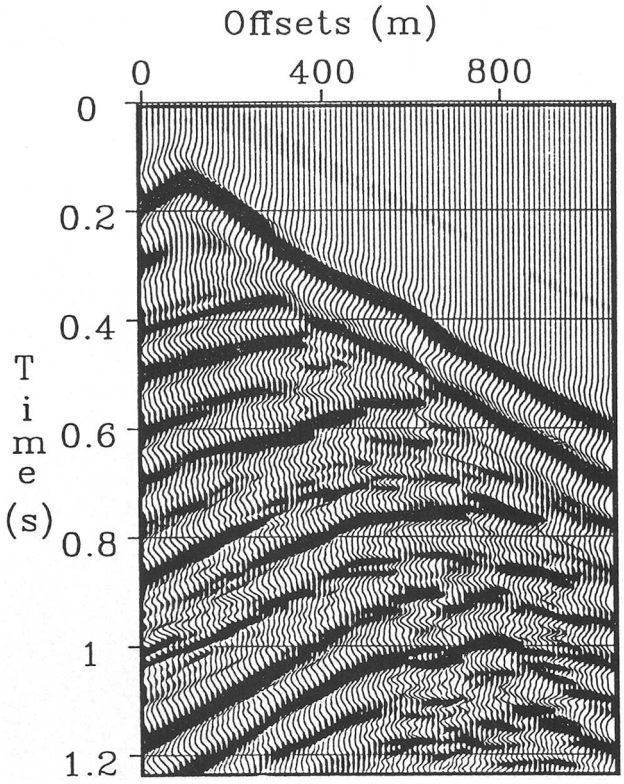

FIG. 11. First shotpoint filtered to 0–10 Hz.

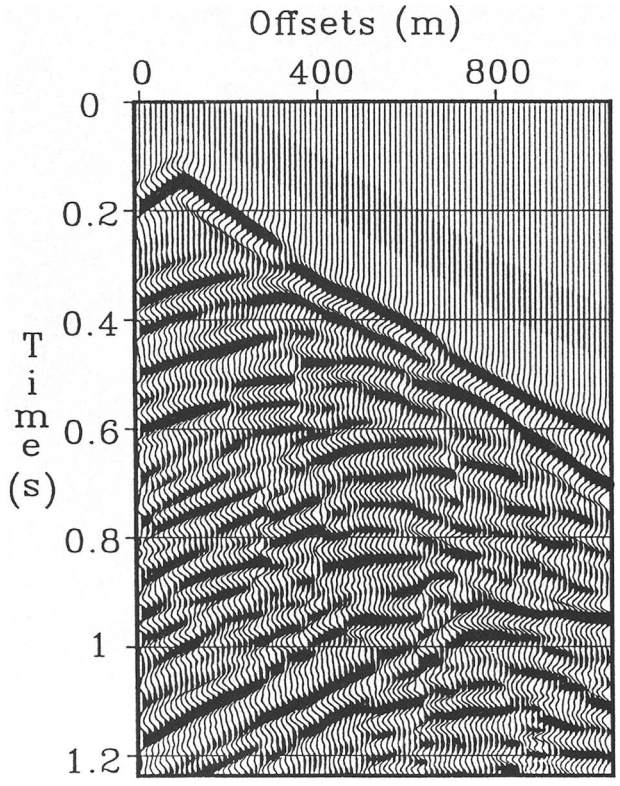

FIG. 10. First shotpoint filtered to 0–15 Hz.

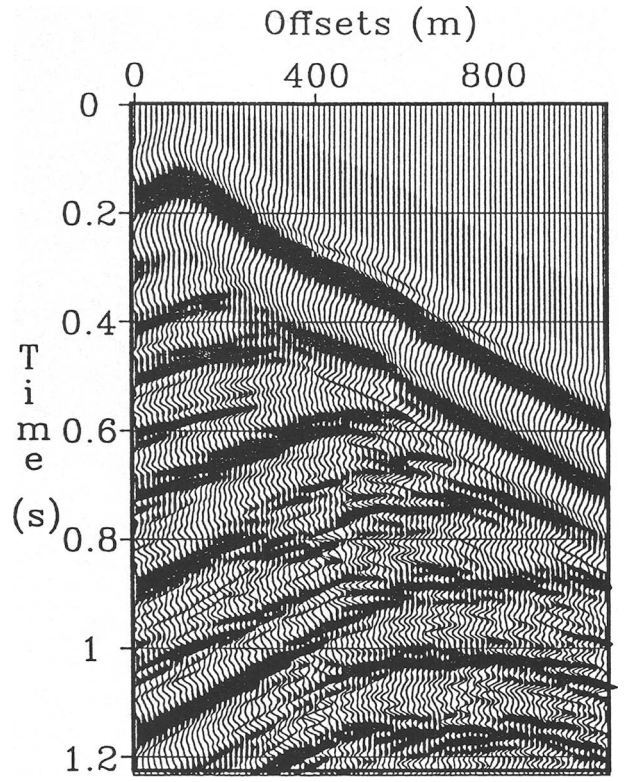

FIG. 12. First shotpoint filtered to 0–7 Hz.

the surface automatically cancels itself (by superposition of the source and its negative reflection from the surface). Our ad-hoc solution to this problem was to add the supplementary water layer so that the source positions would never be located at the surface of the earth. For practical implementations this is obviously not acceptable and another solution needs to be found. However, this example does show that subsampling can greatly accelerate the computation of the inversion.

The details of the modified model are recapitulated by Figure 18 and Table 1. Using a multigrid strategy similar to that used in the first example the inversion was performed at scales 0–7 Hz, 0–15 Hz, and finally with the full scale data (approximately 0–30 Hz). An exhaustive search for a vertical gradient model was performed on the 0–7 Hz data followed by fifty iterations of quasi-Newton at each scale with an unconstrained velocity model. As indicated in Figure 18, there are 129 × 129 points in the velocity grid at the shortest scale length with $dx = dz = 6$ m. At the scale length obtained by filtering the data to 0–15 Hz, the finite-difference grid of the velocity model is subsampled to a size of 65 × 65 points with $dx = dz = 12$ m. Similarly, for the scale length obtained by filtering the data to 0–7 Hz, the finite-difference grid of the velocity model is subsampled to a size of 33 × 33 points with $dx = dz = 24$ m.

The resulting inverted velocity model is illustrated in Figure 19 (to be compared with the extracted part of the real Marmousi model shown in Figure 20). The inverted model is not quite as good as that for the first example because of the coarse spatial restriction and injection (see discussion in the section on the multigrid method and Figure 5) of the data and the velocity model, respectively. However, the example

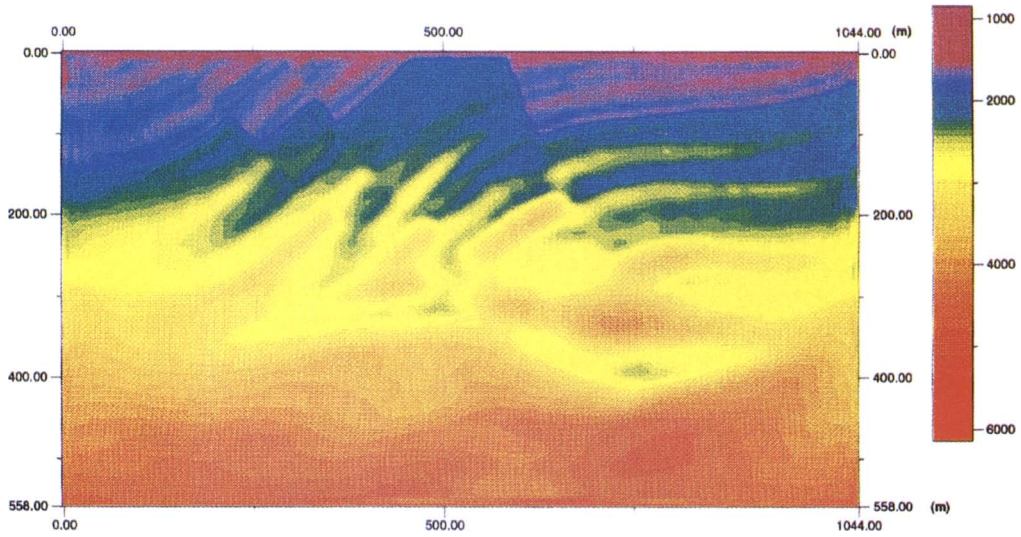

FIG. 13. Best estimate of velocity model after use of the 0–10 Hz data. The velocity color bar has units of m/s.

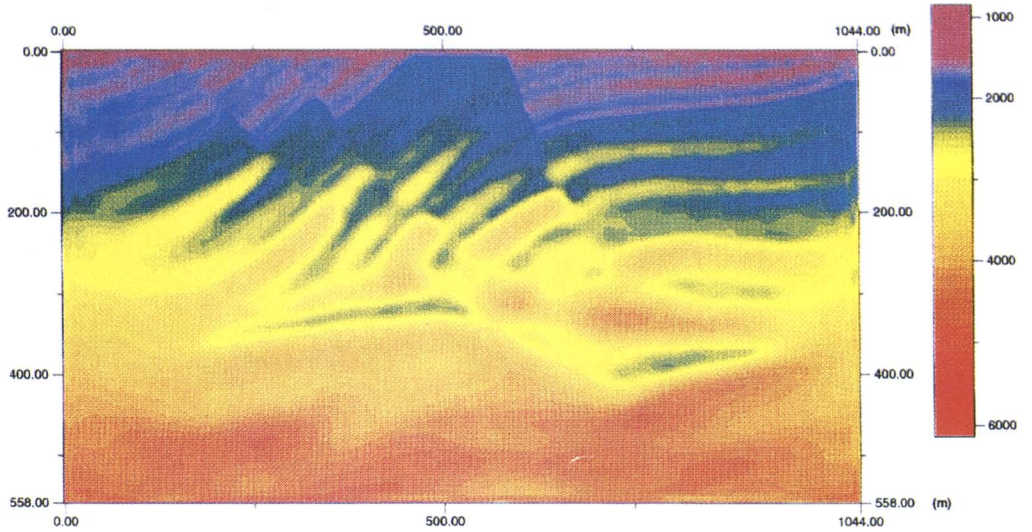

FIG. 14. Best estimate of velocity model after use of the 0–15 Hz data. The velocity color bar has units of m/s.

illustrates the savings in CPU time that can be obtained by subsampling. The most expensive iterations in CPU time were spent on the full scale data that required 22 minutes per iteration (note that the second example is smaller than the first example). The problem scaled to 0–15 Hz required 6 minutes per iteration, and the problem scaled to 0–10 Hz required 1.5 minutes per iteration. Finally, the exhaustive search for the best vertical gradient for the problem scaled to 0–7 Hz required 35 minutes of CPU time (for a total of 25.2 hours of CPU for the inversion). The savings in CPU obtained by changing scale and subsampling should be by a factor of eight, however, in this example the factor is closer to four. Nevertheless, an impressive acceleration of the computation per iteration occurred at the longest scale where the most important iterations of the inversion are calculated.

DISCUSSION AND CONCLUSIONS

The objective of this paper is to show that the multiscale approach to the seismic inversion problem is effective in diminishing the problem of local minima for full waveform seismic inversion. To this end the examples illustrated in this paper were constructed in such a way as to eliminate the dependence on the elements extraneous to this question. The use of the crude second-order acoustic finite-difference modeler presented in the section on classic seismic waveform inversion is justified in that the observed and modeled

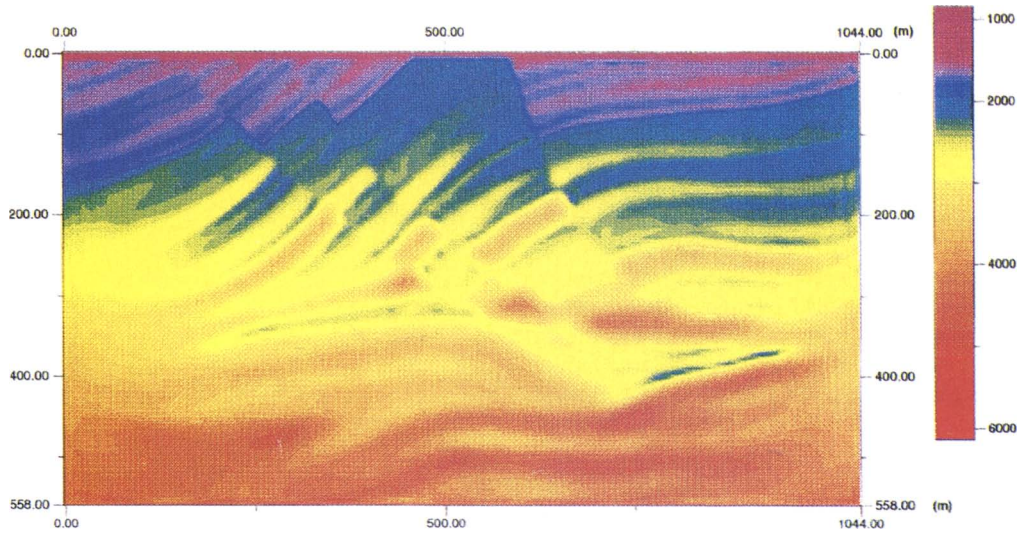

FIG. 15. Final best estimate of velocity model using the multigrid method for example 1 (to be compared to Figure 16). The velocity color bar has units of m/s.

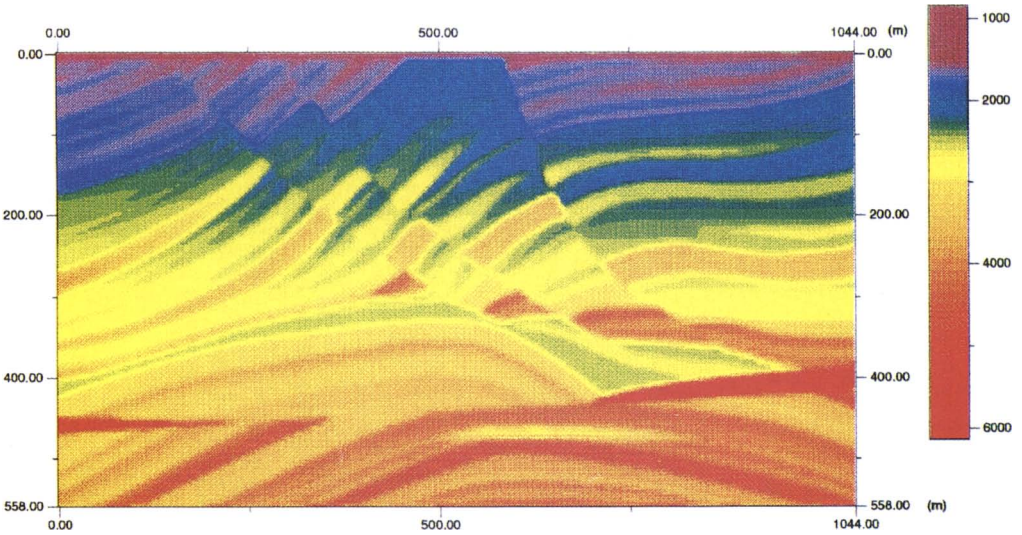

FIG. 16. Extracted part of real Marmousi velocity model for example 1. The velocity color bar has units of m/s.

data were both generated using this modeler. Furthermore, the source wavelet was presumed known, which eliminated another potential difficulty not related to the problem of local minima in the objective function.

As seen in the examples, the method proposed in this paper has been successful at inverting the Marmousi model from the regenerated data. This demonstrates that the multiscale approach is effective in diminishing the number of local minima that has been the main impediment to full waveform seismic inversion being a potentially feasible approach for seismic exploration. The regenerated data used in the examples, however, has much lower frequencies than is normally available in realistic seismic data sets. Consequently, it is not possible to draw definitive conclusions about the efficacy of the method for practical seismic exploration.

Nevertheless, we feel optimistic about the potential of the method for two reasons. First, the lack of low frequencies in the observed seismic data should not be a problem if the region of exploration is of sufficient complexity (as shown in Mora, 1989). If the region is not complex, then the enlargement of the effective acquisition surface should solve the problem. Second, the method should perform well even if the solution to the long-scale part of the problem finds only a local minimum. This is true since a local minimum found at a long scale of the problem should be reasonably close to the global minimum, for reasons pointed out in the section of this paper describing the multigrid method. The results in Versteeg (1991) show that for the band-limited Marmousi data, that velocity models quite distant from the true model yield very good migrated images. Consequently, the local minima found at the long-scale component of the seismic

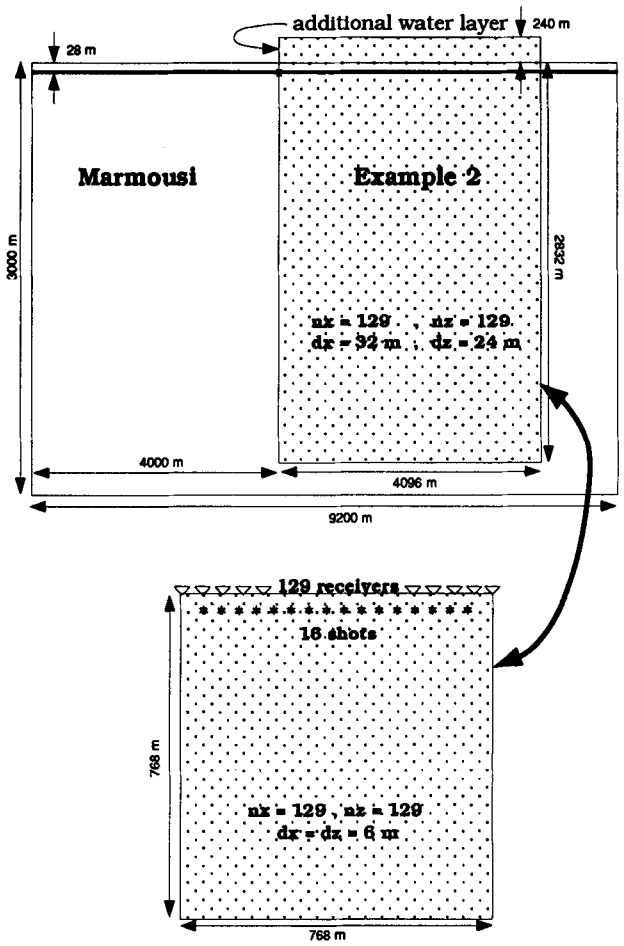

FIG. 18. Geometry of modified Marmousi velocity model and acquisition geometry for second example.

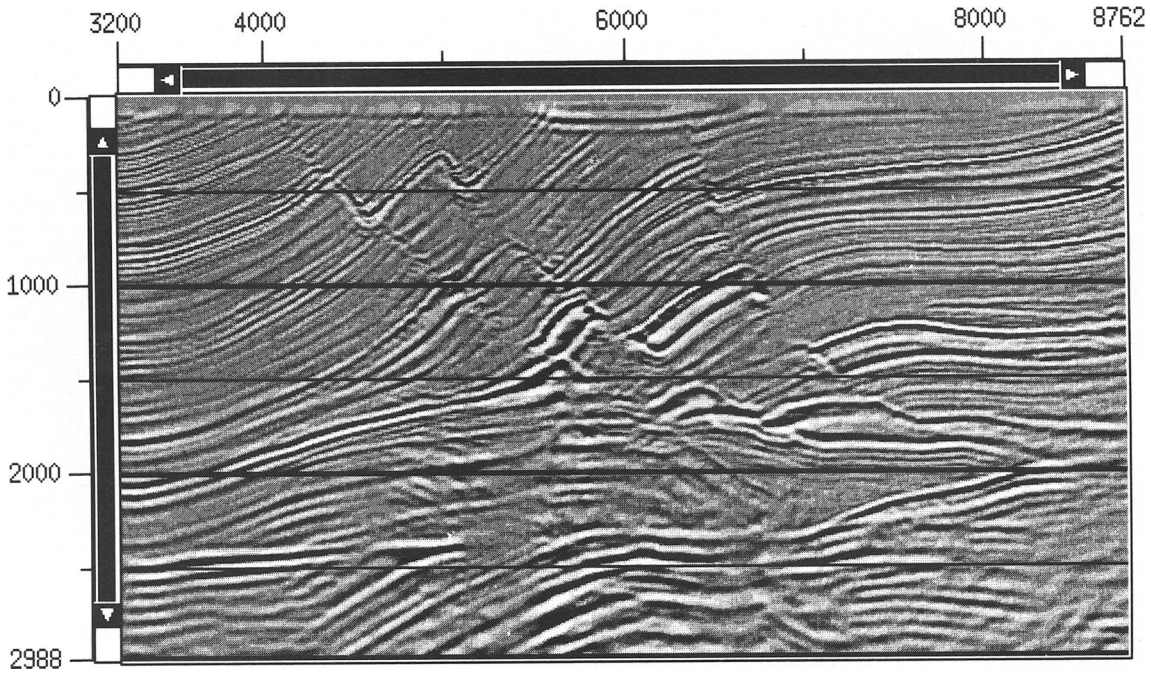

FIG. 17. Migrated image obtained using the best estimate of the velocity model in Figure 15 with a prestack finite-difference migration program.

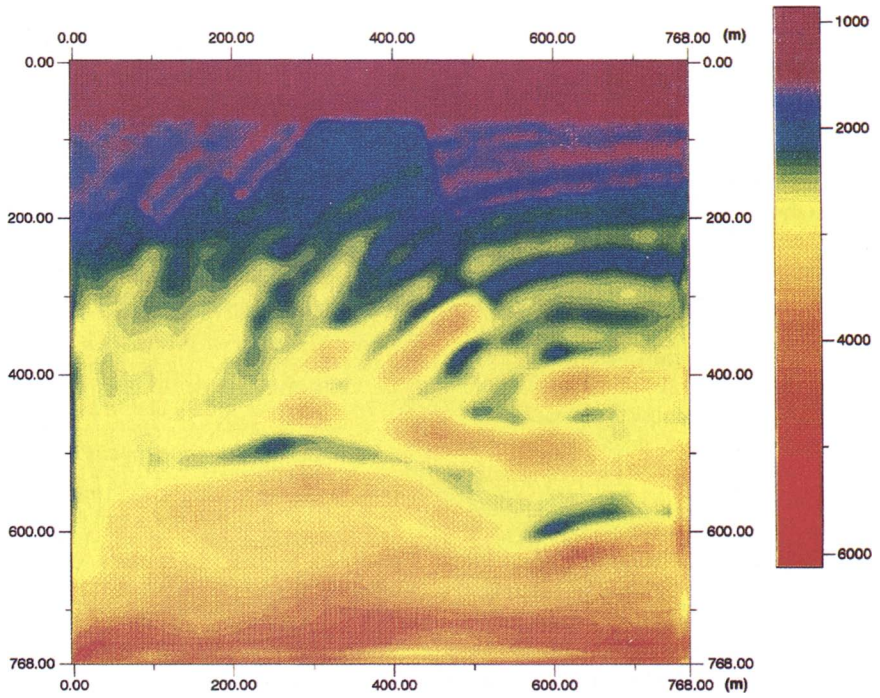

FIG. 19. Final best estimate of the velocity model using the multigrid method and subsampling for example 2 (to be compared to Figure 20). The velocity color bar has units of m/s.

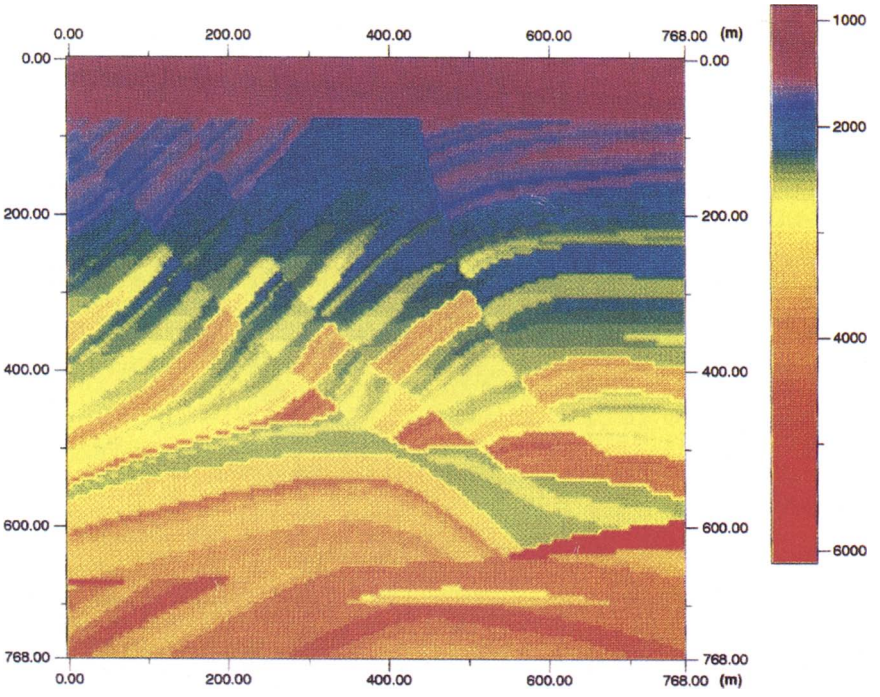

FIG. 20. Extracted part of real Marmousi velocity model for example 2. The velocity color bar has units of m/s.

inversion problem could be sufficiently close to the global minimum to yield good migrated images.

Also shown in this paper is the result that an important acceleration of the inversion procedure is obtained when Nyquist's theorem is used to subsample the longer scale components of the problem. Some practical implementation issues still need to be resolved for this aspect of the problem. Nevertheless, the computational acceleration obtained (on the order of several orders of magnitude) means that there is good potential that the method can be employed to solve more complicated wave equations (e.g., using elasticity and anisotropy) in 2-D and 3-D.

An important complement to this work is to test the sensitivity of the multiscale method in the presence of modeling and source function inaccuracies. There is reason to believe a priori that the multiscale approach to seismic inversion may be robust to modeling inaccuracies. Finite-difference modeling schemes are more accurate for low frequencies than for high frequencies since these frequencies are less sensitive to problems of numerical dispersion. Consequently, the inversion of the long scale components of the velocity model, which are the most important parts for accurate imaging, could be effective regardless of the order of the finite-difference modeler. A natural approach to resolving modeling and source function inaccuracies (if the multiscale approach does not turn out to be inherently robust to these problems) would be to explicitly model these phenomena as noise processes.

Finally, as has been discussed in the section on the multigrid method, the implementation of the elements of the multigrid approach can have an important effect on the efficacy of the method. For the problem of seismic inversion, more investigation is necessary to better understand what aspects of the problem are critical and how to optimize the implementation issues. An important direction for future research must include a thorough exploration of different types of multigrid schemes and a study of how automatic and adaptive methods can be obtained for determining the optimum scale decomposition and the optimum number of relaxation iterations per scale.

ACKNOWLEDGMENTS

The authors are very grateful to F. Bonnans and his research group at the Institut National de Recherche en Informatique et en Automatique in France for the permission to use their quasi-Newton optimization program for this work in seismic inversion.

REFERENCES

Al-Yahya, K. M., 1987, Velocity analysis by iterative profile migration: Ph.D. thesis, Stanford University.

Alexéev, V., Tikhomirov, V., and Fomine, S., 1982. Commande optimale: Éditions Mir.

Bamberger, A., Chavent, G., and Lailly, P., 1980, Étude de schémas numériques pour les équations de l'élastodynamique linéaire: Rapport INRIA France, **41**.

Berkhout, A. J., 1984, Multidimensional linearized inversion and seismic migration: Geophysics, **49**, 1881–1895.

Biondi, B. 1992. Solving the frequency-dependent eikonal equation: 62nd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1315–1319.

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

Bording, R. P., Gersztendorn, A., Lines, L. R., Scales, J. A., and Trietel, S., 1987, Applications of seismic traveltime tomography: Geophys. J. Roy. Astr. Soc., **90**, 285–303.

Brandt, A., 1977, Multi-level adaptive solutions to boundary value problems: Math. Comp., **13**, 333–390.

Briggs, W., 1987, A multigrid tutorial: S.I.A.M.

Bunks, C., 1991, Optimal velocity model inversion from analysis of paraxial wave equation focusing spots: 61st Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts, 1226–1229.

——— 1992, Paraxial wave-equation inversion with geometric constraints: 62nd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts, 901–904.

Chavent, G., and Jacewitz, C., 1990, Automatic determination of background velocities by multiple migration fitting: 60th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts, 1263–1266.

Daubechies, I., 1988, Orthonormal bases of compactly supported wavelets: Commun. in Pure and Applied Math., **41**, 901–996.

Devaney, A. J., 1984, Geophysical diffraction tomography: IEEE Trans. on Geoscience and Remote Sensing, **GE-22**, 1, 3–13.

Esmersoy, C., 1986, The backpropagated field approach to multidimensional velocity inversion: Ph.D. thesis, Massachusetts Institute of Technology.

Geman, S., and Geman, D., 1984, Stochastic relaxation, gibb's distributions, and the bayesian restoration of images: IEEE trans. Patt. Anal. and Mach. Intell., **PAMI-6**, 6, 721–741.

Geoltrain, S., and Brac, J., 1993, Can we image complex structures with first-arrival traveltime?: Geophysics, **58**, 564–575.

Hadjee, Y., 1988, Analyse du problème d'estimation de paramètres en sismique marine: Ph.D. thesis, Université Paris IX Dauphine.

Hildebrand, F. B., 1965, Methods of applied mathematics: Prentice-Hall, Inc.

Kelly, K. R., Ward, R. W., Treitel, S., and Alford, R. M., 1976, Synthetic seismograms: A finite-difference approach: Geophysics, **41**, 1, 2–27.

Kirkpatrick, S., Gellatt, C. D., and Vecchi, M. P., 1983, Optimization by simulated annealing: Science, **220**, 4598, 671–680.

Kolb, P., Collino, F., and Lailly, P., 1986, Pre-stack inversion of a 1-D medium: Proc. IEEE, **74**, 3, 498–508.

Lailly, P., 1984, The seismic inverse problem as a sequence of before stack migrations: Bednar, J. B., Robinson, E., and Weglein, A., Eds., SIAM Conf. Proc. on Inverse Scattering: Theory and Applications, 206–220.

Lanczos, C., 1962, The variational principles of mechanics: University of Toronto Press, second edition.

Levy, B., and Esmersoy, C., 1988, Variable background born inversion by wavefield back propagation: Siam J. Appl. Math., **48**, 4, 952–972.

Lindgren, J., 1992, Waveform inversion of seismic reflection data through local optimization methods: Ph.D. thesis, The Geophysical Tomography Group, Periodical Rep., no. 21, Institut de Physique du Globe, Paris, France.

Lindgren, J., Mora, P., and Tarantola, A., 1989, Nonlinear waveform inversion of seismic data: Retrieving both long and short wavelength velocity information: The Geophysical Tomography Group Periodical Rep., Institut de Physique du Globe, Paris, France, **17**, 2.1–2.23.

Luenberger, D. G., 1969, Optimization by vector space methods: John Wiley & Sons, Inc.

——— 1989, Linear and nonlinear programming: Addison-Wesley Publishing Co.

Mallat, S. G., 1989, A theory for multiresolution signal decomposition: The wavelet representation: IEEE trans. Patt. Anal. and Mach. Intell., **PAMI-11**, 7, 674–693.

Marroquin, J. L., 1985, Probabilistic solution of inverse problems: Ph.D. thesis, Massachusetts Institute of Technology.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of state calculations by fast computing machines: J. Chemical Physics, **21**, 1087–1092.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 9, 1211–1228.

——— 1989, Inversion=migration+tomography: Geophysics, **54**, 12, 1575–1586.

Oppenheim, A. V., and Schafer, R. W. 1975. Digital signal processing: Prentice-Hall Inc.

Pica, A., Tarantola, A., and Diet, J. P., 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 3, 284–292.

Press, W. H., and Teukolsky, S. A., 1991, Multigrid methods for boundary value problems I & II: Computers in Physics.

Rabiner, L. R., and Gold, B., 1975, Theory and application of digital signal processing: Prentice-Hall Inc.
Reynolds, A. C., 1978, Boundary conditions for the numerical solution of wave propagation problems: Geophysics, **43**, 1099–1110.
Saleck, F. M., Bunks, C., Zaleski, S., and Chavent, G., 1993, Combining the multigrid and gradient methods to solve the seismic inversion problem: 63rd Ann. Intern. Mtg, Soc. Expl. Geophys., Expanded Abstracts, 688–691.
Symes, W. W., and Carazzone, J. J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.
Tarantola, A. 1984. Linearized inversion of seismic reflection data: Geophysical Prosp., **32**, 998–1015.
Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903.
——— 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation: Pure and Applied Geophysics, **128**, 1/2, 365–399.
van Trier, J. A., 1990, Tomographic determination of structural velocities from depth-migrated seismic data: Ph.D. thesis, Stanford University.
Versteeg, R. J. 1991. Analysis of the velocity model determination problem for seismic imaging: Ph.D. thesis, Université Paris VII Jussieu.
Versteeg, R. V., and Grau, G. (eds.), 1991, The marmousi experience: Proc. 1990 EAEG Workshop on Practical Aspects of Seismic Data Inversion.