AN ADAPTIVE MULTI-SCALE CONJUGATE
GRADIENT METHOD FOR DISTRIBUTED
PARAMETER ESTIMATION OF 2-D
WAVE EQUATION*

Feng Guofeng(冯国峰)  Han Bo (韩 波)

Abstract  An adaptive multi-scale conjugate gradient method for distributed parameter estimations (or inverse problems) of wave equation is presented. The identification of the coefficients of wave equations in two dimensions is considered. First, the conjugate gradient method for optimization is adopted to solve the inverse problems. Second, the idea of multi-scale inversion and the necessary conditions that the optimal solution should be the fixed point of multi-scale inversion method is considered. An adaptive multi-scale inversion method for the inverse problem is developed in conjunction with the conjugate gradient method. Finally, some numerical results are shown to indicate the robustness and effectiveness of our method.

Key words  Multi-scale; conjugate gradient method; wave equation.

AMS(2000) subject classifications  65M99, 47J06

1  Introduction

Designing effective numerical methods for the nonlinear inverse wave equation is a very significant challenge. Especially for large-scale inverse problems, the number of parameters to be estimated is mesh-dependent. As we know, there are some reasons:

First, conventional nonlinear solvers such as Newton’s method entail difficulties for large-size problems. The Hessian matrices for Newton’s method are formally dense and are of the order of the number of the inversion parameters. Their construction involves numerous solutions of the direct problem. Therefore, the straightforward implementations based on grid require prohibitive memory and computational cost. The denseness, large scale, and construction of the

* Received: Sep. 1, 2004.
linearized inverse operators rule out matrix-based iterative solvers.

Secondly, the objective functions of the nonlinear inverse problems are often non-convex. In other words, Newton’s method will diverge if globalization is neglected. Even worse, for many nonlinear inverse problems, the objective functions to be minimized may have many local minima. Newton’s method as well as other local methods will be trapped at a local minimum unless the objective function can be convexified\cite{1}.

Given the above difficulties faced by the Newton-like methods, it is not surprising that many large-scale PDE inverse methods are based on gradients alone in order to avoid Hessian matrices. The simplest of such techniques are the steepest decent-type methods, in which the current solution is improved by moving in the direction of the negative gradient of the objective function. Better directions can be constructed based on gradient information, resulting in the nonlinear conjugate gradient (CG) method. All of these gradient-based methods avoid the use of Newton Hessians.

Indeed, any large-scale nonlinear inverse problems defined in physical space can have an approximated description at any given scale of the same space. The multi-scale algorithm recursively constructs a sequence of such descriptions with increasingly larger (coarser) scales, and combines local processing at each scale with various inter-scale interactions. As a result of such multilevel interactions, the fine scales can be employed efficiently. Moreover, the inter-scale interaction can eliminate various kinds of difficulties, such as: slow convergence in minimization processes; ill-posedness of inverse problems; large-scale attraction basin traps in global optimization; etc\cite{1}.

The multi-scale algorithms have been applied to the inverse 2-D Wave Equation; for example, E. Gelman and J. Mandel\cite{2}, J. Thomas King\cite{9}, U. M. Ascher and E.Haber\cite{3} performed the analysis of multi-scale inversion algorithms. Y. M. Chen\cite{4}, W. H. Chen\cite{5}, J.L. Qian etc\cite{6,7,8} improved the efficiency of the generalized pulse-spectrum technique (GPST) inversion algorithm by using multi-scale algorithms. Some researchers combined the multi-scale idea with various iterative solvers for solving inverse problems\cite{9}. Another popular approach is the multi-resolution wavelet theory\cite{10,11,12,13,14,15}.

Once the scale has been chosen, the actual level of refinement is difficult to choose: if the scale is too poor, the inversion parameter will not honor properly the data. If it is too rich, it will lead to overparameterization. The adaptability of multi-scale algorithms is hence of crucial importance.

The existing multi-scale methods often employ adjoint equations at each scale\cite{4}, which increase the computational work. In the paper, an adaptive multi-scale conjugate gradient method
An Adaptive Multi-scale Conjugate Gradient Method for Distributed Parameter Estimation

for distributed parameter estimation (or inverse problems) for wave equation is presented. The coefficient identification of wave equations in two dimensions is considered as our model problem. First, the conjugate gradient method for optimization is adopted to solve the inverse problems in Section 2. Second, the idea of multi-scale inversion and the necessary condition that the optimal solution should be the fixed point of multi-scale inversion method is considered. An adaptive multi-scale inversion method for the inverse problem is developed in conjunction with the conjugate gradient method in Section 3. Third, some numerical results are shown to indicate the robustness and effectiveness of our method in Section 4. Finally, some conclusions are drawn in Section 5.

2 The conjugate gradient method for inversion of 2-D wave equation

The acoustic wave equation for a 2-D medium, describing the propagation of the seismic wave, can be written as

\[
\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{v^2(x, z)} \frac{\partial^2 u}{\partial t^2} = 0, \tag{1}
\]

where \(u(x, z, t)\) is the particle displacement field, \(v(x, z)\) is the velocity at \((x, z)\). If the source is a plane wave, the equation can be written as

\[
\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{v^2(x, z)} \frac{\partial^2 u}{\partial t^2} = g(t)\delta(x), \tag{2}
\]

where \(\delta(\cdot)\) is the Dirac-\(\delta\) function, and \(g(t)\) is the source time function that obeys \(g(t) = 0\) for \(t < 0\). We study the problem in a domain \(\Omega = [0, L] \times [0, H]\), with the boundary conditions

\[
\left. \frac{\partial u(x, z, t)}{\partial x} \right|_{x=0,L} = 0, \tag{3}
\]

\[
\left. \frac{\partial u(x, z, t)}{\partial z} \right|_{z=H} = 0, \tag{4}
\]

and the initial conditions

\[
u(x, z, 0) = \frac{\partial u(x, z, 0)}{\partial t} = 0. \tag{5}\]

Equations (2) – (5) form the forward problem of the acoustic wave equation. Given the velocity \(v(x, z)\), the seismograms \(u(x, 0, t) = f(x, t)\) can be calculated from the equations (2) – (5), where \(f(x, t)\) is seismogram on the earth’s surface. The inverse 2-D acoustic wave problem is to solve for the unknown \(v(x, z)\) from \(f(x, z)\). Equations (2) – (5) and \(u(x, 0, t) = f(x, t)\) can be
approximated by the second-order finite difference equations

\[
\begin{align*}
\begin{split}
u_{i,j}^{k+1} & = \frac{\tau^2}{h_x^2} (u_{i-1,j}^k + u_{i+1,j}^k - 2u_{i,j}^k) + \frac{\tau^2}{h_z^2} (u_{i,j+1}^k + u_{i,j-1}^k - 2u_{i,j}^k) - u_{i,j}^{k-1} \\
i & = 1, 2, \ldots, m - 1, \quad j = 1, 2, \ldots, n - 1. \\
u_{i,0}^{k+1} & = \frac{\tau^2}{h_z^2} (u_{i+1,0}^k + u_{i-1,0}^k - 2u_{i,0}^k) + \frac{\tau^2}{h_x^2} (u_{i,1}^k - u_{i,0}^k) + 2u_{i,0}^k - u_{i,0}^{k-1} - \beta \tau^2 v_{i,0}^2 \\
i & = 0, 1, \ldots, m.
\end{split}
\end{align*}
\]

where \(u_{i,j}^k = u(ih_x, jh_z, k\tau)\), \(v_{i,j} = v(ih_x, jh_z)\), \(m = \frac{h_x}{h_z}, n = \frac{h_z}{\tau}\) is the time step length, and \(h_x, h_z\) are the spacings of the grid in the \(x-\) and \(z-\) direction, respectively. The difference equations (6) thus define a vector-valued nonlinear function \(A : V \rightarrow F\), where \(V\) denotes a vector which is composed of \(v_{i,j}\) in a suitable sequence and \(F\) denotes a vector which is composed of \(f_i^k\) in a corresponding suitable sequence. Let \(\hat{f}^k\) denote the observed data and vector \(\hat{F}\) in the same sequence as \(F\).

Let \(V^T = (v_{1,0}, v_{1,1}, \ldots, v_{1,n-1}, v_{2,0}, v_{2,1}, \ldots, v_{2,n-1}, \ldots, v_{m-1,0}, v_{m-1,1}, \ldots, v_{m-1,n-1})\),

\[
F^T = (f_1^1, f_2^1, \ldots, f_{m-1}^1, f_1^2, f_2^2, \ldots, f_{m-1}^2, \ldots, f_1^m, f_2^m, \ldots, f_{m-1}^m),
\]

\[
\hat{F}^T = (\hat{f}_1^1, \hat{f}_2^1, \ldots, \hat{f}_{m-1}^1, \hat{f}_1^2, \hat{f}_2^2, \ldots, \hat{f}_{m-1}^2, \ldots, \hat{f}_1^m, \hat{f}_2^m, \ldots, \hat{f}_{m-1}^m).
\]

The inversion for \(V\) can be transformed into solving the following minimum problem

\[
\min \| A(V) - \hat{F} \|^2.
\]

It is well known that the problem (7) is ill-posed. For the numerical stability of the algorithm, the smoothing functional is introduced followed by the Tikhonov regularization instead of direct solution of (7). We consider another minimum problem

\[
\min \left[ \| A(V) - \hat{F} \|^2 + \lambda_1 \| M_1 V \|^2 + \lambda_2 \| M_2 V \|^2 \right],
\]

where \(M_1, M_2\) are the second-order smoothing matrices in the \(x-\) and \(z-\) direction, respectively, and \(\lambda_1\) and \(\lambda_2\) are the regularization parameters.

Let

\[
J(V) = \left[ \| A(V) - \hat{F} \|^2 + \lambda_1 \| M_1 V \|^2 + \lambda_2 \| M_2 V \|^2 \right],
\]

We will construct conjugate gradient method for the minimum of the objective function,
where the gradient of the objective function is
\[ \nabla J(V^k) = \frac{\partial J(V^k)}{\partial V^k} = 2[A(V^k) - \hat{F}]^T \frac{\partial A(V^k)}{\partial V^k} + 2\lambda_1 M_1^T M_1 V_k + 2\lambda_2 M_2^T M_2 V_k. \tag{10} \]

In principle, the length \( \alpha^k \) at the \( k \)th iteration is derived via minimum of objective function,
\[ \frac{\partial J(V^k - \alpha^k d^k)}{\partial \alpha^k} = 0. \tag{11} \]

Consider the Taylor’s expansion of \( A(V_k - \alpha^k d^k) \) at \( V^k \) and ignore the terms of second or higher orders, we can get
\[ A(V^k - \alpha^k d^k) = A(V^k) - \frac{\partial A(V^k)}{\partial V^k} \alpha^k d^k. \]

Combining the expression with (8) and (10), we arrive at:
\[ [A(V^k) - \hat{F}]^T \frac{\partial A(V^k)}{\partial V^k} d^k + \lambda_1 [M_1 V^k]^T d^k + \lambda_2 [M_2 V^k]^T d^k \]
\[ = \alpha^k \left\{ \left[ \frac{\partial A(V^k)}{\partial V^k} d^k \right]^T \left[ \frac{\partial A(V^k)}{\partial V^k} d^k \right] + \lambda_1 [M_1 d^k]^T d^k + \lambda_2 [M_2 d^k]^T d^k \right\}. \tag{12} \]

The conjugate gradient algorithm for inversion of 2-D wave equation is thus constructed:

1. Let \( k = 0 \), give an initial guess \( V^0 \) and \( d^0 = \nabla J(V^0) \);
2. Solve the forward problem for \( u(x, z, t) \);
3. Calculate the objective function \( J(V^k) \) in (9). Stop iterating if termination rule is valid; otherwise, go to the next step.
4. Calculate \( \frac{\partial A(V^k)}{\partial V^k} \);
5. Calculate the gradient vector \( \nabla J(V^k) \) in (10);
6. Calculate conjugate coefficient \( \beta^k = \frac{\nabla J^T(V^k) \nabla J(V^k)}{\nabla J^T(V^k-1) \nabla J(V^k-1)} \) and descent direction \( d^k = \nabla J^T(V^k) + \beta^k d^{k-1} \);
7. Calculate the length \( \alpha^k \) in (12);
8. Let \( k = k + 1 \) and \( V^{k+1} = V^k - \alpha^k d^k \), return Step 2.

The above algorithm avoids distinctly the calculation cost required for the conventional adjoint methods by introducing the formula that chooses length \( \alpha^k \). At the same time, the choice of length is dependent on the calculated direction and the first order derivative of the objective function only. The ease of applying the algorithm to multi-scale inversion is obvious.
3 An adaptive multi-scale conjugate gradient method

As discussed in Section 1, the size of real-life inverse problems is often large scale. Thus, finding solutions to such inverse problems requires sufficient computational work. Application of the multi-scale algorithm can help reducing the computational cost intensively. On the other hand, if a multi-scale algorithm is applied directly, unnecessary computational work is spent in some regions where the inversion parameter is unchanged or changed faintly. In this section, we present an adaptive multi-scale conjugate gradient method for distributed parameter estimation of 2-D wave equation, that modifies the conjugate gradient method proposed in Section 2. At each scale, with the necessary condition of well-posedness of multi-scale inversion, the new algorithm chooses the region to be refined adaptively. The algorithm is listed below:

The model problem is approximated by finite difference and is transferred into a minimum problem

\[
\min_{V \in \Omega} \| A(V) - \hat{F} \|^2.
\]  

(13)

(1) On a grid \( \Omega_1 = \{(ih_x, jh_z): i = 0, 1, \cdots, I; j = 0, 1, \cdots, J\} \) as coarse as possible on which the stability condition holds.

(2) On the coarsest grid \( \Omega_1 \), let \( k = 0 \) iterate with the proposed conjugate gradient method start from initial gauss value \( V^0 \);

(3) Stop iteration until \( \| \delta V^k \| / \| V^k \| < \epsilon \), where \( \epsilon \) is a given relative error bound; otherwise, interpolate derived \( V_{\Omega_1}^{k+1} \) into the next finer grid \( \Omega_2 = \{(i\frac{h_x}{2}, j\frac{h_z}{2}): i = 0, 1, \cdots, 2I; j = 0, 1, \cdots, 2J\} \), \( V_{\Omega_2}^0 = I_{\Omega_2}^{\Omega_1} V_{\Omega_1}^{k+1} \), where \( I_{\Omega_2}^{\Omega_1} \) is a interpolation operator from \( \Omega_1 \) to \( \Omega_2 \).

(4) The gradient \( \frac{\partial J(V_{\Omega_2}^0)}{\partial V_{\Omega_2}^0} \) of the objective function is solved on \( \Omega_2 \). The necessary condition that the optimal solution should be fixed point of multi-scale inversion algorithm is considered:

\[
\nabla J(V_{\Omega_2}^0) = \nabla J(V_{\Omega_1}^{k+1}) I_{\Omega_2}^{\Omega_1}.
\]  

(14)

Compare the elements at the both sides of (14) correspondingly. If the equation is valid at some locations, iterates considered have provided local resolution well enough at these locations on \( \Omega_1 \) to ensure the solution of multi-scale methods a fixed point, and grids at these locations keep unchanged; otherwise, the level of refinement is considered not enough well and local refinement is performed with the use of the new grid point from \( \Omega_2 \) and the unchanged grid point from \( \Omega_1 \), whose values form new initial guess \( V^0 \), go back to Step 3.

The multi-scale algorithm reduced the large-scale inversion problem to some medium-scale sub-problems by refining some sub-regions in the whole region adaptively. The region in which
coefficients of wave equation change intensively is thus found.

4 Numerical simulations

In order to test the feasibility and the general characteristics of the method in Section 3, the following numerical simulations are carried out. Two numerical examples are presented.

In both examples, the source function is \( g(t) = 60\sin(2\pi ft^2)\exp(-2\pi ft^2) \), where \( f = 60\)Hz. The whole surveyed region is 120m \(	imes\) 120m. There are 33 points in horizontal and vertical direction, respectively. The time spacing is \( t = 1\)ms. The regularization parameters are \( \lambda_1 = \lambda_2 = 10^{-4} \).

The first exact velocity model is shown in figure 1(a), which is a layered model. In the inversion process, choose the initial estimates \( V_{i,j} = 2400\)m/s in whole region and exact parameters are \( V_{i,j} = 2500\)m/s out abnormity and \( V_{i,k} = 2600\)m/s in abnormity. The inversion result is plotted in figure 1(b). We find that our adaptive multi-scale method saves more than 25% computational cost in comparison with mono-scale method under same relative error bound (10%).

![Figure 1](image1.png)

Figure 1 A layered model. (a) The exact velocity mode. (b) The inversion result.

![Figure 2](image2.png)

Figure 2 A multi-abnormity model. (a) The exact velocity mode. (b) The inversion result.
The second velocity model is shown in figure 2(a), which is a multi-abnormity one. The source function and parameters are the same as the first model. In the inversion process, the initial estimates are $V_{i,j} = 2400 \text{m/s}$. The exact parameters are $V_{i,j} = 2500 \text{m/s}$ out abnormity, $V_{i,j} = 2600 \text{m/s}$ in two left abnormities and $V_{i,j} = 2700 \text{m/s}$ in right abnormity. The inversion result is plotted in figure 2(b). In this example, Our adaptive multi-scale method help reducing more than 15% computational cost in comparison with the mono-scale method under same relative error bound (10%).

5 Conclusions

In this paper, we present numerical simulations, suggest that: (1) The multi-scale conjugate gradient inversion algorithm is a stable and fast convergent iterative method, that avoids the computation of adjoint systems. (2) Adaptability reduces computational cost while keeping the well-posedness of the algorithm, precision of inversion, and virtues of multi-scale method, which greatly improves the computational efficiency and synthetically overcomes the trapping to local minima.

References


**Feng Guofeng**  Department of Mathematics, Harbin Institute of Technology, Harbin 150001, PRC.

**Han Bo**  Dept. of Mathematics, Harbin Institute of Technology, Harbin 150001, PRC.