Computers & Geosciences 60 (2013) 88–97

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Wave-equation traveltime inversion: Comparison of three numerical optimization methods

Yikang Zheng [a,b], Yibo Wang [a,*], Xu Chang [a]

[a] Key Laboratory of Engineering Geomechanics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 22 February 2013
Received in revised form
9 July 2013
Accepted 15 July 2013
Available online 23 July 2013

Keywords:
Wave-equation traveltime inversion
Steepest descent method
Conjugate gradient method
L-BFGS method

ABSTRACT

Compared to traditional traveltime inversion methods, no traveltime picking, high-frequency assumption or ray tracing is necessary in wave-equation traveltime inversion (WT). Another merit of WT is that it is insensitive to the starting model. Although WT offers less detailed parts of velocity model than the full waveform inversion (FWI), it can provide a good initial velocity model for FWI. In this paper, the steepest descent, conjugate gradient and limited-memory BFGS (L-BFGS) optimization methods are used in the implementation of acoustic WT. We use synthetic crosswell data for testing and the numerical results show that L-BFGS has a faster rate of convergence and offers a reconstructed velocity model with better resolution compared to the other two gradient methods.

1. Introduction

The various seismic velocity inversion methods have two major categories, traveltime inversion (Dines and Lytle, 1979; Bishop et al., 1985; Ivansson, 1985; Paulsson et al., 1985; Pyun et al., 2005) and full waveform inversion (FWI) (Tarantola, 1986; Mora, 1987; Crase et al., 1992; Shin and Cha, 2008). Both methods have complementary strong and weak points (Zhou et al., 1995).

Traditional traveltime inversion can fail when the velocity variations have nearly the same wavelength as does the source wavelet. On the other side, the traveltime misfit function to be minimized can be quasi-linear with respect to the variation between the presumed and actual velocity model. High quality inversion results can be attained even if the starting model is far from the actual model (Luo and Schuster, 1991). Full waveform inversion can reconstruct a highly resolved velocity model, but the problem is that its misfit function can be highly nonlinear with respect to the velocity model (Gauthier et al., 1986; Luo and Schuster, 1991). A gradient method will tend to get stuck in local minima if the starting model is quite different from the actual model (Zhou et al., 1995). Wave-equation traveltime inversion (WT) minimizes traveltime residuals using traveltimes and derivatives computed from solutions to the wave equation (Luo and Schuster, 1991). WT is quickly convergent for starting model far from the actual model and the reconstructed velocity model has a relatively high resolution.

Seismic inversion is essentially an optimization problem thus the choice of numerical algorithm is important. To solve problems, researchers usually choose iterative methods that generate a sequence of improving approximate solutions. The methods differ according to whether they evaluate Hessians, gradients, or only functions value. For the 2nd derivatives optimizer, such as Newton's method, the number of function calls in each iteration is in the order of N^2, but for a gradient optimizer it is only N. However, gradient optimizers usually need more iterations than Newton's algorithm. In large scale problems, the computational complexity may be excessively high. Quasi-Newton methods are more often used. In the implementation of WT, we use two typical gradient methods, the steepest descent and conjugate gradient, and one quasi-Newton method (L-BFGS).

Waveform inversion has emerged as a key tool in petroleum and natural gas exploration. However, waveform inversion tends to get stuck in local minima and requires sufficient low-frequency components in data. WT can recover long-wavelength structures of the velocity model and offer a high quality starting model for full waveform inversion (FWI). Successful applications of WT can be achieved to the field data which contains unreliable components less than 5 Hz. Many researchers has presented that quasi-Newton method outperforms the classic gradient methods in FWI (Sheen et al., 2006; Brossier et al., 2009; Virieux and Operto, 2009). As WT has become a hot spot for its merits in geophysical prospecting, a significant question to be settled is whether quasi-Newton algorithm still provides higher convergence rate and resolving power over the gradient methods.

This paper first presents the theory of WT and the numerical optimization methods we choose for WT. We then apply WT to 2D

* Corresponding author. Tel./fax: +86 10 82998624.
E-mail address: wangyibo@mail.iggcas.ac.cn (Y. Wang).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.07.010

synthetic crosswell data and compare the results of three methods. Finally, a conclusion is given.

2. Wave-equation traveltime inversion

Let $p(x_r,t;x_s)_{obs}$ denotes the observed pressure seismograms measured at the receiver location $x_r(s=1, 2,\ldots, N_r)$ due to a source at $x_s(s=1, 2,\ldots, N_r)$. The source is always assumed to be excited at $t=0$. For a given velocity model, $p(x,t;x_s)_{cal}$ denotes the calculated seismogram that honors the 2-D wave equation.

$$\frac{1}{c^2(x)}\frac{\partial^2 p(x,t;x_s)}{\partial t^2} - \rho(x)\nabla\left[\frac{1}{\rho(x)}\nabla p(x,t;x_s)\right] = s(t;x), \tag{1}$$

where $\rho(x)$ is the density, $c(x)$ is the wave velocity and $s(t;x)$ is the source function.

The forward modeling problem is defined as finding the pressure field that satisfies Eq. (1) with the given boundary and initial conditions. The inverse problem is defined as finding the velocity model which is in accordance with the observed seismograms and minimizes the misfit function (Zhou et al., 1995), the misfit function is defined as (Luo and Schuster, 1991)

$$S = \frac{1}{2}\sum_s\sum_r \Delta\tau(x_r,x_s)^2, \tag{2}$$

where $\Delta\tau(x_r,x_s)=\tau_{obs}(x_r,x_s)-\tau_{cal}(x_r,x_s)$ is the difference between the observed and calculated first arrival times for a source at x_s and a receiver at x_r.

We choose a steepest descent method to minimize traveltime residual S. To update the velocity model, the steepest descent method gives

$$c_{k+1}(x) = c_k(x) + \alpha_k\gamma_k(x), \tag{3}$$

where $\gamma_k(x)$ is the steepest descent direction, α_k is the step length and x represents any location between the wells, k denotes the kth iteration.

Taking the Fréchet derivative of S,

$$\gamma(x) = -\frac{\partial S}{\partial c(x)} = -\sum_s\sum_r \frac{\partial(\Delta\tau)}{\partial c(x)}\Delta\tau(x_r,x_s) \tag{4}$$

From Eq. (8) in Luo and Schuster (1991)

$$\gamma(x) = \frac{1}{c^3(x)}\sum_s\int dt\, \dot{p}(x,t;x_s)_{cal}\dot{p}'(x,t;x_s), \tag{5}$$

where $\dot{p}'(x,t;x_s)=\sum g(x,-t;x_r,0)*\delta\tau(x_r,t;x_s)$, the symbol * represents temporal convolution, $\dot{p}$ represents the time derivative of p, and $g(x,-t;x_r,0)$ is the Green's function associated with Eq. (1) for the velocity field $c_k(x)$. Here $\delta\tau$ is the pseudo-traveltime residual defined by

$$\delta\tau(x_r,t;x_s) = -\frac{2}{E}\dot{p}(x_r,t+\Delta\tau;x_s)_{obs}\Delta\tau(x_r,x_s), \tag{6}$$

where E is defined as

$$E = \int dt\, \dot{p}(x_r,t+\Delta\tau;x_s)_{obs}\dot{p}(x_r,t;x_s)_{cal}, \tag{7}$$

For a single source, the interpretation of Eq. (5) is that the gradient at x is obtained from the cross-correlation of the forward-modeled field $\dot{p}(x,t;x_s)_{cal}$ with the back-projected field $\dot{p}'(x,t;x_s)$. We weight the observed seismograms at the x_r receiver with its associated traveltime residual $\Delta\tau(x_r,x_s)$ and normalization value E to get the pseudo-traveltime residual $\delta\tau(x_r,t;x_s)$, then back-project it to find the $\dot{p}'(x,t;x_s)$ (Zhou et al., 1995).

3. Numerical optimization methods

We try to minimize the traveltime residual function

$$S = \frac{1}{2}\sum_s\sum_r \Delta\tau(x_r,x_s)^2 \tag{8}$$

An iterative method uses an initial assumed model to generate successive approximations to a solution. Let α_k denotes the step

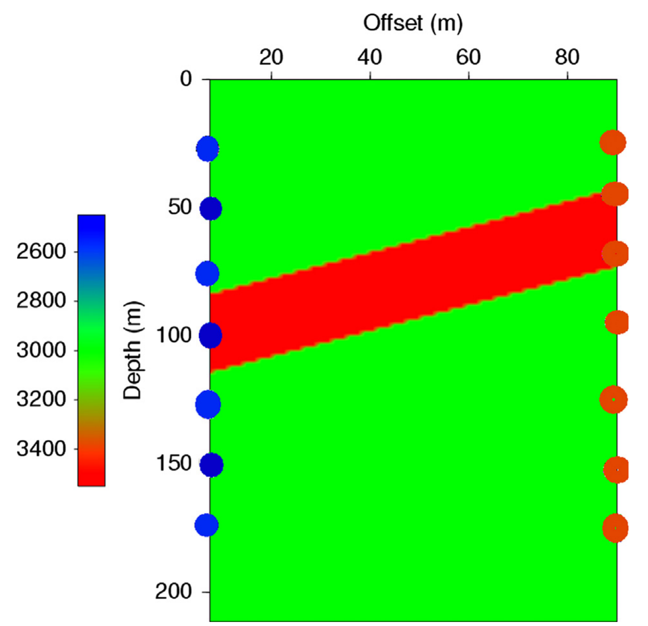

Fig. 1. A fault model with a dipping layer, the blue dots along the left well represent the sources and the red dots along the right well represent the receivers (not all the sources and receivers are marked). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

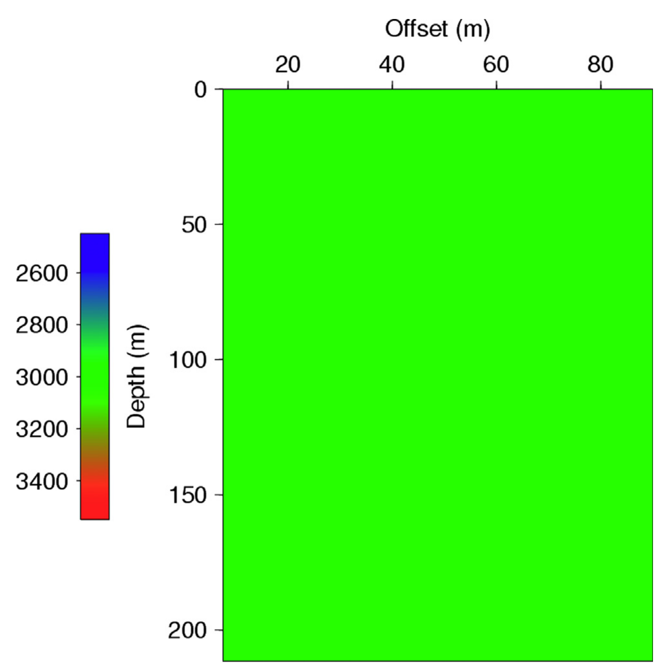

Fig. 2. The starting model, it is a constant-velocity model.