Journal of Applied Geophysics 152 (2018) 9–16

Contents lists available at ScienceDirect

Journal of Applied Geophysics

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

Mini-batch optimized full waveform inversion with geological constrained gradient filtering

Hui Yang [b], Junxiong Jia [a], Bangyu Wu [a,*], Jinghuai Gao [c,d]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China
[b] Earth and Space Sciences, Southern University of Science and Technology of China, Shenzhen, Guangdong 518055, PR China
[c] School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China
[d] National Engineering Laboratory for Offshore Oil Exploration, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China

ARTICLE INFO

Article history:
Received 6 May 2017
Received in revised form 7 December 2017
Accepted 19 February 2018
Available online 16 March 2018

ABSTRACT

High computation cost and generating solutions without geological sense have hindered the wide application of Full Waveform Inversion (FWI). Source encoding technique is a way to dramatically reduce the cost of FWI but subject to fix-spread acquisition setup requirement and slow convergence for the suppression of cross-talk. Traditionally, gradient regularization or preconditioning is applied to mitigate the ill-posedness. An isotropic smoothing filter applied on gradients generally gives non-geological inversion results, and could also introduce artifacts. In this work, we propose to address both the efficiency and ill-posedness of FWI by a geological constrained mini-batch gradient optimization method. The mini-batch gradient descent optimization is adopted to reduce the computation time by choosing a subset of entire shots for each iteration. By jointly applying the structure-oriented smoothing to the mini-batch gradient, the inversion converges faster and gives results with more geological meaning. Stylized Marmousi model is used to show the performance of the proposed method on realistic synthetic model.

1. Introduction

Full Waveform Inversion (FWI) has the potential to meet high resolution object of subsurface property characterization in seismic exploration. It depicts the frame of matching modeling data, in terms of misfit function, with the field data by finding optimal subsurface parameters. In general, the solutions to this problem can be categorized as local optimization (Lailly, 1983; Tarantola, 1984; Pratt et al., 1998; Wu et al., 2014; Luo et al., 2016) or global optimization methods (Afanasiev et al., 2014; Gao et al., 2016; Datta and Sen, 2016) according to whether or not the gradient of the misfit function for model updating is used. The gradient based iterative updating strategy has overwhelming dominance with affordable computing cost since the descent direction can be calculated by cross correlation of the source wavefield and reverse propagation of the misfit function residuals (Tarantola, 1984). Especially, the advance in high-performance computing propels the development of Reverse Time Migration (RTM) which can be an efficient engine to speed up FWI. However, this progress has increased the demand for larger surveys, with a significant growth of source number for wide azimuth acquisition in 3D which slows down the iterative updating process significantly. The computation cost for FWI with tens or hundreds of prestack RTM is still one of the obstacles for practical application.

The frequency domain direct solver technique is efficient with the cost of one complex LU decomposition or matrix inversion and the cost of a matrix-vector multiplication for each source (Krebs et al., 2009). Frequency domain FWI is stable and efficient with good results by inverting a small number of frequencies recursively (Pratt, 1999), but has memory limit to tackle the large matrix system for 3D problems. Source encoding, on the other hand, assembling several sources together for seismic migration/imaging, allows for the reduction of simulation quantity so as to speed up prestack data processing (Morton and Ober, 1998; Jing et al., 2000; Romero et al., 2000; Neelamani et al., 2008; Dai et al., 2013). Source encoding waveform inversion is investigated by the geophysical community with immense enthusiasm (Krebs et al., 2009; Ben-Hadj-Ali et al., 2011; Li et al., 2012; Moghaddam et al., 2013; Anagaw and Sacchi, 2014; Son et al., 2014). It reduces the cost of FWI dramatically since the number of seismic simulations for the misfit function is proportional to the number of sources. By assuming zero mean, the cross-talks for both the misfit function and its gradient are suppressed by the specially devised encoding technique at each iteration. However, such methods are usually fast in initial iterations, the need to average out the cross-talk makes them slow to converge later on (van Leeuwen and Herrmann, 2013). Instead of source encoding, van Leeuwen and Herrmann (2013) investigated a hybrid optimization

* Corresponding author.
E-mail address: bangyuwu@xjtu.edu.cn (B. Wu).

https://doi.org/10.1016/j.jappgeo.2018.02.011

strategy developed by Friedlander and Schmidt (2012), which is the stochastic optimization method using a different, randomly chosen, sequential sources for updating. It ensures the reduction of synthetic computation time while free of cross-talk. In machine learning, this hybrid scheme shares the similar idea as mini-batch gradient descent optimization. While this does not have the trouble of source cross-talk, the acquisition footprints have to be dealt with because of the lack of shot data. In general, Gaussian smoothing filter can be applied to the gradient as a way to mitigate the footprints. However, an isotropic Gaussian filter smears the boundary of structures in the model which leads to a model with less geological features. Guitton et al. (2012) applied directional Laplacian filters to the model reparameterization, leading to a geological constrained full-waveform inversion. Not only did it achieve faster convergence at early iterations but it generated models with geological meaning. Mao et al. (2016) applied image-guided smoothing to reduce the crossline footprint effects for marine Narrow Azimuth acquisition (NAZ) data and tests showed that the proposed FWI generated sharp contrast update around key structures.

This paper studies the feasibility of FWI in combination with mini-batch gradient optimization and structure oriented filter, to accelerate iterations and constrain model updates. Structure-oriented bilateral filtering (Hale, 2011) is adopted for the mini-batch gradient smoothing. In the following, the theory of FWI is first reviewed as the basis for later sections. Then the mini-batch misfit and conjugate gradient FWI is formulated. In the third section, the geological constrained structure-oriented filter smoothing is presented. Finally, the modified shallow water Marmousi-II model is used to illustrate the performance of this method, both on efficiency and inversion quality. We also illustrate the inversion results of isotropic and geological constrained anisotropic mini-batch gradient smoothing. The last part is the discussions and conclusions.

2. Iterative approximation solution of the full waveform inversion

FWI is a highly nonlinear optimization problem, for which iterative numerical methods are typical solutions for the efficiency and easiness fitting on large scale computing platforms. There are many ways to define the misfit functions, for instance, l_1 norm, dynamic warping, correlation and adaptive Wiener filter et al. However, the most widely used misfit function is the summing squares of the wavefield amplitude subtraction, also known as least-square optimization. Taking a two-dimension problem as example, the objective function can be formulated as

$$J(m) = \frac{1}{2}\sum_{x_s}\sum_{x_r}\int_0^T \left\|d_{obs}(t,x_s,x_r) - d_{syn}(t,x_s,x_r,m)\right\|_2^2 dt, \tag{1}$$

where $d_{obs}(t,x_s,x_r)$ and $d_{syn}(t,x_s,x_r,m)$ denote the observed and synthetic data (pressure filed) by subsurface parameter m, in which x_s, x_r and t are the indexes of source location, receiver location and recording time respectively. Here $\|\bullet\|_2^2$ is the square of the l_2 norm. In Eq. (1), the data residual integral is conducted over time for each source-receiver pair. The cost for the misfit function is a full simulation of all shots, data residual summation, with the amount of calculation proportional to shooting number.

The objective of FWI is to find an optimal model m^* which gets the minimum value of the misfit function $J(m^*) = \min J(m)$. Mathematically, it means the gradient of the objective function with respect to model is zero

$$\frac{\partial J(m)}{\partial m} = 0. \tag{2}$$

However, Eq. (2) is not equivalent to the minimization process of the misfit function, since the wave evolution is nonlinear with respect to the subsurface medium parameters. It makes the objective function develop multiple minima. The local differential Eq. (2) cannot guarantee a solution to the global minima. In fact, for the real cases, plenty of factors can make the FWI easily converge to one of the local minima, such as the lack of low frequencies, the presence of noise, and the approximate modeling of the wave propagation in real media. It is still a great challenge for the seismic exploration community to develop a FWI algorithm with global convergence from poor starting model. In the following, we test our method with a smoothed true velocity as the initial model m_0.

A first order Taylor series expansion of the objective function gradient in the vicinity of m_0 gives the following expression

$$\frac{\partial J(m)}{\partial m} \approx -\frac{\partial J(m_0)}{\partial m} + \frac{\partial^2 J(m_0)}{\partial m^2}(m - m_0). \tag{3}$$

Combine Eqs. (2) and (3), we get

$$\Delta m = -\left[\frac{\partial^2 J(m_0)}{\partial m^2}\right]^{-1} \frac{\partial J(m_0)}{\partial m}. \tag{4}$$

with $\Delta m = m - m^*$. The first and second order derivative of the misfit function with respect to model are named as Fréchet derivative and Hessian respectively.

Inserting Eq. (1) into the gradient of the misfit function, we get

$$\frac{\partial J(m_0)}{\partial m} = -\sum_{x_s}\sum_{x_r}\int_0^T (d_{obs}(t,x_s,x_r) - d_{syn}(t,x_s,x_r,m)) \frac{\partial d_{syn}(t,x_s,x_r,m)}{\partial m} dt. \tag{5}$$

And the Hessian matrix is expressed as

$$\begin{aligned}\frac{\partial^2 J(m_0)}{\partial m^2} = &\sum_{x_s}\sum_{x_r}\int_0^T \left[\frac{\partial d_{syn}(t,x_s,x_r,m)}{\partial m}\right]^T \frac{\partial d_{syn}(t,x_s,x_r,m)}{\partial m} dt \\ &+ \sum_{x_s}\sum_{x_r}\int_0^T \frac{\partial^2 d_{syn}(t,x_s,x_r,m)}{\partial^2 m}(d_{obs}(t,x_s,x_r) - d_{syn}(t,x_s,x_r,m)) dt.\end{aligned} \tag{6}$$

Eq. (4) uses both the Fréchet and Hessian for the model perturbation calculation, and methods used in this form are called Newton methods. The newton methods not only make use of the objective gradient information but also the second derivative (Hessian). It is locally quadratically convergent, in theory, with the fastest convergence rate. The Hessian can be separated into two standalone parts, as shown in Eq. (6). The first part is the product of the synthetic data derivative with the model and can be considered as the contribution of the single scattering, that is, Born approximation. It provides proper weighting to focus the energy similar to the illumination compensation. The second part of the Hessian in Eq. (6), including second order derivative of the synthetic data with respect to the model, takes the multi scattering in the wavefields into account for the model updating. However, considering the high dimensionality of the model, the cost for evaluating the Hessian (mainly the second term) is prohibitive. And for high nonlinear objective function of the complex model and approximate synthetic data modeling for the real media etc., the complete Hessian has limited contribution to the inversion convergence and inversion quantity. In all practical application, people tend to approximate the Hessian with the balance between cost and benefits. For the typical cases, substituting the inverse Hessian by a scalar α leads to the steepest-descent method and only keeping the first term of the Hessian is the Gauss-Newton method.

Over the last decade, the most popular local optimization algorithm for solving FWI problems is based on the conjugate gradient method