Computers and Geosciences 114 (2018) 11–21

Contents lists available at ScienceDirect

Computers and Geosciences

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

Research paper

Incoherent dictionary learning for reducing crosstalk noise in least-squares reverse time migration

Juan Wu, Min Bai *

School of Resources and Environment, North China University of Water Resources and Electric Power, Zhengzhou 450045, China

ARTICLE INFO

Keywords:
Seismic imaging
Computational seismology
Regularization
Reverse time migration
LSRTM
Simultaneous source

ABSTRACT

We propose to apply a novel incoherent dictionary learning (IDL) algorithm for regularizing the least-squares inversion in seismic imaging. The IDL is proposed to overcome the drawback of traditional dictionary learning algorithm in losing partial texture information. Firstly, the noisy image is divided into overlapped image patches, and some random patches are extracted for dictionary learning. Then, we apply the IDL technology to minimize the coherency between atoms during dictionary learning. Finally, the sparse representation problem is solved by a sparse coding algorithm, and image is restored by those sparse coefficients. By reducing the correlation among atoms, it is possible to preserve most of the small-scale features in the image while removing much of the long-wavelength noise. The application of the IDL method to regularization of seismic images from least-squares reverse time migration shows successful performance.

1. Introduction

Seismic migration techniques play an important role in exploration for hydrocarbons (Lindstrom et al., 2016; Zhang et al., 2016b, 2016c; Chang et al., 2016; Ren and Tian, 2016; Wu et al., 2016b; Liu et al., 2017a; Fabien-Ouellet et al., 2017; Rastogi et al., 2017; Li et al., 2017; Chen et al., 2017a, 2017b; Bucha, 2017; Huang et al., 2017a; Shabani and Vilcáez, 2018; Xu et al., 2018; Wang et al., 2018). However, conventional migration operator is the adjoint of the forward modelling operator, rather than the exact inverse of the forward modelling operator. This approximation suffers from migration artifacts, which are caused by bandwidth, under-sampled acquisition geometry, limited recording aperture, etc. These artifacts can be mitigated by taking the inverse Hessian matrix into account. However, directly calculating the explicit Hessian matrix requires huge memory storages which prevents the usage in practice. The indirect way to account for the effects of the inverse Hessian matrix is by using a migration followed by applying an approximation of the inverse Hessian matrix or through an inversion-based iterative algorithm (Jiao et al., 2015; Zu et al., 2016a; Lines et al., 2016; Xie et al., 2018), which is also known as least-squares migration (LSM) (Nemeth et al., 1999; Xue et al., 2016c). Using the Born approximation, LSM is equivalent to a linearized inversion, where the reflectivity model is updated at every iteration.

In LSRTM, we assume that the background velocity is smoothly varying, e.g., from NMO-based velocity analysis (Ebrahimi et al., 2017), and use Born modelling to predict the primary reflected waves. However, when the velocity field has sharp contrasts, such as the velocity at the boundary of salt, Born approximation, which assumes that the perturbation of velocity model is small, is violated. The forward- and back-propagated wavefields travelling through the background velocity model will generate perceptible backscattered energy. In this case, cross-correlation of these two wavefields will generate some noise in the gradient. This issue in conventional RTM is well addressed (Guitton et al., 2006). However, the standard gradient formula used in LSRTM has some differences with the cross-correlation imaging condition (Chattopadhyay and McMechan, 2008) normally used in conventional RTM. Although some methods, such as the high-pass spatial filter method (Plessix and Mulder, 2004), Poynting vectors method (Yoon and Marfurt, 2006) and wavefield decomposition method (Liu et al., 2011), can be directly used in LSRTM, other methods, such as Laplacian operator (Youn and Zhou, 2001), inverse scattering imaging condition (Stolk et al., 2009), and energy imaging condition (Rocha et al., 2016), still need further extension to the LSRTM or need to construct the connection with the standard gradient formula. Even using the inversion-based LSM method, seismic images are still usually corrupted by various types of noise during the collection and transmission processes (Rashed and Rashed, 2017). The noise greatly reduces the reliability and effectiveness of image processing, such as feature extraction, target detection and recognition. Hence,

* Corresponding author.
E-mail addresses: wujuan@ncwu.edu.cn, wujuan_2016@126.com (J. Wu), baimin2016@126.com (M. Bai).

https://doi.org/10.1016/j.cageo.2018.01.010
Received 5 September 2017; Received in revised form 14 January 2018; Accepted 17 January 2018
Available online 5 February 2018

the noise must be eliminated to improve the quality of image. In this context, noise elimination has been a major research topic in seismic imaging and tomography (Chen et al., 2017g).

Currently, denoising algorithms can be broadly classified into three categories: the denoising method based on spatial-domain filtering (Chen and Ma, 2014; Huang et al., 2016, 2017e, 2017f, 2017c; Zhou et al., 2017b), the denoising method based on transform-domain filtering (Kong and Peng, 2015; Gan et al., 2015; Zhou et al., 2018) and the denoising method based on learning (Chen, 2017; Siahsar et al., 2017a, 2017b, 2017c; Zhou et al., 2017a). Examples of those methods based on spatial-domain filtering include the Gaussian filtering, empirical mode decomposition (EMD) (Gan et al., 2016a; Chen et al., 2017d), ensemble empirical mode decomposition (EEMD) (Chen et al., 2017f), improved complete ensemble empirical mode decomposition (ICEEMD) (Chen et al., 2016a, 2017c), variational mode decomposition (VMD) (Liu et al., 2016c, 2017b, 2017c, 2018), bilateral filtering, guided filtering (Mortazavi et al., 2017), median filtering (Gan et al., 2016d; Chen et al., 2017e; Huang et al., 2018), morphological filtering (Li et al., 2016a, b), principal component analysis (PCA) filtering (Ikelle, 2016; Naghadeh and Morley, 2016; Xie et al., 2017), and the non-local means filtering methods (Buades et al., 2005; Mairal et al., 2009; Yang et al., 2015). The basic idea of these algorithms is to eliminate the noise using local or non-local self-similarity of the image. The first category of algorithms is computationally efficient, but the denoised image is usually too smooth. Examples of those methods based on transform-domain filtering include the Fourier transform (Zhong et al., 2016; Shen et al., 2016; Li et al., 2016c), Wavelet transform and BM3D methods (Dabov et al., 2007; Burger et al., 2012), Curvelet transform (Liu et al., 2016e), Shearlet transform (Liu et al., 2016a; Kong et al., 2016), Synchrosqueezing Transform (Liu et al., 2016d, 2016g), Radon transform (Xue et al., 2016b, 2017; Sun and Wang, 2016), Seislet transform (Gan et al., 2016b, 2016c; Liu et al., 2016f; Wu et al., 2016a), and EMD-seislet transform (Chen and Fomel, 2018). The basic idea of these methods is to eliminate the noise via thresholding and it also exploits the observation that the transforms lead to different energy distributions of noise coefficients and image coefficients. In the case of BM3D algorithm, block matching is performed on the image to convert the 2-D image blocks with similar structures into 3D data through 3-D transform; then, it is subjected to Wiener filtering. Rank-reduction based methods can also be implemented in a transformed domain (Zhang et al., 2016a, 2017; Wang et al., 2017), e.g., the randomized-order multichannel singular spectrum analysis (Huang et al., 2017d), damped multichannel singular spectrum analysis (Chen et al., 2016b, 2016c), the double least squares projections method (Huang et al., 2017b), empirical low-rank approximation method (Chen et al., 2017h). The learning-based denoising algorithm includes K-SVD (Elad and Aharon, 2006; Aharon et al., 2006; Romano and Elad, 2013), LSSC (Mairal et al., 2009) and CSR (Dong et al., 2013; Romano et al., 2014; Li et al., 2011; Lu et al., 2013). The basic idea of this type of algorithm is to eliminate the noise using local sparsity of the image. In K-SVD, the mutually overlapping small image blocks are learned to yield the self-adaptive redundant dictionary, which is then used to obtain the sparse representation of image blocks, thereby achieving noise elimination. Chen (2017) addressed the computational efficiency problem in K-SVD, and proposed a fast dictionary learning approach based on the sequential generalized K-means (SGK) algorithm for denoising multidimensional seismic data. The SGK algorithm updates each dictionary atom by taking an arithmetic average of several training signals instead of calculating an SVD as used in K-SVD algorithm. More advanced denoising methods are being developed to best preserve the useful signals while maintaining the noise removal ability, e.g., the signal-and-noise orthogonalization method (Chen and Fomel, 2015), waveform-shaping method (Chen and Jin, 2015), and time-frequency-peak filtering method (Zhang et al., 2015).

It has been reported that if the non-correlation degree between dictionary atoms is increased, the redundant dictionary obtained via learning can fully describe the information of image texture (Lin et al., 2012; Abolghasemi et al., 2015; Liu et al., 2016b). Motivated by this observation and in order to address problems of existing algorithms, we propose a novel image denoising algorithm based on non-correlated dictionary learning. The basic idea of our proposed algorithm is to reduce correlation between dictionary atoms using the non-correlated dictionary learning technology and improve the ability of redundant dictionary to represent information of image texture. By reducing the correlation among atoms, it is possible to preserve most of the small-scale features in the image while removing much of the long-wavelength noise. In this way, we can eliminate the noise while maintaining image details. First, the input noise-corrupted image is divided into mutually overlapping image blocks, and some image blocks are chosen randomly as the sample. Next, a non-correlated dictionary learning algorithm is proposed to obtain a redundant dictionary with slightly correlated atoms. We provide a brief introduction of the regularization least-squares reverse time migration (LSRTM) and introduce in detail the incoherent dictionary learning algorithm and its application to LSRTM. We apply the incoherent dictionary learning algorithm to reduce migration crosstalk caused from the simultaneous-source acquisition (Berkhout, 2008; Chen et al., 2014; Qu et al., 2014, 2015, 2016; Xue et al., 2016a; Zu et al., 2016b, 2017c; Zhou, 2017; Chen, 2015). In this paper, we use two numerical examples to show the successful performance of the presented algorithm.

2. Method

2.1. Incoherent dictionary learning

Consider a set of samples $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K] \in \mathbf{R}^{n \times K}$. The aim of dictionary learning is to obtain redundant dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_m] \in \mathbf{R}^{n \times m}$ via learning, so that each sample $y_k (k = 1, \cdots, K)$ can be represented with a sparse vector $\mathbf{x}_k (k = 1, \cdots, K)$. The dictionary learning problem can be formulated as

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2, \\ &\text{s.t.} \|\mathbf{d}_i\|_2 = 1, \|\mathbf{x}_k\|_0 \le T_0, \forall i, k. \end{aligned} \tag{1}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K) \in \mathbf{R}^{m \times K}$ is the coefficient matrix, T_0 is the level of sparsity. $[\cdot]_0$ denotes the L_0-norm of an input vector. The problem in equation (1) can be solved using the K-SVD (Aharon et al., 2006) or online dictionary learning (Lu et al., 2013) algorithms.

Next, we will describe the incoherent dictionary learning (or non-correlated dictionary learning) algorithm (Lin et al., 2012; Abolghasemi et al., 2015; Liu et al., 2016b). The degree of correlation between atoms in the redundant dictionary is an import metric of the dictionary's representation ability. The lower the degree of correlation between atoms, the greater the dictionary's representation ability. Existing models for dictionary learning cannot ensure slight correlation between atoms in the obtained redundant dictionary, and this affects performance of the redundant dictionary. In this paper, we intend to introduce the non-correlation constraint to the dictionary leanring model to guarantee slight correlation between dictionary atoms.

A non-correlated dictionary learning model will be constructed in this sub-section. The degree of correlation between dictionary atoms is defined as (Lin et al., 2012; Abolghasemi et al., 2015):

$$R(\mathbf{d}_i, \mathbf{d}_j) = \frac{|<\mathbf{d}_i, \mathbf{d}_j>|}{\|\mathbf{d}_i\|_2 \|\mathbf{d}_j\|_2} = \frac{\mathbf{d}_i^T \mathbf{d}_j}{\|\mathbf{d}_i\|_2 \|\mathbf{d}_j\|_2}. \tag{2}$$

From the definition above, it can be seen that $\mathbf{R}(\mathbf{d}_i, \mathbf{d}_j) \in [0, 1]$. If $\mathbf{d}_i$ and $\mathbf{d}_j$ are orthogonal. $\mathbf{R}(\mathbf{d}_i, \mathbf{d}_j) = 0$. If $\mathbf{d}_i = \rho \mathbf{d}_j$ (ρ is a non-zero constant), $\mathbf{R}(\mathbf{d}_i, \mathbf{d}_j) = 1$. Based on the definition in equation (2), the correlation degree of the dictionary $\mathbf{D}$ can be defined as (Lin et al., 2012; Abolghasemi et al., 2015)