SPARSE DICTIONARY LEARNING FOR SEISMIC NOISE ATTENUATION USING A FAST ORTHOGONAL MATCHING PURSUIT ALGORITHM

YATONG ZHOU[1], SHUHUA LI[1], JIANYONG XIE[2], DONG ZHANG[2] and YANGKANG CHEN[3*]

[1] *School of Electronic and Information Engineering, Hebei University of Technology, Xiping Road 5340, Beichen District, Tianjin 300401, P.R. China. zyt_htu@126.com*
[2] *State Key Laboratory of Petroleum Resources and Prospecting University of Petroleum-Beijing, Fuxue Road 18, Beijing 102249, P.R. China. xjyshl@sina.com; zhangdongconan@163.com*
[3] *Bureau of Economic Geology, John A. and Katherine G. Jackson School of Geosciences, The University of Texas at Austin, University Station, Box X, Austin, TX 78713-8924, U.S.A. ykchen@utexas.edu*
[*] *Present address: National Center for Computational Sciences, Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN 37831-6008, U.S.A.*

(Received November 9, 2016; revised version accepted July 26, 2017)

ABSTRACT

Zhou, Y., Li, S., Xie, J., Zhang, D. and Chen, Y., 2017. Sparse dictionary learning for seismic noise attenuation using a fast orthogonal matching pursuit algorithm. *Journal of Seismic Exploration*, 26: 433-454.

Attenuation of random noise is a long-standing problem in seismic data processing. One of the most widely used approaches is based on sparse transforms. In the geophysics community, most of the currently used sparse transforms have fixed bases, which we call analytical transforms. In this paper, we seek a different type of sparse transform, with variant bases, to attenuate random noise. We call this type of transform dictionary learning-based (DLB) sparse transforms, because it can adaptively train a sparse dictionary from the observed data to adapt to different seismic data. To increase the efficiency of sparse dictionary learning, we propose to apply a fast orthogonal matching pursuit (OMP) algorithm for sparse coding. We use both synthetic and field data examples to show the superior performance of the dictionary learning-based transform over fixed-basis transforms, and much improved efficiency in sparse coding associated with the fast OMP algorithm, which is one of the two steps in the DLB transform.

KEY WORDS: sparse dictionary learning, denoising, fast OMP algorithm.

0963-0651/17/$5.00 © 2017 Geophysical Press Ltd.

INTRODUCTION

Random noise attenuation plays an indispensable role in seismic data processing. The useful signal that is mixed with the ambient random noise is often neglected and thus may cause confusion between seismic events and artifacts in the final migrated image. Enhancing the useful signal while attenuating random noise can help reduce interpretation difficulties and risks for oil & gas detection (Yang et al., 2014, 2015; Li et al., 2016a,b; Gan et al., 2016e; Chen and Jin, 2015).

The widely used frequency-space prediction filtering (Canales, 1984a) can achieve good results for linear events but may fail in handling complex or hyperbolic events. A mean or median filter (Liu et al., 2009b; Chen et al., 2015; Chen, 2015a; Gan et al., 2016d) is often used to attenuate specific types of random noise, e.g., a mean filter is effective in attenuating Gaussian white noise, and a median filter can remove random spikes with excellent performance. An eigenimage based approach (Bekara and van der Baan, 2007), sometimes referred to as global singular value decomposition (SVD), is effective for horizontal-events in seismic profiles, but cannot be adapted to geologically complicated structures. An enhanced version of this method turns global SVD to local SVD (Bekara and van der Baan, 2007), where a dip steering process is performed in each local processing window to enhance the locally coherent events. The problem with local SVD is that only one slope component for each processing window is allowed, and also the optimal size of each processing window is often difficult to select. Structure-oriented SVD is designed specifically for seismic data by applying the SVD filtering along the morphological structure direction of seismic data (Gan et al., 2015a). Matrix completion via f-x domain multichannel singular spectrum analysis (MSSA) can handle complex dipping events well by extracting the first several eigen-components after SVD for each frequency slice (Huang et al., 2016a,b,c; Chen et al., 2016b,c; Xue et al., 2016a; Zhang et al., 2016a,b; Huang et al., 2017). The f-x MSSA approach is based on a pre-defined rank of the seismic data. The rank here denotes the number of linear components in the seismic data. However, for complex seismic data, the rank is hard to select, and for curved events, the rank tends to be high and thus will involve a serious rank-mixing problem. Chen and Fomel (2015b) proposed a two-step processing strategy to guarantee no coherent signal is lost in the removed noise.

Another common denoising approach is the three-step sparsity-promoting transform based method (Liu et al., 2016a; Kong et al., 2016; Liu et al., 2016e,c). The data are forward-transformed from time-space domain to the transformed domain, and then a thresholding operator is applied in the transform domain, followed by an inverse transform of the data back to the time-space domain. Because of its superb performance and convenient implementation, it has been one of the most popular methods (Wu et al., 2016; Zhong et al., 2016;

Liu et al., 2016a; Kong et al., 2016). Sparsity-promoting transforms can be generally divided into two categories (Chen et al., 2016a): a fixed basis approach or a learning-based approach. A number of fixed basis sparsity-promoting transforms are proposed in the literature for processing seismic data including the Fourier transform (Duijndam et al., 1999; Naghizadeh, 2012), the Radon transform (Yu et al., 2007; Wang et al., 2010; Xue et al., 2016b, 2017), the curvelet transform (Shahidi et al., 2013; Zu et al., 2016a,b; Liu et al., 2016d) and the seislet transform (Fomel and Liu, 2010; Chen et al., 2014a; Chen, 2015b; Gan et al., 2015b, 2016c,a; Chen, 2016). Wang et al. (2008) used the second- generation wavelet transform, which is based on the lifting scheme, to denoise seismic data with a percentile thresholding strategy. Hennenfent and Herrmann (2006) and Neelamani et al. (2008) applied the curvelet transform to attenuate both random and coherent noise in seismic data. Fomel and Liu (2010) designed a sparse seislet transform that is tailored specifically for seismic data, including seismic denoising. Chen and Fomel (2015a) used the adaptive separation properties of empirical mode decomposition (EMD) (Huang et al., 1998; Chen et al., 2014b; Gan et al., 2016b; Liu et al., 2016b,e) for preparing the stable input for the non-stationary 1D seislet transform and proposed a new EMD-seislet transform to denoise seismic data with strong spatial heterogeneity. Recently, Kong and Peng (2015) applied the shearlet transform to seismic random noise attenuation.

The learning-based approach utilizes machine learning techniques to infer a dictionary (Chen, 2017). Instead of fixing the basis for the transform, the basis is adaptively learned from the observed data. Thus, it can adapt to different complicated seismic data. There are some initial results regarding the dictionary-learning-based (DLB) transforms in the geophysics community but, as it is a relatively new concept, these methods have not been widely tested and investigated (Chen et al., 2016a). In this paper, we first introduce the mathematics related to sparse dictionary learning, then apply the DLB approach to random noise attenuation in seismic data. We demonstrate that the DLB denoising approach can obtain much better performance than fixed-basis wavelet and curvelet transforms. Considering that the DLB approaches are more computationally expensive, we introduce a fast orthogonal matching pursuit (OMP) algorithm to accelerate the sparse coding to reduce the computational cost. Here, sparse coding refers to one of the key steps in DLB transform; the other is dictionary updating. The numerical tests show that the fast OMP algorithm can indeed significantly accelerate the DLB process.

THEORY

Sparse dictionary learning

Sparse representation via learning based dictionary consists of two main steps, namely sparse coding and dictionary updating.

- Sparse coding. Given the observed data vector **b**, which can denote a 1D seismic signal (e.g., a trace), sparse coding aims at solving the optimization problem:

$$\mathbf{x}^n = \arg\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}^n\mathbf{x}\|_2^2, \text{ s.t. } \|\mathbf{x}\|_0 \leq L \ , \tag{1}$$

where $\|\cdot\|_2$ and $\|\cdot\|_0$ denote the L_2 and L_0 norms of an input vector, respectively. $\mathbf{x}^n$ denotes the sparse coefficients after nth iterations. L is the number of non-zero coefficients in **x**. **A** is the learned dictionary, with each column in **A** denoting a basis and **x** is the sparse representation of **b** in the sparse transformed domain of **A**. We will provide some illustrations in the EXAMPLES section.

- Dictionary updating. For the obtained $\mathbf{x}^n$, update $\mathbf{A}^n$ such that

$$\mathbf{A}^{n+1} = \arg\min_{\mathbf{A}} \|\mathbf{b} - \mathbf{A}^n\mathbf{x}\|_2^2 \ . \tag{2}$$

Eqs. (1) and (2) are iterated N*iter* times to learn the optimal dictionary and the sparsest representation. The iterations terminate when N*iter* is reached or when convergence of the sparse dictionary is obtained, i.e., when further iterations do not produce any significant change.

The multidimensional seismic data **D** is first reformulated into patch form **B**. Each column vector in **B** is extracted from the multidimensional seismic data matrix. For example, a 3×3 window in a 2D seismic data can be extracted and reshaped as a 9×1 vector, which is stored as a column in **B**. Eqs. (1) and (2) then become

$$\forall \mathbf{x}_i^n = \arg\min_{\mathbf{x}_i} \|\mathbf{B} - \mathbf{A}^n\mathbf{X}\|_F^2, \text{ s.t. } \|\mathbf{x}_i\|_0 \leq L \ , \tag{3}$$

$$\mathbf{A}^{n+1} = \arg\min_{\mathbf{A}} \|\mathbf{B} - \mathbf{A}\mathbf{X}^n\|_F^2 \ , \tag{4}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of an input matrix, $\mathbf{x}_i$ denotes the i-th column in **X**, or i-th sparse coefficient vector corresponding to the i-th column in the data patch **B**.

Dictionary updating by K-SVD

We will first introduce the K-SVD method used for solving the dictionary updating eq. (4). The dictionary update is performed one atom at a time (Aharon et al., 2006). The objective function is minimized for each atom individually while keep the other atoms fixed. Atom here denotes each column in the dictionary matrix **A**. To achieve this, the update step uses only signals in **B** whose sparse representations use the current atom. Letting J denote the indices of the signals in **B** which uses the current atom. The update is obtained by minimizing the following objective function

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \| \mathbf{B}_J - \mathbf{A}\mathbf{X}_J \|_F^2 , \tag{5}$$

over both the atom and its associated coefficient row in $\mathbf{X}_J$. The resulting problem is a simple rank-1 approximation problem expressed as

$$\{a,c\} = \arg\min_{\mathbf{a},\mathbf{c}} \| \mathbf{E} - \mathbf{a}\mathbf{c}^T \|_F^2, \text{ s.t. } \| \mathbf{a} \|_2 = 1 , \tag{6}$$

where $\mathbf{E} = \mathbf{B}_J - \Sigma_{i \neq j}\mathbf{a}_i\mathbf{X}_{i,J}$ is the error matrix without the current atom (j), **a** is the updated atom (a column in **A**), and $\mathbf{x}^T$ is the new coefficients row in $\mathbf{X}_i$. s.t. denotes subject to. **c** denotes the coefficient row corresponding to **a** (or i-th row in $\mathbf{X}_J$). $\mathbf{X}_{i,J}$ simply denotes the i-th row in **X** but truncated by the indices vector **J**. The problem can be solved directly via an SVD decomposition, or other more efficient numerical algorithms. To make it clearer, when updating the i-th column in dictionary **A**, after SVD the i-th row in **X** will also be modified. To retain the sparse property of **X**, we need to restrict the modification to those coefficients in the i-th row (of **X**) which are not zero. Those indices where entries are not zero are denoted by **J**.

The algorithm workflow can be shown as

K-SVD ALGORITHM($\mathbf{B}$,$\mathbf{A}_0$,L,N*iter*)

$\mathbf{A} \leftarrow \mathbf{A}_0$

for n $\leftarrow$ 1,2,…,N*iter*

  do

    $\forall i{:}\mathbf{X}_i = \arg\min_{\mathbf{x}} \| \mathbf{d}_i - \mathbf{A}\mathbf{x} \|_2^2, \text{ s.t. } \| \mathbf{x} \|_0 \leq L$

    for j $\leftarrow$ 1,2,…,K

      do

        $\mathbf{A}_j \leftarrow 0$

        $\mathbf{E} \leftarrow \mathbf{B}_I - \mathbf{A}\mathbf{X}_I$

        $\{\mathbf{a},\mathbf{c}\} = \arg\min_{\mathbf{a},\mathbf{c}} \| \mathbf{E} - \mathbf{a}\mathbf{c}^T \|_F^2, \text{ s.t. } \| \mathbf{a} \|_2 = 1$

$\mathbf{A}_j \leftarrow \mathbf{a}$

$\mathbf{X}_{j,I} = \mathbf{c}^T$

return A

return A,X

In the algorithm above, K denotes the number of columns (atoms) in **A** and the number of SVD calculations when updating A, it explains "K" in the so-called "K-SVD" algorithm.

Sparse coding by OMP

The problem as expressed in (1) is an NP-hard problem, and directly finding the truly optimal **X** is impossible and is usually solved by an approximation pursuit method, such as the orthogonal matching pursuit (OMP) algorithm. The greedy OMP algorithm selects, at each step, the atom with the highest correlation to the current residual. Once the atoms are selected, the signal is orthogonally projected to the span of the selected atoms. Then, the residual and the process are repeated. The output of the sparse coding procedure is the sparse coefficient vector, as shown in eq. (1), or each column in the sparse coefficient matrix, as shown in eq. (3). The algorithm workflow of OMP is as follows

ORTHOGONAL MATCHING PURSUIT (**A**,**b**,L)

Set $I \leftarrow ()$

$\mathbf{r} \leftarrow \mathbf{b}$

$\mathbf{x} \leftarrow \mathbf{0}$

for $it \leftarrow 1,2,\ldots,L$

do

$\hat{k} \leftarrow \arg\max_k |\mathbf{a}_k^T \mathbf{r}|$

$I \leftarrow (I,\hat{k})$

$\mathbf{x}_I \leftarrow (\mathbf{A}_I)^{\dagger}\mathbf{b}$

$\mathbf{r} \leftarrow \mathbf{b} - \mathbf{A}_I\mathbf{x}_I$

return x

where $(\mathbf{A}_I)^{\dagger} = (\mathbf{A}_I^T\mathbf{A}_I)^{-1}\mathbf{A}_I^T$. In the above algorithm, I denotes the vector of indices corresponding to the non-zero entries in vector $\mathbf{x}$. $\mathbf{a}_k^T$ denotes the transpose of the k-th column in dictionary $\mathbf{A}$.

Sparse coding by fast OMP

In the greedy OMP algorithm, the computation $\mathbf{x}_I = (\mathbf{A}_I)^{\dagger}\mathbf{b}$ requires the inversion of matrix $\mathbf{A}_I^T\mathbf{A}_I$, which remains non-singular due to the orthogonalization process which ensures the selection of linearly independent atoms. The matrix $\mathbf{A}_I^T\mathbf{A}_I$ is a symmetric positive-definite matrix and is updated every iteration by simply appending a single row or column to it, and therefore its Cholesky factorization requires only the computation of its last row (Rubinstein et al., 2008).

It can be proven that if $\mathbf{A}$ and $\tilde{\mathbf{A}}$ have the following relation:

$$\mathbf{A} = \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{v} \\ \mathbf{v}^T & c \end{pmatrix}, \tag{7}$$

where c is a constant, $\mathbf{v}$ is an arbitrary column vector, then the Choleskey factorization of $\mathbf{A}$ can be expressed as

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T , \tag{8}$$

$$\mathbf{L} = \begin{pmatrix} \tilde{\mathbf{L}} & \mathbf{0} \\ \mathbf{w}^T & \sqrt{(c - \mathbf{w}^T\mathbf{w})} \end{pmatrix}, \tag{9}$$

$$\mathbf{w} = \tilde{\mathbf{L}}^{-1}\mathbf{v} , \tag{10}$$

where $\tilde{\mathbf{L}}$ is the triangular matrix from the Choleskey factorization of $\tilde{\mathbf{A}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T$. This method for inverting the matrix $\mathbf{A}^T\mathbf{A}$ is called the progressive Cholesky factorization method.

When large numbers of signals must be coded over the same dictionary, it is worthwhile to consider pre-computation of the Choleskey factorization to reduce the total amount of work involved in coding the entire set. It is obvious that the atom selection step at each iteration does not require knowing $\mathbf{r}$ and $\mathbf{x}$

explicitly, but only $\mathbf{A}^{\mathrm{T}}\mathbf{r}$. So we can reduce the computational cost by replacing the explicit computation of $\mathbf{r}$ and its multiplication by $\mathbf{A}^{\mathrm{T}}$ with a lower-cost computation of $\mathbf{A}^{\mathrm{T}}\mathbf{r}$.

The $\mathbf{r}$ can be removed from the equations by simple derivation as follows:

$$\begin{aligned}\alpha &= \mathbf{A}^{\mathrm{T}}[\mathbf{d} - \mathbf{A}_{\mathrm{I}}(\mathbf{A}_{\mathrm{I}})^{\dagger}\mathbf{d}] \\ &= \alpha^0 - \mathbf{G}_{\mathrm{I}}(\mathbf{A}_{\mathrm{I}})^{\dagger}\mathbf{d} \\ &= \alpha^0 - \mathbf{G}_{\mathrm{I}}(\mathbf{A}_{\mathrm{I}}^{\mathrm{T}}\mathbf{A}_{\mathrm{I}})^{-1}\mathbf{A}_{\mathrm{I}}^{\mathrm{T}}\mathbf{d} \\ &= \alpha^0 - \mathbf{G}_{\mathrm{I}}(\mathbf{G}_{\mathrm{I,I}})^{-1}\alpha_{\mathrm{I}}^0 \ ,\end{aligned} \tag{11}$$

where $\alpha = \mathbf{A}^{\mathrm{T}}\mathbf{r}$, $\alpha^0 = \mathbf{A}^{\mathrm{T}}\mathbf{x}$, and $\mathbf{G} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$. Eq. (11) means that we can compute α each iteration instead of explicitly computing $\mathbf{r}$. The computational cost can be greatly reduced by multiplication with $\mathbf{G}_{\mathrm{I}}$ instead of $\mathbf{A}^{\mathrm{T}}$. The matrix $\mathbf{G}_{\mathrm{I,I}}$ can also be inverted using the progressive Cholesky factorization method. The matrix $\mathbf{G}_{\mathrm{I,I}}$ indicates that the columns and rows of $\mathbf{G}$ are both restricted by the vector $\mathbf{I}$ of indices. The complete fast OMP algorithm is as follows:

FAST ORTHOGONAL MATCHING PURSUIT(α^0,$\mathbf{G}$,L)

Set I $\leftarrow$ ()

$\mathbf{L} \leftarrow [1]$

$\alpha \leftarrow \alpha^0$

$\mathbf{x} \leftarrow \mathbf{0}$

for it $\leftarrow 1,2,\ldots,L$

  do

    $\hat{k} \leftarrow \arg\max_k |\alpha_k|$

    if it > 1

    then

      $\mathbf{w} \leftarrow$ Solution to $\{\mathbf{Lw} = \mathbf{G}_{\mathrm{I},\hat{k}}\}$

$$\mathbf{L} = \leftarrow \begin{pmatrix} \tilde{\mathbf{L}} & \mathbf{0} \\ \mathbf{w}^{\mathrm{T}} & \sqrt{(1-\mathbf{w}^{\mathrm{T}}\mathbf{w})} \end{pmatrix}$$

$\mathrm{I} \leftarrow (\mathrm{I}, \hat{\mathrm{k}})$

$\mathbf{x}_{\mathrm{I}} \leftarrow$ Solution to $\{\mathbf{LL}^{\mathrm{T}}\mathbf{c} = \alpha_{\mathrm{I}}^{0}\}$

$\beta \leftarrow \mathbf{G}_{\mathrm{I}}\mathbf{x}_{\mathrm{I}}$

$\alpha \leftarrow \alpha^{0} - \beta$

return x

The $\mathbf{w}$ in the above algorithm is detailed in eqs. (9) and (10).

EXAMPLES

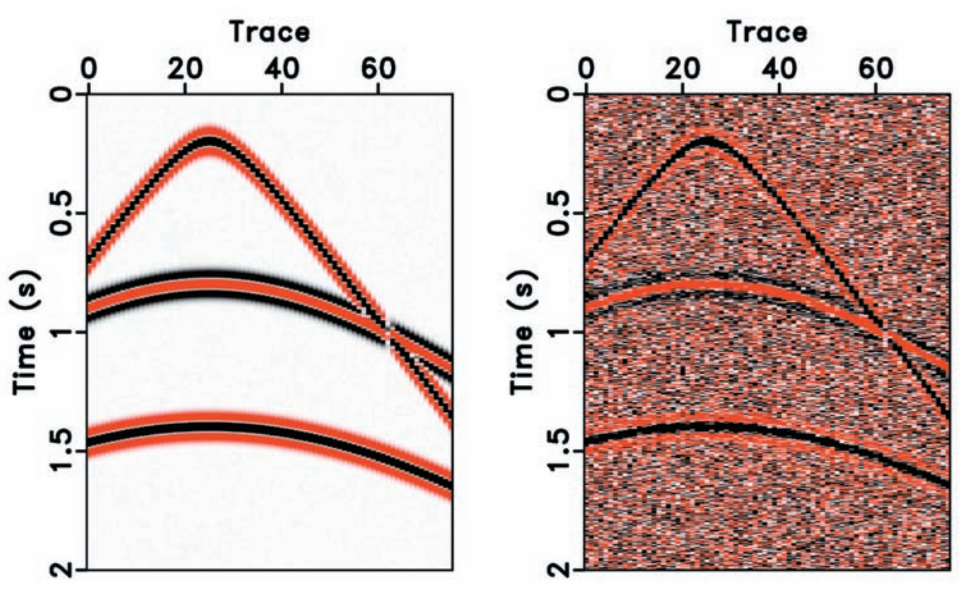
In this section, we will use both synthetic and field data examples to demonstrate the performance of the DLB denoising method. For measuring the denoising performance of synthetic data examples, where one knows the clean data, we use the signal-to-noise ratio (SNR) (Liu et al., 2009a; Huang et al., 2015, 2016a) measurement and the formula is expressed as follows:

$$\mathrm{SNR} = 10\log_{10} \| \mathbf{x}_{\mathrm{true}} \|_2^2 / \| \mathbf{x}_{\mathrm{true}} - \hat{\mathbf{x}} \|^2 \ , \tag{12}$$

where $\mathbf{x}_{\mathrm{true}}$ denotes the clean data and $\hat{\mathbf{x}}$ denotes the denoised data.

The first example (Fig. 1) contains three hyperbolic events, all of which are considered to be useful signals. Because of the high curvature of the first hyperbolic event, and the crossing of first and second events, it is difficult to denoise for many traditional methods. The spatially incoherent components are the random noise, which should be rejected before subsequent seismic data process procedures, e.g., such as migration, velocity analysis, and amplitude-versus-offset (AVO) inversion. The criterion to judge the denoising performance is to maximize the noise removal while minimizing the signal damage. Figs. 2a and 2b show the denoised results using a wavelet transform and a curvelet transform, respectively. It is obvious that both methods cause more or less damage to the useful events, and the curvelet thresholding is generally more effective than wavelet transform in preserving more signals and removing more noise. Figs. 2c and 2d show the denoised results using the learned sparse dictionary via traditional OMP and fast OMP. Note that Figs. 2c and 2d are exactly the same while the computing time for the traditional OMP is 81.74 seconds and for the fast OMP is 33.26 seconds. We show the same results to confirm the correctness of the fast OMP algorithm. Fig. 3 show the removed noise sections corresponding to the four denoised results as shown in Fig. 2. The SNRs of the noisy data, wavelet denoised data, curvelet denoised data, and sparse dictionary denoised data are −5.18 dB, −0.1837 dB, 1.171 dB, and 1.23 dB, respectively, which confirms the best performance using the DLB method. In this example, since the hyperbolic events are deemed to be signals, both the wavelet and curvelet methods tend to damage a large portion of the signals. When the first hyperbolic event is considered to be coherent noise, the performance of both wavelet and curvelet methods is better than the DLB approach. However, the DLB transform can also be used to remove coherent noise, but with special treatment of the dictionaries, as introduced in Kaplan et al. (2009).

To better explain the DLB algorithm, and correlate the numerical experiments with the theory. We plot the key matrices that are introduced above. Fig. 4 shows the dictionary matrices **A** in eqs. (3) and (4) before and after dictionary training. The size of the matrix is 64×256, since we use a 8×8 patch size and 256 dictionary atoms, which means that each 8×8 window is extracted from the seismic data and reformulated into a 1D column. It can be



(a) (b)

Fig. 1. (a) Clean data. (b) Noisy data.

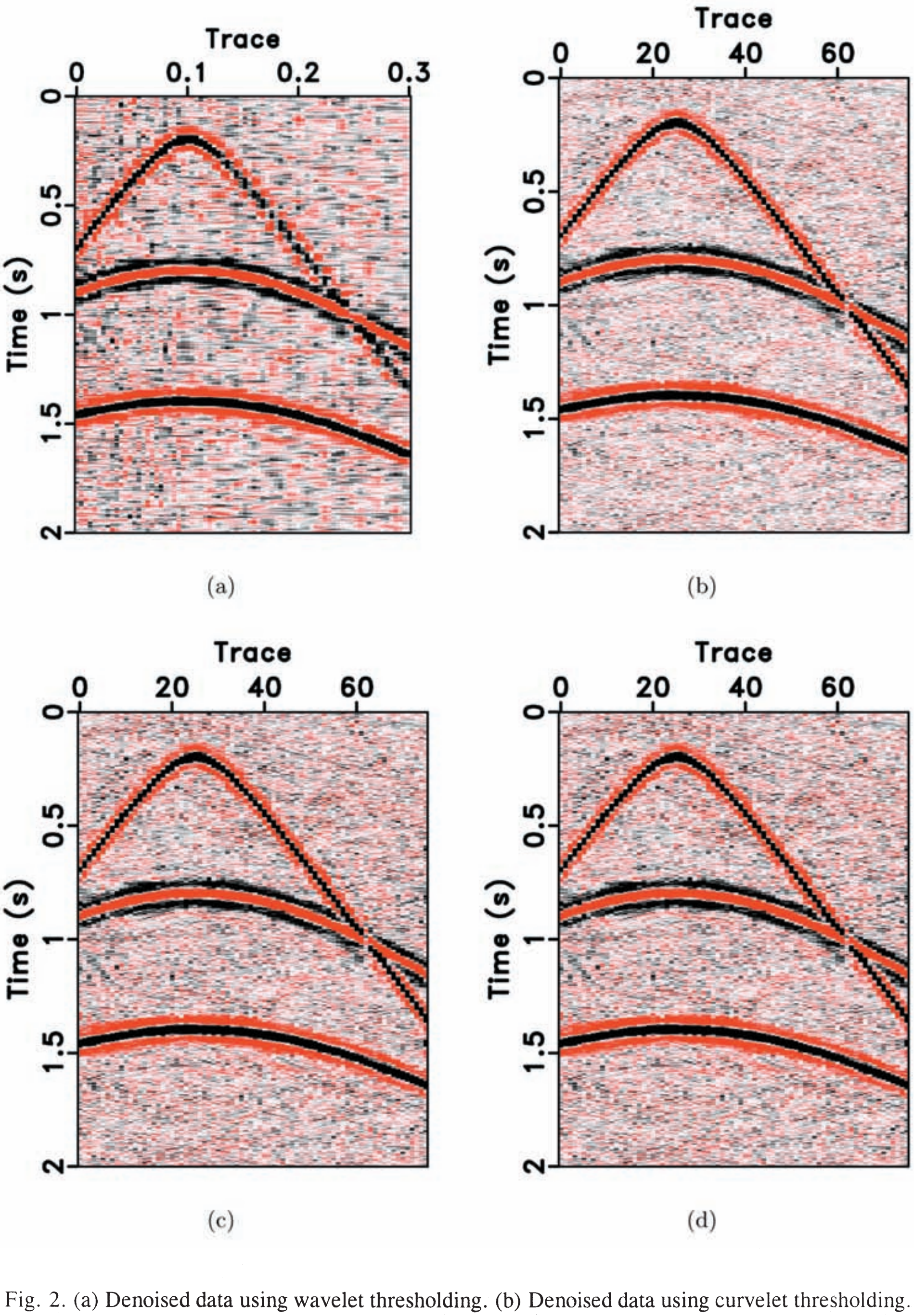


Fig. 2. (a) Denoised data using wavelet thresholding. (b) Denoised data using curvelet thresholding. (c) Denoised data using DLB transform via traditional OMP algorithm. (d) Denoised data using dictionary DLB via fast OMP algorithm.

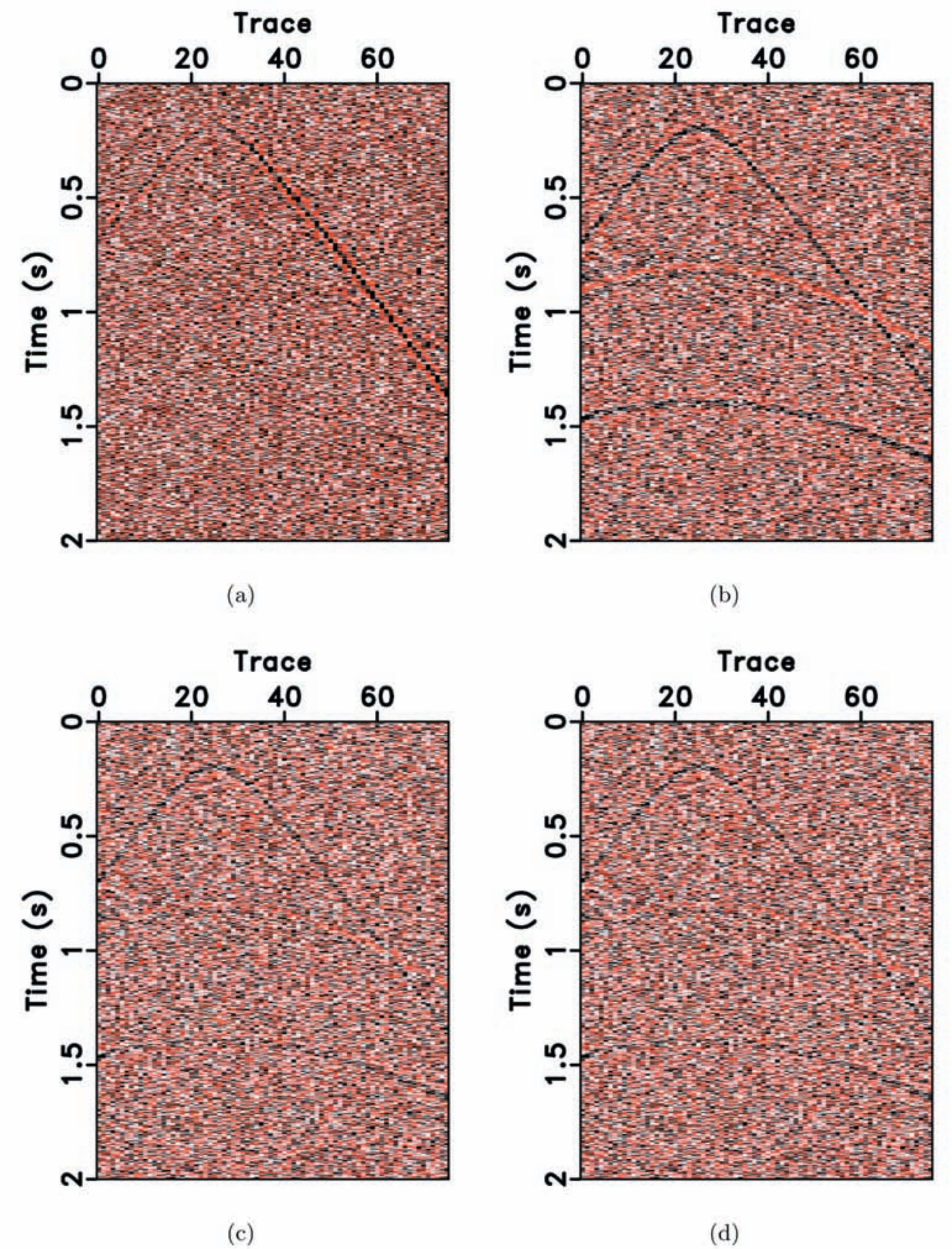


Fig. 3. Removed noise using (a) wavelet thresholding, (b) curvelet thresholding, (c) DLB transform via a traditional OMP algorithm, (d) DLB transform via a fast OMP algorithm.

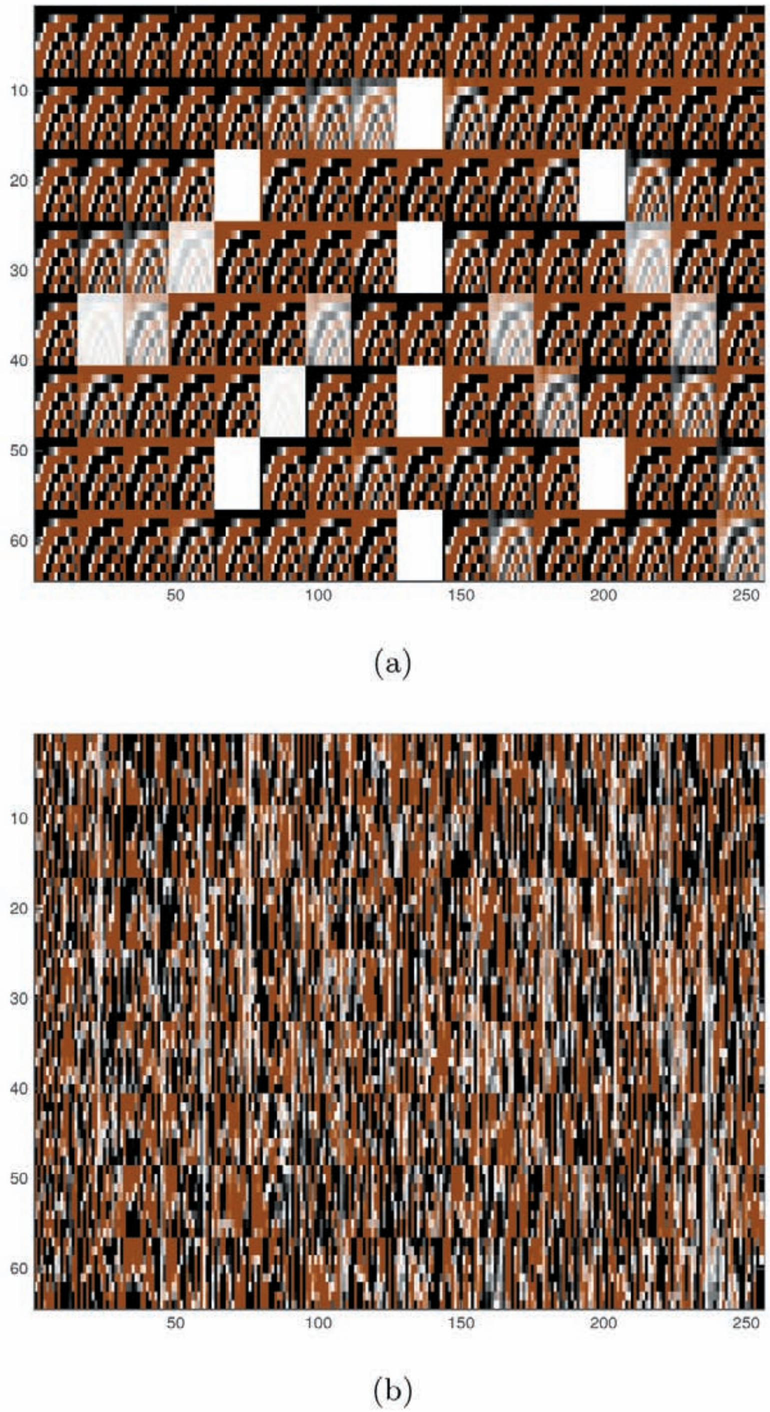
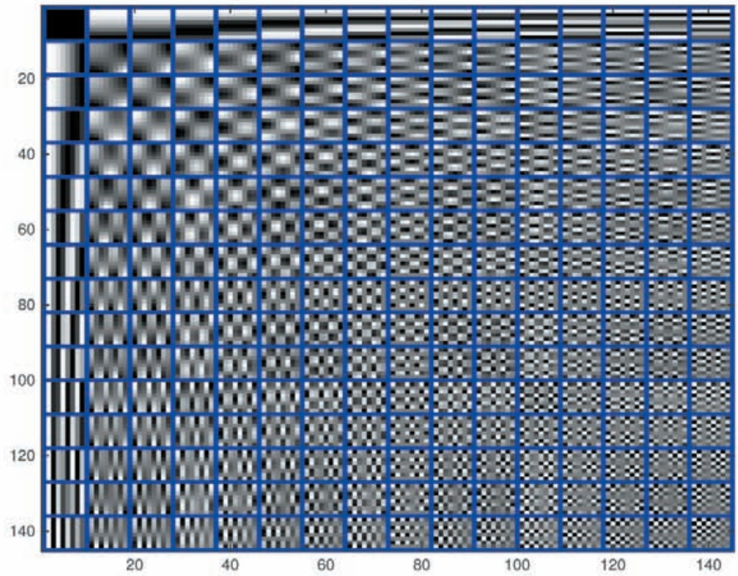
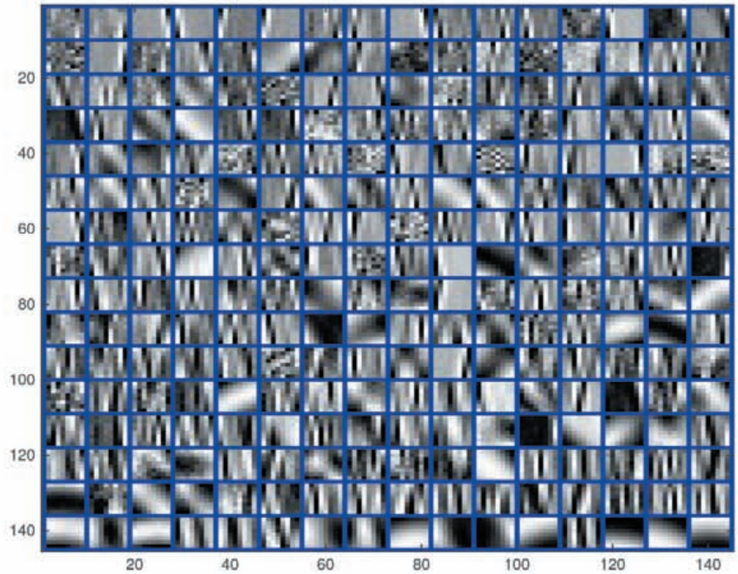


Fig. 4. Dictionary matrix **A** of (a) the discrete cosine transform (DCT) and (b) DLB transform.



(a)



(b)

Fig. 5. (a) Reshaped dictionary (with each atom reshaped into a 8×8 matrix) of the discrete cosine transform (DCT). (b) Reshaped dictionary (with each atom reshaped into a 8×8 matrix) of the DLB transform, each square contains a 2D basis of the DLB transform.

observed that dictionary matrix before training is simple while the dictionary after training is very complicated. The initial dictionary is also called the discrete cosine transform (DCT). To better view the 2D representation of each dictionary atom, we can reshape each column of **A** into a 8×8 matrix and plot all the reshaped 2D atoms together. Fig. 5 shows the reshaped dictionary before and after dictionary learning; the dictionary after training better represents the seismic event than that before training. It also demonstrates that the basis in the DLB transform best matches the local structures of the seismic events and explains why the DLB based method can obtain a better separation between signal and random noise. Fig. 6 shows the matrices **B** and **X** as expressed in eqs. (3) and (4). Since two neighbor patches (8×8 windows) have a 7-point overlap, and the data size of the hyperbolic-events example is 501×76, the size of **B** is 64×34086 and the size of **X** is 256×34086. The matrix **B** is noisy and the matrix **X** is very sparse. Fig. 7 shows the reconstructed data patches, $\hat{\mathbf{B}} = \hat{\mathbf{A}}\hat{\mathbf{X}}$, where $\hat{\cdot}$ denotes estimated matrix. Fig. 7 is much cleaner than Fig. 6a, as a consequence of removal of the random noise.

We further test the effectiveness of the DLB sparse transform in denoising a complicated field data example. The field data are shown in Fig. 8. In this example, we also compare the results of sparse transform based methods with those from the f-x predictive filtering method (Canales, 1984b; Chen and Ma, 2014), wavelet transform filtering, and curvelet transform filtering (Fig. 9). Fig. 10 show their corresponding noise sections. To compare fairly, the results shown in Figs. 9 and 10 are all the best results that can be obtained by each of the individual methods. We try to minimize signal damage in the noise section to make each result acceptable, while judging the performance via the noise level in the difference section. From both the denoised results and the removed noise sections, it is clearly observed that while the wavelet and curvelet transforms fail to remove a large amount of noise, the DLB removes most of noise without damaging the useful energy. The f-x predictive filtering method removes more noise, but also causes some damage to the signals and so is thought to be the least effective method.

CONCLUSION

A sparse DLB denoising approach is relatively new to seismic data processing. We have introduced in detail the mathematical background of the methodology, and a fast orthogonal matching pursuit (OMP) algorithm to accelerate the sparse coding process, which is one of the two key steps in sparse dictionary learning. Both synthetic and field data examples show that the DLB sparse transform obtains obviously better performance in attenuating random noise, even when the data structure is very complicated. Numerical tests also confirm the computational speedup obtained using the proposed fast OMP algorithm.

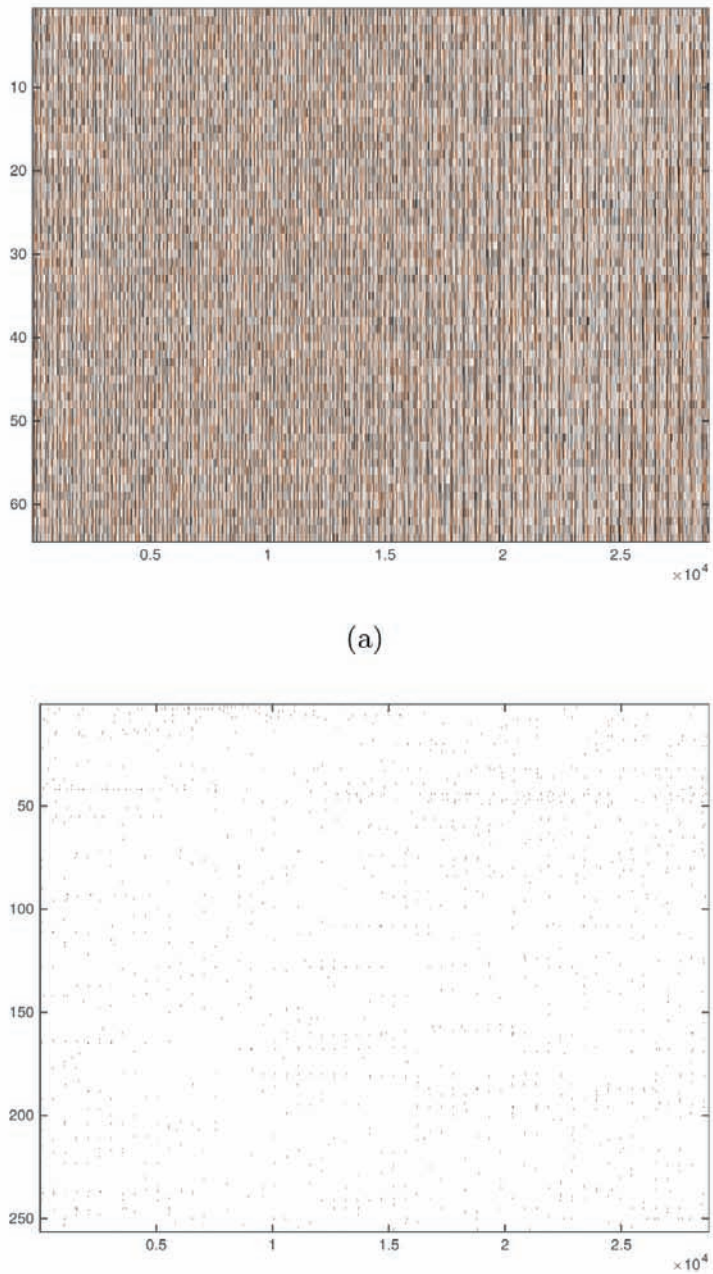


Fig. 6. (a) Matrix **B**. (b) Matrix **X**. Note that the matrix **X** is the sparse coefficient matrix, and is very sparse.

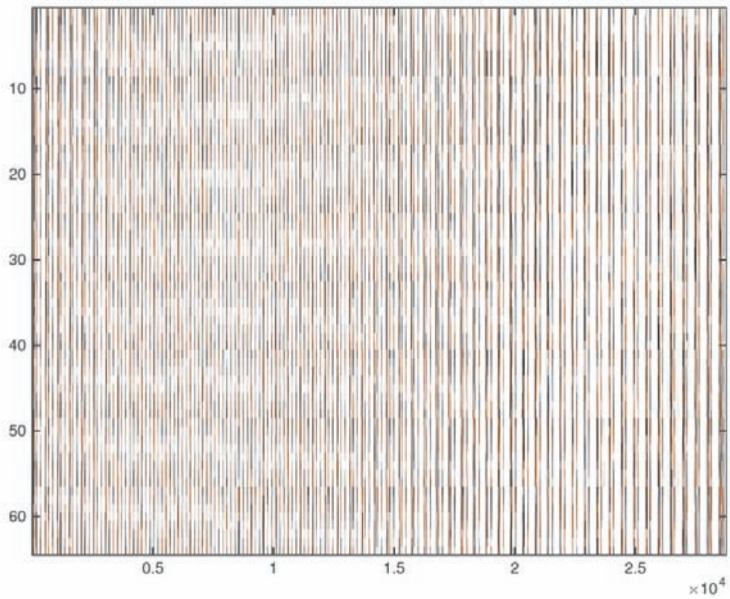


Fig. 7. Denoised patches $\hat{\mathbf{B}} = \hat{\mathbf{A}}\hat{\mathbf{X}}$.

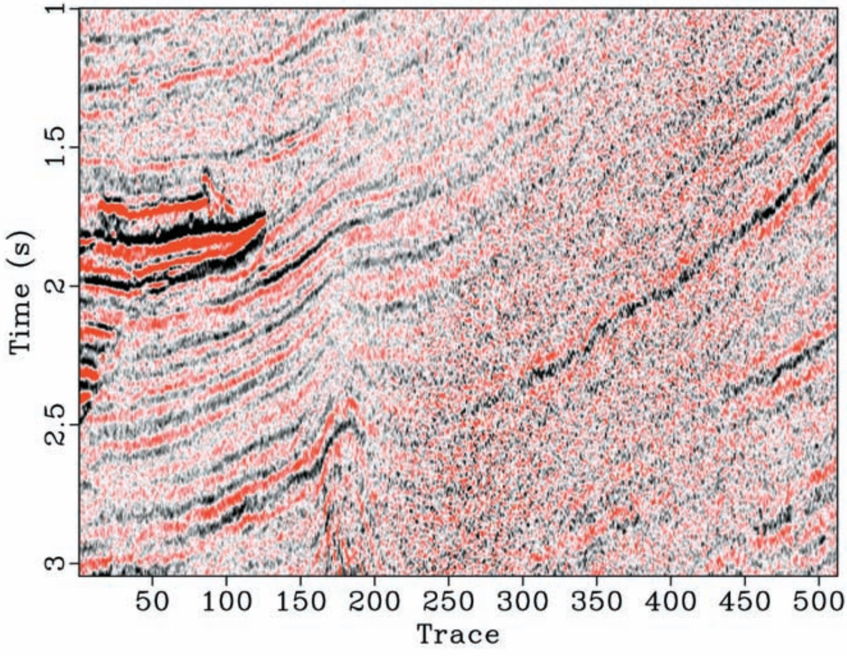


Fig. 8. Field data example.

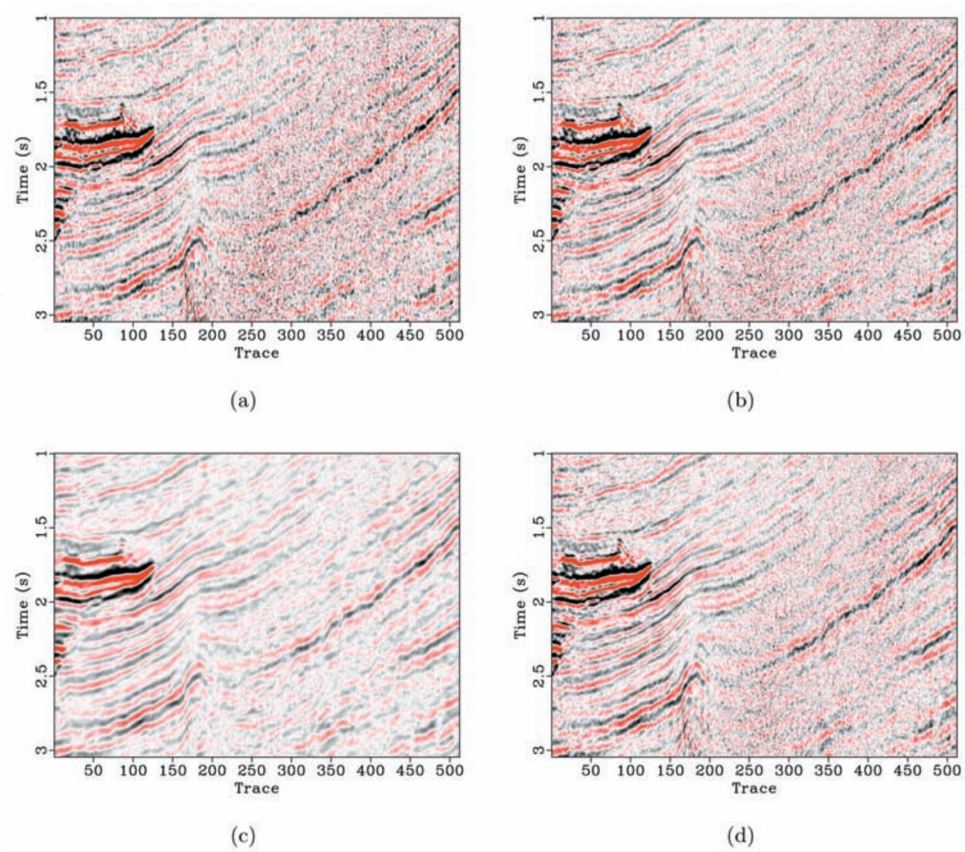


Fig. 9. Denoised data using (a) wavelet thresholding, (b) curvelet thresholding, (c) f-x predictive filtering, and the DLB transform via the fast OMP algorithm.

ACKNOWLEDGMENTS

We would like to thank Shuwei Gan, Weilin Huang, Shaohuan Zu, Wei Chen for inspiring discussions. This work was supported by the China Postdoctoral Science Foundation (No.2014M561053), Humanity and Social Science Foundation of Ministry of Education of China (No.15YJA630108), and Hebei Province Natural Science Foundation (No. E2016202341).

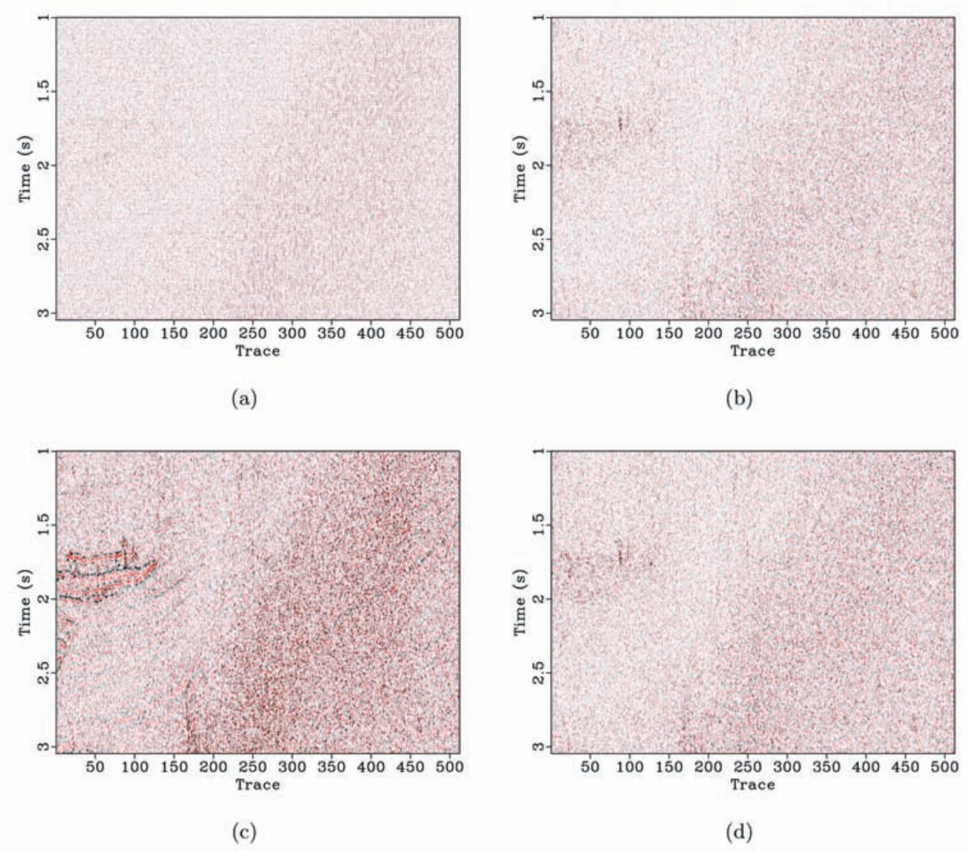


Fig. 10. Removed noise using (a) wavelet thresholding, (b) curvelet thresholding, (c) f-x predictive filtering, and (d) DLB transform via the fast OMP algorithm.

REFERENCES

Aharon, M., Elad, M. and Bruckstein, A.M., 2006. The K-SVD: An algorithm for designing of ever-complete dictionaries for sparse representation. IEEE Transact. Sign. Process., 54: 4311-4322.

Bekara, M. and van der Baan, M., 2007. Local singular value decomposition for signal enhancement of seismic data. Geophysics, 72: V59-V65.

Canales, L., 1984. Random noise reduction. Expanded Abstr., 54th Ann. Internat. SEG Mtg., Atlanta: 525-527.

Chen, Y., 2015a. Deblending using a space-varying median filter. Explor. Geophys., 46: 332-341.

Chen, Y., 2015b. Iterative deblending with multiple constraints based on shaping regularization. IEEE Geosci. Remote Sens. Lett., 12: 2247-2251.

Chen, Y., 2016. Dip-separated structural filtering using seislet thresholding and adaptive empirical mode decomposition based dip filter. Geophys. J. Internat., 206: 457-469.

Chen, Y., 2017. Fast dictionary learning for noise attenuation of multidimensional seismic data. Geophys. J. Internat., 209: 21-31.

Chen, Y. and Fomel, S., 2015a. EMD-seislet transform. Expanded Abstr., 85th Ann. Internat. SEG Mtg., New Orleans: 4775-4778.

Chen, Y. and Fomel, S., 2015b. Random noise attenuation using local signal-and-noise orthogonalization. Geophysics, 80: WD1-WD9.

Chen, Y., Fomel, S. and Hu, J., 2014a. Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization. Geophysics, 79: V179-V189.

Chen, Y. and Jin, Z., 2015. Simultaneously removing noise and increasing resolution of seismic data using waveform shaping. IEEE Geosci. Remote Sens. Lett., 13: 102-104.

Chen, Y., Jin, Z., Gan, S., Yang, W., Xiang, K., Bai, M. and Huang, W., 2015. Iterative deblending using shaping regularization with a combined PNMO-MF-FK coherency filter. J. Appl. Geophys., 122: 18-27.

Chen, Y. and Ma, J., 2014. Random noise attenuation by f-x empirical mode decomposition predictive filtering. Geophysics, 79: V81-V91.

Chen, Y., Ma, J. and Fomel, S., 2016a. Double-sparsity dictionary for seismic noise attenuation. Geophysics, 81: V17-V30.

Chen, Y., Zhang, D., Huang, W. and Chen, W., 2016b. An open-source matlab code package for improved rank-reduction 3D seismic data denoising and reconstruction. Comput. Geosci., 95: 59-66.

Chen, Y., Zhang, D., Jin, Z., Chen, X., Zu, S., Huang, W. and Gan, S., 2016c. Simultaneous denoising and reconstruction of 5D seismic data via damped rank-reduction method. Geophys. J. Internat., 206: 1695-1717.

Chen, Y., Zhou, C., Yuan, J. and Jin, Z., 2014. Application of empirical mode decomposition to random noise attenuation of seismic data. J. Seismic Explor., 23: 481-495.

Duijndam, A.J., Schonewille, M.A. and Hindriks, C.O.H., 1999. Reconstruction of band-limited signals, irregularly sampled along one spatial direction. Geophysics, 64: 524-538.

Fomel, S. and Liu, Y., 2010. Seislet transform and seislet frame. Geophysics, 75: V25-V38.

Gan, S., Chen, Y., Zu, S., Qu, S. and Zhong, W., 2015a. Structure-oriented singular value decomposition for random noise attenuation of seismic data: J. Geophys. Engineer., 12: 262-272.

Gan, S., Wang, S., Chen, Y., Zhang, Y. and Jin, Z., 2015b. Dealiased seismic data interpolation using seislet transform with low-frequency constraint. IEEE Geosci. Remote Sens. Lett., 12: 2150-2154.

Gan, S., Chen, Y., Wang, S., Chen, X., Huang, W. and Chen, H., 2016a. Compressive sensing for seismic data reconstruction using a fast projection onto convex sets algorithm based on the seislet transform. J. Appl. Geophys., 130: 194-208.

Gan, S., Wang, S., Chen, Y., Chen, J., Zhong, W. and Zhang, C., 2016b. Improved random noise attenuation using f-x empirical mode decomposition and local similarity. Appl. Geophys., 13: 127-134.

Gan, S., Wang, S., Chen, Y. and Chen, X., 2016c. Simultaneous-source separation using iterative seislet-frame thresholding. IEEE Geosci. Remote Sens. Lett., 13: 197-201.

Gan, S., Wang, S., Chen, Y., Chen, X. and Xiang, K., 2016d. Separation of simultaneous sources using a structural-oriented median filter in the flattened dimension. Comput. Geosci., 86: 46-54.

Gan, S., Wang, S., Chen, Y., Qu, S. and Zu, S., 2016e., Velocity analysis of simultaneous-source data using high-resolution semblance-coping with the strong noise. Geophys. J. Internat., 204: 768-779.

Hennenfent, G. and Herrmann, F., 2006. Seismic denoising with nonunformly sampled curvelets. Comput. Sci. Engineer., 8: 16-25.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C. and Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. London, Series A, 454: 903-995.

Huang, W., Wang, R., Chen, Y., Li, H. and Gan, S., 2016a. Damped multichannel singular spectrum analysis for 3D random noise attenuation. Geophysics, 81: V261-V270.

Huang, W., Wang, R., Chen, Y., Yuan, Y. and Zhou, Y., 2016b. Randomized-order multi-channel singular spectrum analysis for simultaneously attenuating random and coherent noise. Expanded Abstr., 86th Ann. Internat. SEG Mtg., Dallas: 4777-4781.

Huang, W., Wang, R., Yuan, Y., Gan, S. and Chen, Y., 2017. Signal extraction using randomized-order multichannel singular spectrum analysis. Geophysics, 82(2): V59-V74.

Huang, W., Wang, R., Zhang, M., Chen, Y. and Yu, J., 2015. Random noise attenuation for 3D seismic data by modified multichannel singular spectrum analysis. Extended Abstr., 77th EAGE Conf., Madrid. DOI.10.3997/2214- 4609.201412830.

Huang, W., Wang, R., Zhou, Y., Chen, Y. and Yang, R., 2016c. Improved principal component analysis for 3D seismic data simultaneous reconstruction and denoising. Expanded Abstr., 86th Ann. Internat. SEG Mtg., Dallas.

Kaplan, S.T., Sacchi, M.D. and Ulrych, T.J., 2009. Sparse coding for data-driven coherent and incoherent noise attenuation. Expanded Abstr., 79th Ann. Internat. SEG Mtg., Houston: 3327-3331.

Kong, D. and Peng, Z., 2015. Seismic random noise attenuation using shearlet and total generalized variation. J. Geophys. Engineer., 12: 1024-1035.

Kong, D., Peng, Z., He, Y. and Fan, H., 2016. Seismic random noise attenuation using directional total variation in the shearlet domain. J. Seismic Explor., 25: 321-338.

Li, H., Wang, R., Cao, S., Chen, Y. and Huang, W., 2016. A method for low-frequency noise suppression based on mathematical morphology in microseismic monitoring. Geophysics, 81(3): V159-V167.

Liu, G., Fomel, S., Jin, L. and Chen, X., 2009a. Stacking seismic data using local correlation. Geophysics, 74: V43-V48.

Liu, Y., Liu, C. and Wang, D., 2009b. A 1D time-varying median filter for seismic random, spike-like noise elimination. Geophysics, 74: V17-V24.

Liu, C., Wang, D., Hu, B. and Wang, T., 2016a. Seismic deconvolution with shearlet sparsity constrained inversion. J. Seismic Explor., 25: 433-445.

Liu, W., Cao, S. and Chen, Y., 2016b. Applications of variational mode decomposition in seismic time-frequency analysis. Geophysics, 81: V365-V378.

Liu, W., Cao, S. and Chen, Y., 2016c. Seismic time-frequency analysis via empirical wavelet transform. IEEE Geosci. Remote Sens. Lett., 13: 28-32.

Liu, W., Cao, S., Chen, Y. and Zu, S., 2016d. An effective approach to attenuate random noise based on compressive sensing and curvelet transform. J. Geophys. Engineer., 13: 135-145.

Liu, W., Cao, S., Liu, Y. and Chen, Y., 2016e. Synchrosqueezing transform and its applications in seismic data analysis. J. Seismic Explor., 25: 27-44.

Naghizadeh, M., 2012. Seismic data interpolation and denoising in the frequency-wavenumber domain. Geophysics, 77: V71-V80.

Neelamani, R., Baumstein, A., Gillard, D., Hadidi, M. and Soroka, W., 2008. Coherent and random noise attenuation using the curvelet transform. The Leading Edge, 27: 240-248.

Rubinstein, R., Zibulevsky, M. and Elad, M., 2008. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. Tech. Rep. Comput. Sci. Dept., Technion Israel Inst. of Technol.

Shahidi, R., Tang, R., Ma, J. and Herrmann, F.J., 2013. Application of randomized sampling schemes to curvelet-based sparsity-promoting seismic data recovery. Geophys. Prosp., 61: 973-997.

Wang, D., Liu, C., Liu, Y. and Liu, G., 2008. Application of wavelet transform based on lifting scheme and percentiles soft-threshold to elimination of seismic random noise. Progr. Geophys. (in Chinese), 23: 1124-1130.

Wang, J., Ng, M. and Perz, M., 2010. Seismic data interpolation by greedy local Radon transform. Geophysics, 75: WB225-WB234.

Wu, J., Wang, R., Chen, Y., Zhang, Y., Gan, S. and Zhou, C., 2016. Multiples attenuation using shaping regularization with seislet domain sparsity constraint. J. Seismic Explor., 25: 1-9.

Xue, Y., Chang, F., Zhang, D. and Chen, Y., 2016a. Simultaneous sources separation via an iterative rank-increasing method. IEEE Geosci. Remote Sens. Lett., 13: 1915-1919.

Xue, Y., Yang, J., Ma, J. and Chen, Y., 2016b. Amplitude-preserving nonlinear adaptive multiple attenuation using the high-order sparse Radon transform. J. Geophys. Engineer., 13: 207-219.

Xue, Y., Man, M., Zu, S., Chang, F. and Chen, Y., 2017. Amplitude-preserving iterative deblending of simultaneous source seismic data using high-order Radon transform. J. Appl. Geophys., 139: 79-90.

Yang, W., Wang, R., Chen, Y. and Wu, J., 2014. Random noise attenuation using a new spectral decomposition method. Expanded Abstr., 84th Ann. Internat. SEG Mtg., Denver: 4366-4370.

Yang, W., Wang, R., Chen, Y., Wu, J. Qu, S., Yuan, J. and Gan, S., 2015. Application of spectral decomposition using regularized non-stationary autoregression to random noise attenuation. J. Geophys. Engineer., 12: 175-187.

Yu, Z., Ferguson, J., McMechan, G. and Anno, P., 2007. Wavelet-Radon domain dealiasing and interpolation of seismic data. Geophysics, 72: V41-V49.

Zhang, D., Chen, Y., Huang, W. and Gan, S., 2016a. Multi-step damped multichannel singular spectrum analysis for simultaneous reconstruction and denoising of 3D seismic data. J. Geophys. Engineer., 13: 704-720.

Zhang, D., Chen, Y., Huang, W. and Gan, S., 2016b. Multi-step reconstruction of 3D seismic data via an improved mass algorithm. Expanded Abstr., CPS/SEG Internat. Geophys. Conf., Beijing: 745-749.

Zhong, W., Chen, Y., Gan, S. and Yuan, J., 2016. $L_{1/2}$ norm regularization for 3D seismic data interpolation. J. Seismic Explor., 25: 257-268.

Zu, S., Zhou, H., Chen, Y., Qu, S., Zou, X., Chen, H. and Liu, R., 2016a. A periodically varying code for improving deblending of simultaneous sources in marine acquisition. Geophysics, 81: V213-V225.

Zu, S., Zhou, H., Chen, Y., Chen, H., Cao, M. and Xie, C., 2016b. A marine field trial for iterative deblending of simultaneous sources. Expanded Abstr., 86th Ann. Internat. SEG Mtg., Dallas: 113-118.