SPARSE DICTIONARY LEARNING FOR NOISE ATTENUATION IN THE EXACTLY FLATTENED DIMENSION

HUI LV

Department of Architecture Engineering and Administration, Jiangxi Institute of Economic Administrators, 269 Wolong Road, Hongjiaozhou District Nanchang, Jiangxi Province, 330088, P.R. China.
Present address: School of Civil Engineering and Architecture, Nanchang Hangkong University, 696 South Fenghe Avenue, Nanchang 330063, Jiangxi Province, P.R. China.
lv_hui2016@126.com

(Received August 16, 2018; revised version accepted July 20, 2019)

ABSTRACT

Lv, H., 2019. Sparse dictionary learning for noise attenuation in the exactly flattened dimension. *Journal of Seismic Exploration*, 28: 449-474.

Seismic noise attenuation is a long-standing and crucial problem in reflection seismic data processing community. In recently years, the dictionary learning based approaches have attracted more and more attention. Dictionary learning provides an adaptive way to optimally represent a given dataset. In dictionary learning, the basis function is adapted according the given data instead of being fixed in many analytical sparse transforms. The application of the dictionary learning techniques in seismic data processing has been popular in the past decade. However, most dictionary learning algorithms are directly taken from the image processing community and thus are not suitable for seismic data. Considering that the seismic data is spatially coherent, the dictionary should better be learned according to the coherency information in the seismic data. We found the dictionary learning performs better when the spatial correlation is stronger and thus we propose to use a flattening operator to help learn the dictionary in the flattened dimension, where the strong spatial coherence helps construct a dictionary that follows better the structural pattern in the seismic data. The presented dictionary learning in the flattened dimension (DLF) thus has a stronger capability in separating signal and noise. We use both synthetic and field data examples to demonstrate the superb performance of the proposed method.

KEY WORDS: noise suppression, dictionary learning, flattening, seismic data, signal-to-noise ratio.

0963-0651/19/$5.00

INTRODUCTION

Seismic data processing is a crucial step for preparing high-quality data that can be used to image the subsurface structure. Seismic noise attenuation is one of the most significant steps in the whole seismic data processing and imaging workflow. It has great influence to many subsequent processing tasks, such as amplitude-variation-offset inversion, reverse time migration, full waveform inversion, and automatic interpretation for oil&gas detection (Gan et al., 2016c; Rastogi et al., 2017; Zhou and Han, 2018; Zhou et al., 2017a; Li et al., 2017a; Fabien-Ouellet et al., 2017; Liu et al., 2017a; Li et al., 2017b; Li and Zhang, 2017; Wang et al., 2017b; Wang et al., 2018;Garcia-Yeguas et al., 2017; Yang and Zhu, 2017; Song et al., 2017; Chen et al., 2018; Chen, 2018a). To enhance the quality of the seismic data, one can apply different mathematical or statistical methods to boost the signal to noise ratio (SNR) of the data.

In the past several decades, a large number of algorithms have been developed for seismic noise attenuation. Stacking the seismic data along the spatial directions, e.g., the offset direction, can enhance the energy of spatially coherent useful waveform signals as well as mitigate the spatially incoherent random noise (Mayne, 1962; Yang et al., 2015; Zu et al., 2016; Xie et al., 2017; Wu and Bai, 2018b). One of the commonly used state-of-the-art algorithms is the prediction-based method, including t-x predictive filtering (Abma and Claerbout, 1995) and f-x deconvolution (Canales, 1984; Chen and Ma, 2014). This type of methods utilize the predictive property of useful signals along spatial direction to create a regression-like model for distinguishing between signal and noise. Another type of commonly used methods are based on data decomposition (Chen and Ma, 2014; Gan et al., 2016a; Chen et al., 2016a, 2017a,d,b; Gan et al., 2015a; Wang et al., 2017a; Liu et al., 2016a, 2017c, 2018; Fomel, 2013; Wu et al., 2018). This type of methods assume that noisy seismic data can be decomposed into different components where signal and noise are separated based on their frequency difference or morphological difference (Li et al., 2016b). Empirical mode decomposition (EMD) (Huang et al., 1998; Chen, 2016) and its improved version, e.g., ensemble empirical mode decomposition (EEMD) (Wu and Huang, 2009), complete ensemble empirical mode decomposition (CEEMD) (Chen et al., 2016a), have been used intensively for reducing the noise in seismic data (Chen et al., 2017b). Variational mode decomposition was proposed by Dragomiretskiy and Zosso (2014) for substituting EMD because of its explicit control on the decomposition performance. It has been utilized for noise attenuation in Liu et al. (2017b) and for time-frequency analysis by Liu et al. (2017c). Regularized non-stationary decomposition (Wu et al., 2018; Chen, 2018b) is another decomposition method which is also based on a solid mathematical model.

Other denoising methods include morphology methods (Li et al., 2016a,b), which are based on the morphology difference of signal and noise, transform methods (Chen et al., 2014; Zu et al., 2016; Zu et al., 2017; Liu et al., 2016b,c; Gan et al., 2016d,e; Zu et al., 2017; Xue et al., 2017; Chen and Fomel, 2018), which are based on the sparsity in a transform domain, median filtering methods (Gan et al., 2016b; Chen et al., 2017c; Bai and Wu, 2017; Xie et al., 2018; Chen et al., 2019), which are especially useful in rejecting spike-like noise, and rank-reduction based methods (Vautard et al., 1992; Gao et al., 2013; Cheng and Sacchi, 2015; Chen et al., 2016d; Zhou and Zhang, 2017; Zhou et al., 2018; Bai et al., 2018a,b). Rank-reduction based approaches assume the seismic data to be low-rank after some data rearrangement steps (Chen et al., 2016d). Such methods include nuclear norm minimization (Zhou and Zhang, 2017), singular spectrum analysis (Vautard et al., 1992; Gao et al., 2013; Cheng and Sacchi, 2015; Zhou et al., 2018), damped singular spectrum analysis (Chen et al., 2016c; Siahsar et al., 2017c), multi-step singular spectrum analysis (Zhang et al., 2016), sparsity-regularized singular spectrum analysis (Siahsar et al., 2016; Zhang et al., 2017), structural low-rank approximation (Zhou et al., 2017b), and empirical low-rank approximation (Chen et al., 2017f). The rank-reduction based methods may require a local processing strategy that can further improve the denoising capability (Chen et al., 2017e). Instead of developing a standalone denoising strategy, Chen and Fomel (2015) proposed a two-step denoising approach that tries to solve a long-existing problem in almost all denoising approaches: the signal leakage problem. By initiating a new concept called local orthogonalization, Chen and Fomel (2015) successfully retrieved the coherent signals from the removed noise section to guarantee no signal leakage in any denoising algorithms. Xue et al. (2016) used the rank-increasing property of noise when iteratively estimating the useful signals from simultaneous-source data.

Dictionary learning (DL) methods (Rubinstein et al., 2008; Romano and Elad, 2013; Wu and Bai, 2018a) are alternatives to predefining a transform. They capture the morphology of the redundant signal present in the data, to provide a dictionary which is optimal to represent this data in a sparse manner. The resulting dictionary is similar to and has the same role as a transform, but it is only physically expressed, i.e., it is stored as a matrix. Since the dictionary is data-driven, it often leads to sparse representations that are closer to the data compared with predefined transforms. In seismic processing, DL can provide state-of-the-art results for random noise attenuation, coherent noise elimination, and reconstruction of randomly missing traces. The sparse dictionary ased methods belong to a machine learning type of denoising method. These methods have not been widely used in seismic data processing until recent years (Siahsar et al., 2017b,a; Chen, 2017). Zhu et al. (2014) developed a parametric dictionary learning scheme which exploits underlying sparse structure constraint to reduce the learned dictionary atoms. Zhou et al. (2016) used an patchwise dictionary learning scheme with an additional regularization term on the dictionary to separate the interference for the simultaneous source data. To combine the advantages of analytic approach and dictionary learning approach, Rubinstein et al. (2010) and Chen et al. (2016b) implemented the double sparsity dictionary to attenuate random noise. To solve the problem of computation, Rubinstein et al. (2008) used an approximate solution to instead the K-SVD algorithm, which requires a few iteration to provide very close results to the whole computation.

In this paper, we propose a dictionary learning algorithm to learn the atoms along an flattened event trajectory. The spatial coherence is thus strengthened and the learned dictionary has a better capability to represent the seismic data. Compared with our previous work (Lv and Bai, 2018), in this paper we develop an exact flattening operator to learn the features in the flattened domain in order to better reject the noise. We first introduce the basics of the dictionary learning theory and introduce the flattening operator. We then use synthetic and field data examples to compare the learned dictionaries using different methods and their corresponding denoising performance. Finally, we draw some key conclusions at the end of the paper.

THEORY

Dictionary learning

Dictionary learning (DL) methods (Romano and Elad, 2013; Rubinstein et al., 2008) are effective tools to automatically find a sparse representation of a data set. They train a set of basis vectors on the data to capture the morphology of the redundant signals. The basis vectors are called atoms, and the set is referred to as the dictionary. This dictionary can be used to represent the data in a sparse manner with a linear combination of few of its atoms. Given noisy data y arranged into the training sample matrix $\mathbf{Y} = R(y) \in R^N$, where R denotes the sampling operator which extracts a block from noisy data y and sort the block into a vector as the column of matrix **Y**. Denoising of **Y** by dictionary learning amounts to find a dictionary D ∈ RN(whose columns are the dictionary atoms) and the sparse coefficients $A \in R^{KxN}$ to represent the training matrix **Y**, which can be expressed as

$$\arg\min_{\mathbf{D},\mathbf{A}} \| \mathbf{Y} - \mathbf{DA} \|_F^2, \text{s.t.} \| \mathbf{A} \|_0 \leq T_0 \quad , \tag{1}$$

where $\|\cdot\|_F$ represents the Frobenius norm of a matrix, $\|\cdot\|_0$ refers to the L_0 norm, which denotes the number of non-zero elements of the input vector.

The multiplication **DA** is bilinear, which means that it is linear in each variable of **D** or **A** if the other one is constant. K-SVD is the most widely used way to solve eq. (1). In the K-SVD algorithm, there are mainly two steps:

1. Sparse-coding the training matrix **Y** using the dictionary.

2. Updating the dictionary atoms using the singular value decomposition (SVD).

The sparse-coding is to solve the optimization problem with the given dictionary **D**, e.g., discrete cosine transform dictionary (DCT):

$$\arg\min_{\mathbf{A}} \parallel \mathbf{Y} - \mathbf{DA} \parallel_F^2, \text{s.t.} \parallel \mathbf{A} \parallel_0 \leq T_0 \quad . \tag{2}$$

Solving eq. (2) is NP-hard and the L_1 norm approximate solution can be considered to solve the similar error-constrained objective function

$$\arg\min_{\mathbf{A}} \parallel \mathbf{A} \parallel_1 \text{ s.t.} \parallel \mathbf{Y} - \mathbf{DA} \parallel_F^2 \leq \epsilon \quad , \tag{3}$$

where ε is the target error, $\|\cdot\|_1$ refers to the L_1 norm. The orthogonal matching pursuit (OMP) method can be effective to solve the minimization problem approximately.

Updating the dictionary aims to solve the following the subproblem

$$\begin{aligned} \parallel \mathbf{Y} - \mathbf{DA} \parallel_F^2 &= \parallel \mathbf{Y} - \sum_{j=1}^{K} \mathbf{d}_j \mathbf{a}_T^j \parallel_F^2, \\ &= \parallel \mathbf{Y} - \sum_{j \neq k} \mathbf{d}_j \mathbf{a}_T^j - \mathbf{d}_k \mathbf{a}_T^k \parallel_F^2, \\ &= \parallel \mathbf{E}_k - \mathbf{d}_k \mathbf{a}_T^k \parallel_F^2 \quad , \end{aligned} \tag{4}$$

where $\mathbf{E}_k$ is the error matrix without the *k*-th column atom, $\mathbf{d}_k$ and $\mathbf{a}^k{}_T$ are updated atoms and its corresponding coefficient row in **A**, respectively. This problem can then be easily solve via SVD. SVD decomposes the error matrix $\mathbf{E}_k$ into $U\Delta V^T$. Thus, the updated dictionary atom is the first column of U, and coefficient $\mathbf{a}_k$ is the multiplication of the first column of V and Δ (1, 1). The product **DA** is decomposed into the sum of K rank-1 matrices, in which the algorithm is called "K-SVD". However, the major disadvantage of K-SVD is the computational demanding, especially when $\mathbf{E}_k$ is large, since it may requires thousands of SVD. Usually, an exact solver is not required. To accelerate the learning process, an approximate solution can be used to replace the exact SVD. The step of updating can be accomplished by an iterative alternating optimization (AO) method given by

$$\mathbf{d}_k = \frac{\mathbf{E}_k \mathbf{a}_k}{\| \mathbf{E}_k \mathbf{a}_k \|_2} \quad , \tag{5}$$

$$\mathbf{a}_k = \mathbf{E}_k^T \mathbf{d}_k \quad . \tag{6}$$

Dictionary learning in the flattened dimension

In conventional DL, the atoms are unstructured, and are only numerically defined over a grid which has the same sampling as the data. Consequently, the atoms are unknown away from this sampling grid, and a sparse representation of the data in the dictionary domain is not sufficient information to guarantee a successful separation between signal and noise. We find that the performance of the dictionary learning highly depends on the spatial coherence of the training data. The worse coherence will make the learned dictionary less capable of representing the complex structure. For aiding the dictionary learning process in the case of structurally complex area, e.g., steeply dipping data, we propose to prepare the data to be more spatially coherent so that the structural complexity is much decreased, which is less demanding for the dictionary training.

To prepare a seismic gather with better spatial coherence, we follow the prediction based method proposed in Liu et al. (2010) and Gan et al. (2016b). Simply speaking, seismic data differs from a general digital image in that it contains events that are coherent along the space direction. Because of such space coherence, the neighbor traces can be predicted. The flattening process is equivalent to applying multiple prediction operations among neighbor traces so that the output data after the flattening process contain only flat/horizontal events. The process can be simply expressed as

$$\mathbf{P}_j \mathbf{D}_j^R = \overline{\mathbf{D}}_j^R \quad , \tag{7}$$

where $\mathbf{D}^R{}_j$ means the j-th local window. j denotes the trace number. R denotes the window length. $\mathbf{D}^R{}_j$ denotes the flattened local window for the j-th trace with a window length as R. P_j denotes the j-th flattening operator.

According to Gan et al. (2016b), eq. (7) can be expressed as the detailed form as follows:

$$\begin{bmatrix} \mathbf{P}_{(1,j)\rightarrow(1+R,j)}(\sigma_{1,j}) & & & & \\ & \ddots & & & \\ & & \mathbf{P}_{(1+R,j)\rightarrow(1+R,j)}(\sigma_{1+R,j}) & & \\ & & & \ddots & \\ & & & & \mathbf{P}_{(1+2R,j)\rightarrow(1+R,j)}(\sigma_{1+2R,j}) \end{bmatrix}$$
$$[\mathbf{d}_{1,j},\cdots,\mathbf{d}_{1+R,j},\cdots,\mathbf{d}_{1+2R,j}]$$
$$= [\overline{\mathbf{d}}_{1,j},\cdots,\overline{\mathbf{d}}_{1+R,j},\cdots,\overline{\mathbf{d}}_{1+2R,j}] \tag{8}$$

$\mathbf{d}_{i,j}$ denotes i-th trace in the j-th window. $\sigma_{i,j}$ denotes the slope at i-th trace in the j-th window. If we need to predict a trace from a relatively distant trace, we have to use a recursive prediction strategy, e.g., (Liu et al., 2010), the prediction from the k-th trace to the first trace can be expressed as

$$\mathbf{P}_{(1,j)\rightarrow(k,j)}(\sigma_{1,j}) = \mathbf{P}_{(k-1,j)\rightarrow(k,j)}(\sigma_{k-1,j})\cdots\mathbf{P}_{(2,j)\rightarrow(3,j)}(\sigma_{2,j})\mathbf{P}_{(1,j)\rightarrow(2,j)}(\sigma_{1,j}) \tag{9}$$

Fig. 1 shows a synthetic example, where (a) denotes a synthetic seismic gather and (b) denotes the estimated slope using the PWD method (Fomel, 2002). Fig. 2 shows a example of the flattened gathers. Fig. 2(a) shows the flattened gather corresponding to the 5-th trace and Fig. 2(b) shows the flattened window corresponding to the 30-th trace. It can demonstrated that with the aforementioned flattening operation, a curved seismic event can be flattened in the local window.

The PWD algorithm that we use to generate the local slope map is currently known as the most robust algorithm in the literature (Zu et al., 2017). However, in the case of extremely strong random noise, the accuracy of the result from the PWD algorithm is still unreliable. Thus, to improve the reliability of the PWD algorithm in extreme cases will be a future topic.

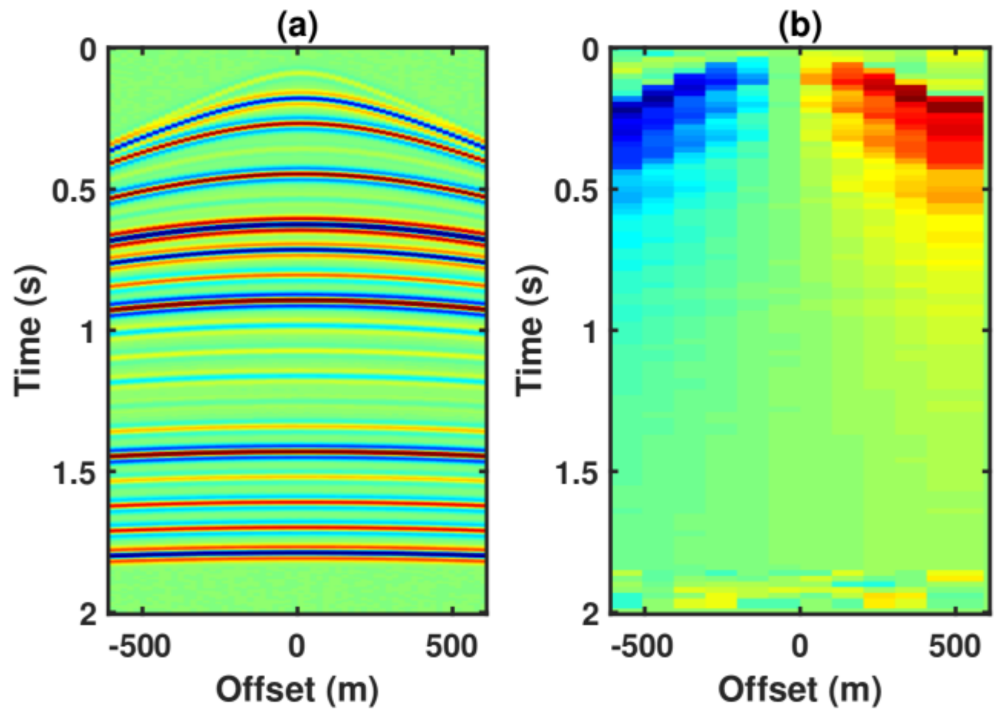


Fig. 1. (a) Synthetic seismic gather. (b) Estimated slope using the PWD method.

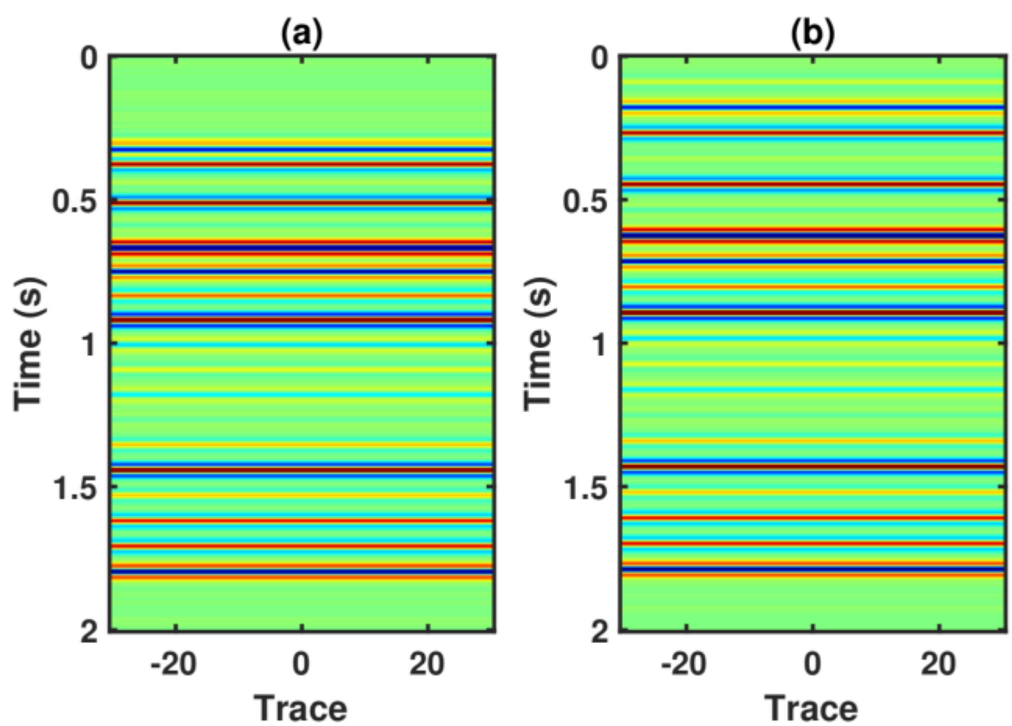


Fig. 2. A demonstration of the flattened gather. (a) Flattened window corresponding to the 5-th trace. (b) Flattened window corresponding to the 30-th trace.

EXAMPLES

In this section, we will use two examples to demonstrate the performance of the dictionary learning method in the flattened dimension (DLF) in separating useful reflection signals and ambient random noise. We would be especially curious on how the flattening operator affect the learned dictionary atoms. To quantitatively compare the denoising performance, we use the signal-to-noise ratio (SNR) metric defined below:

$$SNR = 10\log_{10}\frac{\|\mathbf{d}_0\|_2^2}{\|\mathbf{d}_0-\mathbf{d}\|_2^2} \quad , \tag{10}$$

where $\mathbf{d}_0$ is the clean data, and $\mathbf{d}$ is the noisy or denoised data. Both $\mathbf{d}_0$ and $\mathbf{d}$ are vectorized data (1D vector).

The first synthetic example is a synthetic example, as shown from Figs. 3 to 8. Fig. 3(a) shows the clean data. The noisy data is shown in Fig. 3(b). The SNR of the noisy data is 3.14 dB. Figs. 3(c) and 3(d) show two zoomed sections from the clean and noisy data. The zooming areas are indicated by the black frameboxes. In this example, we first compare the learned dictionaries for different methods. Fig. 4a shows the initial input dictionary, which corresponds to the simple discrete cosine transform (DCT). Fig. 4b shows the traditionally learned dictionary, which is much different from Fig. 4a. Fig. 4c shows the learned dictionary using the proposed method. The dictionary using the proposed method is learned from the flattened data (to enhance spatial coherence). Comparing each figure in Fig.4, we find that the learned dictionary adapts to the input training data better than the initial dictionary (DCT basis functions). The proposed method differs from the traditional DL method in that the obtained dictionary atoms using the proposed method is much more structurally simple, and more uniform, e.g., with small dip angle and being less spatially aliasing. The traditionally trained dictionary, however, contains atoms with much larger irregularities. Fig. 5 shows the comparison between different denoised data. In addition to the dictionary learning method, we also compare the performance with the f-x deconvolution method. Figs. 5(a)-(b) correspond to the denoised data and Figs. 5(c)-(d) correspond to the zoomed data. It is very clear that the result from the proposed method is much cleaner and closer to the exact solution shown in Fig. 3(a). The SNRs for the f-x deconvolution method, the traditional DL method, and the proposed method are 6.75 dB, 8.34 dB, and 14.76 dB, respectively. Fig. 6 shows a comparison between the noise rejection performance. The noise rejection performance is defined as the difference between noisy data and the denoised data using different methods. Note that the proposed method removes obviously more noise than the traditional DL method. Here, it is worth mentioning that for the sake for losing useful reflection energy, we use

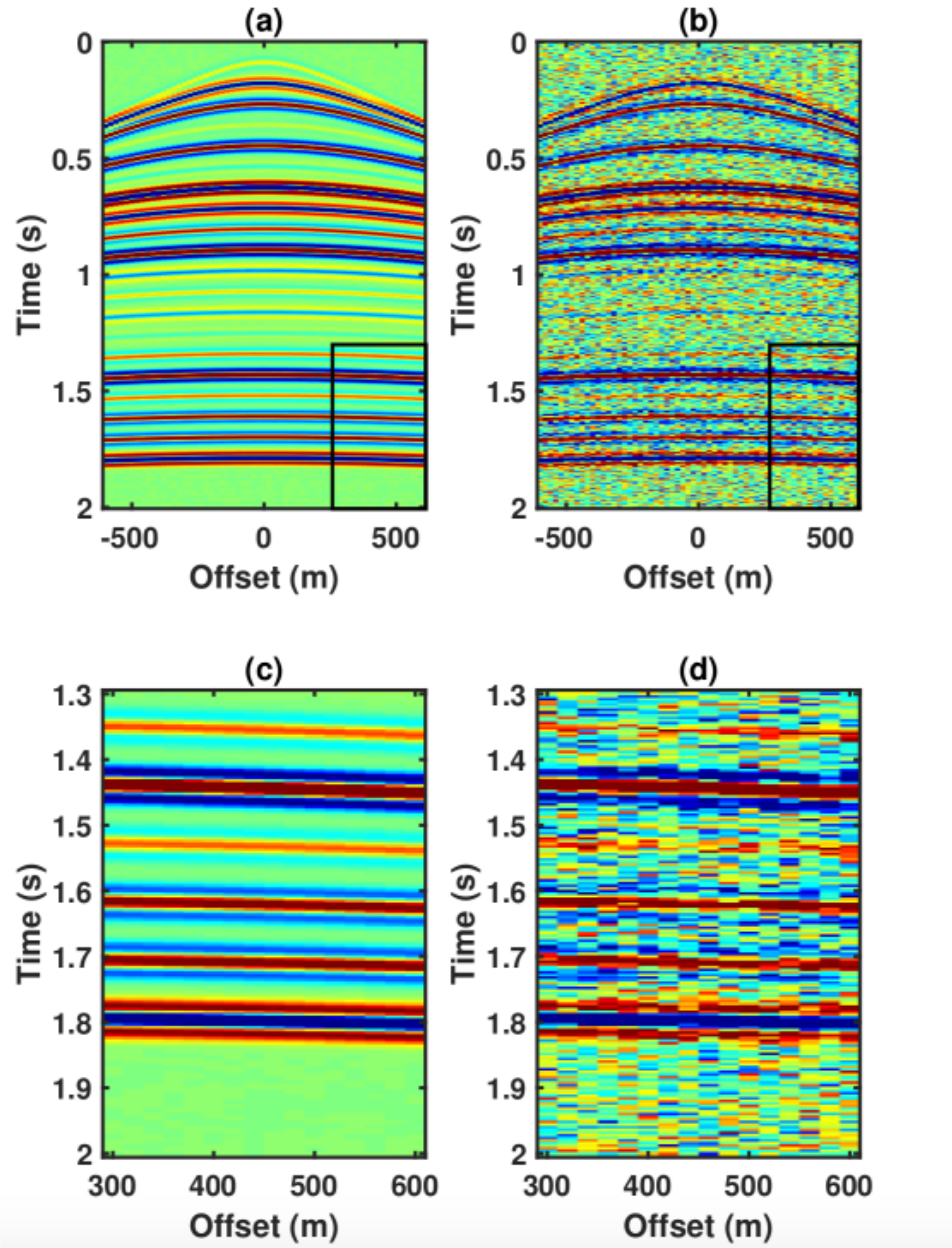


Fig. 3. Synthetic example. (a) Clean data. (b) Noisy data. (c) and (d) Zoomed areas from the frame boxes.

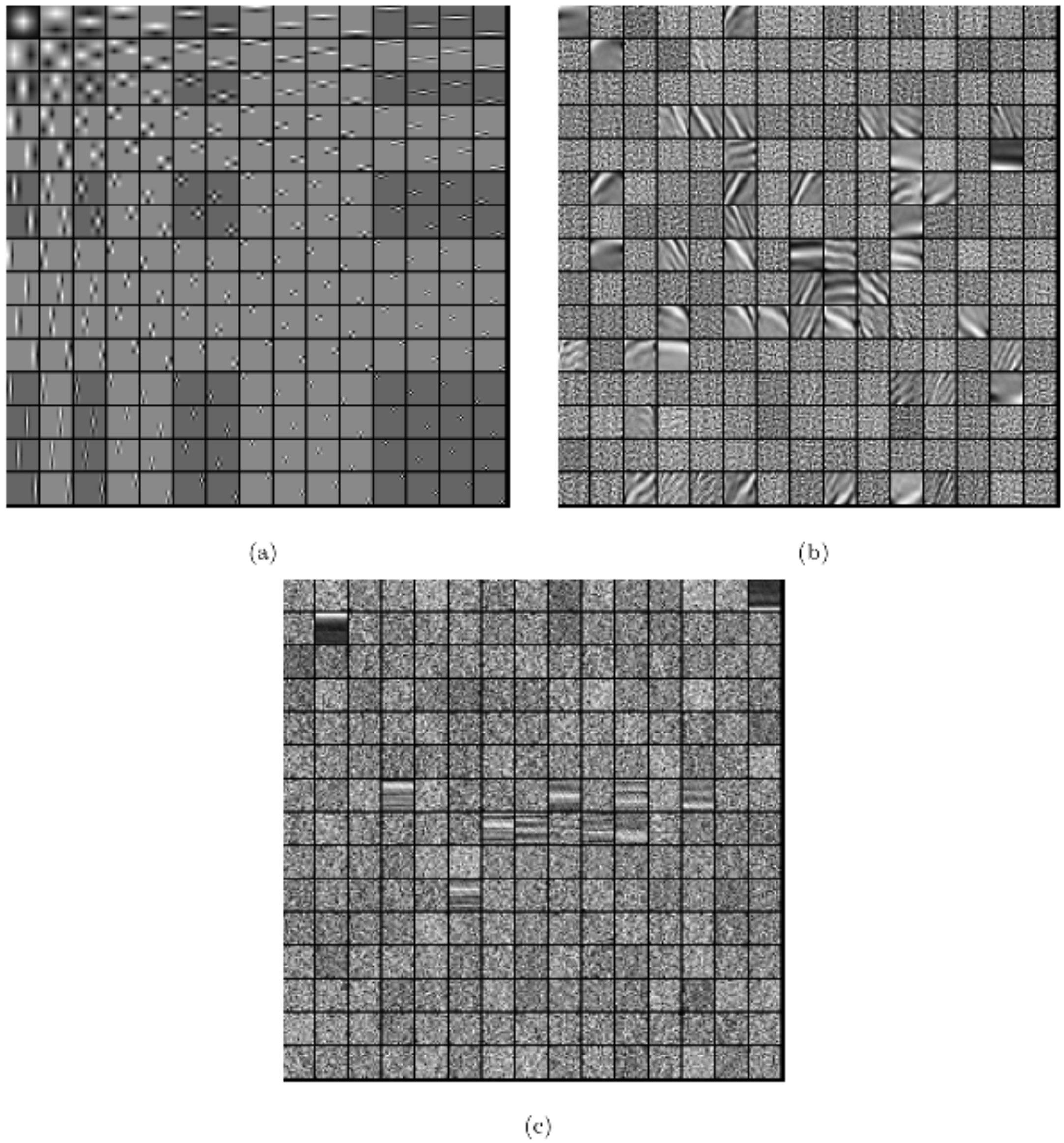


Fig. 4. Comparison between different dictionaries. (a) Initial dictionary. (b) Traditionally learned dictionary. (c) Learned dictionary in the flattened dimension.

a relatively conservative threshold in the transform domain. Because of the less sparsifing performance, we need to use a relatively high threshold value for the traditional DL method, which results a much weaker removal of noise. Fig. 8a shows a comparison of trace-by-trace amplitude for different data. The proposed method appears to be the closest to the clean data. In a better comparison way, we plot the trace-by-trace error for the three methods in Fig. 8b, where we can see that the f-x deconvolution method causes the largest error and the proposed method causes the smallest error.

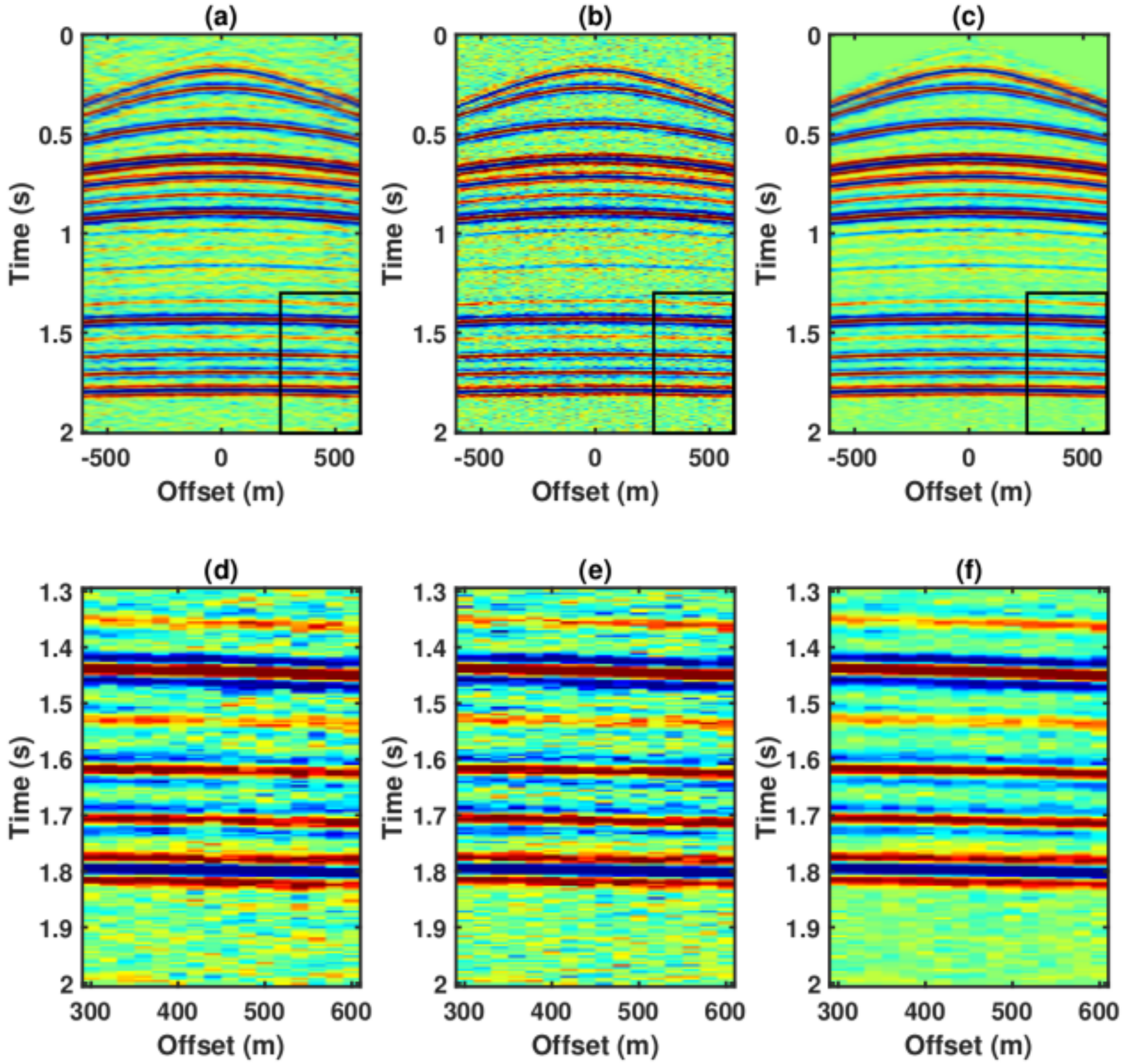


Fig. 5. Synthetic example. (a) Denoised data using the f-x deconvolution method. (b) Denoised data using the traditional DL method. (c) Denoised data using the proposed method. (d)-(f) Zoomed areas from the frame boxes.

It is worth noting that the PWD algorithm is the most accurate and robust slope estimation method to date. It can also work well in the case of strong noise. Thus, in most cases, the flattening operator can also work well and thus the proposed method can work effectively in most situations. However, in order to test the influence of the slope estimation to the final learned dictionary atoms. We implement a synthetic test. We let the estimated slope shown in Fig. 1(b) contain different levels of errors, and then learned dictionary atoms from the noisy data using the inaccurate slopes. The learned dictionaries in different cases are shown in Fig. 9, where (a)-(d) shows the learned dictionary atoms when the local slope contains 20%, 10%, 5%, and 2% errors, respectively. Note that the as the slope becomes more and more accurate, the learned dictionary atoms become more and more uniform and flatter. The flatter atoms indicate better spatial coherency and a better representative capability of the atoms, which further brings a better signal-and-noise separation performance.

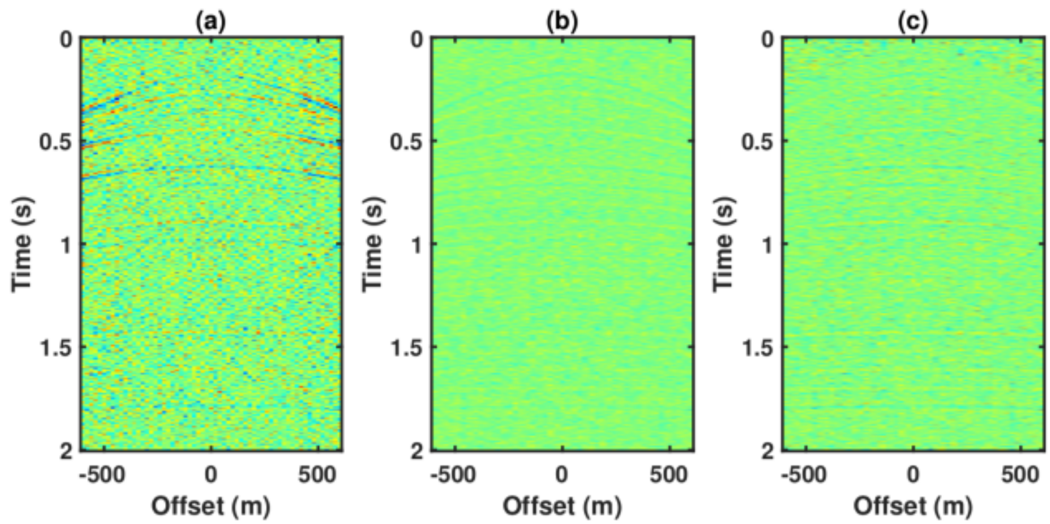


Fig. 6. Synthetic example. (a) Removed noise using the f-x deconvolution method. (b) Removed noise using the traditional DL method. (c) Removed noise using the proposed method.

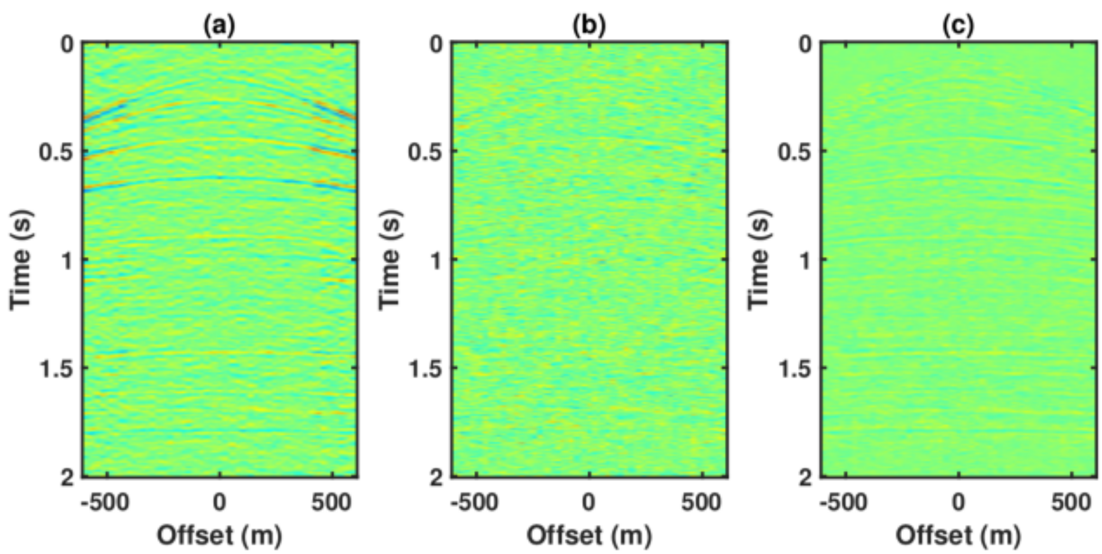
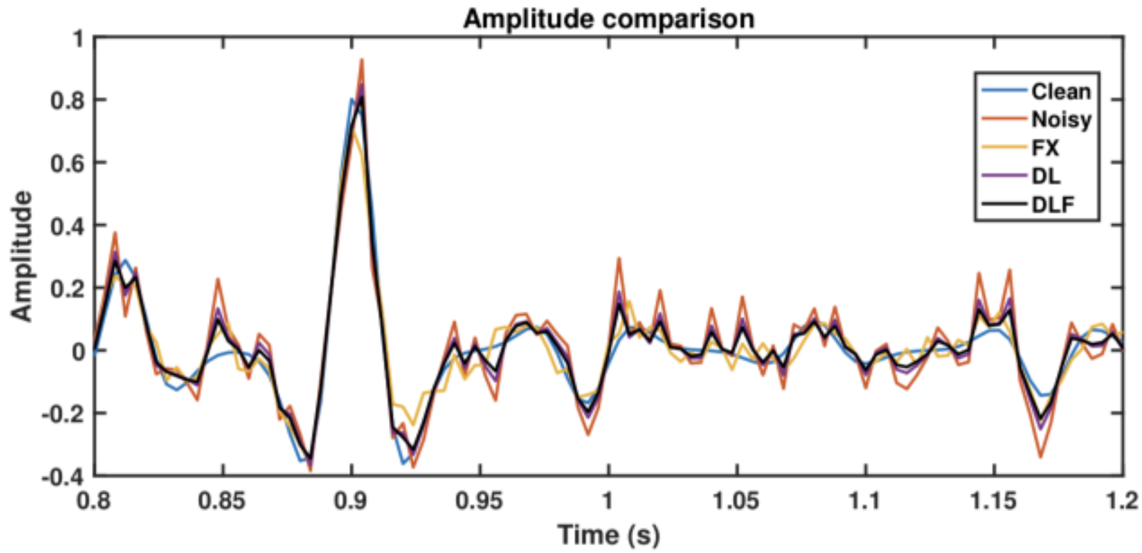
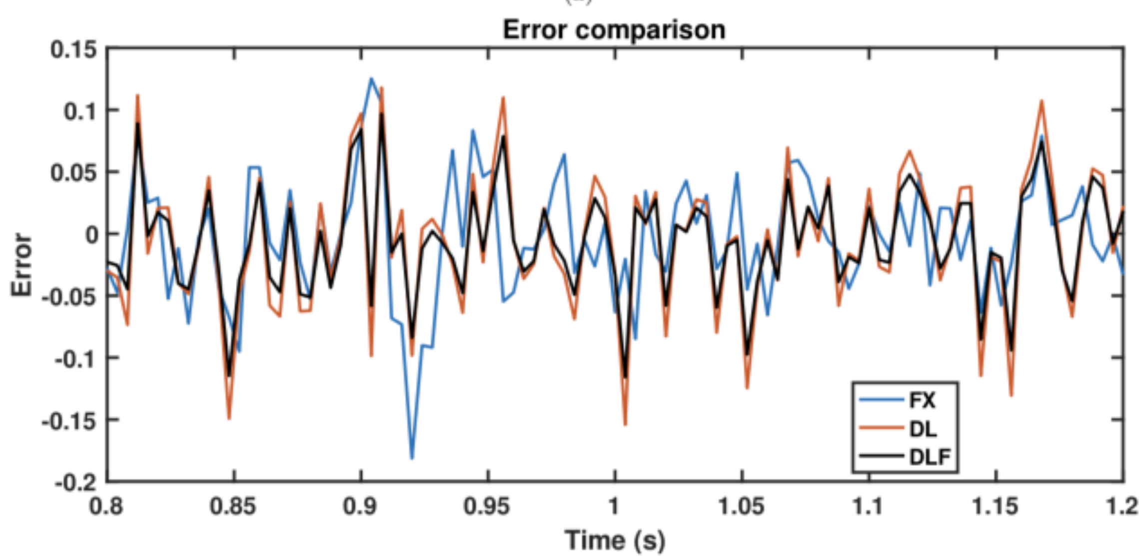


Fig. 7. Synthetic example. (a) Denoising error using the f-x deconvolution method. (b) Denoising error using the traditional DL method. (c) Denoising error using the proposed method.



(a)



(b)

Fig. 8. Amplitude and error comparison. (a) Trace-by-trace amplitude comparison. (b) Error comparison. Note that the proposed DLF method obtains the least error.

To test the effective performance of the proposed method on more complicated examples, we create another more realistic synthetic example shown in Fig. 10. In this example, there is an curving and highly oscillating events, which mimics a common-offset seismic gather. Fig. 10a shows the clean data. Fig. 10b shows the noisy data by adding some random noise. Fig. 10c shows the denoised data using the proposed dictionary learning method. Fig. 10d shows the removed noise corresponding to the proposed method. It is obvious that the proposed method can work robustly even in very complicated dataset.

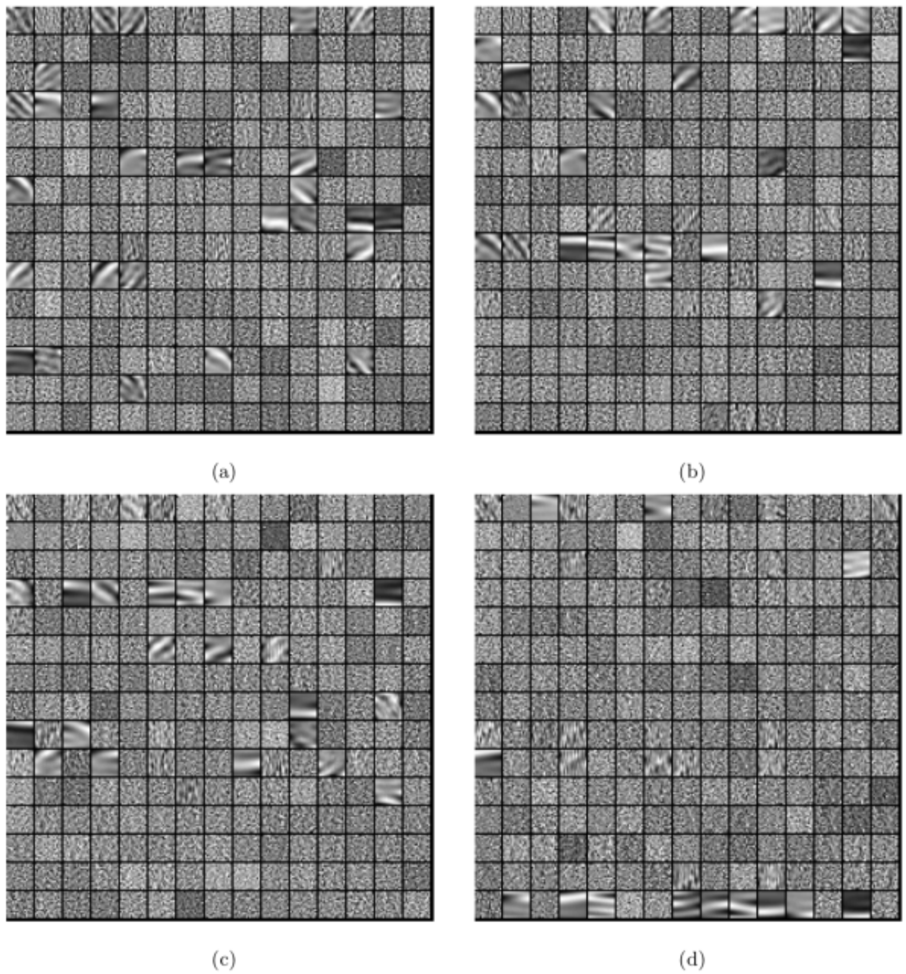


Fig. 9. Learned dictionaries when the local slope contains (a) 20%, (b) 10%, (c) 5%, (d) 2% error. Note that the as the slope becomes more and more accurate, the learned dictionary atoms become more and more uniform and flatter.

Next, we test the proposed method in a real data example. Fig. 11 shows a comparison between the clean and noisy data. Note that in this real data example, in order to quantitatively compare the performance for different methods, we select a common midpoint gather with the highest SNR. We use this data as the clean data and then add band-limited random noise to generate the noisy data (SNR = –1.31 dB). Figs. 11(a) and 11(b) show the clean and noisy field data. Figs. 11(c) and 11(d) show the two zoomed sections. We also compare the learned dictionaries for different methods in this example. Fig. 12a shows the basis functions for the DCT. Fig. 12b shows the learned dictionary using the traditional method. Fig. 12c shows the learned dictionary using the proposed method. We can obtain a similar conclusion from this comparison. The atoms shown in Fig. 12b have obviously large dip angle but the atoms obtained using the proposed method are relatively flatter. The flatter atoms have a better representation of the reflection seismic data because a fewer number of atoms are required to represent the data, which results in a sparser coefficient domain. The denoised comparison shown in Fig. 13 is consistent to the performance of learned dictionary. The proposed method obtains a much cleaner reconstruction of the reflection data, as shown in Fig. 13(c). In Fig. 13, the SNRs of the f-x deconvolution method, the traditional DL method, and the proposed method are 5.35 dB, 7.34 dB, 11.78 dB, respectively. Fig. 14 shows a comparison of noise removal, where we can see a large amount of coherent energy existing in Fig. 14(a), indicating a great damage to useful data using the f-x method. The signal damages between the traditional DL and the proposed DL are similar but the proposed method removes much more noise. Fig. 16a shows a detailed amplitude comparison for different data. It is salient that the proposed method is the closest to the clean data and causes the smallest error.

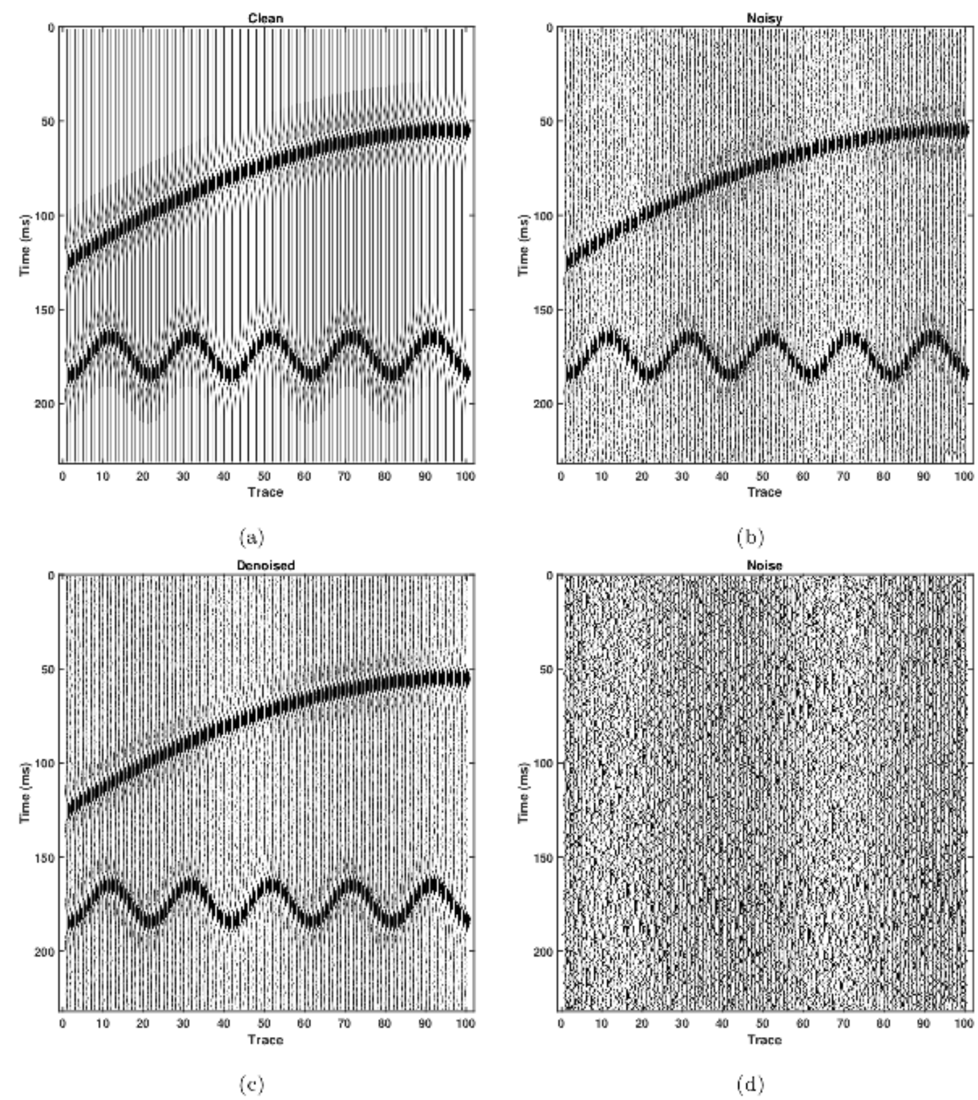


Fig. 10. Synthetic example with more complicated structures. (a) Clean data. (b) Noisy data. (c) Denoised data. (d) Removed noise.

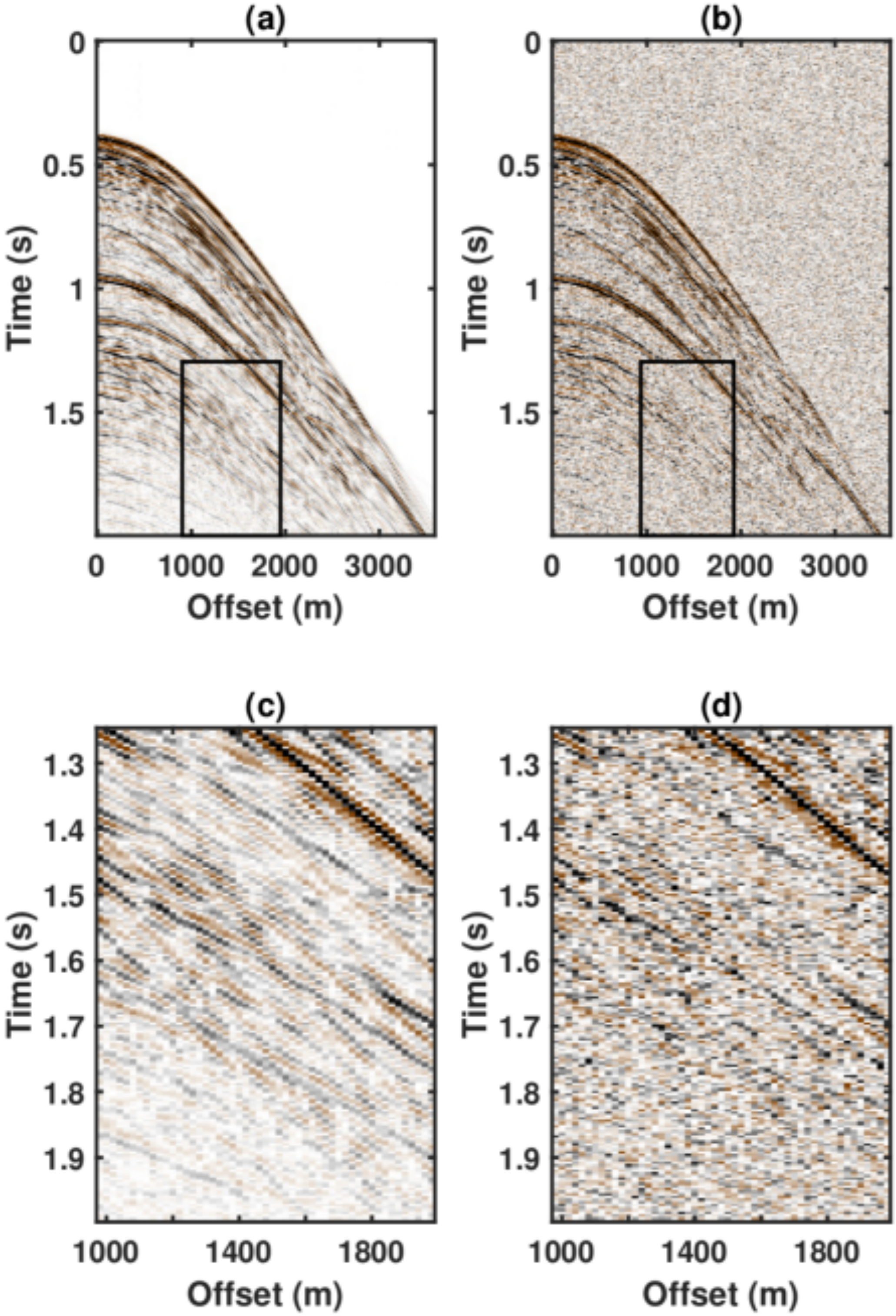


Fig. 11. Real data example. (a) Clean data. (b) Noisy data. (c) and (d) Zoomed areas from the frame boxes.

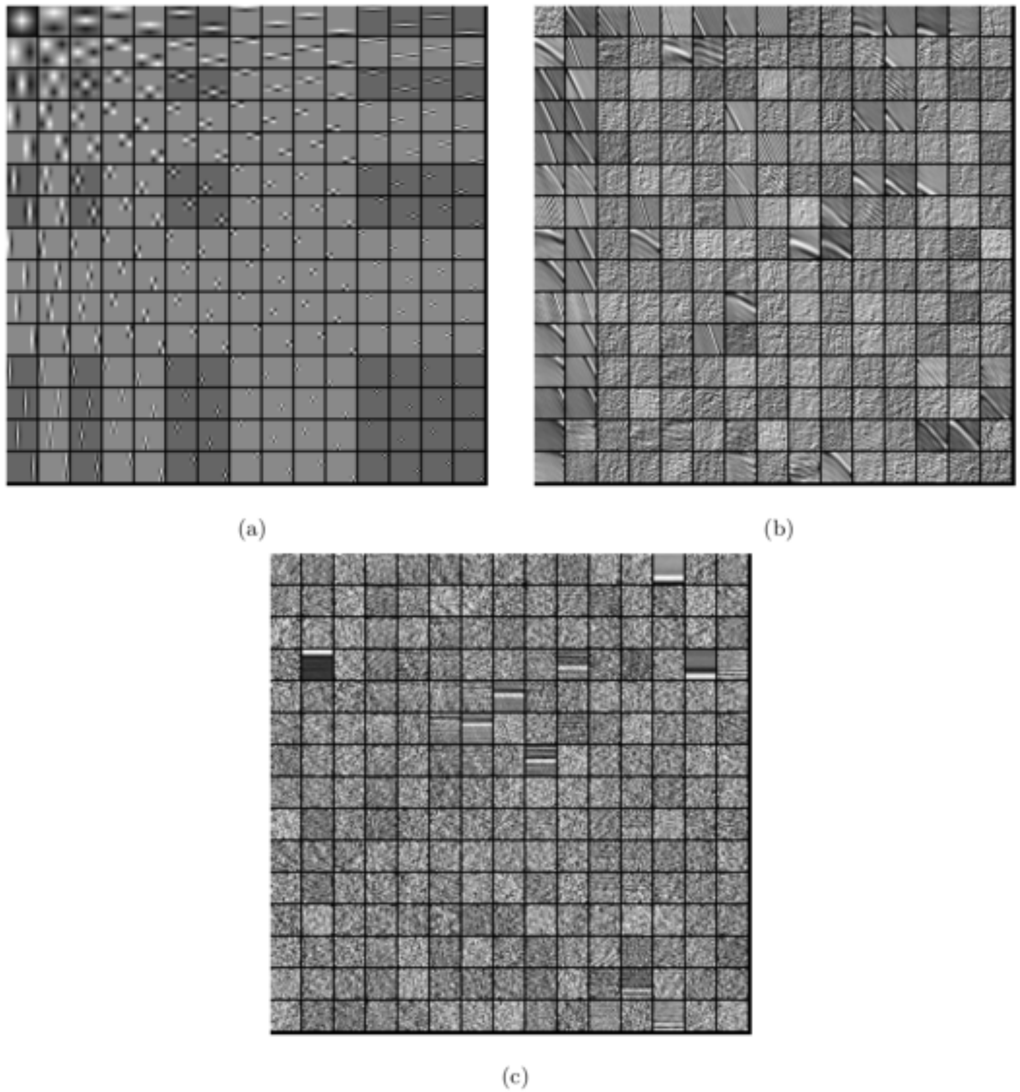


Fig. 12. Comparison between different dictionaries for the real data example. (a) Initial dictionary. (b) Traditionally learned dictionary. (c) Learned dictionary in the flattened dimension.

CONCLUSIONS

Sparse dictionary learning method can use a linear combination of dictionary (atom signals) and sparse coefficients to optimally and adaptively represent the useful reflection signals in seismic data. The spatial coherence affects the learned dictionary greatly in that worse spatial coherence results in steeply dipping atoms, which are more aliasing and less representative for the primary energy. We have introduced a flattening operator to prepare a better-structured dataset that is easier for dictionary learning. Learning the dictionary in the flattened dimension can obtain much flatter atoms and are better in sparsifing the re flections signals. The flatter atoms bring a better separability between signal and noise. We have used both synthetic and field data examples to demonstrate the superior performance of the proposed method compared with the f-x deconvolution method and the traditional dictionary learning method.

ACKNOWLEDGEMENTS

I am grateful to Min Bai for helpful discussions on the topic of dictionary learning.

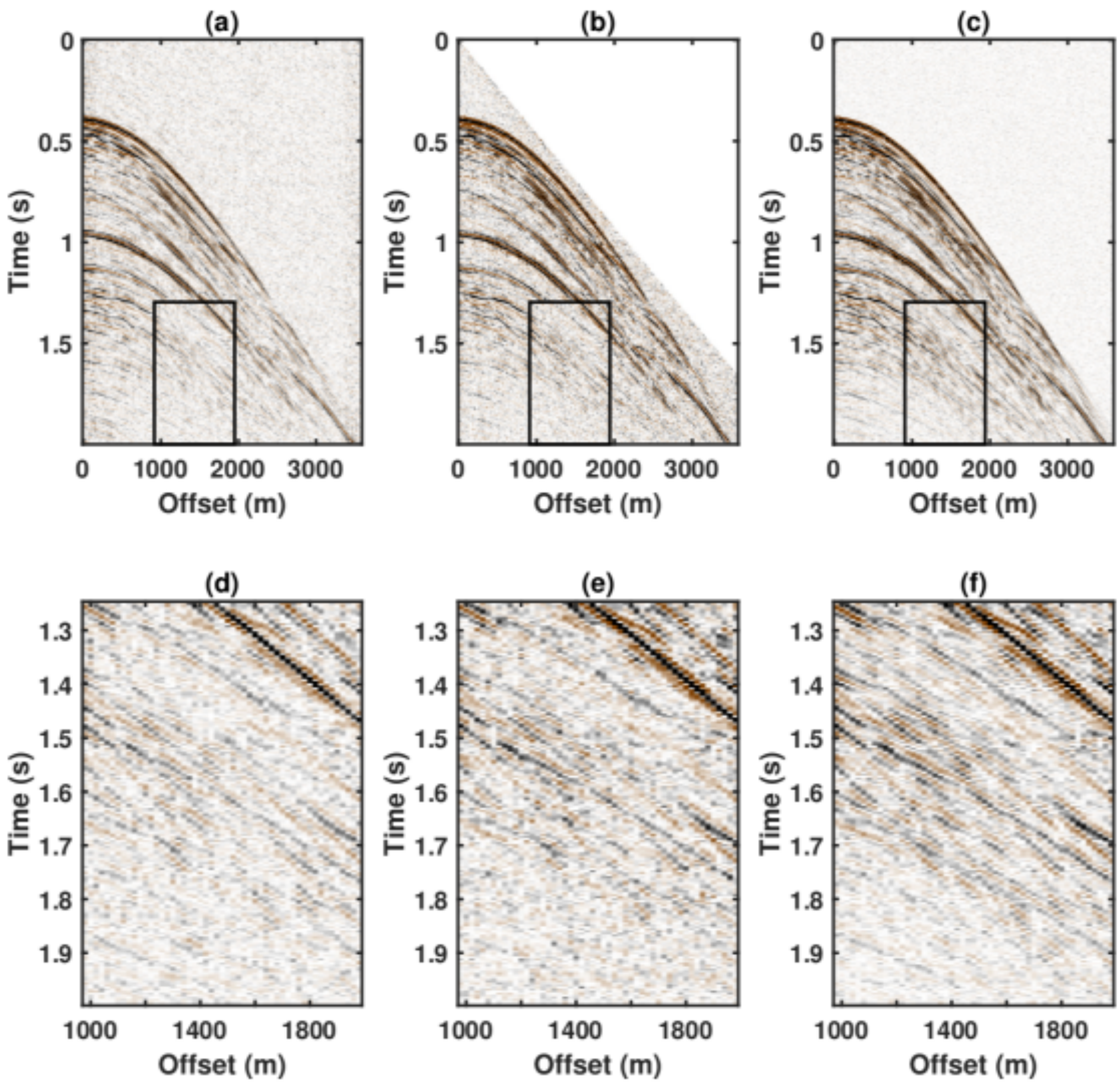


Fig. 13. Real data example. (a) Denoised data using the f-x deconvolution method. (b) Denoised data using the traditional DL method. (c) Denoised data using the proposed method. (d)-(f) Zoomed areas from the frame boxes.

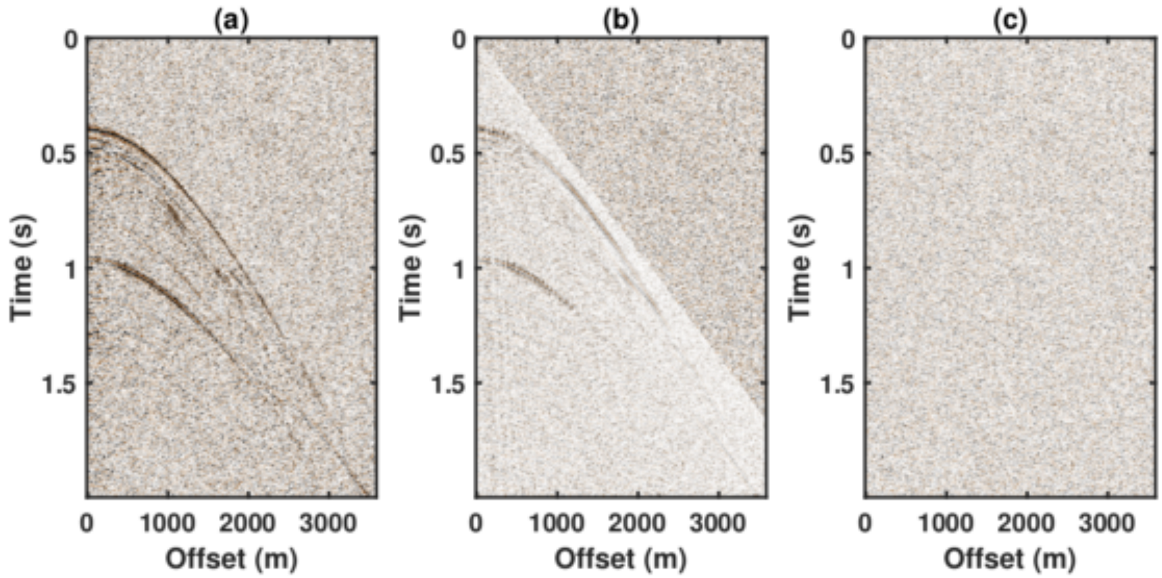


Fig. 14. Real data example. (a) Removed noise using the f-x deconvolution method. (b) Removed noise using the traditional DL method. (c) Removed noise using the proposed method.

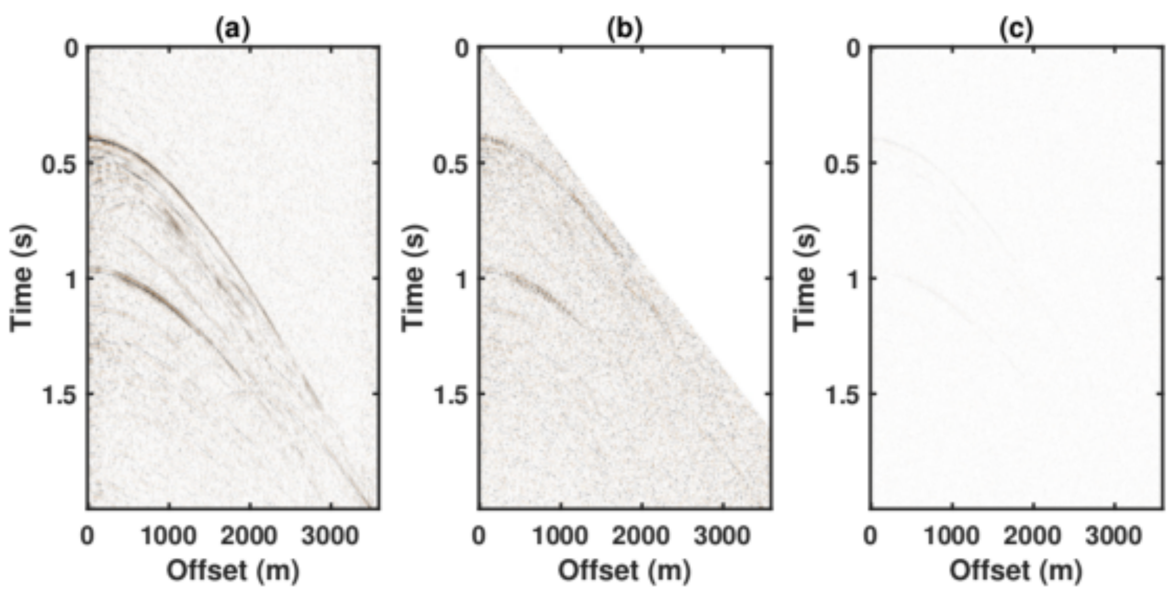
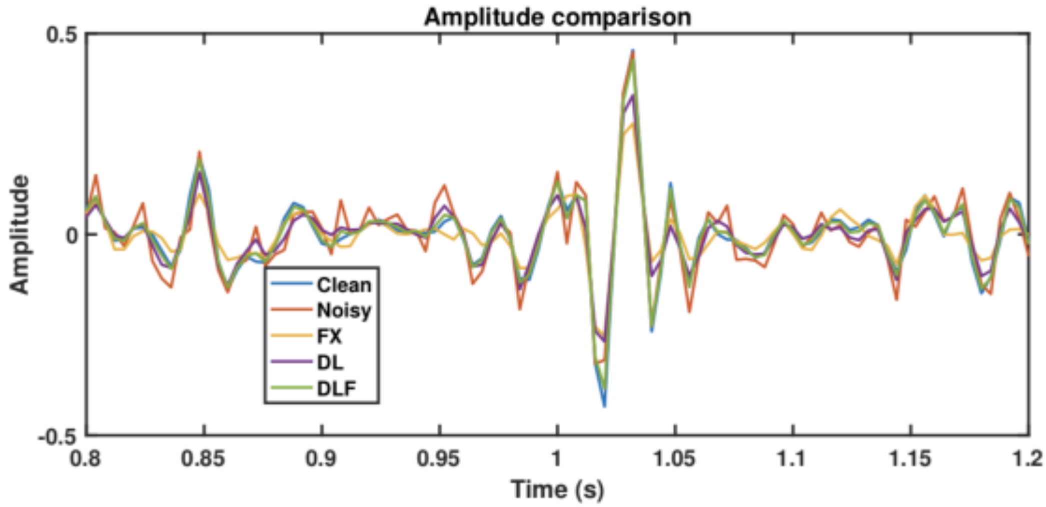
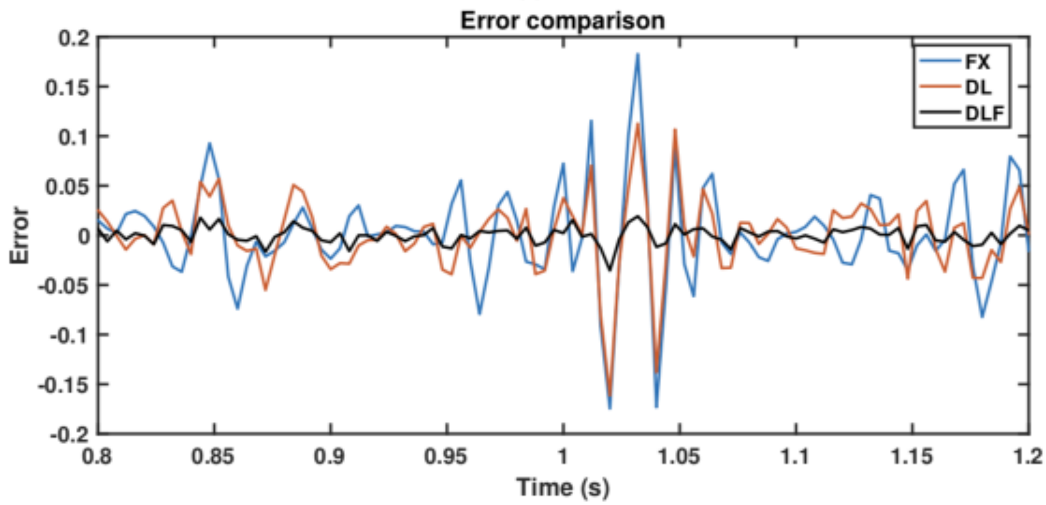


Fig. 15. Real data example.



(a)



(b)

Fig. 16. Amplitude and error comparison. (a) Trace-by-trace amplitude comparison. (b) Error comparison.

REFERENCES

Abma, R. and Claerbout, J., 1995. Lateral prediction for noise attenuation by t-x and f-x techniques. Geophysics, 60: 1887-1896.

Bai, M. and Wu, J., 2017. Efficient deblending using median filtering without correct normal moveout - with comparison on migrated images. J. Seismic Explor., 26: 455-479.

Bai, M., Wu, J., Xie, J. and Zhang, D., 2018a. Least-squares reverse time migration of blended data with low-rank constraint along structural direction. J. Seismic Explor., 27: 29-48.

Bai, M., Wu, J., Zu, S. and Chen, W., 2018b. A structural rank reduction operator for removing artifacts in least-squares reverse time migration. Comput. Geosci., 117: 9-20.

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

Chen, W., Chen, Y. and Cheng, Z., 2017a. Seismic time-frequency analysis using an improved empirical mode decomposition algorithm. J. Seismic Explor., 26: 367-380.

Chen, W., Chen, Y. and Liu, W., 2016a. Ground roll attenuation using improved complete ensemble empirical mode decomposition. J. Seismic Explor., 25: 485-495.

Chen, W., Xie, J., Zu, S., Gan, S. and Chen, Y., 2017b. Multiple reflections noise attenuation using adaptive randomized-order empirical mode decomposition. IEEE Geosci. Remote Sens. Lett., 14: 18-22.

Chen, W., Yuan, J., Chen, Y. and Gan, S., 2017c. Preparing the initial model for iterative deblending by median filtering. J. Seismic Explor., 26: 25-47.

Chen, W., Zhang, D. and Chen, Y., 2017d. Random noise reduction using a hybrid method based on ensemble empirical mode decomposition. J. Seismic Explor., 3: 227-249.

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., 2018a. Fast waveform detection for microseismic imaging using unsupervised machine learning. Geophys. J. Internat., 215: 1185-1199.

Chen, Y., 2018b, Non-stationary least-squares complex decomposition for microseismic noise attenuation. Geophys. J. Internat., 213: 1572-1585.

Chen, Y., Chen, H., Xiang, K. and Chen, X., 2017e. Preserving the discontinuities in least-squares reverse time migration of simultaneous-source data. Geophysics, 82(3): S185– S196.

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

Chen, Y. and Fomel, S., 2018. EMD-seislet transform. Geophysics, 83(1): A27-A32.

Chen, Y., Fomel, S. and Hu, J., 2014. Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: Geophysics, 79(5): V179-V189.

Chen, Y., Huang, W., Zhou, Y., Liu, W. and Zhang, D., 2018. Plane-wave orthogonal polynomial transform for amplitude-preserving noise attenuation. Geophys. J. Internat., 214: 2207-2223.

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

Chen, Y., J. Ma, and S. Fomel, 2016b, Double-sparsity dictionary for seismic noise attenuation. Geophysics, 81(2): V17-V30.

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

Chen, Y., Zhou, Y., Chen, W., Zu, S., Huang, W. and Zhang, D., 2017f,.Empirical low rank decomposition for seismic noise attenuation. IEEE Transact. Geosci. Remote Sens., 55: 4696-4711.

Chen, Y., Zu, S., Wang, Y. and Chen, X., 2019. Deblending of simultaneous-source data using a structure-oriented space-varying median filter. Geophys. J. Internat., 216: 1214-1232.

Cheng, J. and Sacchi, M., 2015. Separation and reconstruction of simultaneous source data via iterative rank reduction. Geophysics, 80(4): V57-V66.

Dragomiretskiy, K. and Zosso, D., 2014. Variational mode decomposition. IEEE Transact. Signal Process., 62: 531-544.

Fabien-Ouellet, G. Gloaguen, E. and Giroux, B., 2017. Time-domain seismic modeling in viscoelastic media for full waveform inversion on heterogeneous computing platforms with opencl (**????**). Comput. Geosci., 100: 142-155.

Fomel, S., 2002. Application of plane-wave destruction filters. Geophysics, 67: 1946-1960.

Fomel, S., 2013. Seismic data decomposition into spectral components using regularized non-stationary autoregression. Geophysics, 78: O69-O76.

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

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

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

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

Gan, S., Chen, Y., Wang, S., Chen, X., Huang, W. and Chen, H., 2016e. 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.

Gao, J., Sacchi, M.D. and Chen, X., 2013. A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions. Geophysics, 78(1,): V21-V30.

Garcıa-Yeguas, A., Ledo, J., Pina-Varas, P., Prudencio, J., Queralt, P., Marcuello, A., Ibanez, J.M., Benjumea, B., Sanchez-Alzola, A. and Perez, N., 2017. A 3D joint interpretation of magnetotelluric and seismic tomographic models: The case of the volcanic island of Tenerife. Comput. Geosci, 109: 95-105.

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.

Li, C., Liu, G. and Li, Y., 2017a,. A practical implementation of 3d TTI reverse time migration with multi-gpus. Comput. Geosci., 102: 68-78.

Li, C. and Zhang, F., 2017. Matching pursuit parallel decomposition of seismic data. Comput. Geosci., 104: 54-61.

Li, H., Tuo, X., Shen, T., Henderson, M.J., Courtois, J. and Yan, M., 2017b. An improved lossless group compression algorithm for seismic data in seg-y and miniseed file formats. Comput. Geosci., 100: 41-45.

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

Li, H., Wang, R., Cao, S., Chen, Y., Tian, N. and Chen, X., 2016b. Weak signal detection using multiscale morphology in microseismic monitoring. J. Appl. Geophys., 133: 39-49.

Liu, S., Yang, D. and Ma, J., 2017a. A modified symplectic prk scheme for seismic wave modeling. Comput. Geosci., 99: 28-36.

Liu, W., Cao, S. and Chen, Y., 2016a. Applications of variational mode decomposition in seismic time-frequency analysis. Geophysics, 81(5): V365-V378.

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

Liu, W., Cao, S., Jin, Z., Wang, Z. and Chen, Y., 2018. A novel hydrocarbon detection approach via high-resolution frequency-dependent AVO inversion based on variational mode decomposition. IEEE Transact. Geosci. Remote Sens., 56: 2007-2024.

Liu, W., Cao, S. and Wang, Z., 2017b. Application of variational mode decomposition to seismic random noise reduction. J. Geophys. Engineer., 14: 888.

Liu, W., Cao, S., Wang, Z., Kong, X. and Chen, Y., 2017c. Spectral decomposition for hydrocarbon detection based on VMD and Teager-Kaiser energy. IEEE Geosci. Remote Sens. Lett., 14: 539-543.

Liu, Y., Fomel, S. and Liu, G., 2010. Nonlinear structure-enhancing filtering using plane-wave prediction. Geophys. Prosp., 58: 415-427.

Lv, H. and Bai, M., 2018. Learning dictionary in the approximately flattened structure domain. J. Appl. Geophys., 159: 522-531.

Mayne, W.H., 1962. Common reflection point horizontal data stacking techniques. Geophysics, 27: 927-938.

Rastogi, R., Londhe, A., Srivastava, A., Sirasala, K.M. and Khonde, K., 2017. 3D Kirchhoff depth migration algorithm: A new scalable approach for parallelization on multicore CPU based cluster. Comput. Geosci., 100: 67-75.

Romano, Y. and Elad, M., 2013. Improving K-VSD denoising by post-processing its method-noise. Image Processing (ICIP), 20th IEEE Internat. Conf.: 435-439.

Rubinstein, R., Zibulevsky, M. and Elad, M., 2008. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. Tech. Report.

Rubinstein, R., Zibulevsky, M. and Elad, M., 2010. Double sparsity: learning sparse dictionaries for sparse signal approximation. IEEE Transact. Signal Process., 58: 1553-1564.

Siahsar, M.A.N., Abolghasemi, V. and Chen, Y., 2017a. Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning. Signal Process., 141: 309-321.

Siahsar, M.A.N., Gholtashi, S., Kahoo, A.R., Chen, W. and Chen, Y., 2017b. Data-driven multi-task sparse dictionary learning for noise attenuation of 3D seismic data. Geophysics, 82(6): V385-V396.

Siahsar, M.A.N., Gholtashi, S., Kahoo, A.R., Marvi, H. and Ahmadifard, A., 2016. Sparse time-frequency representation for seismic noise reduction using low-rank and sparse decomposition. Geophysics, 81: V117-V124.

Siahsar, M.A.N., S. Gholtashi, E. Olyaei, W. Chen, and Y. Chen, 2017c, Simultaneous denoising and interpolation of 3D seismic data via damped data-driven optimal singular value shrinkage. IEEE Geosci. Remote Sens. Lett., 14: 1086–1090.

Song, C., Liu, Z., Wang, Y., Li, X. and Hu, G., 2017. Multi-waveform classification for seismic facies analysis. Comput. Geosci., 101: 1-9.

Vautard, R., Yiou, P. and Ghil, M., 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. Physica D: Nonlinear Phenom., 58: 95-126.

Wang, Y., Zhou, H., Zu, S., Mao, W. and Chen, Y., 2017a. Three-operator proximal splitting scheme for 3D seismic data reconstruction. IEEE Geosci. Remote Sens. Lett., 14: 1830-1834.

Wang, Y., Zhou, H., Chen, H. and Chen, Y., 2017b. Adaptive stabilization for Q-compensated reverse time migration. Geophysics, 83: S15-S32.

Wang, Y., Zhou, H., Ma, X. and Chen, Y., 2018. L1-2 minimization for exact and stable seismic attenuation compensation, Geophys. J. Internat., 213: 1629-1646.

Wu, G., Fomel, S. and Chen, Y., 2018. Data-driven time-frequency analysis of seismic data using non-stationary prony method. Geophys. Prosp., 66: 85-97.

Wu, J. and Bai, M., 2018a. Incoherent dictionary learning for reducing crosstalk noise in l east-squares reverse time migration. Comput. Geosci., 114: 11-21.

Wu, J. and Bai, M., 2018b. Stacking seismic data based on principal component analysis. J. Seismic Explor., 27: 331-348.

Wu, Z. and Huang, N.E., 2009. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis, 1: 1-41.

Xie, J., Chen, W., Zhang, D., Zu, S. and Chen, Y., 2017. Application of principal component analysis in weighted stacking of seismic data. IEEE Geosci. Remote Sens. Lett., 14: 1213-1217.

Xie, J., Di, B., Schmitt, D., Wei, J. and Chen, Y., 2018. Estimation of δ and C_{13} of organic-rich shale from laser ultrasonic technique (LUT) measurement. Geophysis, 83(4): C137-C152.

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

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., Wu, J., Chen, Y., Gan, S. and Zhong, W., 2015. An efficient and effective common reflection surface stacking approach using local similarity and plane-wave flattening. J. Appl. Geophys., 117: 67-72.

Yang, X. and Zhu, P., 2017. Stochastic seismic inversion based on an improved local gradual deformation method. Comput. Geosci., 109: 75-86.

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

Zhang, D., Zhou, Y., Chen, H., Chen, W., Zu, S. and Chen, Y., 2017. Hybrid rank-sparsity constraint model for simultaneous reconstruction and denoising of 3D seismic data. Geophysics, 82(5): V351-V367.

Zhou, Y., Gao, J., Chen, W. and Frossard, P., 2016. Seismic simultaneous source separation via patchwise sparse representation. IEEE Transact. Geosci. Remote Sens., 54: 5271-5284.

Zhou, Y. and Han, W., 2018. Multiples attenuation in the presence of blending noise. J. Seismic Explor., 27: 69-88.

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

Zhou, Y., Li, S., Zhang, D. and Chen, Y., 2018. Seismic noise attenuation using an online subspace tracking algorithm. Geophys. J. Internat., 212: 1072-1097.

Zhou, Y., Shi, C., Chen, H., Xie, J., Wu, G. and Chen, Y., 2017b. Spike-like blending noise attenuation using structural low-rank decomposition. IEEE Geosci. Remote Sens. Lett., 14: 1633-1637.

Zhou, Y. and S. Zhang, S., 2017. Robust noise attenuation based on nuclear norm minimization and a trace prediction strategy. J. Appl. Geophys., 147: 52-67.

Zhu, L., Liu, E. and Mcclellan, J.H., 2014. Seismic data denoising through multiscale and sparsity-promoting dictionary learning. Geophysics, 80: WD45-WD57.

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

Zu, S., Zhou, H., Chen, Y., Pan, X., Gan, S. and Zhang, D., 2016. Interpolating big gaps using inversion with slope constraint. IEEE Geosci. Remote Sens. Lett., 13: 1369-1373.

Zu, S., Zhou, H., Mao, W., Zhang, D., Li, C., Pan, X. and Chen, Y., 2017. Iterative deblending of simultaneous-source data using a coherency-pass shaping operator. Geophys. J. Internat., 211: 541-557.

Zu, S., Zhou, H., Chen, H., Gong, F., Li, C., Zheng, H. and Chen, Y., 2017. Two field trials for deblending of simultaneous source: why we failed and why we succeeded? J. Appl. Geophys., 143: 182-194.

Zu, S., Zhou, H., Chen, H., Li, Q., Zhang, Q. and Chen, Y., 2017. Shot-domain deblending using least-squares inversion. Geophysics, 82: V241-V256.