SEISMIC RANDOM NOISE ATTENUATION USING MULTI-SCALE SPARSE DICTIONARY LEARNING

JINWEI FANG[1], LIANG ZHANG[2]*, HUI ZHOU[3], SHENGDONG LIU[1], BO WANG[1] and WENJIE CHEN[4]

[1]*State Key Laboratory of Deep Geomechanics & Underground Engineering, School of Resource and Earth Science, School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, P.R. China.*
[2]*Institute of Geosciences and Info-Physics, Central South University, Changsha, Hunan, 410083, P.R. China. 175050483@qq.com.*
[3]*State Key Laboratory of Petroleum Resources and Prospecting, CNPC Key Lab of Geophysical Exploration, China University of Petroleum, Changping 102249, Beijing, P.R. China.*
[4]*College of Information Technology, Guangxi Police College, Nanning, Guangxi 530028, P.R. China.*

(Received August 27, 2021; revised version accepted February 1, 2022)

ABSTRACT

Fang, J.W., Zhang, L., Zhou, H., Liu, Z.D., Wang, B. and Chen, W.J., 2022. Seismic random noise attenuation using multi-scale sparse dictionary learning. *Journal of Seismic Exploration*, 31: 177-202.

Seismic data contain random noise, which affects data processing and interpretation and possibly limits the use of seismic data in parameter building and attribute prediction. To effectively remove such noise, we develop a denoising workflow based on multi-scale sparse dictionary learning. The multi-scale sparse dictionary learning method decreases the complexity of data by two approaches. One is seismic data-sparse representation from the data domain to the wavelet domain by wavelet bases. The other is denoising by sparse dictionary learning in a certain frequency band without the noise effect from other frequency bands. Meanwhile, the wavelet transform can also reduce the dimension of the data, which ensures the computational efficiency of the proposed method. After analyzing the effects of sparse dictionary learning parameters on seismic data denoising, we test the proposed method on two synthetic and two field datasets. We learn from the examples that our approach can effectively recover signals from simulated and real noisy data, as well as time-variant noisy data. Compared with the sparse dictionary learning, K-singular value decomposition dictionary learning, and double-sparsity dictionary (DSD), our method can obtain the best-denoised result with less effective signals leaking in noise sections.

KEY WORDS: sparse dictionary learning, wavelet transform, multi-scale denoising.

0963-0651/22/$5.00 © 2022 Geophysical Press Ltd.

INTRODUCTION

Sparse representations aim to obtain the sparse linear coefficients of original signals by given or learned basic signals and play important roles in seismic data denoising, interpolation, migration, and image applications. Random noise occurs in many stages of seismic acquisition and processing, which conceals the information of weak signals and reduces the signal-to-noise ratio (S/N) of seismic data. Except for direct filtering (Bednar, 1983; Yilmaz, 2001), data transforms, e.g., wavelet transform (WT; Ioup and Ioup, 1998; Zhang and Ulrych, 2003), curvelet transform (Hennenfent and Herrmann, 2006; Wang et al., 2008; Naghizadeh and Sacchi, 2010; Neelamani et al., 2010), seislet transform (Fomel et al., 2013), and Radon transform (Zhang et al., 2020), have been widely used as sparsity-promoting transforms to remove such noise and improve the quality of seismic data. Because signals and noise have different characteristics in various domains, the data transform from one domain to another can implement the extraction of effective information.

In recent years, sparse representations based on dictionary learning have been proposed to represent the most useful information from noisy data. Dictionary learning can be divided into two general categories: analytic atom approach and learning-based atom approach. Some standard dictionaries, such as discrete cosine transform (DCT) dictionary, Harr dictionary, Gabor dictionary, and so on, are selected for analytic atom signals as a fixed basis. However, the fixed dictionary may not obtain the best sparse representation of original signals in some cases because the fixed dictionary cannot adapt to the characteristics of the signals. Therefore, a learning-based dictionary approach can adaptively find the required sparse basis by training. One of popular learning-based K-singular value decomposition (K-SVD) dictionary learning methods offers refined dictionary that adapts the structure of the data (Aharon et al., 2006; Rubinstein et al., 2009; Beckouche and Ma, 2014; Liu et al., 2017; Lv and Bai, 2018; Zu et al., 2018). In order to enhance the K-SVD, Rubinstein et al. (2009) believed that dictionary atoms themselves might have some underlying sparse structure over a more fundamental dictionary and proposed a double learning sparse dictionary. However, the relatively large computation amount limits its application to big seismic data, especially for three-dimensional (3D) seismic data. The computation amount is caused by updating the dictionary one column by one column and operating SVD in each column. Recently, the data-driven tight frame (DDTF) method was proposed by Cai et al. (Cai et al., 2014; Liang et al., 2014; Yu et al., 2015). The different point is that this method is constrained by a tight frame, and the DDTF dictionary is updated by a single SVD; thus, the computation cost of this method makes it far more practical than K-SVD. Owing to its higher efficiency than traditional dictionary learning methods, Yu et al. (2016) proposed a more adaptive Monte Carlo DDTF in patch selection in seismic data recovery, and Zu et al. (2019) adopted the dip patch selection in dictionary learning. Chen (2017) and Feng (2021) also proposed a fast dictionary learning approach based on the sequential generalized K-means algorithm to accelerate multi-dimensional seismic data denoising.

Although a learning-based dictionary can be more adaptive than fixed-basis transforms, the complexity of seismic data and no prior-constraint structural information involved in the dictionary's construction are also challenging problems. To improve the robustness of the learning-based dictionary when applied to seismic data denoising, Ophir et al. (2011) introduced the idea of a multi-scale K-SVD dictionary learning algorithm based on wavelet transform. Zhu et al. (2015) also proposed a method of seismic data denoising through multi-scale and sparsity-promoting K-SVD dictionary learning. Besides, Chen et al. (2016) introduced a double-sparsity dictionary method (DSD) through seislet transform-based DDTF. A coherence-constrained dictionary learning method has been developed by Turquais et al. (2017) to ease the prior information on the noise variance.

Because of the effects of attenuation, geometric divergence, and frequency-dispersion, the acquired seismic data are time-variant, i.e., amplitude attenuation and phase distortion during seismic wave propagation. Compared with the global Fourier transform, wavelet transform can represent the local features of signals in the time and frequency domain. What's more, the wavelet transform is invertible, and there are no truncation artifacts displayed in scales. In this paper, we plan to develop a multi-scale sparse dictionary learning method based on wavelet transform. The proposed method follows the main idea of Ophir et al. (2011) and Zhu et al. (2015). The different point is that our approach employs more effective sparse dictionary learning rather than K-SVD or K-SVD-based method in the wavelet domain (frequency domain) to deal with time-variant seismic data. By wavelet transform, we can focus on the signal analysis of a particular seismic frequency band rather than directly using image denoising. What's more, we also discuss how to choose suitable parameters involved in sparse dictionary learning. In the proposed method, the wavelet transform base not only provides good time-frequency analysis coefficients, but also decreases the complexity of original signals, and the sparse dictionary learning provides the multi-level sparsity on WT coefficients. We analyze its denoising ability in terms of the S/N and signal events, as well as its efficiency.

This paper is organized as follows. After this section, we introduce the basic knowledge of sparse dictionary learning and multi-scale waveform transform. Then we give the workflow of seismic data denoising using multi-scale sparse dictionary learning. Denoising experiments on two synthetic and two real seismic data sets are provided in the next section, followed by the Discussion and Conclusions sections.

THEORY

Sparse dictionary learning

In this section, we briefly give the introduction of sparse dictionary learning. Readers may refer to literature (Chen et al., 2001; Daubechies et al., 2003; Cai et al., 2014) for some terminologies, such as tight frame, atoms. To obtain a good filter matrix D, which is constrained by a tight frame such that $D^T D = I$ for given seismic data *d*, we solve the following optimization problem,

$$\arg\min_{m,D} \left\| D^T d - m \right\|_F^2 + \alpha \left\| m \right\|_0, \text{s.t.} D^T D = I, \tag{1}$$

where m is the sparse representation of data *d,* $\|\cdot\|_F$ and $\|\cdot\|_0$ denotes the Frobenius and L_0-norm, and superscript T stands for the adjoint operator, α is a coefficient to balance the fidelity term and the regularization term. The first term in eq. (1) is fidelity and the second term is regularization. The decomposition operation can be expressed as $m = D^T d$, and the reconstruction operation can be written as $d = Dm$. Usually, we select the enough patches of *d* (Zhan and Dong, 2016) as the input data, and solve eq.(1) by the following two steps:

Step 1: Dictionary updating. By giving an initial canonical filter matrix D, we aim at seeking for the tight frame to maximize its sparsity. The data-driven filter matrix D is trained by minimizing

$$\arg\min_{D} \left\| D^T d - m \right\|_F^2 + \alpha \left\| m \right\|_0, \text{s.t.} D^T D = I. \tag{2}$$

An efficient algorithm to solve eq. (2) is a thresholding denoising method under a tight frame (Schnemann, 1966; Cai et al., 2014; Bao et al., 2015; Sezer et al., 2015.). D and *m* can be efficiently computed by

$$\begin{cases} D^{\gamma+1} = XY^T, \\ m^{\gamma+1} = \tau_\lambda\left(\left(D^{\gamma+1}\right)^T d\right), \end{cases} \tag{3}$$

where X and Y are obtained by operating SVD for $d\left(m^\gamma\right)^T$, i.e. $d\left(m^\gamma\right)^T = X\Delta Y$, Δ is a diagonal matrix, γ denotes the γ-th iteration, and τ_λ is the hard-thresholding operator given by

$$\tau_\lambda\left(P_{p,q}\right) = \begin{cases} 0, \text{if} \left|P_{p,q}\right| < \lambda, \\ P_{p,q}, \text{otherwise}, \end{cases} \tag{4}$$

where λ is closely related to the noise variance and the desired sparsity degree of the image, and $P = D^T d$. One thing to be mentioned is that eq. (3) can update all elements of D in one SVD operation.

Step 2: Sparse coding by the learned dictionary. We fix the tight frame D and solve the following optimization problem,

$$\arg\min_m \|d - Dm\|_F^2 + \alpha\|m\|_0. \tag{5}$$

As we know, the l_0-norm problem of optimization is an NP-hard problem and cannot be solved directly. We solve eq. (5) by classical orthogonal matching pursuit (OMP) (Tropp, 2007). We need to set the number N of maximal sparse solutions, i.e., sparsity level, for a specific example. We recover the denoised data approximately by

$$\tilde{d} \approx Dm, \tag{6}$$

and we reshape patch data $\tilde{d}$ to the original data size.

A multi-scale approach based on wavelet transform

Wavelet transform has been widely used in seismic data noise attenuation (Ioup and Ioup, 1998; Miao and Cheadle, 1998; Yuan and Simons, 2014). We use two sets of scaling functions (Jawerth and Sweldens, 1994), $\tilde{\phi}_k^j$ for the analysis and ϕ_k^j for the synthesis, as well as two sets of wavelet functions (Daubechies, 1992), $\tilde{\psi}_k^j$ for analysis and ψ_k^j for synthesis, where j denotes a particular scale and k denotes translation in time (Daubechies, 1992; Mallat, 1999). Then a signal $s(t)$ from a trace of data d can be divided into an approximation coefficient of a_k^J and a series of detail coefficients of d_k^j given by,

$$\begin{cases} a_k^J = \langle s(t), \tilde{\phi}_k^j \rangle, \\ d_k^j = \langle s(t), \tilde{\psi}_k^j \rangle, \end{cases} \tag{7}$$

where J denotes the maximal scale for decomposing the signal, and $\langle , \rangle$ represents the inner product operation. The signal $s(t)$ can be perfectly reconstructed by summing over all translators k and scales $j = 1,2,...,J$ as

$$s(t) = \sum_k a_k^J \phi_k^J + \sum_{j=1}^{J} \sum_k d_k^j \psi_k^j. \tag{8}$$

Regarding discrete seismic signals, supposing that $c_0(n)$ is the discrete signal of $s(t)$, we can use filter bank (Strang and Nguyen, 1996; Mallat, 1999) to implement eqs. (7) and (8). The analysis band has two filters, low-pass $\tilde{g}$ and high-pass $\tilde{h}$, which correspond to analysis scaling function $\tilde{\phi}_k^j$ and wavelet function $\tilde{\psi}_k^j$, respectively. Similarly, the synthesis band also has two filters, low-pass g and high-pass h, which correspond to the synthesis ϕ_k^j scaling function and ψ_k^j wavelet function, respectively. Then the decomposition of signals can be formed as

$$\begin{cases} c_{j+1}(k) = \sum_k \tilde{g}_{k-2n} c_j(n), \\ d_{j+1}(k) = \sum_k \tilde{h}_{k-2n} c_j(n), \end{cases} \tag{9}$$

where scales $j = 1,2,...,J-1$. Then the reconstruction of signals can be formed as

$$c_j(n) = \sum_k g_{n-2k} c_{j+1}(k) + \sum_k h_{n-2k} d_{j+1}(k), \tag{10}$$

where scales $j = J-1, J,...,0$, $k = 0,1,...,S_j - 1$ and the S_j stands for the length of j-th scale signals. Different wavelet functions and scaling functions correspond to different combinations of filters.

Multi-scale sparse dictionary learning

Based on the previous knowledge, the multi-scale sparse dictionary learning is given as follows. For seismic data *d,* we use the wavelet transform to obtain the coefficients *c* by

$$c = Wd, \tag{11}$$

where $c = [c_1, c_2,..., c_J]$ and W stands for the forward wavelet transform which is described in eq. (9). After dividing the seismic data into different scale sparse coefficients with different frequency bands, we first remove the sparse scale coefficients $c' = [c_1, c_2,..., c_{J'}]$ beyond the maximum useful frequency, then perform the dictionary learning on remaining sparse coefficients $c'' = [c_{J'+1}, c_{J'+2},..., c_J]$ by

$$\arg\min_{m_j, D_j} \left\| D_j^T c_j - m_j \right\|_F^2 + \alpha \left\| m_j \right\|_0, \text{s.t.} D_j^T D_j = I, \tag{12}$$

where $c_j \subseteq c''$ and $c = c' + c''$. m_j is the dictionary sparse coefficient corresponding to c_j. The denoised wavelet coefficients $\tilde{c}_j$ with respect to c'' can be obtained by

$$\tilde{c}_j = D_j m_j. \tag{13}$$

Finally, the data after removing noise by multi-scale sparse dictionary learning can be resolved by wavelet reconstruction

$$\tilde{d} = \dddot{W} \tilde{c}, \tag{14}$$

where $\ddot{W}$ stands for the inverse wavelet transform which is described in eq.(10), and $\tilde{d}$ is the noise removed seismic data.

There are three potential advantages of the multi-scale sparse dictionary learning. One is that the wavelet coefficients have more focused energy than that the original signal. Consequently, the dictionary learning process of multi-scale sparse dictionary learning performs well on the low-complexity coefficient data. The second is that WT can act as a useful time-frequency analysis tool to handle with the seismic time series. As a result, we can denoise by dictionary learning in a specific frequency band and avoid the effect of the noise of other frequency bands. The last is that WT can reduce the data size, specifically, each WT operation can decrease data size to about half of its input, which leads to train dictionary on a small dataset. As a result, the multi-scale sparse dictionary learning can ensure the efficiency by multi-scale wavelet transform.

The choices of parameters for the proposed algorithm

The ideal wavelets for seismic data decomposition possess two main points. One is that seismic data achieve a sparser representation under the wavelet expansion. The other is that the wavelets have proper vanishing moments because decomposition using the wavelets with large vanishing moments costs more than those using fewer vanishing moments (Yuan and Simons, 2014). Besides these, the commonly used Daubechies wavelet family can use the same wavelet function and scale function both for the wavelet analysis and synthesis. In this paper, we employ the wavelet db4 from Daubechies wavelet for WT operation, and the filters corresponding to wavelet function and scale function are $\widetilde{h}$ =[-0.0106, 0.0329, 0.0308, -0.1870, -0.0280, 0.6309, 0.7148, 0.2304] and $\widetilde{g}$ =[-0.2304, 0.7148, -0.6309, -0.0280, 0.1870, 0.0308, -0.0329, -0.0106], respectively.

There are five parameters to be appropriately set for an excellent denoised result. The parameters are the maximal decomposition scale *J*, the number *N* of sparsity levels, hard-thresholding operator τ_λ, filter size $L \times L$, and the number *K* of dictionary updates. The maximal decomposition scale *J* depends on the temporal sampling intervals, record time, and the signal's dominant frequency. The parameter *N* (Lopes, 2016) has a close relation to the structure in seismic data, i.e., it is no less than the number of main events in the data if there exists a good transform base. The hard-thresholding operator τ_λ balances the initial dictionary and the trained dictionary, and it has a close relation to the variance σ of noise. We will analyze this in the example section. Generally, the larger *L* in filter size obtains better-denoised results. Still, the cost is more expensive than that with a small *L*. Regarding parameter *K,* the quality of denoised results obviously improves at the beginning and gradually becomes stable with the increase of *K* since the learned dictionary gradually captures enough features of data.

NUMERICAL EXAMPLES

In this section, two synthetic and two field data examples are chosen to test the denoising performance of the proposed method. All noisy data are filtered within 5-80 Hz to focus on the seismic frequency band. The noisy data is expressed as

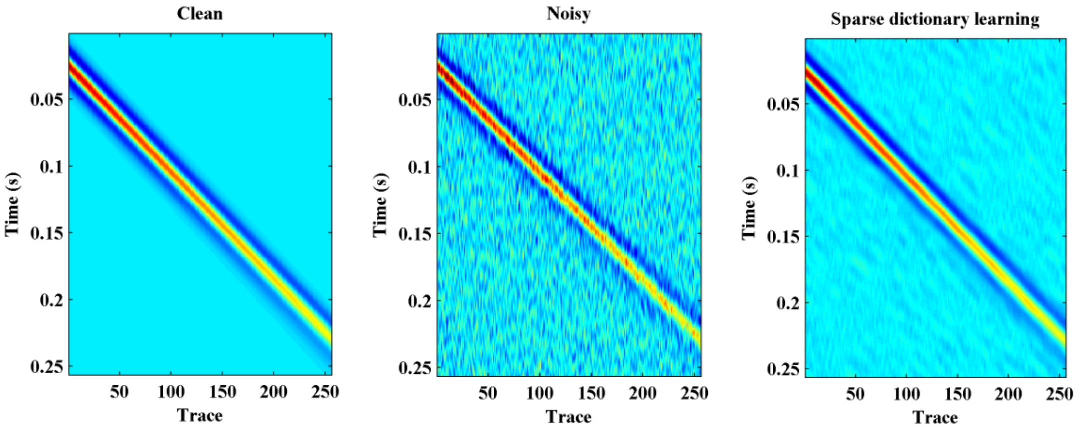
$$d = \bar{d} + \varepsilon, \tag{15}$$

where $\bar{d}$ denotes the clean data and ε denotes the Gaussian band-limited noise with zero mean. Given the denoised signal $\bar{d}$, the S/N is defined by

$$S/N = 10\log_{10} \frac{\left\|\bar{d}\right\|_2^2}{\left\|\bar{d} - \tilde{d}\right\|_2^2}. \tag{16}$$

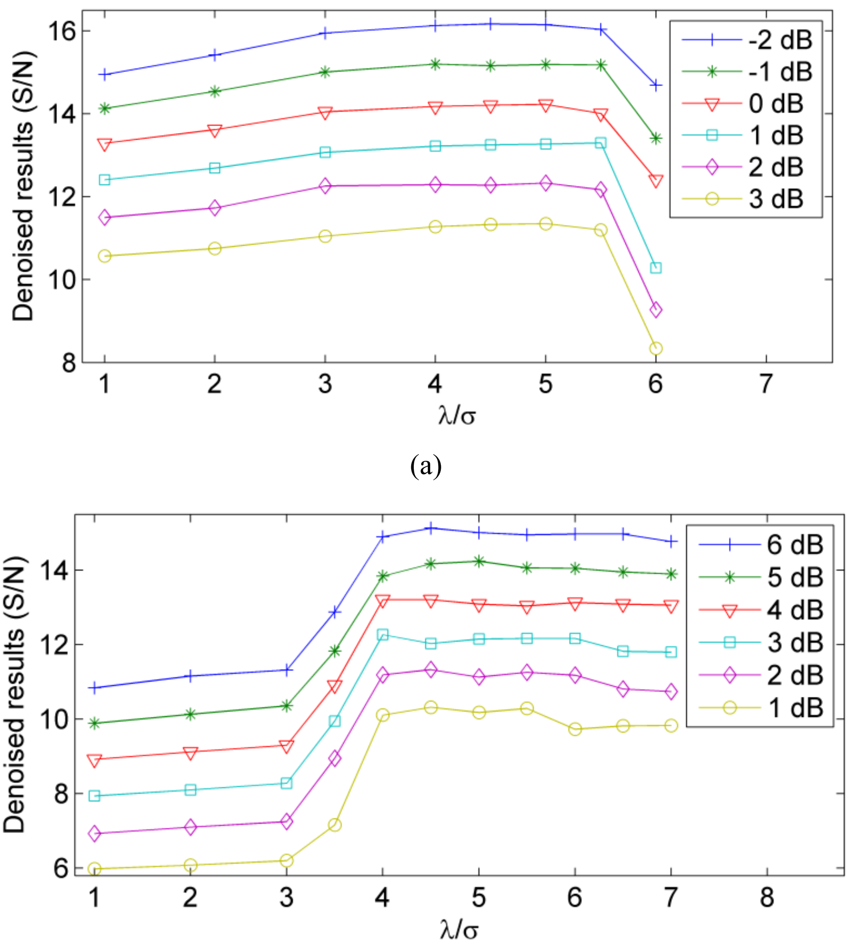
Synthetic examples

First, we use seismic data, named example 0, shown in Fig. 1a from a simple structure to observe the parameters' performance. Fig. 1b displays noisy data with S/N = 2 dB, and Fig. 1c depicts denoised result with a set of parameters. To study operator τ_λ, we use a series of λ, varying from 1.0σ to 7.0σ for different S/N noisy data, where σ is the noise variance. Fig. 2a shows the denoised results using full frequency band noisy data, and Fig. 2b shows the denoised results using the noisy data within 5-80 Hz. When we focus on the denoising results in Fig. 2a, we note that the S/Ns of the results increases gradually with the increase of λ, and the high S/Ns appears from 4.5 to 5.5 times of the variance of Gaussian white noise, then the S/N decreases dramatically with the increase of λ. More interestingly, we note from Fig. 2b that the S/Ns of the results increase greatly when λ varies in [3σ, 4σ]. Then the S/Ns keep stable when λ varies in about [4σ, 5.5σ]. Finally, the S/Ns decreases slightly with the increase of λ. Therefore, the range [4.5, 5.5] may be a good choice for λ in the process of seismic data denoising.



(a) (b) (c)

Fig. 1. (a) Clean data, (b) noisy data, and (c) denoised section. The frequency range of noisy data is 5-80 Hz. The S/N of the original data is 2 dB and that of denoised result is 14.48 dB. The sparsity level is $N = 3$, hard-thresholding operator $\tau_\lambda = 5\sigma$, filter size $L = 16$, and the number of dictionary updates $K = 20$. This is "example 0".



(a)

(b)

Fig. 2. The S/N of denoised data corresponding to different τ_λ operators and S/N of the input data. (a) Denoised results with full frequency band data and (b) denoised results with 5-80 Hz data. Different lines stand for the different original S/N varying from (a) -2 dB to 3 dB and (b) 1 dB to 6 dB. This is "example 0".

Furthermore, the filter matrix D using a filter size of 16×16 by setting $\lambda = 2.0\sigma, 4.5\sigma, 7.0\sigma$ are shown in Figs. 3b-3d. Compared with an initial dictionary, it is noted that the matrix is gradually updated as a symmetric matrix under the constraint condition $D^T D = I$ with the increase of λ. Smaller λ leads to more bases, and larger λ causes fewer bases in the filter matrix. Too many bases mean the current bases are not updated enough to obtain the data's sparse representation. Also, few bases suggest that the current filter matrix only captures some of the data's main structures. Two cases cannot learn a suitable filter matrix for sparse representation. Therefore, it is crucial to choose a proper λ to train a suitable filter matrix.

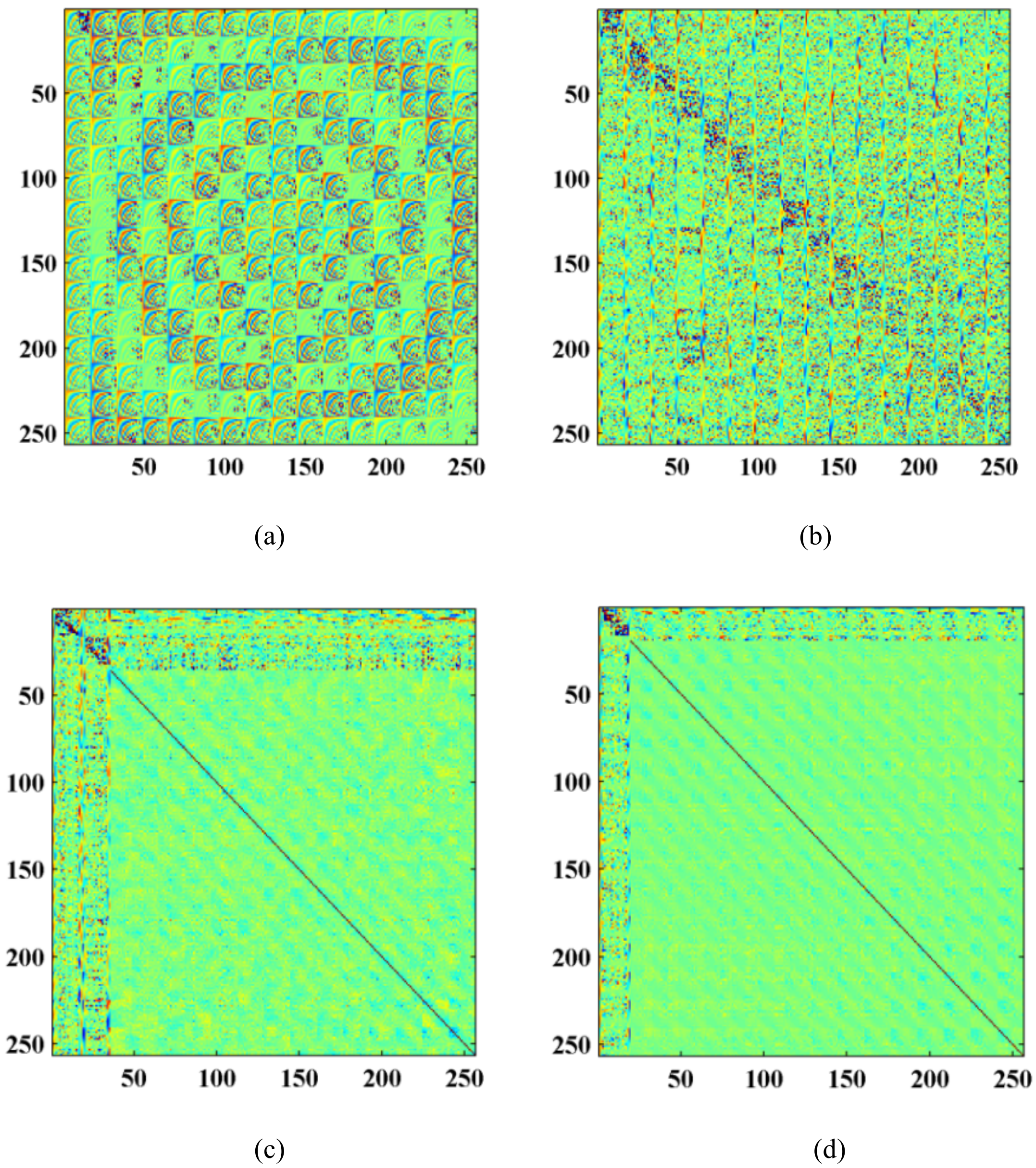


Fig. 3. (a) Initial filter matrix ***G*** and updated filter matrices using the filter size of 16×16 by setting (b) $\lambda = 2.0\sigma$, (c) $\lambda = 4.5\sigma$, (d) $\lambda = 7.0\sigma$. The noisy data with S/N=2 dB are filtered in 5-80 Hz. The maximum iteration number is 20. This is “example 0”.

We also investigate the effect of iterations and filter size on denoising results. We note from Fig. 4 that the filter size has a remarkable effect on denoising. When using a larger filter size, we can obtain a higher S/N. The reason is that large filter size makes the learned filter matrix capture the abundant structures in seismic data more likely. However, the running time increases with the size of the filter. The running times of 11.3 s, 12.99 s, 19.93 s, and 24.23 s correspond to filter sizes of 4×4, 8×8, 12×12, and 16×16 for the test data of the size of 256×256. We also note that when the iteration increases, the S/Ns improve at first and then reach stability, especially for the larger filter size cases. In general, about 20 iterations of dictionary updates are enough for sparse dictionary learning by balancing the efficiency and denoising quality.

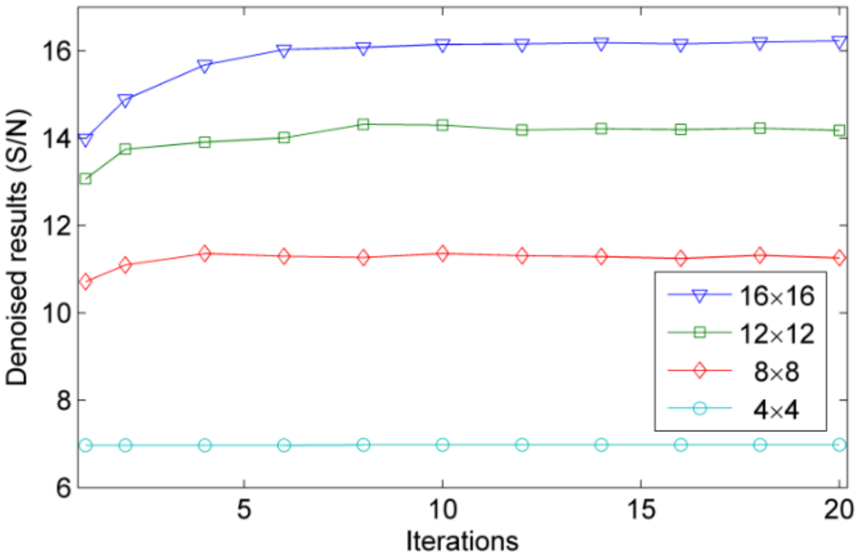


Fig. 4. The S/Ns of denoised results of sparse dictionary learning method. Different lines stand for the different filter sizes of 4×4, 8×8, 12×12, and 16×16. The noisy data with S/N = 2 dB are filtered between 5-80 Hz, and the maximum iteration number is 20. This is "example 0".

We now conduct tests on a synthetic seismic data set (example 1) shown in Fig. 5a with five events generated by a time-invariant Ricker wavelet. To test the denoising performance for different energy events, we add some Gaussian white noise on the signals and show the noisy data in Fig. 5b. We use a 5-80 Hz band-pass filter to obtain noisy data in Fig. 5c. Since the dominant frequency of Ricker wavelet is 40 Hz and the temporal sampling interval is 1 ms, we set J as 6 to divide data into different frequency bands, the lowest band is 0-7.8 Hz, and the highest band is 250-500 Hz. Before we conduct sparse dictionary learning on wavelet coefficient, we first analyze the decomposition coefficients shown in Fig. 6, from which we note that the signals mainly distribute on scales 2-6, and the coefficients of scale 1 are zeros because of filtering, thus we perform denoising on scales 2-6.

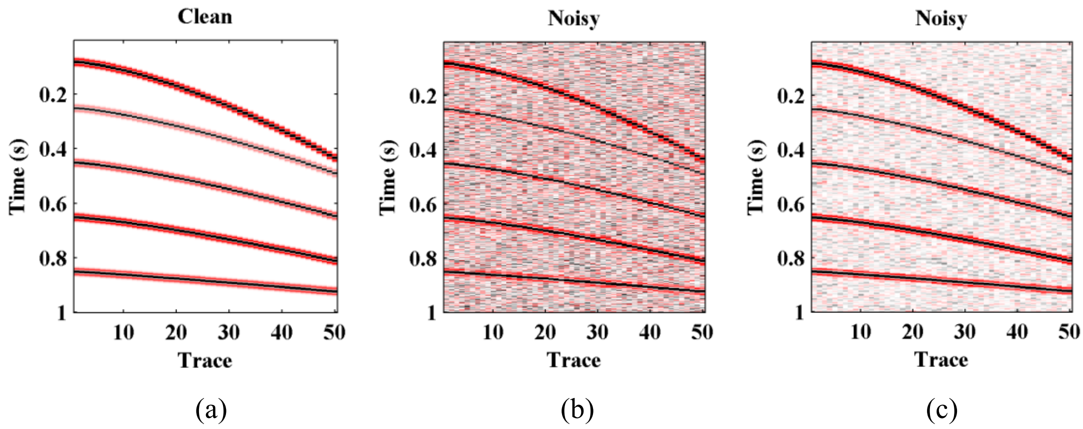


Fig. 5. (a) Clean data, (b) S/N = 2 dB noisy data, and (c) noisy data after filtered by using 5-80 Hz band-pass filter. This is "example 1" with time-invariant data.

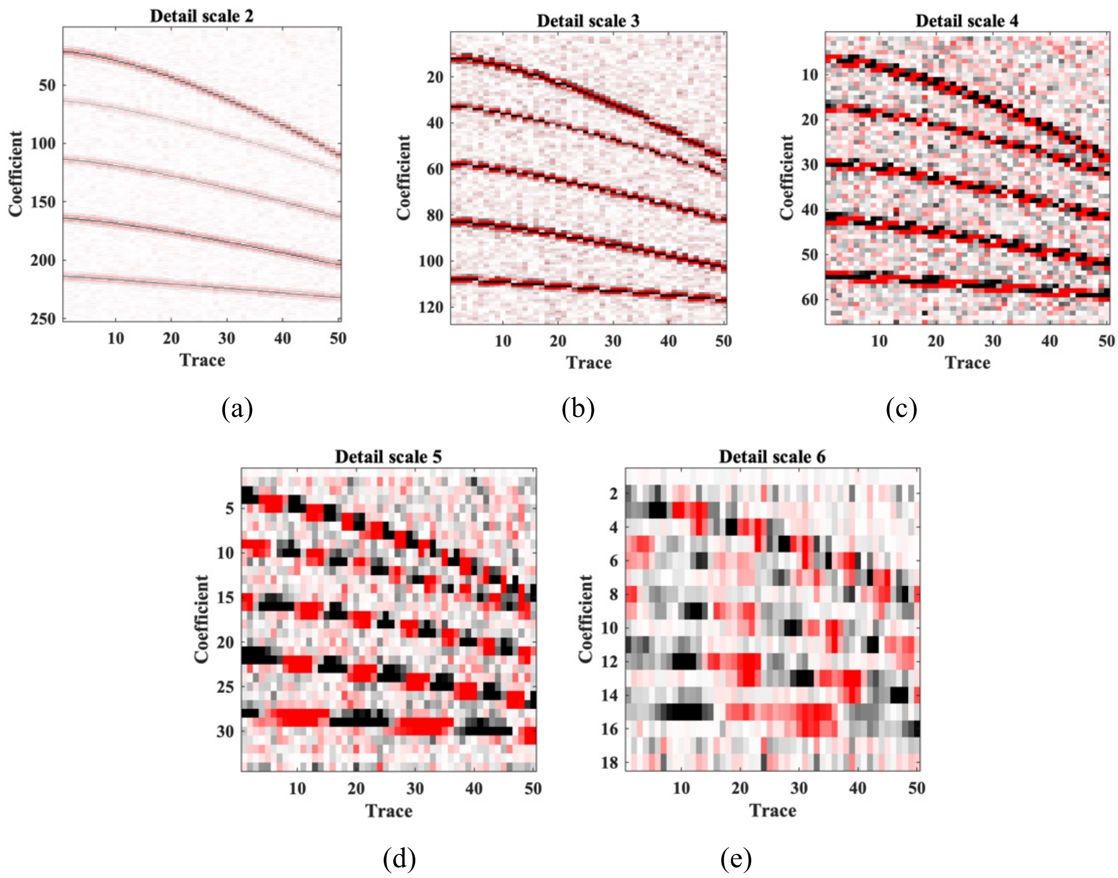


Fig. 6. Detail coefficient panels corresponding to scales 2 to 6 obtained by WT. These results are WT coefficients of example 1 and the size of data in time dimension decreases from 255 on scale 2 to 22 on scale 6, which is a decrease in a factor of about 45 from the original data (of 1001 samples). This is "example 1".

The sparsity level N is set as 5 in this example. We use the maximum filter size of 16 × 16 to conduct 20 iterations on the coefficients shown in Fig. 6. Since the wavelet transform decreases the data size, it is necessary to use a variable filter size. The WT obtains coefficients with different sparsity, i.e., the signals are more concentrated to the dominant frequency of signal than noise. Therefore, we adopt the median filtering method to get rough noise coefficients and then calculate the variances of noise coefficients for determining a suitable τ_{λ}. Figs. 7a-7e show the denoised coefficient panels of scales 2-6 with the same magnitude range as Fig. 5. The panels in Figs. 7a-7e reveal clearer structures than those in Figs. 6a-6e. We show the differences between Figs. 6b-6e and Figs. 7a-7d in Figs. 8a-8e with the same magnitude range for a detailed comparison demonstrating effective noise separation.

The final denoised result is shown in Fig. 7f, and the S/N corresponding to this result is listed in Table 1. It is noted from Table 1 that the S/N increases from 1 dB of the original noisy data to 13.55 dB of the final denoised data. It is also observed that fewer signals or noise leak in the error section (Fig. 8) which is the difference between Fig. 7f and Fig. 5a. Sparse dictionary learning and WT using a fixed threshold (Donoho, 1995) are performed to make comparisons. The same parameters, including the maximal sparse solutions, the iterations for filter dictionary update, and the maximum filter size, are set for the direct sparse dictionary learning and the multi-scale sparse dictionary learning. The waveforms of clean and noisy data are depicted in Figs. 9a and 9b, the denoised results are shown in Figs. 9c-9e, and corresponding removed data are illustrated in Figs. 9f-9h. We note that the WT with a hard threshold can effectively remove the noise (Fig. 9c) but damage some weak useful signals (Fig. 9f). If we operate sparse dictionary learning directly on the noisy data, the denoised result (Fig. 9d) contains some small noise. However, the proposed method can recover the signals more effectively, as shown in Fig. 9e. We conclude that the multi-scale sparse dictionary can relieve the effect of small noise compared with the direct dictionary learning. The S/Ns of the results are listed in Table 1, which also indicates these features. In multi-scale sparse dictionary learning, when we conduct WT on the data, we obtain wavelet coefficients, then we use sparse dictionary learning to recover the sparse coefficient of signals from them. This is a multi-scale sparse process, therefore it is reasonably able to obtain a better result.

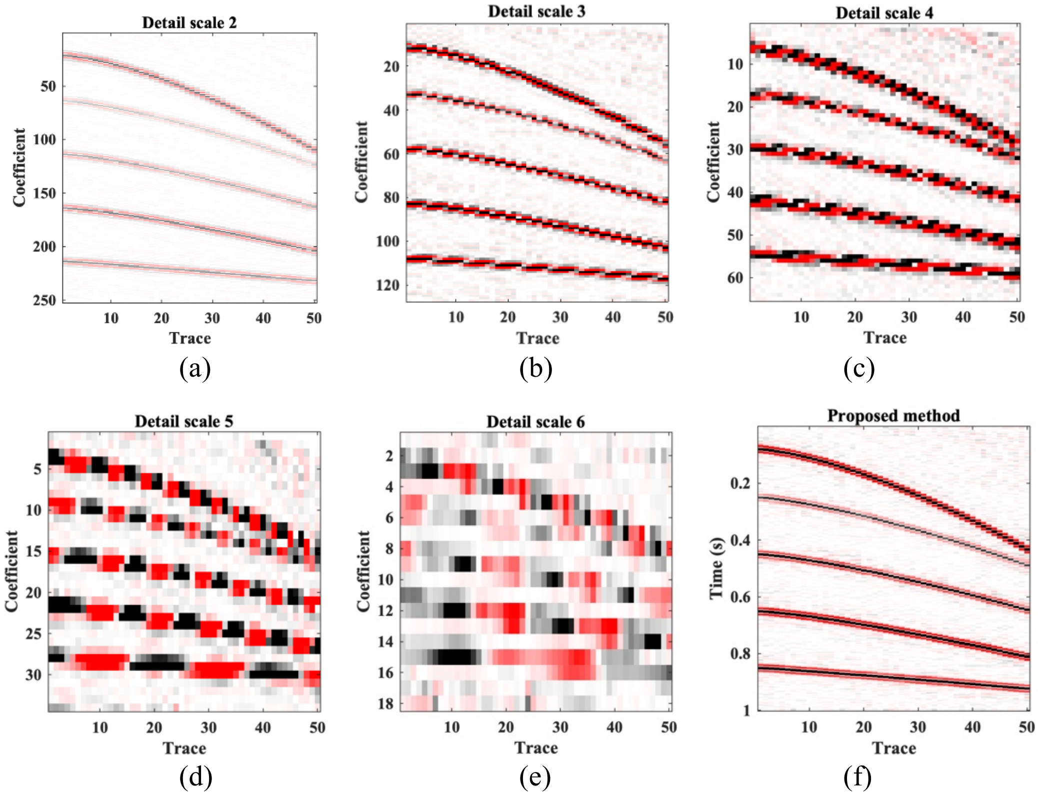


Fig. 7. (a)-(e) Denoised detail coefficient panels of scales 2 to 6, (f) reconstructed signals by using the scales shown in (a)-(e). In this example, the sparsity level is 5, filter size is 16, and the number of dictionary updates is 20 on each coefficient panel. This is "example 1".

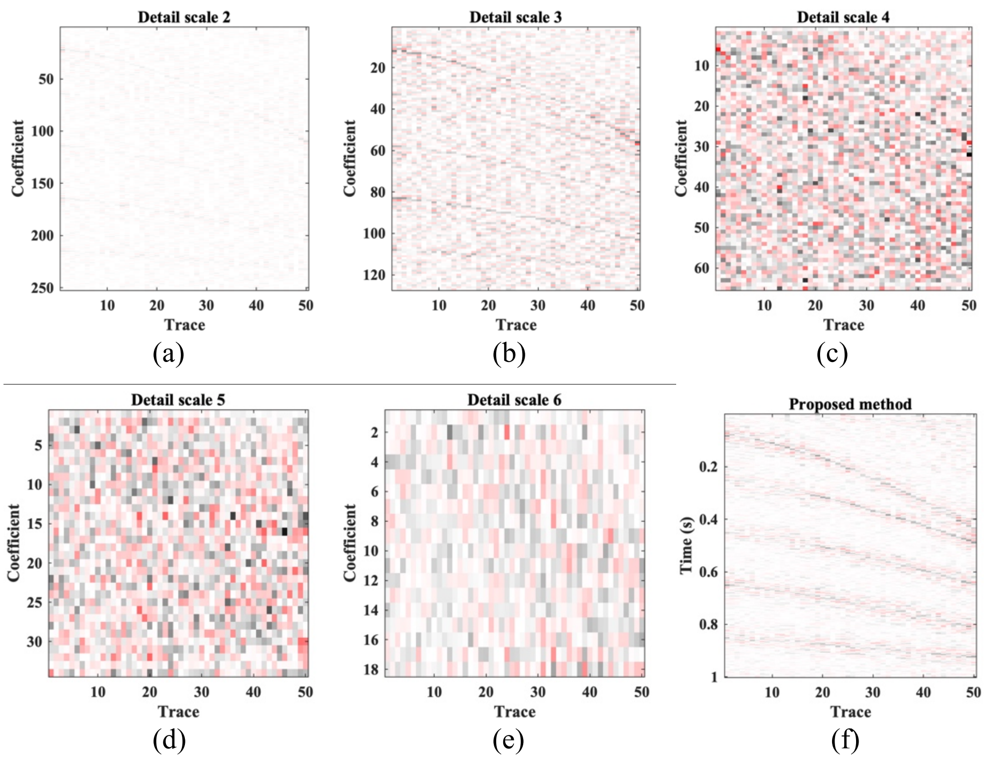


Fig. 8. (a)-(e) The difference panels between noisy coefficients shown in Fig. 6 and denoised coefficients shown in Fig. 7 of scales 2-6. (f) The difference between Fig. 5(a) and Fig. 7(f). This is "example 1".

We also compare the dictionaries shown in Fig. 10, where Fig. 10a displays the dictionary trained by sparse dictionary learning on the seismic data, and Fig. 10b displays that by sparse dictionary learning on the wavelet coefficient (detail scale 3 for example). We note that the dictionary atoms trained on the wavelet coefficient are simpler, i.e., we can use very sparse and concise bases to represent the useful information. We also display the curves of S/Ns with the number of iterations in Fig. 11a. We note from Fig. 11a that the S/Ns obtained by multi-scale sparse dictionary learning become higher than that by sparse dictionary learning, and its S/Ns gradually keep stable near the 20th iteration. For further analyzing the denoising ability of different approaches, we test on the data shown in Fig. 5a with about 3-13 dB S/Ns and show the S/Ns of denoised data in Fig. 11b. We note from Fig. 11b that the proposed method has the ability to achieve the highest S/N, especially for low S/N data.

Next, we exhibit a time-variant example (example 2), in which we utilize the same parameter setting as the last example. We adopt a quality factor Q = 60 to generate the clean time-variant (Margrave, 1998) seismic data shown in Fig. 12a. We note an apparent amplitude attenuation and phase changes from Fig. 12a, comparing with Fig. 9a. Fig. 12b shows the noisy data with the same S/N as used in Fig. 9b. Figs. 12c-12e depict denoised results obtained by WT threshold, sparse dictionary learning and multi-scale sparse dictionary learning, and Figs. 12f-12h show the corresponding error sections. Figs. 12c and 12f show that the WT threshold seems to get clearer data but damages the signals. Sparse dictionary learning retains some small noise (Fig. 12d) and damages a strong event (Fig. 12g). From S/Ns of denoised results (Table 1), we also note that the denoising ability of sparse dictionary learning becomes weaker for time-variant signals than for time-invariant signals. However, multi-scale sparse dictionary learning can overcome this weakness and has better performance. From the final error profiles shown in Figs. 12f-12h, we also know that few noise and signals are in the error section corresponding to multi-scale sparse dictionary learning. Similarly, we display the curves of the S/Ns with iteration in Fig. 13a and note that the S/Ns of the proposed method exceeds that of sparse dictionary learning. We also compare the S/Ns of the denoised results corresponding to the proposed method with those of WT and sparse dictionary learning in Fig. 13b. The data with about 3-13 dB S/N Gaussian band-limited noise are for testing. It is noted from Fig. 13b that our proposed method has an obvious advantage in achieving the highest S/N.

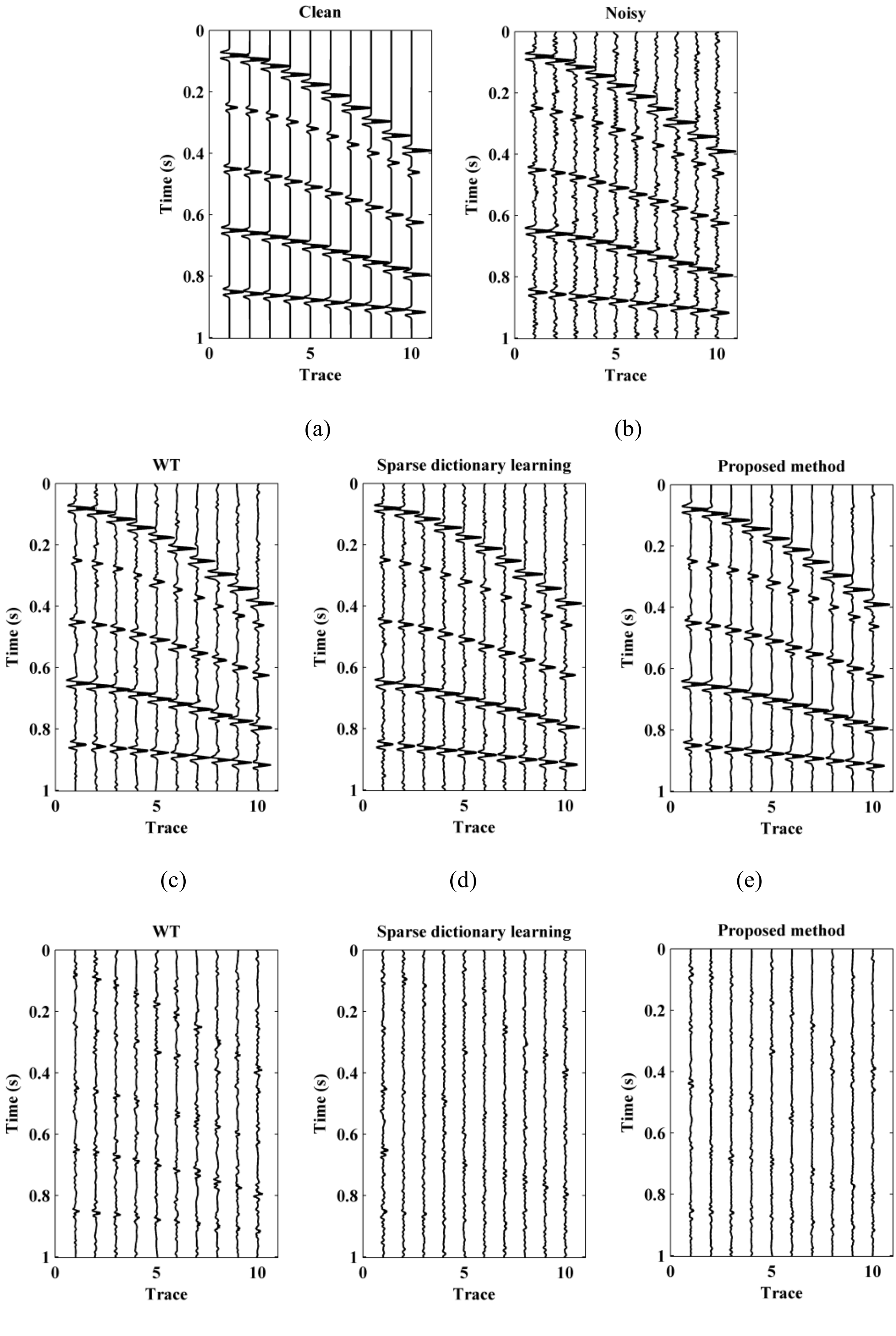
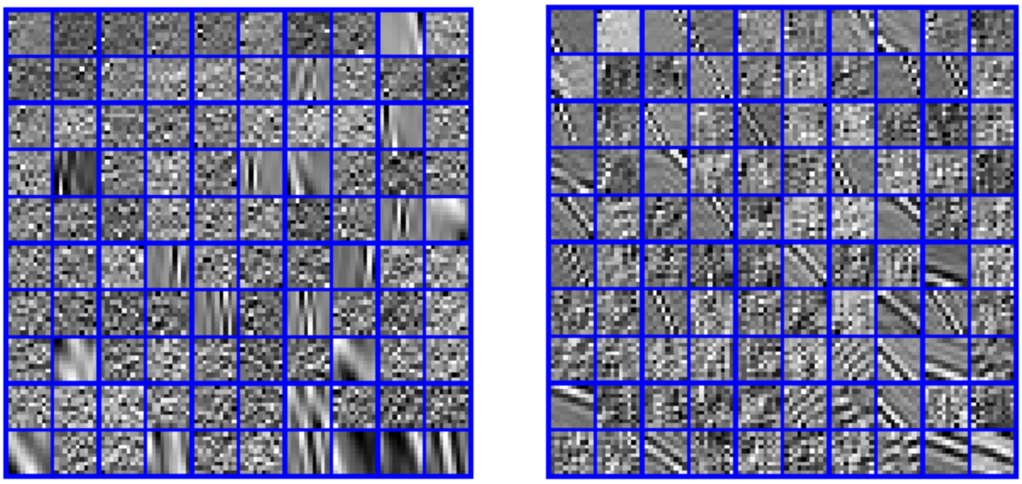


Fig. 9. (a) Clean data, (b) noisy data within 5-80 Hz, and denoised profiles using different approaches of (c) WT with hard threshold, (d) sparse dictionary learning, and (e) multi-scale sparse dictionary learning. Error sections (f)-(h) corresponding to (c)-(e). This is "example 1".

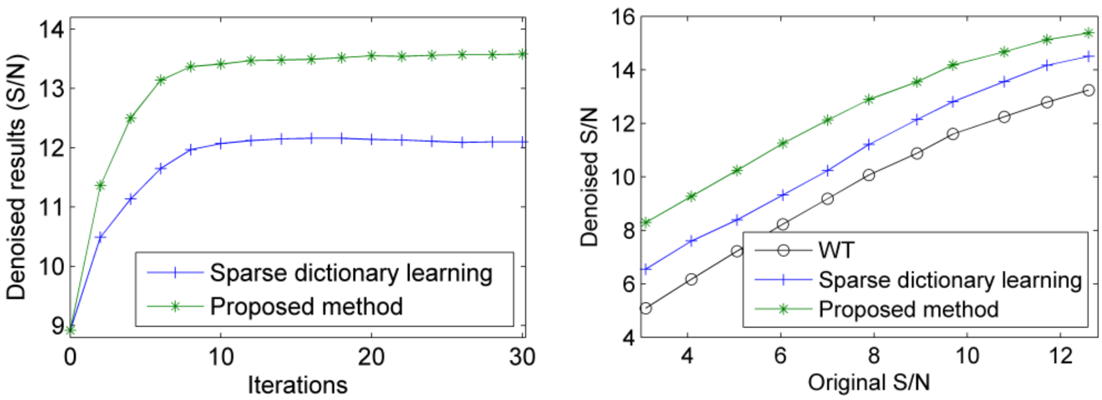
Table 1. The S/N of denoised results in decibels obtained by using different approaches. SDL represents sparse dictionary learning and MSSDL represents our multi-scale sparse dictionary learning.

Data	Original	Filtered	WT	SDL	MSSDL
Example 1	1.0	8.93	10.88	12.14	13.55
Example 2	0.76	8.93	11.16	12.10	13.59
2D Field	–6.0	2.11	3.79	6.74	7.47



(a) (b)

Fig. 10. Dictionaries from (a) sparse dictionary learning on seismic data and (b) sparse dictionary learning on sparse wavelet coefficient of scale 3. The dictionaries are obtained by displaying the learned filter matrix as atom blocks. This is "example 1".



(a) (b)

Fig. 11. (a) The S/N of denoised results with iterations, the S/N of the original data is 8.93 dB, and (b) the S/N of denoised results obtained by the three methods with different original S/N. This is "example 1".

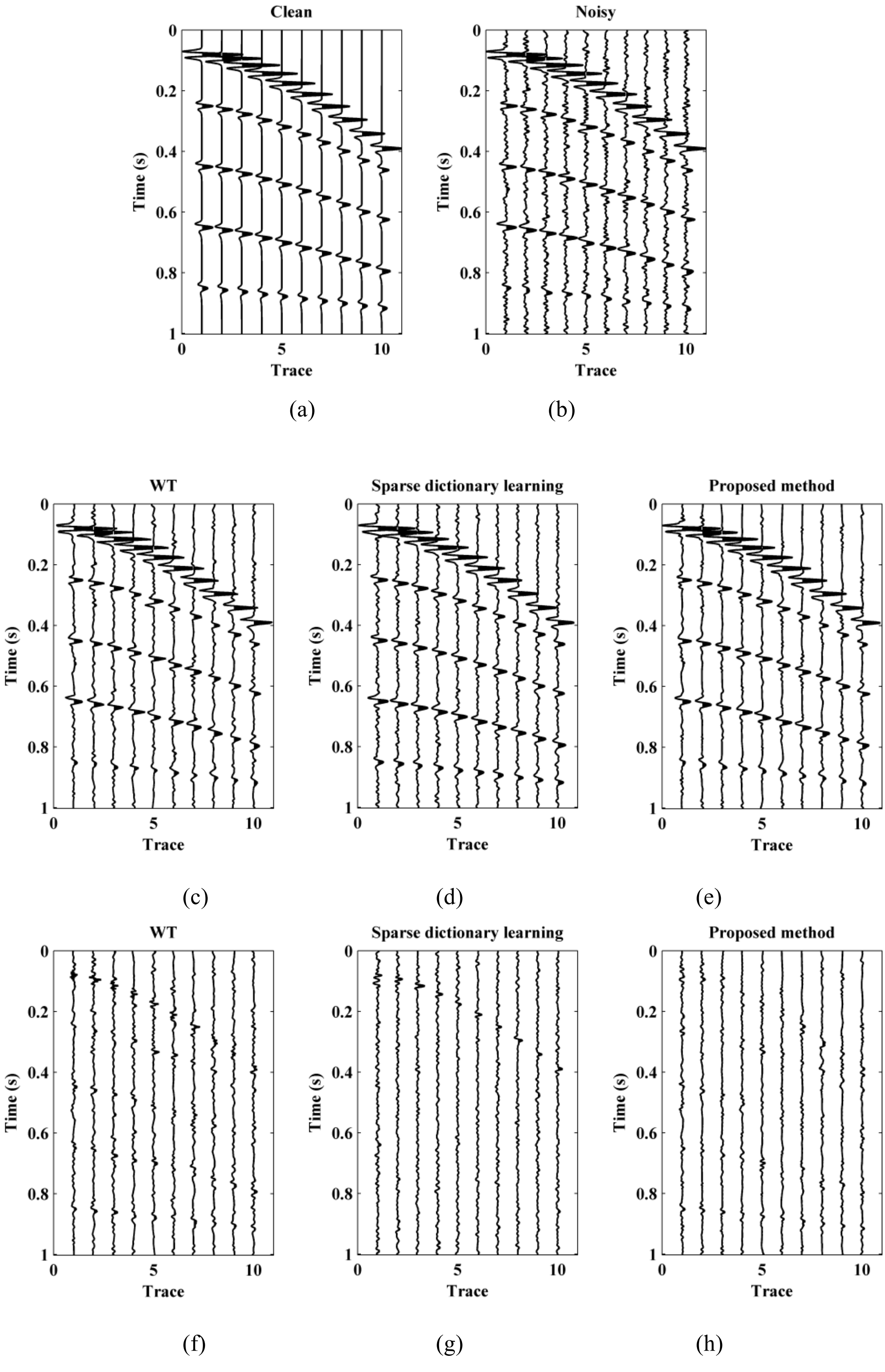
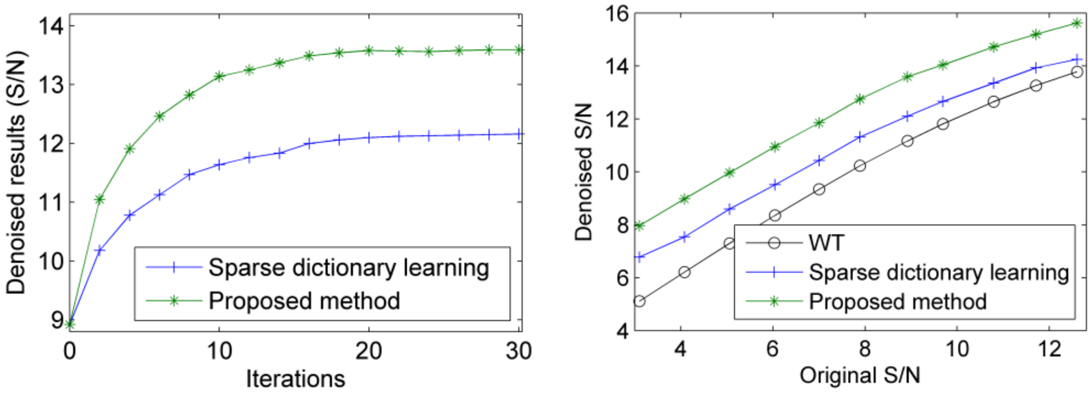


Fig. 12. (a) Clean data and (b) noisy data. Denoised sections using (c) WT threshold, (d) sparse dictionary learning, and (e) multi-scale sparse dictionary learning. (f)-(h) Error sections corresponding to (c)-(e). This is “example 2” with time-variant data and the reflectors are the same as the case depicted in Fig. 9.



(a) (b)

Fig. 13. (a) The S/N of denoised results with iterations, the S/N of original data is 8.93. (b) The S/N of denoised results obtained by the three methods with different original S/N. This is "example 2".

Field data examples

We first use a 2D seismic image shown in Fig. 14a to verify the proposed method. We add Gaussian band-limited noise to the section to make its S/N = 2.11 dB and display it in Fig. 14b. We use the three learning-based approaches to conduct denoising. In this example, we use the maximum filter size of 12×12 for sparse dictionary learning and run 20 iterations for updating the dictionary. The denoised results and corresponding noise sections are shown in Figs. 14c-14f and 15a-15d, respectively. We note that the result obtained by sparse dictionary learning still retains much small noise (Fig. 14c), and the K-SVD method causes some signal leakage shown in Fig. 15b. DSD can remove noise thoroughly, however, we find many weak useful signals in the noise section (Fig. 15c). The result obtained by the proposed method is the cleanest one with the acceptable signals left in the noise section (Fig. 15d).

Our second field data example is a 3D seismic image, as shown in Fig. 16a. First, we divide the 3D seismic dataset into a series of 2D data pieces along inlines, then we apply the denoising methods to each 2D data slice. The maximum filter for sparse dictionary learning is 16×16, and dictionary update times are 20. We adopt 4-scale wavelet coefficients for multi-scale sparse dictionary learning denoising. The denoised results using sparse dictionary learning, K-SVD, DSD, and multi-scale sparse dictionary learning are shown in Figs. 16b-16d. We also extract profiles on the same crossline from Fig. 16, as well as their noise sections, and display them in Fig. 17. We note that the sparse dictionary learning and K-SVD method does not act well because some noise left in the shallow zones of denoised result (Figs. 16b and 16c), especially for the top slice of the data cube in Fig. 16. The reason is that sparse dictionary learning and K-SVD treat some noise as useful information in the presence of complex structures and strong noise in the shallow zone. DSD can obtain clearer results, however, it removes some valuable events in the deep zones, especially for the rectangular regions marked in Fig. 17a. By comparison, the proposed method exceeds the above three methods.

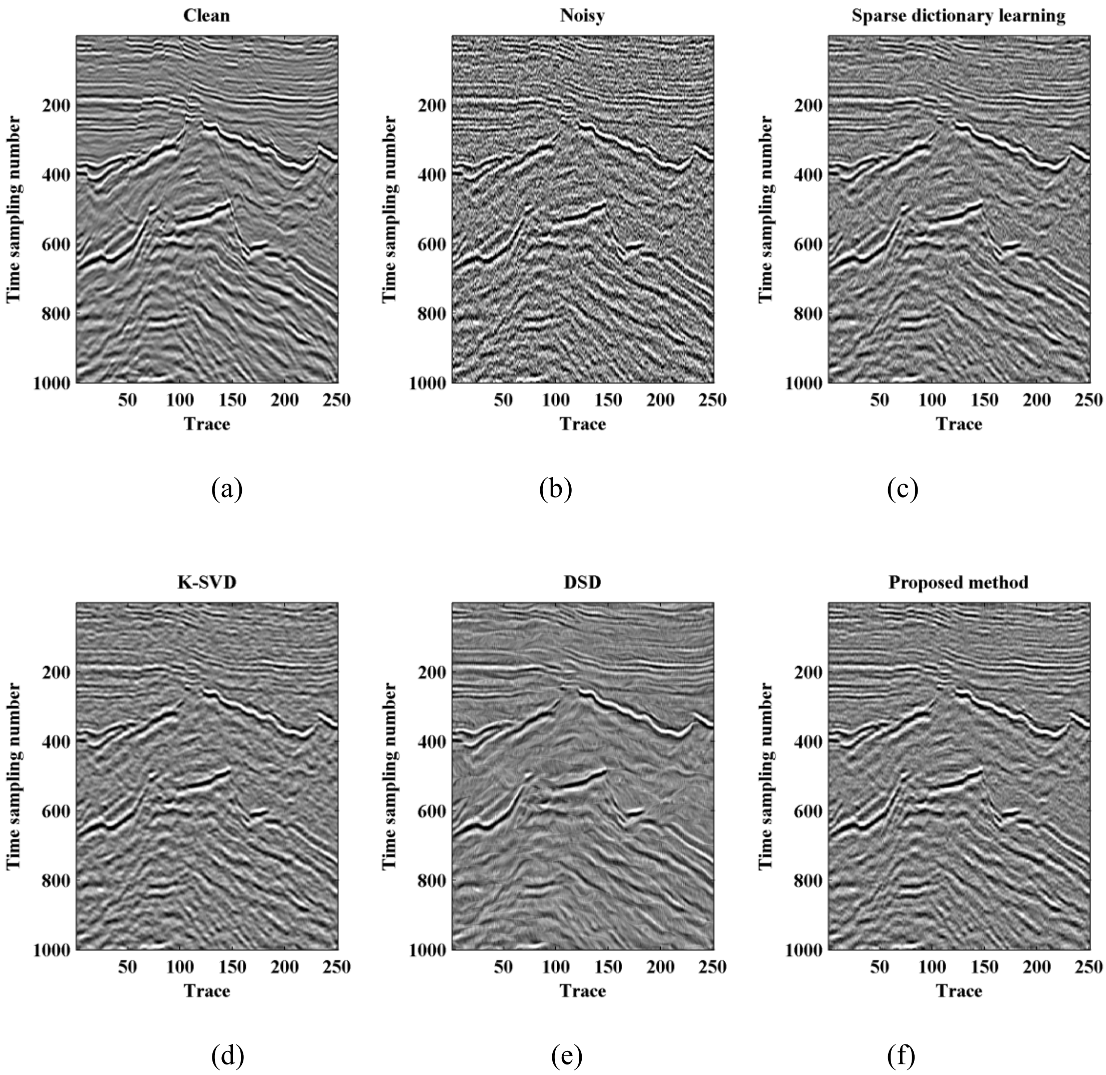


Fig. 14. (a) A real section and (b) noisy section of S/N = 2.11 dB. Denoised sections obtained using different approaches of (c) sparse dictionary learning, (d) K-SVD, (e) DSD, and (f) multi-scale sparse dictionary learning. This is “2D field data”.

We also compare the running time of sparse dictionary learning and the multi-scale sparse dictionary learning in dealing with the above synthetic and field data examples. We need to mention here that in our 3D example, we denoise the profiles along the crossline direction one by one, so we perform a parallel scheme in the program. Table 2 lists the running times taken by the four examples. We conclude that the proposed method is more efficient than direct sparse dictionary learning because WT decreases the size of data, resulting in a short running time of sparse dictionary learning. Owing to the fast WT transform, we can quickly decrease data amount without much price. In addition, the WT can help reduce the complexity of the original data.

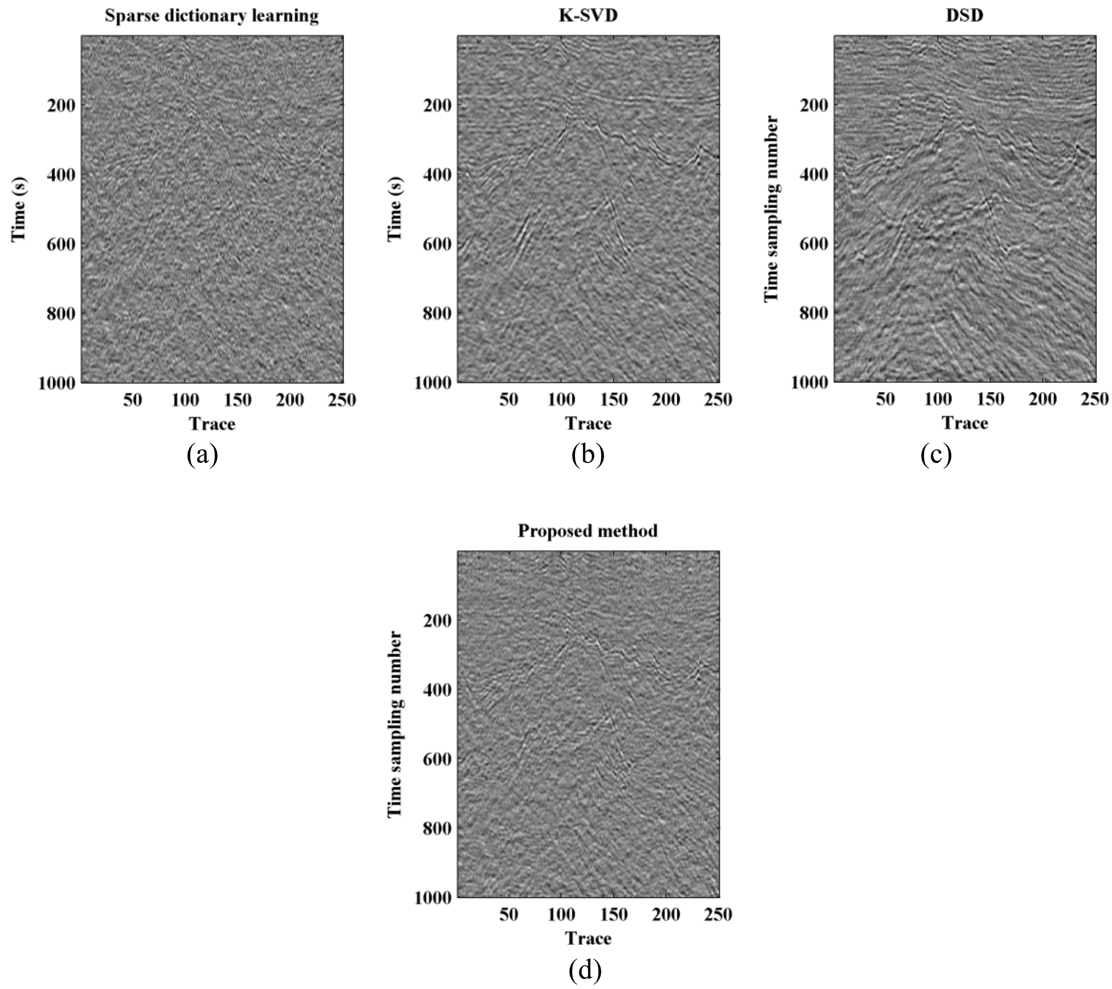


Fig. 15. (a)-(d) Removed noise sections corresponding to Figs. 14(c)-14(f). This is “2D field data”.

Table 2. Comparison of running time in seconds using different approaches. SDL represents sparse dictionary learning and MSSDL represents our multi-scale sparse dictionary learning.

Data	Points	WT	SDL	MSSDL
Example 1	1001×50	0.43	19.85	14.01
Example 2	1001×50	0.46	19.44	18.37
2D Field	1001×251	1.07	113.11	88.35
3D Field	511×133×16	8.06	273.47	213.45

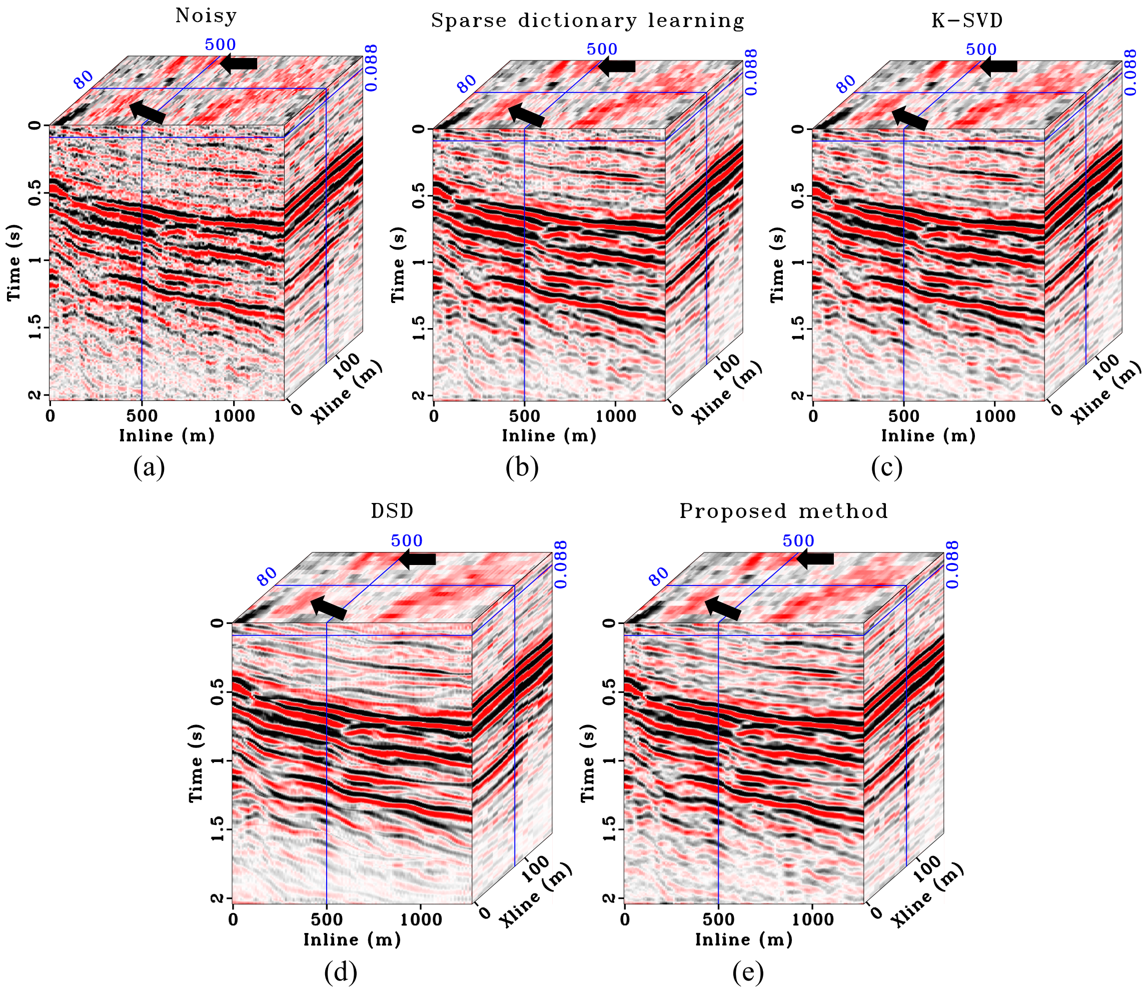


Fig. 16. (a) A real noisy data, denoised sections obtained using different approaches of (b) sparse dictionary learning, (c) K-SVD, (d) DSD, and (e) multi-scale sparse dictionary learning. This is “3D field data”.

DISCUSSION

The sparse dictionary learning algorithm used in this paper makes up two processes: training the dictionary constrained by a data-driven tight frame and using it for sparse representation. When obtaining a filter matrix, the operator τ_λ is a vital parameter to capture information from input data. On the one hand, small λ results in in the insufficient information captured by the dictionary, on the contrary, large λ results in no enough information learned from the seismic data. These two situations do not help get a reasonable tight frame. By numerical tests, we find that 4.5 to 5.5 times the variance of noise may be the right choice for λ to achieve high S/N of the denoised result. Regarding the WT sparse coefficients, the variance of prior noise needs to be estimated. Comparing Fig. 2a with Fig. 2b, the frequency-band-limited noisy data have larger tolerance than white-noise data, thus, it can relax the accuracy of noise variance estimation.

Seismic data suffer from the earth filtering, which results in the signals being time-variant, so we choose wavelet transform as a useful tool to accomplish time-variant signal analysis. Removing noise beyond seismic frequency band directly and learning dictionaries in remaining sparse coefficients are helpful to dictionary learning, we can decrease the effect from other frequency-band noise. Comparing Fig. 13a with Fig. 13b, we find that despite the same S/N for time-invariant and time-variant data, the convergence speed of time-variant data is slower than that of time-invariant data. Therefore, it is necessary to employ more iterations to update the filter matrix.

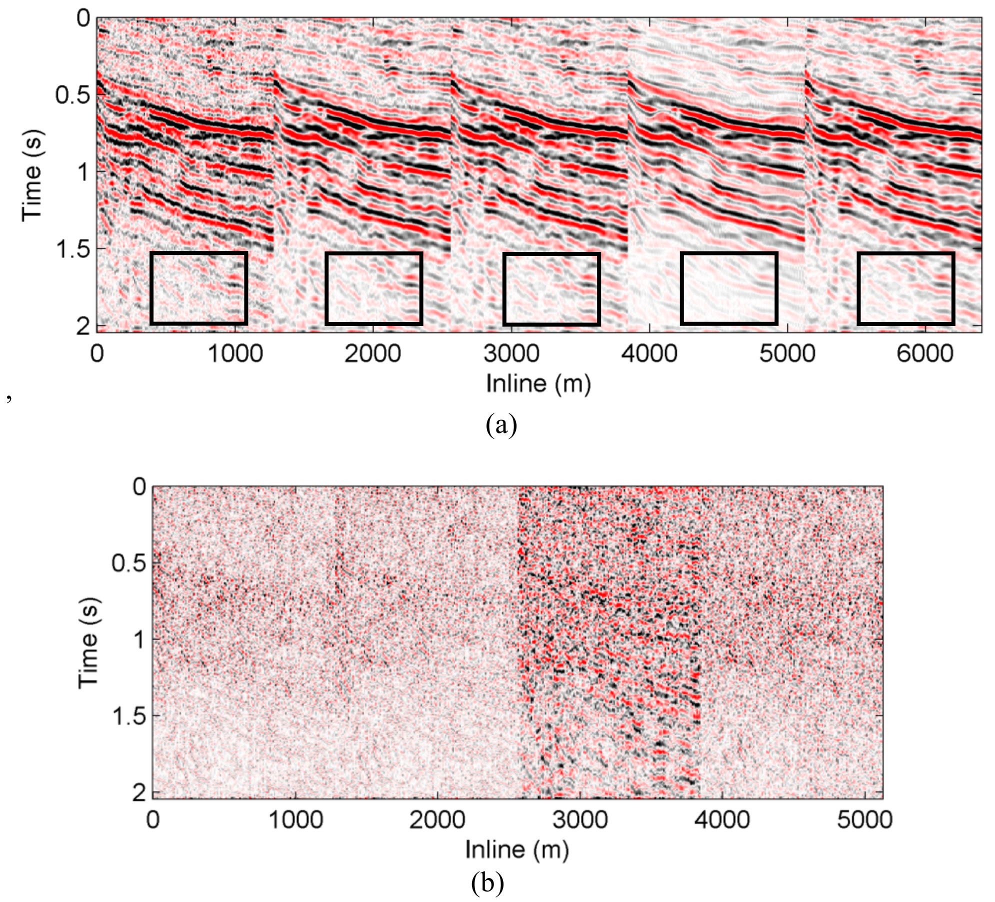


Fig. 17. Profiles of a same crossline. (a) Sections from the left to the right are the original data, denoising results using sparse dictionary learning, K-SVD, DSD, and multi-scale sparse dictionary learning, and (b) removed noise sections corresponding to these methods in (a). This is "3D field data".

We find from the numerical examples that wavelet transform with a threshold denoising method may not obtain a good result since the fixed wavelet base is not suitable for all seismic data characteristics. Regarding direct dictionary learning, its excellent denoising ability decreases when complex structures exist in seismic data and strong noise aliases in the signals. However, these drawbacks can be relaxed by the more robust multi-scale sparse dictionary learning.

CONCLUSION

We have presented a multi-scale sparse dictionary learning approach to remove random noise from the time-variant seismic data and discussed how to obtain suitable parameters involved in the algorithm. The proposed method achieves three advantages. First, through the wavelet transform, the sparse dictionary learning can easily capture the features of each sparse coefficient and represent the sparse coefficients by the corresponding learning-based dictionaries. Secondly, the wavelet transform and multi-scale sparse dictionary learning complete a multi-scale sparse representation conducive to the separation of noise and signals. Finally, the wavelet transform is a down-sampling process, which can reduce the dimensionality of data, thus making multi-scale dictionary learning more effective. By comparing the denoised results, we find that the wavelet transform with a threshold denoising method removes the noise but damages signals simultaneously since it does not take the characteristic of data into account. Direct dictionary learning and K-SVD might fail to describe complex seismic data efficiently. However, multi-scale sparse dictionary learning can separate signals from noisy data effectively for its multi-scale sparse strategy.

ACKNOWLEDGMENTS

This work is partly supported by National Natural Science Foundation of China (42104113, 42174165), Natural Science Foundation of Jiangsu Province of China (BK20210526), Innovation-Driven Project of Central South University (2021zzts0257), and Fundamental Research Funds for the Central Universities (2021QN1099). We thank Mr. Xiong Ma for providing the method to generate time-variant seismic data.

REFERENCES

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

Bao, C., Ji, H. and Shen, Z., 2015. Convergence analysis for iterative data-driven tight frame construction scheme. Appl. Computat. Harmon. Analys., 38: 510-523.

Beckouche, S. and Ma, J., 2014. Simultaneous dictionary learning and denoising for seismic data. Geophysics, 79(3): A27-A31.

Bednar, J.B., 1983. Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data. Geophysics, 48: 1598-1610.

Cai, J.F., Ji, H., Shen, Z. and Ye, G.B., 2014. Data-driven tight frame construction and image denoising. Appl. Computat. Harmon. Analys., 37: 89-105.

Chen, S.S., Donoho, D.L. and Saunders, M.A., 2001. Atomic decomposition by basis pursuit. SIAM review, 43: 129-159.

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

Chen, Y., Ma, J. and Fomel, S., 2016. Double sparsity dictionary for seismic noise attenuation. Geophysics, 81(2): V103-V116.

Daubechies, I., 1992. Ten lectures on wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics, 61. SIAM Publications, Philadelphia.
Daubechies, I., Han, B., Ron, A. and Shen, Z., 2003. Framelets: MRA-based constructions of wavelet frames. Appl. Computat. Harmon. Analys., 14: 1-46.
Donoho D.L., 1995. Denoising by soft-thresholding. IEEE Transact. Informat. Theory, 41: 613-627.
Fomel, S., Sava, P., Vlad, I., Liu, Y. and Bashkardin, V., 2013. Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments. J. Open Res. Softw., 1, e8.
Feng, Z., 2021. Seismic random noise attenuation using effective and efficient dictionary learning. J. Appl. Geophys., 104258.
Hennenfent, G. and Herrmann, F.J., 2006. Seismic denoising with nonuniformly sampled curvelets. Comput. Sci. Engineer., 8(3): 16-25.
Ioup, J.W. and Ioup, G.E., 1998. Noise removal and compression using a wavelet transform. Expanded Abstr., 68th Ann. Internat. SEG Mtg. New Orleans: 1076-1079.
Jawerth, B. and Sweldens, W., 1994. An overview of wavelet based multiresolution analyses. SIAM review, 36: 377-412.
Liang, J., Ma, J. and Zhang, X., 2014. Seismic data restoration via data-driven tight frame. Geophysics, 79(3): V65-V74.
Liu, C., Song, C. and Lu, Q., 2017. Random noise de-noising and direct wave eliminating based on SVD method for ground penetrating radar signals. J. Appl. Geophys., 144: 125-133.
Lopes, M.E., 2016. Unknown sparsity in compressed sensing: Denoising and inference. IEEE Transact. Informat. Theory, 62: 5145-5166.
Lv, H. and Bai, M., 2018. Learning dictionary in the approximately flattened structure domain. J. Appl. Geophys., 159: 522-531.
Mallat, S., 1999. A Wavelet Tour of Signal Processing. Elsevier Science Publishers, Amsterdam.
Margrave, G.F., 1998. Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering. Geophysics, 63: 244-259.
Miao, X. and Cheadle, S., 1998. Noise attenuation with wavelet transforms. Expanded Abstr., 68th Ann. Internat. SEG Mtg., New Orleans: 1072-1075.
Naghizadeh, M. and Sacchi, M.D., 2010. Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data. Geophysics, 75(6): WB189-WB202.
Neelamani, R., Baumstein, A. and Ross, W.S., 2010. Adaptive subtraction using complex-valued curvelet transforms. Geophysics, 75(4): V51-V60.
Ophir, B., Lustig, M. and Elad, M., 2011. Multi-scale dictionary learning using wavelets. IEEE J. Select. Topics Sign. Process., 5: 1014-1024.
Rubinstein, R., Zibulevsky, M. and Elad, M., 2009. Double sparsity: learning sparse dictionaries for sparse signal approximation. IEEE Transact. Signal Process., 58: 1553-1564.
Schnemann, P.H., 1966. A generalized solution of the orthogonal procrustes problem. Psychometrika, 31: 1-10.
Sezer, O.G., Guleryuz, O.G. and Altunbasak, Y., 2015. Approximation and compression with sparse orthonormal transforms. IEEE Transact. Image Process., 24: 2328-2343.
Strang, G. and Nguyen, T., 1996. Wavelets and Filter Banks, Wellesley-Cambridge Press, New York.
Tropp, J.A. and Gilbert, A.C., 2007. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transact. Informat. Theory, 53: 4655-4666.
Turquais, P., Asgedom, E.G. and Söllner, W., 2017. A method of combining coherence-constrained sparse coding and dictionary learning for denoising. Geophysics, 82(3): V137-V148.
Wang, D., Saab, R., Yilmaz, Ö. and Herrmann, F.J., 2008. Bayesian wavefield separation by transform-domain sparsity promotion. Geophysics, 73(5): A33-A38.
Yilmaz, Ö. and Donherty, S., 2001. Seismic Data Analysis. SEG, Tulsa, OK.
Yu, S., Ma, J. and Osher, S., 2016. Monte Carlo data-driven tight frame for seismic data recovery. Geophysics, 81(4): V327-V340.

Yu, S., Ma, J., Zhang, X. and Sacchi, M.D., 2015. Interpolation and denoising of high-dimensional seismic data by learning a tight frame. Geophysics, 80(5): V119-V132.

Yuan, Y.O. and Simons, F.J., 2014. Multi-scale adjoint waveform-difference tomography using wavelets. Geophysics, 79(3): WA79-WA95.

Zhan, R. and Dong, B., 2016. CT image reconstruction by spatial-Radon domain data-driven tight frame regularization. SIAM J. Imaging Sci., 9: 1063-1083.

Zhang, R. and Ulrych, T.J., 2003. Physical wavelet frame denoising. Geophysics, 68: 225-231.

Zhang, Q., Wang, H., Chen, W. and Huang, G., 2020. A robust method for random noise suppression based on the Radon transform. J. Appl. Geophys., 104183.

Zhu, L., Liu, E. and Mcclellan, J.H., 2015. Seismic data denoising through multiscale and sparsity-promoting dictionary learning. Geophysics, 80(6): WD45-WD57.

Zu, S., Zhou, H., Wu, R., Jiang, M. and Chen, Y., 2019. Dictionary learning based on dip patch selection training for random noise attenuation. Geophysics, 84(3): V169-V183.

Zu, S., Zhou, H., Wu, R., Mao, W. and Chen, Y., 2018. Hybrid-sparsity constrained dictionary learning for iterative deblending of extremely noisy simultaneous-source data. IEEE Transact. Geosci. Remote Sens., 57: 2249-2262.