DESERT SEISMIC EXPLORATION LOW-FREQUENCY NOISE ATTENUATION BASED ON IMPROVED CO-SPARSE ANALYSIS MODEL

YI ZHAO[1], YUE LI[1]*, DAN SHAO[1] and BAOJUN YANG[2]

[1] *Department of Information, Jilin University, Changchun 130012, P.R. China.*
** liyue@jlu.edu.cn*
[2] *Department of Geophysics, Jilin University, Changchun 130026, P.R. China.*

(Received June 20, 2019; revised version accepted April 23, 2020)

ABSTRACT

Zhao, Y., Li, Y., Shao, D. and Yang, B.J., 2020. Desert seismic exploration low-frequency noise attenuation based on improved co-sparse analysis model. *Journal of Seismic Exploration*, 29: 505-525.

Desert low-frequency noise is a kind of noise in desert seismic exploration records, with significant low-frequency characteristics. Severe frequency aliasing occurs because the noise is in the same frequency band as the seismic signal. In addition, the interference noise has strong energy over whole time period of seismic records, which makes the signal easily submerged in the noise. These characteristics of desert low-frequency noise challenge traditional denoising methods. Aiming at the noise attenuation of low signal-to-noise ratio (SNR) seismic exploration records in desert areas, first half-quadratic optimization approach is proposed to solve the energy minimization problem instead of common optimization methods in the co-sparse analysis model. And then, shrinkage function is introduced into the model by the additive form of half-quadratic optimization, which makes the model remove the restriction of sparsity promoting function. Finally, according to the characteristics of seismic exploration records, the training data are normalized and then trained. Both the synthetic and real data experiments prove that the improved model can better overcome the frequency aliasing and more thoroughly remove the low-frequency noise compared with the traditional co-sparse analysis model.

KEY WORDS: co-sparse analysis model, half-quadratic optimization, shrinkage function, desert low-frequency noise, noise attenuation.

0963-0651/20/$5.00

INTRODUCTION

Seismic exploration records collected in the field are usually interfered by various noise sources. The characteristics of low SNR and difficult to distinguish effective signals have brought adverse effects on subsequent processing. Desert low-frequency noise is a kind of noise commonly found in desert seismic exploration records, which is mainly composed of natural noise and human noise. In human noise, near-field human noise dominates. The energy distribution of desert low-frequency noise is mainly concentrated in the low-frequency band. Therefore, the noise has significant low-frequency characteristics (Li et al., 2017), causing it to be in the same frequency band as the seismic signal and resulting in severe frequency aliasing. Moreover, desert low-frequency noise also has low stationarity and weak Gaussianity (Zhong et al., 2015). Traditional denoising methods are difficult to attenuate such noise thoroughly, or it is difficult to remove noise without damaging the signal. The energy of the low-frequency noise is also strong, causing the seismic signal to have a large deviation or even completely submerged in the noise. It puts higher requirements on the performance of denoising methods.

In recent years, many experts and scholars have developed various seismic random noise attenuation methods, typically *f-x* prediction filtering (Harris and White, 2010), wavelet transform (Pati et al., 2002), curvelet transform (Emmanuel and Donoho, 2000), sparse dictionary learning (Candès and Donoho, 2010; Ma et al., 2010; Rodriguez et al., 2012), complex diffusion filtering (Gilboa et al., 2004), time-frequency peak filtering (TFPF) (Thode, 1987) and so on. Although these methods have achieved good results in removing seismic random noise, they also have certain limitations. For example, traditional noise attenuation methods such as curvelet transform and wavelet transform are more suitable for removing Gaussian noise. While the time-frequency peak filtering and the complex diffusion filtering have certain advantages in attenuating desert low-frequency noise, they are all subject to certain assumptions or conditions.

Sparse representation is to find a suitable over-complete dictionary for ordinary densely expressed signals, and to represent the signal with as few atoms as possible in the dictionary, thus reducing the complexity of the model. Due to the sparsity of effective seismic signals and the non-sparsity of random noise, sparse dictionary learning has been widely used in random noise attenuation for seismic exploration records in the past decade. In general, the methods based on sparse representation belong to sparse synthesis model. However, there is another model for sparse representation, known as the co-sparse analysis model (Yaghoobi et al., 2012, 2015; Elad, 2007). In the co-sparse analysis model, an over-complete analysis operator $A \in R^{n \times l}$ is given, which can map the signal x from high-dimensional space $R^{m \times n}$ to low-dimensional space $R^{m \times l}$. Let $z = Ax$, $z \in R^{m \times l}$ be a sparse signal and have more zero elements. The number p of zero elements is called the co-sparsity of the signal x, which is important for signal recovery in the linear inverse problem. Compared with the sparse synthesis model,

there are few studies on the co-sparse analysis model. However, in recent years, the rapid development of deep learning has provided conditions for learning the model. Some scholars have applied it to image denoising and achieved good results (Hawe et al., 2013).

In this paper, based on the extension of half-quadratic optimization, shrinkage function is introduced into the co-sparse analysis model. By re-modeling the shrinkage function with the Gaussian kernel function RBF, the restriction of the sparsity promoting function on the model is removed. In the case of being able to effectively learn all the model parameters, the flexibility of the model is improved, and the internal features of the seismic signal can be captured more finely, so that the distribution of the seismic signal is specifically fitted to remove the low-frequency noise. The synthetic and real data experiments also prove that the proposed method has better noise attenuation effect and higher amplitude preservation than the traditional co-sparse analysis model.

WORKING THEORY

Co-sparse Analysis Model

Compared with the well-known sparse synthesis model, the co-sparse analysis model has received less attention. However, in recent years, more and more scholars have been working on it. Previous studies on the co-sparse analysis model were mostly based on image patches, and few literature discussed the entire image. Still, patch-based model can only be used to process image patches but not the whole natural image directly, which ignores the similarity between patches and the global characteristics of the entire image (Elad and Aharon, 2006). Therefore, Chen et al. (2014) proposed a co-sparse analysis model based on global image, which fully utilized the global prior information. Moreover, the equivalence between this model and FOE model (Roth and Black, 2009) is proved.

In the co-sparse analysis model, the prior information of the signal is learned with the sparsity promoting function and the linear operator A which can be decomposed into a series of filters. Therefore, given the observation signal y, the probability density function (PDF) of the pure signal $x \in R^{m \times n}$ can be written as:

$$p(x|\mathrm{y}) = \frac{1}{Z(\Theta)} \exp(-E(x \mid y)) \quad , \tag{1}$$

with

$$E(x \mid y) = \sum_{p=1}^{N_p} \phi(A \cdot P_P \cdot \mathrm{x}) \quad , \tag{2}$$

where $Z(\Theta)$ is the normalized parameter, ϕ is the sparsity promoting

function, and $A \in R^{\mathrm{n}\times\mathrm{m}}$ is the over-complete analysis operator, or linear operator, that needs to be learned in the co-sparse analysis model. $P_P \in R^{m\times N_P}$ ($N_P = n\times m$) is a three-dimensional sampling matrix, and its function is to extract a signal patch centered on point *p* in the signal *x*.

Analysis operator $A \in R^{\mathrm{n}\times\mathrm{m}}$ can be decomposed into *n* row vectors $\{A_1, A_2 \ldots\ldots A_n\}$, where A_i is the *i*-th row vector of *A*. Treating the row vector A_i as a filter with dimension $\sqrt{m}\times\sqrt{m}$, then $A_i \cdot x$ is equivalent to the convolution of the row vector A_i with the signal *x*, i.e., $A_i \cdot x \Leftrightarrow A_i * x$. To this end, the Gibbs energy eq. (2) can be rewritten as follows:

$$E(x \mid y) = \sum_{p=1}^{N_P}\sum_{i=1}^{n}\phi(A_i \cdot P_P \cdot x) = \sum_{p=1}^{N_p}\sum_{i=1}^{n}\phi(A_i * x_p) \quad , \qquad (3)$$

where $x_p = P_P \cdot x$.

In the framework of Bayesian theory, the denoising problem is equivalent to the MAP estimate. Because of the posterior distribution $p(x \mid y) \propto \exp(-E(x \mid y))$, it allows to predict the signal *x* by finding the energy minimum:

$$\hat{x} = \arg\max_{x\in R^{m\times\mathrm{n}}} p(x \mid y) = \arg\min_{x\in R^{m\times\mathrm{n}}} E(x \mid y)$$

$$E(x \mid y) = \sum_{p=1}^{N_p}\sum_{i=1}^{n}\phi(A_i x_p) + \frac{\lambda}{2}\|\mathrm{y} - K\mathrm{x}\|_2^2 \quad . \qquad (4)$$

A large number of experiments have proved that the choice of the sparsity promoting function ϕ in eq. (4) plays a decisive role in the co-sparse analysis model. Chen et al. (2014) proposed the view of using potential function to construct sparsity promoting function, and verified the relationship between them: $\phi_i = -\log\rho_i$, where ρ_i is the potential function. Since the desert seismic signal will show a heavy-tailed distribution [Fig.1(a)] after being filtered, the potential function ρ_i usually also denotes a function that exhibits a heavy-tailed distribution, such as the ST distribution and the GLP distribution. The functional form is:

ST distribution： $\rho_i(z_i; p_i) = (1+z^2)^{-p_i}$

GLP distribution： $\rho_i(z_i; p_i) = e^{-|z|^{r_1}}$

The resulting sparsity promoting function is:

$$\phi_i = p_i \log(1+z^2)$$

$$\phi_i = |z|^{p_i} \quad .$$

Half-Quadratic Optimization

To find the optimal solution of eq. (4), various optimization algorithms, such as gradient descent method, Newton method, quasi-Newton method and so on, are often involved. The optimization algorithm used in this paper is the half-quadratic optimization algorithm first proposed by Geman and Reynolds (2002), and by Geman and Yang (1994) in the field of machine vision. The main content is to minimize the energy function by introducing an independent auxiliary variable z_{ip} related to the filter response. The problem is rewritten as the following mathematical form:

$$\arg\min E(x \mid y) = \arg\min E(x, z \mid y)$$

Then use the block coordinate descent method to optimize x and z respectively. The optimization of the function $E(x, z \mid y)$ can be divided into two parts: one is to optimize the quadratic function $E(x \mid z, y)$ to obtain x under the condition of z and y; the other part is to update the value of independent variable z through optimizing $E(z \mid x, y)$ under the condition that x and y are known. The essence of half-quadratic optimization is a special quadratic relaxation method.

Half-quadratic approach is categorized into additive (Geman and Yang, 1994) and multiplicative forms (Geman and Reynolds, 2002) according to two different extension of $E(x \mid z, y)$. The two forms are expressed as follows:

$$\text{Additive form: } \arg\min E(x \mid z, y) = \Omega(z, y)^{-1}\eta(y)$$

$$\text{Multiplicative form: } \arg\min E(x \mid z, y) = \Omega(y)^{-1}\eta(z, y)$$

where Ω is a highly sparse matrix and η is a vector. These two extended forms simplify the optimization problem into a mathematical problem for solving linear equations. In the additive form, the variable z can only affect the sparse matrix Ω , which ensures that the right side $\eta(y)$ of the equation remains unchanged in the iterative, thus reducing the computational difficulty. Therefore, the additive form of the half-quadratic optimization has more practical significance.

Since the additive form has more advantages than the multiplicative form, we consider using it to solve the energy minimization problem. However, due to the relationship between sparsity promoting function and potential function, the additive form of half-quadratic optimization cannot be directly applied to the optimal solution of the co-sparse analysis model. In response to this problem, Wang et al. (2016) proposed to introduce parameter $\beta \to \infty$ into the half-quadratic optimization. Thus, the equivalent form of eq. (4) is obtained:

$$\hat{x} = \arg\min_{x} E(x \mid y) = \arg\min_{x,z} E_{\beta}(x, z \mid y)$$

$$E_{\beta}(x, z \mid y) = \sum_{p=1}^{N_p} \sum_{i=1}^{n} \left(\frac{\beta}{2}(A_i x_p - z_{ip})^2 + \phi(z_{ip})\right) + \frac{\lambda}{2}\|\mathrm{y} - K\mathrm{x}\|_2^2 \quad , \qquad (5)$$

when $\beta \to \infty$, the independent auxiliary variable $z_{ip} \to A_i x_p$, eq. (5) is equivalent to eq. (4). Krishnan and Fergus (2009) applied this method to the deconvolution of non-blind images; Schmidt and Roth (2014) used this method in the CSFs.

Because $E_{\beta}(x, z \mid y)$ can be transformed into two separate parts of $E_{\beta}(z \mid x, y)$ and $E_{\beta}(x \mid z, y)$, the alternate optimization of *x* and *z* can be achieved by minimizing the above two functions. The processes are summarized as the following mathematical problems:

$$z^* = \arg\min_{z} E_{\beta}(z \mid x, y) = f_{i,\beta}(A_i x_p)$$

$$x^* = \arg\min_{x} E_{\beta}(x \mid z, y) = g_{\beta}(z)$$

with

$$f_{i,\beta}(v) = \arg\min_{z}\left(\phi(z) + \frac{\beta}{2}(v - z)^2\right) \quad , \qquad (6)$$

$$g_{\beta}(z) = \mathcal{F}^{-1}\left[\frac{\mathcal{F}\left(\frac{\lambda}{\beta} K^T y + \sum_{i=1}^{N} \mathcal{A}_i \mathrm{z}_i\right)}{\frac{\lambda}{\beta} \breve{K}^* \circ \breve{K} + \sum_{i=1}^{N} \breve{\mathcal{A}}_i^* \circ \breve{\mathcal{A}}_i}\right] \quad . \qquad (7)$$

In the case of noise attenuation, the value of *K* is generally equal to 1. Since each row of *A* corresponds to a two-dimensional filter of size $\sqrt{\mathrm{m}} \times \sqrt{\mathrm{m}}$, $\mathcal{A}\mathrm{z}=[\mathrm{Az}_1, \cdots, \mathrm{Az}_{N_n}]^T \equiv A \otimes \mathrm{z}$ is used to represent the convolution of filters *A* and *z*. The condition of the convolution is the circular (periodic) boundary condition. $\circ$ denotes division or multiplication. $\breve{\mathcal{A}} = \mathcal{F}(A)$ denotes the DFT of *A* and $\breve{\mathcal{A}}^*$ denotes the conjugate of $\breve{\mathcal{A}}$.

Because (6) is a function only related to *v*, we can store all possible values of *v* in a table to facilitate the next calculation, which reducing the occupation of computing resources. Moreover, as shown in Fig. 1(b), the half-quadratic optimization of the sparsity promoting function fits the signal distribution more than the function itself. Eq. (7) uses DFT to map convolution calculation to multiplication operation in the frequency domain, provides a closed formula, and completes the solution of the linear equations in the transform domain. These greatly simplify the calculation

and improve the precision of the operation in the same time. It can be seen that the additive form of half-quadratic optimization has great appeal in solving the optimization problem.

Shrinkage Function

Shrinkage function (Hel-Or and Shaked, 2008) is to make the filter coefficients gradually approach zero in the iterative, because the coefficients of filter are generally considered to be caused by noise. $f_{i,\beta}(v)$ in eq. (6) is the shrinkage function. At present, it has been proved that the shrinkage function is related to the signal prior, and the relationship between half-quadratic optimization and shrinkage function is given in the literature (Wang et al., 2016).

As can be seen from eq. (6), shrinkage function $f_{i,\beta}(v)$ is only related to the sparsity promoting function, as shown in Fig. 1(c). In practice, $f_{i,\beta}(v)$ is often restricted by the sparsity promoting function. This leads to the fact that the same mode is presented for signals with different characteristics, which seriously affects the flexibility of the model. Based on this, we can get rid of this limitation by directly modeling the shrinkage function (Schmidt and Roth, 2014). We choose to remodel $f_{i,\beta}(v)$ with a series of linear Gaussian kernels:

$$f_{\pi_i}(v)=\sum_{j=1}^{M}\pi_{ij}\exp(-\frac{\gamma}{2}(v-\mu_j)^2) \quad . \tag{8}$$

Set M to 53. π_{ij} denotes the weight of different Gaussian kernel, which is learned from training. γ denotes the shared precision. μ_j denotes a fixed distance between the cores.

Direct modeling of shrinkage function has two main advantages: First, $f_{\pi_i}(v)$ is free from the restriction of sparsity promoting function, which makes the model more flexible and accurate in capturing the fine features of the seismic signals. Second, the original optimization problem is decomposed into two optimization processes. After re-modeling, the two optimization processes are reduced to one optimization process, which improves the computational efficiency.

The function $g(x)$ changes accordingly after re-modeling $f_{i,\beta}(v)$:

$$g_\theta(x) = \Omega^{-1}\eta = \mathcal{F}^{-1}\left[\frac{\mathcal{F}(\lambda K^T y + \sum_{i=1}^{N} \mathcal{A}_i f_{\pi_i}(\mathcal{A}_i^T x)}{\lambda \breve{K}^* \circ \breve{K} + \sum_{i=1}^{N} \breve{\mathcal{A}}_i^* \circ \breve{\mathcal{A}}_i}\right] \quad . \tag{9}$$

Note, parameter β are summarized into the parameters π_i and λ. The above equation is a closed form, which allows for efficient training. Because of $x^* = g_\theta(x)$, $g_\theta(x)$ is called the prediction of signal x. In practice, multiple predictions are often chained together to form a cascade model.

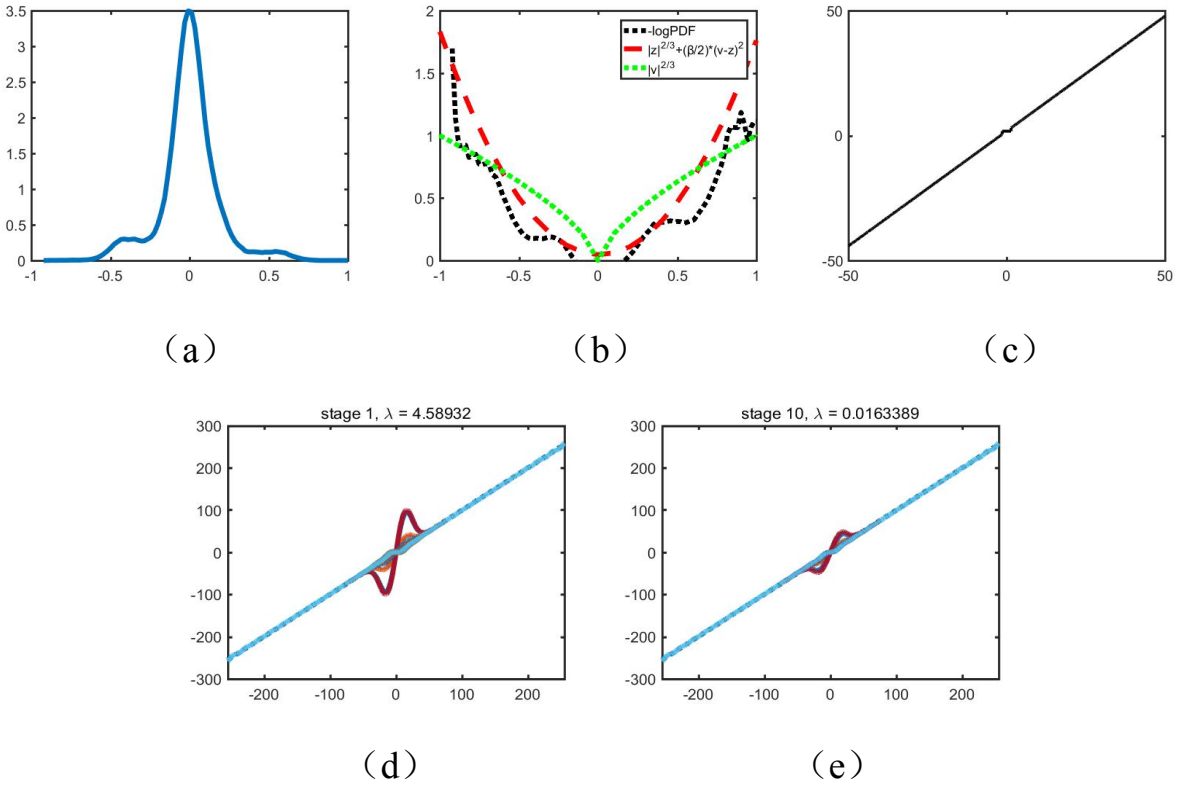


Fig. 1. (a) The PDF of the filtered signal. (b) Negative log PDF of the filtered signal (black, dashed line) and sparsity promoting function $\phi(v) = |v|^{2/3}$ (green, dotted line) and its quadratic relaxation $\phi(z) + \frac{\beta}{2}(v-z)^2$ when $\beta = 3.5, z = 0.01$ (red, double line). (c) Associated shrinkage function $f_\beta(v) = \arg\min_z(\phi(z) + \frac{\beta}{2}(v-z)^2)$ for $\phi(z) + \frac{\beta}{2}(v-z)^2$ when $\beta = 3.5, z = 0.01$. (d,e) Learned shrinkage functions $f_{\pi_i}(v) = \sum_{j=1}^{M} \pi_{ij} \exp(-\frac{\gamma}{2}(v-\mu_j)^2)$ as linear combination of Gaussian RBF kernels.

Learning

The purpose of training is to find the most sparse linear operator A, so that Ax is sparse. However, in the actual training process, the situation $A \equiv 0$ often occurs. A common solution to this problem is to add some additional constraints to A. For example, Yaghoobi et al. (2012) used the l_1 norm as the sparsity promoting function and added the constraint of the Uniform Normalized Tight Frame (UNTF). Hawe et al. (2013) took the mixed (p,q)-pseudo-norm as the sparsity promoting function and the full rank matrices with normalized rows as the constraint. However, this training method is difficult to accurately learn the parameters, so Chen et al. (2014) proposed a bilevel optimization framework without constraints. Although the bilevel optimization framework has further improved the training accuracy, every sample of this framework needs to be bilevel optimized in each iteration, which results in too few training data allowed.

In this paper, the training method is loss-minimization. The main purpose of training is to learn the two parameters π_i, λ and the linear operator A. It is known that each row of linear operator A is equivalent to a two-dimensional filter that can be further decomposed into a series of base filters $\{B_1, B_2, \cdots, B_{N_B}\}$ with zero mean (Fig. 2):

$$A_i = \sum_{j=1}^{N_B} \beta_{ij} B_j \quad .$$

Therefore, A can be obtained by learning the parameters β_{ij}.

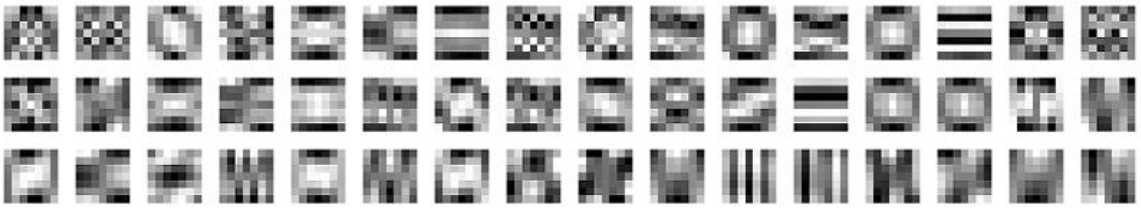


Fig. 2. 7 × 7 basic filters learned in the training.

In practice, if T predictions are chained together, this cascade model has T stages. In the t ($1 \le t \le T$) stage, let $\Theta_t = \{\pi_{ti}, \lambda_t, \beta_{tij}\}$ as a parameter set, then the purpose of training is to learn the optimal parameter Θ_t by minimizing the loss function in the t stage. For a given training data $X = [x^{(1)}, x^{(2)}, \cdots x^{(s)}]$, where $x^{(i)}$ is the i-th noisy signal, the cost function for the t stage is:

$$J(\Theta_t) = \sum_{i=1}^{S} \ell(\hat{x}_t^{(i)}; x_t^{(i)}) \quad , \qquad (10)$$

where $\hat{x}^{(i)} = g_\theta(x^{(i)})$, $\ell(\hat{x}; x)$ is the loss function. Since the loss function

must be a continuously differentiable function, we choose the peak signal-to-noise ratio (PSNR) as $\ell(\hat{x};x)$:

$$\ell(\hat{x};x) = -20\log_{10}(\frac{R\sqrt{D}}{\|\hat{x}-x\|}) \quad , \tag{11}$$

where R is the maximum intensity level (we set $R = 10$) and D is the amount of data that x has.

In each stage, we have to minimize the loss function, and the optimization method used is gradient-based L-BFGS method (Liu and Nocedal, 1989). Therefore, the minimization process in the t stage needs to first calculate the gradient of the loss function with respect to the parameters Θ_t:

$$\frac{\partial\ell(\hat{x}_t;x_t)}{\partial\Theta_t} = \frac{\partial\ell(\hat{x}_t;x_t)}{\partial\hat{x}_t}\cdot\frac{\partial\Omega_t^{-1}\eta_t}{\partial\Theta_t} = \hat{c}_t^T[\frac{\partial\eta_t}{\partial\Theta_t} - \frac{\partial\Omega_t}{\partial\Theta_t}\hat{x}_t] \quad , \tag{12}$$

with

$$\hat{c}_t = \Omega_t^{-1}[\frac{\partial\ell(\hat{x}_t;x_t)}{\partial\hat{x}_t}]^T \quad . \tag{13}$$

Similar to the calculation method of $\hat{x}_t$, $\hat{c}_t$ is also obtained by solving linear equations in the transform domain. For the sake of brevity, gradient formulas for specific parameters are no longer given. For details, refer to the supplementary material in Schmidt and Roth (2014), which also has a detailed discussion on how to remove edge artifacts from periodic convolution.

The training based on loss function can be applied not only to each stage, but also to the joint T stages. The joint cost function of T stages is:

$$J(\Theta_{1,\ldots,T}) = \sum_{i=1}^{S}\ell(\hat{x}_T^{(i)};x^{(i)}) \quad . \tag{14}$$

This shows that there is a correlation between the loss functions of each stage, so the gradient of the joint cost function can be expressed as

$$[\frac{\partial J(\Theta_{1,\ldots,T})}{\partial\Theta_1},\ldots,\frac{\partial J(\Theta_{1,\ldots,T})}{\partial\Theta_T}] \quad .$$

In joint training, the gradient in each stage is not only related to the gradient $\frac{\partial\ell(\hat{x}_t;x)}{\partial\Theta_t}$ of this stage, but also related to the gradient $\frac{\partial\ell(\hat{x}_t;x)}{\partial\Theta_{t\text{-}s}}$ of

other stages. The gradient of other stages can be summarized by recursive method:

$$\frac{\partial \ell(\hat{x}_t; x)}{\partial \Theta_{\text{t-s}}} = \hat{c}_{t-s}^{T}\left[\frac{\partial \eta_{t-s}}{\partial \Theta_{\text{t-s}}} - \frac{\partial \Omega_{t-s}}{\partial \Theta_{\text{t-s}}}\hat{x}_{t-s}\right] \tag{15}$$

with

$$\hat{c}_{t-s}^{T} = \Omega_{t-s}^{-1}\sum_{i=1}^{N} \mathcal{A}_{\text{t-s+1},i} f'_{\pi_{\text{t-s+1}}}(\mathcal{A}_{\text{t-s+1},i}\hat{x}_{t-s})\mathcal{A}_{\text{t-s+1},i}^{T}\hat{c}_{t-s+1} \quad . \tag{16}$$

In this way, the parameters of all the stages in the model can be precisely trained.

The above principle process can be summarized as a flow chart (Fig. 3).

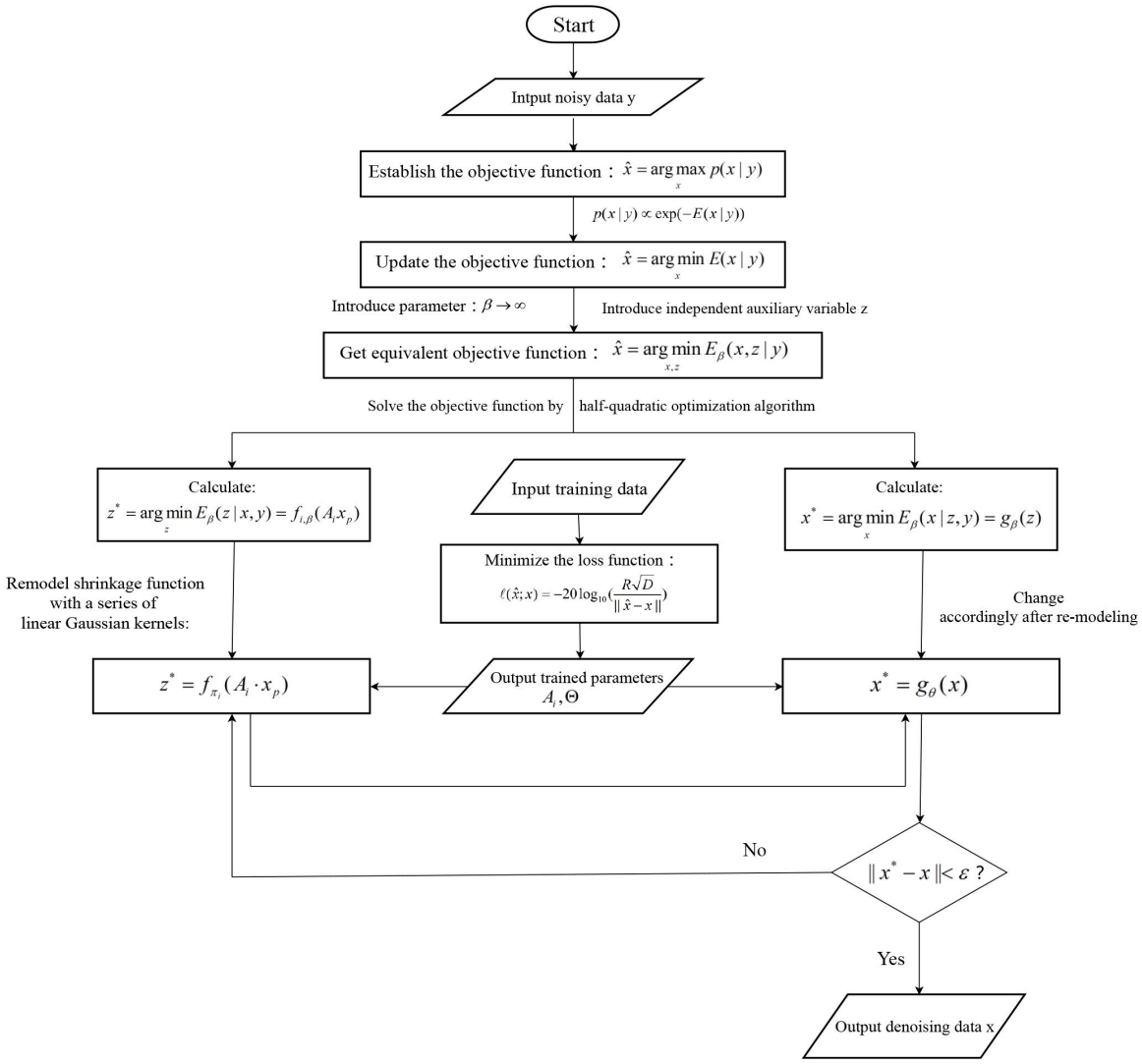


Fig. 3. Principle flow chart.

EXAMPLES

Training model

In the experiment, in order to more specifically remove desert low-frequency noise and recover seismic signals, 85 seismic synthetic records with the size of 800×90 are selected, and 12 times of desert low-frequency noise is added to train. However, due to the unstable intensity level of the noisy seismic exploration records with low SNR, the training process is complicated and difficult to converge. Therefore, we normalize the training data and then train it. For the comprehensive consideration of training time and effect, the number of model stages is 10. In Figs. 1(d) and 1(e), the shrinkage functions learned in the first and tenth stages are given. According to the filter size used in the training, it is mainly divided into the following four models:

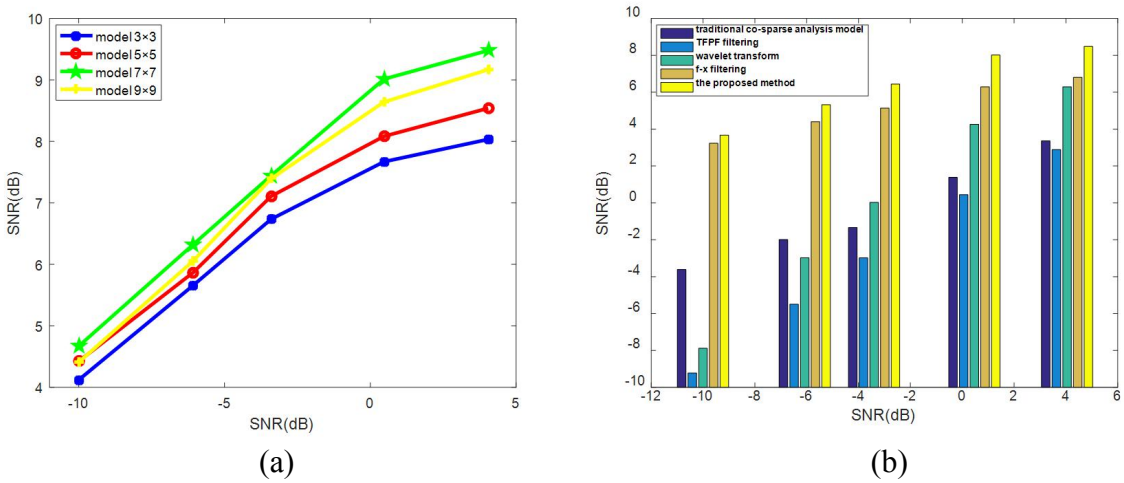
model 3×3	Trained model with 8 filters of size 3×3.
model 5×5	Trained model with 24 filters of size 5×5.
model 7×7	Trained model with 48 filters of size 7×7.
model 9×9	Trained model with 80 filters of size 9×9.

the number of filters l is limited by the filter size $m \times m : l = m^2 - 1$.

Table 1 shows the results of the four models after processing noisy seismic exploration records with different SNR. It can be seen from the data in the table that the model 7×7 has the best noise attenuation effect and the improvement of SNR is also the highest - when processing the seismic exploration record with a high SNR of 4.0443 dB, the SNR before and after processing is about 5.43 dB; When processing the seismic exploration record with a low SNR of -10.0016 dB, the SNR increases by about 14.67dB. Among the other three models, the model 3×3 has the worst effect, and the SNR of processed signal is about 1 to 1.5 dB lower than that of model 7×7. The effect of model 5×5 is better than that of model 3×3, but it is worse than that of model 9×9. The comparative analysis of these data shows that not the larger the size and the more the number of filters, the stronger the learning ability of the model. The effect of model 9×9 is also good, but still is 0.2~0.4 dB lower than that of model 7×7, and the training time of model 9×9 is almost twice as long as that of model 7×7. Consequently, the SNR can not be improved by increasing the number and size of the filter without limitation. In Fig. 4, a broken line chart is given to compare the effects of the four models. Compared with the table, it is more intuitive.

Table 1. Denoising results of different models and different methods.

SNR	model3×3	model5×5	model7×7	model9×9	Traditional co-sparse analysis model	TFPF	Wavelet denoising	F-X
4.0443	8.0331	8.5398	9.4792	9.1678	4.3631	3.8857	7.2928	7.8092
0.4688	7.6696	8.0829	9.0138	8.6412	2.3838	1.4408	5.2587	7.2909
-3.3991	6.7357	7.1083	7.4449	7.3939	-0.3402	-1.9817	1.0286	6.1399
-6.0753	5.6583	5.8677	6.3224	6.0566	-0.9909	-4.4961	-1.9772	5.4022
-10.0016	4.1198	4.4230	4.6694	4.4070	-2.6139	-8.2239	-6.8841	4.2318



(a) (b)

Fig. 4. (a) Comparisons of four co-sparse analysis models. (b) Comparing the results of the proposed method with four traditional methods.

Synthetic data experiments

In order to verify the effectiveness of the proposed model for low-frequency noise attenuation, we designed a synthetic record for denoising testing. The pure seismic data consisting of multiple Ricker wavelets is given in Fig. 5(a). There are 120 tracks, 800 sampling points per track, and the sampling interval is 2 ms. The main frequency of the signal is 25 Hz, 30 Hz and 35 Hz. The events in this record are complex and have different characteristics such as crossover, linearity and non-linearity. At the top of the record are curved events with different curvatures and below are the intersecting events of one large oblique event and the other broken event. The broken events in this record include a slanted broken one and a straight broken one.

Fig. 5(b) is the synthetic desert low-frequency noise, and the frequency distribution is mainly below 25 Hz. Fig. 5(c) is the low SNR seismic data synthesized by (a) and (b). The SNR is -6.2357 dB. It can be seen from the synthetic record that the desert low-frequency noise has strong randomness and weak Gaussianity. The seismic signal produces a large deviation after adding the noise, and the straight event at the lower left is even almost submerged in noise. Fig. 5(d), 5(e) is the denoising result and the denoising residual by proposed method. The co-sparse analysis model adopted is model 7×7 trained with 12 times of synthetic desert low-frequency noise.

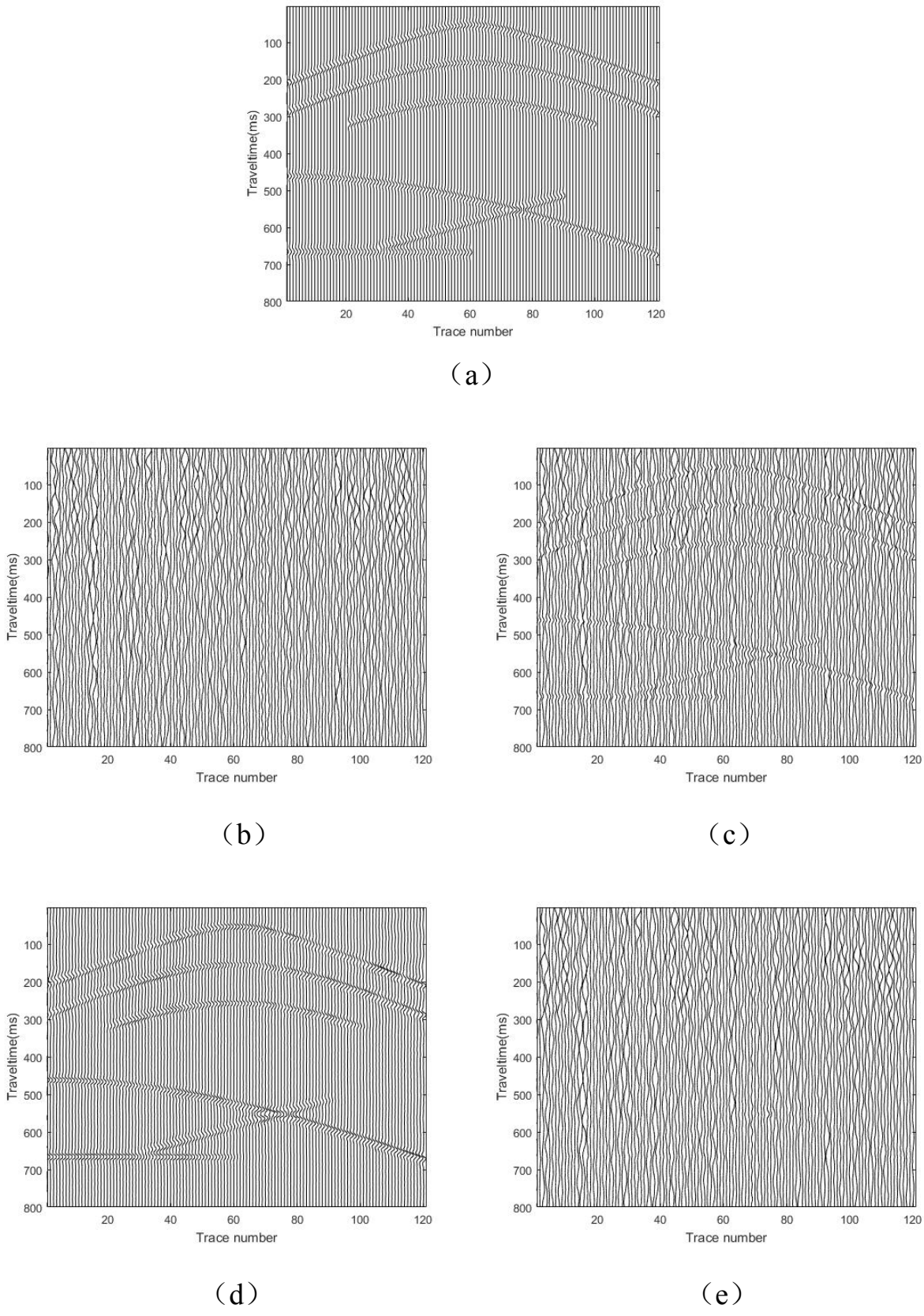


Fig. 5. (a) Synthetic seismic data.（b）Synthetic desert seismic exploration low-frequency noise. (c) Noisy synthetic seismic data (SNR = -6.2357 dB) with low-frequency noise. (d) Denoising result in the proposed method (SNR = 6.2062 dB).（e）Denoising residual.

Fig. 6 shows the frequency-wavenumber (FK) spectrum of Fig. 5. The frequency aliasing of seismic signals and low-frequency noise can be clearly seen in Fig. 6(c).

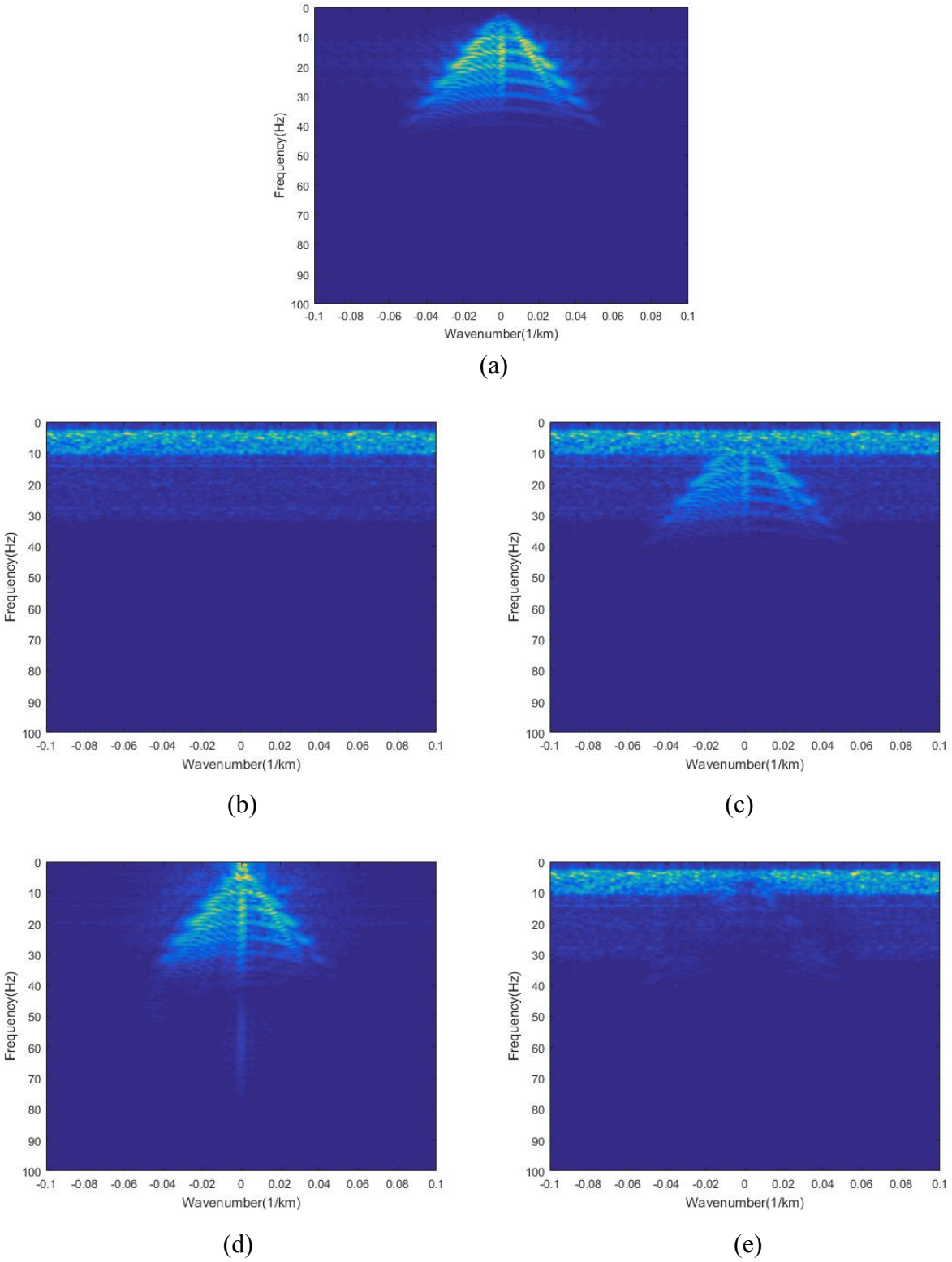


Fig. 6. (a) FK spectrum of synthetic seismic data. （b） FK spectrum of synthetic desert seismic exploration low-frequency noise.（c）FK spectrum of noisy synthetic seismic data. (d) FK spectrum of denoising result in the proposed method. （ e） FK spectrum of denoising residual.

It can be seen from Fig. 5(d) and Fig. 6(d) that the proposed method is effective for removing low-frequency noise. And for seismic events with different characteristics, the method restores them completely. In the residual diagram of Fig. 5(e), the residual signal is almost invisible. This shows that the method has a good amplitude-preserving ability, while the signal amplitude can be kept above 80%.

In order to further prove the advantages of the proposed method in removing low-frequency noise, we choose the traditional co-sparse analysis model, TFPF filtering, wavelet transform and *f-x* filtering to do the comparative experiments of the method, as shown in Figs. 7 and 8.

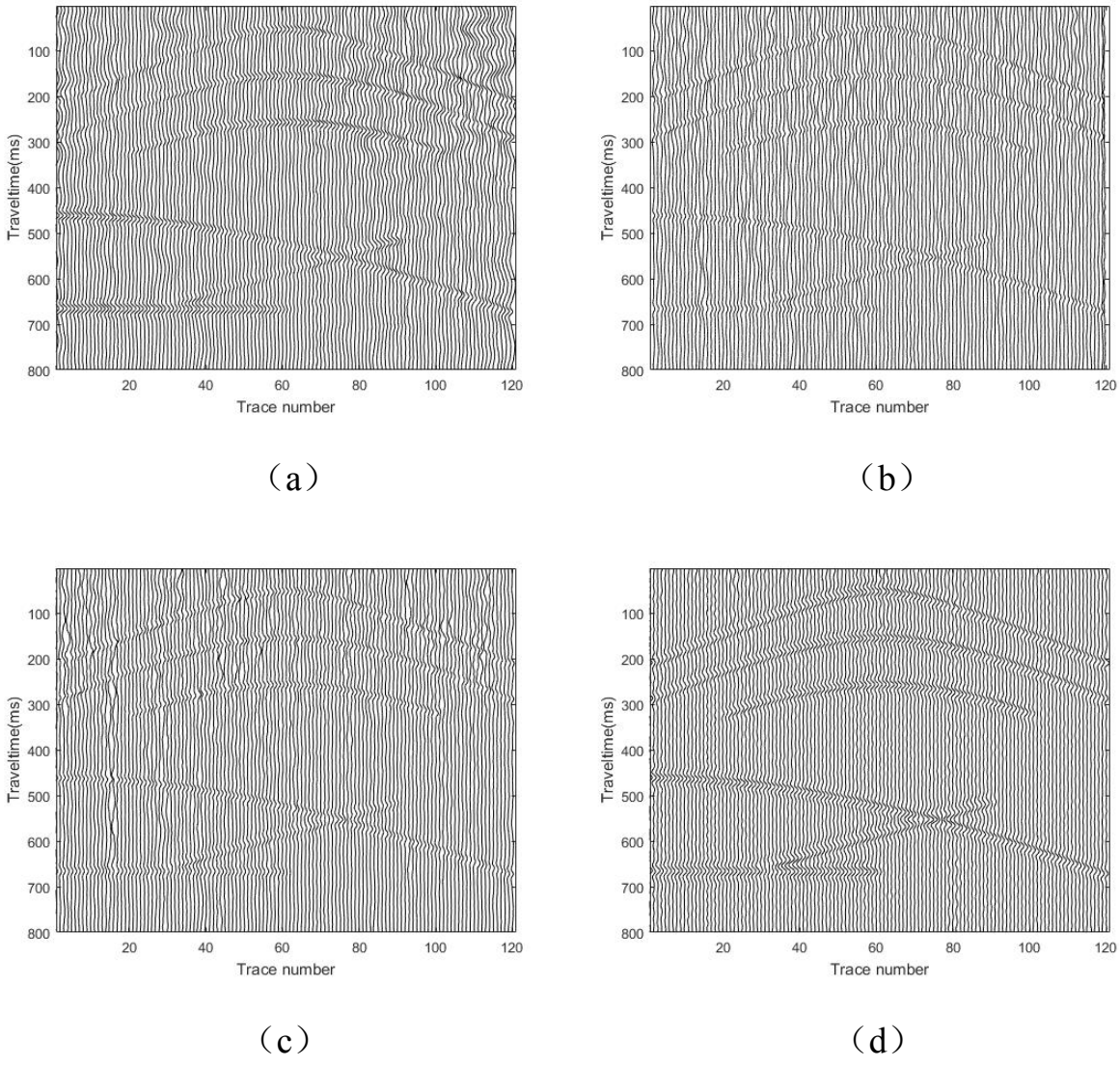


Fig. 7. (a) Denoising result after using traditional co-sparse analysis model.
(b) Denoising result after using TFPF filtering.（c）Denoising result after using wavelet transform.（d）Denoising result after using *f-x* filtering.

The denoising results of the traditional co-sparse analysis model are given in Fig. 7(a) and Fig. 8(a). Although the traditional co-sparse analysis model has a certain attenuation effect on the desert low-frequency noise, it also attenuates the signal largely. And for the low-frequency noise which overlaps with the signal frequency, it does not remove completely. The attenuation effect of TFPF on low-frequency noise seen from Fig. 7(b) and Fig. 8(b) is extremely limited - the noise and signal are not well separated. Wavelet denoising (Fig. 7(c) and Fig. 8(c)) is more suitable for removing noise in the middle and high frequency bands, and there is still more low-frequency noise in the processed synthetic record. It can be seen from Fig. 7(d) and Fig. 8(d) that *f-x* filtering is the best one for attenuating low-frequency noise compared to the other three methods, but it also has the disadvantage of removing the noise and signal in the low-frequency band together without extracting signal. Moreover, for the middle and high frequency band noise, it is not completely removed.

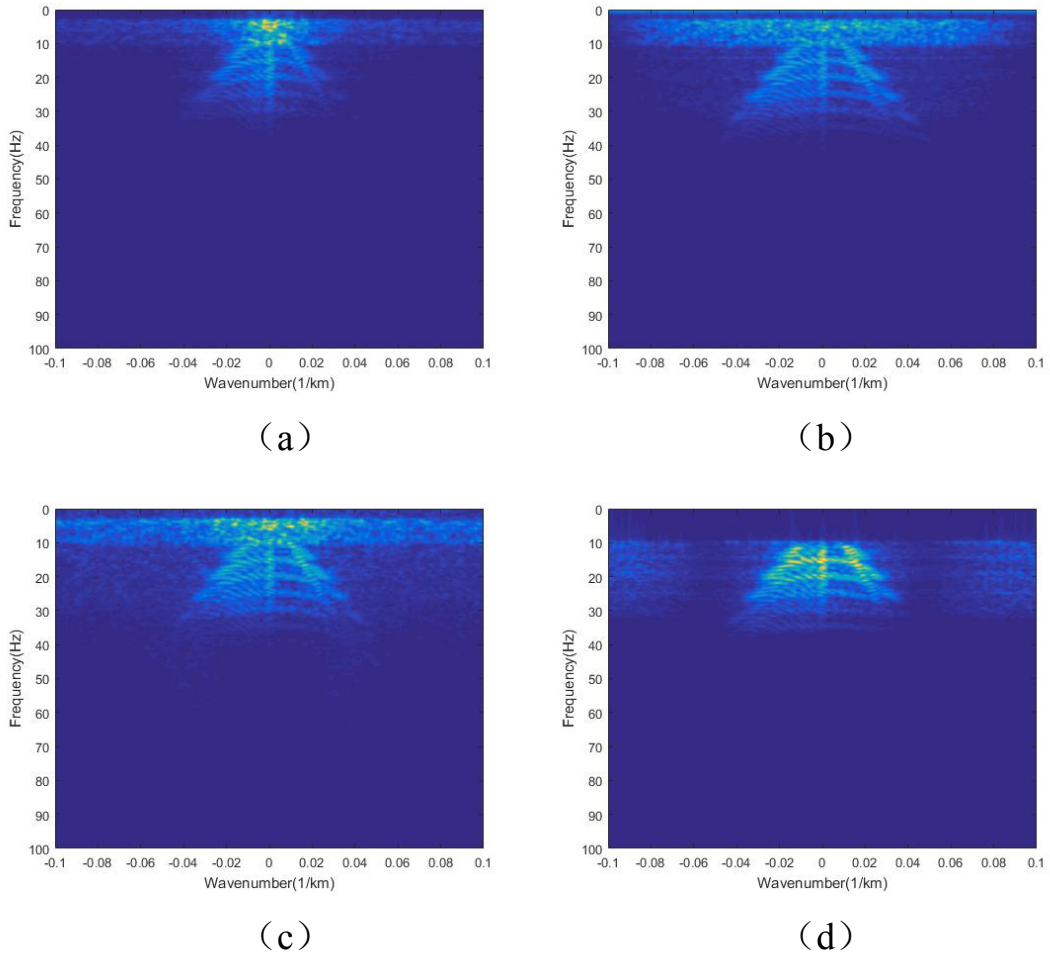


Fig. 8. (a) FK spectrum of denoising result after using traditional co-sparse analysis model.（b）FK spectrum of denoising result after using TFPF filtering.（c）FK spectrum of denoising result after using wavelet transform.（d）FK spectrum of denoising result after using *f-x* filtering.

In addition, more experimental data are given in detail in Table 1. From the comparison of these experimental data, it can also be concluded that the proposed co-sparse analysis model can improve the higher SNR and has greater competitiveness. This histogram in Fig. 4(b) shows this intuitively.

Real data experiments

We also selected a real seismic exploration record (Fig. 9(a)) for processing. The noise in real records is more complex - not only the desert low-frequency noise, but also some regular noise (surface wave) and high frequency random noise. It undoubtedly puts forward higher requirements for the performance of this method. Fig. 9(b) shows the result of the proposed method. At this point, the model 7×7 we used was trained by adding 12 times of real desert low-frequency noise. Figs. 9(c), 9(d), 9(e), and 9(f) are the results of the traditional co-sparse analysis model, TFPF filtering, wavelet transform and *f-x* filtering, respectively. As can be seen from Fig. 9(b), the events processed by proposed method are more clear and coherent. In the part of the upper right frame, the weaker events are

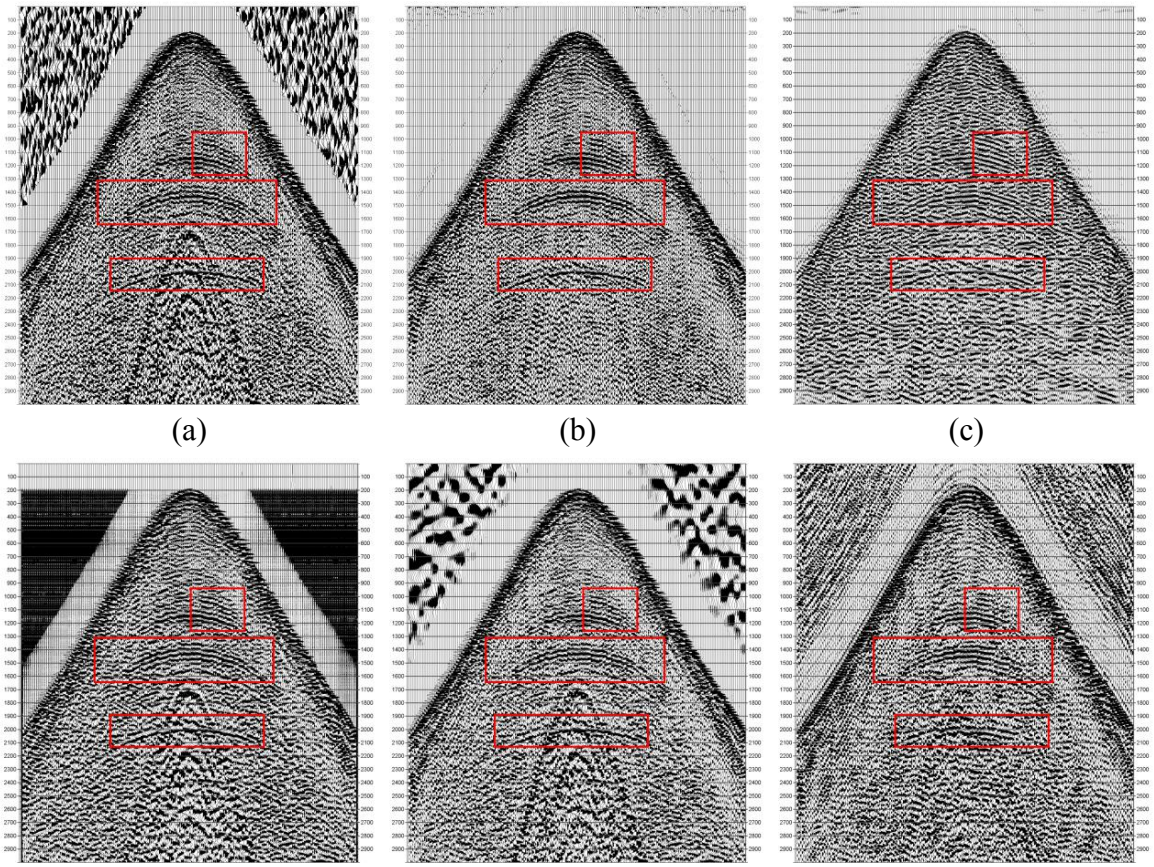


Fig. 9. (a) Real seismic data.（b）Denoising result in the proposed method.（c）Denoising result after using the traditional co-sparse analysis model.（d）Denoising result after using TFPF filtering.（e）Denoising result after using wavelet transform.（f）Denoising result after using *f-x* filtering.

completely recovered. And the proposed method also removes the surface wave relatively clean.

The residual diagram in Fig. 11(a) shows that the method not only attenuates the real desert noise and surface wave, but also has good amplitude preservation. There is no obvious signal residual in the residual diagram. In contrast, in Figs. 11(b), (c), (e), there are obvious signal residues. The result of the traditional co-sparse analysis model [Fig. 9(c)] still has obvious pseudo-axis problem. Although there is no signal residue in Fig. 11(d) of wavelet transform, its ability to remove noise is not strong, and it can not remove surface wave. So, unlike the proposed method, traditional methods are hard to give attention to two.

In order to highlight the advantages of the proposed method, the five methods are compared in the areas circled by black rectangles in Figs. 10 and 12. The areas contain a large number of effective signals. It can be seen from the processing results that although the traditional co-sparse model and *f-x* filtering suppress noise well [Figs. 10(c) and 10(f)], they have a great attenuation of the seismic signal [Figs. 12(b) and 12(e)]. For wavelet denoising and TFPF filtering, the denoising effect is weak [Figs. 10(d) and 10(e)]. Especially for the former, the noise at the white rectangle [Fig. 12(d)] can hardly be removed. Only the proposed method suppresses the noise and improve the continuity of seismic signal.

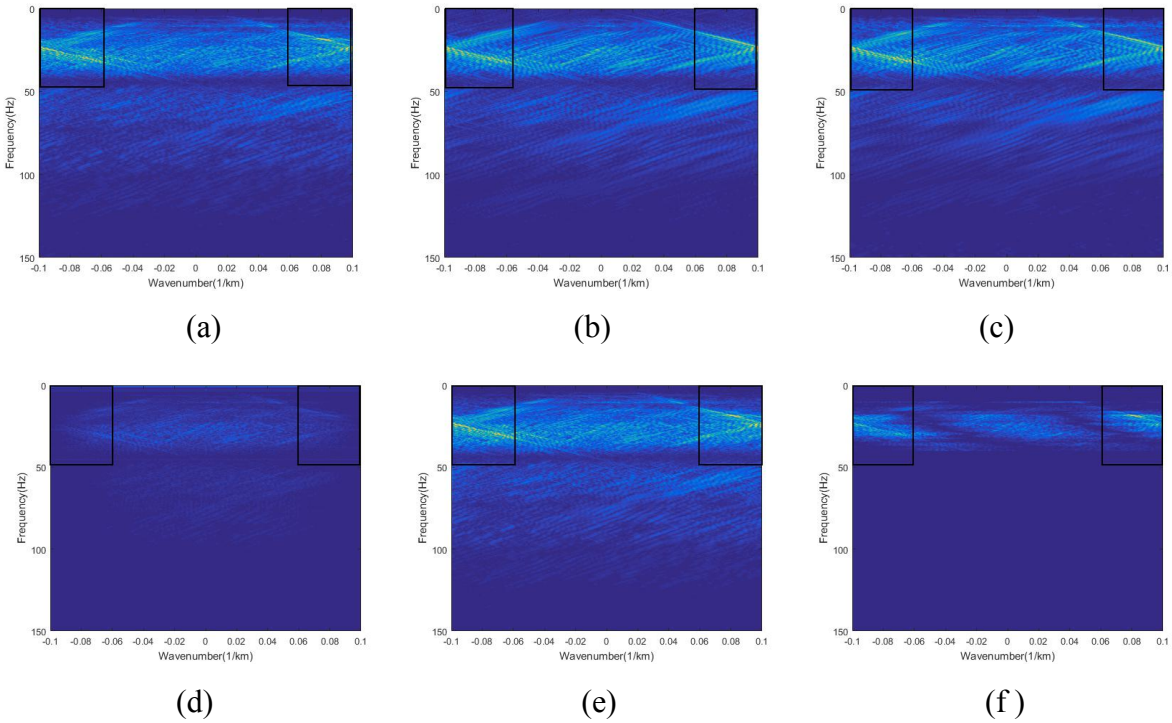
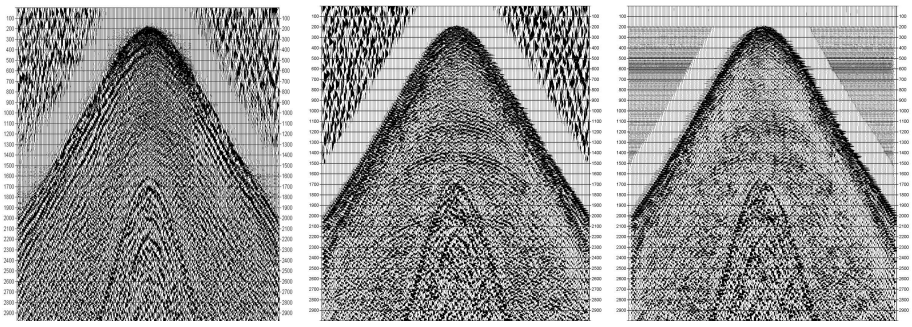
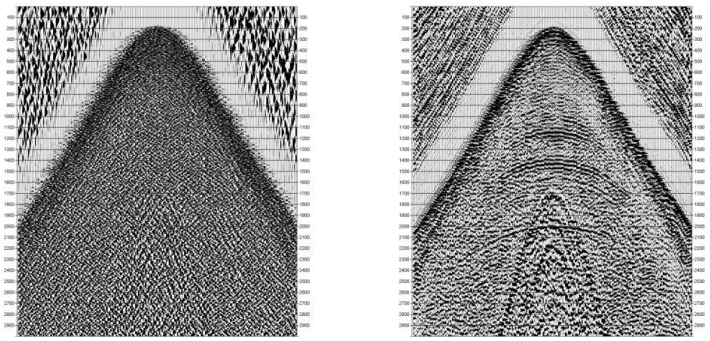


Fig. 10. (a) FK spectrum of real seismic data. （b）FK spectrum of denoising result in the proposed method. (c) FK spectrum of denoising result after using the traditional co-sparse analysis model. (d) FK spectrum of denoising result after using TFPF filtering. (e) FK spectrum of denoising result after using wavelet transform. （f） FK spectrum of denoising result after using *f-x* filtering.

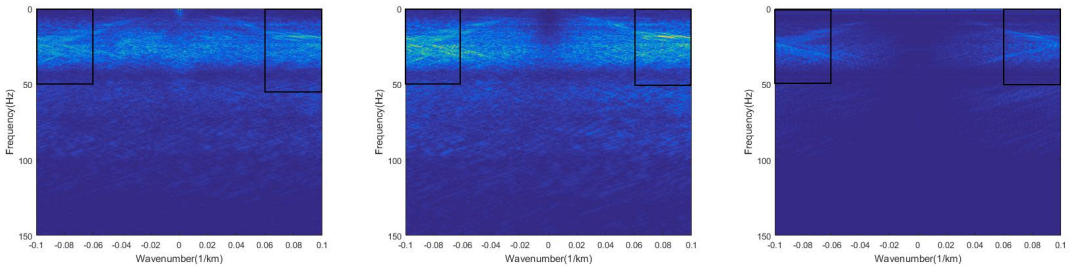


(a) (b) (c)

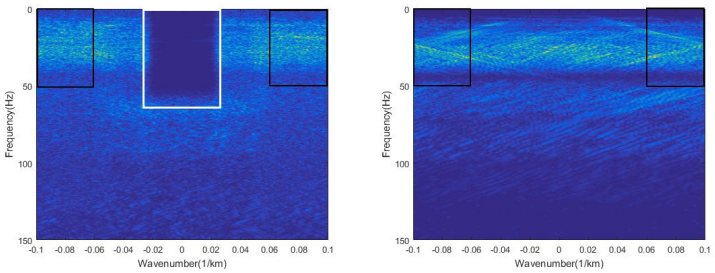


(d) (e)

Fig. 11. (a) Denoising residual in the proposed method. (b) Denoising residual after using traditional co-sparse analysis model. (c) Denoising residual after using TFPF filtering. (d) Denoising residual after using wavelet transform. (e) Denoising residual after using *f-x* filtering.



(a) (b) (c)



(d) (e)

Fig. 12. FK spectrum of denoising residual in the proposed method.（b）FK spectrum of denoising residual after using traditional co-sparse analysis model.（c）FK spectrum of denoising residual after using TFPF filtering.（d）FK spectrum of denoising residual after using wavelet transform.（e）FK spectrum of denoising residual after using *f-x* filtering.

CONCLUSION and DISCUSSION

We propose an improved co-sparse analysis model, which uses half-quadratic optimization to solve the problem of energy minimization at first. Then, according to the additive extension of half-quadratic optimization, shrinkage function is introduced into the co-sparse analysis model. By replacing the traditional sparsity promoting function with the shrinkage function, we improve the flexibility of the model and enable the model to learn the internal characteristics of the seismic signal more efficiently. This will accurately reconstruct the signal. Finally, the synthetic data show that the co-sparse analysis model can not only remove the desert low-frequency noise effectively when the signal and noise are severely frequency aliasing, but also has little attenuation to the seismic signal. The experimental results of real seismic exploration records show that the proposed model also has some advantages in attenuating high frequency random noise and regular noise. Obviously, compared with the traditional co-sparse analysis model, the improved model has better performance and stronger competitiveness in processing low SNR seismic exploration records.

Future work: Although the co-sparse analysis model proposed in this paper has a certain advantage in the attenuation of desert low-frequency noise, it still has the disadvantage of insufficient learning ability which can only train 90 training data. Therefore, the future work will focus on how to improve the learning ability of the model to achieve better effect.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundations of China under Grants No. 41730422.

REFERENCES

Candès, E.J. and Donoho, D.L., 2010. New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities. Communicat. Pure Appl. Mathemat., 57: 219-266.

Chen, Y., Ranftl, R. and Pock, T., 2014. Insights into analysis operator learning: from patch-based sparse models to higher order MRFs. IEEE Transact. Image Process., 23:1060-72.

Elad, M. and Aharon, M., 2006. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transact. Image Process., 15: 3736-3745.

Elad, M., Milanfar, P. and Rubinstein, R., 2007. Analysis versus synthesis in signal priors. Inverse Probl., 23: 947-968.

Emmanuel, J.C. and Donoho, D.L., 2000. Curvelets and reconstruction of images from noisy radon data. Proc. SPIE - The International Society for Optical Engineering, 4119.

Geman, D. and Yang, C., 1994. Nonlinear image recovery with half-quadratic regularization. IEEE Transact. Image Process., 4: 932.

Geman, D. and Reynolds, G., 2002. Constrained restoration and the recovery of discontinuities. IEEE Transact. Patt. Analys. Mach. Intellig., 14: 367-383.

Gilboa, G., Sochen, N. and Zeevi, Y.Y., 2004. Image enhancement and denoising by complex diffusion processes. IEEE Transact. Patt. Analys. Mach. Intellig., 26: 1036.

Harris, P.E. and White, R.E., 2010. Improving the performance of f-x prediction filtering at low signal-to-noise ratios. Geophys. Prosp., 45: 269-302.

Hawe, S. , Kleinsteuber, M. and Diepold, K., 2013. Analysis operator learning and its application to image reconstruction. IEEE Transact. Image Process., 22: 2138-2150.

Hel-Or, Y. and Shaked, D., 2008. A discriminative approach for wavelet denoising. IEEE Transact. Image Process., 17: 443-457.

Krishnan, D. and Fergus, R., 2009. Dark flash photography. ACM Transact. Graphics, 28: 1- 11.

Li, G., Yue, L. and Yang, B., 2017. Seismic exploration random noise on land: modeling and application to noise suppression. IEEE Transact. Geosci. Remote Sens., 99: 1-14.

Liu, D.C. and Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Mathemat. Progr., 45: 503-528.

Ma, J., Plonka, G. and Chauris, H., 2010. A new sparse representation of seismic data using adaptive easy-path wavelet transform. IEEE Geosci. Remote Sens. Lett., 7: 540-544.

Pati, Y.C., Rezaiifar, R. and Krishnaprasad, P.S., 2002. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. Conf. Signals, Systems Computing, Pacific Grove, CA.

Rodriguez, I.V., Sacchi, M.D. and Gu, Y.J., 2012. Simultaneous recovery of origin time, hypocentre location and seismic moment tensor using sparse representation theory. Geophys. J. Internat., 188:1188-1202.

Roth, S. and Black, M.J., 2009. Fields of experts. Internat. J. Comput. Vis., 82: 205.

Schmidt, U. and Roth, S., 2014. Shrinkage Fields for Effective Image Restoration. IEEE Conf. Computer Vision Pattern Recognition (CVPR). IEEE Computer Society, Hilton Head Island, SC.

Thode, H.C., 1987. Statistical tools for simulation practitioners. Technometr., 30: 464-464.

Wang, Y., Yang, J., Yin, W. and Zhang, Y., 2016. A new alternating minimization algorithm for total variation image reconstruction. SIAM J. Imag. Sci., 1: 248-272.

Yaghoobi, M., Nam, S., Gribonval, R. and Davies, M., 2012. Noise aware analysis operator learning for approximately cosparse signals. IEEE Internat. Conf. Acoustics, Speech and Signal Processing (ICASSP), Kyoto: 5409-5412.

Yaghoobi, M., Nam, S., Gribonval, R. and Davies, M., 2015. Analysis operator learning for overcomplete cosparse representations. 19th European Signal Processing IEEE Conf., Barcelona: 1470-1474.

Zhong, T., Li, Y., Wu, N., Nie, P. and Yang, B., 2015. Statistical analysis of background noise in seismic prospecting. Geophys. Prosp., 63: 1161-1174.