Artificial Intelligence in Geosciences 2 (2021) 223–233

Contents lists available at ScienceDirect

Artificial Intelligence in Geosciences

journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences

# Sparse inversion-based seismic random noise attenuation via self-paced learning

Yang Yang [a,b], Zhiguo Wang [b,c,*], Jinghuai Gao [a,b,**], Naihao Liu [a,b], Zhen Li [a,b]

[a] The School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China
[b] National Engineering Laboratory for Offshore Oil Exploration, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China
[c] The School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China

A B S T R A C T

Seismic random noise reduction is an important task in seismic data processing at the Chinese loess plateau area, which benefits the geologic structure interpretation and further reservoir prediction. The sparse inversion is one of the widely used tools for seismic random noise reduction, which is often solved via the sparse approximation with a regularization term. The $\ell_1$ norm and total variation (TV) regularization term are two commonly used regularization terms. However, the $\ell_1$ norm is only a relaxation of the $\ell_0$ norm, which cannot always provide a sparse result. The TV regularization term may provide unexpected staircase artifacts. To avoid these disadvantages, we proposed a workflow for seismic random noise reduction by using the self-paced learning (SPL) scheme and a sparse representation (i.e. the continuous wavelet transform, CWT) with a mixed norm regularization, which includes the $\ell_p$ norm and the TV regularization. In the implementation, the SPL, which is inspired by human cognitive learning, is introduced to avoid the bad minima of the non-convex cost function. The SPL can first select the high signal-to-noise ratio (SNR) seismic data and then gradually select the low SNR seismic data into the proposed workflow. Moreover, the generalized Beta wavelet (GBW) is adopted as the basic wavelet of the CWT to better match for seismic wavelets and then obtain a more localized time-frequency (TF) representation. It should be noted that the GBW can easily constitute a tight frame, which saves the calculation time. Synthetic and field data examples are adopted to demonstrate the effectiveness of the proposed workflow for effectively suppressing seismic random noises and accurately preserving valid seismic reflections.

## 1. Introduction

The loess plateau of the Ordos Basin in northwest China is one of the main target regions for oil and gas exploration, especially the shale oil in China. The topography and landform of the loess plateau are complex and its surface layer contains too thick Quaternary loess, which results in serious noises in the collected 3D field data. Several types of incoherence (random) noises have a bad influence on the successive seismic signal processing and interpretation (Gu et al., 2021; Qu et al., 2021; Li et al., 2021). Therefore, how to improve the signal-to-noise ratio (SNR) of the collected field data is one of the main research contents for the loess plateau area in China.

During past decades, many denoising methods have been proposed for filtering seismic data, such as the predictive filtering-based methods (Chen and Sacchi, 2017), the mode decomposition-based methods (Yu and Ma, 2018), the rank reduction-based methods (Zhang et al., 2020), the sparse transform-based methods (Wang et al., 2019) and machine learning-based methods (Zhong et al., 2021). Among these, the sparse transform-based methods are proposed to reduce seismic random noise based on various time-frequency (TF) transforms (Liu et al., 2016, 2019; Xue et al., 2021). For the noisy seismic data, the TF energy of the random noises distributes among the whole transform space, while the TF energy of the valid seismic data is often restricted in a small TF space (Wang and Gao, 2013). In a word, the TF spectrum of the valid seismic data is sparse. Therefore, the sparse inversion with a sparse regularization is widely used in the sparse transform-based random noise methods (Yuan et al., 2012). The $\ell_1$ norm, as one of the commonly utilized sparse regularization, has been applied in many sparse transforms to reduce seismic random noises, containing the wavelet transform (Wang and Gao, 2013), the curvelet transform (Wang et al., 2010), the synchrosqueezing transform (Anvari et al., 2017). However, the $\ell_1$ norm, which is only a relaxed version of the $\ell_0$ norm, may underestimate the high-amplitude components of the sparse results. Meanwhile, it may lead to abnormal results due to the soft-thresholding strategy (Ding and Selesnick, 2015).

To alleviate the above disadvantages, TV regularization is introduced in sparse inversion, which can accurately maintain the structure edges. Ma and Plonka (2007) proposed a curvelet-based denoising method with a projected TV diffusion. Lari and Gholami (2014) proposed a curvelet-TV regularized seismic random noise reduction method, in which the Bregman iteration method is introduced to address this

* Corresponding author. The School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China.
** Corresponding author. The School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China.
*E-mail addresses:* emailwzg@gmail.com (Z. Wang), jhgao@xjtu.edu.cn (J. Gao).

https://doi.org/10.1016/j.aiig.2022.03.003
Received 19 January 2022; Received in revised form 22 March 2022; Accepted 22 March 2022
Available online 19 April 2022

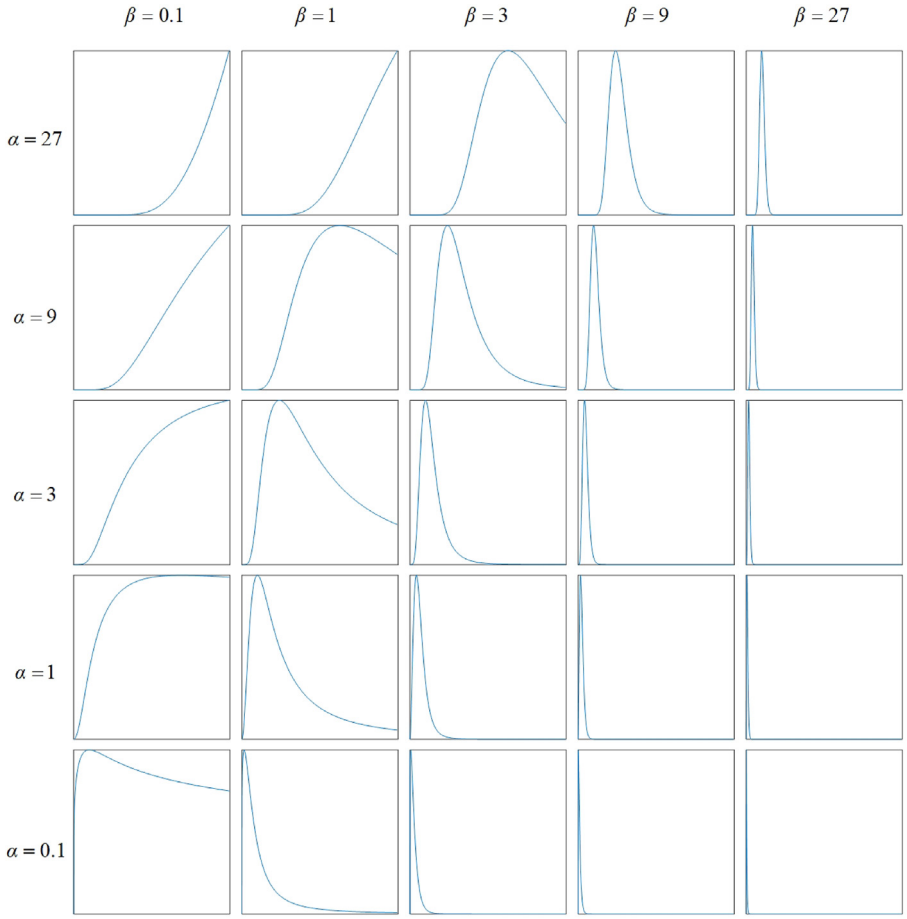


**Fig. 1.** The Fourier spectrum of the GBW with different $a_{\alpha,\beta}$.

optimization model. However, the TV regularization, as a piece-wise regularization, may produce undesirable staircase artifacts (Ma and Plonka, 2007; Selesnick et al., 2020).

Recently, the $\ell_p$ norm regularization Gonin and Money (2017), which can provide the sparser solution, is widely used in the sparse inversion-based random noise reduction method. However, the $\ell_p$ norm-based cost function is non-convex and the solution will suffer from a bad minima. To alleviate this problem, curriculum learning (CL), which is inspired by the humans and animals learning way, is proposed (Bengio et al., 2009; Kumar et al., 2010). The CL will first train the easy samples and then select the complex samples into learning. Although the CL can obtain an accurate solution for the non-convex optimization problem, it has a well-known limitation, which needs prior knowledge to define the easy and complex samples. In practice, the definitions of easy and complex samples are difficult. Self-paced learning (SPL) is present to identify the easy and complex samples adaptively (Meng et al., 2017). Self-paced learning can first select the simple samples into the training scheme and then automatically learn the complex samples. In every training scheme, the simple and complex samples are determined by the SPL model itself. Thus, the SPL can avoid the bad minima for the above problem and obtain an accurate result.

According to the above analysis, we proposed a seismic random noise attenuation workflow based on the SPL and a mixed norm (named SRNA-SPL-MN). The mixed norm, including the TV regularization and the $\ell_p$ norm, can not only address the main drawbacks of the traditional regularization but also generate accurate filtered results. The SPL scheme is introduced in this workflow to overcome the bad minima for the non-convex optimization problem. In the implementation, The CWT with the generalized Beta wavelet (GBW), which matches better for different seismic wavelets, is introduced in this study to generate a localized TF representation (Gao et al., 2006). Moreover, the GBW can constitute a tight frame, which benefits from saving the expensive calculation.

This paper is organized as follows. We first give a brief introduction of sparse theory. Then, we propose a random noise reduction model called SRNA-SPL-MN and then an effective split Bregman-based optimization algorithm is applied to obtain accurate filtered results. Next, synthetic and post-stacked field data examples are used on our proposed workflow, respectively. Finally, the conclusion is given.

## 2. Theory of sparse representation

In this section, we first review the sparse representation based on the GBW. Then, we describe the theory of the sparsity-based seismic random noise reduction method.

### 2.1. The generalized beta wavelet

In the past years, although several sparse transforms have been developed, the CWT still has an important impact on seismic data processing and interpretation (Wang and Gao, 2013). For a given signal $s(t)$, the CWT $WT(a, b)$ is defined as:

$$\begin{aligned} WT(a,b) &= \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} s(t)\psi^*\left(\frac{t-b}{a}\right)dt \\ &= \int_{-\infty}^{+\infty} s(t)\psi^*_{a,b}(t)dt \end{aligned} \tag{1}$$

where $\psi(t)$ is the mother wavelet, $\psi^*(t)$ is the conjugate of $\psi(t)$. $a$ and $b$ are frequency and time parameters, respectively.

For the CWT, selecting an appropriate basic wavelet is a key point, which should match for different seismic wavelets. The CWT with the best basic wavelet will provide a high localized TF representation. The Morlet wavelet is commonly used in seismic data processing and interpretation, which is given in the frequency domain as

$$\Psi_{\sigma}(\omega) = a_{\sigma} e^{-\frac{1}{2}(\omega-\sigma)^2}\left[1 - e^{-\omega\sigma}\right] \tag{2}$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. $\sigma$ is the radian frequency. $a_{\sigma}$ is a constant to maintain the normalization of the energy. Clearly, the Morlet wavelet is a wavelet with a Gaussian envelope. For a larger $\sigma(\sigma > 5.33)$, the second term of equation (2) can be ignored and the Morlet wavelet can be regarded as an analytic wavelet. However, the Morlet wavelet is not entirely analytic and does not match for the real seismic wavelets, which may lead to unsatisfactory TF localization (Wang and Gao, 2013).

To achieve a high TF localization, we utilize the recently proposed GBW as the new mother wavelet in this work. The GBW can achieve an adaptive trade-off between time and frequency localization, which is given in the frequency domain (Wang et al., 2017).

$$\Psi_{\alpha,\beta}(\omega) = U(\omega) a_{\alpha,\beta}\left(\frac{2\arctan(\omega)}{\pi}\right)^{\alpha}\left(1 - \frac{2\arctan(\omega)}{\pi}\right)^{\beta}, \tag{3}$$

where $\alpha$ and $\beta$ are parameters to control the shape of the GBW. $a_{\alpha,\beta}$ is a normalization constant and $U(\omega)$ is the unit step function. To achieve an adaptive trade-off between time and frequency localization, the parameters $\alpha$ and $\beta$ are selected to match the real seismic wavelet according to

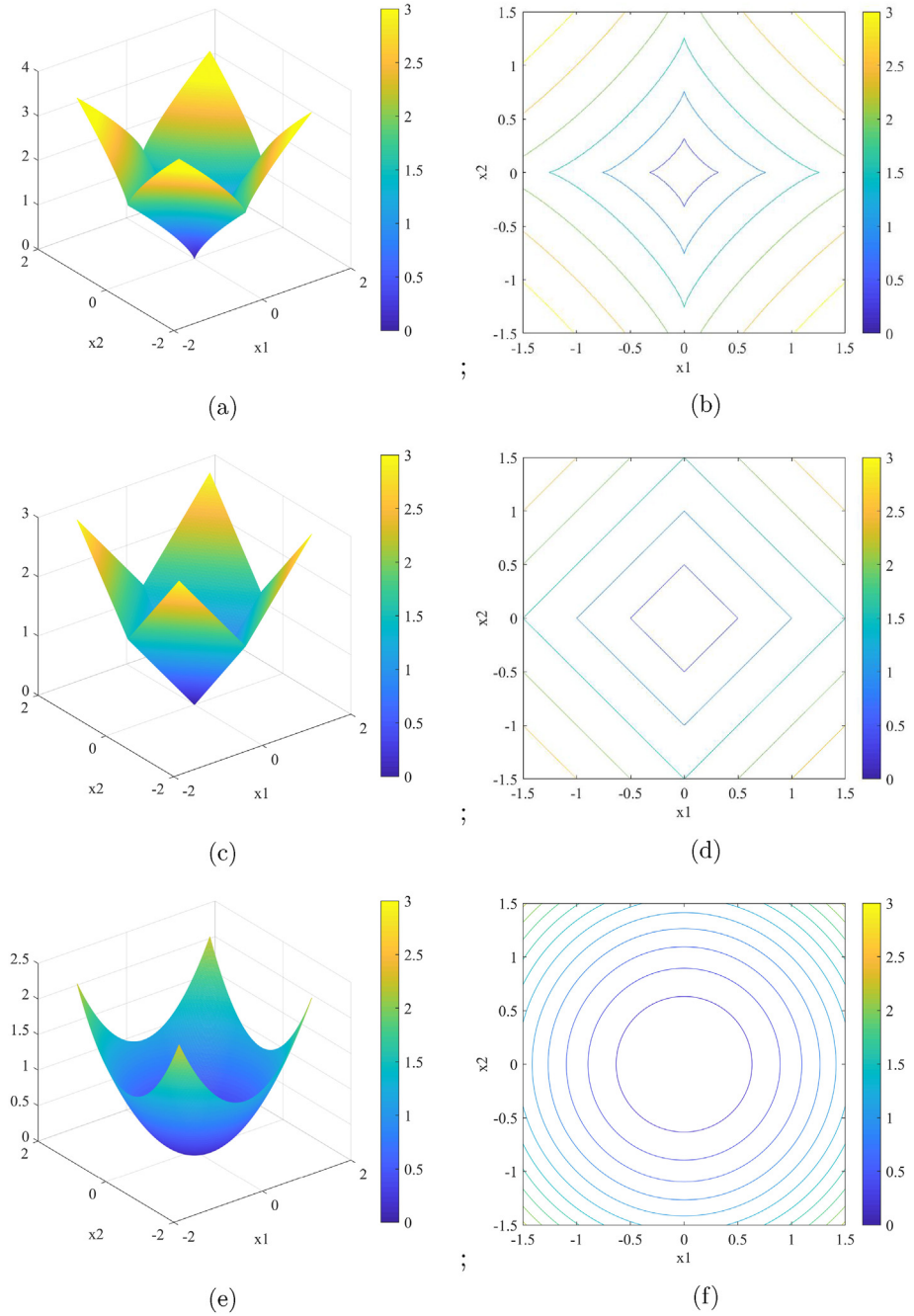


**Fig. 2.** The geometric sketch maps of different norms. (a) The $\ell_{p,p=0.8}$ norm, (b) the contour line of (a). (c) The $\ell_1$ norm, (d) the contour line of (c). (e) The $\ell_2$ norm, (f) the contour line of (e).

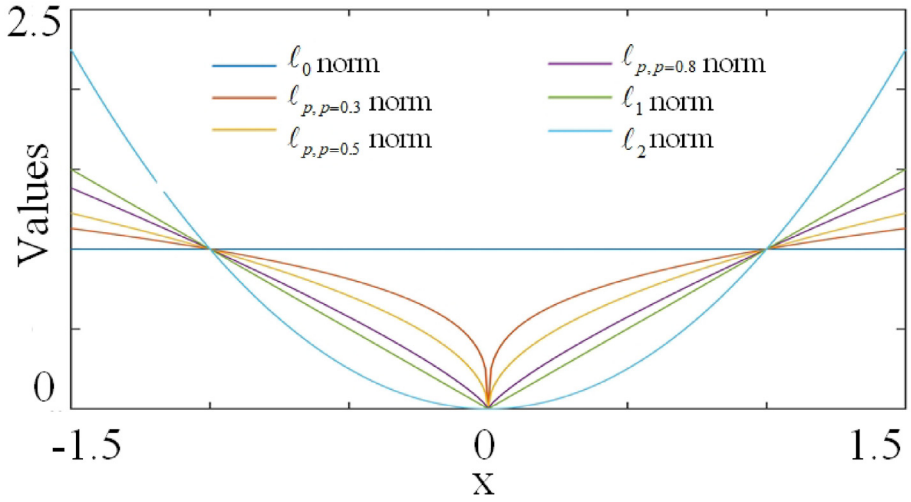


**Fig. 3.** The comparisons between different norm.

the estimation of the seismic amplitude spectrum (Gao et al., 2006, 2017). Fig. 1 displays the Fourier spectrum of the GBW with different $a_{\alpha,\beta}$. Clearly, selecting proper $a_{\alpha,\beta}$ can match different seismic wavelets. Then, a high localized TF representation can be obtained by the GBW.

Furthermore, the CWT can construct a tight frame in the certain condition based on the frame theory in Mallat (1999), which may benefit for the proposed seismic random noise reduction workflow in this paper (Wang and Gao, 2013). Thus, the CWT coefficients $\mathbf{x} = (x_{m,n})_{m,n\in N}$ of a signal $\mathbf{s}$ and the reconstruction CWT are defined as

$$\begin{aligned} \mathbf{x} &= \langle \mathbf{s}, \psi_{m,n} \rangle = \mathbf{F}^* \mathbf{s} \\ \mathbf{s} &= \mathbf{F}\mathbf{x} = \mathbf{F}\mathbf{F}^* \mathbf{s} \end{aligned}, \tag{4}$$

where $\psi_{m,n} = [\psi_{m,n}(t)]_{m,n\in\mathbb{N}}$ denotes the wavelet family. $\langle \cdot, \cdot \rangle$ is the inner product. The adjoint operator $\mathbf{F}^*$ is defined from $L^2(\mathbb{R})$ to $\ell^2(\mathbb{N}^2)$, while the frame operator $\mathbf{F}$ represents from $\ell^2(\mathbb{N}^2)$ to $L^2(\mathbb{R})$.

### 2.2. The sparsity-based seismic random noise reduction

In seismic data processing, the observed noisy seismic signal includes the noise-free seismic data and the random noise. Thus, the noisy seismic signal $\mathbf{y} \in \mathbb{R}^{N\times 1}$ can be modeled as

$$\mathbf{y} = \mathbf{s} + \mathbf{n}, \tag{5}$$

where $\mathbf{s} \in \mathbb{R}^{N\times 1}$ is the noise-free seismic signal and $\mathbf{n} \in \mathbb{R}^{N\times 1}$ is the white Gaussian noise. A great challenge of random noise reduction method is how to accurate estimate the noise-free seismic signal from the noisy seismic signal.

From above analysis, the noise-free seismic signal can been recovered from its inner products based on the tight frame analysis operator (Mallat, 1999). Thus, based on equation (4) and sparse representation theory, the approximation $\tilde{\mathbf{x}}$ can be obtained by a minimized optimization problem

$$\mathbf{x}_{opt} = \arg\min \frac{1}{2}\|\mathbf{y} - \mathbf{F}\mathbf{x}\|_2^2 + \lambda\varphi(\mathbf{x}), \tag{6}$$

where $\varphi(\mathbf{x})$ is the regularization term, which maintains some characters, such as sparsity and smoothness. $\lambda$ is a regularization parameter. Obviously, due to the sparsity of the noise-free seismic signal in transform domain, the $\ell 1$ norm regularization term is commonly used in this problem. Then, the minimization problem can be a lasso problem,

$$\mathbf{x}_{opt} = \arg\min \frac{1}{2}\|\mathbf{y} - \mathbf{F}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{7}$$

where $\|\mathbf{x}\|_1 = \sum_i |x_i|$ is the $\ell 1$ norm and $x_i$ denotes the $i$th element of $\mathbf{x}$. Many optimization methods have been utilized to solve such problem (Daubechies et al., 2004; Beck and Teboulle, 2009) and these algorithms are implemented by an iterative updating scheme and a soft-thresholding strategy, which is given by

$$soft(x,\lambda) = \begin{cases} x - \lambda\dfrac{x}{|x|}, & if|x| > \lambda \\ 0, & if|x| < \lambda \end{cases}. \tag{8}$$

However, the $\ell_1$ norm has some drawbacks. One is that the $\ell_1$ norm is the approximate representation of the $\ell_0$ norm and cannot generate the sparsest solution. Therefore, the $\ell_1$-minimization problem will lead to underestimated results for the high-amplitude component (Selesnick, 2017). On the other hand, the thresholding strategy is always introduced to solve the $\ell_1$-minimization problem and an inappropriate thresholding parameter will lead to undesirable results.

To alleviate the disadvantage of $\ell_1$ norm, the TV regularization, which can preserve edges, is addressed in seismic random noise attenuation. The TV regularization optimization problem is given

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{F}\mathbf{x}\|_2^2 + \lambda_2\|\mathbf{F}\mathbf{x}\|_{TV}, \tag{9}$$

where $\lambda_2$ is the regularization parameter and $\|\cdot\|_{TV}$ denotes the TV regularization term, which is defined as

$$\|\mathbf{x}\|_{TV} = \|\mathbf{D}\mathbf{x}\|_1 \tag{10}$$

where $\mathbf{D}$ represents

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}. \tag{11}$$

Although the TV regularization has good performance to reduce seismic random noise, it may provide unexpected staircase artifacts. Meanwhile, according to equation (9), the TV regularization is generated by the $\ell_1$ norm, which suffers from the similar limitation of the $\ell_1$ norm.

## 3. The proposed seismic random noise reduction method based on self-paced learning

### 3.1. The $\ell_p$ norm regularization

According to above analysis, the $\ell_0$ norm is used to provide the sparsest solution of equation (6). However, the $\ell_0$ norm-based optimization problem is a non-deterministic polynomial-time hardness (NP-Hard) problem, which is difficult to solve. Recently, the $\ell_p$, $0 < p < 1$-based problem is proposed, which is defined as

$$\ell_p = \|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}, \tag{12}$$

where $x_i$ is the $i$th element of $\mathbf{x}$. Fig. 2 describes geometric sketch maps of the $\ell_{p,p=0.8}$, the $\ell_1$ norm and the $\ell_2$ norm, respectively. Clearly, the $\ell_2$ norm only provides smoothness and the $\ell_{p,p=0.8}$ norm maintain much sparser than the $\ell_1$ norm. Fig. 3 also gives the comparison for different norms. Obviously, the $\ell_{p,0<p<1}$ norm can preserve the sparsest result, which can be introduced in the sparsity-based model.

### 3.2. The self-paced learning

Recently, a machine learning algorithm, called curriculum learning (CL), is proposed. The CL, inspired by the humans and animals learning scheme, can learn the easy samples and then select the complex samples. Many empirical evaluations have demonstrated that the CL can avoid the bad minima for a non-convex problem. However, the CL needed a prior to determine the simple and complex samples, which was difficult in the real application.

To overcome above disadvantage, the SPL was addressed by Kumar et al. (2010), which formulated the key principle of CL as an optimal problem. The SPL can select the simple and complex samples automatically, which is defined as

$$\min_{\mathbf{x},\mathbf{w}} \sum_{i=1}^{n} \omega_i e(y_i, f(x_i)) + h(\mathbf{w}, \eta) \tag{13}$$

where $e$ is the loss function. $\mathbf{w} = [\omega_i]$ and $\omega_i$ is the $i$th weighted value on samples. $h(\mathbf{w}, \eta)$ is the SPL regularization to calculate the weighted values. $\eta$ is the age parameter to control the learning pace. With the parameter $\eta$ increasing in every iteration, more complex samples will be introduced in the learning scheme. The SPL regularization $h(\mathbf{w}, \eta)$ has a general definition, which is given as follows:

**Definition** (SPL regularization term (Zhao et al., 2015)) Suppose $w$ is the weighted values, $e$ is the loss function, $\eta$ is the age parameter, $h(w, \eta)$ is the SPL regularization term, if.

1. $h(w, \eta)$ is convex with respect to $w \in [0, 1]$.
2. $w(\eta, e)$ is monotonically decreasing with respect to $e$, and it holds that $\lim_{e\to 0} w(\eta, e) = 1$, $\lim_{e\to\infty} w(\eta, e) = 0$.
3. $w(\eta, e)$ is monotonically increasing with respect to $\eta$, and it holds that $\lim_{\eta\to 0} w(\eta, e) = 0$, $\lim_{\eta\to\infty} w(\eta, e) \le 1$.

where $w(\eta, e) = \arg\min_{w\in[0,1]} we + h(w, \eta)$.

The above definition gives an easy understanding for the SPL. The condition 2 demonstrates that the SPL can first select easy samples. The condition 3 indicates that when the age parameter becomes larger, the SPL may include more complex samples into training. Compared with the CL, the SPL can automatically select the easy or complex samples by learner itself.

### 3.3. The proposed SRNA-SPL-MN model

According to the advantage of the $\ell_{p,0<p<1}$ norm and the SPL, we proposed a mixed $\ell_p$ norm-based seismic random noise reduction method to avoid drawbacks of the single regularization. In our proposed method, the mixed norm, including the $\ell_{p,0<p<1}$ norm and the TV regularization, can boost the advantages of the sparse regularization and the TV regularization, while provide more accurate denoised results. The SPL is also applied in this model to avoid the bad minima, which is introduced by the $\ell_p$ norm regularization term. Thereby, the proposed optimization model is given by

$$\begin{aligned} \min_{\mathbf{x},\mathbf{w}} \frac{1}{2}\left\|\sqrt{\mathbf{w}} \odot (\mathbf{y} - \mathbf{F}\mathbf{x})\right\|_2^2 + \lambda_1\|\mathbf{x}\|_p^p + \lambda_2\|\mathbf{F}\mathbf{x}\|_{TV} + h(\mathbf{w}, \eta), \\ s.t. \mathbf{w} \in [0, 1], \end{aligned} \tag{14}$$

where $\lambda_1$ and $\lambda_2$ are regularization parameters. $\odot$ represents the Hadamard product of matrices. $\sqrt{\mathbf{w}}$ is the element-wise square root of $\mathbf{w}$. The second term $\|\mathbf{x}\|_p^p$ represents the $\ell_{p,0<p<1}$ norm. $\|\mathbf{F}\mathbf{x}\|_{TV}$ is the TV regularization. $h(\mathbf{w}, \eta)$ demotes the SPL regularization term to select the training samples. The detailed properties of the SPL regularization term is given in (Yang et al., 2019).

To further understand the SPL training processing, a simple SPL regularization, called hard regularization, is used, which is defined as

$$h(\omega, \eta) = -\eta\omega \tag{15}$$

Then, we introduce this SPL regularization into equation (14). When $\mathbf{x}$ is fixed, the equation (14) can be rewritten as

$$\begin{aligned} \min_{\mathbf{w}} \sum_i \omega_i e_i + \left(-\sum_i \eta\omega_i\right), \\ s.t. \mathbf{w} \in [0, 1], \end{aligned} \tag{16}$$

where $e_i = 0.5^*(y_i - (\mathbf{F}\mathbf{x})_i)$ is the loss function. Therefore, the solution $\omega_i^*(e_i, \eta)$ is given as

$$\omega_i^*(e_i, \eta) = \begin{cases} 1, e_i \le \eta \\ 0, \text{else} \end{cases}. \tag{17}$$

According to equation (16), the binary values are given for every sample. When the loss function $e_i$ is less than $\eta$, the samples are regarded as simple samples. With the $\eta$ increasing, more complex samples will be included in the training scheme.

### 3.4. The algorithm of the proposed model

Obviously, the proposed SRNA-SPL-MN model in equation (14) is non-convex and non-smooth, which is challenging to solve. However, when the parameter $\mathbf{x}$ is fixed, the equation (14) is a convex problem, which is easy to solve. While the parameter $\mathbf{w}$ is fixed, the equation (14) become a mixed norm-based optimal problem, which can be solved by the alternative direction method of multipliers (ADMM) algorithm (Shen and Cheng, 2016) and the split Bregman algorithm (Yin et al., 2008). Thus, we first divide the proposed SRNA-SPL-MN model into two sub-problem and then optimize each of them with other fixed.

At first, when the $\mathbf{x}$ is fixed, the equation (14) can be a regarded as

$$\min_{\mathbf{w}\in[0,1]} \sum_i \omega_i e_i + \sum_i h(\omega_i, \eta), \tag{18}$$

where $e_i = 0.5(y_i - (\mathbf{F}\mathbf{x})_i)$.

Recently, many SPL regularizations have been proposed (Yang et al., 2019). In this paper, we only consider the soft SPL regularization, which is defined by

$$h(\omega, \eta) = \frac{\gamma^2\eta}{\gamma + \omega\eta}, \tag{19}$$

where $\gamma$ is a constant. Therefore, the solution of equation (18) is given as

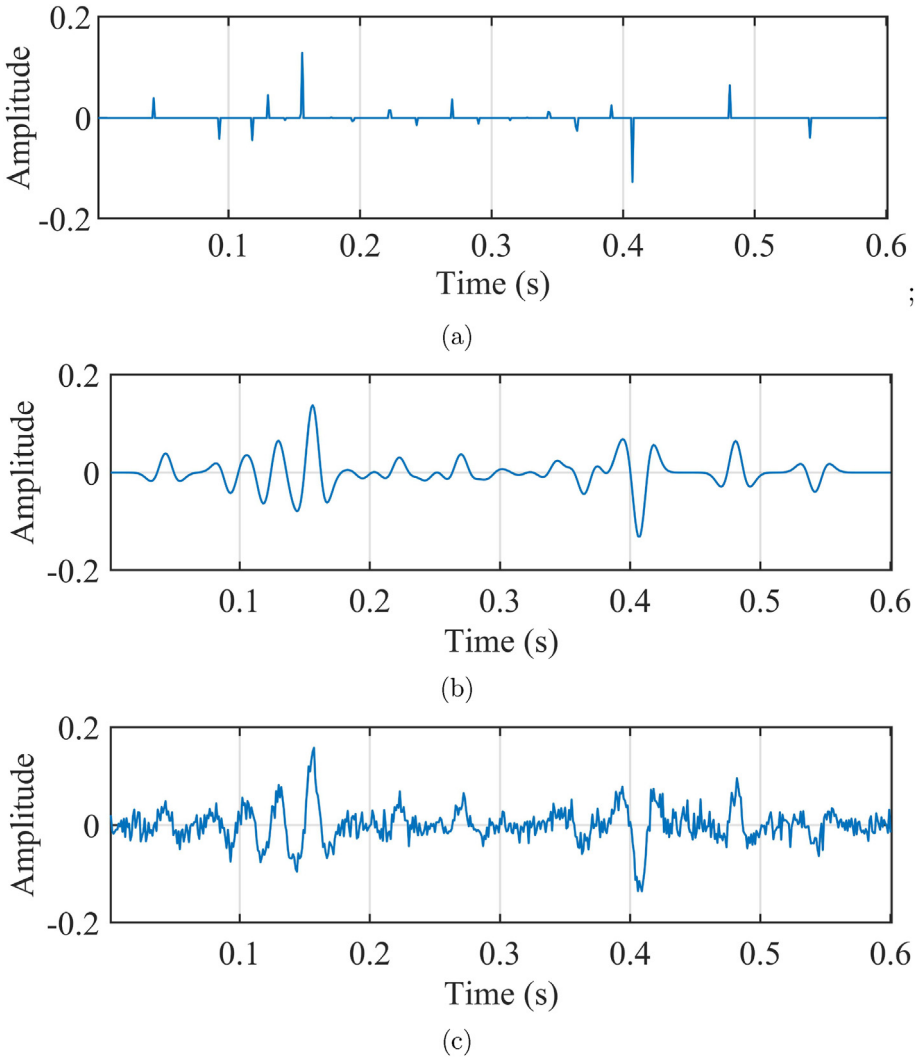


**Fig. 4.** (a) The reflection coefficient from the Marmousi model. (b) The synthetic noise-free seismic signal. (c) The noisy synthetic seismic signal.

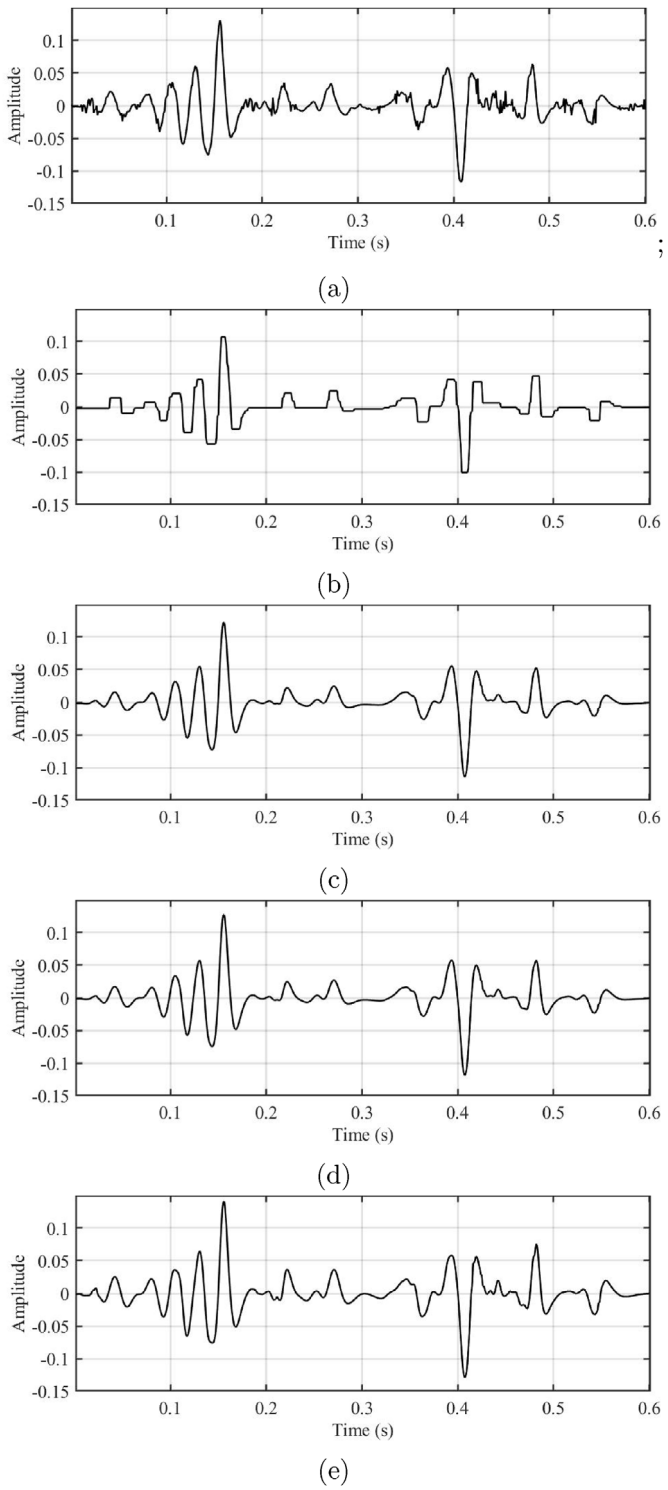


**Fig. 5.** The denoised results by (a) The L1-based method with SNR = 10.89 dB, (b) The TV denoising with SNR = 9.46 dB, (c) The mixed norm with p = 1 and SNR = 11.74 dB, (d) The mixed norm with p = 1 and SNR = 12.58 dB, (e) The proposed method with p = 0.5 and SNR = 13.48 dB.

$$\omega_i(e_i,\eta)=\begin{cases}1, \text{if} e_i \le \left(\dfrac{\gamma\eta}{\gamma+\eta}\right)^2,\\ 0, \text{if} e_i \ge (\eta)^2,\\ \gamma\left(\dfrac{1}{\sqrt{e_i}}-\dfrac{1}{\eta}\right), \text{otherwise.}\end{cases} \tag{20}$$

**Table 1**
The denoised results of Fig. 5 by different methods.

| Method | SNR (dB) |
|---|---|
| GBW-L1 | 10.89 |
| TV | 9.46 |
| GBW-M with $p = 1$ | 11.74 |
| GBW-M with $p = 0.5$ | 12.58 |
| SRNA-SPL-MN with $p = 0.5$ | 13.48 |

Secondly, when the parameter $\mathbf{w}$ is fixed, the equation (14) can be rewritten as

$$\min_{\mathbf{x}} \frac{1}{2}\left\|\sqrt{\mathbf{w}}\odot(\mathbf{y}-\mathbf{F}\mathbf{x})\right\|_2^2+\lambda_1\|\mathbf{x}\|_p^p+\lambda_2\|\mathbf{F}\mathbf{x}\|_{TV}, \tag{21}$$

Many method have been proposed to solve this non-convex problem, such as the alternative direction method of multipliers (ADMM) algorithm (Shen and Cheng, 2016) and the split Bregman algorithm (Yin et al., 2008). In this paper, we use the split Bregman algorithm to solve the equation (21). Based on the theory of the split Bregman algorithm (Yin et al., 2008), two variables $\mathbf{u}$ and $\mathbf{v}$ are introduced and then the equation (21) can be rewritten as

$$\begin{aligned}&\min_{\mathbf{x}} \frac{1}{2}\left\|\sqrt{\mathbf{w}}\odot(\mathbf{y}-\mathbf{F}\mathbf{x})\right\|_2^2+\lambda_1\|\mathbf{u}\|_p^p+\lambda_2\|\mathbf{v}\|_{TV}\\ &\quad s.t.\quad \mathbf{u}=\mathbf{x},\mathbf{v}=\mathbf{F}\mathbf{x}\end{aligned} \tag{22}$$

Then, equation (22) can be split as three sub-optimization problems based on Karush–Kuhn–Tucker (KKT) conditions (Yin et al., 2008), which is implemented as

$$\begin{cases}\mathbf{x}^{k+1}=\arg\min_{\mathbf{x}} \frac{1}{2}\left\|\sqrt{\mathbf{w}}\odot(\mathbf{y}-\mathbf{F}\mathbf{x})\right\|_2^2+\frac{\mu_1}{2}\left\|\mathbf{u}-\mathbf{x}-\mathbf{b}^k\right\|_2^2\\ \qquad\qquad +\frac{\mu_2}{2}\left\|\mathbf{v}-\mathbf{F}\mathbf{x}-\mathbf{c}^k\right\|_2^2\\ \mathbf{u}^{k+1}=\arg\min_{\mathbf{u}} \lambda_1\|\mathbf{u}\|_p^p++\frac{\mu_1}{2}\left\|\mathbf{u}-\mathbf{x}^{k+1}-\mathbf{b}^k\right\|_2^2\\ \mathbf{v}^{k+1}=\arg\min_{\mathbf{v}} \lambda_2\|\mathbf{v}\|_{TV}+\frac{\mu_2}{2}\left\|\mathbf{v}-\mathbf{F}\mathbf{x}^{k+1}-\mathbf{c}^k\right\|_2^2\\ \mathbf{b}^{k+1}=\mathbf{b}^k+(\mathbf{x}^{k+1}-\mathbf{u}^{k+1})\\ \mathbf{c}^{k+1}=\mathbf{c}^k+(\mathbf{F}\mathbf{x}^{k+1}-\mathbf{v}^{k+1})\end{cases} \tag{23}$$

where $\mu_1$ and $\mu_2$ are regularization parameters.

Obviously, the first sub-optimization problem in equation (23) is a least square issue, which is convex and easy to be solve. We set as

$$\begin{aligned}\Omega(x)&=\frac{1}{2}\left\|\sqrt{\mathbf{w}}\odot(\mathbf{y}-\mathbf{F}\mathbf{x})\right\|_2^2+\frac{\mu_1}{2}\left\|\mathbf{u}-\mathbf{x}-\mathbf{b}^k\right\|_2^2+\frac{\mu_2}{2}\left\|\mathbf{v}-\mathbf{F}\mathbf{x}-\mathbf{c}^k\right\|_2^2\\ &=\frac{1}{2}\left\|\tilde{\mathbf{w}}(\mathbf{y}-\mathbf{F}\mathbf{x})\right\|_2^2+\frac{\mu_1}{2}\left\|\mathbf{u}-\mathbf{x}-\mathbf{b}^k\right\|_2^2+\frac{\mu_2}{2}\left\|\mathbf{v}-\mathbf{F}\mathbf{x}-\mathbf{c}^k\right\|_2^2\end{aligned} \tag{24}$$

where $\tilde{\mathbf{w}}=diag(\sqrt{\mathbf{w}})$ is a diagonal matrix. Then, we set $\frac{\partial\Omega(\mathbf{x})}{\partial\mathbf{x}}=0$ and the solution is given as

$$\mathbf{x}^{k+1}=\mathbf{B}^{-1}((\tilde{\mathbf{w}}\mathbf{F})^T\tilde{\mathbf{w}}\mathbf{y}+\mu_1\mathbf{e}_1^k+\mu_2\mathbf{F}^T\mathbf{e}_2^k) \tag{25}$$

where $\mathbf{e}_1^k=\mathbf{u}^k-\mathbf{b}^k$, $\mathbf{e}_2^k=\mathbf{v}^k-\mathbf{c}^k$ and $\mathbf{B}=(\tilde{\mathbf{w}}\mathbf{F})^T(\tilde{\mathbf{w}}\mathbf{F})+\mu_2(\mathbf{F})^T\mathbf{F}+\mu_1\mathbf{I}$.

The second sub-optimization issue in equation (23) is a $\ell_p$ minimization issue and the shrinkage strategy is utilized to solve above sub-optimization issue, which is defined by

$$\mathbf{u}^{k+1}=shrink_p\left(\mathbf{x}^{k+1}+\mathbf{b}^k,\frac{\lambda_1}{\mu_1}\right) \tag{26}$$

Here, the result $\mathbf{u}$ of second sub-optimization issue is a complex number. For a complex number $\mathbf{u}=|\mathbf{u}|e^{j\theta}$, the shrinkage strategy is defined as follows,

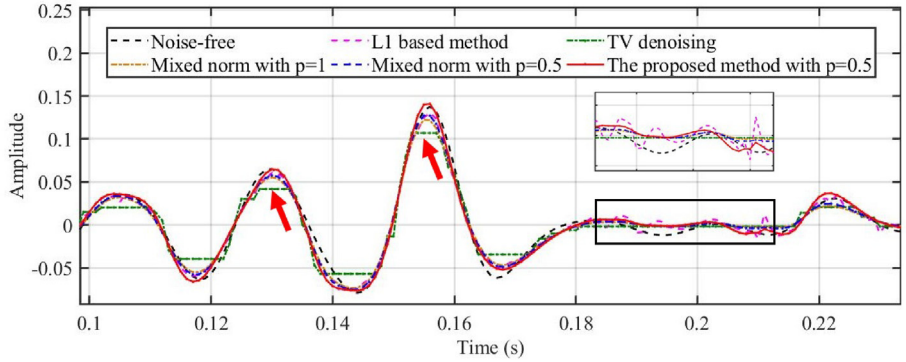


**Fig. 6.** The enlarged denoised results from 0.1s to 0.23s.

$$shrink_p(\mathbf{u},\lambda) = \max(|\mathbf{u}| - \lambda^{2-p}|\mathbf{u}|^{p-1}, 0)e^{i\theta}, \tag{27}$$

where $\theta$ is the phase of this complex number.

At last, the third one is also the TV minimization problem, which can be solved by (Condat, 2013),

$$\mathbf{v}^{k+1} = tvd\left(\mathbf{F}\mathbf{x}^{k+1} + \mathbf{c}^k, \frac{\lambda_2}{\mu_2}\right) \tag{28}$$

where the detailed definition of the $tvd(\cdot,\cdot)$ can be found in (Condat, 2013)).

Totally, our proposed SRNA-SPL-MN algorithm is illustrated in Algorithm 1. The regularization parameters in Algorithm 1 is still an open issue and thus these parameters are given by experimental verification.

**Algorithm 1.** The proposed SRNA-SPL-MN algorithm

**Algorithm 1** The proposed SRNA-SPL-MN algorithm

**Input:** $\mathbf{y}$, $\lambda_1$, $\lambda_2$, $\mu_1$, $\mu_2$, $\mathbf{B} = (\widetilde{\mathbf{w}}\mathbf{F})^T(\widetilde{\mathbf{w}}\mathbf{F}) + \mu_2(\mathbf{F})^T\mathbf{F} + \mu_1\mathbf{I}$, **the SPL parameter** $\eta$.

**Output:** $\hat{\mathbf{y}}$.

calculate $e_i^0 = \frac{1}{2}(y_i^0 - [\mathbf{F}^0\mathbf{x}^0]_i)^2$ and $\mathbf{w}^0$

**while** $\eta_k < \eta_{end}$ **do**

  $\mathbf{e}_1^k = \mathbf{u}^k - \mathbf{b}^k$ and $\mathbf{e}_2^k = \mathbf{v}^k - \mathbf{c}^k$

  $\mathbf{x}^{k+1} = \mathbf{B}^{-1}((\widetilde{\mathbf{w}}\mathbf{F})^T\widetilde{\mathbf{w}}\mathbf{y} + \mu_1\mathbf{e}_1^k + \mu_2\mathbf{F}^T\mathbf{e}_2^k)$

  $\mathbf{u}^{k+1} = shrink_p(\mathbf{x}^{k+1} + \mathbf{b}^k, \frac{\lambda_1}{\mu_1})$

  $\mathbf{v}^{k+1} = tvd(\mathbf{F}\mathbf{x}^{k+1} + \mathbf{c}^k, \frac{\lambda_2}{\mu_2})$

  $\mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{x}^{k+1} - \mathbf{u}^{k+1})$

  $\mathbf{c}^{k+1} = \mathbf{c}^k + (\mathbf{F}\mathbf{x}^{k+1} - \mathbf{v}^k)$

**end while**

Denoised results $\widehat{\mathbf{y}} = \mathbf{F}\widehat{\mathbf{x}}$

## 4. Synthetic examples

In this section, a synthetic seismic signal is introduced to demonstrate the performance of our proposed seismic random noise reduction method. We utilized the SNR to measure the effectiveness of our proposed method, which is given by

$$SNR = 10\log_{10}\left(\frac{\|\mathbf{s}\|^2}{\|\mathbf{y} - \mathbf{s}\|^2}\right)(dB), \tag{29}$$

where $\mathbf{y}$ represents the noisy seismic signal and $\mathbf{x}$ is the denoised result.

The noise-free synthetic seismic signal is shown in Fig. 4. Fig. 4(a) is

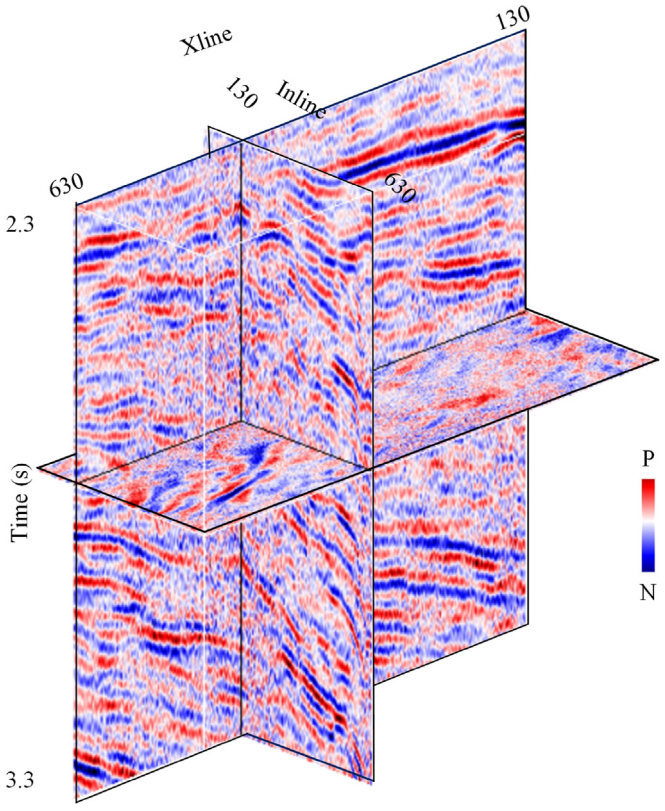


**Fig. 7.** The 3D real seismic data, which includes 300 Inline traces and 300 Xline traces, respectively. The time sample interval is 2 ms and the time sample number is 1000 at every trace.

the reflection, which is from the Marmousi model. Fig. 4(b) represents the noise-free seismic signal, which is given by convolving a 35 Hz Ricker wavelet on Fig. 4(a). The time sampling number is 601 and the time sampling interval is 1 ms. The noisy synthetic seismic signal, which is generated by adding White Gaussian noise, is given in 4(c) and the SNR value is 5 dB. To demonstrate the performance of the proposed seismic random noise reduction method, three comparison methods are applied in this example, including the $\ell_1$ norm regularized sparse random noise reduction method (named GBW-L1), the TV denoising method, the mixed norm-based denoising method (called GBW-M). The regularization parameter of the GBW-L1 is selected based on the noise level (Yang et al., 2019). The TV regularization parameter selection strategy is given in Ding and Selesnick (2015). For the GBW-M method, we also use the Bregman algorithm to solve the optimal solution. The parameter $p$ of the GBW-M is set as $p = 1$ and $p = 0.5$, respectively. Unfortunately, the regularization parameters selection in the GBW-M and our proposed SRNA-SPL-MN method are still an open problem and thus these parameters are obtained by empirical validation.

Fig. 5 describes the denoised results by above-mentioned methods and the SNR values of these methods are listed in table Table 1. Clearly, the GBW-M method and our proposed method performer better than the TV denoised method and GBW-L1 method. In order to describe the comparison among these denoised methods in detail, Fig. 6 gives the enlarged denoised results of Fig. 5 from 0.1s to 0.23s and the denoised results in the black box is enlarged. The TV denoised method, which has the lowest SNR value (SNR = 11.92 dB), produces the unexpected staircase artifacts denoted by the red arrows, while some effective signals are leakage. The black box indicates that the GBW-L1 has some ineffective denoised results. For the GBW-M, when the parameter $p$ is set as 0.5, this method may lead to less effective signal leaking than the GBW-M with $p = 1$, which can be also demonstrated by the SNR value of the denoised results in table ?? Thus, the parameter $p$ of the proposed SRNA-SPL-MN is set 0.5 and the denoised result is represented by the red line in Fig. 5. Obviously, the proposed SRNA-SPL-MN with the highest SNR (SNR = 13.48 dB) can reduce the random noise more effectively and also ensure a more significant signal.

## 5. Field data applications

To further demonstrate the effectiveness of our proposed SRNA-SPL-MN method, an example of 3D real seismic data is given in this section. The 3D real seismic is shown in Fig. 7, which comes from the loess plateau area in China. The real 3D seismic data includes 500 Inline traces and 500 Xline traces, respectively. The time sample interval is 2 ms and the time sample number is 500. We also select GBW-L1 and GBW-M with $p = 0.5$ methods as the comparison methods. Following the analysis of the synthetic example, we set $p = 0.5$ in this application.

The denoised results and the corresponding difference sections of these three methods are shown in Figs. 8 and 9, respectively. From the difference sections in Fig. 9, these three mentioned methods make little effective seismic signal leaking. Unfortunately, the GBW-$\ell_1$ attenuates a little random noise especially within the black arrows in Fig. 8(c). From Fig. 8(a) and (b), our proposed method has the best performance to reduce random noises and maintain the effective signals. Furthermore, the seismic traces of Xline 300 and Inline 300 are given in Fig. 10. Clearly, the proposed method can ensure more effective seismic signal than the other two denoised methods. Then, the normalized Fourier spectrum of these four seismic traces is illustrated in Fig. 11 and the results reveal that the proposed method attenuate more random noise, especially in the high frequency part (shown in the black box).

The denoised results.

## 6. Conclusion and discussion

We propose a sparse transform-based seismic random noise reduction method via mixed norm regularization and self-paced learning in this study. The regularized term is consisted of a mixed norm, including the $\ell_p$ norm and the TV regularization. This mixed norm can not only boost the advantages of the $\ell_1$ norm and the TV regularization but also provide sparser results than the above two traditional regularizations. Meanwhile, SPL, which can avoid the bad minima for the mixed norm regularized optimization problem, is introduced in the proposed method. In the implementation, the SPL selects the high SNR seismic data first into the mixed norm regularized problem and then learns low SNR seismic data. Moreover, an enhanced Bregman iteration algorithm is addressed to solve the above optimization model. In a word, the proposed random noise reduction method has the best performance for improving the SNR values for the seismic data, especially for the complex environment. To show the performance of our proposed workflow, a synthetic seismic signal and a 3D real loess plateau data examples are successfully applied. Compared with the two traditional denoising methods, our proposed workflow can reduce more random noises, while maintaining the valid seismic signals.

Due to the introduction of three regularization terms in the proposed model, the proposed method should take more experimental to set the regularization parameters. Although the proposed method boosts the performance of seismic data denoising, it is a significant task to improve the parameters selection scheme. Furthermore, the proposed model is only focus on the seismic random noise reduction in this paper, which can be extended to solve seismic data reconstruction and seismic random noise reduction simultaneously. Another topic of the future work could be to remove the seismic coherent noise.

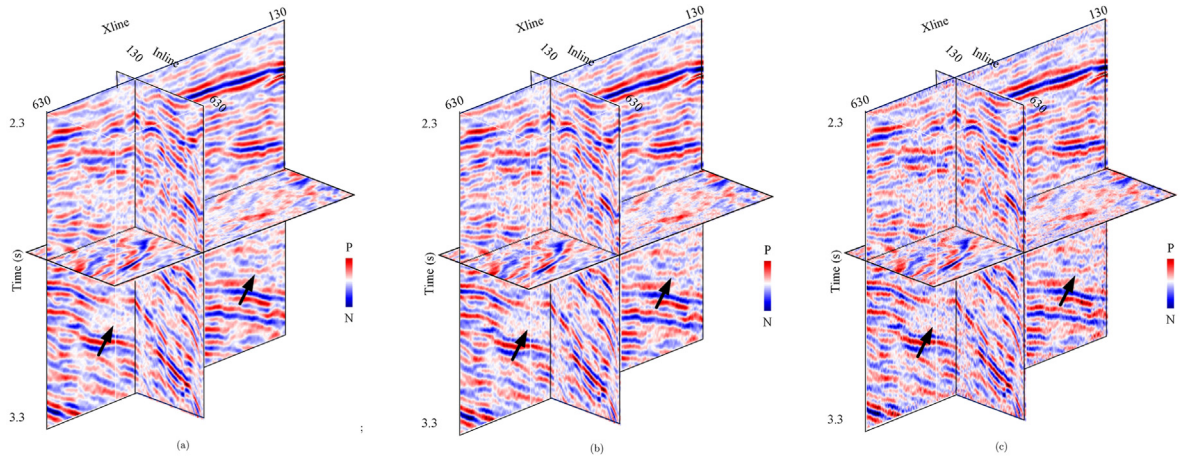


**Fig. 8.** The denoised results by (a) the proposed method, (b) the mixed norm-based method, (c) the $\ell_1$ norm-based method.

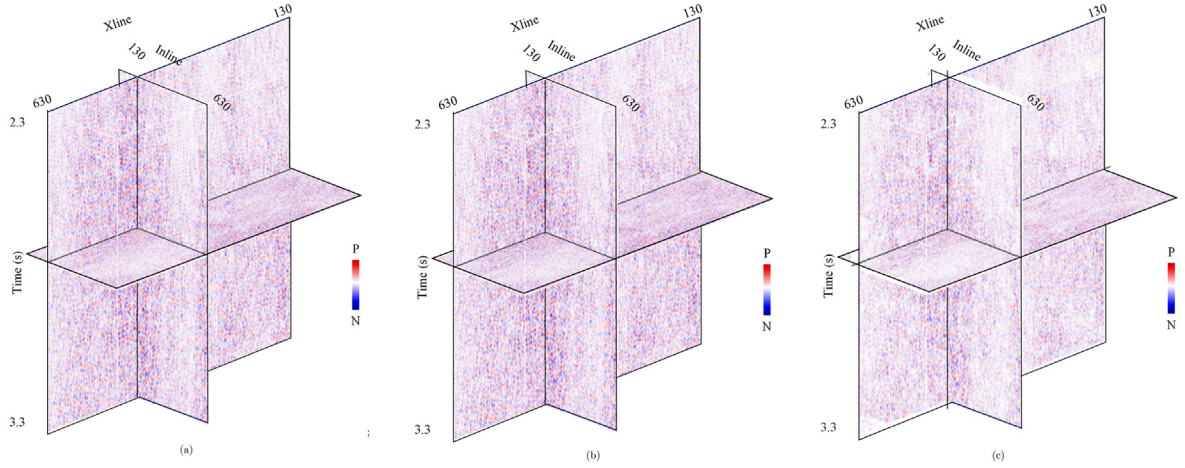


**Fig. 9.** The difference between the noisy real data and the denoised result by (a) the proposed method, (b) the mixed norm-based method, (c) the $\ell_1$ norm-based method.

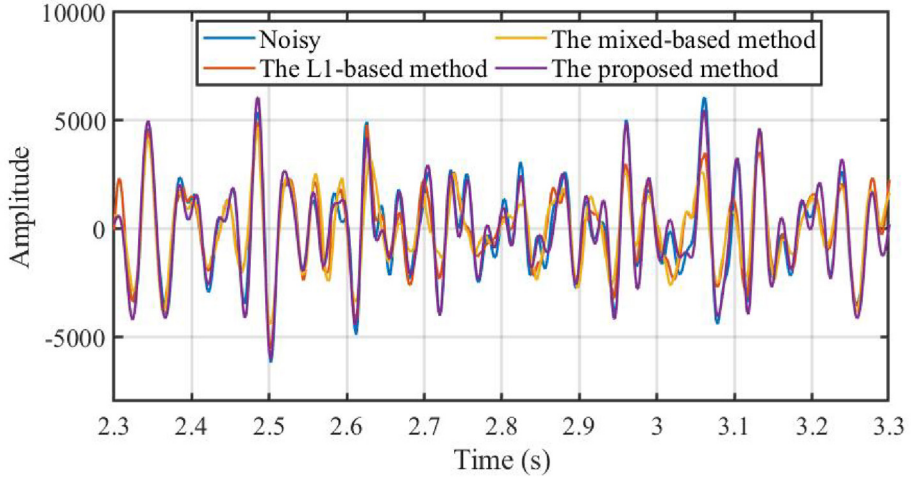


**Fig. 10.** The seismic traces of Xline 300 and Inline 300 by the noisy real data and the denoised results, including the proposed method, the mixed norm-based method, and the $\ell_1$ norm-based method.

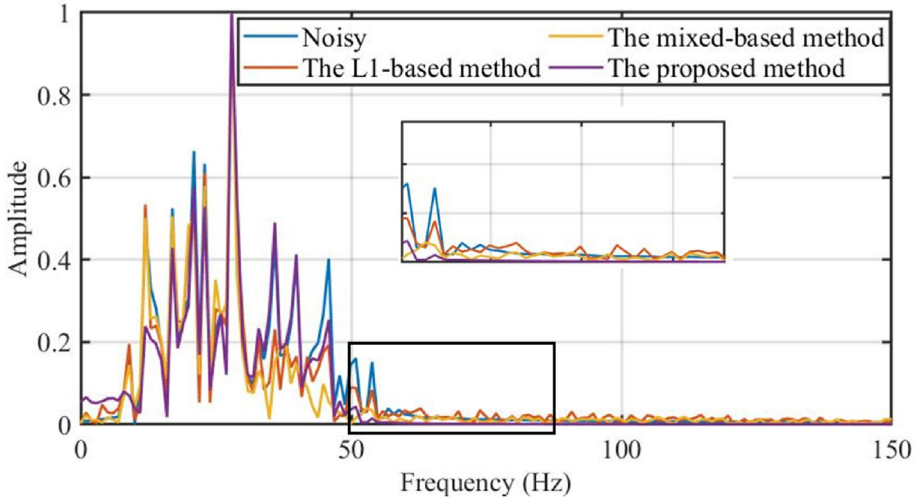


**Fig. 11.** The normalized Fourier spectrum by the noisy real data and the denoised results, including the proposed method, the mixed norm-based method, and the $\ell_1$ norm-based method.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 41974137, in part by the National Key R&D Program of China under Grants 2020YFA0713400.

## References

Anvari, R., Siahsar, M.A.N., Gholtashi, S., Kahoo, A.R., Mohammadi, M., 2017. Seismic random noise attenuation using synchrosqueezed wavelet transform and low-rank signal matrix approximation. IEEE Trans. Geosci. Rem. Sens. 55, 6574–6581.
Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. 2, 183–202.
Bengio, Y., Louradour, J., Collobert, R., Weston, J., 2009. Curriculum Learning: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 41–48.
Chen, K., Sacchi, M.D., 2017. Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation. Geophys. Prospect. 65, 650–668.
Condat, L., 2013. A direct algorithm for 1d total variation denoising. IEEE Signal Process. Lett. 20, 1054–1057.
Daubechies, I., Defrise, M., De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. 57, 1413–1457.
Ding, Y., Selesnick, I.W., 2015. Artifact-free wavelet denoising: non-convex sparse regularization, convex optimization. IEEE Signal Process. Lett. 22, 1364–1368.
Gao, J., Wan, T., Chen, W., Mao, J., 2006. Three parameter wavelet and its applications to seismic data processing. Chin. J. Geophys. Chin Ed. 49, 1802–1812.
Gao, J., Zhang, B., Han, W., Peng, J., Xu, Z., 2017. A new approach for extracting the amplitude spectrum of the seismic wavelet from the seismic traces. Inverse Probl. 33, 085005.
Gonin, R., Money, A.H., 2017. Nonlinear Lp-Norm Estimation. CRC Press.
Gu, M., Xie, R., Xiao, L., 2021. A novel method for nmr data denoising based on discrete cosine transform and variable length windows. J. Petrol. Sci. Eng. 207, 108852.
Kumar, M.P., Packer, B., Koller, D., 2010. Self-paced Learning for Latent Variable Models, vol. 2. NIPS.
Lari, H.H., Gholami, A., 2014. Curvelet-tv regularized bregman iteration for seismic random noise attenuation. J. Appl. Geophys. 109, 233–241.
Li, F., Sun, F., Liu, N., Xie, R., 2021. Denoising seismic signal via resampling local applicability functions. Geosci. Rem. Sens. Lett. IEEE 19, 7501605.
Liu, N., Gao, J., Zhang, B., Wang, Q., Jiang, X., 2019. Self-adaptive generalized s-transform and its application in seismic time–frequency analysis. IEEE Trans. Geosci. Rem. Sens. 57, 7849–7859.
Liu, W., Cao, S., Chen, Y., Zu, S., 2016. An effective approach to attenuate random noise based on compressive sensing and curvelet transform. J. Geophys. Eng. 13, 135–145.
Ma, J., Plonka, G., 2007. Combined curvelet shrinkage and nonlinear anisotropic diffusion. IEEE Trans. Image Process. 16, 2198–2206.
Mallat, S., 1999. A Wavelet Tour of Signal Processing. Elsevier.
Meng, D., Zhao, Q., Jiang, L., 2017. A theoretical understanding of self-paced learning. Inf. Sci. 414, 319–328.
Qu, F., Jiang, Q., Jin, G., Wei, Y., Wang, Z., 2021. Noise cancellation for continuous wave mud pulse telemetry based on empirical mode decomposition and particle swarm optimization. J. Petrol. Sci. Eng. 200, 108308.
Selesnick, I., 2017. Sparse regularization via convex analysis. IEEE Trans. Signal Process. 65, 4481–4494.
Selesnick, I., Lanza, A., Morigi, S., Sgallari, F., 2020. Non-convex total variation regularization for convex denoising of signals. J. Math. Imag. Vis. 62, 825–841.
Shen, Z., Cheng, L., 2016. Coupled image restoration model with non-convex non-smooth p wavelet frame and total variation regularisation. IET Image Process. 10, 926–935.
Wang, Zhiguo, Zhang, Bing, Gao, Jinghuai, Wang, Qingzhen, Liu, Qinghuo, 2017. Wavelet transform with generalized beta wavelets for seismic time-frequency analysis. Geophysics 82, O47–O56.
Wang, C., Zhu, Z., Gu, H., Wu, X., Liu, S., 2019. Hankel low-rank approximation for seismic noise attenuation. IEEE Trans. Geosci. Rem. Sens. 57, 561–573.
Wang, D.-L., Tong, Z.-F., Tang, C., Zhu, H., 2010. An iterative curvelet thresholding algorithm for seismic random noise attenuation. Appl. Geophys. 7, 315–324.
Wang, P., Gao, J., 2013. Extraction of instantaneous frequency from seismic data via the generalized morse wavelets. J. Appl. Geophys. 93, 83–92.
Xue, Y.-j., Cao, J.-x., Wang, X.-j., Du, H.-k., 2021. Reservoir permeability estimation from seismic amplitudes using variational mode decomposition. J. Petrol. Sci. Eng. 208, 109293.
Yang, Y., Gao, J., Wang, Z., Zhang, G., Zhu, X., 2019. 2-d seismic random noise attenuation via self-paced nonnegative dictionary learning. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 12, 5391–5401.
Yin, W., Osher, S., Goldfarb, D., Darbon, J., 2008. Bregman iterative algorithms for l1-minimization with applications to compressed sensing. SIAM J. Imag. Sci. 1, 143–168.
Yu, S., Ma, J., 2018. Complex variational mode decomposition for slop-preserving denoising. IEEE Trans. Geosci. Rem. Sens. 56, 586–597.
Yuan, S., Wang, S., Li, G., 2012. Random noise reduction using bayesian inversion. J. Geophys. Eng. 9, 60–68.
Zhang, M., Liu, Y., Zhang, H., Chen, Y., 2020. Incoherent noise suppression of seismic data based on robust low-rank approximation. IEEE Trans. Geosci. Rem. Sens. 58, 8874–8887.
Zhao, Q., Meng, D., Jiang, L., Xie, Q., Xu, Z., Hauptmann, A.G., 2015. Self-paced Learning for Matrix Factorization: Presented at the Twenty-Ninth AAAI Conference on Artificial Intelligence.
Zhong, T., Cheng, M., Dong, X., Li, Y., Wu, N., 2021. Seismic random noise suppression by using deep residual u-net. J. Petrol. Sci. Eng. 209, 109901.