Damped multichannel singular spectrum analysis for 3D random noise attenuation[a]

[a]Published in Geophysics, 81, V261-V270, (2016)

Weilin Huang, Runqiu Wang*, Yangkang Chen†, Huijian Li*, and Shuwei Gan**

ABSTRACT

Multichannel singular spectrum analysis (MSSA) is an effective algorithm for random noise attenuation in seismic data, which decomposes the vector space of the Hankel matrix of the noisy signal into a signal subspace and a noise subspace by the truncated singular value decomposition (TSVD). However, this signal subspace actually still contains residual noise. We derived a new formula of low-rank reduction, which is more powerful in distinguishing between signal and noise compared with the traditional TSVD. By introducing a damping factor into the traditional MSSA for damping the singular values, we proposed a new algorithm for random noise attenuation. The proposed modified MSSA is named as the damped MSSA. The denoising performance is controlled by the damping factor and the proposed approach reverts to the traditional MSSA approach when the damping factor is sufficiently large. Application of the damped MSSA algorithm on synthetic and field seismic data demonstrates a superior performance compared with the conventional MSSA algorithm.

INTRODUCTION

High signal-to-noise ratio (SNR) is necessary for many procedures in seismic exploration such as amplitude variation with offset (AVO) analysis, seismic attribute analysis and microseismic monitoring. The attenuation of random noise is an important subject in improving the signal-to-noise ratio (SNR) (Chen and Ma, 2014). If signal is eliminated with the noise, continuity of seismic events may be disrupted and the resulting image quality will be negatively affected. Enhancing the useful signal while preserving edge properties of the seismic profiles by attenuating random noise can help reduce interpretation difficulties and risks for oil & gas detection (Gan et al., 2015; Yang et al., 2015).

There are several classical ways for random noise attenuation: prediction based noise-attenuation approaches (Abma and Claerbout, 1995; Liu et al., 2011) such as the $f-x$ deconvolution (Canales, 1984; Gulunay, 1986) and $f-xy$ deconcolution (Gulunay, 2000; Liu et al., 2012), median filtering (Liu, 2013; Chen, 2015; Gan et al., 2016), Karhunen-Loeve transform (Jones and Levy, 1987), sparse transform domain

thresholding strategies (Donoho, 1995; Neelamani et al., 2008; Fomel and Liu, 2010; Chen et al., 2014, 2016) and Cadzow filtering (Trickett, 2008; Trickett and Burroughs, 2009) or multichannel singular spectrum analysis (Oropeza and Sacchi, 2011; Chiu, 2013; Gao et al., 2013). The principle of all these denoising approaches is to distinguish between noise and signal based on certain characteristics, such as the spatial coherency or the sparsity in a sparse transform domain. Chen and Fomel (2015) proposed a two-step denoising approach in order to retrieve the lost useful information from the removed noise based on local signal-and-noise orthogonalization.

In this paper, we focused on attenuating random noise in 3D seismic data using the multichannel singular spectrum analysis (MSSA) algorithm. MSSA is a data-driven algorithm developed from research on alternative tools for the analysis of multichannel time series, which is based on the truncated singular value decomposition (TSVD) (Golub and Loan, 1996) of the Hankel matrix. MSSA is also an extension of singular spectrum analysis (SSA), which is used to analyze 1D time series. Like other noise-attenuation methods, MSSA transforms the data into a domain where signal and noise are mapped onto separate subspaces and then removes the noise. However, many numerical experiments suggest that the random noise can not be completely removed using the MSSA algorithm (Huang et al., 2015). One perspective is that the TSVD can only decompose the data into a noise subspace and a signal-plus-noise subspace. In this paper, we analyzed how to theoretically decompose the input data into signal subspace and noise subspace and proposed a practical solution to apply a variable damping factor to different singular values to obtain results with higher SNR. We used both synthetic and field 3D seismic datasets to demonstrate our proposed approach.

TRADITIONAL MSSA BY TSVD

Consider a block of 3D data $\mathbf{D}_{time}(x, y, t)$ of N_x by N_y by N_t samples ($x = 1 \cdots N_x, y = 1 \cdots N_y, t = 1 \cdots N_t$). The MSSA (Oropeza and Sacchi, 2011) operates on the data in the following way: first, MSSA transforms $\mathbf{D}_{time}(x, y, t)$ into $\mathbf{D}_{freq}(x, y, w)(w = 1 \cdots N_w)$ with complex values in the frequency domain. Each frequency slice of the data, at a given frequency w_0, can be represented by the following matrix:

$$\mathbf{D}(w_0) = \begin{pmatrix} D(1,1) & D(1,2) & \cdots & D(1,N_x) \\ D(2,1) & D(2,2) & \cdots & D(2,N_x) \\ \vdots & \vdots & \ddots & \vdots \\ D(N_y,1) & D(N_y,2) & \cdots & D(N_y,N_x) \end{pmatrix}. \tag{1}$$

To avoid notational clutter we omit the argument w_0. Second, MSSA constructs

a Hankel matrix for each row of $\mathbf{D}$; the Hankel matrix $\mathbf{R}_i$ for row i of $\mathbf{D}$ is as follows:

$$\mathbf{R}_i = \begin{pmatrix} D(i,1) & D(i,2) & \cdots & D(i,m) \\ D(i,2) & D(i,3) & \cdots & D(i,m+1) \\ \vdots & \vdots & \ddots & \vdots \\ D(i,N_x-m+1) & D(i,N_x-m+2) & \cdots & D(i,N_x) \end{pmatrix}. \tag{2}$$

Then MSSA constructs a block Hankel matrix $\mathbf{M}$ for $\mathbf{R}_i$ as:

$$\mathbf{M} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_n \\ \mathbf{R}_2 & \mathbf{R}_3 & \cdots & \mathbf{R}_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{N_y-n+1} & \mathbf{R}_{N_y-n+2} & \cdots & \mathbf{R}_{N_y} \end{pmatrix}. \tag{3}$$

The size of $\mathbf{M}$ is $I \times J$, $I = (N_x - m + 1)(N_y - n + 1)$, $J = mn$. m and n are predifined integers chosen such that the Hankel maxtrix $\mathbf{R}_i$ and the block Hankel matrix $\mathbf{M}$ are close to square matrices, for example, $m = N_x - \lfloor \frac{N_x}{2} \rfloor$ and $n = N_y - \lfloor \frac{N_y}{2} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of the argument. We assume that $I > J$. The filtered data are recovered with random noise attenuated by properly averaging along the anti-diagonals of the low-rank reduction matrix of $\mathbf{M}$ via TSVD. Next, we would like to briefly discuss the TSVD to introduce our work. In general, the matrix $\mathbf{M}$ can be represented as

$$\mathbf{M} = \mathbf{S} + \mathbf{N}, \tag{4}$$

where $\mathbf{S}$ and $\mathbf{N}$ denote the block Hankel matrix of signal and of random noise, respectively. We assume that $\mathbf{M}$ and $\mathbf{N}$ have full rank, $rank(\mathbf{M})$=$rank(\mathbf{N}) = J$ and $\mathbf{S}$ has deficient rank, $rank(\mathbf{S}) = K < J$. The singular value decomposition (SVD) of $\mathbf{M}$ can be represented as:

$$\mathbf{M} = [\mathbf{U}_1^M \quad \mathbf{U}_2^M] \begin{bmatrix} \Sigma_1^M & \mathbf{0} \\ \mathbf{0} & \Sigma_2^M \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^M)^H \\ (\mathbf{V}_2^M)^H \end{bmatrix}, \tag{5}$$

where Σ_1^M $(K \times K)$ and Σ_2^M $((I-K) \times (J-K))$ are diagonal matrices and contain, respectively, larger singular values and smaller singular values. $\mathbf{U}_1^M$ $(I \times K)$, $\mathbf{U}_2^M$ $(I \times (I-K))$, $\mathbf{V}_1^M$ $(J \times K)$ and $\mathbf{V}_2^M$ $(J \times (J-K))$ denote the associated matrices with singular vectors. The symbol $[\cdot]^H$ denotes the conjugate transpose of a matrix. Generally, the signal is more energy-concentrated and correlative than the random noise. Thus, the larger singular values and their associated singular vectors represent the signal, while the smaller values and their associated singular vectors represent the random noise. We let Σ_2^M be $\mathbf{0}$ to achieve the goal of attenuating random noise as follows:

$$\tilde{\mathbf{M}} = \mathbf{U}_1^M \Sigma_1^M (\mathbf{V}_1^M)^H. \tag{6}$$

Equation 6 is referred to as the TSVD.

DAMPED MSSA

However, $\tilde{\mathbf{M}}$ is actually still mixed with residual noise. In the following part, we will first analyse the reason why $\tilde{\mathbf{M}}$ also contains noise component, and then introduce the modified MSSA method by damping the singular values, which we call the damped MSSA.

The singular value decomposition (SVD) of $\mathbf{S}$ can be represented as:

$$\mathbf{S} = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \Sigma_1^S & \mathbf{0} \\ \mathbf{0} & \Sigma_2^S \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^S)^H \\ (\mathbf{V}_2^S)^H \end{bmatrix}. \tag{7}$$

The corresponding matrices in equations 5 and 7 have the same size.

Because of the deficient rank, the matrix $\mathbf{S}$ can be written as

$$\mathbf{S} = \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H. \tag{8}$$

Combining equations 4, 7, and 8, we can factorize $\mathbf{M}$ as follows:

$$\mathbf{M} = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} (\Sigma_1)^{-1}(\mathbf{N}^H \mathbf{U}_1^S + \mathbf{V}_1^S \Sigma_1^S)^H \\ (\Sigma_2)^{-1}(\mathbf{N}^H \mathbf{U}_2^S)^H \end{bmatrix}, \tag{9}$$

where Σ_1 and Σ_2 denote diagonal and positive definite matrices. Please note that M is constructed such that it is close to a square matrix (Oropeza and Sacchi, 2011), and thus the Σ_1 and Σ_2 are assumed to be square matrices for derivation convenience. The appendix A provides the derivation of how we factorize $\mathbf{M}$ into the form of equation 9. We observe that the left matrix has orthonormal columns and the middle matrix is diagonal. It can be proven that the right matrix also has orthonormal columns. The proof is provided in appendix A. Thus, equation 9 is an SVD of $\mathbf{M}$. According to the TSVD method, we let Σ_2 be $\mathbf{0}$ and then the following equation holds:

$$\begin{aligned} \tilde{\mathbf{M}} &= \mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N} + \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H \\ &= \mathbf{S} + \mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N}. \end{aligned} \tag{10}$$

It is clear that $\tilde{\mathbf{M}} \neq \mathbf{S}$. Because the matrices $\mathbf{U}_1^S$ and $\mathbf{N}$ are unknown, we cannot use equation 10 directly to attenuate the residual noise. However, by combining equations 6, 9 and 10, we can derive $\mathbf{S}$ as:

$$\mathbf{S} = \mathbf{U}_1^M \left\{ \Sigma_1^M (\mathbf{V}_1^M)^H - \left[\Sigma_1 (\mathbf{V}_1^M)^H - \Sigma_1^S (\mathbf{V}_1^S)^H \right] \right\} \tag{11}$$

The appendix B gives a detailed derivation to obtain equation 11 from equations 6, 9 and 10.

For simplification, we assume that there exist such $\mathbf{A}$ and $\mathbf{B}$ that $\mathbf{V}_1^S = \mathbf{V}_1^M \mathbf{A}$ and $\Sigma_1 = \Sigma_1^M \mathbf{B}$. $\mathbf{A}$ is a square matrix of $K \times K$ and $\mathbf{B}$ is a diagonal matrix of $K \times K$. Then we can simplify $\mathbf{S}$ as:

$$\mathbf{S} = \mathbf{U}_1^M \Sigma_1^M \mathbf{T} \left(\mathbf{V}_1^M\right)^H . \tag{12}$$

$$\mathbf{T} = \mathbf{I} - \mathbf{B}\left(\mathbf{I} - (\Sigma_1)^{-1}\Sigma_1^S \mathbf{A}^H\right), \tag{13}$$

where $\mathbf{I}$ is a unit matrix and here we name $\mathbf{T}$ the damping operator. In fact, from equation 9, we can also approximate $\mathbf{A}$ and $\mathbf{B}$ as follows:

$$\mathbf{A} \approx (\mathbf{I} - \Gamma)\Sigma_1 \left(\Sigma_1^S\right)^{-1}, \tag{14}$$

$$\mathbf{B} = \mathbf{I}. \tag{15}$$

where $\Gamma = (\mathbf{V}_1^M)^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}$ is an unknown matrix. $(\mathbf{V}_1^M)^o$ can be regarded as an approximate inverse of $\mathbf{V}_1^M$, which satisfies that $\| \mathbf{I} - \mathbf{V}_1^M (\mathbf{V}_1^M)^o \| \rightarrow 0$. The appendix B provides the detailed derivation for obtaining equations 14 and 15.

Inserting equations 14 and 15 into equation 13, we can obtain a simplified formula:

$$\begin{aligned} \mathbf{T} &= \mathbf{I} - \mathbf{I}\left(\mathbf{I} - (\Sigma_1)^{-1}\Sigma_1^S \mathbf{A}^H\right) \\ &= (\Sigma_1)^{-1}\Sigma_1^S \mathbf{A}^H \\ &\approx (\Sigma_1)^{-1}\Sigma_1^S \left((\mathbf{I} - \Gamma)\Sigma_1 \left(\Sigma_1^S\right)^{-1}\right)^H \\ &= \mathbf{I} - \Gamma. \end{aligned} \tag{16}$$

Combing equations 12 and 16, we can conclude that the true signal is a damped version of the previous TSVD method (equation 6), with the damping operator defined by equation 16. Right now, there is still one unknown parameter needed to be defined: Γ. Although we have a potential selection $\Gamma = (\mathbf{V}_1^M)^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}$, as defined during the derivation of $\mathbf{A}$, we cannot calculate it because we do not know $\mathbf{N}$ and $\mathbf{U}_1^S$.

Instead, we seek the form of Γ from a different way. We treat Γ as a whole instead of paying attention to each detailed component in $\Gamma = (\mathbf{V}_1^M)^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}$. We have known that the true signal is a damped version of the TSVD method from the previous derivation, and the damping operator equals to $\mathbf{I} - \Gamma$, acting on the diagonal matrix Σ_1^M. We can begin our search for an approximation of Γ based on the following conditions:

1. As we know from the previous derivation, the truncating point of TSVD is the rank of the signal matrix $\mathbf{S}$. In other words, the rank of $\tilde{\mathbf{M}} = \mathbf{U}_1^M \Sigma_1^M (\mathbf{V}_1^M)^H$ equals to the rank of $\mathbf{S}$. The rank of $\tilde{\mathbf{M}}$ should remain unchanged when we damp Σ_1^M. Therefore, the damping operator should be a diagonal and positive definite matrix.

2. Each element of the damping operator should be in the interval $(0, 1]$.
3. The lower the SNR is, the stronger the damping should be, because the energy of random noise in the signal-noise space is relatively stronger.
4. In the case of zero random noise, the damping operator should be a unit matrix.
5. Since we always hope to preserve the main components of seismic data, the damping operator should have a weaker effect on the larger singular value.
6. The power of the damping operator can be controlled by one coefficient and the damped MSSA can revert to the traditional MSSA.

From the above analysis, we propose to use a Γ of the following form:

$$\Gamma = \begin{pmatrix} a_1/b_1 & & & \\ & a_2/b_2 & & \\ & & \ddots & \\ & & & a_K/b_K \end{pmatrix}, \tag{17}$$

where $\{a\}$ contains the information of random noise, $\{b\}$ contains the information of signal, $a_1/b_1 < a_2/b_2 < \cdots < a_K/b_K$ and $0 < a_i < b_i$, $i = 1, 2, \cdots, K$.

We tried a lot of numerical experiments and found that a very pleasant denoising performance can be obtained when Γ is chosen as

$$\Gamma \approx \hat{\delta}^N \left(\Sigma_1^M\right)^{-N}, \tag{18}$$

where $\hat{\delta}$ denotes the maximum element of Σ_2^M and N denotes the damping factor. We use such approximation because of three reasons. (1) $\hat{\delta}$ reflects the energy of random noise and Σ_1^M contains the information of signal. (2) Because the diagonal elements of Σ^M are in a descending order, $\hat{\delta}$ is certainly smaller than every diagonal element of Σ_1^M, and $\hat{\delta}/\delta_1 < \hat{\delta}/\delta_2 < \cdots < \hat{\delta}/\delta_K$, where δ_i denotes ith diagonal entry in Σ_1^M. (3) $\hat{\delta}$ is zero in the zero random noise situation. Besides, We introduce the parameter N to control the strength of damping operator, the greater the N, the weaker the damping, and the damped MSSA reverts to the basic MSSA when $N \to \infty$.

Combining equations 12, 16, and 18, we conclude the approximation of $\mathbf{S}$ as:

$$\mathbf{S} = \mathbf{U}_1^M \Sigma_1^M \mathbf{T} (\mathbf{V}_1^M)^H, \tag{19}$$

$$\mathbf{T} = \mathbf{I} - (\Sigma_1^M)^{-N} \hat{\delta}^N. \tag{20}$$

We call the newly developed algorithm (equations 19 and 20) damped MSSA algorithm. In the proposed algorithm, we only need to decide two parameters: the rank K and the damping factor N. It is worth mentioning that the damping operator $\mathbf{T}$ (equation 20) also behaves as a soft-thresholding operator applied on the singular

matrix $\mathbf{\Sigma_1^M}$. The damping (thresholding) step may thus cause slight damage to useful signal while more noise is scaled down, and the final S/N can still be greatly improved.

The parameterization for the DMSSA approach is quite convenient. Although the traditional MSSA has a rank with broad range according to the data size and data complexity, the damping factor is usually chosen between 2-5. When damping factor is chosen as 1, the damping is very strong and will cause some useful energy loss, but when the damping factor is chosen as 2, 3, or even larger value, the compromise between preservation of useful signals and removal of random noise are much improved. The implementation of the DMSSA approach can be straightforwardly based on the MSSA framework, except for the slight difference, which introduces the damping factor.

EXAMPLES

We first test the damped MSSA algorithm proposed in this paper on a 3D synthetic data set which is composed of three linear events in $x - y - t$ domain. The data contain $N_x \times N_y$ traces, with $N_x = 20$ and $N_y = 20$. The size of the block Hankel matrix for this example is 121×100. The temporal length of the window is 600 ms. The sampling interval is 2ms. The clean and noisy data are shown in Figures 1a and 1d, respectively. In this example, we add bandlimited random noise with bandwidth close to the seismic bandwidth. The lines indicate the displaying slices for the front, right, and top views. After using the conventional MSSA and the damped MSSA, the results are shown in Figures 1b and 1e, respectively. Figures 1c and 1f show the removed noise that correspond to the conventional MSSA and the damped MSSA, respectively. Figure 2 demonstrates the denoising comparison of the 5th crossline section of the synthetic example. Figure 3 demonstrates the denoising comparison of the 5th inline section of the synthetic example. From Figures 1, 2, and 3, it is clear that although the traditional MSSA approach performs well and suppresses the random noise, the proposed damped MSSA approach obtains even better results. In this example, the rank $K = 3$ and the damping factor $N = 4$. Figure 4 shows two denoising error comparison of the 5th inline section of the synthetic example. The error section denotes the difference section between the clean and denoised section. It is obvious that both MSSA and DMSSA approaches will not cause obvious damage to the useful events while the DMSSA approach cause less denoising error because of a larger amount of removed noise. In order to test the sensitivity of the denoising performance on the damping factor N, we tried different N and compared their performance. Figure 5 shows the denoising performance of the 5th crossline section using different N. It is clear that as N increases, the residual noise becomes stronger and stronger. When $N = 40$, the denoising performance is nearly the same as the traditional MSSA, as shown in Figure 2b. When the N is too small, say $N = 1$, although the denoised section (Figure 5a) is very clean, there is some signal loss on the events. We can increase N from 1 to 4 to obtain a compromise between the

removal of residual noise and the preservation of useful signals, as shown in Figures 2e and 5c.

To demonstrate how the proposed algorithm works in practice, we apply the damped MSSA algorithm on two 3D post-stack field datasets. The first field data is shown in Figure 6a. Figures 6b and 6c show the denoised data using the traditional and damped MSSAs, respectively. Figures 6d and 6e show the removed noise using the traditional and damped MSSAs, respectively. In this test, in order to preserve the main features of the original data, we use a more conservative combination of K and N ($K = 25$ and $N = 5$ in this example). A denoising comparison of the 5th crossline section is shown in Figure 7. The top row of Figure 7 shows the comparison between the raw data and the denoised data using the traditional MSSA and the damped MSSA. The bottom row of Figure 7 shows a comparison of the removed noise sections using two methods. It can be observed that the denoised section using the damped MSSA (Figure 7c) is cleaner than that of the traditional MSSA (Figure 7b). The noise section of the damped MSSA is obviously noisier than that of the traditional MSSA, which indicates that the damped MSSA can attenuate more random noise. Since there is almost no coherent energy in the noise sections, both denoising results would be acceptable, but the proposed approach obtains a better performance. Figure 8 shows three zoomed parts from the crossline sections of the raw data and two denoised data, as highlighted by the frameboxes in Figure 7, which gives us a more detailed comparison. The MSSA approach obtains a big improvement of the image considering the signal-to-noise ratio (S/N) while the DMSSA approach obtains a even better performance. The events using the DMSSA approach become more continuous than the MSSA approach, without remaining any random noise, and both approaches well preserve the discontinuities in the image.

The second field data example is shown in Figure 9a. The reflectors in this example are more complex with steeper dips compared with the first field data example. The two denoised results are shown in Figures 9b and 9c. Figures 9d and 9e show the removed noise cubes using the traditional MSSA and the damped MSSA, respectively. It is clear that the proposed approach can remove stronger noise compared with the traditional MSSA approach, as can be see from Figure 9. In this example, because of the steeper dip of reflectors, we use higher K compared with the previous field data example. In practice, when the seismic data becomes more complicated, a higher K value should be used to account for the higher rank caused by the larger number of dip components. For this example, $K = 40$ and $N = 2$. The 10th crossline sections from the raw data, the denoised data using the traditional MSSA approach, and the denoised data using the damped MSSA approach, are shown in Figures 10a, 10b, and 10c, respectively. The bottom row of Figure 10 shows the zoomed sections that correspond to the frameboxes in the top row. It is clear that the proposed approach can obtain smoother and cleaner reflections than the traditional MSSA approach. For all the three examples, we do not use local processing windows to apply the MSSA or DMSSA algorithm, because the data size is not big and data structure is not very complicated, which makes the performance using MSSA still acceptable.

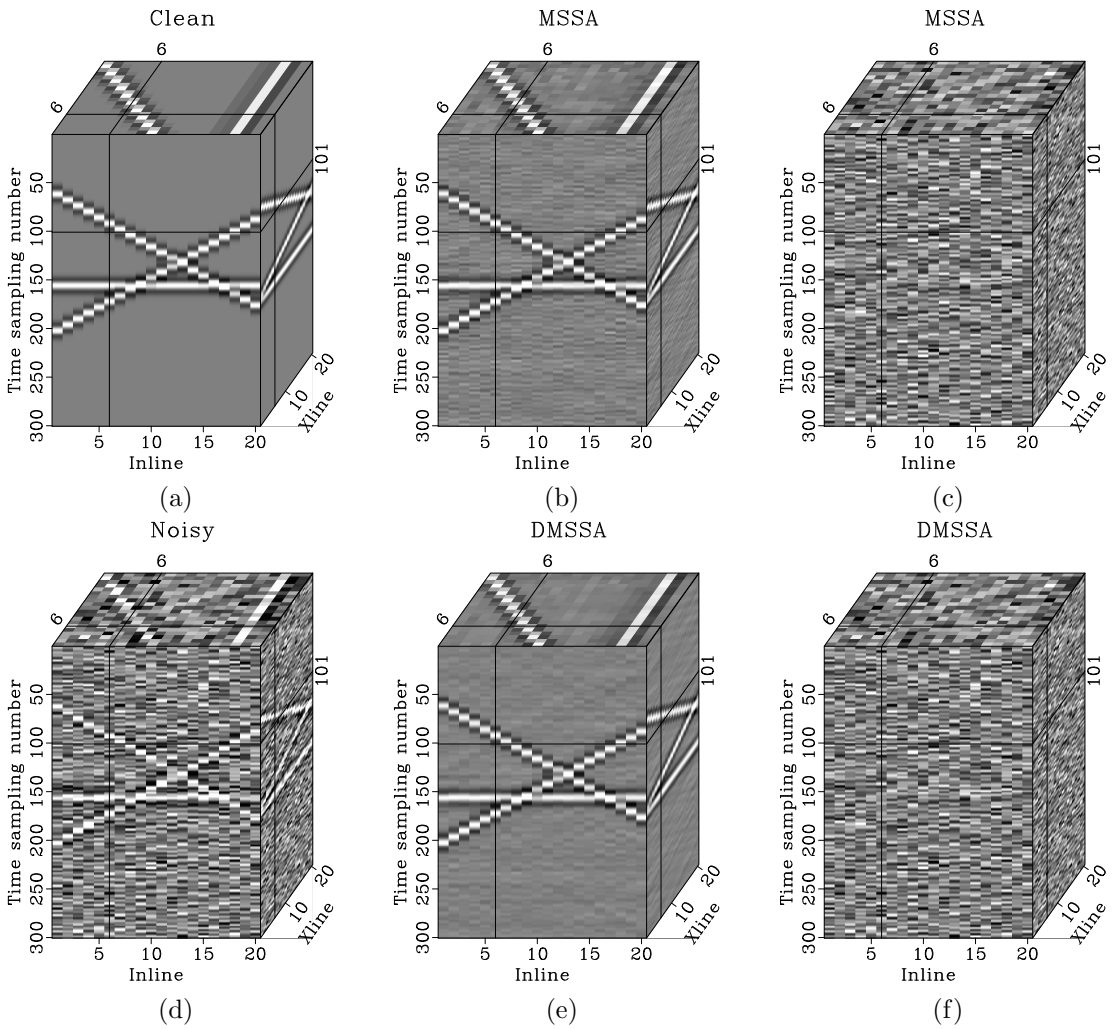


Figure 1: Denoising comparison of synthetic example. (a) & (d) Clean and noisy data, respectively. (b) & (e) Denoised data using the traditional MSSA and the damped MSSA, respectively. (c) & (f) Removed noise using the traditional MSSA and the damped MSSA, respectively. In the example, $K = 3$, $N = 4$.

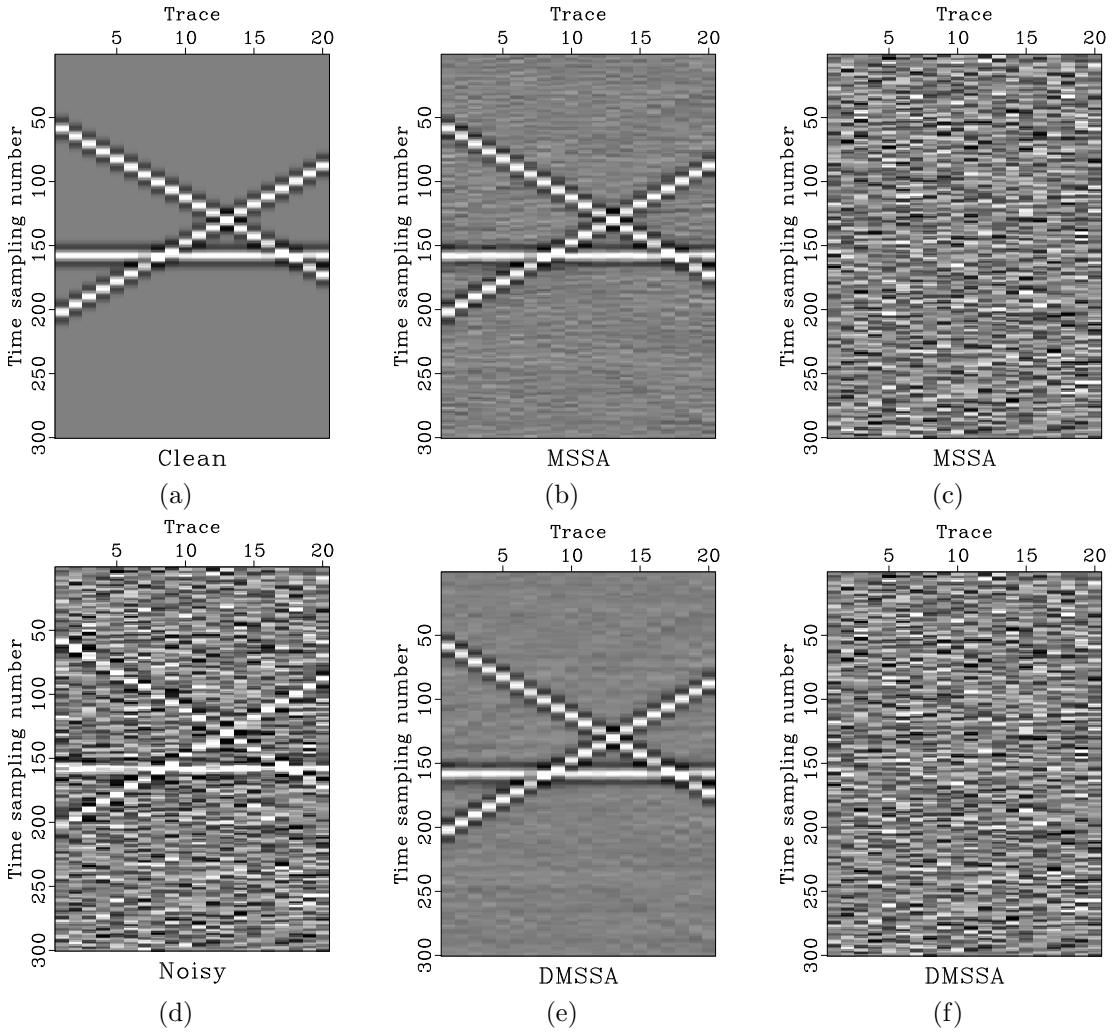


Figure 2: Denoising comparison of the 5th crossline section of the synthetic example. (a) & (d) Clean and noisy data, respectively. (b) & (e) Denoised data using the traditional MSSA and the damped MSSA, respectively. (c) & (f) Removed noise using the traditional MSSA and the damped MSSA, respectively.

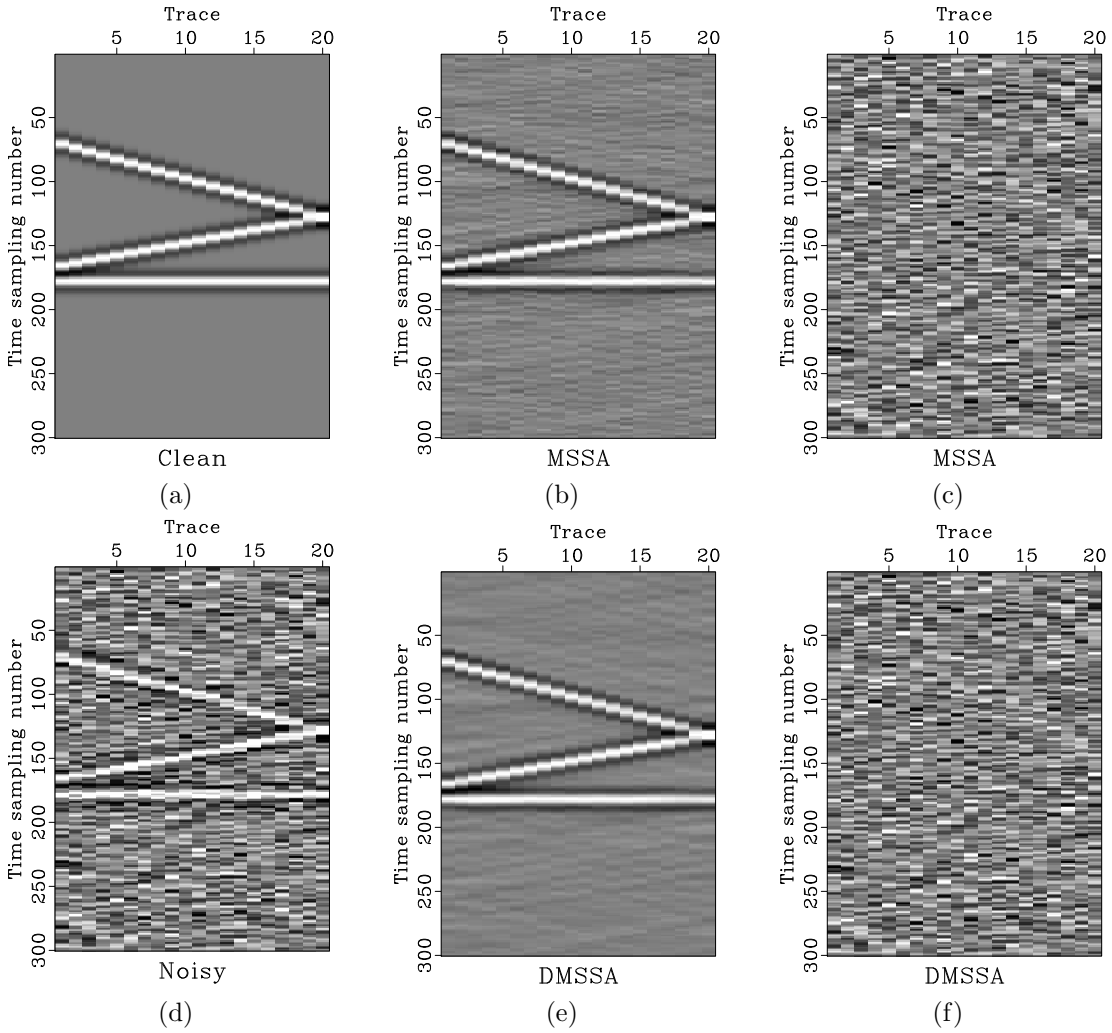
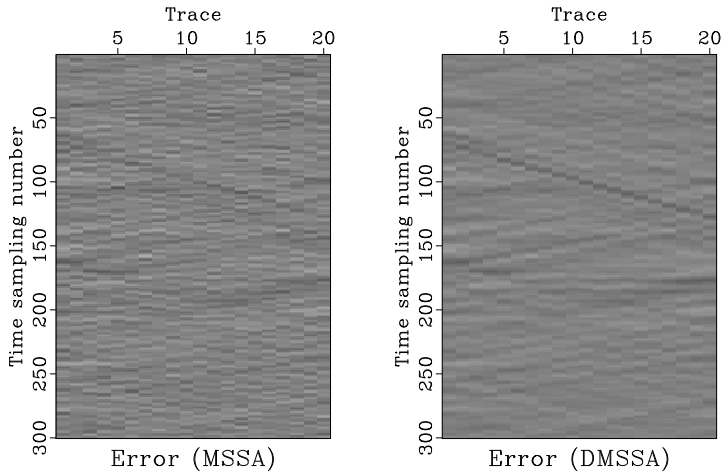


Figure 3: Denoising comparison of the 5th inline section of the synthetic example. (a) & (d) Clean and noisy data, respectively. (b) & (e) Denoised data using the traditional MSSA and the damped MSSA, respectively. (c) & (f) Removed noise using the traditional MSSA and the damped MSSA, respectively.



(a) (b)

Figure 4: Denoising error comparison of the 5th inline section of the synthetic example. (a) Error of the traditional MSSA. (b) Error of the damped MSSA.

CONCLUSION

We have proposed a novel modified multichannel singular spectrum analysis (MSSA) algorithm to attenuate random noise with a new formula of low-rank reduction, which is named as the damped MSSA algorithm. Compared with the traditional truncated singular value decomposition (TSVD) formula, we introduced a damping factor to damp the singular values that correspond to the signal in order to attenuate the residual noise appearing in the traditional approach. The preservation of useful signals and the removal of random noise are compromised through the introduced damping factor. While the rank in MSSA has a big range considering the data size and data complexity, the damping factor in the damped MSSA is usually chosen as an integer that is slightly larger than 1 (such as 2, 3, or 4) to obtain a sufficient improvement. From the synthetic and field data examples, it is obvious that the proposed damped MSSA algorithm can obtain cleaner denoised image compared with the traditional MSSA algorithm.

ACKNOWLEDGEMENT

We would like to thank Ming Zhang, Valentina Socco, Kris Innanen, Mauricio Sacchi, and three anonymous reviewers for constructive suggestions. This work is supported by the National Basic Research Program of China (grant NO: 2013 CB228602) and Texas Consortium for Computational Seismology (TCCS).

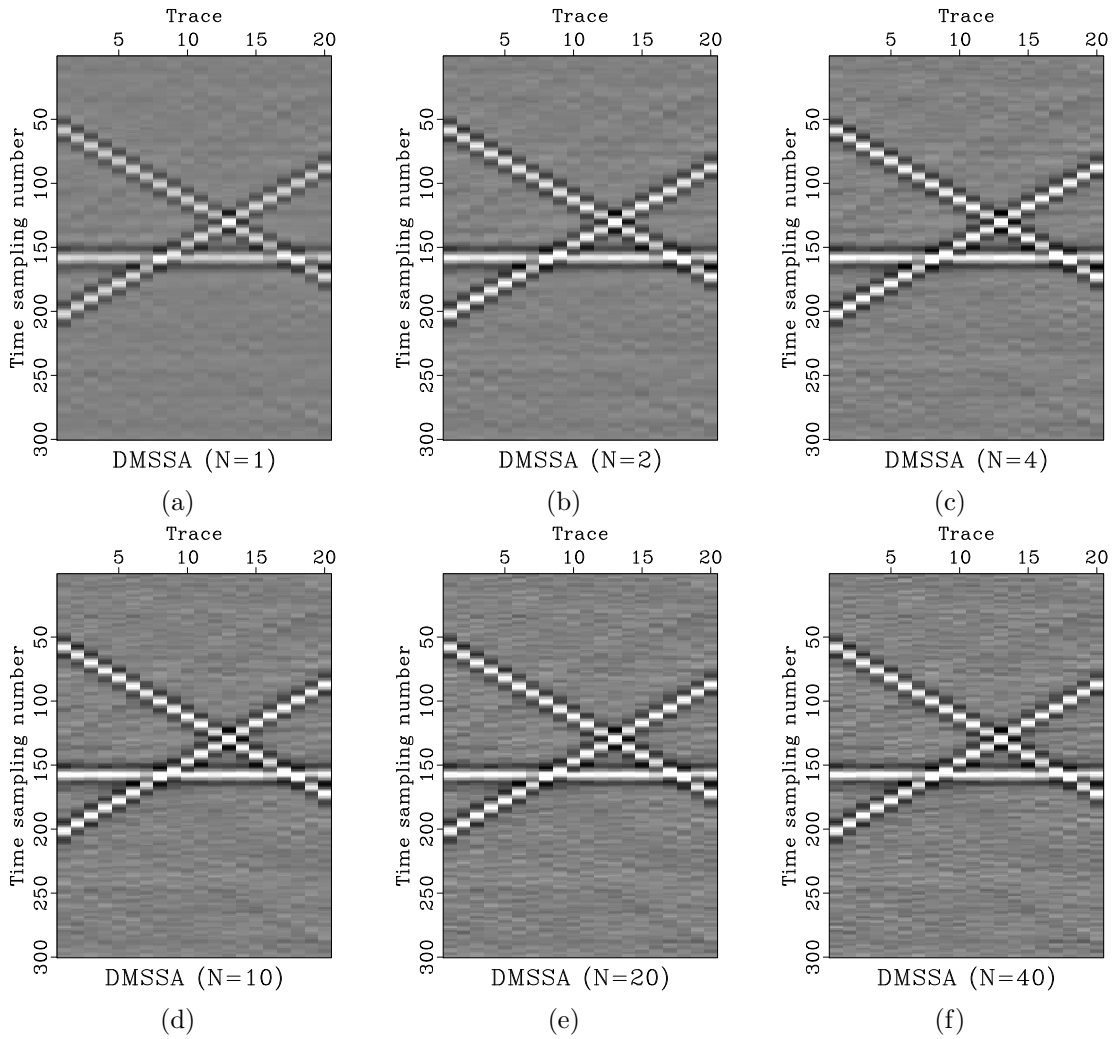


Figure 5: Demonstration of the performance with same $K = 3$ and different N for the 5th crossline section. (a) N=1. (b) N=2. (c) N=4. (d) N=10. (e) N=20. (f) N=40.

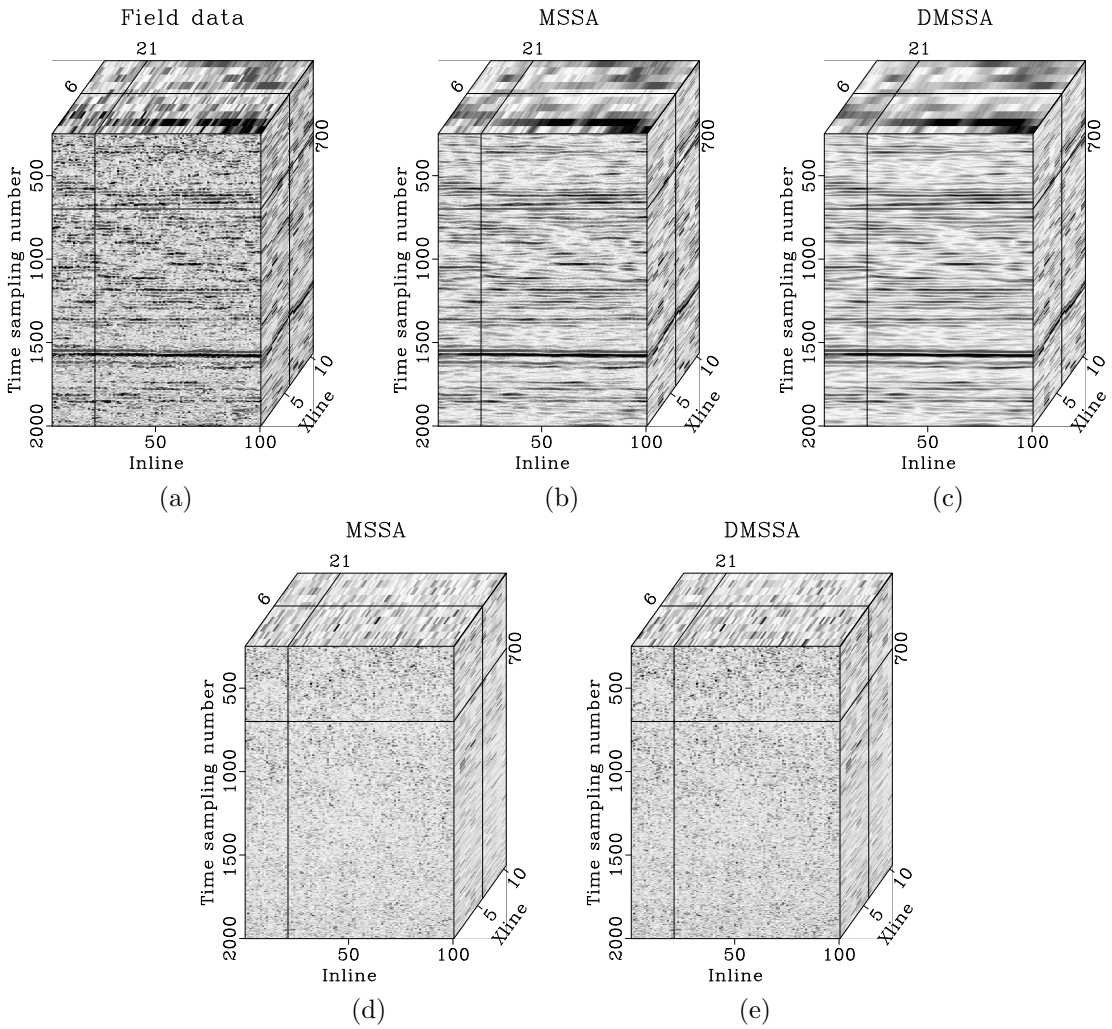


Figure 6: Denoising comparison of first field data example. (a) Noisy 3D field data. (b) Denoised data using the traditional MSSA. (c) Denoised data using the damped MSSA. (d) Removed noise using the traditional MSSA. (e) Removed noise using the damped MSSA. In the example, $K = 25$, $N = 5$.

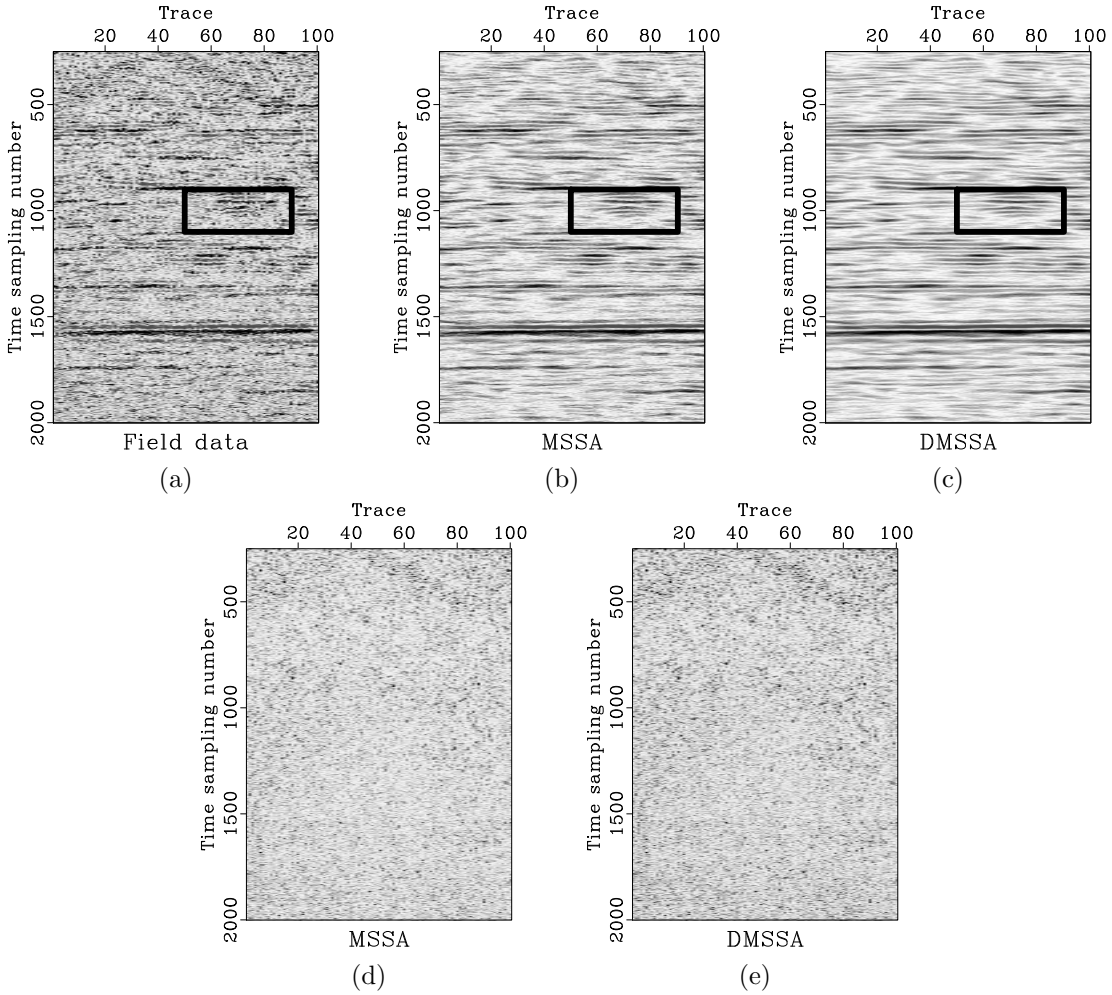


Figure 7: Denoising comparison of the 5th crossline section of field data example. (a) Noisy 3D field data. (b) Denoised data using the traditional MSSA. (c) Denoised data using the damped MSSA. (d) Removed noise using the traditional MSSA. (e) Removed noise using the damped MSSA.

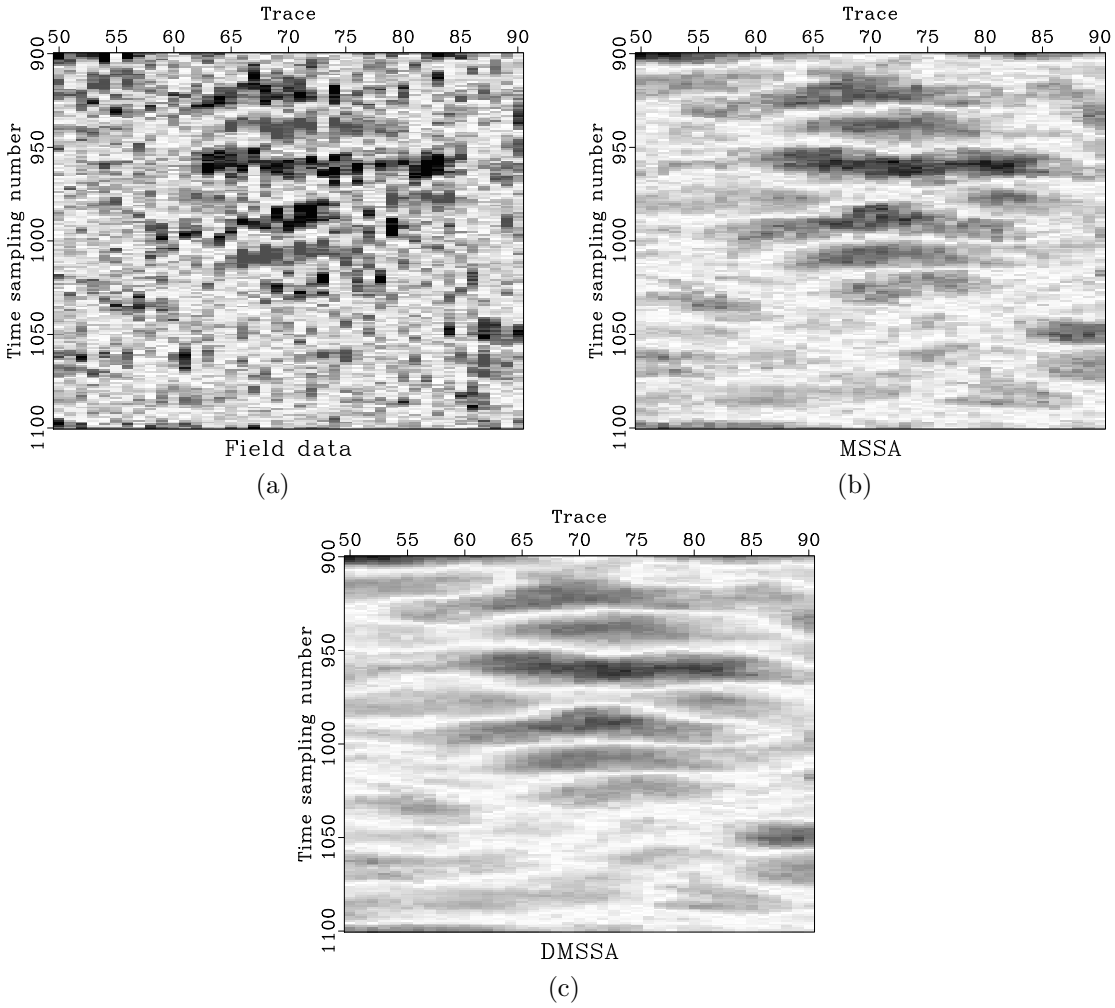


Figure 8: (a) Zoomed section from Figure 7a. (b) Zoomed section from Figure 7b. (c) Zoomed section from Figure 7c.

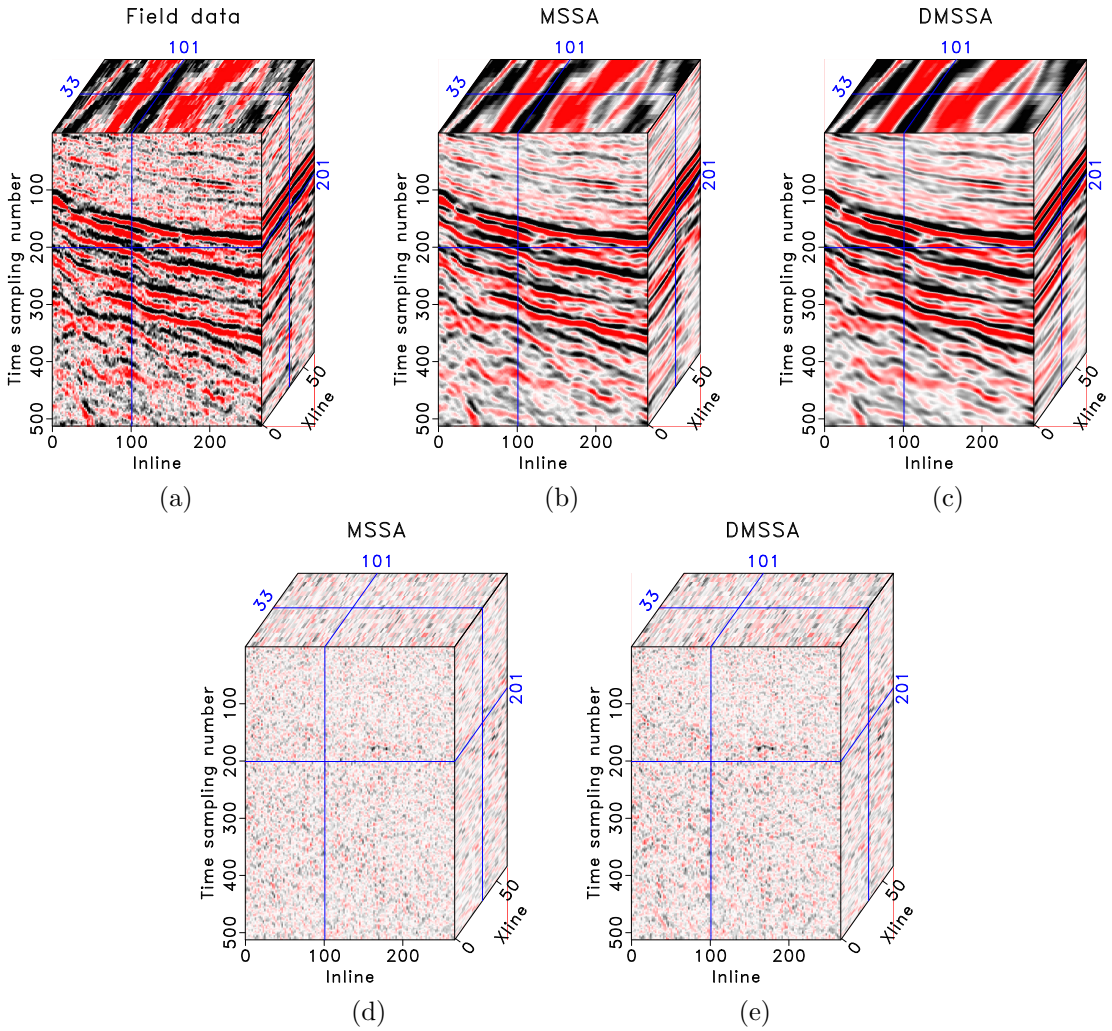


Figure 9: Denoising comparison of second field data example. (a) Noisy 3D field data. (b) Denoised data using the traditional MSSA. (c) Denoised data using the damped MSSA. (d) Removed noise using the traditional MSSA. (e) Removed noise using the damped MSSA. In the example, $K = 40$, $N = 2$.

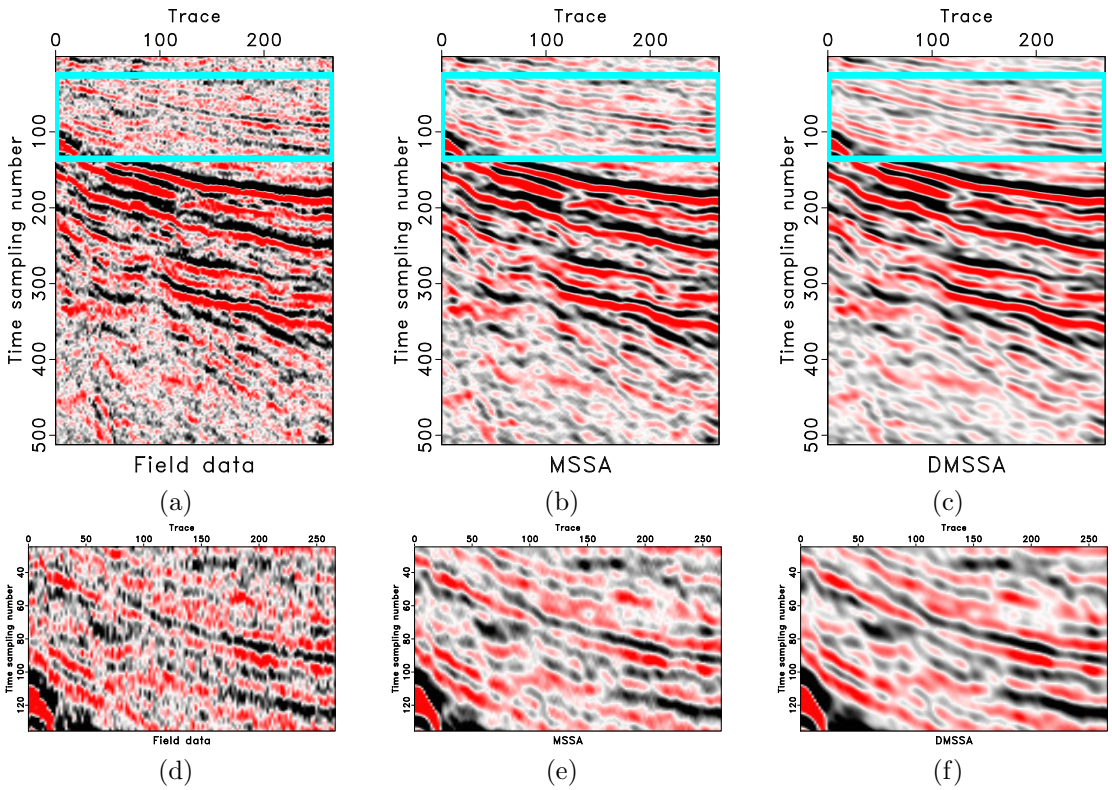


Figure 10: Denoising comparison of the 5th crossline section of field data example. (a) Noisy 3D field data. (b) Denoised data using the traditional MSSA. (c) Denoised data using the damped MSSA. (d) Zoomed section from (a). (e) Zoomed section from (b). (f) Zoomed section from (c).

APPENDIX A: FACTORIZATION OF THE DATA MATRIX M

Because equation 7 is a singular value decomposition (SVD) of the signal matrix $\mathbf{S}$, the left matrix in equation 7 is a unitary matrix:

$$\mathbf{I} = \mathbf{U}^S(\mathbf{U}^S)^H = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} (\mathbf{U}_1^S)^H \\ (\mathbf{U}_2^S)^H \end{bmatrix}. \tag{21}$$

Combining equations 4, 8, and 21, we can derive:

$$\begin{aligned}
\mathbf{M} &= \mathbf{S} + \mathbf{N} \\
&= \mathbf{S} + \mathbf{I}\mathbf{N} \\
&= \mathbf{S} + [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} (\mathbf{U}_1^S)^H \\ (\mathbf{U}_2^S)^H \end{bmatrix} \mathbf{N} \\
&= \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H + \left(\mathbf{U}_1^S(\mathbf{U}_1^S)^H + \mathbf{U}_2^S(\mathbf{U}_2^S)^H\right)\mathbf{N} \\
&= \mathbf{U}_1^S \left((\mathbf{U}_1^S)^H\mathbf{N} + \Sigma_1^S(\mathbf{V}_1^S)^H\right) + \mathbf{U}_2^S(\mathbf{U}_2^S)^H\mathbf{N} \\
&= \mathbf{U}_1^S \left(\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S\right)^H + \mathbf{U}_2^S(\mathbf{N}^H\mathbf{U}_2^S)^H \\
&= [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)^H \\ (\mathbf{N}^H\mathbf{U}_2^S)^H \end{bmatrix}, \\
&= [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} (\Sigma_1)^{-1}(\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)^H \\ (\Sigma_2)^{-1}(\mathbf{N}^H\mathbf{U}_2^S)^H \end{bmatrix},
\end{aligned} \tag{22}$$

where Σ_1 and Σ_2 are introduced matrices and are diagonal and positive definite.

In order to make the right matrix orthonormal, we make two assumptions:

- The noise is close to white noise in the sense that $\mathbf{N}\mathbf{N}^H = \lambda\mathbf{I}$.
- The signal is orthogonal to the noise in the sense that $\mathbf{S}\mathbf{N}^H = \mathbf{0}$.

We let $\mathbf{P}^H$ denote the right matrix of the last equation in 22, then

$$\begin{aligned}
&\mathbf{P}^H\mathbf{P} \\
&= \begin{bmatrix} (\Sigma_1)^{-1}(\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)^H \\ (\Sigma_2)^{-1}(\mathbf{N}^H\mathbf{U}_2^S)^H \end{bmatrix} \begin{bmatrix} (\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} & (\mathbf{N}^H\mathbf{U}_2^S)(\Sigma_2)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (\Sigma_1)^{-1}((\mathbf{U}_1^S)^H\mathbf{N} + \Sigma_1^S(\mathbf{V}_1^S)^H) \\ (\Sigma_2)^{-1}((\mathbf{U}_2^S)^H\mathbf{N}) \end{bmatrix} \begin{bmatrix} (\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} & (\mathbf{N}^H\mathbf{U}_2^S)(\Sigma_2)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}
\end{aligned} \tag{23}$$

where

$$\begin{aligned}
p_{11} &= (\Sigma_1)^{-1}((\mathbf{U}_1^S)^H\mathbf{N} + \Sigma_1^S(\mathbf{V}_1^S)^H)(\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} \\
&= (\Sigma_1)^{-1}((\mathbf{U}_1^S)^H\mathbf{N}\mathbf{N}^H\mathbf{U}_1^S + \Sigma_1^S(\mathbf{V}_1^S)^H\mathbf{N}^H\mathbf{U}_1^S + (\mathbf{U}_1^S)^H\mathbf{N}\mathbf{V}_1^S\Sigma_1^S + \Sigma_1^S(\mathbf{V}_1^S)^H\mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} \\
&= (\Sigma_1)^{-1}(\lambda\mathbf{I} + \mathbf{0} + \mathbf{0} + (\Sigma_1^S)^2)(\Sigma_1)^{-1} \\
&= (\Sigma_1)^{-1}(\lambda\mathbf{I} + (\Sigma_1^S)^2)(\Sigma_1)^{-1} \\
&= \mathbf{I}
\end{aligned} \tag{24}$$

when $\Sigma_1 = \sqrt{\lambda\mathbf{I} + (\Sigma_1^S)^2}$.

$$\begin{aligned}
p_{12} &= (\Sigma_1)^{-1}((\mathbf{U}_1^S)^H\mathbf{N} + \Sigma_1^S(\mathbf{V}_1^S)^H)(\mathbf{N}^H\mathbf{U}_2^S)(\Sigma_2)^{-1} \\
&= (\Sigma_1)^{-1}((\mathbf{U}_1^S)^H\mathbf{N}\mathbf{N}^H\mathbf{U}_2^S + \Sigma_1^S(\mathbf{V}_1^S)^H\mathbf{N}^H\mathbf{U}_2^S)(\Sigma_2)^{-1}
\end{aligned} \tag{25}$$

Since $\mathbf{U}^S$ is an orthogonal matrix, then $(\mathbf{U}_1^S)^H\mathbf{U}_2^S = \mathbf{0}$. Since $\mathbf{S}\mathbf{N}^H = \mathbf{0}$, then $\mathbf{U}_1^S\Sigma_1^S(\mathbf{V}_1^S)^H\mathbf{N}^H = \mathbf{0}$, thus $(\mathbf{V}_1^S)^H\mathbf{N}^H = \mathbf{0}$. In the same way, since $\mathbf{N}\mathbf{S}^H = \mathbf{0}$, thus $\mathbf{N}\mathbf{V}_1^S = \mathbf{0}$. Then,

$$p_{12} = \mathbf{0}. \tag{26}$$

$$\begin{aligned}
p_{21} &= (\Sigma_2)^{-1}((\mathbf{U}_2^S)^H\mathbf{N})(\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} \\
&= (\Sigma_2)^{-1}((\mathbf{U}_2^S)^H\mathbf{N})\mathbf{N}^H\mathbf{U}_1^S + (\mathbf{U}_2^S)^H\mathbf{N})\mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} \\
&= \mathbf{0}.
\end{aligned} \tag{27}$$

$$\begin{aligned}
p_{22} &= (\Sigma_2)^{-1}((\mathbf{U}_2^S)^H\mathbf{N})(\mathbf{N}^H\mathbf{U}_2^S)(\Sigma_2)^{-1} \\
&= (\Sigma_2)^{-1}(\lambda\mathbf{I})(\Sigma_2)^{-1} \\
&= \mathbf{I},
\end{aligned} \tag{28}$$

when $\Sigma_2 = \sqrt{\lambda\mathbf{I}}$. Thus, we prove that $\mathbf{P}^H\mathbf{P} = \mathbf{I}$ when Σ_1 and Σ_2 are appropriately chosen, and $\mathbf{P}$ is orthonormal.

APPENDIX B: DERIVATION FOR EQUATIONS 11, 14 AND 15

We first reformulate equation 10 as

$$\mathbf{S} = \tilde{\mathbf{M}} - \mathbf{U}_1^S(\mathbf{U}_1^S)^H\mathbf{N}. \tag{29}$$

Inserting equation 6 into equation 29, we can further derive

$$\mathbf{S} = \mathbf{U}_1^M\Sigma_1^M(\mathbf{V}_1^M)^H - \mathbf{U}_1^S(\mathbf{U}_1^S)^H\mathbf{N}. \tag{30}$$

Because equations 5 and 9 are both SVDs of $\mathbf{M}$, we let

$$\mathbf{U}_1^S = \mathbf{U}_1^M, \tag{31}$$

and

$$\Sigma_1 = \Sigma_1^M, \tag{32}$$

and

$$(\mathbf{N}^H \mathbf{U}_1^S + \mathbf{V}_1^S \Sigma_1^S)(\Sigma_1)^{-1} = \mathbf{V}_1^M. \tag{33}$$

Considering $\Sigma_1 = \Sigma_1^M \mathbf{B}$,

$$\mathbf{B} = \mathbf{I}. \tag{34}$$

From equation 33,

$$\begin{aligned} \mathbf{V}_1^S &= (\mathbf{V}_1^M - \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}) \Sigma_1 (\Sigma_1^S)^{-1} \\ &\approx \mathbf{V}_1^M (\mathbf{I} - (\mathbf{V}_1^M)^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}) \Sigma_1 (\Sigma_1^S)^{-1} \\ &= \mathbf{V}_1^M \mathbf{A}, \end{aligned} \tag{35}$$

where $(\mathbf{V}_1^M)^o$ satisfies that $\| \mathbf{I} - \mathbf{V}_1^M (\mathbf{V}_1^M)^o \| \rightarrow 0$.

Considering $\mathbf{V}_1^S = \mathbf{V}_1^M \mathbf{A}$,

$$\begin{aligned} \mathbf{A} &\approx (\mathbf{I} - (\mathbf{V}_1^M)^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}) \Sigma_1 (\Sigma_1^S)^{-1} \\ &= (\mathbf{I} - \Gamma) \Sigma_1 (\Sigma_1^S)^{-1}, \end{aligned} \tag{36}$$

where $\Gamma = (\mathbf{V}_1^M)^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}$.

Inserting equations 31 and 33 into equation 30, we can obtain:

$$\begin{aligned} \mathbf{S} &= \mathbf{U}_1^M \Sigma_1^M (\mathbf{V}_1^M)^H - \mathbf{U}_1^S \left(\mathbf{U}_1^S\right)^H \mathbf{N} \\ &= \mathbf{U}_1^M \Sigma_1^M (\mathbf{V}_1^M)^H - \mathbf{U}_1^M \left(\mathbf{U}_1^S\right)^H \mathbf{N} \\ &= \mathbf{U}_1^M \left[\Sigma_1^M (\mathbf{V}_1^M)^H - (\mathbf{U}_1^S)^H \mathbf{N}\right] \\ &= \mathbf{U}_1^M \left[\Sigma_1^M (\mathbf{V}_1^M)^H - (\mathbf{N}^H \mathbf{U}_1^S)^H\right] \\ &= \mathbf{U}_1^M \left[\Sigma_1^M (\mathbf{V}_1^M)^H - (\mathbf{N}^H \mathbf{U}_1^S + \mathbf{V}_1^S \Sigma_1^S - \mathbf{V}_1^S \Sigma_1^S)^H\right] \\ &= \mathbf{U}_1^M \left[\Sigma_1^M (\mathbf{V}_1^M)^H - (\mathbf{V}_1^M \Sigma_1 - \mathbf{V}_1^S \Sigma_1^S)^H\right] \\ &= \mathbf{U}_1^M \left\{\Sigma_1^M (\mathbf{V}_1^M)^H - \left[\Sigma_1 (\mathbf{V}_1^M)^H - \Sigma_1^S (\mathbf{V}_1^S)^H\right]\right\}. \end{aligned} \tag{37}$$

REFERENCES

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

Canales, L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527.

Chen, Y., 2015, Deblending using a space-varying median filter: Exploration Geophysics, **46**, 332–341.

Chen, Y., and S. Fomel, 2015, Random noise attenuation using local signal-and-noise orthogonalization: Geophysics, **80**, WD1–Wd9.

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

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

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

Chiu, S. K., 2013, Coherent and random noise attenuation via multichannel singular spectrum analysis in the randomized domain: Geophysical Prospecting, **61**, 1–9.

Donoho, D. L., 1995, De-noising by soft-thresholding: IEEE transactions on information theory, **41**, 613–627.

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

Gan, S., Y. Chen, S. Zu, S. Qu, and W. Zhong, 2015, Structure-oriented singular value decomposition for signal enhancement of seismic data: Journal of Geophysics and Engineering, **12**, 262–272.

Gan, S., S. Wang, Y. Chen, X. Chen, and K. Xiang, 2016, Separation of simultaneous sources using a structural-oriented median filter in the flattened dimension: Computers & Geosciences, **86**, 46–54.

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

Golub, G. H., and C. F. V. Loan, 1996, Matrix computations (third edition): John Hopkins University Press.

Gulunay, N., 1986, Fx decon and complex wiener prediction filter: 56th Annual International Meeting, SEG, Expanded Abstracts, 279–281.

———, 2000, Noncausal spatial prediction filtering for random noise reduction on 3-d poststack data: Geophysics, **65**, 1641–1653.

Huang, W., R. Wang, M. Zhang, Y. Chen, and J. Yu, 2015, Random noise attenuation for 3d seismic data by modified multichannel singular spectrum analysis: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, DOI:10.3997/2214–4609.201412830.

Jones, I. F., and S. Levy, 1987, Signal-to-noise ratio enhancement in multichannel seismic data via the karhunen-loeve transform: Geophysical Prospecting, **35**, 12–32.

Liu, G., X. Chen, J. Du, and J. Song, 2011, Seismic noise attenuation using nonstationary polynomial fitting: Applied Geophysics, **8**, 18–26.

Liu, G., X. Chen, J. Du, and K. Wu, 2012, Random noise attenuation using *f-x* regularized nonstationary autoregression: Geophysics, **77**, V61–V69.

Liu, Y., 2013, Noise reduction by vector median filtering: Geophysics, **78**, V79–V87.

Neelamani, R., A. Baumstein, D. Gillard, M. Hadidi, and W. Soroka, 2008, Coherent

and random noise attenuation using the curvelet transform: The Leading Edge, **27**, 240–248.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.

Trickett, S., 2008, F-xy cadzow noise suppression: CSPG CSEG CWLS Convention, 303–306.

Trickett, S. R., and L. Burroughs, 2009, Prestack rank-reducing noise suppression: Theory: 79th Annual International Meeting, SEG, Expanded Abstracts, 3332–3336.

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