Seismic signal enhancement based on the lowrank methods[a]

[a]Published in Geophysical Prospecting, 68, 2783-2807, (2020)

Min Bai[*†], *Guangtan Huang*[*], *Hang Wang*[*], *and Yangkang Chen*[*]

ABSTRACT

Based on the fact that the Hankel matrix constructed by noise-free seismic data is lowrank (LR), LR approximation (or rank-reduction) methods have been widely used for removing noise from seismic data. Due to the linear-event assumption of the traditional LR approximation method, it is difficult to define a rank that optimally separates the data subspace into signal and noise subspaces. For preserving the most useful signal energy, a relatively large rank threshold is often chosen, which inevitably leaves residual noise. To reduce the energy of residual noise, we propose an optimally damped rank-reduction method. The optimal damping is applied via two steps. In the first step, a set of optimal damping weights is derived. In the second step, we derive an optimal singular-value damping operator. We review several traditional lowrank methods and compare their performance with the new one. We also compare these lowrank methods with two sparsity-promoting transform methods. Examples demonstrate that the proposed optimally damped rank-reduction method could get significantly cleaner denoised images compared with the state-of-the-art methods.

INTRODUCTION

Random noise can seriously affect the stability and precision of seismic data processing and imaging steps including inversion-based migration, full waveform inversion, AVO inversion, and post-stack seismic interpretation. Thus, its removal is very important (Galbraith, 1991; Amani et al., 2017; Zhao et al., 2018; Li et al., 2020).

The sparse transform based denoising methods assume the seismic data to be sparse in the transform domain and the spreading noise in the transform domain can be suppressed by applying a thresholding operation. These methods include those based on the Fourier transform (Bracewell and Bracewell, 1986), Radon transform (Beylkin, 1987), seislet transform (Fomel and Liu, 2010; Chen and Fomel, 2018), curvelet transform (Candès et al., 2006), wavelet transform (Gilles, 2013; Mousavi et al., 2016), dreamlet transform (Huang et al., 2018), dictionary-learning based sparse transform (Zhou et al., 2016; Siahsar et al., 2017a,b; Zu et al., 2019). Prediction based methods are another group of popular denoising methods, such as time-space domain

prediction method (Abma and Claerbout, 1995), frequency-space predictive filtering (Canales, 1984), regularized non-stationary prediction method (Liu et al., 2012; Liu and Chen, 2013), and the polynomial fitting method (Liu et al., 2011). The decomposition based denoising methods consider the separability of seismic signal and random noise and attempt to extract useful information from the principal components of the noisy data. Typical methods include the empirical-mode decomposition (EMD) related methods (Huang et al., 1998), e.g., the ensemble EMD (Wu and Huang, 2009), complete ensemble empirical-mode decomposition (Colominas et al., 2012), improved complete ensemble empirical-mode decomposition (Colominas et al., 2012), singular-value decomposition (SVD) related methods (Bekara and van der Baan, 2007), and non-stationary decomposition with regularization (Li et al., 2018).

In this paper, we aim to improve on the multi-dimensional Cadzow filter applied to constant-frequency slices (Cadzow, 1988; Trickett, 2008), also referred to as multi-channel singular spectrum analysis (MSSA) (Oropeza and Sacchi, 2011; Chiu, 2013; Qiao et al., 2016). This filter has been widely adopted for seismic data processing due to its good performance (Ginolhac et al., 2013; Gao et al., 2017; Wang et al., 2018). This algorithm is based on lowrank (LR) matrix approximation. The main requirement of the LR methods is the low rank of the frequency-domain Hankel matrix. The rank of the Hankel matrix equals the number of distinct dips (Oropeza and Sacchi, 2011; Chen et al., 2016; Wang et al., 2020). However, the real seismic data are complicated, where the linear-events assumption is not met. To apply the LR methods, one needs to divide the field data into small time-space windows for separate processing (Zhang et al., 2017; Zu et al., 2017). Nevertheless, it will cause another problem in local processing windows, i.e., if we use a fixed rank for all the local windows, then it is possible that this rank is too large for some windows (so that the LR approximation keeps too much noise) and is too small for some other windows (such that the method loses the useful information). Thus, to optimize the denoising performance, it is desirable to find the appropriate rank for each local window. The rank can also be adaptively selected according to the ratio of two consecutive singular-values (Wu and Bai, 2018). However, all these strategies work only when the data structure is not complex and may not be applicable when noise is extremely strong.

In practice, the predefined rank is usually set large enough to preserve the useful signals without damaging weak and curving energy. The selection of a large rank could leave significant residual noise in the filtered data. One possible solution is to use a threshold to further suppress the residual noise after the rank-reduction step. This second-step thresholding can be understood conveniently in the framework of nuclear norm minimization (Zhou and Zhang, 2017). This strategy brings another challenging question on how to optimally choose the threshold for damping the residual noise. Considering that the thresholding step can be interpreted as a re-weighting process for the singular-values, we apply an adaptive singular-value weighting method following Nadakuditi (2013). This weighting method is an adaptive to shrinkage the singular values as compared with the direct truncation strategy. Aharchaou et al. (2017) introduced this adaptive weighting method to seismic data

reconstruction problems. Some other alternatives to the presented approach in this paper, such as those automated methods in (Gavish and Donoho, 2014) or Trickett (2015). One can further improve the optimal weighting based rank-reduction method by cascading the weighting strategy into the damed rank-reduction framework. The resulting algorithm is referred to as the optimal damped rank-reduction (ODRR) method. It has the potential to make the damped rank-reduction method effective for a wide range of rank selection in an adaptive way. We use different synthetic and field seismic datasets to show the advantages of the presented algorithm.

THEORY

Hankel matrix embedding

The rank-reduction based methods discussed in this paper deal with a block Hankel matrix (Inline and Xline) in the frequency-space domain. Let $\mathbf{D}(t,x,y)$ (of size $N_t \times N_x \times N_y$) represent a 3D seismic dataset. First, we transform $\mathbf{D}(t,x,y)$ in time-space domain to $\mathbf{D}(w,x,y)(w=1\cdots N_w)$ in the frequency-space domain. At a given frequency w_0 slice, the 2D data can be expressed as (Oropeza and Sacchi, 2011):

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

From here on, w_0 is omitted for notational convenience. A Hankel matrix is then constructed from $\mathbf{D}$. We first construct a Hankel matrix $\mathbf{R}_i$ as:

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

and then construct the block Hankel matrix as:

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

Parameters m and n are chosen to make $\mathbf{R}_i$ and $\mathbf{M}$ close to square matrices, e.g., $m = N_y - \lfloor \frac{N_y}{2} \rfloor$ and $n = N_x - \lfloor \frac{N_x}{2} \rfloor$. The symbol $\lfloor \cdot \rfloor$ outputs the integer of an input value. The matrix $\mathbf{M}$ is of size $I \times J$, with $I = (N_y - m + 1)(N_x - n + 1)$, $J = mn$. The block Hankel matrix $\mathbf{M}$ is considered to be lowrank (Trickett, 2008; Oropeza and Sacchi, 2011; Huang et al., 2016; Chen et al., 2019a), i.e., it can be approximated by a small number of eigen-images.

Rank-reduction method

Regularization can be implemented as a rank-based constraint (Trickett, 2008; Oropeza and Sacchi, 2011; Cheng and Sacchi, 2015):

$$\begin{aligned} \hat{\mathbf{X}} &= \arg\min_{\mathbf{x}} \parallel \mathbf{D} - \mathbf{X} \parallel_F^2, \\ \text{subject to} \quad & \text{rank}(\mathcal{H}(\mathbf{X})) = N, \end{aligned} \tag{4}$$

where $\| \cdot \|_F$ denotes the Frobenius norm. $\mathbf{D}$ denotes a matrix constructed from the frequency slice corresponding to w_0. $\mathbf{X}$ denotes the noise-free data to be estimated. $\hat{\mathbf{X}}$ denotes the estimated signal. $\mathcal{H}(\mathbf{X})$ denotes the Hankel matrix constructed from $\mathbf{X}$. Let $\mathbf{M} = \mathcal{H}(\mathbf{D})$, the singular value decomposition of $\mathbf{M}$ can be expressed as

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{5}$$

where $\mathbf{U}$ and $\mathbf{V}$ are referred to as the left and right singular vector matrices, respectively. $\mathbf{\Sigma}$ is the singular value matrix. $(\cdot)^H$ denotes complex tranpose. According to equation 4, the rank of the signal component that is embedded in the Hankel matrix $\mathbf{M}$ is assumed to be N. The traditional rank-reduction method based on the truncated singular value decomposition (TSVD) (Oropeza and Sacchi, 2011) can be briefly expressed as

$$\hat{\mathbf{S}} = \mathbf{U}_N\mathbf{\Sigma}_N\mathbf{V}_N^H, \tag{6}$$

which is a solution to equation 4 according to the Eckart-Young-Mirsky theorem (Eckart and Young, 1936). $\hat{\mathbf{S}}$ denotes the denoised signal. $\mathbf{U}_N$ and $\mathbf{V}_N$ are matrices composed of the left N singular vectors in $\mathbf{U}$ and $\mathbf{V}$, respectively. $\mathbf{\Sigma}_N$ is the truncated singular value matrix with the first N singular values preserved. Although theoretically N equals to the number of distinct dipping components, it is practically defined as a relatively large number considering data complexity, otherwise signal energy can be damaged. The algorithm workflow for the traditional rank-reduction method (RR) is outlined in Algorithm 1. $\mathcal{A}$ in the algorithm workflow denotes an averaging operator along anti-diagonals.

Damped rank-reduction method

The estimated signal using the TSVD method, however, still contains non-negligible residual noise components, as explained in Huang et al. (2016). The problem of residual noise can be solved to some extent by further applying a thresholding operator to the singular value matrix according to the nuclear-norm minimization model (Zhou and Zhang, 2017):

$$\hat{\mathbf{S}} = \mathbf{U}_N\mathcal{T}(\mathbf{\Sigma}_N, \tau)\mathbf{V}_N^H, \tag{7}$$

where $\mathcal{T}$ denotes the thresholding operator for the singular-values, and τ denotes the threshold of the operator. Defining an optimal threshold τ, however, is often challenging because a constant threshold is inadequate to handle an inhomogeneous

distribution of noise energy. Huang et al. (2016) developed a more elegant way to shrinkage the singular values by deriving a damping operator:

$$\mathbf{P} = \mathbf{I} - \mathbf{\Gamma}, \tag{8}$$

$$\mathbf{\Gamma} = \hat{\sigma}^K \left(\mathbf{\Sigma_N}\right)^{-K}, \tag{9}$$

where $\mathbf{P}$ is called the damping matrix, $\mathbf{I}$ is an identity matrix, and $\mathbf{\Gamma}$ is referred to as the damping threshold matrix. $\hat{\sigma}$ is the $(N+1)$th singular value in the un-truncated singular matrix $\mathbf{\Sigma}$. K in equation 9 is called the damping factor, which is used to control the strength of the damping operator. In a special case, when $K \to \infty$, $\mathbf{P} \to \mathbf{I}$, indicating that the damped rank-reduction method reverts to the traditional rank-reduction method. The mathematical details to derive the damping operator can be found in Huang et al. (2016) and Huang et al. (2017). The damping operator is used to shrinkage the singular values in the TSVD formula to reduce the residual noise:

$$\hat{\mathbf{S}} = \mathbf{U}_N \mathbf{P} \mathbf{\Sigma}_N \mathbf{V}_N^H. \tag{10}$$

Compared with the thresholding method in equation 7, where a rigid threshold τ is used to shrinkage the singular values, the damping operator applies variable thresholds to different singular values. According to equation 8, the threshold decreases as the residual noise becomes less dominant, i.e., the singular value becomes larger. The algorithm workflow for the damped rank-reduction method (DRR) is outlined in Algorithm 2.

Optimally damped rank-reduction method

A shortcoming of the traditional rank-eduction and damped rank-reduction methods is that the output perfomance is quite sensitive to the input parameter N. The rank parameter varies greatly for different datasets, especially for field datasets. In order to alleviate the influence of the rank parameter, we take advantage of an adaptive singular-value weighting algorithm developed by Benaych-Georges and Nadakuditi (2012) and Nadakuditi (2013). An adaptive singular-value weighting matrix can be obtained by solving the following problem:

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \| \sum_{i=1}^{N} \sigma_i \mathbf{u}_i^S (\mathbf{v}_i^S)^H - \omega_i \sigma_i \mathbf{u}_i \mathbf{v}_i^H \|_F, \tag{11}$$

where $\mathbf{u}_i^S$ and $\mathbf{v}_i^S$ denote ith left and right singular vectors corresponding to the estimated signal component, σ_i denotes the ith singular value in $\mathbf{\Sigma}$. $\boldsymbol{\omega}$ is the adaptive weight vector, which can be used to construct the adaptive weighting matrix $\hat{\mathbf{W}} = \mathrm{diag}(\hat{\boldsymbol{\omega}})$. The solution to the optimization problem can be obtained as:

$$\hat{w}_i = \left(-\frac{2}{\sigma_i} \frac{\mathcal{D}(\sigma_i; \mathbf{\Sigma})}{\mathcal{D}'(\sigma_i; \mathbf{\Sigma})} \right). \tag{12}$$

where $\mathcal{D}$ represents a transform expressed as:

$$\mathcal{D}(\sigma;\mathbf{\Sigma}) = \frac{1}{N^2}\mathrm{Tr}\left(\sigma(\sigma^2\mathbf{I}-\mathbf{\Sigma}\mathbf{\Sigma}^H)^{-1}\right)\mathrm{Tr}\left(\sigma(\sigma^2\mathbf{I}-\mathbf{\Sigma}^H\mathbf{\Sigma})^{-1}\right), \tag{13}$$

and $\mathcal{D}'$ denotes its derivative. The expression of $\mathcal{D}'$ can be expressed as:

$$\begin{aligned} D'(\sigma;\mathbf{\Sigma}) &= 2\left[\frac{1}{N}\mathrm{Tr}\left(\sigma(\sigma^2\mathbf{I}-\mathbf{\Sigma}^2)^{-1}\right)\right]\left[\frac{1}{N}\mathrm{Tr}\left((\sigma^2\mathbf{I}-\mathbf{\Sigma}^2)^{-1}-2\sigma(\sigma^2\mathbf{I}-\mathbf{\Sigma}^2)^{-2}\sigma\right)\right] \\ &= \frac{2}{N^2}\left[\mathrm{Tr}\left(\sigma(\sigma^2\mathbf{I}-\mathbf{\Sigma}^2)^{-1}\right)\right]\left[\mathrm{Tr}\left((\sigma^2\mathbf{I}-\mathbf{\Sigma}^2)^{-1}-2\sigma^2(\sigma^2\mathbf{I}-\mathbf{\Sigma}^2)^{-2}\right)\right]. \end{aligned} \tag{14}$$

The symbol $\mathrm{Tr}(\cdot)$ denotes the trace of the input. The trace of a square matrix $\mathbf{X}$ is defined to be the sum of elements on the main diagonal of $\mathbf{X}$.

$$\mathrm{Tr}(\mathbf{X}) = \sum_{i=1}^{N} X_{i,i}, \tag{15}$$

where $X_{i,i}$ denotes the main diagonal elements of $\mathbf{X}$.

The adaptive weighting operator can be applied to make the traditional rank-reduction method adaptive:

$$\hat{\mathbf{S}} = \mathbf{U}_N\hat{\mathbf{W}}\mathbf{\Sigma}_N\mathbf{V}_N^H. \tag{16}$$

To incorporate the adaptive weighting operator into the damped rank-reduction framework, we can introduce intermediate variables as

$$\mathbf{U}_N^Q = \mathbf{U}_N, \tag{17}$$

$$\mathbf{\Sigma}_N^Q = \hat{\mathbf{W}}\mathbf{\Sigma}_N, \tag{18}$$

$$\mathbf{V}_N^Q = \mathbf{V}_N, \tag{19}$$

then equation 16 turns into

$$\hat{\mathbf{S}} = \mathbf{U}_N^Q\mathbf{\Sigma}_N^Q\mathbf{V}_N^Q. \tag{20}$$

It can be derived that the damping formula (equation 8) also holds for equation 20 but with the damping threshold matrix expressed as

$$\mathbf{\Gamma} = \hat{\sigma}^K\left(\mathbf{\Sigma}_{\boldsymbol{N}}^{\boldsymbol{Q}}\right)^{-K}, \tag{21}$$

where the subscript N denotes a sufficiently large rank parameter, as required by the derivations detailed in Chen et al. (2019b). The resulted final form of the optimally damped rank-reduction method then can be expressed as

$$\hat{\mathbf{S}} = \mathbf{U}_N\mathbf{P}\hat{\mathbf{W}}\mathbf{\Sigma}_N\mathbf{V}_N^H. \tag{22}$$

The algorithm workflow for the optimally damped rank-reduction method (ODRR) is outlined in Algorithm 3. Due to the existence of a weighting matrix in the proposed

method, it is more convenient to choose the input rank parameter N empirically in practice. As mentioned previously, the rank parameter N is in practice set to a relatively large value to forestall damaging the signal. But in the proposed algorithm, even if a large N is used, the algorithm can adaptively shrink the singular-values. Thus, its performance is not sensitive to the input rank parameter, in contrast to other related approaches (Oropeza and Sacchi, 2011). Because of the insensitivity, one can use a sufficiently large N in processing complicated datasets without leaving strong residual noise in the result. The convenience in tuning parameters makes the rank-reduction related methods more computationally feasible in large-scale data processing, e.g., the 5D reconstruction problem (Chen et al., 2019b), since one no longer needs to tune the parameters many times while one trial is already computationally demanding. A glossary describing the main mathematical notations are presented in Table 1. The three methods are all variations of how the singular-values are thresholded. Simply speaking, the DRR method improves the RR method by introducing a damping operation and the ODRR method improves DRR method by further introducing a weighting operation.

RESULTS

In this section, several synthetic and field data examples will be used to test the validity of the proposed method. We will use examples containing both crossing linear events and hyperbolic events to test the sensitivity of methods to different data structures. For synthetic examples, we use the signal-to-noise ratio (SNR) metric for quantitative evaluation, which is defined as follows:

$$\mathrm{SNR(dB)} = 10\log_{10}\frac{\|\mathbf{S}\|_F^2}{\|\mathbf{S}-\hat{\mathbf{S}}\|_F^2}. \tag{23}$$

where $\hat{\mathbf{S}}$ and $\mathbf{S}$ denote the estimated signal and exact solution, respectively. In addition, we use the local similarity attribute (Chen and Fomel, 2015) to measure the damage that a denoising method can cause to seismic data. In general, higher local similarity between the denoised data and removed noise indicates greater damages. The local similarity attribute allows to quantify the effectiveness of a denoising algorithm on datasets when the exact solution is unknown. The local similarity can also measure the local orthogonality between the separated signal and noise components (Chen and Fomel, 2015). Here, we assume that the signal and noise components should be locally orthogonal, which is indicated by low anomalies in the local similarity maps. Processing pre-stack seismic data is much more challenging than that for post-stack seismic data. Subtle diffraction signals could be treated as noise by the algorithm and be attenuated, potentially damaging images along faults. In this paper, we only focus on the denoising of post-stack seismic data, but further work could focus on noise attenuation of pre-stack data.

Synthetic examples

The first example is shown in Figure 1 including both clean and noisy data. The noisy data has an SNR of -8.39 dB. In this test, we compare the denoising performance of four different methods: the frequency-wavenumber (FK) domain thresholding method (Mahdad, 2012; Abma et al., 2015), the rank-reduction method (RR), the damped rank-reduction method (DRR), and the proposed method (ODRR). In this example, we do not use local windows. We use the Ricker wavelet with a dominant frequency of 40 Hz to generate the clean data. We add Gaussian white noise with variance of 0.2 (as compared with the normalized clean data). Since the proposed method is an improved version of the DRR method, we compare the proposed method with two other rank-reduction methods. Since the FK method is one of most commonly used approaches to attenuate high-frequency high-wavenumber random noise, we also treat it as a competing method for this comparison. Figure 2 shows denoised data using different methods. The top row of Figure 2 shows the the dataset denoised by the four methods, while the middle row of Figure 2 presents the noise extracted by the four methods. In this test, we preserve the 10% largest coefficients for the FK method to obtain Figure 2a. We use a rank of $N = 3$ for all three rank-reduction based methods. By observing Figure 2a, it is clear that the FK thresholding causes strong artifacts around the edges. Both FK thresholding and the rank-reduction methods leaves a large amount of residual noise in the denoised results. The calculated SNRs of the four methods are 6.03 dB for the FK thresholding method, 6.57 dB for the rank-reduction method, 10.29 dB for the damped rank-reduction method, and 11.27 dB for the proposed method. It is clear that the SNR comparison quantitatively confirms our initial observation that the proposed method obtains the best result among the four methods. In this test we use a very simple synthetic example that contains three planar components. In the rank-reduction method, it can be proved that the number of the planar/linear events is equal to the input rank parameter, i.e., $N = 3$ in this test (Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011). Because we know the ground truth in this test, we the same rank for all methods, and the proposed method obtains the best performance but is slightly better than the damped method. However, in the case of the more complex data, where the exact number of events with distinct dips is unknown, we prefer to choose a more conservative rank to avoid losing too many useful details. The bottom row of Figure 2 shows the comparison of local similarity, where we can find that the local similarity of the ODRR method is much smaller. The spectrum comparison when $N = 3$ is plotted in Figure 3, which further verify the best performance of the proposed method.

Next, based on the same example, we use a larger rank $N = 6$. We also use a more conservative threshold value for the FK thresholding method, e.g., we preserve the 20% largest coefficients in the transformed domain. The results of the four methods are shown in the top row of Figure 4. The middle row of Figure 4 shows the corresponding noise cubes. It is clear that the denoised results of FK thresholding method, rank-reduction method, damped rank-reduction method are all noisier than the corresponding results presented in Figure 2. However, the result from the

proposed method, as shown in Figure 4d, is less affected. The calculated SNRs in this case are 4.31 dB for the FK method, 3.45 dB for the RR method, 8.35 dB for the DRR method, and 10.86 dB for the proposed method. The differences of SNR with respect to the previous example for the three methods are 1.72 dB for the FK thresholding method, 3.12 dB for the rank-reduction method, 1.94 dB for the damped rank-reduction method, and 0.41 dB for the proposed method. Comparison of SNRs demonstrates that while the other three methods are sensitive to the input parameter, the proposed method is much less sensitive, because it can still find the appropriate rank by using weights. The bottom row of Figure 4 shows the comparison of local similarity, where we find that the local similarity of the presented method is distinctly smaller. The spectrum comparison when $N = 6$ is plotted in Figure 5, which further verify the best performance of the proposed method.

For a better comparison, we extract the 5th Xline slice from Figures 1 and 4 and show the slices in Figure 6. In this display, it is even more evident that the proposed method produces the cleanest result while minimizing signal damage. Figure 7 plots the SNR diagrams of the different methods with respect to the input parameters. The input parameters for the rank-reduction based methods and transform based methods are the selected rank and the percentage of sparse coefficients, respectively. When the rank is chosen large enough, e.g., larger than 8, the proposed method is less sensitive in all methods. The two transform based methods are also sensitive to the percentage of selected coefficients.

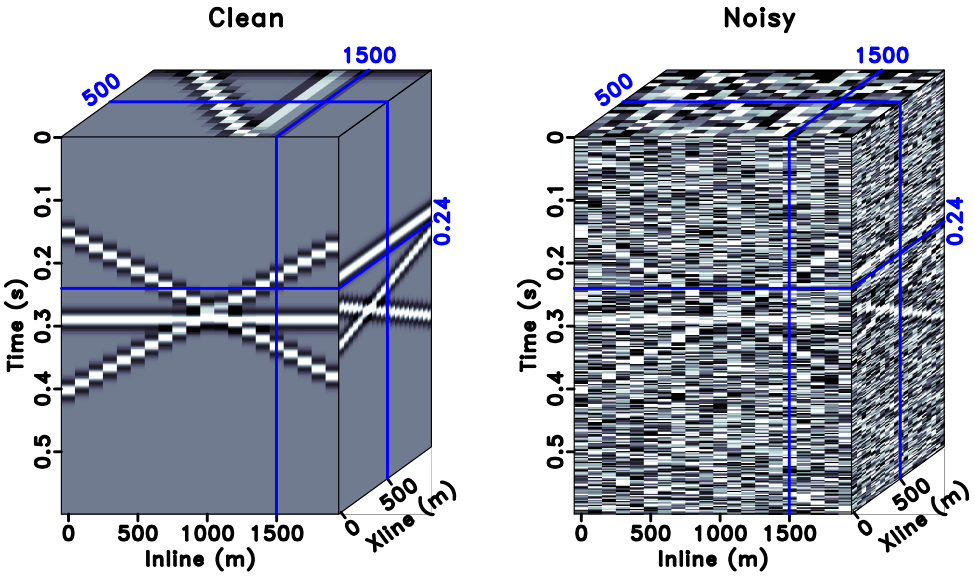
To test the performance of the proposed method in denoising high-frequency band. We create a slightly different example shown in Figure 8. We increase the dominant frequency of the Ricker wavelet to 60 Hz and then extract the frequency band of 60-100 Hz. We also extract the same frequency band of the Gaussian white noise, and add the high-frequency noise to the clean data to generate the noisy data. The denoising comparison of different methods for the high-frequency band is shown in Figure 9. The spectrum comparison is shown in Figure 10. This test demonstrates the lowrank methods also work well in high-frequency band.

The next example is a synthetic dataset with hyperbolic events. In this example, we use the Ricker wavelet with a dominant frequency of 10 Hz to generate the clean data. We add Gaussian white noise with variance of 0.2 (as compared with the normalized clean data). The clean and noisy data are shown in Figures 11a and 11b, respectively. The SNR of the noisy data is -2.17 dB. We apply FK thresholding method and the three rank-reduction based methods to this example and show the results in Figure 12. For the FK thresholding method, we preserve 20% largest coefficients in the transformed domain. For all the rank-reduction based methods, we use $N = 10$. Since the hyperbolic events no longer satisfy the assumption, i.e., being lowrank, the hyperbolic events in this example are over-smoothed when $N = 10$. It is also clear that both FK thresholding and the rank-reduction methods have significant residual noise. The damped rank-reduction method and the proposed method are both very clean, but the proposed method is slightly smoother. In this example, because of the hyperbolic events and their small rank, the removed noise cubes as

shown in the bottom row of Figure 12 contain a small amount of spatially coherent energy, which indicates signal leakage (Chen and Fomel, 2015). The calculated SNRs in this example are 7.05 dB for the FK thresholding method, 8.27 dB for the rank-reduction method, 9.58 dB for the damped rank-reduction method, and 9.65 dB for the proposed method. In this example, we do not use local windows to locally pretend that hyperbolic events act as linear events. When applying local windows, additional parameters (e.g., the window size) need to be compared and considered. To avoid the this step, we can use a relatively large rank to avoid the signal damage. From the comparison of local similarity, significant damage is highlighted as high similarity anomalies. We increase the rank from $N = 10$ to $N = 20$ and show the results in Figure 14. For the FK thresholding method, we increase the threshold percentage from 20% to 40%. We find that in this test, both FK thresholding method and the rank-reduction method leave more residual noise while the results from the damped rank-reduction method and the proposed method are still very smooth. However, when $N = 20$, the rank-reduction methods do not produce significant damage to useful signals but they also do not attenuate the noise very well. In addition, when $N = 20$, the proposed method becomes obviously smoother than the damped rank-reduction method. The comparison is more noticeable when observing the local similarity maps. The calculated SNRs in this example are 5.85 dB for the FK thresholding method, 7.04 dB for the rank-reduction method, 10.08 dB for the damped rank-reduction method, and 11.00 dB for the proposed method. Figure 16 shows a single slice comparison of this example (5th Xline slice) when $N = 20$, where it is more noticeable that the proposed method obtains the best result. The SNRs for all the aforementioned tests are given in Table 2 for a detailed comparison. Figure 17 plots the SNR diagrams of the different methods with respect to the input parameters for the hyperbolic example. When the rank is sufficiently large, e.g., larger than 18, the proposed approach is clearly more insensitive to the rank when compared to the other rank-reduction based methods due to the calculation of adaptive weights for the singular-values. Table 3 compares the computational costs for all these tests. All three methods are comprable but the proposed method is slightly more expensive.

Field data example

Next we apply the proposed method to the real migrated 3D land seismic dataset shown in Figure 18. The 3D seismic image corresponds to a land dataset after time migration. The field dataset goes through a normal seismic processing workflow, e.g., muting the dead traces, ground roll removal, surface-consistent deconvolution, pre-stack noise attenuation by the FX method, NMO-based velocity analysis, and kirchhoff time migration. The temporal sampling interval is 4 ms. There are 200 inlines and 50 crosslines, with trace spacings of 5 m and 10 m, respectively. Figure 19 shows the denoising comparison using different methods. From the previous synthetic examples, we understand the differences for different rank-reduction related methods. Therefore, in this example we examine the denoising performance using the seislet transform (Fomel and Liu, 2010). The seislet transform is deemed to be the sparsest



(a) (b)

Figure 1: Synthetic data examples with linear/planar events. (a) Clean data. (b) Noisy data. The events shown on the outside of the cube are situated at the blue lines within the cube.

transform for seismic data. The denoised data using the seislet thresholding method, the rank-reduction method, and the proposed method are shown in the left column of Figure 19, respectively. The middle column of Figure 19 shows the corresponding noise cubes of the three methods. Since for the field data example we do not have the pure signal for calculating the SNR, we can only use the local similarity metric to evaluate the denoising performance. The general criterion is that the local similarity between the denoised data and removed noise should be negligible provided that there is no signal leakage in the removed noise. The local similarity cubes corresponding to the three methods are shown in the right column of Figure 19. To make the removed noise comparably strong, we preserve 8% largest coefficients in the seislet domain. We use rank $N = 21$ for the rank-reduction method and use rank $N = 30$ for the proposed method. Comparison of the local similarity shows that the seislet transform and the rank-reduction methods both cause significant signal leakage, while the proposed method is almost damage-free for the useful signals. It is clear that when removing the same amount of random noise, it is able to preserve the most signal energy for the proposed method. At the same time, the new method can get a smoother result than the rank-reduction method, and same smoothness level compared with the seislet thresholding method. The spectrum comparison for the field data example is plotted in Figure 20, where the proposed method preserves more signal spectra than the seislet method, but removes more noise spectra than the rank-reduction method. The zoomed comparison in Figure 21 makes this even more obvious.

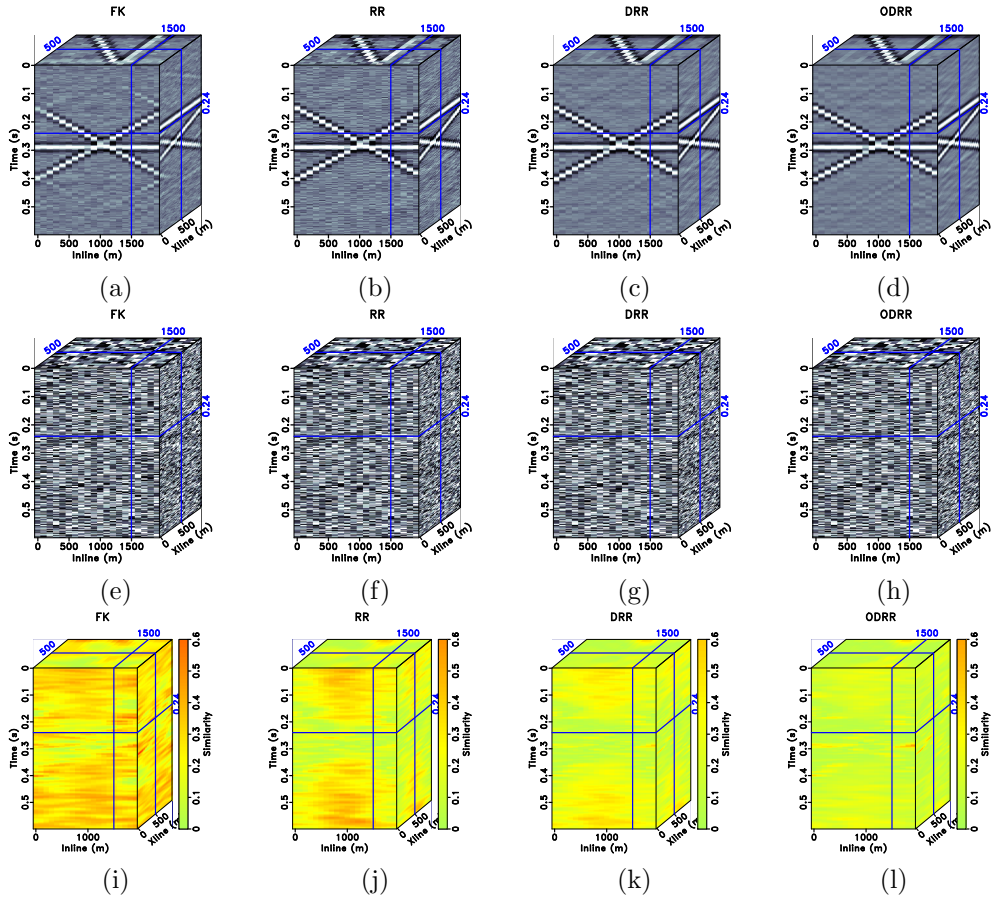


Figure 2: Denoising comparison ($N = 3$). Top row: denoised results using (a) FK method with 10% largest coefficients, (b) rank-reduction method, (c) damped rank-reduction method, and (d) the proposed method. Middle row: separated noise corresponding to the top row. Bottom row: local similarity corresponding to the top row.

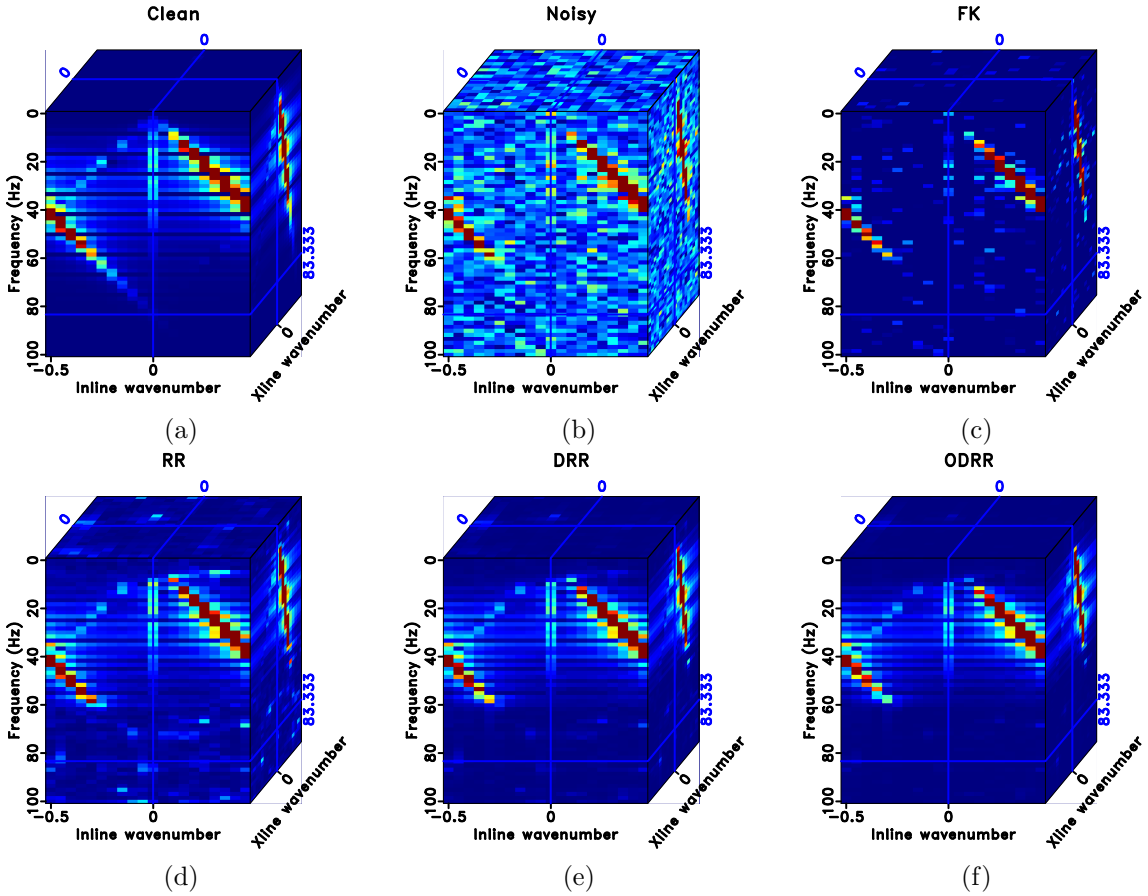


Figure 3: Spectrum comparison ($N = 3$). FK spectrum of (a) clean data, (b) noisy data, (c) FK method, (d) rank-reduction method, (e) damped rank-reduction method, and (f) the proposed method.

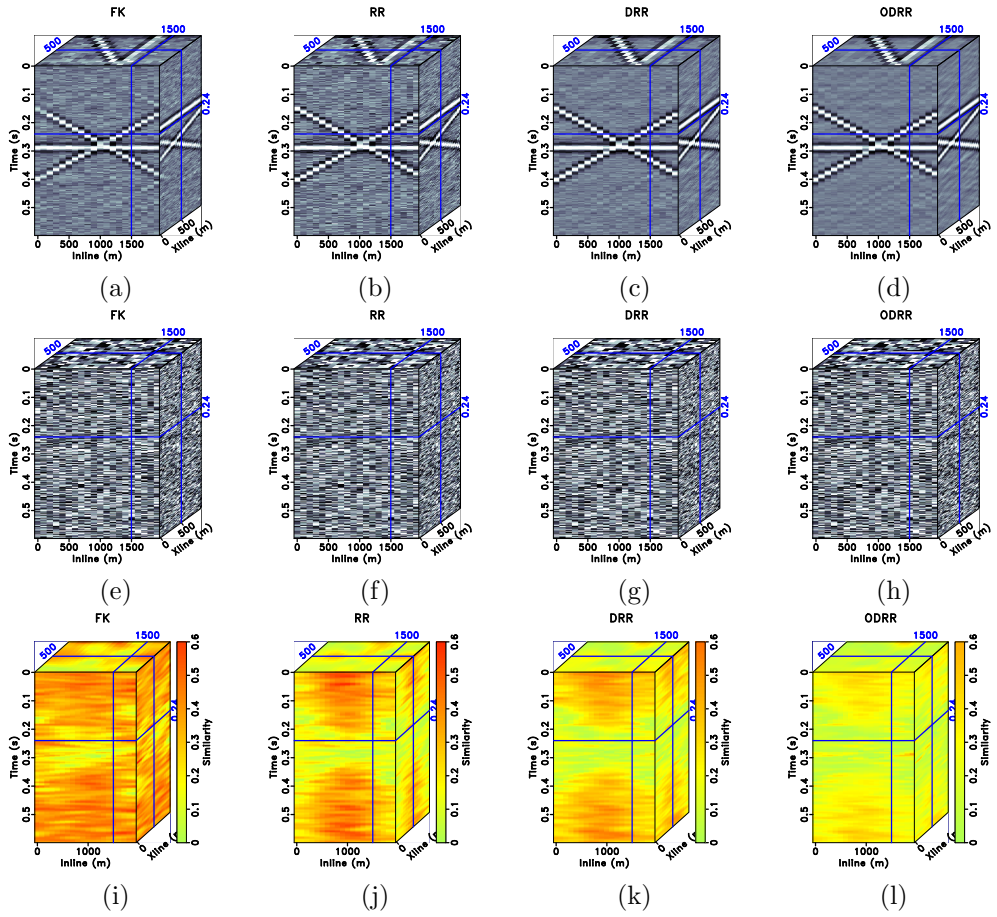


Figure 4: Denoising comparison ($N = 6$). Top row: denoised results using (a) FK method with 20% largest coefficients, (b) rank-reduction method, (c) damped rank-reduction method, and (d) the proposed method. Middle row: separated noise corresponding to the top row. Bottom row: local similarity corresponding to the top row.

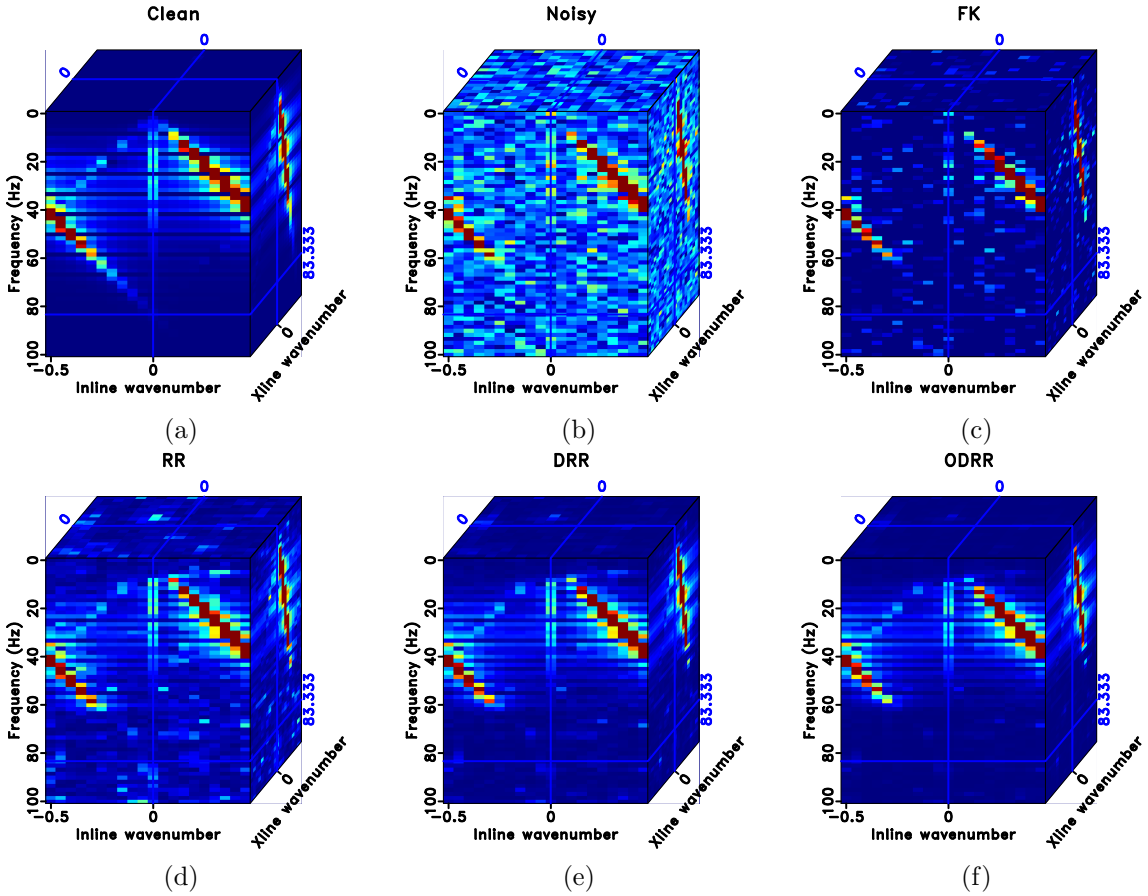


Figure 5: Spectrum comparison ($N = 6$). FK spectrum of (a) clean data, (b) noisy data, (c) FK method, (d) rank-reduction method, (e) damped rank-reduction method, and (f) the proposed method.

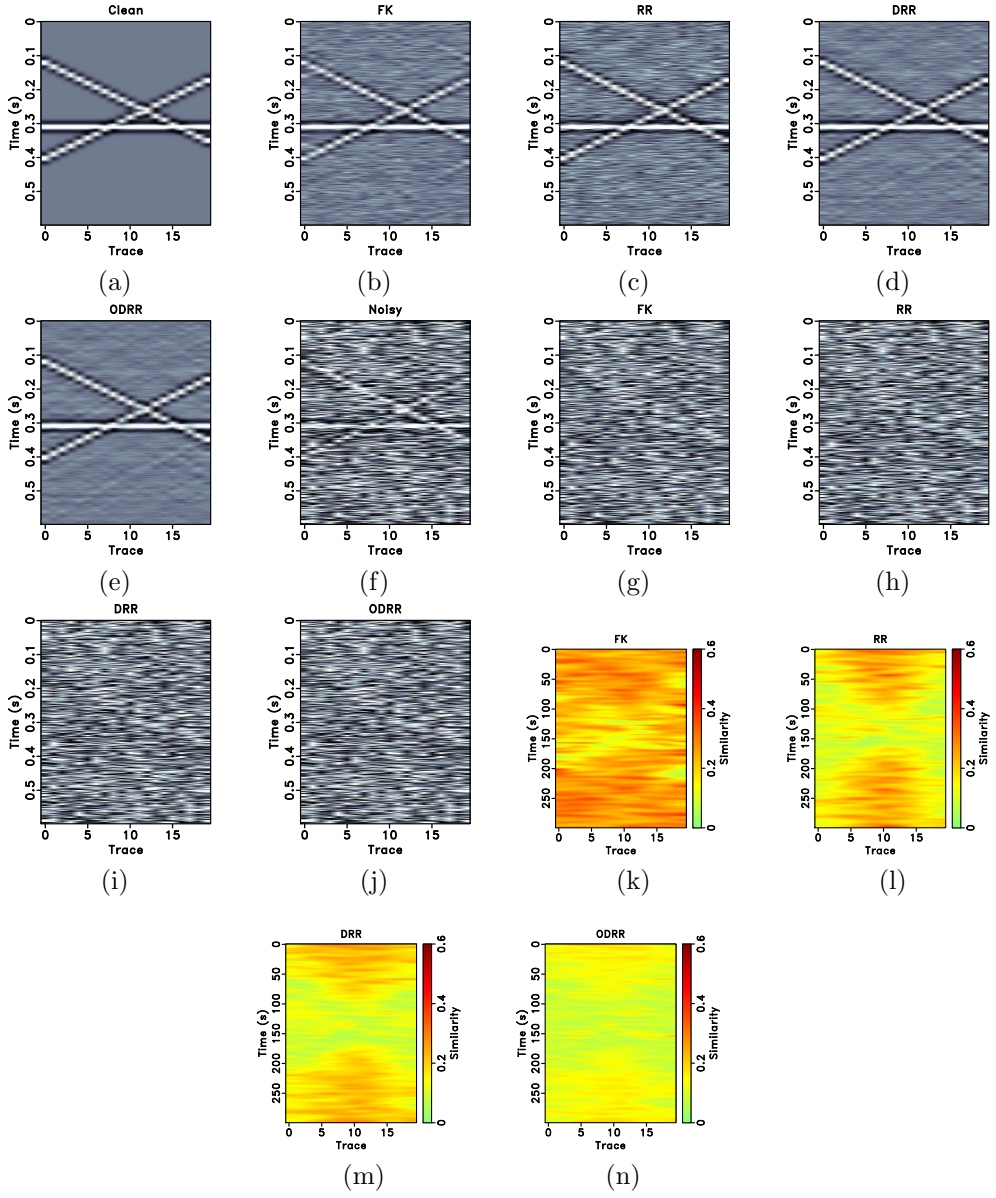


Figure 6: 2D slice view of denoising comparison ($N = 6$). (a) Clean data. (b)-(e): denoised results using FK method, (b) rank-reduction method, (c) damped rank-reduction method, and the proposed method, respectively. (f) Noisy data. (g)-(j) Separated noise corresponding to (b)-(e), respectively. (k)-(n) Local similarity corresponding to (b)-(e), respectively.

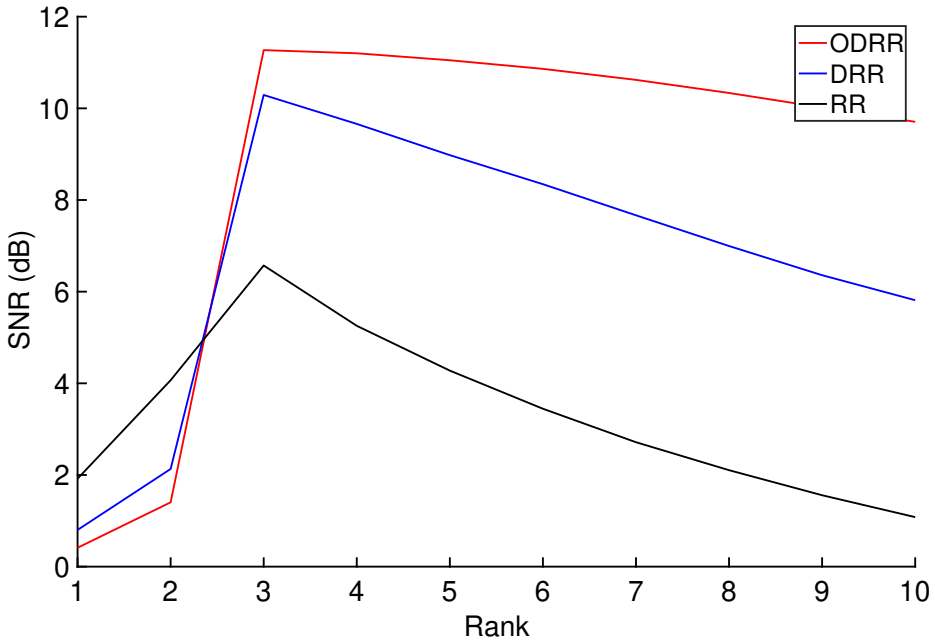
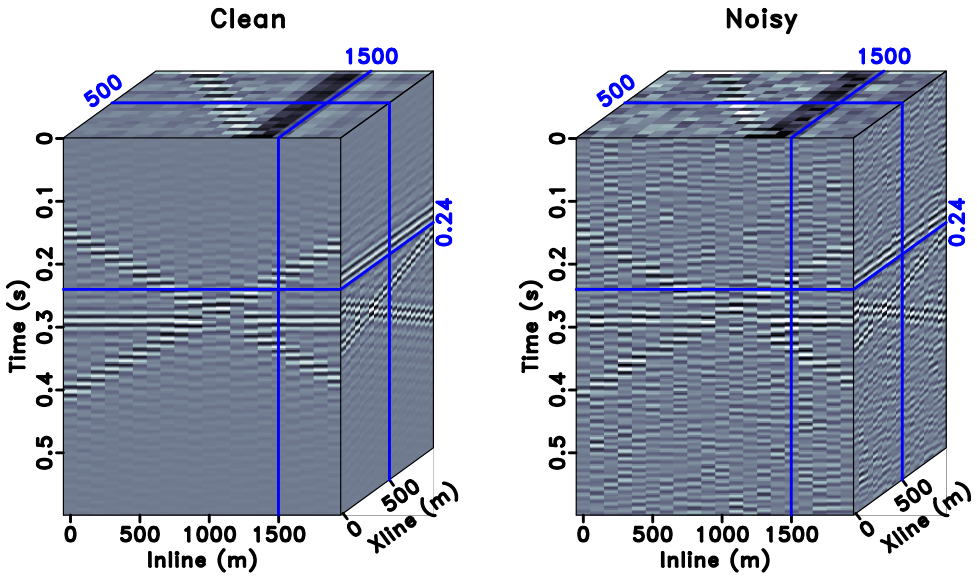


Figure 7: SNR diagrams of different lowrank approaches with respect to the selected rank parameters for the linear synthetic example.



(a) (b)

Figure 8: Synthetic data examples with linear/planar events for the high-frequency denoising test (60-100 Hz). (a) Clean data. (b) Noisy data. The events shown on the outside of the cube are situated at the blue lines within the cube.

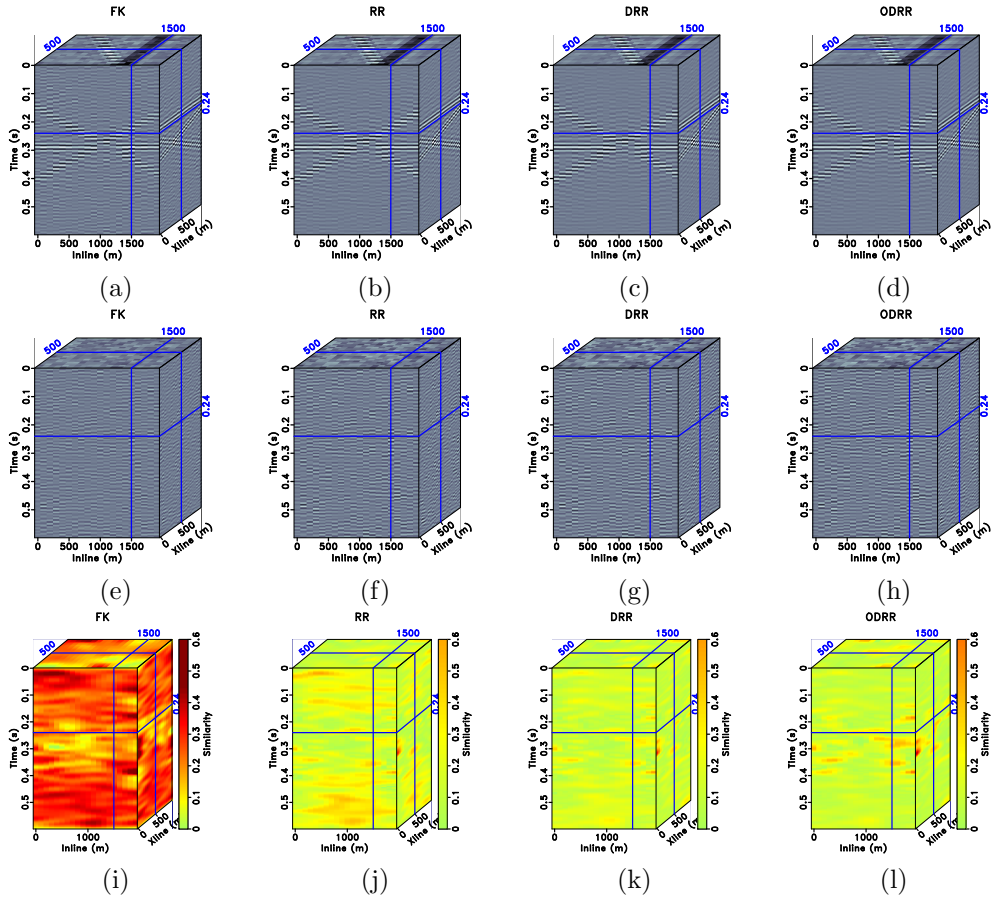


Figure 9: Denoising comparison ($N = 6$) for the high-frequency denoising test (60-100 Hz). Top row: denoised results using (a) FK method with 10% largest coefficients, (b) rank-reduction method, (c) damped rank-reduction method, and (d) the proposed method. Middle row: separated noise corresponding to the top row. Bottom row: local similarity corresponding to the top row.

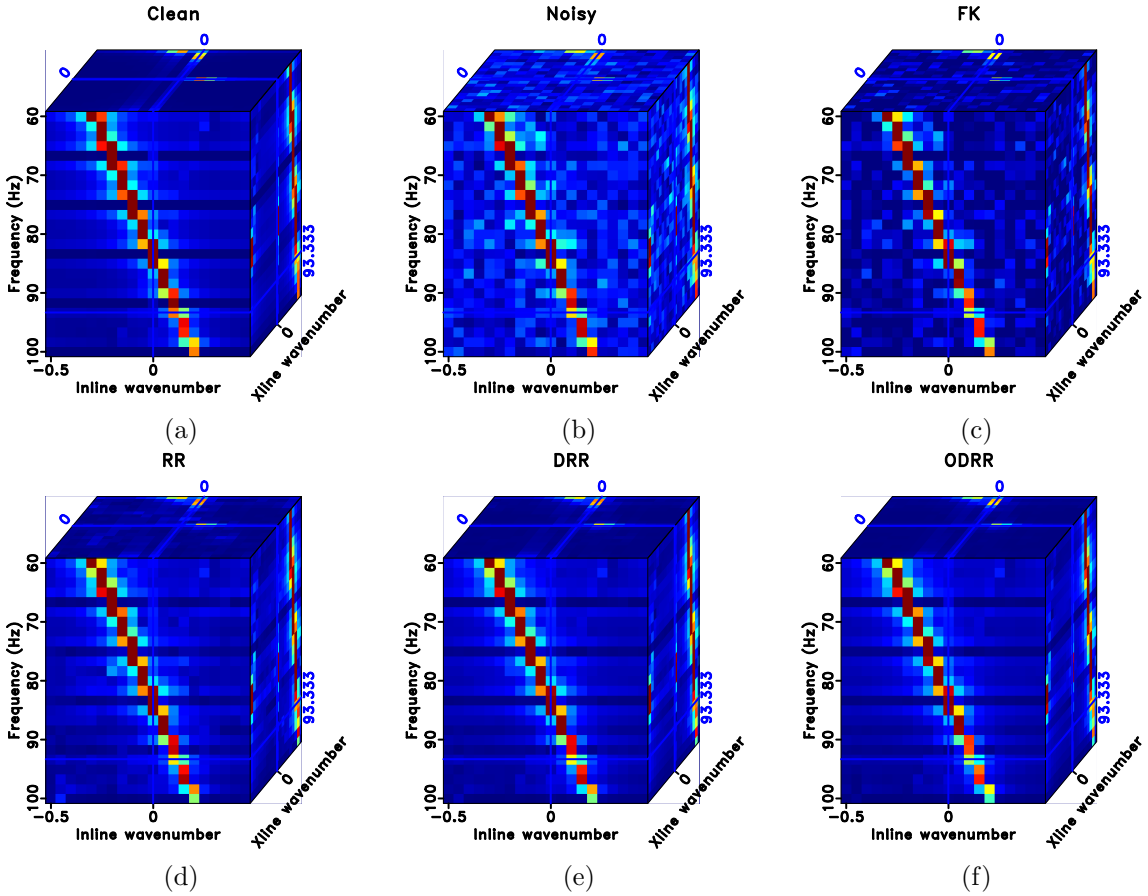
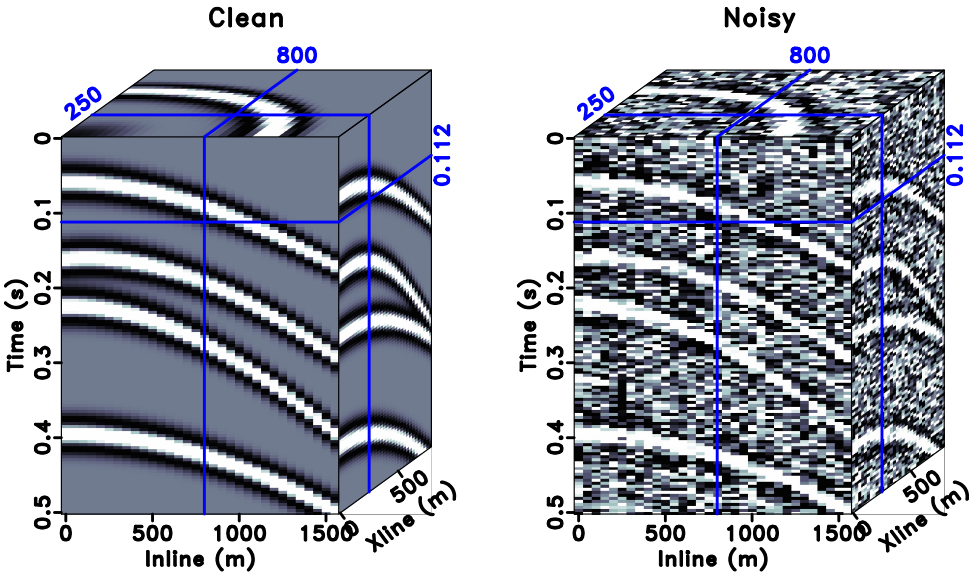


Figure 10: Spectrum comparison ($N = 6$) for the high-frequency denoising test (60-100 Hz). FK spectrum of (a) clean data, (b) noisy data, (c) FK method, (d) rank-reduction method, (e) damped rank-reduction method, and (f) the proposed method.



(a) (b)

Figure 11: Synthetic data examples with hyperbolic events. (a) Clean data. (b) Noisy data. The gray scales for all of the images shown in Figures 11-16 are the same.

CONCLUSIONS

The rank-reduction method for seismic noise suppression based on the nuclear norm minimization requires a carefully selected threshold value. We have developed an optimal weighting strategy to obtain better shrinkage of the singular-values, bypassing the need for manual selection of the rank as a priori information. Considering the issue of residual noise after the optimal rank-reduction method, we further introduce an optimal way to damp the remaining noise. The proposed method can separate the noise subspace and the signal subspace in an optimal way. Detailed analyses on the proposed algorithm via two synthetic datasets and one field example demonstrate that the proposed method can obtain better denoising performance regarding the noise removal and signal preservation than the widely used methods in terms of SNR and local similarity measurements. More importantly, because of the optimally damped singular-values, the proposed method is an adaptive method, i.e., the performance is not sensitive to the predefined rank parameter.

DATA AVAILABILITY STATEMENT

Datasets and source codes associated with this research will be made available online (www.ahay.org) in the format of reproducible research.

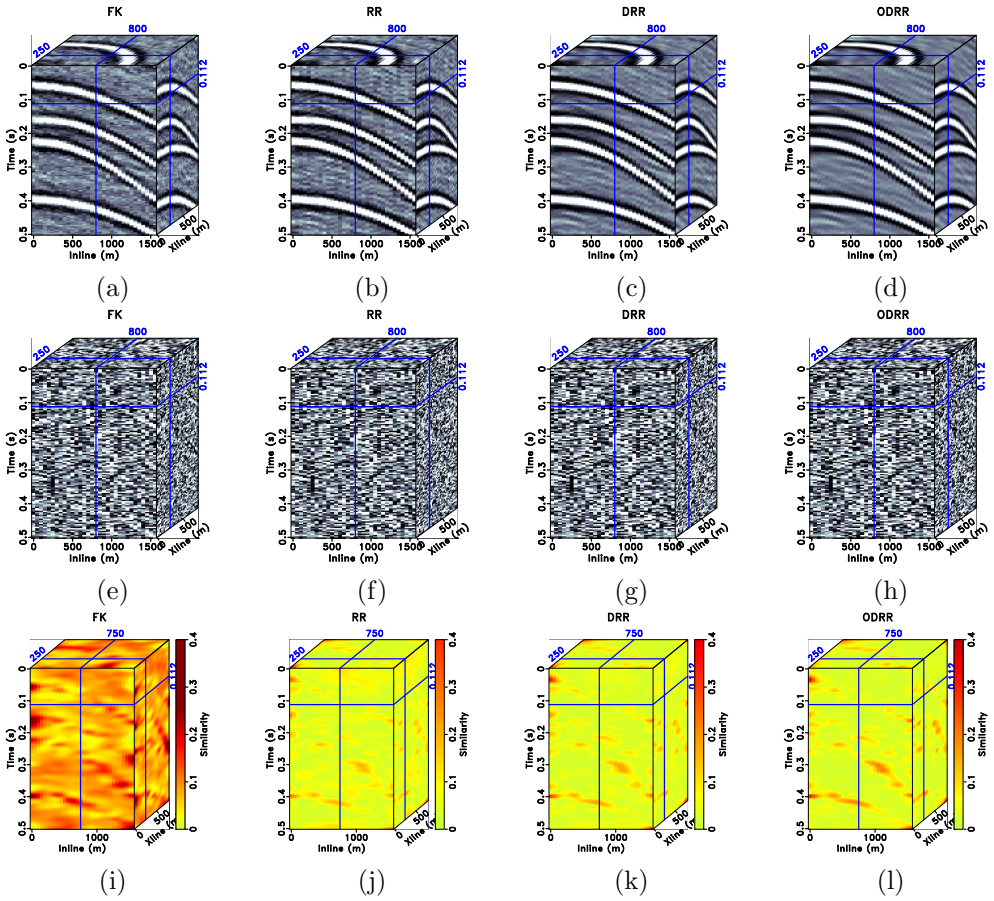


Figure 12: Denoising comparison ($N = 10$). The top row: denoised results using (a) FK method, (b) rank-reduction method, (c) damped rank-reduction method, and (d) the proposed method. The middle row: separated noise corresponding to the top row. The bottom row: local similarity corresponding to the top row.

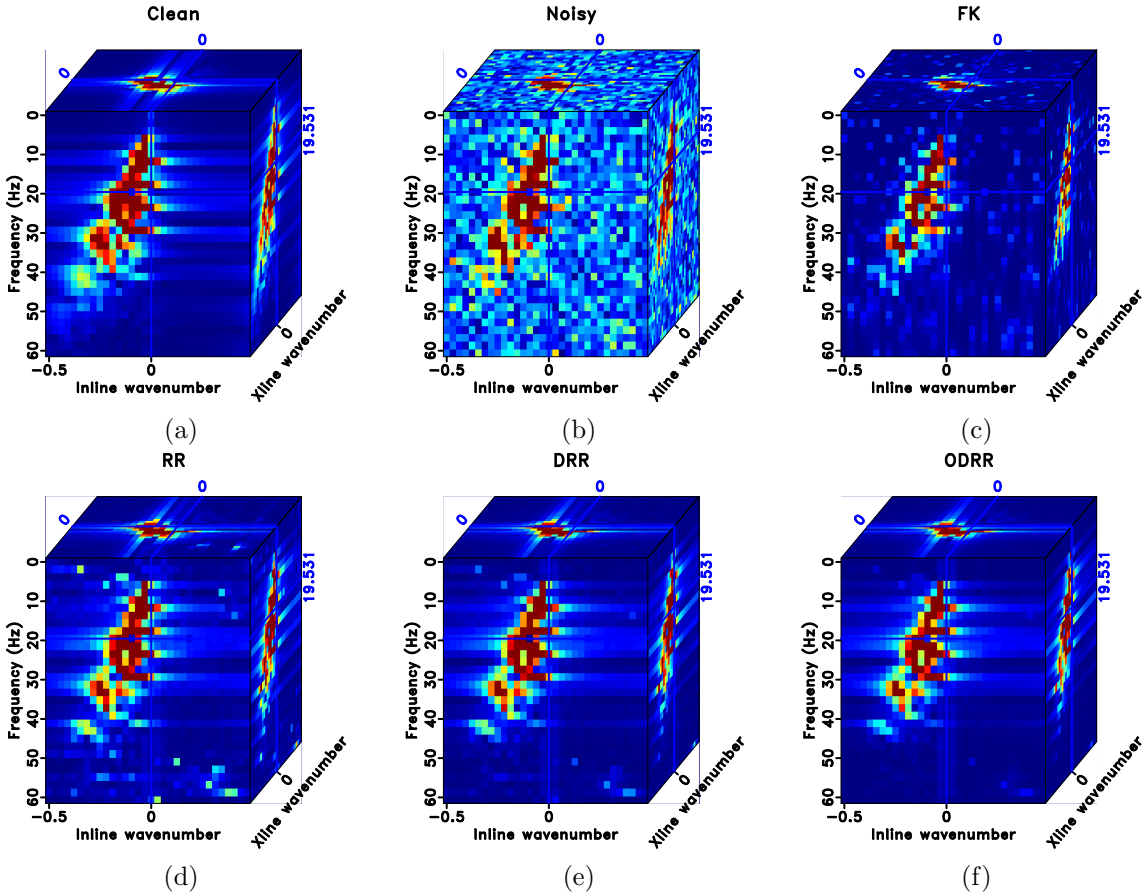


Figure 13: Spectrum comparison ($N = 10$). FK spectrum of (a) clean data, (b) noisy data, (c) FK method, (d) rank-reduction method, (e) damped rank-reduction method, and (f) the proposed method.

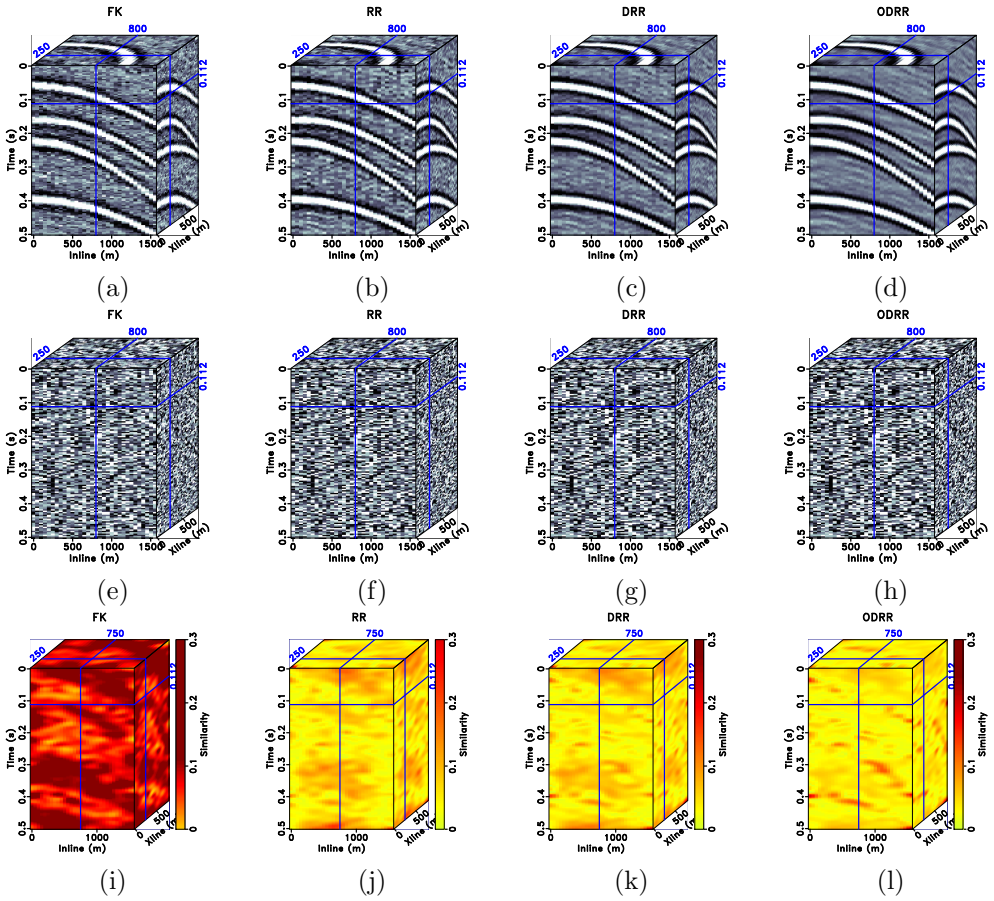


Figure 14: Denoising comparison ($N = 20$). The top row: denoised results using (a) FK method, (b) rank-reduction method, (c) damped rank-reduction method, and (d) the proposed method. The middle row: separated noise corresponding to the top row. The bottom row: local similarity corresponding to the top row.

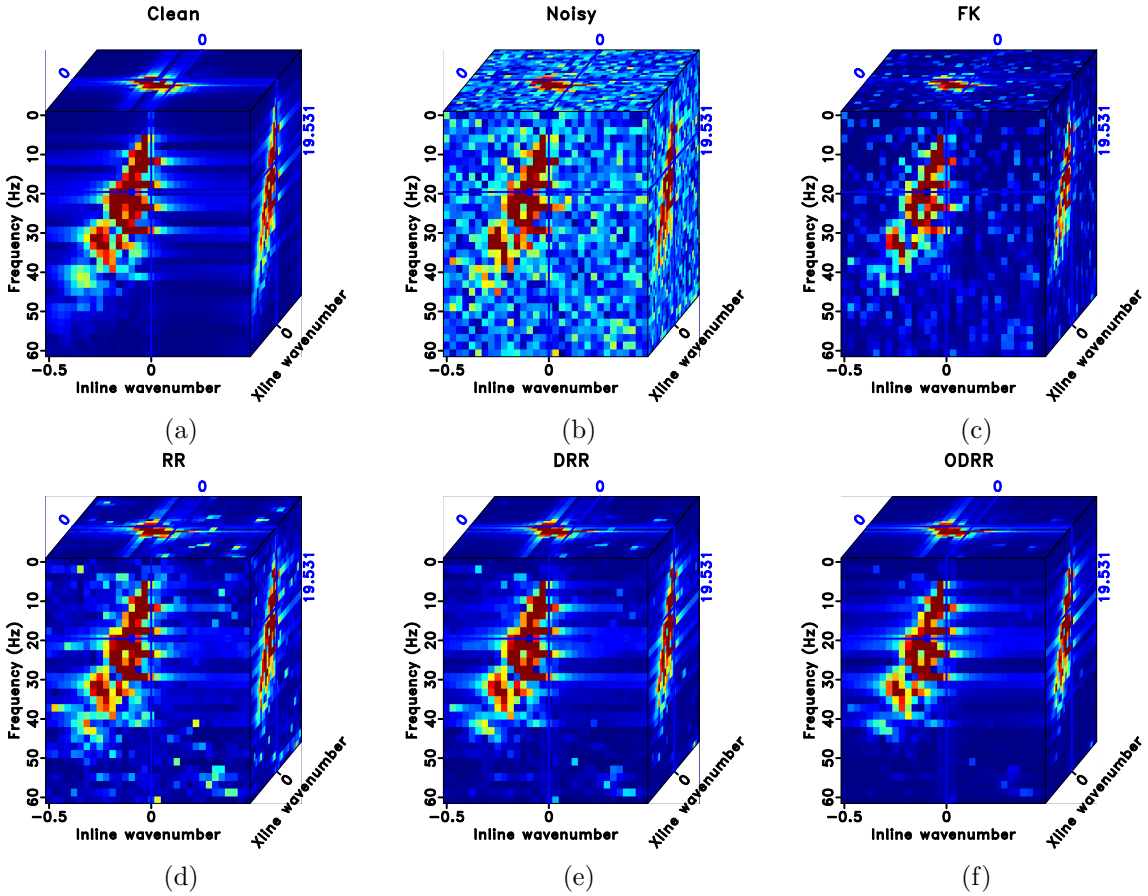


Figure 15: Spectrum comparison ($N = 20$). FK spectrum of (a) clean data, (b) noisy data, (c) FK method, (d) rank-reduction method, (e) damped rank-reduction method, and (f) the proposed method.

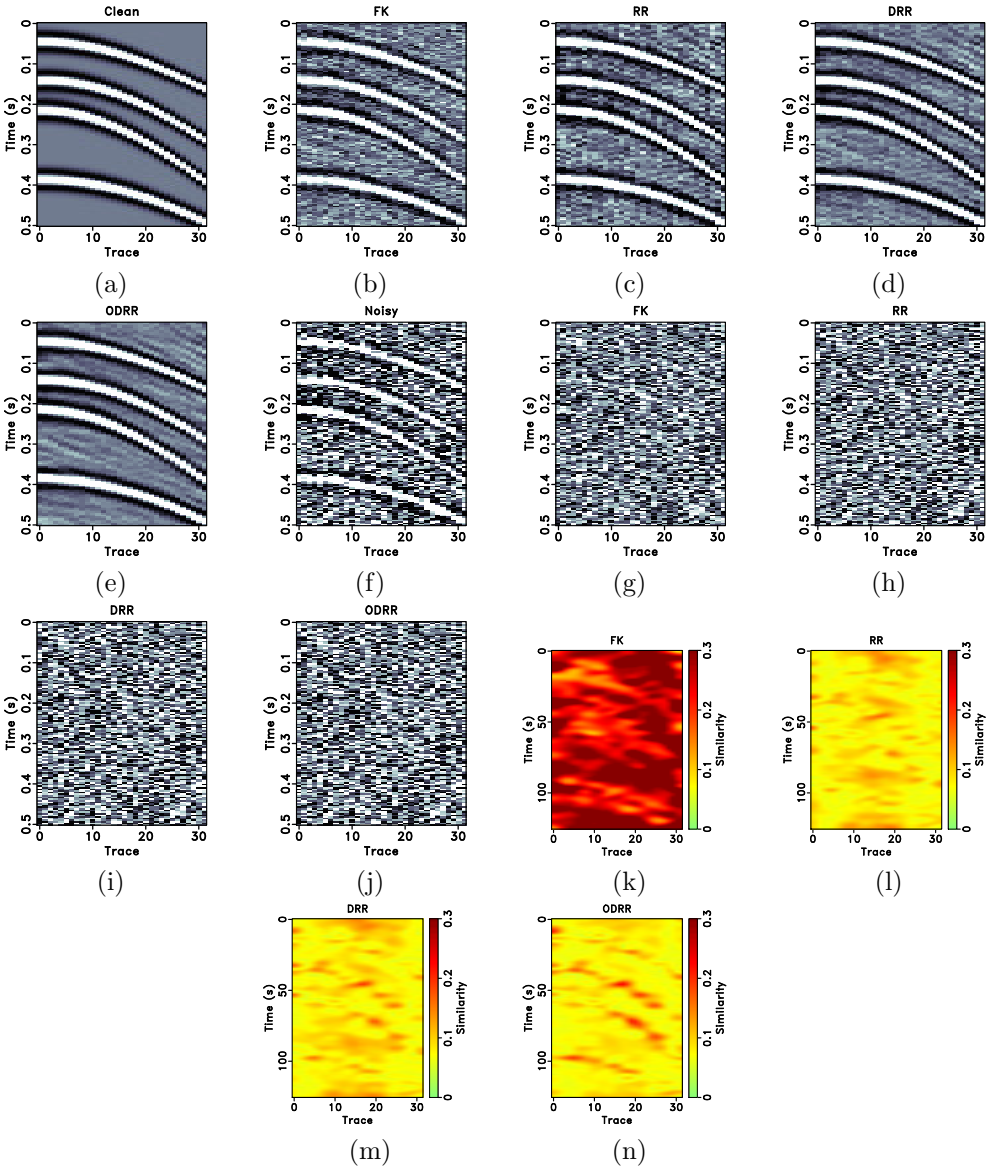


Figure 16: Denoising comparison ($N = 20$). (a) Clean data. (b)-(e): denoised results using FK method, (b) rank-reduction method, (c) damped rank-reduction method, and the proposed method, respectively. (f) Noisy data. (g)-(j) Separated noise corresponding to (b)-(e), respectively. (k)-(n) Local similarity corresponding to (b)-(e), respectively.

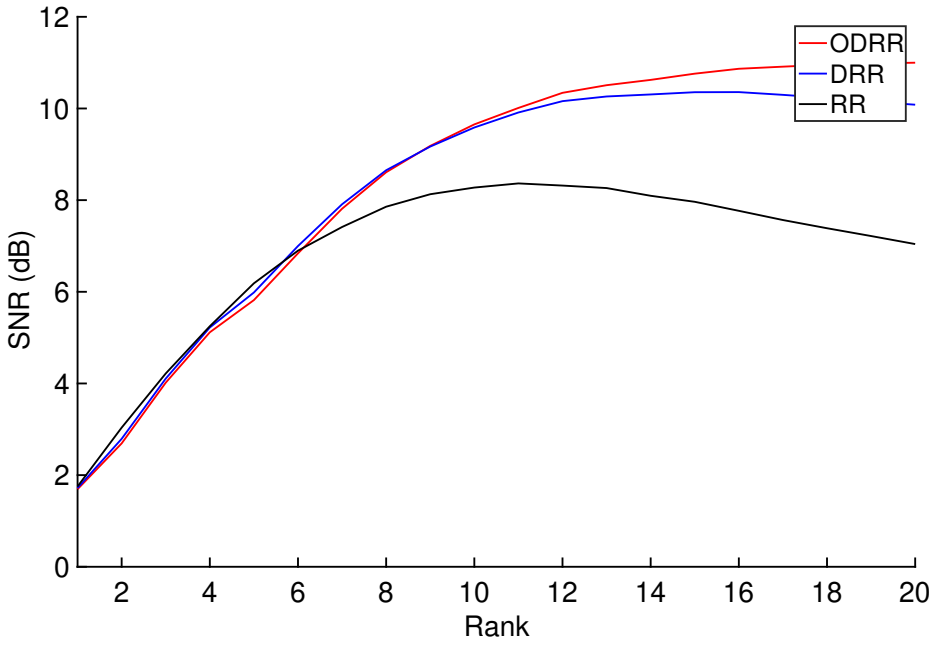


Figure 17: SNR diagrams of different lowrank approaches with respect to the selected rank parameters for the hyperbolic synthetic example.

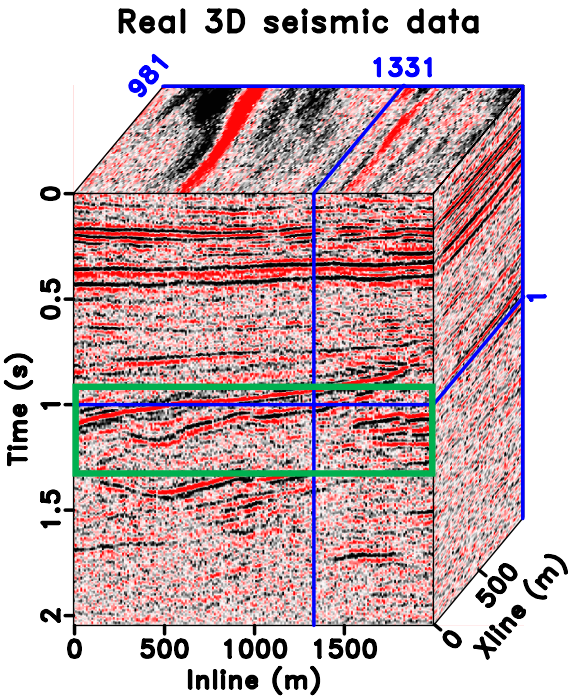


Figure 18: Real 3D seismic data. The green box highlights the zooming area for detailed comparison.

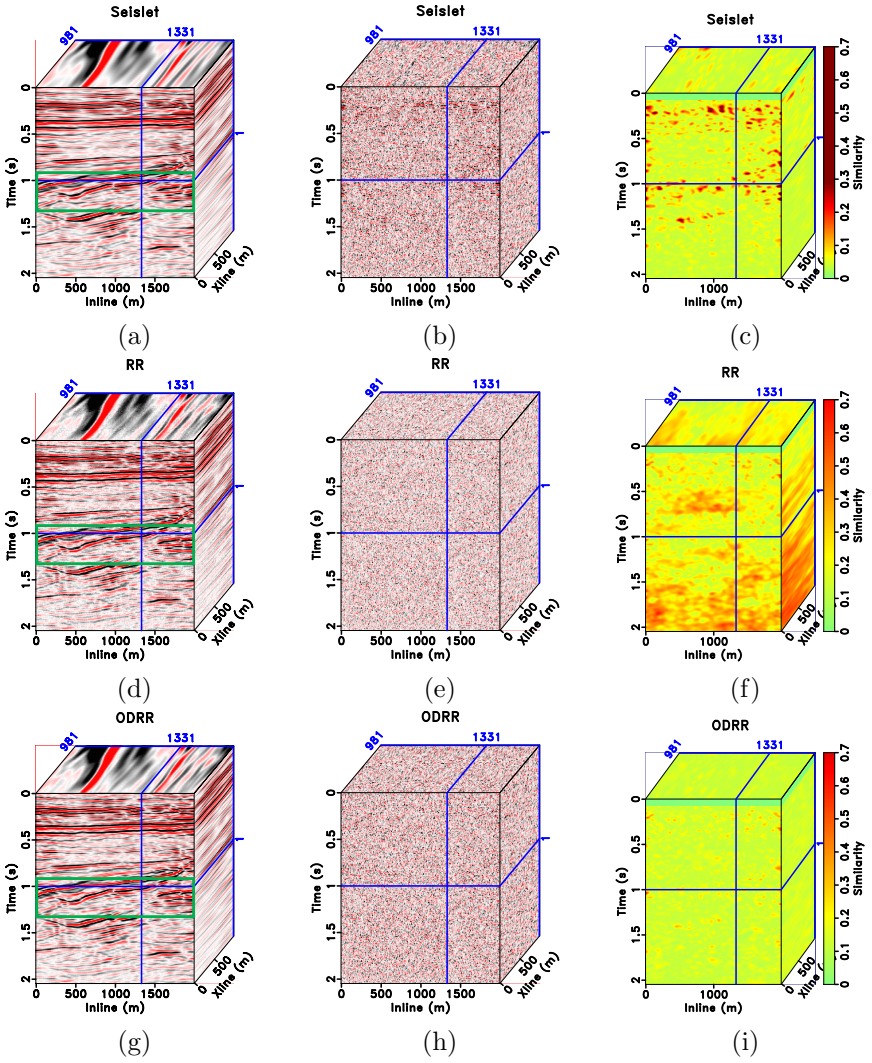


Figure 19: Denoising comparison. (a) Result using the seislet thresholding method. (b) Removed noise corresponding to (a). (c) Local similarity between (a) and (b). (d) Removed noise using the rank-reduction method. (e) Removed noise corresponding to (d). (f) Local similarity between (d) and (e). (g) Result using the proposed method. (h) Removed noise using the proposed method. (i) Local similarity between (g) and (h).

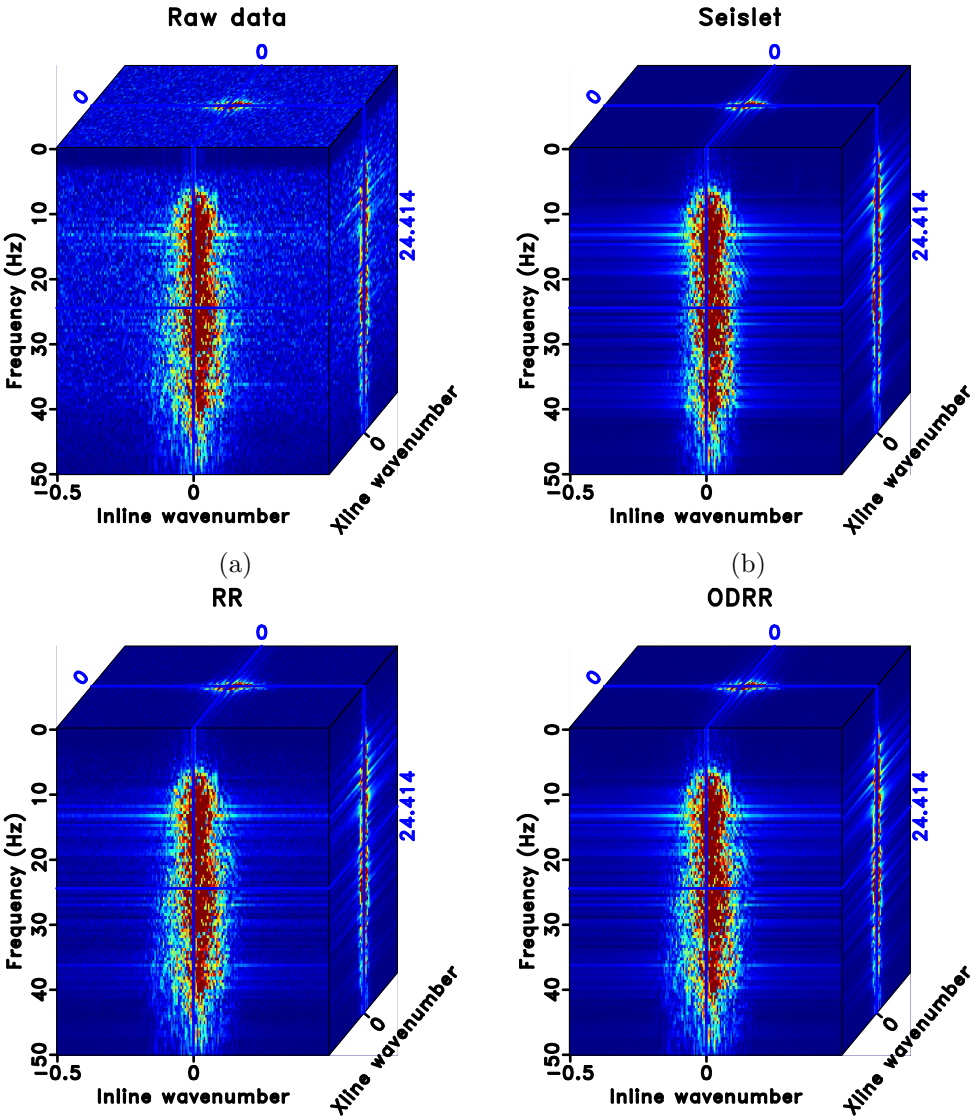


Figure 20: Spectrum comparison for the field data example. Spectrum of (a) real seismic data, (b) result using the seislet thresholding method, (c) result using the rank-reduction method, and (d) result using the proposed method.

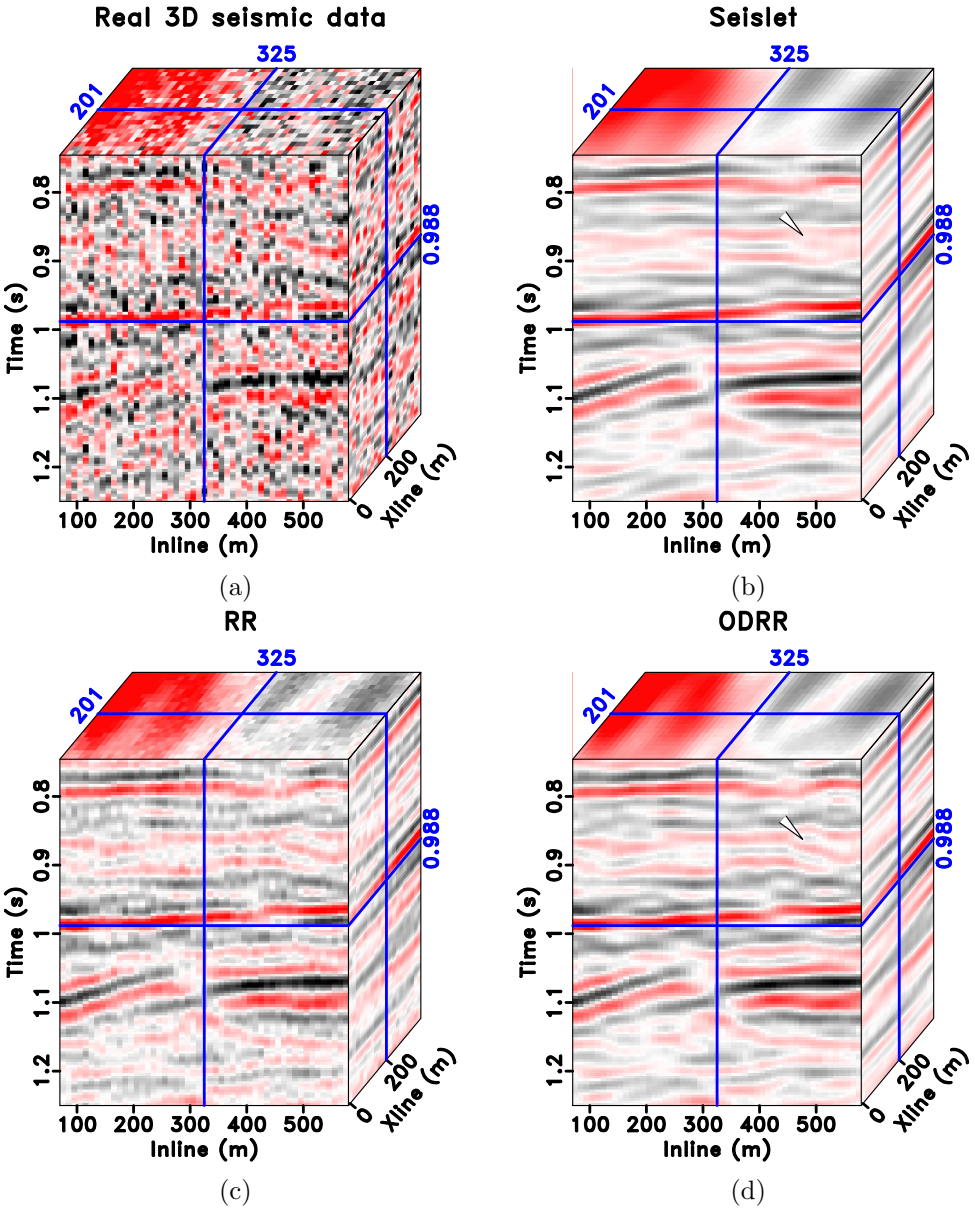


Figure 21: Zoomed denoising comparison for the field data example. (a) Real seismic data. (b) Result using the seislet thresholding method. (c) Result using the rank-reduction method. (d) Result using the proposed method. The arrows highlight the difference between (b) and (d).

REFERENCES

Abma, R., and J. Claerbout, 1995, Lateral prediction for noise attenuation by $t-x$ and $f-x$ techniques: Geophysics, **60**, no. 6, 1887–1896.

Abma, R., D. Howe, M. Foster, I. Ahmed, M. Tanis, Q. Zhang, A. Arogunmati, and G. Alexander, 2015, Independent simultaneous source acquisition and processing: Geophysics, **80**, no. 6, WD37–WD44.

Aharchaou, M., J. Anderson, S. Hughes, and J. Bingham, 2017, Singular-spectrum analysis via optimal shrinkage of singular values, *in* SEG Technical Program Expanded Abstracts 2017: Society of Exploration Geophysicists, 4950–4954.

Amani, S., A. Gholami, and A. J. Niestanak, 2017, Seismic random noise attenuation via 3D block matching: Journal of Applied Geophysics, **136**, 353–363.

Bekara, M., and M. van der Baan, 2007, Local singular value decomposition for signal enhancement of seismic data: Geophysics, **72**, no. 2, V59–V65.

Benaych-Georges, F., and R. R. Nadakuditi, 2012, The singular values and vectors of low rank perturbations of large rectangular random matrices: Journal of Multivariate Analysis, **111**, 120–135.

Beylkin, G., 1987, Discrete radon transform: IEEE transactions on acoustics, speech, and signal processing, **35**, no. 2, 162–172.

Bracewell, R. N., and R. N. Bracewell, 1986, The fourier transform and its applications: McGraw-Hill New York, **31999**.

Cadzow, J. A., 1988, Signal enhancement - a composite property mapping algorithm: IEEE Transactions on Acoustics, Speech, and Signal Processing, **1**, 49–62.

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

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: SIAM, Multiscale Modeling and Simulation, **5**, 861–899.

Chen, Y., M. Bai, and Y. Chen, 2019a, Obtaining free USArray data by multi-dimensional seismic reconstruction: Nature Communications, **10**, no. 1, 4434.

Chen, Y., M. Bai, Z. Guan, Q. Zhang, M. Zhang, and H. Wang, 2019b, Five-dimensional seismic data reconstruction using the optimally damped rank-reduction method: Geophysical Journal International, **218**, no. 1, 224–246.

Chen, Y., and S. Fomel, 2015, Random noise attenuation using local signal-and-noise orthogonalization: Geophysics, **80**, no. 6, WD1–WD9.

———, 2018, EMD-seislet transform: Geophysics, **83**, no. 1, A27–A32.

Chen, Y., D. Zhang, Z. Jin, X. Chen, S. Zu, W. Huang, and S. Gan, 2016, Simultaneous denoising and reconstruction of 5D seismic data via damped rank-reduction method: Geophysical Journal International, **206**, no. 3, 1695–1717.

Cheng, J., and M. Sacchi, 2015, Separation and reconstruction of simultaneous source data via iterative rank reduction: Geophysics, **80**, no. 4, V57–V66.

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

Colominas, M. A., G. Schlotthauer, M. E. Torres, and P. Flandrin, 2012, Noise-assisted EMD methods in action: Adv. Adapt. Data Anal., **4**, 1250025.

Eckart, C., and G. Young, 1936, The approximation of one matrix by another of lower

rank: Psychometrika, **1**, 211–218.
Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, no. 3, V25–V38.
Galbraith, M., 1991, Random noise attenuation by F-X prediction: A tutorial: 61st Annual International Meeting, SEG, Expanded Abstracts, 1428–1431.
Gao, J., J. Cheng, and M. D. Sacchi, 2017, Five-dimensional seismic reconstruction using parallel square matrix factorization: IEEE Transactions on Geoscience and Remote Sensing, **55**, 2124–2135.
Gavish, M., and D. L. Donoho, 2014, The optimal hard threshold for singular values is 4/sqrt(3): IEEE Transactions on Information Theory, **60**, 5040–5053.
Gilles, J., 2013, Empirical wavelet transform: IEEE Transaction on Signal Process, **61**, 3999–4010.
Ginolhac, G., P. Forster, F. Pascal, and J.-P. Ovarlez, 2013, Performance of two low-rank stap filters in a heterogeneous noise: IEEE Transactions on Signal Processing, **61**, 57–61.
Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: Proceeding of the Royal Society of London Series A, **454**, 903–995.
Huang, W., R. Wang, X. Chen, and Y. Chen, 2017, Double least squares projections method for signal estimation: IEEE Transactions on Geoscience and Remote Sensing, **55**, no. 7, 4111–4129.
Huang, W., R. Wang, Y. Chen, H. Li, and S. Gan, 2016, Damped multichannel singular spectrum analysis for 3D random noise attenuation: Geophysics, **81**, no. 4, V261–V270.
Huang, W., R.-S. Wu, and R. Wang, 2018, Damped dreamlet representation for exploration seismic data interpolation and denoising: IEEE Transactions on Geoscience and Remote Sensing, **56**, no. 6, 3159–3172.
Li, C., G. Liu, Z. Hao, S. Zu, F. Mi, and X. Chen, 2018, Multidimensional seismic data reconstruction using frequency-domain adaptive prediction-error filter: IEEE Transactions on Geoscience and Remote Sensing, **56**, no. 4, 2328–2336.
Li, S., B. Liu, Y. Ren, Y. Chen, S. Yang, Y. Wang, and P. Jiang, 2020, Deep learning inversion of seismic data: IEEE Transactions on Geoscience and Remote Sensing, **58**, no. 3, 2135–2149.
Liu, G., and X. Chen, 2013, Noncausal f-x-y regularized nonstationary prediction filtering for random noise attenuation on 3D seismic data: Journal of Applied Geophysics, **93**, 60–66.
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**, no. 2, V61–V69.
Mahdad, A., 2012, Deblending of seismic data: PhD thesis, Delft University of Technology.
Mousavi, S. M., C. A. Langston, and S. P. Horton, 2016, Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet trans-

form: Geophysics, **81**, no. 4, V341–V355.

Nadakuditi, R. R., 2013, Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage: IEEE Transactions on Information Theory, **60**, 3002–3018.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, V25–V32.

Qiao, T., J. Ren, Z. Wang, J. Zabalza, M. Sun, H. Zhao, S. Li, J. A. Benediktsson, and Q. Dai, 2016, Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis: IEEE Transactions on Geoscience and Remote Sensing, **99**, 10.1109/TGRS.2016.2598065.

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

Siahsar, M. A. N., S. Gholtashi, A. R. Kahoo, W. Chen, and Y. Chen, 2017b, Data-driven multi-task sparse dictionary learning for noise attenuation of 3D seismic data: Geophysics, **82**, no. 6, V385V396.

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

———, 2015, Preserving signal: Automatic rank determination for noise suppression, *in* SEG Technical Program Expanded Abstracts 2015: Society of Exploration Geophysicists, 4703–4707.

Trickett, S., and L. Burroughs, 2009, Prestack rank reducing noise suppression: Theory: SEG Technical Program Expanded Abstracts, 3332–3336.

Wang, C., Z. Zhu, and H. Gu, 2020, Low-rank seismic denoising with optimal rank selection for hankel matrices: Geophysical Prospecting, **68**, 892–909.

Wang, C., Z. Zhu, H. Gu, X. Wu, and S. Liu, 2018, Hankel low-rank approximation for seismic noise attenuation: IEEE Transactions on Geoscience and Remote Sensing, 1–13.

Wu, J., and M. Bai, 2018, Adaptive rank-reduction method for seismic data reconstruction: Journal of Geophysics and Engineering, **15**, 1688.

Wu, Z., and N. E. Huang, 2009, Ensemble empirical mode decomposition: A noise-assisted data analysis method: Advances in Adaptive Data Analysis, **1**, no. 1, 1–41.

Zhang, D., Y. Zhou, H. Chen, W. Chen, S. Zu, , and Y. Chen, 2017, Hybrid rank-sparsity constraint model for simultaneous reconstruction and denoising of 3D seismic data: Geophysics, **82**, no. 5, V351–V367.

Zhao, Q., Q. Du, X. Gong, and Y. Chen, 2018, Signal-preserving erratic noise attenuation via iterative robust sparsity-promoting filter: IEEE Transactions on Geoscience and Remote Sensing, **56**, no. 6, 1558–0644.

Zhou, Y., J. Gao, W. Chen, and P. Frossard, 2016, Seismic simultaneous source separation via patchwise sparse representation: IEEE Transactions on Geoscience and Remote Sensing, **54**, no. 9, 5271–5284.

Zhou, Y., and S. Zhang, 2017, Robust noise attenuation based on nuclear norm minimization and a trace prediction strategy: Journal of Applied Geophysics, **147**,

52–67.

Zu, S., H. Zhou, W. Mao, D. Zhang, C. Li, X. Pan, and Y. Chen, 2017, Iterative deblending of simultaneous-source data using a coherency-pass shaping operator: Geophysical Journal International, **211**, no. 1, 541–557.

Zu, S., H. Zhou, R. Ru, M. Jiang, and Y. Chen, 2019, Dictionary learning based on dip patch selection training for random noise attenuation: Geophysics, **84**, no. 3, V169–V183.

Table 1: Glossary of main mathematical notations used in this paper.

Symbols	Meanings
$\mathbf{D}(t/w, x, y)$	3D noisy seismic data
$\mathbf{D}(w_0)$ or $\mathbf{D}$	abbreviated notation of 2D frequency slice
$\mathbf{R}_i$	ith Hankel matrix
$\mathbf{M}$	block Hankel matrix
$\mathbf{S}$	signal component
$\hat{\mathbf{S}}$	estimated signal component
$\mathbf{U}$	left singular vector matrix
$\mathbf{\Sigma}$	singular-value matrix
$\mathbf{V}$	right singular vector matrix
$\hat{\mathbf{W}}$	weighting matrix
$\mathbf{P}$	damping matrix
$\mathbf{\Gamma}$	damping threshold matrix
$\mathbf{Q}$	temporary variable
$\mathcal{D}$	D-transform
$\mathcal{H}$	Hankelization operator
$\mathcal{T}$	thresholding operator
$\mathcal{A}$	averaging operator
N	rank
K	damping factor

Table 2: Comparison of SNRs in dB for different rank-reduction based methods. RR denotes the rank-reduction method. DRR denotes the damped rank-reduction method. ODRR denotes the optimally damped rank-reduction method.

Tests	Noisy (dB)	RR (dB)	DRR (dB)	ODRR (dB)
Linear synthetic (N=3)	-8.39	6.57	10.29	11.27
Linear synthetic (N=6)	-8.39	3.45	8.35	10.86
Hyperbolic synthetic (N=10)	-2.17	8.27	9.58	9.65
Hyperbolic synthetic (N=20)	-2.17	7.04	10.08	11.00

Table 3: Comparison of computational costs in seconds for different rank-reduction based methods.

Tests	FK (s)	RR (s)	DRR (s)	ODRR (s)
Linear synthetic (N=3)	0.13	2.27	2.45	2.43
Linear synthetic (N=6)	0.17	3.02	3.04	3.17
Hyperbolic synthetic (N=10)	0.43	1.69	1.71	1.83
Hyperbolic synthetic (N=20)	0.49	2.75	2.69	2.83