Simultaneous denoising and reconstruction of 5D seismic data via damped rank-reduction method[a]

[a]Published in Geophysical Journal International, 206, 1695-1717, (2016)

Yangkang Chen, Dong Zhang†, Zhaoyu Jin‡, Xiaohong Chen†, Shaohuan Zu†, Weilin Huang† , and Shuwei Gan†*

ABSTRACT

The Cadzow rank-reduction method can be effectively utilized in simultaneously denoising and reconstructing 5D seismic data that depends on four spatial dimensions. The classic version of Cadzow rank-reduction method arranges the 4D spatial data into a level-four block Hankel/Toeplitz matrix and then applies truncated singular value decomposition (TSVD) for rank-reduction. When the observed data is extremely noisy, which is often the feature of real seismic data, traditional TSVD cannot be adequate for attenuating the noise and reconstructing the signals. The reconstructed data tends to contain a significant amount of residual noise using the traditional TSVD method, which can be explained by the fact that the reconstructed data space is a mixture of both signal subspace and noise subspace. In order to better decompose the block Hankel matrix into signal and noise components, we introduced a damping operator into the traditional TSVD formula, which we call the damped rank-reduction method. The damped rank-reduction method can obtain a perfect reconstruction performance even when the observed data has extremely low signal-to-noise ratio (SNR). The feasibility of the improved 5D seismic data reconstruction method was validated via both 5D synthetic and field data examples. We presented comprehensive analysis of the data examples and obtained valuable experience and guidelines in better utilizing the proposed method in practice. Since the proposed method is convenient to implement and can achieve immediate improvement, we suggest its wide application in the industry.

INTRODUCTION

Field obstacles and economic costs in seismic acquisition often result in irregular and incomplete seismic field data (Chiu, 2014). Seismic interpolation is such a processing step to provide the regularly sampled seismic data for the following workflows like high-resolution processing, wave-equation migration, multiple suppression, amplitude-versus-offset (AVO) or amplitude-versus-azimuth (AVAZ) analysis, and time-lapse studies, thus plays a fundamental role in seismic data processing (Por-

sani, 1999; Abma and Kabir, 2005, 2006; Wang et al., 2010a; Li et al., 2012; Chen et al., 2015; Gan et al., 2016, 2015c).

During the past several decades of continuous research from both academia and industry, there have been many effective methods for seismic data interpolation. In prediction based approaches, a prediction error filter is designed such that the predicted data and the existing data has the minimum misfit by solving a least-squares linear inverse problem (Spitz, 1991; Fomel, 2002). Considering that the irregularly sampled seismic data with large number of missing traces usually suffers from the aliasing problem, the prediction error filter is obtained by solving the inverse problem from low-frequency data that is less aliasing and then applied to high-frequency components (Naghizadeh and Sacchi, 2007). The sparse transform based interpolation assumes that the seismic data is sparse in a certain transformed domain. By iteratively thresholding the transformed domain of the incomplete seismic data, one can gradually attenuate the artifacts in the sparse domain that are caused by the missing traces in the time-space domain, and then gradually recover the missing information (Chen et al., 2014a; Gan et al., 2015b,c; Wang et al., 2015b,a). An important component in this type of methods is the sparse transform. Generally speaking, sparse transforms can be divided into two categories (Chen et al., 2016): analytical transforms, which have fixed basis, and learning-based dictionaries, which iteratively update the basis. A number of fixed basis sparse transforms have been proposed in the literature for restoring seismic data, which include Fourier transform (Sacchi et al., 1998; Duijndam et al., 1999; Chen et al., 2014a), Radon transform (Trad et al., 2002; Yu et al., 2007; Wang et al., 2010b), curvelet transform (Herrmann and Hennenfent, 2008; Shahidi et al., 2013; Liu et al., 2015, 2016) and seislet transform (Fomel and Liu, 2010; Chen et al., 2014b; Gan et al., 2015c). The learning-based approach usually utilizes machine learning techniques to construct the dictionary. Another group of methods is based on wave equation. The wave equation theory inherently connects the seismic trace recorded at a specific location from a known shot position to all the other traces assuming the subsurface velocity model is known. Thus, it depends on a prior information that is usually difficult to obtain and it is also limited by the large computational cost aroused from solving the wave equation (Ronen, 1987; Canning and Gardner, 1996; Fomel, 2003).

Recently, rank-reduction based seismic interpolation algorithms become popular (Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011; Gao et al., 2015b). The rank-reduction methods for seismic data reconstruction can be divided into two main categories. The first category applies rank-reduction to multilevel block Hankel/-Toeplitz matrices formed from the entries of the tensor. In other words, multidimensional seismic data is rearranged into a block Hankel or Toeplitz matrix, and a rank-reduction algorithm is used to improve the signal-to-noise ratio (SNR) and to reconstruct the data. Such algorithm is usually named as the Cadzow method (Trickett and Burroughs, 2009) or multichannel singular spectrum analysis (MSSA) method (Oropeza and Sacchi, 2011; Huang et al., 2015, 2016). The other category of rank-reduction methods encompasses techniques that are based on dimensionality reduction of multi-linear arrays or tensors. In this case, the multichannel seismic

data is viewed as a multi-linear array, and dimensionality reduction techniques are directly applied to the multi-linear array (Gao et al., 2015b). Simply speaking, rank-reduction methods are based on the fact that missing traces and random noise will increase the rank of a matrix or tensor. By applying a rank-reduction step, such as truncated singular value decomposition (TSVD), we can intuitively reduce the negative effects caused by the missing traces and random noise. It is worth noting that rank-reduction or low-rank matrix recovery is closely related to sparse transform and compressed sensing (CS) theory because of their utilization of low-dimensional signal structures (Recht et al., 2010; Candès and Tao, 2010; Plan, 2011). Specifically, Recht et al. (2010) discussed the bridge between CS and low-rank matrix recovery. Considering that the multi-dimensional seismic data usually contains significant amount of random noise. A lof of efforts have been put in the area of simultaneous interpolation and denoising of seismic data. Oropeza and Sacchi (2011) introduced a POCS-like iteration into the MSSA algorithm with a variable "alpha" to weight the addition of original data with estimated signal. While this method is effective in attenuating noise while interpolating seismic records, the mathematical relationship between "alpha" and the SNR was not properly understood. Following Oropeza and Sacchi (2011), the weighting strategy was also adopted by Kreimer and Sacchi (2012). Recently the connection between parameter "alpha" and a trade-off parameter "mu" was made in Gao et al. (2015b). Sternfels et al. (2015) solved the simultaneous interpolation and denoising problem via convex optimization, where the problem was formulated as a regularized inverse problem with a sparsity promotion term and solved using the alternating direction method of multipliers (ADMM) algorithm.

In recent years, 5D seismic interpolation has become popular since it takes all physical dimensions of seismic data into consideration for the interpolation, which brings us a stronger constraint when inverting a highly underdetermined inverse problem and thus helps us obtain a much improved reconstruction performance in the case of high missing-traces ratio compared with 2D and 3D interpolation (Liu and Sacchi, 2004; Kreimer et al., 2013; Chiu, 2014; Gao et al., 2015b). The rank-reduction based seismic interpolation has been shown to be an effective method for 5D seismic data interpolation (Gao et al., 2015b). Our recent study shows that when the observed data is extremely noisy, which is often the feature of real seismic data, the reconstructed data using the rank-reduction based algorithm still contains significant residual noise. The residual noise is caused from the fact that the rank-reduced signal space via TSVD is a mixture of signal and noise subspace. We propose an effective method for improving the traditional Cadzow rank-reduction method by modifying the TSVD formula. In order to better decompose the noisy data space into signal and noise subspace, we propose to apply a damping operator to the block Hankel matrix after TSVD. The damping operator is controlled by adjusting an introduced parameter called damping factor. In the special case, when the damping factor is sufficiently large, the improved rank-reduction method reverts the traditional Cadzow rank-reduction method. The proposed method is compared with the traditional method via both synthetic and field 5D seismic dataset and is demonstrated to obtain much better performance.

THEORY

Block Hankel matrix formulation for 5D seismic data

Consider a block of 5D data $\mathbf{D}_{time}(t, hx, hy, x, y)(t = 1 \cdots N_t)$, where hx, hy, x, y denote x and y offsets and x and y midpoints. The rank-reduction method operates on the 5D seismic data in the following way: first, rank-reduction method transforms $\mathbf{D}_{time}(t, hx, hy, x, y)$ into $\mathbf{D}_{freq}(w, hx, hy, x, y)(w = 1 \cdots N_w)$ of complex values in the frequency domain. In the condition that the spatial variables are regularly sampled, each observation at a given frequency slice can be represented via a 4D spatial hypercube $\mathbf{D}_{k1,k2,k3,k4}$ with $k_i = 1 \cdots X_i$, $i = 1, 2, 3, 4$. In order to make our target matrices close to square matrices, parameters Y_i are defined as $\lfloor \frac{X_i}{2} \rfloor + 1$, $i = 1, 2, 3, 4$. Gao et al. (2015a) outlined the usefulness of square Hankel matrices in rank reduction based regularization and described why square matrices are preferred. The symbol $\lfloor \cdot \rfloor$ denotes the integer part of its argument. The 4D spatial hypercube $\mathbf{D}_{k1,k2,k3,k4}$ can be embedded in a level-four block Hankel matrix as follows. We have to address in advance that each frequency slice of data is operated the exactly same way. We first embed the seismic data in a level-one Hankel matrix using all data components in the first dimension of the tensor $\mathbf{D}_{k1,k2,k3,k4}$ at a given frequency w_0. This generates the following Hankel matrices,

$$\mathbf{M}^{(1)}_{k2,k3,k4} = \begin{pmatrix} D_{1,k2,k3,k4} & D_{2,k2,k3,k4} & \cdots & D_{X_1-Y_1+1,k2,k3,k4} \\ D_{2,k2,k3,k4} & D_{3,k2,k3,k4} & \cdots & D_{X_1-Y_1+2,k2,k3,k4} \\ \vdots & \vdots & \ddots & \vdots \\ D_{Y_1,k2,k3,k4} & D_{Y_1+1,k2,k3,k4} & \cdots & D_{X_1,k2,k3,k4} \end{pmatrix}. \tag{1}$$

Matrices in Equation 1 are of size $Y_1 \times (X_1 - Y_1 + 1)$. These matrices are embedded in a level-two block Hankel matrix:

$$\mathbf{M}^{(2)}_{k3,k4} = \begin{pmatrix} M^{(1)}_{1,k3,k4} & M^{(1)}_{2,k3,k4} & \cdots & M^{(1)}_{X_2-Y_2+1,k3,k4} \\ M^{(1)}_{2,k3,k4} & M^{(1)}_{3,k3,k4} & \cdots & M^{(1)}_{X_2-Y_2+2,k3,k4} \\ \vdots & \vdots & \ddots & \vdots \\ M^{(1)}_{Y_2,k3,k4} & M^{(1)}_{Y_2+1,k3,k4} & \cdots & M^{(1)}_{X_2,k3,k4} \end{pmatrix}. \tag{2}$$

Matrices in Equation 2 are of size $(Y_1 Y_2) \times (X_1 - Y_1 + 1)(X_2 - Y_2 + 1)$. These matrices are now embedded in a level-three block Hankel matrix:

$$\mathbf{M}^{(3)}_{k4} = \begin{pmatrix} M^{(2)}_{1,k4} & M^{(2)}_{2,k4} & \cdots & M^{(2)}_{X_3-Y_3+1,k4} \\ M^{(2)}_{2,k4} & M^{(2)}_{3,k4} & \cdots & M^{(2)}_{X_3-Y_3+2,k4} \\ \vdots & \vdots & \ddots & \vdots \\ M^{(2)}_{Y_3,k4} & M^{(2)}_{Y_3+1,k4} & \cdots & M^{(2)}_{X_3,k4} \end{pmatrix}. \tag{3}$$

Matrices in Equation 3 are of size $(Y_1 Y_2 Y_3) \times (X_1 - Y_1 + 1)(X_2 - Y_2 + 1)(X_3 - Y_3 + 1)$.

At last, these matrices are embedded in our final target level-four block Hankel matrix:

$$\mathbf{M}^{(4)} = \begin{pmatrix} M_1^{(3)} & M_2^{(3)} & \cdots & M_{X_4-Y_4+1}^{(3)} \\ M_2^{(3)} & M_3^{(3)} & \cdots & M_{X_4-Y_4+2}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ M_{Y_4}^{(3)} & M_{Y_4+1}^{(3)} & \cdots & M_{X_4}^{(3)} \end{pmatrix}. \tag{4}$$

Equation 4 is the target matrix of our problem. The size of level-four block Hankel matrix $\mathbf{M}^{(4)}$ is $(Y_1Y_2Y_3Y_4) \times (X_1 - Y_1 + 1)(X_2 - Y_2 + 1)(X_3 - Y_3 + 1)(X_4 - Y_4 + 1)$. If X_i is an odd integer, the size of $\mathbf{M}^{(4)}$ is $(Y_1Y_2Y_3Y_4) \times (Y_1Y_2Y_3Y_4)$.

Rank-reduction method for 5D seismic data restoration

The transformation of the 4D data hypercube at a single frequency into a level-four block Hankel matrix can be represented in operator notation as follows:

$$\mathbf{M}^{(4)} = \mathcal{H}\mathbf{D}, \tag{5}$$

where $\mathcal{H}$ denotes the Hankelization operator. To avoid notational clutter, we omit the argument $k1, k2, k3, k4$ of $\mathbf{D}_{k1,k2,k3,k4}$ for convenience. In general, the level-four block Hankel matrix $\mathbf{M}^{(4)}$ can be represented as:

$$\mathbf{M}^{(4)} = \mathbf{S} + \mathbf{N}, \tag{6}$$

where $\mathbf{S}$ and $\mathbf{N}$ denote the level-four block Hankel matrices of signal and random noise, respectively. Here, missing signal can be regarded as a condition where signal and noise are summed to zero, and is distributed randomly due to the irregular sampling. Therefore, missing data represents itself by the noise-like characteristics.

We assume that $\mathbf{M}^{(4)}$ and $\mathbf{N}$ have full rank. The size of $\mathbf{M}^{(4)}$ is $I \times J, \quad J < I$. $rank(\mathbf{M}^{(4)})$=$rank(\mathbf{N}) = J$, $I = Y_1Y_2Y_3Y_4$, $J = (X_1 - Y_1 + 1)(X_2 - Y_2 + 1)(X_3 - Y_3 + 1)(X_4 - Y_4 + 1)$, and $\mathbf{S}$ has deficient rank, $rank(\mathbf{S}) = K < J$. The singular value decomposition (SVD) of $\mathbf{M}^{(4)}$ can be represented as:

$$\mathbf{M}^{(4)} = [\mathbf{U}_1^{M^{(4)}} \quad \mathbf{U}_2^{M^{(4)}}] \begin{bmatrix} \Sigma_1^{M^{(4)}} & \mathbf{0} \\ \mathbf{0} & \Sigma_2^{M^{(4)}} \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^{M^{(4)}})^H \\ (\mathbf{V}_2^{M^{(4)}})^H \end{bmatrix}, \tag{7}$$

where $\Sigma_1^{M^{(4)}}$ $(K \times K)$ and $\Sigma_2^{M^{(4)}}$ $((I-K) \times (J-K))$ are diagonal matrices and contain, respectively, larger singular values and smaller singular values. $\mathbf{U}_1^{M^{(4)}}$ $(I \times K)$, $\mathbf{U}_2^{M^{(4)}}$ $(I \times (I-K))$, $\mathbf{V}_1^{M^{(4)}}$ $(J \times K)$ and $\mathbf{V}_2^{M^{(4)}}$ $(J \times (J-K))$ 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 substitute $\Sigma_2^{M^{(4)}}$ with $\mathbf{0}$ to achieve the goal of attenuating random noise while recovering the missing data during the first iteration in reconstruction process as follows:

$$\tilde{\mathbf{M}}^{(4)} = \mathbf{U}_1^{M^{(4)}} \Sigma_1^{M^{(4)}} (\mathbf{V}_1^{M^{(4)}})^H. \tag{8}$$

Equation 8 is referred to as the TSVD, which is used in the conventional rank-reduction approach for 5D seismic data interpolation.

Nevertheless, $\tilde{\mathbf{M}}^{(4)}$ is actually still contaminated with residual random noise. Following Huang et al. (2016), here we briefly clarify the reason why the $\tilde{\mathbf{M}}^{(4)}$ is still corrupted with some residual random noise. We first consider the SVD of the signal component $\mathbf{S}$ in a way similar to equation 7:

$$\mathbf{S} = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \Sigma_1^S & \mathbf{0} \\ \mathbf{0} & \Sigma_2^S \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^S)^H \\ (\mathbf{V}_2^S)^H \end{bmatrix}. \tag{9}$$

Different from $\tilde{\mathbf{M}}^{(4)}$, the signal matrix should be of deficient rank, then Σ_2^S is a zero diagonal matrix, and $\mathbf{S}$ can be denoted in a concise form:

$$\mathbf{S} = \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H. \tag{10}$$

Combining equations 6, 9, and 10, it straightforwardly derive the following equation:

$$\tilde{\mathbf{M}}^{(4)} = \mathbf{S} + \mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N}. \tag{11}$$

The appendix A gives a detailed derivation for obtaining equation 11, which we call exact formulation of the block Hankel matrix. From equation 11, we can intuitively know that TSVD result $\tilde{\mathbf{M}}^{(4)}$ still contains significant random noise $\mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N}$, or in other words, the traditional TSVD can only decompose the data space into a mixture of signal and noise subspace.

Damped rank-reduction method for 5D seismic data restoration

Huang et al. (2016) introduced a damping factor into TSVD to effectively attenuate the residual random noise for 3D dataset. Here, we further extend this damped algorithm to rank-reduction method for 5D seismic data interpolation in the presence of random noise. Since the missing data at each frequency slice can be regarded as the random noise, the derivation is similar to Huang et al. (2016) except for the extra reconstruction process based on the weighted projection onto convex set (POCS) like method in a level-four block Hankel matrix. Appendix B gives a detailed introduction on how to derive a damped TSVD in order to reduce the residual noise existing in the rank-reduced matrix $\tilde{\mathbf{M}}^{(4)}$. Here, we just conclude the approximation of $\mathbf{S}$ as follows:

$$\mathbf{S} = \mathbf{U}_1^{M^{(4)}} \Sigma_1^{M^{(4)}} \mathbf{T} (\mathbf{V}_1^{M^{(4)}})^H, \tag{12}$$

$$\mathbf{T} = \mathbf{I} - (\Sigma_1^{M^{(4)}})^{-N} \hat{\delta}^N, \tag{13}$$

where $\hat{\delta}$ denotes the maximum element of $\Sigma_2^{M^{(4)}}$ and N denotes the damping factor. It is worth mentioning that the greater the N, the weaker the damping is, and equation 12 degrades to equation 8 when $N \to +\infty$.

The damped TSVD of rank-reduction algorithm for 5D seismic data interpolation can be represented in operator notation as follows:

$$\mathbf{S} = \mathcal{R}_d \mathbf{M}^{(4)}. \tag{14}$$

where we use $\mathcal{R}_d$ as the rank-reduction operator using the damped TSVD (equations 12 and 13) while $\mathcal{R}$ is used for conventional TSVD. According to our experience, the damped TSVD is powerful for attenuating random noise while interpolating missing data. However, it is still SVD-based algorithm that takes the largest portion of computational time. The randomized singular value decomposition (RSVD) strategy (Rokhlin et al., 2009; Halko et al., 2011; Oropeza and Sacchi, 2011) can be used to accelerate the proposed method for better efficiency. Specifically according to our numerical tests, RSVD can save at least 45% computational cost compared with the classic TSVD, and cost saving is different with data dimensions, the bigger the dimension is, the faster the RSVD is. Thus RSVD can increase the efficiency significantly.

The filtered data is recovered with random noise attenuated and missing traces reconstructed by properly averaging along the anti-diagonals of the low-rank matrix $\mathbf{S}$ obtained via the damped TSVD:

$$\hat{\mathbf{D}} = \mathcal{A}\mathbf{S} = \mathcal{A}\mathcal{R}_d\mathbf{M}^{(4)} = \mathcal{A}\mathcal{R}_d\mathcal{H}\mathbf{D} = \mathcal{F}_d\mathbf{D}. \tag{15}$$

$\mathcal{A}$ denotes the averaging operator. The operator $\mathcal{F}_d$ represents improved rank-reduction filter. If $\mathcal{R}_d$ is substituted by the traditional TSVD operator $\mathcal{R}$, $\mathcal{F}$ can represent conventional rank-reduction filter similarly.

In this paper, we pay our attention to the reconstruction of noisy data. Our final goal is to recover the useful signal from the observed noisy and incomplete data. It is worth mentioning that this type of problem often arises in the field of collaborative filtering (**CF**). Similarly to those arising in the field of **CF**, seismic data reconstruction can be interpreted as a matrix completion problem. Missing traces in seismic data will introduce missing value in the observed data $\mathbf{D}_{k1,k2,k3,k4}$. The corresponding level-four block Hankel matrix $\mathbf{M}^{(4)}$ is expected to have rank K. Random noise and missing data will increase the rank of $\mathbf{M}^{(4)}$. Consequently, rank-reduction via the damped TSVD with the damping factor can be implemented to recover the useful signal and to reconstruct the missing records for 5D seismic volumes.

The modified weighted POCS-like algorithm based on conventional rank-reduction method for 5D seismic data interpolation can be shown as $\mathbf{D}_n = a_n\mathbf{D}_{obs} + (1 - a_n\mathcal{S}) \circ \mathcal{F}\mathbf{D}_{n-1}, \quad n = 1, 2, 3, \cdots, n_{max}$. Revised by the improved rank-reduction filter $\mathcal{F}_d$, therefore our proposed improved rank-reduction algorithm for simultaneous 5D seismic data reconstruction and denoising can be formulated as follows:

$$\mathbf{D}_n = a_n\mathbf{D}_{obs} + (1 - a_n\mathcal{S}) \circ \mathcal{F}_d\mathbf{D}_{n-1}, \qquad n = 1, 2, 3, \cdots, n_{max} \tag{16}$$

where $\mathbf{D}_0 = \mathbf{D}_{obs}$. a_n is an iteration-dependent scalar that linearly decreases from $a_1 = 1$ to $a_{n_{max}} = 0$. $\mathcal{S}$ denotes the sampling factor. $\mathcal{S}$ equals 1 at the observation point while 0 at the missing traces. Symbol $\circ$ denotes the Hadamard product (entry-wise product) of two matrices with the same size.

A complete and detailed algorithm workflow of the proposed damped rank-reduction algorithm for 5D seismic data reconstruction and denoising is provided as follows:

IMPROVED 5D SEISMIC DATA RANK-REDUCTION ALGORITHM$(\mathcal{S}, \mathcal{F}_d, \mathbf{D}_{obs}, a_n, \epsilon, n_{max}, F)$

$\mathbf{D}_{obs}(f, hx, hy, x, y) \leftarrow \mathbf{D}_{obs}(t, hx, hy, x, y)$ by 1D forward FFT
$\mathbf{D}_0 \leftarrow \mathbf{D}_{obs}$
for $f \leftarrow 1, 2, \ldots, F$
    do
        for $n \leftarrow 1, 2, \ldots, n_{max}$
            do
                $\mathbf{D}_n^f \leftarrow a_n \mathbf{D}_{obs}^f + (1 - a_n)\mathcal{S}\mathcal{F}_d \mathbf{D}_{n-1}^f + (1 - \mathcal{S})\mathcal{F}_d \mathbf{D}_{n-1}^f$
                if $\|\mathbf{D}_n^f - \mathbf{D}_{n-1}^f\|_F^2 \leq \epsilon$
                  then return $\mathbf{D}_n^f$
        return $\mathbf{D}_N^f$
return $\mathbf{D}_{recoverd}$
$\mathbf{D}_{recoverd}(t, hx, hy, x, y) \leftarrow \mathbf{D}_{recoverd}(f, hx, hy, x, y)$ by 1D inverse FFT

The iteration terminates after all F frequencies are finished. For each frequency slice, the algorithm stops when either a maximum number of iterations n_{max} is reached or $\|\mathbf{D}_n - \mathbf{D}_{n-1}\|_F^2 \leq \epsilon$, where $\|\cdot\|_F$ denotes the Frobenius norm and ϵ denotes a small tolerance value.

EXAMPLES

We apply the proposed improved rank-reduction method on a 5D synthetic seismic dataset, as shown in Figures 1 and 3. Figure 1 shows the performance in one common offset gather ($hx = 5, hy = 5$) and Figure 3 shows the performance in one common midpoint gather ($x = 5, y = 5$). Figures 1a and 3a show the true data. Figures 1b and 3b show the noisy data by adding some bandlimited noise to the clean data. The noise is so strong that the useful signals can be hardly detected. Figures 1c and 3c show the observed incomplete and noisy data. 70% of total traces are missing in this example. The signal-to-noise ratio (SNR) of the observed data is exetremely low and we can hardly see any coherent useful events in the observed data. Figures 1d and 3d show the reconstructed data using traditional method, which is quite acceptable since almost all useful signals are recovered and the noise has been greatly suppressed. However, there is still some residual noise in the reconstructed data, which can be explained by the non-optimal decomposition via traditional TSVD as analyzed in the theoretic section. Figures 1e and 3e show the reconstructed data using the proposed method. In this example, the rank $K = 3$ and the damping factor $N = 3$, and we use 10 weighted POCS-like iterations. It is obvious that the proposed method obtains a much cleaner result, and to some extent, the recovery of the useful signals are nearly

perfect. Figure 2 shows two reconstruction error cubes in common offset domain (the difference between the true data and the recovered data) using the traditional and the proposed approaches, respectively. Similarly, Figure 4 shows the reconstruction error cubes in common midpoint domain. Both Figures 2 and 4 clearly show that the proposed approach can obtain the least reconstruction error, while the error in the traditional approach is mainly due to the residual noise. In order to compare the amplitude between different datasets in detail, we conduct a single trace amplitude comparison for each data cubes shown in Figure 3. The single trace is chosen as the middle trace in each cubes of Figure 3. The comparison is presented in Figure 5. The black line is from the clean data. The red line is from the noisy data. The observed data in this case is zero and thus cannot be seen from the figure. The blue line corresponds to the traditional rank-reduction method. The green line corresponds to the proposed method. It is apparent that the green line is very close to the black line while the blue line deviates from the black line too much in most areas. This trace amplitude comparison further confirms the superior performance of the proposed algorithm.

In order to show how the rank-reduced block Hankel matrix looks like and to more intuitively demonstrate the influence of the introduced damping operator to the target level-four block Hankel matrix in the first 5D seismic data example, we plot the 26 Hz block Hankel matrix for different cases in Figures 6, 7, and 8. Figure 6a shows the 26 Hz block Hankel matrix for the clean data. Figure 6b shows the 26 Hz block Hankel matrix for the observed data, which is very different from that of the true data due to the 70% missing traces. Figure 6c shows rank-reduced block Hankel matrix via the traditional TSVD after 10 iterations. The rank-reduced block Hankel matrix via the damped TSVD is shown in Figure 6d. Figures 6c and 6d are both close to the true block Hankel matrix and are very similar to each other. If we look carefully or zoom in the figure a bit, we can see that the block Hankel matrix via the traditional TSVD is much rougher than the Hankel matrix via the damped TSVD. Considering that each subfigure in Figure 6 is a 1296×625 matrix, which contains a lot of information, we zoom one block Hankel matrix, as highlighted by the black frame boxes in Figure 6 and make a more detailed comparison in Figure 7. Figure 7 clearly shows that the rank-reduced block Hankel matrix via traditional TSVD still contains some artifacts, especially on the right bottom side, when compared with the true block Hankel matrix. The block Hankel matrix from the proposed approach is, however, much smoother, and closer to the true block Hankel matrix. We also show the comparison of two block Hankel matrices using rank-reduction methods after the first iteration and their zoomed matrices in Figure 8. It can be observed that even after 1 iteration, the rank-reduced block Hankel matrix via the damped TSVD is much smoother, which indicates that more noise in each frequency slice is suppressed. Please note that all figures in Figures 6, 7, and 8 use the same plotting scale. It can also be seen that the block Hankel matrix after 1 iteration is far from the true block Hankel matrix, however, during the weighted POCS-like iterations, the block Hankel matrix is gradually converging to the true block Hankel matrix.

In order to numerically compare the denoising performance, we use the commonly

used SNR defined as follows to quantitatively measure the performance (Chen and Fomel, 2015; Yang et al., 2015; Gan et al., 2015a):

$$SNR = 10\log_{10}\frac{\|\mathbf{s}\|_2^2}{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2}. \tag{17}$$

where $\mathbf{s}$ denotes the vectorized clean data and $\hat{\mathbf{s}}$ denotes the vectorized reconstructed data. In this example, the SNRs of the observed data, the reconstructed result via the traditional rank-reduction method, and the reconstructed result via the damped rank-reduction method are -4.59 dB, 8.22 dB, and 11.62dB, respectively. The SNR comparison shows that the damped rank-reduction can result in a huge reconstruction improvement. In addition to SNR comparison, we also use the local similarity to measure the reconstruction performance. As claimed in Chen and Fomel (2015), reconstruction performance via local similarity has several advantages than the commonly used SNR measure, such as the local measurement of the reconstruction performance. Since the local similarity is seldom used in evaluating seismic data reconstruction, we provide a brief review of the theory of local similarity in Appendix C. The range of local similarity is between 0 and 1. The higher local similarity corresponds to the better reconstruction performance. Figure 9 shows the local similarity comparison in common offset domain and Figure 10 shows the local similarity comparison in common midpoint domain. Figures 9a, 9b, 9c, and 9d show the local similarity between the true data and the noisy data, observed data, result from the traditional method, and result from the proposed method, respectively, in common offset domain. Figures 10a, 10b, 10c, and 10d show the counterpart comparisons in common midpoint domain. Because of the strong random noise in the noisy data, the local similarity between noisy data and the true data is relatively low. It is surprising that the local similarity successfully captures the position of the useful signals, which is however not possible from the original amplitude cubes. A potential application of such byproduct in future research may be utilizing the detected useful signals from the noisy data as a prior information that can be put into the inversion framework to improve the final inversion performance. The local similarity between the observed data and the true data correlates well with the position of the available traces, since all missing traces have a zero local similarity with the true data. The local similarity between the result from traditional rank-reduction method is much higher than that of the noisy data and the similarity value is approaching 1, which is the highest possible value of local similarity. The proposed approach can make the local similarity between the result and true data even higher, as can be seen from Figures 9 and 10 that the similarity color that corresponds to the proposed method is redder than that of the traditional approach.

In order to demonstrate the reconstruction performance of the two rank-reduction methods with respect to noise level. We conduct 6 experiments of the 5D data reconstruction based on the first synthetic example, in which we use different noise level in different experiments. We plot the final SNRs using two different approaches after 10 iterations with respect to the noise variances in Figure 11. The red line shows the SNR diagram of the proposed approach and green line shows the SNR

diagram of the traditional approach. The noise variance of 6 experiments are 0.05, 0,1, 0.25, 0.5, 0.75, and 1. Please note that all presented examples above are based on the noise variance 0.25. It can be intuitively inferred that noise variance above 0.25 can make the data even noisier. From Figure 11, it is clear that the proposed approach is always superior than the traditional approach in terms of SNR. It is more important to observe that as the noise level increases, the difference in terms of SNR of the two approaches also increases. It offers us a practical guideline that the proposed approach can be best utilized in the condition of strong noise corruption (or in other words on observed data with low SNR), which is usually the feature of most land seismic data. In order to demonstrate the reconstruction performance with respect to sampling ratio of the observed data, we did 8 experiments with different sampling ratios. The SNRs using two approaches after 10 iterations with respect to the sampling ratios are shown in Figure 12. Red line denotes the proposed approach and green line denotes the traditional approach. The proposed approach is always better than the traditional approach and as the sampling ratio increases (in other words we get more and more available traces), the difference between the proposed and traditional approaches increases too.

In order to test the effectiveness of the rank-reduction method on data containing curved events and to compare the performances of different methods in this case, we then use the second example to show the performance. In Figure 13, we show a common midpoint gather comparison of the second synthetic example. Here we reshape the 3D data cube into a 2D matrix and compare the performance based on the 2D matrix. Figure 13a shows the clean data. Figure 13b shows the observed data with 70% missing traces and strong random noise. The useful signals are hardly seen from the observed data. After reconstruction using the traditional rank-reduction method with 10 POCS-like iterations, we obtain the reconstructed result as shown in Figure 13c. The result using the proposed approach is shown in Figure 13d. The reconstruction result using the traditional method seems to have significant residual noise and even causes artifacts in some parts of the data. The data highlighted by the green frame boxes are zoomed and shown in Figure 14, where we can have a better view to compare the performance in detail and to see how the proposed method helps recover almost the true model. The data highlighted by the black frame boxes are zoomed and shown in Figure 15, where we can see the artifacts caused by the traditional approach clearly and see the almost perfect performance from the proposed approach. We apply the proposed approach with $K = 12$ and $N = 3$ for this example. In this dataset, the original SNR of the observed is 0.34 dB. The SNR of the result from traditional rank-reduction approach is 13.56 dB while the SNR of the reconstruction result from the proposed damped rank-reduction method approach is 17.23 dB. In this example, we also calculate the local similarity between different datasets and the clean dataset and show them in Figure 16. Figure 16a shows the local similarity between the observed data and the clean data. Figure 16b shows the local similarity between the result from traditional method and the clean data and Figure 16c shows the local similarity between the result from the proposed method and the clean data. The local similarity vividly demonstrates that the proposed approach can obtain the

most similar result to the true model.

The next example is a field data example. The data has been binned onto a regular grid and a common offset gather of the field data is shown in Figure 17a. Although we are only allowed to show a small portion of the field data and to give a limited discussion based on the performance, the comparison between traditional and proposed rank-reduction methods is adequate to show the superior performance of our proposed method. In the example, about 80% traces are missing from the regular grids. It is obvious that the traditional rank-reduction method obtains a very successful recovery of the missing data, as shown in Figure 17b. However, there are still some residual noise in the reconstructed data, since we use a relatively large $K = 15$ to account for the curved events. We apply the proposed approach with $K = 15$ and $N = 3$ and obtain an almost perfect reconstruction performance, as shown in Figure 17c. The events become more coherent and the data becomes much cleaner. We zoom a part from each subfigure in Figure 17 and show the comparison in Figure 18. The zoomed data are highlighted by the red frame boxes shown in Figure 17. The zoomed sections shown in Figure 18 show a clearer comparison between the traditional and the proposed methods, and confirm that the proposed method can obtain a very clean and spatially coherent reconstruction result.

The selection of rank affects the performance of the traditional rank-reduction based approaches greatly in that smaller value of rank will cause significant damages to the useful signals while larger value of rank will cause serious residual noise left in the reconstructed data. However, it is even impossible to choose an optimal value of rank to stand for the exact signals in practice. In order to preserve as many useful signals as possible, we tend to set a relatively higher value of rank. The proposed approach relieves the dependence of the rank-reduction method on the rank selection since we can use a safe value of rank without compromising the SNR of the final reconstructed data by damping the residual noise that is caused from the increase of rank. In a nutshell, the proposed approach is less sensitive to the value of rank.

CONCLUSION

Traditional rank-reduction method cannot obtain acceptable result for 5D seismic data denoising and reconstruction when the signal-to-noise ratio (SNR) is extremely low. The traditional truncated singular value decomposition (TSVD) formula will cause residual noise in the seismic data after iterative reconstruction since the denoised data space is a mixture of the signal subspace and noise subspace. In order to better decompose the noisy data into signal and noise components, we introduce a damping operator into the traditional TSVD formula to dampen the singular values that contain significant information of residual noise. The improvement via the damped rank-reduction method is shown to be very appealing from both 5D synthetic and field data examples and can be achieved conveniently following the introduced damped TSVD formula. As shown from the comprehensive analysis of different 5D seismic data examples, we get some extra insightful conclusions in addition to the re-

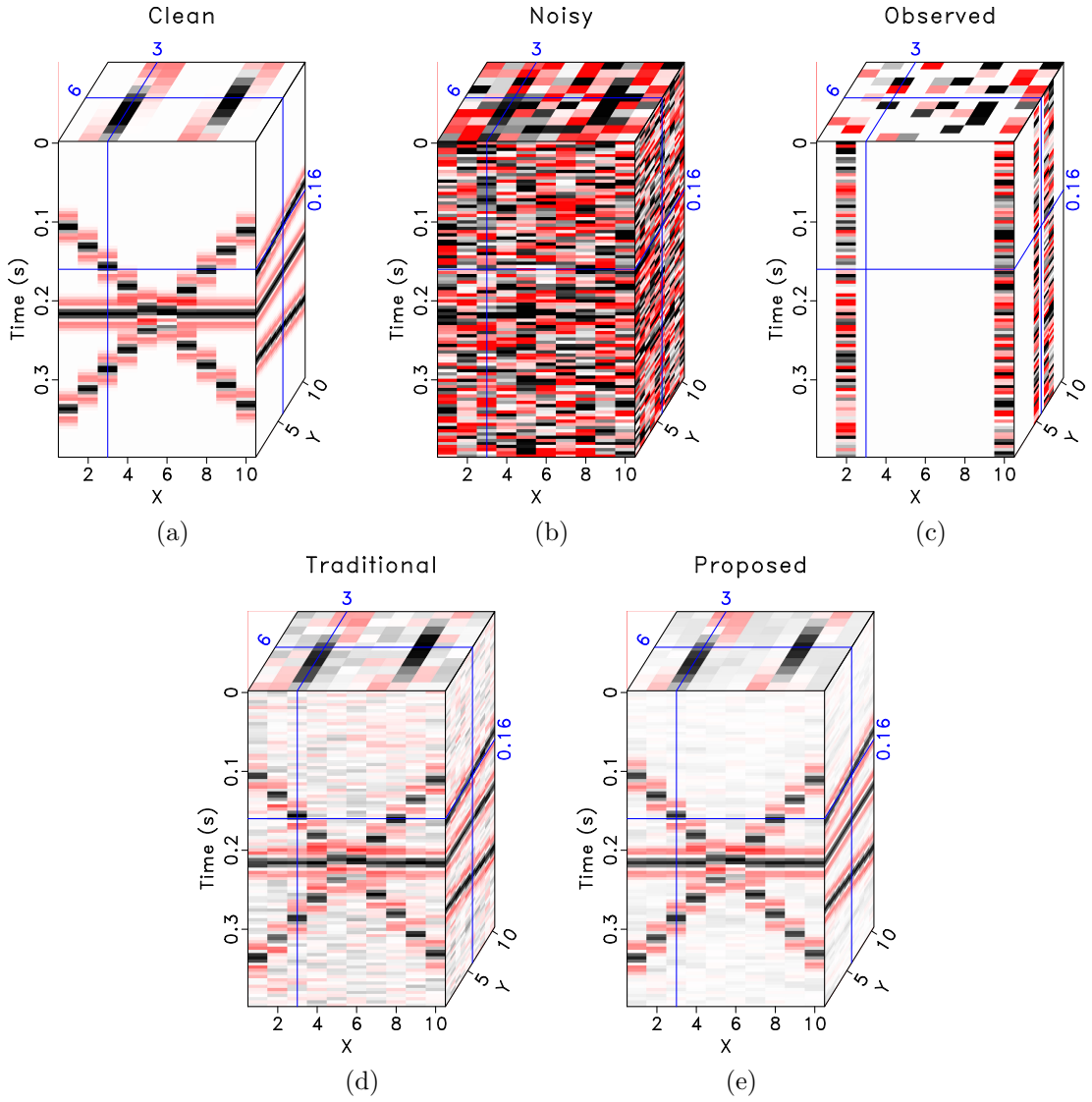


Figure 1: Common offset 3D cubes comparison ($hx = 5, hy = 5$). (a) Clean synthetic data. (b) Noisy synthetic data. (c) Observed synthetic data (with 70% missing traces). (d) Reconstructed data using the traditional approach. (e) Reconstructed data using the proposed approach.

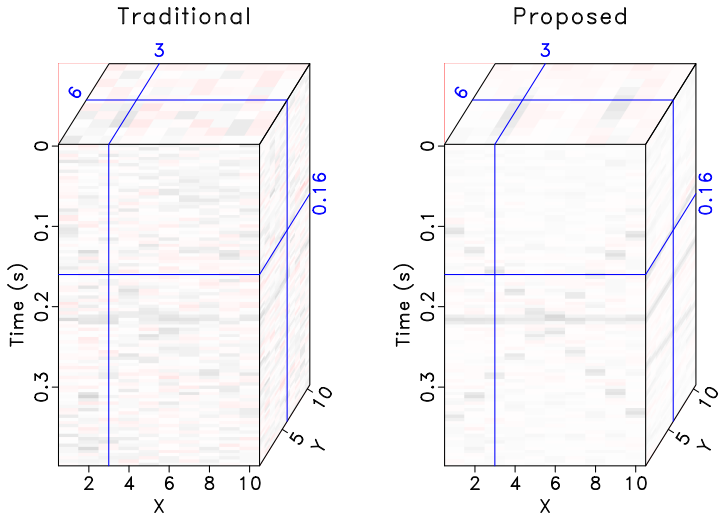


Figure 2: Reconstruction error comparison of common offset 3D cubes ($hx = 5, hy = 5$). (a) Reconstruction error using the traditional approach. (b) Reconstruction error using the proposed approach.

construction result itself. First, the block Hankel matrix after damped rank-reduction is smoother than that of traditional rank-reduction, which is caused by damping the singular values and matches with the overall superior performance of the proposed approach well when the block Hankel matrix is transformed back to time-space domain. Second, the advantage of the proposed rank-reduction method over the traditional method becomes more dominant as the noise level in the observed data increases, which makes the proposed approach very attractive in processing 5D land seismic data. We suggest to replace the current 5D seismic data interpolation framework with our proposed framework for better dealing with the highly noisy incomplete dataset.

ACKNOWLEDGMENTS

We would like to thank Shan Qu, Jiang Yuan, Mauricio Sacchi, Jean Virieux and Aaron Stanton for constructive suggestions that improved the manuscript greatly. We are grateful to developers of the Madagascar software package for providing corresponding codes for testing the algorithms and preparing the figures. This work is partially supported by National Basic Research Program of China (Grant NO: 2013 CB228602), National Natural Science Foundation of China (Grant No.41274137), National Science and Technology of Major Projects of China (Grant No. 2011ZX05019-006), National Engineering Laboratory of Offshore Oil Exploration, and the Texas Consortium for Computational Seismology (TCCS).

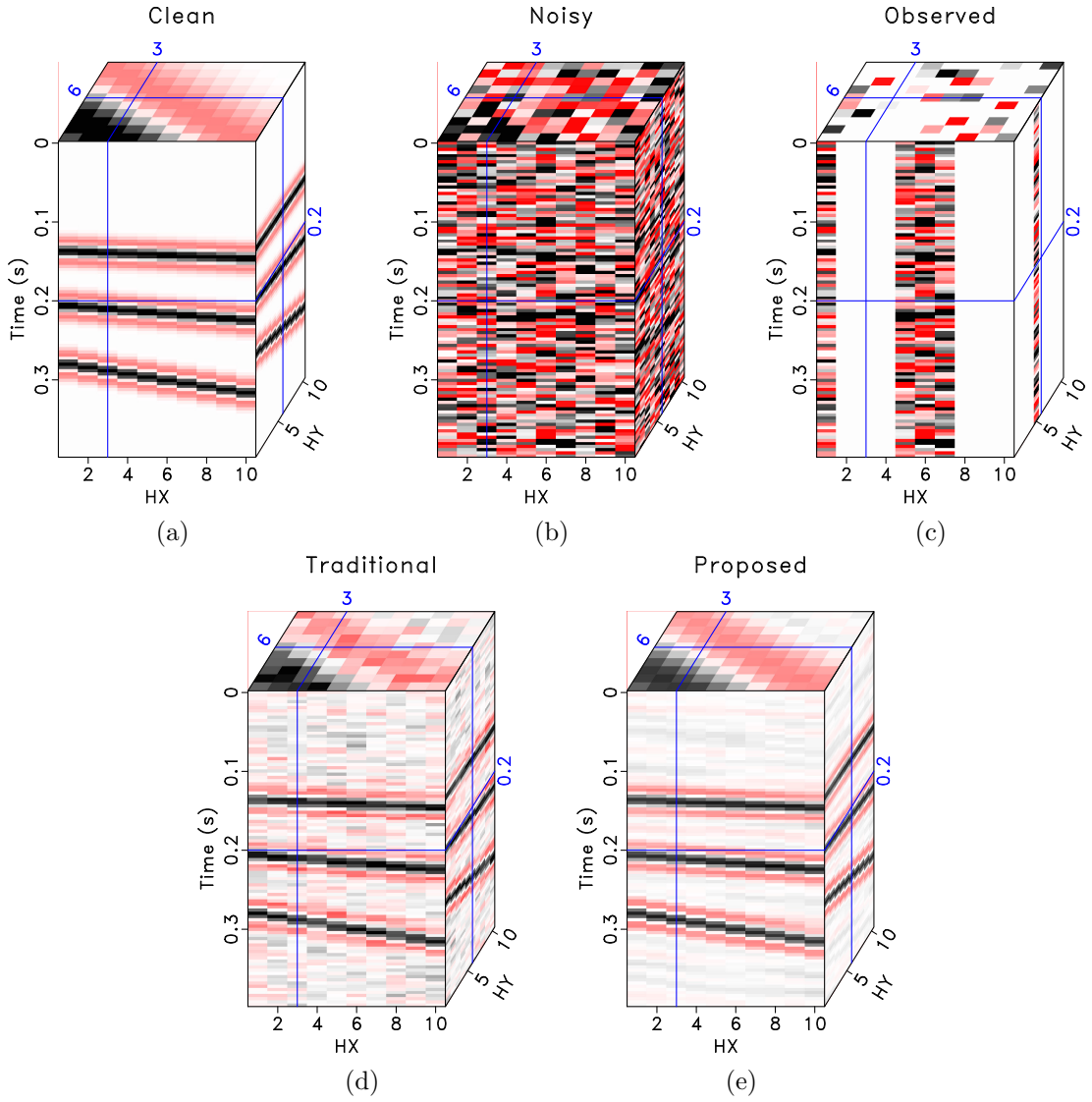


Figure 3: Common midpoint 3D cubes comparison ($x = 5, y = 5$). (a) Clean synthetic data. (b) Noisy synthetic data. (c) Observed synthetic data (with 70% missing traces). (d) Reconstructed data using the traditional approach. (e) Reconstructed data using the proposed approach.

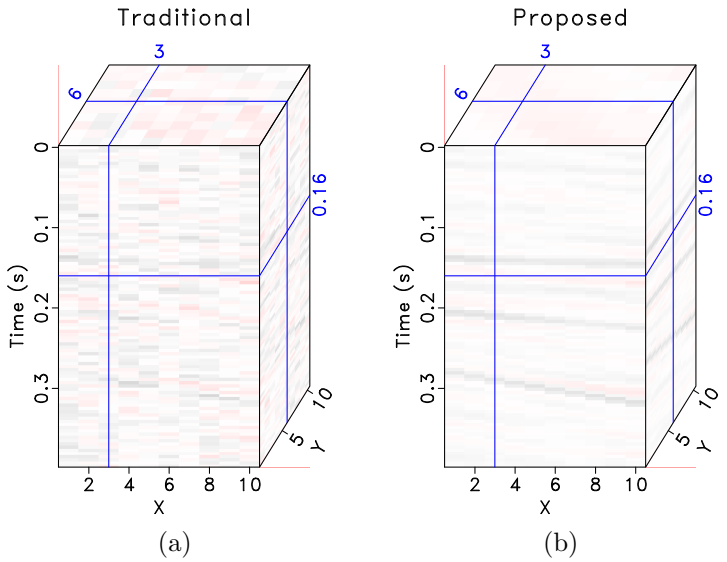


Figure 4: Reconstruction error comparison of common midpoint 3D cubes ($x = 5, y = 5$). (a) Reconstruction error using the traditional approach. (b) Reconstruction error using the proposed approach.

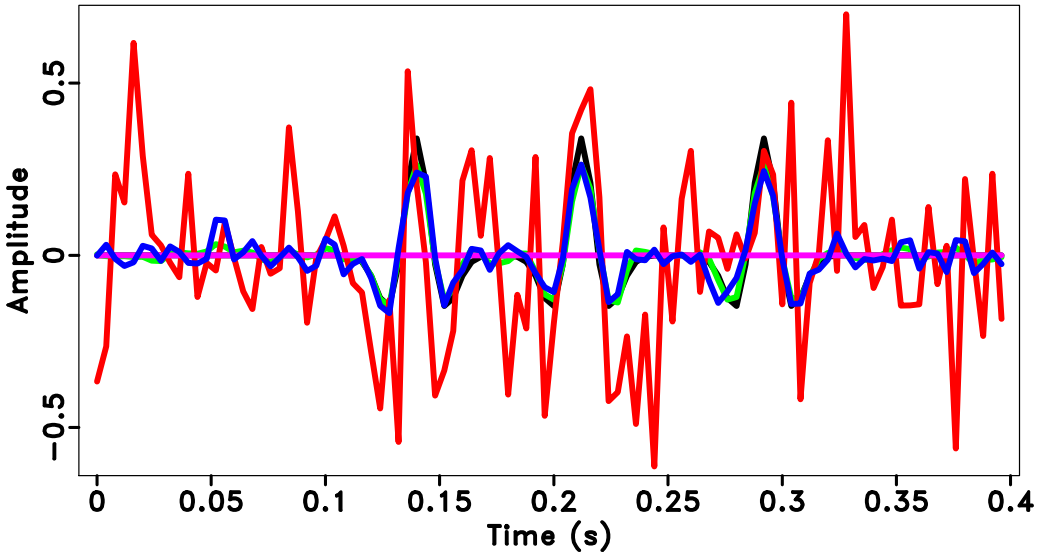


Figure 5: Comparison of the middle trace amplitude of each cubes in Figure 3. The black line is from the clean data. The red line is from the noisy data. The observed data in this case is zero and thus cannot be seen from the figure.The blue line corresponds to the traditional rank-reduction method. The green line corresponds to the proposed method. Note that the black and green lines are very close to each other, thus the reconstruction error using the proposed approach is much less than the traditional method for most parts.

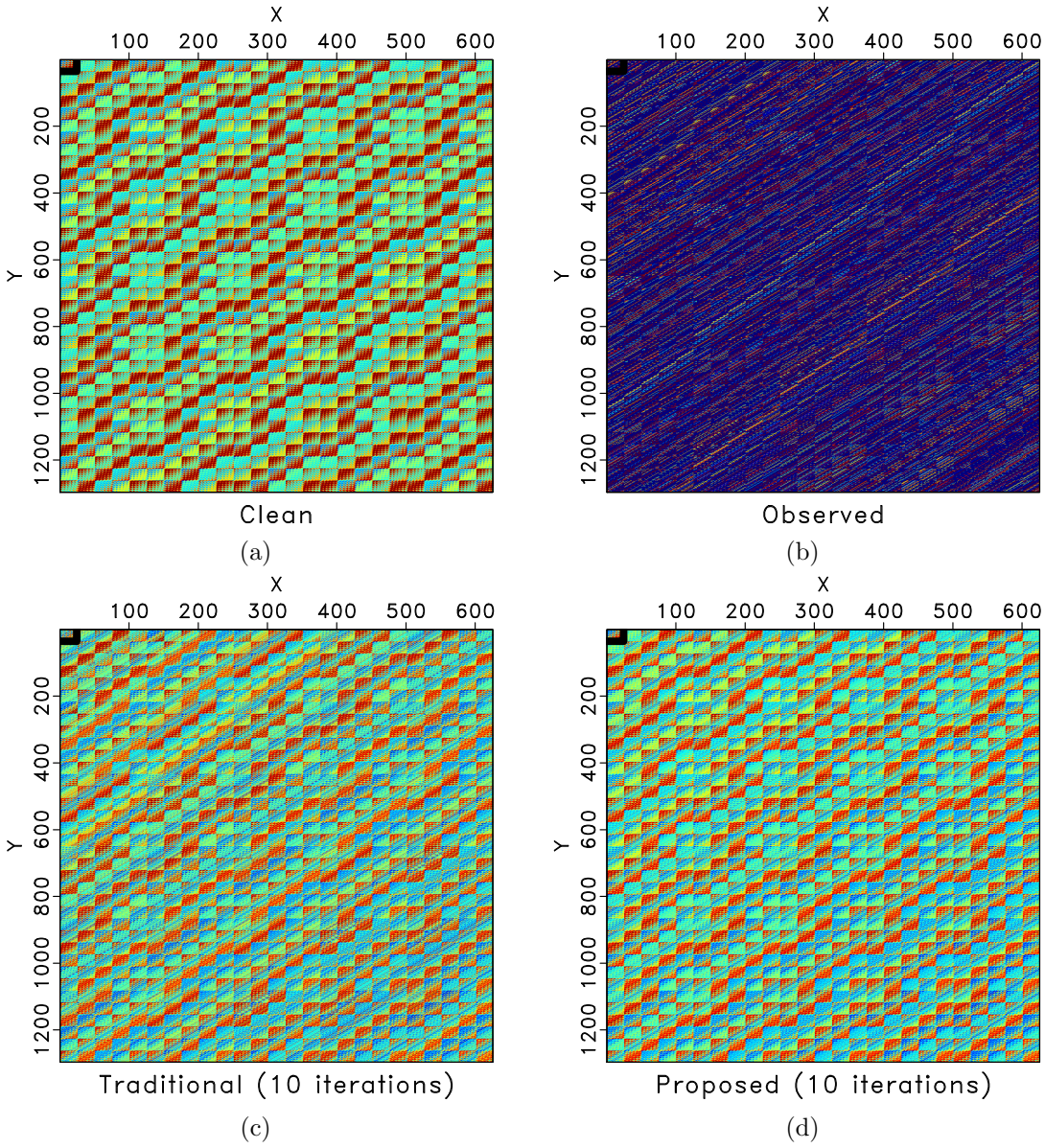


Figure 6: 26 Hz block Hankel matrix comparison. (a) Hankel matrix of the clean data. (b) Hankel matrix of the noisy data. (c) Hankel matrix of traditional method after 10 iterations. (d) Hankel matrix of the proposed method after 10 iterations. Please note the erratic artifacts shown in (c), which cannot be handled using the traditional rank-reduction method. The Hankel matrix of the proposed approach is, however, much closer to the true Hankel matrix.

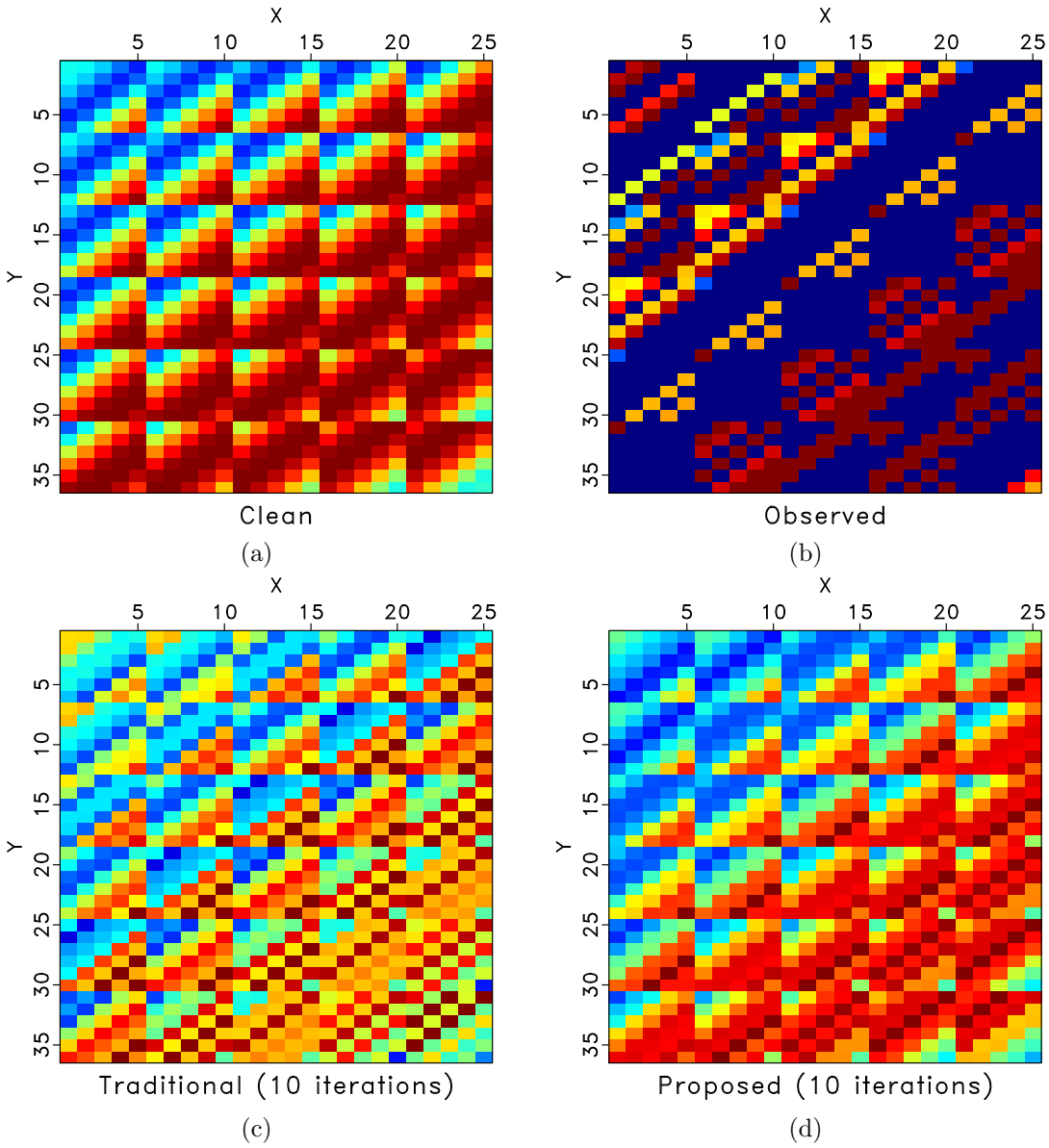


Figure 7: A close up look of the Hankel matrix (corresponding to the block Hankel matrix highlighted by the green frame box in Figure 6. (a) Hankel matrix of the clean data. (b) Hankel matrix of the noisy data. (c) Hankel matrix of traditional method after 10 iterations. (d) Hankel matrix of the proposed method after 10 iterations.

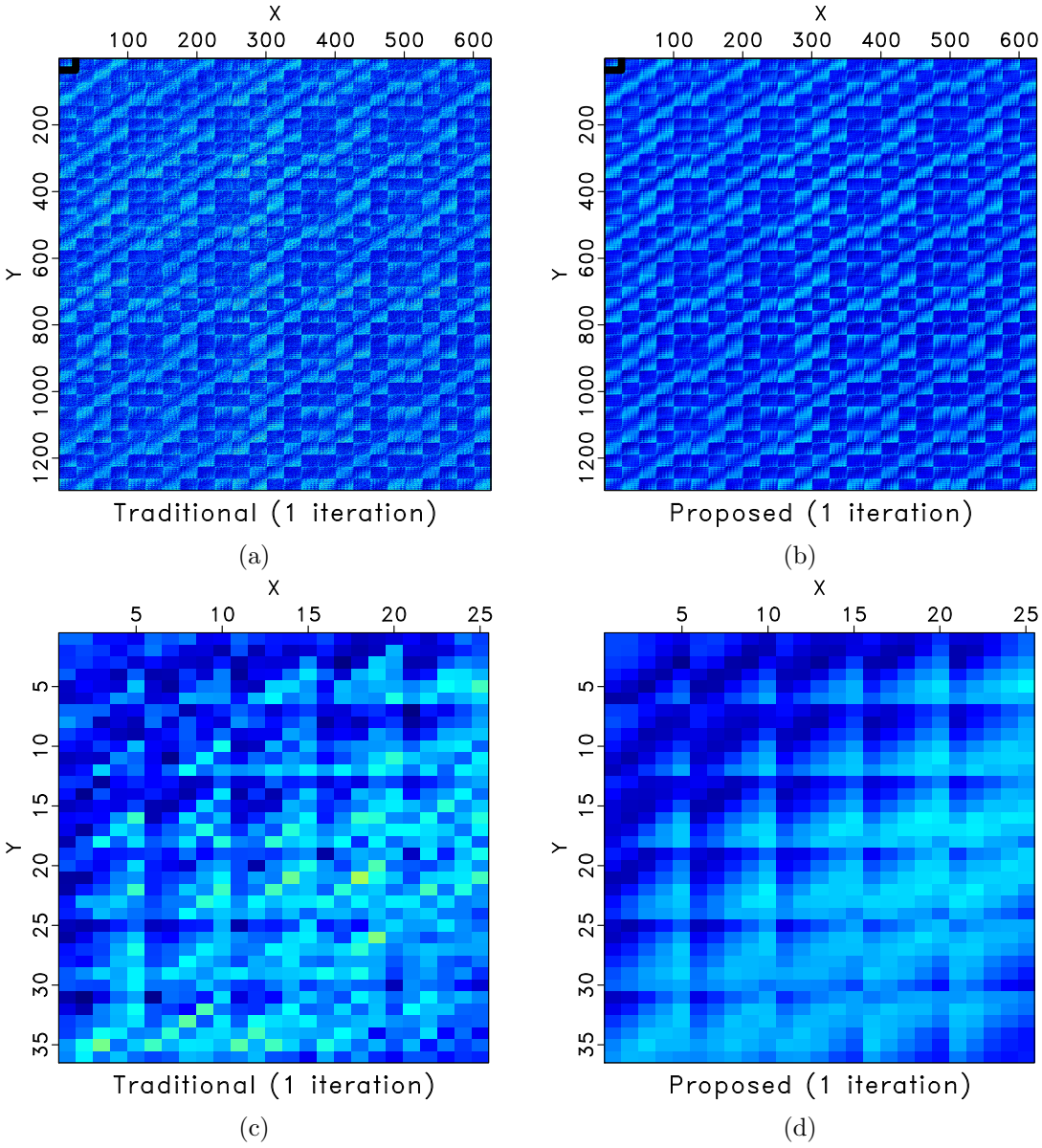


Figure 8: 26 Hz block Hankel matrix comparison after 1 iteration between two approaches. (a) Hankel matrix of traditional method after 1 iteration. (b) Hankel matrix of the proposed method after 1 iteration. (c)&(d) A close up look of the two frame boxes in (a) and (b).

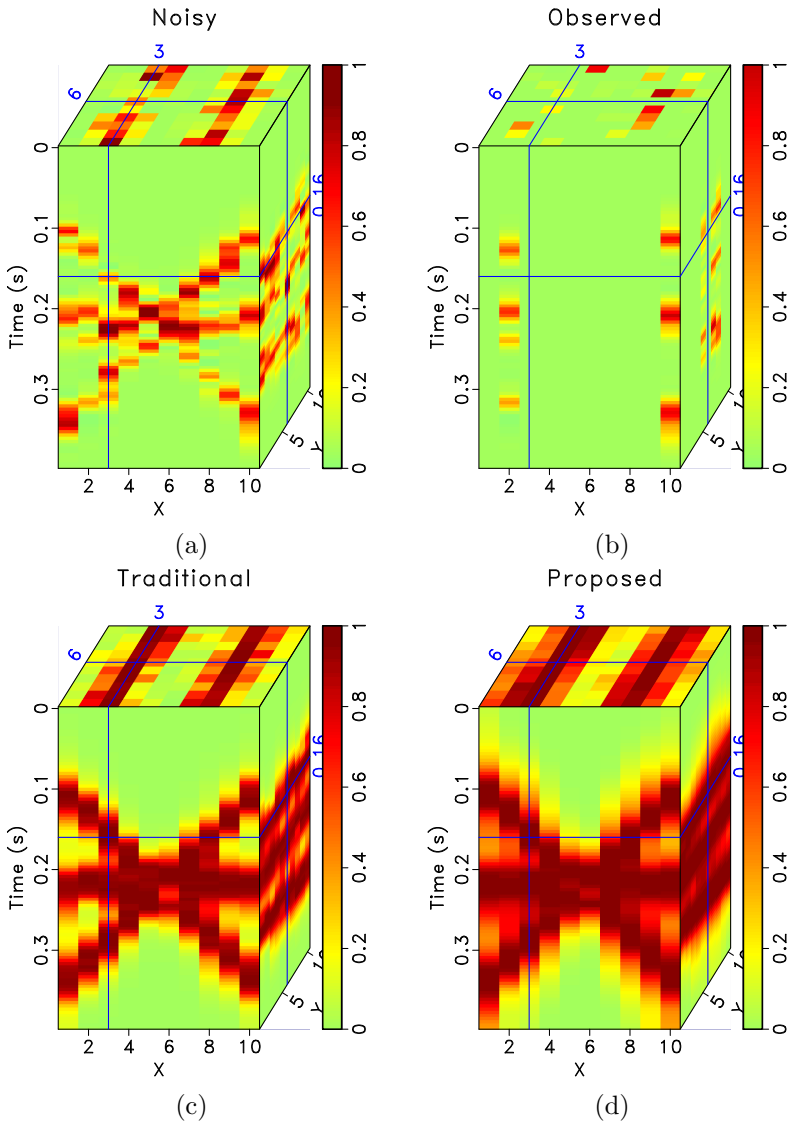


Figure 9: Local similarity comparison of common offset 3D cubes comparison ($hx = 5, hy = 5$). (a) Noisy synthetic data. (b) Observed synthetic data (with 70% missing traces). (c) Reconstructed data using the traditional approach. (d) Reconstructed data using the proposed approach.

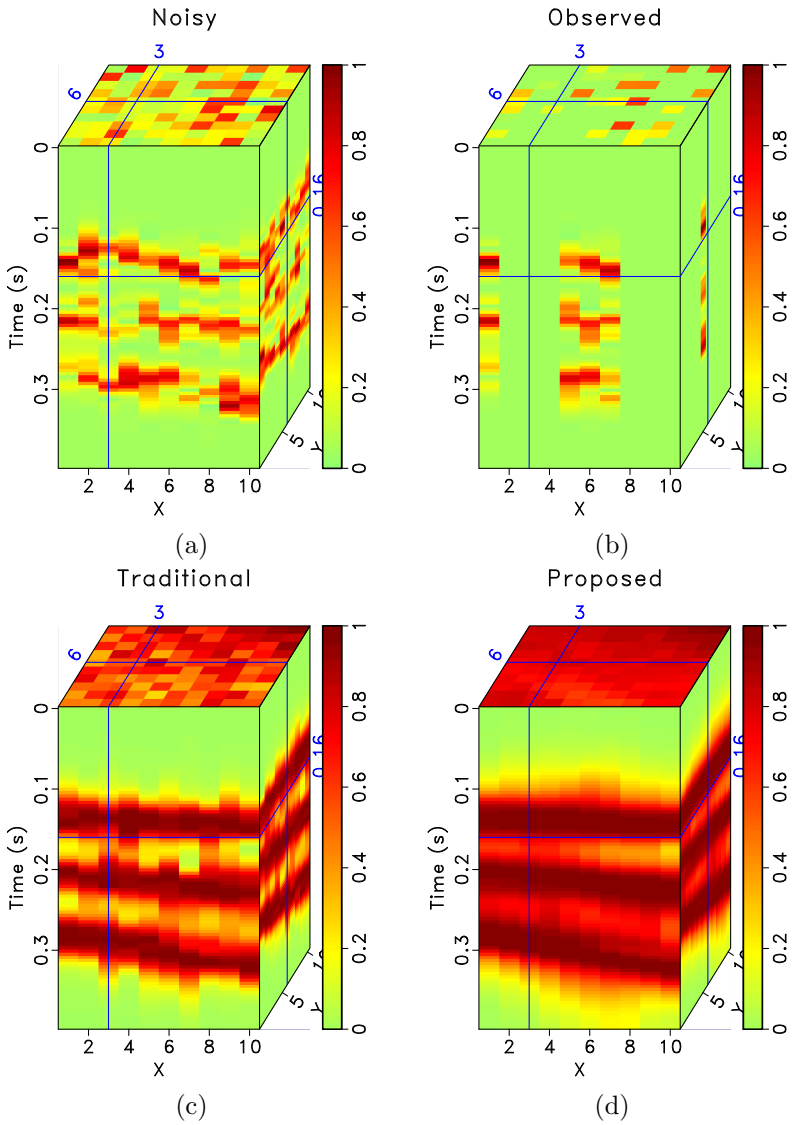


Figure 10: Local similarity comparison of common midpoint 3D cubes ($x = 5, y = 5$). (a) Noisy synthetic data. (b) Observed synthetic data (with 70% missing traces). (c) Reconstructed data using the traditional approach. (d) Reconstructed data using the proposed approach.

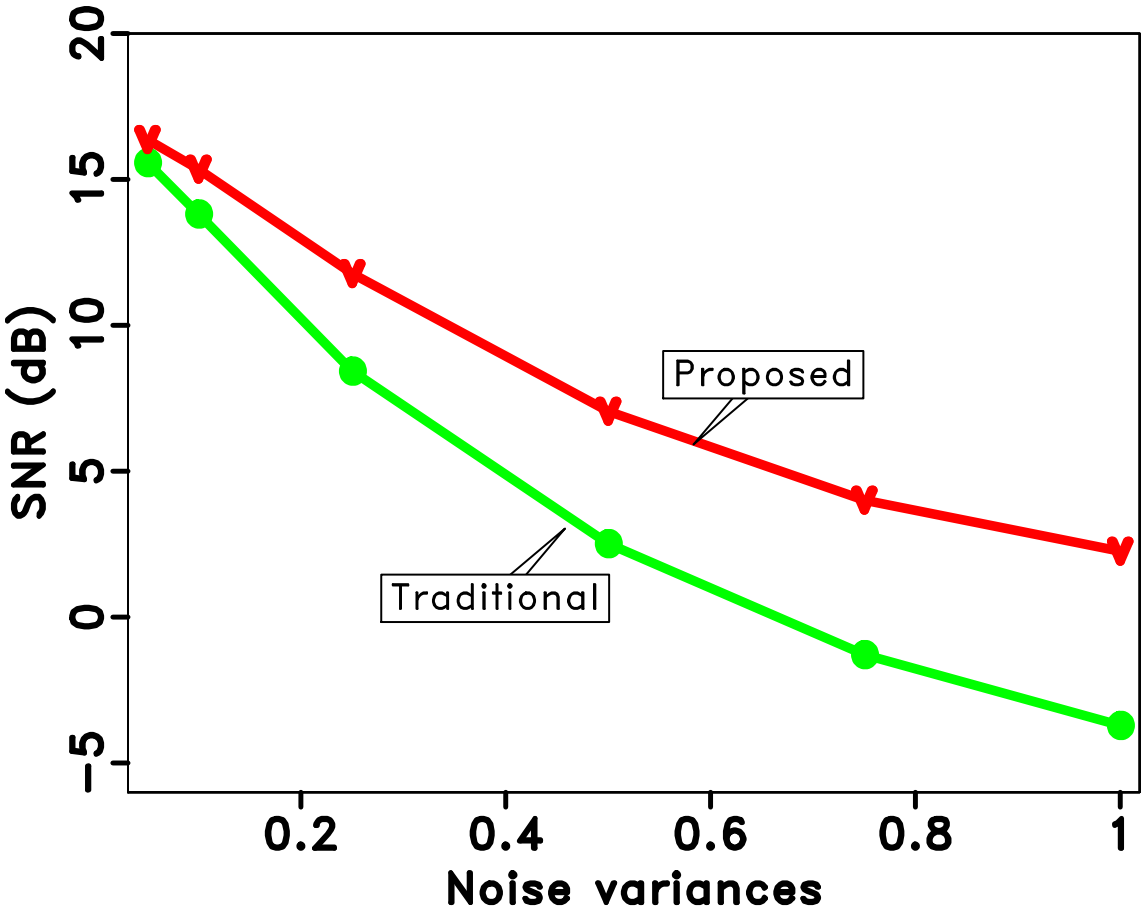


Figure 11: SNR diagrams of the different approaches with respect to the noise level (variance value). Note that the proposed approach outperforms the traditional methods more and more as the noise level increases.

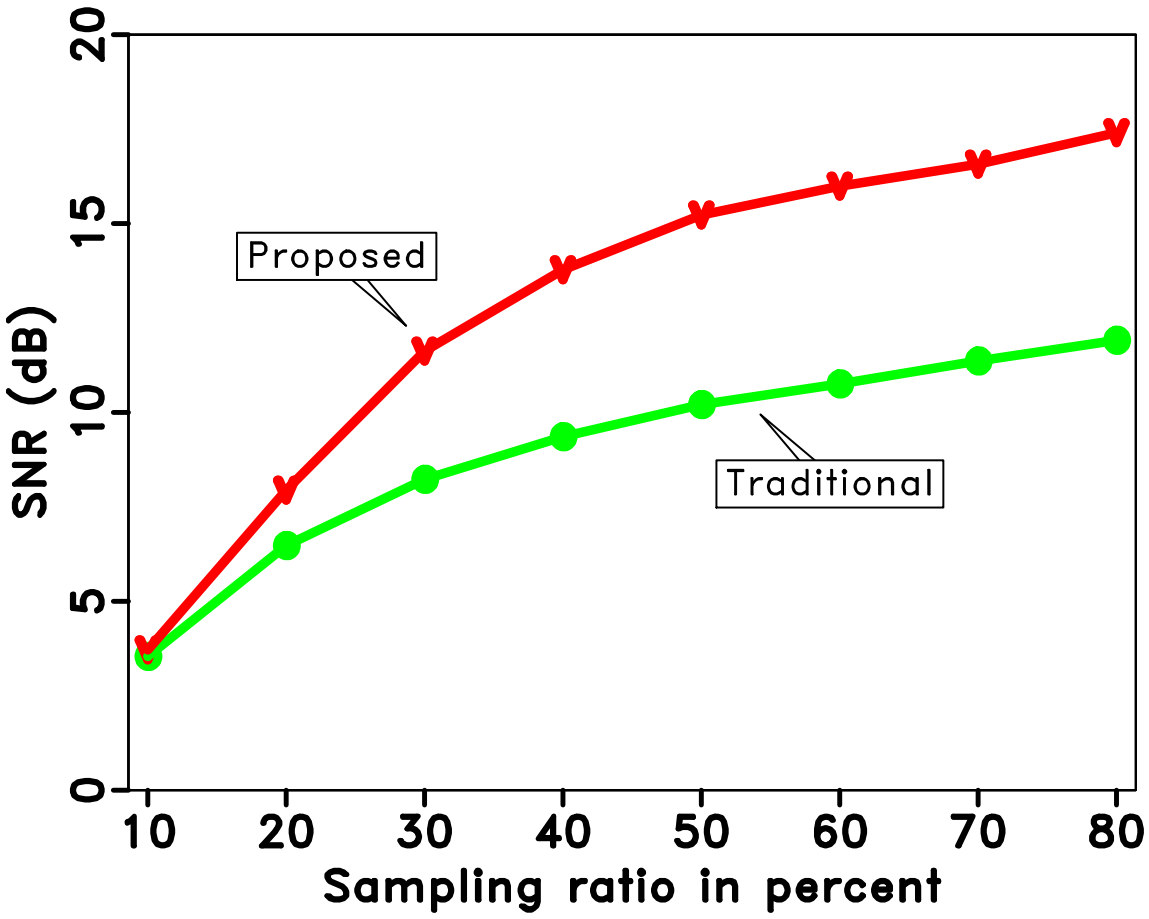


Figure 12: SNR diagrams of the different approaches with respect to the sampling ratio (in percent).

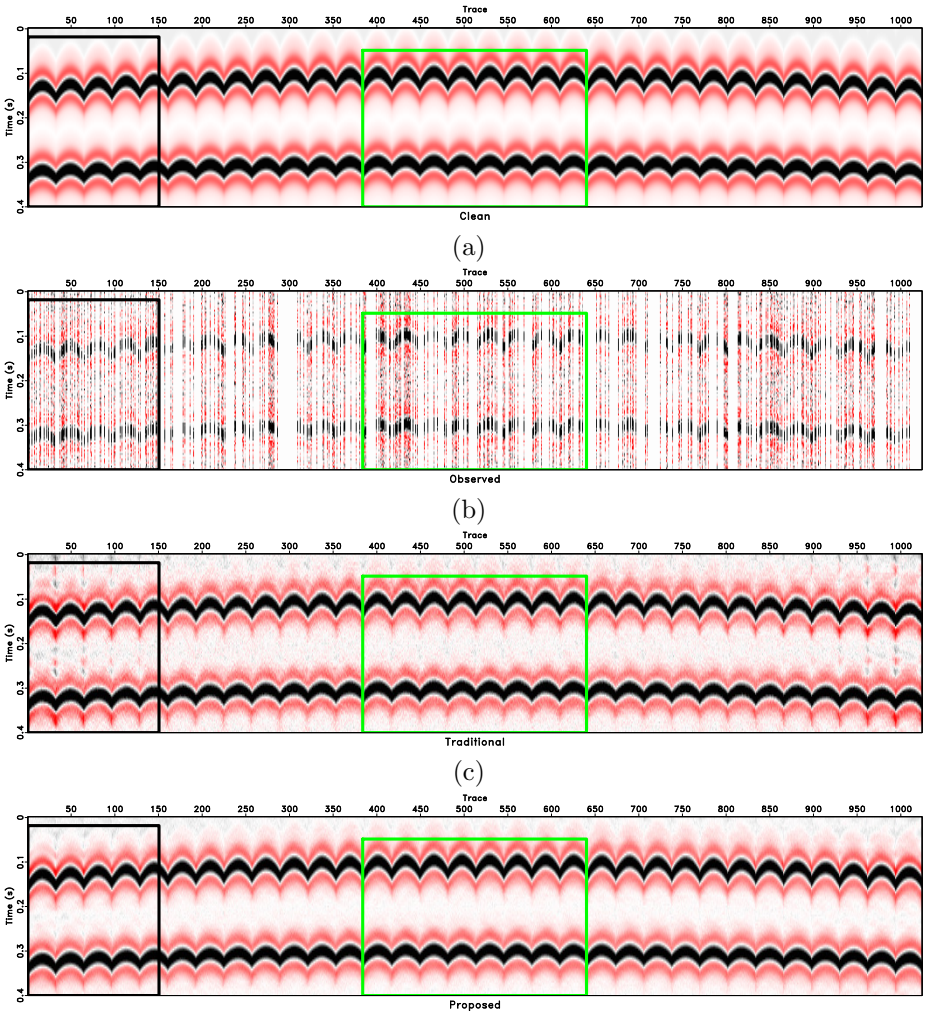


Figure 13: Common midpoint gather comparison (reshaped into a 2D matrix). (a) Clean synthetic data with curved events. (b) Observed synthetic data (with 70% missing traces and strong noise). (c) Reconstructed data using the traditional approach. (d) Reconstructed data using the proposed approach.

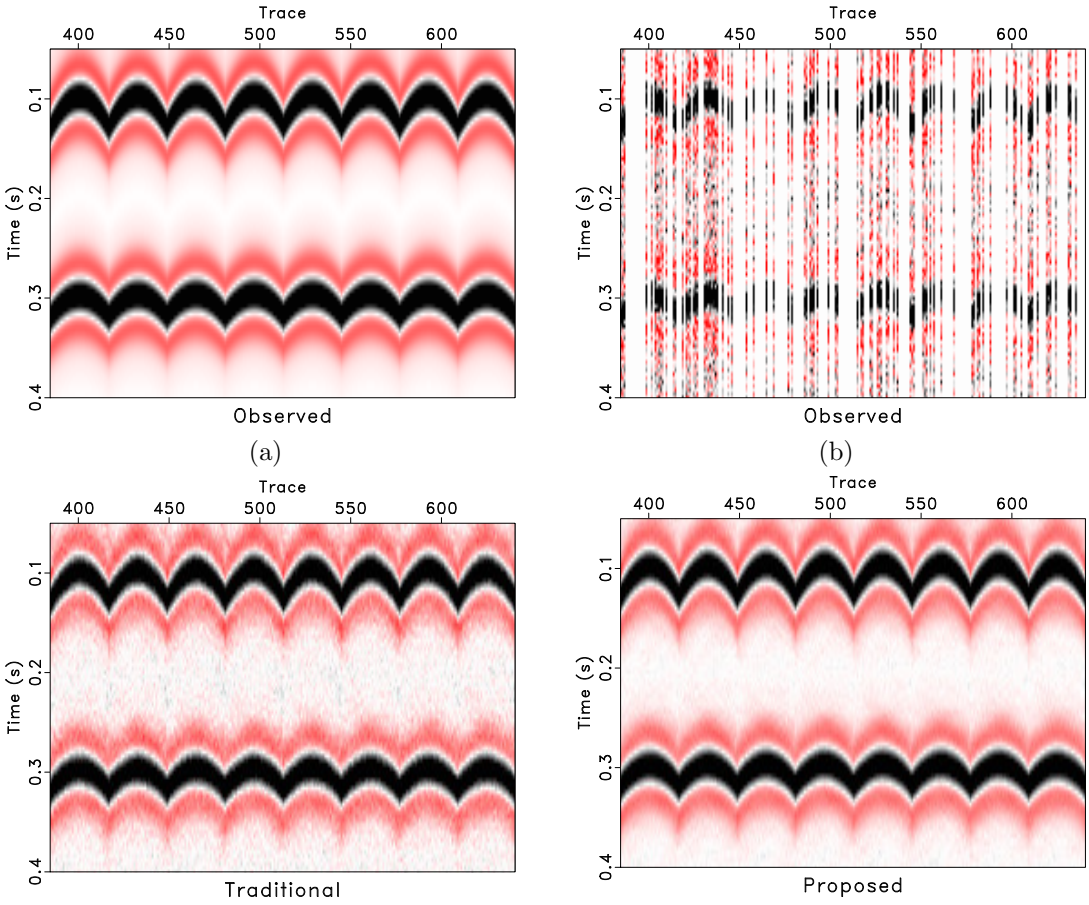


Figure 14: Zoomed sections from Figure 13 (the green frame box). (a) Clean data. (b) Observed data. (b) Reconstructed data using the traditional approach. (c) Reconstructed data using the proposed approach.

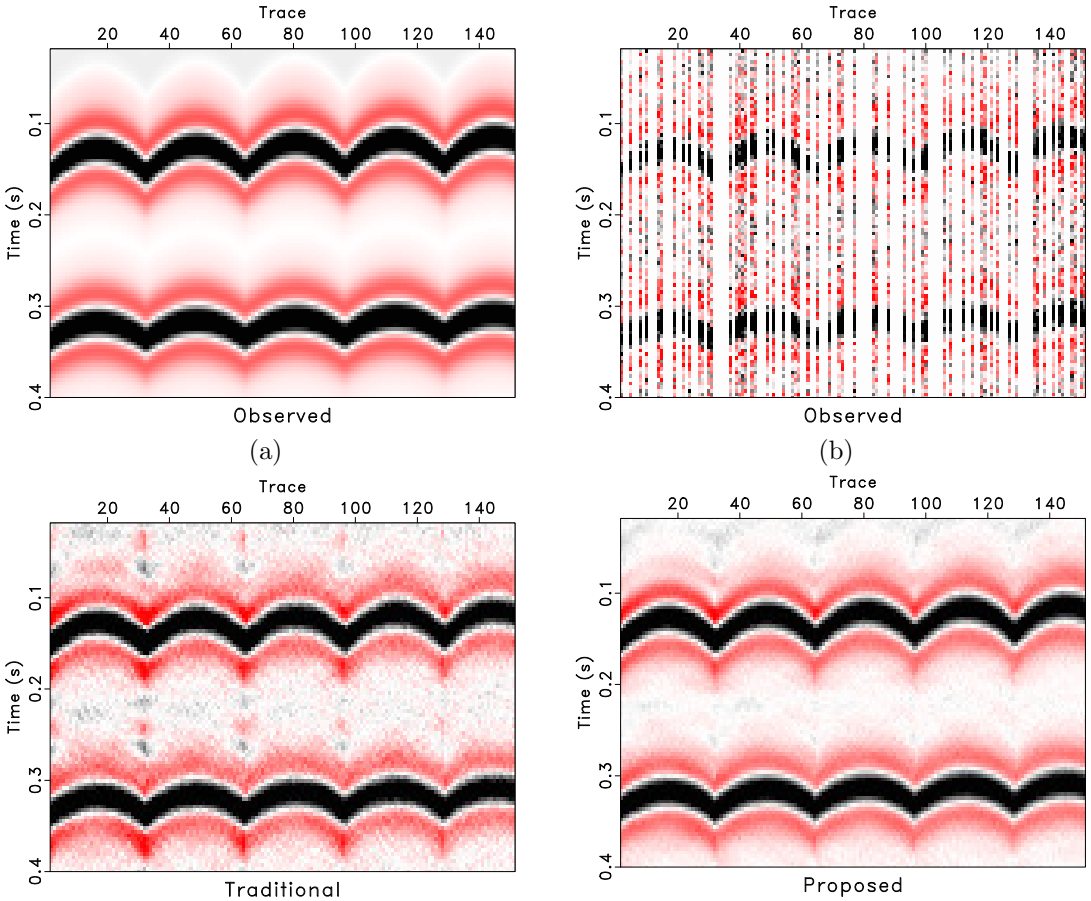
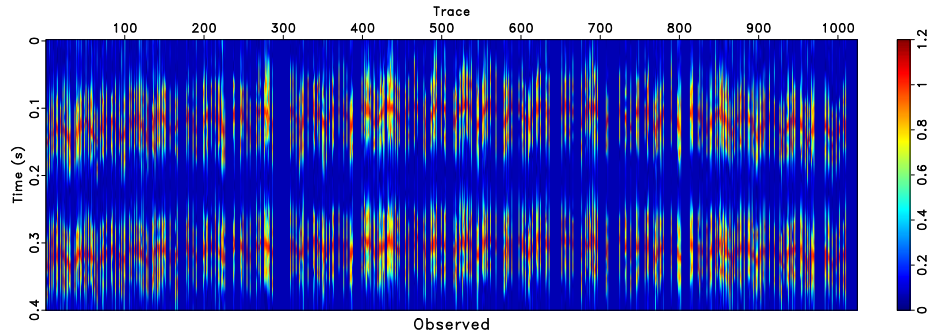
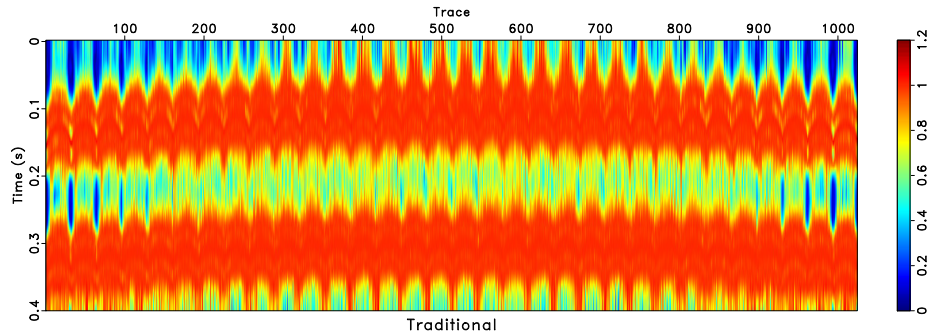


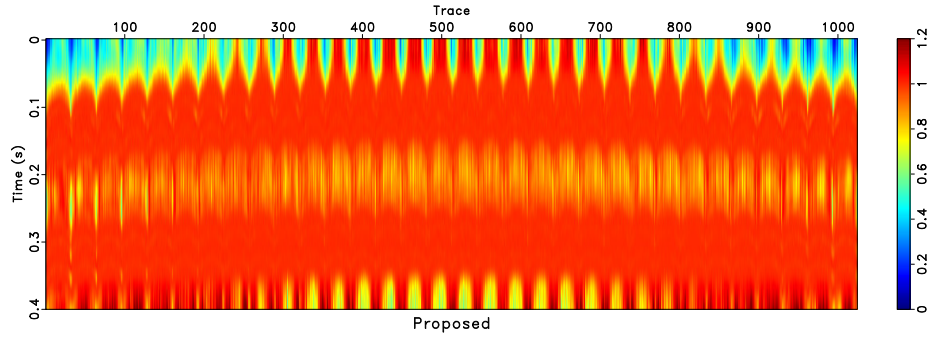
Figure 15: Zoomed sections from Figure 13 (the black frame box). (a) Clean data. (b) Observed data. (b) Reconstructed data using the traditional approach. (c) Reconstructed data using the proposed approach. Note the strong artifacts caused from the traditional approach.



(a)

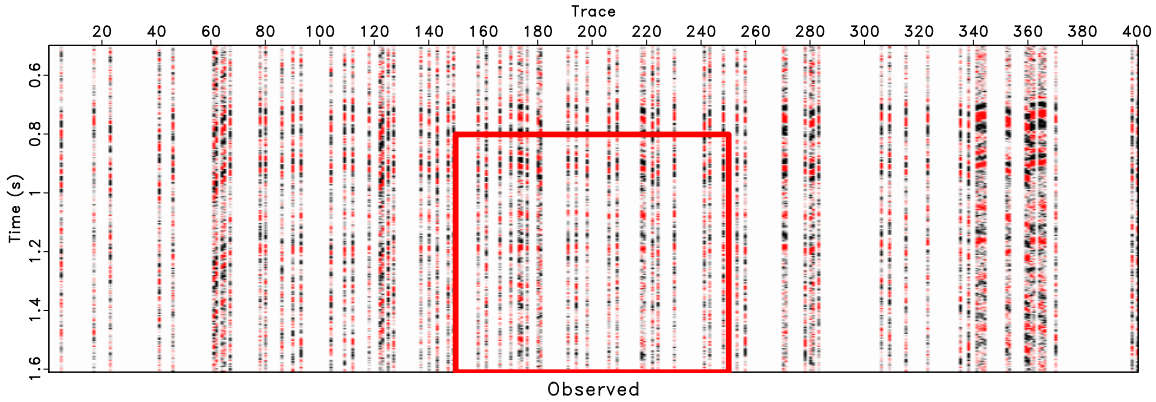


(b)

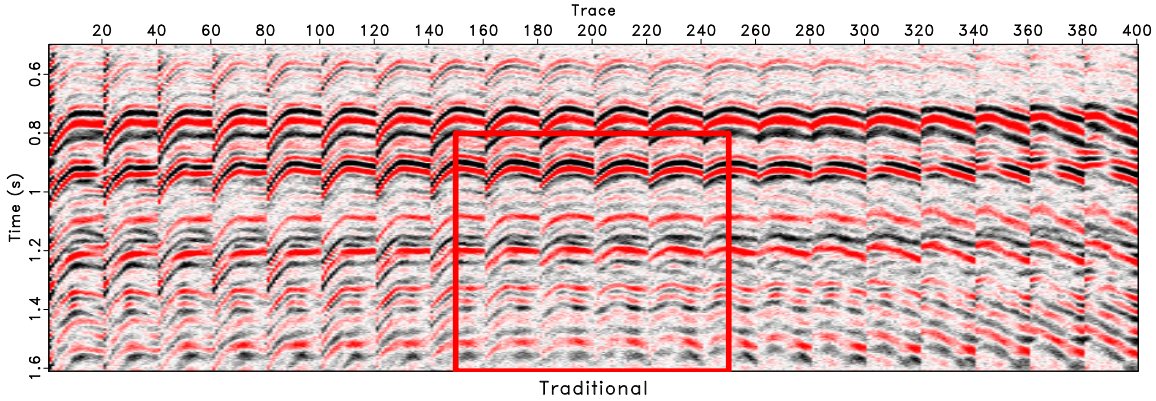


(c)

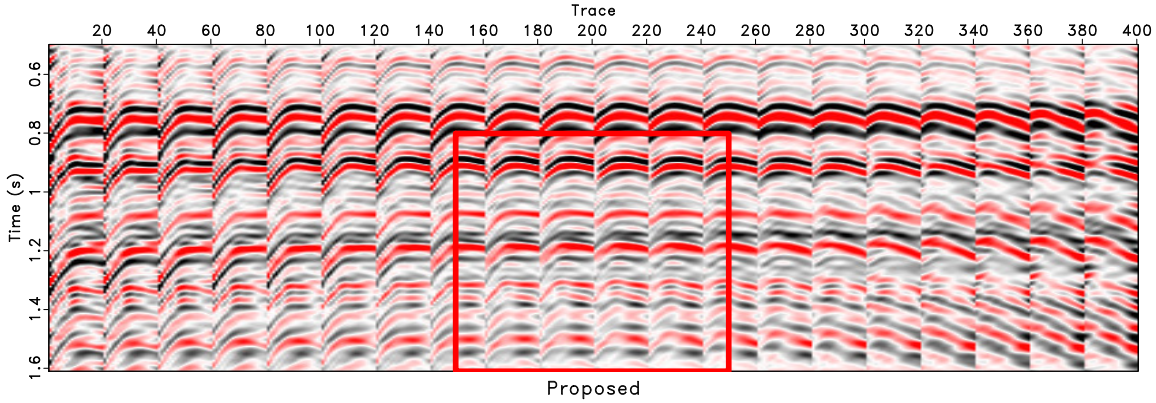
Figure 16: Local similarity comparison with respect to the clean data. (a) Similarity of the Observed synthetic data. (b) Similarity of the reconstructed data using the traditional approach. (c) Similarity of the reconstructed data using the proposed approach.



(a)



(b)



(c)

Figure 17: Common offset gather comparison (reshaped into a 2D matrix). (a) Observed field data (with 80% missing traces). (b) Reconstructed data using the traditional approach. (c) Reconstructed data using the proposed approach.

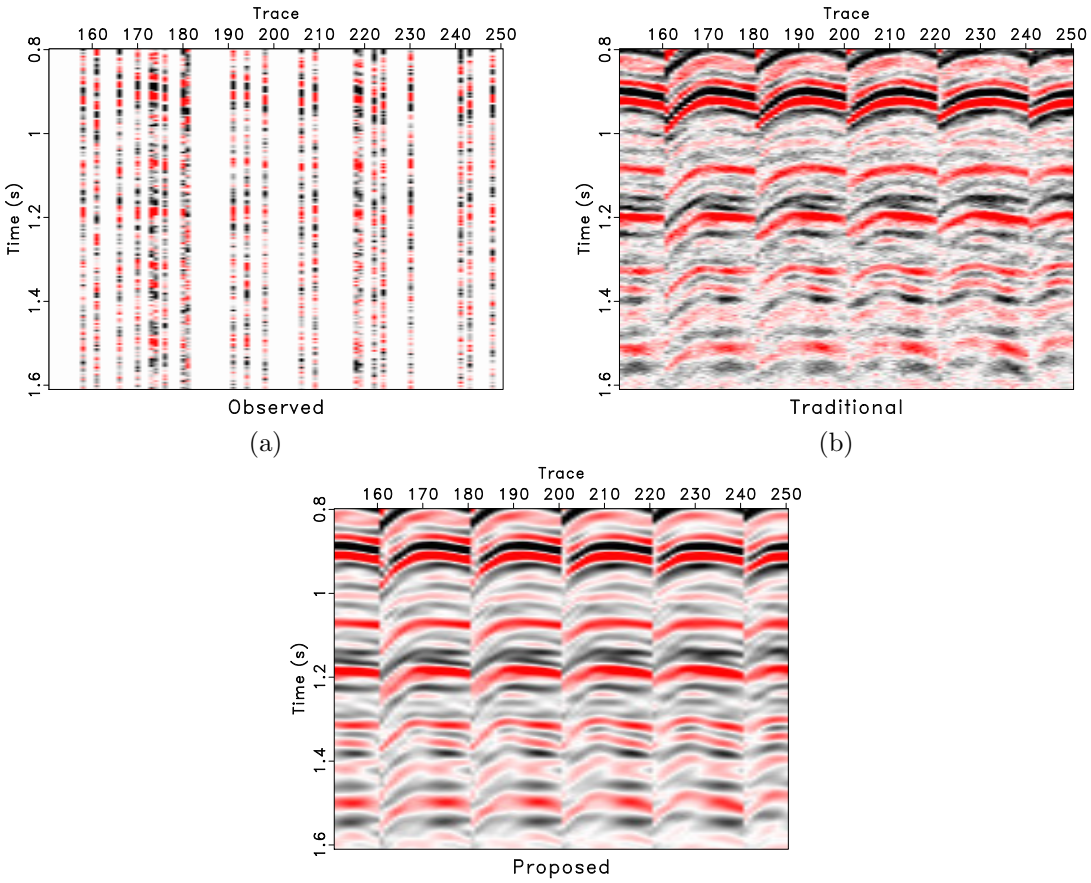


Figure 18: Zoomed sections from Fig 17. (a) Observed field data. (b) Reconstructed data using the traditional approach. (c) Reconstructed data using the proposed approach.

APPENDIX A: EXACT FORMULATION OF THE BLOCK HANKEL MATRIX

The block Hankel matrix $\mathbf{M}^{(4)}$ can be expressed as

$$\mathbf{M}^{(4)} = \mathbf{S} + \mathbf{N}, \tag{A-1}$$

where $\mathbf{S}$ and $\mathbf{N}$ denote signal component and the random noise component that are both in $\mathbf{M}^{(4)}$, respectively. The singular value decomposition (SVD) of $\mathbf{S}$ can be represented as:

$$\mathbf{S} = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \Sigma_1^S & \mathbf{0} \\ \mathbf{0} & \Sigma_2^S \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^S)^H \\ (\mathbf{V}_2^S)^H \end{bmatrix}. \tag{A-2}$$

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

$$\mathbf{S} = \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H. \tag{A-3}$$

Because equation A-2 is a SVD of the signal matrix $\mathbf{S}$, the left matrix in equation A-2 is a unitary matrix:

$$\mathbf{I} = \mathbf{U}^S (\mathbf{U}^S)^H = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} (\mathbf{U}_1^S)^H \\ (\mathbf{U}_2^S)^H \end{bmatrix}. \tag{A-4}$$

Combining equations A-1, A-3, and A-4, we can derive:

$$\begin{aligned} \mathbf{M}^{(4)} &= \mathbf{S} + [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} (\mathbf{U}_1^S)^H \\ (\mathbf{U}_2^S)^H \end{bmatrix} \mathbf{N} \\ &= \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H + \left(\mathbf{U}_1^S (\mathbf{U}_1^S)^H + \mathbf{U}_2^S (\mathbf{U}_2^S)^H \right) \mathbf{N} \\ &= \mathbf{U}_1^S \left((\mathbf{U}_1^S)^H \mathbf{N} + \Sigma_1^S (\mathbf{V}_1^S)^H \right) + \mathbf{U}_2^S (\mathbf{U}_2^S)^H \mathbf{N} \\ &= \mathbf{U}_1^S \left(\mathbf{N}^H \mathbf{U}_1^S + \mathbf{V}_1^S \Sigma_1^S \right)^H + \mathbf{U}_2^S (\mathbf{N}^H \mathbf{U}_2^S)^H \\ &= [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{N}^H \mathbf{U}_1^S + \mathbf{V}_1^S \Sigma_1^S)^H \\ (\mathbf{N}^H \mathbf{U}_2^S)^H \end{bmatrix}, \end{aligned} \tag{A-5}$$

we can further derive equation A-5 and factorize $\mathbf{M}^{(4)}$ as follows:

$$\mathbf{M}^{(4)} = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} (\Sigma_1)^{-1} (\mathbf{N}^H \mathbf{U}_1^S + \mathbf{V}_1^S \Sigma_1^S)^H \\ (\Sigma_2)^{-1} (\mathbf{N}^H \mathbf{U}_2^S)^H \end{bmatrix}, \tag{A-6}$$

where Σ_1 and Σ_2 denote diagonal and positive definite matrices. Please note that $\mathbf{M}^{(4)}$ 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. We observe that the left matrix has orthonormal columns and the middle matrix is diagonal. It can be shown that the right matrix also has orthonormal columns. Thus, equation A-6 is an SVD of $\mathbf{M}^{(4)}$. According to the TSVD method, we let Σ_2 be $\mathbf{0}$ and then the following equation holds:

$$\begin{aligned} \tilde{\mathbf{M}}^{(4)} &= \mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N} + \mathbf{U}_1^S \Sigma_1^S (\mathbf{V}_1^S)^H \\ &= \mathbf{S} + \mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N}. \end{aligned} \tag{A-7}$$

APPENDIX B: DAMPED TSVD FORMULA

We first reformulate equation A-7 as

$$\mathbf{S} = \mathbf{M}^{(4)} - \mathbf{U}_1^S(\mathbf{U}_1^S)^H\mathbf{N}. \tag{B-1}$$

Inserting equation 8 into equation B-1, we can further derive

$$\mathbf{S} = \mathbf{U}_1^{M^{(4)}}\Sigma_1^{M^{(4)}}(\mathbf{V}_1^{M^{(4)}})^H - \mathbf{U}_1^S(\mathbf{U}_1^S)^H\mathbf{N}. \tag{B-2}$$

The SVD of $\mathbf{M}^{(4)}$ can be expressed as

$$\mathbf{M}^{(4)} = [\mathbf{U}_1^{M^{(4)}} \quad \mathbf{U}_2^{M^{(4)}}] \begin{bmatrix} \Sigma_1^{M^{(4)}} & \mathbf{0} \\ \mathbf{0} & \Sigma_2^{M^{(4)}} \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^{M^{(4)}})^H \\ (\mathbf{V}_2^{M^{(4)}})^H \end{bmatrix}. \tag{B-3}$$

Because equations B-3 and A-6 are both SVDs of $\mathbf{M}^{(4)}$, we let

$$\mathbf{U}_1^S = \mathbf{U}_1^{M^{(4)}}, \tag{B-4}$$

$$\Sigma_1 = \Sigma_1^{M^{(4)}}, \tag{B-5}$$

$$(\mathbf{N}^H\mathbf{U}_1^S + \mathbf{V}_1^S\Sigma_1^S)(\Sigma_1)^{-1} = \mathbf{V}_1^{M^{(4)}}. \tag{B-6}$$

Inserting equations B-4 and B-6 into equation B-2, we can derive:

$$\mathbf{S} = \mathbf{U}_1^{M^{(4)}}\left\{\Sigma_1^{M^{(4)}}(\mathbf{V}_1^{M^{(4)}})^H - \left[\Sigma_1(\mathbf{V}_1^{M^{(4)}})^H - \Sigma_1^S(\mathbf{V}_1^S)^H\right]\right\}. \tag{B-7}$$

For simplification, we assume that there exist such $\mathbf{A}$ and $\mathbf{B}$ that $\mathbf{V}_1^S = \mathbf{V}_1^{M^{(4)}}\mathbf{A}$ and $\Sigma_1 = \Sigma_1^{M^{(4)}}\mathbf{B}$. $\mathbf{A}$ is a square matrix and $\mathbf{B}$ is a diagonal matrix. Then we can simplify $\mathbf{S}$ as:

$$\mathbf{S} = \mathbf{U}_1^{M^{(4)}}\Sigma_1^{M^{(4)}}\mathbf{T}\left(\mathbf{V}_1^{M^{(4)}}\right)^H. \tag{B-8}$$

$$\mathbf{T} = \mathbf{I} - \mathbf{B}\left(\mathbf{I} - (\Sigma_1)^{-1}\Sigma_1^S\mathbf{A}^H\right), \tag{B-9}$$

where $\mathbf{I}$ is a unit matrix and here we name $\mathbf{T}$ the damping operator.

From equation B-5, $\mathbf{B} = \mathbf{I}$. From equation B-6, it is straightforward to derive

$$\begin{aligned} \mathbf{V}_1^S &= (\mathbf{V}_1^{M^{(4)}} - \mathbf{N}^H\mathbf{U}_1^S(\Sigma_1)^{-1})\Sigma_1(\Sigma_1^S)^{-1} \\ &\approx \mathbf{V}_1^{M^{(4)}}(\mathbf{I} - (\mathbf{V}_1^{M^{(4)}})^o\mathbf{N}^H\mathbf{U}_1^S(\Sigma_1)^{-1})\Sigma_1(\Sigma_1^S)^{-1} \end{aligned} \tag{B-10}$$

where $(\mathbf{V}_1^{M^{(4)}})^o$ satisfies that $\parallel \mathbf{I} - \mathbf{V}_1^{M^{(4)}}(\mathbf{V}_1^{M^{(4)}})^o \parallel \rightarrow 0$. Considering $\mathbf{V}_1^S = \mathbf{V}_1^{M^{(4)}}\mathbf{A}$, then the following relation holds

$$\begin{aligned} \mathbf{A} &\approx (\mathbf{I} - (\mathbf{V}_1^{M^{(4)}})^o\mathbf{N}^H\mathbf{U}_1^S(\Sigma_1)^{-1})\Sigma_1(\Sigma_1^S)^{-1} \\ &= (\mathbf{I} - \Gamma)\Sigma_1(\Sigma_1^S)^{-1}, \end{aligned} \tag{B-11}$$

where $\Gamma = (\mathbf{V}_1^{M^{(4)}})^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}$.

Inserting equation B-11 and $\mathbf{B} = \mathbf{I}$ into equation B-9, we can obtain a simplified formula:

$$\mathbf{T} = \mathbf{I} - \Gamma. \tag{B-12}$$

Combing equations B-8 and B-12, we can conclude that the true signal is a damped version of the previous TSVD method (equation 8), with the damping operator defined by equation B-12. Right now, there is still one unknown parameter needed to be defined: Γ. Although we have a potential selection $\Gamma = (\mathbf{V}_1^{M^{(4)}})^o \mathbf{N}^H \mathbf{U}_1^S (\Sigma_1)^{-1}$, as defined during the derivation of $\mathbf{A}$, we cannot calculate it because we do not know $\mathbf{N}$ and $\mathbf{U}_1^S$. A very pleasant denoising performance can be obtained when Γ is chosen as

$$\Gamma \approx \hat{\delta}^N \left(\Sigma_1^{M^{(4)}} \right)^{-N}, \tag{B-13}$$

where $\hat{\delta}$ denotes the maximum element of $\Sigma_2^{M^{(4)}}$ and N denotes the damping factor. We use such approximation because of three reasons. (1) $\hat{\delta}$ reflects the energy of random noise and $\Sigma_1^{M^{(4)}}$ contains the information of signal. (2) Because the diagonal elements of $\Sigma^{M^{(4)}}$ are in a descending order, $\hat{\delta}$ is certainly smaller than every diagonal element of $\Sigma_1^{M^{(4)}}$, and $\hat{\delta}/\delta_1 < \hat{\delta}/\delta_2 < \cdots < \hat{\delta}/\delta_K$, where δ_i denotes ith diagonal entry in $\Sigma_1^{M^{(4)}}$. (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 B-8, B-12, and B-13, we conclude the approximation of $\mathbf{S}$ as:

$$\mathbf{S} = \mathbf{U}_1^{M^{(4)}} \Sigma_1^{M^{(4)}} \mathbf{T} (\mathbf{V}_1^{M^{(4)}})^H, \tag{B-14}$$

$$\mathbf{T} = \mathbf{I} - (\Sigma_1^{M^{(4)}})^{-N} \hat{\delta}^N. \tag{B-15}$$

APPENDIX C: REVIEW OF LOCAL SIMILARITY

We begin our review based on 1D vectors $\mathbf{a}$ and $\mathbf{b}$. A common way to measure the similarity between two signals is to calculate the correlation coefficient,

$$c = \frac{\mathbf{a}^T\mathbf{b}}{\| \mathbf{a} \|_2 \| \mathbf{b} \|_2}, \tag{C-1}$$

where c is the correlation coefficient, $\mathbf{a}^T\mathbf{b}$ denotes the dot product of $\mathbf{a}$ and $\mathbf{b}$. $\| \cdot \|_2$ denotes the L_2 norm of the input vector. The correlation coefficient c can measure the correlation, or in other words similarity, between two vectors within the selected window in a global manner. However, in some specific applications, a local measurement for the similarity is more demanded. In order to measure the local similarity of two vectors, it is straightforwardly to apply local windows to the target vectors:

$$c(i) = \frac{\mathbf{a}_i^T\mathbf{b}_i}{\| \mathbf{a}_i \|_2 \| \mathbf{b}_i \|_2}, \tag{C-2}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ denote the localized vectors centered on the ith entry of input long vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. The windowing is sometime troublesome, since the measured similarity is largely dependent on the windowing length and the measured local similarity might be discontinuous because of the separate calculations in windows. To avoid the negative performance caused by local windowing calculations, Fomel (2007) proposed an elegant way for calculating smooth local similarity via solving two inverse problems.

Let us first rewrite equation C-1 in a different form. Getting the absolute value of both sides of equation C-1, considering $\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a}$, we have

$$|c| = \sqrt{\frac{(\mathbf{a}^T\mathbf{b})(\mathbf{b}^T\mathbf{a})}{\| \mathbf{a} \|_2^2 \| \mathbf{b} \|_2^2}} = \sqrt{\frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}} \frac{\mathbf{b}^T\mathbf{a}}{\mathbf{b}^T\mathbf{b}}} \tag{C-3}$$

We let $c_1 = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$ and $c_2 = \frac{\mathbf{b}^T\mathbf{a}}{\mathbf{b}^T\mathbf{b}}$, equation C-3 turns to

$$|c| = \sqrt{c_1 c_2}. \tag{C-4}$$

It is obvious that scalars c_1 and c_2 come from two least-squares inverse problem:

$$c_1 = \arg\min_{\tilde{c}} \| \mathbf{a} - \mathbf{b}\tilde{c} \|_2^2 \tag{C-5}$$

$$c_2 = \arg\min_{\tilde{c}} \| \mathbf{b} - \mathbf{a}\tilde{c} \|_2^2 \tag{C-6}$$

The local similarity attribute is based on equations C-5 and C-6, but is in a localized version:

$$\mathbf{c}_1 = \arg\min_{\tilde{\mathbf{c}}_1} \|\mathbf{a} - \mathbf{B}\tilde{\mathbf{c}}_1\|_2^2, \tag{C-7}$$

$$\mathbf{c}_2 = \arg\min_{\tilde{\mathbf{c}}_2} \|\mathbf{b} - \mathbf{A}\tilde{\mathbf{c}}_2\|_2^2, \tag{C-8}$$

where $\mathbf{A}$ is a diagonal operator composed from the elements of $\mathbf{a}$: $\mathbf{A} = diag(\mathbf{a})$ and $\mathbf{B}$ is a diagonal operator composed from the elements of $\mathbf{b}$: $\mathbf{B} = diag(\mathbf{b})$. Equations C-7 and C-8 are solved via shaping regularization

$$\mathbf{c}_1 = [\lambda_1^2\mathbf{I} + \mathcal{T}(\mathbf{B}^T\mathbf{B} - \lambda_1^2\mathbf{I})]^{-1}\mathcal{T}\mathbf{B}^T\mathbf{a}, \tag{C-9}$$

$$\mathbf{c}_2 = [\lambda_2^2\mathbf{I} + \mathcal{T}(\mathbf{A}^T\mathbf{A} - \lambda_2^2\mathbf{I})]^{-1}\mathcal{T}\mathbf{A}^T\mathbf{b}, \tag{C-10}$$

where $\mathcal{T}$ is a smoothing operator, and λ_1 and λ_2 are two parameters controlling the physical dimensionality and enabling fast convergence when inversion is implemented iteratively. These two parameters can be chosen as $\lambda_1 = \|\mathbf{B}^T\mathbf{B}\|_2$ and $\lambda_2 = \|\mathbf{A}^T\mathbf{A}\|_2$ (Fomel, 2007).

After the $\mathbf{c}_1$ and $\mathbf{c}_2$ are solved, the localized version of $|c|$ is

$$\mathbf{c} = \sqrt{\mathbf{c}_1 \circ \mathbf{c}_2}, \tag{C-11}$$

where $\mathbf{c}$ is the output local similarity and $\circ$ denotes element-wise product.

To calculate the 2D and 3D versions of local similarity, as used in demonstrating the reconstruction performance in the main contents of the paper, one needs to first reshape a 2D matrix or a 3D tensor into a 1D vector and then to use the formula shown above to calculate the local similarity vector. Since the local similarity is of the same size of the input two vectors, it can be easily reshaped into the 2D or 3D form for display purposes.

REFERENCES

Abma, R., and N. Kabir, 2005, Comparisons of interpolation methods: The Leading Edge, **24**, 984–989.

———, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, E91–E97.

Candès, E. J., and T. Tao, 2010, The power of convex relaxation: Near-optimal matrix completion: IEEE Trans. Inf. Theory, **56**, 2053–2080.

Canning, A., and G. H. F. Gardner, 1996, Reducing 3D acquisition footprint for 3D dmo and 3D prestack migration: Geophysics, **63**, 1177–1183.

Chen, Y., K. Chen, P. Shi, and Y. Wang, 2014a, Irregular seismic data reconstruction using a percentile-half-thresholding algorithm: Journal of Geophysics and Engineering, **11**, 065001.

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, 2014b, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: Geophysics, **79**, V179–V189.

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

Chen, Y., L. Zhang, and L. Mo, 2015, Seismic data interpolation using nonlinear shaping regularization: Journal of Seismic Exploration, **24**, 327–342.

Chiu, S. K., 2014, Multidimensional interpolation using a model-constrained minimum weighted norm interpolation: Geophysics, **79**, V191–V199.

Duijndam, A. J., M. A. Schonewille, and C. O. H. Hindriks, 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: Geophysics, **64**, 524–538.

Fomel, S., 2002, Application of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

———, 2003, Seismic reflection data interpolation with differential offset and shot continuation: Geophysics, **68**, 733–744.

———, 2007, Local seismic attributes: Geophysics, **72**, A29–A33.

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

Gan, S., Y. Chen, S. Wang, X. Chen, W. Huang, and H. Chen, 2016, Compressive sensing for seismic data reconstruction using a fast projection onto convex sets algorithm based on the seislet transform: Journal of Applied Geophysics, **130**, 194–208.

Gan, S., Y. Chen, S. Zu, S. Qu, and W. Zhong, 2015a, 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, and X. Chen, 2015b, Seismic data reconstruction via fast projection onto convex sets in the seislet transform domain: SEG expanded abstracts: 85th Annual international meeting, 3814–3819.

Gan, S., S. Wang, Y. Chen, Y. Zhang, and Z. Jin, 2015c, Dealiased seismic data interpolation using seislet transform with low-frequency constraint: IEEE Geoscience and remote sensing letters, **12**, 2150–2154.

Gao, J., J. Cheng, and M. Sacchi, 2015a, A new 5D seismic reconstruction method based on a parallel square matrix factorization algorithm: SEG expanded abstracts: 85th Annual international meeting, 3784–3788.

Gao, J., A. Stanton, and M. Sacchi, 2015b, Parallel matrix factorization algorithm and its application to 5D seismic reconstruction and denoising: Geophysics, **80**, V173–V187.

Halko, N., P. Martinsson, and J. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: SIAM Review, **53**.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.

Huang, W., R. Wang, Y. Chen, H. Li, and S. Gan, 2016, Damped multichannel singular spectrum analysis for 3D random noise attenuation: Geophysics, **81**, V261–V270.

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.

Kreimer, N., and M. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, V113–

V122.

Kreimer, N., A. Stanton, and M. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, V273–V284.

Li, C., C. C. Mosher, and S. Kaplan, 2012, Interpolated compressive sensing for seismic data reconstruction: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–6.

Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, 1560–1568.

Liu, W., S. Cao, Y. Chen, and S. Zu, 2016, An effective approach to attenuate random noise based on compressive sensing and curvelet transform: Journal of Geophysics and Engineering, **13**, 135–145.

Liu, W., S. Cao, G. Li, and Y. He, 2015, Reconstruction of seismic data with missing traces based on local random sampling and curvelet transform: Journal of Applied Geophysics, **115**, 129.

Naghizadeh, M., and M. D. Sacchi, 2007, Multistep autoregressive reconstruction of seismic records: Geophysics, **72**, V111–V118.

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

Plan, Y., 2011, Compressed sensing, sparse approximation, and low-rank matrix estimation: PhD thesis, California Institute of Technology.

Porsani, M. J., 1999, Seismic trace interpolation using half-step prediction filters: Geophysics, **64**, 1461–1467.

Recht, B., M. Fazel, and P. A. Parrilo, 2010, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization: SIAM Review, **52**.

Rokhlin, V., A. Szlam, and M. Tygert, 2009, A randomized algorithm for principle component analysis: SIAM Journal on Matrix Analysis and Applications, **31**, 1100–1124.

Ronen, J., 1987, Wave-equation trace interpolation: Geophysics, **52**, 973–984.

Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: IEEE Transactions on Signal Processing, **46**, 31–38.

Shahidi, R., R. Tang, J. Ma, and F. J. Herrmann, 2013, Application of randomized sampling schemes to curvelet-based sparsity-promoting seismic data recovery: Geophysical Prospecting, **61**, 973–997.

Spitz, S., 1991, Seismic trace interpolation in the f-x domain: Geophysics, **56**, 785–794.

Sternfels, R., G. Viguier, R. Gondoin, and D. L. Meur, 2015, Multidimensional simultaneous random plus erratic noise attenuation and interpolation for seismic data by joint low-rank and sparse inversion: Geophysics, **80**, WD129–WD141.

Trad, D. O., T. J. Ulrych, and M. D. Sacchi, 2002, Accurate interpolation with high-resolution time-variant radon transforms: Geophysics, **67**, 644–656.

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

Wang, B., J. Li, and X. Chen, 2015a, A novel method for simultaneous seismic data interpolation and noise removal based on the l_0 constraint: Journal of Seismic Exploration, 187–204.

Wang, B., R. Wu, X. Chen, and J. Li, 2015b, Simultaneous seismic data interpolation and denoising with a new adaptive method based on dreamlet transform: Geophysical Journal International, **201**, 1180–1192.

Wang, J., M. Ng, and M. Perz, 2010a, Seismic data interpolation by greedy local radon transform: Geophysics, **80**, WB225–WB234.

———, 2010b, Seismic data interpolation by greedy local radon transform: Geophysics, **75**, WB225–WB234.

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.

Yu, Z., J. Ferguson, G. MeMechan, and P. Anno, 2007, Wavelet-radon domain dealiasing and interpolation of seismic data: Geophysics, **72**, V41–V49.