Robust reduced-rank filtering for erratic seismic noise attenuation

Ke Chen[1] and Mauricio D. Sacchi[1]

ABSTRACT

Singular spectrum analysis (SSA) or Cadzow reduced-rank filtering is an efficient method for random noise attenuation. SSA starts by embedding the seismic data into a Hankel matrix. Rank reduction of this Hankel matrix followed by antidiagonal averaging is utilized to estimate an enhanced seismic signal. Rank reduction is often implemented via the singular value decomposition (SVD). The SVD is a nonrobust matrix factorization technique that leads to suboptimal results when the seismic data are contaminated by erratic noise. The term erratic noise designates non-Gaussian noise that consists of large isolated events with known or unknown distribution. We adopted a robust low-rank factorization that permitted use of the SSA filter in situations in which the data were contaminated by erratic noise. In our robust SSA method, we replaced the quadratic error criterion function that yielded the truncated SVD solution by a bisquare function. The Hankel matrix was then approximated by the product of two lower dimensional factor matrices. The iteratively reweighed least-squares method was used to approximately solve for the optimal robust factorization. Our algorithm was tested with synthetic and real data. In our synthetic examples, the data were contaminated with band-limited Gaussian noise and erratic noise. Then, denoising was carried out by means of f-x deconvolution, the classical SSA method, and the proposed robust SSA method. The f-x deconvolution and the classical SSA method failed to properly eliminate the noise and to preserve the desired signal. On the other hand, the robust SSA method was found to be immune to erratic noise and was able to preserve the desired signal. We also tested the robust SSA method with a data set from the Western Canadian Sedimentary Basin. The results with this data set revealed improved denoising performance in portions of data contaminated with erratic noise.

INTRODUCTION

The improvement of the signal-to-noise ratio (S/N) of seismic records is an important topic in seismic data processing. Incoherent noise attenuation can be carried out via prediction error filters in the f-x (Canales, 1984) and t-x (Abma and Claerbout, 1995) domains. Incoherent noise attenuation can also be implemented via rank reduction methods. Rank reduction methods can be grouped into different categories. For instance, eigenimage filtering (Ulrych et al., 1988), similar to filtering via the Karhunen-Loève transform (Jones and Levy, 1987), can operate directly on the seismic data in the t-x or f-x-y domain (Trickett, 2003). Recently, the singular spectrum analysis (SSA) method (Sacchi, 2009; Oropeza and Sacchi, 2011), also known as Cadzow filtering (Trickett, 2008; Trickett and Burroughs, 2009), was introduced to attenuate incoherent noise and for seismic data reconstruction (Oropeza and Sacchi, 2011; Gao et al., 2013). It is also important to note that reduced-rank filtering based on SSA has been also used to suppress coherent noise (Nagarajappa, 2012; Chiu, 2013).

SSA operates in the frequency-space domain (f-x) by embedding spatial data at a given monochromatic temporal frequency into a Hankel matrix. Then, the ideal Hankel matrix that one would have formed in the absence of noise can be estimated from the low-rank approximation of the Hankel matrix of the noisy observations (Oropeza and Sacchi, 2011). SSA reconstruction can also be applied to multidimensional seismic data by forming multilevel Hankel matrices. In this case, rank reduction of large multilevel Hankel matrices is carried out via Lanczos bidiagonalization (Gao et al., 2013) or via fast randomized singular value decomposition (SVD) (Oropeza and Sacchi, 2011).

Our paper proposes a robust SSA method based on robust matrix rank reduction for removing Gaussian and erratic noise (Chen and Sacchi, 2013). The rank reduction in the newly proposed algorithm is implemented via robust low-rank matrix factorization (Gabriel

[1]University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: ke7@ualberta.ca; msacchi@ualberta.ca.

and Zamir, 1979; De la Torre and Black, 2003; Maronna and Yohai, 2008). The Hankel matrix of the data is decomposed into the product of two low-dimensional factor matrices. The bisquare function is used to obtain a robust metric to approximate the Hankel matrix by one of lower rank. The iteratively reweighted least-squares (IRLS) method (De la Torre and Black, 2003) is used to factorize the original Hankel matrix in terms of the two low-rank matrices.

Finally, it is important to mention that our algorithm complements the work of Trickett et al. (2012) on robust matrix rank reduction for denoising. However, Trickett et al. (2012) adopt an imputation algorithm to downweight erratic errors that adopts the nonrobust SVD matrix factorization. Our method, on the other hand, estimates a robust low-rank matrix factorization that replaces the nonrobust SVD solution.

THEORY

Singular spectrum analysis

This section provides a short review of the basic idea of the SSA method, also called Cadzow filtering. Details pertaining to the implementation of SSA for seismic noise attenuation and seismic data reconstruction can be found in Oropeza and Sacchi (2011). We discuss the 2D (t-x) implementation of SSA. However, we stress that SSA for 3D and 5D volumes has been extensively discussed in Oropeza and Sacchi (2011) and Gao et al. (2013), respectively.

Seismic data in a small window can be represented in the frequency-space domain via the superposition of plane waves

$$D_j(\omega) = \sum_{k=1}^{K} A_k(\omega) e^{-i\omega P_k(j-1)\Delta x}, \qquad (1)$$

where $i = \sqrt{-1}$, $j = 1, 2, \ldots, N$ is the trace index in the spatial axis and ω represents temporal frequency. In this equation, we assume that the data are composed of K linear events with distinct ray parameters P_k. Here, $A_k(\omega)$ denotes the complex amplitude of the kth plane wave and Δx indicates the spatial interval between seismograms. The SSA method constructs a trajectory matrix by embedding signal at one frequency $\mathbf{D}(\omega) = (D_1(\omega), D_2(\omega), \cdots, D_N(\omega))^T$ in the following Hankel matrix:

$$\mathbf{M}(\omega) = \mathcal{H}[\mathbf{D}(\omega)]$$
$$= \begin{pmatrix} D_1(\omega) & D_2(\omega) & \cdots & D_{N-L+1}(\omega) \\ D_2(\omega) & D_3(\omega) & \cdots & D_{N-L+2}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ D_L(\omega) & D_{L+1}(\omega) & \cdots & D_N(\omega) \end{pmatrix}, \quad (2)$$

where $\mathcal{H}$ is used to indicate the Hankel operator. For convenience, we choose $L = \lfloor \frac{N}{2} \rfloor + 1$ to make the Hankel matrix approximately square (Trickett, 2008; Oropeza and Sacchi, 2011), $\mathbf{M}(\omega) \in \mathbb{C}^{L \times (N-L+1)}$. We will omit the symbol ω and understand that the analysis is carried out for all frequencies. For a superposition of K plane waves, one can show that rank($\mathbf{M}$) $= K$ (Hua, 1992; Yang and Hua, 1996). Additive noise in $\mathbf{D}$ will increase the rank of matrix $\mathbf{M}$. One way of attenuating the additive noise is via rank reduction. The SSA filter can be represented via the following expression:

$$\hat{\mathbf{D}} = \mathcal{A}[\mathcal{R}_K[\mathcal{H}[\mathbf{D}]]], \qquad (3)$$

where $\mathcal{A}$ is the antidiagonal averaging operator, $\mathcal{R}_K[\mathbf{M}]$ is the rank reduction operator that approximates $\mathbf{M}$ by a rank-K matrix, and $\mathcal{H}$ is the Hankel operator. The operator $\mathcal{A}$ transforms back a Hankel form into a vector by averaging across antidiagonals. It is important to stress that a similar analysis is valid for multidimensional signals in which one adopts block Hankel matrices and block antidiagonal averaging operators (Gao et al., 2013). The rank reduction step ($\mathcal{R}_K$) can be implemented via the truncated SVD (Sacchi, 2009), the randomized SVD (Oropeza and Sacchi, 2011), or by a fast algorithm that adopts the Lanczos bidiagonalization method and highly efficient matrix-vector multiplications implemented via the fast Fourier transform (Gao et al., 2013).

The ℓ_2 low-rank approximation

The rank K approximation of the matrix $\mathbf{M}$ can be found by solving the following problem

$$\mathbf{M}_K = \mathcal{R}_K[\mathbf{M}] = \underset{\hat{\mathbf{M}}}{\mathrm{argmin}} \|\mathbf{M} - \hat{\mathbf{M}}\|_F^2$$
$$\text{subject to rank}(\hat{\mathbf{M}}) = K, \qquad (4)$$

where $\| \cdot \|_F$ is the Frobenius norm, $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij}|^2}$ of the matrix $\mathbf{X} \in \mathbb{C}^{m \times n}$. The problem in expression 4 has an unique local minimum that is global (Srebro and Jaakkola, 2003). This solution has analytic expression and is given by truncated singular value decomposition (TSVD)

$$\mathbf{M}_K = \mathcal{R}_K[\mathbf{M}] = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^H = \mathbf{U}_K \mathbf{U}_K^H \mathbf{M}, \qquad (5)$$

where $\mathbf{U}_K \in \mathbb{C}^{m \times K}$ and $\mathbf{V}_K \in \mathbb{C}^{n \times K}$ are matrices containing singular vectors associated to the K-largest singular values $\sigma_q, q = 1 \ldots K$. These singular values are also the diagonal elements of the matrix $\mathbf{\Sigma}_K \in \mathbb{R}^{K \times K}$. The latter is also known as the Eckart-Young theorem (Eckart and Young, 1936). Even though the solution of problem 4 is simple, the quadratic misfit functional makes the solution quite sensitive to non-Gaussian noise (De la Torre and Black, 2003; Cabral et al., 2013; Meng and De la Torre, 2013). This drawback could limit the application of the SSA method in situations in which the data contain outliers. In this article, we investigate a robust measure of distance between the matrices $\mathbf{M}$ and $\hat{\mathbf{M}}$ and an algorithm to estimate a low-rank approximation under the new distance.

Robust low-rank approximation

We replace the ℓ_2 distance between two matrices by a norm $\|.\|_\rho$ (the Frobenius norm in the preceding section) and present a new problem similar to expression 4 (De la Torre and Black, 2003). In other words, we would like to find a matrix $\hat{\mathbf{M}}$ of rank K that minimizes $\|\mathbf{M} - \hat{\mathbf{M}}\|_\rho$, expressed as

$$\mathbf{M}_K = \mathcal{R}_K[\mathbf{M}] = \underset{\hat{\mathbf{M}}}{\mathrm{argmin}} \|\mathbf{M} - \hat{\mathbf{M}}\|_\rho$$
$$\text{subject to rank}(\hat{\mathbf{M}}) = K, \qquad (6)$$

where $\|\mathbf{M} - \hat{\mathbf{M}}\|_\rho = \sum_{i=1}^{m} \sum_{j=1}^{n} \rho\left(\frac{m_{ij} - \hat{m}_{ij}}{\sigma}\right)$. Here, m_{ij} indicates one element of $\mathbf{M}$ and σ is a scale parameter for the function ρ. We use the same scale parameter σ for all the elements in the residual matrix because the noise in $\mathbf{D}$ is assumed to be independent

and identically distributed. When ρ is a nonquadratic function, the minimization of expression 6 is a nonconvex optimization problem. There is no closed-form solution for this problem. It is only when we adopt an ℓ_2 norm that the solution has a closed-form that is given by the TSVD. We have used synthetic simulations to try different distances for robust estimation and concluded that our best results were obtained with Tukey's bisquare function (Beaton and Tukey, 1974). Tukey's bisquare function is given by the following expression

$$\rho(x) = \begin{cases} \frac{1}{6}\alpha^2\left\{1 - \left[1 - \left(\frac{|x|}{\alpha}\right)^2\right]^3\right\} & |x| \le \alpha, \\ \frac{1}{6}\alpha^2 & |x| > \alpha \end{cases} \quad (7)$$

where α is a tunable parameter. In Figure 1, we provide a comparison of Tukey's bisquare functional ρ versus the ℓ_2 norm used by the classical rank reduction method that yields the TSVD solution. For the ℓ_2 norm (in our case, the Frobenious norm), ρ is given by $\rho(x) = \frac{1}{2}|x|^2$. In both cases, x is a normalized residual, $x = r/\sigma$, where σ is the scale parameter.

The low-rank approximation problem (equation 6) can be solved by representing the unknown matrix via a factorization $\hat{\mathbf{M}} = \mathbf{U}\mathbf{V}^H$, where $\mathbf{U} \in \mathbb{C}^{m\times K}$ and $\mathbf{V} \in \mathbb{C}^{n\times K}$ (Gabriel and Zamir, 1979) and solving

$$\begin{aligned}(\mathbf{U}_K, \mathbf{V}_K) &= \underset{\mathbf{U},\mathbf{V}}{\text{argmin}}\, E(\mathbf{U}, \mathbf{V}) \\ &= \underset{\mathbf{U},\mathbf{V}}{\text{argmin}}\|\mathbf{M} - \mathbf{U}\mathbf{V}^H\|_\rho. \end{aligned} \quad (8)$$

Equation 8 is a bilinear regression problem with two unknowns: $\mathbf{U}$ and $\mathbf{V}$. This approach of low-rank matrix decomposition is referred to as bilinear factorization (Cabral et al., 2013). The basic idea behind the selection of the bisquare norm ρ (or of any robust norm) is that less weight is assigned to large gross errors, and therefore, the final solution is not severely affected by gross errors (Huber, 1981; Hampel et al., 1986; De la Torre and Black, 2003).

Minimization of $E(\mathbf{U}, \mathbf{V})$ via the iteratively reweighed least-squares method

We now describe a method to solve equation 8. In other words, one needs to minimize the cost function

$$\begin{aligned}E(\mathbf{U}, \mathbf{V}) &= \|\mathbf{M} - \mathbf{U}\mathbf{V}^H\|_\rho \\ &= \sum_{i=1}^m \sum_{j=1}^n \rho\left(\frac{m_{ij} - \sum_{q=1}^K u_{iq}v_{jq}^*}{\sigma}\right) \\ &= \sum_{i=1}^m \sum_{j=1}^n \rho\left(\frac{r_{ij}}{\sigma}\right), \end{aligned} \quad (9)$$

where r_{ij} are the residuals. By taking the derivative of scalar function E in equation 9 with respect to $\mathbf{U}^*$ and $\mathbf{V}^*$, respectively (details are given in Appendix A), we obtain the following equations

$$\sum_{j=1}^n w\left(\frac{r_{aj}}{\sigma}\right) r_{aj} v_{jb} = 0, \quad a = 1, \ldots, m, \quad b = 1, \ldots, K, \quad (10)$$

$$\sum_{i=1}^m w\left(\frac{r_{ic}}{\sigma}\right) r_{ic}^* u_{id} = 0, \quad c = 1, \ldots, n, \quad d = 1, \ldots, K, \quad (11)$$

where the weight function is given by $w(x) = \frac{\partial \rho(x)}{\partial |x|} \frac{1}{|x|}$ ($x = \frac{r}{\sigma}$). In the particular case of the bisquare function (equation 7), the weight function is given by

$$w(x) = \begin{cases} \left[1 - \left(\frac{|x|}{\alpha}\right)^2\right]^2 & |x| \le \alpha \\ 0 & |x| > \alpha \end{cases}. \quad (12)$$

The selection of α and σ is discussed in the "Parameter selection and initialization" section.

Note that equations 10 and 11 are a set of nonlinear equations, the weights w depend on the model parameters, and the model parameters conversely depend on the weights. This nonlinear problem can be approximately solved by the IRLS approach (Holland and Welsch, 1977) in which the weights and the model parameters are alternately updated. In each IRLS iteration, for a given σ, the weights w can be calculated from the residuals r_{ij} from the previous IRLS iteration via equation 12. With this in mind, for an iteration of IRLS, the cost function (equation 9) can be transformed into a weighted least-squares form and expressed via the following expression

$$\begin{aligned}E_W(\mathbf{U}, \mathbf{V}) &= \|\mathbf{W}^{\frac{1}{2}}\odot(\mathbf{M} - \mathbf{U}\mathbf{V}^H)\|_F^2 \\ &= \sum_{i=1}^m \sum_{j=1}^n w_{ij}\left|m_{ij} - \sum_{q=1}^K u_{iq}v_{jq}^*\right|^2, \end{aligned} \quad (13)$$
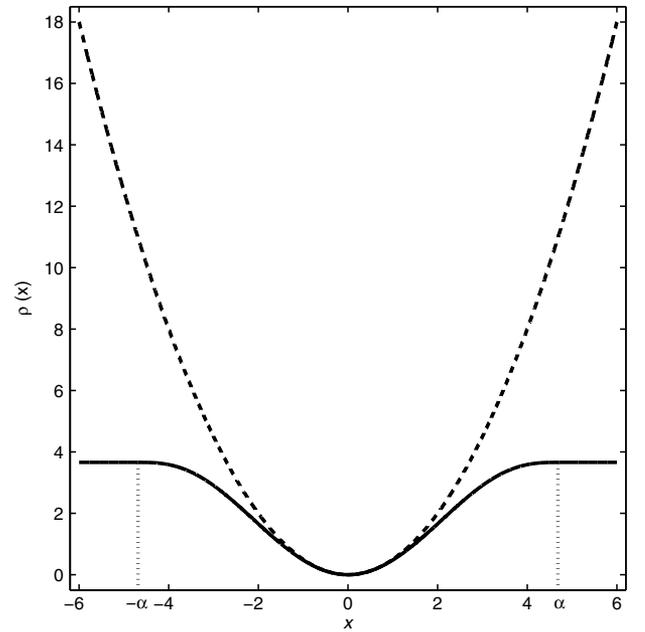
Figure 1. The dashed line is the quadratic function $\frac{1}{2}|x|^2$. The solid line is Tukey's bisquare robust metric (equation 7) adopted in this article for robust matrix factorization.

where the superscript $\frac{1}{2}$ indicates an elementwise square root of the matrix. The matrix $\mathbf{W} \in \mathbb{R}^{+m \times n}$ is the matrix of weights calculated from the residuals of the previous IRLS iteration. The symbol $\odot$ represents the Hadamard product (elementwise product). In each IRLS iteration, the model parameters $\mathbf{U}$ and $\mathbf{V}$ can be updated with fixed $\mathbf{W}$. The weighted cost function $E_W(\mathbf{U}, \mathbf{V})$ is a function of $\mathbf{U}$ and $\mathbf{V}$. It can be solved by the alternating minimization method (Gabriel and Zamir, 1979; Roweis, 1998; Tipping and Bishop, 1999; Niesen et al., 2009). In the alternating minimization method, when $\mathbf{U}$ is fixed, equation 13 is a linear regression problem in $\mathbf{V}$. Similarly, when $\mathbf{V}$ is fixed, equation 13 is a linear regression problem in $\mathbf{U}$. The alternating minimization algorithm can be expressed by alternately minimizing

$$
\begin{aligned}
E_W(\mathbf{V}) &= \|\mathbf{W}^{\frac{1}{2}} \odot (\mathbf{M} - \mathbf{U}\mathbf{V}^H)\|_F^2 \\
&= \sum_{j=1}^{n} (\mathbf{m}^j - \mathbf{U}\mathbf{v}_j)^H \mathbf{W}^j (\mathbf{m}^j - \mathbf{U}\mathbf{v}_j)
\end{aligned} \quad (14)
$$

$$
\begin{aligned}
E_W(\mathbf{U}) &= \|\mathbf{W}^{\frac{1}{2}} \odot (\mathbf{M} - \mathbf{U}\mathbf{V}^H)\|_F^2 \\
&= \sum_{i=1}^{m} (\mathbf{m}_i - \mathbf{V}\mathbf{u}_i)^H \mathbf{W}_i (\mathbf{m}_i - \mathbf{V}\mathbf{u}_i),
\end{aligned} \quad (15)
$$

where

$$
\mathbf{M} = (\,\mathbf{m}^1 \quad \mathbf{m}^2 \quad \cdots \quad \mathbf{m}^n\,) = (\,\mathbf{m}_1 \quad \mathbf{m}_2 \quad \cdots \quad \mathbf{m}_m\,)^H. \quad (16)
$$

Note that all of the vectors are column vectors. For instance, $\mathbf{m}^j$ is the jth column of $\mathbf{M}$. Similarly, $\mathbf{m}_i$ is the conjugate transpose of the ith row of $\mathbf{M}$. The latter leads to the following representation of factors $\mathbf{U}$ and $\mathbf{V}$

$$
\mathbf{U} = (\,\mathbf{u}^1 \quad \mathbf{u}^2 \quad \cdots \quad \mathbf{u}^K\,) = (\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m\,)^H \quad (17)
$$

$$
\mathbf{V} = (\,\mathbf{v}^1 \quad \mathbf{v}^2 \quad \cdots \quad \mathbf{v}^K\,) = (\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,)^H. \quad (18)
$$

The weights are given by the diagonal matrix $\mathbf{W}^j = \text{diag}\{\mathbf{w}^j\} \in \mathbb{R}^{+n \times n}$ that contains the jth column of $\mathbf{W}$. Similarly, the matrix $\mathbf{W}_i = \text{diag}\{\mathbf{w}_i\} \in \mathbb{R}^{+m \times m}$ is the diagonal weight matrix containing the ith row of $\mathbf{W}$. Equations 14 and 15 can be broken up into smaller weighted least-squares problems that alternately update rows of $\mathbf{U}$ and rows of $\mathbf{V}$. The alternating minimization algorithm in each IRLS iteration can be reexpressed as

$$
\text{for } i = 1, 2 ..., m \qquad \min_{\mathbf{u}_i}(\mathbf{m}_i - \mathbf{V}\mathbf{u}_i)^H \mathbf{W}_i (\mathbf{m}_i - \mathbf{V}\mathbf{u}_i) \quad (19)
$$

$$
\text{for } j = 1, 2, ..., n \qquad \min_{\mathbf{v}_j}(\mathbf{m}^j - \mathbf{U}\mathbf{v}_j)^H \mathbf{W}^j (\mathbf{m}^j - \mathbf{U}\mathbf{v}_j). \quad (20)
$$

The QR factorization (Golub and Van Loan, 1996) is used to solve the weighted least-squares minimization problems in equations 19 and 20. The order of operations of the robust low-rank factorization is given by $O(mnK^2 N_a N_{irls})$, where N_a and N_{irls} are the total number of iterations of the alternating minimization and of the IRLS, respectively. The non-robust solution via the truncated SVD needs $O(mn^2)$ operations.

Iterative algorithm for robust low-rank matrix factorization

The robust low-rank approximation algorithm can be summarized as follows:

1) Start with initial factors $\mathbf{U}$ and $\mathbf{V}$.
2) Select parameter α and fix it for the rest of the iterations.
3) Calculate the residual matrix $\mathbf{R} = \mathbf{M} - \mathbf{U}\mathbf{V}^H$.
4) Update the scale parameter σ.
5) Calculate the weight matrix $\mathbf{W}$ using equation 12.
6) Update the factor matrix $\mathbf{U}$ by solving problem 19 via the QR factorization.
7) Update the factor matrix $\mathbf{V}$ by solving problem 20 via the QR factorization.
8) Iterate steps (6)–(7) until convergence or a maximum iteration number is reached (alternating minimization).
9) Iterate steps (3)–(8) until convergence or a maximum iteration number is reached (IRLS).

Last, we point out that our algorithm requires two stopping criteria. One for the number of IRLS iterations and one for the number of updates of the alternating minimization scheme that is needed to estimate the factor matrices $\mathbf{U}$ and $\mathbf{V}$. In each case, we monitor the reduction of the cost function between consecutive iterations and stop either when the reduction has become smaller than a tolerance value or a maximum number of iterations is reached. In general, we have found that a practical strategy entails setting large tolerance values and a small number of maximum iterations for the IRLS and the alternating minimization steps. The latter gives us control on the expected computational cost of the whole denoising process. Our tests with synthetic data indicate that about 10 IRLS iterations and five updates of the factors $\mathbf{U}$ and $\mathbf{V}$ (alternating minimization) are sufficient to consistently achieve acceptable results. We use these parameters for all our synthetic and real data tests.

Parameter selection and initialization

We adopt the normalized median absolute deviation (MAD) as the robust scale parameter for the bisquare function (Holland and Welsch, 1977)

$$
\sigma = 1.4826 \text{ MAD} = 1.4826 \text{ median}|\mathbf{r} - \text{median}|\mathbf{r}||, \quad (21)
$$

where $\mathbf{r}$ is the residual vector obtained by reshaping the residual matrix $\mathbf{R}$ from the previous IRLS iteration. Holland and Welsch (1977) recommend to fix the robust scale σ during the iterations until the IRLS converges. In our paper, σ is updated in each iteration using equation 21. This scheme is stable according to our simulation tests. We also follow Holland and Welsch (1977) and Maronna et al. (2006) to choose the tuning parameter α. The value α in Tukey's bisquare function controls the so-called asymptotic efficiency of the M-estimate (in our case, the bisquare estimate). The asymptotic efficiency of an M-estimate is the ratio of the asymptotic variance of the maximum likelihood estimate and the asymptotic variance of the M-estimate. It reflects how efficient the M-estimate is for Gaussian data (Maronna et al., 2006). The value of the product $\alpha\sigma$ actually acts as the threshold to distinguish outliers and inliers.

Smaller values of $\alpha\sigma$ will add penalization to the outliers resulting in a more robust estimation.

One can adopt the least-squares solution obtained via the TSVD of the data as initial solution for the factor matrices. This will work if a small number of low-amplitude outliers are presented in the data. In general, starting with the least-squares solution can lead to solutions that are skewed by large outliers. The factor matrices $\mathbf{U}$ and $\mathbf{V}$ were initialized with the TSVD solution of an $m \times n$ random matrix to circumvent the aforementioned problem (Chen et al., 2008; Unkel and Trendafilov, 2010). This initialization strategy is used throughout this paper.

EXAMPLES

We present synthetic examples and one field data example to illustrate the proposed algorithm. We compare the performance of the robust SSA, the classical SSA method (Oropeza and Sacchi, 2011), and f-x deconvolution (Canales, 1984) for data that contain erratic and Gaussian noise.

Synthetic example

Figure 2b shows a 2D synthetic t-x data set. The data are composed of 40 traces with a total time of 1.2 s and a sampling interval of 0.004 s. The data are contaminated with band-limited (in time) Gaussian noise with a S/N equal to three and isolated noisy traces (erratic noise). The S/N is defined as the ratio between the maximum amplitude of the clean data and the standard deviation of the band-limited Gaussian noise. The amplitude of the two erratic traces

are two and three times of the maximum amplitude of the uncorrupted data. The wiggles have been clipped to allow for the visualization of the data in the presence of the large amplitude erratic noisy signals. The processing frequency band ranges from 1 to 40 Hz. We select the size of the subspace of the reconstructed data in SSA and for the proposed robust SSA methods to $K = 3$. We choose the number of external iterations (for updating weights) equal to 10 and the number of internal iterations (for alternating minimization) equal to five. The tuning constant α for the bisquare function is set to 4.685. The f-x deconvolution method was run with a prediction filter of length 10 traces and a trade-off parameter of 0.001. The results of f-x deconvolution, SSA, and robust SSA are compared. Figure 2a is the noise-free data, Figure 2b is the data contaminated with noise, and Figure 2c is the noise term. Figure 3a shows the result of f-x deconvolution. Large-amplitude noise leaks over several traces in the output section. A shorter prediction filter can remove more noise, but it will also increase the distortion of the filtered signal. We tested f-x deconvolution with a variety of parameters, and we have never managed to produce fully satisfying results when the data are contaminated by high-amplitude erratic noise. Figure 3b shows the result of the classical nonrobust SSA implemented via the TSVD. Again, we observe that the erratic noise has not been properly removed and noticeable artifacts are present in the output gather. The output of the robust SSA method is shown in Figure 3c. In this case, the Gaussian and erratic noise were successfully suppressed. By examining the difference sections (differences between original input data and filtered results) of the three methods (Figure 4), one observes an important amount of energy leaking in the difference section Figure 4a. However, the proposed robust SSA
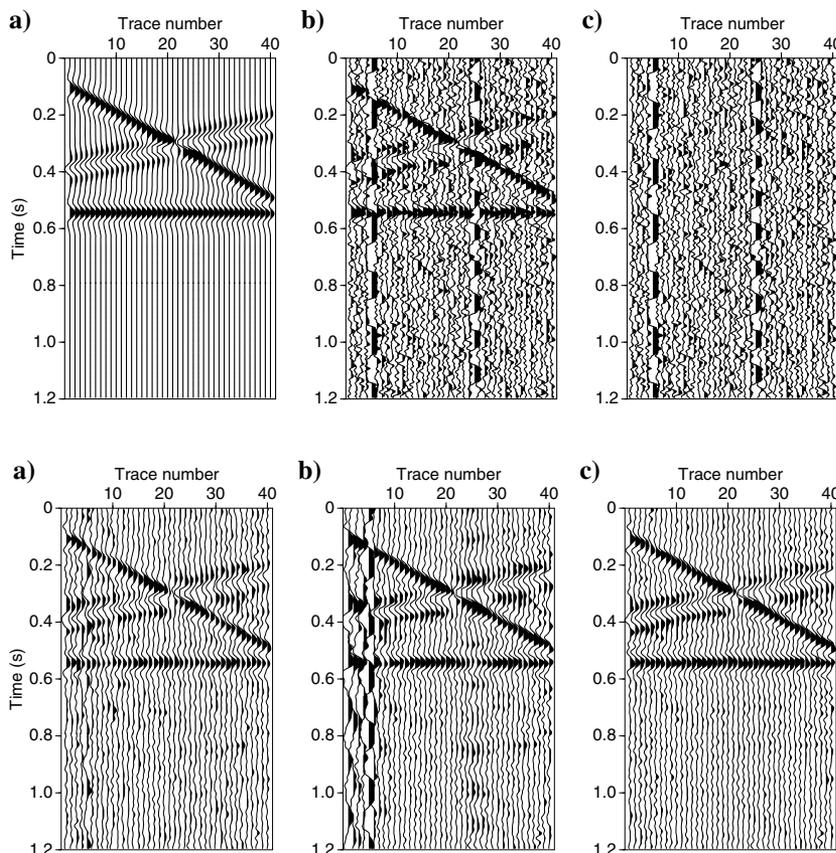
Figure 2. The synthetic data with three linear events. (a) Clean data. (b) Data with Gaussian noise and erratic noise. (c) The noise added to the data.

Figure 3. (a) Data in Figure 2b after f-x deconvolution filtering. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.

method can preserve the amplitudes of the noise-free original signal (Figure 4c). We also compared the result of the robust SSA method on data corrupted with erratic noise and Gaussian noise (Figure 3c) with the result of classical SSA on data corrupted by only Gaussian

noise (Figure 5b). Note that the Gaussian noise in Figure 5a is the same noise used in Figure 2b. The two results are quite similar to each other. We also evaluate the denoising performance by evaluating the factor $Q = 10 \log \frac{\|d^0\|_F^2}{\|d^0 - \tilde{d}\|_F^2}$, where d^0 is the noise-free data,

Figure 4. Difference sections of (a) f-x deconvolution, (b) SSA, and (c) robust SSA.
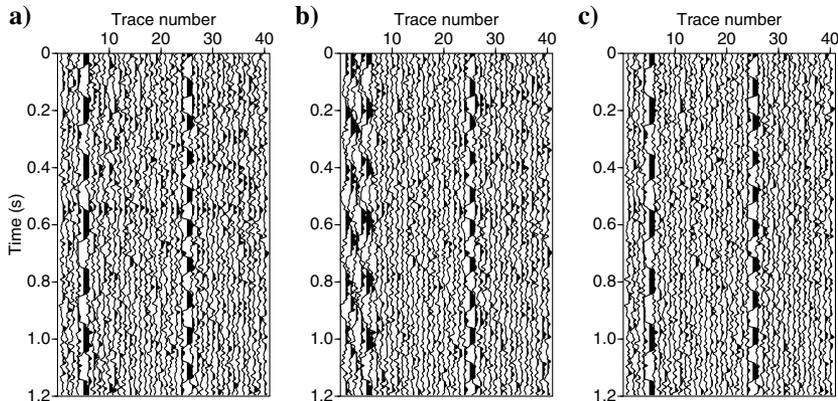
Figure 5. (a) Data corrupted with only Gaussian noise. (b) Data after classical SSA filtering. (c) Difference section of classical SSA.
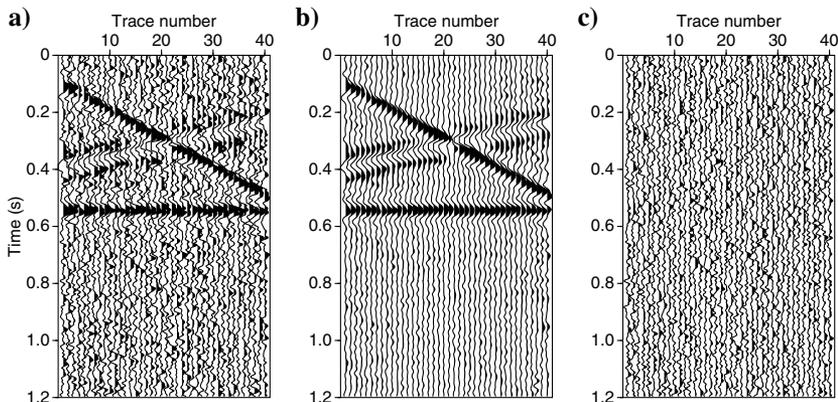
Figure 6. Poststack field data. (a) The whole data set. (b) The data in the left rectangular window. (c) The data in the right rectangular window. (CDP, common depth point.)
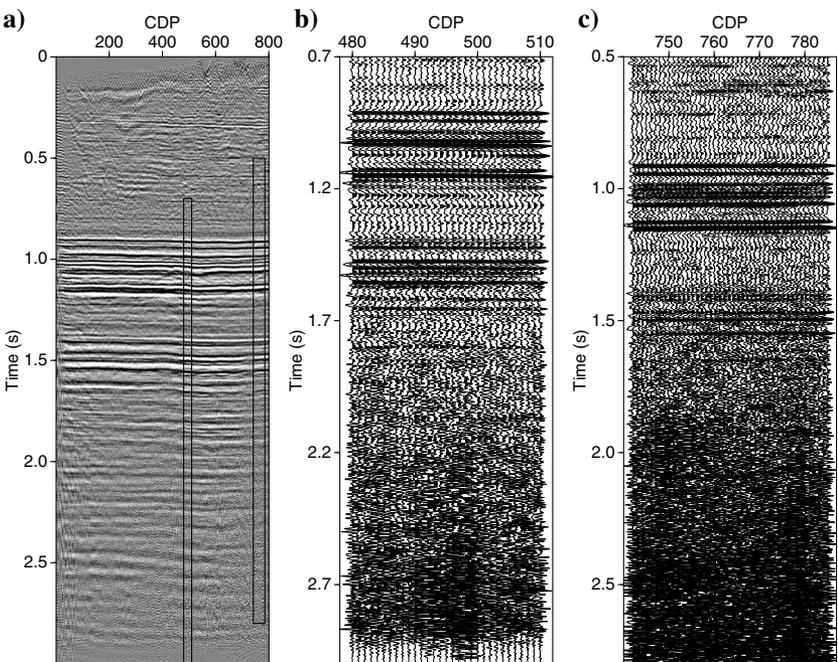
and $\hat{d}$ is the reconstructed data. A larger value of Q means better denoising performance. The Q-value of the f-x deconvolution is $Q_{fx} = 7.7$, the Q-value for SSA is $Q_{ssa} = -2.8$, and the Q-value of robust SSA is $Q_{rssa} = 12.8$. The Q-value of the classical SSA on data with only Gaussian noise (Figure 5) is $Q_{ssaG} = 13.1$. These values indicate that the robust SSA method offers a good alternative to SSA and f-x deconvolution when the data are contaminated by erratic noise.

Field data example

Figure 6a is a poststack data section from a survey in the Western Canadian Sedimentary Basin. It has 800 traces and 1500 time sam-

ples per trace with a time sampling interval of 2 ms. Figure 6b and 6c is magnified portions of data in the left and right rectangular windows highlighted in Figure 6a, respectively. We observe high-amplitude noise in this data set. The whole data are divided into overlapping windows with suitable size. Then, all windows are filtered and added back to recover the clean data. In the spatial direction, each window has 50 traces, and the overlap between two adjacent windows is 25 traces. In the temporal direction, each window has 300 samples (0.6 s), and the overlap between two adjacent windows is 100 samples (0.2 s). The three filtering methods are applied for frequencies in the band of 1–80 Hz. The size of the reconstructed subspace in the SSA and the robust SSA method is $K = 2$. In the robust SSA case, the external iterations (for updating
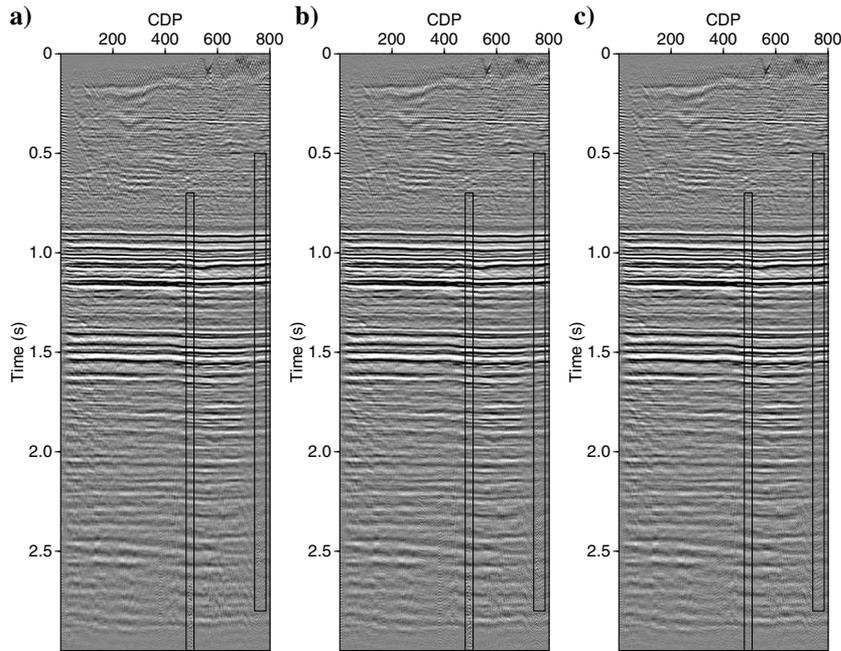
Figure 7. The comparison of the results of three different methods. (a) Data after f-x deconvolution filtering. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.
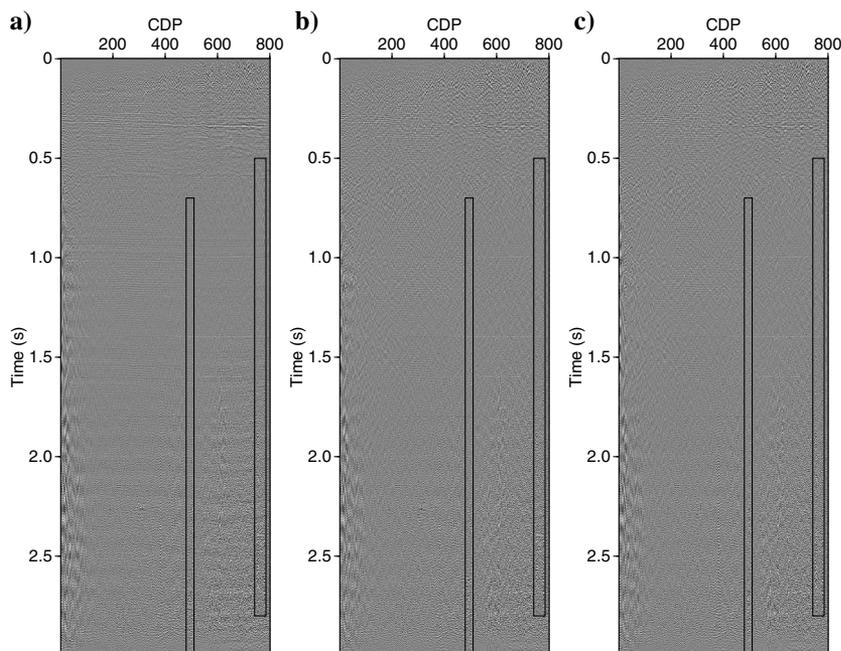
Figure 8. The comparison of difference sections of three different methods. Difference sections of (a) f-x deconvolution, (b) SSA, and (c) robust SSA.

the weights) is set to 10 and the number of internal iterations (for the alternating minimization) is set to five. The tuning constant α for the bisquare function is set to 3.3. For the f-x filter, we set the length of the operator equal to six traces and the trade-off parameter to 0.001. We use the same parameters for the whole data set. Again, we compare the performance of f-x deconvolution, SSA, and robust SSA on these data. To provide a fair comparison, the results for the three methods (Figures 6a, 7, and 8) have been clipped to the same value. The wiggle plots corresponding to the left rectangular window (Figures 6b and 9) have been clipped to the same value. The difference

sections (Figure 10) have been clipped to the same value as well to better compare the estimated noise by the three methods. The results of the three methods applied on the whole data set are shown in Figure 7. The proposed robust SSA method suppresses much more high-amplitude erratic noise than the f-x deconvolution method and the classical SSA algorithm. The comparison of difference sections (Figure 8) shows that the f-x deconvolution leaks more signal energy into the noise section than the robust SSA. We show the zoomed results for the window on the left of Figure 6a in Figure 9. The results for the window to the right of Figure 6a are shown in

Figure 9. The comparison of results of the data in the left rectangular window by three different methods. (a) Data after f-x deconvolution filtering. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.
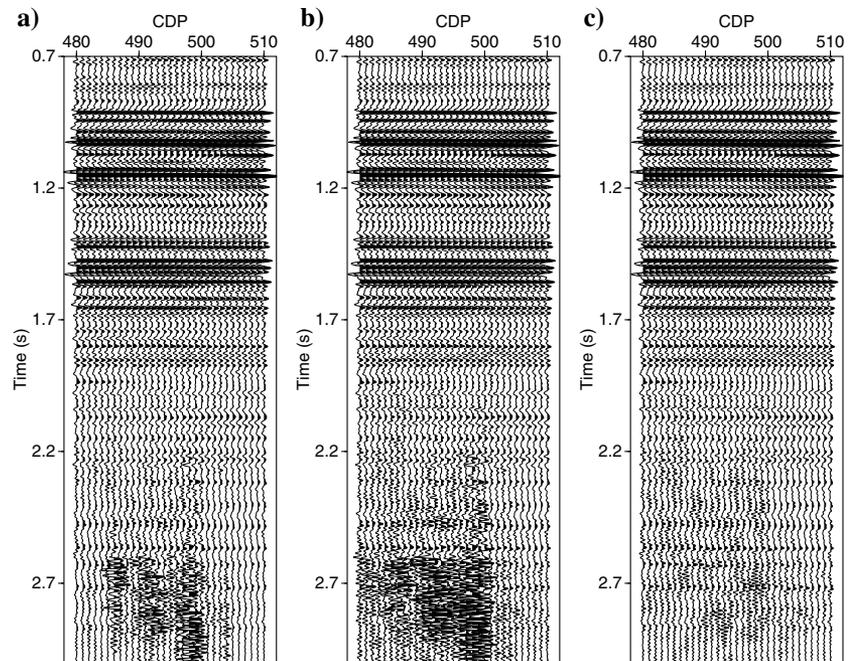
Figure 10. The comparison of difference sections of three different methods in the left rectangular window. Difference sections of (a) f-x deconvolution, (b) SSA, and (c) robust SSA.
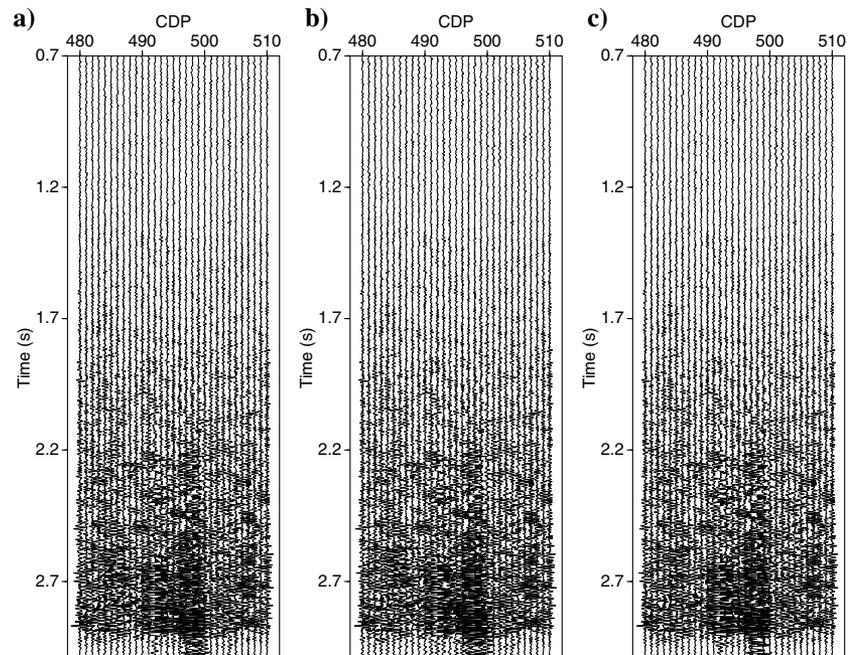
Figure 11. Similarly, difference sections are provided by Figure 12. Again, we note that the robust SSA method is more effective than f-x deconvolution and the classical SSA algorithm.

DISCUSSION

The rank K should be equal to the number of dips in the window of analysis (Oropeza and Sacchi, 2011). However, real data are not composed of a superposition of a limited number of perfect linear events. One can say that parameter K depends on the complexity of the data. For real data scenarios, parameter K can be determined via trial and error by examining the complexity of the input data and by visualizing the resulting enhanced section and the estimated noise section. The choice of the rank controls the balance between denoising performance and acceptable preservation of amplitudes in the denoised data. A small rank is used if data are structurally simple. A relatively large rank should be adopted if the data contain a large number of space-variant dips. In our algorithm, the user must specify the rank K. An alternative is to adopt a strategy that adopts nuclear norm optimization (Recht et al., 2010). The latter has been recently
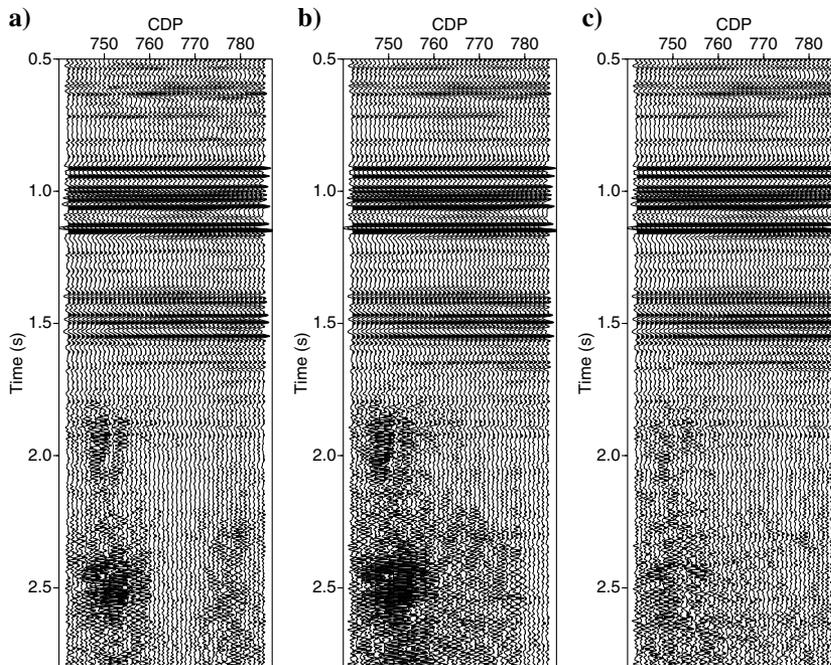
Figure 11. The comparison of results of the data in the right rectangular window by three different methods. (a) Data after f-x deconvolution filtering. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.
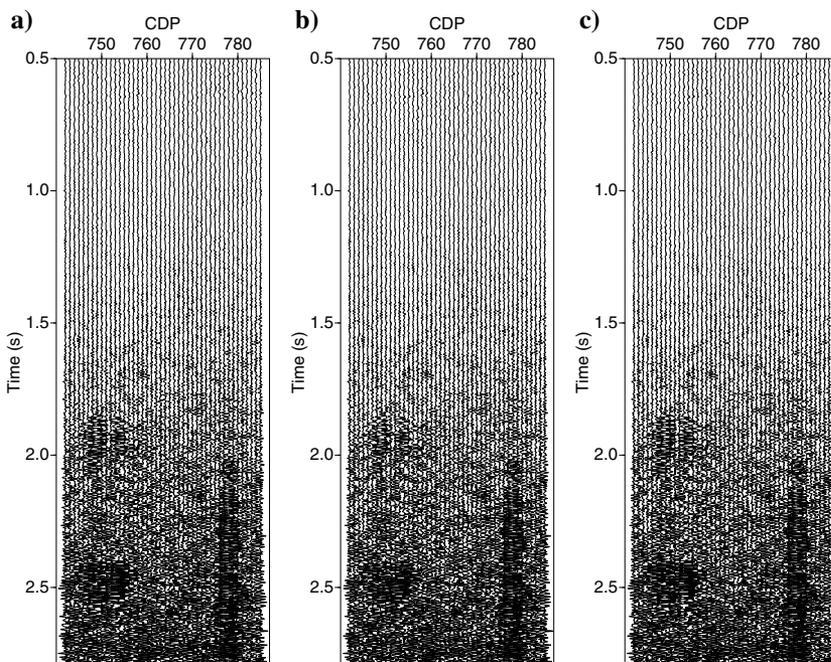
Figure 12. The comparison of difference sections of three different methods in the right rectangular window. Difference sections of (a) f-x deconvolution, (b) SSA, and (c) robust SSA.

explored for matrix and tensor completion in applied mathematics (Candès et al., 2011; Gandy et al., 2011; Tao and Yuan, 2011) and in seismic data processing (Kreimer et al., 2013). However, a formulation of our problem in terms of nuclear norm optimization will require a new scalar trade-off parameter that controls noise rejection versus amplitude preservation in the denoised signal. In other words, we will be replacing rank (K) by a scalar and, again, the complexity of the data will dictate the selection of the scalar parameter.

CONCLUSIONS

We have proposed a robust version of singular spectrum analysis that can remove Gaussian and non-Gaussian noise simultaneously. The robust low-rank approximation was adopted instead of the truncated SVD. The robust low-rank approximation used a robust measure of misfit (bisquare function) and the bilinear factorization model to approximate the Hankel matrix of the data by one of lower rank. The IRLS method is used to solve the optimization problem. Synthetic and real data examples show that the proposed algorithm can cope with erratic noise. One possible concern is the computational cost of the proposed robust algorithm. In general, the robust SSA method can be more expensive to run than the classical SSA method based on the truncated SVD depending on the size of the data, the number of iterations of the alternating minimization and IRLS. This could be problematic for industrial processes. However, there could be situations in which having access to robust denoising methods might offer benefits that surpass computational considerations. The latter is a facet that is always present in seismic data processing.

ACKNOWLEDGMENTS

APPENDIX A

DERIVATIVE OF SCALAR FUNCTION WITH RESPECT TO COMPLEX-VALUED MATRIX VARIABLE

We describe how to use complex differentiation theory to compute the weight function from the cost function 9

$$E(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \rho \left(\frac{m_{ij} - \sum_{q=1}^{K} u_{iq} v_{jq}^*}{\sigma} \right)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \rho \left(\frac{r_{ij}}{\sigma} \right). \tag{A-1}$$

If complex-valued matrix variable $\mathbf{V}$ is fixed, scalar E is a function of $\mathbf{U}$. Similarly, if $\mathbf{U}$ is fixed, E is a function of $\mathbf{V}$. We will first analyze the derivative of E with respect to $\mathbf{U}$ with $\mathbf{V}$ fixed. By using Wirtinger calculus (Brandwood, 1983), we should regard $\mathbf{U}$ and $\mathbf{U}^*$ as independent variables ($E(\mathbf{U}, \mathbf{U}^*)$). Either setting the gradient $\frac{\partial E}{\partial \mathbf{U}}$ or $\frac{\partial E}{\partial \mathbf{U}^*}$ to zero leads to stationary points of function E. Usually, $\frac{\partial E}{\partial \mathbf{U}^*}$ is preferred because it gives the direction in which the cost function E

has the maximum rate of change with respect to $\mathbf{U}$ (Brandwood, 1983). The partial derivative of E with respect to one element of $\mathbf{U}^*$ is given by

$$\frac{\partial E}{\partial u_{ab}^*} = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial \rho \left(\frac{r_{ij}}{\sigma}, \frac{r_{ij}^*}{\sigma} \right)}{\partial u_{ab}^*} = \sum_{j=1}^{n} \frac{\partial \rho \left(\frac{r_{aj}}{\sigma}, \frac{r_{aj}^*}{\sigma} \right)}{\partial u_{ab}^*}$$

$$= \sum_{j=1}^{n} \left(\frac{\partial \rho \left(\frac{r_{aj}}{\sigma}, \frac{r_{aj}^*}{\sigma} \right)}{\partial r_{aj}} \frac{\partial r_{aj}}{\partial u_{ab}^*} + \frac{\partial \rho \left(\frac{r_{aj}}{\sigma}, \frac{r_{aj}^*}{\sigma} \right)}{\partial r_{aj}^*} \frac{\partial r_{aj}^*}{\partial u_{ab}^*} \right)$$

$$= \sum_{j=1}^{n} \frac{\partial \rho \left(\frac{r_{aj}}{\sigma}, \frac{r_{aj}^*}{\sigma} \right)}{\partial r_{aj}^*} \frac{\partial r_{aj}^*}{\partial u_{ab}^*} = - \sum_{j=1}^{n} \frac{\partial \rho \left(\frac{r_{aj}}{\sigma}, \frac{r_{aj}^*}{\sigma} \right)}{\partial r_{aj}^*} v_{jb}$$

$$= - \sum_{j=1}^{n} \frac{\partial \rho \left(\frac{r_{aj}}{\sigma}, \frac{r_{aj}^*}{\sigma} \right)}{\partial r_{aj}^*} \frac{1}{r_{aj}} r_{aj} v_{jb}$$

$$= - \frac{1}{2\sigma^2} \sum_{j=1}^{n} w \left(\frac{r_{aj}}{\sigma} \right) r_{aj} v_{jb}, \tag{A-2}$$

where $w(x) = \frac{\partial \rho(x,x^*)}{\partial x^*} \frac{2}{x}$ ($x = \frac{r_{aj}}{\sigma}$) is the weight function, $a = 1, \ldots, m$, and $b = 1, \ldots, K$. Function ρ is a function of complex variable r_{aj}, but it is not analytic in r_{aj}. So, ρ is assumed to be a function of two independent variables r_{aj} and r_{aj}^*. Among them, r_{aj}^* depends on u_{ab}^*. So, the chain rule is used in the above derivation. Due to the relationship $\frac{\partial |x|}{\partial x^*} = \frac{1}{2} \frac{x}{|x|}$, we have that $w(x) = \frac{\partial \rho(x,x^*)}{\partial x^*} \frac{2}{x} = \frac{\partial \rho(x,x^*)}{\partial x} \frac{2}{x^*} = \frac{\partial \rho(x)}{\partial |x|} \frac{1}{|x|}$.

Similarly, the partial derivative of E with respect to one element of $\mathbf{V}^*$ is

$$\frac{\partial E}{\partial v_{cd}^*} = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial \rho \left(\frac{r_{ij}}{\sigma}, \frac{r_{ij}^*}{\sigma} \right)}{\partial v_{cd}^*} = \sum_{i=1}^{m} \frac{\partial \rho \left(\frac{r_{ic}}{\sigma}, \frac{r_{ic}^*}{\sigma} \right)}{\partial v_{cd}^*}$$

$$= \sum_{i=1}^{m} \left(\frac{\partial \rho \left(\frac{r_{ic}}{\sigma}, \frac{r_{ic}^*}{\sigma} \right)}{\partial r_{ic}} \frac{\partial r_{ic}}{\partial v_{cd}^*} + \frac{\partial \rho \left(\frac{r_{ic}}{\sigma}, \frac{r_{ic}^*}{\sigma} \right)}{\partial r_{ic}^*} \frac{\partial r_{ic}^*}{\partial v_{cd}^*} \right)$$

$$= \sum_{i=1}^{m} \frac{\partial \rho \left(\frac{r_{ic}}{\sigma}, \frac{r_{ic}^*}{\sigma} \right)}{\partial r_{ic}} \frac{\partial r_{ic}}{\partial v_{cd}^*} = - \sum_{i=1}^{m} \frac{\partial \rho \left(\frac{r_{ic}}{\sigma}, \frac{r_{ic}^*}{\sigma} \right)}{\partial r_{ic}} u_{id}$$

$$= - \sum_{i=1}^{m} \frac{\partial \rho \left(\frac{r_{ic}}{\sigma}, \frac{r_{ic}^*}{\sigma} \right)}{\partial r_{ic}} \frac{1}{r_{ic}^*} r_{ic}^* u_{id}$$

$$= - \frac{1}{2\sigma^2} \sum_{i=1}^{m} w \left(\frac{r_{ic}}{\sigma} \right) r_{ic}^* u_{id}, \tag{A-3}$$

where $w(x) = \frac{\partial \rho(x,x^*)}{\partial x} \frac{2}{x^*} = \frac{\partial \rho(x,x^*)}{\partial x^*} \frac{2}{x} = \frac{\partial \rho(x)}{\partial |x|} \frac{1}{|x|}$ ($x = \frac{r_{ic}}{\sigma}$), $c = 1, \ldots, n$, and $d = 1, \ldots, K$.

REFERENCES

Abma, R., and J. Claerbout, 1995, Lateral prediction for noise attenuation by t-x and f-x techniques: Geophysics, **60**, 1887–1896, doi: 10.1190/1.1443920.

Beaton, A. E., and J. W. Tukey, 1974, The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data: Technometrics, **16**, 147–185, doi: 10.1080/00401706.1974.10489171.

Brandwood, D. H., 1983, A complex gradient operator and its application in adaptive array theory: IEE Proceedings F: Communications, Radar and Signal Processing, **130**, 11–16, doi: 10.1049/ip-f-1.1983.0003.

Cabral, R., F. De la Torre, J. Costeira, and A. Bernardino, 2013, Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition: Presented at IEEE International Conference on Computer Vision, 2488–2495.

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

Candès, E. J., X. Li, Y. Ma, and J. Wright, 2011, Robust principal component analysis?: Journal of the ACM, **58**, 1–37, doi: 10.1145/1970392.1970395.

Chen, C., X. He, and Y. Wei, 2008, Lower rank approximation of matrices based on fast and robust alternating regression: Journal of Computational and Graphical Statistics, **17**, 186–200, doi: 10.1198/106186008X286806.

Chen, K., and M. D. Sacchi, 2013, Robust reduced-rank seismic denoising: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4272–4277.

Chiu, S. K., 2013, Coherent and random noise attenuation via multichannel singular spectrum analysis in the randomized domain: Geophysical Prospecting, **61**, 1–9, doi: 10.1111/j.1365-2478.2012.01090.x.

De la Torre, F., and M. J. Black, 2003, A framework for robust subspace learning: International Journal of Computer Vision, **54**, 117–142, doi: 10.1023/A:1023709501986.

Eckart, C., and G. Young, 1936, The approximation of one matrix by another of lower rank: Psychometrika, **1**, 211–218, doi: 10.1007/BF02288367.

Gabriel, K. R., and S. Zamir, 1979, Lower rank approximation of matrices by least squares with any choice of weights: Technometrics, **21**, 489–498, doi: 10.1080/00401706.1979.10489819.

Gandy, S., B. Recht, and I. Yamada, 2011, Tensor completion and low-n-rank tensor recovery via convex optimization: Inverse Problems, **27**, 025010, doi: 10.1088/0266-5611/27/2/025010.

Gao, J., M. D. Sacchi, and X. Chen, 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: Geophysics, **78**, no. 1, V21–V30, doi: 10.1190/geo2012-0038.1.

Golub, G., and C. Van Loan, 1996, Matrix computations, 3rd ed.: Johns Hopkins University Press.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, 1986, Robust statistics: The approach based on influence functions: John Wiley & Sons.

Holland, P. W., and R. E. Welsch, 1977, Robust regression using iteratively reweighted least-squares: Communications in Statistics — Theory and Methods, **6**, 813–827, doi: 10.1080/03610927708827533.

Hua, Y., 1992, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil: IEEE Transactions on Signal Processing, **40**, 2267–2280, doi: 10.1109/78.157226.

Huber, P. J., 1981, Robust statistics: John Wiley & Sons.

Jones, I., and S. Levy, 1987, Signal-to-noise ratio enhancement in multichannel seismic data via the Karhunen-Loève transform: Geophysical Prospecting, **35**, 12–32, doi: 10.1111/j.1365-2478.1987.tb00800.x.

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, no. 6, V273–V284, doi: 10.1190/geo2013-0022.1.

Maronna, R. A., D. R. Martin, and V. J. Yohai, 2006, Robust statistics: Theory and methods: John Wiley & Sons.

Maronna, R. A., and V. J. Yohai, 2008, Robust low-rank approximation of data matrices with elementwise contamination: Technometrics, **50**, 295–304, doi: 10.1198/004017008000000190.

Meng, D., and F. De la Torre, 2013, Robust matrix factorization with unknown noise: Presented at The IEEE International Conference on Computer Vision, 1337–1344.

Nagarajappa, N., 2012, Coherent noise estimation by adaptive Hankel matrix rank reduction: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0063.1.

Niesen, U., D. Shah, and G. W. Wornell, 2009, Adaptive alternating minimization algorithms: IEEE Transactions on Information Theory, **55**, 1423–1429, doi: 10.1109/TIT.2008.2011442.

Oropeza, V., and M. D. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, V25–V32, doi: 10.1190/1.3552706.

Recht, B., M. Fazel, and P. A. Parrilo, 2010, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization: SIAM Review, **52**, 471–501, doi: 10.1137/070697835.

Roweis, S. T., 1998, EM algorithms for PCA and SPCA, *in* M. I. Jordan, M. J. Kearns, and S. A. Solla, eds., Advances in neural information processing systems, **10**, MIT Press, 626–632.

Sacchi, M. D., 2009, FX singular spectrum analysis: Presented at CSPG CSEG CWLS Convention.

Srebro, N., and T. Jaakkola, 2003, Weighted low-rank approximations: Presented at 20th International Conference on Machine Learning, 720–727.

Tao, M., and X. Yuan, 2011, Recovering low-rank and sparse components of matrices from incomplete and noisy observations: SIAM Journal on Optimization, **21**, 57–81, doi: 10.1137/100781894.

Tipping, M. E., and C. M. Bishop, 1999, Probabilistic principal component analysis: Journal of the Royal Statistical Society Series B, **61**, 611–622, doi: 10.1111/1467-9868.00196.

Trickett, S., 2003, F-xy eigenimage noise suppression: Geophysics, **68**, 751–759, doi: 10.1190/1.1567245.

Trickett, S., 2008, F-xy Cadzow noise suppression: 78th Annual International Meeting, SEG, Expanded Abstracts, 2586–2590.

Trickett, S., and L. Burroughs, 2009, Prestack rank reduction-based noise suppression: CSEG Recorder, **34**, 3193–3196.

Trickett, S., L. Burroughs, and A. Milton, 2012, Robust rank reduction filtering for erratic noise: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0129.1.

Ulrych, T., S. Freire, and P. Siston, 1988, Eigenimage processing of seismic sections: 58th Annual International Meeting, SEG, Expanded Abstracts, 1261–1265.

Unkel, S., and N. T. Trendafilov, 2010, A majorization algorithm for simultaneous parameter estimation in robust exploratory factor analysis: Computational Statistics and Data Analysis, **54**, 3348–3358, doi: 10.1016/j.csda.2010.02.003.

Yang, H. H., and Y. Hua, 1996, On rank of block Hankel matrix for 2-D frequency detection and estimation: IEEE Transactions on Signal Processing, **44**, 1046–1048, doi: 10.1109/78.492568.