Robust reduced-rank seismic denoising

Ke Chen and Mauricio D. Sacchi, University of Alberta*

SUMMARY

Singular Spectrum Analysis (SSA) or Cadzow reduced-rank filtering is an efficient method for random noise attenuation when the data are contaminated by Gaussian noise. SSA starts by embedding the seismic data into a Hankel matrix. Rank-reduction of this Hankel matrix followed by anti-diagonal averaging is utilized to estimate an enhanced seismic signal. The rank-reduction step in the SSA filter is often implemented via the truncated Singular Value Decomposition (TSVD). The TSVD is a non-robust matrix factorization that often leads to suboptimal results when the seismic data are contaminated by erratic noise. We propose to adopt a *robust matrix factorization* that permits to utilize the SSA filter in situations where the data are contaminated by noise bursts, outliers and/or isolated anomalous traces.

INTRODUCTION

The improvement of the signal-to-noise ratio of seismic records is an important topic in seismic data processing. Incoherent noise attenuation can be carried out via prediction error filters in f-x (Canales, 1984) and t-x (Abma and Claerbout, 1995) domains. Incoherent noise can also be attenuated via rank-reduction methods. Rank-reduction methods can be grouped into several categories. For instance, eigenimage filtering (Freire and Ulrych, 1988), similar to filtering via the Karhunen-Loeve transform (Al-Yahya, 1991; Ulrych et al., 1999) can operate directly on the seismic data in the t-x, f-x or f-x-y domains (Trickett, 2003). Recently, the Singular Spectrum Analysis (SSA) method (Sacchi, 2009; Oropeza and Sacchi, 2011), also known as Cadzow filtering (Trickett, 2008), was introduced to attenuate incoherent noise and as an alternative to f-x prediction error methods. 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 is found via the low rank approximation of the Hankel matrix of the noisy observations (Oropeza and Sacchi, 2011).

In this article, we propose a robust SSA method for removing Gaussian and erratic noise. Rank-reduction is implemented via robust matrix factorization. The Hankel matrix of the data is approximated by the product of two low-dimensional factor matrices. The bisquare function is used to obtain a robust metric to approximate the Hankel matrix by a matrix of lower rank. The Iteratively Re-weighted Least Squares (IRLS) method (De la Torre and Black, 2003; Maronna and Yohai, 2008) is used to optimize the two low rank factor matrices. Our synthetic and real data examples show that the new robust SSA method can easily cope with non-Gaussian erratic noise.

THEORY

Singular Spectrum Analysis

This section provides a short review of the basic idea of the SSA method, which is also called Cadzow filtering. Details pertaining the implementation of SSA for seismic noise attenuation and seismic data reconstruction can be found in (Oropeza and Sacchi, 2011). We discuss the 2-D (t-x) implementation of SSA. However, we stress that SSA for 3D and 5D volumes have been extensively discussed in Oropeza and Sacchi (2011) and Gao et al. (2011), respectively. Seismic data in a small window can be represented in the frequency-space domain via the superposition of plane waves

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

where $i = \sqrt{-1}$, $j = 1, 2, ..., N$ 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. We denote $A_k(\omega)$ the complex amplitude of the k-th plane wave and Δx indicates the spatial interval between seismograms. The SSA method constructs a trajectory matrix by embedding spatial data at one frequency, i.e. $\mathbf{D}(\omega) = [D_1(\omega), D_2(\omega), \cdots, D_N(\omega)]^T$ into the following Hankel matrix

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

where the symbol $\mathscr{H}$ 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), $\mathbf{M}_\omega \in \mathbb{C}^{L\times(N-L+1)}$. 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). Additive noise in $\mathbf{D}$ will increase the rank of matrix $\mathbf{M}$. Then, one way of attenuating additive noise is via Rank-reduction. The SSA filter can be represented via the following expression

$$\hat{\mathbf{D}} = \mathscr{A}[\mathscr{R}_K[\mathscr{H}[\mathbf{D}]]], \tag{3}$$

where $\mathscr{A}$ is the anti-diagonal averaging operator, $\mathscr{R}_K[\mathbf{M}]$ is the rank-reduction operator that approximates $\mathbf{M}$ by a rank-K matrix and $\mathscr{H}$ is the Hankel operator. The operator $\mathscr{A}$ transforms back a Hankel form into a vector by averaging across anti-diagonals. It is important to stress that a similar analysis is valid for multidimensional signals where one must adopt block Hankel matrices and block anti-diagonal averaging operators (Trickett, 2008). The rank-reduction step ($\mathscr{R}_K$) of the method can be implemented via the truncated SVD (Trickett, 2008),

DOI http://dx.doi.org/10.1190/segam2013-0822.1

the randomized SVD (Oropeza and Sacchi, 2011) or by fast algorithms that adopt Lanczos bidiagonalization and fast Fourier Transforms for matrix-times-vector multiplications (Gao et al., 2011, 2013). All these Rank-reduction methods are non-robust and therefore, they are prone to degradation in the presence of outliers.

The ℓ_2 Low Rank Approximation

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

$$\begin{aligned} \mathbf{M}_K = \mathscr{R}_K(\mathbf{M}) = \underset{\hat{\mathbf{M}}}{\operatorname{argmin}} \|\mathbf{M}-\hat{\mathbf{M}}\|_F^2, \\ \text{subject to } rank(\hat{\mathbf{M}}) = K, \end{aligned} \tag{4}$$

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

$$\begin{aligned} \mathbf{M}_K =& \mathscr{R}_K(\mathbf{M}) = \mathbf{U}_K\mathbf{S}_K\mathbf{V}_K^H \\ =& \mathbf{U}_K\mathbf{U}_K^H\mathbf{M}, \end{aligned} \tag{5}$$

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 first K-largest singular values $s_j, j = 1\ldots K$ which are also the diagonal elements of the matrix $\mathbf{S}_K\in\mathbb{R}^{K\times K}$. The latter is also known as the Eckart-Young theorem (Eckart and Young, 1936). Rank-reduction via TSVD is quite simple to implement and solution is unique. However, the quadratic misfit functional makes the solution quite sensitive to non-Gaussian noise. This drawback could limit the application of the SSA method in situations where the data are contaminated by outliers. In this article, we investigate a robust measure of distance between the matrices $\mathbf{M}$ and $\mathbf{M}_K$ and an algorithm to estimate a low-rank approximation under the new distance.

Robust Low Rank Approximation

We now propose to replace the Frobenius metric for distance between two matrices in equation (4) by a robust metric. The new problem becomes

$$\begin{aligned} \mathbf{M}_K = \mathscr{R}_K(\mathbf{M}) = \underset{\hat{\mathbf{M}}}{\operatorname{argmin}} \|\mathbf{M}-\hat{\mathbf{M}}\|_\rho \\ \text{subject to } rank(\hat{\mathbf{M}}) = K, \end{aligned} \tag{6}$$

where $||\mathbf{M}-\hat{\mathbf{M}}||_\rho = \sum_{i=1}^{m}\sum_{j=1}^{n}\rho(\frac{m_{ij}-\hat{m}_{ij}}{\sigma})$, m_{ij} is the element at i-th row and j-th column of $\mathbf{M}$, σ is a scale parameter for function ρ. When ρ is non-quadratic, the problem (Equation (6)) is non-convex. We have tried different metrics for robust estimation and concluded that good results are attainable via Tukey's bisquare function (Beaton and Tukey, 1974)

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

The bisquare functional is portrayed in Figure 1 in conjunction to the classical ℓ_2 metric utilized by the Frobenius norm. The

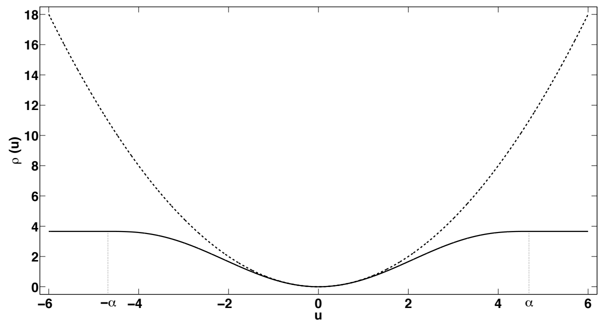

Figure 1: Dashed line is the quadratic function $\frac{1}{2}|u|^2$. Solid line is Tukey's bisquare robust metric adopted in this article for robust matrix factorization.

constant α is a tuneable parameter. Holland and Welsch (1977) recommend to take $\alpha = 4.685$ for Tukey's bisquare function. The value $\alpha\sigma$ performs as a transition point that permits to distinguish outliers from inliers. Smaller $\alpha\sigma$ will penalize the outliers more heavily which results in a more robust estimation.

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

$$\left(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right) = \underset{\mathbf{X},\mathbf{Y}}{\operatorname{argmin}}\|\mathbf{M}-\mathbf{X}\mathbf{Y}^H\|_\rho. \tag{8}$$

The problem (8) is solved via Iteratively Reweighted Least Squares (IRLS) (De la Torre and Black, 2003; Maronna and Yohai, 2008). The iterations for updating model weights are referred to as *external* iterations. The weighting function for bisquare function is

$$w(r) = \begin{cases} \left[1-\left(\frac{|r|}{\alpha\sigma}\right)^2\right]^2 & \left|\frac{r}{\sigma}\right|\le\alpha \\ 0 & \left|\frac{r}{\sigma}\right|>\alpha \end{cases}, \tag{9}$$

where r is residual. In each external iteration, the two unknowns $\mathbf{X}$ and $\mathbf{Y}$ are still coupled together. This can be addressed by the alternating minimization method (Gabriel and Zamir, 1979; Roweis, 1997; Tipping and Bishop, 1999). The iterations for alternating minimization are referred to as *internal* iterations. In t-th external iteration, the alternating minimization algorithm (internal iterations) is briefly expressed as follows

$$\begin{aligned} \mathbf{Y}_t^l &= \underset{\mathbf{Y}_t}{\operatorname{argmin}}\,||\mathbf{W}_{t-1}^{\frac{1}{2}}\odot(\mathbf{M}-\mathbf{X}_t^{l-1}\mathbf{Y}_t^H)||_F^2, \\ \mathbf{X}_t^l &= \underset{\mathbf{X}_t}{\operatorname{argmin}}\,||\mathbf{W}_{t-1}^{\frac{1}{2}}\odot(\mathbf{M}-\mathbf{X}_t\mathbf{Y}_t^{l^H})||_F^2, \end{aligned} \tag{10}$$

where t indicates external iteration index, l indicates internal iteration index, $\mathbf{W}_{t-1}$ is weighting matrix calculated from $(t-1)$-th external iteration, $\odot$ indicates elementwise product operator, $\frac{1}{2}$ indicates elementwise square root operator. The internal iterations are performed until $\|\mathbf{X}_t^l\mathbf{Y}_t^{l^H}-\mathbf{X}_t^{l-1}\mathbf{Y}_t^{l-1^H}\|_F\le$

DOI http://dx.doi.org/10.1190/segam2013-0822.1

ε. The robust scale parameter σ can be estimated from the residuals (difference between the matrix $\mathbf{M}$ and $\mathbf{XY}^H$) via the expression

$$\sigma = 1.4826\,\text{MAD} \; = 1.4826\,\text{med}\;|\,\mathbf{r} - \text{med}\;|\,\mathbf{r}\,||, \qquad (11)$$

where MAD indicates the median absolute deviation and $\mathbf{r}$ are the residuals in vector form (Holland and Welsch, 1977). We choose to adopt a random initialization of the factor matrices $\mathbf{X}$ and $\mathbf{Y}$ to start the robust matrix factorization algorithm and we update σ using equation (11) in each external iteration.

EXAMPLES

We present synthetic examples and also a real data example to illustrate the proposed algorithm. We compared the performance of the robust SSA method with the performance of classical SSA and f-x deconvolution.

Synthetic Example

The synthetic example is used to test the algorithm robustness with respect to outliers. Figure 2 (b) shows a 2-D t-x data set, which has 40 traces and a total time of 1.2 s with sampling interval 0.004 s. It contains Gaussian noise with signal to noise ratio (SNR) equal to 1, and isolated noisy traces. The amplitude of the erratic noise traces is 3 and 2 times of the maximum amplitude of the uncorrupted data. The processing frequency band ranges from 1 to 40 Hz. We select the size of subspace of the reconstructed data to be $K = 3$. The algorithm is started with a random initialization of the factor matrices $\mathbf{X}$ and $\mathbf{Y}$. We choose the number of external iterations (for updating weights) equal to 10 and number of internal iterations (for alternating minimization) equal to 5. The results of f-x deconvolution, SSA and robust SSA are compared. The length of prediction filter in f-x deconvolution is 10, the trade-off parameter is 0.001. Figure 2 (a) is the noise free data, Figure 2 (b) is the contaminated noisy data and Figure 2 (c) is the added noise. Figure 3 (a) shows the result of f-x deconvolution, we can see that the result is not very good because large amplitude noise leaks over several traces in the output panel. Figure 3 (b) shows the result of the classical non-robust SSA implemented via the SVD. Again, we observe that the erratic noise has not been properly removed and noticeable artifacts are present in the output gather. The robust SSA method is shown in Figure 3 (c). In this case, the Gaussian and erratic noise were successfully suppressed. We evaluate the denoising performance by evaluating the factor $Q = 10\,\log\,\dfrac{||d^0||_F^2}{||d^0 - \hat{d}||_F^2}$, where d^0 is the noise free data, $\hat{d}$ is the reconstructed data. Larger value of Q means better denoising performance. The Q value of 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 $Q_{rssa} = 12.4$. 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

The proposed robust SSA algorithm is also tested on a real data set. The data are CDP stack gathers with 800 traces. The time sample interval is 0.002 s, and the total recorded time is 3 s. Figure 4 (a) shows this field data set. Figure 4 (b) and (c) shows the data in the left and right rectangular windows in Figure 4 (a), respectively. We divide the data set into overlapping windows, process each window and then add them back. This operation is used to speed up the method and also to preserve details in the data. The size of the window is 1500 time samples times 80 CDP gathers. The processing band for all test ranges from 1 Hz to 80 Hz. The rank K of the reconstructed low rank matrix is chosen to be 2 for both SSA and the robust SSA method. The external iteration number of the robust SSA method is set to 10 and the internal iteration number is set to 5. The length of prediction filter in f-x deconvolution is set to 10 and the trade-off parameter is set to 0.001. We show the results for data in two windows highlighted Figure 4(a). The results for window to the left in Figure 4(a) are shown in Figure 5. The results for window to the right of Figure 4(a) are shown in Figure 6. We observe that the robust SSA method has successfully removed noise bursts present in the data.

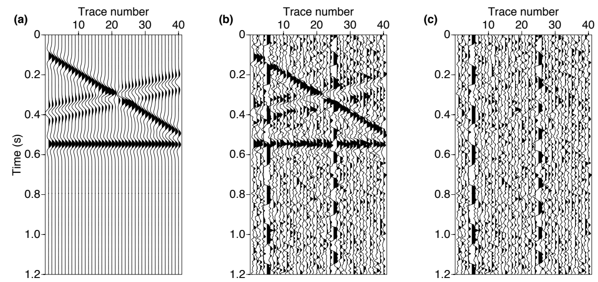

Figure 2: The synthetic data with three linear events. (a) Clean data. (b) Data with Gaussian noise and erratic spatial noise. (c) The noise added to the data.

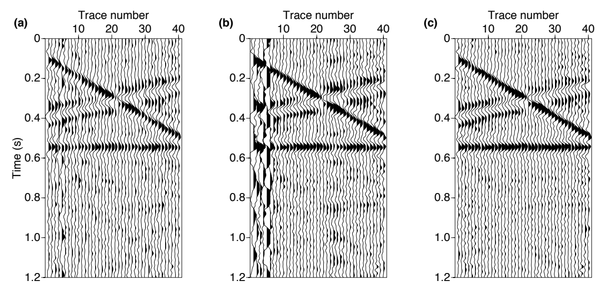

Figure 3: (a) Data in Figure 2(b) after f-x deconvolution. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.

CONCLUSIONS

In this paper, we propose a robust version of the SSA method which can remove Gaussian and non-Gaussian (erratic) noise. The robust matrix factorization is used in the new method instead of the truncated SVD. Synthetic and real data examples were used to analyze the performance of the new algorithm. One possible concern is the computation cost of the robust

DOI http://dx.doi.org/10.1190/segam2013-0822.1

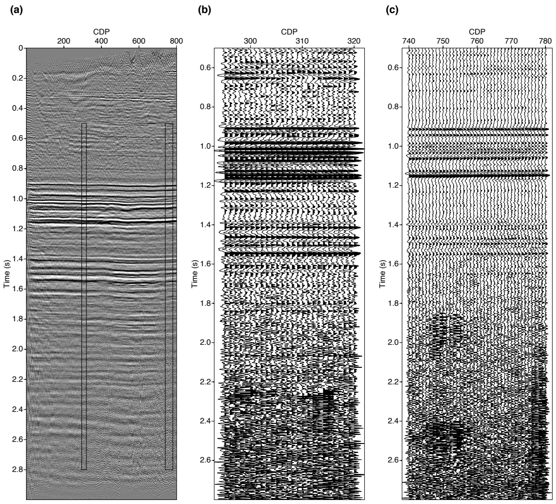

Figure 4: Poststack field data. (a) The whole data set. (b) The data in the left rectangular window. (c) The data in the right rectangular window.

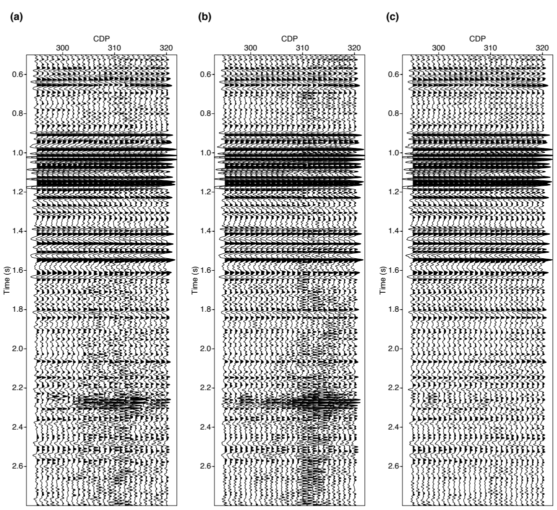

Figure 5: The comparison of results of the data in the left rectangular window by three different methods. (a) Data after f-x deconvolution filtering. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.

algorithm. Computational time can be reduced by adopting windowing strategies to minimize the size of the Hankel matrices to factorize. Another strategy is to truncate the number of iterations of the alternating minimization algorithm and IRLS solvers in a way that an inexact factorization is estimated. We have noticed that an inexact factorization can yield better results than conventional non-robust Rank-reduction via the truncated SVD.

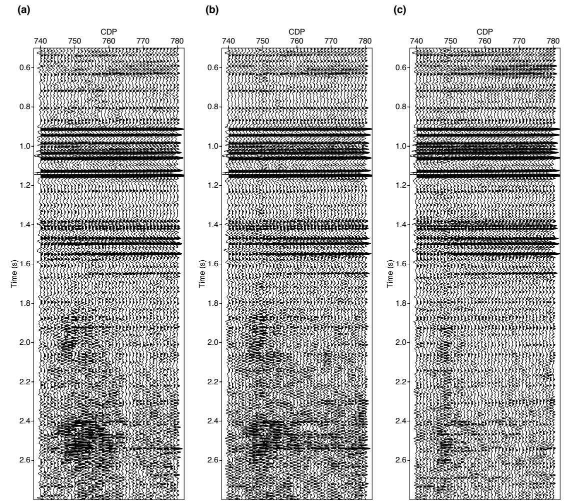

Figure 6: The comparison of results of the data in the right rectangular window by three different methods. (a) Data after f-x deconvolution filtering. (b) Data after classical SSA filtering. (c) Data after robust SSA filtering.

ACKNOWLEDGMENTS

We wish to thank the sponsors of the Signal Analysis and Imaging Group (SAIG) at the University of Alberta for supporting this research.

DOI http://dx.doi.org/10.1190/segam2013-0822.1

http://dx.doi.org/10.1190/segam2013-0822.1

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2013 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

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

Al-Yahya, K. M., 1991, Application of the partial Karhunen-Love transform to suppress random noise in seismic sections: Geophysical Prospecting, **39**, no. 1, 77–93, http://dx.doi.org/10.1111/j.1365-2478.1991.tb00302.x.

Beaton, A. E., and J. W. Tukey, 1974, The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data: Technometrics, **16**, no. 2, 147–185, http://dx.doi.org/10.1080/00401706.1974.10489171.

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

De la Torre, F., and M. J. Black, 2003, A framework for robust subspace learning: International Journal of Computer Vision, **54**, no. 1, 117–142, http://dx.doi.org/10.1023/A:1023709501986.

Eckart, C., and G. Young, 1936, The approximation of one matrix by another of lower rank: Psychometrika, **1**, no. 3, 211–218, http://dx.doi.org/10.1007/BF02288367.

Freire, S. L. M., and T. J. Ulrych, 1988, Application of singular value decomposition to vertical seismic profiling: Geophysics, **53**, 778–785, http://dx.doi.org/10.1190/1.1442513.

Gabriel, K. R., and S. Zamir, 1979, Lower rank approximation of matrices by least squares with any choice of weights: Technometrics, **21**, no. 4, 489–498, http://dx.doi.org/10.1080/00401706.1979.10489819.

Gao, J., M. 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, http://dx.doi.org/10.1190/geo2012-0038.1.

Gao, J., M. D. Sacchi, and X. Chen, 2011, A fast rank reduction method for the reconstruction of 5D seismic volumes: 81st Annual International Meeting, SEG, Expanded Abstracts, **30**, 3622–3627.

Holland, P. W., and R. E. Welsch, 1977, Robust regression using iteratively reweighted least-squares: Communications in Statistics Theory and Methods, **6**, no. 9, 813–827, http://dx.doi.org/10.1080/03610927708827533.

Hua, Y., 1992, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil: Signal Processing: IEEE Transactions on, **40**, no. 9, 2267–2280, http://dx.doi.org/10.1109/78.157226.

Maronna, R. A., and V. J. Yohai, 2008, Robust low-rank approximation of data matrices with elementwise contamination: Technometrics, **50**, no. 3, 295–304, http://dx.doi.org/10.1198/004017008000000190.

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

Roweis, S., 1997, EM algorithms for PCA and SPCA: Neural Information Processing Systems, 626–632.

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

Srebro, N., and T. Jaakkola, 2003, Weighted low-rank approximation: Proceedings of the 20th International Conference on Machine Learning.

Tipping, M. E., and C. M. Bishop, 1999, Probabilistic principal component analysis: Journal of the Royal Statistical Society, Series B, Statistical Methodology, **61**, no. 3, 611–622, http://dx.doi.org/10.1111/1467-9868.00196.

Trickett, S., 2008, F-xy Cadzow noise suppression: 78th Annual International Meeting, SEG, Expanded Abstracts, **27**, 2586–2590.

Trickett, S. R., 2003, F-xy eigen image noise suppression: Geophysics, **68**, 751–759, http://dx.doi.org/10.1190/1.1567245.

Ulrych, T., M. Sacchi, and S. Freire, 1999, Eigenimages processing of seismic sections: SEG.

DOI http://dx.doi.org/10.1190/segam2013-0822.1