GeoConvention 2015
GEOSCIENCE
new horizons
Telus Convention Centre • 4-8 May 2015 • Calgary, AB Canada

5D Tensor-based seismic data completion: The Parallel Matrix Factorization (PMF) algorithm

Jianjung Gao [(1),(2)], *Aaron Stanton* [(2)] *and Mauricio D. Sacchi* [(2)]

[(1)] *China University of Geosciences, Beijing*
[(2)] *Department of Physics, University of Alberta*

Summary

We discuss the implementation of the Parallel Matrix Factorization (PMF) algorithm, an SVD-free tensor completion method that is applied to 5D seismic data reconstruction. The Parallel Matrix Factorization (PMF) algorithm complements and expands our first generation of 5D tensor completion codes based on High Order SVD and Nuclear norm minimization. We review the PMF method and explore its applicability via tests with a 5D field data reconstruction example.

Introduction

Reconstruction methods that are based on rank reduction techniques are becoming popular in seismic data processing. These methods can mainly be divided into two sub-categories. One category of methods applies rank reduction to a block Hankel matrices formed by the entries of observed seismic data in the frequency-space domain. Methods in this subcategory are often named Cadzow (Trickett et al., 2010) or Multichannel Singular Spectrum Analysis reconstruction (Oropeza and Sacchi, 2011; Gao et al., 2013). A second category of methods are based on dimensionality reduction of multilinear arrays or tensors. Example of the latter are High Order SVD (HOSVD) reconstruction (Kreimer and Sacchi, 2011, 2012), Tucker decomposition (Herrmann and Silva, 2013), nuclear norm minimization method (Kreimer et al., 2013) and tensor SVD (Ely et al., 2013). The common feature of these method is that they all utilize the SVD algorithm to reduce the rank of the data tensor. For large-scale seismic data reconstruction problems, the cost of the SVD algorithm inhibits the more widely usage of low rank tensor completion methods for industrial applications.

We analyze the Parallel Matrix Factorization (PMF) algorithm proposed by Xu et al. (2013). The PMF method does not utilize SVDs. We show that PMF is an effective algorithm to recover missing traces from large 5D volumes.

Theory

The PMF reconstruction method is implemented in midpoint-offset frequency domain. We denote the data by $D(\omega,x,y,h_x,h_y)$, where x, y, h_x and h_y indicate the spatial coordinates and in the inline midpoint, crossline midpoint, in-line offset and cross-line offset. After binning the data in midpoint-offset domain, single frequency data $D(\omega,x,y,h_x,h_y)$ can be represented by a 4th-order tensor $\mathscr{D}$, with elements D_{i_1,i_2,i_3,i_4}, where i_1,i_2,i_3,i_4 are bins indices for the spatial coordinates x,y,h_x and h_y, respectively. We remove the dependency on ω to simplify the notation. The reconstructed data are obtained by minimizing a cost function of the form

$$\Phi \;=\; \Phi_C + \mu\,\Phi_M\,, \tag{1}$$

where, Φ_M is the data misfit term, $\Phi_M = \frac{1}{2}\|\mathscr{P}\circ\mathscr{X}-\mathscr{D}\|_F^2$, $\mathscr{P}$ is the Nth-order sampling operator tensor with elements 1 for the observed samples and 0 for the missing samples. $\mathscr{X}$ is the Nth-order low rank tensor representing the reconstructed data (the unknown of our problem). The functional Φ_C is the low rank constraint term that is defined as follows

$$\Phi_C = \frac{1}{2}\sum_{k=1}^{N} \|\mathbf{X}_{(k)}\mathbf{Y}_{(k)} - \mathbf{Z}_{(k)}\|_F^2 , \tag{2}$$

where, $\mathbf{Z}_{(k)}$ is the mode-k unfolding matrix of the tensor $\mathscr{Z}$. The low rank matrix factorization is applied to each mode unfolding of $\mathscr{Z}$ by seeking matrices $\mathbf{X}_{(k)} \in \mathbb{C}^{I_k \times r_k}$ and $\mathbf{Y}_{(k)} \in \mathbb{C}^{r_k \times I_1 \ldots I_{k-1} I_{k+1} \ldots I_N}$ such that $\mathbf{Z}_{(k)} \approx \mathbf{X}_{(k)}\mathbf{Y}_{(k)}$ for $k = 1,\ldots,N$, where r_k is the rank of the unfolding matrix $\mathbf{Z}_{(k)}$. In order to solve $\mathbf{X}_{(k)}$, $\mathbf{Y}_{(k)}$ and $\mathscr{Z}$, we minimize the cost function Φ and apply the alternating least-squares algorithm

$$\mathbf{X}_{(k)}^{i+1} = \mathbf{Z}_{(k)}^{i}(\mathbf{Y}_{(k)}^{i})^H,\ k = 1,\ldots,N, \tag{3a}$$

$$\mathbf{Y}_{(k)}^{i+1} = ((\mathbf{X}_{(k)}^{i+1})^H \mathbf{X}_{(k)}^{i+1})^{\dagger}(\mathbf{X}_{(k)}^{i+1})^H \mathbf{Z}_{(k)}^{i},\ k = 1,\ldots,N, \tag{3b}$$

$$\mathscr{Z}^{i+1} = (\mathscr{I} - \alpha\mathscr{P}) \circ \mathscr{C} + \alpha\mathscr{D} , \tag{3c}$$

where, the parameter $\alpha = \frac{\mu}{N+\mu}$, $\mathscr{I}$ is the Nth order tensor with all entries equal to 1 and $\mathscr{C}$ is given by

$$\mathscr{C} = \frac{1}{N}\sum_{k=1}^{N} \mathrm{fold}_k[\mathbf{X}_{(k)}^{i+1}\mathbf{Y}_{(k)}^{i+1}] . \tag{4}$$

The preceding analysis corresponds to the case where data are contaminated with noise. The noise-free data reconstruction case is tackled by finding the minimum of the following cost function

$$\Phi = <\mathscr{W}, \mathscr{P} \circ \mathscr{Z} - \mathscr{D}> + \Phi_M \tag{5}$$

where $<\mathscr{A},\mathscr{B}> = \sum_{i_1=1}^{I_1} \ldots \sum_{i_N=1}^{I_N} \bar{A}_{i_1 \ldots i_N} B_{i_1 \ldots i_N}$. Using the method of Lagrange multipliers, the solution of equation 5 is given by

$$\mathscr{Z} = (\mathscr{I} - \mathscr{P}) \circ \mathscr{C} + \mathscr{D}. \tag{6}$$

Expression 6 is equal to expression 3c for the particular case when α=1. It is interesting to no mention that equation 3c resembles the typical imputation algorithm used for reconstrustruction via POCS (Abma and Kabir, 2005) and Cazdow (Trickett et al., 2010; Gao et al., 2013) methods.

Synthetic example

The first example is a 5D seismic data that consist of $I_1 \times I_2 \times I_3 \times I_4$ spatial traces with $I_k = 6, 8, 10, 12, 14$, $k = 1,2,3,4$ and 301 time samples per trace. The data include three linear events and S/N = ∞. We randomly remove 50% of the traces and perform the reconstruction using the proposed PMF algorithm, the HOSVD algorithm and the nuclear norm minimization method. For the PMF and HOSVD methods, we adopt a rank r_k=3 for all modes ($k = 1,2,3,4$), the maximum number of iterations is set to $N_{iter} = 100$, and an iteration stopping error $tol = 10^{-4}$ is adopted for each frequency, respectively. For the nuclear norm method, we set N_{iter}=100, $tol = 10^{-4}$ and the parameters $\lambda = 2.5$, $\beta = 15$ (see, Kreimer et al. (2013)) Table 1 shows the compassion of the computational cost of the three methods. For each iteration, the computational cost of PMF method is $O(N(3mnr + mr^2 + \frac{2nr^3}{3}))$. For the nuclear norm method, the cost is $O(N(2m^2n + 2m^3))$ per iteration and for the case of the HOSVD, the computational cost is $O(N(4m^2n + 13m^3 + mr^{3(N-1)}))$. Where $m = I_k = max\{I_1, I_2, \ldots, I_N\}$, $r = r_k = \max\{r_1, r_2, \ldots, r_N\}$, $n = I_1 I_2 \ldots I_{k-1} I_{k+1} \ldots I_N$ and N represents the order of seismic data tensor. From table 1, we observe that the PMF algorithm is faster than the nuclear norm minimization algorithm and the HOSVD algorithm. We also choose a synthetic data model containing $12 \times 12 \times 12 \times 12$ traces in the spatial directions and 301 time samples per trace which is also used in table 1 to examine the reconstruction quality of the proposed PMF algorithm, HOSVD reconstruction and the nuclear norm minimization reconstruction method. We define the reconstruction quality $Q = 10\log_{10}(\frac{\|\mathscr{D}^{true}\|^2}{\|\mathscr{D}^{true}-\mathscr{D}^{recon}\|^2})$ where $\mathscr{D}^{true}$ and $\mathscr{D}^{recon}$ represent the true noise-free complete data and reconstructed data in the time-space domain. Table 2 shows the comparison of the reconstruction quality versus the percentage of missing traces. From Table 2, we find that the reconstruction quality obtained by the proposed PMF method and HOSVD algorithm are very similar. They both perform better than the nuclear norm minimization method. For the third example, we synthesize a noise-free data with four events with strong curvature. The spatial size of the data is $12\times12\times12\times12$ with 301 time samples per trace and S/N $=\infty$. We randomly decimated 90% of the traces and set the rank $r_1 = r_2 = r_3 = 5$ for modes $1,2,3$ and $r_4 = 4$ for mode 4. We also set $N_{iter} = 300$, $tol < 10^{-4}$ and $\alpha = 1$. Figure 1 shows the reconstruction result. From error section in Figure 1d, one can observe that missing traces were accurately recovered. We also add random noise to the noise-free data in Figure 1 to analyze the reconstruction capability our the algorithm in the presence of noise. We set S/N $= 1$, $N_{iter} = 300$, $tol < 10^{-4}$ and $\alpha = 0.51$. Figure 2 shows the reconstruction result.

I_k	Cost (secs)		
	PMF	HOSVD	Nuclear norm
8	49.8	814.1	74.1
10	74.7	919.3	159.3
12	117.9	1077.1	307.4
14	195.1	1259.4	569.6

Table 1: Computational time comparison of the proposed PMF reconstruction method, the HOSVD method and nuclear norm method for different 5D volumes with size of $301\times I_1\times I_2\times I_3\times I_4$, $I_k = 8,10,12,14$, $k = 1,2,3,4$.

Decimation [%]	Reconstruction quality Q		
	PMF	HOSVD	Nuclear norm
60	62.6	62.7	16.8
70	61.1	61.3	9.7
80	60.6	60.7	6.5
90	39.9	31.3	3.0

Table 2: Reconstruction quality Q versus percentage of missing traces for the PMF, HOSVD and Nuclear norm reconstruction methods with data size $I_k = 12$.

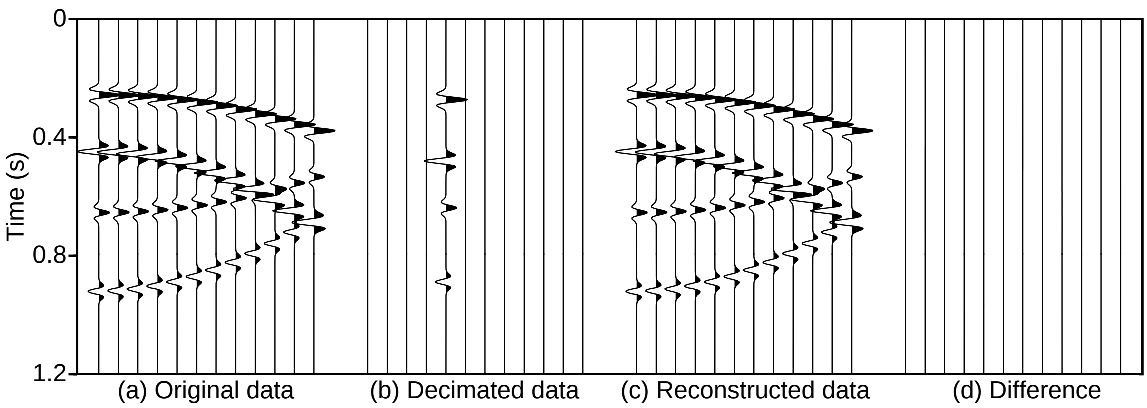


Figure 1: Data reconstruction results with rank $r_1 = r_2 = r_3 = 5$, r_4=4, α=1.0 and 90% missing traces. A slice of the synthetic 5D volume with size $301\times12\times12\times12\times12$ is portrayed. In this case $SNR = \infty$.

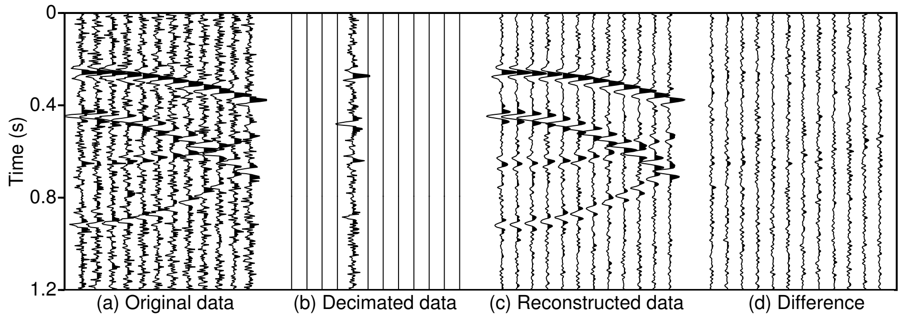
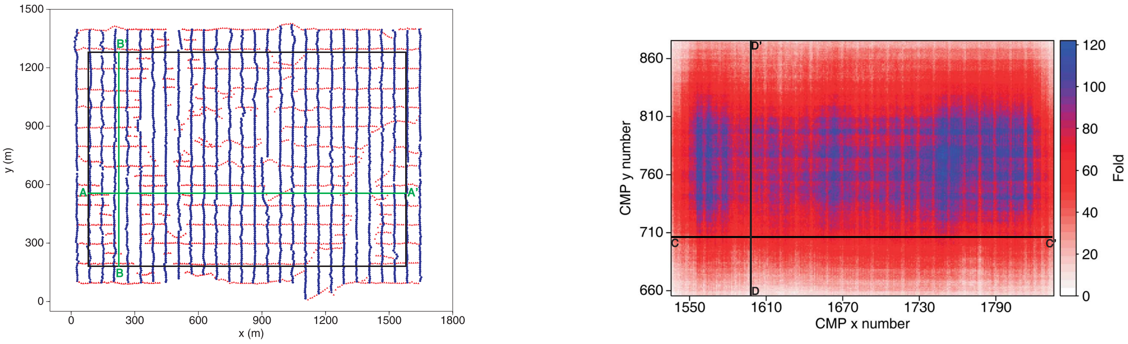


Figure 2: Noisy data reconstruction result with S/N=1.0, rank $r_1 = r_2 = r_3 = 5$, r_4=4, α=0.51 and 90% missing traces. A slice of the synthetic 5D volume of size $301 \times 12 \times 12 \times 12 \times 12$ is portrayed.



(a) (b)

Figure 3: Field data example (WCB). (a) Distribution of sources and receivers. (b) Fold map.

Field data example

Based on the above synthetic data analysis, we tested the performance of PMF reconstruction method on a land data set obtained from a heavy oil field in the WCB (Figure 3). The data are first binned on a $5m \times 5m$ CMP grid and a $100m \times 100m$ Offset-x-y grid prior to interpolation. The reconstruction area includes 300 CMP$_x$ bins and 220 CMP$_y$ bins. We divide the whole survey data into 2640 overlapping blocks. Each block has about 85% missing traces. We set $r_k = 4,\ k = 1,2,3,4,$ $N_{iter} = 100$ and $\alpha = 0.40$ for the PMF reconstruction. Figure 4a and 4 b shows a zero offset slice of 5D volume by fixing CMP$_y$ bin 730, h_x bin 4 and h_y bin 11. Figure 4c and 4 d show a zero offset slice of 5D volume by fixing CMP$_x$ bin 1555, h_x bin 4 and h_y bin 11 before and after reconstruction.

Conclusions

We have presented a SVD-free method for multidimensional seismic data reconstruction. The proposed PMF method applies low rank matrix factorization to mode unfoldings of the seismic data tensor and applies an alternating minimization algorithm to estimate the complete data tensor. Contrary to other low rank reconstruction methods, PMF does not require the SVD algorithm. The latter makes the PMF algorithm attractive for industrial implementations. We compared the proposed method to two methods that were developed by our group (HOSVD and minimum Nuclear Norm reconstruction). We conclude that the proposed 5D data completion PMF method is faster than our previously reported algorithms for tensor completion.

Acknowledgements

We thank the sponsors of the Signal Analysis and Imaging Group (SAIG) at the University of Alberta and NSERC (Discovery Grant, Fundamental and Applied Studies in Seismic Data Preconditioning

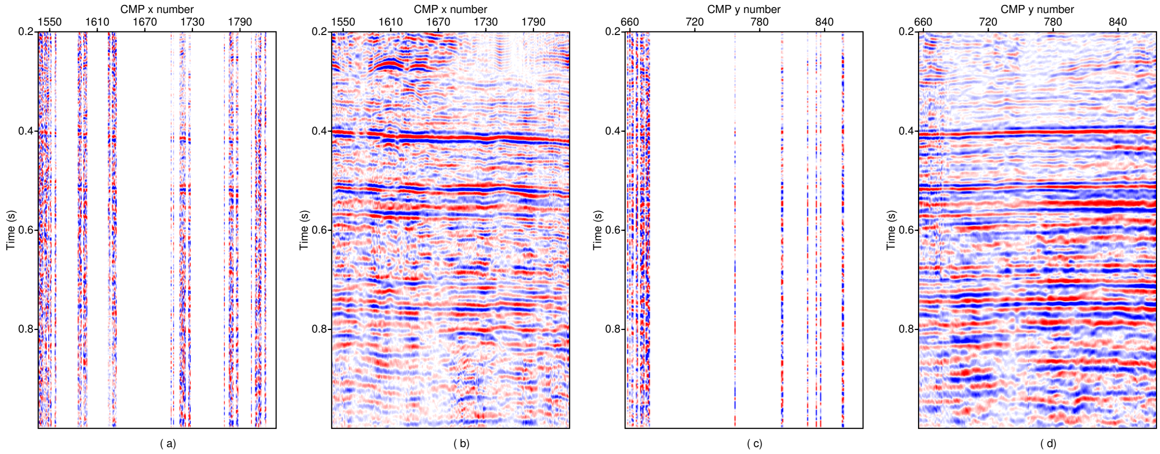


Figure 4: Real data reconstruction result for zero offset. A subset of the 5D data is portrayed. (a) CMP_x gather at CMP_y bin 730, h_x bin 4 and h_y bin 11 before reconstruction. (b) The reconstructed data of (a) via PMF method. (c) CMP_y gather at CMP_x bin 1555, h_x bin 4 and h_y bin 11 before reconstruction. (d) The reconstructed data via the PMF method.

and Inversion). Jianjun Gao also thanks financial support from the National Natural Science Foundation of China (Grant No. 41304102), Fundamental Research Fund for Central Universities (Grant No. 2652013048) and Fundamental Research Fund (Grant No. GDL1206) for the Key Laboratory of Geo-detection (China University of Geosciences, Beijing), Ministry of Education. We also thank Nadia Kreimer for implementing HOSVD and Minimum Nuclear Norm reconstruction methods.

References

Abma, R. and N. Kabir, 2005, Comparison of interpolation algorithms: The Leading Edge, **24**, no. 10, 984–989.

Ely, G., S. Aeron, N. Hao, and M. Kilmer, 2013, 5D and 4D pre-stack seismic data completion using tensor nuclear norm (TNN): SEG Technical Program Expanded Abstracts, 3639–3644.

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.

Herrmann, F. J. and C. D. Silva, 2013, Structured tensor missing-trace interpolation in the hierarchical Tucker format: SEG Technical Program Expanded Abstracts, **702**, 3623–3627.

Kreimer, N. and M. D. Sacchi, 2011, Evaluation of a new 5d seismic volume reconstruction method - Tensor Completion versus Fourier reconstruction: CSPG CSEG CWLS Convention, 5 pages.

——— 2012, A tensor higher-order singular value (HO-SVD) decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, V113–V122.

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

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

Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: SEG, Expanded Abstracts, **29**, no. 1, 3829–3833.

Xu, Y. Y., R. R. Hao, W. T. Yin, and Z. X. Su, 2013, Parallel matrix factorization for low rank tensor completion: UCLA Computational and Applied Mathematics, 13–77.