A new 5D seismic reconstruction method based on a Parallel Square Matrix Factorization algorithm

Jianjun Gao, Key Laboratory of Geo-detection (Ministry of Education), China University of Geosciences (Beijing), Jinkun Cheng, University of Alberta, Mauricio D. Sacchi, University of Alberta*

SUMMARY

Multidimensional seismic data reconstruction can be viewed as a low rank matrix or tensor completion problem. Different rank-reduction approaches can be employed to perform seismic data interpolation and denoising. For these methods, the computational cost and reconstruction quality are two important aspects that must be carefully considered. In this paper, we present a new fast and economic tensor completion method named Parallel Square Matrix Factorization (PSMF). We apply the algorithm to the ubiquitous 5D seismic data regularization problem. 5D reconstruction entails reconstructing a series 4th-order multilinear arrays (tensors) in the frequency domain. For this purpose we transform the data to the frequency domain and 4D spatial volumes in midpoint-offset are reshaped into matrices. Rank-reduction of these matrices is at the core of our reconstruction algorithms. We show that properly reshaping the data tensor into almost square matrices lead to an improved tensor completion algorithm. We demonstrate the effectiveness of the proposed approach via synthetic examples and by a data set from Western Canadian Sedimentary Basin.

INTRODUCTION

In the frequency-space domain, properly sampled seismic data can be represented by a low rank matrix or tensor. Decimation of traces and additive noise increase the rank of the matrix or tensor. Hence, matrix and tensor completion methods which are widely applied in computer vision and recommendation systems (Liu et al., 2013) can be adopted to recover missing traces and to enhance the SNR of the seismic volume. Different reduced-rank methods have been adopted for seismic data processing. For instance, Trickett et al. (2010) proposed a matrix rank-reduction method based on Cadzow Filtering (CF) to interpolate missing traces. Similarly, Oropeza and Sacchi (2011) proposed a Multichannel Singular Spectrum Analysis (MSSA) method to reconstruct the 3D data and adopted a Randomized SVD algorithm to speed up the rank reduction filter required by their algorithm. Gao et al. (2013) expanded the MSSA method to reconstruct 4D spatial data and adopted fast multilevel Toeplitz matrix-vector multiplication algorithms to improve the computational efficiency of the original MSSA algorithm. Kreimer and Sacchi (2012) introduced a low rank tensor completion method, named High Order SVD (HOSVD), to reconstruct 5D seismic volumes. Kreimer et al. (2013) proposed a nuclear norm minimization method which does not require the provision of a priori rank estimates. The common place of these methods is that they all reduce the rank of the tensor via the SVD algorithm or via the Lanczos bidiagonalization technique. Kumar et al. (2013) proposed a robust nuclear norm minimization method and adopted iterative low rank matrix factorization as an alternative to the SVD. Recently, Gao et al. (2015) introduced a fast SVD-free tensor completion approach, named the Parallel Matrix Factorization (PMF) method, to reconstruct multidimensional seismic data. Although the PMF method can significantly improve the computational efficiency of low rank reconstruction methods, it adopts a tensor unfolding procedure that leads to unbalanced "long strip" matrices. We will show in this paper that the reconstruction quality and ability of the "long strip" matrix rank reduction method is not as good as rank reduction methodologies applied on a balanced square matrix. We propose a fast and economic tensor completion method, called Parallel Square Matrix Factorization (PSMF) to improved the quality of the reconstruction of seismic tensors. We first reshape the original tensor of seismic data into balanced square matrices and perform low rank matrix factorization to recover missing traces. We use synthetic data sets and a field data to examine the reconstruction efficacy of the proposed method.

THEORY

Previous studies show that fully sampled seismic data can be represented via low rank matrices or tensors. The latter is more obvious in the frequency domain than in time domain. Furthermore, it has been shown that the singular values of unfolded tensors decay faster in the midpoint-offset domain than in the source-receiver domain (Kreimer, 2013; Aleksandr et al., 2014). Hence, we transform the subsampled seismic data from source-receiver domain to midpoint-offset domain and then implement the low rank tensor completion in frequency-midpoint-offset domain. We denote the prestack seismic volume by $D(\omega,x,y,h_x,h_y)$, where x, y, h_x and h_y represent the spatial coordinates of the inline, crossline midpoint, inline, crossline offset. We bin the data in midpoint-offset domain and define the seismic volume for one single frequency slice via a 4th-order tensor $\mathscr{D}^{obs}$ with elements D_{i_1,i_2,i_3,i_4}, where the indices i_1, i_2, i_3 and i_4 represent the spatial coordinates x, y, h_x and h_y, respectively. The symbol ω is dropped to gain concision but it is clear that the algorithm is run for frequencies in the seismic band $\omega \in [\omega_{min}, \omega_{max}]$.

Traditional Matrix Factorization (TMF) Algortihm

For N-dimensional seismic data reconstruction, one would like to recover a low rank tensor $\mathscr{Z} \in \mathbb{C}^{I_1\times I_2\times I_3\times I_4}$ from the partially observed data tensor $\mathscr{D}^{obs} = \mathscr{P} \circ \mathscr{Z}$, where $\mathscr{P}$ is the sampling operator with elements 1 and 0. '$\circ$' is an element-wise product operator. Following the work of Wen et al. (2012), we convert the tensor $\mathscr{Z}$ to a matrix $Z_{(k)}$ along an arbitrary mode-k and build the following cost function that one must

DOI http://dx.doi.org/10.1190/segam2015-5842738.1

minimize in order to recover the missing entries of the tensor

$$\min_{X,Y,\mathscr{Z}} \| XY - Z_{(k)} \|_F^2 + \frac{\mu}{2} \| \mathscr{P} \circ \mathscr{Z} - \mathscr{D}^{obs} \|_F^2, \quad k \in \{1,2,\ldots,N\}, \tag{1}$$

where, $\|\cdot\|_F^2$ denotes the Frobenius norm, μ is a regularization factor that balances the weight of the low rank constraint term $\| XY - Z_{(k)} \|_F^2$ and the data misfit term $\| \mathscr{P} \circ \mathscr{Z} - \mathscr{D}^{obs} \|_F^2$. We call the model in equation 2 the Traditional Matrix Factorization (TMF) method. In essence, this model belongs to the family of matrix completion methods. The mode-i unfolding matrix $Z_{(k)}$ of size $I_k \times I_1 \cdots I_{k-1} I_{k+1} \cdots I_N$ is usually an unbalanced "long strip" matrix. Recent research by Mu et al. (2014) indicates that $Z_{(k)}$ could be reshaped into a more balanced square matrix and, as a consequence, the tensor $\mathscr{Z}$ can be better recovered.

Square Matrix Factorization (SMF) model

We modify the TMF model in equation 2 and propose a Square Matrix Factorization model (SMF),

$$\min_{X,Y,\mathscr{Z}} \| XY - \hat{Z}_{[j]} \|_F^2 + \frac{\mu}{2} \| \mathscr{P} \circ \mathscr{Z} - \mathscr{D}^{obs} \|_F^2, \tag{2}$$

where, $\hat{Z}_{[j]}$ is a balanced square matrix (close to a square matrix) with size of $I_{i_1} \cdots I_{i_j} \times I_{i_{j+1}} \cdots I_{i_N}$, $\hat{Z}_{[j]}$=reshape$(\hat{Z}_{(1)}, \prod_{k=1}^{j} I_{i_k}, \prod_{k=j+1}^{N} I_{i_k})$, $j \in \{1,2,\ldots,N\}$. The matrix $\hat{Z}_{(1)}$ is the mode-1 unfolding of the tensor $\hat{\mathscr{Z}}$. The tensor $\hat{\mathscr{Z}}$ is obtained by relabelling the mode-i_k of tensor $\mathscr{Z}$ to k for $k = 1,2,\ldots,N$. The permutation $\{i_1, i_2, \ldots i_N\}$ is chosen to make $\prod_{k=1}^{j} I_{i_k}$ as close possible to $\prod_{k=j+1}^{N} I_{i_k}$. After the reshaping operation, $\hat{Z}_{[j]}$ becomes a more balanced matrix than the unfolded tensor $Z_{(k)}$ in equation 2 (Mu et al., 2014).

Parallel Matrix Factorization (PMF) model

In our simulations we show that the recovery ability of SMF model outperforms the TMF method. However, both TMF and SMF belong to the category of matrix completion methods. In this section we propose to exploit all the modes in which one can unfold the seismic tensor (Xu et al., 2013; Gao et al., 2015). In the Parallel Matrix Factorization (PMF) method, the 4D spatial seismic data reconstruction is reconstructed by minimizing the following cost function

$$\min_{X_{(k)},Y_{(k)},\mathscr{Z}} \sum_{k=1}^{N} \| X_{(k)} Y_{(k)} - Z_{(k)} \|_F^2 + \frac{\mu}{2} \| \mathscr{P} \circ \mathscr{Z} - \mathscr{D}^{obs} \|_F^2, \tag{3}$$

where, $Z_{(k)}, k = 1,2,3,4$ are usually is a "long strip" matrices of size $I_k \times I_1 \ldots I_{(k-1)} I_{(k+1)} \ldots I_N$. Compared with the tensor completion methods via SVD algorithm, the PMF method is an SVD-free approach, and can significantly decrease the computation cost of the reconstruction while attaining similar reconstruction quality as SVD-based methods (Gao et al., 2015).

Parallel Square Matrix Factorization (PSMF) model

Unfortunately, $Z_{(k)}$ in the PMF method are unbalanced matrices and we will show that by simple reorganizing the tensor into balanced matrices one can improve the reconstruction quality. For this purpose, modify the PMF model and introduce the square matrix factorization technique. For 4D spatial data reconstruction, the new model named Parallel Square Matrix Factorization (PSMF) is given by,

$$\min_{X_{(j)},Y_{(j)},\mathscr{Z}} \sum_{j=1}^{3} \| X_{(j)} Y_{(j)} - \hat{Z}_{[j]} \|_F^2 + \frac{\mu}{2} \| \mathscr{P} \circ \mathscr{Z} - \mathscr{D}^{obs} \|_F^2, \tag{4}$$

where, the balanced square matrices are now given by $\hat{Z}_{[1]}$= reshape$(Z_{(1)}, I_1 I_2, I_3 I_4)$, $\hat{Z}_{[2]}$=reshape$(Z_{(1)}, I_1 I_3, I_2 I_4)$ and $\hat{Z}_{[3]}$=reshape$(Z_{(1)}, I_1 I_4, I_2 I_3)$. Compared with the original PMF model (Gao et al., 2015) in equation 3, the PSMF model only contains three low rank constraint term. This is due to the dimensional permutation $\{I_1, I_2, I_3, I_4\}$ that was chosen in such a way that the row or column index of the balanced matrix $\hat{Z}_{[j]}$ only has three different patterns $\{I_1 I_2\}$, $\{I_1 I_3\}$ and $\{I_1 I_4\}$. The comparisons of reconstruction performance for the TMF, SMF, PMF and PSMF methods are shown in tables 1 and 2. Clearly, by simple balancing the size of the unfolded tensors that are needed by the PMF algorithm, we increase the quality of reconstruction by $5 - 10$dB.

TESTS

Results with synthetic data

We construct a series of 5D seismic data models to examine the reconstruction performance of TMF, SMF, PMF and PSMF methods. The first model is designed to test the computation cost and reconstruction quality for the noise-free data case. We synthesize a 5D seismic volume that consist of $12 \times 12 \times 12 \times 12$ traces and 301 time samples per trace. The data include 4 curve events and $S/N = \infty$. We randomly remove 20%, 50% and 80% traces from the complete data. Then, we apply the alternating least squares method to solve the four tensor completion problems defined in equation 2, 2, 3 and 4. We set the maximum number of iterations to $N_{iter} = 300$, the iteration stopping error $tol = 10^{-4}$, temporal frequencies for reconstruction in the band are $1 - 70$ Hz, and we vary the rank value $r_1 = r_2 = r_3 = r_4 = r$ from 1 to 12 with increment 1. We define the reconstruction quality factor $Q = 10 \log_{10} \frac{\| \mathscr{D}^{true} \|_F^2}{\| \mathscr{D}^{recon} - \mathscr{D}^{true} \|_F^2}$ in dB units, where $\mathscr{D}^{true}$ is the complete data and $\mathscr{D}^{recon}$ is the reconstructed data. For each decimated volume reconstruction case, we run the code 5 times for each method and record the computation time and best rank value r which corresponds to the largest Q. The average computation time corresponding to the best rank r for each method is shown in Table 1. We point out that the best rank value for TMF, SMF, PSMF and PMF is 4, 4, 5 and 5. Table 1 illustrates that the SMF method has the best performance in terms of computational efficiency, the PSMF method performs second best, and both of them are faster than the PMF method. Table 2 shows the reconstruction quality of the four methods. We observe that the PSMF method outperform the other three methods. It is interesting to observe a noticeable improvement when the data are severely decimated (80% missing traces). Figure 1 shows a slice of the original complete data with $S/N = \infty$ and decimated data with 80% traces missing. Figure 2 displays a slice of the reconstructed results and the difference. From Figure 2 e to 2 h, we observe that the TMF method fails to reconstruct the data with 80% missing traces. The other three methods are able to recover the data.

The second model is used to examine the denoising and reconstruction ability for data contaminated with noise. We add random noise to the noise-free data model with $S/N = 1$. Figure 3 shows a slice of the complete data and decimated data. We set the regularization factor $\mu = 0.1$ and keep all the other parameters as in the noise-free example. Figure 4 shows the reconstruction results. From the difference sections, 4e to 4h, we can see that the error for the PSMF method is smaller than for the other methods.

Field data example

Based on the above synthetic data analysis, we test PSMF method on a real prestack data set obtained by a survey from the Western Canadian Sedimentary Basin. The data volume contains 15×15 CMP bins and 13×13 offsets per bin and 351 time samples per trace. We set rank r_k=5, k=1, 2, 3, 4., N_{iter}=100 and μ=0.9 for the PSMF and PMF reconstruction. Figure 5 shows the reconstruction result of a CMPx gather obtained by fixing CMPy bin=7, hy bin=7. Figure 6 shows the reconstruction result of a CMPy gather obtained by fixing CMPx bin=6, hx=6. From Figure 5 and 6, we conclude that the seismic events are better recovered by the PSMF than by the PMF method.

Decimation	Computation time (s)			
	TMC	SMF	PSMF	PMF
20%	38.7	7.6	**21.3**	88.4
50%	42.1	12.0	**30.0**	174.6
80%	38.3	45.3	**141.8**	354.6

Table 1: Computational time comparison of the proposed PSMF reconstruction method, TMF method, SMF method and PMF method for 5D volumes. The best rank value r adopted in TMF, SMF, PSMF and PMF methods is 4, 4, 5 and 5, respectively.

Decimation	Quality factor Q			
	TMC	SMF	PSMF	PMF
20%	35.7	55.6	**60.8**	60.4
50%	11.4	50.1	**57.2**	55.6
80%	-3.0	40.5	**50.6**	39.1

Table 2: Reconstruction quality Q versus percentage of missing traces for the TMF, SMF, PSMF and PMF methods. The best rank r used in the TMF, SMF, PSMF and PMF methods is 4, 4, 5 and 5, respectively.

CONCLUSIONS

We have presented a new fast and accurate low rank tensor completion method and apply it to reconstruct multidimensional seismic data. The proposed PSMF method reshapes the

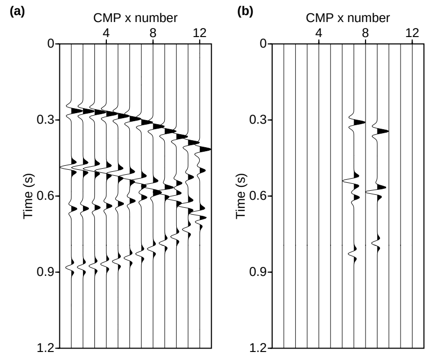

Figure 1: Slice view of original noise-free complete data a) and decimated data with 80% missing traces b).

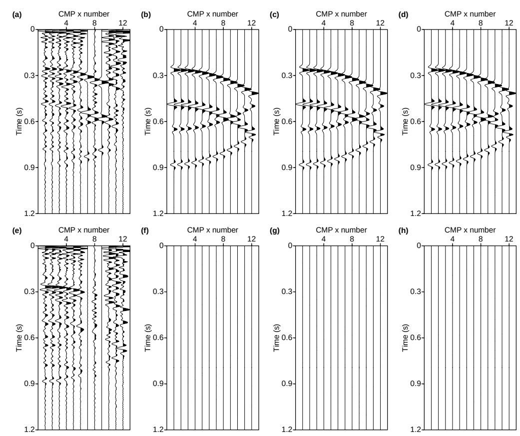

Figure 2: Slice view of reconstructed data for the noise-free data. a) TMF method. b) SMF method. c) PSMF method. d) PMF method. e), f), g) and h) are the difference of the subtraction of the true complete data and reconstructed data results.

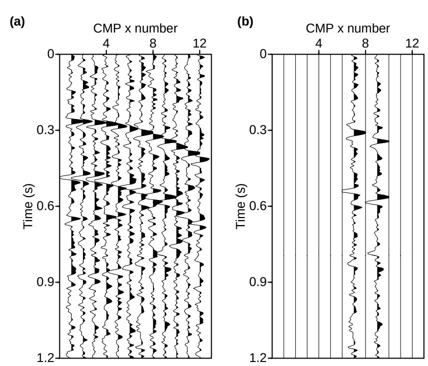

Figure 3: Slice view of noisy complete data with S/N=1.0 a) and decimated data with 80% missing traces b).

DOI http://dx.doi.org/10.1190/segam2015-5842738.1

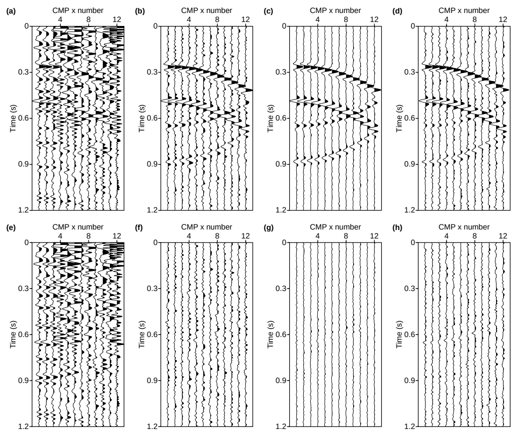

Figure 4: Slice view of reconstruction results for the noisy data set with 80% missing traces and μ=0.1. a) TMF method. b) SMF method. c) PSMF method. d) PMF method. e), f), g) and h) are the difference of the subtraction of the true complete data and reconstructed data.

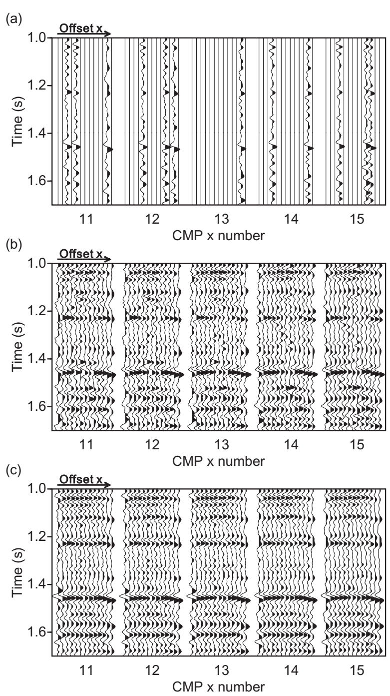

Figure 5: A CMPx gather reconstruction result of a real data for fixed CMPy bin=7, hy bin=7 and μ=0.9. a) Input gathers. b) Reconstructed gathers using the PMF method. c) Reconstructed gathers using the PSMF method.

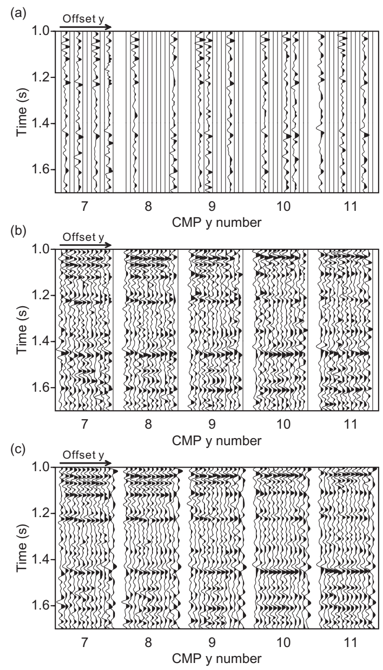

Figure 6: A CMPy gather reconstruction result of a real data for fixed CMPx bin=6, hx bin=6. a) Input gathers. b) Reconstructed gathers using the PMF method. c) Reconstructed gathers using the PSMF method.

"long strip" unfolded matrices into more balanced square matrices and utilize a SVD-free parallel matrix factorization algorithm to reduce the rank of the seismic tensor and recover the missing samples. The synthetic and field data tests indicate that the PSMF method not only improves the reconstruction ability of tensor completion methods for sparsely sampled data but also decreases computational cost in comparison to the PFM method.

ACKNOWLEDGMENTS

We would like to thank all the sponsors of the Signal Analysis and Imaging Group (SAIG) at University of Alberta and by an NSERC Discovery Grant (Fundamental and Applied Studies in Seismic Data Preconditioning and Inversion). Jianjun Gao also thanks the financial support from the National Natural Science Foundation of China (NO. 41304102), Fundamental Research Fund for Central Universities (NO. 2652013048) and Fundamental Research Fund (No. GDL1206) for the Key Laboratory of Geo-detection (China University of Geosciences, Beijing), Ministry of Education.

DOI http://dx.doi.org/10.1190/segam2015-5842738.1

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2015 SEG Technical Program Expanded Abstracts have been copyedited 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

Aravkin, A., R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, 2014, Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation: SIAM Journal on Scientific Computing, **36**, no. 5, S237–S266. http://dx.doi.org/10.1137/130919210.

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. J., and A. Stanton, and S. M., 2015, 5D seismic data reconstruction and denoising using the parallel matrix factorization method: Presented at the 77th Annual International Conference and Exhibition, EAGE.

Kreimer, N., 2013, Multidimensional seismic data reconstruction using tensor analysis: Ph.D. thesis, University of Alberta.

Kreimer, N., and M. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, no. 3, V113–V122. http://dx.doi.org/10.1190/geo2011-0399.1.

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

Kumar, R., A. Aravkin, H. Mansour, B. Recht, and F. Herrmann, 2013, Seismic data interpolation and denoising using SVD-free low-rank matrix factorization: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.20130388.

Liu, J., P. Musialski, P. Wonka, and J. Ye, 2013, Tensor completion for estimating missing values in visual data: IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**, no. 1, 208–220. http://dx.doi.org/10.1109/TPAMI.2012.39 PMID:22271823

Mu, C., B. Huang, J. Wright, and D. Goldfarb, 2014, Square deal: Lower bound and improved relaxations for tensor recovery: International Conference on Machine Learning, 73–81.

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.

Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: 80th Annual International Meeting, SEG, Expanded Abstracts, 3829–3833.

Wen, Z., W. Yin, and Y. Zhang, 2012, Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm: Mathematical Programming Computation, **4**, no. 4, 333–361. http://dx.doi.org/10.1007/s12532-012-0044-1.

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.

DOI http://dx.doi.org/10.1190/segam2015-5842738.1