Comparative Study of Complex Parallel Factor Analysis and Parallel Factor Analysis

Huilu ZENG, Zhinong LI, Zewen ZHOU
Key Laboratory of Nondestructive Testing, Ministry of Education, Nanchang Hangkong University, Nanchang 330063, China
Email: lizhinong@tsinghua.org.cn

***Abstract*—The running time and the convergence between traditional parallel factor trilinear alternating least squares algorithm (TALS) algorithm and complex parallel factor (COMFAC) algorithm is compared by the experiment. The experiment result shows that both methods can obtain good separation performance. However, the traditional parallel factor separation algorithm has the higher complexity and the slower convergence. The complex parallel factor analysis can improve the convergence of the the traditional parallel factor analysis. The solution of complex parallel factor is usually very close to the least squares solution with only a few iterations.**

Keywords—Complex parallel factor analysis, Parallel factor analysis, trilinear alternating least squares algorithm (TALS), Blind source separation, Separation performance.

I. INTRODUCTION

In recent years, parallel factor analysis has been widely concerned and successfully applied in the environmental monitoring, brain signal processing, communication processing, etc. Seredyńska [1] used parallel factor analysis to monitor organic pollution in swimming pools by means of three-dimensional fluorescence maps (EEMs). Shutova [2] used parallel factor analysis to achieve on-line monitoring of soluble organic matter in drinking water. Yang [3] summarized the applications of EEM-PARAFAC in drinking water and waste water treatment, and analyzed the feasibility of on-line monitoring of PARAFAC in drinking water and waste water treatment. Ref.[4] used PARAFAC to analyze the driver's brain activity during driving period in order to understand their relevant information and facilitate better interaction between people and cars. Spanier [5] used CANDECOMP/PARAFAC to analyze brain activity of schizophrenia, and find abnormal brain activity areas in patients with schizophrenia. Sidropoulos [6] introduced the PARAFAC model into the field of signal processing, and proposed a receiver for DS-CDMA system based on parallel factor model.

However, With the development of parallel factor model, some shortcomings of traditional parallel factor separation method are gradually revealed, such as high complexity, slow convergence, etc. To overcome the above shortcomings, Bro [7] proposed a fast parallel factor analysis algorithm, i,e. complex parallel factor Algorithms (COMFAC).COMFAC algorithm improved the convergence speed of TALS effectively. In this paper, the TALS algorithm and COMFAC algorithm are described. Their convergence and running time of two algorithms is compared by the experiment. The research of this paper provides theoretical support for the application of the parallel factor theory into the engineering field.

II. TRADITIONAL PARALLEL FACTOR THEORY AND ALGORITHM

Parallel factor (PARAFAC) analysis, also known as trilinear decomposition or canonical decomposition. The specific model diagram can be as shown in Fig. 1.

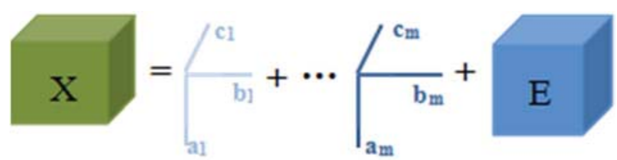

Fig. 1 Parallel factor model

In Fig.1, $I\times J\times K$ three-dimensional array $\underline{X}$ can be expressed as

$$X_{i,j,n}=\sum_{m=1}^{M}A_{i,m}B_{j,m}C_{n,m}+E_{i,j,n} \quad (1)$$

where $i=1,2,\ldots,I$, $j=1,2,\ldots,J$, $n=1,2,\ldots,N$, $m=1,2,\ldots,M$; $A\in \mathrm{C}^{I\times M}$, $B\in \mathrm{C}^{J\times M}$, and $C\in \mathrm{C}^{N\times M}$ are the three load matrices of the model respectively. $E\in \mathrm{C}^{I\times J\times N}$

This work was supported by National Natural Science Foundation of China (51675258, 51261024), the Science and Technology Research Project of Education Department of Jiangxi Province (No. GJJ150699) and Research Project of State Key Laboratory of Mechanical System and Vibration (No. MSV201914).

is a noise matrix. Thus, the model constructed in Eq. (1) can be divided into three-dimensional matrix X along three dimensions.

$$\begin{cases} X_{\mathrm{i::}} = BD_i(A)C^T + E_{\mathrm{i::}} \\ X_{\mathrm{:j:}} = AD_j(B)C^T + E_{:j:} \\ X_{\mathrm{::n}} = AD_n(C)B^T + E_{::n} \end{cases} \tag{2}$$

where $\mathrm{D_i}(A)$ is a diagonal matrix composed of the i th row elements of matrix A .The parallel factor is different from the low-rank decomposition of the two-dimensional matrix. Without any addition constrains, the PARAFAC model decomposition is unique. Therefore, in order to discuss whether PARAFAC model is identifiable, the definition of matrix *k*-rank and the uniqueness theorem of PARAAFAC model decomposition are introduced. The definition of *k*-rank is as follows [8].

For a given matrix $A \in \mathrm{C}^{I \times M}$, If and only if A contains at least r independent column, the rank of A is

$$\mathrm{r}_A = Rank(A) = r \tag{3}$$

If any *k*-columns of matrix A is independent, the *k*-rank of A is $k_A = k$. where $k_A \le r_A \le \min(I, M), \forall A$.

The PARAFAC uniqueness theorem [9] is described as follows.

if the k -ranks of matrices A , B , and C satisfied

$$k_A + k_B + k_c \ge 2(M+1) \tag{4}$$

the three load matrices, obtained by the decomposition of tensor X , are unique in column transformation and scale transformation. If all three load matrices satisfy the k-rank condition, the identifiable sufficient condition is described as

$$\min(I, M) + \min(J, M) + \min(N, M) \ge 2M + 2 \tag{5}$$

In parallel factor decomposition, trilinear alternating least square (TALS) is a common method for data detection in trilinear models. The basic idea of TALS is simple, the updating algorithm of matrix in TALS algorithm is that a matrix is updated at each step, the remaining matrix is updated with Least Square (LS) according to the last estimation. The other matrices continue to update. The above steps are repeated until the algorithm converges.

III. COMPLEX PARALLEL FACTOR THEORY AND ALGORITHM

Generally speaking, Trilinear Alternating Least Square (TALS) algorithm is a common algorithm for parallel factorization. However the step of the TALS algorithm is very cumbersome, and leads to many problems, such as more iteration times, slower convergence, high complexity and larger errors. Under the premise of ensuring convergence accuracy, the COMFAC (Complex Parallel Factor Analysis) algorithm can overcome the deficiency in the TALS algorithm.

The step of the COMFAC algorithm is described as follows.

1. the data matrix is compressed to reduce the redundant space.

2. data matrix is initialized and matched by parallel factor model.

3. data matrix is decompressed in the original space.

The above three steps are explained respectively. The explanation is as follows.

For step 1, the data model constructed by the parallel factor is a three-dimensional matrix $X(I \times J \times K)$, this three-dimensional matrix can be transformed into three matrix *U*, *V* and *Z* after compression, where *U* is a *I*-by-*P* matrix, *V* is a *J*-by-*Q* matrix, *Z* is a *K*-by-*R* matrix. The obtained compressed matrix is a $P \times Q \times R$ matrix.

the dimension of the compressed matrix is less than that of the original matrix. The specific process can be expressed in Eq.(6)

$$G^{(P \times QR)} = U^H X^{(I \times JK)} (Z \otimes V) \tag{6}$$

Thus step 3 can be inversely deduced and restored to the decompression mathematical expression.

$$\hat{X}^{(I \times JK)} = UG^{(P \times QR)} (Z^H \otimes V^H) \tag{7}$$

Here, the Tucker 3-ALS algorithm is used to compress the matrix. The specific steps are as follows

(1) Initialize the matrix V and the matrix Z

(2) Determine the matrix *U*, *U* is the first *P* left singular value vectors of $X^{(I \times JK)}(Z \otimes V)$.

(3) Determine the matrix *V*, *V* is the first *Q* left singular value vectors of the matrix $X^{(J \times IK)}(Z \otimes U)$.

(4) Determine the matrix *Z*, *Z* is the first *R* left singular value vectors of the matrix $X^{(K \times IJ)}(V \otimes U)$.

(5) Repeat step (2), until the convergence.

(6) Calculate the compression matrix $G^{(P \times QR)}$.

The initialization and matching Parallel factor model is as follows the parallel factor model of G is initialized. Each mold use different singular value vectors. For $G^{(P\times QR)}$ $G^{(Q\times RP)}$, and $G^{(R\times PQ)}$, the first F left singular value vectors are obtained by singular value decomposition respectively.

Slice form of G is

$$G_r = LD_r(N)M^T + E_r \tag{8}$$

where L is a P-by-F factor matrix, M is a Q-by-F factor matrix, N is a R-by-F factor matrix. Er is a error matrix. After using these symbols, the parallel factor can be expressed as.

$$\hat{G}^{(P\times QR)} = L(N \odot \mathrm{M})^{\mathrm{T}} \tag{9}$$

In Eq.(1), The minimization of loss function is the sum of squared residuals (SSR). Therefore SSR can be expressed as

$$SSR = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} e_{i,j,k}^2 \tag{10}$$

where $e_{i,j,k}$ represents the ith, jth, kth elements of the error matrix E. The sum of squared residuals in COMFAC algorithm is much smaller than that of tradition PARAFAC algorithm. The smaller the SSR, the better the fitting effect.

IV. COMPARATIVE EXPERIMENT

In order to compare the effectiveness of the algorithm, two sets of source signals are used as follows

$$S = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} (t+0.12)^{0.07}\cdot(1+\sin(2\pi f_{1b}t))\sin(2\pi f_{1r}t) \\ (t+0.24)^{0.08}\cdot(1+\sin(2\pi f_{2b}t))\sin(2\pi f_{2r}t) \end{pmatrix} \tag{11}$$

where f_{1b} = f_{2b} =1Hz, $f_{1r} = 30$ Hz, $f_{2r} = 20$ Hz, Sampling frequency $f_s = 5000$ Hz. Sampling length is N=40960. The time waveform and amplitude spectrum of the simulated source signal are shown in Fig. 1 and Fig. 2.

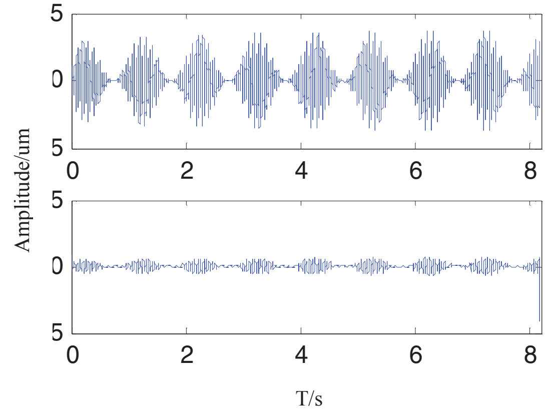

Fig. 1 The time waveform of source signal

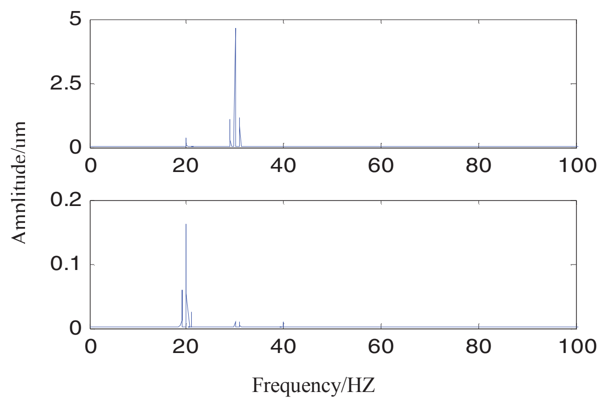

Fig. 2 The spectrum of source signal

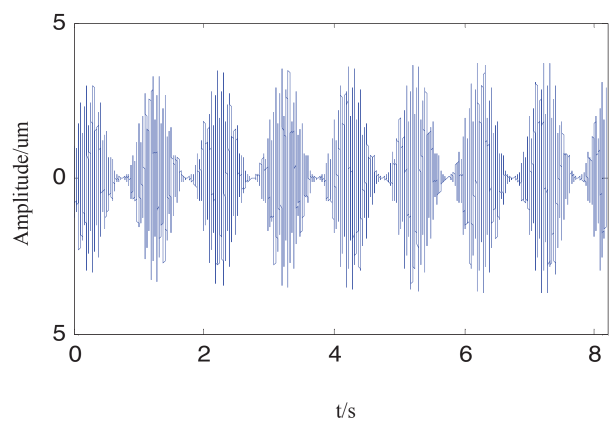

Fig. 3 The time domain waveform of observation signal

In order to obtain a virtual observation signal. an arbitrary random matrix A is selected to obtained an observation signal according to x(t) = As(t). The time domain waveform and spectrum of observation signal are shown in Fig. 3 and Fig. 4. From Fig.3 and Fig.4, the two source signals are mixed into a observation signal, which is an underdetermined blind separation problem. This two vibration source signals are completely interfered with each other.

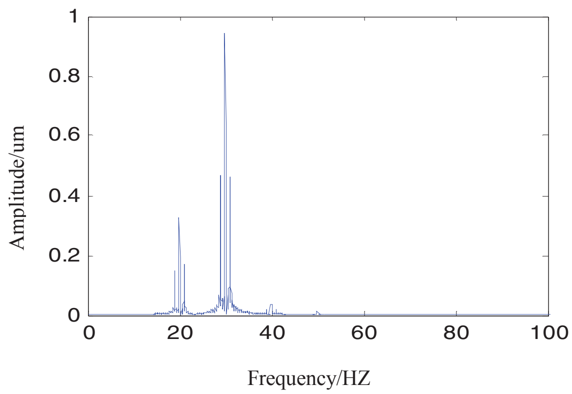

Fig. 4 The spectrum of observation signal

In order to solve the underdetermined blind separation problem, the production function (PF) component and original observed signal constitutes a new observed signal, where production function (PF) is obtained by local mean decomposition (LMD) of original observed signal [10]. Thus the underdetermined blind source separation problem is transformed into an over-determined blind source separation problem. TALS algorithm and COMFAC algorithm, which combined with LMD, is used to the blind separation of observed signal respectively. Fig. 5 and Fig. 6 shows the time domain waveform and its spectrum of estimated signal obtained by traditional PARAFAC algorithm. Fig. 7and Fig. 8 shows the time domain waveform and its spectrum of estimated signal obtained by COMFAC algorithm.

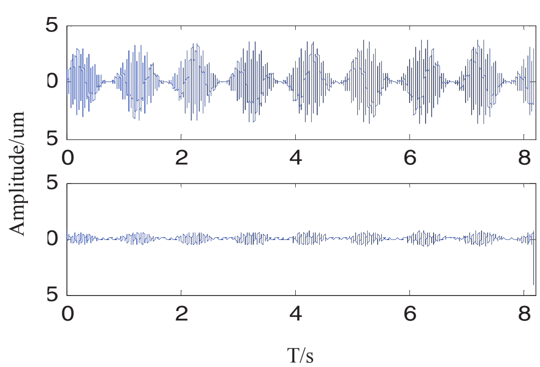

Fig. 5 The estimated signal obtained by traditional

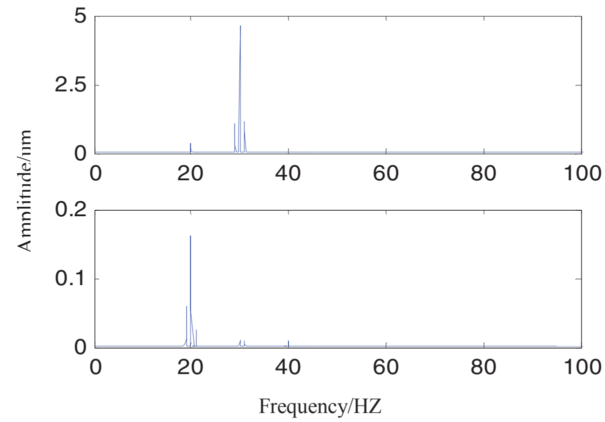

Fig. 6 The spectrum of estimated signal obtained

by traditional PARAFAC method

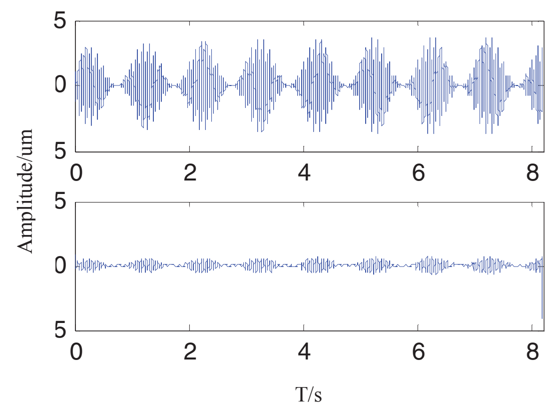

Fig. 7 The estimated signal obtained by COMFAC

method

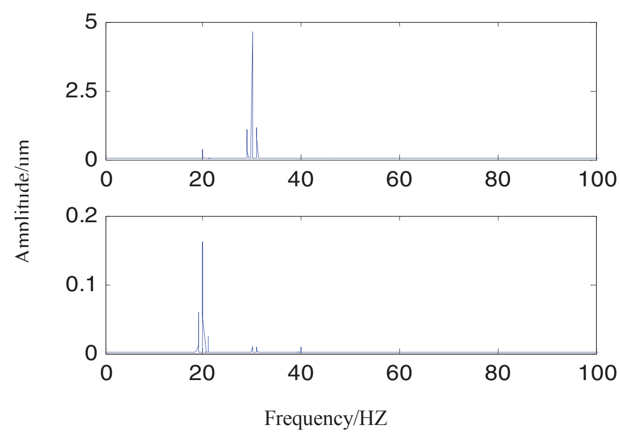

Fig. 8 The spectrum of estimated signal obtained by COMFAC method

From Fig. 5, Fig. 6 and Fig. 1, Fig. 2, Fig. 7, Fig. 8 and Fig. 1 and Fig. 2, two separation algorithms can obtain a good separation effect, and fully reflect the characteristic frequency of source signals. However, the TALS algorithm is used in the traditional parallel factor analysis, TALS algorithm has some obvious deficiency, such as large calculation amount, many iterations, and long running time etc. The COMFAC algorithm is a fast parallel factor algorithm. The TUCK3-ALS algorithm in the COMFAC algorithm replaces TALS algorithm in the traditional parallel factor, and the convergence speed is obviously improved. Therefor The COMFAC algorithm is obviously superior to the TALS algorithm.

In order to further compare the separation performance of the two algorithms, here the two algorithms is compared in the minimum loss function (SSR) and the running time. The comparison results are shown in Table 1.

TABLE I. THE COMPARE OF TWO ALGORITHMS IN SSR AND RUNNING TIME

Separation algorithm	COMFAC	TALS
Running time	80.6s	110.8s
SSR	10^{-4}	10^{0}

From Table1, the separation time is 110.8s in the traditional PARAFAC algorithm. However in the COMFAC algorithm, the separation time is 80.6s , the number of iterations is greatly reduced. From the performance index SSR, the SSR in the COMFAC algorithm is also much smaller than that in the TALS algorithm. Therefore the COMFAC algorithm is superior to the traditional PARAFAC algorithm. the COMFAC algorithm is an effective fast PARAFAC algorithm.

V. SUMMARY

In this paper, the basic principle and algorithm of traditional parallel factor analysis and complex parallel factor analysis are discussed in detail. On this basis, the separation performance of the two algorithms is compared by experiment. the experiment results show that two algorithms can obtain good separation results. However the two algorithms have obvious differences in running time and SSR. Traditional parallel factor analysis has high complexity, slow convergence and long running time than COMFAC algorithm. The SSR in TALS algorithm is also much larger than that in COMFAC algorithm. The complex parallel factor is a fast parallel factor algorithm, overcomes the deficiency in the traditional parallel factor analysis. The convergence speed in COMFAC algorithm is obviously improved, and the SSR is also greatly reduced.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (51675258, 51261024), the Science and Technology Research Project of Education Department of Jiangxi Province (No. GJJ150699) and Research Project of State Key Laboratory of Mechanical System and Vibration (No. MSV201914).

REFERENCES

[1] Seredyńska-Sobecka B, Stedmon CA, Boe-Hansen R, et al. Monitoring organic loading to swimming pools by fluorescence excitation–emission matrix with parallel factor analysis (PARAFAC)[J]. water research, 2011, 45 (6): 2306-2314.

[2] Shutova Y, Baker A, Bridgeman J, et al. Spectroscopic characterisation of dissolved organic matter changes in drinking water treatment: from PARAFAC analysis to online monitoring wavelengths[J]. Water research, 2014, 54: 159-169.

[3] Yang L, Hur J, Zhuang W. Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review[J]. Environmental Science and Pollution Research, 2015, 22 (9): 6500-6510.

[4] Ikenishi T, Kamada T, Nagai M. Parallel factor analysis of driver's EEG in lane change maneuver for cooperative driver assistance systems[C], 2010: 105-110.

[5] Spanier A, Joskowicz L, Moshel S, et al. Schizophrenia patients differentiation based on MR vascular perfusion and volumetric imaging[C]: International Society for Optics and Photonics, 2015: 94132Y-94132Y-94137.

[6] Sidiropoulos ND, Giannakis GB, Bro R. Blind PARAFAC receivers for DS-CDMA systems[J]. IEEE Transactions on Signal Processing, 2000, 48 (3): 810-823.

[7] Bro R, Sidiropoulos ND, Giannakis GB. A fast least squares algorithm for separating trilinear mixtures[C]. Proceedings of International Workshop on Independent Component Analysis and Blind Separation, Aussois, France, 1999: 289-294.

[8] Sidiropoulos ND, Bro R, Giannakis GB. Parallel factor analysis in sensor array processing[J]. IEEE transactions on Signal Processing, 2000, 48 (8): 2377-2388.

[9] Sidiropoulos ND, Giannakis GB, Bro R. Blind PARAFAC receivers for DS-CDMA systems[J]. IEEE Transactions on Signal Processing, 2000, 48 (3): 810-823.

[10] Zhinong LI, Weibing LIU, Xiaobing YI. Research on blind source separation method for mechanical fault underdetermination based on local mean decomposition[J]. Journal of Mechanical Engineering, 2011,07(47): 97-102.