



# An efficient PCA based method for Tucker3 model selection

Zhaoshui He, Andrzej Cichocki, *Senior Member, IEEE*

**Abstract**—There has been a growing interest in Tucker3 analysis recently. One of the biggest challenges in Tucker3 analysis is the model selection problem: how to choose the number of components in each mode of an observed tensor. An alternative Tucker3 model selection approach is developed based on principal component analysis (PCA) for this problem in this paper. It is computationally efficient and straightforward to implement. Its effectiveness is demonstrated by the experiments.

**Index Terms**—Tucker3 decomposition, model selection, principal component analysis (PCA)

## I. INTRODUCTION

TUCKER3 analysis arises in many applications, including chemometrics [1]–[3], signal processing [4]–[6], telecommunication [7], image processing [8], etc. Given an observed tensor  $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ , its Tucker3 approximation can be formulated as

$$y_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}, \quad (1)$$

where  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  and  $k = 1, \dots, K$  ( $P < I$ ,  $Q < J$  and  $R < K$ ), or in a matrix form as

$$\underline{\mathbf{Y}} = \underline{\mathbf{G}} \times_1 \underline{\mathbf{A}} \times_2 \underline{\mathbf{B}} \times_3 \underline{\mathbf{C}} + \underline{\mathbf{E}}, \quad (2)$$

where  $\times_A$ ,  $\times_B$  and  $\times_C$  denote the  $m_A$ -mode product,  $m_B$ -mode product and  $m_C$ -mode product, respectively [9]. The core tensor  $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times R}$ , the noise/error tensor  $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$  and the factor  $\underline{\mathbf{A}} \in \mathbb{R}^{I \times P}$ ,  $\underline{\mathbf{B}} \in \mathbb{R}^{J \times Q}$ ,  $\underline{\mathbf{C}} \in \mathbb{R}^{K \times R}$  are unknown. The objective is to choose the number ( $P, Q, R$ ) of components of tensor  $\underline{\mathbf{Y}}$  [10], [11].

There have been many methods to Tucker3 model selection. Roughly speaking, these methods can be categorized into two groups: the Tucker fitting based methods and the PCA based methods. The Tucker fitting based methods choose the parameters ( $P, Q, R$ ) by computing the Tucker fit values for the Tucker3 model (1) for all possible  $P = 1, \dots, I$ ,  $Q = 1, \dots, J$  and  $R = 1, \dots, K$  (e.g., Timmerman and Kiers's DIFFIT method [10]–[12]). Due to the repeated Tucker3 fitting

computations involved, the Tucker fitting based methods are expensive computationally so that it is usually suitable for small scale problems. The PCA based methods first unfolds the observed tensor  $\underline{\mathbf{Y}}$  to the large matrices in an appropriate way and then separately chooses the number of components via PCA for each mode. The typical PCA based methods include Cattell's "scree-test" [13], Bartlett's chi-square test [14], Niesing's QDA method [15], etc.

Since the PCA based methods do not need any Tucker3 decompositions, they can rapidly choose the number of components even for some middle scale problems. For this reason, although the Tucker fitting based methods are more precise, the PCA based methods are often used in practice. In this paper, we further discuss Tucker3 model selection under separate PCA based framework and develop more reliable method for Tucker3 model selection.

## II. THE SEPARATE PCAS BASED TUCKER3 MODEL SELECTION

The tensor  $\underline{\mathbf{Y}}$  can be unfolded into a matrix in different ways. Here we particularly consider a special matricization  $\mathbf{Y} \in \mathbb{R}^{I \times JK}$  given by

$$(\mathbf{Y})_{i, j+(k-1)K} = (\underline{\mathbf{Y}})_{i, j, k} = y_{i, j, k}, \quad (3)$$

where  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  and  $k = 1, \dots, K$ . The unfolded matrix  $\mathbf{Y}$  is shown in (4). From [16], [17], we have

$$\mathbf{Y} = \underline{\mathbf{A}} \cdot \underline{\mathbf{G}} \cdot (\underline{\mathbf{C}}^T \otimes \underline{\mathbf{B}}^T) + \underline{\mathbf{E}}, \quad (5)$$

where the operator  $\otimes$  is the Kronecker product, the matrices  $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times R}$  and  $\underline{\mathbf{E}} \in \mathbb{R}^{I \times JK}$  are analogously unfolded from the tensors  $\underline{\mathbf{G}}$  and  $\underline{\mathbf{E}}$ , respectively. The expression (5) can be re-written as

$$\mathbf{Y} = \underline{\mathbf{A}} \cdot \mathbf{X} + \underline{\mathbf{E}}, \quad (6)$$

where  $\mathbf{X} = \underline{\mathbf{G}} \cdot (\underline{\mathbf{C}}^T \otimes \underline{\mathbf{B}}^T) \in \mathbb{R}^{P \times JK}$ . Note that (6) is separable in column, i.e.,

$$\mathbf{y}_t = \underline{\mathbf{A}} \cdot \mathbf{x}_t + \mathbf{e}_t, \quad t = 1, \dots, JK. \quad (7)$$

### A. Noisy PCA

Consider a noisy PCA model:

$$\mathbf{y} = \underline{\mathbf{A}} \cdot \mathbf{x} + \mathbf{e}, \quad (8)$$

where the source vector  $\mathbf{x} = (x_1, \dots, x_P)^T$ , the noise vector  $\mathbf{e} = (e_1, \dots, e_I)^T$ . Let's make the following assumptions:

1) The noise components  $e_1, \dots, e_I$  are mutually independent and identically distributed to  $N(0, \sigma^2)$ ;

Z. He is with Lab. for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, 3510198, Japan, and School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510641, China (E-mail: he\_shui@tom.com).

A. Cichocki is with Lab. for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, 3510198, Japan, System Research Institute, Polish Academy of Sciences (PAN), Warsaw, 00-901, Poland, and Warsaw University of Technology, Warsaw, 00-661, Poland (E-mail: cia@brain.riken.jp).

$$\mathbf{Y} = \begin{bmatrix} y_{1,1,1} & y_{1,2,1} & \cdots & y_{1,J,1} & y_{1,1,2} & y_{1,2,2} & \cdots & y_{1,J,2} & \cdots & y_{1,1,K} & y_{1,2,K} & \cdots & y_{1,J,K} \\ y_{2,1,1} & y_{2,2,1} & \cdots & y_{2,J,1} & y_{2,1,2} & y_{2,2,2} & \cdots & y_{2,J,2} & \cdots & y_{2,1,K} & y_{2,2,K} & \cdots & y_{2,J,K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ y_{I,1,1} & y_{I,2,1} & \cdots & y_{I,J,1} & y_{I,1,2} & y_{I,2,2} & \cdots & y_{I,J,2} & \cdots & y_{I,1,K} & y_{I,2,K} & \cdots & y_{I,J,K} \end{bmatrix}. \quad (4)$$

2) The noise components  $e_1, \dots, e_I$  are also statistically independent to the source components  $x_1, \dots, x_P$ .

From above two assumptions, we have

$$\mathbf{V} = \text{cov}(\mathbf{y}, \mathbf{y}) = \mathbf{A} \cdot \text{cov}(\mathbf{x}, \mathbf{x}) \cdot \mathbf{A}^T + \sigma^2 \mathbf{I}, \quad (9)$$

where  $\mathbf{I}$  is an identity matrix. From (8), we have  $P = \text{rank}(\mathbf{A})$  if the noise is free (i.e.,  $e = 0$ ). Apply eigenvalue decomposition (EVD) on matrix  $\mathbf{V}$  as

$$\text{EVD}(\mathbf{V}) = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (10)$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_I)$ ,  $\lambda_1, \dots, \lambda_I$  are the eigenvalues and  $\mathbf{U}$  is the corresponding eigenvector matrix. Since  $\text{rank}(\mathbf{A}) = P < I$ , we can obtain

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P > \lambda_{P+1} = \dots = \lambda_I = \sigma^2. \quad (11)$$

### B. The existing PCA based methods for Tucker3 model selection

Starting from (11), a variety of PCA based methods have been proposed to determine the number  $P$  of nonzero-eigenvalue components in Tucker3 analysis (see [10]). For instance, Tucker [18] suggested directly using the ‘‘scree-test’’. The scree-test plots the magnitude of eigenvalues in a descending order against their ordinal numbers. Then it discards the components according to the Kaiser criterion [10], which is to retain only those components corresponding to the eigenvalues greater than the mean eigenvalue.

The second method is based on Bartlett’s chi-square test [10], [19], where the *null hypothesis* that the remaining eigenvalues are equal is tested. Each eigenvalue, ordered from the largest to the smallest, is excluded sequentially until the null hypothesis is not rejected. Then the components related to the excluded eigenvalues are retained.

The third alternative method uses the minimum average partial rule (MAP) to choose the number of components [20], [21], which employs a matrix of partial correlations between the variables with components partialled out. The components are partialled out sequentially, beginning with the component explaining most of the variance, and each time the average of the squared partial correlations is computed. The MAP method chooses the number of components based on the property that this average partial correlation usually first decreases, but then, after reaching a minimum, it starts increasing again.

In addition, Niesing [15] proposed to solve this problem by finding an appropriate value of  $p$  to maximize the following quotient of differences in additional values (QDA)

$$\text{QDA}(p) = \frac{\lambda_p - \lambda_{p+1}}{\lambda_{p+1} - \lambda_{p+2}}, \quad (12)$$

subject to the Kaiser constraint  $\lambda_p > \frac{1}{P} \sum_{p=1}^P \lambda_p$ .

Above four methods were compared by Monte Carlo studies in [10], [15], [21]. It was reported that QDA performed better than the other methods.

### III. IMPROVED PCA BASED METHOD FOR TUCKER3 MODEL SELECTION

The QDA method works well in many situations. However, it tends to underestimate the number of components. Consider the following problem:

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_P > \lambda_{P+1} = \dots = \lambda_I = \sigma^2. \quad (13)$$

The QDA method is likely to find the wrong estimation  $\hat{P} = 1 < P$  in this case. To avoid this problem, we propose a GAP measure as follows:

$$\text{GAP}(p) = \begin{cases} \frac{\text{var}\{\bar{\lambda}_i\}_{i=p+1}^{I-1}}{\text{var}\{\bar{\lambda}_i\}_{i=p}^{I-1}}, & \text{var}\{\bar{\lambda}_i\}_{i=p}^{I-1} \neq 0 \\ +\infty, & \text{var}\{\bar{\lambda}_i\}_{i=p}^{I-1} = 0 \end{cases} \quad (14)$$

where  $p = 1, \dots, I-2$ ,  $\text{var}(\{\bar{\lambda}_i\}_{i=p}^{I-1})$  denotes the sample variance of the sequence  $\{\bar{\lambda}_i\}_{i=p}^{I-1}$  and  $\bar{\lambda}_i = \lambda_i - \lambda_{i+1}$ ,  $i = 1, \dots, I-1$ . We choose the number of mode- $m_A$  components of tensor  $\underline{\mathbf{Y}}$  by the criterion:

$$\hat{P} = \arg \min_{p=1, \dots, I-2} \text{GAP}(p). \quad (15)$$

Only three continuous eigenvalues  $\lambda_p, \lambda_{p+1}, \lambda_{p+2}$  are taken into account in the QDA expression (12), which causes the underestimation of the component number sometimes. In contrast, all smaller eigenvalues from  $\lambda_p$  to  $\lambda_I$  are involved in (14). In addition, the Kaiser constraint is removed in the GAP method. It is easy to check that the GAP method can exactly find  $\hat{P} = P$  in (13).

From above discussions, the GAP algorithm can be briefly described as follows:

#### Choosing the number of components in $m_A$ -mode for Tucker3 model (1)

$$\hat{P} = \text{GAP}(\underline{\mathbf{Y}}, m_A)$$

**Step 1:** Unfold the tensor  $\underline{\mathbf{Y}}$  to an  $I \times JK$  matrix by (3).

**Step 2:** Compute the sample covariance matrix  $\hat{\mathbf{V}} = \frac{1}{JK} \sum_{t=1}^{JK} \mathbf{y}_t \mathbf{y}_t^T$  from (7).

**Step 3:** Compute the eigenvalue sequence  $\{\lambda_i\}_{i=1}^I$  of  $\hat{\mathbf{V}}$  such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_I$ .

**Step 4:** Compute the difference quotient GAPs:  $\text{GAP}(p)$ ,  $p = 1, \dots, I-2$  by (14).

**Step 5:** Choose the number  $\hat{P}$  of components of  $m_A$ -mode by (15).

**Step 6:** output  $\hat{P}$ .

To choose the number  $Q$  of  $m_B$ -mode components of tensor  $\underline{\mathbf{Y}}$ , instead of tensor matricization (3), we can unfold  $\underline{\mathbf{Y}}$  to be a  $J \times KI$  matrix arranged as

$$(\mathbf{Y})_{j,k+(i-1)I} = (\underline{\mathbf{Y}})_{i,j,k} = y_{i,j,k}, \quad (16)$$

where  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  and  $k = 1, \dots, K$ . Then we obtain  $\hat{Q} = \text{GAP}(\underline{\mathbf{Y}}, m_B)$  in the same way as the selection of  $\hat{P}$ . By analogy, we can choose the number  $R$  of  $m_C$ -mode components by  $\hat{R} = \text{GAP}(\underline{\mathbf{Y}}, m_C)$ , where the corresponding tensor matricization is

$$(\mathbf{Y})_{k,i+(j-1)J} = (\underline{\mathbf{Y}})_{i,j,k} = y_{i,j,k}, \quad (17)$$

where  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  and  $k = 1, \dots, K$ .

#### IV. EXPERIMENTAL EXAMPLES

In this section, we test the performance of the GAP method and compare it with other methods. Since all methods can rapidly find the results, we only compare their percentage of correct selection in different noise levels.

*Example 1:* Consider a noisy Tucker3 example with the following parameters:  $(I, J, K) = (32, 33, 34)$ ,  $(P, Q, R) = (28, 29, 30)$ . Totally 500 Monte Carlo trials were conducted. In each Monte Carlo trial, the core tensor  $\underline{\mathbf{G}}$  and three factor matrices  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  were randomly drawn from a uniform distribution  $U[-0.5, 0.5]$ . Then we obtain the observations by  $\underline{\mathbf{Y}} = \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} + \underline{\mathbf{E}}$  in each Monte Carlo trial, where  $\underline{\mathbf{E}}$  is the white Gaussian noise. The noise levels are shown in Fig 1(a).

We respectively applied the MAP method [20]–[22], the QDA method [15] and the GAP method on the tensor  $\underline{\mathbf{Y}}$  in each Monte Carlo trial. The percentages of correct selection were plotted in Fig 1(a), from which we can see that the MAP method was sensitive to noise compared with the other two methods, and the GAP method significantly outperformed the QDA method when  $\text{SNR} < 15\text{dB}$ .

*Example 2:* In this example, we consider the scenario where the core tensor  $\underline{\mathbf{G}}$  is much smaller the observed tensor  $\underline{\mathbf{Y}}$ , i.e.,  $(P, Q, R) \ll (I, J, K)$ . Both the core tensor  $\underline{\mathbf{G}}$  and the noise tensor  $\underline{\mathbf{E}}$  as well as three matrices  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  were randomly generated in the same manner as Example 1. The dimension of  $\underline{\mathbf{Y}}$  is the same as Example 1, i.e.,  $(I, J, K) = (32, 33, 34)$ . The difference is that here  $(P, Q, R) = (2, 3, 4) \ll (I, J, K) = (32, 33, 34)$ . Similarly, 500 Monte Carlo trials were conducted.

Fig.1(b) plots the detailed results. It is shown that the problem of Tucker3 model selection is relatively easy when  $(P, Q, R) \ll (I, J, K)$ . For example, the GAP method succeeded in all Monte Carlo trials even in the very noisy cases (e.g.,  $\text{SNR} = -1\text{dB}$ ).

#### V. CONCLUSIONS

The Tucker3 model selection has been discussed under the PCA based framework in this paper. An improved PCA based Tucker3 model selection algorithm has been developed. Since it does not involve any complex Tucker3 decompositions at all during the model selection, it is fast and easy to implement.

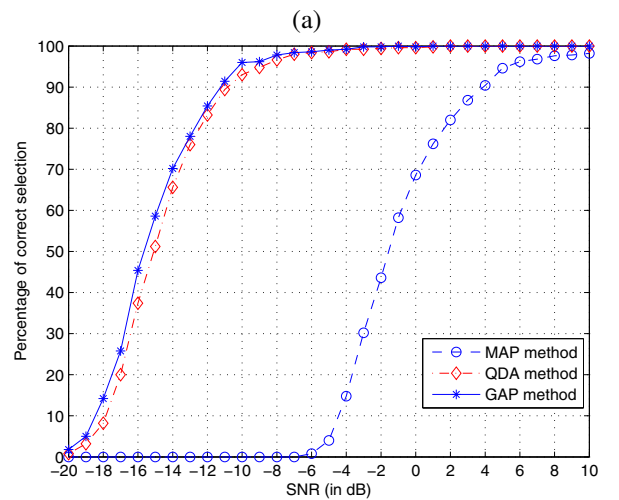
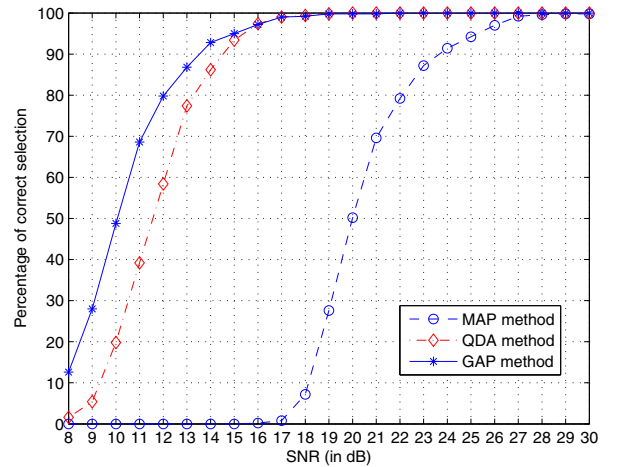


Fig. 1. The percentage of correct selection versus the noise level (a)  $(P, Q, R) = (28, 29, 30)$ ; (b)  $(P, Q, R) = (2, 3, 4)$ .

In practice, it can play an important role in Tucker3 analysis where the number of components is completely unknown. At least, it can be used as a preprocessing/initialization tool to detect the number of salient components in such applications.

#### REFERENCES

- [1] H. A. Kiers, "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.
- [2] R. Bro, "PARAFAC. tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149–171, October 1997.
- [3] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*. New York: John Wiley & Sons, August 2005.
- [4] L. D. Lathauwer, J. Castaing, and J.-F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2965–2973, June 2007.
- [5] L. D. Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1096–1105, 2008.
- [6] M. Haardt, F. Roemer, and G. D. Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, July 2008.
- [7] N. Sidiropoulos, G. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 810–823, March.

- [8] A. Shashua, R. Zass, and T. Hazan, "Multi-way clustering using supersymmetric non-negative tensor factorization," *Lecture Notes in Computer Science*, vol. 3954, pp. 595–608, July 2006.
- [9] R. Badeau and R. Boyer, "Fast multilinear singular value decomposition for structured tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1008–1021, 2008.
- [10] M. E. Timmerman and H. A. L. Kiers, "Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima," *British journal of mathematical and statistical psychology*, vol. 53, no. 1, pp. 1–16, 2000.
- [11] H. A. L. Kiers and A. der Kinderen, "A fast method for choosing the numbers of components in Tucker3 analysis," *British journal of mathematical and statistical psychology*, vol. 56, no. 1, pp. 119–125, May 2003.
- [12] E. Ceulemans and H. A. L. Kiers, "Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method," *British Journal of Mathematical and Statistical Psychology*, vol. 59, no. 1, pp. 133–150, May 2006.
- [13] R. B. Cattell, "The meaning and strategic use of factor analysis," *Handbook of multivariate experimental psychology*, pp. 174–243, 1966, Chicago: Rand McNally.
- [14] M. S. Bartlett, "Tests of significance in factor analysis," *British journal of mathematical and statistical psychology*, vol. 3, pp. 77–85, 1950.
- [15] J. Niesing, *Simultaneous component and factor analysis methods for two or more groups: a comparative study*, 2nd ed. Leiden: The Netherlands: DSWO Press, Leiden University, 1997.
- [16] A. Kapteyn, H. Neudecker, and T. Wansbeek, "An approach to  $n$ -mode components analysis," *Psychometrika*, vol. 51, no. 2, pp. 269–275, June 2003.
- [17] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, to appear (accepted June 2008).
- [18] L. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, September 1966.
- [19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1999.
- [20] W. Velicer, "Determining the number of components from the matrix of partial correlations," *Psychometrika*, vol. 41, no. 3, pp. 321–327, September 1976.
- [21] W. R. Zwick and W. F. Velicer, "Comparison of five rules for determining the number of components to retain," *Psychological Bulletin*, vol. 99, no. 3, pp. 432–442, May 1986.
- [22] W. F. Velicer, C. Eaton, and J. Fava, "Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components," in *Problems and Solutions in Human Assessment: Honoring Douglas Jackson at Seventy*, R. D. Goffin and E. Helmes, Eds. Boston: Kluwer Aca, 2000, pp. 41–71.