Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 3, pp. 519–527, 2022

TENSOR SYMMETRIZATION AND ITS APPLICATIONS IN GENERALIZED PRINCIPAL COMPONENT ANALYSIS*

Changqing Xu

Abstract: Generalized Principal Component Analysis or GPCA, is an algebraic geometric approach to estimate a mixture of linear subspaces provided that some data points are given. The GPCA is equivalent to factoring a homogeneous polynomial with factors representing normal vectors to each subspace. A formula for the number of subspaces and an analytic solution to the factorization problem are given in this paper in tensor form. We present a necessary and sufficient condition for the GPCA with two subspaces of uniform dimension. We also use the symmetrization of a rank-1 tensor to present a necessary and sufficient condition for the genral problem of GPCA.

Key words: *tensor, generalized principal component analysis, subspace, symmetrization*

Mathematics Subject Classification: *15A69, 15A86*

1 Introduction

Principal Component Analysis (PCA) is a standard tool in data analysis– in diverse fields such as the neuroscience, computer graphics, pattern recognition, data compression, image analysis and regressions etc.. As a simple non-parametric method for extracting relevant information from confusing data sets, PCA provides a roadmap to reduce a complex data set to a lower dimension to reveal the latent structure of the observations. Several kinds of approaches for the dimension reduction have been developed since 1901 when PCA was first formulated in statistics by S. Pearson[12] in 1901. PCA was also mentioned by Fisher and MacKenzie [5] in 1923 for the modeling of the response. The current form of PCA shall be owned to Hotelling [8] who developed PCA to its present stage in 1933. In the 1930s Thurstone and other psychologists developed the theory of factor analysis (FA), which is closely related to PCA. The utility of PCA has been rediscovered in many fields and goes with different titles e.g. singular value decomposition (SVD) in numerical analysis and the Karhunen-LoCve expansion in electrical engineering, eigenvector analysis and characteristic vector analysis in the physical sciences. In image analysis, the term Hotelling transformation is often used for a principal component projection. In chemistry, PCA was introduced by Malinowski around 1960 under the name principal factor analysis. In geology, PCA has lived a more secluded life, partly overshadowed by its twin brother FA.

As a simple popular tool for data mining, PCA has several different versions. A traditional one is to calculate the leading components of the data set based on the covariance

*This work was partially supported by the program No. 412011102 from Jiangsu province.

matrix A of the dataset. The geometric interpretation of PCA can be reformulated as the finding of an alternative basis for the space of the data set such that the data points under the new basis can be best expressed. While it is possible that all data points are drawn from the same subspace, which is exactly the assumption in PCA, there are cases when the sample points may be collected from several different subspaces. In this situation, the PCA may be invalid. The generalized PCA, or briefly, GPCA, is such an extended PCA where multiple subspaces are allowed.

Throughout the paper we denote $R, \mathcal{C}$ resp. for the field of real numbers and complex numbers and write $[n]$ for the set $\{1, 2, \ldots, n\}$ for any given positive integer n. A tensor $\mathcal{A} = (A_\sigma)$ of size $\mathrm{I} := \mathrm{I}_1 \times \mathrm{I}_2 \times \cdots \times \mathrm{I}_\mathrm{m}$ is defined as an *m-array* or a *multiway matrix* ($A_{i_1 i_2 \ldots i_m} := A_\sigma$ when $\sigma := (i_1, \ldots, i_m) \in \mathrm{I}$). A (column) vector corresponds to $m = 1$ and a matrix corresponds to $m = 2$. An m-order tensor $\mathcal{A} = (A_\sigma)$ is said to be indexed by I if $\sigma = (i_1, i_2, \ldots, i_m) \in \mathrm{I}$. $\mathcal{A}$ indexed by I is called an m-order n-dimensional real tensor or an $m \times n$ tensor provided that $I_1 = I_2 = \cdots = I_m = [n]$. Denote the set of all the m-order tensors indexed by I by $\mathcal{T}(\mathrm{I})$ and the set of all m-order n-dimensional real tensors by $\mathcal{T}_{m;n}$. An m-tuple $\sigma \in \mathrm{I}$ is sometimes identified with an m-multiset or an m-permutation chosen from $[n]$ with displacement allowed. Denote by $\mathcal{T}_m$ the set of all m-order tensors. An $m \times n$ tensor $\mathcal{A}$ is called a *symmetric tensor* if each entry A_σ is invariant under any permutation on its indices. Note that the degree of freedom (dof) of a symmetric $m \times n$ tensor is $\binom{n+m-1}{m}$ while the dof of a general $m \times n$ tensor is n^m. An m-order n-dimensional real tensor $\mathcal{A}$ is associated with an m-order homogeneous polynomial $f_\mathcal{A}(\mathbf{x}) := \mathcal{A}\mathbf{x}^m$, or more specifically

$$f_\mathcal{A}(\mathbf{x}) := \sum_{i_1, i_2, \ldots, i_m} A_{i_1 i_2 \ldots i_m} x_{i_1} x_{i_2} \ldots x_{i_m}. \tag{1.1}$$

$\mathcal{A}$ is called *positive definite* or pd (*positive semidefinite* or psd) if

$$f_\mathcal{A}(\mathbf{x}) := \mathcal{A}\mathbf{x}^m > 0 (\geq 0), \forall \mathbf{x} \in R^n \backslash \{0\} \, (\forall \mathbf{x} \in R^n). \tag{1.2}$$

A polynomial $f_A(\mathbf{x})$ is uniquely determined by the coefficient matrix A when A is symmetric.

Given N data points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in R^d$. It is assumed in the generalized PCA (GPCA) that all data points fall into some subspaces say

$$S_1, S_2, \ldots, S_n \subset R^d,$$

where the number n of the subspaces shall be estimated before the determination of the subspaces $S_1, S_2, \ldots, S_n$. Note that the problem of GPCA is reduced to PCA when $n = 1$. In the next section we mainly discuss the issue under the restriction $n \leq 2$. The problem in this case is transformed into the decomposition of a polynomial with the symmetrization of its coefficient tensor. In the third section, we take care of a more general case where the decomposition of its coefficient tensor is much more complicated.

2 GPCA with Two Subspaces

Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in R^d$. Denote $X = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Then $n \leq d$ if X is linearly independent. However, n cannot be upper bounded if X is pairwise linearly independent.

Lemma 2.1. *Let $n, d > 1$ be positive integers. Then there exist n vectors $\alpha_1, \alpha_2, \ldots, \alpha_n \in R^d$ such that α_i, α_j are linearly independent for any distinct pair $i, j \in [n]$.*

Proof. We may choose $\mathbf{u}, \mathbf{v} \in R^d$ such that $\mathbf{u}, \mathbf{v}$ are linearly independent. Now we choose n distinct real numbers say $\lambda_1, \lambda_2, \ldots, \lambda_n$, and let $\alpha_j = \mathbf{u} + \lambda_j \mathbf{v}$ for all $j \in [n]$. Then for any pair $i, j \in [n], i \neq j$, α_i, α_j are linearly independent. □

We now introduce the homogeneous coordinate before we go to the concept of affine space. Homogeneous coordinates are generally found in designing and construction applications as well as in computer vision. It is used to combine the geometric transformations such as the translation and the rotation into a single transformation. In homogeneous coordinate system, a 2-dimensional coordinate position (x, y) is represented by a triple-coordinate (X, Y, Z) where $x = X/Z, y = Y/Z$ when $Z \neq 0$ and $(X, Y, 0)$ refers to the point at infinity along the direction of the ray originated from $(0, 0)$ to point (X, Y). An affine set (space) $\mathcal{F}$ in vector space R^d is the set consisting of all possible affine combinations of some points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in R^d$, i.e., $\mathbf{x} = \sum_{j=1}^{k} \theta_j \mathbf{x}_j$ with all $\theta_j \in R$ satisfying $\theta_1 + \theta_2 + \cdots + \theta_k = 1$. An affine set is a lift of a linear space V, i.e., $\mathcal{F} = \alpha + V$ where $\alpha \in \mathcal{F}$ is a vector in $\mathcal{F}$. The dimension of an affine space F is defined as the dimension of V, denoted by $\dim(F) = \dim(V)$. Under the homogeneous coordinate system, each hyperplane H in R^d can be expressed as the set $\left\{\bar{\mathbf{x}} \in R^{d+1} : \bar{\alpha}^\top \bar{\mathbf{x}} = 0\right\}$ where $\bar{\alpha}^\top = (\alpha, 1)$ and α is the normal vector of H.

Now we assume that the data points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in R^d$ be on affine subspaces S_1, $S_2, \ldots, S_n$ of R^d where $n << N$ (n is usually very small) and $\dim(S_i) < d$ for each i. A simple case is when all S_i's are of the same dimension, say k, with $k \in [d-1]$, i.e.,

$$1 \leq \dim(S_1) = \dim(S_2) = \cdots = \dim(S_n) = k < d.$$

When $k = d-1$, each S_i is an affine subspace of R^d which can be characterized by a normal vector $\mathbf{w}_i \in R^d$, i.e.,

$$S_i = \left\{\mathbf{x} \in R^d : \mathbf{x}^\top \mathbf{w}_i = c_i\right\}, i \in [n],$$

with each $c_i \in R$ being a real number. The homogeneous representation of S_i is given by

$$S_i = \left\{\mathbf{y} \in R^{d+1} : \mathbf{y}^\top \mathbf{u}_i = 0\right\}, i \in [n]. \tag{2.1}$$

where $\mathbf{u}_i = (\mathbf{w}_i^\top, c_i)^\top, i \in [n] \in R^{d+1}$ and $\mathbf{y}$ is the homogeneous representation of $\mathbf{x}$ which corresponds to point $\mathbf{x} = (y_1/y_{d+1}, y_2/y_{d+1}, \ldots, y_d/y_{d+1})^\top \in R^d$ for nonzero y_{d+1} ($\mathbf{y}$ corresponds to a point at infinity if $y_{d+1} = 0$). The assumption of the distinction of $S_1, S_2, \ldots, S_n$ is equivalent to pairwise linear independency of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ in R^{d+1}. Now we form the data matrix

$$X = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in R^{(d+1)\times N} \tag{2.2}$$

in terms of the given points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in R^d$ and denote

$$W = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n] \in R^{(d+1)\times n} \tag{2.3}$$

(we may assume that $n << N$) and let

$$f(x) = \prod_{i=1}^{n} (\mathbf{y}^\top \mathbf{u}_i) \tag{2.4}$$

For case $n = 1$, all the observations must lie in an affine plane, which corresponds to a linear form f. The optimal rank-1 approximation $X \approx \lambda \mathbf{u}\mathbf{v}^\top$ can be achieved by choosing $\lambda = \sigma_1$, i.e., the largest singular value of X, and $\mathbf{u}, \mathbf{v}$ being respectively the normalized left and right singular eigenvector corresponding to σ_1.

In order to investigate the case $n = 2$, we present some basic terminology and some lemmas. Given a matrix $A = (a_{ij}) \in \mathcal{C}^{n\times n}$ and let $A_s := \frac{1}{2}(A + A^\top)$. Then A_s is the

symmetrical part of A. Now we let $\mathcal{A} = (A_{i_1 i_2 \ldots i_m})$ be any m-order n-dimensional real tensor . The *symmetrization* of $\mathcal{A}$, also denoted by $\mathcal{A}_s = (a_{i_1 i_2 \ldots i_m})$, is defined by

$$a_{i_1 i_2 \ldots i_m} := \frac{1}{m!} \sum_{\sigma \in S_m} A_{i_{\sigma(1)} i_{\sigma(2)} \ldots i_{\sigma(m)}},$$

where the summation is taken over all elements in the permutation group S_m imposed on set $[m]$. Obviously that $\mathcal{A}_s$ is a symmetric tensor by definition. More specifically an m-order n-dimensional real tensor $\mathcal{A}$ can be written as

$$\mathcal{A} = \alpha_1 \times \alpha_2 \times \cdots \times \alpha_m \tag{2.5}$$

if and only if $\mathcal{A}$ is a rank-1 tensor ($0 \neq \alpha_j \in \mathcal{C}^n$ for each $j \in [m]$). We say $\mathcal{A}$ is *separable* if α_j's are mutually independent. A separable rank-1 tensor in form (2.5) can be symmetrized in the following way:

$$\mathcal{A}_s = \frac{1}{m!} \sum_{j_1, j_2, \ldots, j_m} \alpha_{j_1} \times \alpha_{j_2} \times \cdots \times \alpha_{j_m}, \tag{2.6}$$

where the summation is taken through all the permutations on $[m]$ and $\alpha_j \in \mathcal{C}^n$. By definition, a separable tensor is a real symmetric tensor generated by the symmetrization of a rank-1 (possibly complex) tensor, and $\mathcal{A}$ and $\mathcal{A}_s$ are associated with the same polynomial, i.e., $f_{\mathcal{A}_s}(\mathbf{x})$ is identical to $f_{\mathcal{A}}(\mathbf{x})$.

There are infinitely many coefficient tensors $\mathcal{A}$ which can be associated with a polynomial $f(x)$ defined by (1.1). But this correspondence becomes unique when the coefficient tensor is required to be symmetric, and $f_{\mathcal{A}}(\mathbf{x}) = f_{\mathcal{A}_s}(\mathbf{x})$. Now conversely we assume that we are given a symmetric tensor $\mathcal{A} \in \mathcal{T}_{m,n}$. We want to know whether it is possible to be regarded as a symmetrization of a separable rank-1 tensor. There is no easy answer to this question. Let us first consider the matrix case, that is,

Problem 2.2. Given a symmetric matrix $A = (a_{ij}) \in R^{n \times n}$. Is that possible to find two linearly independent vectors $\alpha, \beta \in \mathcal{C}^n$ such that

$$A = \frac{1}{2}(\alpha \times \beta + \beta \times \alpha). \tag{2.7}$$

Problem (2.2) can be reformulated as

Problem 2.3. Given a quadratic form $f(x)$ of n-variate. Is that possible to factorize f into the product of two different linear forms, i.e.,

$$f(x) = l_1(x) l_2(x), \tag{2.8}$$

where $x = (x_1, x_2, \ldots, x_n)^\top \in R^n$ is the vector of variables and $l_i(x) = \sum_{j=1}^{n} b_{ij} x_j$, and $b_i = (b_{i1}, \ldots, b_{in})^\top$ are linearly independent.

Note that when $l_1(x) \equiv l_2(x)$ in (2.8), $f(x) = l_1^2(x)$ which corresponds to the case $n = 1$. To present the necessary and sufficient condition for a symmetric matrix to be separable, we recall that the *inertia* of a real symmetric matrix A, denoted $\texttt{inert}(A)$, is a triple (p, q, s) , representing respectively the number of the positive eigenvalues, negative eigenvalues and the eigenvalues equal to zero. It is commonly known that a real symmetric

matrix $A \in R^{n\times n}$ is positive semidefinite (psd) if and only if $\texttt{inert}(A) = (r, 0, n-r)$ where $r = \texttt{rank}(A)$. Furthermore, A is positive definite (pd) if and only if $\texttt{inert}(A) = (n, 0, 0)$. From the knowledge of linear algebra, A preserves the inertia under the transformation of the congruence, i.e., for any invertible matrix $U \in R^{n\times n}$, we have $\texttt{inert}(UAU^\top) = \texttt{inert}(A)$.

Now we state the following result.

Theorem 2.4. *Given a nonzero symmetric matrix $A = (a_{ij}) \in R^{n\times n}$. Then A is separable if and only if $\texttt{inert}(A) = (1, 1, n-2)$.*

Proof. To prove the necessity, we suppose that A is separable and let

$$A = \frac{1}{2}(\alpha \times \beta + \beta \times \alpha) \tag{2.9}$$

with $\alpha, \beta \in R^n$ linearly independent. Then $\texttt{rank}(A) \le 2$. We denote $U = [\alpha_1, \alpha_2, \ldots, \alpha_n] \in R^{n\times n}$ where $\alpha_1 = \alpha, \alpha_2 = \beta$, and $\alpha_3, \ldots, \alpha_n$ are the extension of α_1, α_2 such that $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ forms a basis of R^n. Then by (2.9) we have $A = UDU^\top$ where $D = \texttt{diag}(D_1, 0) \in R^{n\times n}$ with $D_1 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$. Then we have $\texttt{inert}(A) = \texttt{inert}(D) = (1, -1, n-2)$.

To prove the sufficiency, we suppose that A is a symmetric matrix with $\texttt{inert}(A) = (1, 1, n-2)$. Then there exists an invertible matrix $U \in R^{n\times n}$ such that $A = UDU^\top$ with $D = \texttt{diag}(1, -1, 0, \ldots, 0)$. Denote $U = [u_1, u_2, \ldots, u_n]$ where $u_j \in R^n$ is the jth column vector of U for $j \in [n]$. Thus

$$A = u_1 u_1^\top - u_2 u_2^\top, \tag{2.10}$$

Now we denote $\alpha = u_1 + u_2, \beta = u_1 - u_2$. Then $\alpha, \beta \in R^n$ are linearly independent due to the independency of u_1, u_2, and by (2.10) we obtain

$$A = \frac{1}{2}(\alpha \times \beta + \beta \times \alpha) = \frac{1}{2}(\alpha\beta^\top + \beta\alpha^\top).$$

The proof is completed. □

The condition $\texttt{inert}(A) = (1, 1, n-2)$ in Theorem 2.4 implies that $\texttt{rank}(A) = 2$. An interesting phenomenon is that A is separable in the field of complex numbers but not in the field of real numbers when the condition $\texttt{inert}(A) = (1, 1, n-2)$ is relaxed to $\texttt{rank}(A) = 2$, as the following corollary shows:

Corollary 2.5. *Let $A = (a_{ij}) \in R^{n\times n}$ be a nonzero symmetric matrix. Then A is separable in $\mathcal{C}$ if and only if $\texttt{rank}(A) = 2$.*

Proof. Let $A = (a_{ij}) \in R^{n\times n}$ be a symmetric matrix. By Theorem 2.4, we need only to show the sufficiency, that is, if $\texttt{rank}(A) = 2$, then there are two linear independent vectors $\alpha, \beta \in \mathcal{C}^n$ such that (2.9) holds. By $\texttt{rank}(A) = 2$ and the symmetry of A, there are two two linear independent vectors $\mathbf{x}, \mathbf{y} \in R^n$ such that $A = \lambda_1 \mathbf{u}\mathbf{u}^\top + \lambda_2 \mathbf{v}\mathbf{v}^\top$ where $\lambda_1, \lambda_2 \in R$ are the first largest eigenvalues of A, i.e., $|\lambda_1| = \max\{|\lambda| : \lambda \in \sigma(A)\}$ where $\sigma(A)$ denotes the spectrum of A. Now we denote

$$\alpha = \sqrt{\lambda_1}\mathbf{u} + \imath\sqrt{\lambda_2}\mathbf{v}, \quad \beta = \sqrt{\lambda_1}\mathbf{u} - \imath\sqrt{\lambda_2}\mathbf{v}.$$

Then $\alpha, \beta \in \mathcal{C}^n$. An easy computation shows that

$$\alpha \times \beta + \beta \times \alpha = 2(\lambda_1 \mathbf{u}\mathbf{u}^\top + \lambda_2 \mathbf{v}\mathbf{v}^\top) = 2A.$$

Hence (2.9) holds. □

Corollary 2.5 presents an easy-to-check condition for a symmetric matrix to be separable in case $n = 2$. It can also be justified by the fundamental algebraic theorem, i..e, a real polynomial of order n always have n roots in the field of complex numbers.

Recall that an $(p, q)-$ commutation matrix $K_{p,q}$ is an $p \times q$ block matrix $[K_{ij}]$ where each block K_{ij} is an $q \times p$ matrix with its unique nonzero entry taking value 1 at position (j, i). This concept is extended to the commutation tensor $\mathcal{K}_{p,q}$ [16] which is a (0,1) tensor of size $p \times q \times q \times p$ whose entry $K_{ijkl} = 1$ only if $j = k, i = l$. The commutation tensor $\mathcal{K}_{p,q}$ transforms a matrix into its transpose as in the following manner [16]:

Lemma 2.6. *$\mathcal{K}_{q,p} \times X = X^\top$ for any matrix $X \in R^{p\times q}$.*

Here $\mathcal{A} \times X$ is defined as a matrix $B = (b_{ij})$ where $b_{ij} = \sum_{l,k} A_{ijkl}X_{kl}$ where $\mathcal{A} \in R^{q\times p\times p\times q}, X \in R^{p\times q}$.

Given a rank-1 m-order n-dimensional real tensor is $\mathcal{A} := \alpha_1 \times \alpha_2 \times \cdots \times \alpha_m$ with $\alpha_k \in R^n$ and a permutation $\tau \in \mathtt{sym}_m$ where $\mathtt{sym}_m$ denotes the pemutation group on the set $\{1, 2, \ldots, m\}$. We define an m-order n-dimensional real tensor $\mathcal{A}^\tau := (A^\tau_{i_1 i_2 \ldots i_m}) \in T_{m;n}$ as

$$A^\tau_{i_1 i_2 \ldots i_m} = A_{i_{\tau(1)} i_{\tau(2)} \ldots i_{\tau(m)}}.$$

The symmetrization of a rank-1 tensor $\mathcal{A} := \alpha_1 \times \alpha_2 \times \cdots \times \alpha_m$ is just the average of all $\mathcal{A}^\tau$s by (2.6). The symmetrization of a rank-1 tensor will be used in the following to establish the method to tackle the problem of GPCA.

3 GPCA in General Case

Given N data points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in R^d$ which are supposed to lie on some subspaces $S_1, S_2, \ldots, S_n$ of R^d. In this paper we only consider the simple case when all S_i's are with the same dimension, say $\dim(S_i) = k \in [d]$ for all i. For $d = 3, k = 2$, each subspaces S_i in R^d is uniquely determined by its normal vector say $\mathbf{u}_i$. The Generalized Principal Component Analysis (GPCA) is an algebraic-geometric approach proposed by Vidal in 2003 [15] to model mixtures of subspaces with a unique global solution to the clustering of the given points based on polynomial decomposition. By the homogeneous coordinate expression, the approach takes the mixture of subspaces as a projective algebraic variety which is estimated from sample data points as a particular case of NLPCA to derive the embedding of the data analytically. The subspaces in GPCA is estimated by using segmentation independent constraints satisfied by all data points, regardless of the subspace to which they belong.

The estimation of the n subspaces can be transformed into that of the algebraic variety defined by a set of polynomials. Moreover, the problem of identifying a collection of hyperplanes, i.e., $\dim(S_i) = d - 1$ for each $i \in [n]$, boils down to the the estimation and the factorization of $p_n(\mathbf{x})$. The polynomial $p_n(x)$ can be retrieved from the data points though we have no knowledge about the clustering of the points. $p_n(x)$ can be determined from the solution of a set of linear equations if n is known. On the other hand, the estimation of the hyperplanes is essentially equivalent to factoring $p_n(x)$ into a product of n linear factors.

Let $S \in R^{N\times N}$ be the similarity matrix generated by the data matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in R^{d\times N}$ and $\mathbf{u} = (u_1, u_2, \ldots, u_N)^\top$ be the eigenvector of S corresponding to the largest eigenvalue of S whose entries take value in $\pi := \{\mu_1, \mu_2, \ldots, \mu_n\}$ where it is assumed μ_i's are labeled in increasing order. Denote $\mathbf{x}^{[m]} := (x_1^m, x_2^m, \ldots, x_N^m)^\top$ and write

$$L_k(\mathbf{x}) := [\mathbf{x}^{[0]}, \mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \ldots, \mathbf{x}^{[k]}] \in R^{d\times(k+1)},$$

where $\mathbf{x}^{[0]} \in R^N$ is the all-ones vector, $\mathbf{x}^{[1]} = \mathbf{x}$. It is not difficult to see that the number n of the distinct entries of $\mathbf{x}$ can be determined by the following result:

Lemma 3.1.

$$n = \arg\min \{k : \mathbf{rank}(L_k(\boldsymbol{x})) = k\}. \tag{3.1}$$

Denote $L_k := L_k(\mathbf{x})$ when there is no risk of confusion. Note that $L_{k+1} = [L_k, \mathbf{x}^{[k+1]}]$, we can computer n iteratively from $k = 1$. Since S is a positive symmetric (and thus irreducible), the $\mathbf{x} \in R^N$ be the normalized largest positive vector with unit length. Denote $m_k := E[x^k] = \frac{1}{N}\sum_{j=1}^{N} x_j^k$, that is, the k-moment of $\mathbf{x}$. Write $A = (a_{ij}) \in R^{n\times n}$ where $a_{ij} = m_{i+j-2}$ and $\beta_{k,n} := (m_k, m_{k+1}, \ldots, m_{k+n-1}) \in R^n$. Then A is a Hankel matrix associated with $\beta_{0,2n-2}$. It is shown by [15] that

Lemma 3.2. *Let $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in R^{d\times N}$ whose coulmns are data points and let $n : 1 < n << N$ be a positive integer which is much smaller than N. Then there are n subspaces $S_1, S_2, \ldots, S_n$ with the same dimensions in R^d such that all $\boldsymbol{x}_j$'s lie on the union of these subspaces if and only if*

$$A\boldsymbol{x} = \beta_{k,n}. \tag{3.2}$$

Suppose there exist unknown n vectors $\mathbf{v}_i \in R^{d+1}$ each standing for a normal vector of subspace S_i for $i \in [n]$. Then a data point $\mathbf{x} \in R^{d+1}$ in homogenuous system lies on S_i if and only if $\mathbf{x}^\top \mathbf{v}_i = 0$. Thus the N data points lying on exactly n hyperspaces of R^d if and only if each data point satisfies

$$\prod_{i=1}^{n} (\mathbf{x}^\top \mathbf{v}_i) = 0. \tag{3.3}$$

Now we denote $\mathcal{Y}$ be an $(n+1)$-order tensor with size $\overbrace{d \times d \times \cdots \times d}^{n} \times N$ where

$$Y(:, :, \ldots, :, j) := \overbrace{\mathbf{x}_j \times \mathbf{x}_j \times \cdots \times \mathbf{x}_j}^{n}, \forall j \in [N],$$

and write $\mathcal{V} = \mathtt{sym}(\mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_n)$. Then we have the result:

Theorem 3.3. *Let $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in R^{d\times N}$ be the matrix whose coulmns are the given data points. Then these data points lie on some n different hypergraphs of R^d if and only if there exists a symmetrization of a rank-1 n-order d-dimensional tensor $\mathcal{V}$ satisfying*

$$\mathcal{Y} \times_{[n]} \mathcal{V} = 0, \tag{3.4}$$

where $\mathcal{Y} \times_{[} n]\mathcal{V}$ denotes the tensor multiplication of $\mathcal{Y}$ with $\mathcal{V}$ along the first n modes, i.e., $(\mathcal{Y} \times_{[} n]\mathcal{V})_j$ is the inner product of $Y(:, :, \ldots, :, j)$ with $\mathcal{V}$.

Proof. It is easy to see that (3.4) is equivalent to (3.3) if $\mathcal{V} = \alpha_1 \times \alpha_2 \times \cdots \times \alpha_n$ for some pairwise linearly independent vectors α_j's since for any given $j \in [N]$ we have

$$\begin{aligned} 0 = (\mathcal{Y} \times_{[n]} \mathcal{V})_j &= Y(:, :, \ldots, :, j) \times \mathcal{V} \\ &= (\mathbf{x}_j \times \mathbf{x}_j \times \cdots \times \mathbf{x}_j) \times (\alpha_1 \times \alpha_2 \times \cdots \times \alpha_n) \\ &= \prod_{i=1}^{n} (\mathbf{x}_j^\top \alpha_i) \end{aligned}$$

which implies that $\mathbf{x}_j$ satisfies condition (3.3). For the general case we can also show that (3.4) is equivalent to (3.3) by the symmetrization of a rank-1 tensor. □

Acknowledgement

The authors would like to thank the referees for their valuable suggestions and remarks which are helpful for the improvement of the writing of the manuscript.

References

[1] D. Cartwright and B. Sturmfels, The number of eigenvalues of a tensor, *Linear Algebra and Its Applications* 438 (2013) 942–952.

[2] H. Chen and L. Qi, Positive definiteness and semi-definiteness of even order symmetric Cauchy tensors, *J. Ind. Manag. Optim.* 11 (2017) 1263–1274.

[3] W.L. Chooi, K.H. Kwa and M.-H. Lim, Coherence invariant maps on tensor products, *Linear Algebra and Its Applications* 516 (2017) 24–46.

[4] P. Comon, G. Golub, L.-H. Lim and B. Mourrain, Symmetric tensors and symmetric tensor rank, *SIAM. J. Matrix Analysis and Applications* 30 (2008) 1254–1279.

[5] R. Fisher and W. MacKenzie, Studies in crop variation II. The manurial response of different potato varieties, *Journal of Agricultural Science* 13 (1923) 311–320.

[6] G. Frobenius, Über die Darstellung der endlichen Gruppen durch lineare Substitutionen, S.-B. Press, Akad. Wiss. Berlin, 1897, pp. 994–1015.

[7] R. Grone, Decomposable tensors as a quadratic variety, *Proc. Amer. Math. Soc.* 64 (1977) 227–230.

[8] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* 24 (1933) 417–441.

[9] Z. Huang and L. Qi, On determinants and eigenvalue theory of tensors, Journal of Symbolic Computation 50 (2013) 508–531.

[10] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed.. Springer Series in Statistics. New York: Springer-Verlag New York, 2002.

[11] T.G. Kolda and B. W. Bader, *Tensor Decompositions and Applications*, SIAM Review, 2009.

[12] S.K. Pearson, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine, Series 6, 2 (1901) 559–572.

[13] L. Qi, Eigenvalues of a real supersymmetric tensor, *Journal of Symbolic Computation* 40 (2005) 1302–1324.

[14] L. Qi, Symmetric nonnegative tensors and copositive tensors, *Linear Algebra and Its Applications* 439 (2013) 228–238.

[15] R. Vidal, *Generalized Principal Component Analysis (GPCA): An Algebraic Geometric Approach to Subspace Clustering and Motion Segmentation*, Ph.D. Thesis, Electrical Engineering and Computer Sciences, University of California at Berkeley, 2003.

[16] C. Xu, L. He and Z. Lin, Commutation matrices and commutation tensors, *Linear and Multilinear Algebra* 68 (2020) 1721–1742.

Manuscript received 12 May 2021
revised 23 June 2021
accepted for publication 3 September 2021

Changqing Xu
School of Mathematics, Suzhou University of Science and Technology, Suzhou, China
E-mail address: cqxurichard@usts.edu.cn