

High Order Tensor Forms of Growth Curve Models

Zerong Lin¹, Dongzhe Liu², Xueying Liu³, Lingling He¹, Changqing Xu^{1*}

¹School of Mathematics and Physics, Suzhou University of Science and Technology, Suzhou, China

²School of Energy and Environment, City University of Hong Kong, Hong Kong, China

³Zhejiang University of Water Resources and Electric Power, Hang Zhou, China

Email: *cqxrurichard@usts.edu.cn

How to cite this paper: Lin, Z.R., Liu, D.Z., Liu, X.Y., He, L.L. and Xu, C.Q. (2018) High Order Tensor Forms of Growth Curve Models. *Advances in Linear Algebra & Matrix Theory*, 8, 18-32.

<https://doi.org/10.4236/alamt.2018.81003>

Received: November 20, 2017

Accepted: February 26, 2018

Published: March 1, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we first study the linear regression model $y = X\beta + \varepsilon$ and obtain a norm-minimized estimator of the parameter vector β by using the g -inverse and the singular value decomposition of matrix X . We then investigate the growth curve model (GCM) and extend the GCM to a generalized GCM (GGCM) by using high order tensors. The parameter estimations in GGCMs are also achieved in this way.

Keywords

Tensor, Generalized Linear Model, Growth Curve Model, Parameter Estimation, Generalized Inverse

1. Introduction

Linear regression model, or called linear model (LM), is one of the most widely used models in statistics. There are many kinds of linear models including simple linear models, general linear models, generalized linear models, mixed effects linear models and some other extended forms of linear models [1] [2] [3] [4] [5]. The growth curve model (GCM) is a special kind of general linear models which have applications in many areas such as the psychology data analysis [6]. The GCMs can be used to handle longitudinal data or missing data or even the hierarchical multilevel mixed case [2] [3] [5] [7]-[12]. There are some variations of GCMs such as the latent GCMs which are also very useful. The traditional treatment of the GCMs in the estimation of the parameters in the case of mixed effects for a single response factor is usually to stack all the dependent observations vertically into a very long column vector, usually denoted by y , and all the design matrices (both the fixed effect design matrix and the random effect design ma-

trix), the random errors are accordingly concatenated to fit the size of y . This treatment makes the implementation of the related programming much slow due to the magnificent dimensions of the data (matrices and vectors). Things will get even worse if we encounter a huge dataset (big data) such as the data of genome, web related, image gallery, or social network etc.

In this paper, we first use the generalized inverse of matrices and the singular value decomposition to obtain the norm-minimized estimation of the parameters in the linear model. Then we introduce some basic knowledge about tensors before we employ tensors to express and extend the multivariate mixed effects linear models. The extended tensor form of the model can be also regarded as a generalization of the GCM.

Let us first begin with some basic linear regression models. We let y be a response variable and x_1, \dots, x_r be independent random variables for explaining y . The most general regression model between y and x_1, \dots, x_r is in form

$$y = f(x_1, \dots, x_r) + \varepsilon \quad (1.1)$$

where ε is the error term, and f is an unknown regression function. In linear regression model, f is assumed to be a linear function, *i.e.*,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir} + \varepsilon \quad (1.2)$$

where all β_i are unknown parameters. Denote $\mathbf{x} = (x_1, \dots, x_r)$ which is called a random vector. Let $P = (y, \mathbf{x})$, an $(r+1)$ -dimensional random vector, which is called an observable vector. Given N observations of P , say $P_i = (y_i, x_{i1}, x_{i2}, \dots, x_{ir})$, $i = 1, 2, \dots, N$. Here y_i stands for the i th observation of the response variable y , and $x_{i1}, x_{i2}, \dots, x_{ir}$ are the corresponding explanatory observations. The sample model of Equation (1.2) turns out to be

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir} + \varepsilon \quad (1.3)$$

or equivalently

$$y = X\beta + \varepsilon \quad (1.4)$$

where $y = (y_1, y_2, \dots, y_N)^T \in \mathbb{R}^N$ (here and throughout the paper T stands for the transpose of a matrix/vector) is the sample vector of the response variable y , $X = (x_{ij}) \in \mathbb{R}^{N \times (r+1)}$ is the data matrix or the design matrix each of whose rows corresponding to an observation of \mathbf{x} , $\beta = (\beta_0, \beta_1, \dots, \beta_r) \in \mathbb{R}^{r+1}$ is the regression coefficient vector, which is to be estimated, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$ is the random error vector. A general linear regression model, abbreviated GLM, is a LM (1.4) with the error terms ε_i satisfying:

- 1) Zero-mean: $E[\varepsilon_i] = 0, \forall i \in [N]$, *i.e.*, the expected value of the error term is zero for all the observations.
- 2) Homoskedasticity: $\text{Var}[\varepsilon] = \sigma^2$, *i.e.*, all the error term are distributed with the same variance.
- 3) Uncorrelation: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for all distinct i, j , *i.e.*, distinct error terms are uncorrelated.

Equations (1)-(3) is called Gauss-Markov assumption [13]. The model (1.4) under the Gauss-Markov assumption is called the Gauss-Markov model. Note that the variance reflects the uncertainty of the model, the zero-mean, homoskedasticity and the uncorrelation of the sample errors form the Gauss-Markov assumption. An alternative form of the Gauss-Markov model is

$$E[Y] = X\beta, \text{Cov}(\varepsilon) = \sigma^2 I_N \tag{1.5}$$

where I_N is the $N \times N$ identity matrix and $\sigma^2 > 0$. In order to investigate the general linear model and extend the properties, we recall some known results concerning the linear combinations of some random variables. Suppose $\alpha \in \mathbb{R}^n$ is a constant vector with the same length as that of y , the random vector under the investigation.

Let $A \in \mathbb{R}^{m \times n}$. The *g-inverse* of A , denoted A^g , is a generalized inverse defined as an $n \times m$ matrix satisfying [4] $AA^gA = A$. An equivalent definition for *g-inverse* is that $x = A^g b$ is always a solution to equation $Ax = b$ whenever $b \in \mathcal{C}(A)$, the column space of A . A well known result is that all the solutions to $Ax = b$ (when compatible) are in form

$$x = A^g b + (I - A^g A)\omega, \forall \omega \in \mathbb{R}^n. \tag{1.6}$$

It is easy to verify that $A^g = A^{-1}$ is unique when A is invertible. The *g-inverse* of a matrix (usually not unique) can be calculated by using singular value decomposition (SVD).

Lemma 1.1. Let $X \in \mathbb{R}^{N \times p}$ with a SVD decomposition $X = UDV^T$ such that $U \in \mathbb{R}^{N \times N}$ and $V \in \mathbb{R}^{p \times p}$ are orthogonal matrices, and $D \in \mathbb{R}^{N \times p}$ is in form $D = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$ where $r = \text{rank}(X) \leq \min(N, p)$, and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$. Then

$$A^g = V \begin{bmatrix} D_r^{-1} & * \\ * & * \end{bmatrix} U^T \in \mathbb{R}^{n \times m} \tag{1.7}$$

where $*$ denotes any matrix of suitable size and $D_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$.

The Gauss-Markov Theorem (e.g. Page 51 of [13]) is stated as:

Lemma 1.2. Suppose that model (1.4) satisfies the Gauss-Markov assumption and $a \in \mathbb{R}^N$ be a constant vector. Then $z = a^T y$ is estimable, and $a^T \hat{\beta}$ is the best (minimum variance) linear unbiased estimator (BLUE) of $a^T \beta$, with $\hat{\beta} = (X^T X)^g X^T y$.

Based on Lemma 1.2, we get

Proposition 1.3. If $\text{rank}(X) = r < \min(N, p)$ in Equation (1.4) and X satisfies condition in Lemma 1.2. Then the estimator of β with minimal 2-norm is in form

$$\hat{\beta} = V \begin{bmatrix} D_r^{-1} \tilde{y}_1 \\ 0 \end{bmatrix} \tag{1.8}$$

where $\tilde{y} = U^T y = [\tilde{y}_1^T, \tilde{y}_2^T]^T$ with $\tilde{y}_1 \in \mathbb{R}^r$, $\tilde{y}_2 \in \mathbb{R}^{N-r}$.

Proposition 1.3 tells us that by taking D as a block upper triangle form in the decomposition

$$(X^T X)^g = V \begin{bmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{bmatrix} V^T.$$

We can reach a norm-minimised estimator of β . Now denote $H := (X^T X)^g X$. By Gauss-Markov Theorem, we have

$$\|\hat{\beta}\|^2 = \langle Hy, Hy \rangle = y^T H^T Hy = y^T X (X^T X)^{2g} X^T y$$

which implies that $\|\hat{\beta}\| = \|(XX^T)^g y\|$.

The generalized linear model (GLM) is a generalization of LM [1]. In a GLM model some basic assumptions in linear regression model are relaxed. Also the fitting values of the response variables are no longer directly expressed as a linear combination of parameters, but rather a function which is usually called a link function. A GLM consists of the independent random components y_i in exponential distribution, the predictive value $X_i^T \beta$, the system components, and the link function f , strictly monotone differentiable function in GLM $\eta_i = f(X_i^T \beta)$. The parameters in a GLM include regression parameters β and the discrete parameters in the covariance matrix, both can be estimated with maximum likelihood method. The estimation of the regression parameter for model (1.4) can be expressed as

$$\beta^{(m)} = (X^T W^{(m-1)} X)^g X^T W^{(m-1)} z^{(m-1)}, W = \text{diag}(W_1, W_2, \dots, W_N)$$

where $W_i = w_i / \left\{ \phi v(\mu_i) [g^T(\mu_i)]^2 \right\}$ with w_i being a known priori weight, ϕ the dispersion parameter, $v(\cdot)$ a variance function, g a link function, and $z \in \mathbb{R}^N$ the *work dependent variable* with $z_i = \eta_i + (y_i - \mu_i) g^T(\mu_i)$. The moment estimation of discrete parameters is

$$\hat{\phi} = \frac{1}{N - k - 1} \sum_{i=1}^N \frac{w_i (y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)}. \quad (1.9)$$

In order to extend the GLMS to more general case, we need some knowledge on tensors. In the next section, we will introduce some basic terminology and operations implemented on tensors, especially on low order tensors.

2. The 3-Order Tensors and Their Applications in GLMs

A tensor is an extension of a matrix in the case of high order, which is an important tool to study high-dimensional arrays. The origin of tensor can be traced back to early nineteenth century when Cayley studied linear transformation theory and invariant representation. Gauss and Riemann et al. promoted the development of tensor in mathematics. In 1915 Albert Einstein used tensor to describe his general relativity, leading tensor calculus more widely accepted. In the early twentieth century, Ricci and Levi-Civita further developed tensor analysis in absolute differential methods and explored their applications [14].

For our convenience we denote $[n] := \{1, 2, \dots, n\}$ and use $S(m, n)$ to denote the index set

$$S(m, n) := \{ \tau = (i_1, i_2, \dots, i_m) : i_k \in [n], \forall k \in [m] \}.$$

Let $I_k (k \in [m])$ be any positive integer (usually larger than 1). Sometimes we abuse I_k as a set $[I_k]$. Denote $I := I_1 \times I_2 \times \dots \times I_m$. If I_k stands for an index set, then I is a tensor product of I_1, I_2, \dots, I_m . An m -order tensor $\mathcal{A} = (A_\sigma)$ of size I is an m -array whose entries are denoted by $A_\sigma := A_{i_1 i_2 \dots i_m}$ with $\sigma = (i_1, i_2, \dots, i_m) \in I$. Note that a vector is a 1-order tensor and an $m \times n$ matrix is a 2-order or second order tensor. An $m \times n$ tensor is a tensor with $I_1 = I_2 = \dots = I_m = n$. We denote by $\mathcal{T}_{m,n}$ the set of all m th order n -dimensional real tensors. An $m \times n$ tensor \mathcal{A} is called *symmetric* if A_σ is constant under any permutation on its index.

An m th order n -dimensional real tensors \mathcal{A} is always associated with an m -order homogeneous polynomial $f_{\mathcal{A}}(\mathbf{x})$ which is defined by

$$f_{\mathcal{A}}(\mathbf{x}) := \mathcal{A}\mathbf{x}^m = \sum_{i_1, i_2, \dots, i_m} A_{i_1 i_2 \dots i_m} x_{i_1} x_{i_2} \dots x_{i_m}. \tag{2.10}$$

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

$$f_{\mathcal{A}}(\mathbf{x}) := \mathcal{A}\mathbf{x}^m \geq 0, \forall \mathbf{x} \in \mathbb{R}^n. \tag{2.11}$$

A nonzero psd tensor must be of an even order. Let \mathcal{A} be of size $m \times n \times p$. Given an r -order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_r}$ and a matrix $U = (u_{ij}) \in \mathbb{R}^{I_k \times J_k}$ where $k \in [r]$. The product of \mathcal{A} with U along k -mode is defined as the r -order tensor $\mathcal{A} \times_k U$ defined as

$$(\mathcal{A} \times_k U)_{i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_m} = \sum_{i=1}^n A_{i_1, \dots, i_{k-1}, i, i_{k+1}, \dots, i_m} u_{ik}. \tag{2.12}$$

Note that $\mathcal{A} \times_k U$ is compressed into an $(m-1)$ -order tensor when $U \in \mathbb{R}^{I_k}$ is a column vector ($J_k = 1$). There are two kinds of tensor decomposition, *i.e.*, the rank-1 decomposition, also called the CP decomposition, and the Tucker decomposition, or HOSVD. The former is the generalization of matrix rank-1 decomposition and the latter is the matrix singular value decomposition in the high order case. A *zero tensor* is a tensor with all entries being zero. A *diagonal tensor* is a tensor whose off-diagonal elements are all zero, *i.e.*, $A_{i_1, i_2, \dots, i_m} = 0$ if i_1, i_2, \dots, i_m are not identical. Thus an $m \times n$ tensor has n diagonal elements. By this way, we can define similarly (and analogous to matrix case) the identity tensor and a scalar tensor.

For any $i \in [n]$, an *i -slice* of an m -order tensor $\mathcal{A} = (A_{i_1 i_2 \dots i_m})$ along mode k for any given $k \in [m]$ is an $(m-1) \times n$ tensor \mathcal{B} with

$$B_{i_1, i_2, \dots, i_{m-1}} = A_{i_1, \dots, i_{k-1}, i, i_{k+1}, \dots, i_m}, i = 1, 2, \dots, n.$$

A slice of 3-order tensor $\mathcal{A} = (A_{ijk}) \in \mathbb{R}^{m \times n \times p}$ along mode-3 is an $m \times n$ matrix $A(:, :, k)$ with $k \in [p]$, and a slice of a 4-order tensor is a 3-order tensor.

Let $\mathcal{A} \in \mathbb{R}^{m \times n \times p}, \mathcal{B} \in \mathbb{R}^{n_1 \times n \times p}$ be two tensors of 3-order. The *slice-wise product* of \mathcal{A}, \mathcal{B} , denoted by $\mathcal{C} = \mathcal{A} \otimes_s \mathcal{B}$, is defined as $\mathcal{C} = (C_{ijk}) \in \mathbb{R}^{m \times n \times p}$ where $C(:, :, k) = A(:, :, k)B(:, :, k)$ for all $k \in [p]$. This multiplication can be used to build a regression model

$$A = \mathcal{X} \otimes_s \mathcal{B} + \varepsilon \tag{2.13}$$

where $A(:, :, k)$ is the matrix consisting of n sample points of size m in class k and $X(:, :, k)$ is the design matrix corresponding to the k th sample (there are n_1 observations in each class in this situation).

Let k be a positive integer. The k -moment of a random variable x is defined as the expectation of x^k , i.e., $m_x^{(k)}(x) = E(x^k)$. The traditional extension of moments to a multivariate case is done by an iterative vectorization imposed on k . This technique is employed not only in the definition of moments but also in other definitions such as that of a characteristic function. By introducing the tensor form into these definitions, we find that the expressions will be much easier to handle than the classical ones. In the next section, we will introduce the tensor form of all these definitions.

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ be a random vector. Denote by \mathbf{x}^m the (symmetric) rank-one m -order tensor with

$$\mathbf{x}_\sigma^m = x_{i_1} x_{i_2} \cdots x_{i_m}, \forall \sigma := (i_1, i_2, \dots, i_m) \in S(m, n).$$

\mathbf{x}^m is called a rank-1 tensor generated by \mathbf{x} which is also symmetric. It is shown by Comon *et al.* [15] that a real tensor \mathcal{A} (with size $I_1 \times I_2 \times \cdots \times I_m$) can always be decomposed into form

$$\mathcal{A} = \sum_{j=1}^r \alpha_1^{(j)} \times \alpha_2^{(j)} \times \cdots \times \alpha_m^{(j)} \tag{2.14}$$

where $\alpha_i^{(j)} \in \mathbb{R}^{I_i}$ for all $j \in [r], i \in [m]$. The smallest positive integer r is called the rank of \mathcal{A} , denoted by $rank(\mathcal{A})$. We note that Equation (2.14) can also be used to define the tensor product of two matrices, which will be used in our next work on the covariance of random matrices. Note that the tensor product of two rank-one matrices is

$$(\alpha_1 \times \beta_1) \times (\alpha_2 \times \beta_2) = \alpha_1 \times \alpha_2 \times \beta_1 \times \beta_2.$$

Now consider two matrices $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times q}$. Then write A, B in a rank-1 decomposition, i.e.,

$$A = \sum_{j=1}^{R_1} \alpha_1^{(j)} \times \beta_1^{(j)}, B = \sum_{k=1}^{R_2} \alpha_2^{(k)} \times \beta_2^{(k)}.$$

Tucker decomposition decomposes the original tensor into a product of the core tensor and a number of unitary matrices in different directions [15] so \mathcal{A} can be decomposed into

$$\mathcal{A} = S \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_N U_N \tag{2.15}$$

where S is the core tensor, and U_1, U_2, \dots, U_N are unitary matrices.

Example 2.1. Let $\mathcal{X} = (X_{ijk})$ be an $2 \times 2 \times 2$ tensor which is defined by

$$X(:, :, 1) = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, X(:, :, 2) = \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}.$$

Then the unfolded matrices along 1-mode, 2-mode and 3-mode are respectively

$$X_1 = \begin{bmatrix} 1 & 5 & 3 & 7 \\ 2 & 6 & 4 & 8 \end{bmatrix}, X_2 = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}, X_3 = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}.$$

3. Application of 3-Order Tensors in GLMs

The growth curve model (GCM) is one of the GLMs introduced by Wishart in 1938 [16] to study the growth situation of animals and plant between different groups. It is a kind of generalized multivariate variance analysis model, and has been widely used in modern medicine, agriculture and biology etc. GCM originally referred to a wide array of statistical models for repeated measures data [2] [14]. The contemporary use of GCM allows the estimation of inter-object variability such as time trends, time paths, growth curves or latent trajectories, in intra-object patterns of change over time [17]. The trajectories are the primary focus of analysis in most cases, whereas in others, they may represent just one part of a much broader longitudinal model. The most basic GCMs contain fixed and random effects that best capture the collection of individual trajectories over time. In a GCM, the fixed effects represent the mean of the trajectory pooling of all individuals, and the random effects represent the variance of the individual trajectories around these group means. For example, the fixed effects in a linear trajectory are estimates of the mean intercept and mean slope that define the underlying trajectory pooling of the entire sample, and the random effects are estimates of the between-person variability in the individual intercepts and slopes. Smaller random effects imply the more similar parameters that define the trajectory across the sample of individuals; at the extreme situation where the random effects equal 0, all individuals are governed by precisely the same trajectory parameters (*i.e.*, there is a single trajectory shared by all individuals). In contrast, larger random effects imply greater individual differences in the magnitude of the trajectory parameters around the mean values.

The analysis of a GCM focuses on the functional relationship among ordered responses. Conventional GCM methods apply to growth data and to other analogs such as dose-response data (indexed by dose), location-response data (indexed by distance), or response-surface data (indexed by two or more variables such as latitude and longitude). The GCM methods mainly focus on longitudinal observations on a one-dimensional characteristic even though they may also be used in multidimensional cases [2].

A general GCM can be indicated by

$$Y = XBT + E \quad (3.16)$$

where $Y \in \mathbb{R}^{N \times p}$ is the random response matrix whose rows are mutually independent and columns correspond to the response variables ordered according to $d = [d_1, d_2, \dots, d_p]^T$; $X \in \mathbb{R}^{N \times q}$ is the fixed design matrix with $r := \text{rank}(X) \leq q \leq N$; The matrix $B \in \mathbb{R}^{q \times m}$ is a fixed parameter matrix whose entries are the regression coefficients; $T \in \mathbb{R}^{m \times p}$ is a within-subject design matrix each of whose entries is a fixed function of d , and $E \in \mathbb{R}^{N \times p}$ is a random error with matrix normal distribution $E \sim \mathcal{N}_{N,p}(0, \Sigma, I_N)$ where $\Sigma \in \mathbb{R}^{p \times p}$ is an

unknown symmetric positive definite matrix. Suppose the samplings corresponding to each object are recorded at p different times (moments) d_1, d_2, \dots, d_p . Consider an example of a pattern of the children's weight. The plotting of the weights against the ages indicates a temporal pattern of growth. A univariate linear model for weight given age could be fitted with a design matrix T expressing the central tendency of the children's weights as a linear or curvilinear function of age. Here T is an example of a within-subject design matrix. If $N > 1$, a separate curve could be fitted for each subject to obtain a separate matrix of regression parameter estimators for each independent sampling units, $\hat{\beta}_i = Y_i T (T^T T)^{-1}$ for $i \in [N]$, and a simple average of the N fitted curves is a proper (if not efficient) estimator of the population growth curve, that is,

$$\hat{B} = \frac{1}{N} \sum_{j=1}^N \hat{B}_j.$$

The efficient estimator has the form

$$\hat{B} = \left[(X^T X)^{-1} X^T \right] \left[T (T^T T)^{-1} \right]. \tag{3.17}$$

If the subjects are grouped in a balanced way, *i.e.*, The N observations are clustered into m groups, each containing the same number, say n , of observations. For simplicity, we may assume that first n each then $X = I_N$, the all-one vector, is the appropriate choice for computing \hat{B} . The choice of T defines the functional form of the population growth curve by describing a function relationship between weight and age.

Example 3.1. *We recorded the heights of n boys and n girls whose ages are 2, 3, 4 and 6 years. From the observations we make an assumption that the average height increases linearly with age. Since the observed data is partitioned into two groups (one is for the heights of n boys and another is for the height of n girls), each consisting of n objects, and $p = 4$ with age vector $d = [2, 3, 4, 6]^T$. Thus the model for the height vs. age shall be $Y = XBT + E$ where*

$$X = \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix}, T = \begin{bmatrix} I_p^T \\ d^T \end{bmatrix}$$

where $I_k \in \mathbb{R}^k$ is an all-ones vector of dimension k .

Here β_{11}, β_{12} are respectively the intercept and the slope for girls and β_{21}, β_{22} are respectively the intercept and the slope for boys. We find that it is not so easy for us to investigate the relationship between the gender, height, weight, and age. In the following we employ tensor expression to deal with this issue.

Using the notation in tensor theory, we rewrite model (3.16) in form

$$Y = X \times_2 B \times_2 T + E$$

or equivalently

$$Y = B \times_1 X^T \times_2 T + E \tag{3.18}$$

where B is regarded as a second order tensor and X, T as two matrices. Ac-

tually, according to Equation (2.12), we have

$$(B \times_1 X^T)_{ij} = \sum_{k=1}^q X_{ik} B_{kj}.$$

Similarly we can define $B \times_2 V$. Note that

$$B \times_1 U \times_2 V = B \times_2 V \times_1 U.$$

Now we extend model (3.16) in a more general form as

$$\mathcal{A} = \mathcal{B} \times_1 X_1 \times_2 X_2 \times_3 X_3 + E \tag{3.19}$$

where $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{B} = (B_{ijk}) \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ is a 3-order tensor, which is usually an unknown constant parameter tensor or the *kernel tensor*, and $X_i \in \mathbb{R}^{n_i \times m_i}$ for $i = 1, 2, 3$. Here the tensor-matrix multiplication is defined by Equation (2.12) according to the dimensional coherence along each mode.

The potential applications of Equation (3.21) are obvious. A HOSVD (high order singular value decomposition) of a 3-order tensor can be regarded as a good example for this model.

Example 3.2. *A sequence of 1000 images extracted from a repository of face images of ten individuals, each with 100 face images. Suppose each face image is of size 256×256 . Then these images can be restored in an $256 \times 256 \times 1000$ tensor \mathcal{A} . Let \mathcal{A} be decomposed as*

$$\mathcal{A} = \mathcal{B} \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T \tag{3.20}$$

where

$$\mathcal{B} \in \mathbb{R}^{16 \times 16 \times 50}, U_1 \in \mathbb{R}^{256 \times 16}, U_2 \in \mathbb{R}^{256 \times 16}, U_3 \in \mathbb{R}^{1000 \times 50}.$$

The decomposition Equation (3.20) yields a set of compressed images, each with size 16×16 . If each individual can be characterized by five images (this is called a *balanced compression*), then the kernel tensor \mathcal{B} consists of 50 compressed images where each U_i is a projection matrix along mode- i ($i = 1, 2, 3$). Specifically, U_1 and U_2 together play a role of compression of each image into an 16×16 image, while U_3 finds the representative elements (here is the 50 images) among a large set of images (the set of 1000 face images).

Analog to GCM, we let Y_{ijk} be the measured value of Index I_k in Class C_i at time T_j . A tensor $Y = (Y_{ijk}) \in \mathbb{R}^{m \times n \times p}$ can be used to express m objects, say P_1, \dots, P_m , each having p indexes I_1, \dots, I_p measured respectively at times t_1, \dots, t_n . For each index $I_k, k \in [p]$, we have GCM form:

$$Y_k = B_k \times_1 X \times_2 T + E_k \tag{3.21}$$

where $Y_k = (y_1^{(k)}, \dots, y_n^{(k)}) \in \mathbb{R}^{m \times n}$. Suppose each row of Y_k stands for a class of individuals, e.g., partitioned by ages. To make things more clear, we consider a concrete example.

Example 3.3. *There are 30 persons under health test, each measured, at time T_1, \dots, T_4 , 10 indexes such as the lower/higher blood pressures, heartbeat rate, urea, cholesterol, bilirubin, etc. We label these indexes respectively by I_1, \dots, I_{10} . Suppose that the 30 people are partitioned into three groups (denoted by C_1, C_2, C_3) with respect to their ages, consisting of 5, 10, 15 individuals respec-*

tively. Denote

$$X = \begin{bmatrix} l_5 & 0 & 0 \\ 0 & l_{10} & 0 \\ 0 & 0 & l_{15} \end{bmatrix}, T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 \\ t_1^2 & t_2^2 & t_3^2 & t_4^2 \end{bmatrix}$$

and

$$B_k = \begin{bmatrix} \beta_{11k} & \beta_{12k} & \beta_{1pk} \\ \beta_{21k} & \beta_{22k} & \beta_{2pk} \\ \beta_{31k} & \beta_{32k} & \beta_{3pk} \end{bmatrix}.$$

Denote by Y_{ijk} the measurement of Index I_k in group C_i at time T_j . Set $Y(:, :, k) = Y_k$, $B(:, :, k) = B_k$ for $k = 1, 2, \dots, 10$. Then we have

$$\mathcal{Y} = \mathcal{B} \times_1 X^T \times_2 T^T + \varepsilon \quad (3.22)$$

where $\mathcal{Y}, \varepsilon \in \mathbb{R}^{30 \times 4 \times 10}$, $X \in \mathbb{R}^{30 \times 3}$, $T \in \mathbb{R}^{4 \times 3}$ and $B \in \mathbb{R}^{3 \times 3 \times 10}$ is an unknown constant parameter tensor to be estimated, where $B(:, :, k) = B_k$ is the parameter matrix corresponding to the k th index model. The model (3.22) can be further promoted to manipulate a balanced linear mixed model when multiple responses are measured for balanced clustered (*i.e.*, there are same number of subjects in each cluster) subjects.

4. Tensor Normal Distributions

In the multivariate analysis, the correlations between the coordinates of a random vector $\mathbf{x} = (x_1, \dots, x_n)^T$ are represented by the covariance matrix $\Sigma := \Sigma(\mathbf{x})$, which is symmetric and positive semidefinite. When the variables are arrayed as a matrix, say $X = (X_{ij}) \in \mathbb{R}^{m \times n}$, which is called a *random matrix*, the correlation between any pair of entries, say $X_{i_1 j_1}$ and $X_{i_2 j_2}$ of matrix X , is represented as an entry of a matrix Σ which is defined as the covariance matrix of the vector. A matrix normal distribution is defined. $\mu \in \mathbb{R}^{m \times n}$, and $\Sigma \in \mathbb{R}^{m \times m}$, $\phi \in \mathbb{R}^{n \times n}$ are two positive definite matrices. A random matrix $X \in \mathbb{R}^{m \times n}$ is said to obey a *matrix normal distribution*, denoted by $X \sim \mathcal{N}_{m,n}(\mu, \Sigma, \phi)$, if it satisfies the following the conditions:

- 1) $E[X] = \mu$, *i.e.*, $E[X_{ij}] = \mu_{ij}$ for each $i \in [m], j \in [n]$.
- 2) Each row $X_{i \cdot}$ of X obeys normal distribution $X_{i \cdot} \sim \mathcal{N}_p(0, \phi)$ for $i \in [m]$.
- 3) Each column $X_{\cdot j}$ obeys normal distribution $X_{\cdot j} \sim \mathcal{N}_q(0, \Sigma)$.

It is easy to show that a matrix normal distribution $X \sim \mathcal{N}_{p,q}(\mu, \Sigma, \phi)$ is equivalent to $\text{vec}(X) \sim \mathcal{N}_{pq}(\text{vec}(\mu), \Sigma \otimes \phi)$ (see e.g. [8]).

We now define the tensor normal distribution. Let $\mathcal{A} = (A_{i_1 i_2 \dots i_m}) \in \mathcal{T}_m$ be an m -order tensor of size $\mathcal{N} := N_1 \times N_2 \times \dots \times N_m$, each of whose entries is a random variable. Let $\mu = (\mu_{i_1 i_2 \dots i_m})$ be an m -order tensor of the same size as that of \mathcal{A} , and $\Sigma_k (k \in [m])$ be an $N_k \times N_k$ positive definite matrix. For convenience, we denote by $I^{(n)}$ the $(m-1)$ -tuple $(i_1, i_2, \dots, i_{n-1}, i_{n+1}, \dots, i_m)$ with $i_k \in [N_k]$. We denote by $A(I^{(n)})$ and $\mu(I^{(n)})$, both in \mathbb{R}^{N_n} respectively the corresponding fibre (vector) of \mathcal{A} and μ , indexed by $I^{(n)}$, *i.e.*,

$$A(I^{(n)}) := A(i_1, i_2, \dots, i_{n-1}, :, i_{n+1}, \dots, i_m), \forall i_k \in N_k, k \in [m] \setminus \{n\}.$$

$A(I^{(n)})\mu(I^{(n)})$ is called a *fibre of \mathcal{A} (μ resp.) along mode- n indexed by $I^{(n)}$. \mathcal{A} is said to obey a *tensor normal distribution* with parameter matrices $(\mu, \Sigma_1, \dots, \Sigma_m)$ or denoted by*

$$\mathcal{A} \sim \mathcal{N}_T(\mu, \Sigma_1, \dots, \Sigma_m)$$

if for any $n \in [m]$, we have

$$A(i_1, \dots, i_{n-1}, :, i_{n+1}, \dots, i_m) \sim \mathcal{N}_{N_n}(\mu(I^{(n)})\Sigma_n).$$

\mathcal{A} is said to follow a *standard tensor normal distribution* if all the Σ_k 's are identity matrices. A model (2.13) with a tensor normal distribution is called a *general tensor normal (GTN) model*.

To show the application of tensor normal distribution, we consider the 3-order tensor. For our convenience, we use (i, j) -value to denote the value related to the i th subject at j th measurement for any $(i, j) \in [m] \times [n]$. For example, the k th response observation Y_{ijk} at (i, j) represents the k th response value measured on the i th subject at time j . Now we let m, n, p be respectively the number of observed objects, number of measuring times for each subject, and the number of responses for each observation. Denote by \mathcal{Y} the response tensor with Y_{ijk} being the k th response at (i, j) , and by \mathcal{X} the covariate tensor with $X_{ij} \in \mathbb{R}^r$ being the covariate vector at (i, j) for fixed effects, and by U_{ij} the covariate vector at (i, j) for random effects. Further, for each $k \in [p]$, we denote by $B_k \in \mathbb{R}^r$ the coefficient vector related to the fixed effects corresponding to the k th response Y_{ijk} at (i, j) for each pair $(i, j) \in [m] \times [n]$, and similarly by $C_k \in \mathbb{R}^q$ the coefficient vector related to the random effects. Now let $B = [B_1, \dots, B_p]$ and $\gamma = [C_1, \dots, C_p]$. Then $B \in \mathbb{R}^{r \times p}, \gamma \in \mathbb{R}^{q \times p}$. We call X, U respectively the *design matrix* for fixed effects and the *design matrix* for random effects. Then we have

$$\mathcal{Y} = \mathcal{X}B + \mathcal{U}\gamma + \varepsilon \tag{4.23}$$

with $\mathcal{Y} = (Y_{ijk}) \in \mathbb{R}^{m \times n \times p}$, $\mathcal{X} = (X_{ijk}) \in \mathbb{R}^{m \times n \times r}$, $B = (\beta_{ij}) \in \mathbb{R}^{r \times p}$, $\mathcal{U} = (U_{ijk}) \in \mathbb{R}^{m \times n \times q}$, $\gamma = (\gamma_{ij}) \in \mathbb{R}^{q \times p}$, $\varepsilon = (\varepsilon_{ijk}) \in \mathbb{R}^{m \times n \times p}$ where ε_{ijk} is the error term. Here the tensor-matrix multiplications $\mathcal{X}B$ and $\mathcal{U}\gamma$ are defined by

$$(\mathcal{X}B)_{ijk} = X(i, j, :)B_k = \sum_{k'=1}^r X_{ijk'}\beta_{k'k}, \forall i \in [m], j \in [n], k \in [p]$$

$$(\mathcal{U}\gamma)_{ijk} = U(i, j, :)\gamma_k = \sum_{k'=1}^q U_{ijk'}\gamma_{k'k}, \forall i \in [m], j \in [n], k \in [p].$$

Denote $B = [\beta_1, \dots, \beta_p], \gamma = [\gamma_1, \dots, \gamma_p]$, and

$$E_i^{(1)} := E(i, :, :), E_j^{(2)} := E(:, j, :), E_k^{(3)} := E(:, :, k), \forall i \in [m], j \in [n], k \in [p]$$

$$E_{jk}^{(1)} := E(:, j, k), E_{ik}^{(2)} := E(i, :, k), E_{ij}^{(3)} := E(i, j, :), \forall i \in [m], j \in [n], k \in [p]$$

where each matrix $E_i^{(s)}$ is called a *slice on mode s* , and each vector $E_{it}^{[s]}$ is called a *fibre along mode- s* . We also use $E^{[s]}$ to denote the set consisting of all

fibres of ε along mode- s , and use notation $E^{[s]} \sim P$ to express that each element of $E^{[s]}$ obeys distribution P where P is a distribution function. For example, $E^{[1]} \sim \mathcal{N}_m(0, I_m)$ means that each 1-mode fibre $E_{jk}^{[1]}$ (there are np 1-mode fibres) obeys a standard normal distribution, i.e., $E_{jk}^{[1]} \sim \mathcal{N}_m(0, I_m)$.

Now for convenience we let $(n_1, n_2, n_3) := (m, n, p)$. We assume that

- 1) γ obeys matrix normal distribution $\gamma \sim \mathcal{N}_{q,p}(0, \Sigma, \phi)$.
- 2) The random vectors in $\{\gamma_1, \gamma_2, \dots, \gamma_p\}$ are independent with $E^{[s]}$ for each $s = 1, 2, 3$.
- 3) For any $s \in \{1, 2, 3\}$, $E^{[s]} \sim \mathcal{N}_{n_s}(0, \Sigma_s)$ with Σ_s being positive definite of size $n_s \times n_s$.

The model (4.23) with conditions (I, II, III) is called a 3-order general mixed tensor (GMT) model. We will generalise this model to a more general case. In the following we first define the standard normal 3-order tensor distribution:

Definition 4.1. Let $\mathcal{X} = (X_{ijk}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a random tensor, i.e., each entry of \mathcal{X} is a random variable. Let $\mu = (A_{ijk}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a constant tensor and $\Sigma_s \in \mathbb{R}^{n_s \times n_s}$ be a positive definite matrix for each $s \in [3]$. Then \mathcal{X} is said to obey 3-order tensor standard normal (TSN) distribution if $X^{[s]} \sim \mathcal{N}_{n_s}(\mu^{[s]}, \Sigma_s)$ for all $s = 1, 2, 3$.

A 3-order random tensor satisfying TSN distribution has the following property:

Theorem 4.2. Let $\mathcal{A} \in \mathbb{R}^{m \times n \times p}$ be an 3-order random tensor which obeys the tensor standard normal (TSN) distribution. Then each slice of \mathcal{A} shall obey a standard matrix normal distribution. Specifically, we have

$$\begin{aligned} A_i^{(1)} &\sim \mathcal{N}_{n,p}(0, I_p, I_n), \forall i \in [m], \\ A_j^{(2)} &\sim \mathcal{N}_{m,p}(0, I_p, I_m), \forall j \in [n], \\ A_k^{(3)} &\sim \mathcal{N}_{m,n}(0, I_n, I_m), \forall k \in [p]. \end{aligned} \tag{4.24}$$

Proof. □

Note that condition (III) is a generalization to the matrix normal distribution, and we denote it by $\varepsilon \sim \mathcal{N}_{m,n,p}(\mu, \Sigma_1, \Sigma_2, \Sigma_3)$,

$$\text{vec}(E_j) \sim \mathcal{N}_{mn}(0, D_2 \otimes D_1).$$

Note that $D := D_2 \otimes D_1 \in \mathbb{R}^{mn \times mn}$ is a diagonal matrix since both D_1 and D_2 are diagonal. Write $\varepsilon^* := \mathcal{U}\gamma + E = \hat{\mathcal{U}}\tilde{E}$ where $\hat{\mathcal{U}}$ is the expansion of \mathcal{U} along the third mode, specifically,

$$\hat{\mathcal{U}}(:, :, 1:q) = U(:, :, 1:q), \hat{\mathcal{U}}(:, :, q+1:q+n) = I(:, :, 1:n) \in \mathbb{R}^{m \times m \times n}$$

here $I(:, :, 1:n)$ is the tensor consisting of n identity matrices of size $m \times m$ stacking along the third mode, thus $\hat{\mathcal{U}} \in \mathbb{R}^{m \times m \times (q+n)}$ and $\tilde{E} \in \mathbb{R}^{m \times n}$. Then we have

$$\mathcal{Y} = \mathcal{X}B + \mathcal{E}. \tag{4.25}$$

Now we unfold \mathcal{Y}, \mathcal{X} along mode-3 to get matrix $Y(3) \in \mathbb{R}^{p \times mn}$ and $X(3) \in \mathbb{R}^{r \times mn}$ respectively. Then Equation (4.25) is equivalent to

$$Y(3)^T = X(3)^T B + U(3)^T \gamma + \tilde{E} \tag{4.26}$$

where $U(3)$ and \tilde{E} are generated similarly as $Y(3)$.

The multivariate linear mixed model (4.23) or (4.25) can be transformed into a general linear model through the vectorization of matrices. Recall that the *vectorization* of a matrix $A \in \mathbb{R}^{m \times n}$ is a vector of dimension \mathbb{R}^{mn} , denoted by $\text{vec}(A)$, formed by vertically stacking the columns of A in order, that is,

$$\text{vec}(A)^\top = [A_1^\top \quad A_2^\top \quad \cdots \quad A_m^\top]$$

where A_1, A_2, \dots, A_m are the column vectors of A . The vectorization is closely related to Kronecker product of matrices. The following lemma presents some basic properties of the vectorization and Kronecker product. We will use the following lemma (Proposition 1.3.14 on Page 89 of [4]) to prove our main result:

Lemma 4.3. *Let A, B, C, X be matrices of appropriate sizes such that all the operations defined in the following are valid. Then*

- 1) $\text{vec}(ABC) = (C^\top \otimes A)\text{vec}(B)$.
- 2) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
- 3) $(A \otimes B)^\top = A^\top \otimes B^\top$.

The following property of the multiplication of a tensor with a matrix is shown by Kolda [18] and will be used to prove our main result.

Lemma 4.4. *Let \mathcal{A} be a real tensor of size $I := I_1 \times I_2 \times I_N$, and $U \in \mathbb{R}^{I_n \times J_n}$ where $n \in [N]$. Then $\mathcal{B} = \mathcal{A} \times_n U$ if and only if*

$$B(n) = U^\top A(n) \tag{4.27}$$

where $A(n), B(n)$ are respectively the flattened matrices of \mathcal{A} and \mathcal{B} along mode- n .

Proof. Let $\mathcal{A} = (A_\sigma), \mathcal{B} = (B_\sigma) = \mathcal{A} \times_n U$. Then for any $\sigma := (i_1, i_2, \dots, i_N)$, we have

$$B_\sigma = B_{i_1 \dots i_{n-1} i_{n+1} \dots i_N} = \sum_{i=1}^{I_n} A_{i_1 \dots i_{n-1} i i_{n+1} \dots i_N} U_{i i_n}. \tag{4.28}$$

From which the result Equation (4.27) is immediate. □

Note that our Formula (4.27) is different from that in Section 2.5 in [18] since the definition of tensor-matrix multiplication is different.

We have the following result for the estimation of the parameter matrix B :

Theorem 4.5. *Suppose $\text{rank}(X) = r \leq mn$ in Equation (4.23). Then the optimal estimation of the parameter matrix B in Equation (4.23) is*

$$\hat{B} = (X(3)X(3)^\top)^{-1} X(3)Y(3)^\top. \tag{4.29}$$

Proof. We first write Equation (4.25) in a matrix-vector form by vectorization by using the first item in Lemma 4.3,

$$\text{vec}(Y(3)^\top) = \text{vec}(X(3)^\top B) + \text{vec}(\hat{E}) \tag{4.30}$$

where $\hat{E} : U(3)^\top \gamma + \tilde{E}$ is the sum of two random terms. By the property of the vectorizations (see e.g. [4]), we know that

$\text{vec}(X(3)^T B) = (I_p \otimes X(3)^T) \text{vec}(B)$. By the ordinary least square solution method we get

$$\begin{aligned} \text{vec}(\hat{B}) &= \left[(I_p \otimes X(3)^T)^T (I_p \otimes X(3)^T) \right]^{-1} (I_p \otimes X(3)^T) \text{vec}(Y(3)^T) \\ &= (I_p \otimes X(3)X(3)^T)^{-1} (I_p \otimes X(3)^T) \text{vec}(Y(3)^T) \\ &= \left[(I_p \otimes X(3)X(3)^T)^{-1} \right] (I_p \otimes X(3)^T) \text{vec}(Y(3)^T) \\ &= \left[(I_p \otimes X(3)X(3)^T)^{-1} X(3)^T \right] \text{vec}(Y(3)^T). \end{aligned}$$

By using (1) of Lemma 4.3 again (this time in the opposite direction), we get result (4.29). \square

For any $\mathcal{X} \in \mathbb{R}^{m \times n \times r}$, we denote $H := X(3)^T (X(3)X(3)^T)^{-1} X(3)$ when $\text{rank}(\mathcal{X}) = r \leq mn$, and $H := (X(3)X(3)^T)^g X(3)$ when $\text{rank}(\mathcal{X}) < r$ (A^g stands for an g -inverse of a matrix A). Then $H \in \mathbb{R}^{r \times mn}$ can be regarded as the projection from \mathbb{R}^{mn} into \mathbb{R}^r since $H^T = H, H^2 = H$. Furthermore, we have $HX(3)^T = X(3)^T$. We now end the paper by presenting the following result as a prediction model which follows directly from Theorem 4.5.

Theorem 4.6. Suppose $\text{rank}(\mathcal{X}) = r \leq mn$ in Equation (4.23). Then the mean of the response tensor \mathcal{Y} in Equation (4.23) is

$$\bar{\mathcal{Y}} = \mathcal{X} \times_3 A(3) \quad (4.31)$$

where $A(3) := (X(3)X(3)^T)^{-1} X(3)Y(3)^T$.

Proof. By Theorem 4.5 and Equation (4.26), we have

$$\begin{aligned} E[Y(3)] &= \bar{\mathcal{Y}}(3) = X(3)^T \hat{B} \\ &= X(3)^T (X(3)X(3)^T)^{-1} X(3)Y(3)^T \\ &= HY(3)^T. \end{aligned}$$

It follows that

$$\bar{Y}(3) = Y(3)X(3)^T (X(3)X(3)^T)^{-1} X(3). \quad (4.32)$$

By employing Lemma 4.4, we get result (4.31). \square

Acknowledgements

This research was partially supported by the Hong Kong Research Grant Council (No. PolyU 15301716) and the graduate innovation funding of USTS. We shall thank the anonymous referees for their patient and elaborate reading and their suggestions which improved the writing of the paper.

References

- [1] Bilodeau, M. and Brenner, D. (1961) Theory of Multivariate Statistics. Springer, New York.

- [2] Bollen, K.A. and Curran, P.J. (2006) Latent Curve Models: A Structure Equation Perspective. Wiley, Hoboken.
- [3] Bryk, A.S. and Raudenbush, S.W. (1987) Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin*, **101**, 147-158.
<https://doi.org/10.1037/0033-2909.101.1.147>
- [4] Kollo, T. and Rosen, D. (2005) Advanced Multivariate Statistics with Matrices. Springer, New York. <https://doi.org/10.1007/1-4020-3419-9>
- [5] Raudenbush, S.W. and Bryk, A.S. (2002) Hierarchical Linear Models: Applications and Data Analysis Methods 2. Sage Publications, Thousand Oaks, CA.
- [6] Bauer, D.J. (2007) Observations on the Use of Growth Mixture Models in Psychological Research. *Multivariate Behavioral Research*, **42**, 757-786.
<https://doi.org/10.1080/00273170701710338>
- [7] Bollen, K.A. and Curran, P.J. (2004) Autoregressive Latent Trajectory (ALT) Models: A Synthesis of Two Traditions. *Sociological Methods and Research*, **32**, 336-383.
<https://doi.org/10.1177/0049124103260222>
- [8] Coffman, D.L. and Millsap, R.E. (2006) Evaluating Latent Growth Curve Models Using Individual Fit Statistics. *Structural Equation Modeling*, **13**, 1-27.
https://doi.org/10.1207/s15328007sem1301_1
- [9] Hedeker, D. and Gibbons, R. (2006) Longitudinal Data Analysis. Wiley Inc., New York.
- [10] Little, R.J. and Rubin, D.B. (1987) Statistical Analysis with Missing Data. Wiley Inc., New York.
- [11] Singer, J.D. (1998) Using SAS Proc Mixed to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics*, **23**, 323-355. <https://doi.org/10.3102/10769986023004323>
- [12] Singer, J.D. and Willett, J.B. (2003) Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York.
<https://doi.org/10.1093/acprof:oso/9780195152968.001.0001>
- [13] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning. 2nd Edition, Springer, New York. <https://doi.org/10.1007/978-0-387-84858-7>
- [14] Bollen, K.A. (2007) On the Origins of Latent Curve Models. In: Cudeck, R. and MacCallum, R., Eds., *Factor Analysis*, at 100, Lawrence Erlbaum Associates, Mahwah, NJ, 79-98.
- [15] Comon, P., Golub, G., Lim, L.-H. and Mourrain, B. (2008) Symmetric Tensors and Symmetric Tensor Rank. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1254-1279. <https://doi.org/10.1137/060661569>
- [16] Wishart, J. (1938) Growth Rate Determinations in Nutrition Studies with the Bacon Pig and Their Analysis. *Biometrika*, **30**, 16-28.
<https://doi.org/10.1093/biomet/30.1-2.16>
- [17] McArdle, J.J. (2009) Latent Variable Modeling of Differences and Changes with Longitudinal Dynamic Structural Analysis. *Annual Review of Psychology*, **60**, 577-605.
<https://doi.org/10.1146/annurev.psych.60.110707.163612>
- [18] Kolda, T.G. and Bader, B.W. (2009) Tensor Decompositions and Applications. *SIAM Review*, **51**, 455-500. <https://doi.org/10.1137/07070111X>