Principal component analysis in the case of multivariate repeated measures data

Karol Deręgowski[1], Mirosław Krzyśko[1,2]

[1]President Stanislaw Wojciechowski Higher Vocational State School in Kalisz, Institute of Management, Poland, e-mail: kadere@o2.pl
[2]Adam Mickiewicz University, Faculty of Mathematics and Computer Science, Poznań, Poland, e-mail: mkrzysko@amu.edu.pl

Dedicated to Professor Tadeusz Caliński for his 80th birthday

SUMMARY

In this paper, we propose the principal components applicable in the case of multivariate repeated measures data under the following assumptions: (1) multivariate normality for the vector of observations $\mathbf{x}_j$, (2) Kronecker product structure of the positive definite covariance matrix Ω. Computational schemes for maximum likelihood estimates of required parameters are also given.

Key words: Principal component analysis; repeated measures data; Kronecker product covariance structure; maximum likelihood estimates

1.   Introduction

Suppose that we have a sample of n objects characterized by p variables, measured in T different time points or physical conditions. Such data are often referred to in the statistical literature as multivariate repeated measures data or doubly multivariate data. Let $\mathbf{X}_{jk}$ be a $p \times 1$ column vector of measurements on the jth individual at the kth time point, let

$$\mathbf{X}_j = (\ \mathbf{X}_{j1}, \mathbf{X}_{j2}, \ldots, \mathbf{X}_{jT}\ )$$

be a $p \times T$ matrix and let $\mathbf{x}_j = \mathrm{vec}\,(\mathbf{X}_j)$, $j = 1, 2, \ldots, n$; $k = 1, 2, \ldots, T$.

The vector $\mathbf{x}_j$ is a $(pT \times 1)$-dimensional column vector obtained by stacking all p variables at the first time point, then stacking all p variables at the second time point below it and so on.

Next, assume that

$$\mathbf{x}_j \sim N_{pT}(\boldsymbol{\mu}, \boldsymbol{\Omega}), j = 1, 2, \ldots, n$$

with pT×pT positive definite covariance matrix $\boldsymbol{\Omega}$.

Our goal is to construct principal components based on these doubly multivariate data.

Principal components analysis (PCA) (Hotelling,1933) was introduced as a technique for deriving a reduced set of orthogonal linear projections of a single collection of correlated variables $\mathbf{x}$, where the projections are ordered by decreasing variances. Variance is a second-order property of a random variable and is an important measurement of the amount of information in that variable. PCA has also been referred to as a method for "decorrelating" $\mathbf{x}$, and as a result the technique has been independently rediscovered by many different fields, with alternative names such as "Karhunen–Loève transform" and "empirical orthogonal functions", which are used in communications theory and atmospheric science respectively. PCA is used primarily as a dimensionality-reduction technique. In this role, PCA is used, for example, in lossy data compression, pattern recognition, and image analysis.

In addition to reducing dimensionality, PCA can be used to discover important features of the data. Discovery in PCA takes the form of graphical displays of the principal component scores. The first few principal component scores can reveal whether most of the data actually live on a linear subspace of $\Re^p$ and can be used to identify outliers, distributional peculiarities, and clusters of points. The last few principal component scores show those linear projections of $\mathbf{x}$ that have the smallest variance; any principal component with zero or near-zero variance is virtually constant, and hence can be used to detect collinearity, as well as outliers that alter the perceived dimensionality of the data.

Principal components are constructed on the basis of the covariance matrix estimator. The covariance matrix $\boldsymbol{\Omega}$ is positive definite. Its estimate $\hat{\boldsymbol{\Omega}}$ is positive definite with probability 1 if and only if n > pT (see, e.g. Giri (1996), p.93). Hence estimation of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ will require a very large sample, which may not always be feasible. Hence we assume $\boldsymbol{\Omega}$ to be of the form (Roy and Khattree, 2005):

$$\boldsymbol{\Omega} = \mathbf{V} \otimes \boldsymbol{\Sigma},$$

where $\mathbf{V}$ is a $T{\times}T$ positive define covariance matrix. The matrix $\mathbf{V}$ represents the covariance between repeated measures on a given object and for a given characteristic. Likewise, $\mathbf{\Sigma}$ represents the covariance between all characteristics on a given object and for a given time point. The above covariance structure is subject to an implicit assumption that for all characteristics, the correlation structure between repeated measures remains the same, and that covariance between all characteristics does not depend on time and remains constant for all time points.

In this case the estimates of the matrices $\mathbf{V}$ and $\mathbf{\Sigma}$ are positive definite with probability 1 if and only if n > max(p, T).

As noted in the literature, e.g. Galecki (1994) and Naik and Rao (2001), since

$$(c\mathbf{V}) \otimes (c^{-1}\mathbf{\Sigma}) = \mathbf{V} \otimes \mathbf{\Sigma},$$

all the parameters of $\mathbf{V}$ and $\mathbf{\Sigma}$ are not defined uniquely. Thus, without loss of generality, assume that for $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{pp} = 1$, or equivalently assume that for $\mathbf{V} = (v_{ij})$, $v_{TT} = 1$ instead of $\sigma_{pp} = 1$.

In this paper, we propose the principal components applicable in the case of multivariate repeated measures data under the following assumptions:
(1) multivariate normality for the vector of observations $\mathbf{x}$,
(2) Kronecker product structure of the positive definite covariance matrix $\mathbf{\Omega}$.

In Section 2 the principal components are presented in the case where no structures whatsoever are imposed on $\mathbf{V}$ and $\mathbf{\Sigma}$ except that they are positive definite and $v_{TT} = 1$. In Section 3, we test the hypothesis that the general $pT{\times}pT$ covariance matrix has the form $\mathbf{\Omega} = \mathbf{V} \otimes \mathbf{\Sigma}$, against the alternative that the covariance matrix is not of Kronecker product structure. In Section 4 we present an example.

2.   Principal Component Analysis

Let

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{X}_j, \tag{2.1}$$

$$\mathbf{X}_{jc} = \mathbf{X}_j - \overline{\mathbf{X}}, \tag{2.2}$$

and

$$\mathbf{X}_{jc} = \left(\mathbf{X}_{jc1}, \mathbf{X}_{jcT}\right), \tag{2.3}$$

where $\mathbf{X}_{jc1}$ is a $p \times (T-1)$ matrix and $\mathbf{X}_{jcT}$ is a $p \times 1$ vector, $j = 1, 2, \ldots, n$. For $\mathbf{V} = (v_{rs})$, we only assume that $v_{TT} = 1$.

In this case the maximum likelihood estimation equations are of the form (Srivastava et al. 2008):

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}},$$

$$\hat{\mathbf{V}} = \frac{1}{np} \begin{bmatrix} \sum_{j=1}^{n} \mathbf{X}_{jc1}^{T} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_{jc1}, & \sum_{j=1}^{n} \mathbf{X}_{jc1}^{T} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_{jcT} \\ \sum_{j=1}^{n} \mathbf{X}_{jcT}^{T} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_{jc1}, & \sum_{j=1}^{n} \mathbf{X}_{jcT}^{T} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_{jcT} \end{bmatrix} = \frac{1}{np} \sum_{j=1}^{n} \mathbf{X}_{jc}^{T} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_{jc} \tag{2.4}$$

where

$$\sum_{j=1}^{n} \mathbf{X}_{jcT}^{T} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_{jcT} = np, \tag{2.5}$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{nT} \sum_{j=1}^{n} \mathbf{X}_{jc} \hat{\mathbf{V}}^{-1} \mathbf{X}_{jc}^{T}. \tag{2.6}$$

In this case no explicit maximum likelihood estimates of $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are available. The MLEs of $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are obtained by solving simultaneously and iteratively the equations (2.4) and (2.6) subject to condition (2.5). This is the so called "flip-flop" algorithm.

Srivastava et. al. (2008) have proved that if $n > \max(p, T)$ then the maximum likelihood estimation equations given by (2.4) and (2.6) subject to the condition (2.5) will always converge to the unique maximum.

The following iterative steps are suggested to obtain the maximum likelihood estimates of $\mathbf{V}$ and $\boldsymbol{\Sigma}$.

Algorithm

Step 1. Get the initial covariance matrix $\boldsymbol{\Sigma}$ of the form

$$\tilde{\boldsymbol{\Sigma}} = \mathbf{S} = \frac{1}{nT} \sum_{j=1}^{n} \left(\mathbf{X}_j - \overline{\mathbf{X}}\right)\left(\mathbf{X}_j - \overline{\mathbf{X}}\right)^T. \tag{2.7}$$

Step 2. Based on the initial covariance matrix $\mathbf{S}$, compute the matrix $\hat{\mathbf{V}}$ given by (2.4) and replace each element $\hat{v}_{rs}$ by $\hat{v}_{rs} / \hat{v}_{TT}$.

Step 3. Compute the matrix $\hat{\boldsymbol{\Sigma}}$ from equation (2.6) using the $\hat{\mathbf{V}}$ obtained in Step 2.

Step 4. Repeat Steps 2 and 3 until convergence is attained.

We selected the following stopping rule: Compute two matrices: (a) a matrix of difference between two successive solutions of (2.4), and (b) a matrix of difference between two successive solutions of (2.6). Continue the iterations until the maxima of the absolute values of the elements of the matrices in (a) and (b) are smaller than the pre-specified quantities.

Principal components are constructed on the basis of the $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ matrix. If $n > \max(p, T)$ then $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ is a positive define matrix with probability 1 and all the characteristic roots/eigenvalues of $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ are real and positive. One of the main reasons for interest in the Kronecker product is the beautifully simple connection between the eigenvalues and the eigenvectors $\hat{\mathbf{V}}$ and $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ (see e.g. Lancaster and Tismenetsky (1995), p.412 or Ortega (1987), p.237). If $\alpha_1, \alpha_2,\ldots, \alpha_T$ are the eigenvalues of $\hat{\mathbf{V}}$ and $\beta_1, \beta_2,\ldots, \beta_p$ are the eigenvalues of $\hat{\boldsymbol{\Sigma}}$, then eigenvalues of $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ are the pT numbers $\alpha_r \beta_s$, $r = 1, 2,\ldots, T$, $s = 1, 2,\ldots, p$.

If $\mathbf{u} = (u_1, u_2,\ldots, u_T)^T$ is an eigenvector of $\hat{\mathbf{V}}$ corresponding to the eigenvalues α, and $\mathbf{w} = (w_1, w_2,\ldots, w_T)^T$ is an eigenvector of $\hat{\boldsymbol{\Sigma}}$ corresponding to the eigenvalues β, then an eigenvector $\boldsymbol{\gamma}$ of $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ associated with $\alpha\beta$ is $\boldsymbol{\gamma} = \mathbf{u} \otimes \mathbf{w} = (u_1\mathbf{w}^T, u_2\mathbf{w}^T,\ldots, u_T\mathbf{w}^T)^T$.

Let $\lambda_1 > \lambda_2 >\ldots> \lambda_{pT} > 0$ be the ordered characteristic roots of $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$. Then there exists an orthogonal matrix $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2,\ldots, \boldsymbol{\gamma}_{pT})$, $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \mathbf{I}_{pT}$, such that $\boldsymbol{\Gamma}^T \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma} = \mathbf{D}_\lambda = \mathrm{diag}(\lambda_1, \lambda_2,\ldots, \lambda_{pT})$.

Hence, if we let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{pT} \end{bmatrix} = \mathbf{\Gamma}^T \mathbf{x} = \begin{bmatrix} \boldsymbol{\gamma}_1^T \\ \boldsymbol{\gamma}_2^T \\ \vdots \\ \boldsymbol{\gamma}_{pT}^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \boldsymbol{\gamma}_1^T \mathbf{x} \\ \boldsymbol{\gamma}_2^T \mathbf{x} \\ \vdots \\ \boldsymbol{\gamma}_{pT}^T \mathbf{x} \end{bmatrix}$$

then Cov($\mathbf{y}$) = $\mathbf{D}_\lambda$, and the components $y_1 = \boldsymbol{\gamma}_1^T \mathbf{x}$, $y_2 = \boldsymbol{\gamma}_2^T \mathbf{x}$,…., $y_{pT} = \boldsymbol{\gamma}_{pT}^T \mathbf{x}$ are uncorrelated. The component $y_1 = \boldsymbol{\gamma}_1^T \mathbf{x}$ is called the first principal component, $y_2 = \boldsymbol{\gamma}_2^T \mathbf{x}$ the second principal component, and so on. The variance of y_i is λ_i. Since $\lambda_1 + \lambda_2 + … + \lambda_{pT}$ = tr($\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$), the sum of the variance of the pT principal components is the same as the sum of the variance of the old variables, measured at T different time points. Thus the components with smaller variances could be ignored without significantly affecting the total variance, thereby reducing the number of variables from pT to, say, $k \le pT$.

Since tr($\hat{\mathbf{V}} \otimes \hat{\mathbf{\Sigma}}$) $- \sum_{j=1}^{k} \lambda_j = \sum_{j=k+1}^{pT} \lambda_j$, the proportion of the total univariate variance accounted for by k PCs is $\sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{pT} \lambda_j$, which may be used as a criterion for selecting a subspace of k components from pT.

3. Testing that the Kronecker model holds

We wish to test the hypothesis H_0: $\mathbf{\Omega} = \mathbf{V} \otimes \mathbf{\Sigma}$ against the alternative H_a: $\mathbf{\Omega}$ unstructured, using the likelihood ratio test (LRT).

The likelihood ratio test compares the maximum value of the likelihood function L restricted to the region defined by the null hypothesis H_0, to the maximum of the likelihood function L in the unrestricted region H_a. Thus the ratio

$$\Lambda = \frac{\max_{H_0} L}{\max_{H_a} L}$$

or a function of it, is used as the best statistic to test the null hypothesis H_0. It is well known that, for large samples and under a normality assumption, $-2\ln\Lambda$ is approximately χ_ν^2 under H_0 where the degrees of freedom ν is equal to the number of parameters estimated under H_a minus the number of parameters

estimated under H_0. The maximum likelihood estimators of $\mathbf{V}$ and $\mathbf{\Sigma}$ have been obtained in Section 2. The MLE of $\mathbf{\Omega}$, under the alternative, is given by

$$\hat{\mathbf{\Omega}} = \frac{1}{n} \sum_{j=1}^{n} (\text{vec}(\mathbf{X}_{jc}))\,(\text{vec}(\mathbf{X}_{jc}))^T , \; n > pT. \tag{3.1}$$

Thus the LRT for H_0 against H_a is given by (Srivastava et. al. 2008)

$$\Lambda = \frac{\left|\hat{\mathbf{\Omega}}\right|^{\frac{1}{2}n}}{\left|\hat{\mathbf{V}}\right|^{\frac{1}{2}np}\left|\hat{\mathbf{\Sigma}}\right|^{\frac{1}{2}nT}} = \frac{\left(\prod_{i=1}^{pT}\delta_i\right)^{\frac{1}{2}n}}{\left(\prod_{i=1}^{T}\alpha_i\right)^{\frac{1}{2}np}\left(\prod_{i=1}^{p}\beta_i\right)^{\frac{1}{2}nT}} , \tag{3.2}$$

where $\delta_1 > \delta_2 > \ldots > \delta_{pT}$ are the eigenvalues of $\hat{\mathbf{\Omega}}$.

From asymptotic theory, $-2\ln\Lambda \sim \chi^2_{\frac{1}{2}pT(pT+1)-\frac{1}{2}p(p+1)-\frac{1}{2}T(T+1)+1}$.

4. Example

A laboratory experiment was set up to investigate the effect of growth of inoculating paspalum grass with a fungal infection applied at four different temperatures (14, 18, 22, 26^{0}C). For each pot of paspalum, measurements were made of three variables:
- fresh weight of roots (gm),
- maximum root length (mm),
- fresh weight of tops (gm).

The inoculated group was compared with a control group and six three-dimensional observations were made of each treatment-temperature combination. These are given in Table 9.10 of the monograph by Seber (1984). In our case $p = 3$, $T = 4$, $n = 12$.

Since $pT = n$, the positive definite estimate of the covariance matrix $\mathbf{\Omega}$ does not exist and hence multivariate normality with $\mathbf{\Omega} = \mathbf{V} \otimes \mathbf{\Sigma}$ is assumed.

The maximum likelihood estimates of the upper triangle of covariance matrices $\mathbf{V}$ and $\mathbf{\Sigma}$ subject to the condition $\hat{v}_{TT} = 1$ are:

$$\hat{\mathbf{V}} = \begin{bmatrix} 0.1094 & -0.0392 & -0.1262 & -0.0230 \\ & 5.7266 & 0.0592 & 0.2925 \\ & & 2.9797 & -0.2706 \\ & & & 1 \end{bmatrix}$$

and

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 3.7203 & 6.2975 & 5.4486 \\ & 26.6551 & 10.5388 \\ & & 10.9115 \end{bmatrix}.$$

The eigenvalues of $\hat{\mathbf{V}}$ are: $\alpha_1 = 5.7456$, $\alpha_2 = 3.0210$, $\alpha_3 = 0.9467$, $\alpha_4 = 0.1023$.
The eigenvectors of $\hat{\mathbf{V}}$ corresponding to the eigenvalues α_1 and α_2 are:

$$\mathbf{u}_1 = \begin{bmatrix} -0.007550 \\ 0.998005 \\ 0.015773 \\ 0.060658 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} -0.041770 \\ -0.007860 \\ 0.990169 \\ -0.133260 \end{bmatrix}.$$

The eigenvalues of $\hat{\boldsymbol{\Sigma}}$ are: $\beta_1 = 34.0928$, $\beta_2 = 6.4274$, $\beta_3 = 0.7668$.
The eigenvectors of $\hat{\boldsymbol{\Sigma}}$ corresponding to the eigenvalues β_1 and β_2 are:

$$\mathbf{w}_1 = \begin{bmatrix} 0.257959 \\ 0.855251 \\ 0.449448 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_2 = \begin{bmatrix} 0.360580 \\ -0.516800 \\ 0.776465 \end{bmatrix}.$$

The eigenvalues of $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ are:

$$\lambda_1 = \alpha_1 \beta_1 = 195.8853, \quad \lambda_2 = \alpha_2 \beta_1 = 102.9950, \quad \lambda_3 = \alpha_1 \beta_2 = 36.9294,$$

$$\lambda_4 = \alpha_3 \beta_1 = 32.2774, \quad \lambda_5 = \alpha_2 \beta_2 = 19.4172, \quad \lambda_6 = \alpha_3 \beta_2 = 6.0851,$$

$$\lambda_7 = \alpha_1 \beta_3 = 4.4056, \quad \lambda_8 = \alpha_4 \beta_1 = 3.4884, \quad \lambda_9 = \alpha_2 \beta_3 = 2.3164,$$

$$\lambda_{10} = \alpha_3 \beta_3 = 0.7259, \quad \lambda_{11} = \alpha_4 \beta_2 = 0.6576, \quad \lambda_{12} = \alpha_4 \beta_3 = 0.07885.$$

The eigenvectors of $\hat{\mathbf{V}} \otimes \hat{\boldsymbol{\Sigma}}$ corresponding to the eigenvalues λ_1 and λ_2 respectively are:

$$\gamma_1 = \mathbf{u}_1 \otimes \mathbf{w}_1 = \begin{bmatrix} -0.00195 \\ -0.00645 \\ -0.00339 \\ 0.25744 \\ 0.85354 \\ 0.44855 \\ 0.00407 \\ 0.01349 \\ 0.00709 \\ 0.01565 \\ 0.05188 \\ 0.02726 \end{bmatrix} \quad \text{and} \quad \gamma_2 = \mathbf{u}_2 \otimes \mathbf{w}_1 = \begin{bmatrix} -0.01077 \\ -0.03572 \\ -0.01877 \\ -0.00203 \\ -0.00673 \\ -0.00353 \\ 0.25542 \\ 0.84684 \\ 0.44503 \\ -0.03438 \\ -0.11397 \\ -0.05989 \end{bmatrix}.$$

The first principal component is $y_1 = \gamma_1^T \mathbf{x}$ and the second principal component is $y_2 = \gamma_2^T \mathbf{x}$, where vector $\mathbf{x}$ is a $(12{\times}1)$-dimensional column vector obtained by stacking the measurements of all three variables at the first temperature, then stacking the measurements of all three variables at the second temperature below it, and so on.

The first two principal components account for $((\lambda_1 + \lambda_2)/\sum_{i=1}^{12}\lambda_i)\times100\%$ $= 73.75\%$ of the total variation. Plotted values of the first two principal components for individual measurements are shown in Figure 1. We see that these two groups are well separated.
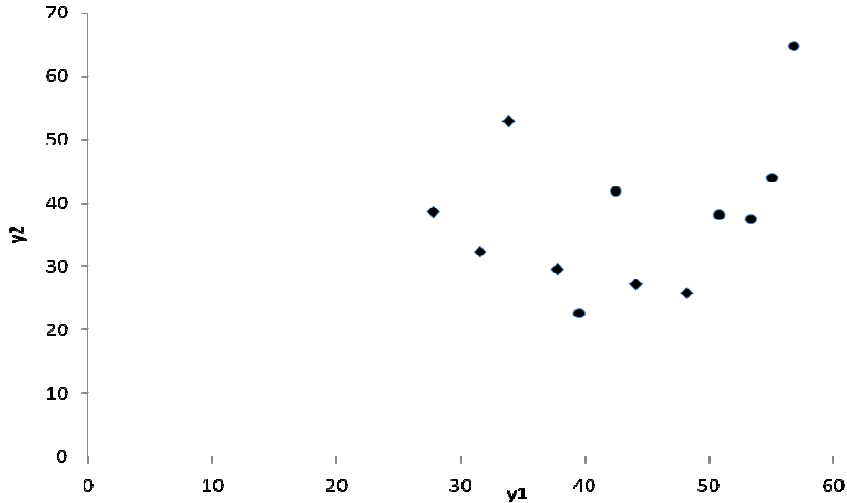

Figure 1. Plotted values of the first two principal components for individual measurements; ♦ – Inoculated ● – Control

REFERENCES

Galecki A.T. (1994): General class of covariance structures for two or more repeated factors in longitudinal data analysis, Comn. Statist. Theory and Methods 23: 3105–3119.

Giri N.C. (1996): Multivariate Statistical Analysis, Marcel Dekker, Inc., New York.

Hotelling H. (1933): Analysis of a complex of statistical variables into principal components, J. Ed. Psych. 24: 417–441, 498–520.

Lancaster P., Tismenetsky M. (1985): The Theory of Matrices, Second Edition with Applications, Academic Press, Orlando.

Naik D.N., Rao S. (2001): Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix, J. App. Statist. 28: 91–105.

Ortega J.M. (1987): Matrix Theory: A Second Course, Plenum Press, New York.

Roy A., Khattree R. (2005): On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data, Statist. Methodology 2: 297–306.

Seber G.A.F. (1984): Multivariate Observations, Wiley, New York.

Srivastava M.S., von Rosen T., von Rosen D. (2008): Models with a Kronecker product covariance structure estimation and testing, Mathematical Method of Statistics 17: 357–370.