Journal of Multivariate Analysis 113 (2013) 37–47

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

The multilinear normal distribution: Introduction and some basic properties

Martin Ohlson [a], M. Rauf Ahmad [a], Dietrich von Rosen [a,b,*]

[a] *Department of Mathematics, Linköping University, Sweden*
[b] *Department of Energy and Technology, Swedish University of Agricultural Sciences, Sweden*

ARTICLE INFO

Article history:
Available online 2 June 2011

AMS subject classification:
62H12
62H99

Keywords:
Flip-flop algorithm
Matrix normal distribution
Marginal and conditional distributions
Maximum likelihood estimators
Moments
Tensor product

ABSTRACT

In this paper, the multilinear normal distribution is introduced as an extension of the matrix-variate normal distribution. Basic properties such as marginal and conditional distributions, moments, and the characteristic function, are also presented. A trilinear example is used to explain the general contents at a simpler level. The estimation of parameters using a flip-flop algorithm is also briefly discussed.

1. Introduction

The matrix normal distribution, being an extension of the ordinary multivariate (vector-) normal distribution, can be regarded as a *bilinear* normal distribution — a distribution of a two-way (two-component) array, each component representing a vector of observations. The complexity of data, which has become a norm of the day for a variety of applied research areas, requires a consideration of extension of the bilinear normal distribution. The present paper presents this extension, correspondingly named *multilinear* normal distribution [20, Ch. 2], based on a parallel extension of bilinear matrices to multilinear tensors [9]. The adjective *multilinear* has not yet found its way into the general statistical literature. One may, however, trace the same or similar nomenclature with reference to the analysis of complicated data structures, with a commonly used alternative expression being *analysis of multiway data* [21, p. 16]. [21] also gives some useful references on multiway analysis, particularly based on tensor algebra; see also [10].

Compared to the multivariate normal distribution, the multilinear distribution has been a relatively uncharted territory of research. Still, however, some interesting and very useful applications of multilinear distribution can be found in the literature. Particularly, the emergence of complicated and enormous data sets in recent decades has given serious impetus for such applied literature to flourish. As a byproduct, this has caused a huge amount of literature on the theory and applications of tensors in statistics.

One of the most important uses of the multilinear normal (MLN) distribution, and hence tensor analysis, is perhaps in magnetic resonance imaging (MRI). A nice work, particularly focusing on the need to go from matrix-variate to tensor-based

* Corresponding author at: Department of Mathematics, Linköping University, Sweden.
E-mail addresses: martin.ohlson@liu.se (M. Ohlson), muahm@mai.liu.se (M. Rauf Ahmad), dietrich.von.rosen@et.slu.se, Dietrich.von.Rosen@bt.slu.se (D. von Rosen).

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.05.015

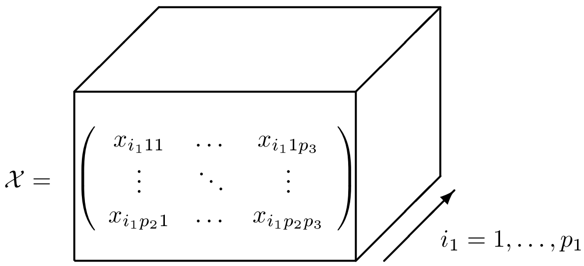

Fig. 1. The box visualizes a three-dimensional data set as a third order tensor.

MLN distribution, is given in [3]. They genuinely argue why a vectorial treatment of a complex data set which actually needs a tensorial treatment and the application of multilinear normality, can lead to wrong or inefficient conclusions. For some more relevant work in the same direction, see [2,4,5], and the references cited therein, whereas a Bayesian perspective is given in [25]; see also [15]. Analysis of multilinear, particularly trilinear data, has a specific attraction in chemometrics and spectroscopy; see for example [23,6]. Other areas of applications include signal processing [18], morphometry [22], geostatistics [24], and statistical mechanics [34], to mention a few. The extensive use of tensor variate analysis in these and other similar fields has generated a special tensorial nomenclature, for example diffusion tensor, dyadic tensor, stress and strain tensors etc. [27]. Similarly, special tensorial decompositions, for example PARAFAC and Tucker decompositions [21], have been developed; for a general comprehensive review of tensor decompositions and their various applications, see [8,19,32].

The use of a tensor, and its associated distributional structure, is even older, and with most frequent applications in the theory of linear models. Some classical treatises on tensors and multilinear algebra are [1,28,7]. For a comprehensive exposition of the use of tensors in statistics, see [27]. In another unique contribution, McCullagh had already introduced tensor notation in statistics with particular reference to the computation of polynomial cumulants [26]; see also [17,11]. The decomposition of ANOVA models into the potential sources of variation is always an important task in the theory of linear models. A tensorial treatment of ANOVA decomposition is given in [36], whereas a study of multilinear skewness and kurtosis in linear models is given in [30]; see also [12]. [14] gives an interesting application in the theory of design of experiments, with particular emphasis on rock magnetism. This paper uncovers some very attractive features of theoretical and geometrical aspects of tensors, when considered from a statistical perspective. The geometrical consideration of tensors in statistics, sometimes even more important than pure theoretical treatment, owes basically to setting the multivariate normality on the Riemannian geometry [33]. As the simplest case of geometrical structure of the parameter space of bivariate normal distribution, see [31], which also uses tensor notation to simplify complicated expressions.

This paper formally introduces MLN distribution, i.e., a normal distribution for the analysis of multiway data, and discusses some basic properties. The rest of the paper is organized as follows. Section 2 introduces the MLN distribution, along with some notation which simplifies the calculations that follow. In Section 3, some properties of the MLN distribution, such as marginal and conditional distributions, moments, and characteristic function, are given. A special case of trilinear normal distribution is interspersed throughout Sections 2 and 3 to explain the theory and notations at a more comprehensible level. Section 4 presents an estimation procedure for the parameters of the distribution.

2. Model

Let $\mathcal{X} = (x_{i_1,\ldots,i_k}) : \times_{i=1}^{k} p_i$ be a tensor of order k, with the dimensions $p_1, p_2, \ldots, p_k$. Fig. 1 shows the special case when $k = 3$. If $p_i = 1, 2 \leq i \leq k$ or $3 \leq i \leq k$ we have the special cases when the tensor equals a vector or a linear mapping.

In order to perform explicit computations, the tensor has to be represented via coordinates. In this paper, the representation will mainly be in vector form. However, the representation of the tensor $\mathcal{X} : \times_{i=1}^{k} p_i$ as a vector can be done in several ways. If we look at the tensor space in Fig. 1, this means that we can look upon the tensor from different directions.

Put $\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}} = \boldsymbol{e}_{i_1}^{p_1} \otimes \cdots \otimes \boldsymbol{e}_{i_k}^{p_k}$, where $\boldsymbol{p} = (p_1, \ldots, p_k)$ and $\otimes$ denotes the Kronecker product. To emphasize the dimension, we will write $\boldsymbol{p}_k$, or $\boldsymbol{p}(1 : k)$, instead of $\boldsymbol{p}$. The vectors $\boldsymbol{e}_j^p : p \times 1$ are the unit basis vectors, i.e., a p-vector with 1 in the jth position, and 0 elsewhere. Further, let

$$p_{j:l}^{*} = \prod_{i=j}^{l} p_i \quad \text{and} \quad p_{j:l}^{+} = \sum_{i=j}^{l} p_i, \tag{1}$$

with the special cases

$$p^{*} = p_{1:k}^{*} \quad \text{and} \quad p^{+} = p_{1:k}^{+}, \tag{2}$$

respectively. When there is no ambiguity, we shall drop the dimension from the basis vectors and write $\boldsymbol{e}_{i_1}^{p_1}$ as $\boldsymbol{e}_{i_1}$, and $\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}}$ as $\boldsymbol{e}_{i_1:i_k}$, etc. We begin with a formal definition of tensor space.

Definition 2.1.

(i) $\mathcal{T}^{\boldsymbol{p}} = \left\{\boldsymbol{x} : \boldsymbol{x} = \sum_{I_{\boldsymbol{p}}} x_{i_1\ldots i_k} \boldsymbol{e}^{\boldsymbol{p}}_{i_1:i_k}\right\}$, where

$$I_{\boldsymbol{p}} = \left\{i_1, \ldots, i_k : 1 \le i_j \le p_j, 1 \le j \le k\right\}$$

is the index set,

(ii) $\mathcal{T}^{\boldsymbol{pq}} = \left\{\boldsymbol{X} : \boldsymbol{X} = \sum_{I_{\boldsymbol{p}} \cup I_{\boldsymbol{q}}} x_{i_1\ldots i_k, j_1,\ldots,j_l} \boldsymbol{e}^{\boldsymbol{p}}_{i_1:i_k} (\boldsymbol{e}^{\boldsymbol{q}}_{j_1:j_l})'\right\}$, where

$$I_{\boldsymbol{q}} = \{j_1, \ldots, j_l : 1 \le j_i \le p_i, 1 \le i \le l\}$$

is another index set, $I_{\boldsymbol{p}}$ being the same as in (i) above,

(iii) $\mathcal{T}^{\boldsymbol{pq}}_{\otimes} = \{\boldsymbol{X} \in \mathcal{T}^{\boldsymbol{pq}} : \boldsymbol{X} = \boldsymbol{X}_1 \otimes \cdots \otimes \boldsymbol{X}_k$, where $\boldsymbol{X}_i : p_i \times q_i\}$,

(iv) $\mathcal{T}^{\boldsymbol{p}}_{\otimes} = \left\{\boldsymbol{X} \in \mathcal{T}^{\boldsymbol{pp}}_{\otimes} : \boldsymbol{X} = \boldsymbol{X}_1 \otimes \cdots \otimes \boldsymbol{X}_k\right.$, where $\left.\boldsymbol{X}_i : p_i \times p_i\right\}$.

Note that, the tensor space in (i) is described using vectors, whereas in (ii) using matrices.

The space $\mathcal{T}^{\boldsymbol{p}}$ defined in Definition 2.1 (i) is the space of vectorized tensors of size $p_1 \times p_2 \times \cdots \times p_k$; see [19] for more details about decompositions of tensors. In the following, we begin with an example of a trilinear normal distribution, which we shall continue to embed with several other main results to follow, to explain the general results at a simpler level.

Example 1. Let $\mathcal{X} = (x_{i_1,i_2,i_3}) : 3 \times 2 \times 2$ be a tensor of order 3. This tensor can be written as

$$\mathcal{X} = \begin{pmatrix} & & & & x_{311} & x_{312} \\ & & & & x_{321} & x_{322} \\ & & x_{211} & x_{212} & & \\ & & x_{221} & x_{222} & & \\ x_{111} & x_{112} & & & & \\ x_{121} & x_{122} & & & & \end{pmatrix},$$

where

$$\boldsymbol{X}_{1::} = \begin{pmatrix} x_{111} & x_{112} \\ x_{121} & x_{122} \end{pmatrix}, \quad \boldsymbol{X}_{:2:} \begin{pmatrix} x_{211} & x_{212} \\ x_{221} & x_{222} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{X}_{::3} = \begin{pmatrix} x_{311} & x_{312} \\ x_{321} & x_{322} \end{pmatrix},$$

are known as the horizontal slices of a tensor. One can similarly define lateral and frontal slices of a third order tensor by fixing the other two indices (see [19]). By Definition 2.1(i), the vectorization of the tensor is given as $\boldsymbol{x} = \mathrm{vec}\mathcal{X} \in \mathcal{T}^{\boldsymbol{p}}$, such that, in lexicographical order,

$$\boldsymbol{x} = (x_{111}, x_{112}, x_{121}, x_{122}, x_{211}, x_{212}, x_{221}, x_{222}, x_{311}, x_{312}, x_{321}, x_{322})'. \quad \square$$

The univariate, multivariate, and matrix normal distributions are well known. We may observe that a matrix-variate normal distribution, $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, can be defined as $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{\Psi}^{1/2}$, where $\boldsymbol{U} = (u_{vl})$, $u_{vl} \sim N(0, 1)$, i.i.d. This can also be written as

$$\sum_{ij} X_{ij} \boldsymbol{e}^p_i (\boldsymbol{e}^n_j)' = \sum_{ij} \mu_{ij} \boldsymbol{e}^p_i (\boldsymbol{e}^n_j)' + \sum_{ik} \sum_{vl} \sum_{mj} \tau_{ik} \delta_{mj} u_{vl} \boldsymbol{e}^p_i (\boldsymbol{e}^p_k)' \boldsymbol{e}^p_v (\boldsymbol{e}^n_l)' \boldsymbol{e}^n_m (\boldsymbol{e}^n_j)',$$

where $\boldsymbol{\Sigma} = \boldsymbol{\tau}\boldsymbol{\tau}'$ and $\boldsymbol{\Psi} = \boldsymbol{\delta}\boldsymbol{\delta}'$. Alternatively,

$$\sum_{ij} X_{ij} \boldsymbol{e}^p_i (\boldsymbol{e}^n_j)' = \sum_{ij} \boldsymbol{\mu}_{ij} \boldsymbol{e}^p_i (\boldsymbol{e}^n_j)' + \sum_{ij} \sum_{kl} \tau_{ik} \delta_{lj} u_{kl} \boldsymbol{e}^p_i (\boldsymbol{e}^n_j)'.$$

Writing the basis vectors as a Kronecker product, i.e., $\boldsymbol{e}^p_i (\boldsymbol{e}^n_j)' \to \boldsymbol{e}^n_j \otimes \boldsymbol{e}^p_i$, we obtain the vec-representation of the matrix normal distribution. This leads to the following extension of the matrix-variate normal distribution.

Definition 2.2. A tensor $\mathcal{X}$ is multilinear normal (MLN) of order k,

$$\mathcal{X} \sim N_{\boldsymbol{p}} (\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

if

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \boldsymbol{u},$$

where $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$, $\boldsymbol{\mu} \in \mathcal{T}^{\boldsymbol{p}}$, $\boldsymbol{\Sigma} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$, $\boldsymbol{p} = (p_1, \ldots, p_k)$, and the elements of $\boldsymbol{u} \in \mathcal{T}^{\boldsymbol{p}}$ are independent standard normally distributed. The square root $\boldsymbol{\Sigma}^{1/2}$ can be any square root.

The dispersion matrix in Definition 2.2, $\boldsymbol{\Sigma} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$, expanded in terms of its component matrices, can be written as $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{1:k} = \boldsymbol{\Sigma}_1 \otimes \cdots \otimes \boldsymbol{\Sigma}_k$. Moreover, $\mathcal{X} \sim N_{\boldsymbol{p}} (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{x} \sim N_{\boldsymbol{p}} (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pertain to the same distribution.

Relieving Definition 2.2 of its basis vectors, a coordinate-free (vector-space) version of the MLN distribution follows immediately as (see [37,13])

$$x_{i_1\dots i_k} = \mu_{i_1\dots i_k} + \sum_{j_1,\dots,j_k} \tau^1_{i_1j_1}\tau^2_{i_2j_2}\cdots\tau^k_{i_kj_k}u_{j_1j_2\dots j_k},$$

where $\boldsymbol{\Sigma}_i = \boldsymbol{\tau}^i(\boldsymbol{\tau}^i)' : p_i \times p_i$ with $\boldsymbol{\tau}^i = (\tau^i_{kl}) : p_i \times p_i$.

Using vector representation, $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$, and the fact that

$$\left|\boldsymbol{\Sigma}_{1:k}\right| = \left|\boldsymbol{\Sigma}_1 \otimes \cdots \otimes \boldsymbol{\Sigma}_k\right| = \prod_{i=1}^{k}\left|\boldsymbol{\Sigma}_i\right|^{p^*/(p_i)},$$

$$\boldsymbol{\Sigma}_{1:k}^{-1} = (\boldsymbol{\Sigma}_1 \otimes \cdots \otimes \boldsymbol{\Sigma}_k)^{-1} = \boldsymbol{\Sigma}_1^{-1} \otimes \cdots \otimes \boldsymbol{\Sigma}_k^{-1},$$

where p^* is defined in (2), we can conveniently write the probability density function (pdf) of an MLN distribution, extending the pdf of ordinary multivariate normal distribution.

Theorem 2.1. *The density function of the MLN distribution (Definition 2.2) is given as*

$$f_{\mathcal{X}}(\boldsymbol{x}) = (2\pi)^{-p^*/2}\left(\prod_{i=1}^{k}|\boldsymbol{\Sigma}_i|^{-p^*/(2p_i)}\right)\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{3}$$

where $\boldsymbol{\Sigma}_{1:k}$ is positive definite, $\boldsymbol{x}, \boldsymbol{\mu} \in \mathcal{T}^{\boldsymbol{p}}$, $\boldsymbol{\Sigma}_{1:k} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$, and p^* is defined in (2).

Example 1 (*Continued*)**.** Simplifying Theorem 2.1 to the notations of Example 1, the pdf of trilinear normal distribution of order $3 \times 2 \times 2$ can be written as follows:

$$f_{\mathcal{X}}(\boldsymbol{x}) = (2\pi)^{-6}|\boldsymbol{\Sigma}_1|^{-2}|\boldsymbol{\Sigma}_2|^{-3}|\boldsymbol{\Sigma}_3|^{-3}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \quad \square$$

We close this section by giving some comments on a special matrix which will be used in Section 4.

Definition 2.3. For some i, j with $i < j$, define

$$\boldsymbol{e}^{\boldsymbol{p}(i:j)}_{i:j} = \boldsymbol{e}^{p_i}_i \otimes \boldsymbol{e}^{p_{i+1}}_{i+1} \otimes \cdots \otimes \boldsymbol{e}^{p_j}_j.$$

The matrix $\boldsymbol{K}_{s,r} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$ is the tensor commutation operator, defined as an orthogonal matrix, satisfying

$$\boldsymbol{K}_{s,r}\boldsymbol{e}^{\boldsymbol{p}}_{i_1:i_k} = \boldsymbol{e}^{\boldsymbol{p}(1:s-1)}_{i_1:i_{s-1}} \otimes \boldsymbol{e}^{p_r}_{i_r} \otimes \boldsymbol{e}^{\boldsymbol{p}(s+1:r-1)}_{i_{s+1}:i_{r-1}} \otimes \boldsymbol{e}^{p_s}_{i_s} \otimes \boldsymbol{e}^{\boldsymbol{p}(r+1:k)}_{i_{r+1}:i_k}, \quad s \le r,$$

i.e., $\boldsymbol{K}_{s,r}$ interchanges basis vectors.

Observe that, multiplying with the commutation matrix $\boldsymbol{K}_{s,r}$ from the left will change rows. For notational convenience, we shall write

$$\boldsymbol{K}_{s,r}\boldsymbol{e}^{\boldsymbol{p}}_{i_1:i_k} = \boldsymbol{e}^{s,r}_{i_1:i_k}.$$

The tensor commutation operator that operates on the same lines as the well-known commutation matrix is used to interchange vectors in a Kronecker product of two vectors. Hence, for $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$, we write

$$\boldsymbol{K}_{s,r}\boldsymbol{x} = \boldsymbol{x}^{s,r}.$$

The following two theorems, about the properties of the tensor commutation operator, follow directly from Definition 2.3, and from properties of the commutation matrix.

Theorem 2.2. *Let $\boldsymbol{K}_{s,r} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$ be the tensor commutation operator (Definition 2.3). Then*

(i) $\boldsymbol{K}_{s,r} = \boldsymbol{K}'_{r,s}$, *and*
(ii) $\boldsymbol{K}_{s,r}\boldsymbol{K}_{r,s} = \boldsymbol{I}_{p^+}$.

Theorem 2.3. *Let $\boldsymbol{K}_{s,r} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$ be the tensor commutation operator, and $\boldsymbol{\Sigma}_{1:k} \in \mathcal{T}^{\boldsymbol{p}}_{\otimes}$. Then*

$$\boldsymbol{K}_{r,s}\boldsymbol{\Sigma}_{1:k}\boldsymbol{K}_{s,r} = \boldsymbol{\Sigma}_{1:s-1} \otimes \boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_{s+1:r-1} \otimes \boldsymbol{\Sigma}_s \otimes \boldsymbol{\Sigma}_{r+1:k},$$

where $\boldsymbol{\Sigma}_{i:j} = \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_{i+1} \otimes \cdots \otimes \boldsymbol{\Sigma}_j$, for some $i < j$.

3. Properties of the MLN distribution

In this section, we establish some properties of the MLN distribution, using the notations introduced in Section 2.

3.1. Moments, characteristic function and cumulants

When comparing the multivariate normal distribution with the MLN distribution, the difference lies in the structure of the parameter space generated by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The elements of both distributions are organized in vectors of non-repeated normal components. Thus, it is easy to imagine that moments for the MLN distribution can be obtained from those of the multivariate distribution. Indeed, the characteristic function and the cumulant generating function for the MLN distribution follow immediately from those of the multivariate normal distribution.

Theorem 3.1. *Let $\boldsymbol{x} \sim N_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$. The characteristic function of $\boldsymbol{x}$ is*

$$\varphi(\boldsymbol{t}) = E[\mathrm{e}^{\mathrm{i}\,\boldsymbol{t}'\boldsymbol{x}}] = \mathrm{e}^{\mathrm{i}\,\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}, \quad \boldsymbol{t} \in \mathcal{T}^{\boldsymbol{p}},$$

and the cumulant generating function is

$$\kappa(\boldsymbol{t}) = \ln E[\mathrm{e}^{\mathrm{i}\,\boldsymbol{t}'\boldsymbol{x}}] = \mathrm{i}\,\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}, \quad \boldsymbol{t} \in \mathcal{T}^{\boldsymbol{p}}.$$

To compute moments, we need a suitable differential operator (matrix derivative). Let $\boldsymbol{Y} \in \mathcal{T}^{\boldsymbol{pq}}$ be a function of $\boldsymbol{X} \in \mathcal{T}^{\boldsymbol{rs}}$, with their vectorized versions $\boldsymbol{y}$ and $\boldsymbol{x}$, defined as

$$\boldsymbol{y} = \sum_{i_1:i_{k_1}} \sum_{j_1:j_{k_2}} y_{i_1:i_{k_1} j_1:j_{k_2}} \boldsymbol{e}_{j_1:j_{k_2}}^{\boldsymbol{q}(1:k_2)} \otimes \boldsymbol{e}_{i_1:i_{k_1}}^{\boldsymbol{p}(1:k_1)},$$

$$\boldsymbol{x} = \sum_{m_1:m_{k_3}} \sum_{n_1:n_{k_4}} x_{m_1:m_{k_3} n_1:n_{k_4}} \boldsymbol{e}_{n_1:n_{k_4}}^{\boldsymbol{s}(1:k_4)} \otimes \boldsymbol{e}_{m_1:m_{k_3}}^{\boldsymbol{r}(1:k_3)},$$

respectively. Then,

$$\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}\boldsymbol{X}} = \frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}\boldsymbol{x}} = \sum_{i_1:i_{k_1}} \sum_{j_1:j_{k_2}} \sum_{m_1:m_{k_3}} \sum_{n_1:n_{k_4}} \frac{\partial y_{i_1:i_{k_1} j_1:j_{k_2}}}{\partial x_{m_1:m_{k_3} n_1:n_{k_4}}} \left(\boldsymbol{e}_{n_1:n_{k_4}}^{\boldsymbol{s}(1:k_4)} \otimes \boldsymbol{e}_{m_1:m_{k_3}}^{\boldsymbol{r}(1:k_3)}\right) \left(\boldsymbol{e}_{j_1:j_{k_2}}^{\boldsymbol{q}(1:k_2)} \otimes \boldsymbol{e}_{i_1:i_{k_1}}^{\boldsymbol{p}(1:k_1)}\right)'. \tag{4}$$

Higher order derivatives may be defined recursively, i.e.,

$$\frac{\mathrm{d}^k\boldsymbol{Y}}{\mathrm{d}\boldsymbol{X}^k} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{X}} \frac{\mathrm{d}^{k-1}\boldsymbol{Y}}{\mathrm{d}\boldsymbol{X}^{k-1}}.$$

Applying (4) to $\varphi(\boldsymbol{t})$, and evaluating the derivatives at $\boldsymbol{t} = \boldsymbol{0}$, we get

$$\left.\frac{\mathrm{d}\varphi(\boldsymbol{t})}{\mathrm{d}\boldsymbol{t}}\right|_{\boldsymbol{t}=\boldsymbol{0}} = (\mathrm{i}\boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t})\varphi(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} = \mathrm{i}\boldsymbol{\mu},$$

which, on further differentiation, gives

$$\left.\frac{\mathrm{d}^2\varphi(\boldsymbol{t})}{\mathrm{d}\boldsymbol{t}^2}\right|_{\boldsymbol{t}=\boldsymbol{0}} = \left[-\boldsymbol{\Sigma} + (\mathrm{i}\boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t})(\mathrm{i}\boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t})'\right]\varphi(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} = -\boldsymbol{\Sigma} - \boldsymbol{\mu}\boldsymbol{\mu}'.$$

The same moments can also be obtained using $\kappa(\boldsymbol{t})$, following the same lines as for multivariate normal distribution, where it can also be shown that all moments of order more than two are zero.

Using the same differential operator directly on the pdf, we get

$$\frac{\mathrm{d}^k f_{\mathcal{X}}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^k} = (-1)^k \boldsymbol{H}_k(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) f_{\mathcal{X}}(\boldsymbol{x}),$$

where $f_{\mathcal{X}}(\boldsymbol{x})$ is the pdf (Eq. (3)), and $\boldsymbol{H}_k(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the Hermite polynomials. Clearly, for $k = 0$, $\frac{\mathrm{d}^0 f_{\mathcal{X}}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{X}^0} = f_{\mathcal{X}}(\boldsymbol{x})$.

Theorem 3.2. *Let $\boldsymbol{x} \sim N_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$. The Hermite polynomials, $\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})k = 0, 1, 2$, are given by*

$$H_0(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 1,$$
$$\boldsymbol{H}_1(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}),$$
$$\boldsymbol{H}_2(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}.$$

One possible application of Hermite polynomials is in Edgeworth expansions. Finally, we state the following theorem which can be trivially proved using the invariance property of normal distribution under linear (in general, affine) transformation.

Theorem 3.3. *Let $\boldsymbol{x} \sim N_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$, and let $\boldsymbol{A} \in \mathcal{T}_{\otimes}^{\boldsymbol{qp}}$ is nonsingular. Then,*

$$\boldsymbol{A}\boldsymbol{x} \sim N_{\boldsymbol{q}}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'),$$

where $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}' \in \mathcal{T}_{\otimes}^{\boldsymbol{q}}$.

By appropriately choosing $\boldsymbol{A}$ in Theorem 3.3, several interesting special cases can be studied.

3.2. Marginal distributions

The matrix

$$\boldsymbol{M}_r = \sum_{I_r} \boldsymbol{e}_{i_1:i_{r-1}}^{\boldsymbol{p}(1:r-1)} \otimes \boldsymbol{e}_{i_r-r^{(1)}+1}^{r^{(2)}-r^{(1)}+1} \otimes \boldsymbol{e}_{i_{r+1}:i_k}^{\boldsymbol{p}(r+1:k)} \left(\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}}\right)',$$

with the index set

$$I_r = \left\{i_1, \dots, i_k : 1 \le i_t \le p_t, t = 1, \dots, r-1, r+1, \dots, k, r^{(1)} \le i_r \le r^{(2)}\right\},$$

facilitates the computation of several marginal distributions. The margins $r^{(1)}, r^{(2)}$ are known values. The index set I_r, by imparting restrictions on $\boldsymbol{M}_r$ through $r^{(1)}$ and $r^{(2)}$, generates marginal distributions, $\boldsymbol{M}_r\boldsymbol{x}$, represented as

$$\boldsymbol{M}_r\boldsymbol{x} \sim N_{p_r(r^{(2)}-r^{(1)}+1)}(\boldsymbol{M}_r\boldsymbol{\mu}, \boldsymbol{M}_r\boldsymbol{\Sigma}\boldsymbol{M}_r'). \tag{5}$$

In the sequel, we shall focus on the specific marginal distributions,

$$\boldsymbol{x}_{\bullet r_l} = \boldsymbol{M}_{r_1}\boldsymbol{x}, \tag{6}$$

$$\boldsymbol{x}_{\bullet r^l} = \boldsymbol{M}_{r_2}\boldsymbol{x}, \tag{7}$$

where

$$\boldsymbol{M}_{r_1} = \sum_{I_{r_1}} \boldsymbol{e}_{i_1:i_{r-1}}^{\boldsymbol{p}(1:r-1)} \otimes \boldsymbol{e}_{i_r}^{m} \otimes \boldsymbol{e}_{i_{r+1}:i_k}^{\boldsymbol{p}(r+1:k)} \left(\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}}\right)', \tag{8}$$

$$\boldsymbol{M}_{r_2} = \sum_{I_{r_2}} \boldsymbol{e}_{i_1:i_{r-1}}^{\boldsymbol{p}(1:r-1)} \otimes \boldsymbol{e}_{i_r-m}^{p_r-m} \otimes \boldsymbol{e}_{i_{r+1}:i_k}^{\boldsymbol{p}(r+1:k)} \left(\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}}\right)', \tag{9}$$

with their respective index sets

$$I_{r_1} = \{i_1, \dots, i_k : 1 \le i_t \le p_t, t = 1, \dots, r-1, r+1, \dots, k, 1 \le i_r \le m\},$$
$$I_{r_2} = \{i_1, \dots, i_k : 1 \le i_t \le p_t, t = 1, \dots, r-1, r+1, \dots, k, m+1 \le i_r \le p_r\}.$$

Example 1 (*Continued*)**.** For the third order tensor in Example 1, we calculate two marginal distributions, one for the slice $\boldsymbol{X}_{1::}$, and one for the the other two slices, $\boldsymbol{X}_{:2:}$ and $\boldsymbol{X}_{::3}$, combined. That is, the tensor is partitioned as

$$\mathcal{X} = \left(\begin{array}{cc|cccc} & & & & x_{311} & x_{312} \\ & & & & x_{321} & x_{322} \\ & & x_{211} & x_{212} & & \\ & & x_{221} & x_{222} & & \\ \hline x_{111} & x_{112} & & & & \\ x_{121} & x_{122} & & & & \end{array}\right),$$

with the index sets for the two partitions are given as

$$I_{r_1} = \{i_1, i_2, i_3 : i_1 = 1, 1 \le i_2 \le 2, 1 \le i_3 \le 2\},$$
$$I_{r_2} = \{i_1, i_2, i_3 : 2 \le i_1 \le 3, 1 \le i_2 \le 2, 1 \le i_3 \le 2\},$$

respectively. Then, from Eqs. (8) and (9), we have

$$\begin{aligned}
\boldsymbol{M}_{r_1} &= \sum_{i_2=1,2}\sum_{i_3=1,2} 1 \otimes \boldsymbol{e}_{i_2}^2 \otimes \boldsymbol{e}_{i_3}^2 \left(\boldsymbol{e}_1^3 \otimes \boldsymbol{e}_{i_2}^2 \otimes \boldsymbol{e}_{i_3}^2\right)' \\
&= \sum_{i_2=1,2}\sum_{i_3=1,2} (\boldsymbol{e}_1^3)' \otimes \boldsymbol{e}_{i_2}^2(\boldsymbol{e}_{i_2}^2)' \otimes \boldsymbol{e}_{i_3}^2(\boldsymbol{e}_{i_3}^2)' = (\boldsymbol{e}_1^3)' \otimes \boldsymbol{I}_4, \\
\boldsymbol{M}_{r_2} &= \sum_{i_1=2,3}\sum_{i_2=1,2}\sum_{i_3=1,2} \boldsymbol{e}_{i_1-1}^2 \otimes \boldsymbol{e}_{i_2}^2 \otimes \boldsymbol{e}_{i_3}^2 \left(\boldsymbol{e}_{i_1}^3 \otimes \boldsymbol{e}_{i_2}^2 \otimes \boldsymbol{e}_{i_3}^2\right)' \\
&= (\boldsymbol{0}_2 : \boldsymbol{I}_2) \otimes \boldsymbol{I}_4,
\end{aligned}$$

where $\boldsymbol{0}_2 = (0\ ,\ 0)'$. This gives the marginal vectors

$$\boldsymbol{x}_{\bullet r_l} = \boldsymbol{M}_{r_1}\boldsymbol{x} = (x_{111}, x_{112}, x_{121}, x_{122})'$$
$$\boldsymbol{x}_{\bullet r^l} = \boldsymbol{M}_{r_2}\boldsymbol{x} = (x_{211}, x_{212}, x_{221}, x_{222}, x_{311}, x_{312}, x_{321}, x_{322})',$$

so that the corresponding marginal distributions can be calculated from (5). □

The following three theorems specify certain independence conditions on the marginals.

Theorem 3.4. *The normal variables* $\boldsymbol{x}_{\bullet r_l}$ *and* $\boldsymbol{x}_{\bullet r^l}$ *are independent, if and only if,* $\boldsymbol{\Sigma}_{12}^r = \mathbf{0}$*, where* $\boldsymbol{\Sigma}_{12}^r$*, of size* $m \times (p_r - m)$*, is the upper right partition of* $\boldsymbol{\Sigma}_r$ *in* $\boldsymbol{\Sigma}_{1:k}$*.*

Proof. Because of normality, independence holds if and only if

$$\begin{aligned}\mathbf{0} &= C[\boldsymbol{x}_{\bullet r_l}, \boldsymbol{x}_{\bullet r^l}] = \boldsymbol{M}_{r_1} C[\boldsymbol{x}, \boldsymbol{x}] \boldsymbol{M}_{r^1}' = \boldsymbol{M}_{r_1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1}' \\ &= \left(\boldsymbol{I}_{p_{1:r-1}^+} \otimes (\boldsymbol{I} : \mathbf{0}) \otimes \boldsymbol{I}_{p_{r+1:k}^+}\right) \boldsymbol{\Sigma}_{1:k} \left(\boldsymbol{I}_{p_{1:r-1}^+} \otimes (\mathbf{0} : \boldsymbol{I})' \otimes \boldsymbol{I}_{p_{r+1:k}^+}\right) \\ &= \boldsymbol{\Sigma}_{1:r-1} \otimes \boldsymbol{\Sigma}_{12}^r \otimes \boldsymbol{\Sigma}_{r+1:k},\end{aligned} \tag{10}$$

where $p_{j:l}^+$ is defined in (1). Since $\boldsymbol{\Sigma}_i$, $i = 1, 2, \dots k$, differ from zero, (10) holds if and only if $\boldsymbol{\Sigma}_{12}^r = \mathbf{0}$. □

Clearly, with $r = 1$ and $r = k$, we can reduce Theorem 3.4 to obvious special cases.

Example 1 (*Continued*)**.** First, we note that the covariance matrix for the tensor $\mathcal{X}$ in Example 1 can be written as

$$\begin{aligned}C(\mathcal{X}) &= C(\mathrm{vec}\mathcal{X}) = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_3 \\ &= \begin{pmatrix} \sigma_{11}^1 \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_3 & (\boldsymbol{\sigma}_{12}^1)' \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_3 \\ \boldsymbol{\sigma}_{12}^1 \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_3 & \boldsymbol{\Sigma}_{22}^1 \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_3 \end{pmatrix},\end{aligned}$$

where $\boldsymbol{\Sigma}_1$ is partitioned according to the margins, i.e., $\sigma_{11}^1 : 1 \times 1$, $(\boldsymbol{\sigma}_{12}^1)' : 1 \times 2$ and $\boldsymbol{\Sigma}_{22}^1 : 2 \times 2$. Then, by the property of normal distribution, $\boldsymbol{x}_{\bullet r_l}$ and $\boldsymbol{x}_{\bullet r^l}$ are independent if and only if $\boldsymbol{\sigma}_{12}^1 = \mathbf{0}$. □

Theorem 3.5. *The normal variables* $\boldsymbol{x}_{\bullet r_l}$ *and* $\boldsymbol{x}_{\bullet s^l}$ *are not independent, if* $s \neq r$*.*

Proof. The statement is evident from Eq. (10), i.e., for any matrices $\boldsymbol{P}_i, \boldsymbol{Q}_i$, $i = 1, 2$, $(\boldsymbol{P}_1 \otimes \boldsymbol{P}_2)(\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2) = \mathbf{0}$ if and only if $\boldsymbol{P}_i \boldsymbol{Q}_i = \mathbf{0}$ for either $i = 1$ or $i = 2$. □

Finally, Theorem 3.4 is the special case of the following theorem.

Theorem 3.6. *Let* $\boldsymbol{A} \in \mathcal{T}^{\boldsymbol{sp}}$ *and* $\boldsymbol{B} \in \mathcal{T}^{\boldsymbol{tp}}$*. Then,* $\boldsymbol{Ax}$ *and* $\boldsymbol{Bx}$ *are independent if and only if* $\boldsymbol{A\Sigma B}' = \mathbf{0}$*, i.e., if for some* r*,* $\boldsymbol{A}_r \boldsymbol{\Sigma}_r \boldsymbol{B}_r' = \mathbf{0}$*.*

3.3. Conditional distributions

Having the joint and marginal densities, we can compute the conditional densities. Define

$$\boldsymbol{\mu}_{\bullet r_l} = \boldsymbol{M}_{r_1} \boldsymbol{\mu}, \tag{11}$$

$$\boldsymbol{\mu}_{\bullet r^l} = \boldsymbol{M}_{r_2} \boldsymbol{\mu}, \tag{12}$$

$$\boldsymbol{\Sigma}_{1\bullet 2}^r = \boldsymbol{\Sigma}_{11}^r - \boldsymbol{\Sigma}_{12}^r (\boldsymbol{\Sigma}_{22}^r)^{-1} \boldsymbol{\Sigma}_{21}^r, \qquad \boldsymbol{\Sigma}_r = (\boldsymbol{\Sigma}_{ij}^r). \tag{13}$$

Then, we have the following theorem.

Theorem 3.7. *Let* $\boldsymbol{x} \sim N_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, where* $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$*. Let* $\boldsymbol{x}_{\bullet r_l}, \boldsymbol{x}_{\bullet r^l}, \boldsymbol{\mu}_{\bullet r_l}, \boldsymbol{\mu}_{\bullet r^l}$*, be as defined in (6), (7), (11) and (12), respectively. Then,* $\boldsymbol{x}_{\bullet r_l} | \boldsymbol{x}_{\bullet r^l}$ *has the same distribution as*

$$\boldsymbol{\mu}_{\bullet r_l} + (\boldsymbol{I}_{p_{1:r-1}^+} \otimes \boldsymbol{\Sigma}_{12}^r (\boldsymbol{\Sigma}_{22}^r)^{-1} \otimes \boldsymbol{I}_{p_{r+1:k}^+})(\boldsymbol{x}_{\bullet r^l} - \boldsymbol{\mu}_{\bullet r^l}) + (\boldsymbol{\Sigma}_{1:r-1} \otimes \boldsymbol{\Sigma}_{1\bullet 2}^r \otimes \boldsymbol{\Sigma}_{r+1:k})^{1/2} \boldsymbol{u}, \tag{14}$$

where $\boldsymbol{u} \sim N_{p^+}(\mathbf{0}, \boldsymbol{I}_{p^+})$*, and* $p_{j:l}^+$ *and* p^+ *are defined in (1) and (2), respectively.*

Proof. The normality of the conditional distribution from the properties of the normal distribution. Then, we only need to compute the conditional mean and dispersion. Since

$$\begin{aligned}D[\boldsymbol{M}_{r_1} \boldsymbol{x}] &= \boldsymbol{M}_{r_1} \boldsymbol{\Sigma} \boldsymbol{M}_{r_1}', \\ D[\boldsymbol{M}_{r^1} \boldsymbol{x}] &= \boldsymbol{M}_{r^1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1}', \\ C[\boldsymbol{M}_{r_1} \boldsymbol{x}, \boldsymbol{M}_{r^1} \boldsymbol{x}] &= \boldsymbol{M}_{r_1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1}',\end{aligned}$$

the conditional mean is

$$E[\boldsymbol{M}_{r_1} \boldsymbol{x} | \boldsymbol{M}_{r^1} \boldsymbol{x}] = \boldsymbol{M}_{r_1} \boldsymbol{\mu} + \boldsymbol{M}_{r_1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1} (\boldsymbol{M}_{r^1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1}')^{-1} (\boldsymbol{M}_{r^1} \boldsymbol{x} - \boldsymbol{M}_{r^1} \boldsymbol{\mu})$$

which is identical to (14). The conditional dispersion is

$$\begin{aligned}D[\boldsymbol{M}_{r_1} \boldsymbol{x} | \boldsymbol{M}_{r^1} \boldsymbol{x}] &= \boldsymbol{M}_{r_1} \boldsymbol{\Sigma} \boldsymbol{M}_{r_1}' - \boldsymbol{M}_{r_1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1} (\boldsymbol{M}_{r^1} \boldsymbol{\Sigma} \boldsymbol{M}_{r^1}')^{-1} \boldsymbol{M}_{r^1} \boldsymbol{\Sigma} \boldsymbol{M}_{r_1} \\ &= \boldsymbol{\Sigma}_{1:r-1} \otimes \boldsymbol{\Sigma}_{1\bullet 2}^r \otimes \boldsymbol{\Sigma}_{r+1:k}. \quad \square\end{aligned}$$

4. Inference

In what follows, it is explained how maximum likelihood estimators of the unknown parameters in the MLN can be obtained. The likelihood equations for the covariance matrices, $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$ of $\boldsymbol{\Sigma}_{1:k}$, will be derived. To achieve unique parametrization, we need to put certain restrictions on a subset of the parameters, for example,

$$\sigma^{(2)}_{p_2p_2} = \sigma^{(3)}_{p_3p_3} = \cdots = \sigma^{(k)}_{p_kp_k} = 1, \tag{15}$$

where $\boldsymbol{\Sigma}_r = \left(\sigma^{(r)}_{ij}\right)$; see [35]. A Bayesian estimation strategy for a multidimensional separable covariance matrix, without restrictions like (15), is given in [16]. Assume that there are n independent tensor observations $\mathcal{X}_j : \times_{i=1}^{k} p_i, j = 1, \ldots, n$, from $f_{\mathcal{X}}(\boldsymbol{x})$ in (3), with vector representations $\boldsymbol{x}_j$. From the joint likelihood function, one can easily see that $\boldsymbol{\mu}$ will be estimated by averaging all vectors of observations. Hence, in the subsequent computations, we assume $\boldsymbol{\mu} = \boldsymbol{0}$, without any loss of generality. Then, the likelihood function for $\boldsymbol{\Sigma}_{1:k}$ is given by

$$L = L(\boldsymbol{\Sigma}_{1:k}) = (2\pi)^{-p^*/2} \prod_{i=1}^{k} |\boldsymbol{\Sigma}_i|^{-p^*n/(2p_i)} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \boldsymbol{x}_j' \boldsymbol{\Sigma}_{1:k}^{-1} \boldsymbol{x}_j \right\}, \tag{16}$$

which can also be written as

$$L = (2\pi)^{-p^*/2} \prod_{i=1}^{k} |\boldsymbol{\Sigma}_i|^{-p^*n/(2p_i)} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \operatorname{tr}\left\{ \boldsymbol{\Sigma}_1^{-1} \left(\boldsymbol{X}_j^{(1)}\right)' \boldsymbol{\Sigma}_{2:k}^{-1} \boldsymbol{X}_j^{(1)} \right\} \right\}, \tag{17}$$

where

$$\begin{aligned} \boldsymbol{X}_j^{(1)} &= \sum_{I_{\boldsymbol{p}}} x_{i_1 \ldots i_k} (\boldsymbol{e}_{i_2}^{p_2} \otimes \cdots \otimes \boldsymbol{e}_{i_k}^{p_k})(\boldsymbol{e}_{i_1}^{p_1})' \\ &= \sum_{I_{\boldsymbol{p}}} x_{i_1 \ldots i_k} \boldsymbol{e}_{i_2:i_k}^{\boldsymbol{p}_{2:k}} (\boldsymbol{e}_{i_1}^{p_1})' : p_{2:k}^* \times p_1, \end{aligned}$$

since $\operatorname{vec} \boldsymbol{X}_j^{(1)} = \boldsymbol{x}_j$. For simplicity, we will omit the upper index "(1)" and write the matrix as $\boldsymbol{X}_j^{(1)} = \boldsymbol{X}_j$. Now the trace in (17) can be rewritten as (see also [29])

$$\begin{aligned} \operatorname{tr}\left\{ \boldsymbol{\Sigma}_1^{-1} \boldsymbol{X}_j' (\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_{3:k})^{-1} \boldsymbol{X}_j \right\} &= \operatorname{tr}\left\{ \boldsymbol{\Sigma}_1^{-1} \boldsymbol{X}_j' \left(\boldsymbol{I}_{p_2} \otimes \boldsymbol{\Sigma}_{3:k}^{-1/2}\right) \left(\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{I}_{p_{3:k}^*}\right) \left(\boldsymbol{I}_{p_2} \otimes \boldsymbol{\Sigma}_{3:k}^{-1/2}\right) \boldsymbol{X}_j \right\} \\ &= \sum_{l=1}^{p_{3:k}^*} \operatorname{tr}\left\{ \boldsymbol{\Sigma}_1^{-1} \boldsymbol{X}_j' \left(\boldsymbol{I}_{p_2} \otimes \left(\boldsymbol{\Sigma}_{3:k}^{-1/2} \boldsymbol{e}_l^{p_{3:k}^*}\right)\right) \boldsymbol{\Sigma}_2^{-1} \left(\boldsymbol{I}_{p_2} \otimes \left(\left(\boldsymbol{e}_l^{p_{3:k}^*}\right)' \boldsymbol{\Sigma}_{3:k}^{-1/2}\right)\right) \boldsymbol{X}_j \right\} \\ &= \sum_{l=1}^{p_{3:k}^*} \operatorname{tr}\left\{ \boldsymbol{\Sigma}_1^{-1} \boldsymbol{Y}_{jl}' \boldsymbol{\Sigma}_2^{-1} \boldsymbol{Y}_{jl} \right\}, \end{aligned} \tag{18}$$

where

$$\boldsymbol{Y}_{jl} = \left(\boldsymbol{I}_{p_2} \otimes \left(\left(\boldsymbol{e}_l^{p_{3:k}^*}\right)' \boldsymbol{\Sigma}_{3:k}^{-1/2}\right)\right) \boldsymbol{X}_j.$$

Hence, given $\boldsymbol{\Sigma}_{3:k}$, we have $n\,p_{3:k}^*$ independent observations $\boldsymbol{Y}_{jl}, j = 1, \ldots, n, l = 1, \ldots, p_{3:k}^*$, respectively.

Under the condition $\sigma^{(2)}_{p_2p_2} = 1$, the likelihood equations for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ follow from [35]

$$\boldsymbol{\Sigma}_1 = \frac{1}{p_{2:k}^* n} \sum_{j=1}^{n} \sum_{l=1}^{p_{3:k}^*} \boldsymbol{Y}_{jl}' \boldsymbol{\Sigma}_2^{-1} \boldsymbol{Y}_{jl}, \tag{19}$$

$$\boldsymbol{\Sigma}_2 = \frac{1}{p_1 p_{3:k}^* n} \sum_{j=1}^{n} \sum_{l=1}^{p_{3:k}^*} \boldsymbol{Y}_{jl} \boldsymbol{\Sigma}_1^{-1} \boldsymbol{Y}_{jl}'. \tag{20}$$

Rewriting (19), we have

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \frac{1}{p_{2:k}^* n} \sum_{j=1}^{n} \sum_{l=1}^{p_{3:k}^*} \boldsymbol{Y}_{jl}' \boldsymbol{\Sigma}_2^{-1} \boldsymbol{Y}_{jl} \\ &= \frac{1}{p_{2:k}^* n} \sum_{j=1}^{n} \sum_{l=1}^{p_{3:k}^*} \boldsymbol{X}_j' \left(\boldsymbol{I}_{p_2} \otimes \left(\boldsymbol{\Sigma}_{3:k}^{-1/2} \boldsymbol{e}_l^{p_{3:k}^*}\right)\right) \boldsymbol{\Sigma}_2^{-1} \left(\boldsymbol{I}_{p_2} \otimes \left(\left(\boldsymbol{e}_l^{p_{3:k}^*}\right)' \boldsymbol{\Sigma}_{3:k}^{-1/2}\right)\right) \boldsymbol{X}_j \end{aligned}$$

$$= \frac{1}{p^*_{2:k}n}\sum_{j=1}^{n} \boldsymbol{X}_j'\boldsymbol{\Sigma}_{2:k}^{-1}\boldsymbol{X}_j. \tag{21}$$

Using the tensor commutation operator (Definition 2.3), we can also write the likelihood function as

$$L = (2\pi)^{-p^*/2}\prod_{i=1}^{k}|\boldsymbol{\Sigma}_i|^{-p^*n/(2p_i)}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{x}_j^{s,r})'\left(\boldsymbol{\Sigma}_{1:k}^{s,r}\right)^{-1}\boldsymbol{x}_j^{s,r}\right\} \tag{22}$$

and, for $s = 2, r \neq 1, 2$,

$$L = (2\pi)^{-p^*/2}\prod_{i=1}^{k}|\boldsymbol{\Sigma}_i|^{-p^*n/(2p_i)}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\mathrm{tr}\left\{\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{X}_j^{2,r}\right)'\left(\boldsymbol{\Sigma}_{2:k}^{2,r}\right)^{-1}\boldsymbol{X}_j^{2,r}\right\}\right\}, \tag{23}$$

where

$$\boldsymbol{X}_j^{2,r} = \sum_{i_1,\ldots,i_k} x^{(j)}_{i_1\ldots i_k}\boldsymbol{e}^{2,r}_{i_2:i_k}(\boldsymbol{e}^{p_1}_{i_1})' : p^*_{2:k}\times p_1,$$

and, for notational convenience, we have used $\boldsymbol{K}_{r,s}\boldsymbol{\Sigma}_{1:k}\boldsymbol{K}_{s,r} = \boldsymbol{\Sigma}^{s,r}_{1:k}$. Now, the trace in (23) equals

$$\mathrm{tr}\left\{\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{X}_j^{2,r}\right)'\left(\boldsymbol{\Sigma}_{2:k}^{2,r}\right)^{-1}\boldsymbol{X}_j^{2,r}\right\} = \sum_{l=1}^{p^*_{2:k}/p_r}\mathrm{tr}\left\{\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{Y}_{jl}^{2,r}\right)'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{Y}_{jl}^{2,r}\right\}, \tag{24}$$

where

$$\boldsymbol{Y}_{jl}^{2,r} = \left(\boldsymbol{I}_{p_r}\otimes\left(\left(\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\right)\right)\boldsymbol{X}_j^{2,r},$$

$$\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r} = \boldsymbol{\Sigma}_3\otimes\cdots\otimes\boldsymbol{\Sigma}_{r-1}\otimes\boldsymbol{\Sigma}_2\otimes\boldsymbol{\Sigma}_{r+1}\otimes\cdots\otimes\boldsymbol{\Sigma}_k, \tag{25}$$

i.e., $\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}$ is $\boldsymbol{\Sigma}^{2,r}_{2:k}$ with $\boldsymbol{\Sigma}_r$ deleted. Using a similar notation for basis vectors, we write $\boldsymbol{e}^{2,r}_{i_1:i_k\backslash i_r}$ as $\boldsymbol{e}^{2,r}_{i_1:i_k}$ with $\boldsymbol{e}_{i_r}$ removed, so that

$$\boldsymbol{e}^{2,r}_{i_1:i_k\backslash i_r} = \boldsymbol{e}^{p_1}_{i_1}\otimes\boldsymbol{e}^{p_3}_{i_3}\otimes\cdots\otimes\boldsymbol{e}^{p_{r-1}}_{i_{r-1}}\otimes\boldsymbol{e}^{p_2}_{i_2}\otimes\boldsymbol{e}^{p_{r+1}}_{i_{r+1}}\otimes\cdots\otimes\boldsymbol{e}^{p_k}_{i_k}. \tag{26}$$

Then, given $\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}$, we have $n\,p^*_{2:k}/p_r = n\,p^*_{1:r-1}p^*_{r+1:k}$ independent observations. Again, since $\sigma^{(r)}_{p_rp_r} = 1$, using the techniques in [35], the following likelihood equations for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_r$ become

$$\boldsymbol{\Sigma}_1 = \frac{1}{p^*_{2:k}n}\sum_{j=1}^{n}\sum_{l=1}^{p^*_{2:k}/p_r}\left(\boldsymbol{Y}_{jl}^{2,r}\right)'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{Y}_{jl}^{2,r}, \tag{27}$$

$$\boldsymbol{\Sigma}_r = \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{l=1}^{p^*_{2:k}/p_r}\boldsymbol{Y}_{jl}^{2,r}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{Y}_{jl}^{2,r}\right)'. \tag{28}$$

The following theorem can now be stated.

Theorem 4.1. *The likelihood equations that are maximizing the likelihood function (16) under the conditions $\sigma^{(2)}_{p_2p_2} = \sigma^{(3)}_{p_3p_3} = \cdots = \sigma^{(k)}_{p_kp_k} = 1$ are given by*

$$\boldsymbol{\Sigma}_1 = \frac{1}{p^*_{2:k}n}\sum_{j=1}^{n}\boldsymbol{X}_j'\boldsymbol{\Sigma}_{2:k}^{-1}\boldsymbol{X}_j \tag{29}$$

and, for $r = 2, \ldots, k$

$$\boldsymbol{\Sigma}_r = \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\left(\boldsymbol{X}_j^{2,r(r)}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{1:k\backslash r}\right)^{-1}\boldsymbol{X}_j^{2,r(r)}, \tag{30}$$

where $\boldsymbol{X}_j^{2,r(r)} = \sum_{I_{\boldsymbol{p}}} x^{(j)}_{i_1\ldots i_k}\boldsymbol{e}^{2,r}_{i_1:i_k\backslash i_r}(\boldsymbol{e}^{p_r}_{i_r})'$, and $\boldsymbol{\Sigma}^{2,r}_{1:k\backslash r}$ and $\boldsymbol{e}^{2,r}_{i_1:i_k\backslash i_r}$ are given above in (25) and (26), respectively.

Proof. $\boldsymbol{\Sigma}_1$ is given in (21). We will now prove (30). From (28), we have

$$\begin{aligned}
\boldsymbol{\Sigma}_r &= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{l=1}^{p^*_{2:k}/p_r}\boldsymbol{Y}^{2,r}_{jl}\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{Y}^{2,r}_{jl}\right)' \\
&= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{l=1}^{p^*_{2:k}/p_r}\left(\boldsymbol{I}_{p_r}\otimes\left(\left(\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\right)\right)\boldsymbol{X}^{2,r}_j\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{X}^{2,r}_j\right)'\left(\boldsymbol{I}_{p_r}\otimes\left(\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)\right) \\
&= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{I_{\boldsymbol{p}}}\sum_{I'_{\boldsymbol{p}}}x^{(j)}_{i_1\ldots i_k}x^{(j)}_{i'_1\ldots i'_k}(\boldsymbol{e}^{p_1}_{i_1})'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}^{p_1}_{i'_1}\left(\sum_{l=1}^{p^*_{2:k}/p_r}\left(\boldsymbol{I}_{p_r}\otimes\left(\left(\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\right)\right)\boldsymbol{e}^{2,r}_{i_2:i_k}\left(\boldsymbol{e}^{2,r}_{i'_2:i'_k}\right)'\right. \\
&\qquad\left.\times\left(\boldsymbol{I}_{p_r}\otimes\left(\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)\right)\right) \\
&= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{I_{\boldsymbol{p}}}\sum_{I'_{\boldsymbol{p}}}x^{(j)}_{i_1\ldots i_k}x^{(j)}_{i'_1\ldots i'_k}(\boldsymbol{e}^{p_1}_{i_1})'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}^{p_1}_{i'_1}\left(\sum_{l=1}^{p^*_{2:k}/p_r}\boldsymbol{e}^{p_r}_{i_r}\left(\boldsymbol{e}^{p_r}_{i'_r}\right)'\otimes\left(\left(\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\boldsymbol{e}^{2,r}_{i_2:i_k\backslash i_r}\right)\right. \\
&\qquad\left.\times\left(\left(\boldsymbol{e}^{2,r}_{i'_2:i'_k\backslash i'_r}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1/2}\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)\right) \\
&= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{I_{\boldsymbol{p}}}\sum_{I'_{\boldsymbol{p}}}x^{(j)}_{i_1\ldots i_k}x^{(j)}_{i'_1\ldots i'_k}(\boldsymbol{e}^{p_1}_{i_1})'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{e}^{p_1}_{i'_1}\left(\boldsymbol{e}^{2,r}_{i_2:i_k\backslash i_r}\right)'\left(\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1}\boldsymbol{e}^{2,r}_{i'_2:i'_k\backslash i'_r}\left(\boldsymbol{e}^{p_r}_{i_r}\left(\boldsymbol{e}^{p_r}_{i'_r}\right)'\right) \\
&= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\sum_{I_{\boldsymbol{p}}}\sum_{I'_{\boldsymbol{p}}}x^{(j)}_{i_1\ldots i_k}x^{(j)}_{i'_1\ldots i'_k}\left(\left(\boldsymbol{e}^{p_1}_{i_1}\otimes\boldsymbol{e}^{2,r}_{i_2:i_k\backslash i_r}\right)\left(\boldsymbol{e}^{p_r}_{i_r}\right)'\right)'\left(\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Sigma}^{2,r}_{2:k\backslash r}\right)^{-1}\left(\boldsymbol{e}^{p_1}_{i'_1}\otimes\boldsymbol{e}^{2,r}_{i'_2:i'_k\backslash i'_r}\right)\left(\boldsymbol{e}^{p_r}_{i'_r}\right)' \\
&= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\left(\boldsymbol{X}^{2,r(r)}_j\right)'\left(\boldsymbol{\Sigma}^{2,r}_{1:k\backslash r}\right)^{-1}\boldsymbol{X}^{2,r(r)}_j,
\end{aligned}$$

since

$$\sum_{l=1}^{p^*_{2:k}/p_r}\boldsymbol{e}_l^{p^*_{2:k}/p_r}\left(\boldsymbol{e}_l^{p^*_{2:k}/p_r}\right)' = \boldsymbol{I}_{p^*_{2:k}/p_r}.$$

Thus the proof is complete. □

The likelihood equations (29) and (30) are nested, for which no explicit solution exists. One way to solve these equations is to use the so-called flip-flop algorithm [35]. We may also note that by using the results of [35], the algorithm converges to a unique solution provided there is enough data, and $\sigma^{(2)}_{p_2p_2} = \sigma^{(3)}_{p_3p_3} = \cdots = \sigma^{(k)}_{p_kp_k} = 1$ is chosen to be a part of the starting values.

References

[1] G.B. Arfken, H.J. Weber, Mathematical Methods for Physicists, fourth ed., Academic Press, New York, 1995.
[2] P.J. Basser, S. Pajevic, Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise, Magnetic Resonance in Medicine 44 (2000) 41–50.
[3] P.J. Basser, S. Pajevic, A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI, IEEE Transactions on Medical Imaging 22 (7) (2003) 785–794.
[4] P.J. Basser, S. Pajevic, Spectral decomposition of a 4th-order covariance tensor: applications to diffusion tensor MRI, Signal Processing 87 (2007) 220–236.
[5] D.L. Bihan, J.-F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, H. Chabriat, Diffusion tensor imaging: concepts and applications, Journal of Magnetic Resonance Imaging 13 (2001) 534–546.
[6] D.S. Burdick, An introduction to tensor products with applications to multiway data analysis, Chemometrics and Intelligent Laboratory Systems 28 (1995) 229–237.
[7] M. Carmeli, S. Milon, Theory of Spinors, World Scientific, Singapore, 2000.
[8] P. Comon, Tensor decompositions: state of the art and applications, in: J.G. McWirter, I.K. Proudler (Eds.), Mathematics in Signal Processing, V, Oxford University Press, Oxford, 2001, pp. 1–24.
[9] P. Comon, 2009. Tensors versus matrices: usefulness and unexpected properties, in: IEEE/SP 15th Workshop on Statistical Signal Processing. Cardiff, UK.
[10] R. Coppi, S. Bolasco (Eds.), Multiway Data Analysis, Elsevier, Amsterdam, 1989.
[11] J. Dauxois, Y. Romain, S. Viguier-Pla, Tensor products and statistics, Linear Algebra and its Applications 210 (1984) 59–88.

[12] H. Drygas, Linear sufficiency and some applications in multilinear estimation, Journal of Multivariate Analysis 16 (1985) 71–84.
[13] M.L. Eaton, Multivariate Statistics: A Vector Space Approach, in: Lecture Notes - Monograph Series, vol. 53, Institute of Mathematical Statistics, Ohio, 2007.
[14] G.R. Hext, The estimation of second-order tensors, with related tests and designs, Biometrika 50 (3/4) (1963) 353–373.
[15] P.D. Hoff, Hierarchical multilinear models for multiway data, Computational Statistics & Data Analysis 55 (2011) 530–543.
[16] P.D. Hoff, Separable covariance arrays via the Tucker product, with applications to multivariate relational data, Bayesian Analysis 6 (1) (2011) 1–18.
[17] E.L. Kaplan, Tensor notation and the sampling cumulants of *k*-statistics, Biometrika 39 (3/4) (1952) 319–323.
[18] E. Kofidis, P.A. Regalia, Tensor approximation and signal processing applications, in: V. Olshevsky (Ed.), Structural Matrices in Mathematics, Computer Science, and Engineering, Vol. I, in: Contemporary Mathematics, vol. 280, American Mathematical Society, Providence, 2001, pp. 103–133.
[19] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.
[20] T. Kollo, D. von Rosen, Advanced Multivariate Statistics with Matrices, Springer, Dordrecht, The Netherlands, 2005.
[21] P.M. Kroonberg, Applied Multiway Data Analysis, Wiley, New York, 2008.
[22] N. Lepore, C. Brun, Y.Y. Chou, M.C. Chiang, R.A. Dutton, K.M. Hayashi, E. Luders, O.L. Lopez, H.J. Aizenstein, A.W. Toga, J.T. Becker, P.M. Thompson, Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics of deformation tensors, IEEE Transactions on Medical Imaging 27 (1) (2008) 129–141.
[23] S. Leurgens, R.T. Ross, Multilinear models: applications in spectroscopy, Statistical Science 7 (3) (1992) 289–319.
[24] C. Liu, K. Koike, Extending multivariate space–time geostatistics for environmental data analysis, Mathematical Geology 39 (2007) 289–305.
[25] M. Martín-Fernández, C.-F. Westin, C. Alberola-López, 3D Bayesian regularization of diffusion tensor MRI using multivariate Gaussian Markov random fields, in: C. Barillot, D.R. Haynor, P. Hellier (Eds.), Lecture Notes in Computer Science, vol. 3216, Springer, Berlin, 2004, pp. 351–359. (MICCAI).
[26] P. McCullagh, Tensor notation and cumulants of polynomials, Biometrika 71 (3) (1984) 461–476.
[27] P. McCullagh, Tensor Methods in Statistics, Chapman & Hall, London, 1987.
[28] D.G. Northcott, Multilinear Algebra, Cambridge University Press, Cambridge, 1984.
[29] M. Ohlson, M. Rauf Ahmad, D. von Rosen, 2010, More on the Kronecker structured covariance matrix. Communications in Statistics - Theory and Methods. Revision submitted.
[30] F. Pukelsheim, Multilinear estimation of skewness and kurtosis in linear models, Metrika 27 (1980) 103–113.
[31] Y. Sato, K. Sugawa, M. Kawaguchi, The geometrical structure of the parameter space of the two-dimensional normal distribution, Reports on Mathematical Physics 16 (1) (1979) 111–119.
[32] A. Shashua, T. Hazan, 2005, Non-negative tensor factorization with applications to statistics and computer vision, in: Proceedings of the 22nd international conference on machine learning, Bonn, Germany.
[33] L.T. Skovgaard, A Riemannian geometry of the multivariate normal model, Scandinavian Journal of Statistics 11 (4) (1984) 211–223.
[34] C. Soize, Tensor-valued random fields for meso-scale stochastic model of anisotropic elastic microstructure and probabilistic analysis of representative volume element size, Probabilistic Engineering Mechanics 23 (2007) 307–323.
[35] M. Srivastava, T. von Rosen, D. von Rosen, Models with a Kronecker product covariance structure: estimation and testing, Mathematical Methods in Statistics 17 (2008) 357–370.
[36] A. Takemura, Tensor analysis of ANOVA decomposition, Journal of the American Statistical Association 78 (384) (1983) 894–900.
[37] M.J. Wichura, The Coordinate-Free Approach to Linear Models, Cambridge University Press, Cambridge, 2006.