THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date May 22, 2012

WP # 0027MSS-253-2012

Classification of higher-order data with separable covariance and structured multiplicative or additive mean models

Ricardo Leiva
Departamento de Matematica
F.C.E., Universidad Nacional de Cuyo
5500 Mendoza, Argentina
Email: rleiva@fcemail.uncu.edu.ar

Anuradha Roy
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249 USA
Email: aroy@utsa.edu
Phone: +00-210-458-6343, Fax: +00-210-458-6350

Classification of higher-order data with separable covariance and structured multiplicative or additive mean models

Ricardo Leiva
Departamento de Matemática
F.C.E., Universidad Nacional de Cuyo
5500 Mendoza, Argentina
Email: rleiva@fcemail.uncu.edu.ar

Anuradha Roy *
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249 USA
Email: aroy@utsa.edu
Phone: +00-210-458-6343, Fax: +00-210-458-6350

Abstract

Although devised in 1936 by Fisher, discriminant analysis is still rapidly evolving, as the complexity of contemporary data sets grows exponentially. Our classification rules explore these complexities by modeling various correlations in higher-order data. Moreover, our classification rules are suitable to data sets where the number of response variables is comparable or larger than the number of observations. We assume that the higher-order observations have a separable covariance matrix and two different Kronecker product structures on the mean vector. In this article we develop quadratic classification rules among g different populations where each individual has κth order ($\kappa \geq 2$) measurements. We also provide the computational algorithm to compute the sample classification rules.

Keywords: Higher-order data; Separable covariance structure; Separable covariance structure; structured additive mean model; Discriminant function; structured multiplicative mean model; Maximum likelihood estimates.

JEL Classification: C10,C13

1 INTRODUCTION

Higher-order data (HOD) analysis is a mathematical challenge of this 21st Century. Higher-order data is data that can be arranged in hypercubes as opposed to matrices. Such HOD spaces are frequently encountered in areas such as engineering, environmental, medical and biomedical sciences. HOD sets often contain many variables, and in most cases the number of variables

*Correspondence to: Anuradha Roy, Department of Management Science and Statistics, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA

exceeds the sample size. Dimensionality is an issue that can arise almost in every scientific field. Traditional modeling does not work quite as well when the order of the is high. Modeling such HOD, poses many challenges, often involving complex data structures. The properties of high dimensionality are often poorly understood or overlooked in data modeling and analysis. HOD are increasingly enveloping, and bringing new problems and opportunities for statisticians and data analysts. Many questions arise with these HOD. One may ask what kind of structure do these HOD have? What kind of models should we select and how to reduce the dimensions of these data? How do they compare to the traditional model? What kind of theory do we have to develop to handle these data? And, the ultimate question is whether we need a supercomputer to find the solutions?

These questions are the motivating force for the study of HOD. Instead of going for analytic closed form solutions which may not even exist for these HOD, we exploit the blessings of high speed computers or a supercomputer to use these tremendous information to develop computational methods to compute the solutions numerically: we do not need any more explicit or exact solutions in our present computer age.

We develop successful algorithms to avoid the curse of Higher-order, also at the same time taken care of the computational efficiency. One of the challenging problems with the HOD is to deal with the estimation of very large variance-covariance matrix, and the analysis of complex dependence structures between the variables and over the time-space point where the number of samples is much less than the number of total dimensions in the HOD. One can achieve this by imposing some appropriate variance-covariance matrix so that it captures the "natural" structure of the data with much less number of samples. This may be achieved by first selecting the essential variables (Yu and Liu, 2003) and then choosing the appropriate covariance structure over time-space points of the data. Reduction of dimensionality over time-space points of the data is also another option. Pavlenko, Björkstrom and Tillander (2011) studied the classification problem in high dimensional data based on exploring sparsity patterns in the data dependence first and then computing the estimate of inverse of the variance-covariance matrix using constrained maximum likelihood. In their study, rather than restricting themselves to methods that completely ignore potential dependence structure they tried to recover it in the data and then used it to their advantage. They used the popular technique *graphical* Lasso or gLasso (Friedman, Hastie and Tibshirani, 2008) in learning the sparsity patterns. Roy and Leiva (2007) and Leiva and Roy (2009, 2011a,b) have introduced many covariance structures for 3rd order (*variables* $\times$ *sites* $\times$ *time points*) high-dimensional data in the context of classification problem. Leiva and Roy (2009) studied the problem of classification of 3rd order or three-level high-dimensional data by using an "equicorrelated (partitioned) matrix" (Leiva, 2007) on the measurement vector over sites in addition to an AR(1) correlation structure on the repeated measurements over time, whereas in Roy and Leiva (2007) and Leiva and Roy (2011a,b) they used doubly exchangeable covariance structure. Doubly exchangeable covariance structure allows to partition a covariance structure into three unstructured covariance matrices, corresponding to each of the three levels. Kroonenberg (2008) discussed practical issues in applying multi-level or multiway component techniques to multiway data with an emphasis on methods for three-way data. Akdemir and Gupta (2011) have developed classification techniques for high dimensional multiway data. In their paper Akdemir and Gupta presented a technique called slicing for obtaining an approximate nonsingular estimate of the covariance matrix for high dimensional data when sample size is less than the dimension of the vector variate random variable. Dudoit et al. (2002) and Lai et al. (2006) developed classification rules for tumor samples using thousands of gene expression profiles with at most hundreds of samples. By allowing the monitoring of expression levels in cells for thousands of genes simultaneously, microarray experiments may lead to a more complete understanding of the molecular variations among tumors and hence to a finer and more reliable classification. Bhattacharya et al. (2003) proposed a classifier called *Liknon* that simultaneously performs classification and relevant gene identification. Liknon is trained by optimizing a linear discriminant function with a penalty constraint via linear programming. Most recently, Kim and Simon (2011) developed probabilistic classifiers which use the probabilities in conjunction with other information such as treatment options and patient preferences for making complex integrated clinical decisions.

However, there has been much less work on methods of discrimination of datasets that are large at all levels, for example, having data points that exist in many variables over many time-space points representing numerous populations. The goal of the study in this article is to develop discriminant functions which are suitable for these kinds of data. For example, in classifying genes in tumors, one can use any of the techniques mentioned in the previous paragraph in identifying genes and then use these identified genes along with the intensity values in other gene probes simultaneous and over a selected period of time. If we introduce more levels of gene expression in our discriminant analysis the classification performance is bound to improve. Gene expression from a diseased tissue would change over the time, whereas gene expression from a healthy tissue would not change. The variance-covariance matrix should capture this information for the analysis of this data set. Our new method with separable covariance structure and a suitably selected structure on mean vector can handle this kind of data for better classification than just using the intensity levels of genes taken at once. More and more HOHDD with different types of structures will come in future and we must analyze them by developing appropriate methodologies for a particular data set. A range of different models with varying complexity should be developed for a specific type of structured data set and a model that is best in some sense (AIC or BIC) needs to be chosen from a set of candidate models. Many of these HOHDD analysis problems require new or different mathematics as well as different computational algorithms for solutions.

In this article we present new discriminant functions for discriminating g populations with $\kappa-$separable covariance structures along with two different structured (additive and multiplicative) mean vectors for κth order high-dimensional data. Recently, Ohlson, Ahmad and von Rosen (2010) have studied separable covariance structure of several matrices, nonetheless separable covariance structure of two matrices was studied by many authors in the past,

2 PROBLEM FORMULATION

Let $\boldsymbol{x}_{r,s}^{(p)}$ be an $m_\kappa-$variate vector of measurements of the r^{th} replicate (individual) in the p^{th} population on the $\mathbf{s} = (s_1, \ldots, s_{\kappa-1})$ time-space point, $r = 1, \ldots, n^{(p)}$, $p = 1, \ldots, g$, $s_i = 1, \ldots, m_i$, with $i = 1, \ldots, \kappa-1$. The first component s_1 of the subindex vector $\mathbf{s} = (s_1, \ldots, s_{\kappa-1})$ indicates the time coordinate of the time-space point $(s_1, \ldots, s_{\kappa-1})$. Let $\boldsymbol{x}_r^{(p)} = (\boldsymbol{x}_{r,(1,\ldots,1)}^{(p)\prime}, \boldsymbol{x}_{r,(1,\ldots,2)}^{(p)\prime}, \ldots, \boldsymbol{x}_{r,(m_1,\ldots,m_{\kappa-1})}^{(p)\prime})'$ be the $m_\bullet$-variate vector, where $m_\bullet = \prod_{i=1}^{\kappa} m_i$, of all measurements corresponding to the r^{th} individual in the p^{th} population obtained by stacking all m_κ responses of the r^{th} individual in the p^{th} population on the first time-space point $(1, 1, \ldots, 1)$, then stacking all its m_κ responses on the second time-space point $(1, 1, \ldots, 2)$, and so on until its m_κ responses on the last time-space point $(m_1, m_3, \ldots, m_{\kappa-1})$ has been stacked. Let $\boldsymbol{x}_1^{(p)}, \ldots, \boldsymbol{x}_{n^{(p)}}^{(p)}$ be a random sample of size $n^{(p)}$ from the p^{th} population with distribution $N_{m_\bullet}(\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}}, \boldsymbol{\Gamma}_{\boldsymbol{x}^{(p)}})$, where $\boldsymbol{\Gamma}_{\boldsymbol{x}^{(p)}}$ has a $\kappa-$separable covariance structure and the mean vector

$$\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}} = (\boldsymbol{\mu}_{(1,\ldots,1)}^{(p)\prime}, \boldsymbol{\mu}_{(1,\ldots,2)}^{(p)\prime}, \ldots, \boldsymbol{\mu}_{(m_1,\ldots,m_{\kappa-1})}^{(p)\prime})'. \tag{1}$$

In this article we develop a discriminant function with $\kappa-$separable covariance structure in addition to two different mean vector structures: $\kappa-$separable additive and $\kappa-$separable multiplicative structures. In the $\kappa-$additive mean structure we have that for each of the $m_\bullet$ random variables $\boldsymbol{x}_r^{(p)}$, measured on the r^{th} individual in the p^{th} population, the mean of each of $m_\bullet$ random variables $\boldsymbol{x}_r^{(p)}$ can be expressed as

$$\mathrm{E}\left[x_r^{(p)}\right] = \mu_{x^{(p)}} = \sum_{i=1}^{\kappa} \sum_{s_i=1}^{m_i} \mu_{is_i}^{(p)} z_{is_i},$$

where $\boldsymbol{\mu}_i^{(p)} = \left(\mu_{i1}^{(p)}, \ldots, \mu_{im_i}^{(p)}\right)' \in \Re^{m_i}$, for $i = 1, \ldots, \kappa$, with some identifiability constraints, for instance $\mu_{i1}^{(p)} = 0$ or $\mu_{im_i}^{(p)} = 0$ for $i = 1, \ldots, \kappa-1$. With the identifiability constraints $\mu_{im_i}^{(p)} = 0$ for $i = 1, \ldots, \kappa-1$ the model reduces to

$$\mathrm{E}\left[x_r^{(p)}\right] = \sum_{i=1}^{\kappa} \sum_{s_i=1}^{m_i-1} \mu_{is_i}^{(p)} z_{is_i} + \mu_{\kappa m_\kappa}^{(p)} z_{\kappa m_\kappa}$$

For each $i = 1, \ldots, \kappa$, the terms $z_{i1}, \ldots, z_{im_i}$ are used to indicate to which of the m_i categories (cells) of the i^{th} level $\mathrm{E}(x_r^{(p)})$ belongs, that is, only one term at a time of these z's is equal to one in each of these κ levels and all the others are zero. For $\kappa = 1$, the above model reduces to the commonly used additive mean model for one-way multivariate data or univariate repeated measures data.

The additive mean model can also be expressed using "Kronecker sum" of two vectors (see (A4)). With this notation the additive mean model can be expressed as

$$\mathrm{E}\left[\boldsymbol{x}_r^{(p)}\right] = \boldsymbol{\mu}_{\boldsymbol{x}^{(p)}} = \bigoplus_{i=1}^{\kappa} \boldsymbol{\mu}_i^{(p)} := \sum_{i=1}^{\kappa} \left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_i^{(p)} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h} \right),$$

where $\boldsymbol{x}_r^{(p)}$ indicates the $m_\bullet-$dimensional vector of all the measurements taken on the r^{th} individual in the p^{th} population. We use the notation $\bigotimes_{h=1}^{0} \mathbf{1}_{m_h} = 1 = \bigotimes_{h=\kappa+1}^{\kappa} \mathbf{1}_{m_h}$.

In the $\kappa-$separable multiplicative mean vector structure we have that the mean of the $m_\bullet-$vector $\boldsymbol{x}_r^{(p)}$ can be expressed as

$$\mathrm{E}\left[\boldsymbol{x}_r^{(p)}\right] = \boldsymbol{\mu}_{\boldsymbol{x}^{(p)}} = \bigotimes_{i=1}^{\kappa} \boldsymbol{\mu}_i^{(p)},$$

with $\mu_{im_i}^{(p)} = 1$, for $i = 1, \ldots, \kappa - 1$, as identifiability conditions. Similarly to the additive mean model, this model can also be written as

$$\mathrm{E}\left[x_r^{(p)}\right] = \prod_{i=1}^{\kappa} \prod_{s_i=1}^{m_i} \left(\mu_{is_i}^{(p)}\right)^{z_{is_i}},$$

where again $x_r^{(p)}$ indicates a generic random variable of the $m_\bullet$ random variables measured on the r^{th} individual in the p^{th} population, and the z exponents are used to indicate to which of the m_i categories (cells) of the i^{th} level $\mathrm{E}(x_r^{(p)})$ belongs.

In both separable mean vector structures (additive and multiplicative), the mean vectors vary over the time-space points with suitably (additive and multiplicative) constants. The number of unknown free parameters in each of these separable mean vector structures in the p^{th} population is $(m_1+m_2+\cdots+m_\kappa)-(\kappa-1)$. In this article we develop discriminant functions with separable mean vector structures: both additive and multiplicative, along with separable covariance structure. We also develop a discriminant function with the unstructured mean vectors along with separable covariance structure.

Section 3 defines separable covariance structure. The maximum likelihood estimates (MLEs) of the mean vector and separable variance covariance matrix in a single population case are obtained in Section 4. The proposed classification rules with structured and unstructured mean vectors are presented in Section 5. Finally, Section 6 concludes with several comments. Technical proofs of the MLEs of all unknown parameters and derivatives of Kronecker sum are presented in four appendices.

3 BASIC RESULTS FOR SEPARABLE COVARIANCE STRUCTURE ANALYSIS

Definition 1. *Let* $\boldsymbol{x}_r = (\boldsymbol{x}'_{r,(1,\ldots,1)}, \ldots, \boldsymbol{x}'_{r,(m_1,\ldots,m_{\kappa-1})})'$ *be an* $m_\bullet-$*variate partitioned real-valued random vector where* $\boldsymbol{x}_{r,\mathbf{s}} = \boldsymbol{x}_{r,(s_1,\ldots,s_{\kappa-1})} = (x_{r,\mathbf{s}1}, \ldots, x_{r,\mathbf{s}m_\kappa})'$ *for* $s_i = 1, \ldots, m_i$*, with* $i =$

$1, \ldots, \kappa-1$. *Let* $\boldsymbol{\mu}_{\boldsymbol{x}} \in \Re^{m_\bullet}$ *be the mean vector, and* $\mathbf{\Gamma}_{\boldsymbol{x}}$ *be the* $(m_\bullet \times m_\bullet)-$*dimensional partitioned covariance matrix* $\mathbf{\Gamma}_{\boldsymbol{x}} = Cov[\boldsymbol{x}]$. *The* $m_\kappa-$*variate vectors* $\boldsymbol{x}_{r,(1,\ldots,1)}, \ldots, \boldsymbol{x}_{r,(m_2,\ldots,m_{\kappa-1})}$ *are said to have a* $\kappa-$*separable covariance structure if* $\mathbf{\Gamma}_{\boldsymbol{x}}$ *is given by*

$$\mathbf{\Gamma}_{\boldsymbol{x}} = \bigotimes_{i=1}^{\kappa} \mathbf{V}_i,$$

where $\mathbf{V}_i,\ i = 1, \ldots, \kappa$ *are* $m_i \times m_i-$*dimensional* κ *unstructured variance-covariance (positive definite and symmetric) matrix at* i^{th} *level for any other fixed* j^{th} $(j \neq i)$ *level.*

Note that if $\mathbf{\Gamma}_{\boldsymbol{x}} = \boldsymbol{V}_1 \otimes \boldsymbol{V}_2 \otimes \cdots \otimes \boldsymbol{V}_\kappa$, then $\mathbf{\Gamma}_{\boldsymbol{x}} = \left(\frac{1}{\alpha_2 \cdots \alpha_\kappa} \boldsymbol{V}_1\right) \otimes (\alpha_2 \boldsymbol{V}_2) \otimes \cdots \otimes (\alpha_\kappa \boldsymbol{V}_\kappa)$, for any non zero real numbers $\alpha_2, \ldots, \alpha_\kappa$, and therefore parameters in each variance-covariance matrix $\boldsymbol{V}_i$ are not jointly identifiable unless we impose an appropriate condition. There are several possible ways to handle this. It is always possible to obtain an estimate of either of $\mathbf{V}_i,\ i = 1, \ldots, \kappa$ by taking one of the diagonal elements of either of the component matrices $\boldsymbol{V}_i$ for $i = 1, 2, \ldots, \kappa$ to be one. Normally, the first or the last diagonal element of $\boldsymbol{V}_i$ is taken to be one. It must be noted that for classification purposes the variance-covariance matrices $\mathbf{V}_i$ for $i = 1, \ldots, \kappa$ do not need to be unique, but $\boldsymbol{V}_1 \otimes \boldsymbol{V}_2 \otimes \cdots \otimes \boldsymbol{V}_\kappa$ does. Therefore, the total number of parameters in $\mathbf{\Gamma}_{\boldsymbol{x}}$ is $\sum_{r=1}^{\kappa} m_i (m_i + 1)/2$, but the total number of free parameters in $\mathbf{\Gamma}_{\boldsymbol{x}}$ is $\sum_{r=1}^{\kappa} (m_i (m_i + 1)/2) - (\kappa - 1)$. This covariance structure can be justified as follows. Let $\mathbf{V}_i$ represent the $m_i \times m_i-$dimensional unstructured variance-covariance (positive definite) matrix on all m_i repeated measurements at the i^{th} level. It is assumed that it does not depend on the j^{th} level, $j \neq i$. The apparent advantage is that the number of parameters to be estimated is greatly reduced, and thus the statistical analysis can be accomplished in small sample set-up.

Result 1: The inverse of $\mathbf{\Gamma}_{\boldsymbol{x}}$ is given by

$$\mathbf{\Gamma}_{\boldsymbol{x}}^{-1} = \bigotimes_{i=1}^{\kappa} \mathbf{V}_i^{-1}, \tag{2}$$

and the determinant of $\mathbf{\Gamma}_{\boldsymbol{x}}$ is given by

$$|\mathbf{\Gamma}_{\boldsymbol{x}}| = \prod_{i=1}^{\kappa} |\mathbf{V}_i|^{\frac{m_\bullet}{m_i}}. \tag{3}$$

4 MAXIMUM LIKELIHOOD ESTIMATES OF THE MEAN VECTOR AND THE COVARIANCE MATRIX IN A SINGLE POPULATION

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be an $m_\bullet-$variate random sample of size n from a population with distribution $N_{m_\bullet}(\boldsymbol{\mu}_{\boldsymbol{x}}, \mathbf{\Gamma}_{\boldsymbol{x}})$. We assume that the covariance matrix $\mathbf{\Gamma}_{\boldsymbol{x}}$ has the $\kappa-$separable structure as defined in Section 3. We consider structured as well as the unstructured mean vectors. The structured additive mean vector case is discussed in Section 4.1, the structured multiplicative mean vector case is discussed in Section 4.2, and the unstructured mean vector case in Section 4.3.

4.1 STRUCTURED ADDITIVE MEAN VECTOR

As in Definition 1 we partition the $m_\bullet-$variate vector $\boldsymbol{x}_r$ as $\boldsymbol{x}_r = (\boldsymbol{x}'_{r,(1,\ldots,1)}, \ldots, \boldsymbol{x}'_{r,(m_1,\ldots,m_{\kappa-1})})'$, for $r = 1, \ldots, n$, where $\boldsymbol{x}_{r,\mathbf{s}} = \boldsymbol{x}_{r,(s_1,\ldots,s_{\kappa-1})} = (x_{r,\mathbf{s}1}, \ldots, x_{r,\mathbf{s}m_\kappa})' \in \Re^{m_\kappa}$, for $s_i = 1, \ldots, m_i$ with $i = 1, \ldots, \kappa-1$. In this case we assume that for $s_\kappa = 1, \ldots, m_\kappa$, the s_κ^{th} component of the mean vector $\mathrm{E}[\boldsymbol{x}_{r,\mathbf{s}}] = \boldsymbol{\mu}_{\mathbf{s}} = (\mu_{\mathbf{s},s_\kappa})_{s_\kappa=1}^{m_\kappa}$ is the sum $\mu_{\mathbf{s},s_\kappa} = \sum_{h=1}^{\kappa} \mu_{hs_h}$, where $\mu_{hs_h} \in \Re$, with $\mu_{im_i} = 0$ for $i = 1, \ldots, \kappa-1$. That is, the s_κ^{th} component $\mu_{\mathbf{s},s_\kappa}$ of the mean $\boldsymbol{\mu}_{\mathbf{s}}$ of $\boldsymbol{x}_{r,\mathbf{s}}$ is discomposed into a sum of κ summands: the s_κ^{th} component of a (base) $m_\kappa-$vector $\boldsymbol{\mu}_\kappa \in \Re^{m_\kappa}$, plus μ_{is_i} (an effect due to the i^{th} level) for each $i = 1, \ldots, \kappa-1$. Therefore, $\boldsymbol{\mu}_{\boldsymbol{x}} = \bigoplus_{i=1}^{\kappa} \boldsymbol{\mu}_i$ where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})' \in \Re^{m_i}$, for $i = 1, \ldots, \kappa$, with $\mu_{im_i} = 0$ for $i = 1, \ldots, \kappa-1$.

4.1.1 MAXIMUM LIKELIHOOD SYSTEM OF EQUATIONS

The following theorem yields a system of equations to obtain the MLEs of the structured additive mean vector $\boldsymbol{\mu}_{\boldsymbol{x}}$ and the separable covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{x}}$.

Theorem 1. *Under the above assumptions and using the notation* $\boldsymbol{\lambda}_i = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{i-1}, \boldsymbol{\mu}_{i+1}, \ldots, \boldsymbol{\mu}_\kappa)$, *the maximum likelihood estimates of* $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_\kappa$ *and* $\boldsymbol{\Gamma}_{\boldsymbol{x}}$ *are given by*

$$\widehat{\boldsymbol{\mu}}_i = \mathbf{D}_{\boldsymbol{\lambda}_\mathbf{i}} \left(\overline{\boldsymbol{x}} - \sum_{i \neq j=1}^{\kappa} \left(\bigotimes_{h=1}^{j-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_j \otimes \bigotimes_{h=j+1}^{\kappa} \mathbf{1}_{m_h} \right) \right), \quad \textit{for } i = 1, \ldots, \kappa, \tag{4}$$

and

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}} = \bigotimes_{i=1}^{\kappa} \widehat{\mathbf{V}}_i, \tag{5}$$

where

$$\begin{aligned} \mathbf{D}_{\boldsymbol{\lambda}_\mathbf{i}} \quad = \quad & \left[\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h} \right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1} \left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h} \right) \right]^{-1} \\ & \cdot \left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h} \right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}, \end{aligned} \tag{6}$$

and

$$\widehat{\mathbf{V}}_i = \frac{m_i}{n m_\bullet} \mathbf{C}_i,$$

where $\mathbf{C}_i$ *is given in (A12) for* $i = 1, \ldots, \kappa$.

The proof of this theorem which is simple but tedious, is given in Appendix B. We see that the MLEs of $(\boldsymbol{\mu}_i, \mathbf{V}_i)$, for $i = 1, \ldots, \kappa$ have implicit equations, and therefore are not tractable analytically. The computation of these MLEs can be carried out by solving the above implicit equations by the following fixed point iteration algorithm.

4.1.2 ITERATION ALGORITHM

We simultaneously calculate the maximum likelihood estimates (MLEs) of a total of $\big(\sum_{r=1}^{\kappa} m_i - (\kappa-1)\big) + \sum_{r=1}^{\kappa} m_i\,(m_i+1)\,/2$ unknown parameters in the separable additive mean vector and the separable variance-covariance matrix. The solutions satisfy the fully implicit and coupled equations with $\boldsymbol{\mu}_i$ and $\mathbf{V}_i$ for $i = 1, 2, \ldots, \kappa$.

Step 1: Calculate the global sample mean $\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{r=1}^{n} \boldsymbol{x}_r$ as

$$\overline{\boldsymbol{x}} = \left(\overline{\boldsymbol{x}}'_{(1,\ldots,1,1)}, \ldots, \overline{\boldsymbol{x}}'_{(1,\ldots,1,m_{\kappa-1})}, \ldots, \overline{\boldsymbol{x}}'_{(1,m_2,\ldots,m_{\kappa-1})}, \ldots, \overline{\boldsymbol{x}}'_{(m_1,m_2,\ldots,m_{\kappa-1})}\right)'.$$

It is a partitioned $m_\bullet-$dimensional vector, where $\overline{\boldsymbol{x}}_{\mathbf{s}}$ denotes the sample mean vector corresponding to time-space point $\mathbf{s} = (s_1, \ldots, s_{\kappa-1})$, and is given by $\overline{\boldsymbol{x}}_{\mathbf{s}} = \frac{1}{n}\sum_{r=1}^{n} \boldsymbol{x}_{r,\mathbf{s}}$. The initial value $\widehat{\boldsymbol{\mu}}_1^0$ of $\widehat{\boldsymbol{\mu}}_1$ is taken as $\widehat{\boldsymbol{\mu}}_1^0 = \overline{\boldsymbol{x}}_{(1,\ldots,1,1)}$. Compute the initial values of $\boldsymbol{\mu}_x^0 = \bigoplus_{i=1}^{\kappa} \boldsymbol{\mu}_i^0$ by assuming the initial values of $\boldsymbol{\mu}_i^0 = \mathbf{1}_{m_i}$ for $i = 2, \ldots, \kappa$.

Step 2: Compute $\mathbf{A}$ from (B3).

Step 3: Compute the initial estimates $\mathbf{V}_i^0$ of $\mathbf{V}_i$ for $i = 1, \ldots, \kappa$ from the data.

Step 4: Compute $\mathbf{W}_s$ from (A8).

Step 5: Compute $\mathbf{A}_s^*$ from (A9) and then compute $\mathbf{C}_s$ from (A12).

Step 6: Compute the revised estimate of $\mathbf{V}_i$ for $i = 1, \ldots, \kappa$ from (B5).

Step 7: Compute the estimate of $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}$ from (2).

Step 8: Compute $\mathbf{D}_{\boldsymbol{\lambda}_\mathbf{i}}$ for $i = 1, \ldots, \kappa$ from (6) using $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}$ in Step 7.

Step 9: Compute the estimate $\widehat{\boldsymbol{\mu}}_i$ from (4).

Step 10: Repeat Steps 2-9 until convergence is attained. This is ensured by verifying if the maximum of the absolute difference among the L_1 distance between two successive values of $\widehat{\boldsymbol{\mu}}_i,\ i = 1, \ldots, \kappa$, and the absolute difference among the two successive values of trace of $\widehat{\mathbf{V}}_i,\ i = 1, \ldots, \kappa$, is less than a pre-determined number ϵ.

4.2 STRUCTURED MULTIPLICATIVE MEAN VECTOR

Using the same notation as in Section 4.1, in this case we assume that the s_κ^{th} component of the mean vector $\mathrm{E}\left[\boldsymbol{x}_{r,\mathbf{s}}\right] = \boldsymbol{\mu}_{\mathbf{s}} = \left(\mu_{\mathbf{s},s_\kappa}\right)_{s_\kappa=1}^{m_\kappa}$ is factorized into κ factors, that is, $\mu_{\mathbf{s},s_\kappa} = \prod_{h=1}^{\kappa} \mu_{hs_h}$, where $\mu_{hs_h} \in \Re$, with $\mu_{im_i} = 1$ for $i = 1, \ldots, \kappa-1$. That is, the s_κ^{th} component $\mu_{\mathbf{s},s_\kappa}$ of the mean $\boldsymbol{\mu}_{\mathbf{s}}$ of $\boldsymbol{x}_{r,\mathbf{s}}$ is the product of the s_h^{th} effect of the h^{th} level for each $h = 1, \ldots, \kappa-1$. Therefore, $\boldsymbol{\mu}_{\boldsymbol{x}} = \bigotimes_{i=1}^{\kappa} \boldsymbol{\mu}_i$ where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})' \in \Re^{m_i}$, for $i = 1, \ldots, \kappa$, with $\mu_{im_i} = 1$ for $i = 1, \ldots, \kappa-1$.

4.2.1 MAXIMUM LIKELIHOOD SYSTEM OF EQUATIONS

The following theorem gives a system of equations whose solutions (iteratively obtained) are the MLEs of the structured multiplicative mean vector $\boldsymbol{\mu}_{\boldsymbol{x}}$ and of the separable covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{x}}$.

Theorem 2. *Under the above assumptions and indicating with* $\boldsymbol{\lambda}_i = \left(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{i-1}, \boldsymbol{\mu}_{i+1}, \ldots, \boldsymbol{\mu}_\kappa\right)$, *the maximum likelihood estimates of* $\boldsymbol{\mu}_i : i = 1, \ldots, \kappa$ *and* $\boldsymbol{\Gamma_x}$ *are given by*

$$\widehat{\boldsymbol{\mu}}_i = \mathbf{E}_{\boldsymbol{\lambda}_\mathbf{i}} \cdot \overline{\boldsymbol{x}},$$

for $i = 1, \ldots, \kappa,$ *and*

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}} = \bigotimes_{i=1}^{\kappa} \widehat{\mathbf{V}}_i,$$

where

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\lambda}_\mathbf{i}} \quad = \quad & \left[\left(\bigotimes_{h=1}^{i-1} \boldsymbol{\mu}_h \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \boldsymbol{\mu}_h\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1} \left(\bigotimes_{h=1}^{i-1} \boldsymbol{\mu}_h \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \boldsymbol{\mu}_h\right)\right]^{-1} \\ & \cdot \left(\bigotimes_{h=1}^{i-1} \boldsymbol{\mu}_h \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \boldsymbol{\mu}_h\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1},\end{aligned}$$

and

$$\widehat{\mathbf{V}}_i = \frac{m_s}{n m_\bullet} \mathbf{C}_i,$$

where $\mathbf{C}_i$ *is given in (C2) for* $i = 1, \ldots, \kappa.$

The proof of this theorem is similar to the one of theorem 1. A sketch of this proof is given in Appendix C. Note that MLEs of $(\boldsymbol{\mu}_i, \mathbf{V}_i)$, for $i = 1, \ldots, \kappa$, have not explicit closed form, and the estimates are obtained by a similar algorithm similar to Algorithm 4.1.2.

4.3 UNSTRUCTURED MEAN VECTOR

In this case we don't assume any structure for the mean vector $\boldsymbol{\mu_x}$, that is, $\boldsymbol{\mu_x}$ could be any vector of $\Re^{m_\bullet}$.

4.3.1 MAXIMUM LIKELIHOOD SYSTEM OF EQUATIONS

Using the same notation as in the above sections, we have the following theorem:

Theorem 3. *Under the above assumptions, the maximum likelihood estimates of* $\boldsymbol{\mu_x}$ *and* $\boldsymbol{\Gamma_x}$ *are given by*

$$\widehat{\boldsymbol{\mu}_{\boldsymbol{x}}} = \overline{\boldsymbol{x}},$$

and

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}} = \bigotimes_{i=1}^{\kappa} \widehat{\mathbf{V}}_i,$$

where

$$\widehat{\mathbf{V}}_i = \frac{m_s}{n m_\bullet} \mathbf{C}_i,$$

where $\mathbf{C}_i$ *is given in (A12) for* $i = 1, \ldots, \kappa.$

A sketch of the proof is given in Appendix D. Note that MLEs of $(\boldsymbol{\mu}_i, \mathbf{V}_i)$, for $i = 1, \ldots, \kappa$, have explicit closed form, and the estimates are obtained easily.

5 DISCRIMINATION WITH SEPARABLE COVARIANCE MATRIX

In this section we derive the Bayesian decision rule for g populations with structured mean vectors. Using the same notations as in the introduction, we assume that the component vectors of the partitioned $m_\bullet$-variate vector $\boldsymbol{x}_r^{(p)} = (\boldsymbol{x}_{r,(1,\ldots,1)}^{(p)\prime}, \boldsymbol{x}_{r,(1,\ldots,2)}^{(p)\prime}, \ldots, \boldsymbol{x}_{r,(m_1,\ldots,m_{\kappa-1})}^{(p)\prime})'$ have jointly separable covariance matrix with factor matrices $\mathbf{V}_i^{(p)}$, for $i = 1, \ldots, \kappa$, with mean vector $\mathrm{E}[\boldsymbol{x}_r^{(p)}] = \boldsymbol{\mu}_{\boldsymbol{x}^{(p)}}$, for $p = 1, \ldots, g$. Three different cases of mean vector are considered:

1. Separable additive model: $\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}} = \bigoplus_{i=1}^{\kappa} \boldsymbol{\mu}_i^{(p)}$, where $\boldsymbol{\mu}_i^{(p)} = \left(\mu_{i1}^{(p)}, \ldots, \mu_{im_i}^{(p)}\right)' \in \Re^{m_i}$, for $i = 1, \ldots, \kappa$, with $\mu_{im_i}^{(p)} = 0$ for $i = 1, \ldots, \kappa - 1$.
2. Separable multiplicative model: $\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}} = \bigotimes_{i=1}^{\kappa} \boldsymbol{\mu}_i^{(p)}$, with $\mu_{im_i}^{(p)} = 1$ for $i = 1, \ldots, \kappa - 1$.
3. Unstructured model: $\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}} \in \Re^{m_\bullet}$.

Let $\boldsymbol{x}_1^{(p)}, \ldots, \boldsymbol{x}_{n^{(p)}}^{(p)}$ be a random sample of size $n^{(p)}$ from the p^{th} population with distribution $N_{m_\bullet}\left(\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}}, \boldsymbol{\Gamma}_{\boldsymbol{x}}^{(p)}\right)$, for $p = 1, \ldots, g$. These g random training samples are independent among each other.

Now we consider the problem of assigning a new individual with $m_\bullet-$variate partitioned measurement vector $\boldsymbol{x}_0$ to one of the g groups in a Bayesian framework. The previous set-up leads to a quadratic discriminant function as follows:
Under the assumptions of equal prior probabilities and equal costs of misclassification, the sample classification rule is given by
Allocate an individual with response $\boldsymbol{x}_0$ to population i if

$$q^{(i)}\ (\boldsymbol{x}_0) = \text{largest of } \left\{q^{(p)}\ (\boldsymbol{x}_0) : p = 1, \ldots, g\right\}, \quad \text{for} \quad i = 1, \ldots, g, \tag{7}$$

where the quadratic score $q^{(p)}$ is defined by

$$q^{(p)}\ (\boldsymbol{x}_0) = -\frac{1}{2} \ln \left|\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}^{(p)}\right| - \frac{1}{2} \left(\boldsymbol{x}_0 - \widehat{\boldsymbol{\mu}}_{\boldsymbol{x}^{(p)}}\right)' \cdot \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}^{(p)-1} \cdot \left(\boldsymbol{x}_0 - \widehat{\boldsymbol{\mu}}_{\boldsymbol{x}^{(p)}}\right),$$

and $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}^{(p)-1}$ and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}^{(p)}}$, for $p = 1, \ldots, i$, are the MLEs of $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{(p)-1}$ and $\boldsymbol{\mu}_{\boldsymbol{x}^{(p)}}$ corresponding to the considered mean model. These estimates are obtained by using a similar fixed point iteration algorithm as described in the respective section. The MLE $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}^{(p)}$ of $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{(p)}$ is given in (5). This quadratic rule has been extensively studied by many authors. See McLachlan (1992).

6 CONCLUDING REMARKS

While discriminant analysis has a long history and a large number of classification rules have already been developed, significant challenges still remain. In this article we establish discriminant functions with separable covariance structure along with different structures on mean vector that are suitable for higher-order high-dimensional data, e.g., for better understanding of progressive diseases. Models for high-dimensional data analysis is a challenging task and discriminating HO-HDD sets is a contemporary challenge. We should always take the design consideration of the data structure for selecting an appropriate classification rule. Since the distribution theory associated with the quadratic rule is particularly difficult, several authors (Park and Kshirsagar, 1994; Paranjpe and Gore, 1994) proposed linear solutions to this problem. This article is restricted to quadratic classification rule, however one can use modified linear classification rule (See Roy and Leiva, 2007) by using the average of the variance-covariance matrices.

A KRONECKER PRODUCT DERIVATIVES AND KRONECKER SUM DERIVATIVES

Definition 2. *Kronecker product: Let* $\underset{m_h\times 1}{\boldsymbol{x}_h} = (x_{h1}, \ldots, x_{hm_h})'$ *be an* $(m_h \times 1)-$*dimensional vector of real variables for* $h = 1, \ldots, n$*, and let* $\boldsymbol{w}$ *be the Kronecker product of them. That is* $\boldsymbol{w}$ *is the* $(\sum_{h=1}^n m_h) \times 1$ *dimensional vector and is given by*

$$\underset{\left(\sum_{h=1}^{n} m_h\right)\times 1}{\boldsymbol{w}} = \bigotimes_{h=1}^{n} \boldsymbol{x}_h = \left(\bigotimes_{h=1}^{n-1} \boldsymbol{x}_h\right) \otimes \boldsymbol{x}_n = \boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_n$$

$$= \left[\left(\bigotimes_{h=1}^{j-1} \boldsymbol{x}_h\right) \otimes \mathbf{I}_{m_j} \otimes \left(\bigotimes_{h=j+1}^{n} \boldsymbol{x}_h\right)\right] \boldsymbol{x}_j \tag{A1}$$

Now, the following formulas can be easily proved:

1. The quantity $\dfrac{\partial \boldsymbol{w}}{\partial \boldsymbol{x}_j}$ can be calculated as follows:

$$\frac{\partial \bigotimes_{h=1}^{n} \boldsymbol{x}_h}{\partial \boldsymbol{x}_j} = \left(\bigotimes_{h=1}^{j-1} \boldsymbol{x}'_h\right) \otimes \mathbf{I}_{m_j} \otimes \left(\bigotimes_{h=j+1}^{n} \boldsymbol{x}'_h\right) \quad \text{for } j = 1, \ldots, n, \tag{A2}$$

   where it is assumed that

$$\bigotimes_{h=k}^{i} \boldsymbol{x}_h = 1 \quad \text{if} \quad k > i.$$

2. Let $\mathbf{D}$ be a $\left(\sum_{h=1}^n m_h \times \sum_{h=1}^n m_h\right)-$dimensional symmetric matrix. Then $\frac{\partial(\mathbf{D}\cdot\boldsymbol{w})}{\partial \boldsymbol{x}_j}$ and $\frac{\partial(\boldsymbol{w}'\cdot\mathbf{D}\cdot\boldsymbol{w})}{\partial \boldsymbol{x}_j}$ are given by

$$\frac{\partial(\mathbf{D}\cdot \bigotimes_{h=1}^{n} \boldsymbol{x}_h)}{\partial \boldsymbol{x}_j} = \left[\left(\bigotimes_{h=1}^{j-1} \boldsymbol{x}'_h\right) \otimes \mathbf{I}_{m_j} \otimes \left(\bigotimes_{h=j+1}^{n} \boldsymbol{x}'_h\right)\right] \cdot \mathbf{D}', \quad \text{for} j = 1, \ldots, n.$$

and

$$\frac{\partial\left(\bigotimes_{h=1}^{n}\boldsymbol{x}_h\right)'\cdot\mathbf{D}\cdot\left(\bigotimes_{h=1}^{n}\boldsymbol{x}_h\right)}{\partial\boldsymbol{x}_j}=2\left[\bigotimes_{h=1}^{j-1}\boldsymbol{x}_h'\otimes\mathbf{I}_{m_j}\otimes\left(\bigotimes_{h=j+1}^{n}\boldsymbol{x}_h'\right)\right]\cdot\mathbf{D}\cdot\left(\bigotimes_{h=1}^{n}\boldsymbol{x}_h\right),$$

for $j=1,\ldots,n$. When $\boldsymbol{y}=\boldsymbol{a}-\boldsymbol{w}=\boldsymbol{a}-\bigotimes_{h=1}^{n}\boldsymbol{x}_h$, where $\boldsymbol{a}$ is a constant vector, and $\boldsymbol{q}=\boldsymbol{y}'\cdot\mathbf{D}\cdot\boldsymbol{y}=(\boldsymbol{a}-\boldsymbol{w})'\cdot\mathbf{D}\cdot(\boldsymbol{a}-\boldsymbol{w})$, then

$$\begin{aligned}\frac{\partial\boldsymbol{q}}{\partial\boldsymbol{x}_j} &= \frac{\partial\left(\boldsymbol{y}'\cdot\mathbf{D}\cdot\boldsymbol{y}\right)}{\partial\boldsymbol{x}_j}=\frac{\partial\left[\left(\boldsymbol{a}-\bigotimes_{h=1}^{n}\boldsymbol{x}_h\right)'\cdot\mathbf{D}\cdot\left(\boldsymbol{a}-\bigotimes_{h=1}^{n}\boldsymbol{x}_h\right)\right]}{\partial\boldsymbol{x}_j}\\ &= \left(\frac{\partial\bigotimes_{h=1}^{n}\boldsymbol{x}_h}{\partial\boldsymbol{x}_j}\right)\cdot\left(\frac{\partial\left[\boldsymbol{y}'\cdot\mathbf{D}\cdot\boldsymbol{y}\right]}{\partial\boldsymbol{y}}\right)\\ &= -2\left[\left(\bigotimes_{h=1}^{j-1}\boldsymbol{x}_h'\right)\otimes\mathbf{I}_{m_j}\otimes\left(\bigotimes_{h=j+1}^{n}\boldsymbol{x}_h'\right)\right]\cdot\mathbf{D}\cdot\left(\boldsymbol{a}-\bigotimes_{h=1}^{n}\boldsymbol{x}_h\right).\end{aligned}\tag{A3}$$

Definition 3. *Kronecker sum: Let* $\underset{m_h\times 1}{\boldsymbol{x}_h}=(x_{h1},\ldots,x_{hm_h})'$ *be an* $m_h\times 1$ *vector of real variables, for* $h=1,\ldots,n$, *then the Kronecker sum of these vectors is the* $\left(\sum_{h=1}^{n}m_h\right)\times 1-$ *vector* $\boldsymbol{v}$ *given by*

$$\begin{aligned}\underset{\left(\sum_{h=1}^{n}m_h\right)\times 1}{\boldsymbol{v}} &= \bigoplus_{k=1}^{n}\boldsymbol{x}_k=\boldsymbol{x}_1\oplus\boldsymbol{x}_2\oplus\cdots\oplus\boldsymbol{x}_n\\ &= \sum_{k=1}^{n}\left(\bigotimes_{h=1}^{k-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_k\otimes\bigotimes_{h=k+1}^{n}\mathbf{1}_{m_h}\right)\\ &= \sum_{k=1}^{n}\left(\bigotimes_{h=1}^{k-1}\mathbf{1}_{m_h}\otimes\mathbf{I}_{m_k}\otimes\bigotimes_{h=k+1}^{n}\mathbf{1}_{m_h}\right)\boldsymbol{x}_k\end{aligned}\tag{A4}$$

Using the previous Kroncker product derivatives, the following formulas can be easily proved:

1. The quantity $\frac{\partial\boldsymbol{v}}{\partial\boldsymbol{x}_j}$ can be calculated as follows:

$$\frac{\partial\bigoplus_{k=1}^{n}\boldsymbol{x}_k}{\partial\boldsymbol{x}_j}=\left(\bigotimes_{k=1}^{j-1}\mathbf{1}_{m_k}'\right)\otimes\mathbf{I}_{m_j}\otimes\left(\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_k}'\right)\quad\text{for } j=1,\ldots,n,\tag{A5}$$

   where again it is assumed that

$$\bigotimes_{h=k}^{i}\boldsymbol{x}_h=1\quad\text{if}\quad k>i.$$

2. Let $\mathbf{D}$ is a $\left(\sum_{h=1}^{n}m_h\times\sum_{h=1}^{n}m_h\right)-$dimensional symmetric matrix. Then $\frac{\partial(\mathbf{D}\cdot\boldsymbol{v})}{\partial\boldsymbol{x}_j}$ and $\frac{\partial(\boldsymbol{v}'\cdot\mathbf{D}\cdot\boldsymbol{v})}{\partial\boldsymbol{x}_j}$ are given by

$$\frac{\partial\mathbf{D}\cdot\bigoplus_{k=1}^{n}\boldsymbol{x}_k}{\partial\boldsymbol{x}_j}=\left[\left(\bigotimes_{k=1}^{j-1}\mathbf{1}_{m_k}'\right)\otimes\mathbf{I}_{m_j}\otimes\left(\bigotimes_{k=j+1}^{n}\mathbf{1}_{m_k}'\right)\right]\cdot\mathbf{D}',\quad\text{for} j=1,\ldots,n,$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{x}_j}\left[\left(\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right)'\cdot\mathbf{D}\cdot\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right]\\
=\;&\frac{\partial}{\partial \boldsymbol{x}_j}\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_j\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)'\cdot\mathbf{D}\cdot\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_j\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)\\
&+2\frac{\partial}{\partial \boldsymbol{x}_j}\sum_{j\neq k=1}^{n}\left(\bigotimes_{h=1}^{k-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_k\otimes\bigotimes_{h=k+1}^{n}\mathbf{1}_{m_h}\right)'\cdot\mathbf{D}\cdot\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_j\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)\\
=\;&2\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\mathbf{I}_{m_j}\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)'\cdot\mathbf{D}\cdot\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_j\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)\\
&+2\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\mathbf{I}_{m_j}\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)'\cdot\mathbf{D}\cdot\sum_{j\neq k=1}^{n}\left(\bigotimes_{h=1}^{k-1}\mathbf{1}_{m_h}\otimes\boldsymbol{x}_k\otimes\bigotimes_{h=k+1}^{n}\mathbf{1}_{m_h}\right)\\
=\;&2\left[\left(\bigotimes_{h=1}^{j-1}\mathbf{1}_{m_h}\otimes\mathbf{I}_{m_j}\otimes\bigotimes_{h=j+1}^{n}\mathbf{1}_{m_h}\right)'\right]\cdot\mathbf{D}\cdot\bigoplus_{k=1}^{n}\boldsymbol{x}_k,
\end{aligned}
$$

for $j=1,\ldots,n$. Therefore, if $\boldsymbol{y}=\boldsymbol{a}-\boldsymbol{v}=\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k$, where $\boldsymbol{a}$ is a constant vector, $\mathbf{D}$ is a $(\sum_{h=1}^{n}m_h\times\sum_{h=1}^{n}m_h)-$dimensional symmetric matrix, and $\boldsymbol{q}=\boldsymbol{y}'\cdot\mathbf{D}\cdot\boldsymbol{y}=(\boldsymbol{a}-\boldsymbol{v})'\cdot\mathbf{D}\cdot(\boldsymbol{a}-\boldsymbol{v})$ then

$$
\begin{aligned}
\frac{\partial\boldsymbol{q}}{\partial\boldsymbol{x}_j} &= \frac{\partial(\boldsymbol{y}'\cdot\mathbf{D}\cdot\boldsymbol{y})}{\partial\boldsymbol{x}_j}=\frac{\partial\left[\left(\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right)'\cdot\mathbf{D}\cdot\left(\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right)\right]}{\partial\boldsymbol{x}_j}\\
&= \frac{\partial\left[\left(\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right)'\cdot\mathbf{D}\cdot\left(\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right)\right]}{\partial\boldsymbol{x}_j}\\
&= \left(\frac{\partial\left(\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right)}{\partial\boldsymbol{x}_j}\right)\cdot\left(\frac{\partial\left[\boldsymbol{y}'\cdot\mathbf{D}\cdot\boldsymbol{y}\right]}{\partial\boldsymbol{y}}\right)\\
&= -2\left[\left(\bigotimes_{k=1}^{j-1}\mathbf{1}'_{m_k}\right)\otimes\mathbf{I}_{m_j}\otimes\left(\bigotimes_{h=j+1}^{n}\mathbf{1}'_{m_k}\right)\right]\cdot\mathbf{D}\cdot\left(\boldsymbol{a}-\bigoplus_{k=1}^{n}\boldsymbol{x}_k\right) \qquad \text{(A6)}
\end{aligned}
$$

If $\underset{m_k\times m_k}{\mathbf{V}_k}=(v_{k,ij})$, for $k=1,\ldots,\kappa$, and $\underset{m_\bullet\times m_\bullet}{\mathbf{A}}$ are symmetric matrices, where $m_\bullet=\prod_{k=1}^{\kappa}m_k$. Let $n_s^-=\prod_{k=1}^{s}m_k$ and $n_s^+=\prod_{k=s}^{\kappa}m_k=\frac{m_\bullet}{n_{s-1}^-}$. The matrix $\mathbf{A}$ will be considered (when necessary) as an appropriate partitioned matrix in square submatrices, that is, for each $k=1,\ldots,\kappa$ we have $\underset{m_\bullet\times m_\bullet}{\mathbf{A}}=\left(\underset{\frac{m_\bullet}{m_k}\times\frac{m_\bullet}{m_k}}{\mathbf{A}_{s,ij}}\right)_{i,j=1,\ldots,m_k}=\left(\underset{n_{k-1}^-n_{k+1}^+\times n_{k-1}^-n_{k+1}^+}{\mathbf{A}_{s,ij}}\right)_{i,j=1,\ldots,m_k}$. With this notation and since

$$\begin{aligned}\bigotimes_{k=1}^{\kappa}\mathbf{V}_k &= \mathbf{K}_{n_{s-1}^-,n_s^+}\left(\bigotimes_{k=s}^{\kappa}\mathbf{V}_k\otimes\bigotimes_{k=1}^{s-1}\mathbf{V}_k\right)\mathbf{K}_{n_s^+,n_{s-1}^-}\\ &= \mathbf{K}_{n_{s-1}^-,n_s^+}\left[\mathbf{V}_s\otimes\left(\bigotimes_{k=s+1}^{\kappa}\mathbf{V}_k\otimes\bigotimes_{k=1}^{s-1}\mathbf{V}_k\right)\right]\mathbf{K}_{n_s^+,n_{s-1}^-},\end{aligned}$$

we have

$$\begin{aligned}&\operatorname{tr}\left[\left(\bigotimes_{k=1}^{\kappa}\mathbf{V}_k\right)\cdot\mathbf{A}\right]\\ =\ &\operatorname{tr}\left[\left(\mathbf{K}_{n_{s-1}^-,n_s^+}\left[\mathbf{V}_s\otimes\left(\bigotimes_{k=s+1}^{\kappa}\mathbf{V}_k\otimes\bigotimes_{k=1}^{s-1}\mathbf{V}_k\right)\right]\mathbf{K}_{n_s^+,n_{s-1}^-}\right)\cdot\mathbf{A}\right]\\ =\ &\operatorname{tr}\left[\left(\mathbf{V}_s\otimes\left(\bigotimes_{k=s+1}^{\kappa}\mathbf{V}_k\otimes\bigotimes_{k=1}^{s-1}\mathbf{V}_k\right)\right)\cdot\left(\mathbf{K}_{n_s^+,n_{s-1}^-}\cdot\mathbf{A}\cdot\mathbf{K}_{n_{s-1}^-,n_s^+}\right)\right]\\ =\ &\operatorname{tr}\left[(\mathbf{V}_s\otimes\mathbf{W}_s)\cdot\mathbf{A}_s^*\right]\end{aligned}\tag{A7}$$

where

$$\mathbf{W}_s=\bigotimes_{k=s+1}^{\kappa}\mathbf{V}_k\otimes\bigotimes_{k=1}^{s-1}\mathbf{V}_k\tag{A8}$$

and

$$\mathbf{A}_s^*=\mathbf{K}_{n_s^+,n_{s-1}^-}\cdot\mathbf{A}\cdot\mathbf{K}_{n_{s-1}^-,n_s^+},\tag{A9}$$

where $\mathbf{W}_1=\bigotimes_{k=2}^{\kappa}\mathbf{V}_k$, $\mathbf{W}_\kappa=\bigotimes_{k=1}^{\kappa-1}\mathbf{V}_k$ and $\mathbf{K}_{n_0^-,n_1^+}=\mathbf{I}_{n_1^+}$. The matrices $\mathbf{K}$ are called vec-permutation matrices or commutation matrices (see Section 16.3, pp. 343-350 in Harville (1997)). In particular $\underset{mn\times mn}{\mathbf{K}_{m,n}}$ is the matrix such that

$$vec\left(\mathbf{H}'\right)=\mathbf{K}_{m,n}vec\left(\mathbf{H}\right),$$

for any given $m\times n-$matrix $\mathbf{H}$, and it turns up to be

$$\mathbf{K}_{m,n}=\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathbf{e}_i\left(m\right)\mathbf{e}_j'\left(n\right)\right]\otimes\left[\mathbf{e}_i\left(m\right)\mathbf{e}_j'\left(n\right)\right]',\tag{A10}$$

where $\mathbf{e}_i\left(h\right)$ is the h^{th} column of the identity matrix $\mathbf{I}_h$. The operation $vec(\cdot)$ stacks the columns of a matrix on top of each other. Then we have

$$\begin{aligned}\frac{\partial}{\partial\mathbf{V}_s}\operatorname{tr}\left[\left(\bigotimes_{k=1}^{\kappa}\mathbf{V}_k\right)\cdot\mathbf{A}\right] &= \frac{\partial}{\partial\mathbf{V}_s}\operatorname{tr}\left[(\mathbf{V}_s\otimes\mathbf{W}_s)\cdot\mathbf{A}_s^*\right]\\ &= \frac{\partial}{\partial\mathbf{V}_s}\operatorname{tr}\left[\mathbf{V}_s\cdot\left(\operatorname{tr}\left[\mathbf{W}_sA_{s,ij}^*\right]\right)_{i,j=1,\ldots,p}\right]\\ &= \frac{\partial}{\partial\mathbf{V}_s}\operatorname{tr}\left[\mathbf{V}_s\cdot\mathbf{C}_s\right]\\ &= 2\mathbf{C}_s-\operatorname{diag}\left(\mathbf{C}_s\right),\end{aligned}\tag{A11}$$

where

$$\mathbf{C}_s = \left(\mathrm{tr}\left[\mathbf{W}_s \mathbf{A}^*_{s,ij}\right]\right)_{i,j=1,\ldots,p}, \tag{A12}$$

where $\mathbf{W}_s$ and $\mathbf{A}^*_s$ are given by (A8) and (A9), respectively.

B MAXIMUM LIKELIHOOD ESTIMATION OF $(\boldsymbol{\mu}_i, \mathbf{V}_i)$, $i = 1, \ldots, \kappa$, FOR THE SEPARABLE ADDITIVE MEAN MODEL.

Proof of Theorem 1. In this case $\boldsymbol{\mu}_{\boldsymbol{x}} = \bigoplus_{i=1}^{\kappa} \boldsymbol{\mu}_i$. The likelihood function $L = L(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Gamma}_{\boldsymbol{x}}) = L((\boldsymbol{\mu}_i, \mathbf{V}_i) : i = 1, \ldots, \kappa)$ is

$$L(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Gamma}_{\boldsymbol{x}}) = \frac{\exp -\frac{1}{2}\sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})}{(2\pi)^{\frac{nm_\bullet}{2}} |\boldsymbol{\Gamma}_{\boldsymbol{x}}|^{\frac{n}{2}}}.$$

Thus, the log likelihood function can be written as

$$\ln(L) = -\frac{nm_\bullet}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Gamma}_{\boldsymbol{x}}| - \frac{1}{2}\sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}}). \tag{B1}$$

Let $\dot{\boldsymbol{x}}_r = \boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{x}_r - \bigoplus_{i=1}^{\kappa} \boldsymbol{\mu}_i$, and let $\overline{\boldsymbol{x}}_{\mathbf{s}}$ denotes the sample mean vector corresponding to time-space point $\mathbf{s} = (s_1, \ldots, s_{\kappa-1})$, that is, $\overline{\boldsymbol{x}}_{\mathbf{s}} = \frac{1}{n}\sum_{r=1}^{n}\boldsymbol{x}_{r,\mathbf{s}}$. Then the sample mean vector $\overline{\boldsymbol{x}}$ can be expressed as $\overline{\boldsymbol{x}} = \left(\overline{\boldsymbol{x}}'_{(1,\ldots,1,1)}, \ldots, \overline{\boldsymbol{x}}'_{(1,\ldots,1,m_{\kappa-1})}, \ldots, \overline{\boldsymbol{x}}'_{(1,m_2,\ldots,m_{\kappa-1})}, \ldots, \overline{\boldsymbol{x}}'_{(m_1,m_2,\ldots,m_{\kappa-1})}\right)'$. It is a partitioned $m_\bullet$−dimensional vector, and $\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{r=1}^{n}\boldsymbol{x}_r$. Since $\dot{\boldsymbol{x}}_r = \boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}} = (\boldsymbol{x}_r - \overline{\boldsymbol{x}}) + (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}})$, the sum of quadratic terms in the above log likelihood function can be written as

$$\sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}}) = \mathrm{tr}\left[\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})'\right] = \mathrm{tr}\left[\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}(\mathbf{U} + \mathbf{Z})\right],$$

where

$$\mathbf{U} = \sum_{r=1}^{n}(\boldsymbol{x}_r - \overline{\boldsymbol{x}})(\boldsymbol{x}_r - \overline{\boldsymbol{x}})', \tag{B2}$$

and

$$\begin{aligned} \mathbf{Z} &= n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}})(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}})', \\ &= n\left(\overline{\boldsymbol{x}} - \bigoplus_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)\left(\overline{\boldsymbol{x}} - \bigoplus_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)'. \end{aligned}$$

Therefore, the log likelihood function (B1) can be written as

$$\begin{aligned} \ln(L) &= -\frac{nm_\bullet}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Gamma}_{\boldsymbol{x}}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\mathbf{U}\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\mathbf{Z}\right), \\ &= -\frac{nm_\bullet}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Gamma}_{\boldsymbol{x}}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\mathbf{U}\right), \\ &\quad -\frac{n}{2}\left(\overline{\boldsymbol{x}} - \bigoplus_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \bigoplus_{i=1}^{\kappa}\boldsymbol{\mu}_i\right). \end{aligned}$$

We first find the maximum likelihood estimates for the parameters $\boldsymbol{\mu}_i : i = 1, \ldots, \kappa$, for a fixed covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{x}}$. For this we find the partial derivatives of $\ln(L)$ with respect of $\boldsymbol{\mu}_i$, for each $i = 1, \ldots, \kappa$. Now, the partial derivative $\frac{\partial}{\partial \boldsymbol{\mu}_i} \ln(L)$ is given by (see (A6) and (A1))

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\mu}_i} \ln(L) \\
= \;& -\frac{n}{2}\frac{\partial}{\partial \boldsymbol{\mu}_i}\left[\left(\overline{\boldsymbol{x}} - \bigoplus_{h=1}^{\kappa} \mu_h\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \bigoplus_{h=1}^{\kappa} \boldsymbol{\mu}_h\right)\right] \\
= \;& n\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \bigoplus_{h=1}^{\kappa} \boldsymbol{\mu}_h\right) \\
= \;& n\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \sum_{j=1}^{\kappa}\left(\bigotimes_{h=1}^{j-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_j \otimes \bigotimes_{h=j+1}^{\kappa} \mathbf{1}_{m_h}\right)\right) \\
= \;& n\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \sum_{i\neq j=1}^{\kappa}\left(\bigotimes_{h=1}^{j-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_j \otimes \bigotimes_{h=j+1}^{\kappa} \mathbf{1}_{m_h}\right)\right) \\
& -n\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_i \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right) \\
= \;& n\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \sum_{i\neq j=1}^{\kappa}\left(\bigotimes_{h=1}^{j-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_j \otimes \bigotimes_{h=j+1}^{\kappa} \mathbf{1}_{m_h}\right)\right) \\
& -n\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)\boldsymbol{\mu}_i
\end{aligned}$$

Equating $\frac{\partial}{\partial \boldsymbol{\mu}_i} \ln(L)$ to 0 we get

$$\widehat{\boldsymbol{\mu}}_i = \mathbf{D}_{\boldsymbol{\lambda}_\mathbf{i}}\left(\overline{\boldsymbol{x}} - \sum_{i\neq j=1}^{\kappa}\left(\bigotimes_{h=1}^{j-1} \mathbf{1}_{m_h} \otimes \boldsymbol{\mu}_j \otimes \bigotimes_{h=j+1}^{\kappa} \mathbf{1}_{m_h}\right)\right),$$

where

$$\begin{aligned}
\mathbf{D}_{\boldsymbol{\lambda}_\mathbf{i}} = \;& \left[\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)\right]^{-1} \\
& \cdot\left(\bigotimes_{h=1}^{i-1} \mathbf{1}_{m_h} \otimes \mathbf{I}_{m_i} \otimes \bigotimes_{h=i+1}^{\kappa} \mathbf{1}_{m_h}\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}, \quad \text{for } i = 1, \ldots, \kappa.
\end{aligned}$$

We now maximize the log likelihood function (B1) with respect to $\mathbf{V}_i$, for $i = 1, \ldots, \kappa$ for fixed $\boldsymbol{\mu}_i : i = 1, \ldots, \kappa$ to get the MLEs of $\mathbf{V}_i$, for $i = 1, \ldots, \kappa$. Since $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}$ and $|\boldsymbol{\Gamma}_{\boldsymbol{x}}|$ can be expressed as a function of $\mathbf{V}_i^{-1}$, for $i = 1, \ldots, \kappa$, maximizing with respect to $\mathbf{V}_i$, for $i = 1, \ldots, \kappa$ is equivalent to maximizing with respect to $\mathbf{V}_i^{-1}$, for $i = 1, \ldots, \kappa$. Now since $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1} = \bigotimes_{i=1}^{\kappa} \mathbf{V}_i^{-1}$ and $|\boldsymbol{\Gamma}_{\boldsymbol{x}}| =$

$\prod_{i=1}^{\kappa} |\mathbf{V}_i|^{\frac{m_\bullet}{m_i}}$, we substitute these expressions in the log likelihood (B1) we get

$$\begin{aligned}
\ln(L) &= -\frac{nm_\bullet}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Gamma}_{\boldsymbol{x}}| - \frac{1}{2}\sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}}) \\
&= -\frac{nm_\bullet}{2}\ln(2\pi) + \sum_{i=1}^{\kappa}\frac{nm_\bullet}{2m_i}\ln\left|\mathbf{V}_i^{-1}\right| - \frac{1}{2}\sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})'\left(\bigotimes_{i=1}^{\kappa}\mathbf{V}_i^{-1}\right)(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}}) \\
&= -\frac{nm_\bullet}{2}\ln(2\pi) + \sum_{i=1}^{\kappa}\frac{nm_\bullet}{2m_i}\ln\left|\mathbf{V}_i^{-1}\right| - \frac{1}{2}\mathrm{tr}\left[\left(\bigotimes_{i=1}^{\kappa}\mathbf{V}_i^{-1}\right)\cdot\mathbf{A}\right],
\end{aligned}$$

where

$$\mathbf{A} = \sum_{r=1}^{n}(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}})'. \tag{B3}$$

Since $\mathbf{V}_s = (v_{s,ij})$ is a symmetric matrix, using Equation 8.12 on pp. 306 in Harville (1997), and Equations (A7) and (A11) of Appendix A, we have

$$\begin{aligned}
\frac{\partial}{\partial\mathbf{V}_s^{-1}}\ln\left(\left|\mathbf{V}_s^{-1}\right|\right) &= 2\mathbf{V}_s - \mathrm{diag}\left(v_{s,11}, v_{s,22}, , \ldots, v_{s,m_sm_s}\right) \\
&= 2\mathbf{V}_s - \mathrm{diag}(\mathbf{V}_s),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial\mathbf{V}_s^{-1}}\mathrm{tr}\left[\left(\bigotimes_{i=1}^{\kappa}\mathbf{V}_i^{-1}\right)\cdot\mathbf{A}\right] &= \frac{\partial}{\partial\mathbf{V}_s^{-1}}\mathrm{tr}\left[\left(\mathbf{V}_s^{-1}\otimes\mathbf{W}_s\right)\cdot\mathbf{A}_s^*\right] \\
&= 2\mathbf{C}_s - \mathrm{diag}(\mathbf{C}_s),
\end{aligned}$$

where

$$\mathbf{W}_s = \bigotimes_{i=s+1}^{\kappa}\mathbf{V}_i \otimes \bigotimes_{i=1}^{s-1}\mathbf{V}_i,$$

$$\mathbf{A}_s^* = \left(\underset{n_{s-1}^- n_{s+1}^+ \times n_{s-1}^- n_{s+1}^+}{\mathbf{A}_{s,ij}^*}\right)_{i,j=1,\ldots,m_s} = \mathbf{K}_{n_s^+, n_{s-1}^-}\cdot\mathbf{A}\cdot\mathbf{K}_{n_{s-1}^-, n_s^+},$$

and

$$\mathbf{C}_s = \left(\mathrm{tr}\left[\mathbf{W}_s\mathbf{A}_{s,ij}^*\right]\right)_{i,j=1,\ldots,m_s},$$

with

$$n_s^- = \prod_{r=1}^{s} m_r, \quad \text{and} \quad n_s^+ = \prod_{r=s}^{\kappa} m_r,$$

and $\mathbf{K}_{n_s^+, n_{s-1}^-}$ the commutation matrix (see Appendix equation (A10)) with $\mathbf{K}_{n_1^+, n_0^-} = \mathbf{I}_{n_1^+}$. Now,

$$\begin{aligned}
\frac{\partial}{\partial\mathbf{V}_s^{-1}}\ln(L) &= \frac{nm_\bullet}{2m_s}\frac{\partial}{\partial\mathbf{V}_s^{-1}}\ln\left(\left|\mathbf{V}_s^{-1}\right|\right) \\
&\quad -\frac{1}{2}\frac{\partial}{\partial\mathbf{V}_s^{-1}}\mathrm{tr}\left[\left(\bigotimes_{i=1}^{\kappa}\mathbf{V}_i^{-1}\right)\cdot\mathbf{A}\right] \\
&= \frac{nm_\bullet}{2m_s}\left[2\mathbf{V}_s - \mathrm{diag}(\mathbf{V}_s)\right] - \frac{1}{2}\left(2\mathbf{C}_s - \mathrm{diag}(\mathbf{C}_s)\right)
\end{aligned}$$

Equating $\frac{\partial}{\partial \mathbf{V}_s^{-1}} \ln(L)$ to 0 we get

$$\mathbf{V}_s - \frac{1}{2}\text{diag}\left(\mathbf{V}_s\right) = \frac{m_s}{nm_\bullet}\mathbf{C}_s - \frac{1}{2}\text{diag}\left(\frac{m_s}{nm_\bullet}\mathbf{C}_s\right). \tag{B4}$$

Which implies

$$\begin{aligned}
\frac{1}{2}\text{diag}\left(\mathbf{V}_s\right) &= \text{diag}\left[\mathbf{V}_s - \frac{1}{2}\text{diag}\left(\mathbf{V}_s\right)\right] \\
&= \text{diag}\left[\frac{m_s}{nm_\bullet}\mathbf{C}_s - \frac{1}{2}\text{diag}\left(\frac{m_s}{nm_\bullet}\mathbf{C}_s\right)\right] \\
&= \frac{1}{2}\text{diag}\left(\frac{m_s}{nm_\bullet}\mathbf{C}_s\right),
\end{aligned}$$

therefore, from (B4), we obtain

$$\mathbf{V}_s = \frac{m_s}{nm_\bullet}\mathbf{C}_s, \quad \text{for } \ s = 1, \ldots, \kappa. \tag{B5}$$

C MAXIMUM LIKELIHOOD ESTIMATION OF $(\boldsymbol{\mu}_i, \mathbf{V}_i)$, FOR $i = 1, \ldots, \kappa$, FOR THE SEPARABLE MULTIPLICATIVE MEAN MODEL.

Proof of Theorem 2. In this case $\boldsymbol{\mu}_{\boldsymbol{x}} = \bigotimes_{i=1}^{\kappa} \boldsymbol{\mu}_i$. The quantity $\dot{\boldsymbol{x}}_r$ now indicates the vector $\dot{\boldsymbol{x}}_r = \boldsymbol{x}_r - \boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{x}_r - \bigotimes_{i=1}^{\kappa} \boldsymbol{\mu}_i$, then following the same logic as in the prof of theorem 1 we obtain the following expression for the log likelihood function

$$\begin{aligned}
\ln(L) &= -\frac{nm_\bullet}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left|\boldsymbol{\Gamma}_{\boldsymbol{x}}\right| - \frac{1}{2}\text{tr}\left(\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\mathbf{U}\right), \\
&\quad -\frac{n}{2}\left(\overline{\boldsymbol{x}} - \bigotimes_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)' \boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}} - \bigotimes_{i=1}^{\kappa}\boldsymbol{\mu}_i\right).
\end{aligned} \tag{C1}$$

where $\mathbf{U}$ is given in (B2). We will first find the maximum likelihood estimates for the parameters $\boldsymbol{\mu}_i : i = 1, \ldots, \kappa$, for a fixed covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{x}}$. For that we will find the first partial derivatives of $\ln(L)$ with respect to each of the $\boldsymbol{\mu}_i : i = 1, \ldots, \kappa$, respectively. Now, the partial derivative

$\frac{\partial}{\partial \boldsymbol{\mu}_i}\ln(L)$ is given by (see (A3))

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\mu}_i}\ln(L) &= -\frac{n}{2}\frac{\partial}{\partial \boldsymbol{\mu}_i}\left[\left(\overline{\boldsymbol{x}}-\bigotimes_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}}-\bigotimes_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)\right]\\
&= n\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}}-\bigotimes_{i=1}^{\kappa}\boldsymbol{\mu}_i\right)\\
&= n\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\overline{\boldsymbol{x}}-\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)\boldsymbol{\mu}_i\right)\\
&= n\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\overline{\boldsymbol{x}}\\
&\quad -n\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)\boldsymbol{\mu}_i
\end{aligned}$$

Equating $\frac{\partial}{\partial \boldsymbol{\mu}_i}\ln(L)$ to 0 we get

$$\widehat{\boldsymbol{\mu}}_i = \mathbf{E}_{\boldsymbol{\lambda}_i}\overline{\boldsymbol{x}},$$

where

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{\lambda}_i} &= \left[\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}\left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)\right]^{-1}\\
&\quad \left(\bigotimes_{h=1}^{i-1}\boldsymbol{\mu}_h\otimes\mathbf{I}_{m_i}\otimes\bigotimes_{h=i+1}^{\kappa}\boldsymbol{\mu}_h\right)'\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1},\quad \text{for } i=1,\ldots,\kappa.
\end{aligned}$$

We now substitute the values of $\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1}$ and $|\boldsymbol{\Gamma}_{\boldsymbol{x}}|$ in the log likelihood (C1), and maximize it, as before, with respect to $\mathbf{V}_i^{-1}$, for $i=1,\ldots,\kappa$, by equating its corresponding partial derivatives to 0. This procedure leads to the same results as before, that is,

$$\mathbf{V}_s = \frac{m_s}{nm_\bullet}\mathbf{C}_s,\quad \text{for } s=1,\ldots,\kappa, \tag{C2}$$

where

$$\mathbf{C}_s = \left(\operatorname{tr}\left[\mathbf{W}_s A^*_{s,ij}\right]\right)_{i,j=1,\ldots,m_s},$$

with

$$\mathbf{W}_s = \bigotimes_{i=s+1}^{\kappa}\mathbf{V}_i\otimes\bigotimes_{i=1}^{s-1}\mathbf{V}_i,$$

and

$$\mathbf{A}^*_s = \left(\underset{n^-_{s-1}n^+_{s+1}\times n^-_{s-1}n^+_{s+1}}{A^*_{s,ij}}\right)_{i,j=1,\ldots,m_s} = \mathbf{K}_{n^+_s,n^-_{s-1}}\cdot\mathbf{A}\cdot\mathbf{K}_{n^-_{s-1},n^+_s},$$

being

$$n_s^- = \prod_{r=1}^{s} m_r, \quad \text{and} \quad n_s^+ = \prod_{r=s}^{\kappa} m_r,$$

and $\mathbf{K}_{n_s^+, n_{s-1}^-}$ the commutation matrix (see Appendix equation (A10)) understanding that $\mathbf{K}_{n_1^+, n_0^-} = \mathbf{I}_{n_1^+}$.

D MAXIMUM LIKELIHOOD ESTIMATION OF $\boldsymbol{\mu_x}$ AND $\mathbf{V}_i$, FOR $i = 1, \ldots, \kappa$, FOR THE UNSTRUCTURED MEAN MODEL.

Proof of Theorem 3. Since, the log likelihood function (B1) can be written as

$$\begin{aligned} \ln(L) \quad = \quad & -\frac{nm_\bullet}{2} \ln(2\pi) - \frac{n}{2} \ln |\boldsymbol{\Gamma_x}| - \frac{1}{2} \text{tr}\left(\boldsymbol{\Gamma_x}^{-1} \mathbf{W}\right), \\ & -\frac{n}{2} \left(\overline{\boldsymbol{x}} - \boldsymbol{\mu_x}\right)' \boldsymbol{\Gamma_x}^{-1} \left(\overline{\boldsymbol{x}} - \boldsymbol{\mu_x}\right), \end{aligned}$$

then, it is easy to show that the MLE of the mean vector $\boldsymbol{\mu_x}$ is $\overline{\boldsymbol{x}}$.

In the other hand, the system equation for obtaining the MLE of $\mathbf{V}_i$, for $i = 1, \ldots, \kappa$ is the same as in the structured cases, but replacing $\boldsymbol{\mu_x}$ by $\overline{\boldsymbol{x}}$, that is,in the unstructured case the matrix $\mathbf{A}$ is $\mathbf{A} = \sum_{r=1}^{n} (\boldsymbol{x}_r - \overline{\boldsymbol{x}})(\boldsymbol{x}_r - \overline{\boldsymbol{x}})'$, and

$$\mathbf{V}_s = \frac{m_s}{nm_\bullet} \mathbf{C}_s,$$

for $s = 1, \ldots, \kappa$.

ACKNOWLEDGEMENTS

The second author would like to thank the College of Business at the University of Texas at San Antonio for their continued support with this publication.

References

[1] D. Akdemir and A. K. Gupta, *Array variate random variables with multiway Kronecker delta covariance matrix structure*, Tech Report. Department of Mathematics and Statistics at the Bowling Green State University, (2011).

[2] C. Bhattacharyya, L. R. Grate, A. Rizki, D. Radisky, F.J. Molina, M.I. Jordan, M. J. Bissell and I. S. Mian, "Simultaneous classification and relevant feature identification in high-dimensional spaces: application to molecular profiling data", Signal Processing 83(4), 729-743 (2003).

[3] S. Dudoit, J. Fridlyand and T. P. Speed, *Comparison of discrimination methods for the classification of tumors using gene expression data*, Journal of the American Statistical Association, 97, 7787 (2002).

[4] J. Friedman, T. Hastie and R. Tibshirani, *Sparce inverse covariance estimation with the graphical lasso*, Biostatistics 9, 432-441 (2008).

[5] D. A. Harville, *Matrix algebra from a Statistician's perspective*, Springer-Verlag NY, (1997).

[6] K. I. Kim and R. Simon, "Probabilistic classifiers with high-dimensional data", Biostatistics 12(3), 399-412 (2011).

[7] P. M. Kroonenberg, *Applied Multiway Data Analysis*, (John Wiley & Sons, Inc., New Jersey, 2008).

[8] M. Krzyśko and M. Skorzybut, "Discriminant analysis of multivariate repeated measures data with a Kronecker product structured covariance matrices", Statistical Papers 50(4), 817-835 (2009).

[9] C. Lai, M. J. Reinders, L. J. VanT Veer and L. F. Wessels, "A comparison of univariate and multivariate gene selection techniques for classification of cancer datasets", BMC Bioinformatics 7, 235 (2006).

[10] R. Leiva, "Linear discrimination with equicorrelated training vectors", J. Multivariate Analysis, 98(2), 384-409 (2007).

[11] R. Leiva and A. Roy, "Classification rules for triply multivariate data with an AR(1) correlation structure on the repeated measures over time", J. of Statistical Planning and Inference 139(8), 2598-2613 (2009).

[12] R. Leiva and A. Roy, "Linear Discrimination for Multi-level Multivariate Data with Separable Means and Jointly Equicorrelated Covariance Structure", Journal of Statistical Planning and Inference 141(5), 1910-1924 (2011a).

[13] R. Leiva and A. Roy, "Linear discrimination for three-level multivariate data with separable additive mean vector and doubly exchangeable covariance structure", Computational Statistics and Data Analysis, DOI: 10.1016/j.csda.2011.10.007 (Published Online, 2011b).

[14] L. M. Lix and T. T. Sajobi, *Discriminant analysis for repeated measures data: a review*, Frontiers in Psychology 1, Article 146 (2010).

[15] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition* (Wiley, New York, 1992).

[16] M. Ohlson, M. R. Ahmad, and D. von Rosen, "More on the Kronecker structured covariance matrix", Communications in Statistics - Theory and Methods, (Acepted, 2011).

[17] S. A. Paranjpe and A. P. Gore, "Selecting variables for discrimination when covariance matrices are unequal", Statistics and Probability Letters 21(5), 417-419 (1994).

[18] P. S. Park and A.M. Kshirsagar, "Distances between normal populations when covariance matrices are unequal", Communications in Statistics-Theory and Methods 23(12), 3549-3556 (1994).

[19] T. Pavlenko, A. Björkstrom and A. Tillander, "Covariance Structure Approximation via gLasso in High-Dimensional Supervised Classification", Journal of Applied Statistics 1-24 (2011).

[20] A. Roy and R. Leiva "Discrimination with jointly equicorrelated multi-level multivariate data", Advances in Data Analysis and Classification 1(3), 175-199 (2007).

[21] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC.