Journal of Multivariate Analysis 143 (2016) 345–361

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Existence and uniqueness of the maximum likelihood estimator for models with a Kronecker product covariance structure

Beata Roś [a], Fetsje Bijma [a], Jan C. de Munck [b], Mathisca C.M. de Gunst [a,*]

[a] Department of Mathematics, Faculty of Exact Sciences, VU University Amsterdam, De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands

[b] Department of Physics and Medical Technology, VU University Medical Center, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 19 July 2013
Available online 21 September 2015

AMS subject classifications:
62F10
62F30
62J10
62H12

Keywords:
Matrix normal model
Covariance matrix
Kronecker product structure
Maximum likelihood estimation
Existence and uniqueness of estimator

ABSTRACT

This paper deals with multivariate Gaussian models for which the covariance matrix is a Kronecker product of two matrices. We consider maximum likelihood estimation of the model parameters, in particular of the covariance matrix. There is no explicit expression for the maximum likelihood estimator of a Kronecker product covariance matrix. We investigate whether the maximum likelihood estimator of the covariance matrix exists and whether it is unique. We consider models with general, with double diagonal, and with one diagonal Kronecker product covariance matrices, and find different results.

1. Introduction

Often data are measured in multiple domains like space, time, frequency, or different versions or scales of one of these. In this paper we consider the case where data are measured in two domains, so that we have matrix-valued observations. We assume that the observations are independent and identically distributed with a matrix normal distribution or, equivalently, that the vectorized observations are independent and identically multivariate normally distributed with a covariance matrix which is the Kronecker product of two positive definite matrices. The focus of this paper is maximum likelihood estimation of the model parameters and, in particular, the existence and uniqueness of the maximum likelihood estimator of the covariance matrix.

Multivariate normal models with a Kronecker product covariance structure have been proposed for many different purposes and applications. For instance, for spatio-temporal estimation of the covariance structure of EEG/MEG data [4,6,10,21] or environmental data [8,18], for missing data imputation for microarray or Netflix movie rating data [1], and for multi-task learning for detecting land mines in multiple fields or recognizing faces between different subjects [25]. Several

* Corresponding author.
E-mail address: m.c.m.de.gunst@vu.nl (M.C.M. de Gunst).

http://dx.doi.org/10.1016/j.jmva.2015.05.019

tests have been developed for checking whether or not the covariance matrix is separable, that is, is a Kronecker product of two matrices [16,20]. Two general references for properties of Kronecker products are [17,22].

In practice maximum likelihood estimates for the Kronecker product covariance matrix are typically obtained by numerical approximation, because no explicit expression for the solution of the likelihood equations exists. However, one should be aware of the fact that even though the numerical approximation algorithm that is used for obtaining the estimate may converge to some value, this value may only be one of many values that maximize the likelihood. In principle, it could also happen that, although all updates of the algorithm produce full-rank Kronecker product matrices, their limiting value is a Kronecker product matrix that does not have full rank and thus is not the estimate one is looking for. This is why it is important to know under which conditions the maximum likelihood estimator exists and, if it exists, whether or not it is unique. Moreover, because in practice Kronecker product covariance models are more and more used in situations where the number of observations is much smaller than the dimension of the covariance matrix, it is essential to not only consider large sample sizes, but to investigate what happens for small sample sizes as well.

We remark that in this paper we do not consider (properties of) algorithms for finding numerical approximations of maximum likelihood estimates of a Kronecker product covariance matrix. This is, for instance, done in [7,15,20,23] where the widely used flip-flop algorithm is discussed, in [11] for the case in which one of the two components of the Kronecker product is a persymmetric matrix, and in [24] for the case where the covariance matrix is a Kronecker product of two Toeplitz matrices. Instead, we study theoretical properties, namely existence and uniqueness, of the maximum likelihood estimator itself. We do this for three different cases: for the maximum likelihood estimator of a general Kronecker product covariance matrix, of a Kronecker product covariance matrix with both composing matrices diagonal, and of a Kronecker product covariance matrix with one of the two composing matrices diagonal. The results presented in this paper show that existence and uniqueness of the maximum likelihood estimator cannot be taken for granted.

Analysis of existence and uniqueness of the maximum likelihood estimator of a Kronecker product covariance matrix is non-trivial, because the parameter space is not convex and the set of multivariate normal distributions with a Kronecker product covariance matrix is a curved exponential family. The topic was also studied in the papers [7] and [20], which have stimulated a lot of research in the area. In [7] a condition is given that is claimed to be necessary and sufficient for existence. However, the proof of the sufficiency part is incomplete. We cannot verify sufficiency of this condition, but instead prove existence under a stronger condition. In [7] and [20] different results about conditions for uniqueness are given. Unfortunately, as we demonstrate below, both results are incorrect. Additionally we give novel proofs for the existence and uniqueness of the maximum likelihood estimator for the cases where only one and where both composing matrices of the Kronecker product covariance matrix is/are constrained to be diagonal.

The remainder of the paper is structured as follows. In the next section the model and the estimation problem are introduced. In Section 3 existence and uniqueness of the maximum likelihood estimator for general Kronecker product covariance matrices are discussed. In Section 4 our results for the model with both components diagonal are presented. These results are extended to the case with only one component diagonal in Section 5. We conclude with a discussion of our results. Some lemmas and proofs are given in the Appendix.

2. Model and estimation problem

Suppose we have n matrix-valued observations $X_1, \ldots, X_n \in \mathcal{M}_{p,q}$, that represent measurements in two domains, say space and time, such that $X_k(i,j)$ is the measurement at the ith point in space and at the jth point in time for the kth observation, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $k = 1, \ldots, n$. Here $\mathcal{M}_{p,q}$ denotes the space of real $p \times q$ matrices. For example, in a study in which a subject is repeatedly exposed to a stimulus and the subject's brain activity is measured by multi-channel EEG for a fixed period of time after each stimulus, p could be the number of EEG channels, q the number of time points per recording, and X_k the signal recorded after the kth stimulus, $k = 1, \ldots, n$, with n being the total number of stimuli the subject is exposed to [6].

We assume that the observations $X_1, \ldots, X_n$ are independent and

$$\operatorname{vec}(X_k) \sim \mathcal{N}(\mu, \Psi \otimes \Gamma), \quad k = 1, \ldots, n, \tag{1}$$

where μ is a vector of length pq, $\Gamma \in \mathcal{M}_{p,p}$ and $\Psi \in \mathcal{M}_{q,q}$ are positive definite matrices, and $\otimes$ denotes the Kronecker product. Here the vec operator vectorizes a matrix by stacking its columns into one vector. We note that for Γ, Ψ positive definite, the Kronecker product $\Psi \otimes \Gamma$ is also a positive definite matrix [9], which implies that model (1) is well defined. The Kronecker product assumption for the covariance matrix means that $\operatorname{Cov}(X_k(i_1,j_1), X_k(i_2,j_2)) = \Gamma(i_1,i_2)\Psi(j_1,j_2)$ for all i_1, i_2, j_1, j_2 and k. The mean vector μ and the covariance matrix $\Psi \otimes \Gamma$ in model (1) are generally unknown and need to be estimated from the observations.

We consider maximum likelihood estimation for μ and $\Psi \otimes \Gamma$. Let $M \in \mathcal{M}_{p,q}$ be such that $\operatorname{vec}(M) = \mu$. The likelihood function for M, Γ and Ψ is

$$L(M, \Gamma, \Psi \mid X_1, \ldots, X_n) = (2\pi)^{-\frac{1}{2}pqn} |\Gamma|^{-\frac{1}{2}qn} |\Psi|^{-\frac{1}{2}pn} \operatorname{etr}\left(-\frac{1}{2}\Psi^{-1} \sum_{k=1}^{n} (X_k - M)^T \Gamma^{-1} (X_k - M)\right), \tag{2}$$