



HAL
open science

Robust Mean and Covariance Matrix Estimation Under Heterogeneous Mixed-Effects Model

Alexandre Hippert-Ferrer, M. N. El Korso, Arnaud Breloy, Guillaume Ginolhac

► **To cite this version:**

Alexandre Hippert-Ferrer, M. N. El Korso, Arnaud Breloy, Guillaume Ginolhac. Robust Mean and Covariance Matrix Estimation Under Heterogeneous Mixed-Effects Model. 2021. hal-03156771

HAL Id: hal-03156771

<https://hal.science/hal-03156771>

Preprint submitted on 2 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust Mean and Covariance Matrix Estimation Under Heterogeneous Mixed-Effects Model

A. Hippert-Ferrer⁺, M. N. El Korso⁺, A. Breloy⁺ and G. Ginolhac^{*a}

^{a+} *Paris-Nanterre University, 50 rue de Sèvres, 92410 Ville d'Avray, France*

^{*} *LISTIC, University Savoie Mont-Blanc, France*

Abstract

In this paper, robust mean and covariance matrix estimation are considered in the context of mixed-effects models. Such models are widely used to analyze repeated measures data which arise in several signal processing applications that need to incorporate a same global individuals behavior with a possible individual variations. In this context, most algorithms are based on Gaussian assumption of the observations. Nevertheless, in certain situations in which there exist outliers within the data set, such assumption is not valid and leads to a dramatic performance loss. To overcome this drawback, we design an expectation-conditional maximization either algorithm in which the heterogeneous component is considered as part of the complete data. Then, the proposed algorithm is cast into a parallel scheme, w.r.t. the individuals, in order to alleviate the computational cost and a possible central processor overload. Finally, the proposed algorithm is extended to deal with missing data which refers to the situation where part of the individual responses are unobserved. Numerical simulations are performed in order to assess the performance of the proposed algorithm in regard to robust regression estimators, probabilistic principal component analysis and its recent robust version.

Keywords: Maximum likelihood, expectation maximization, robust mean estimation.

1. Introduction

The use of mixed-effects models has become popular in a wide range of signal processing applications [1]. Specifically, it is mainly used in applications which require to model

[☆]This work was supported by ANR ASTRID project MARGARITA (ANR-17-ASTR-0015).

a same global individuals behavior with a possible individual variations. This reads,

$$\mathbf{y}_{ij} = \mathbf{A}_i \mathbf{m} + \mathbf{Z}_i \mathbf{t}_{ij} + \mathbf{n}_{ij}, \quad \forall j = 1, \dots, q \text{ and } i = 1, \dots, s, \quad (1)$$

where $\mathbf{y}_{ij} \in \mathbb{C}^{n_i}$ denotes the complex vector of the j -th observation at the i -th individual, $\mathbf{t}_{ij} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{n}_{ij} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ characterize the individual errors. Finally, $\mathbf{A}_i \in \mathbb{C}^{n_i \times p}$, $\mathbf{Z}_i \in \mathbb{C}^{n_i \times q_i}$ and $\mathbf{m} \in \mathbb{C}^p$ denote the known design matrix related to the fixed-effect \mathbf{m} , the unknown design matrix related to the random-effect \mathbf{t}_{ij} and the common unknown vector mean, respectively.

As an example in signal processing, we can cite direction-dependent calibration in the context of partially calibrated arrays, in which each individual represents a sub-array [2, 3, 4]. The design matrix is the array response (which can be known in a data-aided scenario) whereas the unknown direction-dependent gain is represented by \mathbf{m} . A second illustration is found in imaging and data processing in the context of the next generation of radio-interferometers [5]. Such instruments are composed of several spaced stations (individuals) in which each station observation represents the visibilities observed within one station. In this case, the design matrix is a known linear operator that maps the image from the space domain to the visibility domain and the mean denotes the intensity vector (i.e., the unknown common image observed by each station) [6]. In both examples, the mixed-effect component, \mathbf{t}_{ij} , represents a local interference concerning each individual. An additional application of the model (1) in the case of only one individual ($s = 1$) can be found in the so-called probabilistic principal component analysis literature [7].

In the aforementioned references, the authors considered a Gaussian distributed observation. Nevertheless, such assumption is not realistic in a plethora of signal processing applications as those related to high resolution sensing systems, non-homogeneous environment or in the possible presence of outliers [8, 9].

In this paper, we consider a parametric model that takes into account the possible presence of outliers by modeling the within-individual error as a mixture of Gaussian process and non-Gaussian distributed noise lying in a low rank covariance matrix. The mixed-effects model gives us the flexibility to assign outliers only to some of the individuals depending on their environment. Consequently, the proposed model becomes

$$\mathbf{y}_{ij} = \mathbf{A}_i \mathbf{m} + \frac{\sqrt{\tau_{ij}}}{2} \mathbf{W}_i \mathbf{t}_{ij} + \mathbf{n}_{ij} \quad (2)$$

in which \mathbf{t}_{ij} represents the low rank heterogeneous random effect error related to a positive texture parameter τ_{ij} and a low rank loading matrix $\mathbf{W}_i \in \mathbb{C}^{n_i \times r_i}$ of rank $r_i \leq n_i$. The heterogeneous component, $\sqrt{\tau_{ij}}\mathbf{W}_i\mathbf{t}_{ij}$, leads to a compound Gaussian random effect error in the case of a random texture parameter, or its counterpart referred to as heteroscedastic Gaussian random effect error if the texture parameter is assumed unknown and deterministic [10]. Such process formulation became popular since it owns a great flexibility allowing to gather several elliptically symmetric distributions, e.g., Gaussian, K- and t-distributions, Cauchy distributions, etc. Again, considering the examples given above, the non-Gaussian random effect error can model the clutter (the external noise), $\sqrt{\tau_{ij}}\mathbf{W}_i\mathbf{t}_{ij}$, plus thermal noise (the internal noise) \mathbf{n}_{ij} which exists in several array/radar processing applications, or the unknown background of power-fluctuating sources present in the imaging process of the recent radio-interferometers [11, 12, 13].

In order to tackle efficiently the estimation procedure related to model (2), we use the Expectation-Conditional Maximisation Either (ECME) algorithm. We select in a proper way the complete data, which is composed of the observed variables \mathbf{y}_{ij} and the latent variables \mathbf{t}_{ij} (and the missing observations in the case of a missing data context) which leads to closed-form expressions in the E-step and M-step. Then, the proposed estimator is cast into a parallel scheme in order to lower the computational cost and avoid a possible central processor overload.

2. ECME based estimation under heterogeneous mixed-effects model

Considering model (2), the unknown vector parameter reads (with some abuse of notation) $\boldsymbol{\theta} = \left[\mathbf{m}^T, \{\boldsymbol{\zeta}_i^T\}_i, \{\tau_{ij}\}_{ij} \right]^T$, in which $\boldsymbol{\zeta}_i$ is the concatenation of the non-redundant elements in \mathbf{W}_i (the estimation of \mathbf{W}_i is known to display a rotational ambiguity which is discussed later in Section 3) and where the texture parameters $\{\tau_{ij}\}_{ij}$ are considered deterministic and unknown. The latter assumption ensures more flexibility as the texture distribution is not specified, which avoids any possible model misspecification. We consider the maximum likelihood principal, that is

$$\boldsymbol{\theta}_{\text{ML}} = \arg \max_{\boldsymbol{\theta}} \mathcal{L}(\{\mathbf{y}_{ij}\}_{ij} | \boldsymbol{\theta}) \quad (3)$$

where $\mathcal{L}(\boldsymbol{\theta} | \{\mathbf{y}_{ij}\}_{ij}) = -\sum_i \sum_j \log |\mathbf{C}_{ij}| + (\mathbf{y}_{ij} - \mathbf{A}_i \mathbf{m})^H \mathbf{C}_{ij}^{-1} (\mathbf{y}_{ij} - \mathbf{A}_i \mathbf{m})$ with $\mathbf{C}_{ij} = \tau_{ij} \mathbf{W}_i \mathbf{W}_i^H + \sigma^2 \mathbf{I}$ in which independence between individuals and between observations is assumed. It is clear that solving (3) is challenging due to the non-convexity of the objective function. Consequently, in the following we propose the use of the ECME algorithm, known to be an efficient extension of the EM scheme with faster monotone convergence [14]. The ECME algorithm is an iterative algorithm whose estimates $\boldsymbol{\theta}^{(m)}$ converge, under certain mild conditions, to the maximum likelihood estimate $\boldsymbol{\theta}_{\text{ML}}$ (m denoting the iteration number). The ECME algorithm is decomposed in two steps: the E-step and the M-step. In the E-step, we derive the surrogate function $Q(\cdot | \cdot)$, which is the expectation of the log-likelihood of the complete data \mathcal{L}_C conditioned on the observed data and the previously computed $\boldsymbol{\theta}^{(m-1)}$. The complete data is a combination of the observed data \mathbf{y}_{ij} and the missing/latent data \mathbf{t}_{ij} . While the classical M-step requires to maximize Q , we will also consider maximizing the loglikelihood \mathcal{L} of the incomplete data [14] depending on the ease of derivation w.r.t a block of the unknown vector parameter $\boldsymbol{\theta}$. This procedure, which is described hereafter, is repeated until convergence.

E-Step: First, the complete data must be specified in order to simplify the M-step while maintaining the derivation of the expectation of \mathcal{L}_C feasible. Based on (2), it seems natural to choose the complete data as $\mathbf{x}_{ij} = [\mathbf{y}_{ij}^T, \mathbf{t}_{ij}^T]^T$. Consequently,

$$\begin{aligned} \mathcal{L}_C(\mathbf{x} | \boldsymbol{\theta}) &= \sum_i \sum_j \log p(\mathbf{y}_{ij} | \mathbf{t}_{ij}, \boldsymbol{\theta}) + \log p(\mathbf{t}_{ij} | \boldsymbol{\theta}) \\ &= -\sum_i \sum_j n_i \log \sigma^2 - \|\mathbf{r}_{ij} - \sqrt{\tau_{ij}} \mathbf{W}_i \mathbf{t}_{ij}\|_{\sigma^2}^2 - \|\mathbf{t}_{ij}\|_2^2 \end{aligned} \quad (4)$$

in which $\mathbf{r}_{ij} = \mathbf{y}_{ij} - \mathbf{A}_i \mathbf{m}$ and the weighted norm reads $\|\mathbf{b}\|_{\mathbf{B}}^2 = \mathbf{b} \mathbf{B}^{-1} \mathbf{b}^H$. Thus, the so-called Q function reads

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) &= E_{\{\mathbf{x}_{ij}\}_{ij} | \{\mathbf{y}_{ij}\}_{ij}, \boldsymbol{\theta}^{(m)}} \{ \mathcal{L}_C(\boldsymbol{\theta} | \mathbf{x}) \} \\ &= E_{\{\mathbf{t}_{ij}\}_{ij} | \{\mathbf{y}_{ij}\}_{ij}, \boldsymbol{\theta}^{(m)}} \{ \mathcal{L}_C(\boldsymbol{\theta} | \mathbf{x}) \} = \sum_i \sum_j Q_{ij}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) \end{aligned}$$

where $Q_{ij}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = E_{\mathbf{t}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\theta}^{(m)}} \{ \log p(\mathbf{x}_{ij} | \boldsymbol{\theta}) \}$. First, let us derive the pdf of the latent variable \mathbf{t}_{ij} conditioned on the observation and $\boldsymbol{\theta}^{(m)}$. Namely, since $p(\mathbf{t}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\theta}^{(m)}) \propto p(\mathbf{y}_{ij} | \mathbf{t}_{ij}, \boldsymbol{\theta}^{(m)}) p(\mathbf{t}_{ij} | \boldsymbol{\theta}^{(m)})$, after some calculus and considering the adequate normaliza-

tion constant, we obtain

$$p(\mathbf{t}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\theta}^{(m)}) \sim \mathcal{CN}\left(\mathbf{V}_{ij}^{-1(m)} \mathbf{u}_{ij}^{(m)}, \mathbf{V}_{ij}^{-1(m)}\right) \quad (5)$$

where

$$\begin{aligned} \mathbf{u}_{ij}^{(m)} &= \frac{\sqrt{\tau_{ij}^{(m)}}}{(\sigma^{(m)})^2} \mathbf{W}_i^H \mathbf{r}_{ij}^{(m)} \\ \mathbf{V}_{ij}^{(m)} &= \mathbf{I} + \frac{\tau_{ij}^{(m)}}{(\sigma^{(m)})^2} \mathbf{W}_i^H \mathbf{W}_i \end{aligned}$$

Consequently, the surrogate function reads

$$\begin{aligned} Q_{ij}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) &= E_{\mathbf{t}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\theta}^{(m)}} \left\{ -n_i \log \sigma^2 - \text{Tr}(\mathbf{t}_{ij} \mathbf{t}_{ij}^H \mathbf{V}_{ij}) - \|\mathbf{r}_{ij}\|_{\sigma^2}^2 + 2 \frac{\sqrt{\tau_{ij}}}{\sigma^2} \Re(\mathbf{r}_{ij}^H \mathbf{W}_i \mathbf{t}_{ij}) \right\} \\ &= -n_i \log \sigma^2 - \text{Tr}(\widehat{\mathbf{T}}_{ij}^{(m)} \mathbf{V}_{ij}) - \|\mathbf{r}_{ij}\|_{\sigma^2}^2 + \frac{2\sqrt{\tau_{ij}}}{\sigma^2} \Re(\mathbf{r}_{ij}^H \mathbf{W}_i \widehat{\mathbf{t}}_{ij}^{(m)}) \end{aligned} \quad (6)$$

in which

$$\widehat{\mathbf{t}}_{ij}^{(m)} = E_{\mathbf{t}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\theta}^{(m)}} \{\mathbf{t}_{ij}\} = \mathbf{V}_{ij}^{-1(m)} \mathbf{u}_{ij}^{(m)} \quad (7)$$

and

$$\widehat{\mathbf{T}}_{ij}^{(m)} = E_{\mathbf{t}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\theta}^{(m)}} \{\mathbf{t}_{ij} \mathbf{t}_{ij}^H\} = \mathbf{V}_{ij}^{-1(m)} + \widehat{\mathbf{t}}_{ij}^{(m)} \widehat{\mathbf{t}}_{ij}^{(m)H} \quad (8)$$

M-Step: This step is carried by block coordinate descent, which has the advantage to lead to closed-form expressions of the unknown parameter when maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ w.r.t. $\boldsymbol{\theta}$. We recall that the EM extension used here is the ECME algorithm in which the estimates of $\{\zeta_i^T\}_i$, $\{\tau_{ij}\}_{ij}$, σ^2 are obtained by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ based on the complete log-likelihood, whereas the \mathbf{m} estimate is obtained by maximizing the log-likelihood of the incomplete data (i.e. observations only). After some calculus, this leads to the following updates (for the sake of clarity, the iteration index m is omitted):

$$\sqrt{\tau_{ij}} = \frac{\Re(\mathbf{r}_{ij}^H \mathbf{W}_i \widehat{\mathbf{t}}_{ij})}{\text{Tr}(\widehat{\mathbf{T}}_{ij} \mathbf{W}_i^H \mathbf{W}_i)} \quad (9)$$

$$\sigma^2 = \frac{1}{\sum_i n_i} \sum_i \sum_j \text{Tr} \left\{ \tau_{ij} \widehat{\mathbf{T}}_{ij} \mathbf{W}_i^H \mathbf{W}_i + \mathbf{r}_{ij}^H \mathbf{r}_{ij} - 2\sqrt{\tau_{ij}} \Re(\mathbf{r}_{ij}^H \mathbf{W}_i \widehat{\mathbf{t}}_{ij}) \right\} \quad (10)$$

$$\mathbf{W}_i = \sum_j \sqrt{\tau_{ij}} \mathbf{r}_{ij} \widehat{\mathbf{t}}_{ij}^H \left(\sum_j \widehat{\mathbf{T}}_{ij} \right)^{-H} \quad (11)$$

$$\mathbf{m} = \left(\sum_i \sum_j \mathbf{A}_i^H \boldsymbol{\Gamma}_{ij} \mathbf{A}_i \right)^{-1} \sum_i \sum_j \mathbf{A}_i^H \boldsymbol{\Gamma}_{ij} \mathbf{y}_{ij} \quad (12)$$

in which $\mathbf{\Gamma}_{ij} = \frac{1}{\sigma^2} \left(\mathbf{I} - \mathbf{W}_i \left(\frac{r_{ij}}{\sigma^2} \mathbf{I} + \mathbf{W}_i^H \mathbf{W}_i \right)^{-1} \mathbf{W}_i^H \right)$. As a summary, the ECME algorithm operates as described in pseudo-code 1.

Algorithm 1 ECME algorithm for parameter estimation under mixed-effects model

- 1: Initialize mean $\mathbf{m}^{(m=0)}$ with robust multivariate mean estimate (e.g. [15, 16, 17]) and loading matrix $\mathbf{W}_i^{(m=0)}$ with PPCA algorithm separately for each individual [7].
 - 2: **repeat**
 - 3: Given current $\boldsymbol{\theta}^{(m)}$, compute $\hat{\mathbf{t}}_{ij}^{(m)}$ and $\hat{\mathbf{T}}_{ij}^{(m)}$ using (7) and (8)
 - 4: Update $\boldsymbol{\theta}^{(m+1)}$ as in (9), (10), (11) followed by (12)
 - 5: $m \leftarrow m + 1$
 - 6: **until** convergence of $\|\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{(m)}\|_F^2$ is met.
 - 7: **return** $\hat{\boldsymbol{\theta}}$
-

3. Discussions and extension

3.1. Parallelization

A parallel scheme naturally appears from (9)–(12). This can be used to enhance the computational cost and avoid a central processor overload. Fig. 1 represents the operation flow and signaling exchange between a local individual’s processor and the fusion center. The consensus step is enforced for deriving the common mean and homogeneous noise power. Note that the update of \mathbf{m} involves the inversion of a $n_i \times n_i$ matrix in classical robust multivariate covariance and mean estimators [10]. Conversely, the ECME algorithm updates \mathbf{m} locally by the inversion of a $r_i \times r_i$ matrix followed, at the fusion center, by an inversion of a $p \times p$ matrix. In array processing we commonly have $p \ll n_i$ (e.g., less sources than sensors or in radio-interferometry imaging), which makes the proposed method computationally efficient.

3.2. Rotational ambiguity of the loading matrix

It has been noted in [7] that the estimates of the loading matrix are generally not orthogonal. It means that if \mathbf{O}_i denotes an orthogonal rotational matrix of an adequate size, then $\widehat{\mathbf{W}}_{i\text{ML}} \mathbf{O}_i$ remains the maximum likelihood estimate. Yet, it is common to be interested in the spanned subspace by the columns of \mathbf{W}_i rather than the loading matrix

itself. On the other hand, if needed, a post-processing of $\widehat{\mathbf{W}}_i$ is proposed in [7], which consists in computing the SVD of $\widehat{\mathbf{W}}_i^H \widehat{\mathbf{W}}_i = \mathbf{O}_i^H \mathbf{\Lambda}_i \mathbf{O}_i$ and rotating according to $\mathbf{O}_i \widehat{\mathbf{W}}_i$.

3.3. Extension to the missing data case

The missing data case refers to the situation where part of the individual responses are unobserved, which is a very common issue when analyzing time series, such as remote sensing data [18] or biochemical data [19]. In the following, we denote the observed responses by \mathbf{y}_{ij}^o and the unobserved responses by \mathbf{y}_{ij}^u for the i -th individual at the j -th observation. Furthermore, the scenario is assumed to be missing at random (i.e. the missing observations do not depend on their values) [20]. In order to use of the results obtained in the previous section, we consider the complete data as $\mathbf{x} = \{\mathbf{x}_{ij}\}_{ij}$ in which $\mathbf{x}_{ij} = [\mathbf{y}_{ij}^{oT}, \mathbf{y}_{ij}^{uT}, \mathbf{t}_{ij}^T]^T$. Consequently,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = E_{\{\mathbf{y}_{ij}^u\}_{ij}|\{\mathbf{y}_{ij}^o\}_{ij}, \boldsymbol{\theta}^{(m)}} \left\{ E_{\{\mathbf{t}_{ij}\}_{ij}|\{\mathbf{y}_{ij}^o\}_{ij}, \{\mathbf{y}_{ij}^u\}_{ij}, \boldsymbol{\theta}^{(m)}} \{ \mathcal{L}_C(\boldsymbol{\theta}|\mathbf{x}) \} \right\} \quad (13)$$

The inner expectation in (13) is given by (6), the outer expectation can be obtained from the classical results of Anderson [21]. Namely, it is fairly easy to see that

$$\widehat{\mathbf{y}}_{ij} \triangleq E_{\mathbf{y}_{ij}^u|\mathbf{y}_{ij}^o, \boldsymbol{\theta}^{(m)}} \{\mathbf{y}_{ij}\} = \left[\mathbf{y}_{ij}^{oT} \widehat{\mathbf{y}}_{ij}^{uT} \right]^T = \begin{bmatrix} \mathbf{y}_{ij}^o \\ \mathbf{m}^u + \boldsymbol{\Sigma}_{ij}^{uo} \boldsymbol{\Sigma}_{ij}^{oo^{-1}} (\mathbf{y}_{ij}^o - \mathbf{m}^o) \end{bmatrix} \quad (14)$$

and

$$\widehat{\mathbf{Y}}_{ij} \triangleq E_{\mathbf{y}_{ij}^u|\mathbf{y}_{ij}^o, \boldsymbol{\theta}^{(m)}} \{\mathbf{y}_{ij} \mathbf{y}_{ij}^H\} = \begin{bmatrix} \mathbf{y}_{ij}^o \mathbf{y}_{ij}^{oH} & \mathbf{y}_{ij}^o \widehat{\mathbf{y}}_{ij}^{uH} \\ \widehat{\mathbf{y}}_{ij}^u \mathbf{y}_{ij}^{oH} & \boldsymbol{\Sigma}_{ij}^{uu} - \boldsymbol{\Sigma}_{ij}^{uo} \boldsymbol{\Sigma}_{ij}^{oo^{-1}} \boldsymbol{\Sigma}_{ij}^{ou} + \widehat{\mathbf{y}}_{ij}^u \widehat{\mathbf{y}}_{ij}^{uH} \end{bmatrix} \quad (15)$$

in which the mean and the covariance matrix of \mathbf{y}_{ij} are decomposed as $\mathbf{m} = [\mathbf{m}^{oT} \mathbf{m}^{uT}]^T$

and $\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} \boldsymbol{\Sigma}_{ij}^{oo} & \boldsymbol{\Sigma}_{ij}^{ou} \\ \boldsymbol{\Sigma}_{ij}^{uo} & \boldsymbol{\Sigma}_{ij}^{uu} \end{bmatrix}$. Then, we deduce the E-step which consists in updating the following expectations (with regard to $\{\mathbf{x}_{ij}\}_{ij} | \{\mathbf{y}_{ij}^o\}_{ij}, \boldsymbol{\theta}^{(m)}$):

$$\widehat{\mathbf{r}}_{ij} \triangleq E \{\mathbf{r}_{ij}\} = \widehat{\mathbf{y}}_{ij} - \mathbf{m} \quad (16)$$

$$\widehat{\mathbf{R}}_{ij} \triangleq E \{\mathbf{r}_{ij} \mathbf{r}_{ij}^H\} \triangleq \widehat{\mathbf{Y}}_{ij} + \mathbf{m} \mathbf{m}^H - 2\Re \{\widehat{\mathbf{y}}_{ij} \mathbf{m}^H\} \quad (17)$$

$$\widehat{\mathbf{t}}_{ij} = E \{\mathbf{t}_{ij}\} \triangleq \frac{\sqrt{\tau_{ij}}}{\sigma^2} \mathbf{V}_{ij}^{-1} \mathbf{W}_i E \{\mathbf{r}_{ij}\} \quad (18)$$

$$\widehat{\mathbf{r}}_{ij} \widehat{\mathbf{t}}_{ij}^H \triangleq E \{\mathbf{r}_{ij} \mathbf{t}_{ij}^H\} = \frac{\sqrt{\tau_{ij}}}{\sigma^2} \widehat{\mathbf{R}}_{ij} \mathbf{W}_i \quad (19)$$

$$\widehat{\mathbf{T}}_{ij} \triangleq E \{\mathbf{t}_{ij} \mathbf{t}_{ij}^H\} = \mathbf{V}_{ij}^{-1} + \frac{\tau_{ij}}{\sigma^4} \mathbf{V}_{ij}^{-1} \mathbf{W}_i^H \mathbf{R}_{ij} \mathbf{W}_i \mathbf{V}_{ij}^{-1} \quad (20)$$

Finally, by plugging (16)–(20) into (13), the M-step of the ECME reads

$$\sqrt{\tau_{ij}} = \frac{\Re\left(\text{Tr}(\widehat{\mathbf{t}}_{ij}\widehat{\mathbf{r}}_{ij}^H\mathbf{W}_i)\right)}{\text{Tr}(\widehat{\mathbf{T}}_{ij}\mathbf{W}_i^H\mathbf{W}_i)} \quad (21)$$

$$\sigma^2 = \frac{1}{\sum_i n_i} \sum_i \sum_j \text{Tr}\left\{\tau_{ij}\widehat{\mathbf{T}}_{ij}\mathbf{W}_i^H\mathbf{W}_i + \widehat{\mathbf{R}}_{ij} - 2\sqrt{\tau_{ij}}\Re\left(\text{Tr}(\widehat{\mathbf{t}}_{ij}\widehat{\mathbf{r}}_{ij}^H\mathbf{W}_i)\right)\right\} \quad (22)$$

$$\mathbf{W}_i = \sum_j \sqrt{\tau_{ij}}\widehat{\mathbf{r}}_{ij}\widehat{\mathbf{t}}_{ij}^H \left(\sum_j \widehat{\mathbf{T}}_{ij}\right)^{-H} \quad (23)$$

$$\mathbf{m} = \left(\sum_i \sum_j \mathbf{A}_i^H \Gamma_{ij} \mathbf{A}_i\right)^{-1} \sum_i \sum_j \mathbf{A}_i^H \Gamma_{ij} \widehat{\mathbf{y}}_{ij} \quad (24)$$

As in the fully observed case, the missing data case can be cast into a parallel scheme (see Fig. 1).

4. Numerical simulations

In this section, we aim at evaluating numerically the performance of the proposed algorithm. Specifically, we consider three different scenarios with rank $r_i = 3$, $\mathbf{A}_i = \mathbf{I}$ and $\mathbf{m} \in \mathbb{C}^{10 \times 1}$ following a Gaussian distribution. For the sake of simplicity, $[\mathbf{W}_i]_{1:r_i;1:r_i} = \mathbf{I} + \mathbf{K}$ with $[\mathbf{K}]_{h,q} \sim \mathcal{CN}(0,1)$ and $[\mathbf{W}_i]_{r_i+1:n;1:r_i} = \mathbf{0}$ such that we still have $\text{rank}(\mathbf{W}_i) = r$. In the first scenario, we focus on the mean estimation, \mathbf{m} , by comparing the ECME proposed algorithm with the minimum covariance determinant [16], the Marona's orthogonalized Gnanadesikan-Kettenring (OGK) [17], the Olive Hawkins estimators [15], the classical probabilistic PCA [7] and the robust student-based probabilistic PCA [22]. Fig. 2 and Fig. 3 show the mean square error (MSE) of \mathbf{m} estimates versus the number of observations in which we consider a 0 dB signal-to-noise ratio. Only one individual with $m \geq 2n$ is considered in order to respect restrictions given by the aforementioned algorithms [7, 15, 16, 17, 22]. In Fig. 2, the texture parameter is a realization of an inverse-Gamma distribution (leading to a t-distributed random effect components) to which three strong outliers are added. We note that the proposed ECME and the robust student-based PPCA exhibit comparable performances, which means that the proposed ECME fits well with the robust student-based PPCA which assumes a perfect knowledge of the texture distribution. In Fig. 3, in which the texture parameter is taken as a

realization of a uniform distribution over $]0, 500]$, the proposed ECME outperforms the compared algorithms since it does not assume any distribution of texture while taking into account the low rank structure of the within-individual error. In the second scenario, we focus on the normalized MSE of $\mathbf{W}^H\mathbf{W}$ versus the number of observations. Again, we notice from Fig. 4 that the proposed algorithm outperforms PPCA and the robust student based PPCA regarding subspace inference. Finally, in the third scenario, we consider a multi-individual case and we show the benefit of considering the extension of the proposed algorithm to the scenario of the possible presence of missing data. Specifically, we plot in Fig. 5 the MSE of the mean estimate in the case of three individuals versus the percentage of missing data (i.e. missing values at individuals i and observations j). We notice that the proposed ECME extension shows a good performance since it remains close to the “clairvoyant” ECME which uses the full original data (i.e. observed and unobserved), whereas the ECME using only the observed data in all individuals exhibits a dramatic performance loss when the missing data ratio exceeds 10%.

5. Conclusion

In this paper we design an ECME-based algorithm for mean and covariance estimation under heterogeneous mixed-effects model. The random effects were modelled as heteroscedastic Gaussian process allowing flexibility of the within-individual error and robustness against any possible elliptically symmetric distribution. In addition, the low rank structure of the loading matrix of each individual was taken into account, leading to a natural parallel scheme enhancing the computational cost. Finally, an extension to the missing data scenario was given and numerical simulations assessing the usefulness of the proposed scheme have been presented, showing a performance gain with existing state-of-the-art estimators.

References

- [1] E. Vonesh, *Mixed models: Theory and applications*, Taylor & Francis, 2006.
- [2] M. Pesavento, A. Gershman, M. Wong, Direction finding in partly calibrated sensor arrays composed of multiple subarrays, *IEEE Trans. Signal Process.* 50 (2002) 2113–2115.
- [3] S. Wijnholds, S. Van Der Tol, R. Nijboer, A. van der Veen, Calibration challenges for future radio telescopes, *IEEE Signal Processing Magazine* 1 (2009) 30–42.

- [4] M. Greco, F. Gini, A. Farina, M. Rangaswamy, Doa estimation and multi-user interference in a two-radar system, *Signal processing* 89 (4) (2009) 355–364.
- [5] A. Leshem, A. Van der Veen, Radio-astronomical imaging in the presence of strong radio interference, *IEEE Transactions on Information Theory* 46 (2001) 1730–1747.
- [6] A. Dabbech, C. Ferrari, D. Mary, E. Slezak, O. Smirnov, J. Kenyon, Moresane: Model reconstruction by synthesis-analysis estimators-a sparse deconvolution algorithm for radio interferometric imaging, *Astronomy & Astrophysics* 576 (2015) A7.
- [7] M. Tipping, C. Bishop, Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (1999) 611–622.
- [8] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, P. Larzabal, Covariance structure maximum likelihood estimates in compound gaussian noise: Existence and algorithm analysis, *IEEE Trans. Signal Process.* 56 (2008) 34–48.
- [9] M. Greco, F. Gini, Cramer-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions, *IEEE Trans. Signal Process.* 61 (2013) 6401–6409.
- [10] E. Ollila, D. Tyler, V. Koivunen, H. Poor, Complex elliptically symmetric distributions: Survey, new results and applications, *IEEE Transactions on Signal Processing* 60 (11) (2012) 5597–5625.
- [11] F. Gini, M. V. Greco, Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter, *Signal Process.* 82 (2002) 1847–1859.
- [12] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal and M. Pesavento, Robust calibration of radio interferometers in non-gaussian environment, *IEEE Trans. Signal Process.* 65 (2017) 5649–5660.
- [13] O. Besson, Y. Abramovich, B. Johnson, Direction-of-arrival estimation in a mixture of K-distributed and Gaussian noise, *Signal Process.* 128 (2016) 512–520.
- [14] C. Liu, D. Rubin, The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence, *Biometrika* 81 (1994) 633–648.
- [15] D. Olive, A resistant estimator of multivariate location and dispersion, *Computational Statistics and Data Analysis*, Elsevier 46 (2004) 99–102.
- [16] P. Rousseeuw, K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, Taylor & Francis 41 (1999) 212–223.
- [17] R. Maronna, R. Zamar, Robust estimates of location and dispersion for high dimensional datasets, *Technometrics*, Taylor & Francis 44 (2002) 307–317.
- [18] H. Shen, X. Li, Q. Chen, C. Zeng, G. Yang, H. Li, L. Zhang, Missing information reconstruction of remote sensing data: A technical review, *IEEE Geosci. Remote Sens. Mag.* 3 (2015) 61–85.
- [19] B. Walczak, D. L. Massart, Dealing with missing data : Part I, *Chemom. Intell. Lab. Syst.* 58 (2001) 15–27.
- [20] R. Little, D. Rubin, *Statistical analysis with missing data*, John Wiley & Sons, Wiley Series in Probability and Statistics, 2002.
- [21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1984.
- [22] T. Chen, E. Martin, G. Montague, Robust probabilistic PCA with missing data and contribution analysis for outlier detection, *Computational Statistics & Data Analysis* 53 (2009) 3706–3716.

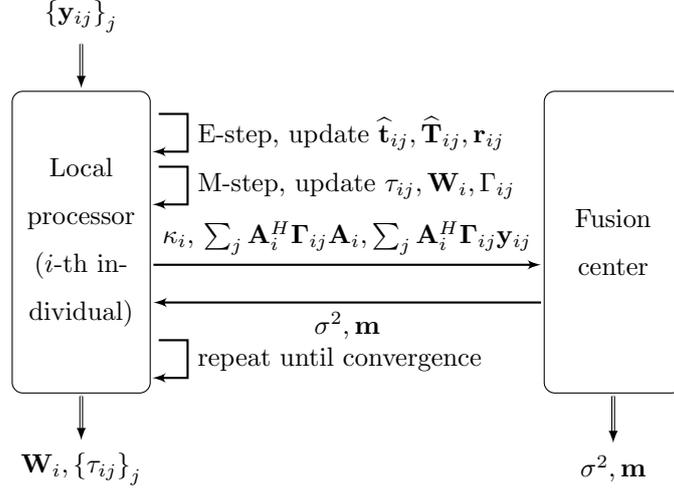


Figure 1: Operation flow and signaling exchange between a local individual's processor and the central processor, in which $\kappa_i = \sum_j \text{Tr}\{\tau_{ij} \hat{\mathbf{T}}_{ij} \mathbf{W}_i^H \mathbf{W}_i + \mathbf{r}_{ij}^H \mathbf{r}_{ij} - 2\sqrt{\tau_{ij}} \Re(\mathbf{r}_{ij}^H \mathbf{W}_i \hat{\mathbf{t}}_{ij})\}$. The consensus step is enforced for deriving the common mean and homogeneous noise power.

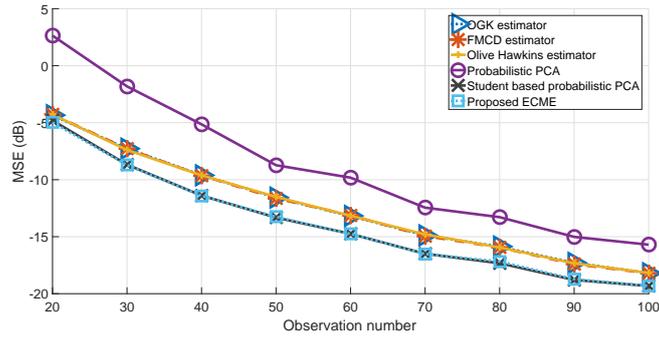


Figure 2: Mean square error vs. number of observations under t-distributed data (1000 runs mean).

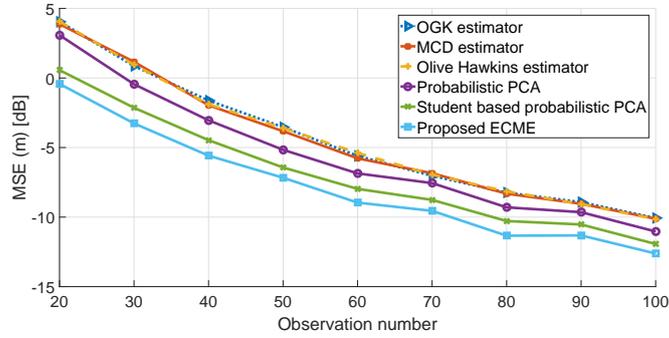


Figure 3: Mean square error vs. number of observations under a uniform distributed within-individual error (1000 runs mean).

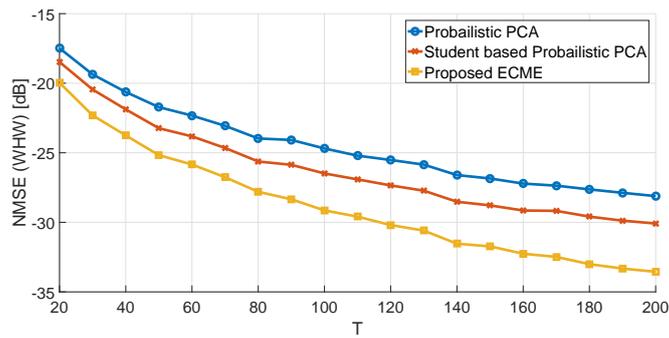


Figure 4: Normalized mean square error vs. number of observations under a uniform distributed within-individual error (1000 runs mean).



Figure 5: Mean square error vs. percentage of unobserved data (1000 runs mean).