Canonical Polyadic Tensor Decomposition in the Presence of Non Gaussian Noise

Rodrigo Cabral Farias, Pierre Comon

CANONICAL POLYADIC TENSOR DECOMPOSITION IN THE PRESENCE OF NON GAUSSIAN NOISE

Rodrigo Cabral Farias, Pierre Comon

GIPSA-Lab, CNRS UMR5216, Grenoble Campus, St. Martin d'Hères, 38402

ABSTRACT

In this paper we describe an estimator for the canonical polyadic (CP) tensor model using order statistics of the residuals. The estimator minimizes in an iterative and alternating fashion a dispersion function given by the weighted ranked absolute residuals. Specific choices of the weights lead to either equivalent or approximate versions of the least squares estimator, least absolute deviation estimator or least trimmed squares estimators. For different noise distributions, we present simulations comparing the performance of the proposed algorithm with the standard least squares estimator. The simulated performance is equivalent in the Gaussian noise case and superior when the noise is distributed according to the Laplacian or Cauchy distributions.

***Index Terms*—** Tensor decomposition, order statistics, non Gaussian noise, robust estimation.

1. INTRODUCTION

In recent years, data models have been extended from the standard one dimensional (vector) and two dimensional models (matrices) to n-dimensional models. A straightforward n-dimensional extension of linear and bilinear models is that of multilinear models. One of the most widely used multilinear model is the canonical polyadic (CP) model, also called the Candecomp/Parafac model, which has been applied extensively in psychometrics [1], chemometrics [2], and more recently, in many other domains such as digital communications [3], array processing [4] and data mining [5].

If the data array is noisy, then finding the CP model is equivalent to a multivariate parameter estimation problem, whose solution can be obtained through maximum likelihood (ML) estimation. In most applications, the noise is considered to be additive and Gaussian, thus leading to a least squares problem. Simple gradient descent or alternating least squares (ALS) techniques [6] can be used to obtain estimates of the parameters and, to assess estimation efficiency, their variances can be compared with the Cramér-Rao bound (CRB) [7]. Even though the Gaussian assumption leads to simple estimation procedures, it leads to poor performances in the presence of outliers [8], [9]. Outliers and impulsive noise can be dealt with in two ways: either by considering robust estimation procedures [10], which disregard the noise distribution, or by assuming that the noise follows a known non Gaussian distribution, with a tail heavier than the Gaussian tail, for example, the Laplacian or Cauchy distributions.

In the robust setting, CP model estimation can be tackled by applying known robust regression techniques in an alternating fashion. For example, in [11], alternating iteratively reweighted least squares (IRLS) using Huber influence functions was considered. In [12] alternating least absolute deviation (LAD) regression was proposed, while in [13] a robust PCA method was adapted to CP estimation. Alternating LAD regression was revisited in [14] and [15]. In [14], each alternating step was evaluated through a majorization-minimization technique, and in [15] an IRLS algorithm with the Tukey loss function and sparse regularization was presented.

In this paper, we propose an alternating regression approach based on order statistics for the estimation of the CP model under non Gaussian additive noise. Instead of focusing on the LAD cost function, we minimize the dispersion of the absolute residuals, which is a weighted sum of the absolute residuals. The weight assigned to each residue depends on its absolute value rank. Since this cost function can be rewritten as a sum of weighted quadratic functions (with non constant weights), we propose an alternating regression scheme based on IRLS, where the reweighting depends on the absolute residuals and their ranks. The main advantage of this approach is that, if the noise distribution is known, it can achieve equal or better performance than LAD regression by careful choice of the weights. Moreover, in the case of known noise distribution and if the data array size is sufficiently large, the weights of the IRLS procedure can be approximated by constants depending on the noise distribution and on the residues rank, not on their values. This leads to a general and simple least squares based alternating procedure, where the influence of the noise distribution is taken into account only through constant predefined weights and not through a specific non linear function (as in M-estimation for example).

The paper is organized as follows: in Section 2 we present the CP model estimation problem and its standard solution in the Gaussian noise case. In Section 3 the alternating robust estimator based on order statistics and its simplified version

This work was funded by the FP7 European Research Council Programme, DECODA project, under grant ERC-AdG-2013-320594.

are presented. Simulations comparing its performance with ALS, LAD regression, and the Cramér-Rao bounds for Gaussian, Laplacian and Cauchy noise are presented in Section 3. Finally, conclusions are drawn in Section 5.

In this paper the following notations are used: scalars and vectors are denoted by lower case x and bold lower case $\boldsymbol{x}$ letters respectively. Matrices are denoted by upper case bold letters $\boldsymbol{X}$, while tensors by calligraphic letters $\boldsymbol{\mathcal{X}}$. Elements of a given array are indicated by subscripts $\boldsymbol{\mathcal{X}}_{ijk}$. The Khatri-Rao product of two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ (column-wise Kronecker product) is denoted by $\boldsymbol{X} \odot \boldsymbol{Y}$ and the tensor product of three vectors or vector outer product by $\boldsymbol{x} \otimes \boldsymbol{y} \otimes \boldsymbol{z}$.

2. CP ESTIMATION AND ALS

The CP model of rank R of a three dimensional real data array of size $I \times J \times K$, is given by (see *e.g.* [16]):

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r, \tag{1}$$

where $\boldsymbol{a}_r$, $\boldsymbol{b}_r$ and $\boldsymbol{c}_r$ are real vectors of parameters. We assume that the measured data array $\boldsymbol{\mathcal{Y}}$ is corrupted by additive independent and identically distributed (i.i.d.) noise

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{V}}, \tag{2}$$

and denote by $f(\mathcal{V}_{ijk})$ the probability density function (PDF) of the noise. The goal is to estimate $\boldsymbol{a}_r$, $\boldsymbol{b}_r$ and $\boldsymbol{c}_r$ based on $\mathcal{Y}$. To solve this problem, we can resort to ML estimation. If we stack all parameters in a single vector $\boldsymbol{\theta}$ then the ML estimator $\hat{\boldsymbol{\theta}}$ is $\arg\max_{\boldsymbol{\theta}}[\mathcal{L}(\boldsymbol{\mathcal{Y}}; \boldsymbol{\mathcal{X}})]$ where $\mathcal{L}(\boldsymbol{\mathcal{Y}}; \boldsymbol{\mathcal{X}})$ is the log-likelihood function

$$\mathcal{L}(\boldsymbol{\mathcal{Y}}; \boldsymbol{\mathcal{X}}) = \sum_{i,j,k}^{I,J,K} \log f[\boldsymbol{\mathcal{Y}}_{ijk} - \boldsymbol{\mathcal{X}}_{ijk}(\boldsymbol{\theta})]. \tag{3}$$

In most applications, the noise is assumed to be Gaussian. As a consequence, $\log f[\boldsymbol{\mathcal{Y}}_{ijk} - \boldsymbol{\mathcal{X}}_{ijk}(\boldsymbol{\theta})] = -\|\boldsymbol{\mathcal{Y}}_{ijk} - \boldsymbol{\mathcal{X}}_{ijk}(\boldsymbol{\theta})\|^2$. If the vectors $\boldsymbol{a}_r$, $\boldsymbol{b}_r$ and $\boldsymbol{c}_r$ are stacked in factor matrices $\boldsymbol{A} = [\boldsymbol{a}_1 \cdots, \boldsymbol{a}_R]$, $\boldsymbol{B} = [\boldsymbol{b}_1 \cdots, \boldsymbol{b}_R]$ and $\boldsymbol{C} = [\boldsymbol{c}_1 \cdots, \boldsymbol{c}_R]$ and if the data array is unfolded into a matrix in one of its three modes (dimensions of the tensor) $\boldsymbol{Y}^{(1)}$, $\boldsymbol{Y}^{(2)}$ or $\boldsymbol{Y}^{(3)}$, then the ML estimator is given as the minimizer of the following cost function

$$\begin{aligned} \Upsilon(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) &= \|\boldsymbol{Y}^{(1)} - \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^\top\|_F^2, \\ &= \|\boldsymbol{Y}^{(2)} - \boldsymbol{B}(\boldsymbol{C} \odot \boldsymbol{A})^\top\|_F^2, \\ &= \|\boldsymbol{Y}^{(3)} - \boldsymbol{C}(\boldsymbol{B} \odot \boldsymbol{A})^\top\|_F^2, \end{aligned} \tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm. There are multiple ways of solving this minimization problem. We can find a local minimum by applying a general purpose joint descent method (*e.g.* a gradient method) with respect to the three factors or we can apply a block descent method. In the second approach, if we choose the blocks to be exactly the factor matrices, then each update is the solution of a linear least squares problem. For iterate k of block descent we have estimates

$$\begin{aligned} \hat{\mathbf{A}}_k &= \boldsymbol{Y}^{(1)}(\hat{\boldsymbol{C}}_{k-1} \odot \hat{\boldsymbol{B}}_{k-1})^\dagger, \\ \hat{\boldsymbol{B}}_k &= \boldsymbol{Y}^{(2)}(\hat{\boldsymbol{C}}_{k-1} \odot \hat{\mathbf{A}}_k)^\dagger, \\ \hat{\boldsymbol{C}}_k &= \boldsymbol{Y}^{(3)}(\hat{\boldsymbol{B}}_k \odot \hat{\mathbf{A}}_k)^\dagger, \end{aligned} \tag{5}$$

where † denotes the pseudoinverse. This is the standard ALS algorithm. Contrary to joint descent methods, it does not enjoy nice local convergence properties, but is still widely used because of its easy implementation. For these reasons, in the next section, we will modify this method to tackle robust estimation and other types of noise.

3. ALTERNATING ROBUST ESTIMATOR BASED ON ORDER STATISTICS

In general, when the noise is not Gaussian, the block updates do not have analytical expressions and we cannot proceed as in ALS. In [17], the problem of estimation of a location parameter from i.i.d data was cast as a weighted least squares, by ordering the samples and evaluating the best linear unbiased estimator of location, based on mean and covariance of the ordered data. Its extension to linear regression was proposed as the minimization of a dispersion function, which takes into account the order of the residuals [18], [19]. If we consider the $\hat{\mathbf{A}}_k$ update, we can minimize the following dispersion:

$$\Upsilon(\boldsymbol{A}) = \sum_{i,j}^{I,J\times K} w\left[\text{rk}\{|\boldsymbol{\mathcal{E}}_{ij}^{(1)}(\boldsymbol{A})|\}\right] \, |\boldsymbol{\mathcal{E}}_{ij}^{(1)}(\boldsymbol{A})|, \tag{6}$$

where $\boldsymbol{\mathcal{E}}^{(1)} = \boldsymbol{Y}^{(1)} - \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^\top$, and $w[i], i \in \{1, .., IJK\}$, are nonnegative coefficients depending on the ranks of the absolute residuals when ordered in increasing order, $\text{rk}\{|\boldsymbol{\mathcal{E}}_{ij}^{(1)}(\boldsymbol{A})|\}$. Remark that this dispersion function is a norm [20] and it is also a convex (non differentiable) function. Therefore, if the goal was not to develop a simple modification of ALS, we could use more complex convex optimization algorithms to solve each block update.

Choice of $w[i]$. A common choice is the Wilcoxon score $w[i] = [i/(IJK + 1)]$ [10, p. 61]. When the goal is to be robust against outliers, then $w[i]$ can be set to zero for large i:

$$w[i] = \begin{cases} 1, & \text{if } i \leq IJK - h, \\ 0, & \text{if } i > IJK - h. \end{cases} \tag{7}$$

This leads to least trimmed squares [21]. If the noise distribution is known, then it is possible to weight the residuals using the score function for the estimation of a location parameter

$$w[i] = f'(F^{-1}(t_i))/f(F^{-1}(t_i)), \tag{8}$$

where $f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x}$, $F^{-1}(x)$ is the inverse cumulative distribution function and $t_i = \frac{i}{2(IJK+1)} + \frac{1}{2}$. This is equivalent to weight the absolute residuals according to the sensitivity of the log-likelihood. However, instead of using the true residual values, we use their corresponding quantile values. When Υ depends on the ranked residuals and not on the ranked absolute values, this choice is asymptotically equivalent to ML estimation, and hence asymptotically efficient [18, 19].

IRLS approach: we can rewrite the minimization problem separately for each row of matrix $\boldsymbol{A}$. Denoting the i-th row of $\boldsymbol{A}$ and $\boldsymbol{Y}^{(1)}$ both as column vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{y}_i^{(1)}$, the cost function for the corresponding row becomes

$$\Upsilon'(\boldsymbol{\alpha}_i) = \sum_{j=1}^{J\times K} w_{ij}(\boldsymbol{\alpha}_i) \left| [\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_j \right|, \qquad (9)$$

where $w_{ij}(\boldsymbol{\alpha}_i) = w\left[\mathrm{rk}\{|\boldsymbol{\mathcal{E}}_{ij}^{(1)}(\boldsymbol{\alpha}_i)|\}\right]$ and $\boldsymbol{M_A} = \boldsymbol{C} \odot \boldsymbol{B}$. If we divide and multiply by the residuals we obtain

$$\Upsilon'(\boldsymbol{\alpha}_i) = \sum_{j=1}^{J\times K} \frac{w_{ij}(\boldsymbol{\alpha}_i)}{|[\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_j|} [\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_j^2. \quad (10)$$

This cost function can be written as a quadratic cost function

$$\Upsilon'(\boldsymbol{\alpha}_i) = [\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]^\top \boldsymbol{W}(\boldsymbol{\alpha}_i)[\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i], \quad (11)$$

where the weighting matrix $\boldsymbol{W}(\boldsymbol{\alpha}_i)$ is diagonal

$$\boldsymbol{W}(\boldsymbol{\alpha}_i) = \mathrm{diag}\left[\frac{w_{i1}(\boldsymbol{\alpha}_i)}{|[\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_1|} \cdots \frac{w_{iJ\times K}(\boldsymbol{\alpha}_i)}{|[\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_{J\times K}|} \right]$$

Since the weights depend on $\boldsymbol{\alpha}_i$, we cannot minimize $\Upsilon(\boldsymbol{\alpha}_i)$ directly using least squares. However, if we have a good initial guess $\hat{\boldsymbol{\alpha}}_i^{l-1}$, we can approximate the weights by $\boldsymbol{W}(\hat{\boldsymbol{\alpha}}_i^{l-1})$ and then use the update formula

$$\hat{\boldsymbol{\alpha}}_i^{l} = [\boldsymbol{M_A}^\top \boldsymbol{W}(\hat{\boldsymbol{\alpha}}_i^{l-1})\boldsymbol{M_A}]^{-1} \boldsymbol{M_A}^T \boldsymbol{W}(\hat{\boldsymbol{\alpha}}_i^{l-1})\boldsymbol{y}_i^{(1)}. \quad (12)$$

These approximations can be iterated, leading to an IRLS based solution for the block updates. This leads to the alternating IRLS approach decribed in **Alg. 1**, which we call alternating robust estimator (ARE).

ARE approximation: if we know the noise distribution, we can simplify the IRLS procedure further by approximating the denominator of the weights by the quantile corresponding to the absolute residual

$$\frac{w_{ij}(\boldsymbol{\alpha}_i)}{|[\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_j|} \approx \frac{w_{ij}(\boldsymbol{\alpha}_i)}{F^{-1}\left(t_{\mathrm{rk}\{|[\boldsymbol{y}_i^{(1)} - \boldsymbol{M_A}\boldsymbol{\alpha}_i]_j|\}}\right)}. \quad (13)$$

The error of this approximation is expected to be small when the estimation error is small and the number of samples is large, so that the denominator is close to the i-th quantile of the absolute value of the noise.

Algorithm 1 ARE

Require: $\boldsymbol{Y}^{(1)}$, $\boldsymbol{Y}^{(2)}$, $\boldsymbol{Y}^{(3)}$ and initial $\hat{\mathbf{A}}_0$, $\hat{\mathbf{B}}_0$ and $\hat{\mathbf{C}}_0$.
1: k:=0,
2: **while** Υ (6) still decreases **do**
3: Update the rows of $\hat{\mathbf{A}}_k$ with IRLS (12) using $\hat{\mathbf{B}}_{k-1}$ and $\hat{\mathbf{C}}_{k-1}$,
4: Update the rows of $\hat{\mathbf{B}}_k$ with IRLS using $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{C}}_{k-1}$,
5: Update the rows of $\hat{\mathbf{C}}_k$ with IRLS using $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{B}}_k$,
6: k:=k+1.
7: **end while**
8: **return** $\hat{\mathbf{A}}_k$, $\hat{\mathbf{B}}_k$ and $\hat{\mathbf{C}}_k$.

Since the approximate weights do not depend directly on the values of the residuals, they can be calculated in advance and stored in a table. If we consider $w[i]$ given by the scores (8), then we can interpret how the residuals are dealt with for different noise distributions. We shall analyze three cases.

1. An example of impulsive noise distribution is the Laplacian: $f_L(x) = (1/2\delta^2))\exp(-|x|/\delta^2)$. For this type of noise the scores are constant. Since the quantiles increase in i, the weights decrease for large residuals, showing that most information is around the median.

2. For Gaussian noise $f_G(x) = (1/\sqrt{\pi\delta^2})\exp(-x^2/\delta^2)$, the scores are linear on the quantiles. As a consequence, large residuals have the same importance as small ones, since the weights are constant. As expected, no gain in performance should be obtained with ARE, as its iterates are exactly equal to ALS in this case.

3. Finally, for the heavy-tailed Cauchy distribution, $f_C(x) = (1/\pi\delta^2)\{1/[1 + (x/\delta)^2]\}$, the weights are almost constant for small residuals and almost zero for large residuals, showing that large residuals are completely discarded.

4. SIMULATIONS

In this section, we show computer results for the ARE approach and its version with approximate weights (labelled as ARE approx.). We start showing the overall effect of impulsive noise both on ALS and ARE realizations. We generate a random CP model of size $(I, J, K) = (10, 10, 10)$ with $R = 2$. The CP model is then corrupted by Cauchy noise with $\delta = 0.005$. We run 10 realizations of ALS with different random initializations. After convergence, we select the result giving the least Υ and use it to initialize the ARE and its approximation, both with weights given by the scores. At each block update, we execute only one iterate of the IRLS procedure. The results for the estimation of the factor C are given in Fig. 1. For ALS, we can see that the impulsive noise spreads through all elements of the factor, while such a problem does not appear for ARE.

In what follows we simulate both algorithms under the three noise distributions previously introduced: Laplacian, Gaussian and Cauchy. In the three cases, we compare the results with the ARE approximation for the Laplacian distribution, denoted by LAD approx., since this is equivalent to LAD estimation which is known to be robust in CP estima-

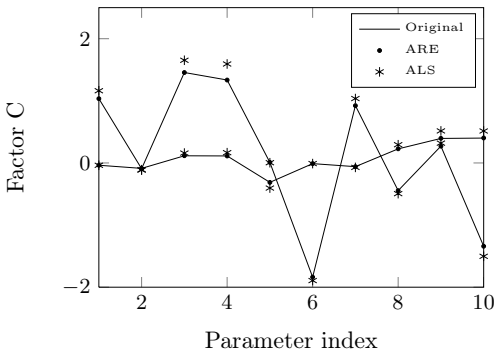

Fig. 1. Realization of the ALS and ARE estimators for the factor C of a rank 2 CP model with dimensions $10 \times 10 \times 10$. The data array is corrupted by Cauchy noise.

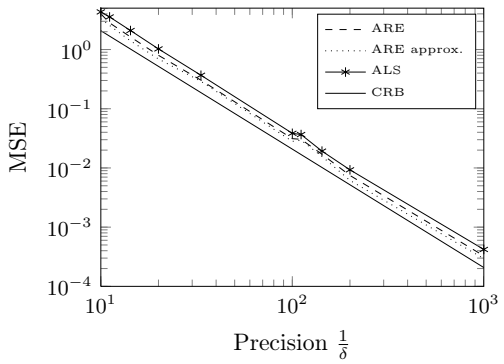

Fig. 2. Total MSE for the estimation of a rank 4 CP model with dimensions $10 \times 10 \times 10$. The data array is corrupted by Laplacian noise.

tion under impulsive non Gaussian noise [12].

The CP model size is $(I, K, K) = (10, 10, 10)$ with $R = 4$. Different precision values $\frac{1}{\delta}$ are considered. For each precision, two random CP models are simulated and 100 different realizations of the noise are generated. ALS is initialized 10 times for each realization and ARE is initialized with the best ALS realization. The realizations of the estimators are used to evaluate the sum of the averaged squared errors of all parameters, which is an approximation of the total mean squared error (MSE). The evaluation of the squared error is carried out after correction of the permutation and scaling ambiguities, using the true CP model.

To have a clear view of the efficiency of the estimators we compare their performance with the trace of Cramér-Rao bound (CRB) $\mathrm{tr}\left[\mathrm{CRB}(\boldsymbol{\theta})\right]$, which is a lower bound on the total MSE. Since two tensors are generated, the performance is compared to the averaged CRB. The CRB for CP model estimation under specific cases of additive non Gaussian (Laplacian and Cauchy) noise was studied in [12] where it was shown that the main difference with the Gaussian case is a scalar factor. In Appendix A, we show that this is true for any distribution and that the scalar factor is the CRB for the estimation of a scalar parameter under non Gaussian noise.

In Fig. 2 the results for the Laplacian distribution are shown. We can observe that the ARE and ARE approx. perform better than ALS. This result was already expected, since ARE is the alternating IRLS version of the maximum likelihood estimator (the LAD estimator), which in theory is expected to be efficient for large data arrays.

Fig. 3 shows the results for the Gaussian distribution. Since both ARE methods are strictly equivalent to ALS we obtain the same results for all these estimators. Note also that their performance is very close to the CRB. The LAD estimator gives the worst performance since it rejects some information contained in the large residuals.

The performance results for the Cauchy distribution are presented in Fig. 3. We can confirm once more, the well-known fact that LAD based estimators are more robust than least squares estimators, when dealing with impulsive noise. This fact was already verified for CP model estimation in [12]. However, we can note that if we have information on the noise

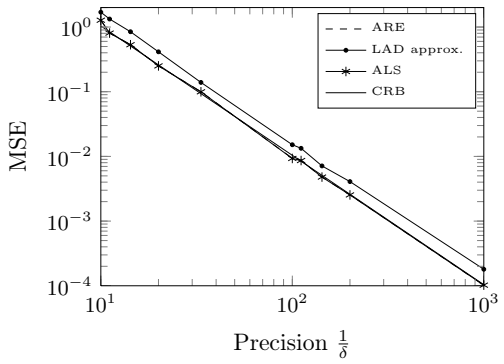

Fig. 3. Total MSE for the estimation of a rank 4 CP model with dimensions $10 \times 10 \times 10$. The data array is corrupted by Gaussian noise.

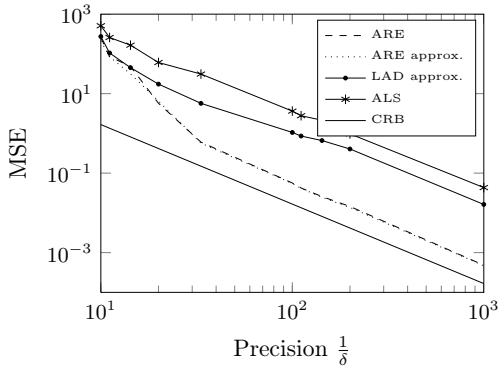

Fig. 4. Total MSE for the estimation of a rank 4 CP model with dimensions $10 \times 10 \times 10$. The data array is corrupted by Cauchy noise.

distribution (through its scores), we can increase the performance even further using ARE, without increasing substantially the complexity.

5. CONCLUSIONS

We have proposed an alternating regression estimator of CP model parameters, based on order statistics. The algorithm is a simple weighted ALS, whose weights depend on the ranks of the residuals. By using information about the noise distribution, we have shown that the algorithm can be superior or equivalent to standard approaches such as ALS and alternating LAD. We focused on a parametric approach, where the noise distributions are known. In future work, we can focus on nonparametric approaches, for example, using an adaptive strategy to estimate jointly the CP model and the score function used to evaluate the weights [22].

A. CRAMER-RAO BOUND FOR THE CP MODEL UNDER NON GAUSSIAN ADDITIVE NOISE

The Cramér-Rao bound is given by the inverse of the Fisher information (FI) matrix $\boldsymbol{I}$, whose elements are given by

$$\boldsymbol{I}_{ij} = \mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_i}\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_j}\right], \tag{14}$$

For two elements of the factor $\boldsymbol{A}$, say $\boldsymbol{A}_{ir}$ and $\boldsymbol{A}_{i'r'}$, the corresponding element of the FI matrix is

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}_{ir}}\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}_{i'r'}}\right] = \sum_{jkj'k'} \frac{\partial \boldsymbol{\mathcal{X}}_{ijk}}{\partial \boldsymbol{A}_{ir}}\frac{\partial \boldsymbol{\mathcal{X}}_{i'jk}}{\partial \boldsymbol{A}_{i'r'}}\mathbb{E}\left[\frac{f'(\boldsymbol{\mathcal{E}}_{ijk})}{f(\boldsymbol{\mathcal{E}}_{ijk})}\frac{f'(\boldsymbol{\mathcal{E}}_{i'j'k'})}{f(\boldsymbol{\mathcal{E}}_{i'j'k'})}\right]$$

where $\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}$. Since the noise samples are independent and $\mathbb{E}\left[\frac{f'(\boldsymbol{\mathcal{E}}_{ijk})}{f(\boldsymbol{\mathcal{E}}_{ijk})}\right] = 0$, we have

$$\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}_{ir}}\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}_{i'r'}}\right] = \begin{cases} 0, & \text{if, } i' \neq i, \\ I_f \sum\limits_{jkj'k'} \boldsymbol{B}_{jr}\boldsymbol{C}_{kr}\boldsymbol{B}_{j'r}\boldsymbol{C}_{k'r}, & \text{otherwise,} \end{cases} \tag{15}$$

where $I_f = \int_{\mathbb{R}}[f'(v)]^2/f(v)\,\mathrm{d}v$ is the FI for the estimation of a location parameter of a distribution $f(y-x)$. Similar expressions can be found for the FI related to factors $\boldsymbol{B}$ and $\boldsymbol{C}$. Using the independence assumption and the fact that the score function has zero mean, we obtain the following crossed FI elements related to factors $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\begin{aligned}\mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}_{ir}}\frac{\partial \mathcal{L}}{\partial \boldsymbol{B}_{jr'}}\right] &= \sum_{jki'k'} \frac{\partial \boldsymbol{\mathcal{X}}_{ijk}}{\partial \boldsymbol{A}_{ir}}\frac{\partial \boldsymbol{\mathcal{X}}_{i'jk'}}{\partial \boldsymbol{B}_{jr'}}\mathbb{E}\left[\frac{f'(\boldsymbol{\mathcal{E}}_{ij'k})}{f(\boldsymbol{\mathcal{E}}_{ij'k})}\frac{f'(\boldsymbol{\mathcal{E}}_{i'jk'})}{f(\boldsymbol{\mathcal{E}}_{i'jk'})}\right] \\ &= I_f \boldsymbol{A}_{ir}\boldsymbol{B}_{jr'}\sum_k \boldsymbol{C}_{kr}\boldsymbol{C}_{kr'}. \end{aligned} \tag{16}$$

The crossed FI elements related to $\boldsymbol{A}$ and $\boldsymbol{C}$ and to $\boldsymbol{B}$ and $\boldsymbol{C}$ have similar expressions. Both expressions (15) and (16) indicate that the only difference between FI matrices for different noise distributions is the scale factor I_f, which is related to the CRB for the estimation of a scalar parameter x of a distribution $f(y-x)$. This fact was partially shown in [12] and it can be understood as a particular case of the behavior of the CRB in additive noise models [23] (cited in [12]).

Acknowledgment

The authors would like to thank Souleymen Sahnoun for his help in the evaluation of the performance bounds.

REFERENCES

[1] R. A. Harshman, "Foundations of the Parafac procedure," *UCLA papers in phonetics*, vol.16, pp. 1-84, 1970.

[2] A. Smilde, R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*, John Wiley & Sons, 2005.

[3] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Sig. Proc.*, vol. 48, no. 3, pp. 810–823, 2000.

[4] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Sig. Proc.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[5] M. Mørup, "Applications of tensor factorizations and decompositions in data mining," *Data Mining and Knowledge Discovery*, Wiley, vol. 1, no. 1, pp. 24–40, 2011.

[6] P. Comon, X. Luciani, and A. L. F. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.

[7] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Sig. Proc.*, vol. 49, no. 9, pp. 2074–2086, 2001.

[8] S. F. Møller, J. von Frese, and R. Bro, "Robust methods for multivariate data analysis," *Journal of Chemometrics*, vol. 19, no. 10, pp. 549–563, 2005.

[9] S. Engelen and M. Hubert, "Detecting outlying samples in a parallel factor analysis model," *Analytica chimica acta*, vol. 705, no. 1, pp. 155–165, 2011.

[10] P. J. Huber, *Robust statistics*, Springer, 2011.

[11] P. Paatero, "Least squares formulation of robust non-negative factor analysis," *Chemo. Intel. Lab. Syst.*, vol. 37, no. 1, pp. 23–35, 1997.

[12] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Sig. Proc.*, vol. 53, no. 8, pp. 2678–2689, 2005.

[13] M. Hubert, J. Van Kerckhoven, and T. Verdonck, "Robust PARAFAC for incomplete data," *Journal of Chemometrics*, vol. 26, no. 6, pp. 290–298, 2012.

[14] E. C. Chi and T. G. Kolda, "Making tensor factorizations robust to non-Gaussian noise," *arXiv:1010.3043*, 2010.

[15] H.-J. Kim, E. Ollila, V. Koivunen, and H. V. Poor, "Robust iteratively reweighted lasso for sparse tensor factorizations," in *IEEE Work. on Stat. Sig. Proc. (SSP)*, 2014, pp. 420–423.

[16] P. Comon, "Tensors: a brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, 2014.

[17] E. H. Lloyd, "Least-squares estimation of location and scale parameters using order statistics," *Biometr.*, pp. 88–95, 1952.

[18] J. Jureckova, "Nonparametric estimate of regression coefficients," *The Annals Math. Stat.*, pp. 1328–1338, 1971.

[19] L. A. Jaeckel, "Estimating regression coefficients by minimizing the dispersion of the residuals," *The Annals Math. Stat.*, pp. 1449–1458, 1972.

[20] J. W. McKean and R. M. Schrader, "The geometry of robust procedures in linear models," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 366–371, 1980.

[21] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, vol. 589, John Wiley & Sons, 2005.

[22] J. D. Naranjo and J. W. McKean, "Rank regression with estimated scores," *Statistics & probability letters*, vol. 33, no. 2, pp. 209–216, 1997.

[23] A. Swami, "Cramér-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Processing*, vol. 53, no. 2, pp. 231–244, 1996.