2004 IEEE 5th Workshop on Signal Processing Advances in Wireless Communications

Robust Fitting of Multilinear Models with Application to Blind Multiuser Receivers: Iterative Weighted Median Filtering Approach

Sergiy A. Vorobyov, Yue Rong
Dept. of Comm. Systems
University of Duisburg-Essen
Bismarckstr. 81, D-47057
Duisburg, Germany
Email: svor@sent5.uni-duisburg.de

Nicholas D. Sidiropoulos
Dept. of ECE
Technical University of Crete
Kounoupidiana Campus
Chania, 731 00 Crete, Greece
Email: nikos@telecom.tuc.gr

Alex B. Gershman[1]
Dept. of ECE
McMaster University
1280 Main Str. West, L8S 4K1
Hamilton, Ontario, Canada
Email: gershman@ieee.org

Abstract — **PARAllel FACtor (PARAFAC) analysis is an extension of low-rank matrix decomposition to higher-way arrays. It decomposes a given array in a sum of multilinear terms. PARAFAC analysis generalizes and unifies common array processing models, like joint diagonalization and ESPRIT. The prevailing fitting algorithm in all these applications is based on Alternating Least Squares (ALS) optimization, which is matched to Gaussian noise. In many cases, however, measurement errors are far from being Gaussian. An iterative algorithm for least absolute error (robust) fitting of general multilinear models based on Linear Programming (LP) has been recently developed. However, the computational complexity of this method remains high. In this paper, we develop a new iterative algorithm for robust fitting of multilinear models based on iterative Weighted Median Filtering (WMF), which is appealing from a simplicity viewpoint. Performance of the proposed method is illustrated with application to the blind multiuser separation-detection problem, and compared to the performance of Trilinear Alternating Least Squares (TALS), Trilinear Alternating Least Absolute Error based on Linear Programming (TALAE-LP), and the pertinent Cramér-Rao Bounds (CRBs) in Laplacian, Cauchy, and Gaussian noise environments.**

I. Introduction

The PARAFAC model is a useful data analysis tool that has recently found applications in array signal processing and communications, e.g., [1], [2]. Generalizing the concept of low-rank decomposition to higher-way arrays or *tensors*, PARAFAC is instrumental in the analysis of data arrays indexed by three or more independent variables, just like Singular Value Decomposition (SVD) is instrumental in ordinary matrix (two-way array) analysis. Unlike SVD, PARAFAC does not impose orthogonality constraints; the reason is that low-rank decomposition of higher-order tensorial data is essentially unique, unlike low-rank matrix decomposition.

[1]The work of Alex Gershman has been supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Premier Research Excellence Award Program of the Ministry of Energy, Science, and Technology (MEST) of Ontario, and Wolfgang Paul Award Program of the Alexander von Humboldt Foundation. Nicholas Sidiropoulos has been supported by the Army Research Laboratory through participation in the ARL-CTA under Cooperative Agreement DADD19-01-2-0011.

Because of its direct link to low-rank decomposition, PARAFAC analysis has found applications in numerous and diverse disciplines [1], [2]. In most applications of PARAFAC analysis, an alternating least squares regression procedure is used to fit the model parameters (e.g., cf. [1], [3]). Least Squares (LS) regression is optimal (in the maximum likelihood sense) when measurement errors are additive i.i.d. Gaussian. However, in many applications [4] the measurement errors are far from being Gaussian random variables.

The Least Absolute Error (LAE) criterion is often used as a robust alternative to LS. LAE regression is optimal (in the maximum likelihood sense) when measurement errors are additive i.i.d. Laplacian (e.g., see [5]). The Laplacian distribution is more heavy-tailed than the Gaussian one; therefore, it is better suited to model impulsive noise and outliers. An easy way to see this is to consider mean estimation under LS and LAE criteria. These correspond to arithmetic mean and median operators, respectively. The median operator rejects impulses regardless of strength[1]; whereas the arithmetic mean is skewed by even one outlying sample. It is therefore of interest to develop PARAFAC regression procedures that optimize the LAE fitting criterion.

One such procedure has been recently proposed in [6] and it is based on Linear Programming (LP). However, the computational complexity of this procedure remains high. In this paper, we develop a new and simpler iterative procedure which makes use of Weighted Median Filtering (WMF) [7]. The relative merits of these two LAE model fitting algorithms, as well as the standard iterative LS algorithm, are investigated numerically with application to the blind multiuser separation-detection problem and compared to the pertinent Cramér-Rao bounds, which are also provided herein.

II. Parallel Factor Analysis

We introduce the notation that will be useful in the sequel. Consider an $I \times J \times K$ *three-way array* $\underline{\mathbf{X}}$ with typical element $x_{i,j,k}$ and the F-*component trilinear decomposition*

$$x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \qquad (1)$$

for all $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Here $a_{i,f}$ stands for the (i,f)-th element of $I \times F$ matrix $\mathbf{A}$, and similarly $b_{j,f}$ and $c_{k,f}$ stand for (j,f)-th and (k,f)-th elements of $J \times F$ and $K \times F$ matrices $\mathbf{B}$ and $\mathbf{C}$, respectively. Matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are in general complex-valued. Equation (1) expresses $x_{i,j,k}$

[1]Up to roughly $K/2$ impulses can be rejected, where K is the sample size.

as a sum of F rank-1 triple products; it is known as trilinear decomposition, or *PARAFAC analysis* of $x_{i,j,k}$.

Let $\mathbf{A}_i = \mathcal{D}_i(\mathbf{A})$ denote the operator which takes the i-th row of matrix $\mathbf{A}$ and produces a diagonal matrix by placing this row on the main diagonal. Then by "slicing" the three-dimensional array $\underline{\mathbf{X}}$ in a series of "slabs" (two-dimensional arrays) we obtain

$$\mathbf{X}_i = \mathbf{B}\mathbf{A}_i\mathbf{C}^T, \qquad i = 1, \ldots, I \tag{2}$$

Here such a slicing is made perpendicular to the ith dimension, i.e., $\mathbf{X}_i := [x_{i,:,:}]$ is the $J \times K$ two-dimensional slice of $\underline{\mathbf{X}}$ corresponding to the given index i. Two other types of slicing of $\underline{\mathbf{X}}$ are useful in understanding the algorithm that will be developed in the next section. They are given by

$$\mathbf{Y}_j = \mathbf{C}\mathbf{B}_j\mathbf{A}^T, \qquad j = 1, \ldots, J \tag{3}$$

$$\mathbf{Z}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}^T, \qquad k = 1, \ldots, K \tag{4}$$

where $\mathbf{B}_j = \mathcal{D}_j(\mathbf{B})$, $\mathbf{C}_k = \mathcal{D}_k(\mathbf{C})$, while the $K \times I$ matrix $\mathbf{Y}_j$ and $I \times J$ matrix $\mathbf{Z}_k$ are defined as $\mathbf{Y}_j := [x_{:,j,:}]$ and $\mathbf{Z}_k := [x_{:,:,k}]$, respectively.

III. Trilinear Alternating Least Absolute Error Regression based on Weighted Median Filtering (TALAE-WMF)

In practice, the three-way array will contain measurement noise, i.e. $\underline{\widetilde{\mathbf{X}}} = \underline{\mathbf{X}} + \underline{\mathbf{V}}$, where the (i,j,k)th element of $\underline{\widetilde{\mathbf{X}}}$ can be written as

$$\tilde{x}_{i,j,k} = x_{i,j,k} + v_{i,j,k} \tag{5}$$

where $v_{i,j,k}$ denotes the additive complex i.i.d. zero-mean measurement noise with statistically independent real and imaginary parts.

The PARAFAC fitting problem is then formulated as follows. We are given the noisy data $\underline{\widetilde{\mathbf{X}}}$ and wish to estimate $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Let us introduce the tall matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_I \end{bmatrix}_{JI \times K} = \begin{bmatrix} \mathbf{B}\mathbf{A}_1 \\ \vdots \\ \mathbf{B}\mathbf{A}_I \end{bmatrix} \mathbf{C}^T = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T \tag{6}$$

where $\odot$ stands for the Khatri-Rao matrix product. Similarly, we introduce the matrix of noisy data

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \widetilde{\mathbf{X}}_1 \\ \vdots \\ \widetilde{\mathbf{X}}_I \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_I \end{bmatrix} + \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_I \end{bmatrix} \tag{7}$$

Then the conditional Maximum Likelihood (ML) estimation problem for the matrix $\mathbf{C}$ given matrices $\mathbf{A}$ and $\mathbf{B}$ and assuming i.i.d. Gaussian measurement noise is the LS fitting problem $\min_{\mathbf{C}} \|\widetilde{\mathbf{X}} - (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T\|_F^2$, where $\|\cdot\|_F^2$ denotes the Frobenius matrix norm.

If the measurement noise is i.i.d. Laplacian (with i.i.d. Laplacian-distributed real and imaginary parts in the complex case), then ML estimation is equivalent to LAE regression. Some manipulations are necessary in order to express the absolute error criterion in the form of a convenient vector ℓ_1 norm. Towards this end, introduce the operator $\mathcal{F}(\cdot)$

$$\mathbf{s} = \mathcal{F}(\mathbf{S}) = \begin{bmatrix} \breve{\mathbf{S}}_{:,1} \\ \vdots \\ \breve{\mathbf{S}}_{:,L} \end{bmatrix}, \qquad \breve{\mathbf{S}}_{:,l} = \begin{bmatrix} \mathrm{Re}\{\mathbf{S}_{:,l}\} \\ \mathrm{Im}\{\mathbf{S}_{:,l}\} \end{bmatrix} \tag{8}$$

where $\mathbf{S}$ is a complex-valued $M \times L$ matrix, and $\mathbf{S}_{:,l}$ denotes its lth column. The following property holds:

$$\mathcal{F}\{\mathbf{DF}\} = (\mathbf{I} \otimes \mathcal{G}\{\mathbf{D}\})\mathcal{F}\{\mathbf{F}\} \tag{9}$$

where $\mathbf{I}$ is the identity matrix of commensurate dimension, $\mathbf{D}$ and $\mathbf{F}$ are any complex-valued matrices of commensurate dimensions, $\otimes$ denotes the Kronecker matrix product, and $\mathcal{G}\{\mathbf{D}\}$ stands for the operator

$$\mathcal{G}\{\mathbf{D}\} = \begin{bmatrix} \mathrm{Re}\{\mathbf{D}\} & -\mathrm{Im}\{\mathbf{D}\} \\ \mathrm{Im}\{\mathbf{D}\} & \mathrm{Re}\{\mathbf{D}\} \end{bmatrix} \tag{10}$$

Using property (9), we find that the absolute error model fitting criterion can be written as

$$\|\tilde{\mathbf{x}} - (\mathbf{I}_K \otimes \mathcal{G}\{\mathbf{A} \odot \mathbf{B}\})\mathbf{c}\|_1 \tag{11}$$

i.e., through the ℓ_1 norm of a real-valued vector. Here, $\tilde{\mathbf{x}} = \mathcal{F}(\widetilde{\mathbf{X}})$ and $\mathbf{c} = \mathcal{F}(\mathbf{C})$.

Using the other two ways of slicing the array $\underline{\mathbf{X}}$, we introduce the matrices

$$\mathbf{Y} = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T, \qquad \mathbf{Z} = (\mathbf{C} \odot \mathbf{A})\mathbf{B}^T \tag{12}$$

and correspondingly

$$\widetilde{\mathbf{Y}} = \begin{bmatrix} \widetilde{\mathbf{Y}}_1 \\ \vdots \\ \widetilde{\mathbf{Y}}_J \end{bmatrix}_{KJ \times I}, \qquad \widetilde{\mathbf{Z}} = \begin{bmatrix} \widetilde{\mathbf{Z}}_1 \\ \vdots \\ \widetilde{\mathbf{Z}}_K \end{bmatrix}_{IK \times J} \tag{13}$$

where $\widetilde{\mathbf{Y}}_j$, $j = 1, \ldots, J$, and $\widetilde{\mathbf{Z}}_k$, $k = 1, \ldots, K$ are the noisy slabs of $\underline{\widetilde{\mathbf{X}}}$ along corresponding dimensions.

We now have all necessary notation to explain the new fitting algorithm. Let us consider the TALAE sub-problems [6], [8]

$$\widehat{\mathbf{c}} = \arg\min_{\mathbf{c}} \|\tilde{\mathbf{x}} - (\mathbf{I}_K \otimes \mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\})\mathbf{c}\|_1, \ \widehat{\mathbf{C}} = (\mathcal{F}^{-1}\{\widehat{\mathbf{c}}\})^T \tag{14}$$

$$\widehat{\mathbf{b}} = \arg\min_{\mathbf{b}} \|\tilde{\mathbf{z}} - (\mathbf{I}_J \otimes \mathcal{G}\{\widehat{\mathbf{C}} \odot \widehat{\mathbf{A}}\})\mathbf{b}\|_1, \ \widehat{\mathbf{B}} = (\mathcal{F}^{-1}\{\widehat{\mathbf{b}}\})^T \tag{15}$$

$$\widehat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\tilde{\mathbf{y}} - (\mathbf{I}_I \otimes \mathcal{G}\{\widehat{\mathbf{B}} \odot \widehat{\mathbf{C}}\})\mathbf{a}\|_1, \ \widehat{\mathbf{A}} = (\mathcal{F}^{-1}\{\widehat{\mathbf{a}}\})^T \tag{16}$$

and show how to solve them iteratively. Here $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse operator to $\mathcal{F}\{\cdot\}$ of (8). In what follows, only the problem (14) will be considered in detail; (15) and (16) can be treated similarly.

We first (e.g., randomly) initialize $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Then, fixing all parameters in (14) except for $c_{k+(f-1)K}$ ($k \in \{1, \ldots, K\}$, $f \in \{1, \ldots, 2F\}$), we can simplify this problem as

$$\min_{c_{k+(f-1)K}} \| \breve{\mathbf{X}}_{:,k} - \sum_{p=1, p \neq f}^{2F} \mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\}_{:,p} c_{k+(p-1)K} - \mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\}_{:,f} c_{k+(f-1)K} \|_1 \tag{17}$$

where $\breve{\mathbf{X}}_{:,k} = [\mathrm{Re}\{\widetilde{\mathbf{X}}_{:,k}\}^T, \ \mathrm{Im}\{\widetilde{\mathbf{X}}_{:,k}\}^T]^T$, and $\mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\}_{:,f}$ stands for the fth column of $\mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\}$.

Defining

$$\mathbf{h}_k = \breve{\mathbf{X}}_{:,k} - \sum_{p=1, p \neq f}^{2F} \mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\}_{:,p} c_{k+(p-1)K} \tag{18}$$

$$\mathbf{m}_f = \mathcal{G}\{\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}}\}_{:,f} \tag{19}$$

the problem in (17) becomes

$$\min_{c_{k+(f-1)K}} \|\mathbf{h}_k - \mathbf{m}_f c_{k+(f-1)K}\|_1 \tag{20}$$

The minimization problem in (20) can be further written as

$$\min_{c_{k+(f-1)K}} \sum_{l=1}^{2IJ} |h_k(l) - m_f(l) c_{k+(f-1)K}| \tag{21}$$

where $h_k(l)$ and $m_f(l)$ are the lth elements of the vectors $\mathbf{h}_k$ and $\mathbf{m}_f$, respectively. Equivalently, (21) can be expressed as

$$\min_{c_{k+(f-1)K}} \sum_{l=1}^{2IJ} |m_f(l)| \left| \frac{h_k(l)}{m_f(l)} - c_{k+(f-1)K} \right| \tag{22}$$

provided that none of the elements $m_f(l)$ is zero. Note that if one of these elements is zero, then the corresponding summand in (22) can be dropped because it becomes a constant in this case (independent of $c_{k+(f-1)K}$). Note that the optimization problem (22) is solved by WMF [7] where $\{h_k(l)/m_f(l)\}_{l=1}^{2IJ}$, $\{|m_f(l)|\}_{l=1}^{2IJ}$, and $c_{k+(f-1)K}$ are the filter inputs, weights, and output value, respectively. The WMF operation boils down to sorting, and can thus be efficiently implemented at a complexity cost of $2IJ\log(2IJ)$.

Iterating the WMF over real and imaginary parts of all elements of the matrix $\mathbf{C}$, e.g., in a circular fashion, and likewise for the elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ involved in the decomposition, one obtains a LAE trilinear regression algorithm that is monotonically convergent in terms of the LAE cost function. The per-iteration complexity of the TALAE-WMF algorithm is then estimated as $\mathcal{O}(FIJK\log(IJK))$, which is much lower than the corresponding complexity of TALAE-LP [8] and is comparable with the complexity of TALS [3]. The per-iteration complexity of the TALAE-LP procedure can be estimated as $\mathcal{O}(I^3J^3 + J^3K^3 + I^3K^3)$ [8], while the complexity of the TALS as $\mathcal{O}(FIJK)$ [3]. However, the overall complexity of any algorithm depends on the number of iterations, which may vary depending on problem-specific parameters and the given batch of data.

IV. Cramér-Rao Bounds

The PARAFAC Cramér-Rao Bound (CRB) for Gaussian noise has been derived in [9]. Corresponding Laplacian and α-stable CRBs are of interest as benchmarks in our present context. Since the symmetric α-stable distribution does not have an analytic expression for its density function, we will derive the CRB for robust fitting of the trilinear PARAFAC model for the special case of Cauchy noise. First, the Cauchy distribution has a closed-form expression for its density function. Second, estimators that perform well under the Cauchy distribution are robust in different impulsive noise environments, i.e., the performance of such estimators does not change significantly when other symmetric α-stable distributions are used [4].

One delicate point regarding the CRB for the trilinear decomposition model is the inherent permutation and scale ambiguity. To derive a meaningful CRB, we assume that the first rows of $\mathbf{A}$ and $\mathbf{B}$ are normalized to $[1,\ldots,1]_{1\times F}$, which resolves the scale ambiguity [9]. Further we assume that the first row of $\mathbf{C}$ is known and consists of distinct elements, which resolves the permutation ambiguity [9]. Then, we can write the $1\times(I+J+K-3)F$ row-vector of unknown complex parameters as

$$[\mathbf{a}_2^T,\ldots,\mathbf{a}_I^T,\mathbf{b}_2^T,\ldots,\mathbf{b}_J^T,\mathbf{c}_2^T,\ldots,\mathbf{c}_K^T] \tag{23}$$

Result: The $(I+J+K-3)F\times(I+J+K-3)F$ Fisher information matrix (FIM) for the estimation of unknown complex parameters (23) in Gaussian, Laplacian and Cauchy noise is given by

$$\text{FIM} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 \\ \mathbf{F}_2^H & \mathbf{F}_{\mathbf{c},\mathbf{c}} \end{bmatrix} \tag{24}$$

where

$$\mathbf{F}_1 = \begin{bmatrix} \mathbf{F}_{\mathbf{a},\mathbf{a}} & \mathbf{F}_{\mathbf{a},\mathbf{b}} \\ \mathbf{F}_{\mathbf{a},\mathbf{b}}^H & \mathbf{F}_{\mathbf{b},\mathbf{b}} \end{bmatrix}, \qquad \mathbf{F}_2 = \begin{bmatrix} \mathbf{F}_{\mathbf{a},\mathbf{c}} \\ \mathbf{F}_{\mathbf{b},\mathbf{c}} \end{bmatrix} \tag{25}$$

$$\mathbf{F}_{\mathbf{a},\mathbf{a}} = \beta\left[\mathbf{I}_{I-1}\otimes((\mathbf{B}\odot\mathbf{C})^H(\mathbf{B}\odot\mathbf{C}))\right] \tag{26}$$

$$\mathbf{F}_{\mathbf{b},\mathbf{b}} = \beta\left[\mathbf{I}_{J-1}\otimes((\mathbf{C}\odot\mathbf{A})^H(\mathbf{C}\odot\mathbf{A}))\right] \tag{27}$$

$$\mathbf{F}_{\mathbf{c},\mathbf{c}} = \beta\left[\mathbf{I}_{K-1}\otimes((\mathbf{A}\odot\mathbf{B})^H(\mathbf{A}\odot\mathbf{B}))\right] \tag{28}$$

$$\mathbf{F}_{\mathbf{a},\mathbf{b}} = \beta\left[\mathbf{e}\otimes(\mathbf{B}\odot\mathbf{C})^H\right]\mathbf{V}_{\mathbf{a},\mathbf{b}}\left[\mathbf{e}^T\otimes(\mathbf{C}\odot\mathbf{A})\right] \tag{29}$$

$$\mathbf{F}_{\mathbf{a},\mathbf{c}} = \beta\left[\mathbf{e}\otimes(\mathbf{B}\odot\mathbf{C})^H\right]\mathbf{V}_{\mathbf{a},\mathbf{c}}\left[\mathbf{e}^T\otimes(\mathbf{A}\odot\mathbf{B})\right] \tag{30}$$

$$\mathbf{F}_{\mathbf{b},\mathbf{c}} = \beta\left[\mathbf{e}\otimes(\mathbf{C}\odot\mathbf{A})^H\right]\mathbf{V}_{\mathbf{b},\mathbf{c}}\left[\mathbf{e}^T\otimes(\mathbf{A}\odot\mathbf{B})\right] \tag{31}$$

$$\beta = \begin{cases} \frac{1}{\sigma_v^2}, & \text{for Gaussian noise} \\ \frac{2}{\sigma_v^2}, & \text{for Laplacian noise} \\ \frac{3}{10\gamma^2}, & \text{for complex Cauchy noise} \\ \frac{1}{2\gamma^2}, & \text{for real Cauchy noise} \end{cases} \tag{32}$$

σ_v is the standard deviation of the Gaussian or Laplacian distribution, $\gamma > 0$ is the *dispersion* of the Cauchy distribution, $\mathbf{e} = [1,1,\ldots,1]^T$, and

$$\mathbf{V}_{\mathbf{a},\mathbf{b}} = \begin{bmatrix} \mathbf{V}_{\mathbf{a},\mathbf{b}}(2,2) & \ldots & \mathbf{V}_{\mathbf{a},\mathbf{b}}(2,J) \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{\mathbf{a},\mathbf{b}}(I,2) & \ldots & \mathbf{V}_{\mathbf{a},\mathbf{b}}(I,J) \end{bmatrix} \tag{33}$$

$$\mathbf{V}_{\mathbf{a},\mathbf{c}} = \begin{bmatrix} \mathbf{V}_{\mathbf{a},\mathbf{c}}(2,2) & \ldots & \mathbf{V}_{\mathbf{a},\mathbf{c}}(2,K) \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{\mathbf{a},\mathbf{c}}(I,2) & \ldots & \mathbf{V}_{\mathbf{a},\mathbf{c}}(I,K) \end{bmatrix} \tag{34}$$

$$\mathbf{V}_{\mathbf{b},\mathbf{c}} = \begin{bmatrix} \mathbf{V}_{\mathbf{b},\mathbf{c}}(2,2) & \ldots & \mathbf{V}_{\mathbf{b},\mathbf{c}}(2,K) \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{\mathbf{b},\mathbf{c}}(J,2) & \ldots & \mathbf{V}_{\mathbf{b},\mathbf{c}}(J,K) \end{bmatrix} \tag{35}$$

Here $\mathbf{V}_{\mathbf{a},\mathbf{b}}(i,j)$, $\mathbf{V}_{\mathbf{a},\mathbf{c}}(i,k)$ and $\mathbf{V}_{\mathbf{b},\mathbf{c}}(j,k)$ are the matrices which contain only zeros and ones. In $\mathbf{V}_{\mathbf{a},\mathbf{b}}(i,j)$, ones are located on the intersection of the rows with numbers $(j-1)K+1,(j-1)K+2,\ldots,(j-1)K+K$ and columns with numbers $i, I+i,\ldots,(K-1)I+i$; in $\mathbf{V}_{\mathbf{a},\mathbf{c}}(i,k)$ – on the intersection of the rows with numbers $k, k+(j-1)K,\ldots,k+(J-1)K$ and columns with numbers $(i-1)J+1,(i-1)J+2,\ldots,(i-1)J+J$; and in $\mathbf{V}_{\mathbf{b},\mathbf{c}}(j,k)$ – on the intersection of rows with numbers $(k-1)I+1,(k-1)I+2,\ldots,(k-1)I+I$ and columns with numbers $j, J+j,\ldots,(I-1)J+j$.

The CRB matrix that corresponds to the unknown elements of $\mathbf{C}$ is given by

$$\text{CRB}_{\mathbf{c},\mathbf{c}} = (\mathbf{F}_{\mathbf{c},\mathbf{c}} - \mathbf{F}_2^H\mathbf{F}_1^{-1}\mathbf{F}_2)^{-1} \tag{36}$$

Proof: See [8].

V. Application to Blind Multiuser Separation-Detection and Simulation Results

In this section, we apply the proposed WMF technique to the problem of blind PARAFAC multiuser separation-detection in

a Direct-Sequence Code-Division Multiple Access (DS-CDMA) communication system [1]. The performance of the proposed TALAE-WMF algorithm is numerically compared with that of the TALAE-LP [6], conventional TALS method [3], and against the pertinent CRB. An example similar to that of [1] is simulated. For the application considered, the elements of the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ have the following meanings

$a_{i,f}$ fading/gain between user f and antenna element i;

$b_{j,f}$ j-th chip of the spreading code of user f;

$c_{k,f}$ k-th symbol transmitted by user f.

Correspondingly, $\mathbf{A}$ is the channel matrix, $\mathbf{B}$ is the spreading code matrix, and $\mathbf{C}$ is the user symbol matrix, all unknown to the receiver. Here also

F number of users;

I total number of antenna elements in the array;

J number of Intersymbol Interference (ISI)-free chips per symbol (ISI-free equivalent spreading gain);

K length of the transmitted sequence of symbols (number of snapshots).

The reader is refereed to [1] for a list of the necessary assumptions which make the PARAFAC model valid in this context.

The data $\underline{\mathbf{X}}$ are contaminated by channel noise. Three models of the channel noise are used. One is Gaussian noise, while the other two are Laplacian and Cauchy noise.

Scale and permutation ambiguities are inherent to this blind separation problem [1]; the scale ambiguity manifests itself as a complex constant that multiplies each individual row of $\mathbf{C}$. For constant-modulus transmissions, this ambiguity can be removed via Automatic Gain Control (AGC) and differential encoding/decoding. We assume differentially-encoded user signals throughout the simulations. For the purpose of performance evaluation only, the permutation ambiguity is resolved using a greedy LS $(\mathbf{C}, \widehat{\mathbf{C}})$ column-matching algorithm.

We present Monte Carlo simulations that are designed to assess the Root Mean Square Error (RMSE) performance of the aforementioned algorithms. The parameters used in the simulations are as follows: N = number of Monte Carlo trials = 100; $I = 8$; $J = 8$; $K = 20$; $F = 2$; and $\alpha = 1$. Here, α is the characteristic exponent which determines the heaviness of the tail of the symmetric α-stable distribution used in our third example ($\alpha = 1$ yields the Cauchy distribution). The associated symmetric α-stable characteristic function is given by

$$\phi(\omega) = \exp\{-\gamma|\omega|^{\alpha}\} \tag{37}$$

where γ is a positive constant related to the scale of the distribution (also known as *dispersion* in the case of the Cauchy distribution). The *geometric* Signal-to-Noise Ratio (SNR) in this case is defined according to [10, p. 68]

$$\text{SNR} = \frac{1}{C_{\text{g}}}\left(\frac{A}{S_0}\right)^2 \tag{38}$$

where $C_{\text{g}} = e^{C_{\text{e}}} \approx 1.78$ is the exponential of the Euler constant $C_{\text{e}} = 0.5772\ldots$, A is the magnitude of the noise-free signal, and S_0 is the geometric power of symmetric α-stable noise [10, p. 38]:

$$S_0 = \frac{(C_{\text{g}}\gamma)^{1/\alpha}}{C_{\text{g}}} \tag{39}$$

Both Gaussian and Cauchy distributions belong to the class of symmetric α-stable distributions. The geometric power of complex Gaussian noise is given by [10]

$$S_{0,\text{Gaussian}} = \frac{\sigma_{\text{v}}}{\sqrt{C_{\text{g}}}} \tag{40}$$

Substituting (40) into (38) we find that for the Gaussian case the geometric SNR is equivalent to the standard SNR. In the complex Cauchy case, the noise power and corresponding geometric SNR can be respectively written as

$$S_{0,\text{Cauchy}} = 2\gamma, \qquad \text{SNR}_{\text{Cauchy}} = \frac{A^2}{4C_{\text{g}}\gamma^2} \tag{41}$$

In the Laplacian case, we use the standard SNR since Laplacian distribution does not belong to the class of symmetric α-stable distributions.

Throughout the simulations, we assume that the noise power is normalized to be equal to 1. User signals are redrawn from an i.i.d. Bernoulli distribution and differentially encoded for each Monte Carlo trial. BPSK modulation is used for all user signals. The gains of the channel matrix $\mathbf{A}$ and the elements of the spreading code matrix $\mathbf{B}$ are generated as i.i.d. Gaussian unit variance random variables and are fixed in each Monte Carlo trial, and re-drawn from one trial to another.

Even though dimensions and ranks are such that algebraic (ESPRIT-like) initialization is possible for all three algorithms in our simulation setup, we initialize all three competing algorithms randomly for each batch of data. The reason is that we wish to assess the global convergence characteristics of the three iterations.

The RMSE for each simulated point and for each method tested is calculated according to the following expression

$$\text{RMSE} = \sqrt{\frac{1}{N(K-1)F}\sum_{n=1}^{N}\|\widehat{\mathbf{C}}(n) - \mathbf{C}\|_F^2} \tag{42}$$

while the (averaged) CRB is calculated as

$$\text{CRB} = \sqrt{\sum_{n=1}^{N}\frac{\text{Tr}\{\text{CRB}_{\mathbf{c},\mathbf{c}}(n)\}}{N(K-1)F}} \tag{43}$$

where $\text{Tr}\{\cdot\}$ stands for the trace operator.

Note that using the RMSE (respectively CRB) as performance measure (respectively benchmark) we ignore the finite alphabet (FA) constraint on the transmitted symbols. This is reasonable because the RMSE is a good indicator of post-detection bit error rate, and all three fitting algorithms do not make use of the FA constraint. Hence the CRB is indeed a pertinent bound on their performance.

Figures 1, 2 and 3 plot the performance of the aforementioned algorithms in terms of RMSE versus the SNR for the case of Gaussian, Laplacian and Cauchy noise, respectively, and compare the performance with the corresponding CRBs.

Figures 1 and 2 demonstrate that in the case of Gaussian noise, the TALS method performs slightly better than the TALAE-LP and TALAE-WMF algorithms, while in the case of Laplacian noise, the latter robust algorithms have slightly better performance as compared to the TALS method. In the case of Cauchy noise (Fig. 3), the TALS method breaks down, while the performance of the TALAE-LP and TALAE-WMF

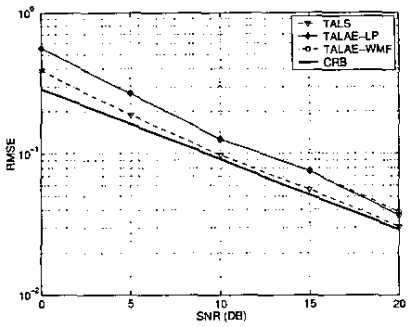

Figure 1: RMSEs versus SNR. Gaussian channel noise.

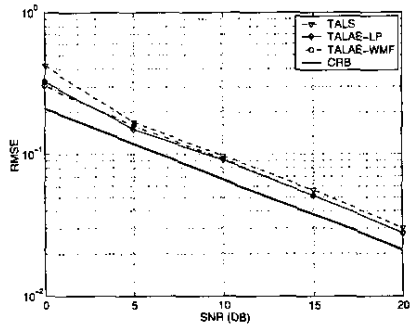

Figure 2: RMSEs versus SNR. Laplacian channel noise.

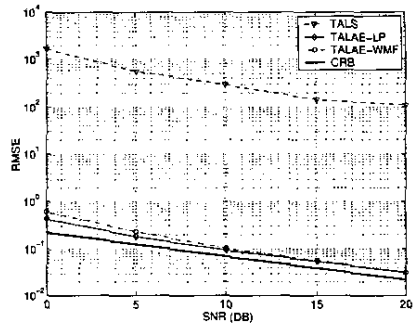

Figure 3: RMSEs versus SNR. Cauchy channel noise.

algorithms is not affected and is close to the CRB (despite the fact that our techniques are designed for Laplacian noise). The degradation in performance relative to TALS in the Gaussian case can be considered as a moderate price for greatly improved robustness against heavy-tailed Cauchy noise.

Comparing the two robust regressions (TALAE-LP versus TALAE-WMF), we see that they behave very similarly performance-wise in all cases considered in our simulations. This was not necessarily expected, because TALAE-LP jointly updates many parameters, and is therefore capable of making "superdiagonal" optimization steps which are not possible with TALAE-WMF. The latter updates one parameter at a time, and thus it may be more easily trapped in ridges which do not allow fit improvements by means of updating only a single parameter. Nevertheless, this possibility does not appear to affect the performance of TALAE-WMF in our simulations.

We have seen that each complete update cycle of TALAE-WMF (in which all parameters are visited for update once, in any order) is computationally much simpler than the corresponding cycle of TALAE-LP. To get a real sense of computational complexity, the number of iterations required for convergence of both methods is also needed. This number varied between 10 and 20 depending on simulated noise model and initialization of matrices $\mathbf{A}$ and $\mathbf{B}$. Throughout the simulations, it was observed that TALAE-WMF requires 2-5 more iterations than TALAE-LP for Gaussian or Laplacian noise, and about the same number of iterations for Cauchy noise. This is indeed a very promising result taking into account that TALAE-WMF can be implemented with simple sorting hardware, whereas TALAE-LP and TALS require a sophisticated computing capability.

VI. Conclusions

A new iterative algorithm for robust fitting of trilinear PARAFAC models has been proposed and applied to the problem of blind multiuser separation-detection. The algorithm relies on alternating optimization using WMF. The proposed algorithm outperforms the workhorse alternating LS PARAFAC fitting procedure and performs as good as TALAE-LP method under heavy-tailed noise. Even though it is matched to the Laplacian distribution, it still performs well under Cauchy noise. Furthermore, its performance degrades only moderately under Gaussian noise and it has much lower complexity than the TALAE-LP procedure. We therefore expect that the new algorithm will prove useful in a variety of applications of PARAFAC analysis.

References

[1] N. D. Sidiropoulos, G. B. Giannakis and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 810-823, Mar. 2000.

[2] N. D. Sidiropoulos, R. Bro and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Processing*, vol. 48, pp. 2377-2388, Aug. 2000.

[3] R. Bro, N. D. Sidiropoulos and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures," in *Proc. Int. Workshop Idependent Component Analysis and Blind Signal Separation*, Aussois, France, Jan. 1999.

[4] P. Tsakalides and C.L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Processing*, vol. 43, pp. 2700-2713, Nov. 1995.

[5] N. D. Sidiropoulos and R. Bro, "Mathematiacl programming algorithms for regression-based filtering in $\mathcal{R}^N$," *IEEE Trans. Signal Processing*, vol. 47, pp. 771-782, Mar. 1999.

[6] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models based on linear programming," *IEEE Int. Conf. Acoust., Speech, and Sig. Proc., ICASSP'2004*, to appear.

[7] R. Yang, L. Yin, M. Gabbouj, J. Astola and Y. Neuvo, "Optimal weighted median filtering under structural constraints," *IEEE Trans. Signal Processing*, vol. 43, pp. 591-603, Mar. 1995.

[8] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Processing*, revised.

[9] X. Liu, N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimantional arrays," *IEEE Trans. Signal Processing*, vol. 49, pp. 2074-2086, Sept. 2001.

[10] J. G. Gonzalez, "Robust techniques for wireless communications in non-Gaussian environments," Ph.D. dissertation, University of Delaware, 1997.