Robust Iterative Fitting of Multilinear Models

Sergiy A. Vorobyov, *Member, IEEE*, Yue Rong, *Student Member, IEEE*,
Nicholas D. Sidiropoulos, *Senior Member, IEEE*, and Alex B. Gershman, *Senior Member, IEEE*

Abstract—**Parallel factor (PARAFAC) analysis is an extension of low-rank matrix decomposition to higher way arrays, also referred to as *tensors*. It decomposes a given array in a sum of multilinear terms, analogous to the familiar bilinear vector outer products that appear in matrix decomposition. PARAFAC analysis generalizes and unifies common array processing models, like joint diagonalization and ESPRIT; it has found numerous applications from blind multiuser detection and multidimensional harmonic retrieval, to clustering and nuclear magnetic resonance. The prevailing fitting algorithm in all these applications is based on (alternating) least squares, which is optimal for Gaussian noise. In many cases, however, measurement errors are far from being Gaussian. In this paper, we develop two iterative algorithms for the least absolute error fitting of general multilinear models. The first is based on efficient interior point methods for linear programming, employed in an alternating fashion. The second is based on a weighted median filtering iteration, which is particularly appealing from a simplicity viewpoint. Both are guaranteed to converge in terms of absolute error. Performance is illustrated by means of simulations, and compared to the pertinent Cramér–Rao bounds (CRBs).**

Index Terms—**Array signal processing, non-Gaussian noise, parallel factor analysis, robust model fitting.**

I. INTRODUCTION

THE parallel factor (PARAFAC) model [1]–[5] is a useful data analysis tool that has recently found applications in array signal processing and communications, e.g., [6] and [7]. 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 under certain conditions [5], [6], in contrast to low-rank matrix decomposition.

Because of its direct link to low-rank decomposition, PARAFAC analysis has found applications in numerous and diverse disciplines, e.g., cf. [6], [7], and references therein. Related work on joint diagonalization, symmetric, super-symmetric, and rank-one tensorial decomposition, has also appeared in the signal processing literature, mostly in the context of higher order statistics (HOS) and independent component analysis (ICA)-based blind source separation [8]–[11]. The papers [8]–[11] do not address the general PARAFAC analysis problem, as we do, but focus on special (symmetric, rank-one) cases which are of interest in their own right.

In most applications of PARAFAC analysis, an alternating least squares regression procedure is used to fit the model parameters (e.g., cf. [6] and [12]). Least squares (LS) regression is optimal (in the maximum likelihood sense) when measurement errors are additive i.i.d. Gaussian. Gaussianity is an often-made assumption, due to the central limit theorem, but also for tractability considerations. However, in many applications (e.g., [13] and [14]), 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 [15]. LAE regression is optimal (in the maximum likelihood sense) when measurement errors are additive i.i.d. Laplacian (e.g., see [16]). 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.

Another distribution commonly used for modeling impulsive noise is the Cauchy, and, more generally, the class of α-stable distributions [13], [14]. Interestingly, as we will see, regression under the LAE criterion often performs well even when measurement errors are not Laplacian, but rather drawn from the class of α-stable distributions. It is, therefore, of interest to develop PARAFAC regression procedures that optimize the LAE fitting criterion.

In this paper, we develop two such iterative procedures. One is based on linear programming (LP); the other makes use of weighted median filtering (WMF) (e.g., [19]). Their relative merits are investigated via simulations and compared to the

S. A. Vorobyov and Y. Rong are with the Department of Communication Systems, Darmstadt University of Technology, Darmstadt D-64283, Germany (e-mail: svor@ieee.org; yue.rong@ieee.org).

N. D. Sidiropoulos is with the Department of Electronic and Computer Engineering, Technical University of Crete, Chania 73100, Greece, and also with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@telecom.tuc.gr).

A. B. Gershman is with the Department of Communication Systems, Darmstadt University of Technology, Darmstadt D-64283, Germany, on leave from the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: gershman@ieee.org).

Digital Object Identifier 10.1109/TSP.2005.850343

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

pertinent Cramér–Rao bounds (CRBs), which are also derived herein.

Yet, another commonly used class of impulsive noise models is Gaussian mixture distributions. Gaussian mixture models can approximate a broad class of non-Gaussian distributions. Estimation algorithms derived under the Gaussian mixture assumption often maintain good performance under symmetric stable distributions (e.g., cf. [17] and [18]). LAE regression affords a similar robustness benefit across stable distributions; however, it is conceptually much simpler. For example, it does not depend on any noise parameters, and, thus, there is no need for a separate parameter estimation step.

This paper is organized as follows. In the next section, the PARAFAC model is introduced. The proposed robust iterative algorithms are developed in Section III and the related CRBs are presented in Section IV. Simulation results are given in Section V and conclusions are drawn in Section VI.

II. PARALLEL FACTOR ANALYSIS

We introduce notation that will be useful in the sequel. Consider an $I \times J \times K$ *three-way array* $\underline{\boldsymbol{X}}$ 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 $\boldsymbol{A}$, 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 $\boldsymbol{B}$ and $\boldsymbol{C}$, respectively. Matrices $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ are, in general, complex valued. Equation (1) expresses $x_{i,j,k}$ as a sum of F rank-1 triple products; it is known as trilinear decomposition, or *PARAFAC analysis* of $x_{i,j,k}$. If $F \le ((k_{\boldsymbol{A}} + k_{\boldsymbol{B}} + k_{\boldsymbol{C}})/2) - 1$, where $k_{\boldsymbol{A}}, k_{\boldsymbol{B}}$ and $k_{\boldsymbol{C}}$ are the Kruskal ranks [5] of the matrices $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$, respectively, then rank-F decomposition of the three-way array $\underline{\boldsymbol{X}}$ is unique [5], [6]. The Kruskal rank of $\mathbf{A}$ is the maximum integer, k, such that *any* k columns drawn from $\mathbf{A}$ are linearly independent.

Let $A_i = \mathcal{D}_i(\boldsymbol{A})$ denote the operator which takes the ith row of matrix $\boldsymbol{A}$ and produces a diagonal matrix by placing this row on the main diagonal. Then, by "slicing" the three-dimensional (3-D) array $\underline{\boldsymbol{X}}$ in a series of "slabs" [two-dimensional (2-D) arrays], we obtain

$$ \boldsymbol{X}_i = \boldsymbol{B} \boldsymbol{A}_i \boldsymbol{C}^T, \quad i = 1, \ldots, I. \qquad (2) $$

Here, such a slicing is made perpendicular to the ith dimension, i.e., $\boldsymbol{X}_i := [x_{i,\cdot,\cdot}]$ is the $J \times K$ 2-D slice of $\underline{\boldsymbol{X}}$ corresponding to the given index i. Two other types of slicing of $\underline{\boldsymbol{X}}$ are useful in understanding the algorithms which will be developed in the next section. They are given by

$$ \boldsymbol{Y}_j = \boldsymbol{C} \boldsymbol{B}_j \boldsymbol{A}^T, \quad j = 1, \ldots, J \qquad (3) $$
$$ \boldsymbol{Z}_k = \boldsymbol{A} \boldsymbol{C}_k \boldsymbol{B}^T, \quad k = 1, \ldots, K \qquad (4) $$

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

III. TRILINEAR ALTERNATING LEAST ABSOLUTE ERROR (TALAE) REGRESSION

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

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

where $v_{i,j,k}$ denote 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 $\underline{\tilde{\boldsymbol{X}}}$ and wish to estimate $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$. Let us introduce the tall matrix

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

where $\odot$ stands for the Khatri–Rao (column-wise Kronecker) matrix product. Similarly, we introduce the matrix of noisy data

$$ \tilde{\boldsymbol{X}} = \begin{bmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_I \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_I \end{bmatrix} + \begin{bmatrix} V_1 \\ \vdots \\ V_I \end{bmatrix}. \qquad (7) $$

Then, the conditional maximum lkelihood (ML) estimation problem for the matrix $\boldsymbol{C}$ given matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and assuming i.i.d. Gaussian measurement noise is the LS fitting problem

$$ \min_{\boldsymbol{C}} \|\tilde{X} - (\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{C}^T\|_F^2 \qquad (8) $$

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. Toward this end, introduce the operator $\mathcal{F}(\cdot)$

$$ \boldsymbol{s} = \mathcal{F}(\boldsymbol{S}) \triangleq \begin{bmatrix} \breve{\boldsymbol{S}}_{\cdot,1} \\ \vdots \\ \breve{\boldsymbol{S}}_{\cdot,L} \end{bmatrix} \qquad (9) $$

$$ \breve{\boldsymbol{S}}_{\cdot,l} \triangleq \begin{bmatrix} \operatorname{Re}\{\boldsymbol{S}_{\cdot,l}\} \\ \operatorname{Im}\{\boldsymbol{S}_{\cdot,l}\} \end{bmatrix} \qquad (10) $$

where $\boldsymbol{S}$ is a complex-valued $M \times L$ matrix and $\boldsymbol{S}_{\cdot,l}$ denotes its lth column. The following property holds (see Appendix A for derivation):

$$ \mathcal{F}\{\boldsymbol{D}\boldsymbol{F}\} = (\boldsymbol{I} \otimes \mathcal{G}\{\boldsymbol{D}\})\mathcal{F}\{\boldsymbol{F}\} \qquad (11) $$

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

$$\mathcal{G}\{D\} \triangleq \begin{bmatrix} \text{Re}\{D\} & -\text{Im}\{D\} \\ \text{Im}\{D\} & \text{Re}\{D\} \end{bmatrix}. \tag{12}$$

Using property (11), we find that the absolute error model fitting criterion can be written as

$$\|\tilde{x} - (I_K \otimes \mathcal{G}\{A \odot B\})c\|_1 \tag{13}$$

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

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

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_J \end{bmatrix}_{KJ \times I} = (B \odot C)A^T \tag{14}$$

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_K \end{bmatrix}_{IK \times J} = (C \odot A)B^T \tag{15}$$

and correspondingly

$$\tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_J \end{bmatrix}, \quad \tilde{Z} = \begin{bmatrix} \tilde{Z}_1 \\ \vdots \\ \tilde{Z}_K \end{bmatrix} \tag{16}$$

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

Now, we have all notations necessary to explain the new fitting algorithms.

A. Trilinear Alternating LAE (TALAE) Regression Based on Linear Programming (TALAE-LP)

The idea behind this algorithm is similar to that of trilinear alternating LS (TALS) regression for Gaussian noise[2] [7], [12] and is as follows. Each time, update a subset of parameters using the LAE criterion, conditioned on previously obtained estimates of the remaining parameters; proceed to update another subset of parameters; repeat until convergence.

In more detail, we first initialize matrices A and B randomly or by single-invariance ESPRIT when applicable [6], [7]. Then, given the matrix $\tilde{X}$, and these initial estimates of A and B (which we denote hereafter as $\hat{A}$ and $\hat{B}$), our purpose is to find the estimate of the matrix C which minimizes the norm (13). Specifically, we have to find the estimate of C by solving the following optimization problem:

$$\hat{c} = \arg\min_{c} \|\tilde{x} - (I_K \otimes \mathcal{G}\{\hat{A} \odot \hat{B}\})c\|_1, \quad \hat{C} = (\mathcal{F}^{-1}\{\hat{c}\})^T \tag{17}$$

[2]However, the norm of type (13) is now used instead of the Frobenius norm of (8).

for given $\tilde{x}, \hat{A}$, and $\hat{B}$. In (17), $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse operator to $\mathcal{F}\{\cdot\}$ of (9). Introducing the vector $e = [1, 1, \ldots, 1]^T$ and the vector of slack variables q_1 (both of commensurate dimensions), we can equivalently write the problem (17) as

$$\min_{c,q_1} e^T q_1 \quad \text{subject to} \quad \tilde{x} - (I_K \otimes \mathcal{G}\{\hat{A} \odot \hat{B}\})c \preceq q_1$$
$$\tilde{x} - (I_K \otimes \mathcal{G}\{\hat{A} \odot \hat{B}\})c \succeq -q_1 \tag{18}$$

where $\succeq$ denotes the usual *pointwise* ordering. The optimization problem in (18) is an LP problem that can be very efficiently solved using interior-point methods [20], [21].

Using the second way of slicing the 3-D array (i.e., working with the data $\tilde{y} = \mathcal{F}(\tilde{Y})$) and exploiting the property (11), we obtain that the estimate of A can be found by solving the following optimization problem:

$$\hat{a} = \arg\min_{a} \|\tilde{y} - (I_I \otimes \mathcal{G}\{\hat{B} \odot \hat{C}\})a\|_1, \quad \hat{A} = (\mathcal{F}^{-1}\{\hat{a}\})^T \tag{19}$$

with given $\tilde{y}$ and previously estimated $\hat{B}$ and $\hat{C}$. This problem can be rewritten as the following LP problem:

$$\min_{a,q_2} e^T q_2 \quad \text{subject to} \quad \tilde{y} - (I_I \otimes \mathcal{G}\{\hat{B} \odot \hat{C}\})a \preceq q_2$$
$$\tilde{y} - (I_I \otimes \mathcal{G}\{\hat{B} \odot \hat{C}\})a \succeq -q_2 \tag{20}$$

where q_2 is the vector of slack variables of commensurate dimension.

Finally, using the third way of slicing the 3-D array and applying the property (11), we obtain that the estimate of B can be found by solving the following optimization problem:

$$\hat{b} = \arg\min_{b} \|\tilde{z} - (I_J \otimes \mathcal{G}\{\hat{C} \odot \hat{A}\})b\|_1, \quad \hat{B} = (\mathcal{F}^{-1}\{\hat{b}\})^T \tag{21}$$

with given $\tilde{z}$ and previously estimated $\hat{A}$ and $\hat{C}$. This problem is equivalent to the following LP problem:

$$\min_{b,q_3} e^T q_3 \quad \text{subject to} \quad \tilde{z} - (I_J \otimes \mathcal{G}\{\hat{C} \odot \hat{A}\})b \preceq q_3$$
$$\tilde{z} - (I_J \otimes \mathcal{G}\{\hat{C} \odot \hat{A}\})b \succeq -q_3 \tag{22}$$

where q_3 is the vector of slack variables of commensurate dimension.

Fitting proceeds by updating one matrix at a time, conditioned on interim estimates of the other two, in a round-robin fashion. Note that the conditional update of any given matrix may either improve or maintain but cannot worsen the current fit. Monotone convergence of the fit (but not necessarily to the global minimum) follows directly from this observation.

Each conditional LAE matrix update entails solving an LP problem whose complexity is higher than that of the corresponding LS matrix update (matrix pseudo-inverse) [21]. If we define one iteration to be a full update cycle of all three matrices, then the per-iteration complexity of TALAE-LP

can be estimated as $\mathcal{O}(I^3J^3 + J^3K^3 + I^3K^3)$, while the corresponding per-iteration complexity of TALS is $\mathcal{O}(FIJK)$. Overall complexity depends on the number of iterations, which varies depending on problem-specific parameters and the given batch of data.

B. TALAE Based on Weighted Median Filtering (TALAE-WMF)

LP yields the optimal solution for each of the conditional optimization problems in (17), (19), and (21), at a relatively modest complexity. In the following, we show how one can *iteratively* solve (17) [and, likewise, (19) and (21)], using simple weighted median filtering. Unlike the LP-based solution, the iterative solution derived below is not necessarily an optimal solution of (17). However, the weighted median filtering iteration is simpler, monotonically convergent (and, thus, maintains monotone convergence of the outer loop as well), and, as we will see in the simulations, it does not appear to affect the performance of the overall model-fitting procedure appreciably.

Fixing all parameters in (17) except for $c_{k+(f-1)K}$ ($k \in \{1, \ldots, K\}, f \in \{1, \ldots, 2F\}$), we can simplify this problem as

$$\min_{c_{k+(f-1)K}} \left\| \breve{X}_{.,k} - \sum_{p=1,p\neq f}^{2F} \mathcal{G}\{\hat{A} \odot \hat{B}\}_{.,p} c_{k+(p-1)K} \right.$$
$$\left. - \mathcal{G}\{\hat{A} \odot \hat{B}\}_{.,f} c_{k+(f-1)K} \right\|_1 \quad (23)$$

where $\breve{X}_{.,k} = [\text{Re}\{\tilde{X}_{.,k}\}^T, \text{Im}\{\tilde{X}_{.,k}\}^T]^T$, and $\mathcal{G}\{\hat{A} \odot \hat{B}\}_{.,f}$ stands for the fth column of $\mathcal{G}\{\hat{A} \odot \hat{B}\}$.

Defining

$$h_k = \breve{X}_{.,k} - \sum_{p=1,p\neq f}^{2F} \mathcal{G}\{\hat{A} \odot \hat{B}\}_{.,p} c_{k+(p-1)K} \quad (24)$$

and

$$m_f = \mathcal{G}\{\hat{A} \odot \hat{B}\}_{.,f} \quad (25)$$

the problem in (23) becomes

$$\min_{c_{k+(f-1)K}} \left\| h_k - m_f c_{k+(f-1)K} \right\|_1. \quad (26)$$

The minimization problem in (26) can be further written as

$$\min_{c_{k+(f-1)K}} \sum_{l=1}^{2IJ} \left| h_k(l) - m_f(l) c_{k+(f-1)K} \right| \quad (27)$$

where $h_k(l)$ and $m_f(l)$ are the lth elements of the vectors h_k and m_f, respectively. Equivalently, (27) can be expressed as

$$\min_{c_{k+(f-1)K}} \sum_{l=1}^{2IJ} |m_f(l)| \left| \frac{h_k(l)}{m_f(l)} - c_{k+(f-1)K} \right| \quad (28)$$

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 (28) can be dropped because it becomes a constant in this case (independent of $c_{k+(f-1)K}$). Note that the optimization problem (28) is solved by WMF (e.g., see [19]) 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 C, e.g., in a circular fashion, and likewise for the elements of the matrices A and 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 close to the per-iteration complexity of TALS, and much lower than the corresponding complexity of TALAE-LP.

Note that the proposed robust PARAFAC algorithms (as any alternating-optimization-based technique) can use any additional side information by keeping the corresponding columns of the respective matrices fixed during iterations or initializing them with preliminary (possibly coarse) estimates. The advantage of PARAFAC fitting versus other approaches in this case is that it uses all the model structure and aims for the ML solution.

IV. CRAMÉR–RAO BOUNDS

The PARAFAC CRB for Gaussian noise has been derived in [22]. 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 [13].

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 A and B are normalized to $[1, \ldots, 1]_{1\times F}$, which resolves the scale ambiguity [22]. Further, we assume that the first row of C is known and consists of distinct elements, which resolves the permutation ambiguity [22]. Then, we can write the $1 \times (I + J + K - 3)F$ row-vector of unknown complex parameters as

$$\theta = \left[a_2^T, \ldots, a_I^T, b_2^T, \ldots, b_J^T, c_2^T, \ldots, c_K^T \right]. \quad (29)$$

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

$$\text{FIM} = \begin{bmatrix} \Psi_1 & \Psi_2 \\ \Psi_2^H & \Psi_{c,c} \end{bmatrix} \quad (30)$$

where

$$\Psi_1 = \begin{bmatrix} \Psi_{a,a} & \Psi_{a,b} \\ \Psi_{a,b}^H & \Psi_{b,b} \end{bmatrix}, \quad \Psi_2 = \begin{bmatrix} \Psi_{a,c} \\ \Psi_{b,c} \end{bmatrix} \quad (31)$$

$$\Psi_{a,a} = \beta[I_{I-1} \otimes ((B \odot C)^H (B \odot C))] \qquad (32)$$

$$\Psi_{b,b} = \beta[I_{J-1} \otimes ((C \odot A)^H (C \odot A))] \qquad (33)$$

$$\Psi_{c,c} = \beta[I_{K-1} \otimes ((A \odot B)^H (A \odot B))] \qquad (34)$$

$$\Psi_{a,b} = \beta \left[e \otimes (B \odot C)^H \right] V_{1,2} \left[e^T \otimes (C \odot A) \right] \qquad (35)$$

$$\Psi_{a,c} = \beta[e \otimes (B \odot C)^H] V_{1,3} [e^T \otimes (A \odot B)] \qquad (36)$$

$$\Psi_{b,c} = \beta[e \otimes (C \odot A)^H] V_{2,3} [e^T \otimes (A \odot B)] \qquad (37)$$

$$\beta = \begin{cases} \frac{1}{\sigma_v^2}, & \text{for Gaussian noise} \\ \frac{2}{\sigma_v^2}, & \text{for Laplacian noise} \\ \frac{3}{10\gamma^2}, & \text{for complex Cauchy noise} \\ \frac{1}{2\gamma^2}, & \text{for real Cauchy noise} \end{cases} \qquad (38)$$

σ_v is the standard deviation of the Gaussian or Laplacian distribution, $\gamma > 0$ is the *dispersion* of the Cauchy distribution and

$$V_{1,2} = \begin{bmatrix} V_{1,2}(2,2) & \cdots & V_{1,2}(2,J) \\ \vdots & \ddots & \vdots \\ V_{1,2}(I,2) & \cdots & V_{1,2}(I,J) \end{bmatrix} \qquad (39)$$

$$V_{1,3} = \begin{bmatrix} V_{1,3}(2,2) & \cdots & V_{1,3}(2,K) \\ \vdots & \ddots & \vdots \\ V_{1,3}(I,2) & \cdots & V_{1,3}(I,K) \end{bmatrix} \qquad (40)$$

$$V_{2,3} = \begin{bmatrix} V_{2,3}(2,2) & \cdots & V_{2,3}(2,K) \\ \vdots & \ddots & \vdots \\ V_{2,3}(J,2) & \cdots & V_{2,3}(J,K) \end{bmatrix}. \qquad (41)$$

We have (42)–(44), shown at the bottom of the page.

Finally, the CRB matrix of the unknown elements of C is given by

$$\text{CRB}_{c,c} = \left(\Psi_{c,c} - \Psi_2^H \Psi_1^{-1} \Psi_2 \right)^{-1}. \qquad (45)$$

Proof: See Appendix B.   □

V. SIMULATIONS

In this section, we compare the performance of the proposed TALAE-LP and TALAE-WMF algorithms with that of the conventional TALS method, and against the pertinent CRB. The example of fully blind PARAFAC multiuser separation-detection for a direct-sequence code-division multiple access

$$V_{1,2}(i,j) = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow (j-1)K+1 \\ \leftarrow (j-1)K+2 \\ \vdots \\ \leftarrow (j-1)K+K \\ \\ \end{matrix} \qquad (42)$$
$$\begin{matrix} \uparrow & & \uparrow & & \uparrow \\ i & & I+i & & (K-1)I+i \end{matrix}$$

$$V_{1,3}(i,k) = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & & & & \ddots & & & & & \\ 0 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & & & & \ddots & & & & & \\ 0 & \cdots & 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & & & & \ddots & & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{matrix} \\ \leftarrow \quad k \\ \vdots \\ \leftarrow \ k+(j-1)K \\ \vdots \\ \leftarrow \ k+(J-1)K \end{matrix} \qquad (43)$$
$$\begin{matrix} \uparrow & \uparrow & & & \uparrow \\ (i-1)J+1 & (i-1)J+2 & & (i-1)J+J \end{matrix}$$

$$V_{2,3}(j,k) = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow (k-1)I+1 \\ \leftarrow (k-1)I+2 \\ \vdots \\ \leftarrow (k-1)I+I \\ \\ \end{matrix} \qquad (44)$$
$$\begin{matrix} \uparrow & & \uparrow & & \uparrow \\ j & & J+j & & (I-1)J+j \end{matrix}$$

(DS-CDMA) communication system [6] is simulated.[3] For the DS-CDMA application, the elements of the matrices A, B, and C have the following meanings:

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

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

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

Correspondingly, matrix A is the channel matrix, B is the spreading code matrix, and C is the user symbol matrix, all unknown to the receiver. Here

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).

See [6] for a list of the necessary assumptions which make the PARAFAC model valid in this context.

The data $\underline{X}$ 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.

For LS fitting, we use the COMFAC algorithm [12], which is a fast implementation of TALS. The MOSEK convex optimization MATLAB toolbox [25] is used to solve LP problems associated with our TALAE-LP algorithm. The interior-point based solver of the MOSEK toolbox is applied. Scale and permutation ambiguities are inherent to this blind separation problem [6]; the scale ambiguity manifests itself as a complex constant that multiplies each individual row of 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 $(C, \hat{C})$ column-matching algorithm.

We present Monte Carlo simulations that are designed to assess the root-mean square error (RMSE) performance of the

[3]Note that, in contrast to the existing robust/blind multiuser detection algorithms (e.g., see [23], [24], and references therein) where the space-time signature of the user-of-interest has to be known, in our example, we assume that the spatio-temporal signatures of all users are *completely* unknown.

aforementioned algorithms. The parameters used in the simulations are as follows: N = number of Monte Carlo trials = 100; $J = 8$; $K = 20$; 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

$$\phi(\omega) = \exp\{-\gamma|\omega|^\alpha\} \tag{46}$$

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 ([26, p. 68])

$$\text{SNR} = \frac{1}{C_\text{g}} \left(\frac{A}{S_0}\right)^2 \tag{47}$$

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 ([26, p. 38])

$$S_0 = \frac{(C_\text{g}\gamma)^{1/\alpha}}{C_\text{g}}. \tag{48}$$

Both Gaussian and Cauchy distributions belong to the class of symmetric α-stable distributions. The geometric power of complex Gaussian noise can be calculated as [26], shown in (49) at at the bottom of the page. Substituting (49) into (47), we find that for the Gaussian case the geometric SNR is equivalent to the standard SNR.

For the complex Cauchy case, the noise power and corresponding geometric SNR can be written as (50) and (51), respectively, shown at the bottom of the page. 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

$$S_{0,\text{Gaussian}} = \exp\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\log\sqrt{\text{Re}\{v_{i,j,k}\}^2 + \text{Im}\{v_{i,j,k}\}^2}}{\pi\sigma_\text{v}^2} \right. \\ \left. \times \exp\left\{-\frac{\text{Re}\{v_{i,j,k}\}^2 + \text{Im}\{v_{i,j,k}\}^2}{\sigma_\text{v}^2}\right\} d\text{Re}\{v_{i,j,k}\}\, d\text{Im}\{v_{i,j,k}\}\right\} = \frac{\sigma_\text{v}}{\sqrt{C_\text{g}}} \tag{49}$$

$$S_{0,\text{Cauchy}} = \exp\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \log\sqrt{\text{Re}\{v_{i,j,k}\}^2 + \text{Im}\{v_{i,j,k}\}^2} \right. \\ \left. \times \frac{\gamma}{2\pi(\text{Re}\{v_{i,j,k}\}^2 + \text{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{3/2}} d\text{Re}\{v_{i,j,k}\}\, d\text{Im}\{v_{i,j,k}\}\right\} = 2\gamma \tag{50}$$

and

$$\text{SNR}_{\text{Cauchy}} = \frac{A^2}{4C_\text{g}\gamma^2} \tag{51}$$

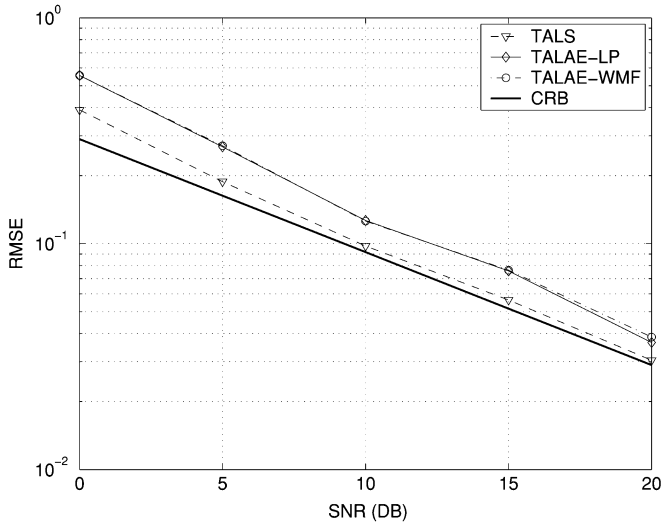Fig. 1.   RMSEs versus SNR. First example; Gaussian channel noise.

Fig. 2.   RMSEs versus SNR. First example; Laplacian channel noise.

Monte Carlo trial. BPSK modulation is used for all user signals. The gains of the channel matrix A and the elements of the spreading code matrix 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 choose to 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:

$$\text{RMSE} = \sqrt{\frac{1}{N(K-1)F} \sum_{n=1}^{N} \|\hat{C}(n) - C\|_F^2} \qquad (52)$$

while the (averaged) CRB is calculated as

$$\text{CRB} = \sqrt{\sum_{n=1}^{N} \frac{\text{Tr}\{\text{CRB}_{c,c}(n)\}}{N(K-1)F}} \qquad (53)$$

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

In the first example, we assume $I = 8$ sensors and $F = 2$ users. Figs. 1–3 display the performance of the aforementioned algorithms in terms of RMSE versus the SNR[4] for the case of Gaussian, Laplacian and Cauchy noise, respectively, and compare the performance with the corresponding CRBs.

A conceptually simple method that is often employed in engineering practice to deal with impulsive noise and outliers is to use a memoryless *clipper* nonlinearity (e.g., [24]) followed by ordinary LS fitting. Fig. 4 shows the performance of the proposed TALAE methods as compared to the TALS procedure
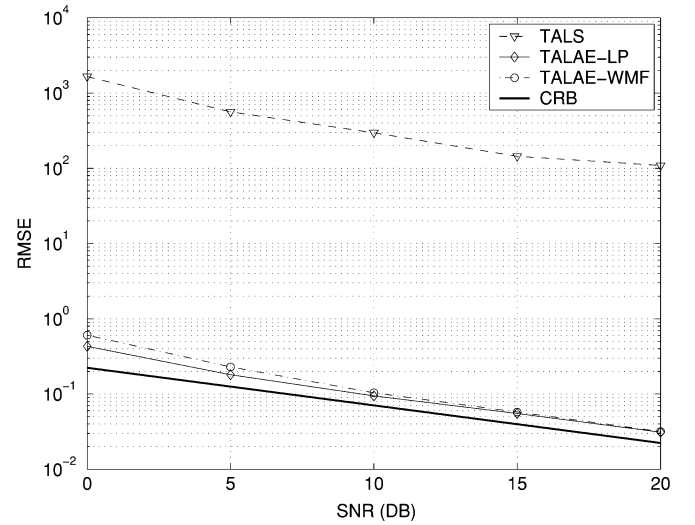

Fig. 3.   RMSEs versus SNR. First example; Cauchy channel noise.

with clipper nonlinearity. The input-output relationship of the clipper nonlinearity is

$$c(x) = \begin{cases} x, & |x| < \tau \\ \tau \text{sgn}(x), & \text{otherwise} \end{cases}$$

where τ denotes the threshold of the clipper.

In the second example, SNR $= 15$ dB is fixed while the numbers of sensors and users are varied. In this example, Cauchy noise is assumed. Figs. 5 and 6 show the RMSEs of the TALAE-LP and TALAE-WMF methods versus the number of antenna elements for $F = 2$ and $F = 4$, respectively.

Figs. 1 and 2 demonstrate that in the case of Gaussian noise, the TALS method performs slightly better than the proposed robust algorithms, while in the case of Laplacian noise, the proposed robust algorithms have slightly better performance as compared to the TALS method. In the case of Cauchy noise, the TALS method breaks down (Figs. 3, 5, and 6), while the performance of the proposed robust algorithms is not affected and is close to the CRB (despite the fact that our techniques
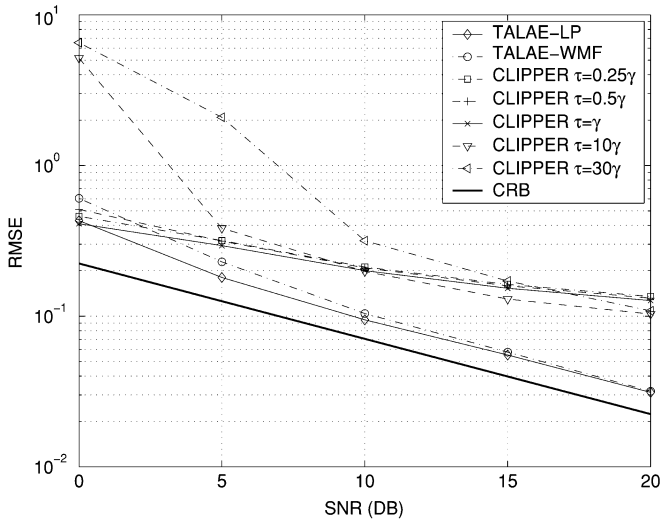
---

[4]Standard SNR in Gaussian and Laplacian case and geometric SNR in Cauchy case.

Fig. 4.   RMSEs versus SNR. First example; Cauchy channel noise. γ is the dispersion of the Cauchy noise.
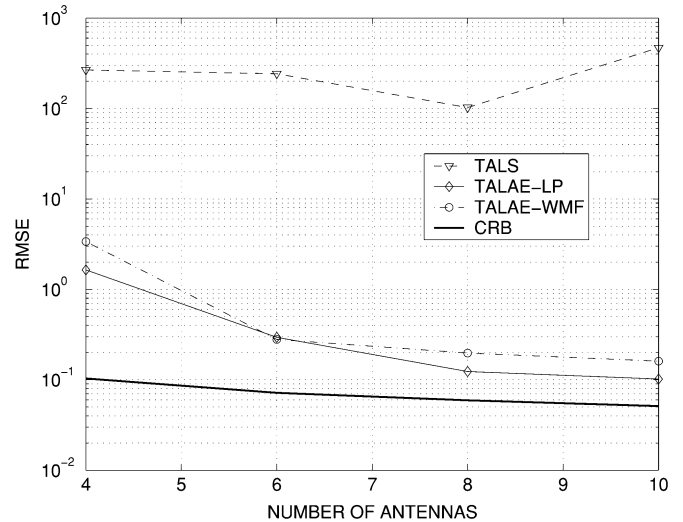

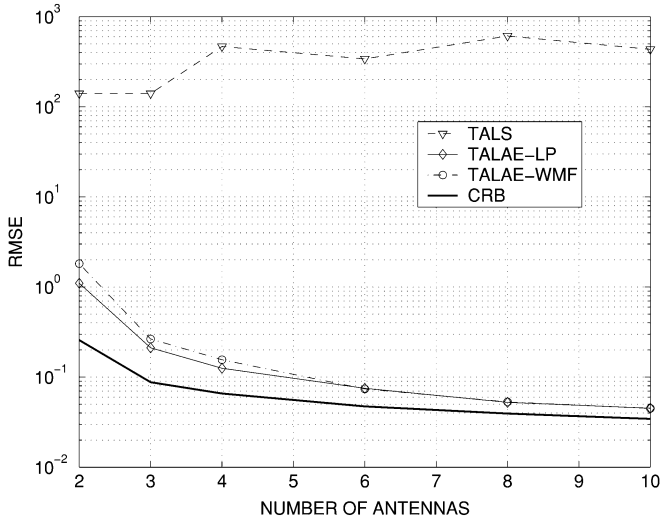Fig. 6.   RMSEs versus I for $F = 4$. Second example; Cauchy channel noise.

Fig. 5.   RMSEs versus I for $F = 2$. Second example; Cauchy channel noise.

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.

The TALS procedure with clipper nonlinearity shows reasonably good performance in the case of Cauchy noise (Fig. 4), although the clipper-plus-TALS results are uniformly worse than those of TALAE-LP and TALAE-WMF for all choices of clipper threshold considered. Notice that the performance of clipper-plus-TALS technique depends critically on the clipper threshold, and different thresholds should be chosen to achieve the best performance at different SNRs. The TALAE-LP/WMF techniques do not require such calibration. Furthermore, TALAE-WMF only incurs a small complexity penalty relative to TALS ($\mathcal{O}(FIJK\log(IJK))$) versus $\mathcal{O}(FIJK)$ per iteration). Based on these observations, we conclude that TALAE-WMF is to be preferred over the clipper-plus-TALS approach.

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 performance in our simulations. Further note that the two TALAE regressions appear robust against random initialization. Intuitively, this fact can be attributed to the fact that we work with a relatively over-determined model.

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 $\boldsymbol{A}$ and $\boldsymbol{B}$. Throughout the simulations it was observed that TALAE-WMF requires two to five more iterations than TALAE-LP for Gaussian or Laplacian noise, and about the same number of iterations for Cauchy noise. This is indeed a positive 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. Conclusion

Two new iterative algorithms for robust fitting of trilinear PARAFAC models have been proposed. The algorithms rely on alternating optimization, using LP or WMF. Our findings show that both algorithms are computationally efficient, but the WMF iteration is particularly appealing from a simplicity viewpoint. The proposed algorithms outperform the workhorse alternating LS PARAFAC fitting procedure under heavy-tailed noise, and,

even though they are matched to the Laplacian distribution, they still perform well under Cauchy noise. Furthermore, their performance degrades only moderately under Gaussian noise. We, therefore, expect that the new algorithms will prove useful in a variety of applications of PARAFAC analysis. Finally, we note that both TALAE-LP and TALAE-WMF can be readily generalized to multilinear LAE regression in higher dimensions (i.e., for $M > 3$-way arrays).

APPENDIX A
PROOF OF PROPERTY (11)

Let $\boldsymbol{D}$ and $\boldsymbol{F}$ be $M \times N$ and $N \times L$, respectively, complex-valued matrices. The lth column of $\boldsymbol{DF}$ satisfies

$$
[\breve{\boldsymbol{DF}}]_{\cdot,l} = \begin{bmatrix} \mathrm{Re}\{[\boldsymbol{DF}]_{\cdot,l}\} \\ \mathrm{Im}\{[\boldsymbol{DF}]_{\cdot,l}\} \end{bmatrix}
$$
$$
= \begin{bmatrix} \mathrm{Re}\{\boldsymbol{D}\} & -\mathrm{Im}\{\boldsymbol{D}\} \\ \mathrm{Im}\{\boldsymbol{D}\} & \mathrm{Re}\{\boldsymbol{D}\} \end{bmatrix} \begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}_{\cdot,l}\} \\ \mathrm{Im}\{\boldsymbol{F}_{\cdot,l}\} \end{bmatrix} = \mathcal{G}\{D\}\breve{\boldsymbol{F}}_{\cdot,l}
$$

(54)

and, therefore

$$
\mathcal{F}\{\boldsymbol{DF}\} = \begin{bmatrix} [\breve{\boldsymbol{DF}}]_{\cdot,1} \\ \vdots \\ [\breve{\boldsymbol{DF}}]_{\cdot,L} \end{bmatrix} = \begin{bmatrix} \mathcal{G}\{D\} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \mathcal{G}\{D\} \end{bmatrix} \begin{bmatrix} \breve{\boldsymbol{F}}_{\cdot,1} \\ \vdots \\ \breve{\boldsymbol{F}}_{\cdot,L} \end{bmatrix}
$$
$$
= (\boldsymbol{I}_L \otimes \mathcal{G}\{D\})\mathcal{F}\{\boldsymbol{F}\}.
$$

(55)

APPENDIX B
PROOF OF CRB RESULT

The proof for the case of Gaussian noise is given in [22]. Here we provide proofs for the Laplacian and Cauchy cases only.

Laplacian Case: The likelihood function in this case can be written as [27]

$$
f_{\mathrm{Laplacian}}(\tilde{\underline{\boldsymbol{X}}} | \boldsymbol{\theta})
$$
$$
= \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \frac{1}{\sigma_{\mathrm{v}}^2}
$$
$$
\times \exp\left\{ -\frac{2}{\sigma_{\mathrm{v}}} \left(\left| \mathrm{Re}\left\{ \tilde{x}_{i,j,k} - \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \right\} \right| \right. \right.
$$
$$
\left. \left. + \left| \mathrm{Im}\left\{ \tilde{x}_{i,j,k} - \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \right\} \right| \right) \right\}.
$$

(56)

The corresponding log-likelihood (LL) function is given by

$$
L_{\mathrm{Laplacian}}(\boldsymbol{\theta})
$$
$$
= IJK \ln \frac{1}{\sigma_{\mathrm{v}}^2}
$$
$$
- \frac{2}{\sigma_{\mathrm{v}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\left| \mathrm{Re}\left\{ \tilde{x}_{i,j,k} - \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \right\} \right| \right.
$$
$$
\left. + \left| \mathrm{Im}\left\{ \tilde{x}_{i,j,k} - \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \right\} \right| \right).
$$

(57)

The complex FIM for log-likelihood $L_{\mathrm{Laplacian}}(\boldsymbol{\theta})$ can be expressed as

$$
\mathrm{FIM} = \mathrm{E}\left\{ \left(\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^H \left(\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right\}
$$

(58)

where $\mathrm{E}\{\cdot\}$ denotes the expectation operator.

Taking partial derivatives of $L_{\mathrm{Laplacian}}(\boldsymbol{\theta})$ with respect to the unknown parameters, we obtain

$$
\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial a_{i,f}} = \frac{1}{2}\left[\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial \mathrm{Re}\{a_{i,f}\}} - \sqrt{-1}\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial \mathrm{Im}\{a_{i,f}\}} \right]
$$
$$
= \frac{1}{\sigma_{\mathrm{v}}} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\mathrm{Re}\{v_{i,j,k}\}b_{j,f}c_{k,f}}{|\mathrm{Re}\{v_{i,j,k}\}|}
$$
$$
- \sqrt{-1}\frac{\mathrm{Im}\{v_{i,j,k}\}b_{j,f}c_{k,f}}{|\mathrm{Im}\{v_{i,j,k}\}|}
$$

(59)

$$
\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial b_{j,f}} = \frac{1}{\sigma_{\mathrm{v}}} \sum_{i=1}^{I} \sum_{k=1}^{K} \frac{\mathrm{Re}\{v_{i,j,k}\}a_{i,f}c_{k,f}}{|\mathrm{Re}\{v_{i,j,k}\}|}
$$
$$
- \sqrt{-1}\frac{\mathrm{Im}\{v_{i,j,k}\}a_{i,f}c_{k,f}}{|\mathrm{Im}\{v_{i,j,k}\}|}
$$

(60)

$$
\frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial c_{k,f}} = \frac{1}{\sigma_{\mathrm{v}}} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\mathrm{Re}\{v_{i,j,k}\}a_{i,f}b_{j,f}}{|\mathrm{Re}\{v_{i,j,k}\}|}
$$
$$
- \sqrt{-1}\frac{\mathrm{Im}\{v_{i,j,k}\}a_{i,f}b_{j,f}}{|\mathrm{Im}\{v_{i,j,k}\}|}.
$$

(61)

Next, let us derive expressions for the elements of FIM, starting from the element

$$
\mathrm{E}\left\{ \frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial a_{i_1,f_1}^*} \frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial a_{i_2,f_2}} \right\}
$$
$$
= \mathrm{E}\left\{ \frac{1}{\sigma_{\mathrm{v}}^2} \sum_{j_1=1}^{J} \sum_{k_1=1}^{K} \sum_{j_2=1}^{J} \sum_{k_2=1}^{K} \left(\frac{\mathrm{Re}\{v_{i_1,j_1,k_1}\}b_{j_1,f_1}^* c_{k_1,f_1}^*}{|\mathrm{Re}\{v_{i_1,j_1,k_1}\}|} \right. \right.
$$
$$
+ \sqrt{-1}\frac{\mathrm{Im}\{v_{i_1,j_1,k_1}\}b_{j_1,f_1}^* c_{k_1,f_1}^*}{|\mathrm{Im}\{v_{i_1,j_1,k_1}\}|} \right)
$$
$$
\times \left(\frac{\mathrm{Re}\{v_{i_2,j_2,k_2}\}b_{j_2,f_2}c_{k_2,f_2}}{|\mathrm{Re}\{v_{i_2,j_2,k_2}\}|} \right.
$$
$$
\left. \left. - \sqrt{-1}\frac{\mathrm{Im}\{v_{i_2,j_2,k_2}\}b_{j_2,f_2}c_{k_2,f_2}}{|\mathrm{Im}\{v_{i_2,j_2,k_2}\}|} \right) \right\}
$$
$$
= \frac{2}{\sigma_{\mathrm{v}}^2} \sum_{j=1}^{J} \sum_{k=1}^{K} b_{j,f_1}^* c_{k,f_1}^* b_{j,f_2} c_{k,f_2} \delta_{i_1,i_2}
$$
$$
= \frac{2}{\sigma_{\mathrm{v}}^2} \boldsymbol{e}_{f_1}^T (\boldsymbol{B} \odot \boldsymbol{C})^H (\boldsymbol{B} \odot \boldsymbol{C}) \boldsymbol{e}_{f_2} \delta_{i_1,i_2}
$$

(62)

where $\boldsymbol{e}_f$ is the fth unit coordinate vector and $\delta_{i,l}$ stands for the Kronecker delta. In (62), we use the fact that the real and imaginary parts of $v_{i,j,k}$ are statistically independent.

Thus, the matrix containing all the elements (62) ($i_1, i_2 = 2, \ldots, I$ and $f_1, f_2 = 1, \ldots, F$) $\boldsymbol{\Psi}_{\boldsymbol{a},\boldsymbol{a}}$ can be expressed in the form of (32) with the coefficient $\beta = 2/\sigma_{\mathrm{v}}^2$. In the same way,

we compute the matrices $\boldsymbol{\Psi}_{b,b}$ and $\boldsymbol{\Psi}_{c,c}$ in the form of (33) and (34), respectively.

Similarly, we can find the element

$$
\begin{aligned}
\mathrm{E}&\left\{ \frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial a_{i,f_1}^*} \frac{\partial L_{\mathrm{Laplacian}}(\boldsymbol{\theta})}{\partial b_{j,f_2}} \right\} \\
&= \frac{2}{\sigma_{\mathrm{v}}^2} \sum_{k=1}^{K} b_{j,f_1}^* c_{k,f_1}^* a_{i,f_2} c_{k,f_2} \\
&= \frac{2}{\sigma_{\mathrm{v}}^2} \boldsymbol{e}_{f_1}^T (\boldsymbol{B} \odot \boldsymbol{C})^H \boldsymbol{V}_{1,2}(i,j)(\boldsymbol{C} \odot \boldsymbol{A}) \boldsymbol{e}_{f_2}.
\end{aligned} \tag{63}
$$

Then, the matrix containing all the elements (63) ($i = 2,\ldots,I, j = 2,\ldots,J$ and $f_1, f_2 = 1,\ldots,F$) $\boldsymbol{\Psi}_{a,b}$ can be written in the form of (35) with the coefficient $\beta = 2/\sigma_{\mathrm{v}}^2$. In the same way, we find the matrices $\boldsymbol{\Psi}_{a,c}$ and $\boldsymbol{\Psi}_{b,c}$ given by (36) and (37), respectively. Applying the matrix inversion lemma to the FIM matrix, we obtain (45).

Cauchy Case: We have seen that the Laplacian CRB for the PARAFAC model only differs by a multiplicative constant from the corresponding Gaussian CRB. In [28], it is shown that this is true for general signal models observed in i.i.d. additive noise, provided that the noise probability density function (pdf) possesses everywhere continuous first and second derivatives. This is not the case for the Laplacian pdf, due to the discontinuity at the origin; however, it is true for the Cauchy. In fact, in [28] and [29], it is shown that the noise pdf-dependent multiplicative constant that appears in the FIM can be computed as

$$
\beta = \int_{-\infty}^{+\infty} \frac{|p'(v)|^2}{p(v)} \, dv \tag{64}
$$

where $p(\cdot)$ is the noise pdf and $p'(\cdot)$ is its first derivative. Hence, we can proceed in this fashion, calculating the integral above for the Cauchy pdf.

The expression for complex isotropic Cauchy density function in our case is given by [13]

$$
\begin{aligned}
&p_{\mathrm{Cauchy}}(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\}) \\
&= \frac{\gamma}{2\pi(\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{3/2}}.
\end{aligned} \tag{65}
$$

The first derivative of the pdf (65) can be easily calculated as

$$
\begin{aligned}
&p_{\mathrm{Cauchy}}'(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\}) \\
&= \frac{1}{2} \left[\frac{\partial p_{\mathrm{Cauchy}}(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\})}{\partial \mathrm{Re}\{v_{i,j,k}\}} \right. \\
&\qquad \left. - \sqrt{-1} \frac{\partial p_{\mathrm{Cauchy}}(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\})}{\partial \mathrm{Im}\{v_{i,j,k}\}} \right] \\
&= -\frac{3\gamma}{4} \left[\frac{2\mathrm{Re}\{v_{i,j,k}\}}{2\pi(\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{5/2}} \right. \\
&\qquad \left. - \sqrt{-1} \frac{2\mathrm{Im}\{v_{i,j,k}\}}{2\pi(\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{5/2}} \right] \\
&= \frac{3\gamma}{4} \frac{-\mathrm{Re}\{v_{i,j,k}\} + \sqrt{-1}\mathrm{Im}\{v_{i,j,k}\}}{\pi(\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{5/2}}
\end{aligned} \tag{66}
$$

and, correspondingly

$$
\begin{aligned}
&\frac{|p_{\mathrm{Cauchy}}'(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\})|^2}{p_{\mathrm{Cauchy}}(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\})} \\
&= \frac{9\gamma}{4} \frac{\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2}{2\pi(\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{7/2}}.
\end{aligned} \tag{67}
$$

Substituting (67) into (64) and calculating the integral, we finally find the coefficient β for the Cauchy noise case, shown in (68) at the bottom of the page, where $r = \sqrt{\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2}$ and $\varphi = \arctan((\mathrm{Im}\{v_{i,j,k}\})/(\mathrm{Re}\{v_{i,j,k}\}))$ are the polar coordinates.

In the proof above, we considered complex noise. However, it is worth noting that in the particular case of real Cauchy noise, the coefficient β is equal to $1/2\gamma^2$. Indeed, the real Cauchy pdf is

$$
p_{\mathrm{Cauchy}}(v_{i,j,k}) = \frac{\gamma}{\pi \left(v_{i,j,k}^2 + \gamma^2 \right)} \tag{69}
$$

and its first derivative is

$$
p_{\mathrm{Cauchy}}'(v_{i,j,k}) = -\gamma \frac{2v_{i,j,k}}{\pi \left(v_{i,j,k}^2 + \gamma^2 \right)^2}. \tag{70}
$$

$$
\begin{aligned}
\beta &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{|p_{\mathrm{Cauchy}}'(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\})|^2}{p_{\mathrm{Cauchy}}(\mathrm{Re}\{v_{i,j,k}\}, \mathrm{Im}\{v_{i,j,k}\})} \, d\mathrm{Re}\{v_{i,j,k}\} \, d\mathrm{Im}\{v_{i,j,k}\} \\
&= \frac{9\gamma}{4} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2}{2\pi(\mathrm{Re}\{v_{i,j,k}\}^2 + \mathrm{Im}\{v_{i,j,k}\}^2 + \gamma^2)^{7/2}} \, d\mathrm{Re}\{v_{i,j,k}\} \, d\mathrm{Im}\{v_{i,j,k}\} \\
&= \frac{9\gamma}{4} \int_{0}^{+\infty} \int_{0}^{2\pi} \frac{r^2}{2\pi(\gamma^2 + r^2)^{7/2}} r \, dr \, d\varphi \\
&= \frac{9\gamma}{4} \int_{0}^{+\infty} \frac{r^3}{(\gamma^2 + r^2)^{7/2}} \, dr = \frac{9\gamma}{4} \frac{2}{15\gamma^3} = \frac{3}{10\gamma^2}
\end{aligned} \tag{68}
$$

Substituting (69) and (70) into (64), we obtain

$$
\begin{aligned}
\beta &= \int_{-\infty}^{+\infty} \frac{(p'_{\text{Cauchy}}(v_{i,j,k}))^2}{p_{\text{Cauchy}}(v_{i,j,k})}\, dv_{i,j,k} \\
&= \frac{4\gamma}{\pi} \int_{-\infty}^{+\infty} \frac{v_{i,j,k}^2}{\left(v_{i,j,k}^2 + \gamma^2\right)^3}\, dv_{i,j,k} = \frac{8\gamma}{\pi}\frac{\pi}{16\gamma^3} = \frac{1}{2\gamma^2}
\end{aligned}
\tag{71}
$$

and the proof is complete.

ACKNOWLEDGMENT

REFERENCES

[1] R. A. Harshman, "Foundations of PARAFAC procedure: Models and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[2] ——, "Determination and proof of minimum uniqueness conditions for PARAFAC1," *UCLA Working Papers in Phonetics*, vol. 22, pp. 111–117, 1972.

[3] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," in *Research Methods for Multimode Data Analysis*, H. G. Law, C. W. Snyder, Jr., J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.

[4] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometr.*, vol. 35, no. 3, pp. 283–319, 1970.

[5] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.

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

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

[8] P. Comon, "Tensor decompositions," in *Mathematics in Signal Processing V*, J. G. McWhirter and I. K. Proudler, Eds. Oxford, U.K.: Clarendon, 2002, pp. 1–24.

[9] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation and applications of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[10] ——, "Independent component analysis and (simultaneous) third-order tensor diagonalization," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2262–2271, Oct. 2001.

[11] E. Kofidis and P. A. Regalia, "On the best rank-1 approximation of higher-order super-symmetric tensors," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 3, pp. 863–884, 2002.

[12] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures," presented at the *Int. Workshop Independent Component Analysis and Blind Signal Separation*, Paris, France, Jan. 1999, http://www.ece.umn.edu/users/nikos/public_html/3SPICE/code.html.

[13] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2700–2713, Nov. 1995.

[14] P. Tsakalides, "Array signal processing with alpha-stable distributions," Ph.D. dissertation, Univ. Southern California, Los Angeles, Dec. 1995.

[15] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[16] N. D. Sidiropoulos and R. Bro, "Mathematical programming algorithms for regression-based filtering in $\mathbb{R}^N$," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 771–782, Mar. 1999.

[17] R. J. Kozick and B. M. Sadler, "Maximum-Likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.

[18] R. S. Blum, R. J. Kozick, and B. M. Sadler, "An adaptive spatial diversity receiver for non-Gaussian interference and noise," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2100–2111, Aug. 1999.

[19] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filtering under structural constraints," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 591–603, Mar. 1995.

[20] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, no. 4, pp. 373–395, 1984.

[21] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

[22] X. Liu and N. D. Sidiropoulos, "Cramér–Rao lower bounds for low-rank decomposition of multidimantional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.

[23] X. Wang and H. V. Poor, "Space-time multiuser detection in multipath CDMA channels," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2356–2374, Sep. 1999.

[24] S. N. Batalama, M. J. Medley, and I. N. Psaromiligkos, "Adaptive robust spread-spectrum receivers," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 905–917, Jun. 1999.

[25] The MOSEK Optimization Tools Version 2.0 (Build 19). User's Manual and Reference (2001). [Online]. Available: http://www.mosek.com

[26] J. G. Gonzalez, "Robust techniques for wireless communications in non-Gaussian environments," Ph.D. dissertation, Univ. Delaware, Newark, 1997.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[28] A. Swami, "Cramér-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Process.*, vol. 53, no. 2–3, pp. 231–244, Sep. 1996.

[29] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Process.*, vol. 82, no. 12, pp. 1829–1846, Dec. 2002.

Sergiy A. Vorobyov (M'02) was born in Ukraine in 1972. He received the diploma (M.S.) and Ph.D. degrees in systems and control from Kharkiv National University of Radioelectronics (KNUR), Kharkiv, Ukraine, in 1994 and 1997, respectively.

From 1995 to 2000, he was with the Control and Systems Research Laboratory at KNUR, where he became a Senior Research Scientist in 1999. From 1999 to 2001, he was with the Brain Science Institute, RIKEN, Tokyo, Japan, as a Research Scientist. From 2001 to 2003, he was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, as a Postdoctoral Fellow. From 2003 to 2004, he was with the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. Since April 2005, he has been a Senior Research Associate with the Communications Systems Group, Darmstadt University of Technology, Darmstadt, Germany. He also held short-term visiting appointments with the Institute of Applied Computer Science, Karlsruhe, Germany, in 1996, Gerhard-Mercator University, Duisberg, in 2002, and McMaster University, Hamilton, ON, Canada and Technion–Israel Institute of Technology, Haifa, Israel, both in 2005. His research interests include control theory, statistical array signal processing, blind source separation, robust adaptive beamforming, and wireless and multicarrier communications.

Dr. Vorobyov was a recipient of the 1996–1998 Young Scientist Fellowship of the Ukrainian Cabinet of Ministers, the 1996 and 1997 Young Scientist Research Grants from the George Soros Foundation, and the 1999 DAAD Fellowship (Germany). He received the 2004 IEEE Signal Processing Society Best Paper Award.

Yue Rong (S'03) was born in 1976 in Jiangsu, China. In 1999, he received the Bachelor degrees from Shanghai Jiao Tong University, Shanghai, China, both in electrical and computer engineering. He received the M.Sc. degree in computer science and communication engineering from the University of Duisburg-Essen, Duisburg, Germany, in 2002. Currently, he is working toward the Ph.D. degree at the Department of Communication Systems, Darmstadt University of Technology, Darmstadt, Germany.

From April 2001 to April 2002, he was a student research assistant at the Fraunhofer Institute of Microelectronic Circuits and Systems. From October 2001 to March 2002, he was with the Application-Specific Integrated Circuit Design Department, Nokia Ltd., Bochum, Germany. His research interests include signal processing for communications, MIMO communication systems, multicarrier communications, statistical and array signal processing, and parallel factor analysis.

Mr. Rong received the 2001–2002 Graduate Sponsoring Asia scholarship of the DAAD/ABB (Germany) and the 2004 Outstanding Ph.D. Student Abroad Award (China).

Nicholas D. Sidiropoulos (M'92–SM'99) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park (UMCP), in 1988, 1990 and 1992, respectively.

From 1988 to 1992, he was a Fulbright Fellow and a Research Assistant at the Institute for Systems Research (ISR), UMCP. From September 1992 to June 1994, he served his military service as a Lecturer in the Hellenic Air Force Academy. From October 1993 to June 1994, he also was a member of the technical staff, Systems Integration Division, G-Systems Ltd., Athens, Greece. He was a Postdoctoral Fellow (1994 to 1995) and Research Scientist (1996 to 1997) at ISR-UMCP, an Assistant Professor with the Department of Electrical Engineering, University of Virginia, Charlottesville, from 1997 to 1999, and an Associate Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, from 2000 to 2002. He is currently a Professor with the Telecommunications Division of the Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Crete, Greece, and Adjunct Professor at the University of Minnesota. His current research interests are primarily in signal processing for communications, and multiway analysis. He is an active consultant for industry in the areas of frequency hopping systems and signal processing for xDSL modems.

Dr. Sidiropoulos is a member of both the Signal Processing for Communications (SPCOM) and Sensor Array and Multichannel Signal Processing (SAM) Technical Committees of the IEEE Signal Processing Society and currently serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. From 2000 to 2002, he also served as Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He received the NSF/CAREER award (Signal Processing Systems Program) in June 1998 and an IEEE Signal Processing Society Best Paper Award in 2001.

Alex B. Gershman (M'97–SM'98) received the Diploma (M.Sc.) and Ph.D. degrees in radiophysics from the Nizhny Novgorod State University, Nizhny Novgorod, Russia, in 1984 and 1990, respectively.

From 1984 to 1989, he was with the Radiotechnical and Radiophysical Institutes, Nizhny Novgorod. From 1989 to 1997, he was with the Institute of Applied Physics, Russian Academy of Science, Nizhny Novgorod, as a Senior Research Scientist. From the summer of 1994 until the beginning of 1995, he was a Visiting Research Fellow at the Swiss Federal Institute of Technology, Lausanne, Switzerland. From 1995 to 1997, he was Alexander von Humboldt Fellow at Ruhr University, Bochum, Germany. From 1997 to 1999, he was a Research Associate at the Department of Electrical Engineering, Ruhr University. In 1999, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he became a Professor in 2002. From 2002 to 1005, he held a visiting professorship at the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. Since April 2005, he has been a professor and Head of the Communications Systems Group, Darmstadt University of Technology, Darmstadt, Germany. His research interests are in the area of signal processing and communications, and include statistical and array signal processing, adaptive beamforming, spatial diversity in wireless communications, multiuser and MIMO communications, parameter estimation and detection, and spectral analysis. He has published over 240 technical papers in these areas.

Dr. Gershman was a recipient of the 1993 International Union of Radio Science (URSI) Young Scientist Award, the 1994 Outstanding Young Scientist Presidential Fellowship (Russia), the 1994 Swiss Academy of Engineering Science and Branco Weiss Fellowships (Switzerland), and the 1995–1996 Alexander von Humboldt Fellowship (Germany). He received the 2000 Premier's Research Excellence Award of Ontario and the 2001 Wolfgang Paul Award from the Alexander von Humboldt Foundation, Germany. He is also a recipient of the 2002 Young Explorers Prize from the Canadian Institute for Advanced Research (CIAR), which has honored Canada's top 20 researchers 40 years of age or under. He received the 2004 IEEE Signal Processing Society Best Paper Award. He is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal on Wireless Communications and Networking*, and the *EURASIP Signal Processing Jopurnal* and is a Member of both the Sensor Array and Multichannel Signal Processing (SAM) and Signal Processing Theory and Methods (SPTM) Technical Committees of the IEEE Signal Processing Society. He was Technical Co-Chair of the Third IEEE International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, in December 2003. He is Technical Co-Chair of the Fourth IEEE Workshop on Sensor Array and Multichannel Signal Processing, to be held in Waltham, MA, in July 2006.