Cramér–Rao Lower Bounds for Low-Rank Decomposition of Multidimensional Arrays

Xiangqian Liu and Nicholas D. Sidiropoulos, *Senior Member, IEEE*

***Abstract*—Unlike low-rank matrix decomposition, which is generically nonunique for rank greater than one, low-rank three- and higher dimensional array decomposition is unique, provided that the array rank is lower than a certain bound, and the correct number of components (equal to array rank) is sought in the decomposition. Parallel factor (PARAFAC) analysis is a common name for low-rank decomposition of higher dimensional arrays. This paper develops Cramér–Rao Bound (CRB) results for low-rank decomposition of three- and four-dimensional (3-D and 4-D) arrays, illustrates the behavior of the resulting bounds, and compares alternating least squares algorithms that are commonly used to compute such decompositions with the respective CRBs. Simple-to-check necessary conditions for a unique low-rank decomposition are also provided.**

***Index Terms*—Cramér–Rao bound, least squares method, matrix decomposition, multidimensional signal processing.**

I. Introduction

FOR matrices [two-dimensional (2-D) or *two-way* arrays], the low-rank property in itself is not enough to guarantee a unique data model, and one has to resort to additional problem-specific structural properties to obtain a unique parameterization. Examples include orthogonality (as in singular value decomposition), Vandermonde, Toeplitz, or finite-alphabet constraints. Notwithstanding the lack of inherent uniqueness, low-rank matrix decomposition plays a key role in modern signal processing.

The concept of rank can be extended to arrays in three or higher dimensions in a natural way [14], [15]. Low-rank decomposition of three-dimensional (3-D) arrays was developed by Harshman [8]–[10] under the name parallel factor (PARAFAC) analysis and independently by Carroll and Chang [6] under the name canonical decomposition (CANDECOMP), building on a inconspicuous principle proposed in the factor analysis literature by Cattell [7]. Interestingly, these developments actually preceded the fundamental work on higher dimensional array rank [14], [15]. Quite unlike low-rank matrix decomposition, which is generically nonunique for any rank greater than one, low-rank, 3-D array decomposition is essentially unique for a meaningful range of low-enough ranks [14], and the situation actually improves in higher dimensions [21]. This remarkable fact was illustrated by Cattell and rigorously established under increasingly milder conditions over a span of some 30 years by various authors [8]–[10], [14], [15], [20], [21], [25].

Low-rank decomposition of multidimensional arrays has been used as as a data analysis tool in an unusual variety of disciplines, e.g., [2], [5], [6], [10], [14], and [16], to mention a few. More recently, it was shown to have extensive application in signal processing and communications, including blind multiuser detection in direct-sequence code-division multiple-access (DS-CDMA) communications [25], multiple-invariance sensor array processing [24], blind beamforming in specular multipath [17], [26], and, in more general terms, blind diversity-combining in communications [22]. Despite these wide-ranging applications, pertinent Cramér–Rao bounds (CRBs) for low-rank decomposition of multidimensional arrays have been missing from the literature. The first contribution of this paper is the development of CRBs for low-rank decomposition of 3-D and four-dimensional (4-D) arrays. Complex model parameters and a complex circularly symmetric white Gaussian noise model are assumed, but some special cases are also considered due to their importance in applications: a 3-D array with Vandermonde structure in one dimension, and the real-parameter 3-D case.

A wide variety of algorithms have been developed for computing low-rank decomposition in three and higher dimensions. These range from eigenvalue-type algebraic techniques [19], [20] to alternating least squares (ALS) algorithms [8]–[10]; see [25] for a readily available reference and [5] and [23] for a tutorial overview. Among these possibilities, ALS is the workhorse technique that is mostly preferred in practice [5], [23]. Even though ALS has been used extensively in this context and it has many desirable properties, its performance has not previously been measured relative to the CRB. This is the second contribution of this paper: putting ALS to the test versus the CRB and verifying that, indeed, the performance of ALS comes close to the CRB. This means that the CRB can be used to predict the average behavior of ALS, which in turn makes the CRB a valuable design tool. For example, in the context of DS-CDMA communications [25] and diversity-combining [22] in more general terms, the CRB allows quantitative evaluation of the diversity tradeoff for a specified performance. The real-parameter Gaussian CRB for low-rank decomposition of 3-D data can be used to improve experimental design in PARAFAC applications in chemistry; see [5] and references therein.

The paper is organized as follows. The model is introduced in Section II, including sufficient uniqueness conditions. Section III provides a starting point and roadmap for the

Manuscript received July 17, 2000; revised May 31, 2001. This work was supported by the National Science Foundation under CAREER grant 0096165 and Wireless grant 0096164. The associate editor coordinating the review of this paper and approving it for publication was Dr. Brian Sadler.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: xiang@ece.umn.edu; nikos@ece.umn.edu).

Publisher Item Identifier S 1053-587X(01)07057-X.

pertinent CRBs (associated derivations are deferred to the Appendices). These bounds are discussed and illustrated in Section IV. Simple-to-check necessary uniqueness conditions are developed in Section V. Multilinear ALS is compared with the CRB in Section VI. Conclusions are drawn in Section VII.

Some notation conventions that will be used in this paper follow.

$\mathbf{A}^{\mathrm{T}}$	Transpose of $\mathbf{A}$.
$\mathbf{A}^{*}$	Complex conjugate of $\mathbf{A}$.
$\mathbf{A}^{\mathrm{H}}$	Conjugate transpose of $\mathbf{A}$.
$\mathbf{A}(i)$	ith column of $\mathbf{A}$.
$\mathbf{a}_i$	ith column of $\mathbf{A}^{\mathrm{T}}$ ($\mathbf{a}_i^{\mathrm{T}}$ is the ith row of $\mathbf{A}$).
$\mathbf{A}\otimes\mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.
$\mathbf{A}\odot\mathbf{B}$	Khatri–Rao (column-wise Kronecker) product of $\mathbf{A}$ and $\mathbf{B}$.
$\mathbf{A}\diamond\mathbf{B}$	the Hadamard (element-wise) product of $\mathbf{A}$ and $\mathbf{B}$: $(\mathbf{A}\diamond\mathbf{B})_{i,j}=\mathbf{A}_{i,j}\mathbf{B}_{i,j}$;
$\mathbf{D}_i(\mathbf{A})$	Diagonal matrix constructed out of the ith row of $\mathbf{A}$.
$\delta_{i,j}$	Kronecker delta: $\delta_{i,j}=1$ when $i=j$, and $\delta_{i,j}=0$ when $i\neq j$, i and j are integers.
$\|\cdot\|_F$	Frobenius norm.
$(\cdot)^{\dagger}$	Matrix pseudo-inverse.

II. Multidimensional Low-Rank Modeling

Consider an $I\times J$ matrix $\mathbf{X}$. $\mathrm{rank}(\mathbf{X})=F$ if and only if $\mathbf{X}$ can be written as a sum of F but no less than F rank-one matrices (vector outer products)

$$\mathbf{X}=\mathbf{A}\mathbf{B}^{\mathrm{T}}=\mathbf{A}(1)\mathbf{B}^{\mathrm{T}}(1)+\cdots+\mathbf{A}(F)\mathbf{B}^{\mathrm{T}}(F) \tag{1}$$

where $\mathbf{A}$ is $I\times F$, and $\mathbf{B}$ is $J\times F$. Note that we may assume without loss of generality that the columns of $\mathbf{A}$ and $\mathbf{B}$ are linearly independent; otherwise, at least one of the F rank-one components can be absorbed in the others. In general, $F\leq\min(I,J)$, but if $F<\min(I,J)$, then $\mathbf{X}=\mathbf{A}\mathbf{B}^{\mathrm{T}}$ constitutes a low-rank decomposition of $\mathbf{X}$. Let $x_{i,j}$ denote the (i,j)th entry of $\mathbf{X}$. A scalar view of the relationship $\mathbf{X}=\mathbf{A}\mathbf{B}^{\mathrm{T}}$ is

$$x_{i,j}=\sum_{f=1}^{F}a_{i,f}b_{j,f} \tag{2}$$

for all $i=1,\ldots,I$, and $j=1,\ldots,J$, with obvious notation. Observe that $x_{i,j}$ is written as a sum of double products, i.e., it admits an F*-component bilinear decomposition*.

Next, consider an $I\times J\times K$ 3-D (also known as *three-way*) array $\underline{\mathbf{X}}$ with typical element $x_{i,j,k}$, and the F *component trilinear decomposition*:

$$x_{i,j,k}=\sum_{f=1}^{F}a_{i,f}b_{j,f}c_{k,f} \tag{3}$$

for all $i=1,\ldots,I$, $j=1,\ldots,J$, and $k=1,\ldots,K$. Equation (3) expresses the three-way array $\underline{\mathbf{X}}$ as a sum of F rank-one three-way factors, each one of which is the "outer product" of three vectors. Analogous to the definition of matrix (two-way array) rank, the rank of a three-way array $\underline{\mathbf{X}}$ is defined as the minimum number of rank-one (three-way) components needed to decompose $\underline{\mathbf{X}}$.

Define an $I\times F$ matrix $\mathbf{A}$ with typical element $\mathbf{A}(i,f):=a_{i,f}$, $J\times F$ matrix $\mathbf{B}$ with $\mathbf{B}(j,f):=b_{j,f}$, $K\times F$ matrix $\mathbf{C}$ with $\mathbf{C}(k,f):=c_{k,f}$, and $J\times K$ matrices $\mathbf{X}_i$ with $\mathbf{X}_i(j,k):=x_{i,j,k}$. Compact matrix representations of the model in (3) are made possible by employing the Khatri–Rao matrix product. For example

$$\mathbf{X}^{(JI\times K)}:=\begin{bmatrix}\mathbf{X}_{i=1}\\ \mathbf{X}_{i=2}\\ \vdots\\ \mathbf{X}_{i=I}\end{bmatrix}=\begin{bmatrix}\mathbf{B}\mathbf{D}_1(\mathbf{A})\\ \mathbf{B}\mathbf{D}_2(\mathbf{A})\\ \vdots\\ \mathbf{B}\mathbf{D}_I(\mathbf{A})\end{bmatrix}\mathbf{C}^{\mathrm{T}}=(\mathbf{A}\odot\mathbf{B})\,\mathbf{C}^{\mathrm{T}}. \tag{4}$$

The superscript $^{(JI\times K)}$ means that the matrix is of size $JI\times K$ and that the j-index (J goes first in the product JI) runs faster than the i-index along its rows. By symmetry

$$\mathbf{X}^{(IK\times J)}:=(\mathbf{C}\odot\mathbf{A})\mathbf{B}^{\mathrm{T}} \tag{5}$$

and

$$\mathbf{X}^{(KJ\times I)}:=(\mathbf{B}\odot\mathbf{C})\mathbf{A}^{\mathrm{T}}. \tag{6}$$

Next, consider an $I\times J\times K\times L$ 4-D array $\underline{\mathbf{X}}$ with typical element

$$x_{i,j,k,l}=\sum_{f=1}^{F}a_{i,f}b_{j,f}g_{k,f}h_{l,f} \tag{7}$$

for $i=1,\ldots,I$, $j=1,\ldots,J$, $k=1,\ldots,K$, and $l=1,\ldots,L$. Define a $K\times F$ matrix $\mathbf{G}$ with typical element $\mathbf{G}(k,f):=g_{k,f}$ and $L\times F$ matrix $\mathbf{H}$ with $\mathbf{H}(l,f):=h_{l,f}$. Similar to the three-way case, the Khatri–Rao product can be used to cast the model in (7) in matrix form. For example

$$\mathbf{X}^{(IKL\times J)}:=(\mathbf{H}\odot\mathbf{G}\odot\mathbf{A})\,\mathbf{B}^{\mathrm{T}}. \tag{8}$$

The Khatri–Rao product has the property $(\mathbf{H}\odot\mathbf{G})\odot\mathbf{A}=\mathbf{H}\odot(\mathbf{G}\odot\mathbf{A})$ [13]; in addition, the order of Khatri–Rao multiplications only affects the order of rows in the final result [13]. In particular, rank is not affected by the order in which the multiplications are carried out.

Further generalizing to N dimensions

$$x_{i_1,\ldots,i_N}=\sum_{f=1}^{F}\prod_{n=1}^{N}a_{i_n,f}^{(n)} \tag{9}$$

for $i_n=1,\ldots,I_n$, $n=1,\ldots,N$, where $a_{i_n,f}^{(n)}\in\mathbb{C}$. Upon defining $I_n\times F$ matrices $\mathbf{A}^{(n)}$ with $\mathbf{A}^{(n)}(i_n,f):=a_{i_n,f}^{(n)}$, many matrix representations are possible, e.g.,

$$\mathbf{X}^{(I_2\ldots I_N\times I_1)}=\left(\mathbf{A}^{(N)}\odot\cdots\odot\mathbf{A}^{(2)}\right)\left(\mathbf{A}^{(1)}\right)^{\mathrm{T}}.$$

A. Uniqueness

Definition 1: The k-rank of a matrix $\mathbf{A}\in\mathbb{C}^{I\times F}$ (which is denoted by $k_{\mathbf{A}}$) is r if and only if *every* r columns of $\mathbf{A}$ are linearly independent and either $\mathbf{A}$ has r columns or $\mathbf{A}$ contains

a set of $r+1$ linearly dependent columns. Note that k-rank is always less than or equal to rank $k_{\mathbf{A}} \leq r_{\mathbf{A}} := \text{rank}(\mathbf{A}) \leq \min(I, F)$, $\forall\ \mathbf{A}$. k-rank stands for *Kruskal*-rank [14]; the term was coined by Harshman and Lundy [11].

Theorem 1: (*Sufficient condition for uniqueness of low-rank decomposition of N-way arrays* [21], [22]; cf. [14] for a basic precursor result, and [25] for its complex counterpart). Consider the F-component N-way array given in (9), and suppose that it is irreducible (i.e., the rank of the N-way array is F). If

$$\sum_{n=1}^{N} k_{\mathbf{A}^{(n)}} \geq 2F + (N-1) \tag{10}$$

then the N-way array $x_{i_1,\ldots,i_N}$, $i_n = 1, \ldots, I_n$, $n = 1, \ldots, N$ has unique rank-one N-way factors

$$\prod_{n=1}^{N} a_{i_n,f}^{(n)}, \quad f = 1, \ldots, F.$$

Note that the three-way array rank F must be low enough relative to the maximum possible sum of k-ranks, e.g., $\min(I,F) + \min(J,F) + \min(K,F) \geq 2F+2$ in the three-way case,[1] and the rank of the decomposition (number of columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$) must be equal to the correct rank F. If the latter is less than F, then exact decomposition is impossible, by definition of array rank. If the rank of the decomposition is greater than F, then the decomposition is not unique. In addition, note that increasing the number of dimensions decreases the k-rank requirement per dimension. In three dimensions, it is necessary for uniqueness that the k-rank of any matrix is at least two. In four dimensions and beyond, even this is not necessary: It is possible to have one matrix of k-rank equal to one and still have uniqueness. Regarding conditions that are necessary for uniqueness, see also Theorem 2 in Section V.

III. Roadmap of CRB Results

As mentioned before, although low-rank decomposition of multidimensional arrays has been used as a data analysis tool in a variety of disciplines, pertinent CRB results have been missing from the literature. The parameters of interest for which the CRB will be established are the elements of the unknown matrices that are involved in the decomposition ($\mathbf{A}, \mathbf{B}, \mathbf{C}$ in the three-way case).

In the Appendixes, we derive CRBs for low-rank decomposition of complex-parameter three-way arrays (Appendix A) and complex-parameter four-way arrays (Appendix B), both in i.i.d. complex circularly symmetric Gaussian noise. These are applicable, e.g., in diversity-combining applications in communications [22], [25]. In Appendix C, we consider the case of a complex-parameter three-way array that exhibits Vandermonde structure in one dimension. This is of interest in sensor array processing applications [24], and it can be viewed as generalizing corresponding results for the two-way case with Vandermonde structure in one dimension [27]. In Appendix D, we consider the special case of real-parameter three-way arrays for which certain simplifications are possible due to the fact that the noise covariance matrix can be written in convenient closed form and an alternative scalar computation approach can be adopted. The real-parameter three-way case is the one that is mostly encountered in applications of PARAFAC in other disciplines. Under the Gaussian assumption, the CRB will be block diagonal in the noise and signal parameters. In the interest of brevity, we therefore assume that the noise variance is known throughout the derivations.

A delicate point regarding these CRBs is the inherent permutation and scale ambiguity. For example, in the three-way case (3), one can clearly reshuffle the F components and/or set $\bar{a}_{i,f} = \lambda_f^{(a)} a_{i,f}$, $\bar{b}_{j,f} = \lambda_f^{(b)} b_{j,f}$, and $\bar{c}_{k,f} = \lambda_f^{(c)} c_{k,f}$ such that $\lambda_f^{(a)} \lambda_f^{(b)} \lambda_f^{(c)} = 1$, $\forall\ f$, and $x_{i,j,k}$ remains unaffected. In other words

$$(\mathbf{A}, \mathbf{B}, \mathbf{C}) \Longleftrightarrow \left(\mathbf{A}\mathbf{\Pi}\mathbf{\Lambda}^{(a)}, \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda}^{(b)}, \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}^{(c)}\right)$$

for any permutation matrix $\mathbf{\Pi}$ and diagonal scaling matrices $\mathbf{\Lambda}^{(a)}$, $\mathbf{\Lambda}^{(b)}$, $\mathbf{\Lambda}^{(c)}$ such that $\mathbf{\Lambda}^{(a)}\mathbf{\Lambda}^{(b)}\mathbf{\Lambda}^{(c)} = \mathbf{I}_{F\times F}$. In fact, low-rank decomposition of multidimensional arrays is unique (under the sum of k-ranks condition) *up to* the above indeterminacy, which is unresolvable but also mostly insignificant. Some means of fixing permutation and scale should be agreed upon in order to obtain a meaningful bound. A simple (but not the only) way of doing this in the three-way case is to fix the first row of $\mathbf{A}$ and $\mathbf{B}$ to $[1 \ldots 1]_{1\times F}$ (this takes care of scale ambiguity) and further assume that the first row of $\mathbf{C}$ is known and consists of distinct elements (which subsequently resolves the permutation ambiguity).[2]

Our convention is plausible in the context of antenna array reception of DS-CDMA. Let

- $\mathbf{A}$ antenna array response (steering) matrix;
- $\mathbf{B}$ code matrix;
- $\mathbf{C}$ symbol matrix.

Since the first element of the antenna array simply serves as an arbitrary frame of reference, the first row of $\mathbf{A}$ can be assumed to be a row vector of all ones, without loss of generality. Similarly, the first row of $\mathbf{B}$ can be assumed to be a row vector of all ones (first chip of all users equal to one) without loss of generality. Then, the scale factor (propagation loss and phase shift times reference antenna gain times first chip) for each user can be absorbed in the corresponding column of the symbol matrix $\mathbf{C}$. It then seems reasonable to assume that the first row of $\mathbf{C}$ consists of distinct elements, although considering it to be known is less appealing in a blind setting. Note, however, that this is just a way of technically fixing an inherently unresolvable ambiguity for the purpose of performance evaluation. Permutation and scale-fixing conventions could be application dependent and, indeed, could influence the final bound. What is less restrictive in one application could be more restrictive in another—we do not have a universal solution, but one could still use the results in the Appendixes to get a head start in computing the pertinent bounds under alternative conventions.

[1] Three-way array rank could be up to $\min(IJ, JK, IK)$; see [15].

[2] It is not enough to assume that the first row of $\mathbf{A}$ is known. The reason is that scale ambiguity has to be fixed prior to resolving the permutation ambiguity; without knowing which column is which, we cannot divide by the first element because we do not know what this element is.

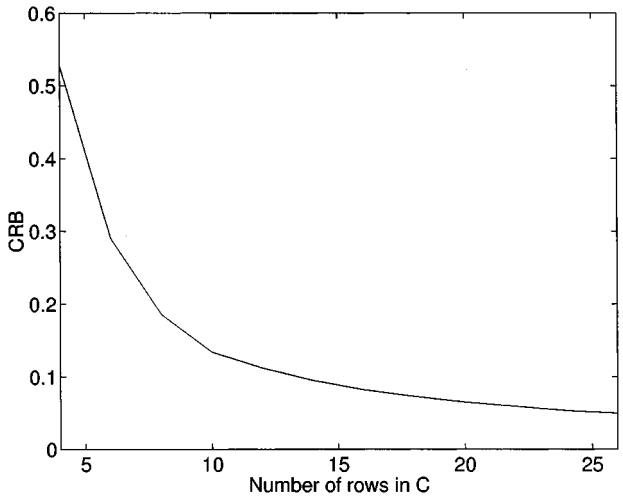

Fig. 1. PARAFAC CRB: Varying sample size along one dimension (number of rows of $\mathbf{C}$).

IV. Discussion and Illustration of CRB Results

In this section, we illustrate the behavior of the CRBs derived in the Appendixes. Throughout, the CRB is first normalized in an element-wise fashion, i.e., each unknown parameter's CRB is divided by the corresponding parameter's modulus square. The average CRB of all the unknown parameters is then used as a single performance measure. Therefore, instead of plotting the trace of the CRB, we plot a weighted trace, with weights proportional to the inverse modulus square of the respective parameters. This measures average variance relative to modulus square across all parameters of interest. Signal-to-noise ratio (SNR) is defined as [cf. (4) and (8)]

$$\text{SNR} = 10\log_{10}\frac{\|\mathbf{X}^{(JI\times K)}\|_F^2}{IJK\sigma^2} \quad \text{(three-way array)}$$

$$\text{SNR} = 10\log_{10}\frac{\|\mathbf{X}^{(IKL\times J)}\|_F^2}{IJKL\sigma^2} \quad \text{(four-way array)}.$$

Unless otherwise specified, the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ were randomly drawn from an i.i.d. standard Gaussian distribution, and the CRB is averaged over all the unknown parameters of all matrices.

Fig. 1 is an illustration of the behavior of the CRB for low-rank decomposition of a complex-parameter three-way array (Appendix A) as the size of one dimension increases, with SNR fixed at 10 dB. In this experiment, $\mathbf{A}$ and $\mathbf{B}$ are 4×3, whereas $\mathbf{C}$ is augmented from 4×3 to 26×3. Notice that the number of parameters also increases with the sample size. However, the number of equations (data points) increases faster than the number of unknown parameters (IJK versus order of $(I+J+K)F$, respectively). One would then intuitively expect that the CRB should decrease with increasing sample size along any dimension. Although this seems to be the trend, it is not true in general. We have seen examples of datasets where the average CRB may actually increase before decreasing again as one adds rows to $\mathbf{C}$.

Fig. 2 illustrates the effect of adding an extra diversity dimension. The CRB for a three-way array (Appendix A) with 4×3 component matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ is compared with the corresponding bound for a four-way array (Appendix B) obtained by augmenting the three-way model by adding another dimension of size 2×3, 4×3, or 8×3. We have seen that higher dimensionality leads to a relaxed uniqueness condition (Theorem 1). Fig. 2 demonstrates that higher dimensionality also benefits in terms of CRB.

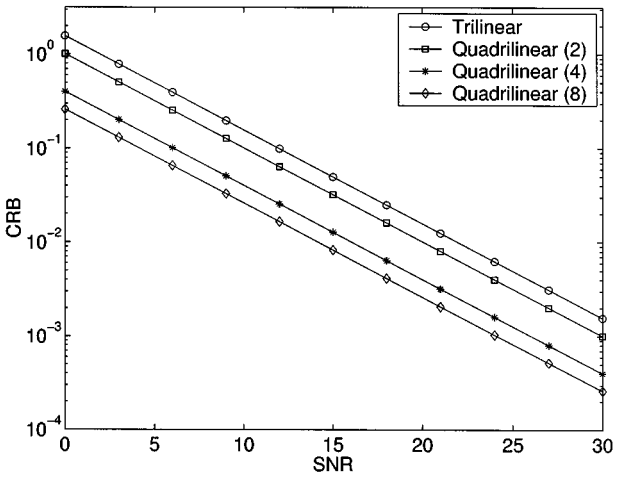

Fig. 2. CRB: Effect of extra diversity dimension.

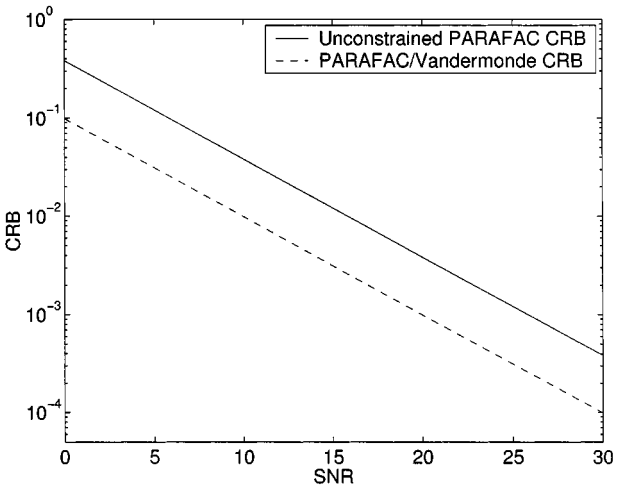

Fig. 3. PARAFAC CRB: With versus without assuming Vandermonde structure.

In Fig. 3, we compare the CRB of Appendix A with that of Appendix C for a given three-way array with Vandermonde structure in one dimension. The plot shows average CRB of the generators of the Vandermonde matrix. The bound in Appendix A ignores the fact that Vandermonde structure is present and is therefore above the bound in Appendix C. Comparisons of this kind can help gauge the performance margin that can be gained by employing and exploiting additional model structure—in this case, Vandermonde. An example could be the choice of complex exponential (OFDMA-like) spreading codes in the context of [25].

A quadrilinear model can always be viewed as a trilinear model, i.e., in (8), let $\mathbf{C} = \mathbf{H} \odot \mathbf{G}$ to obtain $\mathbf{X}^{(IKL\times J)} = (\mathbf{C} \odot \mathbf{A})\mathbf{B}^{\mathrm{T}}$. This "unfolding" into a lower dimensional model

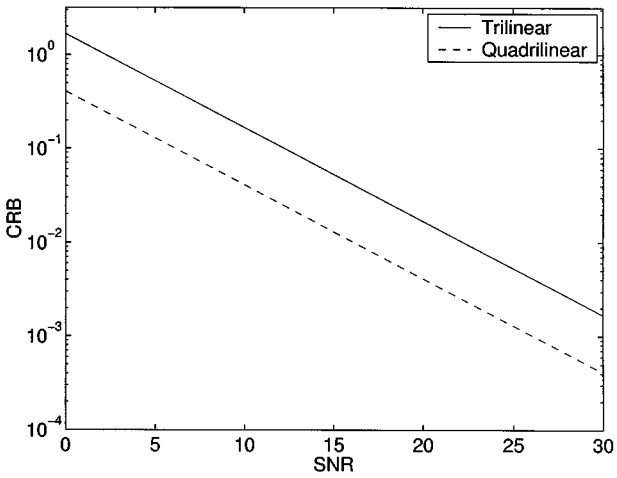

Fig. 4. PARAFAC CRB: Model viewed as trilinear versus quadrilinear.

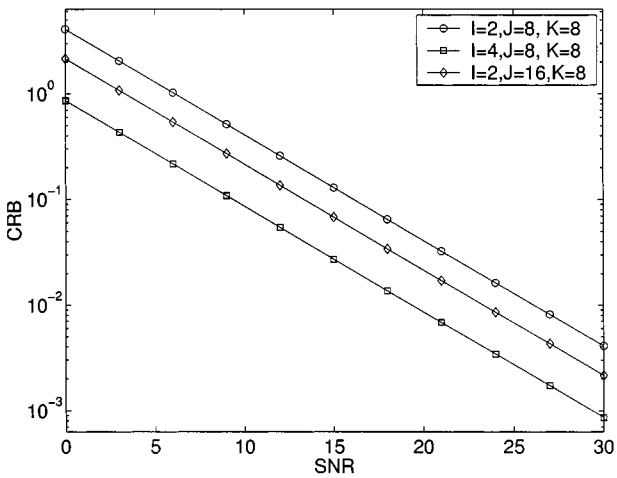

Fig. 5. PARAFAC CRB: CDMA diversity-combining tradeoff.

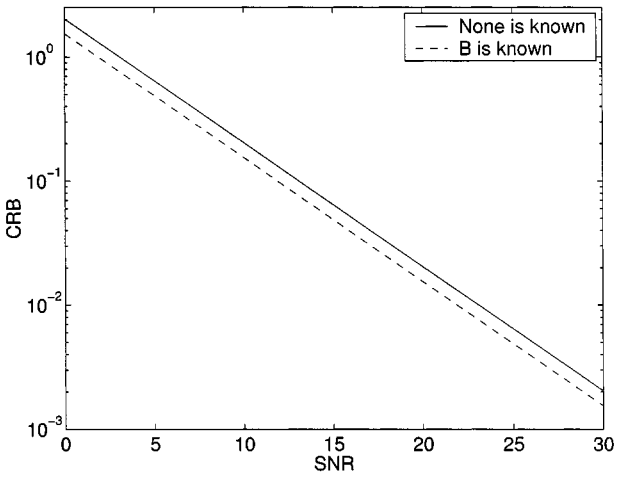

Fig. 6. PARAFAC CRB: performance differential when one matrix is known.

ignores part of the model structure. Even if the lower dimensional model is uniquely parameterized, the loss of structure will result in a loss of statistical efficiency. This is illustrated in Fig. 4, which plots the CRB in Appendix B for a quadrilinear data model with $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{H}$ all 4×3 and the CRB in Appendix A corresponding to unfolding it into a trilinear model. Notice that the difference is significant.

The bounds in Appendices A–D can be used to gauge at the diversity-combining tradeoff for system design purposes. As an example, consider a DS-CDMA system with $F = 4$ users, spread using Walsh–Hadamard codes of length $J = 8$ chips/symbol, and received at the base station using a ULA consisting of $I = 2$ antennas spaced half a wavelength apart. Collect $K = 8$ symbols worth of samples from the baseband outputs of the antenna elements sampled at the chip rate. As shown in [25], the resulting data can be arranged into an $I \times J \times K$ three-way array of rank F, and hence, the steering vectors in $\mathbf{A}$, spreading codes in $\mathbf{B}$, and transmitted symbols in $\mathbf{C}$ can all be blindly recovered. Now, suppose that one wishes to improve blind estimation performance and that two options are available: increase spreading, [going from $J = 8$ to $J = 16$ chips/symbol (at the expense of doubling the bandwidth)] or double the number of receive antennas in the ULA from $I = 2$ to $I = 4$ (paying the cost of extra hardware). Aside from other considerations, which of the two options is best performance-wise? Using the CRB in Appendix C (notice that the steering vectors in $\mathbf{A}$ are Vandermonde, due to the ULA assumption), Fig. 5 shows that for directions 20°, 40°, 60°, and 80° relative to the array broadside, it is best to double the number of antennas, rather than the spreading gain. Note that the CRB in Fig. 5 is averaged over 100 random realizations of the symbol matrix $\mathbf{C}$, and the situation reverses for closely spaced directions, as expected. The reason we average the CRB over random realizations of $\mathbf{C}$ is as follows. Let θ denote the parameter vector to be estimated. Any estimation algorithm that treats θ as a deterministic unknown is bounded by CRB(θ) pointwise in θ. Hence, the average performance of any such algorithm over a collection of $\theta \in \Theta$ is bounded by the average of the respective CRBs over Θ. For a large number of representative θ drawn via Monte Carlo simulation, such averages can be used to gauge the achievable performance of estimation algorithms that treat θ as a deterministic unknown.

In some instances, the signals along one of the diversity dimensions (one of the three matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$) may be known. Examples include known spreading codes or estimated signatures or training signals. One may then wonder how this knowledge affects the estimation of the remaining unknowns. Interestingly, the answer is often *not much*. Fig. 6 depicts the average CRB for the free elements of $\mathbf{A}$ and $\mathbf{C}$ when $\mathbf{B}$ is assumed known versus unknown. All three matrices are 6×4, and the CRB curves are averaged over 1000 realizations of randomly drawn $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. The curves come closer for higher I, J, and K relative to F. We conclude that there is often enough structure in the trilinear model itself so that knowledge of the signals in one dimension does not significantly improve the estimation of signals in the other dimensions.

In general, uniqueness neither implies nor is implied by a finite CRB. A simple example in [1] can be used to clarify this. Consider the estimation of parameter x from the measurement $y = x^2 + n$, where n is a zero-mean, Gaussian random variable of variance σ^2. The Fisher information is proportional to x^2/σ^2, which is 0 at $x = 0$, although $x = 0$ is the only identifiable value of x. Hence, uniqueness does not imply a finite CRB. For

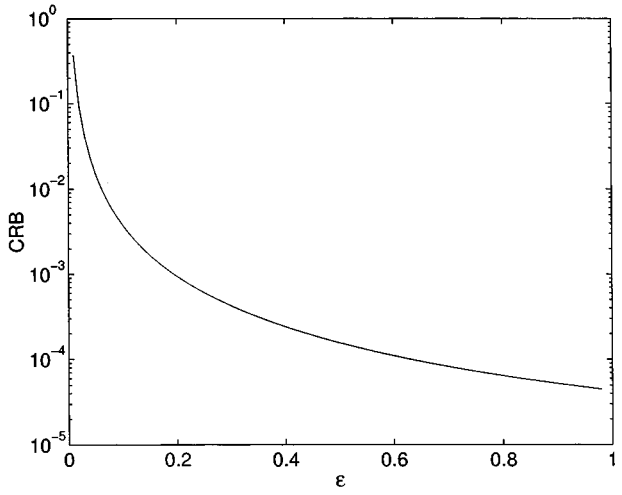

Fig. 7. CRB behavior near nonidentifiability.

any $x \neq 0$, the Fisher information is nonzero, and hence, the CRB is finite, whereas these x are not identifiable. Therefore, uniqueness is not implied by a finite CRB. However, it is natural to expect that as one moves closer to nonuniqueness, the CRB should increase. This is the subject of Fig. 7. Let $\mathbf{A}$ and $\mathbf{B}$ be 3×2 matrices with full k rank, and let

$$\mathbf{C} = \begin{bmatrix} 1+\sqrt{-1} & 0 \\ 1+\sqrt{-1} & \epsilon \end{bmatrix}$$

where $\epsilon \in [0, \ 1]$. When $\epsilon = 1$, $k_{\mathbf{C}} = 2$, and $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} = 2F + 2$, the model is unique according to (10). $\epsilon = 0$ implies $k_{\mathbf{C}} = 0$, and hence, the sufficient condition (10) is violated: $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} < 2F + 2$. Fig. 7 depicts the behavior of the CRB as ϵ approaches zero. Observe that the CRB remains stable for a wide range of values and then increases sharply in the vicinity of $\epsilon = 0$. This is consistent with practical experience involving ALS applied to real data sets exhibiting near-collinearity [5].

V. Necessary Uniqueness Conditions

Condition (10) is sufficient but not necessary for uniqueness. When low-rank modeling is utilized as an exploratory design tool, it is useful to have *simple* tests that quickly rule out nonidentifiable models. This is the subject of this section.

Theorem 2: For $N = 3$, a necessary condition for uniqueness of the model in (9) is

$$\min(r_{\mathbf{A}\odot\mathbf{B}}, r_{\mathbf{C}\odot\mathbf{A}}, r_{\mathbf{B}\odot\mathbf{C}}) = F. \tag{11}$$

For $N = 4$, a necessary condition is

$$\min(r_{\mathbf{B}\odot\mathbf{G}\odot\mathbf{A}}, r_{\mathbf{G}\odot\mathbf{H}\odot\mathbf{B}}, r_{\mathbf{H}\odot\mathbf{A}\odot\mathbf{G}}, r_{\mathbf{A}\odot\mathbf{B}\odot\mathbf{H}}) = F.$$

The result is further generalized for any N: The Khatri–Rao product of any leave-one-out selection of matrices from $\{\mathbf{A}^{(n)}\}_{n=1}^{N}$ must be of rank F for the model in (9) to be unique.

Proof: From (4), if $r_{\mathbf{A}\odot\mathbf{B}} < F$, then

$$\mathbf{X}^{(JI\times K)} = (\mathbf{A}\odot\mathbf{B})\mathbf{C}^{\mathrm{T}} = (\mathbf{A}\odot\mathbf{B})\left(\mathbf{C}^{\mathrm{T}} + \mathbf{N}^{\mathrm{T}}\right)$$

for any $\mathbf{N} \in \mathbb{C}^{K\times F}$ whose rows are in the null space of $\mathbf{A}\odot\mathbf{B}$. It follows that $\mathbf{C}$ is not unique. Similarly, from (5) and (6), we conclude that if $r_{\mathbf{C}\odot\mathbf{A}} < F$ or $r_{\mathbf{B}\odot\mathbf{C}} < F$, then $\mathbf{B}$ or $\mathbf{A}$ cannot be unique. For uniqueness of the model in (9), all three conditions need to hold; hence, condition (11) follows. For the $N = 4$ case, consider (8). There are a total of four leave-one-out selections from $\{\mathbf{A}, \mathbf{B}, \mathbf{G}, \mathbf{H}\}$, and the Khatri–Rao product of the three selected matrices must be full rank for the left-out matrix to be unique (recall that Khatri–Rao product rank is not affected by the order in which the multiplications are carried out). ■

Remark 1: In order to see that the sufficient condition (10) in Theorem 1 implies the necessary condition (11) in Theorem 2, consider (10) with $N = 3$, and note that for any $\mathbf{C} \in \mathbb{C}^{K\times F}$, it holds that $k_{\mathbf{C}} \leq r_{\mathbf{C}} \leq \min(K, F) \leq F$; hence, (10) implies

$$k_{\mathbf{A}} + k_{\mathbf{B}} \geq F + 2 \tag{12}$$

which in turn implies

$$\min(k_{\mathbf{A}} + k_{\mathbf{B}} - 1, F) = F. \tag{13}$$

It has been shown in [26] that $k_{\mathbf{A}\odot\mathbf{B}} \geq \min(k_{\mathbf{A}} + k_{\mathbf{B}} - 1, F)$; thus, (13) yields $k_{\mathbf{A}\odot\mathbf{B}} = F$, which implies $r_{\mathbf{A}\odot\mathbf{B}} = F$. The remaining two implications follow by symmetry of (9) and (10).

Remark 2: Testing whether the necessary condition is valid is much simpler than checking the sufficient condition, which involves k-rank rather than rank. Checking the k-rank of a matrix involves sequentially checking all possible selections of n columns (n going from F to 2 or vice-versa) for linear independence. In the worst case, the calculation of $k_{\mathbf{A}}$, $k_{\mathbf{B}}$, and $k_{\mathbf{C}}$ by singular value decomposition requires up to $O\left(2^{F-2}F(F+1)(I+J+K)\right)$ floating-point operations (flops), whereas the calculation of $r_{\mathbf{A}\odot\mathbf{B}}$, $r_{\mathbf{C}\odot\mathbf{A}}$, and $r_{\mathbf{B}\odot\mathbf{C}}$ takes only $O\left((3F + IJ + JK + IK)F^2\right)$ flops.

The Khatri–Rao product can be viewed as a selection of columns from the Kronecker product, whose rank is the product of ranks of its constituent matrix factors. Hence

$$k_{\mathbf{A}\odot\mathbf{B}} \leq r_{\mathbf{A}\odot\mathbf{B}} \leq r_{\mathbf{A}\otimes\mathbf{B}} = r_{\mathbf{A}} r_{\mathbf{B}}.$$

Therefore, $k_{\mathbf{A}\odot\mathbf{B}} = F$ requires $r_{\mathbf{A}} r_{\mathbf{B}} \geq F$, and we obtain the following further simplified condition.

Corollary 1: For $N = 3$, a further simplified necessary condition for uniqueness of the model in (9) is

$$\min(r_{\mathbf{A}} r_{\mathbf{B}}, r_{\mathbf{C}} r_{\mathbf{A}}, r_{\mathbf{B}} r_{\mathbf{C}}) \geq F. \tag{14}$$

The simplified condition (14) can also be generalized to higher dimensions.

VI. Multilinear ALS Performance

The principle of ALS can be used to fit low-rank models in any dimension. A trilinear ALS (TALS) algorithm can be found in [25], but the idea of using ALS to fit low-rank three-way models goes back to Harshman [8]–[10]. The basic idea behind ALS is simple: each time update *one* matrix, using least

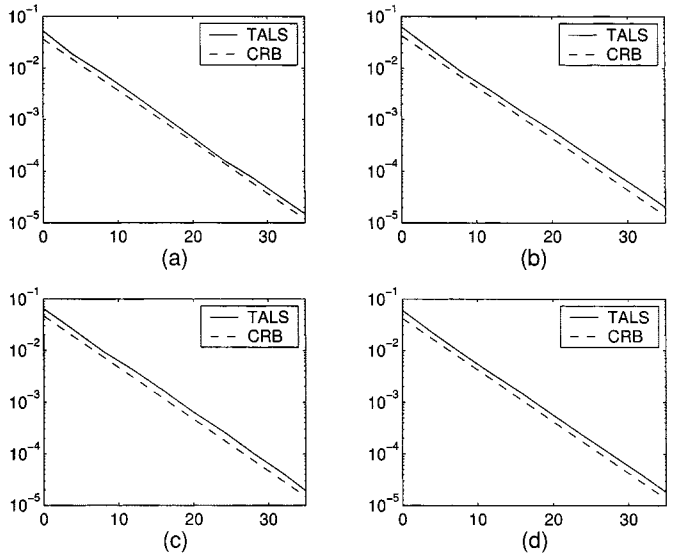

Fig. 8. TALS performance vesus CRB. (a) $\mathbf{A}$. (b) $\mathbf{B}$. (c) $\mathbf{C}$. (d) Average.

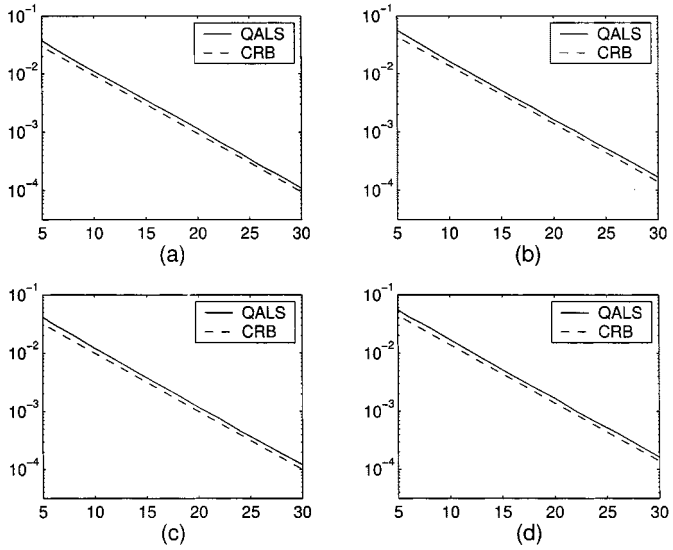

Fig. 9. QALS performance versus CRB. (a) $\mathbf{A}$. (b) $\mathbf{B}$. (c) $\mathbf{G}$. (d) $\mathbf{H}$.

squares (LS) *conditioned* on previously obtained estimates for the remaining matrices; proceed to update the other matrices; repeat until convergence of the LS cost function (this is guaranteed for ALS algorithms, e.g., [25]). For example, quadrilinear ALS (QALS) aims to

$$\min_{\mathbf{A},\mathbf{B},\mathbf{G},\mathbf{H}} \left\| \tilde{\mathbf{X}}^{(IKL\times J)} - (\mathbf{H}\odot\mathbf{G}\odot\mathbf{A})\,\mathbf{B}^{\mathrm{T}} \right\|_F^2$$

where $\tilde{\mathbf{X}}^{(IKL\times J)}$ is the noisy data matrix. Due to the complete symmetry of the quadrilinear model [cf. (7)], the conditional least square updates are

$$\hat{\mathbf{A}}^{\mathrm{T}} = \left[\hat{\mathbf{G}}\odot\hat{\mathbf{B}}\odot\hat{\mathbf{H}}\right]^{\dagger}\tilde{\mathbf{X}}^{(LJK\times I)}$$
$$\hat{\mathbf{B}}^{\mathrm{T}} = \left[\hat{\mathbf{H}}\odot\hat{\mathbf{G}}\odot\hat{\mathbf{A}}\right]^{\dagger}\tilde{\mathbf{X}}^{(IKL\times J)}$$
$$\hat{\mathbf{G}}^{\mathrm{T}} = \left[\hat{\mathbf{A}}\odot\hat{\mathbf{H}}\odot\hat{\mathbf{B}}\right]^{\dagger}\tilde{\mathbf{X}}^{(JLI\times K)}$$
$$\hat{\mathbf{H}}^{\mathrm{T}} = \left[\hat{\mathbf{B}}\odot\hat{\mathbf{A}}\odot\hat{\mathbf{G}}\right]^{\dagger}\tilde{\mathbf{X}}^{(KIJ\times L)}$$

where $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{G}}$, and $\hat{\mathbf{H}}$ denote running estimates of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{H}$.

For zero-mean white (in all dimensions) Gaussian noise, ALS yields maximum likelihood (ML) estimates, provided the global minimum is reached. Under mild regularity conditions, ML is *asymptotically* (in sample size) unbiased and asymptotically achieves the CRB [12, Ch. 7]. For signal-in-noise problems, ML also achieves the CRB for high-enough signal-to-noise ratios [12, Ch. 7]. It therefore makes sense to compare trilinear and quadrilinear ALS against the respective CRBs.

In all of our Monte Carlo results, ALS is randomly initialized once per trial and then iterated until convergence. No reinitializations are used.

Fig. 8 depicts simulation results comparing TALS performance to the CRB for a unconstrained trilinear model. In this simulation, $F = 3$, $I = J = K = 20$, and 200 Monte Carlo trials per datum have been conducted. The lower right plot presents average mean squared error (MSE) for all free

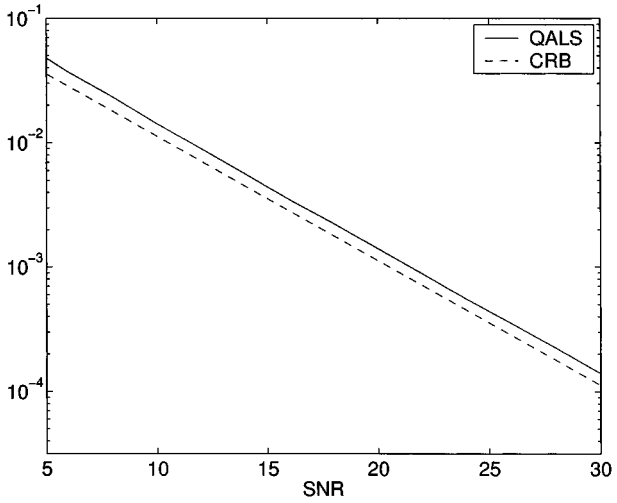

Fig. 10. QALS performance versus CRB. Average of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{H}$.

model parameters versus trace(CRB)$/(I+J+K-3)F$. The remaining three plots present average MSE for the free elements of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ respectively, versus the corresponding bound.

Figs. 9 and 10 depict Monte Carlo simulation results comparing QALS performance to the CRB for a typical scenario. For this simulation, $F = 5$, $I = J = K = L = 6$, and 800 Monte Carlo trials per datum have been conducted. Fig. 9 plots the estimation MSE for the free elements of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{H}$, respectively, versus the corresponding bound. Fig. 10 presents average MSE for all free model parameters versus the CRB. Observe that both TALS and QALS remain close to the CRB for a wide range of SNRs and reasonably small sample sizes in all dimensions—but note that total sample size is the product of sample sizes along the individual dimensions; hence, we have 8000 samples for the trilinear example in Fig. 8 and 1296 samples for the more structured quadrilinear model in Figs. 9 and 10. These are *typical* sample sizes in applications, wherein the number of samples in one dimension may be related to spreading gain, number of symbols collected, etc. [5], [25]. Total sample size is what matters for the good asymptotic properties of ML to come

into play, and this explains the good behavior of ALS, which aims for the ML solution. Observe that the gap between ALS and the CRB reduces quickly with increasing SNR (note that the plots are in log-log scale). The reasons for the nonzero gap at finite SNR are as follows.

- Practical experience [5] with trilinear ALS shows that the global minimum is usually reached as long as the constituent rank-one factors are not essentially collinear along any dimension. Nevertheless, for finite SNR, there exist a few inevitable bad runs, which hit a local minimum.
- Even if the right minimum is reached, scale and, more importantly, permutation matching may fail due to residual noise. This effect is particularly pronounced at low SNR.
- In practice, some prespecified tolerance threshold is used to terminate the ALS iteration. In certain situations, convergence can be relatively slow, which may lead to premature termination.

Even though an analytical verification of these findings would be highly desirable from a theoretical standpoint, it seems to be very difficult to carry through without making unrealistic assumptions, due to the iterative nature of ALS. The above Monte Carlo results are perhaps sufficient from an applications viewpoint.

VII. Conclusion

Low-rank decomposition of multidimensional arrays (in $N \geq 3$ dimensions) has the remarkable property of uniqueness on the basis of low-rank structure alone. It has found extensive applications in diverse areas, including signal processing and communications. This paper has contributed pertinent CRBs for low-rank decomposition of 3-D and 4-D arrays. Aside from performance evaluation, these can be used as design tools for diversity tradeoff studies. Easy-to-check necessary uniqueness conditions for low-rank decomposition of multidimensional arrays have also been developed. Finally, the performance of ALS algorithms that are commonly used to compute low-rank decomposition of multidimensional arrays has been assessed relative to the respective CRBs. It was demonstrated by means of Monte Carlo simulation that for typical problem sizes, ALS stays close to what is theoretically achievable performance-wise.

Appendix A
CRB for Low-Rank Decomposition of 3–D Arrays

Consider an $I \times J \times K$ three-way array $\underline{\mathbf{X}}$ with typical element

$$x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} + n_{i,j,k} \tag{15}$$

where $n_{i,j,k}$ is i.i.d. (in i, j, and k) zero mean, circularly symmetric complex Gaussian noise, of variance σ^2. Unfolding as in (4)–(6), define

$$\mathbf{X}_1 = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^{\mathrm{T}} + \mathbf{N}_1, \ : KJ \times I \tag{16}$$
$$\mathbf{X}_2 = (\mathbf{C} \odot \mathbf{A})\mathbf{B}^{\mathrm{T}} + \mathbf{N}_2, \ : IK \times J \tag{17}$$
$$\mathbf{X}_3 = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^{\mathrm{T}} + \mathbf{N}_3, \ : JI \times K \tag{18}$$

where $\mathbf{N}_1$, $\mathbf{N}_2$, and $\mathbf{N}_3$ are the corresponding noise matrices. Therefore, we can write the likelihood function of $\underline{\mathbf{X}}$ in three equivalent ways:

$$\begin{aligned} L(\underline{\mathbf{X}}) &= \frac{1}{(\pi\sigma^2)^{IJK}} \exp\left\{ -\frac{1}{\sigma^2} \sum_{i=1}^{I} \|\mathbf{X}_1(i) - (\mathbf{B} \odot \mathbf{C})\mathbf{a}_i\|^2 \right\} \\ &= \frac{1}{(\pi\sigma^2)^{IJK}} \exp\left\{ -\frac{1}{\sigma^2} \sum_{j=1}^{J} \|\mathbf{X}_2(j) - (\mathbf{C} \odot \mathbf{A})\mathbf{b}_j\|^2 \right\} \\ &= \frac{1}{(\pi\sigma^2)^{IJK}} \exp\left\{ -\frac{1}{\sigma^2} \sum_{k=1}^{K} \|\mathbf{X}_3(k) - (\mathbf{A} \odot \mathbf{B})\mathbf{c}_k\|^2 \right\} \end{aligned}$$

where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}^{\mathrm{T}}$ and similarly for $\mathbf{b}_j$ and $\mathbf{c}_k$.

As explained in Section III, a 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 row of $\mathbf{A}$ and $\mathbf{B}$ is fixed (or normalized) to $[1 \cdots 1]_{1\times F}$ (this takes care of scale ambiguity) and further assume that the first row of $\mathbf{C}$ is known and consists of distinct elements (which subsequently resolves the permutation ambiguity). This way, the number of unknown complex parameters is $(I+J+K-3)F$ instead of $(I+J+K)F$. Similar to [3] and [28], to simplify the CRB derivation, it is useful to introduce the equivalent $1 \times 2(I+J+K-3)F$ complex parameter vector

$$\boldsymbol{\theta} = \left[\mathbf{a}_2^{\mathrm{T}}, \ldots, \mathbf{a}_I^{\mathrm{T}}, \mathbf{b}_2^{\mathrm{T}}, \ldots, \mathbf{b}_J^{\mathrm{T}}, \mathbf{c}_2^{\mathrm{T}}, \ldots, \mathbf{c}_K^{\mathrm{T}}, \mathbf{a}_2^{\mathrm{H}}, \ldots, \mathbf{c}_K^{\mathrm{H}}\right].$$

Let us write the log-likelihood function as $f(\boldsymbol{\theta}) = \ln L(\underline{\mathbf{X}})$

$$\begin{aligned} f(\boldsymbol{\theta}) &= -IJK\ln\left(\pi\sigma^2\right) - \frac{1}{\sigma^2} \sum_{i=1}^{I} \|\mathbf{X}_1(i) - (\mathbf{B} \odot \mathbf{C})\mathbf{a}_i\|^2 \\ &= -IJK\ln\left(\pi\sigma^2\right) - \frac{1}{\sigma^2} \sum_{j=1}^{J} \|\mathbf{X}_2(j) - (\mathbf{C} \odot \mathbf{A})\mathbf{b}_j\|^2 \\ &= -IJK\ln\left(\pi\sigma^2\right) - \frac{1}{\sigma^2} \sum_{k=1}^{K} \|\mathbf{X}_3(k) - (\mathbf{A} \odot \mathbf{B})\mathbf{c}_k\|^2. \end{aligned} \tag{19}$$

Then, the complex Fisher information matrix (FIM) is given by (see, e.g., [18])

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathrm{E}\left\{ \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathrm{H}} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \right\}. \tag{20}$$

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

$$\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}} = \frac{1}{\sigma^2} \left(\mathbf{X}_1(i) - (\mathbf{B} \odot \mathbf{C})\mathbf{a}_i\right)^{\mathrm{H}} (\mathbf{B} \odot \mathbf{C})\mathbf{e}_f \tag{21a}$$
$$\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}} = \frac{1}{\sigma^2} \left(\mathbf{X}_2(j) - (\mathbf{C} \odot \mathbf{A})\mathbf{b}_j\right)^{\mathrm{H}} (\mathbf{C} \odot \mathbf{A})\mathbf{e}_f \tag{21b}$$
$$\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f}} = \frac{1}{\sigma^2} \left(\mathbf{X}_3(k) - (\mathbf{A} \odot \mathbf{B})\mathbf{c}_k\right)^{\mathrm{H}} (\mathbf{A} \odot \mathbf{B})\mathbf{e}_f \tag{21c}$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}^*} = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}}\right)^* \tag{21d}$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}^*} = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}}\right)^* \tag{21e}$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f}^*} = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f}}\right)^* \tag{21f}$$

where $\mathbf{e}_f$ is the fth unit coordinate vector.

Although the computation of the FIM may look cumbersome, some simplifications are possible. Note that although $n_{i,j,k}$ is independent in three dimensions, i.e.,

$$E\left\{n_{i_1,j_1,k_1} n_{i_2,j_2,k_2}^*\right\} = \delta_{i_1,i_2}\delta_{j_1,j_2}\delta_{k_1,k_2} \tag{22}$$

$\mathbf{N}_1(i)$, $\mathbf{N}_2(j)$, and $\mathbf{N}_3(k)$ are nevertheless correlated. Using the fact that (e.g., cf. [27])

$$E\left\{n_{i_1,j_1,k_1} n_{i_2,j_2,k_2}\right\} = E\left\{n_{i_1,j_1,k_1}^* n_{i_2,j_2,k_2}^*\right\} = 0 \tag{23}$$

it can be shown that the FIM can be written as

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^* \end{bmatrix} \tag{24}$$

where the matrix $\boldsymbol{\Psi}$ is of size $(I+J+K-3)F \times (I+J+K-3)F$. Given this block diagonal structure, we need to compute and invert the submatrix $\boldsymbol{\Psi}$. Based on (22) and (23), it can be shown that the elements of $\boldsymbol{\Psi}$ can be computed as

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i_1,f_1}^*}\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i_2,f_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{B}\odot\mathbf{C})^{\mathrm{H}}(\mathbf{B}\odot\mathbf{C})\mathbf{e}_{f_2}\delta_{i_1,i_2}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j_1,f_1}^*}\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j_2,f_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{C}\odot\mathbf{A})^{\mathrm{H}}(\mathbf{C}\odot\mathbf{A})\mathbf{e}_{f_2}\delta_{j_1,j_2}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k_1,f_1}^*}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k_2,f_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{A}\odot\mathbf{B})^{\mathrm{H}}(\mathbf{A}\odot\mathbf{B})\mathbf{e}_{f_2}\delta_{k_1,k_2}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f_1}^*}\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f_2}}\right\} = \frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{B}\odot\mathbf{C})^{\mathrm{H}}E\left\{\mathbf{N}_1(i)\mathbf{N}_2^{\mathrm{H}}(j)\right\} \cdot(\mathbf{C}\odot\mathbf{A})\mathbf{e}_{f_2}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f_1}^*}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f_2}}\right\} = \frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{B}\odot\mathbf{C})^{\mathrm{H}}E\left\{\mathbf{N}_1(i)\mathbf{N}_3^{\mathrm{H}}(k)\right\} \cdot(\mathbf{A}\odot\mathbf{B})\mathbf{e}_{f_2}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f_1}^*}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f_2}}\right\} = \frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{C}\odot\mathbf{A})^{\mathrm{H}}E\left\{\mathbf{N}_2(j)\mathbf{N}_3^{\mathrm{H}}(k)\right\} \cdot(\mathbf{A}\odot\mathbf{B})\mathbf{e}_{f_2}.$$

where the covariance matrix of $\mathbf{N}_1(i)$ and $\mathbf{N}_2(j)$ is given by (25), shown at the bottom of the page. In (25), the row index and column index of nonzero elements of the covariance matrix are indicated explicitly. $E\{\mathbf{N}_1(i)\mathbf{N}_3^{\mathrm{H}}(k)\}$ and $E\{\mathbf{N}_2(j)\mathbf{N}_3^{\mathrm{H}}(k)\}$ can be obtained similarly.

Furthermore, from (20) and (24), $\boldsymbol{\Psi}$ can be written as

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ab} & \boldsymbol{\Psi}_{ac} \\ \boldsymbol{\Psi}_{ab}^{\mathrm{H}} & \boldsymbol{\Psi}_{bb} & \boldsymbol{\Psi}_{bc} \\ \boldsymbol{\Psi}_{ac}^{\mathrm{H}} & \boldsymbol{\Psi}_{bc}^{\mathrm{H}} & \boldsymbol{\Psi}_{cc} \end{bmatrix}$$

with obvious notation. Let

$$\boldsymbol{\Psi}_1 = \begin{bmatrix} \boldsymbol{\Psi}_{bb} & \boldsymbol{\Psi}_{bc} \\ \boldsymbol{\Psi}_{bc}^{\mathrm{H}} & \boldsymbol{\Psi}_{cc} \end{bmatrix}, \quad \boldsymbol{\Psi}_2 = \begin{bmatrix} \boldsymbol{\Psi}_{ab} & \boldsymbol{\Psi}_{ac} \end{bmatrix}. \tag{26}$$

Then

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_2 \\ \boldsymbol{\Psi}_2^{\mathrm{H}} & \boldsymbol{\Psi}_1 \end{bmatrix}.$$

The CRB on the variance of any unbiased estimator of the trilinear model (15) is the inverse of $\boldsymbol{\Omega}$, which is given in terms of

$$\begin{bmatrix} \mathrm{CRB}_{aa} & \mathrm{CRB}_{ab} & \mathrm{CRB}_{ac} \\ \mathrm{CRB}_{ab}^{\mathrm{H}} & \mathrm{CRB}_{bb} & \mathrm{CRB}_{bc} \\ \mathrm{CRB}_{ac}^{\mathrm{H}} & \mathrm{CRB}_{bc}^{\mathrm{H}} & \mathrm{CRB}_{cc} \end{bmatrix} = \boldsymbol{\Psi}^{-1}. \tag{27}$$

To find the CRB of the unknown elements of $\mathbf{A}$, we use the following formula for the inverse of a partitioned Hermitian matrix:

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_3 \\ \mathbf{P}_3^{\mathrm{H}} & \mathbf{P}_2 \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{P}_1 - \mathbf{P}_3\mathbf{P}_2^{-1}\mathbf{P}_3^{\mathrm{H}}\right)^{-1} & -\mathbf{P}_1^{-1}\mathbf{P}_3\mathbf{P}_4 \\ -\mathbf{P}_4\mathbf{P}_3^{\mathrm{H}}\mathbf{P}_1^{-1} & \mathbf{P}_4 \end{bmatrix}$$

where $\mathbf{P}_4 = (\mathbf{P}_2 - \mathbf{P}_3^{\mathrm{H}}\mathbf{P}_1^{-1}\mathbf{P}_3)^{-1}$. It easily follows that

$$\mathrm{CRB}_{aa} = \left(\boldsymbol{\Psi}_{aa} - \boldsymbol{\Psi}_2\boldsymbol{\Psi}_1^{-1}\boldsymbol{\Psi}_2^{\mathrm{H}}\right)^{-1}.$$

$$E\left\{N_1(i)N_2^{\mathrm{H}}(j)\right\} = \sigma^2 \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ & & & \ddots & & & & & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{matrix} \\ \\ \longleftarrow (j-1)K+1 \\ \longleftarrow (j-1)K+2 \\ \vdots \\ \longleftarrow (j-1)K+K \\ \\ \end{matrix} \tag{25}$$

$$\begin{matrix} & & \uparrow & & \uparrow & \cdots & \uparrow & & \\ & & i & & I+i & & (K-1)I+i & & \end{matrix}$$

APPENDIX B
CRB FOR LOW-RANK DECOMPOSITION OF 4–D ARRAYS

Consider an $I \times J \times K \times L$ array $\underline{\mathbf{X}}$ with typical element

$$x_{i,j,k,l} = \sum_{f=1}^{F} a_{i,f} b_{j,f} g_{k,f} h_{l,f} + n_{i,j,k,l} \tag{28}$$

for $i = 1,\ldots,I$, $j = 1,\ldots,J$, $k = 1,\ldots,K$, $l = 1,\ldots,L$, with $a_{i,f}, b_{j,f}, g_{k,f}, h_{l,f} \in \mathbb{C}$. $n_{i,j,k,l}$ is i.i.d. (in i, j, k, and l) zero mean, complex Gaussian noise of variance σ^2. Define $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{H}$ as in Section II, and unfold $\underline{\mathbf{X}}$ in four equivalent matrix representations:

$$\mathbf{X}_1 = (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})\mathbf{A}^{\mathrm{T}} + \mathbf{N}_1, \ : LJK \times I \tag{29}$$
$$\mathbf{X}_2 = (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})\mathbf{B}^{\mathrm{T}} + \mathbf{N}_2, \ : IKL \times J \tag{30}$$
$$\mathbf{X}_3 = (\mathbf{A} \odot \mathbf{H} \odot \mathbf{B})\mathbf{G}^{\mathrm{T}} + \mathbf{N}_3, \ : JLI \times K \tag{31}$$
$$\mathbf{X}_4 = (\mathbf{B} \odot \mathbf{A} \odot \mathbf{G})\mathbf{H}^{\mathrm{T}} + \mathbf{N}_4, \ : KIJ \times L. \tag{32}$$

Correspondingly, we can write the likelihood function in four equivalent ways as

$$\begin{aligned} L(\underline{\mathbf{X}}) =& \frac{1}{(\pi\sigma^2)^{IJKL}} \\ & \times \exp\left\{ -\frac{1}{\sigma^2} \sum_{i=1}^{I} \|\mathbf{X}_1(i) - (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})\mathbf{a}_i\|^2 \right\} \\ =& \frac{1}{(\pi\sigma^2)^{IJKL}} \\ & \times \exp\left\{ -\frac{1}{\sigma^2} \sum_{j=1}^{J} \|\mathbf{X}_2(j) - (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})\mathbf{b}_j\|^2 \right\} \\ =& \frac{1}{(\pi\sigma^2)^{IJKL}} \\ & \times \exp\left\{ -\frac{1}{\sigma^2} \sum_{k=1}^{K} \|\mathbf{X}_3(k) - (\mathbf{A} \odot \mathbf{H} \odot \mathbf{B})\mathbf{g}_k\|^2 \right\} \\ =& \frac{1}{(\pi\sigma^2)^{IJKL}} \\ & \times \exp\left\{ -\frac{1}{\sigma^2} \sum_{l=1}^{L} \|\mathbf{X}_4(l) - (\mathbf{B} \odot \mathbf{A} \odot \mathbf{G})\mathbf{h}_l\|^2 \right\} \end{aligned}$$

where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}^{\mathrm{T}}$ and similarly for $\mathbf{b}_j$, $\mathbf{g}_k$ and $\mathbf{h}_l$. As mentioned in Section III, for the purpose of removing sale and permutation ambiguity to obtain a meaningful bound, we assume that the first row of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{H}$ are known. Therefore, the unknown $1 \times 2(I+J+K+L-4)F$ complex parameter vector is

$$\boldsymbol{\theta} = \Big[\mathbf{a}_2^{\mathrm{T}}, \ldots, \mathbf{a}_I^{\mathrm{T}}, \mathbf{b}_2^{\mathrm{T}}, \ldots, \mathbf{b}_J^{\mathrm{T}}, \mathbf{g}_2^{\mathrm{T}}, \ldots \mathbf{g}_K^{\mathrm{T}}, \mathbf{h}_2^{\mathrm{T}}, \ldots, \mathbf{h}_L^{\mathrm{T}}, \mathbf{a}_2^{\mathrm{H}}, \ldots, \mathbf{h}_L^{\mathrm{H}} \Big]$$

and the log-likelihood function $f(\boldsymbol{\theta}) = \ln L(\underline{\mathbf{X}})$ is given by

$$\begin{aligned} f(\boldsymbol{\theta}) =& -IJKL\ln(\pi\sigma^2) \\ & - \frac{1}{\sigma^2} \sum_{i=1}^{I} \|\mathbf{X}_1(i) - (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})\mathbf{a}_i\|^2 \\ =& -IJKL\ln(\pi\sigma^2) \\ & - \frac{1}{\sigma^2} \sum_{j=1}^{J} \|\mathbf{X}_2(j) - (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})\mathbf{b}_j\|^2 \\ =& -IJKL\ln(\pi\sigma^2) \\ & - \frac{1}{\sigma^2} \sum_{k=1}^{K} \|\mathbf{X}_3(k) - (\mathbf{A} \odot \mathbf{H} \odot \mathbf{B})\mathbf{g}_k\|^2 \\ =& -IJKL\ln(\pi\sigma^2) \\ & - \frac{1}{\sigma^2} \sum_{l=1}^{L} \|\mathbf{X}_4(l) - (\mathbf{B} \odot \mathbf{A} \odot \mathbf{G})\mathbf{h}_l\|^2. \end{aligned}$$

The complex FIM is

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = E\left\{ \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{H}} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right\}. \tag{33}$$

The partial derivatives of $f(\boldsymbol{\theta})$ with respect to the unknown parameters are given by

$$\begin{aligned} \frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}} =& \frac{1}{\sigma^2} (\mathbf{X}_1(i) - (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})\mathbf{a}_i)^{\mathrm{H}} (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})\mathbf{e}_f \\ \frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}} =& \frac{1}{\sigma^2} (\mathbf{X}_2(j) - (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})\mathbf{b}_j)^{\mathrm{H}} (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})\mathbf{e}_f \\ \frac{\partial f(\boldsymbol{\theta})}{\partial g_{k,f}} =& \frac{1}{\sigma^2} (\mathbf{X}_3(k) - (\mathbf{A} \odot \mathbf{H} \odot \mathbf{B})\mathbf{g}_k)^{\mathrm{H}} (\mathbf{A} \odot \mathbf{H} \odot \mathbf{B})\mathbf{e}_f \\ \frac{\partial f(\boldsymbol{\theta})}{\partial h_{l,f}} =& \frac{1}{\sigma^2} (\mathbf{X}_4(k) - (\mathbf{B} \odot \mathbf{A} \odot \mathbf{G})\mathbf{h}_l)^{\mathrm{H}} (\mathbf{B} \odot \mathbf{A} \odot \mathbf{G})\mathbf{e}_f \\ \frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}^*} =& \left(\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}} \right)^*, \quad \frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}^*} = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}} \right)^* \\ \frac{\partial f(\boldsymbol{\theta})}{\partial g_{k,f}^*} =& \left(\frac{\partial f(\boldsymbol{\theta})}{\partial g_{k,f}} \right)^*, \quad \frac{\partial f(\boldsymbol{\theta})}{\partial h_{l,f}^*} = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial h_{l,f}} \right)^* \end{aligned} \tag{34}$$

where $\mathbf{e}_f$ is the fth unit coordinate vector.

Similar to (24), the FIM $\boldsymbol{\Omega}$ in (33) can be written as

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^* \end{bmatrix} \tag{35}$$

whose size is $2(I+J+K+L-4)F \times 2(I+J+K+L-4)F$. The elements of $\boldsymbol{\Psi}$ can be computed as in

$$\begin{aligned} E\left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial a_{i_1,f_1}^*} \frac{\partial f(\boldsymbol{\theta})}{\partial a_{i_2,f_2}} \right\} =& \frac{1}{\sigma^2} \mathbf{e}_{f_1}^{\mathrm{H}} (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})^{\mathrm{H}} \\ & \times (\mathbf{G} \odot \mathbf{B} \odot \mathbf{H})\mathbf{e}_{f_2} \delta_{i_1,i_2} \end{aligned} \tag{36a}$$

$$\begin{aligned} E\left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial b_{j_1,f_1}^*} \frac{\partial f(\boldsymbol{\theta})}{\partial b_{j_2,f_2}} \right\} =& \frac{1}{\sigma^2} \mathbf{e}_{f_1}^{\mathrm{H}} (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})^{\mathrm{H}} \\ & \times (\mathbf{H} \odot \mathbf{G} \odot \mathbf{A})\mathbf{e}_{f_2} \delta_{j_1,j_2} \end{aligned} \tag{36b}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial g^*_{k_1,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial g_{k_2,f_2}}\right\}=\frac{1}{\sigma^2}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{A}\odot\mathbf{H}\odot\mathbf{B})^{\mathrm{H}}\times(\mathbf{A}\odot\mathbf{H}\odot\mathbf{B})\mathbf{e}_{f_2}\delta_{k_1,k_2} \tag{36c}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial h^*_{l_1,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial h_{l_2,f_2}}\right\}=\frac{1}{\sigma^2}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{B}\odot\mathbf{A}\odot\mathbf{G})^{\mathrm{H}}\times(\mathbf{B}\odot\mathbf{A}\odot\mathbf{G})\mathbf{e}_{f_2}\delta_{l_1,l_2} \tag{36d}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{i,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f_2}}\right\}=\frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{G}\odot\mathbf{B}\odot\mathbf{H})^{\mathrm{H}}\times E\left\{\boldsymbol{N}_1(i)\boldsymbol{N}_2^{\mathrm{H}}(j)\right\}\times(\mathbf{H}\odot\mathbf{G}\odot\mathbf{A})\mathbf{e}_{f_2} \tag{36e}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{i,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial g_{k,f_2}}\right\}=\frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{G}\odot\mathbf{B}\odot\mathbf{H})^{\mathrm{H}}\times E\left\{\boldsymbol{N}_1(i)\boldsymbol{N}_3^{\mathrm{H}}(k)\right\}\times(\mathbf{A}\odot\mathbf{H}\odot\mathbf{B})\mathbf{e}_{f_2} \tag{36f}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{i,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial h_{l,f_2}}\right\}=\frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{G}\odot\mathbf{B}\odot\mathbf{H})^{\mathrm{H}}\times E\left\{\boldsymbol{N}_1(i)\boldsymbol{N}_4^{\mathrm{H}}(l)\right\}\times(\mathbf{B}\odot\mathbf{A}\odot\mathbf{G})\mathbf{e}_{f_2} \tag{36g}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial b^*_{j,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial g_{k,f_2}}\right\}=\frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{H}\odot\mathbf{G}\odot\mathbf{A})^{\mathrm{H}}\times E\left\{\boldsymbol{N}_2(j)\boldsymbol{N}_3^{\mathrm{H}}(k)\right\}\times(\mathbf{A}\odot\mathbf{H}\odot\mathbf{B})\mathbf{e}_{f_2} \tag{36h}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial b^*_{j,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial h_{l,f_2}}\right\}=\frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{H}\odot\mathbf{G}\odot\mathbf{A})^{\mathrm{H}}\times E\left\{\boldsymbol{N}_2(j)\boldsymbol{N}_4^{\mathrm{H}}(l)\right\}\times(\mathbf{B}\odot\mathbf{A}\odot\mathbf{G})\mathbf{e}_{f_2} \tag{36i}$$

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial g^*_{k,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial h_{l,f_2}}\right\}=\frac{1}{\sigma^4}\mathbf{e}_{f_1}^{\mathrm{H}}(\mathbf{A}\odot\mathbf{H}\odot\mathbf{B})^{\mathrm{H}}\times E\left\{\boldsymbol{N}_3(k)\boldsymbol{N}_4^{\mathrm{H}}(l)\right\}\times(\mathbf{B}\odot\mathbf{A}\odot\mathbf{G})\mathbf{e}_{f_2}. \tag{36j}$$

It is cumbersome to write out the covariance matrix of $\mathbf{N}_1(i)$ and $\mathbf{N}_2(j)$, but its computation can be done in the following simple way. First, construct an auxiliary four-way array $\mathbf{U}$ and fill it with integers from 1 to $IJKL$ as

$$u_{i,j,k,l}=(i-1)JKL+(j-1)KL+(k-1)L+l.$$

By data rearrangement, we can express this four-way array in four different matrix forms, say, $\mathbf{U}_1^{(LJK\times I)}$, $\mathbf{U}_2^{(IKL\times J)}$, $\mathbf{U}_3^{(JLI\times K)}$ and $\mathbf{U}_4^{(KIJ\times L)}$ corresponding to (29)–(32), respectively. Now, to obtain $E\{\mathbf{N}_1(i)\mathbf{N}_2^{\mathrm{H}}(j)\}$, we can simply compare $\mathbf{U}_1(i)$ and $\mathbf{U}_2(j)$. If $\mathbf{U}_1(m,i)$ and $\mathbf{U}_2(n,j)$ contains the same integer, then $[E\{\mathbf{N}_1(i)\mathbf{N}_2^{\mathrm{H}}(j)\}]_{m,n}=\sigma^2$; otherwise, $[E\{\mathbf{N}_1(i)\mathbf{N}_2^{\mathrm{H}}(j)\}]_{m,n}=0$. In this way, we can easily obtain all necessary covariance matrices.

The CRB on the variance of any unbiased estimator of the quadrilinear model (28) is then given by the inverse of $\boldsymbol{\Omega}$ (35):

$$\mathrm{CRB}(\boldsymbol{\theta})=\boldsymbol{\Omega}^{-1}=\begin{bmatrix}\boldsymbol{\Psi}^{-1} & \mathbf{0}\\ \mathbf{0} & \left(\boldsymbol{\Psi}^{-1}\right)^*\end{bmatrix}.$$

Further simplification can be achieved as in Appendix A but is omitted here for brevity.

Appendix C
CRB for Low-Rank Decomposition of 3-D Arrays with Vandermonde Structure in One Dimension

Consider the $I\times J\times K$ three-way array $\underline{\mathbf{X}}$ in (15), and assume without loss of generality that $\mathbf{A}$ is Vandermonde:

$$\mathbf{A}=\begin{bmatrix}1 & 1 & \cdots & 1\\ a_1 & a_2 & \cdots & a_F\\ a_1^2 & a_2^2 & \cdots & a_F^2\\ & & \vdots & \\ a_1^{I-1} & a_2^{I-1} & \cdots & a_F^{I-1}\end{bmatrix}. \tag{37}$$

The log-likelihood function is the same as (19), whereas the $1\times 2(J+K-1)F$ complex parameter vector is given by

$$\boldsymbol{\theta}=\left[\mathbf{a}_2^{\mathrm{T}},\mathbf{b}_2^{\mathrm{T}},\ldots,\mathbf{b}_J^{\mathrm{T}},\mathbf{c}_2^{\mathrm{T}},\ldots,\mathbf{c}_K^{\mathrm{T}},\mathbf{a}_2^{\mathrm{H}},\ldots,\mathbf{c}_K^{\mathrm{H}}\right].$$

The partial derivatives of $f(\boldsymbol{\theta})$ with respect to a_f and a_f^* are

$$\frac{\partial f(\boldsymbol{\theta})}{\partial a_f}=\sum_{i=1}^{I}\frac{(i-1)}{\sigma^2}a_f^{i-2}\left[\mathbf{X}_1(i)-(\mathbf{B}\odot\mathbf{C})\mathbf{a}_i\right]^{\mathrm{H}}(\mathbf{B}\odot\mathbf{C})\mathbf{e}_f$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial a_f^*}=\left(\frac{\partial f(\boldsymbol{\theta})}{\partial a_f}\right)^*.$$

The derivatives with respect to $b_{j,f}$, $c_{k,f}$, $b^*_{j,f}$ and $c^*_{k,f}$ are the same as in (21b), (21c), (21e), and (21f). The remaining procedures to obtain the CRB are similar to those in Appendix A and, hence, are omitted for brevity.

Appendix D
CRB for Real-Parameter Low-Rank Decomposition of 3-D Arrays

The real case yields simpler CRB expressions, due to the fact that the noise covariance matrix can be written in convenient closed form, and an alternative scalar computation approach can be adopted. Consider the $I\times J\times K$ three-way array $\underline{\mathbf{X}}$ with typical element

$$x_{i,j,k}=\sum_{f=1}^{F}a_{i,f}b_{j,f}c_{k,f}+n_{i,j,k}$$

where $n_{i,j,k}$ is zero mean real-valued Gaussian i.i.d. in i, j, k of variance σ^2.

The likelihood function of $\underline{\mathbf{X}}$ is given by

$$L(\underline{\mathbf{X}})=\prod_{i,j,k}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp-\frac{1}{2\sigma^2}\left(x_{i,j,k}-\sum_{f=1}^{F}a_{i,f}b_{j,f}c_{k,f}\right)^2\right].$$

Thus, the log-likelihood function is

$$f(\boldsymbol{\theta}) = -\frac{IJK}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i,j,k}\left(x_{i,j,k} - \sum_{f=1}^{F} a_{i,f}b_{j,f}c_{k,f}\right)^2$$

where

$$\boldsymbol{\theta} = \left[\mathbf{a}_2^{\mathrm{T}},\ldots,\mathbf{a}_I^{\mathrm{T}},\mathbf{b}_2^{\mathrm{T}},\ldots,\mathbf{b}_J^{\mathrm{T}},\mathbf{c}_2^{\mathrm{T}},\ldots,\mathbf{c}_K^{\mathrm{T}}\right].$$

The FIM associated with the estimation of these parameters is given by

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = E\left\{\left(\frac{\partial f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^{\mathrm{T}}\left(\frac{\partial f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right\}. \tag{38}$$

First, we calculate the derivatives of $f(\boldsymbol{\theta})$ with respect to $a_{i,f}$, $b_{j,f}$, and $c_{k,f}$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f}} = \frac{1}{\sigma^2}\sum_{j,k} n_{i,j,k}b_{j,f}c_{k,f} \tag{39}$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f}} = \frac{1}{\sigma^2}\sum_{i,k} n_{i,j,k}a_{i,f}c_{k,f} \tag{40}$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f}} = \frac{1}{\sigma^2}\sum_{i,j} n_{i,j,k}a_{i,f}b_{j,f}. \tag{41}$$

From (39)–(41), we can obtain

$$E\left[\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i_1,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i_2,f_2}}\right] = \frac{1}{\sigma^2}\sum_{j,k}(b_{j,f_1}b_{j,f_2}c_{k,f_1}c_{k,f_2})\delta_{i_1,i_2}$$

$$E\left[\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j_1,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j_2,f_2}}\right] = \frac{1}{\sigma^2}\sum_{i,k}(a_{i,f_1}a_{i,f_2}c_{k,f_1}c_{k,f_2})\delta_{j_1,j_2}$$

$$E\left[\frac{\partial(\boldsymbol{\theta})}{\partial c_{k_1,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k_2,f_2}}\right] = \frac{1}{\sigma^2}\sum_{i,j}(a_{i,f_1}a_{i,f_2}b_{j,f_1}b_{j,f_2})\delta_{k_1,k_2}$$

$$E\left[\frac{\partial(\boldsymbol{\theta})}{\partial a_{i,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f_2}}\right] = \frac{1}{\sigma^2}\sum_{k}(c_{k,f_1}c_{k,f_2})a_{i,f_2}b_{j,f_1}$$

$$E\left[\frac{\partial f(\boldsymbol{\theta})}{\partial a_{i,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f_2}}\right] = \frac{1}{\sigma^2}\sum_{j}(b_{j,f_1}b_{j,f_2})a_{i,f_2}c_{k,f_1}$$

$$E\left[\frac{\partial f(\boldsymbol{\theta})}{\partial b_{j,f_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{k,f_2}}\right] = \frac{1}{\sigma^2}\sum_{i}(a_{i,f_1}a_{i,f_2})b_{j,f_2}c_{k,f_1}.$$

Define the matrices

$$\mathbf{H}_a = \sum_{i=1}^{I}\mathbf{a}_i\mathbf{a}_i^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$$

$$\mathbf{H}_b = \sum_{j=1}^{J}\mathbf{b}_j\mathbf{b}_j^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{B}$$

$$\mathbf{H}_c = \sum_{k=1}^{K}\mathbf{c}_k\mathbf{c}_k^{\mathrm{T}} = \mathbf{C}^{\mathrm{T}}\mathbf{C}.$$

Then, it can be shown that the FIM in (38) is given by

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\begin{bmatrix}\boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ab} & \boldsymbol{\Psi}_{ac}\\ \boldsymbol{\Psi}_{ab}^{\mathrm{T}} & \boldsymbol{\Psi}_{bb} & \boldsymbol{\Psi}_{bc}\\ \boldsymbol{\Psi}_{ac}^{\mathrm{T}} & \boldsymbol{\Psi}_{bc}^{\mathrm{T}} & \boldsymbol{\Psi}_{cc}\end{bmatrix} \tag{42}$$

where $\boldsymbol{\Psi}_{aa}$ is an $(I-1)F\times(I-1)F$ block diagonal matrix containing $(I-1)$ replicas of $\mathbf{H}_b\diamond\mathbf{H}_c$, and $\boldsymbol{\Psi}_{bb}$, $\boldsymbol{\Psi}_{cc}$ are constructed similarly, i.e.,

$$\begin{aligned}\boldsymbol{\Psi}_{aa} &= \mathbf{I}_{I-1}\otimes(\mathbf{H}_b\diamond\mathbf{H}_c)\\ \boldsymbol{\Psi}_{bb} &= \mathbf{I}_{J-1}\otimes(\mathbf{H}_c\diamond\mathbf{H}_a)\\ \boldsymbol{\Psi}_{cc} &= \mathbf{I}_{K-1}\otimes(\mathbf{H}_a\diamond\mathbf{H}_b)\end{aligned}$$

where $\mathbf{I}_{I-1}$ denotes an $(I-1)\times(I-1)$ identity matrix. $\boldsymbol{\Psi}_{ab}$, $\boldsymbol{\Psi}_{ac}$, and $\boldsymbol{\Psi}_{bc}$ in (42) are given by (43)–(45), as follows:

$$\boldsymbol{\Psi}_{ab} = \begin{bmatrix}\left(\mathbf{b}_2\mathbf{a}_2^{\mathrm{T}}\right)\diamond\mathbf{H}_c & \cdots & \left(\mathbf{b}_J\mathbf{a}_2^{\mathrm{T}}\right)\diamond\mathbf{H}_c\\ \vdots & \ddots & \vdots\\ \left(\mathbf{b}_2\mathbf{a}_I^{\mathrm{T}}\right)\diamond\mathbf{H}_c & \cdots & \left(\mathbf{b}_J\mathbf{a}_I^{\mathrm{T}}\right)\diamond\mathbf{H}_c\end{bmatrix} \tag{43}$$

$$\boldsymbol{\Psi}_{ac} = \begin{bmatrix}\left(\mathbf{c}_2\mathbf{a}_2^{\mathrm{T}}\right)\diamond\mathbf{H}_b & \cdots & \left(\mathbf{c}_K\mathbf{a}_2^{\mathrm{T}}\right)\diamond\mathbf{H}_b\\ \vdots & \ddots & \vdots\\ \left(\mathbf{c}_2\mathbf{a}_I^{\mathrm{T}}\right)\diamond\mathbf{H}_b & \cdots & \left(\mathbf{c}_K\mathbf{a}_I^{\mathrm{T}}\right)\diamond\mathbf{H}_b\end{bmatrix} \tag{44}$$

$$\boldsymbol{\Psi}_{bc} = \begin{bmatrix}\left(\mathbf{c}_2\mathbf{b}_2^{\mathrm{T}}\right)\diamond\mathbf{H}_a & \cdots & \left(\mathbf{c}_K\mathbf{b}_2^{\mathrm{T}}\right)\diamond\mathbf{H}_a\\ \vdots & \ddots & \vdots\\ \left(\mathbf{c}_2\mathbf{b}_J^{\mathrm{T}}\right)\diamond\mathbf{H}_a & \cdots & \left(\mathbf{c}_K\mathbf{b}_J^{\mathrm{T}}\right)\diamond\mathbf{H}_a\end{bmatrix}. \tag{45}$$

The CRB on the variance of any unbiased estimator is then given by the inverse of $\boldsymbol{\Omega}$ (42), provided it exists.

ACKNOWLEDGMENT

The authors would like to thank an anonymous reviewer for constructive comments that helped improve the quality and readability of this paper.

REFERENCES

[1] S. Basu and Y. Bresler, "The stability of nonlinear least squares problems and the Cramér–Rao bound," *IEEE Trans. Signal Processing*, vol. 48, pp. 3426–3436, Dec. 2000.

[2] J. Beltran, J. Guiteras, and R. Ferrer, "Three-way multivariate calibration procedures applied to high-performance liquid chromatography coupled with fast-scanning spectrometry (HPLC-FSFS) detection. Determination of aromatic hydrocarbons in water samples," *Anal. Chem.*, vol. 70, pp. 1949–1955, 1998.

[3] A. van den Bos, "A Cramér-Rao lower bound for complex parameters," *IEEE Trans. Signal Processing*, vol. 42, p. 2859, Oct. 1994.

[4] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 772–781, Sept. 1978.

[5] R. Bro, "PARAFAC: tutorial and applications," *Chemometr. Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.

[6] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[7] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267–283, 1944.

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

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

[10] ——, "PARAFAC2: mathematical and technical notes," *UCLA Working Papers Phonetics*, vol. 22, pp. 33–44, 1972.
[11] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," in *Res. Methods Multimode Data Anal.*, H. G. Law, C. W. Snyder, Jr, J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.
[12] S. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1993, vol. I, Estimation Theory.
[13] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhya, Se. A*, vol. 30, pp. 167–180, 1968.
[14] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Lin. Algebra Applicat.*, vol. 18, pp. 95–138, 1977.
[15] ——, "Rank, decomposition, and uniqueness for 3-way and N-way arrays," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 7–18.
[16] J. K. Lee, R. T. Ross, S. Thampi, and S. E. Leurgans, "Resolution of the properties of hydrogen-bonded tyrosine using a trilinear model of fluorescence," *J. Phys. Chem.*, vol. 96, pp. 9158–6192, 1992.
[17] X. Liu and N. Sidiropoulos, "PARAFAC methods for blind beamforming: multilinear ALS performance and CRB," in *Proc ICASSP*, Istanbul, Turkey, June 5–9, 2000.
[18] C. R. Rao, "Minimum variance and the estimation of several parameters," in *Proc. Cambridge Phil. Soc.*, vol. 43, 1946, pp. 280–283.
[19] E. Sanchez and B. R. Kowalski, "Generalized rank annihilation factor analysis," *Anal. Chem.*, vol. 58, pp. 496–499, 1986.
[20] P. H. Schonemann, "An algebraic solution for a class of subjective metrics models," *Psychometrika*, vol. 37, p. 441, 1972.
[21] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chem.*, vol. 14, pp. 229–239, May 2000.
[22] ——, "On communication diversity for blind identifiability and the uniqueness of low-rank decomposition of N-way arrays," in *Proc ICASSP*, Istanbul, Turkey, June 5–9, 2000.
[23] ——, "PARAFAC techniques for signal separation," in *Signal Processing Advances in Wireless Communications*, P. Stoica, G. Giannakis, Y. Hua, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2000, vol. 2, ch. 4.
[24] 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.
[25] 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.
[26] N. D. Sidiropoulos and X. Liu, "Identifiability results for blind beamforming in incoherent multipath with small delay spread," *IEEE Trans. Signal Processing*, vol. 49, pp. 228–236, Jan. 2001.
[27] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
[28] S. F. Yau and Y. Bresler, "A compact Cramér–Rao bound expression for parametric estimation of superimposed signals," *IEEE Trans. Signal Processing*, vol. 40, pp. 1226–1229, May 1992.

Xiangqian Liu received the B.S. degree from Beijing Institute of Technology, Beijing, China, in 1997 and the M.S. degree from University of Virginia, Charlottesville, in 1999, both in electrical engineering. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

His research interests are in the area of signal processing for communications with focus on wireless multiuser communications.

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 with the Institute for Systems Research (ISR), UMCP. From September 1992 to June 1994, he served his military service as a Lecturer with the Hellenic Air Force Academy. From October 1993 to June 1994, he was a Member of Technical Staff, Systems Integration Division, G-Systems Ltd., Athens, Greece. He has held Postdoctoral (from 1994 to 1995) and Research Scientist (from 1996 to 1997) positions at ISR-UMCP, before joining the Department of Electrical Engineering, University of Virginia , Charlottesville, in July 1997 as an Assistant Professor. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His current research interests are primarily in multi-way analysis and its applications in signal processing for communications and networking.

Dr. Sidiropoulos is a member of the Signal Processing for Communications Technical Committee (SPCOM-TC) of the IEEE SP Society and currently serves as Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He received the NSF/CAREER award for the Signal Processing Systems Program in June 1998.