LOW-RANK DECOMPOSITION OF MULTI-WAY ARRAYS: A SIGNAL PROCESSING PERSPECTIVE

N. D. Sidiropoulos

Dept. of ECE, Tech. Univ. of Crete,
73100 Chania - Crete, Greece
&
Dept. of ECE, Univ. of Minnesota - Minneapolis, U.S.A.
nikos@telecom.tuc.gr

ABSTRACT

In many signal processing applications of linear algebra tools, the signal part of a postulated model lies in a so-called *signal subspace*, while the parameters of interest are in one-to-one correspondence with a certain *basis* of this subspace. The signal subspace can often be reliably estimated from measured data, but the particular basis of interest cannot be identified without additional problem-specific structure. This is a manifestation of rotational indeterminacy, i.e., non-uniqueness of low-rank matrix decomposition. The situation is very different for three- or higher-way arrays, i.e., arrays indexed by three or more independent variables, for which low-rank decomposition is unique under mild conditions. This has fundamental implications for DSP problems which deal with such data. This paper provides a brief tour of the basic elements of this theory, along with many examples of application in problems of current interest in the signal processing community.

Keywords: Three-way analysis, low-rank decomposition, parallel factor analysis (PARAFAC), canonical decomposition (CANDECOMP)

1. INTRODUCTION

Three-way and higher-way data (that is, arrays indexed by three or more independent variables) abound in modern applications of signal processing. Here are some examples:

- A video signal.
- Blind multiuser detection-estimation in Direct-Sequence Code-Division Multiple-Access (DS-CDMA) communication, using a receive antenna array [42, 41, 36]. Here, the dimensions of the array will be (number of - #) chips per symbol $\times\#$ symbols $\times\#$ antennas.
- Multiple-invariance sensor array processing (MI-SAP) [40]: # displaced subarrays $\times\#$ subarray elements $\times\#$ temporal snapshots.
- Joint detection-estimation for OFDM systems subject to CFO, using receive diversity [19]: # tones $\times\#$ symbols $\times\#$ receive diversity branches.
- Blind decoding of a class of linear space-time codes [39].
- 3-D Radar clutter modeling and mitigation [25]: # range gates $\times\#$ Doppler bins $\times\#$ spatial or beamspace sectors.
- Exploratory data analysis (e.g., [2, 3, 5]): clustering, scatter plots, multi-dimensional scaling [8, 24], ... with applications to chromatography, spectroscopy, and magnetic resonance. These applications parallel classical applications of principal component analysis, but with the added benefit of naturally fixed axes.

In addition to the above applications, some of which may sound exotic, there are more bread-and-butter signal processing applications of multi-way analysis. Multi-dimensional harmonic retrieval and joint diagonalization problems can be viewed as structured higher-way analysis problems [34, 20, 27, 29, 28, 38]. Blind spatial signature estimation from covariance matrices, using time-varying power loading [31], spectral color, or multiple covariance lags, can also be cast in the multi-way analysis framework. The latter class of problems includes, e.g., blind source separation for multi-channel speech signals. Certain higher-order statistics based parameter estimation and signal separation problems also fall under multi-way analysis [10, 11, 12].

At this point, it should be clear that our subject matter is not only about fascinating mathematics. However, it does not hurt that the underlying theory is genuinely interesting, exhibiting many surprising twists, as we will see. This has kept me going over the years, and I hope it will also get you interested.

While signal processing researchers often work with three-way data sets, they usually convert them to more familiar matrix representations prior to any processing - as in working with specific video frames, or "unfolding" the higher-dimensional structure by stacking frames into a tall matrix. Subsequent processing usually ignores the underlying algebraic structure of the original higher-way array, which is often crucial for proper interpretation of the measured data. This practice is mostly due to the fact that algebraic analysis of higher-way arrays is quite different from matrix algebra, and largely unknown to the signal processing community. The purpose of this paper is to shed light in this direction.

We will note several similarities and differences between three-way arrays and ordinary matrices (two-way arrays). In particular, we will focus on the concept of rank, and how it naturally generalizes to three- and higher-way arrays. Mastering three-way notation is very important for getting into the subject matter, and it

Supported in part by ARL Cooperative Agreement DADD19-01-2-0011, and in part by a GSRT collaborative research grant between the Technical University of Crete and the University of Minnesota.

usually is the stumbling block that dissuades newcomers. For this reason, we will carefully develop the associated notation, making extensive use of the *Khatri-Rao* (column-wise Kronecker) product, which enables compact matrix representation of three- and higher-way data. Symmetry will also play a key role; look out for symmetries in the various equivalent representations of the data. After these preliminaries, we will take a closer look at data modeling aspects of various applications. Uniqueness (i.e., model identifiability) is central to three-way analysis; accordingly, much of our exposition will be devoted to it. We will not lose sight of algorithms though, and performance analysis will also be touched upon. Essential web pointers and a judiciously compiled bibliography will be provided at the end of the paper. While far from being comprehensive, these references do provide a good roadmap for further study. We will wrap things up with a discussion of open issues and some problems for further research.

2. THREE-WAY ARRAYS

Consider a matrix (that is, two-way array) $\mathbf{X} := [x_{i,j}] : (I \times J)$. A three-way array will be denoted by $[x_{i,j,k}] : (I \times J \times K)$.

2.1. Three-Way Versus Two-Way Arrays

A rank-one matrix is generated as the outer product of two vectors, and it contains all double products of elements drawn from the respective vectors. The rank of a matrix can then be defined as the smallest number of rank-one matrices that is needed for exact additive decomposition of the given matrix. This concept naturally generalizes to three-way arrays [22, 23]: The rank of a three-way array is the smallest number of rank-one three-way arrays needed for its decomposition. The only extension here is that a rank-one three-way array is the outer product of *three* vectors, containing all triple products.

So far so good, but this is essentially where the similarity ends. Whereas for two-way $(I \times J)$ arrays it holds that the row-rank (# linearly independent rows) = column-rank (# linearly independent columns) = rank $\leq \min(I, J)$, for three-way arrays the row-rank (# linearly independent "fibers" along the row dimension) $\neq$ column-rank $\neq$ "tube"-rank $\neq$ rank. Furthermore, if we draw a random Gaussian matrix in Matlab, then rank(randn(I,J))=min(I,J) with probability (w.p.) 1; whereas for a random Gaussian $2 \times 2 \times 2$ three-way array, rank(randn(2,2,2)) is a random variable (2 w.p. 0.3, 3 w.p. 0.7) [23]. Thus, while maximal rank and typical rank coincide for matrices, this is not the case for three-way arrays ([23]; see also [44, 30, 49]). The rank of a real matrix remains the same if we decompose it over the complex field; not so for real three-way arrays, whose decomposition is sensitive to whether or not the underlying field is open or closed ($\mathbb{R}$ versus $\mathbb{C}$) [22]. Furthermore, except for loose bounds and special cases (e.g., [23, 44]), general results for maximal rank and typical rank are sorely missing for decomposition over $\mathbb{R}$, although the theory is more developed for decomposition over $\mathbb{C}$ [7].

2.2. Khatri-Rao Product

The Khatri-Rao product [21] of two matrices $\mathbf{A}$ $(I \times F)$ and $\mathbf{B}$ $(J \times F)$ is defined as

$$\mathbf{A} \odot \mathbf{B} := [\mathbf{a}_1 \otimes \mathbf{b}_1 \cdots \mathbf{a}_F \otimes \mathbf{b}_f] \,,$$

where $\otimes$ stands for the Kronecker product [1]. The Khatri-Rao product has the following useful properties [1]:

$$vec(\mathbf{A}\mathbf{D}\mathbf{B}^T) = (\mathbf{B} \odot \mathbf{A})\mathbf{d}(\mathbf{D}),$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{d}(\mathbf{D})$ is a column vector containing the diagonal of $\mathbf{D}$. Furthermore,

$$\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}$$

2.3. Decomposition of Three- and Higher-Way Arrays

In *scalar* form, a rank-F three way array can be written as a sum of triple products

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

$$i = 1, \cdots, I, \; j = 1, \cdots, J, \; k = 1, \cdots, K$$

which can also be expressed in so-called *slab* form

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}^T, \; k = 1, \cdots, K$$

where $\mathbf{A} := [a_{i,f}]$ $(I \times F)$, $\mathbf{B} := [b_{j,f}]$ $(J \times F)$, $\mathbf{C} := [c_{k,f}]$ $(K \times F)$, and $\mathbf{D}_k(\mathbf{C})$ is a diagonal matrix holding the k-th row of $\mathbf{C}$ along its diagonal. We can combine all the above equations for $k = 1, \cdots, K$ in a single tall *matrix* equation

$$\mathbf{X}^{(KJ \times I)} = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T$$

and further recast the latter in *vector* form, using the vectorization property of the Khatri-Rao product

$$\mathbf{x}^{(KJI)} := vec\left(\mathbf{X}^{(KJ \times I)}\right) = (\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}))\, \mathbf{1}_{F \times 1} =$$

$$(\mathbf{A} \odot \mathbf{B} \odot \mathbf{C})\, \mathbf{1}_{F \times 1}$$

The model in either one of the above representations is variably known as *Parallel Factor Analysis* (PARAFAC) [13, 14, 15], *Canonical Decomposition* (CANDECOMP) [8], and *triple-product* (or, *trilinear*) decomposition [22, 23]. We will use the shorthand $\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ to denote the three-way array generated by any of the above equations for given $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

Jos ten Berge proposed the compact acronym CP (Candecomp - Parafac), which we will also adopt for brevity. If F is smaller than the maximal possible rank, then it also makes sense to think of CP as *low-rank decomposition*. In applications, the CP model is usually fitted to noisy measured data, and F is generally much lower than what is needed for exact decomposition of the measured data array. This corresponds to *low-rank approximation* of the measured data array.

CP was independently introduced by Harshman [13, 14, 15] and Carroll & Chang [8], inspired by early exploratory work by Cattell [9] (see also Schonemann [33]).

For $(N > 3)$-way arrays, the situation is similar: we have the *scalar* sum of N-fold products equation

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

the *matrix* equation

$$\mathbf{X}^{(I_1 \cdots I_{N-1} \times I_N)} = \left(\mathbf{A}^{(N-1)} \odot \mathbf{A}^{(N-2)} \odot \cdots \odot \mathbf{A}^{(1)}\right) \left(\mathbf{A}^{(N)}\right)^T$$

and the *vector* equation

$$\mathbf{x}^{(I_1 \cdots I_N)} := vec\left(\mathbf{X}^{(I_1 \cdots I_{N-1} \times I_N)}\right) =$$

$$\left(\mathbf{A}^{(N)} \odot \mathbf{A}^{(N-1)} \odot \mathbf{A}^{(N-2)} \odot \cdots \odot \mathbf{A}^{(1)}\right) \mathbf{1}_{F \times 1}$$

2.4. What is the Physical Meaning of Multi-Way Array Rank?

The meaning of $I, J, K, \ldots$ is easy to see in the context of any particular application - I is the sample length along one measurement dimension, J is the sample length along another, and so on. However, the physical meaning of array rank may not be as clear. For this, we will need to take a closer look at some applications.

Consider, for example, the problem of blind multiuser detection-separation in the context of a synchronous DS-CDMA system with symbol-periodic spreading. Let (i, j, k, f) stand for (receive antenna, symbol snapshot, chip, user). Then the discrete-time baseband-equivalent chip-rate samples can be arranged in a three-way structure [42]

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

$$i = 1, \cdots, I,\ j = 1, \cdots, J,\ k = 1, \cdots, K$$

where I is the number of receive antennas, J is the number of symbols, K is the number of chips per symbol (spreading gain), and F is the number of users in the system. Here, $a_{i,f}$ is the response of antenna i to user f, $b_{j,f}$ is the j-th symbol emmitted by user f, and $c_{k,f}$ is the k-th chip of the spreading code of user f. We see that F corresponds to the number of users in the system, while $I \times J \times K$ corresponds to the total number of measurements. Also note that, the rank of the three-way array generated as $\sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f}$ is generally $\leq F$; in order to be equal to F, we need certain uniqueness conditions, to be discussed later.

In the context of MI-SAP, the slab notation is better suited to explain the link between low-rank three-way decomposition and the MI-SAP baseband-equivalent model. Let $\mathbf{A}$ be the response of the reference subarray, $\mathbf{B}^T$ the temporal signal matrix (usually denoted $\mathbf{S}$), and let $\mathbf{D}_k(\mathbf{C})$ hold the phase shifts for the k-th displaced but otherwise identical subarray. Then [40]

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}^T,\ k = 1, \cdots, K$$

Clearly then, F stands for the number of narrowband sources. Note that rank can be augmented in the presence of incoherent multipath, up to the total number of signal paths.

Let us now turn to a Single-Input Multiple-Output (SIMO) Orthogonal Frequency-Division Multiplexing (OFDM) data model. At the transmitter, N symbols are first linearly combined using an inverse $N \times N$ FFT matrix $\mathbf{F}^H$, and then a Cyclic Prefix (CP) of length L is added before transmission. Thus the signal

$$\mathbf{x}(k) = \mathbf{T}_{cp}\mathbf{F}^H \mathbf{s}^T(k)$$

is serially transmitted through a multipath channel of order $\leq L$. At receive antenna element i, $i = 1, \cdots, I \geq 2$, after suitable processing (using that circular matrices are diagonalized by the FFT):

$$\mathbf{y}_i(k) = \mathbf{P}\mathbf{F}^H \mathbf{H}_i \mathbf{s}^T(k) e^{j\phi((k+1)(N+L)-N)} + \mathbf{w}_i(k)$$

where $\mathbf{H}_i$ is $N \times N$ diagonal, containing the channel's frequency response, and

$$\mathbf{P} := \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & e^{j(n-1)\phi} & \vdots \\ 0 & \cdots & e^{j(N-1)\phi} \end{bmatrix} \in \mathbb{C}^{N \times N}$$

models the Carrier Frequency Offset (CFO). Collecting K such snapshots, we write

$$\mathbf{Y}_i = \mathbf{P}\mathbf{F}^H \mathbf{H}_i (\mathbf{Q}\mathbf{S})^T + \mathbf{W}_i =: \mathbf{A}\mathbf{D}_i\mathbf{B}^T + \mathbf{W}_i, i = 1, \cdots, I$$

This is a CP decomposition of order N (the number of tones), with extra structure on the factor matrices [19].

3. UNIQUENESS

It is easy to see from the scalar view of the CP model that a permutation of the F rank-one terms will not affect the data. Furthermore, it clear that it is possible to scale, e.g., the first column of $\mathbf{A}$, and, independently, the first column of $\mathbf{B}$, and, so long as the first column of $\mathbf{C}$ is counter-scaled, this will have no effect on the data. From this, it is clear that the CP model entails an inherent column permutation ambiguity that is common to all three component matrices, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and column scaling / counter-scaling ambiguity on each triple of corresponding columns. The key question then is under which conditions, if any, these trivial ambiguities are the *only* model ambiguities? This is the subject of CP uniqueness, i.e., CP model identifiability.

The systematic study of CP uniqueness goes back to the 70's. The first results are due to Jennrich and Harshman [13, 14]). The most general sufficient condition for essential uniqueness is due to Kruskal [22]. In his seminal paper [22], Kruskal showed that for $N = 3$ and decomposition over $\mathbb{R}$, if $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2F + 2$, where $k_{\mathbf{A}}$ is the *k-rank* of $\mathbf{A}$ = maximum r such that *every* r columns of $\mathbf{A}$ are linearly independent, then CP decomposition is unique, except for the trivial permutation and scaling ambiguity. Sidiropoulos *et al*, further qualified Kruskal's condition for decomposition over $\mathbb{C}$ [42]. Sidiropoulos & Bro [35] further generalized the condition to any N.

Kruskal built his original proof on the following cornerstone result:

• Kruskal's Permutation Lemma [22]: Consider $\mathbf{A}$ ($I \times F$) having no zero column, and $\bar{\mathbf{A}}$ ($I \times \bar{F}$). Let $w(\cdot)$ be the *weight* (# of nonzero elements) of its argument, and $r_{\mathbf{A}}$ stand for the rank of matrix $\mathbf{A}$. If for any vector $\mathbf{x}$ such that

$$w(\mathbf{x}^H \bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}} + 1,$$

we have

$$w(\mathbf{x}^H \mathbf{A}) \leq w(\mathbf{x}^H \bar{\mathbf{A}}),$$

then $F \leq \bar{F}$; if also $F \geq \bar{F}$, then $F = \bar{F}$, and there exist a permutation matrix $\mathbf{\Pi}$ and a non-singular diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \bar{\mathbf{A}}\mathbf{\Pi}\mathbf{D}$.

It is easy to demonstrate an *elementary* version of the result, for a pair of square nonsingular matrices (using the rows of the pseudo-inverse). The interesting case, however, is for fat matrices, for which the result is deep and difficult to show. On this, see a recent clear-cut proof in [18], based solely on induction and contradiction (but still lengthy).

Instead of asking for faith in the permutation Lemma, we shall take a different, perhaps more pedagogical route. We start with

a very simple (albeit also very restricted) proof of uniqueness, that parallels the original arguments of Harshman and Jennrich [13, 14, 33], and is based solely on eigenvalue and singular value decomposition (EVD, SVD, respectively). Then, we will present another proof that is closer in both claim and spirit to Kruskal's proof - thus allowing the reader to develop an appreciation of the mathematical underpinnings of the area.

Consider the 2-slab model, with square/tall and full column rank $\mathbf{A}(I \times F)$ and $\mathbf{B}(J \times F)$, and distinct elements along the diagonal of $\mathbf{D}$:

$$\begin{array}{rcl} \mathbf{X}_1 & = & \mathbf{A}\mathbf{B}^T, \\ \mathbf{X}_2 & = & \mathbf{A}\mathbf{D}\mathbf{B}^T. \end{array}$$

Introduce the compact F-component SVD of

$$\left[\begin{array}{c} \mathbf{X}_1 \\ \mathbf{X}_2 \end{array}\right] = \left[\begin{array}{c} \mathbf{A} \\ \mathbf{A}\mathbf{D} \end{array}\right] \mathbf{B}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$

$rank(\mathbf{B}) = F$ implies that span$(\mathbf{U})$ = span$(\left[\begin{array}{c} \mathbf{A} \\ \mathbf{A}\mathbf{D} \end{array}\right])$; hence there exists a nonsingular matrix $\mathbf{P}$ such that

$$\mathbf{U} = \left[\begin{array}{c} \mathbf{U}_1 \\ \mathbf{U}_2 \end{array}\right] = \left[\begin{array}{c} \mathbf{A} \\ \mathbf{A}\mathbf{D} \end{array}\right] \mathbf{P}.$$

Next, construct the auto- and cross-product matrices

$$\begin{array}{rclclcl} \mathbf{R}_1 & = & \mathbf{U}_1^H \mathbf{U_1} & = & \mathbf{P}^H \mathbf{A}^H \mathbf{A}\mathbf{P} & =: & \mathbf{Q}\mathbf{P}, \\ \mathbf{R}_2 & = & \mathbf{U}_1^H \mathbf{U_2} & = & \mathbf{P}^H \mathbf{A}^H \mathbf{A}\mathbf{D}\mathbf{P} & = & \mathbf{Q}\mathbf{D}\mathbf{P}. \end{array}$$

and note that the matrices on the right hand side are square and full-rank. Solving the first equation and substituting to the second, yields the eigenvalue problem

$$\left(\mathbf{R}_1^{-1}\mathbf{R}_2\right)\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{D}$$

From which the sought matrices can be recovered, up to a common column permutation, and column scaling / counter-scaling, which carry over from the solution of the eigenvalue problem.

As mentioned earlier, the proof of Kruskal's result relies on his *Permutation Lemma* (in its general form), and is beyond our scope. The following is more palatable:

Theorem: Given $\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C})$, with $\mathbf{A}: \ I \times F$, $\mathbf{B}: \ J \times F$, and $\mathbf{C}: \ K \times F$, it is *necessary* for uniqueness of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ that $\min(r_{\mathbf{A}\odot\mathbf{B}}, r_{\mathbf{C}\odot\mathbf{A}}, r_{\mathbf{B}\odot\mathbf{C}}) = F$. If $F > 1$, then it is also necessary that $\min(k_{\mathbf{A}}, k_{\mathbf{B}}, k_{\mathbf{C}}) \geq 2$.

If, in addition $r_{\mathbf{C}} = F$, and $k_{\mathbf{A}} + k_{\mathbf{B}} \geq F + 2$, then $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are unique up to permutation and scaling of columns, meaning that if $\underline{\mathbf{X}} = (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$, for some $\bar{\mathbf{A}}: \ I \times F$, $\bar{\mathbf{B}}: \ J \times F$, and $\bar{\mathbf{C}}: \ K \times F$, then there exists a permutation matrix $\mathbf{\Pi}$ and diagonal scaling matrices $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mathbf{\Lambda}_3$ such that

$$\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}\mathbf{\Lambda}_1, \ \bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda}_2, \ \bar{\mathbf{C}} = \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}_3, \ \mathbf{\Lambda}_1\mathbf{\Lambda}_2\mathbf{\Lambda}_3 = \mathbf{I}.$$

Proof: Necessary conditions [26] follow from

$$\mathbf{X}^{(JI\times K)} = (\mathbf{A} \odot \mathbf{B})\,\mathbf{C}^T$$

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

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

If any of the three Khatri-Rao products fails to be full column rank, then adding a null vector to any of the rows of the corresponding third matrix does not affect the data. Necessity of $\min(k_{\mathbf{A}}, k_{\mathbf{B}}, k_{\mathbf{C}}) \geq 2$ can be proven by explicit rotation of the corresponding factors.

Sufficiency will be proven by contradiction. Without loss of generality, we may assume that $\mathbf{C}$ is square nonsingular, and $r_{\mathbf{A}\odot\mathbf{B}} = F$

We will use the *elementary* version of the *Permutation Lemma*: For square nonsingular $\mathbf{C}$, $\bar{\mathbf{C}}$, if

$$w(\mathbf{v}^T\mathbf{C}) = 1, \ \forall \mathbf{v} \ \mid \ w(\mathbf{v}^T\bar{\mathbf{C}}) = 1$$

then $\bar{\mathbf{C}} = \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}$

Suppose $\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) = (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$; then

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

Since $r_{\mathbf{A}\odot\mathbf{B}} = r_{\mathbf{C}} = F$, it follows that $r_{\bar{\mathbf{A}}\odot\bar{\mathbf{B}}} = r_{\bar{\mathbf{C}}} = F$

Taking linear combinations of the slabs along the third mode, we obtain

$$\sum_{k=1}^{K} v_k \mathbf{X}_k = \mathbf{A}diag(\mathbf{v}^T\mathbf{C})\mathbf{B}^T = \bar{\mathbf{A}}diag(\mathbf{v}^T\bar{\mathbf{C}})\bar{\mathbf{B}}^T,$$

for all $\mathbf{v} := [v_1, \cdots, v_F]^T \in \mathbb{C}^F$.

The rank of a matrix product is always less than or equal to the rank of any factor, and thus

$$w(\mathbf{v}^T\bar{\mathbf{C}}) = r_{diag(\mathbf{v}^T\bar{\mathbf{C}})} \geq r_{\bar{\mathbf{A}}diag(\mathbf{v}^T\bar{\mathbf{C}})\bar{\mathbf{B}}^T} = r_{\mathbf{A}diag(\mathbf{v}^T\mathbf{C})\mathbf{B}^T}$$

Assume $w(\mathbf{v}^T\bar{\mathbf{C}}) = 1$; then $r_{\mathbf{A}diag(\mathbf{v}^T\mathbf{C})\mathbf{B}^T} \leq 1$, and we wish to show that $w(\mathbf{v}^T\mathbf{C}) = 1$.

Let us use the shorthand $w := w(\mathbf{v}^T\mathbf{C})$. Using Sylvester's inequality and the definition of k-rank:

$$r_{\mathbf{A}diag(\mathbf{v}^T\mathbf{C})\mathbf{B}^T} \geq \min(k_{\mathbf{A}}, w) + \min(k_{\mathbf{B}}, w) - w.$$

Hence

$$\min(k_{\mathbf{A}}, w) + \min(k_{\mathbf{B}}, w) - w \leq 1$$

Consider cases:

1. Case of $w \leq \min(k_{\mathbf{A}}, k_{\mathbf{B}})$: then $w \leq 1$, and hence $w = 1$, because $\mathbf{C}$ is nonsingular and $\mathbf{v}$ is nonzero;
2. Case of $\min(k_{\mathbf{A}}, k_{\mathbf{B}}) \leq w \leq \max(k_{\mathbf{A}}, k_{\mathbf{B}})$: then $\min(k_{\mathbf{A}}, k_{\mathbf{B}}) \leq 1$, which contradicts our hypothesis; and
3. Case of $w \geq \max(k_{\mathbf{A}}, k_{\mathbf{B}})$: then $w \geq k_{\mathbf{A}} + k_{\mathbf{B}} - 1$: yet another contradiction under our hypothesis, as it requires that $w \geq F + 1$ (max possible $w = w(\mathbf{v}^T\mathbf{C})$ is F)

We conclude that $w(\mathbf{v}^T\bar{\mathbf{C}}) = 1$ implies $w(\mathbf{v}^T\mathbf{C}) = 1$. Thus $\bar{\mathbf{C}} = \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}$, and the rest follows easily; we're home.

An important question is whether or not Kruskal's sufficient condition is also necessary for uniqueness. A long-held affirmative conjecture by Kruskal [23] was recently refuted by ten Berge and Sidiropoulos [48], who showed that his condition is indeed necessary for uniqueness for $F \in \{2, 3\}$, but not for $F > 3$. Jiang & Sidiropoulos [18] subsequently provided further insights on this issue, and have shown that, for $r_{\mathbf{C}} = F$, the following condition is *necessary and sufficient* for CP uniqueness:

> No linear combination of two or more columns of $\mathbf{A} \odot \mathbf{B}$ can be written as a tensor (Kronecker, Khatri-Rao) product of two vectors

3.1. Significance

We have seen that CP decomposition is essentially unique (i.e., except for the inherent permutation and scaling ambiguity), provided rank is "low enough". This often works for rank much greater than 1, in contrast to low-rank decomposition of matrices, which is only unique, up to scaling, when the rank is *equal* to 1. This sounds fascinating, but what is its significance in terms of signal processing applications?

This can be appreciated by going back to some of our application examples. In particular:

- In the CDMA application, fully blind multiuser separation is possible from the multiple-sensor mixture data [42];
- In the MI-SAP application, blind identification of all source signals and associated parameters is possible [40];
- In the OFDM application, it is possible to recover the transmitted symbols, up to a common scale factor, despite the presence of the unknown frequency offset, which can also be estimated along with other pertinent signal parameters [19].
- More generally, if the number of users (sources, transmitters) is not too big, completely blind identification is possible, and the resulting identifiability conditions are the best available to date for the respective problems.

4. ALGORITHMS

Many algorithms have been developed for fitting the CP model. The SVD/EVD approach used in the first constructive proof of uniqueness in this paper, is perhaps the simplest way to "fit" certain restricted CP models. Being similar to the well-known ESPRIT algorithm in array processing, this algebraic procedure works reasonably well in the high-SNR regime, but can be quite suboptimal even at moderate SNR.

The workhorse CP fitting algorithm is *Trilinear Alternating Least Squares* (TALS), and it goes back to Harshman [13, 14, 15]. TALS is a Least-Squares (LS)-driven iterative algorithm, that is usually initialized using the SVD/EVD solution, when applicable, and/or multiple random "cold starts".

TALS is based on the matrix view of the CP model:

$$\mathbf{X}^{(KJ\times I)} = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T$$

The basic idea behind TALS is this: Given interim estimates of $\mathbf{B}$, $\mathbf{C}$, solve for the conditional Least-Squares (LS) update of $\mathbf{A}$:

$$\mathbf{A}_{CLS} = \left((\mathbf{B} \odot \mathbf{C})^{\dagger}\mathbf{X}^{(KJ\times I)}\right)^T$$

LS updates of $\mathbf{B}$, $\mathbf{C}$ can be easily derived based on symmetry. Then repeat in a circular fashion until convergence in fit.

Like any alternating optimization algorithm, TALS enjoys monotone convergence of the fit, but not necessarily to the global minimum - although, due to the highly structured nature of the CP model, convergence to the global minimum is often attained (especially when a few random cold starts are employed).

Cramer-Rao bounds for the CP model have been developed in [26, 17, 50]. These include:

- The real i.i.d. Gaussian, 3-way, case [26];
- The complex circularly symmetric i.i.d. Gaussian, 3-way & 4-way case [26];
- Asymptotic CRB when one mode length goes to infinity [17];
- Laplacian and Cauchy cases [50].

It has been shown [26], by extensive simulations, that TALS generally stays close to the pertinent Cramer-Rao bound, provided that the problem is sufficiently over-determined (that is, $F << IJK$). To understand this, note that, with 3-way data it is easy to get to the large-samples regime: e.g., $10 \times 10 \times 10 = 1000$

Performance is worse (and further from the CRB) when operating close to the identifiability boundary; but TALS *works* under model identifiability conditions only, which means that at high SNR the parameter estimates are still accurate

The main shortcoming of TALS is its high computational cost, and occasionally slow convergence. For difficult data sets, so-called *swamps* may emerge: progress towards convergence becomes extremely slow. A cold re-start is the usual way around this.

There are many alternative algorithms for CP fitting, e.g., performing rank-1 three-way array updates instead of matrix updates, using Gauss-Newton iterations, etc. While these alternatives may indeed work better than TALS in certain cases, the overall perception in the multi-way community is that TALS is the best jack-of-all-trades across the spectrum of CP applications considered to date.

Robust regression algorithms have also been developed for situations wherein one has to deal with heavy-tailed measurement noise and outliers. Recent work in this direction [50] has focused on the Least Absolute Error (LAE) criterion, which is optimal (ML) for Laplacian noise, and also robust across α-stable noise processes. The corresponding robust regression algorithms are similar to TALS, except that each conditional matrix update is now equivalent to a Linear Programming (LP) problem. This gives rise to an alternating LP procedure [50]; alternatively, very simple element-wise updating using *weighted median filtering* can do the job [50], at a much reduced complexity and often negligible performance loss. These robust versions of TALS also work well under Gaussian noise.

TALS assumes uncorrelated errors; significant noise color throws it off track. While it is easy to handle noise color *along one mode only*, via whitening, things become much more complicated when there is 2-way or 3-way noise correlation. In this case, one can use the vectorized model

$$\mathbf{x}^{(KJI)} = vec\left(\mathbf{X}^{(KJ\times I)}\right) = (\mathbf{A} \odot \mathbf{B} \odot \mathbf{C})\,\mathbf{1}_{F\times 1}$$

and the principle of *Majorization* to come up with a suitable ML fitting algorithm [6]. Majorization can be implemented as a simple outer loop over TALS, and is thus easy to program.

5. LEARN MORE

In addition to the selected bibliography in the end of this paper, the following web pointers provide a wealth of information related to the CP model and multi-way analysis in general. These links include free on-line multi-way tutorials (including the Tucker class of models, the other main-stay of multi-way analysis, which unfortunately we could not touch upon), software, paper reprints, and other current information.

Richard Harshman's homepage (http://publish.uwo.ca/ harshman/) contains many hard-to-find reprints of early seminal pa-

pers. The web-site of Peter Kroonenburg's *3-Mode Company* contains many useful links, and perhaps the most extensive multi-way bibliography that is available as of this writing. Rasmus Bro's page (http://www.models.kvl.dk/users/rasmus/) contains links to free interactive multi-way tutorials and software, in addition to a free monograph on multi-way methods. My own page (still at www.ece.umn.edu/users/nikos/) contains software and several reprints. The main scientific forum for cross-disciplinary basic research in this area is the TRICAP (Three-Way Analysis in Chemistry and Psychology) Workshop, which is held every three years. It has been held in Faaborg, Denmark in 2000; Lexington, KY, U.S.A., in 2003; and will be held in Crete - Greece in 2006. Search for 'TRICAP' on the web for more information.

6. WHAT LIES AHEAD

The key *take home point* is that $(N > 3)$-way arrays *are* different, in many ways. Perhaps the most important is that low-rank models are unique, which has many applications in signal processing, and beyond.

What are the major challenges for future research in this area? In my opinion, the major challenge with respect to uniqueness research is to find *easy to check* necessary and sufficient conditions, even if these hold almost-surely. Higher-way models are another challenge. For $N > 3$, the existing sufficient conditions are much further from necessity than those for $N = 3$. The investigation of uniqueness under application-specific constraints (e.g., Toeplitz) is also of interest. Symmetric and super-symmetric CP models have important applications in joint diagonalization and higher-order statistics; however, symmetry has been largely left unexplored in so far as strengthening uniqueness conditions is concerned.

In terms of algorithms, faster fitting methods that only incur a small performance loss relative to TALS are sorely needed, in order to bring on-line implementation closer to reality. The incorporation of application-specific constraints could help in this direction.

In the end, the sustained growth of a scientific area rests largely on finding new motivating applications, which attract talented young researchers into it. I hope that you will be one of those.

7. ACKNOWLEDGMENTS

I owe a great deal to Rasmus Bro, who has been my mentor in three-way analysis, and Jos M.F. ten Berge, who has instilled in me the importance of considering special cases *prior* to seeking general results. I would also like to thank my former Ph.D. students, Tao Jiang and Xiangqian Liu, who have contributed to my understanding of CP.

8. REFERENCES

[1] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory," *IEEE Trans. on Circuits and Systems*, 25(9):772–781, Sept. 1978.

[2] R. Bro, "PARAFAC: Tutorial and Applications", *Chemometrics and Intelligent Laboratory Systems*, 38:149–171, 1997.

[3] Bro, R. and Sidiropoulos, N. D., "Least squares algorithms under unimodality and non-negativity constraints, *Journal of Chemometrics*, 12:223-247, 1998.

[4] Bro, R. and Smilde, A. K. and De Jong, S., "On the difference between low-rank and subspace approximation: improved model for multi-linear PLS Regression", *Chemometrics and Intelligent Laboratory Systems*, 58:3-13, 2001.

[5] Bro, R. and Workman Jr, J. J. and Mobley, P. R. and Kowalski, B., "Review of chemometrics applied to spectroscopy: 1985-95, Part 3 - Multi-way analysis", *Applied Spectroscopy Reviews*, 32:237-261, 1997.

[6] R. Bro, N.D. Sidiropoulos, and A.K. Smilde, "Maximum Likelihood Fitting Using Ordinary Least Squares Algorithms", *J. of Chemometrics*, 16:387-400, Sep. 2002.

[7] Burgisser, P., Clausen, M., and Shokrollahi, M.A., *Algebraic complexity theory (Grundlehren der mathematischen Wissenschaften 315)*, Berlin, Springer, 1997.

[8] J. D. Carroll, J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition", *Psychometrika*, 35(3):283–319, 1970.

[9] R.B. Cattell, "Parallel Proportional Profiles and Other Principles for Determining the Choice of Factors by Rotation", *Psychometrika*, 9:267–283, 1944.

[10] De Lathauwer L., De Moor B., Vandewalle J., "A multilinear singular value decomposition", *SIAM J. Matrix Anal. Appl.*, 21(4):1253-1278, 2000.

[11] De Lathauwer L., De Moor B., Vandewalle J., "On the Best rank-1 and Rank-$(R_1, R_2, ..., R_N)$ Approximation and Applications of Higher-Order Tensors", *SIAM J. Matrix Anal. Appl.*, 21(4):1324-1342, 2000.

[12] De Lathauwer L., De Moor B., Vandewalle J., "Independent component analysis and (simultaneous) third-order tensor diagonalization", *IEEE Transactions on Signal Processing*, 49(10):2262-2271, 2001.

[13] R. Harshman, "Foundations of PARAFAC procedure: Models and Conditions for an 'Explanatory' Multi-Mode Factor Analysis", *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[14] R.A. Harshman, "Determination and Proof of Minimum Uniqueness Conditions for PARAFAC1", *UCLA Working Papers in Phonetics*, 22:111-117, 1972.

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

[16] T. Jiang and N. D. Sidropoulos "Blind Identification of Out-of-cell Users in DS-CDMA: An Algebraic Approach," *ICASSP 2002*, Orlando, FL, May. 13 - 17, 2002.

[17] T. Jiang, and N.D. Sidiropoulos, "Blind Identification of Out of Cell Users in DS-CDMA", *EURASIP Journal on Applied Signal Processing (JASP)*, special issue on Advances in Smart Antennas, to appear.

[18] T. Jiang, and N.D. Sidiropoulos, "Kruskal's Permutation Lemma and the Identification of CANDECOMP/PARAFAC and Bilinear Models with Constant Modulus Constraints", *IEEE Trans. on Signal Processing*, to appear.

[19] T. Jiang and N. D. Sidropoulos "A Direct Semi-Blind Receiver for SIMO and MIMO OFDM Systems Subject to Frequency Offset," in *Proc. SPAWC 2003*, Rome, Italy, Jun. 15 - 18, 2003.

[20] T. Jiang, N.D. Sidiropoulos, and J. M.F. ten Berge, "Almost Sure Identifiability of Multidimensional Harmonic Retrieval", *IEEE Trans. on Signal Processing*, 49(9):1849-1859, Sep. 2001.

[21] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions", *Sankhya, Se. A*, 30:167-180, 1968.

[22] J.B. Kruskal, "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics", *Linear Algebra and Its Applications*, 18:95–138, 1977.

[23] J.B. Kruskal, "Rank Decomposition, and Uniqueness for 3-way and N-way Arrays", *Multiway Data Analysis*, R. Coppi and S. Bolasco (*Eds.*), pp. 8–18, Elsevier North-Holland, Amsterdam, 1989.

[24] J.B. Kruskal, and M. Wish, *Multidimensional Scaling*, Sage University Papers, Quantitative Applications in the Social Sciences, Sage Publications, Inc., Newbury Park, CA, 1978.

[25] T. Li, N.D. Sidiropoulos, and G. Giannakis, "PARAFAC STAP for the UESA Radar", in *Proc. ASAP2000*, Mar. 14-15, 2000, MIT Lincoln Laboratory, Lexington, Mass.

[26] X. Liu and N. D. Sidiropoulos, "Cramer-Rao Lower Bounds for Low-rank Decomposition of Multidimensional Arrays," *IEEE Trans. on Signal Processing*, 49:2074–2086, 2001.

[27] X. Liu, and N.D. Sidiropoulos, "Almost Sure Identifiability of Constant Modulus Multidimensional Harmonic Retrieval", *IEEE Trans. on Signal Processing*, 50(9):2366-2368, Sep. 2002.

[28] X. Liu, and N.D. Sidiropoulos, "PARAFAC Techniques for High Resolution Array Processing", in *High Resolution Signal Processing*, Y. Hua, and A. Gershman, Eds., Marcel-Dekker, Oct. 2003.

[29] X. Liu, N. D. Sidiropoulos, and A. Swami, "High Resolution Localization and Tracking of Multiple Frequency Hopped Signals," *IEEE Trans. on Signal Processing*, 50:891–901, Apr. 2002.

[30] Rocci, R. and Ten Berge, J. M. F., "A simplification of a result by Zellini on the maximal rank of symmetric three-way arrays", *Psychometrika*, 59:377-380, 1994.

[31] Y. Rong, S.A. Vorobyov, A.B. Gershman, and N.D. Sidiropoulos, "Blind Spatial Signature Estimation via Time-Varying User Power Loading and Parallel Factor Analysis", submitted to *IEEE Trans. on Signal Processing*, July 2003.

[32] E. Sanchez, and B. R. Kowalski, "Generalized rank annihilation factor analysis," *Analytical Chemistry*, vol. 58, pp. 496–499, 1986.

[33] P. H. Schonemann, "An Algebraic Solution for a Class of Subjective Metrics Models", *Psychometrika*, 37:441, 1972.

[34] N.D. Sidiropoulos, "Generalizing Carathéodory's Uniqueness of Harmonic Parameterization to N Dimensions", *IEEE Trans. Information Theory*, 47(4):1687-1690, May 2001.

[35] N. D. Sidiropoulos and R. Bro, "On the Uniqueness of Multilinear Decomposition of N-way Arrays," *J. Chemometrics*, special cross-disciplinary issue on multi-way analysis, 14(3):229-239, May 2000.

[36] N.D. Sidiropoulos, and R. Bro, "User Separation in DS-CDMA Systems with Unknown Long PN Spreading Codes", in *Second IEEE-SPS Workshop on Signal Processing Advances in Wireless Communications (SPAWC99)*, May 9-12, 1999, Annapolis, MD.

[37] N.D. Sidiropoulos, and R. Bro, "On Communication Diversity for Blind Identifiability and the Uniqueness of Low-Rank Decomposition of N-way Arrays," in *Proc ICASSP*, Istanbul, Turkey, Jun. 5–9, 2000.

[38] N.D. Sidiropoulos, and R. Bro, "PARAFAC Techniques for Signal Separation," Chap. 4 in *Signal Processing Advances in Wireless Communications*, Vol. 2, P. Stoica, G. Giannakis, Y. Hua, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2000.

[39] N.D. Sidiropoulos, and R. Budampati, "Khatri-Rao Space-Time Codes", *IEEE Trans. on Signal Processing*, special issue on signal processing techniques for space-time coded transmissions, 50(10):2396-2407, Oct. 2002.

[40] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel Factor Analysis in Sensor Array Processing," *IEEE Trans. Signal Processing*, 48(8):2377–2388, Aug. 2000.

[41] N.D. Sidiropoulos, and G. Z. Dimic, "Blind Multiuser Detection in W-CDMA Systems with Large Delay Spread", *IEEE Signal Proc. Letters*, 8(9):87-89, Mar. 2001.

[42] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC Receivers for DS-CDMA Systems," *IEEE Trans. Signal Processing*, 48(3):810–823, Mar. 2000.

[43] N.D. Sidiropoulos, and X. Liu, "Identifiability Results for Blind Beamforming in Incoherent Multipath with Small Delay Spread", *IEEE Trans. Signal Processing*, 49(1):228-236, Jan. 2001.

[44] Ten Berge, J.M.F. (2000), "The typical rank of tall three-way arrays", *Psychometrika*, 65:525-532.

[45] Ten Berge, J. M. F. and Kiers, H. A. L., "Some clarifications of the CANDECOMP algorithm applied to INDSCAL", *Psychometrika*, 56:317-326, 1991.

[46] Ten Berge, J. M. F. and Kiers, H. A. L., "Some uniqueness results for PARAFAC2", *Psychometrika*, 61:123-132, 1996.

[47] Ten Berge, J. M. F. and Kiers, H. A. L., "Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times q \times 2$ arrays", *Linear Algebra and its Applications*, 294:169-179, 1999.

[48] J. M. F. ten Berge and N. D. Sidiropoulos, "On Uniqueness in CANDECOMP/PARAFAC," *Psychometrika*, 67(3):399-409, Sep. 2002.

[49] J. M.F. ten Berge, N.D. Sidiropoulos, and R. Rocci, "Typical Rank and INDSCAL Dimensionality for Symmetric Three-way Arrays of Order Ix2x2 or Ix3x3", *Linear Algebra and its Applications*, to appear.

[50] S.A. Vorobyov, Y. Rong, N.D. Sidiropoulos, and A.B. Gershman, "Robust Iterative Fitting of Multilinear Models", submitted to *IEEE Trans. on Signal Processing*, July 2003.