Sparse Representations and Low-Rank Tensor Approximation

Pierre Comon, Lek-Heng Lim

LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

SPARSE REPRESENTATIONS AND LOW-RANK TENSOR APPROXIMATION

Pierre Comon, Lek-Heng Lim

Equipe SIGNAL

Rapport de recherche
ISRN I3S/RR–2011-02–FR

Février 2011

Laboratoire d'Informatique de Signaux et Systèmes de Sophia Antipolis - UNSA-CNRS
2000, rte.des Lucioles – Les Algorithmes – Bât Euclide B – B.P. 121 – 06903 Sophia-Antipolis Cedex – France
Tél.: 33 (0)4 92 94 27 01 – Fax: 33 (0)4 92 94 28 98 – www.i3s.unice.fr
UMR6070

RÉSUMÉ :

MOTS CLÉS :
Identification aveugle ; mélanges linéaires sous-déterminés ; séparation aveugle de sources ; décompositions polyadiques tensorielles ; tenseurs, rang tensoriel ; localisation ; meilleure approximation de rang faible ; représentation parcimonieuse ; spark ; échantillonnage compressé ; rang de Kruskal ; cohérence ; antennes multiples ; capteurs multiples

---

ABSTRACT:
Approximating a tensor by another of lower rank is in general an ill posed problem. Yet, this kind of approximation is mandatory in the presence of measurement errors or noise. We show how tools recently developed in compressed sensing can be used to solve this problem. More precisely, a minimal angle between the columns of loading matrices allows to restore both existence and uniqueness of the best low rank approximation. We then show how these results can be applied to perform jointly localization and extraction of multiple sources from the measurement of a noisy mixture recorded on multiple sensors, in an entirely deterministic manner. The main interest in deterministic approaches is that they can be followed in the presence of strong channel nonstationarities.

KEY WORDS :
Blind identification; under-determined linear mixtures; blind source separation; polyadic tensor decompositions; tensors; tensor rank; localization; best low rank approximations; sparse representations; spark; compressed sensing; Kruskal's rank; coherence; multiarrays; multisensors

Sparse Representations and Low-Rank Tensor Approximation

Pierre Comon*, *Fellow, IEEE*, and Lek-Heng Lim‡

Abstract

Approximating a tensor by another of lower rank is in general an ill posed problem. Yet, this kind of approximation is mandatory in the presence of measurement errors or noise. We show how tools recently developed in compressed sensing can be used to solve this problem. More precisely, a minimal angle between the columns of loading matrices allows to restore both existence and uniqueness of the best low rank approximation. We then show how these results can be applied to perform jointly localization and extraction of multiple sources from the measurement of a noisy mixture recorded on multiple sensors, in an entirely deterministic manner. The main interest in deterministic approaches is that they can be followed in the presence of strong channel nonstationarities.

Index Terms

Blind identification; under-determined linear mixtures; blind source separation; polyadic tensor decompositions; tensors; tensor rank; localization; best rank-r approximations; sparse representations; spark; compressed sensing; Kruskal's rank; coherence; multiarrays; multisensors

I. Introduction

Tensor decomposition and approximation models arise naturally in multiarray multisensor signal processing, as already demonstrated in [1], [2], [3], [4], [5] when high-order statistics are used, and in [6] when sensor arrays enjoy particular geometrical properties. However, the fact that approximating a tensor by another of lower rank is generally an ill-posed problem has not

* Pierre Comon is with Lab. Informatique Signaux et Systèmes de Sophia-Antipolis (I3S), UMR6070 CNRS-UNS, 2000 route des Lucioles, BP.121, F06903 Sophia Antipolis cedex, France, and with INRIA, Galaad, 2004 route des Lucioles, BP.93, F06902 Sophia Antipolis cedex, France. ‡ Lek-Heng Lim is with the Department of Statistics, University of Chicago.

been taken into account in the latter works. This explains why numerical algorithms sometimes do not converge to the expected solution, and that they often converge quite slowly.

We explain in Section II why the problem is ill-posed, and what remedies have already been proposed to face it. Then we see in Section III how contributions borrowed from compressed sensing can be used to address the problem in a more convenient manner. The term *compressed sensing* should be understood in a broad sense, encompassing not only the ideas covered in [7], [8], [9], [10], [11], [12] but also in [13], [14], [15], [16]. Then, the usefulness of the proposed approach is demonstrated in Section IV, where several applications are pointed out, with an emphasis on the problem of joint localization and estimation of radiating sources with short data lengths, which can be solved deterministically.

II. Problem position and first remedies

A tensor of order D is an object defined on a product of D linear spaces, $\mathcal{S}_d$, $1 \le d \le D$. Such a tensor may represent a map from $\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_\gamma$ onto $\mathcal{S}_{\gamma+1} \otimes \cdots \otimes \mathcal{S}_D$, for some γ, $0 < \gamma < D$. If $\gamma = D$, one has a D-linear form defined on $\otimes_{d=1}^{D} \mathcal{S}_d$. Once the basis of each linear space $\mathcal{S}_i$ is fixed, such a tensor can be represented by a D-way array of coordinates, $\mathbf{T} = [\![T_{ij..k}]\!]$. The background field is assumed to be $\mathbb{R}$ or $\mathbb{C}$.

When a change of basis is operated in each linear space $\mathcal{S}_d$, defined by a matrix $\mathbf{A}^{(d)}$, the array of coordinates $\mathbf{T}$ must be modified accordingly into an array $\mathbf{T}'$. We shall be concerned by the so-called *contravariant* tensors, which enjoy the multilinearity property below:

$$T'_{ij..k} = \sum_i A^{(1)}_{ip} \sum_j A^{(2)}_{jq} \cdots \sum_k A^{(D)}_{kr}\, T_{pq..r}$$

which we shall denote compactly as

$$\mathbf{T}' = (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(D)}) \cdot \mathbf{T} \tag{1}$$

A. Canonical Polyadic decomposition

A tensor $\mathbf{E}$ of order D is said to be *decomposable* if it can be written as the tensor product of D vectors:

$$\mathbf{E} = \mathbf{u}^{(1)} \otimes \mathbf{u}^{(2)} \otimes \ldots \otimes \mathbf{u}^{(D)}$$

In other words, its array of coordinates can be written as $E_{ij..k} = u_i^{(1)}\, u_j^{(2)} .. \, u_k^{(1)}$. Any tensor $\mathbf{T}$ can always be decomposed in a sum of decomposable tensors [17]:

$$\begin{aligned} \mathbf{T} &= \sum_{r=1}^{R} \lambda_r \, \mathbf{E}(r), \\ \mathbf{E}(r) &= \mathbf{u}_r^{(1)} \otimes \mathbf{u}_r^{(2)} \otimes \ldots \otimes \mathbf{u}_r^{(D)} \end{aligned} \tag{2}$$

In addition, λ_r can be imposed to be real nonnegative, and vectors $\mathbf{u}_r^{(d)}$ can be imposed to be of unit norm, for some suitably chosen norm.

Definition 1. *The minimal number of decomposable terms necessary to meet the exact fit in equation (2) is referred to as the* rank *of tensor* $\mathbf{T}$*:*

$$\mathrm{rank}\{\mathbf{T}\} = min \left\{ R \,\middle|\, \mathbf{T} = \sum_{r=1}^{R} \lambda_r \, \mathbf{E}(r) \right\} \tag{3}$$

In particular, decomposable tensors have a rank equal to one. When minimal, decomposition (2) reveals the rank and is often called the *Canonical Polyadic* (CP) decomposition of $\mathbf{T}$. Other terminologies have been used in various communities, including *CanDecomp* [18] and *Parafac* [19] in Psychometrics. The Linear Algebra community has taken the habit to use the acronym CP, standing for CanDecomp/Parafac, which fortunately coincides with the former.

Note that the CP decomposition can be written in compact form as

$$\mathbf{T} = (\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(D)}) \cdot \boldsymbol{\Lambda} \tag{4}$$

if $\boldsymbol{\Lambda}$ denotes the D-way diagonal array, whose sole nonzero entries are $\Lambda_{rr..r} = \lambda_r$, and matrices $\mathbf{U}^{(d)}$ are each built with the $R = \mathrm{rank}\{\mathbf{T}\}$ column vectors $\mathbf{u}_r^{(d)}$, $1 \le d \le D$, $1 \le r \le R$.

At this stage, it is convenient to make a comparison with matrix decompositions, which are better known. Let a matrix $\mathbf{M}$ of rank $R > 1$. Then it can be decomposed in infinitely many ways into a sum of rank-1 terms as

$$\mathbf{M} = \sum_{r=1}^{R} \lambda_r \, \mathbf{u}_r \, \mathbf{v}_r^{\mathsf{T}}$$

where $\mathbf{U} = \{\mathbf{u}_r\}$ and $\mathbf{V} = \{\mathbf{v}_r\}$ are collections of R unit-norm vectors. The Singular Value decomposition (SVD) of $\mathbf{M}$ yields one such decomposition, where vectors of $\mathbf{U}$ (resp. $\mathbf{V}$) are orthogonal to each other.

In (2-4), no orthogonality constraint is imposed. However, a sufficient condition for uniqueness has still been derived when $D > 2$ as now pointed out. We first need a definition

Definition 2. *The Kruskal's rank of a matrix* $\mathbf{M}$*, or* $\mathrm{krank}\{\mathbf{M}\}$ *in brief, is the maximal number* κ *such that any subset of* κ *columns of* $\mathbf{M}$ *are linearly independent.*

From this definition, originally introduced in [20], it is clear that $\mathrm{rank}\{\mathbf{M}\} \geq \mathrm{krank}\{\mathbf{M}\}$. Also note that the notion of *spark* introduced in compressed sensing [11], [12] is related to Kruskal's rank, since spark$\{\mathbf{M}\} = \mathrm{krank}\{\mathbf{M}\} + 1$. Let's now turn to the following result, generally referred to as *Kruskal's lemma* [20], [21], [22]:

Lemma 3. *Let* $\mathbf{T}$ *be a tensor of order* D*. Then its CP decomposition (4) is unique if*

$$2\ \mathrm{rank}\{\mathbf{T}\} + 2 \leq \sum_{d=1}^{D} \mathrm{krank}\{\mathbf{U}^{(d)}\}. \tag{5}$$

This condition has been proved to be sufficient but not proved to be necessary. We insist that uniqueness is here to be understood *up to a scale factor*. More precisely, there still remains a whole equivalence class of CP decompositions in the sense that one can replace each $\mathbf{U}^{(d)}$ by $\mathbf{U}^{(d)}\boldsymbol{\Delta}^{(d)}$, where matrices $\boldsymbol{\Delta}^{(d)}$ are diagonal invertible, and satisfy the constraint $\prod_{d=1}^{D} \boldsymbol{\Delta}^{(d)} = \mathbf{I}_R$, the $R \times R$ identity matrix.

B. Ill-posedness of the best low-rank approximation

The best rank-R approximate is defined by the minimum of the objective

$$\Upsilon(\mathbf{U}_r^{(1)}, .., \mathbf{U}_r^{(D)}, \boldsymbol{\Lambda}) = \left\| \mathbf{T} - (\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(D)}) \cdot \boldsymbol{\Lambda} \right\|_F^2 \tag{6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm defined by $\|\mathbf{T}\|_F^2 = \sum_{ij..k} |T_{ij..k}|^2$, and matrices $\mathbf{U}_r^{(d)}$ have R columns, $R < \mathrm{rank}\{\mathbf{T}\}$. It turns out that this best approximate may not exist for tensors of order $D > 2$. In fact, the set of tensors of rank at most β is not closed if $\beta > 1$, except when the latter set is the whole space, *i.e.* when β is maximal (but there is then no approximation).

This lack of closeness is now well known, and examples have been provided in the literature [23], [24], which suffice to prove it.

Example 4. *Let's for instance consider two non collinear vectors,* $\mathbf{a}$ *and* $\mathbf{b}$*, and define the sequence of rank-2 tensors:*

$$\mathbf{T}(n) = n\left[(\mathbf{a} + \frac{1}{n}\mathbf{b})^{\otimes 4} - \mathbf{a}^{\otimes 4}\right]$$

As n *tends to infinity, this sequence converges towards*

$$\mathbf{T}_\infty = \mathbf{a}^{\otimes 3} \otimes \mathbf{b} + \mathbf{a}^{\otimes 2} \otimes \mathbf{b} \otimes \mathbf{a} + \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{a}^{\otimes 2} + \mathbf{b} \otimes \mathbf{a}^{\otimes 3}$$

which may be shown to be of rank 4. This demonstrates that the set of tensors of rank at most 2 is not closed.

The limit of tensors of rank R is said to be of *border rank* R. In general, the actual tensor rank is larger than the border rank, as in the above example. However, there are cases where they always coincide. In particular, this is the case of real tensors with nonnegative entries, if they are decomposed into a sum of real nonnegative decomposable tensors [25]. But in the present framework, they differ, so that the lower rank approximation problem is ill-posed.

C. Searching a compact set via constrained optimization

The most natural way to face this problem is to change the set we are searching into a compact set. This can be done in several ways. In [26], loading marices $\mathbf{U}^{(d)}$ are imposed to be orthogonal. This solution is acceptable only in very restrictive conditions; in particular the rank of $\mathbf{T}$ must be smaller than its dimensions. In [27], it has been proposed to impose orthogonality between the decomposable tensors; this constraint is less restrictive, but quite difficult to impose and still too restrictive. The first available practical technique was that proposed in [28], consisting of minimizing the objective $\Upsilon(\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(D)}) + \beta \sum_d \|\mathbf{U}^{(d)}\|_F^2$, where β is an arbitrarily chosen regularization parameter. It can be seen that this is equivalent to constrain matrices $\mathbf{U}^{(d)}$ to lie on a sphere $\sum_{d=1}^{D} \|\mathbf{U}^{(d)}\|_F^2 = \rho$, ρ being determined a posteriori from β. The drawback of this efficient constraint is that ρ and β are arbitrary, and that they generally have no physical meaning.

We posed the problem slightly differently in (6), where each column vector is imposed to have a unit norm, which permits to define scale coefficients λ_r properly. In other words, we minimize the Lagrangian

$$\Upsilon(\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(D)}, \mathbf{\Lambda}) + \sum_{d=1}^{D} \sum_{r=1}^{R} \beta_{d,r} \, \|\mathbf{u}_r^{(d)}\|_2^2$$

where $\beta_{d,r}$ denote Lagrange multipliers, and $\|\cdot\|_2$ the L^2 norm. However, contrary to [28], the set we are searching is not compact anymore, because of the presence of unbounded variables λ_r (the entries of $\mathbf{\Lambda}$) in the objective Υ. The goal of the next section is to define physically meaningful constraints, which will ensure the existence of a unique minimum, even if $\mathbf{\Lambda}$ is unbounded.

III. Angular constraint

A. Existence

The goal is to prevent the phenomenon we observed in Example 4 to occur, by imposing natural and weak constraints; we do not want to reduce the search to a compact set. It is clear that the objective (6) is not coercive, which explains why the minimum may not exist. But with an additional condition on the *coherence*, we shall be able to prove existence thanks to coercivity.

Definition 5. *Let $\mathbb{H}$ be a Hilbert space provided with scalar product $\langle\cdot\rangle$, and let $\mathbf{V} = \{\mathbf{v}_1, .., \mathbf{v}_R\}$ be a finite collection of vectors of unit norm. The coherence of the collection $\mathbf{V}$ is defined as $\mu(\mathbf{V}) \stackrel{\text{def}}{=} \max_{p\neq q} |\langle \mathbf{v}_p, \mathbf{v}_q\rangle|$.*

This notion has received different names in the literature: muual incoherence of two dictionaries [11], mutual coherence of two dictionaries [8], the coherence of a subspace projection [14], etc. The version here follows that of [12]. We are interested in the case when $\mathbb{H}$ is finite dimensional, namely $\mathbb{C}^N$. Usually, dictionaries are finite or countable, but we have here a continuum of atoms. Clearly, $0 \leq \mu(\mathbf{V}) \leq 1$, and $\mu(\mathbf{V}) = 0$ iff $\mathbf{v}_1, \dots, \mathbf{v}_R$ are orthonormal, and $\mu(\mathbf{V}) = 1$ iff $\mathbf{V}$ contains at least a pair of collinear vectors.

The following shows that a solution to the bounded coherence best rank-R approxiamtion problem always exists:

Proposition 6. *Le $\mathbf{T}$ be a tensor of order D and dimensions N_d, $1 \leq d \leq D$, and define the sets of dictionaries of unit vectors of coherence not larger than μ_d:*

$$\mathcal{U}^{(d)} = \{\mathbf{U}^{(d)} \in \mathbb{C} \mid \mu(\mathbf{U}^{(d)}) \leq \mu_d\} \tag{7}$$

If $\prod_{d=1}^{D} \mu_d < \frac{1}{R}$, then

$$\eta = \inf\left\{ \|\mathbf{T} - (\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(D)}) \cdot \mathbf{\Lambda}\| \;\middle|\; \boldsymbol{\lambda} \in \mathbb{C}^R, \mathbf{U}^{(d)} \in \mathcal{U}^{(d)} \right\} \tag{8}$$

is attained, where $||\cdot||$ *denotes any norm on* $\otimes_{d=1}^{D}\mathbb{C}^{N_d}$*. Above, vector* $\boldsymbol{\lambda}$ *contains the entries* λ_d *of the diagonal tensor* $\boldsymbol{\Lambda}$*.*

Proof: Since all norms are equivalent on a finite dimensional space, we may assume the Frobenius norm $\|\cdot\|_F$. We have the following inequalities

$$\begin{aligned} \|(\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(D)})\cdot\boldsymbol{\Lambda}\| &= \sum_{p,q=1}^{R}\lambda_p\bar{\lambda}_q\prod_{d=1}^{D}\langle\mathbf{u}_p^{(d)},\mathbf{u}_q^{(d)}\rangle \\ \geq \sum_{p=1}^{R}\lambda_p\bar{\lambda}_p\prod_{d=1}^{D}||\mathbf{u}_p^{(d)}||^2 &- \sum_{p\neq q}^{R}\left|\lambda_p\bar{\lambda}_q\prod_{d=1}^{D}\langle\mathbf{u}_p^{(d)},\mathbf{u}_q^{(d)}\rangle\right| \\ \geq \sum_{p=1}^{R}|\lambda_p|^2 &- \prod_{d=1}^{D}\mu_d\sum_{p\neq q}|\lambda_p\bar{\lambda}_q| \\ \geq ||\boldsymbol{\lambda}||_2^2 &- \prod_{d=1}^{D}\mu_d||\boldsymbol{\lambda}||_1^2 \end{aligned}$$

Now, use the fact that $||\boldsymbol{\lambda}||_1^2 \leq \sqrt{R}\,||\boldsymbol{\lambda}||_2^2$ for any vector $\boldsymbol{\lambda}$ of size R, to get eventually

$$\|(\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(D)})\cdot\boldsymbol{\Lambda}\| \geq \left(1 - R\prod_{d=1}^{D}\mu_d\right)||\boldsymbol{\lambda}||_2^2 \tag{9}$$

Since by assumption $R\prod_{d=1}^{D}\mu_d < 1$, it is clear that the left hand side of (9) tends to infinity as $||\boldsymbol{\lambda}||_2 \to \infty$. And because $||\mathbf{T}||$ is fixed, $\|\mathbf{T} - (\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(D)})\cdot\boldsymbol{\Lambda}\|$ also tends to infinity. This proves coercivity, and hence the proposition. ■

B. Uniqueness

In order to prove uniqueness, we shall call for Kruskal's lemma. For that purpose, the following lemma is needed.

Lemma 7. *Let* $\mathbb{H}$ *be a Hilbert space and let* $\mathbf{V} = \{\mathbf{v}_1,..,\mathbf{v}_R\}$ *be a finite collection of vectors of unit norm. Then*

$$\text{krank}\{\mathbf{V}\} \geq \frac{1}{\mu(\mathbf{V})} \tag{10}$$

Proof: Let $s = \text{krank}\{\mathbf{V}\}+1$, the spark of $\mu(\mathbf{V})$. Then there exists a s-uple of distinct unit vectors in $\mathbf{V}$, $\{\mathbf{v}_1,\ldots,\mathbf{v}_s\}$ such that $\alpha_1\mathbf{v}_1+\cdots+\alpha_s\mathbf{v}_s = 0$ with $|\alpha_1| = \max\{|\alpha_1|,\ldots,|\alpha_s|\} > 0$. Taking inner product with $\mathbf{v}_1$ we get $\alpha_1 = -\alpha_2\langle\mathbf{v}_2,\mathbf{v}_1\rangle - \cdots - \alpha_s\langle\mathbf{v}_s,\mathbf{v}_1\rangle$ and so $|\alpha_1| \leq (|\alpha_2|+\cdots+|\alpha_s|)\mu(V)$. Dividing by $|\alpha_1|$ then yields $1 \leq (s-1)\mu(\mathbf{V})$. ■

Definition 8. *We shall say that the CP decomposition (4) is unique up to* unimodulus scaling *if each vector* $\mathbf{u}_r^{(d)}$ *can be multiplied by a scalar factor* $\alpha_r^{(d)}$ *of unit modulus, such that* $\prod_{d=1}^D \alpha_r^{(d)} = 1, \forall r, 1 \leq r \leq R$.

We now characterize the uniqueness of the CP decomposition in terms of coherence introduced in Definition 5.

Proposition 9. *Let* $\mathbf{T}$ *be a tensor of order* D *and a decomposition* $\mathbf{T} = \sum_{r=1}^R \lambda_r \mathbf{E}(r)$ *into* R *decomposable tensors* $\mathbf{E}(r) = \mathbf{u}_r^{(1)} \otimes \mathbf{u}_r^{(2)} \otimes \ldots \otimes \mathbf{u}_r^{(D)}$*, where* $\mathbf{u}_r^{(d)}$ *are of unit norm. Denote* $\mathbf{U}_r^{(d)}$ *the matrices with columns* $\mathbf{u}_r^{(d)}$*. If*

$$\frac{1}{2} \sum_{d=1}^{D} \frac{1}{\mu(\mathbf{U}_r^{(d)})} \geq R+1 \tag{11}$$

then $R = \operatorname{rank}\{\mathbf{T}\}$ *and the decomposition is unique up to unimodulus scaling.*

Proof: If inequality (11) is satisfied, then so is Kruskal's condition (5) thanks to Lemma 7. The results hence directly follows from Lemma 3. ■

Note that unlike the k-ranks in (5), the coherences in (11) are trivial to compute. In addition to uniqueness, an easy but important consequence of Proposition 9 is that it provides a readily checkable sufficient condition for tensor rank, which is NP-hard over any field [29], [30].

C. Existence and uniqueness

Now the following existence and uniqueness sufficient condition can be deduced from Propositions 6 and 9.

Corollary 10. *If* $D \leq 5$ *and if coherences* $\mu(\mathbf{U}^{(d)})$ *satisfy*

$$\left(\prod_{d=1}^{D} \mu(\mathbf{U}^{(d)})\right)^{1/D} \leq \frac{D}{2R+2} \tag{12}$$

then the bounded coherence best rank-R *approximation problem has a unique solution up to unimodulus scaling.*

Proof: The existence in the case $R = 1$ is ensured, because the set of tensors of rank 1 is closed (it is in fact a determinantal variety). Consider thus the case $R \geq 2$. Since the function $f(x) = \frac{1}{x} - (\frac{D}{2x+2})^D$ is strictly positive for $x \geq 2$ and $D \leq 5$, condition (12) implies that

$\prod_{d=1}^{D} \mu(\mathbf{U}^{(d)})$ is smaller than $1/R$, which permits to claim that the solution exists by calling for Proposition 6.

Next in order to prove uniqueness, we use the inequality between harmonic and geometric means: if (12) is verified, then we also necessarily have $D\,[\sum_{d=1}^{D} \mu(\mathbf{U}^{(d)})^{-1}]^{-1} \leq \frac{D}{2R+2}$. Hence $\sum_{d=1}^{D} \mu(\mathbf{U}^{(d)})^{-1} \geq 2R+2$ and we can apply Proposition 9. ■

IV. Applications

The goal of this section is two-fold. First we want to show the usefulness of the CP decomposition in real world problems, amd second we want to know the meaning of the coherence conditions in terms of physical quantities.

A. Joint channel and source estimation

Consider a narrow band transmission problem in the far field. We assume here that we are in the context of wireless telecommunications, but the same principle could apply in other fields. Let P signals impinge on an array, so that their mixture is recorded. It is wished to recover the original signals, and to estimate their directions of arrival and respective powers at the receiver. If the channel is specular, some of these signals can correspond to different propagation paths of the same radiating source, and are hence correlated. In other words, P does not denote the number of sources, but the total number of distinct paths viewed from the receiver.

In the present framework, we assume that channels can be time-varying, but that they can be assumed constant over a sufficiently short observation length. The goal is hence to be able to work with extremely short samples.

In order to face this challenge, we assume that the sensor array is structured, as in [6]. More precisely, the sensor array is composed of a *reference array* containing I sensors, whose location is defined by a vector $\mathbf{b}_i \in \mathbb{R}^3$, and $J-1$ other subarrays, deduced from the reference array by a translation in space defined by a vector $\boldsymbol{\Delta}_j \in \mathbb{R}^3$, $1 < j \leq J$. The reference subarray is numbered with $j = 1$ in the remainder.

Under these assumptions, the signal received at discrete time k on the ith sensor of the reference subarray can be written as:

$$s_{i,1}(k) = \sum_{p=1}^{P} \sigma_p(k)\, \exp(\psi_{i,p})$$

with $\psi_{i,p} = \jmath\frac{\omega}{C}\,(\mathbf{b}_i^{\mathsf{T}}\mathbf{d}_p)$ where the dotless $\jmath$ denotes $\sqrt{-1}$, vector $\mathbf{d}_p$ is unit norm and denotes the direction of arrival of the pth path. Next, on the jth subarray, $j > 1$, we have

$$s_{i,j}(k) = \sum_{p=1}^{P} \sigma_p(k)\,\exp(\psi_{i,j,p}) \tag{13}$$

with $\psi_{i,j,p} = \jmath\frac{\omega}{C}\,(\mathbf{b}_i^{\mathsf{T}}\mathbf{d}_p + \mathbf{\Delta}_j^{\mathsf{T}}\mathbf{d}_p)$. If we let $\mathbf{\Delta}_1$ be the null vector, then (13) also applies for the reference subarray. The interest of this structure is that variables i and j decouple in function $\exp(\psi_{i,j,p})$, yielding a relation resembling the CP decomposition:

$$s_{i,j}(k) = \sum_{p=1}^{P} \lambda_p\, U_{ip}^{(1)} U_{jp}^{(2)} U_{kp}^{(3)} \tag{14}$$

where $U_{ip}^{(1)} = \exp(\jmath\frac{\omega}{C}\,\mathbf{b}_i^{\mathsf{T}}\mathbf{d}_p)$, $U_{jp}^{(2)} = \exp(\jmath\frac{\omega}{C}\,\mathbf{\Delta}_j^{\mathsf{T}}\mathbf{d}_p)$ and $U_{kp}^{(3)} = \sigma_p(k)/||\boldsymbol{\sigma}_p||$, $\lambda_p = ||\boldsymbol{\sigma}_p||$.

Hence, by computing the CP decomposition of the $I \times J \times K$ tensor $\mathbf{S} = [\![s_{i,j}(k)]\!]$, it is possible to jointly estimate: (i) signal waveforms $\sigma_p(k)$, and (ii) the directions of arrival $\mathbf{d}_p$ of each propagation path if $\mathbf{b}_i$ or $\mathbf{\Delta}_j$ are known.

However, the observation model (13) is not realistic, and an additional error term should be added in order to stand for modeling inaccuracies and background noise. It is customary (and realistic thanks to the central limit theorem) to assume that this additive error has a continuous probability distribution, so that tensor $\mathbf{S}$ has a *generic rank*. Yet, the generic rank takes values at least as large as $\lceil IJK/(I+J+K-2)\rceil$, which is always larger than Kruskal's bound [24]. Therefore, we have to face the problem of approximating tensor $\mathbf{S}$ by another of rank P. And we have seen that the angular constraint imposed in Section III permits to deal with a well-posed problem. In order to see the physical meaning of this constraint, it is convenient to define first the tensor product between subarrays.

B. Towards the concept of tensor product between sensor subarrays

The sensor arrays we cope with are structured, in the sense that the whole array is generated by one subarray, defined by the collection of vector locations $\{\mathbf{b}_i \in \mathbb{R}^3,\, 1 \le i \le I\}$, and a collection of translations in space, $\{\mathbf{\Delta}_j \in \mathbb{R}^3,\, 1 \le j \le J\}$. If we define vectors

$$\begin{aligned}
\mathbf{u}_p^{(1)} &= [\![\exp(\jmath\frac{\omega}{C}\,\mathbf{b}_i^{\mathsf{T}}\mathbf{d}_p)]\!]_{i=1}^{I}/\sqrt{I} \\
\mathbf{u}_p^{(2)} &= [\![\exp(\jmath\frac{\omega}{C}\,\mathbf{\Delta}_j^{\mathsf{T}}\mathbf{d}_p)]\!]_{j=1}^{J}/\sqrt{J} \\
\mathbf{u}_p^{(3)} &= \boldsymbol{\sigma}_p/||\boldsymbol{\sigma}_p||
\end{aligned} \tag{15}$$

then this means that we may see all measurements as the superimposition of decomposable tensors:

$$\lambda_p \, \mathbf{u}_p^{(1)} \otimes \mathbf{u}_p^{(2)} \otimes \mathbf{u}_p^{(3)}$$

The geometry of the sensor array is contained in $\mathbf{u}_p^{(1)} \otimes \mathbf{u}_p^{(2)}$, whereas λ_p and $\mathbf{u}_p^{(3)}$ contain energy and time information on each path p, respectively. Note that the reference subarray and the set of translations play symmetric roles, in the sense that $\mathbf{u}_p^{(1)}$ and $\mathbf{u}_p^{(2)}$ could be interchanged without changing the whole array. This will become clear with a few examples.

When we are given a structured sensor array, there can be several ways of splitting it into a tensor product of two (or more) subarrays, as now shown by simple examples.

Example 11. *Define the matrix of sensor locations*

$$[\mathbf{b}_1,\ \mathbf{b}_2,\ \mathbf{b}_3] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

This subarray is depicted in Figure 1.b. By translating it according to the translation defined in Figure 1.c one obtains another subarray. The union of the two subarrays yields the array of Figure 1.a. The same array is obtained by interchanging the roles of the two subarrays, i.e. three subarrays of two sensors deduced from each other by two translations.

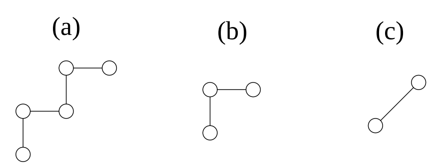
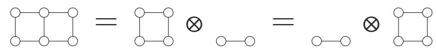


Fig. 1. Antenna array (a) is obtained as the tensor product between subarrays (b) and (c)

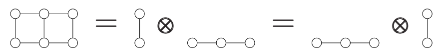
Example 12. *Define the array by*

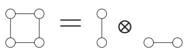
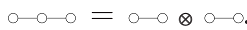
$$[\mathbf{b}_1,\ \mathbf{b}_2, \ldots \mathbf{b}_6] = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

This array, depicted in Figure 2.a, can be obtained either by the union of subarray of Figure 2.b and its translation defined by Figure 2.c, or by the array of Figure 2.c translated three times *according to Figure 2.b. We agree to express this relationship by the equation:*



Another decomposition may be obtained as



In fact,  *and* *. However, it is important to stress that the various decompositions of the whole array into tensor products of subarrays are not equivalent from the point of view of performance. In particular, the Kruskal's bound can be different, as will be pointed out next.*

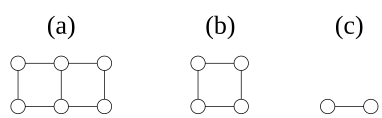
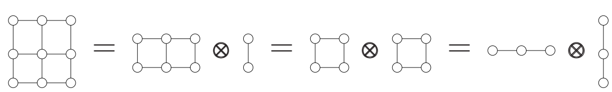


Fig. 2. Antenna array (a) is obtained as the tensor product between subarrays (b) and (c)

Similar obsevations can be made for grid arrays in general.

Example 13. *Take an array of 9 sensors located at* $(x, y) \in \{1, 2, 3\} \times \{1, 2, 3\}$*. We have the relations*

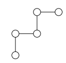
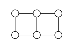
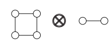
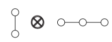
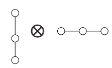
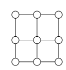
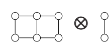
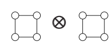


among others.

Let now have a look at the maximal number of sources P_{max} that can be extracted from a tensor of size $I \times J \times \; K$. A sufficient condition is that the total number of paths, P, is smaller than Kruskal's bound (5). We shall simplify the bound by making two assumptions: (a) the loading matrices are generic, that is, they are full rank, and (b) the number of paths is larger than the sizes I and J of the two subarrays entering the array tensor product, and smaller than the number of time samples, K. Under these simplifying assumptions, Kruskal's bound becomes $2P \leq I + J + P - 2$, or:

$$P_{max} = I + J - 2 \tag{16}$$

The table below illustrates the fact that the choice of subarrays has an impact on this bound.

Array	Subarray product	I	J	P_{max}
		3	2	3
		4	2	4
		2	3	3
		3	3	4
		6	2	6
		4	4	6

C. *Signification of the angular constraint*

We are now in a position to interpret the meaning of angular constraints proposed in Section III. According to the notations given in (15), the first coherence

$$\mu^{(1)} = \max_{p\neq q} |\mathbf{u}_p^{(1)\mathsf{H}}\mathbf{u}_q^{(1)}|$$

corresponds to the angular separation viewed from the reference subarray. In fact, vectors $\mathbf{b}_i$ and $\mathbf{d}_p$ having a unit norm, as well as vectors $\mathbf{u}_p$, the quantity $|\mathbf{u}_p^{\mathsf{H}}\mathbf{u}_q|$ may be seen as a measure of angular separation between $\mathbf{d}_p$ and $\mathbf{d}_q$, as we shall now subsequently show in Proposition 15.

Definition 14. *We shall say that a collection of vectors $\{\mathbf{b}_i\}_{1\leq i\leq I}$ is resolvent w.r.t. direction $\mathbf{b}_k - \mathbf{b}_{k0}$ if*

$$0 < \|\mathbf{b}_k - \mathbf{b}_{k0}\| < \frac{\lambda}{2} \tag{17}$$

where $\lambda = \frac{2\pi C}{\omega}$ denotes the wavelength.

Let $\mathbf{b}_i$, $\mathbf{d}_p$ and $\mathbf{u}_q$ be defined as in (15), $1 \leq i \leq I$, $1 \leq p, q \leq P$.

Proposition 15. *If $\{\mathbf{b}_i\}_{1\leq i\leq I}$ is resolvent w.r.t. three linearly independent directions, then*

$$|\mathbf{u}_p^{\mathsf{H}}\mathbf{u}_q| = 1 \Leftrightarrow \mathbf{d}_p = \mathbf{d}_q \tag{18}$$

Proof: Assume $|\mathbf{u}_p^{\mathsf{H}}\mathbf{u}_q| = 1$. Then because they are unit norm, vectors $\mathbf{u}_p$ and $\mathbf{u}_q$ are collinear with a unit modulus proportionality factor. Hence from (15), for all i, k, $1 \leq i, k \leq I$, $(\mathbf{b}_i - \mathbf{b}_k)^{\mathsf{T}}(\mathbf{d}_p - \mathbf{d}_q) \in \lambda\mathbb{Z}$, where λ is defined in Definition 17. Since $\{\mathbf{b}_i\}$ is resolvant, there exist $(i, i0)$ such that $0 < \|\mathbf{b}_i - \mathbf{b}_{i0}\| < \lambda/2$. Hence, because vectors $\mathbf{d}_p$ are unit norm, $\|\mathbf{d}_p - \mathbf{d}_q\| \leq 2$ so that we necessarily have that $(\mathbf{b}_i - \mathbf{b}_{i0})^{\mathsf{T}}(\mathbf{d}_p - \mathbf{d}_q) = 0$. Vector $(\mathbf{d}_p - \mathbf{d}_q)$ is consequently orthogonal to $(\mathbf{b}_i - \mathbf{b}_{i0})$. The same reasoning can be carried out with two other independent vectors. Eventually, vector $(\mathbf{d}_p - \mathbf{d}_q)$ is null because it is orthogonal to three linearly independent vectors in $\mathbb{R}^3$. The converse is immediate, by the definition of $\mathbf{u}_q$. ■

Note that the condition of Definition 17 is not very restrictive, since sensor arrays usually contain sensors separated by half a wavelength or less.

From Section IV-B, one can claim that a similar interpretation can be put forward for the second coherence, which measures the minimal angular separation between paths, viewed from the subarray defining translations.

The third coherence is nothing else but the maximal correlation coefficient between signals received from various paths on the array:

$$\mu^{(3)} = \max_{p \neq q} \frac{|\boldsymbol{\sigma}_p^{\mathsf{H}}\boldsymbol{\sigma}_q|}{||\boldsymbol{\sigma}_p|| \cdot ||\boldsymbol{\sigma}_q||}$$

As a conclusion, the tensor approximation exists and is unique if either signals propagating through various paths are not too much correlated, or if their direction of arrival are not too close. By “not too” it should be understood that the product of coherencies need to satisfy inequality (12) of Corollary 10. In other words, one can separate paths with high correlation provided they are sufficiently well separated in space.

Hence, the decomposition of an array into a tensor product of two (or more) subarrays should not only take into account Kruskal’s bound, as elaborated in Section IV-B, but also the ability of the latter subarrays to separate two distinct directions of arrival (cf. Proposition 15).

REFERENCES

[1] J. C. Pesquet and E. Moreau, “Cumulant based independence measures for linear mixtures,” *IEEE Trans. Information Theory*, pp. 1947–1956, March 2001.

[2] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, “Blind channel identification algorithms based on the Parafac decomposition of cumulant tensors: the single and multiuser cases,” *Signal Processing, Elsevier*, 2008.

[3] P. Comon, “Contrasts, independent component analysis, and blind deconvolution,” *Int. Journal Adapt. Control Sig. Proc.*, vol. 18, no. 3, pp. 225–243, Apr. 2004.

[4] J. F. Cardoso and E. Moulines, “Asymptotic performance analysis of direction-finding algorithms based on fourth-order cumulants,” *IEEE Trans. Sig. Proc.*, vol. 43, no. 1, pp. 214–224, Jan. 1995.

[5] S. Shamsunder and G. B. Giannakis, “Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors,” *Signal Processing, Elsevier*, vol. 30, no. 3, pp. 279–297, Feb. 1993.

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

[7] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] E. J. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, pp. 969–985, 2007.

[9] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best k-term approximation,” *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.

[10] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[11] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization,” *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[12] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Trans. Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[13] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[14] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Information Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[15] M. Fazel, H. Hindi, and S. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in *Proc. American Control Conf.*, vol. 6, 2001, pp. 4734–4739.

[16] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, 2009, preprint. arxiv.org/abs/0910.1879.

[17] F. L. Hitchcock, “Multiple invariants and generalized rank of a p-way matrix or tensor,” *J. Math. and Phys.*, vol. 7, no. 1, pp. 40–79, 1927.

[18] J. D. Carroll and J. J. Chang, “Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition,” *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sep. 1970.

[19] R. A. Harshman, “Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis,” *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970, http://publish. uwo. ca/ harshman.

[20] J. B. Kruskal, “Three-way arrays: Rank and uniqueness of trilinear decompositions,” *Linear Algebra and Applications*, vol. 18, pp. 95–138, 1977.

[21] T. Jiang and N. Sidiropoulos, “Kruskal’s permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models,” *IEEE Trans. Sig. Proc.*, vol. 52, no. 9, pp. 2625–2636, Sep. 2004.

[22] A. Stegeman and N. D. Sidiropoulos, “On Kruskal’s uniqueness condition for the Candecomp/Parafac decomposition,” *Linear Algebra and Appl.*, vol. 420, pp. 540–552, 2007.

[23] V. D. Silva and L.-H. Lim, “Tensor rank and the ill-posedness of the best low-rank approximation problem,” *SIAM Journal on Matrix Analysis Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.

[24] P. Comon, X. Luciani, and A. L. F. De Almeida, “Tensor decompositions, alternating least squares and other tales,” *Jour. Chemometrics*, vol. 23, pp. 393–405, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1002/cem.1236

[25] L.-H. Lim and P. Comon, “Nonnegative approximations of nonnegative tensors,” *Jour. Chemometrics*, vol. 23, pp. 432–441, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1002/cem.1244

[26] P. Comon, “Independent Component Analysis,” in *Higher Order Statistics*, J.-L. Lacoume, Ed. Amsterdam, London: Elsevier, 1992, pp. 29–38. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00346684/fr

[27] T. G. Kolda, “Orthogonal tensor decomposition,” *SIAM Jour. Matrix Ana. Appl.*, vol. 23, no. 1, pp. 243–255, 2001.

[28] P. Paatero, “A weighted non-negative least squares algorithm for three-way Parafac factor analysis,” *Chemometrics Intell. Lab. Syst.*, vol. 38, pp. 223–242, 1997.

[29] J. Håstad, “Tensor rank is np-complete,” *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.

[30] C. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, 2009, preprint. arxiv.org/abs/0911.1393.