



Multiarray signal processing: Tensor decomposition meets compressed sensing

Traitement du signal multi-antenne : Les décompositions tensorielles rejoignent l'échantillonnage compressé

Lek-Heng Lim^a, Pierre Comon^{b,*}

^a Department of Mathematics, University of California, Berkeley, CA 94720-3840, United States

^b Laboratoire I3S, CNRS UMR6070, University of Nice, 06903, Sophia-Antipolis, France

ARTICLE INFO

Article history:

Received 16 December 2009

Accepted after revision 29 June 2010

Available online 24 July 2010

Presented by Odile Macchi

Keywords:

Signal processing

Blind source separation

Blind channel identification

Tensors

Tensor rank

Polyadic tensor decompositions

Best rank- r approximations

Sparse representations

Spark

k -rank

Coherence

Multiarrays

Multisensors

Mots-clés:

Traitement de signal

Séparation aveugle de sources

Identification aveugle de canal

Tenseurs

Rang tensoriel

Décompositions tensorielles polyadiques

Meilleure approximation de rang r

Représentations parcimonieuses

Spark

k -rang

Cohérence

ABSTRACT

We discuss how recently discovered techniques and tools from compressed sensing can be used in tensor decompositions, with a view towards modeling signals from multiple arrays of multiple sensors. We show that with appropriate bounds on a measure of separation between radiating sources called coherence, one could always guarantee the existence and uniqueness of a best rank- r approximation of the tensor representing the signal. We also deduce a computationally feasible variant of Kruskal's uniqueness condition, where the coherence appears as a proxy for k -rank. Problems of sparsest recovery with an infinite continuous dictionary, lowest-rank tensor representation, and blind source separation are treated in a uniform fashion. The decomposition of the measurement tensor leads to simultaneous localization and extraction of radiating sources, in an entirely deterministic manner.

© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

RÉSUMÉ

Nous décrivons comment les techniques et outils d'échantillonnage compressé récemment découverts peuvent être utilisés dans les décompositions tensorielles, avec pour illustration une modélisation des signaux provenant de plusieurs antennes multicapteurs. Nous montrons qu'en posant des bornes appropriées sur une certaine mesure de séparation entre les sources rayonnantes (appelée cohérence dans le jargon de l'échantillonnage compressé), on pouvait toujours garantir l'existence et l'unicité d'une meilleure approximation de rang r du tenseur représentant le signal. Nous en déduisons aussi une variante calculable de la condition d'unicité de Kruskal, où cette cohérence apparaît comme une mesure du k -rang. Les problèmes de récupération parcimonieuse avec un dictionnaire infini continu, de représentation tensorielle de plus bas rang, et de séparation aveugle de sources sont ainsi abordés d'une seule et même façon. La décomposition du tenseur de mesures conduit à la localisation et à l'extraction simultanées des sources rayonnantes, de manière entièrement déterministe.

© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

* Corresponding author.

E-mail addresses: lekheng@math.berkeley.edu (L.-H. Lim), pcomon@unice.fr (P. Comon).

Version française abrégée

Nous expliquons comment les décompositions tensorielles et les modèles d'approximation apparaissent naturellement dans les signaux multicapteurs, et voyons comment l'étude de ces modèles peut être enrichie par des contributions provenant de l'échantillonnage compressé. Le vocable *échantillonnage compressé* est à prendre au sens large, englobant non seulement les idées couvertes par [1–6], mais aussi les travaux sur la minimisation du rang et la complétion de matrice [7–12].

Nous explorons notamment deux thèmes : (1) l'utilisation de dictionnaires redondants avec des bornes sur les produits scalaires entre leurs éléments ; (2) le recours à la cohérence ou au spark pour prouver l'unicité. En particulier, nous verrons comment ces idées peuvent être étendues aux tenseurs, et appliquées à leur décomposition et leurs approximations. Si nous qualifions les travaux [1–6] d'«échantillonnage compressé de formes linéaires» (variables vectorielles) et [7–12] d'«échantillonnage compressé de formes bilinéaires» (variables matricielles), alors cet article porte sur l'échantillonnage compressé de formes multilinéaires (variables tensorielles).

Les approximations tensorielles recèlent des difficultés dues à leur caractère mal posé [14,15], et le calcul de la plupart des problèmes d'algèbre multilinéaire sont de complexité non polynomiale (NP-durs) [16,17]. En outre, il est souvent difficile ou même impossible de répondre dans le cadre de la géométrie algébrique à certaines questions fondamentales concernant les tenseurs, cadre qui est pourtant usuel pour formuler ces questions (cf. Section 4). Nous verrons que certains de ces problèmes pourraient devenir plus abordables si on les déplace de la géométrie algébrique vers l'analyse harmonique. Plus précisément, nous verrons comment les concepts glanés auprès de l'échantillonnage compressé peuvent être utilisés pour atténuer certaines difficultés.

Enfin, nous montrons que si les sources sont suffisamment séparées, alors il est possible de les localiser et de les extraire, d'une manière complètement déterministe. Par «suffisamment séparées», on entend que certains produits scalaires soient inférieurs à un seuil, qui diminue avec le nombre de sources présentes. Dans le jargon de l'échantillonnage compressé, la «cohérence» désigne le plus grand de ces produits scalaires. En posant des bornes appropriées sur cette cohérence, on peut toujours garantir l'existence et l'unicité d'une meilleure approximation de rang r d'un tenseur, et par conséquent l'identifiabilité d'un canal de propagation d'une part, et l'estimation des signaux source d'autre part.

1. Introduction

We discuss how tensor decomposition and approximation models arise naturally in multiarray multisensor signal processing and see how the studies of such models are enriched by mathematical innovations coming from compressed sensing. We interpret the term *compressed sensing* in a loose and broad sense, encompassing not only the ideas covered in [1–6] but also the line of work on rank minimization and matrix completion in [7–12]. We explore two themes in particular: (1) the use of overcomplete dictionaries with bounds on coherence; (2) the use of spark or coherence to obtain uniqueness results. In particular, we will see how these ideas may be extended to tensors and applied to their decompositions and approximations. If we view [1–6] as 'compressed sensing of linear forms' (vector variables) and [7–12] as 'compressed sensing of bilinear forms' (matrix variables), then this article is about 'compressed sensing of multilinear forms' (tensor variables), where these vectors, matrices, or tensors are signals measured by sensors or arrays of sensors.

Tensor approximations are fraught with ill-posedness difficulties [14,15] and computations of most multilinear algebraic problems are NP-hard [16,17]. Furthermore even some of the most basic questions about tensors are often difficult or even impossible to answer within the framework of algebraic geometry, the usual context for formulating such questions (cf. Section 4). We will see that some of these problems with tensors could become more tractable when we move from algebraic geometry to slightly different problems within the framework of harmonic analysis. More specifically we will show how wisdom gleaned from compressed sensing could be used to alleviate some of these issues.

This article is intended to be a short communication. Any result whose proof requires more than a few lines of arguments is not mentioned at all but deferred to our full paper [18]. Relations with other aspects of compressed sensing beyond the two themes mentioned above, most notably exact recoverability results under the restricted isometry property [19] or coherence assumptions [20], are also deferred to [18]. While the discussions in this article are limited to order-3 tensors, it is entirely straightforward to extend them to tensors of any higher order.

2. Multisensor signal processing

Tensors are well-known to arise in signal processing as higher order *cumulants* in independent component analysis [21] and have been used successfully in blind source separation [22]. The signal processing application considered here is of a different nature but also has a natural tensor decomposition model. Unlike the amazing *single-pixel* camera [23] that is celebrated in compressed sensing, this application comes from the opposite end and involves *multiple arrays of multiple sensors* [24].

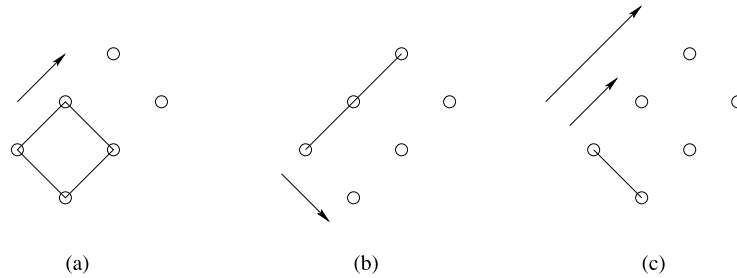


Fig. 1. From the same array of sensors, several subarrays can be defined that differ from each other by translation: (a) two (overlapping) subarrays of 4 sensors, (b) two subarrays of 3 sensors, (c) three subarrays of 2 sensors.

Consider an array of l sensors, each located in space at a position defined by a vector $\mathbf{b}_i \in \mathbb{R}^3$, $i = 1, \dots, l$. Assume this array is impinged by r narrow-band waves transmitted by independent radiating sources through a linear stationary medium. Denote by $\sigma_p(t_k)$ the complex envelope of the p th source, $1 \leq p \leq r$, where t_k denotes a point in time and $k = 1, \dots, n$. If the location of source p is characterized by a parameter vector θ_p , the signal received by sensor i at time t_k can be written as

$$s_i(k) = \sum_{p=1}^r \sigma_p(t_k) \varepsilon_i(\theta_p) \tag{1}$$

where ε_i characterizes the response of sensor i to external excitations.

Such multisensor arrays occur in a variety of applications including acoustics, neuroimaging, and telecommunications. The sensors could be antennas, EEG electrodes, microphones, radio telescopes, etc., capturing signals in the form of images, radio waves, sounds, ultrasounds, etc., emanating from sources that could be cell phones, distant galaxies, human brain, party conversations, etc.

Example 2.1. For instance, if one considers the transmission of narrowband electromagnetic waves over air, $\varepsilon_i(\theta_p)$ can be assimilated to a pure complex exponential (provided the differences between time delays of arrival are much smaller than the inverse of the bandwidth):

$$\varepsilon_i(\theta_p) \approx \exp(\psi_{i,p}), \quad \psi_{i,p} := i \frac{\omega}{c} \left(\mathbf{b}_i^\top \mathbf{d}_p - \frac{1}{2R_p} \|\mathbf{b}_i \wedge \mathbf{d}_p\|_2^2 \right) \tag{2}$$

where the p th source location is defined by its direction $\mathbf{d}_p \in \mathbb{R}^3$ and distance R_p from an arbitrarily chosen origin O , ω denotes the central pulsation, c the wave celerity, $i^2 = -1$, and \wedge the vector wedge product. More generally, one may consider $\psi_{i,p}$ to be a sum of functions whose variables separate, i.e. $\psi_{i,p} = \mathbf{f}(i)^\top \mathbf{g}(p)$, where $\mathbf{f}(i)$ and $\mathbf{g}(p)$ are vectors of the same dimension. Note that if sources are in the far field ($R_p \gg 1$), then the last term in the expression of $\psi_{i,p}$ in (2) may be neglected.

2.1. Structured multisensor arrays

We are interested in sensor arrays enjoying an invariance property. We assume that there are m arrays, each having the same number l of sensors. They do not need to be disjoint, that is, two different arrays may share one or more sensors.

From (1), the signal received by the j th array, $j = 1, \dots, m$, takes the form

$$s_{i,j}(k) = \sum_{p=1}^r \sigma_p(t_k) \varepsilon_{i,j}(\theta_p) \tag{3}$$

The invariance property¹ that we are interested in can be expressed as

$$\varepsilon_{i,j}(\theta_p) = \varepsilon_{i,1}(\theta_p) \varphi(j, p) \tag{4}$$

In other words, variables i and j decouple.

This property is encountered in the case of arrays that can be obtained from each other by a translation (see Fig. 1). Assume sources are in the far field. Denote by Δ_j the vector that allows deduction of the locations of sensors in the j th

¹ So called as the property follows from translation invariance: angles of arrival remain the same. The term was probably first used in [25].

array from those of 1st array. Under these hypotheses, we have for the first array, $\psi_{i,p,1} = \iota \frac{\omega}{c} (\mathbf{b}_i^\top \mathbf{d}_p)$. By a translation of Δ_j we obtain the phase response of the j th array as:

$$\psi_{i,p,j} = \iota \frac{\omega}{c} (\mathbf{b}_i^\top \mathbf{d}_p + \Delta_j^\top \mathbf{d}_p)$$

Observe that indices i and j decouple upon exponentiation and that we have $\varphi(j, p) = \exp(\iota \frac{\omega}{c} \Delta_j^\top \mathbf{d}_p)$.

Now plug the invariance expression (4) into (3) to obtain the observation model:

$$s_{i,j}(k) = \sum_{p=1}^r \varepsilon_{i,1}(\theta_p) \varphi(j, p) \sigma_p(t_k), \quad i = 1, \dots, l; \quad j = 1, \dots, m; \quad k = 1, \dots, n$$

This simple multilinear model is the one that we shall discuss in this article. Note that the left-hand side is measured, while the quantities on the right-hand side are to be estimated. If we rewrite $a_{ijk} = s_{i,j}(k)$, $u_{ip} = \hat{\varepsilon}_{i,1}(\theta_p)$, $v_{jp} = \hat{\varphi}(j, p)$, $w_{kp} = \hat{\sigma}_p(t_k)$ (where the ‘hat’ indicates that the respective quantities are suitably normalized) and introduce a scalar λ_p to capture the collective magnitudes, we get the *tensor decomposition* model

$$a_{ijk} = \sum_{p=1}^r \lambda_p u_{ip} v_{jp} w_{kp}, \quad i = 1, \dots, l; \quad j = 1, \dots, m; \quad k = 1, \dots, n$$

with $\|\mathbf{u}_p\| = \|\mathbf{v}_p\| = \|\mathbf{w}_p\| = 1$. In the presence of noise, we often seek a *tensor approximation* model with respect to some measure of nearness, say, a sum-of-squares loss that is common when the noise is assumed white and Gaussian:

$$\min_{\lambda_p, \mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p} \sum_{i,j,k=1}^{l,m,n} \left[a_{ijk} - \sum_{p=1}^r \lambda_p u_{ip} v_{jp} w_{kp} \right]^2$$

Our model has the following physical interpretation: if a_{ijk} is the array of measurements recorded from sensor i of subarray j at time k , then it is ideally written as a sum of r individual source contributions $\sum_{p=1}^r \lambda_p u_{ip} v_{jp} w_{kp}$. Here, u_{ip} represent the transfer functions among sensors of the same subarray, v_{jp} the transfer between subarrays, and w_{kp} the discrete-time source signals. All these quantities can be identified. In other words, the exact way one subarray can be deduced from the others does not need to be known. Only the existence of this geometrical invariance is required.

3. Tensor rank

Let $\mathbb{V}_1, \dots, \mathbb{V}_k$ be vector spaces over a field, say, \mathbb{C} . An element of the tensor product $\mathbb{V}_1 \otimes \dots \otimes \mathbb{V}_k$ is called an *order- k tensor* or *k -tensor* for short. Scalars, vectors, and matrices may be regarded as tensors of order 0, 1, and 2 respectively. For the purpose of this article and for notational simplicity, we will limit our discussions to 3-tensors. Denote by l, m, n the dimensions of $\mathbb{V}_1, \mathbb{V}_2$, and \mathbb{V}_3 , respectively. Up to a choice of bases on $\mathbb{V}_1, \mathbb{V}_2, \mathbb{V}_3$, a 3-tensor in $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \mathbb{V}_3$ may be represented by an $l \times m \times n$ array of elements of \mathbb{C} ,

$$A = (a_{ijk})_{i,j,k=1}^{l,m,n} \in \mathbb{C}^{l \times m \times n}$$

These are sometimes called *hypermatrices*² and come equipped with certain algebraic operations inherited from the algebraic structure of $\mathbb{V}_1 \otimes \mathbb{V}_2 \otimes \mathbb{V}_3$. The one that interests us most is the decomposition of $A = (a_{ijk}) \in \mathbb{C}^{l \times m \times n}$ as

$$A = \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p, \quad a_{ijk} = \sum_{p=1}^r \lambda_p u_{ip} v_{jp} w_{kp} \tag{5}$$

with $\lambda_p \in \mathbb{C}$, $\mathbf{u}_p \in \mathbb{C}^l$, $\mathbf{v}_p \in \mathbb{C}^m$, $\mathbf{w}_p \in \mathbb{C}^n$. For $\mathbf{u} = [u_1, \dots, u_l]^\top$, $\mathbf{v} = [v_1, \dots, v_m]^\top$, $\mathbf{w} = [w_1, \dots, w_n]^\top$, we write $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} := (u_i v_j w_k)_{i,j,k=1}^{l,m,n} \in \mathbb{C}^{l \times m \times n}$. This generalizes $\mathbf{u} \otimes \mathbf{v} = \mathbf{uv}^\top$ in the case of matrices.

A different choice of bases on $\mathbb{V}_1, \dots, \mathbb{V}_k$ would lead to a different hypermatrix representation of elements in $\mathbb{V}_1 \otimes \dots \otimes \mathbb{V}_k$. For the more pedantic readers, it is understood that what we call a tensor in this article really means a hypermatrix. The decomposition of a tensor into a linear combination of rank-1 tensors was first studied in [26].

Definition 3.1. A tensor that can be expressed as an outer product of vectors is called **decomposable** (or **rank-one** if it is also nonzero). More generally, the **rank** of a tensor $A = (a_{ijk})_{i,j,k=1}^{l,m,n} \in \mathbb{C}^{l \times m \times n}$, denoted $\text{rank}(A)$, is defined as the minimum r for which A may be expressed as a sum of r rank-1 tensors,

² The subscripts and superscripts will be dropped when the range of i, j, k is obvious or unimportant.

$$\text{rank}(A) := \min \left\{ r \mid A = \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p \right\} \tag{6}$$

We will call a decomposition of the form (5) a *rank-revealing decomposition* when $r = \text{rank}(A)$. The definition of rank in (6) agrees with matrix rank when applied to an order-2 tensor.

$\mathbb{C}^{l \times m \times n}$ is a Hilbert space of dimension lmn , equipped with the *Frobenius* (or *Hilbert–Schmidt*) norm, and its associated scalar product:

$$\|A\|_F = \left[\sum_{i,j,k=1}^{l,m,n} |a_{ijk}|^2 \right]^{\frac{1}{2}}, \quad \langle A, B \rangle_F = \sum_{i,j,k=1}^{l,m,n} a_{ijk} \bar{b}_{ijk}$$

One may also define tensor norms that are the ℓ^p equivalent of Frobenius norm [27] and tensor norms that are analogous to operator norms of matrices [17].

4. Existence

The problem that we consider here is closely related to the best r -term approximation problem in nonlinear approximations, with one notable difference – our dictionary is a continuous manifold, as opposed to a discrete set, of atoms. We approximate a general signal $\mathbf{v} \in \mathbb{H}$ with an r -term approximant over some *dictionary* of atoms \mathcal{D} , i.e. $\mathcal{D} \subseteq \mathbb{H}$ and $\text{span}(\mathcal{D}) = \mathbb{H}$. We refer the reader to [3] for a discussion of the connection between compressed sensing and nonlinear approximations. We denote the set of r -term approximants by $\Sigma_r(\mathcal{D}) := \{\lambda_1 \mathbf{v}_1 + \dots + \lambda_r \mathbf{v}_r \in \mathbb{H} \mid \mathbf{v}_1, \dots, \mathbf{v}_r \in \mathcal{D}, \lambda_1, \dots, \lambda_r \in \mathbb{C}\}$. Usually \mathcal{D} is finite or countable but we have a continuum of atoms comprising all decomposable tensors. The set of decomposable tensors

$$\text{Seg}(l, m, n) := \{A \in \mathbb{C}^{l \times m \times n} \mid \text{rank}(A) \leq 1\} = \{\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} \mid \mathbf{x} \in \mathbb{C}^l, \mathbf{y} \in \mathbb{C}^m, \mathbf{z} \in \mathbb{C}^n\}$$

is known in geometry as the *Segre variety*. It has the structure of both a smooth manifold and an algebraic variety, with dimension $l + m + n$ (whereas finite or countable dictionaries are 0-dimensional). The set of r -term approximants in our case is the r th *secant quasiprojective variety* of the Segre variety, $\Sigma_r(\text{Seg}(l, m, n)) = \{A \in \mathbb{C}^{l \times m \times n} \mid \text{rank}(A) \leq r\}$. Such a set may not be closed nor irreducible. In order to study this set using standard tools of algebraic geometry [28–30], one often considers a simpler variant called the r th *secant variety* of the Segre variety, the (Zariski) closure of $\Sigma_r(\text{Seg}(l, m, n))$. Even with this simplification, many basic questions remain challenging and open: For example, it is not known what the value of the generic rank³ is for general values of l, m, n [28]; nor are the polynomial equations⁴ defining the r th secant variety known in general [29].

The seemingly innocent remark in the preceding paragraph that for $r > 1$, the set $\{A \in \mathbb{C}^{l \times m \times n} \mid \text{rank}(A) \leq r\}$ is in general not a closed set has implication on the model that we proposed. Another way to view this is that tensor rank for tensors of order 3 or higher is not an upper semicontinuous function [15]. Note that tensor rank for order-2 tensors (i.e. matrix rank) is upper semicontinuous: if A is a matrix and $\text{rank}(A) = r$, then $\text{rank}(B) \geq r$ for all matrices B in a sufficiently small neighborhood of A . As a consequence, the best rank- r approximation problem for tensors,

$$\underset{\|\mathbf{u}_p\|_2 = \|\mathbf{v}_p\|_2 = \|\mathbf{w}_p\|_2 = 1}{\text{argmin}} \|A - \lambda_1 \mathbf{u}_1 \otimes \mathbf{v}_1 \otimes \mathbf{w}_1 - \dots - \lambda_r \mathbf{u}_r \otimes \mathbf{v}_r \otimes \mathbf{w}_r\|_F \tag{7}$$

unlike that for matrices, does not in general have a solution. The following is a simple example taken from [15]:

Example 4.1. Let $\mathbf{u}_i, \mathbf{v}_i \in \mathbb{C}^m, i = 1, 2, 3$. Let $A := \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \mathbf{v}_3 + \mathbf{u}_1 \otimes \mathbf{v}_2 \otimes \mathbf{u}_3 + \mathbf{v}_1 \otimes \mathbf{u}_2 \otimes \mathbf{u}_3$ and for $n \in \mathbb{N}$, let

$$A_n := n \left(\mathbf{u}_1 + \frac{1}{n} \mathbf{v}_1 \right) \otimes \left(\mathbf{u}_2 + \frac{1}{n} \mathbf{v}_2 \right) \otimes \left(\mathbf{u}_3 + \frac{1}{n} \mathbf{v}_3 \right) - n \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \mathbf{u}_3$$

One may show that $\text{rank}(A) = 3$ iff $\mathbf{u}_i, \mathbf{v}_i$ are linearly independent, $i = 1, 2, 3$. Since it is clear that $\text{rank}(A_n) \leq 2$ by construction and $\lim_{n \rightarrow \infty} A_n = A$, the rank-3 tensor A has no best rank-2 approximation. Such a tensor is said to have border rank 2.

This phenomenon where a tensor fails to have a best rank- r approximation is much more widespread than one might imagine, occurring over a wide range of dimensions, orders, and ranks; happens regardless of the choice of norm (or even Brègman divergence) used. These counterexamples occur with positive probability and in some cases with certainty (in $\mathbb{R}^{2 \times 2 \times 2}$ and $\mathbb{C}^{2 \times 2 \times 2}$, no tensor of rank-3 has a best rank-2 approximation). We refer the reader to [15] for further details.

³ Roughly speaking, this is the value of r such that a randomly generated tensor will have rank r . For $m \times n$ matrices, the generic rank is $\min\{m, n\}$ but $l \times m \times n$ tensors in general have generic rank $> \min\{l, m, n\}$.

⁴ For matrices, these equations are simply given by the vanishing of the $k \times k$ minors for all $k > r$.

Why not consider approximation by tensors in the closure of the set of all rank- r tensors, i.e. the r th secant variety, instead? Indeed this was the idea behind the *weak solutions* suggested in [15]. The trouble with this approach is that it is not known how one could parameterize the r th secant variety in general: While we know that all elements of the r th secant quasiprojective variety $\Sigma_r(\text{Seg}(l, m, n))$ may be parameterized as $\lambda_1 \mathbf{u}_1 \otimes \mathbf{v}_1 \otimes \mathbf{w}_1 + \dots + \lambda_r \mathbf{u}_r \otimes \mathbf{v}_r \otimes \mathbf{w}_r$, it is not known how one could parameterize the limits of these, i.e. the additional elements that occur in the closure of $\Sigma_r(\text{Seg}(l, m, n))$, when $r > \min\{l, m, n\}$. More specifically, if $r \leq \min\{l, m, n\}$, Terracini’s Lemma [30] provides a way to do this since generically a rank- r tensor has the form $\lambda_1 \mathbf{u}_1 \otimes \mathbf{v}_1 \otimes \mathbf{w}_1 + \dots + \lambda_r \mathbf{u}_r \otimes \mathbf{v}_r \otimes \mathbf{w}_r$ where $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$, $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ are linearly independent; but when $r > \min\{l, m, n\}$, this generic linear independence does not hold and there are no known ways to parameterize a rank- r tensor in this case.

We propose that a better way would be to introduce natural *a priori* conditions that prevent the phenomenon in Example 4.1 from occurring. An example of such conditions is nonnegativity restrictions on λ_i , \mathbf{u}_i , \mathbf{v}_i , examined in our earlier work [27]. Here we will impose much weaker and more natural restrictions motivated by the notion of *coherence*. Recall that a real valued function f with an unbounded domain $\text{dom}(f)$ and $\lim_{\mathbf{x} \in \text{dom}(f), \|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty$ is called *coercive* (or *0-coercive*) [31]. A nice feature of such functions is that the existence of a global minimizer is guaranteed. The objective function in (7) is not coercive in general but we will show here that a mild condition on coherence, a notion that frequently appears in recent work on compressed sensing, allows us to obtain a coercive function and therefore circumvent the non-existence difficulty. In the context of our application in Section 2, coherence quantifies the minimal angular separation in space or the maximal cross correlation in time between the radiating sources.

Definition 4.2. Let \mathbb{H} be a Hilbert space and $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{H}$ be a finite collection of unit vectors, i.e. $\|\mathbf{v}_p\|_{\mathbb{H}} = 1$. The **coherence** of the collection $V = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is defined as $\mu(V) := \max_{p \neq q} |\langle \mathbf{v}_p, \mathbf{v}_q \rangle|$.

This notion has been introduced in slightly different forms and names: mutual incoherence of two dictionaries [5], mutual coherence of two dictionaries [2], the coherence of a subspace projection [8], etc. The version here follows that of [6]. We will be interested in the case when \mathbb{H} is finite dimensional (in particular, $\mathbb{H} = \mathbb{C}^{l \times m \times n}$ or \mathbb{C}^m). When $\mathbb{H} = \mathbb{C}^m$, we often regard V as an $m \times r$ matrix whose column vectors are $\mathbf{v}_1, \dots, \mathbf{v}_r$. Clearly $0 \leq \mu(V) \leq 1$, $\mu(V) = 0$ iff $\mathbf{v}_1, \dots, \mathbf{v}_r$ are orthonormal, and $\mu(V) = 1$ iff V contains at least a pair of collinear vectors.

While a solution to the best rank- r approximation problem (7) may not exist, the following shows that a solution to the bounded coherence best rank- r approximation problem (8) always exists.

Theorem 4.3. Let $A \in \mathbb{C}^{l \times m \times n}$ and let $\mathcal{U} = \{U \in \mathbb{C}^{l \times r} \mid \mu(U) \leq \mu_1\}$, $\mathcal{V} = \{V \in \mathbb{C}^{m \times r} \mid \mu(V) \leq \mu_2\}$, $\mathcal{W} = \{W \in \mathbb{C}^{n \times r} \mid \mu(W) \leq \mu_3\}$, be families of dictionaries of unit vectors of coherence not more than μ_1, μ_2, μ_3 respectively. If

$$\mu_1 \mu_2 \mu_3 < \frac{1}{r}$$

then the infimum η defined as

$$\eta = \inf \left\{ \left\| A - \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p \right\| \mid \lambda \in \mathbb{C}^r, U \in \mathcal{U}, V \in \mathcal{V}, W \in \mathcal{W} \right\} \tag{8}$$

is attained. Here $\|\cdot\|$ denotes any norm on $\mathbb{C}^{l \times m \times n}$.

Proof. Since all norms are equivalent on a finite dimensional space, we may assume that $\|\cdot\| = \|\cdot\|_F$, the Frobenius norm. Let the objective function $f : \mathbb{C}^r \times \mathcal{U} \times \mathcal{V} \times \mathcal{W} \rightarrow [0, \infty)$ be

$$f(\lambda, U, V, W) := \left\| A - \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p \right\|_F^2 \tag{9}$$

Let $\mathcal{E} = \mathbb{C}^r \times \mathcal{U} \times \mathcal{V} \times \mathcal{W}$. Note that \mathcal{E} as a subset of $\mathbb{C}^{r(1+l+m+n)}$ is noncompact (closed but unbounded). We write $T = (\lambda, U, V, W)$ and let the infimum in question be $\eta := \inf\{f(T) \mid T \in \mathcal{E}\}$. We will show that the sublevel set of f restricted to \mathcal{E} , defined as $\mathcal{E}_\alpha = \{T \in \mathcal{E} \mid f(T) \leq \alpha\}$, is compact for all $\alpha > \eta$ and thus the infimum of f on \mathcal{E} is attained. The set $\mathcal{E}_\alpha = \mathcal{E} \cap f^{-1}(-\infty, \alpha]$ is closed since \mathcal{E} is closed and f is continuous (by the continuity of norm). It remains to show that \mathcal{E}_α is bounded. Suppose the contrary. Then there exists a sequence $(T_k)_{k=1}^\infty \subset \mathcal{E}$ with $\|T_k\|_2 \rightarrow \infty$ but $f(T_k) \leq \alpha$ for all k . Clearly, $\|T_k\|_2 \rightarrow \infty$ implies that $\|\lambda^{(k)}\|_2 \rightarrow \infty$. Note that

$$f(T) \geq \left[\|A\|_F - \left\| \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p \right\|_F \right]^2$$

We have

$$\begin{aligned} \left\| \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p \right\|_F^2 &= \sum_{p,q=1}^r \lambda_p \bar{\lambda}_q \langle \mathbf{u}_p, \mathbf{u}_q \rangle \langle \mathbf{v}_p, \mathbf{v}_q \rangle \langle \mathbf{w}_p, \mathbf{w}_q \rangle \\ &\geq \sum_{p=1}^r |\lambda_p|^2 \|\mathbf{u}_p\|_2^2 \|\mathbf{v}_p\|_2^2 \|\mathbf{w}_p\|_2^2 - \sum_{p \neq q} |\lambda_p \bar{\lambda}_q \langle \mathbf{u}_p, \mathbf{u}_q \rangle \langle \mathbf{v}_p, \mathbf{v}_q \rangle \langle \mathbf{w}_p, \mathbf{w}_q \rangle| \\ &\geq \sum_{p=1}^r |\lambda_p|^2 - \mu_1 \mu_2 \mu_3 \left(\sum_{p \neq q} |\lambda_p \bar{\lambda}_q| \right) \\ &\geq \|\boldsymbol{\lambda}\|_2^2 - \mu_1 \mu_2 \mu_3 \|\boldsymbol{\lambda}\|_1^2 \geq (1 - r \mu_1 \mu_2 \mu_3) \|\boldsymbol{\lambda}\|_2^2 \end{aligned}$$

The last inequality follows from $\|\boldsymbol{\lambda}\|_1 \leq \sqrt{r} \|\boldsymbol{\lambda}\|_2$ for any $\boldsymbol{\lambda} \in \mathbb{C}^r$. By our assumption $1 - r \mu_1 \mu_2 \mu_3 > 0$ and so as $\|\boldsymbol{\lambda}^{(k)}\|_2 \rightarrow \infty$, $f(T_k) \rightarrow \infty$, which contradicts the assumption that $f(T_k) \leq \alpha$ for all k . \square

5. Uniqueness

While never formally stated, one of the main maxims in compressed sensing is that ‘uniqueness implies sparsity’. For example, this is implicit in various sparsest recovery arguments in [2,5,6] where, depending on context, ‘sparsest’ may also mean ‘lowest rank’. We state a simple formulation of this observation for our purpose. Let \mathcal{D} be a dictionary of atoms in a vector space \mathbb{V} (over an infinite field). We do not require \mathcal{D} to be finite or countable. In almost all cases of interest \mathcal{D} will be overcomplete with high redundancy. For $\mathbf{x} \in \mathbb{V}$ and $r \in \mathbb{N}$, by a \mathcal{D} -representation, we shall mean a representation of the form $\mathbf{x} = \alpha_1 \mathbf{x}_1 + \dots + \alpha_r \mathbf{x}_r$ where $\mathbf{x}_1, \dots, \mathbf{x}_r \in \mathcal{D}$ and $\alpha_1 \dots \alpha_r \neq 0$ ($\mathbf{x}_1, \dots, \mathbf{x}_r$ are not required to be distinct).

Lemma 5.1. *Let $\mathbf{x} = \alpha_1 \mathbf{x}_1 + \dots + \alpha_r \mathbf{x}_r$ be a \mathcal{D} -representation.*

- (i) *If this is the unique \mathcal{D} -representation with r terms, then $\mathbf{x}_1, \dots, \mathbf{x}_r$ must be linearly independent.*
- (ii) *If this is the sparsest \mathcal{D} -representation, then $\mathbf{x}_1, \dots, \mathbf{x}_r$ must be linearly independent.*
- (iii) *If this is unique, then it must also be sparsest.*

Proof. Suppose $\beta_1 \mathbf{x}_1 + \dots + \beta_r \mathbf{x}_r = \mathbf{0}$ is a nontrivial linear relation. (i) Since not all β_i are 0 while all $\alpha_i \neq 0$, for some $\theta \neq 0$ we must have $(\alpha_1 + \theta \beta_1) \dots (\alpha_r + \theta \beta_r) \neq 0$, which yields a different \mathcal{D} -representation $\mathbf{x} = \mathbf{x} + \theta \mathbf{0} = (\alpha_1 + \theta \beta_1) \mathbf{x}_1 + \dots + (\alpha_r + \theta \beta_r) \mathbf{x}_r$. (ii) Say $\beta_r \neq 0$, then $\mathbf{x} = (\alpha_1 - \beta_r^{-1} \beta_1) \mathbf{x}_1 + \dots + (\alpha_{r-1} - \beta_r^{-1} \beta_{r-1}) \mathbf{x}_{r-1}$ is a sparser \mathcal{D} -representation. (iii) Let $\mathbf{x} = \gamma_1 \mathbf{y}_1 + \dots + \gamma_s \mathbf{y}_s$ be a \mathcal{D} -representation with $s < r$. Write $\mathbf{y}_1 = \sum_{k=1}^{r-s+1} \theta_k \mathbf{x}_k$ with $\sum_{k=1}^{r-s+1} \theta_k = 1$. Then we obtain an r -term \mathcal{D} -representation $\sum_{k=1}^{r-s+1} \gamma_1 \theta_k \mathbf{x}_k + \sum_{i=2}^s \gamma_i \mathbf{y}_i$, different from the given one. They are different since $\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_s$ are linearly dependent, whereas (i) implies that $\mathbf{x}_1, \dots, \mathbf{x}_r$ are linearly independent. \square

We will now discuss a combinatorial notion useful in guaranteeing uniqueness or sparsity of \mathcal{D} -representations. The notion of the *girth* of a circuit [32] is standard and well-known in graphical matroids – it is simply the length of a shortest cycle of a graph. However the girth of a circuit in vector matroids, i.e. the cardinality of the smallest linearly dependent subset of a collection of vectors in a vector space, has rarely appeared in linear algebra. This has led to it being reinvented multiple times under different names, most notably as *Kruskal rank* or *k-rank* in tensor decompositions [33], as *spark* in compressed sensing [5], and as *k-stability* in coding theory [12]. The notions of girth, spark, *k-rank*, and *k-stability* [12] are related as follows:

Lemma 5.2. *Let \mathbb{V} be a vector space over a field \mathbb{F} and $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a finite subset of \mathbb{V} . Then*

$$\text{girth}(X) = \text{spark}(X) = \text{krank}(X) + 1$$

and furthermore X is k -stable iff $\text{krank}(X) = n - k$.

Proof. These follow directly from the respective definitions. \square

These notions are unfortunately expected to be difficult to compute because of the following result [34].

Theorem 5.3 (Vardy). *It is NP-hard to compute the girth of a vector matroid over a finite field of two elements, \mathbb{F}_2 .*

A consequence is that spark, *k-rank*, *k-stability* are all NP-hard if the field is \mathbb{F}_2 . We note here that several authors have assumed that spark is NP-hard to compute over $\mathbb{F} = \mathbb{R}$ or \mathbb{C} (assuming \mathbb{Q} or $\mathbb{Q}[i]$ inputs) but this is actually unknown. In particular, it does not follow from [35]. While it is clear that computing spark via a naive exhaustive search has complexity

$O(2^n)$, one may perhaps do better with cleverer algorithms when $\mathbb{F} = \mathbb{R}$ or \mathbb{C} ; in fact it is unknown in this case whether the corresponding decision problem (Given finite $X \subseteq \mathbb{V}$ and $s \in \mathbb{N}$, is $\text{spark}(X) = s$?) is NP-hard. On the other hand it is easy to compute coherence. Even a straightforward search for an off-diagonal entry of $X^\top X$ of maximum magnitude is of polynomial complexity. An important observation of [5] is that coherence may sometimes be used in place of spark.

One of the early results in compressed sensing [5,6] on the uniqueness of the sparsest solution is that if

$$\frac{1}{2} \text{spark}(X) \geq \|\beta\|_0 = \text{card}\{\beta_i \neq 0\} \tag{10}$$

then $\beta \in \mathbb{C}^n$ is a unique solution to $\min\{\|\beta\|_0 \mid X\beta = \mathbf{x}\}$.

For readers familiar with Kruskal’s condition that guarantees the uniqueness of tensor decomposition, the parallel with (10) is hard to miss once we rewrite Kruskal’s condition in the form

$$\frac{1}{2} [\text{krank}(X) + \text{krank}(Y) + \text{krank}(Z)] > \text{rank}(A) \tag{11}$$

We state a slight variant of Kruskal’s result [33] here. Note that the scaling ambiguity is unavoidable because of the multilinearity of \otimes .

Theorem 5.4 (Kruskal). *If $A = \sum_{p=1}^r \mathbf{x}_p \otimes \mathbf{y}_p \otimes \mathbf{z}_p$ and $\text{krank}(X) + \text{krank}(Y) + \text{krank}(Z) \geq 2r + 2$, then $r = \text{rank}(A)$ and the decomposition is unique up to scaling of the form $\alpha \mathbf{x} \otimes \beta \mathbf{y} \otimes \gamma \mathbf{z} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$ for $\alpha, \beta, \gamma \in \mathbb{C}$ with $\alpha\beta\gamma = 1$. This inequality is also sharp in the sense that $2r + 2$ cannot be replaced by $2r + 1$.*

Proof. The uniqueness was Kruskal’s original result in [33]; alternate shorter proofs may be found in [36,13,37]. That $r = \text{rank}(A)$ then follows from Lemma 5.1(iii). The sharpness of the inequality is due to [38]. \square

Since spark is expected to be difficult to compute, one may substitute coherence to get a condition [5,6] that is easier to check

$$\frac{1}{2} \left[1 + \frac{1}{\mu(X)} \right] \geq \|\beta\|_0 \tag{12}$$

Eq. (12) relaxes (10) because of the following result of [5,6]:

Lemma 5.5. *Let \mathbb{H} be a Hilbert space and $V = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ be a finite collection of unit vectors in \mathbb{H} . Then*

$$\text{spark}(V) \geq 1 + \frac{1}{\mu(V)} \quad \text{and} \quad \text{krank}(V) \geq \frac{1}{\mu(V)}$$

Proof. Let $\text{spark}(V) = s = \text{krank}(V) + 1$. Assume without loss of generality that $\{\mathbf{v}_1, \dots, \mathbf{v}_s\}$ is a minimal circuit of V and that $\alpha_1 \mathbf{v}_1 + \dots + \alpha_s \mathbf{v}_s = \mathbf{0}$ with $|\alpha_1| = \max\{|\alpha_1|, \dots, |\alpha_s|\} > 0$. Taking inner product with \mathbf{v}_1 we get $\alpha_1 = -\alpha_2 \langle \mathbf{v}_2, \mathbf{v}_1 \rangle - \dots - \alpha_s \langle \mathbf{v}_s, \mathbf{v}_1 \rangle$ and so $|\alpha_1| \leq (|\alpha_2| + \dots + |\alpha_s|) \mu(V)$. Dividing by $|\alpha_1|$ then yields $1 \leq (s - 1) \mu(V)$. \square

We now characterize the uniqueness of tensor decompositions in terms of coherence. Note that \mathbb{C} may be replaced by \mathbb{R} . By “unimodulus scaling”, we mean scaling of the form $e^{i\theta_1} \mathbf{u} \otimes e^{i\theta_2} \mathbf{v} \otimes e^{i\theta_3} \mathbf{w}$ where $\theta_1 + \theta_2 + \theta_3 \equiv 0 \pmod{2\pi}$.

Theorem 5.6. *Let $A \in \mathbb{C}^{l \times m \times n}$ and $A = \sum_{p=1}^r \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p$, where $\lambda_p \in \mathbb{C}$, $\lambda_p \neq 0$, and $\|\mathbf{u}_p\|_2 = \|\mathbf{v}_p\|_2 = \|\mathbf{w}_p\|_2 = 1$ for all $p = 1, \dots, r$. We write $U = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, $V = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$, $W = \{\mathbf{w}_1, \dots, \mathbf{w}_r\}$. If*

$$\frac{1}{2} \left[\frac{1}{\mu(U)} + \frac{1}{\mu(V)} + \frac{1}{\mu(W)} \right] > r \tag{13}$$

then $r = \text{rank}(A)$ and the rank revealing decomposition is unique up to unimodulus scaling.

Proof. If (13) is satisfied, then Kruskal’s condition for uniqueness (11) must also be satisfied by Lemma 5.5. \square

Note that unlike the k -ranks in (11), the coherences in (13) are trivial to compute. In addition to uniqueness, an easy but important consequence of Theorem 5.6 is that it provides a readily checkable sufficient condition for tensor rank, which is NP-hard over any field [16,17].

6. Conclusion

The following existence and uniqueness result may be deduced from Theorems 4.3 and 5.6:

Corollary 6.1. Let $A \in \mathbb{C}^{l \times m \times n}$. If $\mu_1, \mu_2, \mu_3 \in (0, \infty)$ satisfy

$$\frac{1}{\sqrt[3]{\mu_1 \mu_2 \mu_3}} > \frac{2}{3} r$$

then the bounded coherence rank- r approximation problem (8) has a solution that is unique up to unimodulus scaling.

Proof. The case $r = 1$ is trivial. For $r \geq 2$, since $\mu_1 \mu_2 \mu_3 < (3/2r)^3 < 1/r$, Theorem 4.3 guarantees that a solution to (8) exists. Let $A_r = \lambda_1 \mathbf{u}_1 \otimes \mathbf{v}_1 \otimes \mathbf{w}_1 + \dots + \lambda_r \mathbf{u}_r \otimes \mathbf{v}_r \otimes \mathbf{w}_r$ be a solution and let $U = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, $V = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$, $W = \{\mathbf{w}_1, \dots, \mathbf{w}_r\}$. Since $\mu(U) \leq \mu_1$, $\mu(V) \leq \mu_2$, $\mu(W) \leq \mu_3$, the harmonic mean-geometric mean inequality yields

$$3 \left[\frac{1}{\mu(U)} + \frac{1}{\mu(V)} + \frac{1}{\mu(W)} \right]^{-1} \leq \sqrt[3]{\mu(U)\mu(V)\mu(W)} < \sqrt[3]{\mu_1 \mu_2 \mu_3} < \frac{3}{2r}$$

the decomposition of A_r is unique by Theorem 5.6. \square

In the context of our application in Section 2.1, this corollary means that radiating sources can be uniquely localized if they are either (i) sufficiently separated in space (angular separation viewed by a subarray, or by the array defined by translations between subarrays), or (ii) in time (small sample cross correlations), noting that the scalar product between two time series is simply the sample cross correlation. Contrary to more classical approaches based on second or higher order moments, both conditions are not necessary here – Corollary 6.1 requires only that the *product* between coherences be small. In addition, there is no need for long data samples since the approach is deterministic; this is totally unusual in antenna array processing. Cross correlations need to be sufficiently small among sources only for identifiability purposes but they are not explicitly computed in the identification process. Hence our model is robust with respect to short record durations. Observe also that the number of time samples can be as small as the number of sources. Lastly, an estimate of source time samples may be obtained from the tensor decomposition as a key byproduct of this deterministic approach.

References

- [1] E.J. Candès, Compressive sampling, Int. Congress Math. III (2006) 1433–1452.
- [2] E.J. Candès, J. Romberg, Sparsity and incoherence in compressive sampling, Inverse Problems 23 (3) (2007) 969–985.
- [3] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k -term approximation, J. Amer. Math. Soc. 22 (1) (2009) 211–231.
- [4] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
- [5] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization, Proc. Nat. Acad. Sci. 100 (5) (2003) 2197–2202.
- [6] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 49 (12) (2003) 3320–3325.
- [7] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717–772.
- [8] E.J. Candès, T. Tao, The power of convex relaxation: Near-optimal matrix completion, preprint, 2009.
- [9] M. Fazel, H. Hindi, S. Boyd, A rank minimization heuristic with application to minimum order system approximation, Proc. Amer. Control Conf. 6 (2001) 4734–4739.
- [10] D. Gross, Recovering low-rank matrices from few coefficients in any basis, preprint, 2009, <http://arxiv.org/abs/0910.1879>.
- [11] B. Recht, M. Fazel, P. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, preprint, 2008, <http://arxiv.org/abs/0706.4138>.
- [12] Z. Zhu, A.M.-C. So, Y. Ye, Fast and near-optimal matrix completion via randomized basis pursuit, preprint, 2009.
- [13] J. Rhodes, A concise proof of Kruskal's theorem on tensor decomposition, Linear Algebra Appl. 432 (7) (2010) 1818–1824.
- [14] P. Comon, G. Golub, L.-H. Lim, B. Mourrain, Symmetric tensors and symmetric tensor rank, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1254–1279.
- [15] V. De Silva, L.-H. Lim, Tensor rank and the ill-posedness of the best low-rank approximation problem, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1084–1127.
- [16] J. Håstad, Tensor rank is NP-complete, J. Algorithms 11 (4) (1990) 644–654.
- [17] C. Hillar, L.-H. Lim, Most tensor problems are NP-hard, preprint, 2009, <http://arxiv.org/abs/0911.1393>.
- [18] P. Comon, L.-H. Lim, Sparse representations and tensor decompositions, preprint, 2010.
- [19] E.J. Candès, The restricted isometry property and its implications for compressed sensing, C. R. Math. Acad. Sci. Paris 346 (9–10) (2008) 589–592.
- [20] J. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (10) (2004) 2231–2242.
- [21] P. Comon, Contrasts, independent component analysis, and blind deconvolution, Int. J. Adapt. Control Signal Process. 18 (3) (2004) 225–243.
- [22] P. Comon, C. Jutten (Eds.), Handbook of Blind Source Separation: Independent Component Analysis and Applications, Academic Press, New York, NY, 2010.
- [23] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, R. Baraniuk, Single-pixel imaging via compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 83–91.
- [24] N.D. Sidiropoulos, R. Bro, G.B. Giannakis, Parallel factor analysis in sensor array processing, IEEE Trans. Signal Process. 48 (8) (2000) 2377–2388.
- [25] R. Roy, A. Paulraj, T. Kailath, ESPRIT – A subspace rotation approach to estimation of parameters of cisoids in noise, IEEE Trans. Acoust. Speech Signal Process. 34 (5) (1986) 1340–1344.
- [26] F.L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, J. Math. Phys. 6 (1) (1927) 164–189.
- [27] L.-H. Lim, P. Comon, Nonnegative approximations of nonnegative tensors, J. Chemometrics 23 (7–8) (2009) 432–441.

- [28] M.V. Catalisano, A.V. Geramita, A. Gimigliano, Ranks of tensors, secant varieties of Segre varieties and fat points, *Linear Algebra Appl.* 355 (1–3) (2002) 263–285.
- [29] J.M. Landsberg, L. Manivel, On the ideals of secant varieties of Segre varieties, *Found. Comput. Math.* 4 (4) (2004) 397–422.
- [30] F.L. Zak, *Tangents and Secants of Algebraic Varieties*, AMS, Providence, RI, 1993.
- [31] J.-B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algorithms, I & II*, Springer-Verlag, Berlin, 1993.
- [32] J.G. Oxley, *Matroid Theory*, Oxford University Press, New York, NY, 1992.
- [33] J.B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, *Linear Algebra Appl.* 18 (2) (1977) 95–138.
- [34] A. Vardy, The intractability of computing the minimum distance of a code, *IEEE Trans. Inform. Theory* 43 (6) (1997) 1757–1766.
- [35] B.K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.* 24 (2) (1995) 227–234.
- [36] J.M. Landsberg, Kruskal's theorem, preprint, 2009, <http://arxiv.org/abs/0902.0543>.
- [37] A. Stegeman, N.D. Sidiropoulos, On Kruskal's uniqueness condition for the CANDECOMP/PARAFAC decomposition, *Linear Algebra Appl.* 420 (2–3) (2007) 540–552.
- [38] H. Derksen, Sharpness of Kruskal's theorem, preprint, 2009.