Deterministic Blind Identification in Antenna Array Processing

Souleymen Sahnoun, Pierre Comon

Deterministic Blind Identification in Antenna Array Processing

Souleymen Sahnoun and Pierre Comon

GIPSA-Lab, CNRS UMR5216,
Grenoble Campus, BP.46,
F-38402 St Martin d'Heres cedex, France

***Abstract*— The estimation of directions of arrival is formulated as the decomposition of a 3-way array into a sum of rank-one terms. However, a low-rank tensor approximation does not always exist. We propose an optimization technique based on differentiable angular constraints on the factors, ensuring the existence of the low-rank tensor decomposition. The efficiency of the proposed algorithm is demonstrated via numerical simulations, and compared to Cramér-Rao bounds.**

I. Introduction

Estimation of Directions of Arrival (DoA) is a central problem in antenna array processing, including in particular Radar, Sonar, or Telecommunications [1]. Traditional approaches are based on low-rank approximation of the covariance matrix of observations, and on detecting points of minimal distance with the so-called array manifold [2] [3]. These approaches hence assume that (i) the measurements are weakly stationary over sufficiently long observation lengths, (ii) the number of sources of interest is smaller than the number of sensors, and (iii) spatial responses of all sensors are known, and in particular their location (in other words, the sensor array needs to be calibrated).

In [4], a deterministic approach has been proposed, which permits not only to work with short data lengths (and hence less stationary sources), but also to localize more sources than sensors present in the reference array. This approach is based on the same rotational invariance as exploited in [5]. It consists in storing the measurements in a 3-way array, and to decompose it into a sum of rank-one terms. One very interesting by-product of [4] is that source copies are also delivered for free, without any further estimation stage.

In this paper we revisit the approach of [4], where an important fact has been neglected: a low-rank tensor approximation does not always exist, so that the latter approach is actually ill posed. This fact has been already pointed out in [6], and additional constraints have been suggested, which enjoy a reliable physical meaning and at the same time ensure existence of a solution. In the present contribution, we shed some light on conditioning and algorithmic issues.

Work funded by the European Research Council under the European Community's Seventh Framework Programme FP7/2007–2013 Grant Agreement no. 320594.

II. Modeling and Notation

In the following, vectors will be denoted by bold lowercases, *e.g.* $\mathbf{a}$, whereas matrices or higher-order arrays will be denoted by bold uppercases, *e.g.* $\mathbf{A}$. Moreover, $\mathbf{a}_r$ will denote the rth column of matrix $\mathbf{A}$.

Suppose that R narrow-band radiating sources impinge on an array of sensors, formed of L subarrays of K sensors each. We make the far-field assumption, that is, we assume that sources are located sufficiently far from the array, compared to the array dimensions, so that waves can be considered as plane. The key assumption made in [5], [4], [6] is that, taking one subarray as reference, every subarray can be deduced from the reference one by an unknown translation in space, defined by some vector $\boldsymbol{\delta}_\ell$ of $\mathbb{R}^3$, $1 \leq \ell \leq L$, $\boldsymbol{\delta}_1 \stackrel{\text{def}}{=} \mathbf{0}$.

Denote $\varsigma_r(t)$ the signal transmitted by the rth source, $\mathbf{d}_r$ its DoA viewed from the array, and $s(\mathbf{y},t)$ the signal measured at a point in space defined by its coordinates $\mathbf{y}$ (we consider complex envelopes about the central frequency). Then we have:

$$s(\mathbf{y},t) = \sum_{r=1}^{R} \varsigma_r(t)\, a_r(\mathbf{y}),\ a_r(\mathbf{y}) \stackrel{\text{def}}{=} \exp\{\jmath \frac{\omega}{C}\, \mathbf{y}^\mathsf{T} \mathbf{d}_r\} \qquad (1)$$

where ω is the central pulsation of the narrow-band waves, C the wave celerity, and $\jmath = \sqrt{-1}$. Because waves are plane and narrow-band, the signal measured at another point $\mathbf{y}+\boldsymbol{\tau}$, deduced from $\mathbf{y}$ by a translation $\boldsymbol{\tau}$ takes the form:

$$s(\mathbf{y},\boldsymbol{\tau},t) = \sum_{r=1}^{R} a_r(\mathbf{y})\, b_r(\boldsymbol{\tau})\, \varsigma_r(t),\ b_r(\boldsymbol{\tau}) \stackrel{\text{def}}{=} \exp\{\jmath \frac{\omega}{C}\, \boldsymbol{\tau}^\mathsf{T} \mathbf{d}_r\} \qquad (2)$$

In other words, function $s(\mathbf{y},\boldsymbol{\tau},t)$ decomposes into a sum of R simpler functions whose variables separate.

Now, if we discretize the $\mathbb{R}^3$ space with the above defined subarrays, and take M time samples, we end up with a multilinear relationship. In fact, let $\mathbf{p}_k$ be the coordinate vector of the kth sensor of the reference subarray, and $\boldsymbol{\delta}_\ell$ the translation defining the location of the ℓth subarray, $1 < \ell \leq L$. Then signal (2) can be stored in a $K \times L \times M$ three-way array, which follows the model below:

$$T_{k\ell m} = \sum_{r=1}^{R} \lambda_r\, A_{kr}\, B_{\ell r}\, S_{mr} \qquad (3)$$

where $A_{kr} = \frac{1}{\sqrt{K}} \exp\left(\jmath \frac{\omega}{C} \mathbf{p}_k^\mathsf{T} \mathbf{d}_r\right)$, $B_{\ell r} = \frac{1}{\sqrt{L}} \exp\left(\jmath \frac{\omega}{C} \boldsymbol{\delta}_\ell^\mathsf{T} \mathbf{d}_r\right)$, $S_{mr} = \varsigma_r(t_m)/\|\boldsymbol{\varsigma}_r\|$, and $\lambda_r = \sqrt{KL}\|\boldsymbol{\varsigma}_r\|$.

Model (3) is related to the Canonical Polyadic decomposition (CP)[1], which consists of decomposing a tensor $\mathbf{T}$ into a sum of *decomposable* tensors. For the sake of convenience, equation (3) is rewritten in vector form as

$$\mathbf{t} = \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{s}_r \tag{4}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{t} = \text{vec}\{\mathbf{T}\}$ is a column vector of dimension KLM containing the entries of the 3-way array $\mathbf{T}$.

III. Existence and Uniqueness

The goal is to identify the directions of arrival (DoA), θ_r, of the R impinging plane waves and to estimate corresponding transmitted source signals $\varsigma_r(t_m)$ up to a scale factor, given the whole array $\mathbf{T}$. Clearly, a sufficient condition is to be able to identify all parameters in the RHS of (4).

A. Low rank approximation

Actually, observations are corrupted by noise, so that (3-4) do not hold exactly. A natural idea is then to fit model (4) by minimizing the error

$$\Upsilon(\mathbf{A}, \mathbf{B}, \mathbf{S}; \boldsymbol{\Lambda}) = \left\| \mathbf{t} - \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{s}_r \right\|^2 \tag{5}$$

where $\boldsymbol{\Lambda}$ denotes the diagonal matrix containing the λ_r's. Minimizing error (5) means finding the best rank-R approximate of $\mathbf{T}$ and its CP decomposition. However, the infimum of Υ may not be reached; see *e.g.* [7], [6] and references therein. The reason is that the set of rank-R tensors is not closed for $R > 1$. The idea we promote here is to impose an additional constraint that will ensure the existence of a minimum, as elaborated in the next section. In addition, from the physical point of view, one can make the following observations:

- sources that are totally correlated need to be localized separately only if they are sufficiently well angularly separated. In that case they correspond to multi-paths.
- sources that are located in the same direction need to be estimated separately only if they are sufficiently decorrelated.
- otherwise, one can assimilate highly correlated sources arriving from close directions to a single *fat source*, spread out in space.

The purpose of the section is to formalize these constraints.

B. Coherences

As in the compressed sensing literature [8], [9], we define the *coherence of a set* of unit norm vectors as the largest value of cross scalar products:

$$\mu_A = \sup_{k \neq \ell} |\mathbf{a}_k^\mathsf{H} \mathbf{a}_\ell| \tag{6}$$

[1]also sometimes called Candecomp/Parafac in Psychometry.

The coherence of matrix $\mathbf{A}$ is defined this way, if $\mathbf{a}_k$ denote its (unit norm) columns. Coherences of matrices $\mathbf{B}$ and $\mathbf{S}$ are defined similarly, and denoted by μ_B and μ_S, respectively. Let $\mathbf{G}$ be the $R \times R$ Gram matrix defined by:

$$G_{pq} = (\mathbf{a}_p^\mathsf{H} \mathbf{a}_q)(\mathbf{b}_p^\mathsf{H} \mathbf{b}_q)(\mathbf{s}_p^\mathsf{H} \mathbf{s}_q)$$

Then for given matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}$, the optimal value $\boldsymbol{\Lambda}^o$ minimizing error Υ is obtained by cancelling the gradient of (5) w.r.t. $\boldsymbol{\Lambda}$, which leads to the linear system:

$$\mathbf{G}\, \boldsymbol{\lambda}^o = \mathbf{f}, \tag{7}$$

where $\boldsymbol{\lambda}^o = \text{diag}(\boldsymbol{\Lambda}^o)$ and vector $\mathbf{f}$ in the right hand side is defined by the contraction $f_r = \sum_{ijk} T_{ijk} \, A_{ir}^* B_{jr}^* S_{kr}^*$, $1 \leq r \leq R$. Equation (7) shows that coherences play a role in the conditioning of the minimization problem. Also note that only the *product* between coherences appears, and not coherences individually.

C. Existence

We are now in a position to state conditions of existence. It has been shown in [6] that if

$$\mu_A \mu_B \mu_S < \frac{1}{R-1} \tag{8}$$

then the infimum of (5) is reached. This happens because error (5) becomes coercive as soon as (8) is satisfied. And it must then reach its minimum since it is continuous.

Constraint (8) needs some care because it involves $\max$ operators, which are not differentiable. For this reason, we propose to use the fact that L^∞ norm can be bounded by $L^{2\rho}$ norms, and approximated for large values of ρ:

$$||\mathbf{z}||_\infty = \max_k \{z_k\} \leq ||\mathbf{z}||_{2\rho} = (\sum_k z_k^{2\rho})^{1/2\rho}, \; \forall \rho \geq 1$$

for $z_k \in \mathbb{R}^+$. Applying this inequality to $z_k \equiv \|\mathbf{a}_p^\mathsf{H} \mathbf{a}_q\|$ allows to bound coherences above by a differentiable quantity, so that another (somewhat more constraining) sufficient condition can be obtained. More precisely:

$$\mu_A \leq \mu(\mathbf{A}, \rho) \stackrel{\text{def}}{=} \left(\sum_{p<q} |\mathbf{a}_p^\mathsf{H} \mathbf{a}_q|^{2\rho} \right)^{1/2\rho}$$

We subsequently call $\mathcal{C}_\rho$ the constraint obtained by replacing the max operators by the $L^{2\rho}$ norms in constraint (8):

$$\mathcal{C}_\rho \stackrel{\text{def}}{=} 1 - R + \mu(\mathbf{A}, \rho)^{-1} \mu(\mathbf{B}, \rho)^{-1} \mu(\mathbf{S}, \rho)^{-1} > 0 \tag{9}$$

D. Uniqueness

There exist sufficient conditions ensuring that the solution of (4) is unique, which involve coherences [6]. However, the condition below is much less constraining [10, p.13]:

$$R \leq M \text{ and } R(R-1) \leq \frac{K(K-1)L(L-1)}{2} \tag{10}$$

and guarantees that there exists almost surely a unique solution. Other sufficient conditions for generic uniqueness exist [11], [12], but are less attractive when one dimension (*i.e.* M) is large.

IV. Computer results

A. Optimization

The constrained optimization is carried out with the help of gradient descent type algorithms, which handle constraints in different manners. Denote for conciseness $\mathbf{x} = \text{vec}\{[\mathbf{A}^\mathsf{T}, \mathbf{B}^\mathsf{T}, \mathbf{S}^\mathsf{T}]\}$ and define the objective function:

$$\mathcal{F}_\rho(\mathbf{x}; \boldsymbol{\lambda}) = \Upsilon(\mathbf{x}; \boldsymbol{\Lambda}) + \eta \, \exp(-\gamma \, \mathcal{C}_\rho(\mathbf{x})) \qquad (11)$$

where η is the penalty weight, γ is introduced to control the importance of penalty $\mathcal{C}_\rho(\mathbf{x})$, and $\boldsymbol{\lambda}$ is defined in (7) and depends on $\mathbf{x}$ and $\mathbf{t}$. This leads to the algorithm below

ALGORITHM

1) Choose R satisfying (10).
2) Initialize $(\mathbf{A}(0), \mathbf{B}(0), \mathbf{S}(0))$ to matrices with unit-norm columns satisfying $\mathcal{C}_\rho \geq 0$.
3) Compute $\mathbf{G}(0)$ and $\mathbf{f}(0)$, and solve $\mathbf{G}(0)\,\boldsymbol{\lambda}(0) = \mathbf{f}(0)$ for $\boldsymbol{\lambda}$, according to (7)
4) For $k \geq 1$ and subject to a stopping criterion, do
   a) Compute the descent direction as the gradient w.r.t. $\mathbf{x}$: $\mathbf{d}(k) = -\nabla \mathcal{F}_\rho(\mathbf{x}(k-1); \boldsymbol{\lambda}(k-1))$
   b) Compute a stepsize $\ell(k)$
   c) Update $\mathbf{x}(k) = \mathbf{x}(k-1) + \ell(k)\,\mathbf{d}(k)$
   d) Extract the 3 blocks from $\mathbf{x}(k)$: $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{S}(k)$
   e) Normalize the columns of $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{S}(k)$
   f) Compute $\mathbf{G}(k)$ and $\mathbf{f}(k)$, and solve $\mathbf{G}(k)\,\boldsymbol{\lambda}(k) = \mathbf{f}(k)$ for $\boldsymbol{\lambda}$, according to (7).

In the algorithm, η is decreased as the reconstruction error $\Upsilon(\mathbf{x}; \boldsymbol{\Lambda})$ decreases, whereas γ is kept fixed.

We give now some gradient expressions[2] necessary to determine the descent direction $\mathbf{d}(k)$ when $\mathcal{F}_\rho$ is used:

$$\frac{\partial \Upsilon}{\partial \mathbf{A}} = 2\,\mathbf{A}\mathbf{M}^A - 2\,\mathbf{N}^A$$

where $M^A_{pq} \stackrel{\text{def}}{=} \sum_{jk} \lambda_p B_{jp} S_{kp} S^*_{kq} B^*_{jq} \lambda^*_q$, $N^A_{ip} \stackrel{\text{def}}{=} \sum_{jk} T_{ijk} B^*_{jp} S^*_{kp} \lambda^*_p$ and

$$\frac{\partial \exp(-\gamma\,\mathcal{C}_\rho)}{\partial \mathbf{A}} = \frac{\gamma}{\exp(\gamma\,\mathcal{C}_\rho)}\,\mathcal{L}^A_\rho\,\mathbf{A}\left[(\mathbf{A}^\mathsf{H}\mathbf{A}) \boxdot \mathbf{Q}^A - \mathbf{I}\right]$$

where $\boxdot$ denotes the Hadamard entry-wise product,

$$\mathcal{L}^A_\rho \stackrel{\text{def}}{=} \left(\sum_{q<p} |\mathbf{a}_p^\mathsf{H}\mathbf{a}_q|^{2\rho}\right)^{\frac{-1}{2\rho}-1} \mu(\mathbf{B},\rho)^{-1}\mu(\mathbf{S},\rho)^{-1},$$

and $Q_{pq} \stackrel{\text{def}}{=} |\mathbf{a}_q^\mathsf{H}\mathbf{a}_p|^{2\rho-2}$. Keep in mind that expressions above hold true because matrix $\mathbf{A}$ has unit-norm columns. And expressions are similar for matrices $\mathbf{B}$ and $\mathbf{S}$, which also have unit-norm columns.

[2]Matrix gradients are written with the conventions described in [13], [14].

B. Advantage of the constraint $\mathcal{C}_\rho(\mathbf{x})$

To see the interest of constraint $\mathcal{C}_\rho(\mathbf{x})$ used in the optimization algorithm, Figure 3 sketches the evolution of the reconstruction error $\Upsilon(\mathbf{x}; \boldsymbol{\Lambda})$, and $\mathcal{C}_\rho(\mathbf{x})$ as a function of iterations. The figure shows that thanks to the constraint $\mathcal{C}_\rho$: (i) iterates are incited to remain/turn back to the feasible region (where existence is guaranteed), (ii) the optimization algorithm converges quickly because iterates are allowed to move away from the feasible region (depending on parameters η and γ).

C. Monte Carlo experiments

In this section, we asses the performances of the proposed algorithm using several sensor subarrays in two different configurations. The scenario on which the algorithm is tested can be of interest in various applications. For instance, consider sonar buoys left floating on the surface and equipped with a device permitting to maintain their orientation towards North. Each subarray is rigid and fixed on a buoy, but its relative location with respect to other subarrays is unknown.

The first configuration consists of two identical subarrays ($L = 2$) separated by a distance of 25λ (i.e. $\boldsymbol{\delta}_2 = [0, 25\lambda, 0]^T$), where $\lambda = \omega/2\pi C$ is the wavelength. The second configuration consists of three identical subarrays ($L = 3$) with the following translations: $\boldsymbol{\delta}_2 = [0, 25\lambda, 0]^T$, $\boldsymbol{\delta}_3 = [0, 37.5\lambda, 5\lambda]^T$. In both configurations, each subarray is an ULA array of 4-element with half-wavelength spacing. For the two configurations, we consider three equal-power narrowband source signals ($R = 3$) impinging respectively from $\theta_1 = 77°, \theta_2 = 55°$ and $\theta_3 = 5°$ (angles w.r.t endfire).

In all experiments, $\gamma = 5$, $M = 200$ time samples are used, and 50 Monte-Carlo simulations are run for each SNR level. The additive noise is complex-valued circular Gaussian. Figures 1 and 2 report the total root-mean square error (RMSE) of the DoA using respectively the first ($L = 2$) and the second ($L = 3$) configuration. The deterministic Cramér-Rao bound (CRB) is reported as a benchmark; exact expressions of DoA

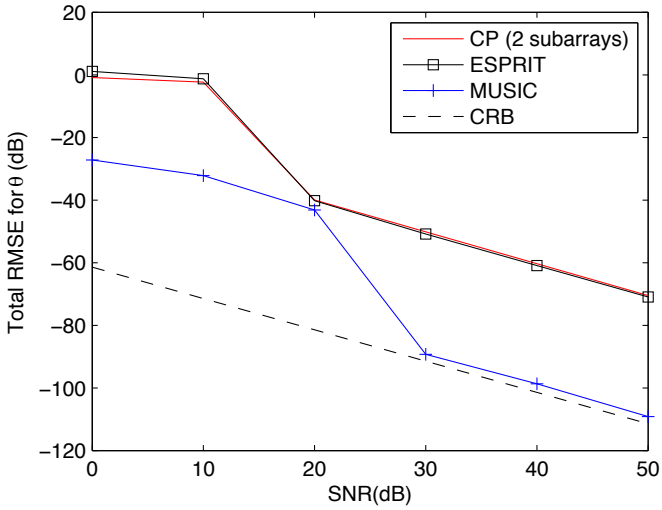


Fig. 1. Total DoA error versus SNR, with $L = 2$ subarrays.

CRB will be detailed in a full length paper. In fact, CRB can be found in [15] for factor matrices, *e.g.* $\mathbf{A}$, and in [16] for DoAs obtained with $L = 2$ subarrays, but DoA CRB are not available for $L > 2$.

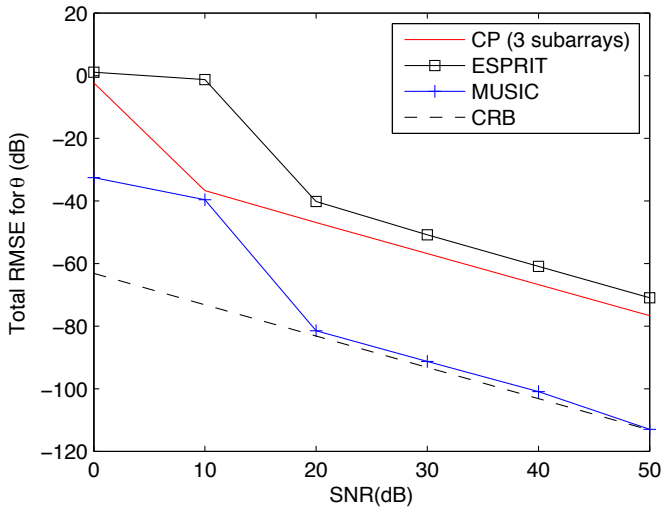
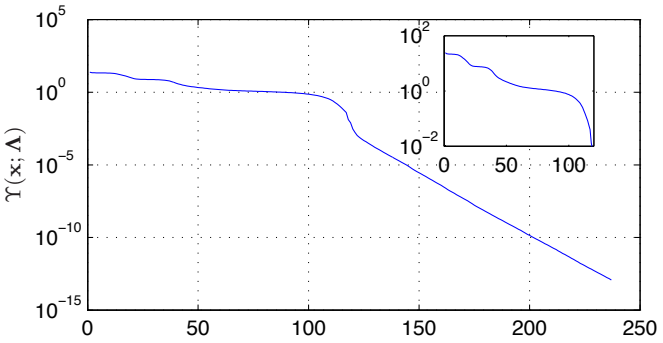
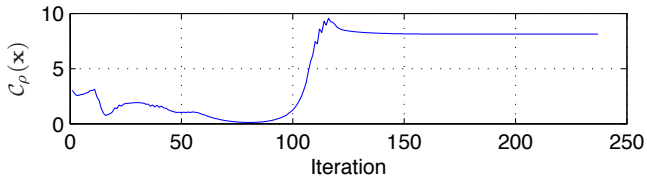


Fig. 2. Total DoA error versus SNR, with $L = 3$ subarrays.



(a) Error $\Upsilon(\mathbf{x};\mathbf{\Lambda})$. The insert shows a zoom between iterations 0 and 120



(b) Constraint $\mathcal{C}_\rho(\mathbf{x})$

Fig. 3. Reconstruction error $\Upsilon(\mathbf{x};\mathbf{\Lambda})$ and $\mathcal{C}_\rho(\mathbf{x})$ as a function of the number of iterations.

Configuration 1. This experiment shows that: (i) the proposed CP algorithm exhibits the same performances as ESPRIT, which makes sense, (ii) MUSIC performs the best, but exploits more information, namely the exact knowledge of sensor locations, whereas this information is actually not available in the present scenario. Hence MUSIC performances just serve as a reference.

Configuration 2. This experiment shows that the proposed algorithm yields better results than EPSRIT. The reason is that ESPRIT uses at most two subarrays, whereas the proposed algorithm uses all of them. Again, MUSIC is reported just as a reference benchmark.

V. Concluding remarks

The DoA estimation is formulated as a CP decomposition under the assumptions presented in § II on the sensor array formed of L identical subarrays. We proposed an optimization algorithm using a new differentiable constraint $\mathcal{C}_\rho(\mathbf{x})$ ensuring existence of the low-rank tensor approximation. It was shown that thanks to $\mathcal{C}_\rho(\mathbf{x})$ the proposed algorithm converges quickly and is prevented to leave for long the feasible region. DoA estimation results show that, as expected, the CP algorithm exhibits better results than reference DoA estimation methods when $L > 2$. Performances of signal estimation will be reported in a full length paper, as well as expressions of CRBs.

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Sig. Proc. Mag.*, pp. 67–95, July 1996.

[2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenna Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[3] H. L. Van Trees, *Optimum Array Processing*, vol. IV, Wiley, New York, 2002.

[4] 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.

[5] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 37, pp. 984–995, July 1989.

[6] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014, open access.

[7] P. Comon, "Tensors, usefulness and unexpected properties," in *15th IEEE Workshop on Statistical Signal Processing (SSP'09)*, Cardiff, UK, Aug. 31 - Sep. 3 2009, pp. 781–788, keynote. hal-00417258.

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

[9] 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.

[10] L. De Lathauwer, "A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM Journal on Matrix Analysis*, vol. 28, no. 3, pp. 642–666, 2006.

[11] M. V. Catalisano, A. V. Geramita, and A. Gimigliano, "Higher secant varieties of the Segre varieties," *Journal of Pure and Applied Algebra*, vol. 201, no. 1-3, pp. 367–380, 2005.

[12] H. Abo, G. Ottaviani, and C. Peterson, "Induction for secant varieties of Segre varieties," *Trans. Amer. Math. Soc.*, pp. 767–792, 2009, arXiv:math/0607191.

[13] P. Comon, "Estimation multivariable complexe," *Traitement du Signal*, vol. 3, no. 2, pp. 97–101, 1986.

[14] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2740–2746, June 2007.

[15] X. Liu and N. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Sig. Proc.*, vol. 49, no. 9, pp. 2074–2086, 2001.

[16] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Sig. Proc.*, vol. 39, no. 5, pp. 1122–1135, 1991.