Int. Conf. on Image and Signal Processing (ICISP), May 3-5, 2001, Agadir, Morocco, pp.20–32

From Source Separation to Blind Equalization
Contrast-based Approaches

Pierre Comon
I3S Laboratory, Algorithmes-Euclide-B, 2000 route des Lucioles
BP 121, F-06903 Sophia-Antipolis cedex, France
comon@unice.fr   www.i3s.unice.fr/~comon

Abstract

Blind deconvolution is currently the subject of growing interest, both in Telecommunications and Biomedical engineering. The reasons for this include a better robustness in specific situations, and an increase in throughput. In a companion paper [19], links with tensor decompositions have been analyzed. Here, we rather focus on the so-called contrast optimization criteria, ranging from the simplest one for Blind Source Separation to the more complicated for Blind Multichannel System Equalization. It is shown that many of the instances of this problem can be solved analytically, often with the help of standard tools such as Eigen Value or Singular Value Decompositions. Other problems that remain inherently of tensor nature are also pointed out.

1 Introduction

1.1 A glance at the state of the art.

The subject of Blind Source Separation (BSS) and Blind System Equalization (BSE) are closely related, and have raised a growing interest since the first papers appeared on the subject around 1989 [29] [1] [4] [42] [36] [7] [63] [55] [12] [37]. Beside the early papers, there has been a abundant literature on the subject [8], which will not be reported here exhaustively. On the other hand, we shall concentrate on the so-called contrast-based approaches, that have the advantage to enjoy some optimality in the presence of additive noise with unknown statistics. In addition, contrast-based approaches of BSS are not well known, and some misunderstanding can be found in the literature, *e.g.* [56] [46].

Even if they have not been called so, contrasts have been first introduced by Donoho [29] for SISO blind equalization. Later, contrasts were introduced for the static blind source separation [14] [15]. Several generalizations have been then proposed, in particular by breaking the symmetry of the variables in the definition of the contrast itself [61] [50]; links with the Joint Approximate Diagonalization of matrices [9] are established in [45].

Blind identification or equalization of MIMO linear systems has been addressed also in the nineties [64] [13] [59] [65], essentially by cumulant matching. The introduction of *contrasts* devoted to MIMO blind equalization of linear dynamic systems is more recent [17] [49] [58] [48]; in particular, a deflation procedure dedicated to non-linear source processes is proposed in [57]; more general forms of contrasts are derived in [47]; links with Joint Approximate Diagonalization of matrices (JAD) are established in [23].

1.2 Contents.

The basic hypotheses that are usually assumed in BSS and BSE problems are first stated. Then we review the definitions of contrast criteria, and give a number of examples. Numerical *analytical* algorithms are listed in detail in a third section. Eventually follows a brief account on special purpose algorithms, dedicated to discrete sources, or to the case of underdetermined mixtures (more sources than sensors). The algorithms presented in this last section are not based on contrast.

1.3 Modeling.

In the present framework, we consider the linear statistical model below:

$$\boldsymbol{y}(n) = [\boldsymbol{H} \star \boldsymbol{x}](n) + \boldsymbol{v}(n) \tag{1}$$

where $\boldsymbol{y}(n)$ is a discrete-time process of dimension K, $\{\boldsymbol{H}(n)\}$ is the $K \times P$ matrix-valued impulse response of a stable and invertible linear time-invariant filter $\boldsymbol{H}[z]$, and $\boldsymbol{x}(n)$ is a so-called source process, of dimension P; $\boldsymbol{v}(n)$ stands for modeling and background noises, and is assumed to be statistically independent of $\boldsymbol{x}(n)$. With linear system terminology, $\boldsymbol{y}(n)$ is the output, and $\boldsymbol{x}(n)$ the input. In this paper we are mainly interested in *blind* identification and/or equalization methods, in the sense that only the output is observed; in other words, from a finite set of observations $\{\boldsymbol{y}(n), 1 \leq n \leq N\}$, it is desired to estimate the impulse response $\boldsymbol{H}(n)$ (blind identification), or to extract the source process $\boldsymbol{x}(n)$ (blind equalization). Throughout the paper, boldfaced letters denote order-one arrays, and uppercase boldfaced letters arrays with two indices or more. Scalar quantities are denoted in plain font.

There are two special instances of this problem. The first one is the Single Input Single Output (SISO) problem, in which $P = 1 = K$. The SISO blind identification problem has been studied since the eighties [29] [3] [42] [53] [35].

The second instance is the static case, where the impulse response reduces to its first coefficient, $\boldsymbol{H}(0)$. This problem is meaningful only when $K > 1$, and the corresponding model is written as:

$$\boldsymbol{y}(n) = \boldsymbol{H}\, \boldsymbol{x}(n) + \boldsymbol{v}(n) \tag{2}$$

Furthermore, the problem is interesting (*i.e.* not trivial) mainly when it is impossible to separate the sources $x_i(n)$ by sole spectral filtering, that is, when the source power spectra do not form a set of linearly independent functions; this occurs for instance when sources are i.i.d. processes. We shall say that the source process has no *time coherency*. Some references can be found in [62] [2] [69] [32] [39].

More generally, the model (2) is meaningful in many other contexts, and in particular in factor analysis. In the latter, index n does not denote necessarily time, but just a realization number. This means that indices n can be sorted differently without changing anything in the problem. There is a whole literature on the subject, and the reader can refer to [38] [10] [19] and references therein.

1.4 Assumptions.

In view of equation (1), it is clear that the problem is ill-posed, for there are infinitely many ways of finding a pair $(\boldsymbol{H}, \boldsymbol{x})$ such that $\boldsymbol{y}(n) \approx [\boldsymbol{H} \star \boldsymbol{x}](n)$. In the frame of source separation, the key property usually assumed is that [15]:

A1. The sources $x_p(n)$ are statistically mutually independent.

However, contrary to what is sometimes believed, this assumption is not necessary but only sufficient. Another sufficient assumption more rarely encountered is that

A1'. The sources $x_p(n)$ have a discrete distribution.

For the time being, let us consider assumption (**A1**). Denote $\boldsymbol{H}[z]$ the $z-$transform of the impulse response $\boldsymbol{H}(n)$. Looking more carefully at (1) shows that if the pair $(\boldsymbol{H}[z], \boldsymbol{x}(n))$ is solution, then so is the pair $(\boldsymbol{H}[z]\boldsymbol{D}[z]^{-1}\boldsymbol{P}^{-1}, \boldsymbol{P}[\boldsymbol{D} \star \boldsymbol{x}](n))$, where $\boldsymbol{D}[z]$ is a diagonal filter and $\boldsymbol{P}$ a constant permutation. In fact, if components of $\boldsymbol{x}(n)$ are independent, then so are those of $\boldsymbol{P}\,[\boldsymbol{D} \star \boldsymbol{x}](n)$.

As a consequence, there is a whole equivalence class of solutions, and not a single one. Depending on the application, this inherent indeterminacy can (or cannot) raise difficulties. Nevertheless, it is often possible to impose reasonable additional constraints in practice in order to get rid of indeterminacies, if necessary. As far as we are concerned, we shall assume that:

A2. Sources $x_i(n)$ are i.i.d. processes of unit variance

This assumption does not restrict the generality of the problem, when sources are *linear processes*. In fact, assumption **A2** considers that the sources are the driving noises of the linear processes; if the linear processes themselves are searched for, then one would alternatively assume that

A2'. The diagonal entries of $\boldsymbol{H}[z]$ are equal to 1.

Assumptions **A2** and **A2'** are equivalent in the sense that they lead to the same equivalence class of solutions. Under assumptions **A1** and **A2**, the only acceptable diagonal filters $\boldsymbol{D}[z]$ have entries of the form $\boldsymbol{D}_{pp}[z] = \lambda_p\, z^{m_p}$; these are indeed the only filters that preserve statistical independence between variables $x_p(n)$ for different p or n. In other words, $\boldsymbol{D}[z]$ induces integer delay and scale factors only.

1.5 Preprocessing.

Under assumption **A1**, the estimated sources are wished to be mutually independent. Under **A2**, this means that their cross-correlation matrix is $\boldsymbol{\Gamma}_{xx}(\tau) = \delta(\tau)\boldsymbol{I}$, $\forall \tau \in \mathbb{Z}$, among others. One first step (not necessary in noisy environment) towards this goal is to implement a filter of impulse response $\boldsymbol{T}[z] = \boldsymbol{\Gamma}_{yy}[z]^{-1/2}$ (for instance minimum phase). Then, the output of the preprocessing, $\bar{\boldsymbol{y}} = \boldsymbol{T} \star \boldsymbol{y}$, has a correlation matrix of the required form, $\delta(\tau)\boldsymbol{I}$. But it does not mean that the sources are separated. Indeed, the mixture is still undefined up to a para-unitary filter, $\boldsymbol{U}[z]$, satisfying $\boldsymbol{U}[z]\,\boldsymbol{U}[1/z^*]^{\mathrm{H}} = \boldsymbol{I}$. This is easily seen when computing the covariance matrix of the output $\boldsymbol{s} = \boldsymbol{U} \star \bar{\boldsymbol{y}}$ in the $z-$domain: $\boldsymbol{\Gamma}_{ss}[z] = \boldsymbol{U}[z]\,\boldsymbol{\Gamma}_{\bar{y}\bar{y}}\,\boldsymbol{U}[1/z^*]^{\mathrm{H}}$. Since $\boldsymbol{\Gamma}_{\bar{y}\bar{y}} = \boldsymbol{I}$, so is $\boldsymbol{\Gamma}_{ss}[z]$. Random variables with unit covariance will be referred to as (spatially) standardized, and random processes with correlation function $\delta(\tau)\boldsymbol{I}$ will be referred to as *spatially* and *temporally* second-order white.

This preprocessing sometimes needs a dimension reduction, *e.g.* when the number of sensors, K, exceeds the number of sources, P, and when the noise is absent. Indeed in that case, matrix $\boldsymbol{\Gamma}_{yy}[z]$ is not invertible, but there still exists a transform $\boldsymbol{T}[z]$ such that $\boldsymbol{\Gamma}_{\bar{y}\bar{y}} = \boldsymbol{T}[z]\,\boldsymbol{\Gamma}_{yy}[z]\,\boldsymbol{T}[1/z^*]^{\mathrm{H}}$ is the identity matrix of size P. So when $P \leq K$, we can assume that the observation has a unit covariance, and that the pre-processing led to the equivalent problem where $P = K$. Note that, by doing so, we may loose the optimality of the whole processing; but we shall make this assumption in the remainder, unless otherwise specified, in order to ease the presentation.

As a conclusion, the (non unique) pre-processing has allowed to whiten the output at the second order, but there remains to find a para-unitary separating filter, $\boldsymbol{U}[z]$. We shall subsequently see (section 2.2) that this filter can be found only thanks to some non Gaussian character of the source vector, $\boldsymbol{x}$, and hence an extraneous assumption.

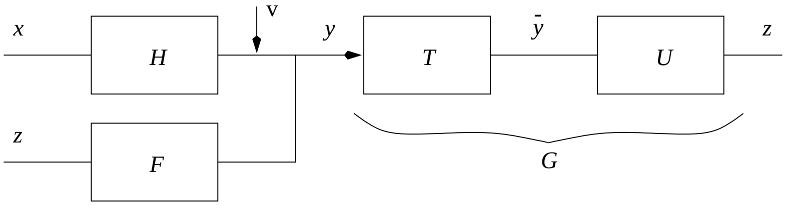

Figure 1: The usual processing line

1.6 BSS and BSE schemes.

Contrast-based approaches are *direct* in the sense that they aim at estimating directly the source vector $\boldsymbol{x}$ by searching for an equalizer $\boldsymbol{G}(n)$ whose output, $\boldsymbol{z} = \boldsymbol{G} \star \boldsymbol{y}$, is as close as possible to the original unknown source vector, $\boldsymbol{x}$, up to indeterminacies inherent in the problem. For this purpose, the best equalizers are the maxima of some output optimization criterion, called contrast function.

On the other hand, *indirect* approaches (that can be based on cumulant matching) aim at identifying the mixing filter $\boldsymbol{H}(n)$. The source inputs are estimated in a second stage, from the estimate $\boldsymbol{F}(n)$ of $\boldsymbol{H}(n)$. For instance, the simplest source separation algorithm in dimension $K = 2$ in the noiseless case proceeds by cumulant matching [39] [13].

As an illustration, the Constant Modulus Algorithms (CMA) are necessarily direct, contrary to algorithms based on the discrete character of the source distribution, that can be either direct or indirect, as will be seen in section 4.1.

2 Contrast functions

Under assumption **A1** that the sources are independent, it is natural to choose an optimization criterion that aims at restoring independence at the output. If the noise distribution is known, one can follow a maximum-likelihood approach [52]. If it is unknown (and this is often the case), one must resort to independence measures. The first tool proposed has been the kurtosis, in the frame of SISO blind identification. In his exceptional paper [29], Donoho demonstrates that there exist an order relation between processes, and that equalizing means maximizing some distance to Gaussianity. This tool has been rediscovered independently ten years later in the context of static mixtures [14] [15]; in the latter approach, the optimization criteria were called *contrasts*, and were based on high-order cumulants. Even if they were not so named in [29], the optimization criteria were actually of the same nature.

2.1 Cumulants

The moments of a $K-$dimensional random variable $\boldsymbol{x}$ with components x_i are defined as the coefficients in the Taylor expansion of its characteristic function, $\Phi_x(\boldsymbol{u})$. The cumulants are the coefficients in the expansion of $\log \Phi_x$ (this is possible because Φ_x is continuous and $\Phi_x = 1$ at the origin). Therefore, moments and cumulants are related to each other. For instance, at order four, and for a zero-mean random variable $\boldsymbol{x}$:

$$\mathcal{C}^{\boldsymbol{x}}_{ijk\ell} \stackrel{\text{def}}{=} \text{Cum}\{x_i, x_j, x_k, x_\ell\} = \mathcal{M}^{\boldsymbol{x}}_{ijk\ell} - \mathcal{M}^{\boldsymbol{x}}_{ij}\mathcal{M}^{\boldsymbol{x}}_{k\ell} - \mathcal{M}^{\boldsymbol{x}}_{ik}\mathcal{M}^{\boldsymbol{x}}_{j\ell} - \mathcal{M}^{\boldsymbol{x}}_{i\ell}\mathcal{M}^{\boldsymbol{x}}_{jk} \tag{3}$$

where

$$\mathcal{M}^{\boldsymbol{x}}_{ij} \stackrel{\text{def}}{=} \text{E}\{x_i\, x_j\},\ \mathcal{M}^{\boldsymbol{x}}_{ijk\ell} \stackrel{\text{def}}{=} \text{E}\{x_i\, x_j\, x_k\, x_\ell\}.$$

General expressions can be generated with the help of specific notations [43], but we shall not reproduce them here. Let us also introduce compact notations for moments and cumulants of complex variables:

$$\mathcal{M}^{\boldsymbol{x}}_{ij,k\ell} \stackrel{\text{def}}{=} \text{E}\{x_i\, x_j\, x_k{}^*\, x_\ell{}^*\}\,, \quad \mathcal{C}^{\boldsymbol{x}}_{ij,k\ell} \stackrel{\text{def}}{=} \text{Cum}\{x_i, x_j, x_k{}^*, x_\ell{}^*\} \tag{4}$$

and:

$$C^{x_k}_{p,q} \stackrel{\text{def}}{=} \text{Cum}\{\underbrace{x_k, \dots x_k}_{p \text{ times}}, \underbrace{x_k{}^*, \dots x_k{}^*}_{q \text{ times}}\}\,, \quad C^{\boldsymbol{x}}_{p,q} \stackrel{\text{def}}{=} \mathbf{vec}\{C^{x_k}_{p,q},\, 1 \le k \le K\} \tag{5}$$

Moments and cumulants enjoy the *multi-linearity* property, which can be summarized by the relation:

$$\text{Cum}\{\alpha\, X + \beta\, Y,\, Z_2, \dots Z_r\} = \alpha\, \text{Cum}\{X,\, Z_2, \dots Z_r\} + \beta\, \text{Cum}\{Y,\, Z_2, \dots Z_r\} \tag{6}$$

for any pair of deterministic parameters (α, β). The interest in using cumulants instead of moments lies in the following property, that we shall refer to as the *independence property* : if a set of random variables $\{x_k\}$, $1 \le k \le K$, can be partitioned into two statistically independent subsets, then their joint cumulants are null. As a corollary, we also have that:

$$\text{Cum}\{x_1 + y_1,\, x_2 + y_2,\, \dots x_r + y_r\} = \text{Cum}\{x_1,\, x_2,\, \dots x_r\} + \text{Cum}\{y_1,\, y_2,\, \dots y_r\} \tag{7}$$

whenever variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are statistically independent. This property is not satisfied by moments.

2.2 Contrasts

Trivial filters. Denote $\mathcal{T}$ the set of so-called trivial filters, of the form $\boldsymbol{PD}[z]$, with the notation of subsection 1.4. In other words, any filter of $\mathcal{T}$ is a combination of a diagonal filter of pure delays, a constant scale diagonal matrix, and a constant permutation. As was already pointed out in an earlier section, a pair of solution $(\boldsymbol{H}, \boldsymbol{x})$ can be obtained only up to filters in $\mathcal{T}$.

Contrast. Denote $\mathcal{X} = \{\boldsymbol{x}(n)\}$ the set of source processes, $\mathcal{H} = \{\boldsymbol{H}(n)\}$ the set of filters entering model (1), and $\mathcal{Y}$ the set of associated observed processes. We assume throughout that filters in $\mathcal{H}$ have a finite L^2 norm, that the identity is a member of $\mathcal{H}$, and that $\mathcal{H}$ is stable under convolution (finitely many times). For instance, picking up a separating filter $\boldsymbol{G}(n)$ in $\mathcal{H}$ means that $\boldsymbol{z} = \boldsymbol{G} \star \boldsymbol{y}$ is in $\mathcal{Y}$. In addition, unless otherwise specified, we assume that processes in $\mathcal{X}$ are stationary up to the required order, and that the cumulants of required order (depending on the contrast definition) are always finite.

A contrast Υ is a mapping from the observation set $\mathcal{Y}$ onto the set of real positive numbers, $\mathbb{R}^+$; $\Upsilon(\cdot)$ depends only on the statistics of $\boldsymbol{y}$, and satisfies the properties below:

1. *Invariance.* Υ is invariant under the action of any trivial filter:

$$\forall \boldsymbol{G} \in \mathcal{T}, \forall \boldsymbol{z} \in \mathcal{Y},\ \Upsilon(\boldsymbol{G} \star \boldsymbol{z}) = \Upsilon(\boldsymbol{z}) \tag{8}$$

2. *Domination.* The contrast should be maximal when sources are separated:

$$\forall \boldsymbol{G} \in \mathcal{H}, \forall \boldsymbol{z} \in \mathcal{X}, \Upsilon(\boldsymbol{G} \star \boldsymbol{z}) \leq \Upsilon(\boldsymbol{z}) \tag{9}$$

3. *Discrimination.* In the absence of noise, all the maxima of the contrast should yield a solution:

$$\forall \boldsymbol{z} \in \mathcal{X}, \Upsilon(\boldsymbol{G} \star \boldsymbol{z}) = \Upsilon(\boldsymbol{z}) \Rightarrow \boldsymbol{G} \in \mathcal{T} \tag{10}$$

Thus by construction, all the absolute maxima of a contrast function are equivalent, because they belong to the same equivalence class spanned by trivial filters of $\mathcal{H} \cap \mathcal{T}$. Unfortunately, like most usual optimization criteria, this does not mean that there does not exist spurious local maxima, even in the noiseless case. With some abuse of notation, we shall frequently write $\Upsilon(\boldsymbol{U})$ instead of $\Upsilon(\boldsymbol{U}\bar{\boldsymbol{y}})$.

Example 1. Let $\mathcal{H}_s$ be the set of invertible constant matrices, and $\mathcal{X}_s$ be the set of $K-$dimensional random variables with unit covariance and with independent components, whose at most one is Gaussian. Then the opposite of the mutual information:

$$\Upsilon_\infty(\boldsymbol{U}) = -\mathcal{I}(\boldsymbol{U}\bar{\boldsymbol{y}}) = -\int f_{\boldsymbol{z}}(\boldsymbol{u}) \log \frac{f_{\boldsymbol{z}}(\boldsymbol{u})}{\prod_i f_{z_i}(u_i)}\, d\boldsymbol{u} \tag{11}$$

is a contrast over $\mathcal{Y}_s$, the subset of $K-$dimensional random variables with finite and invertible covariance, spanned by $\mathcal{X}_s$ under the action of $\mathcal{H}_s$.

Example 2. Let $\mathcal{X}_s^{(r)}$ be the set of $K-$dimensional random variables with invertible diagonal covariance matrix, with finite marginal cumulants of order r, $C^{x_p}_{r-q,q}$, whose at most one is null, and with null cross cumulants of order r. Then, since the mixing and separating matrices are both imposed to belong to $\mathcal{H}_s$, as defined previously, the following optimization criterion is a contrast, for any fixed triplet (r, q, α):

$$\Upsilon^{(\alpha)}_{r-q,q}(\boldsymbol{U}) = \sum_{p=1}^{P} |C^{z_p}_{r-q,q}|^\alpha,\ \alpha \geq 1,\ r > 2,\ 1 \leq q \leq r. \tag{12}$$

This contrast family has been shown to be extendible to any convex function of cumulants, even non symmetric in its P arguments [50].

Example 3. Under the same hypotheses as in the previous example, if $r = 2q$ is even and if source marginal cumulants $C^{x_p}_{q,q}$ have the same sign ϵ, then Moreau showed [44] that the absolute value can be dropped, and that the following is also a contrast:

$$\Upsilon^{(1)}_{q,q}(\boldsymbol{U}) = \epsilon \sum_{p=1}^{P} C^{z_p}_{q,q}\,,\ q > 1 \tag{13}$$

Expression as a function of $\boldsymbol{U}$. In the contrast expressions above, the separating matrix $\boldsymbol{U}$ that we are searching for does not explicitly appear. In order to do this, it suffices to replace the output cumulants as functions of the observation cumulants with the help of the multilinearity property (6). Let us take the example 3 with $r = 4$ to make it simple. The contrast is then the sum of the output *kurtoses*. It can be rewritten as a function of $\boldsymbol{U}$:

$$\Upsilon^{(1)}_{2,2}(\boldsymbol{U}) = \epsilon \sum_{p=1}^{P} \sum_{ijk\ell} U_{pi}\, U_{pj}\, U^*_{pk}\, U^*_{p\ell}\, \mathcal{C}^{\boldsymbol{y}}_{ij,k\ell}$$

Example 4. Now consider dynamic mixtures. As noticed by Donoho [29], the output kurtosis is an admissible contrast for SISO deconvolution. In the MIMO case, one can combine the results obtained above with those valid for SISO dynamic systems [17]. In particular, let $\mathcal{X}_D$ be the set of stochastic processes of dimension P, stationary up to order r, with null cross cumulants, having at most one null marginal cumulant of order r, $C^{x_p}_{r-q,q}$. Also define $\mathcal{H}_D$, the set of para-unitary filters, and $\mathcal{Y}_D$ the set of processes spanned by $\mathcal{X}_D$ under the action of $\mathcal{H}_D$. Then the following mapping is again a contrast, for any fixed triplet (r, q, α):

$$\Upsilon^{(\alpha)}_{r-q,q}(\boldsymbol{U}) = \sum_{p=1}^{P} |C^{z_p}_{r-q,q}|^{\alpha},\ \alpha \geq 1,\ r > 2,\ 1 \leq q \leq r. \tag{14}$$

This kind of result has been extended lately to arbitrary convex functions [47] and to non-symmetric mappings [48].

2.3 Links with joint diagonalization

Despite their formal sub-optimality, the Joint Approximate Diagonalization (JAD) formulations have the advantage to yield more conventional numerical algorithms, insofar as they belong to linear algebra (but there exist other contrasts yielding simple solutions, as will be reported in section 3). Some instances of non symmetric contrast functions can be linked to the algebraic problem of JAD, and are worth mentioning in more detail.

Satic mixtures. The first one, proposed in [9], is inspired from the fact that the sum of all squared cumulants is invariant by unitary transform [14]. In other words, particularizing the reasoning to $\Upsilon^{(2)}_{2,2}$, the quantity:

$$\Omega = \sum_{ijk\ell} |\mathcal{C}^{\boldsymbol{y}}_{ij,k\ell}|^2 = \sum_{ijk\ell} |\mathcal{C}^{\boldsymbol{z}}_{ij,k\ell}|^2$$

is a constant under the action of $\boldsymbol{U} \in \mathcal{H}$. This means that minimizing all the squared moduli of *cross-cumulants* of $\boldsymbol{z}$ amounts to maximizing the sum of the marginal ones, $\Upsilon^{(2)}_{2,2}$. In [9], only part of the set of cross-cumulants is minimized in order to deflate the original problem (of tensor nature) to a joint matrix diagonalization. The consequence is that the equivalent contrast now contains more terms than in (12):

$$\Upsilon^{JAD}_{2,2}(\boldsymbol{U}) = \sum_{ijk} |\mathcal{C}^{\boldsymbol{z}}_{ij,ik}|^2. \tag{15}$$

The interest in doing so appears when expressing this quantity as a function of $\boldsymbol{U}$. First use the multilinearity property, based on the definition $\boldsymbol{z} = \boldsymbol{U}\,\bar{\boldsymbol{y}}$:

$$\Upsilon^{JAD}_{2,2}(\boldsymbol{U}) = \sum_{ijk} |\sum_{pqmn} U_{ip}\, U_{jq}\, U^*_{im}\, U^*_{kn}\, \mathcal{C}^{\boldsymbol{y}}_{pq,mn}|^2$$

Because of the unitarity of $\boldsymbol{U}$, the expansion of the square modulus simplifies. After some manipulations, we get:

$$\Upsilon^{JAD}_{2,2}(\boldsymbol{U}) = \sum_{qn} \sum_{i} |\sum_{pm} U_{ip} \mathcal{C}^{\boldsymbol{y}}_{pq,mn} U^*_{im}|^2 \tag{16}$$

It can now be seen that this criterion is nothing else but that of a Joint Approximate Diagonalization of a set of matrices $\boldsymbol{N}(q,n)$ whose entries are $N(q,n)_{pm} = \mathcal{C}^{\bar{\boldsymbol{y}}}_{pq,mn}$, by a unitary change of basis defined by matrix $\boldsymbol{U}$. In fact, this contrast selects all the squares of diagonal terms of every matrix $\boldsymbol{U}\boldsymbol{N}(q,n)\,\boldsymbol{U}^{\mathrm{H}}$.

This type of contrast can also be defined for even orders $r = 2q > 2$ when data are in the complex field, even if it is given here only for $r = 4$. Further extensions to JAD of tensors of order higher than 2 are proposed in [25] [45]. For instance, in [25, ch.9], the STOTD contrast takes the form:

$$\Upsilon^{STO}_{2,2}(\boldsymbol{U}) = \sum_{pq} |\mathcal{C}^{\boldsymbol{z}}_{pq,pp}|^2 \tag{17}$$

Dynamic mixtures. Now, despite much more cumbersome calculations and heavier notations, this type of result, derived for BSS, can be extended to BSE [23]. We just report here the contrast expression eventually obtained:

$$\Upsilon^{JAD}_{2,r-2}(\boldsymbol{U}[z]) = \sum_i \sum_{\boldsymbol{j}} \sum_{\boldsymbol{\ell}} |\mathcal{C}^{\boldsymbol{z}}_{2,r-2}[i,\boldsymbol{j},\boldsymbol{\ell}]|^2, \tag{18}$$

where $\mathcal{C}^{\boldsymbol{z}}_{2,r-2}[i,\boldsymbol{j},\boldsymbol{\ell}]$ are $r-$th order cumulant multi-correlation slices defined as

$$\mathcal{C}^{\boldsymbol{z}}_{p,q}[i,\boldsymbol{j},\boldsymbol{\ell}] \stackrel{\text{def}}{=} \text{Cum}\{\underbrace{z_i(n),\ldots,z_i(n)}_{p\ \text{terms}}, \underbrace{z_{j_1}(n-\ell_1),\ldots,z_{j_q}(n-\ell_q)}_{q=r-p\ \text{terms}}\}$$

and where $\boldsymbol{j} = (j_1,\ldots j_q) \in \{1,\ldots,P\}^q$ and $\boldsymbol{\ell} = (\ell_1,\ldots \ell_q) \in \mathbb{Z}^q$. One can then show that this contrast can be formulated as a JAD criterion, similar to that of the previous paragraph:

$$\Upsilon^{JAD}_{2,r-2}(\boldsymbol{U}[z]) = \sum_{\boldsymbol{b}} \sum_{\gamma} ||\mathbf{Diag}\{\boldsymbol{V}\,\boldsymbol{N}(\boldsymbol{b},\gamma)\,\boldsymbol{V}^{\mathrm{H}}\}||^2 \tag{19}$$

where $\boldsymbol{V}$ is a $N \times NL$ semi-unitary matrix, if L denotes the number of matrix taps in the MIMO equalizer.

3 General-purpose algorithms in the square case

Since the proposal made in [12], consisting of processing the entries in a tensor by pairs, in the same manner as in the Jacobi algorithm for diagonalizing matrices, several efficient numerical algorithms have been published, including [15] [9] [25]. This sweeping procedure allows to reduce the dimension of the tensor diagonalization problem down to 2.

So let's consider the 2×2 BSS problem, and write the output as:

$$\boldsymbol{z} = \begin{pmatrix} c & s\,e^{\jmath\phi} \\ -s\,e^{-\jmath\phi} & c \end{pmatrix} \bar{\boldsymbol{y}} \tag{20}$$

where $c = \cos\alpha$ is real positive, $s = \sin\alpha$ is real, and $\alpha,\ \phi \in [-\pi/2,\ \pi/2[$. Parameters α and ϕ fully characterize the 2×2 unitary matrix $\boldsymbol{U}$.

Analytical maximization of $\Upsilon^{(1)}_{r,0}$ in the real 2×2 case. In the real case, matrix $\boldsymbol{U}$ depends on a single parameter α. In that particular case, finding the extrema of the contrast $\Upsilon^{(1)}_{4,0}$ amounts to maximizing a quadratic form in dimension 2, since, after some manipulations, we can show that $\mathcal{C}^{\boldsymbol{z}}_{1111} + \mathcal{C}^{\boldsymbol{z}}_{2222}$ reduces to:

$$\Upsilon^{(1)}_{4,0}(\boldsymbol{U}) = \epsilon(\cos 2\alpha,\, \sin 2\alpha) \begin{pmatrix} \mathcal{C}^{\bar{\boldsymbol{y}}}_{1111} + \mathcal{C}^{\bar{\boldsymbol{y}}}_{2222} & \mathcal{C}^{\bar{\boldsymbol{y}}}_{1112} - \mathcal{C}^{\bar{\boldsymbol{y}}}_{1222} \\ \mathcal{C}^{\bar{\boldsymbol{y}}}_{1112} - \mathcal{C}^{\bar{\boldsymbol{y}}}_{1222} & \frac{1}{2}(\mathcal{C}^{\bar{\boldsymbol{y}}}_{1111} + \mathcal{C}^{\bar{\boldsymbol{y}}}_{2222}) + 3\,\mathcal{C}^{\bar{\boldsymbol{y}}}_{1122} \end{pmatrix} \begin{pmatrix} \cos 2\alpha \\ \sin 2\alpha \end{pmatrix} \tag{21}$$

The dominant eigenvector, normed to one, yields immediately α up to an indetermination of π, inherent in the BSS problem. In dimension 2, this vector can be computed analytically [16].

Analytical maximization of $\Upsilon^{(2)}_{r,0}$ in the real 2×2 case. Now when $\alpha > 1$, the quartic form is no longer quadratic. In fact, it can be shown that:

$$\Upsilon^{(2)}_{4,0} \stackrel{\text{def}}{=} \sum_{p=1}^{2} (C^{z_p}_{4,0})^2 = \boldsymbol{v}^{\mathrm{T}} \boldsymbol{A}\, \boldsymbol{v} + \boldsymbol{w}^{\mathrm{T}} \boldsymbol{B}\, \boldsymbol{w} \tag{22}$$

where $\boldsymbol{v}^{\mathrm{T}} = [\cos 4\alpha,\ \sin 4\alpha]$, $\boldsymbol{w}^{\mathrm{T}} = [\cos 2\alpha,\ \sin 2\alpha]$, and, denoting for compactness $(\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5) \stackrel{\text{def}}{=} (\mathcal{C}^{\boldsymbol{y}}_{1111}, \mathcal{C}^{\boldsymbol{y}}_{1112}, \mathcal{C}^{\boldsymbol{y}}_{1122}, \mathcal{C}^{\boldsymbol{y}}_{1222}, \mathcal{C}^{\boldsymbol{y}}_{2222})$:

$$\begin{aligned}
A_{11} &= \tfrac{1}{16}\left[\kappa_1 + \kappa_5\right]^2 + \tfrac{9}{4}\kappa_3^2, \\
A_{12} = A_{21} &= \tfrac{1}{8}\left[-\kappa_1\kappa_4 + \kappa_2\kappa_5 - \kappa_4\kappa_5 + \kappa_1\kappa_2\right] + \tfrac{3}{4}\left[\kappa_3\kappa_4 - \kappa_2\kappa_3\right], \\
A_{22} &= \tfrac{1}{2}\left[\kappa_2 - \kappa_4\right]^2 + \tfrac{1}{32}\left[\kappa_1 + \kappa_5\right]^2 + \tfrac{3}{8}\left[\kappa_1\kappa_3 + 3\kappa_3^2 + \kappa_3\kappa_5\right] \\
B_{11} &= \tfrac{1}{16}\left[15\kappa_1^2 - 2\kappa_1\kappa_5 + 15\kappa_5^2\right] - \tfrac{9}{4}k_3^2 \\
B_{12} = B_{21} &= \tfrac{1}{4}\left[\kappa_1\kappa_4 - \kappa_2\kappa_5\right] + \tfrac{7}{4}\left[\kappa_1\kappa_2 - \kappa_4\kappa_5\right] + \tfrac{3}{2}\left[\kappa_2\kappa_3 - \kappa_3\kappa_4\right] \\
B_{33} &= \tfrac{1}{16}\left[\kappa_1 + \kappa_5\right]^2 + 2\left[\kappa_2^2 + \kappa_4^2\right] + \tfrac{3}{2}\left[\kappa_1\kappa_3 + \kappa_3\kappa_5\right] + \tfrac{9}{4}\kappa_3^2
\end{aligned}$$

However, it has been shown in [14] [15] that the extrema of this contrast criterion can still be obtained by rooting the fourth degree polynomial:

$$\omega(\xi) = \sum_{i=0}^{4} c_i\, \xi^i \tag{23}$$

where $\xi = \tan\alpha - \frac{1}{\tan\alpha}$ and

$$\begin{gathered}
c_4 = \kappa_1\kappa_2 - \kappa_4\kappa_5,\ c_3 = \nu - 4(\kappa_2^2 + \kappa_4^2),\ c_2 = 3\mu, \\
c_1 = 3\nu - 2\kappa_1\kappa_5 - 32\kappa_2\kappa_4 - 36\kappa_2^2,\ c_0 = -4(\mu + 4c_4) \\
\mu = (\kappa_1 + \kappa_5 - 6\kappa_3)(\kappa_2 - \kappa_4),\ \nu = \kappa_1^2 + \kappa_5^2
\end{gathered}$$

Once the value ξ_o yielding the absolute maximum has been selected, the two equivalent solutions in α can then be obtained by solving the trinomial [15]: $\tan^2\alpha - \xi_o \tan\alpha - 1 = 0$.

Similarly, it can be shown that the extrema of $\Upsilon^{(2)}_{3,0}$ are the roots of a polynomial of degree 2 in ξ [16].

Analytical maximization of $\Upsilon^{(1)}_{q,q}$ in the complex 2×2 case. Let $r = 2q$ be even. It has been shown in [25, ch.9] that the maximization of contrast (17) amounts to diagonalizing simultaneously a set of third-order tensors. For this purpose, it has been shown that $|\mathcal{C}^{\boldsymbol{z}}_{12,1}|^2 + |\mathcal{C}^{\boldsymbol{z}}_{21,2}|^2$ is a quadratic form in variable $\boldsymbol{w} = [\cos 2\alpha, \sin 2\alpha\cos\phi, \sin 2\alpha \sin\phi]^{\mathrm{T}}$. Here we report that a similar result holds true for contrast $\Upsilon^{(1)}_{2,2}$:

$$\Upsilon^{(1)}_{2,2}(\boldsymbol{U}) = \boldsymbol{w}^{\mathrm{T}}\, \boldsymbol{B}\, \boldsymbol{w}, \tag{24}$$

where $\boldsymbol{B}$ is a 3×3 real symmetric matrix with entries

$$\begin{gathered}
B_{11} = \mathcal{C}_{11,11} + \mathcal{C}_{22,22},\ B_{21} = \Re[\mathcal{C}_{11,12} - \mathcal{C}_{12,22}],\ B_{31} = \Im[\mathcal{C}_{11,12} - \mathcal{C}_{12,22}], \\
B_{22} = \tfrac{1}{2}(\mathcal{C}_{11,11} + \mathcal{C}_{22,22}) + 2\mathcal{C}_{12,12} + \Re[\mathcal{C}_{11,22}],\ B_{23} = \Im[\mathcal{C}_{11,22}],\ B_{33} = \tfrac{1}{2}(\mathcal{C}_{11,11} + \mathcal{C}_{22,22}) + 2\mathcal{C}_{12,12} - \Re[\mathcal{C}_{11,22}]
\end{gathered}$$

where $\mathcal{C}_{ij,k\ell}$ stands for $\mathcal{C}^{\bar{\boldsymbol{y}}}_{ij,k\ell}$. This result is new and unpublished. In [22], it had been proved constructively that finding the absolute maxima of $\Upsilon^{(1)}_{2,2}$ with respect to (α, ϕ) was feasible entirely analytically. The above result gives another proof, less constructive however, because eigenvectors of a real 3×3 symmetric matrix are known to be obtainable analytically (the characteristic polynomial is of degree 3, thus smaller than 4).

Analytical maximization of $\Upsilon^{JAD}_{2,2}$ in the complex 2×2 case. Take now $\Upsilon^{JAD}_{2,2} = \sum_{ik\ell} |\mathcal{C}_{ik,i\ell}|^2$, that we can rewrite, according to (16), $\Upsilon^{JAD}_{2,2} = \sum_t ||\mathbf{Diag}\{\boldsymbol{U}\boldsymbol{N}(t)\boldsymbol{U}^{\mathrm{H}}\}||^2$. Then it has been shown [9] that:

$$\frac{1}{2}\,\Upsilon^{JAD}_{2,2} = \boldsymbol{w}^{\mathrm{T}}\, \Re[\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}]\, \boldsymbol{w} + constant, \tag{25}$$

where the t-th row of $\boldsymbol{B}$ is $[a_t - d_t,\ b_t + c_t,\ \jmath(c_t - b_t)]$, and $\begin{pmatrix} a_t & b_t \\ c_t & d_t \end{pmatrix} \stackrel{\text{def}}{=} \boldsymbol{U}\,\boldsymbol{N}(t)\,\boldsymbol{U}^{\mathrm{H}}$. As a consequence, finding the maxima of $\Upsilon^{JAD}_{2,2}$ amounts to maximizing a real quadratic form in 3 variables.

4 Special-purpose algorithms

4.1 Discrete sources

When it is possible, it may be interesting to exploit the fact that sources take their values in a known finite alphabet. For computational reasons, and in order to avoid to take care of a possible carrier residual, the case of sources uniformly distributed on the unit circle has been first addressed (which does not correspond to a finite alphabet *stricto sensu*).

It is important to stress that the two approaches described in this section 4.1 do not require sources to be statistically independent. In fact, the separation is based upon the discrete character, so that correlated sources can be separated, provided that some weak identifiability conditions are verified [67] [34] [60].

Constant modulus sources. Van der Veen was the first to propose an *analytical* algorithm to separate CM sources [67]. The principle of his algorithm is best understood through the Noiseless Case. Assume again the observations have been spatially pre-whitened, so that the data matrix is of size $P \times N$ and full rank.

One attempts to extract each source one by one with a linear filter $\boldsymbol{f}$ as: $s(n) = \boldsymbol{f}^{\mathrm{H}}\boldsymbol{y}(n)$. Thus the constraint of unit modulus can be written as:

$$\boldsymbol{f}^{\mathrm{H}}\boldsymbol{y}(n)\boldsymbol{y}(n)^{\mathrm{H}}\boldsymbol{f} = 1, \forall n. \tag{26}$$

This yields $\boldsymbol{f}^{\otimes\,\mathrm{H}}\boldsymbol{y}(n)^{\otimes} = 1$, for all time samples n, $1 \leq n \leq N$, where vector $\boldsymbol{u} \otimes \boldsymbol{u}^*$ is denoted $\boldsymbol{u}^{\otimes}$. Stacking all these equations one above the other, we get a $N \times P$ formally linear system: $\boldsymbol{Y}^{\otimes}\boldsymbol{f}^{\otimes} = \mathbf{1}$, where $\mathbf{1}$ is a vector formed of ones. It can be shown that, in the absence of noise, and if $N \geq P^2$, the null space of $\boldsymbol{Y}^{\otimes}$ is of dimension $P-1$. Among the infinite number of solutions to this system, only P have the structure requested by vectors of the form $\boldsymbol{f}^{\otimes}$. In order to calculate them, simply construct the null space equation $\boldsymbol{N}\boldsymbol{f}^{\otimes} = \mathbf{0}$, where $\boldsymbol{N}$ is a given $(N-1) \times P^2$ matrix. Let $\{\boldsymbol{u}^{(1)}, \ldots \boldsymbol{u}^{(P)}\}$ be a basis of $\ker\{\boldsymbol{N}\}$. Then any of the P extractors $\boldsymbol{f}_j^{\otimes}$ is a linear combination of the $\boldsymbol{u}^{(p)}$'s, and vice-versa. Rearranging the system of equations leads eventually to $\boldsymbol{u}^{(p)} = \sum_{j=1}^{P} \lambda_j^{(p)} \boldsymbol{f}_j^{\otimes}$, or, in closed form, by applying the operator $\mathbf{Unvec}\{\cdot\}$:

$$\boldsymbol{U}^{(p)} = \boldsymbol{F}\Lambda^{(p)}\boldsymbol{F}^{\mathrm{H}} \tag{27}$$

Thus, the problem amounts to jointly diagonalizing P matrices; the difference with [9] is that here $\boldsymbol{F}$ is not constrained to be unitary. The numerical algorithm proposed in [67] is iterative, and is inspired from the QZ iteration usually dedicated to the diagonalization of matrix pencils. Another analytical solution to this type of problem, applicable to CM sources, has been introduced in [34] for tensors, and is described in the next section.

Discrete sources In [66], the above algorithm has been specialized to real data, that is, to discrete BPSK sources. This algorithm cannot handle complex constellations however. Here, we examine a related procedure for $q-$PSK modulations, where the support of the distribution reduces to q masses, c_j.

In the present case, we impose the constraint $[\boldsymbol{f}^{\mathrm{T}}\boldsymbol{y}(n)]^q = 1$, so that:

$$\boldsymbol{y}(n)^{\oslash q\,\mathrm{T}}\boldsymbol{f}^{\oslash q} = 1 \tag{28}$$

where $\boldsymbol{u}^{\oslash}$ stands for $\boldsymbol{u} \oslash \boldsymbol{u}$, and contains all *distinct* cross products between entries of $\boldsymbol{u}$ (as opposed to $\boldsymbol{u} \otimes \boldsymbol{u}$). For $q = 2$, this is still different from equation (26) because of the absence of complex conjugates, and because of the smaller vector dimension.

By stacking equations (28) for $1 \leq n \leq N$, one obtains a formally linear system $\boldsymbol{Y}^{\oslash q}\boldsymbol{f}^{\oslash q} = \mathbf{1}$. Matrix $\boldsymbol{Y}^{\oslash q}$ is generically of rank $P-1$ if P sources are present [34]; there are thus again infinitely many solutions. Denote $\boldsymbol{f}_{min}^{\oslash q}$ the minimum norm solution, and $\{\boldsymbol{u}^{(p)}\}$ a basis of the null space. Then extractors take the form: $\boldsymbol{f}^{\oslash q} = \boldsymbol{f}_{min}^{\oslash q} + \sum_{p=1}^{P-1} \lambda_p \boldsymbol{u}^{(p)}$. The coefficients λ_p are determined by imposing the structure on $\boldsymbol{f}^{\oslash q}$. Unlike the algorithm described in the previous section, here one uses the property that $\mathbf{Unvecs}_q\{\boldsymbol{f}^{\oslash q}\}$ must have a rank 1. An analytical solution is proposed in [34].

In the presence of noise, one can attempt to satisfy equations (28) in the LS sense. Actually, since source distributions are known, a Bayesian approach is possible. With this goal, it has been proved in [33] that the Maximum A Posteriori (MAP) criterion is asymptotically equivalent to the Mean Square Error (MSE) $\sum_n \prod_j |\boldsymbol{f}^{\mathrm{T}}\boldsymbol{y}(n) - c_j|^2$, which is itself equal to $\sum_n |\boldsymbol{f}^{\oslash q\,\mathrm{T}}\boldsymbol{y}(n)^{\oslash q} - 1|^2$.

4.2 More sources than sensors: under-determined mixtures

Mixtures in which the number of sources, P exceeds the available diversity, K, are referred to as under-determined [18]; some authors also refer to over-complete representations [41].

The reasons for using higher order arrays are three-fold. First, the sole use of second order statistics is often (but not always) insufficient to obtain identifiability in BSS problems. Second, data are often arranged in many-way arrays (*e.g.* in factor analysis), for instance three-way [25] [6]; the reduction to a 2-way array represents a loss of information. Third, the number of factors that can be identified is limited to the rank of the data matrix, itself bound by the smallest dimension. Yet, there may very well be more factors than the smallest dimension.

One should distinguish again between identification of mixture $\boldsymbol{H}$, and source extraction, that can obviously not been carried out linearly (the mixture being rectangular, it does not admit an inverse). The identification problem can be addressed in a rather general framework under assumptions **A1** and **A2**. On the other hand, the extraction of sources needs a stronger knowledge on the source distribution, such as their discrete character [18] [21]. The identification of under-determined systems is a much more complicated instance of the BSS problem, and is partly covered in a companion paper [19]. One of the striking limitation is that one should not resort to a second-order preprocessing $\boldsymbol{T}$ as in the previous sections, because it would then limit the rank to the dimension.

Nevertheless, one can briefly summarize the state of the art by saying that blind identification of such systems is equivalent to decomposing a tensor into a sum of rank-one terms. This kind of decomposition is not equivalent to known problems in linear algebra. At the third order for instance, the extension of the SVD concept does not yield a diagonal tensor [26]. Even the definition of the rank itself raises big difficulties [5]. Some sub-optimal decompositions have been proposed for 3-way arrays in the seventies [10] [38] but are not entirely satisfactory. Results about identifiability are more recent [40].

Fortunately, symmetric arrays of order r and dimension K can be associated bijectively to homogeneous polynomials of degree r in K variables. Based on this remark, decomposing a symmetric array is equivalent to decomposing a homogeneous polynomial into a sum of linear forms raised to the $q-$th power [24]. This remark allows to connect the problem to early works in invariant theory [51] and multilinear algebra [28]. The first results go back to the beginning of the century with the works of Sylvester and Wakeford [68]. Also related are the works of Rota [30] on binary quantics, and those of Reznick on quantics of even degree [54]. Readers can refer to [25] or [19] for more details.

5 Concluding remarks

In this paper, we attempted to summarize what is known about MIMO blind deconvolution of linear systems (including static ones), essentially based on contrast maximization procedures. We tried to gather some examples of numerical algorithms, most of them proceeding pairwise in order to compute the solutions analytically. However, in some cases, numerical difficulties imposed approximations in order to deflate the problem to an addressable one; for instance, the analytical maximization of $\Upsilon^{(2)}_{2,2}$ is still considered as difficult, even if it is of polynomial complexity. One can notice that the static instance (BSS), now often referred to as *Independent Component Analysis*, is very important in several fields, including Factor Analysis, and should not be overlooked.

References

[1] Y. BAR-NESS, J. W. CARLIN, M. L. STEINBERGER, "Bootstrapping adaptive interference cancelers: Some practical limitations", in *Proc. The Globecom. Conference*, Miami, Nov. 1982, pp. 1251–1255, paper No F3.7.

[2] A. BELOUCHRANI, K ABEDMERAIM, J. F. CARDOSO, E. MOULINES, "Second-order blind separation of correlated sources", in *Proc. Int. Conf. Digital Signal Processing*, Cyprus, 1993, pp. 346–351.

[3] A. BENVENISTE, M. GOURSAT, "Blind equalizers", *IEEE Trans. Com.*, vol. 32, no. 8, pp. 871–883, Aug. 1984.

[4] A. BENVENISTE, M. GOURSAT, G. RUGET, "Robust identification of a non-minimum phase system", *IEEE Trans. Auto. Contr.*, vol. 25, no. 3, pp. 385–399, June 1980.

[5] G. M. BERGMAN, "Ranks of tensors and change of base field", *Journal of Algebra*, vol. 11, pp. 613–621, 1969.

[6] R. BRO, N. SIDIROPOULOS, G. GIANNAKIS, "A fast least squares algorithm for separating trilinear mixtures", in *First Int. Work. Indep. Comp. Anal.*, Aussois, France, Jan. 1999, pp. 289–294.

[7] J. F. CARDOSO, "Source separation using higher order moments", in *Proc. ICASSP* Glasgow, 1989, pp. 2109–2112.

[8] J. F. CARDOSO, "High-order contrasts for independent component analysis", *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan. 1999.

[9] J. F. CARDOSO, A. SOULOUMIAC, "Blind beamforming for non-Gaussian signals", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 362–370, Dec. 1993, Special issue on Applications of High-Order Statistics.

[10] J. D. CAROLL, 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, Sept. 1970.

[11] P. CHEVALIER, "Optimal separation of independent narrow-band sources - concept and performance", *Signal Processing, Elsevier*, vol. 73, no. 1, Feb. 1999, special issue on blind separation and deconvolution.

[12] P. COMON, "Separation of stochastic processes", in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 28-30 1989, IEEE-ONR-NSF, pp. 174–179.

[13] P. COMON, "Analyse en Composantes Indépendantes et identification aveugle", *Traitement du Signal*, vol. 7, no. 3, pp. 435–450, Dec. 1990, Numero special non lineaire et non gaussien.

[14] P. COMON, "Independent Component Analysis", in *Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics*, Chamrousse, France, July 10-12 1991, pp. 111–120, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 29–38.

[15] P. COMON, "Independent Component Analysis, a new concept ?", *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, Apr. 1994, Special issue on Higher-Order Statistics.

[16] P. COMON, "Tensor diagonalization, a useful tool in signal processing", in *IFAC-SYSID, 10th IFAC Symposium on System Identification*, M. Blanke, T. Soderstrom, Eds., Copenhagen, Denmark, July 4-6 1994, vol. 1, pp. 77–82, invited session.

[17] P. COMON, "Contrasts for multichannel blind deconvolution", *Signal Processing Letters*, vol. 3, no. 7, pp. 209–211, July 1996.

[18] P. COMON, "Blind channel identification and extraction of more sources than sensors", in *SPIE Conference*, San Diego, July 19-24 1998, pp. 2–13, keynote address.

[19] P. COMON, "Tensor decompositions, state of the art and applications", in *IMA Conf. Mathematics in Signal Processing*, Warwick, UK, Dec. 18-20, 2000, keynote address.

[20] P. COMON, P. CHEVALIER, "Source separation: Models, concepts, algorithms and performance", in *Unsupervised Adaptive Filtering, Vol. I, Blind Source Separation*, pp. 191–236. Wiley, 2000.

[21] P. COMON, O. GRELLIER, "Non linear inversion of underdetermined mixtures", in *ICA99, IEEE Workshop on Indep. Comp. Anal. and Blind Source Separation*, Aussois, France, Jan 11–15 1999, pp. 461–465.

[22] P. COMON, E. MOREAU, "Improved contrast dedicated to blind separation in communications", in *ICASSP*, Munich, April 20-24 1997, pp. 3453–3456.

[23] P. COMON, E. MOREAU, "Blind MIMO equalization and joint-diagonalization criteria", in *ICASSP'01*, Salt Lake City, May 7-11, 2001.

[24] P. COMON, B. MOURRAIN, "Decomposition of quantics in sums of powers of linear forms", *Signal Processing, Elsevier*, vol. 53, no. 2, pp. 93–107, Sept. 1996, special issue on High-Order Statistics.

[25] L. DeLATHAUWER, *Signal Processing based on Multilinear Algebra*, Doctorate, Katholieke Universiteit Leuven, Sept. 1997.

[26] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, "Blind sources separation by higher-order singular value decomposition", in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 175–178.

[27] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, "Independent component analysis based on higher-order statistics only", in *8th IEEE SP Workshop on Statistical Signal and Array Processing*, Corfu, Greece, June 1996, pp. 356–359.

[28] J. A. DIEUDONNÉ, J. B. CARREL, "Invariant theory, old and new", in *Computing Methods in Optimization Problems-2*, San Remo, Italy, Sept 1968, 1970, pp. 1–80, Academic Press.

[29] D. DONOHO, "On minimum entropy deconvolution", in *Applied time-series analysis II*, pp. 565–609. Academic Press, 1981.

[30] R. EHRENBORG, G. C. ROTA, "Apolarity and canonical forms for homogeneous polynomials", European Journal of Combinatorics, 1993.

[31] A. FERREOL, P. CHEVALIER, "On the behavior of current second and higher order blind source separation methods for cyclostationary sources", *IEEE Trans. Sig. Proc.*, vol. 48, pp. 1712–1725, June 2000.

[32] S. Van GERVEN, D. Van COMPERNOLLE, "Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness", *IEEE Trans. Sig. Proc.*, vol. 43, no. 7, pp. 1602–1612, July 1995.

[33] O. GRELLIER, P. COMON, "Blind separation of discrete sources", *IEEE Signal Processing Letters*, vol. 5, no. 8, pp. 212–214, Aug. 1998.

[34] O. GRELLIER, P. COMON, "Analytical blind discrete source separation", in *Eusipco*, Tampere, Finland, 5-8 sept. 2000.

[35] S. HAYKIN, *Unsupervised Adaptive Filtering*, vol. 2, Wiley, 2000, series in Adaptive and Learning Systems for Communications, Signal Processing, and Control.

[36] C. JUTTEN, J. HÉRAULT, "Independent component analysis versus PCA", in *Proc. EUSIPCO,* Grenoble, France, 1988, pp. 643–646.

[37] C. JUTTEN, J. HÉRAULT, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture", *Signal Processing, Elsevier*, vol. 24, no. 1, pp. 1–20, 1991.

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

[39] J. L. LACOUME, P. O. AMBLARD, P. COMON, *Statistiques d'ordre supérieur pour le traitement du signal*, Collection Sciences de l'Ingénieur. Masson, 1997.

[40] S. LEURGANS, R. T. ROSS, R. B. ABEL, "A decomposition for three-way arrays", *SIAM Jour. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, Oct. 1993.

[41] M. LEWICKI, T. J. SEJNOWSKI, "Learning non-linear overcomplete representations for efficient coding", in *Advances in Neural Information Processing Systems*, 1998, pp. 815–821.

[42] O. MACCHI, E. EWEDA, "Convergence analysis of self-adaptive equalizers", *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 161–176, Mar. 1984.

[43] P. McCULLAGH, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.

[44] E. MOREAU, "Criteria for complex sources separation", in *VIII European Signal Processing Conference (EUSIPCO'96)*, Trieste, Italy, September 1996, pp. 931–934.

[45] E. MOREAU, "A generalization of joint-diagonalization criteria for source separation", *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 530–541, March 2001.

[46] E. MOREAU, P. COMON, "Comments on Blind Beamforming for Multiple Non-Gaussian Signals and the Constant Modulus Algorithm", *IEEE Trans. Signal Processing*, vol. 48, no. 11, pp. 3248–3250, Nov. 2000.

[47] E. MOREAU, J. C. PESQUET, "Generalized contrasts for multichannel blind deconvolution of linear systems", *IEEE Signal Processing Letters*, vol. 4, no. 6, pp. 182–183, June 1997.

[48] E. MOREAU, J. C. PESQUET, N. THIRION, "An equivalence between non symmetrical contrasts and cumulant matching for blind signal separation", in *ICA99, IEEE Workshop on Indep. Comp. Anal. and Blind Source Separation*, Aussois, France, Jan 11–15 1999, pp. 301–306.

[49] E. MOREAU, N. THIRION, "Multichannel blind signal deconvolution using high order statistics", in *IEEE Signal Processing Workshop on Statistical Signal and Array Processing (SSAP'96)*, Corfu, Greece, June 1996, pp. 336–339.

[50] E. MOREAU, N. THIRION, "Non symmetrical contrasts for sources separation", *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2241–2252, August 1999.

[51] K. V. PARSHALL, "The one-hundred anniversary of the death of invariant theory", *The Mathematical Intelligencer*, vol. 12, no. 4, pp. 10–16, 1990.

[52] D. T. PHAM, P. GARAT, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach", *IEEE Trans. Sig. Proc.*, vol. 45, no. 7, pp. 1712–1725, July 1997.

[53] J. G. PROAKIS, C. L. NIKIAS, "Blind equalization", in *SPIE Adaptive Signal Processing*, 1991, vol. 1565, pp. 76–88.

[54] B. REZNICK, "Sums of even powers of real linear forms", *Memoirs of the AMS*, vol. 96, no. 463, Mar. 1992.

[55] P. RUIZ, J.L. LACOUME, "Extraction of independent sources from correlated inputs", in *Proc. Workshop on Higher-Order Spectral Analysis,* Vail, Colorado, June 1989, pp. 146–151.

[56] J. SHEINVALD, "On blind beamforming for multiple non-gaussian signals and the constant-modulus algorithm", *IEEE Trans. Signal Processing*, vol. 46, no. 7, pp. 1878–1885, July 1998.

[57] C. SIMON, P. LOUBATON, C. JUTTEN, "Separation of a class of convolutive mixtures: A contrast function approach", in *ICASSP*, Phoenix, May 1999.

[58] C. SIMON, P. LOUBATON, C. VIGNAT, C. JUTTEN, G. d'URSO, "Blind source separation of convolutive mixtures by maximization of fourth-order cumulants : the non i.i.d. case", in *ICASSP*, Seattle, May 12-15 1998.

[59] A. SWAMI, G. GIANNAKIS, S. SHAMSUNDER, "Multichannel ARMA processes", *IEEE Trans. Sig. Proc.*, vol. 42, no. 4, pp. 898–913, Apr. 1994.

[60] S. TALWAR, M. VIBERG, A. PAULRAJ, "Blind estimation of multiple co-channel digital signals arriving at an antenna array: Part I, algorithms", *IEEE Trans. Sig. Proc.*, pp. 1184–1197, May 1996.

[61] N. THIRION, E. MOREAU, "New criteria for blind signal separation", in *IEEE Workshop on Statistical Signal and Array Processing*, Pocono Manor, Pennsylvania, Usa, Aug. 14-16, 2000, pp. 344–348.

[62] L. TONG, R. LIU, V. C. SOON, "Indeterminacy and identifiability of blind identification", *IEEE Trans. Cir. Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.

[63] J. K. TUGNAIT, "Approaches to FIR system identification with noisy data using higher-order statistics", in *Proc. Workshop on Higher-Order Spectral Analysis,* Vail, Colorado, June 1989.

[64] J. K. TUGNAIT, "Approaches to FIR system identification with noisy data using higher-order statistics", *IEEE Trans. Acoust. Speech Sig. Proc.*, pp. 1307–1317, July 1990.

[65] J. K. TUGNAIT, "Identification and deconvolution of multichannel non-gaussian processes using higher order statistics and inverse filter criteria", *IEEE Trans. Sig. Proc.*, vol. 45, pp. 658–672, Mar. 1997.

[66] A. J. van der VEEN, "Analytical method for blind binary signal separation", *IEEE Trans. Sig. Proc.*, vol. 45, no. 4, pp. 1078–1082, Apr. 1997.

[67] A. J. van der VEEN, A. PAULRAJ, "An analytical constant modulus algorithm", *IEEE Trans. Sig. Proc.*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[68] E. K. WAKEFORD, "Apolarity and canonical forms", in *Proceedings of the London Mathematical Society*, 1918, pp. 403–410.

[69] E. WEINSTEIN, M. FEDER, A. V. OPPENHEIM, "Multi-channel signal separation by decorrelation", *IEEE Trans. on Speech and Audio Processing*, vol. 1, no. 4, pp. 405–413, Oct. 1993.