

Contrasts, Independent Component Analysis, and Blind Deconvolution

P. Comon

I3S, CNRS-UNSA, Sophia-Antipolis

SUMMARY

A general definition of contrast criteria is proposed, which induces the concept of trivial filters. These optimization criteria enjoy identifiability properties, and aim at delivering outputs satisfying specific properties, such as statistical independence or a discrete character. Several ways of building new contrast criteria are described. It is then briefly elaborated on practical numerical algorithms. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: Independent Component Analysis (ICA), Contrast criteria, Blind Deconvolution and Equalization, MIMO linear systems

1. INTRODUCTION

The paper is organized as follows. In this section, the MIMO Blind Deconvolution problem is introduced; notations and a brief bibliographical survey are also included. Section 2 is a reminder about whitening operations and cumulants. Assumptions and definitions of Trivial Filters are given in section 3. Section 4 defines a wide class of contrast functionals and gives several examples; some of them are new. Practical algorithms are tackled in section 5.

Blind equalization or identification schemes have been the subject of intense interest since the works of Sato [55] in 1975, and then Godard [25] a few years later. One of the main advantages of blind techniques is that training sequences are not required; by deleting pilot sequences, one can thus increase the transmission rate. But there are other advantages, which stem from limitations of classical approaches. In fact, techniques based on pilot sequences are difficult to use when channel responses are long, or fast varying, compared to the length of the pilot sequence. The presence of a carrier residual can also make the equalization task more difficult [32] [13].

Instead of basing the identification or equalization schemes on input-output measurements (data-aided approaches), some properties about the inputs are exploited (blind approaches), as is now explained.

*Correspondence to: P. Comon, Lab. I3S, 2000 route des Lucioles, B.P.121, F-06903 Sophia-Antipolis cedex

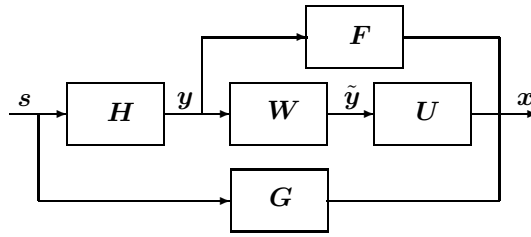


Figure 1. The transmitted sequence $\mathbf{s}[n]$ propagates through a channel $\check{\mathbf{H}}[z]$, is then whitened by a filter $\check{\mathbf{W}}[z]$, and is eventually deconvolved by a paraunitary equalizer $\check{\mathbf{U}}[z]$ to yield output $\mathbf{x}[n]$; $\check{\mathbf{G}}[z]$ denotes the global filter.

1.1. Modeling

Limiting our discussion to linear modulations, the complex envelope of a transmitted signal $s(t)$ takes the form in baseband [52]: $s(t) = \sum_k g(t - kT) s[k]$. Note the distinction between discrete-time and continuous time processes via brackets and parentheses: $s[k] = s(kT_s)$, where T_s is the symbol period. After propagation through the channel and the receive filter, the signal received on the antenna may be written as:

$$y(t) = \sum_k h(t - kT_s) s[k]$$

where h is the convolution of the transmit filter, the channel, and the receive filter. If the received signal is sampled at the rate $1/T_s$, it can be modeled as:

$$y[n] = \sum_k h[n - k] s[k] \quad (1)$$

with $h[k] \stackrel{\text{def}}{=} h(kT_s)$. For Multiple Input Multiple Output (MIMO) systems, the transmitted signal $\mathbf{s}[k]$ and the received signal $\mathbf{y}[k]$ may be considered as vector-valued discrete-time processes; their dimension is denoted by P and K , respectively. Note the boldface that emphasizes the multi-dimensional character of the processes. Model (1) can then be rewritten as:

$$\mathbf{y}[n] = \sum_k \mathbf{H}[n - k] \mathbf{s}[k] \quad (2)$$

where the global channel impulse response $\mathbf{H}[k]$ is now a sequence of $K \times P$ matrices. Its z -transform is denoted as

$$\check{\mathbf{H}}[z] \stackrel{\text{def}}{=} \sum_k \mathbf{H}[k] z^{-k}$$

In the present context, inputs $s_j[k]$ are often referred to as *sources*.

The case where source symbol rates are different or unknown is not addressed in this paper, although it is an interesting issue, in surveillance or interception contexts for instance. However, one can still say that the output appears as a convolution, but not as a *discrete convolution* anymore. In fact, if the sample rate is $1/T'$ at the receiver, we have:

$$\mathbf{y}[n] = \sum_k \mathbf{H}(nT' - kT_s) \mathbf{s}[k]$$

Another important issue is that of carrier mismatch. If the carrier frequency at the receiver is slightly different from the modulating carrier, then there is a *carrier residual*, which one can merely represent in baseband by a multiplicative exponential. For a SISO channel, this can be written as:

$$y[n] = \sum_k h[n-k] s[k] e^{j k \delta} \quad (3)$$

where δ is a small real number, and the dotless j denotes $\sqrt{-1}$.

The goal of blind equalization is to yield an estimate of input sequences $s_j[k]$ from the sole observation of output sequences $y_i[n]$. In order to solve this problem, some additional assumptions are necessary, as explained in section 3.

1.2. Noise

The modeling described above does not include noise, which is not very satisfactory. In fact, to some extent, an additive noise is a good means to take into account modeling errors. For instance, the presence of weak extraneous sources (if they are fewer than $K - P$) can be taken into account this way. Yet, the core of the contrast setting relies on the existence of some inverse filter, able to extract exactly all sources [11] [2] [7] [33]. For this reason, noise can only be treated as a nuisance, and the number of sources cannot in principle exceed the number of sensors.

Nevertheless, it is always possible to design a filter extracting one or several sources by maximizing some optimization criterion. If at most K sources are attempted to be simultaneously extracted, then this criterion may be a contrast. As an example, the contrast criteria described in the present paper can be utilized in order to find the best MISO filter extracting any single source, individually, regardless of the number of sources present. Each MISO extracting filter is then hopefully a local maximum of the criterion. But the question that then remains to be answered is: are there spurious local maxima, or does every maximum correspond to an extracting filter? In its general form, this still remains an open question when non Gaussian noise or extraneous sources are present, even if some works have already been initiated in this direction [54].

1.3. Bibliographical comments

The exploitation of source color and spectral differences has been investigated in [5] [51], among others. The case of non stationary or cyclo-stationary sources is studied for instance in [50] [23]. MIMO blind identification/equalization via subspace fitting is described in [45] [40]. The alternative approach via linear prediction has been first proposed in [10], and then independently addressed in [56] [21] [1] [26] and in numerous subsequent works; linear prediction approaches have long been recognized to be more robust to channel/equalizer length misadjustment. This has been revisited in [39]. Deflation approaches have the advantage not to require prior space-time whitening [20] [60] [61] [47], because sources are extracted one by one; this also allows to write an ascent algorithm as a fixed point search [33]. Contrast-based approaches can be found in [61] [22] [17] [12] [16]. For more references on the use of the discrete character of sources in MIMO over-determined mixtures, either static or convolutive, see [29] [64] [62] [57] [59] [63] [38] [24] [37] [30]. Underdetermined mixtures are addressed in [14] and references therein. For a general account on blind techniques, see [31] [33] [36] [22] [7].

2. STATISTICAL TOOLS

2.1. Whitening

The whitening operation is useful in SISO systems, and in over-determined ($P \leq K$) MIMO systems. In fact, if outputs aim at being independent, then they should at least be uncorrelated at order 2. This is what the whitening operation does.

Theorem 1. Any matrix $\check{H}[z]$, whose entries are rational functions, can be factored (generally in a non unique manner) as $\check{H}[z] = \check{R}[z]\check{U}[z]$, where $\check{U}[z]$ is paraunitary and $\check{R}[z]$ is triangular minimum phase.

The paraunitarity means that

$$\check{U}[z]\check{U}^H[1/z^*] = \mathbf{I} \quad (4)$$

where \mathbf{I} is the identity. This property characterizes lossless filters. In the scalar case for instance, $\check{U}[z]$ is all-pass and $\check{R}[z]$ has all its roots in the unit disk.

One possibility is thus to assume a 3-step processing: (i) whitening with the help of second-order statistics, which can be done in many ways, (ii) estimation of a paraunitary separating filter, (iii) Space-Time Matched Filtering (STMF), based on the estimated channel [15]. From now on, we shall retain this procedure, although it is not always optimal, because it puts at lot more weight on second-order statistics than on higher orders.

2.2. Cumulants

Whereas moments can be defined as the coefficients in the expansion of the first characteristic function, cumulants are the coefficients in the expansion of the second one (the logarithm of the first) [42]. For a zero-mean real random variable, the fourth order cumulant can be expressed as a function of second and fourth order moments: $\mathcal{C}_y^{(4)} = \mu_y^4 - 3\mu_y^2$. The Leonov-Shiryaev formula allows to express any cumulant as a function of moments of lower or equal orders [42].

For a stationary process $\mathbf{y}[n]$, we define the i th marginal r th order cumulant as:

$$\mathcal{C}_{y_i}^{(r,p)} \stackrel{\text{def}}{=} \text{Cum}\left\{\underbrace{y_i, \dots, y_i}_p, \underbrace{y_i^*, \dots, y_i^*}_{r-p}\right\} \quad (5)$$

The marginal cumulant with no complex conjugation, $\mathcal{C}_{y_i}^{(r,r)}$, sometimes called the *non circular* cumulant, will be denoted $\mathcal{C}_{y_i}^{(r)}$, in short. The generic entry of the 4th order circular cumulant tensor is denoted $\mathcal{C}_{\mathbf{y};i,j,k,\ell} = \text{Cum}\{y_i, y_j, y_k^*, y_\ell^*\}$.

One key property enjoyed by cumulants, and that is not shared by moments, is the following:

Property 2. If x_1 and x_2 are independent random variables, then $\forall r, p: \mathcal{C}_{x_1+x_2}^{(r,p)} = \mathcal{C}_{x_1}^{(r,p)} + \mathcal{C}_{x_2}^{(r,p)}$.

In digital communications, an important property is that of *circularity*. In particular, for PSK- m modulations, elements of the complex constellation satisfy $x^m = 1$. As a consequence, $\mathbf{E}\{x^m\} = 1$, of course. But more interestingly, $\mathbf{E}\{x^p\} = 0$ and $\mathcal{C}_x^{(p)} = 0$, $\forall p < m$. We shall say that x is circular up to order $m - 1$. General definitions can be provided [35]:

Definition 3. A random variable is said to be strongly circular if its distribution is invariant by rotation in the complex plane. It is said to be circular at order $m-1$, $m > 1$, if its distribution is invariant by a rotation of angle $2\pi/m$.

One must pay attention to the fact that by *circular*, it is often understood *circular at order 2*. For instance, BPSK variables are *non circular*.

3. TRIVIAL FILTERS

3.1. Assumptions and taxonomy.

In order to blindly equalize convolutive models, the most widely used assumption is the *statistical independence* between successive symbols.

Hypothesis H1. Sources $s_j[k]$ are all *i.i.d.* sequences.

For MIMO models, the independence assumption between sources is often utilized:

Hypothesis H2. Sources $s_j[k]$ are mutually statistically independent.

These hypotheses can generally be deflated to less strong whiteness/independence properties, because moments of finite orders are used [15]. Let us stress that the case where sources are *linear processes* can also be treated in a similar manner as *i.i.d.* sources; Hypothesis H1 is thus not very restrictive. A particular case however raises problems, namely that of Gaussian sources. In that case, all the information is contained in moments up to order 2, which is not sufficient to establish identifiability. For this reason, it is necessary to resort to a third hypothesis, along with hypotheses H1 and H2:

Hypothesis H3. At most one source is Gaussian.

Again, this assumption can often be deflated to a weaker one. For instance, when the contrast optimization criterion is based on cumulants, then it is sufficient to consider the assumption

Hypothesis H4. At most one source has a null marginal cumulant.

On the other hand, there exist other frameworks in which hypotheses are different. For instance, if sources have different spectra, or if they are non stationary, or cyclo-stationary, then they can be separated with the help of appropriate techniques (cf. section 1.3). These three cases are not addressed in the present paper, and yield quite different (easier) theoretical problems. Nevertheless, the special framework of *discrete sources* is relevant in the context of digital communications.

The interest of exploiting the discrete character lies not only in a more accurate characterization of the desired output (than just non Gaussian or CM), but also in the fact that some other assumptions can be dropped. In this section, the sole assumption used is

Hypothesis H5. The sources $s_j[n]$ belong to a known finite alphabet \mathcal{A} characterized by the d distinct complex roots of a polynomial $Q(z) = 0$.

For instance, sources can be correlated and non stationary. In fact, the criterion proposed above is entirely algebraic and deterministic, so that no statistical tool is required. The simplest case is $Q(z) = z^q - 1$, for which we have a PSK- q constellation.

Next, for identifiability reasons, another hypothesis will be needed:

Hypothesis H6. *Sources $s_j[n]$ are sufficiently exciting*

By *sufficiently exciting*, it is meant that sufficiently many distinct states (among the d^P possible ones) of the P -uplet \mathbf{s} are present in the source data matrix. For general alphabets, as shown in the proof of Theorem 16, it is sufficient that all binary sequences are present, which means that the observation length should be at least of $\binom{d}{2}2^P$ symbols. For PSK alphabets, this minimal length can be deflated to $\lceil \frac{d}{2} \rceil 2^P$.

Depending on hypotheses (independence, discrete character, SISO/MIMO, $P > K$ or not...), a whole variety of problems can be stated [15]. From now on, and unless otherwise specified, we shall concentrate only on Hypotheses H1 to H6, and on the case where $P \leq K$ (over-determined mixtures). When more sources than sensors are present, the problem becomes more complicated (under-determined mixtures), and specific tools generally need to be utilized; see [14] and references therein.

3.2. Trivial filters.

The separating linear filter, $\check{\mathbf{F}}[z]$, if it exists, aims at delivering an output, $\mathbf{x}[n]$, which should satisfy as well as possible hypotheses H1 and H2. But it is clear that there exist some filters that do not affect them. These are called the trivial filters, and we can prove

Lemma 4. *Under hypotheses H1 to H3, trivial filters are of the form $\check{\mathbf{T}}[z] = \mathbf{P}\check{\mathbf{D}}[z]$, where \mathbf{P} is a permutation, and $\check{\mathbf{D}}[z]$ a diagonal filter. In addition, because of the i.i.d. property of hypothesis H1, entries of $\check{\mathbf{D}}[z]$ must be of the form $\check{D}_{pp}[z] = \lambda_p z^{\delta_p}$, where δ_p is an integer.*

Consequently, it is hopeless to estimate the pair $(\check{\mathbf{H}}[z], \mathbf{s}[k])$. One should rather try to estimate one representative of the equivalence class of solutions. Once one solution is found, all the others can be generated by trivial filtering.

Example. Assume the model is MIMO static. Then $\mathbf{y}[n] = \mathbf{H}\mathbf{s}[n]$, where $\mathbf{y}[n]$ and $\mathbf{s}[n]$ are realizations of random variables. In that case, hypothesis H1 is not mandatory anymore. The estimation of the pair $(\mathbf{H}, \mathbf{s}[n])$ from the sole observations $\mathbf{y}[n]$ under hypotheses H2 and H3, has been introduced originally by Jutten and Comon [34] [10], and is now called *Independent Component Analysis* (ICA) [11] [49] [33] [36] [31] [14].

Consider now hypotheses H5 and H6 instead of hypotheses H1 to H3. In the PSK- q case, we have the following

Lemma 5. *Under hypothesis H5 and H6 and with $Q(z) = z^q$, trivial filters are of the form $\mathbf{P}\check{\mathbf{D}}[z]$, where $\check{D}_{pp}[z]$ are rotations in the complex plane of an angle multiple of $2\pi/q$ combined with a pure delay, and \mathbf{P} are permutations.*

For more general alphabets (called constellations in the framework of digital linear modulations), we need to define the following set

Definition 6. *Let \mathcal{A} be a finite alphabet not reduced to $\{0\}$, defined by $Q(x) = 0$, where Q is a polynomial of degree d with $d > 1$ distinct roots, and let \mathcal{G} be the subset of complex numbers γ , such that $\gamma\mathcal{A} \subset \mathcal{A}$.*

Note that, because \mathcal{A} is finite, numbers of \mathcal{G} are necessarily of unit modulus, and integer roots of unity (i.e., for every $\gamma \in \mathcal{G}$, there exists an integer $q(\gamma)$ such that $\gamma^{q(\gamma)} = 1$); this is true regardless of polynomial Q . In other words, $\gamma \mathcal{A} = \mathcal{A}, \forall \gamma \in \mathcal{G}$. Then, for every $\gamma \in \mathcal{G}$, the equation $\gamma \mathcal{A} = \mathcal{A}$ implies that $\gamma^{-1} \in \mathcal{G}$. Now, a similar statement as lemma 5 holds:

Lemma 7. *Trivial filters associated with hypotheses H5 and H6 are of the form $P \check{D}[z]$, where the entries of $\check{D}[z]$ can be written as $D_{pp}[z] = \gamma_p z^n$, with $\gamma_p \in \mathcal{G}$ and $n \in \mathbb{Z}$.*

It can thus be seen that the size of the set of trivial filters (characterizing the inherent indeterminacy of the problem) depends on the nature of the alphabet.

4. CONTRASTS

When noise is present in model (2), the estimation of inputs can be carried out according to a Maximum Likelihood (ML) or a Maximum A Posteriori (MAP) procedure if the noise has a known distribution. If this is not the case, noise must be considered as a nuisance. Contrast criteria are dedicated to this kind of situation.

4.1. Definition

Let \mathcal{H} be a set of filters, and denote $\mathcal{H} \cdot \mathcal{S}$ the set of processes obtained by operation of filters of \mathcal{H} on processes of \mathcal{S} . Denote \mathcal{T} the subset of \mathcal{H} of trivial filters, defined in lemma 4. An optimization criterion, $\Upsilon(\mathbf{H}; \mathbf{x})$, is referred to as a contrast, defined on $\mathcal{H} \times \mathcal{H} \cdot \mathcal{S}$, if it satisfies the three properties below [12]:

- P1 Invariance:** The contrast should not change within the set of acceptable solutions, which means that
 $\forall \mathbf{H} \in \mathcal{T}, \forall \mathbf{x} \in \mathcal{H} \cdot \mathcal{S}, \Upsilon(\mathbf{H}; \mathbf{x}) = \Upsilon(\mathbf{I}; \mathbf{x})$.
- P2 Domination:** If sources are already separated, any filter should decrease the contrast. In other words,
 $\forall \mathbf{s} \in \mathcal{S}, \forall \mathbf{H} \in \mathcal{H}$, then $\Upsilon(\mathbf{H}; \mathbf{s}) \leq \Upsilon(\mathbf{I}; \mathbf{s})$.
- P3 Discrimination:** The maximum contrast should be reached only for filters linked to each other via trivial filters:
 $\forall \mathbf{s} \in \mathcal{S}, \Upsilon(\mathbf{H}; \mathbf{s}) = \Upsilon(\mathbf{I}; \mathbf{s}) \Rightarrow \mathbf{H} \in \mathcal{T}$.

The most natural criterion to measure the statistical mutual independence between P variables z_p is the divergence between the joint probability density and the product of the marginal ones. If we assume the Kullback-Leibler divergence, we end up with the Mutual Information (MI):

$$I(pz) = \int pz(\mathbf{u}) \log \frac{pz(\mathbf{u})}{\prod_p p_{z_p}(u_p)} d\mathbf{u}. \tag{6}$$

This quantity is indeed positive, and vanishes if and only if random variables z_p are mutually independent. It is also invariant by scale change. The MI is thus a first possible contrast function [11] [49].

However, its practical use is rather difficult, especially in large dimension (e.g. convolutive mixtures), even if some iterative algorithms have been devised [51]; the MI is nevertheless

a powerful tool for separating Non Linear mixtures [58] [3]. Therefore, contrasts based on cumulants are often preferred in the context of linear mixtures.

4.2. Construction of contrast functionals based on cumulants

In this section, we restrict our attention to the construction of optimization criteria based on marginal standardized cumulants of the equalizer output.

Lemma 8. *If ϕ is a real convex function of a real variable, and if a_j are real positive numbers, $1 \leq j \leq P$, $\sum_j a_j \leq 1$, then $\phi(\sum_j a_j x_j) \leq \sum_j a_j [\phi(x_j) - \phi(0)] + \phi(0)$*

Proof. It suffices to add the coefficient $a_0 = 1 - \sum_j a_j$ to weight the origin. \square

Lemma 9. *Let $\{\alpha_{ij}\}$ be a set of positive real numbers, $1 \leq i, j \leq P$, such that **(A1)** $\sum_j \alpha_{ij} \leq 1$ and **(A2)** $\sum_i \alpha_{ij} \leq 1$ and let ϕ be convex as above. Then $\sum_i \phi(\sum_j \alpha_{ij} x_j) \leq \sum_j \phi(x_j)$.*

Proof. From lemma 8 and **(A1)**, one gets $\phi(\sum_j \alpha_{ij} x_j) \leq \sum_j \alpha_{ij} [\phi(x_j) - \phi(0)] + \phi(0)$. By summing this over i , and then using **(A2)**, one obtains $\sum_i \phi(\sum_j \alpha_{ij} x_j) \leq \sum_j [\phi(x_j) - \phi(0)] + P\phi(0)$. \square

Lemma 10. *Let $\{\alpha_{ij}\}$ be as in lemma 9, and ψ be a real concave function of a real variable, then $\sum_i \psi(\sum_j \alpha_{ij} x_j) \geq \sum_j \psi(x_j)$.*

The proof is similar.

Consider now a paraunitary equalizer, $\check{U}[z]$. The lemma below, already proved in [44], defines a class of contrast criteria, encompassing those proposed in [12], allowing to optimize $\check{U}[z]$:

Lemma 11. *If ϕ is monotonous strictly increasing convex and positive, then $\sum_i \phi(|\mathcal{C}_{x_i}^{(r,p)}|) \leq \sum_j \phi(|\mathcal{C}_{s_j}^{(r,p)}|)$, $\forall p$.*

Proof. To prove this, one assumes the input of the equalizer is the source process itself, $\mathbf{s}[n]$. We have $\mathbf{x}[n] = \sum_k \mathbf{U}[k] \mathbf{s}[n-k]$. This yields $\mathcal{C}_{x_i}^{(r,p)} = \sum_{j,k} U_{ij}[k]^r U_{ij}[k]^{(r-p)*} \mathcal{C}_{s_j}^{(r,p)}$. Then using the triangle inequality and the monotonicity of ϕ leads to: $\phi(|\mathcal{C}_{x_i}^{(r,p)}|) \leq \phi(\sum_{j,k} |U_{ij}[k]|^r |\mathcal{C}_{s_j}^{(r,p)}|)$. Now, because $\check{U}[z]$ is paraunitary, $\sum_{j,k} |U_{ij}[k]|^2 = 1$ and $\sum_{i,k} |U_{ij}[k]|^2 = 1$. The first consequence is that $\sum_{j,k} |U_{ij}[k]|^r \leq \sum_{j,k} |U_{ij}[k]|^2$; and the second is that the coefficients $\alpha_{ij} \stackrel{\text{def}}{=} \sum_k |U_{ij}[k]|^r$ satisfy the assumptions of lemma 9. Applying this lemma yields $\sum_i \phi(|\mathcal{C}_{x_i}^{(r,p)}|) \leq \sum_j \phi(|\mathcal{C}_{s_j}^{(r,p)}|)$. This proves requirement **P2**.

It remains to prove that the equality can hold only when the filter is trivial. This is visible from the equality $\sum_{j,k} |U_{ij}[k]|^r = \sum_{j,k} |U_{ij}[k]|^2$, which can hold true only when only one entry of the polynomial matrix $\check{U}[z]$ is nonzero in each row and each column, hence the result **P3**. \square

Theorem 12. *Let ϕ be real positive strictly increasing convex, and ψ positive decreasing concave. If \mathcal{H} denotes the group of paraunitary filters, and \mathcal{S} the set of processes satisfying*

hypotheses H1, H2, and H4, then the functional below is a contrast defined on $\mathcal{H} \times \mathcal{H} \cdot \mathcal{S}$:

$$\Upsilon(\mathbf{x}) \stackrel{\text{def}}{=} \frac{\sum_i \phi(|\mathcal{C}_{x_i}^{(r,p)}|)}{\sum_i \psi(|\mathcal{C}_{x_i}^{(r,p)}|)}$$

Proof. Obviously, in the same manner as lemma 11, one can state another lemma for concave decreasing functions ψ . If ψ is not strictly decreasing, one loses the discrimination **P3**, but we still have $\sum_i \psi(\mathcal{C}_{x_i}^{(r,p)}) \geq \sum_j \psi(\mathcal{C}_{s_j}^{(r,p)})$. This inequality combined with lemma 11 proves that $\Upsilon(\mathbf{x}) \leq \Upsilon(\mathbf{s})$.

If equality holds, we must necessarily have $\sum_i \psi(\mathcal{C}_{x_i}^{(r,p)}) = \sum_j \psi(\mathcal{C}_{s_j}^{(r,p)})$ and $\sum_i \phi(\mathcal{C}_{x_i}^{(r,p)}) = \sum_j \phi(\mathcal{C}_{s_j}^{(r,p)})$. The strict monotonicity of ϕ then suffices to prove **P3**, as in the proof of lemma 11. \square

Now we come to a nice simple theorem allowing to combine criteria based on cumulants of possibly different orders, which allows better discrimination properties:

Theorem 13. *If $\Upsilon_k(\mathbf{y})$ are contrasts defined on $\mathcal{H} \cdot \mathcal{S}_k$, and $\{a_k\}$ are strictly positive numbers, then $\Upsilon(\mathbf{y}) \stackrel{\text{def}}{=} \sum_k a_k \Upsilon_k(\mathbf{y})$ is a contrast on $\mathcal{H} \cdot \bigcup_k \mathcal{S}_k$.*

Proof. Requirement **P2** is obtained immediately, because all terms are positive: $\Upsilon(\mathbf{x}) = \sum_k a_k \Upsilon_k(\mathbf{x}) \leq \sum_k a_k \Upsilon_k(\mathbf{s}) = \Upsilon(\mathbf{s})$. If equality holds, then $\sum_k a_k [\Upsilon_k(\mathbf{s}) - \Upsilon_k(\mathbf{x})] = 0$, which is possible only if every term vanishes because they are all positive. Thus $\Upsilon_k(\mathbf{x}) = \Upsilon_k(\mathbf{s})$, $\forall k$. But $\mathbf{s} \in \mathcal{S}_k$ for some k , by hypothesis. And since Υ_k is a contrast, one can conclude that $\mathbf{x} = \mathbf{T} \star \mathbf{s}$, for some trivial filter \mathbf{T} of \mathcal{H} . This proves the theorem. \square

Lastly, the theorem below is an obvious corollary of the definition of contrasts; Corollary 15 can be obtained by merely combining theorems 13 and 14, and using the same reasoning as in the proof of Theorem 12.

Theorem 14. *If $\Upsilon(\mathbf{y})$ is a contrast, then so is $\phi(\Upsilon(\mathbf{x}))$, for any positive strictly increasing function ϕ .*

Corollary 15. *Let ϕ be positive strictly increasing and ψ positive decreasing. Then, if $\Upsilon_k(\mathbf{x})$ are contrasts defined on $\mathcal{H} \cdot \mathcal{S}_k$, and $\{a_k\}$ and $\{b_k\}$ are positive numbers, $a_k > 0$, then*

$$\Upsilon(\mathbf{x}) \stackrel{\text{def}}{=} \frac{\phi(\sum_k a_k \Upsilon_k(\mathbf{x}))}{\psi(\sum_k b_k \Upsilon_k(\mathbf{x}))}$$

is a contrast on $\mathcal{H} \cdot \bigcup_k \mathcal{S}_k$.

4.3. Contrasts for discrete sources

Let us now turn to discrete inputs. We now assume Hypotheses H5 and H6 instead of H3 or H4. For general alphabets (as defined in hypothesis H5), trivial filters are given by Lemma 7, and a result similar to the above still holds true:

Theorem 16. *Let \mathcal{S} be the set of processes taking their values in alphabet \mathcal{A} , and \mathcal{H} the set of $P \times P$ invertible FIR filters. Then criterion*

$$\Upsilon(\mathbf{A}; \mathbf{x}) = - \sum_n \sum_i |Q(x_i(n))|^2 \tag{7}$$

is a contrast under hypotheses H5 and H6.

The proof of the theorem needs the lemma below.

Lemma 17. *Let $\mathcal{A} = \{x_n, 1 \leq n \leq d\}$ be a given finite set of complex numbers not reduced to $\{0\}$, and $\{c_i, 1 \leq i \leq P\}$ non zero complex coefficients. Then, if $\sum_{i=1}^P c_i x_{\sigma(i)} \in \mathcal{A}$, for all mappings σ , not necessarily injective, from $\{1, \dots, P\}$ to $\{1, \dots, d\}$, only one component c_k is non zero.*

The proof of the Lemma is given in appendix. Let us now prove theorem 16.

Proof. Let us prove property **P1**. For any trivial filter \mathbf{T} defined by lemma 7, we have $-\Upsilon(\mathbf{T}; \mathbf{x}) = \sum_i \sum_n |Q(\gamma x_{\sigma(i)}[n + \tau])|^2$, with $\tau \in \mathbf{Z}$, $\gamma \in \mathcal{G}$, and σ a permutation. Because of the sums, this can also be simply written $-\Upsilon(\mathbf{T}; \mathbf{x}) = \sum_j \sum_m |Q(\gamma x_j[m])|^2$. If $x_j[m]$ is in \mathcal{S} , then its value is in \mathcal{A} , and so is that of $\gamma x_j[m]$, by definition of \mathcal{G} . Thus $Q(\gamma x_j[m]) = 0$.

Property **P2** is obvious. In fact, $\sum_i \sum_n |Q(y_i[n])|^2 \geq 0$. Hence it is larger than or equal to $\sum_i \sum_n |Q(s_i[n])|^2$, because the latter is null when $\mathbf{s}[n] \in \mathcal{S}$. We have indeed $-\Upsilon(\mathbf{H}; \mathbf{s}) \geq -\Upsilon(\mathbf{T}; \mathbf{s})$.

Lastly, in order to prove property **P3**, we must show that if we have the equality $\sum_i \sum_n |Q(y_i[n])|^2 = 0$, then \mathbf{T} is trivial. Consider one component of vector $\mathbf{y}[n]$, and drop its subscript for the sake of convenience. Denote $y[n] = \sum_j c_j s_j[n]$, with $s_j[n] \in \mathcal{A}$, and where c_j corresponds to the j th row of filter \mathbf{H} . Then we have $\forall n, Q(y[n]) = 0$. From hypothesis H6, we thus have that $Q(\sum_j c_j x_j) = 0$, for all P -uplet $\{x_1, \dots, x_P\}$. We are in the conditions of Lemma 17, and we may conclude that a single c_k is non zero. In addition, this c_k is necessarily in \mathcal{G} since $c_k x$ must be in \mathcal{A} for any $x \in \mathcal{A}$. By proceeding in the same way for every component of vector $\mathbf{y}[n]$, we end up with an impulse response $\mathbf{H}[n]$ having only one non zero entry in every row. Because \mathbf{H} is invertible, it must eventually be of the form given in lemma 7. \square

4.4. Examples

- For SISO blind equalization, the Constant Modulus (CM) criterion is

$$J_{CM}(x) = \mathbb{E}\{(1 - x^2)^2\} \quad (8)$$

whereas the criterion of the Maximal kurtosis (KM) is

$$J_{KM}(x) = \frac{\mathbb{E}\{|x|^4\} - \mathbb{E}\{x^2\}^2}{\mathbb{E}\{|x|^2\}} - 2 = \frac{\mathcal{C}_x^{(4,2)}}{[\mathcal{C}_x^{(2,1)}]^2} \quad (9)$$

Under mild conditions (circularity at order 2), these two criteria are equivalent, because related by a monotonous decreasing function [22, ch.4] [53].

- Deflation approaches consist of extracting estimated sources one by one. Therefore, the equalizer has K inputs but only one output. Criteria as J_{CM} or J_{KM} can be used for that purpose [20] [60] [61] [47].

- For MIMO mixtures, the following criteria are contrasts, for any $r > 2$, $\alpha \geq 1$, $0 \leq p \leq r$ [12]:

$$J_\alpha^{(r,p)}(\mathbf{x}) = \sum_i |\mathcal{C}_{x_i}^{(r,p)}|^\alpha \quad (10)$$

Next, the Edgeworth expansion of the MI of a standardized (i.e. after whitening) real variable is given as follows, up to an additive constant I_o , as a function of standardized cumulants (i.e.

cumulants of the whitened variable) [11]:

$$-I = I_o + \sum_i 4 [\mathcal{C}_{x_i}^{(3)}]^2 + [\mathcal{C}_{x_i}^{(4)}]^2 + 7 [\mathcal{C}_{x_i}^{(3)}]^4 - 6 [\mathcal{C}_{x_i}^{(3)}]^2 \mathcal{C}_{x_i}^{(4)}$$

Note that a similar expression holds true in the complex case, though more complicated.

This approximation is not proved to be a contrast. But from theorem 13 and (10), the criterion below is a contrast:

$$J_2^{(3+4)}(\mathbf{x}) = \sum_i 4 |\mathcal{C}_{x_i}^{(3)}|^2 + |\mathcal{C}_{x_i}^{(4)}|^2 \tag{11}$$

and has the advantage to be discriminating for either 3rd or 4th order circular variables [11] [6], and still approximates the MI in some way. This kind of idea has also been recently exploited by jointly diagonalizing two cumulant tensors of different orders in [43]; the optimization criterion has nevertheless not been proved to be a contrast by the author.

- Again thanks to theorem 13, one can combine contrasts based on cumulants and others based on the discrete character. This has been for instance proposed in [41], even if the criterion used there was not proved to be a contrast by the authors.

- For discrete PSK- q inputs, $Q(z) = z^q$, and trivial filters are those given by Lemma 5 and criterion

$$\Upsilon(\mathbf{x}) \stackrel{\text{def}}{=} - \sum_i \sum_n |x_i[n]^q - 1|^2 \tag{12}$$

is a contrast if the matrix $(s_i[n])$ is full rank.

- Instead of maximizing marginal cumulants through criteria such as $J_2^{(4,2)}$, one can try to Jointly Approximately Diagonalize (JAD) several cumulant matrices. The advantage is that linear algebra tools can be utilized; One can prove that, under certain conditions, these JAD criteria are indeed contrast functions [8] [16] [17]. Consider the static MIMO case for instance.

Because the sum of all squared cumulants is invariant by unitary transform [11], minimizing all the squared moduli of *cross-cumulants* of \mathbf{x} amounts to maximizing the sum of the marginal ones. In [8], 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 (10):

$$\Upsilon_{JAD}^{(4,2)}(\mathbf{U}) = \sum_{ijk} |\mathcal{C}_{\mathbf{x};ij,ik}|^2. \tag{13}$$

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

$$\Upsilon_{JAD}^{(4,2)}(\mathbf{U}) = \sum_{ijk} \left| \sum_{pqmn} U_{ip} U_{jq} U_{im}^* U_{kn}^* \mathcal{C}_{\bar{\mathbf{y}};pq,mn} \right|^2$$

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

$$\Upsilon_{JAD}^{(4,2)}(\mathbf{U}) = \sum_{qn} \sum_i \left| \sum_{pm} U_{ip} \mathcal{C}_{\bar{\mathbf{y}};pq,mn} U_{im}^* \right|^2 \tag{14}$$

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

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 [18] [43]. For instance, in [18, ch.9], the STOTD contrast takes the form:

$$\Upsilon_{STO}^{(4,2)}(\mathbf{U}) = \sum_{pq} |\mathcal{C}_{\mathbf{x}; pq, pp}|^2 \quad (15)$$

• Let us turn to dynamic MIMO systems. After space-time prewhitening, the MIMO equalizer is paraunitary, as already pointed out earlier. One can then try to approximate this paraunitary constraint by a unitary constraint in a JAD framework [17]; the criterion proposed therein is also a contrast [16]. Another idea is to sweep the delays [4]. For the former, the contrast takes the form [16]:

$$\Upsilon_{JAD}^{(r,2)}(\check{\mathbf{U}}[z]) = \sum_i \sum_j \sum_{\ell} |\mathcal{C}_{\mathbf{x}}^{(r,2)}[i, \mathbf{j}, \ell]|^2, \quad (16)$$

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

$$\mathcal{C}_{\mathbf{x}}^{(r,p)}[i, \mathbf{j}, \ell] \stackrel{\text{def}}{=} \text{Cum}\left\{ \underbrace{x_i(n), \dots, x_i(n)}_{p \text{ terms}}, \underbrace{x_{j_1}(n - \ell_1), \dots, x_{j_q}(n - \ell_q)}_{q = r - p \text{ terms}} \right\}$$

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

$$\Upsilon_{JAD}^{(r,2)}(\check{\mathbf{U}}[z]) = \sum_{\mathbf{b}} \sum_{\gamma} \|\text{Diag}\{\mathbf{V}\mathbf{N}(\mathbf{b}, \gamma)\mathbf{V}^H\}\|^2 \quad (17)$$

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

4.5. Pathological cases: carrier mismatch

In the presence of carrier residual, real modulations (such as PAM or BPSK) raise problems, because they appear as being circular (even if they are not originally circular: $\mathbf{E}\{s[n]^2\} \neq 0$). In these particular cases, one can show that the kurtosis criterion can reach a maximum for filters that contain two masses [32], instead of one as desired. In all other cases, one can prove that most results for contrast-based SISO blind equalization hold true [9].

Another class of pathological cases encompasses the non Gaussian random variables that have a null kurtosis. One could think these variables are rare and never encountered; this is untrue. Take the example of a variable uniformly distributed on the unit circle with probability 1/2, and null with probability 1/2. This variable is encountered when the on-off modulation $\{0, 1\}$ suffers from a strong carrier offset [48]. This kind of variable is viewed as Gaussian by the algorithms based on cumulants or moments of order lower than or equal to 4.

5. NUMERICAL ALGORITHMS

5.1. Iterative ascent

Criteria J_{CM} and J_{KM} do not need prewhitening. J_{CM} can be minimized directly, and J_{KM} should be maximized under a norm constraint on the equalizer tap vector. The usual practice, at least in the SISO and deflation cases, is to run a gradient ascent of J_{KM} :

$$\mathbf{v} = \mathbf{f}(k) + \eta \mathbf{g}(k); \mathbf{f}(k+1) = \mathbf{v}/\|\mathbf{v}\| \tag{18}$$

where $\mathbf{g}(k)$ denotes the gradient of $J_{KM}(\mathbf{f})$ calculated at $\mathbf{f}(k)$, and η the step size. Standard gradient implementations, especially with a fixed step, perform poorly because of the shape of the criterion, which contains many saddle points. The way the step size is adjusted (e.g. quasi-Newton) does not improve anything with this respect: if the algorithm is initialized near a saddle point, the iterations can stay a long time in its neighborhood, and suddenly burst out far away from the attraction basin, and take again a long time to come back. Yet, a significant improvement can be brought to this.

In fact, assume $J(\mathbf{f})$ is a rational function in the f_i 's. This is the case for instance for KM and CM criteria. Then, $J(\mathbf{f}(k) + \eta \mathbf{g}(k))$ is a rational function in variable η . As a consequence, all its stationary points can be explicitly computed, as roots of a polynomial in a single variable, and the absolute minimum/maximum easily found [13][†] This kind of algorithm may sometimes give the possibility to leave the attraction basin of a local minimum, if any.

The contrast criteria that we proposed for discrete sources, e.g. (12), are of polynomial nature. Hence, they can also be maximized with the help of our algorithm above, searching for absolute maxima. The same remarks already made still hold. In the present case however, algebraic solutions may be computed, as reported in [28] [27], among others.

Other approaches, that are not based on contrast maximization, exist in the literature, including [64] [57] [62] [59] [63] [38] [28] [29] [19].

5.2. Quasi algebraic algorithms for static MIMO

Several algorithms have been proposed in the static (i.e. non frequency selective, encountered in flat fading channels) MIMO case (ICA). Roughly speaking, one can divide them into two families. First, the deflation type algorithms try to extract one source, and then subtract it in the observation by linear regression [20]; we already talked about that. Second, direct separating algorithms impose some structure on the filter \mathbf{F} in order to make sure it does not become singular, neither deliver several times the same source; this can be implemented adaptively or not [11] [46].

• As an illustration, consider the 2×2 separating filter, which can often appear as a building block of MIMO equalizers.

$$\mathbf{x} = \begin{pmatrix} c & s e^{j\varphi} \\ -s e^{-j\varphi} & c \end{pmatrix} \mathbf{y} \tag{19}$$

where the input \mathbf{y} has been whitened, $c = \cos \alpha$ is real positive, $s = \sin \alpha$ is real, and $\alpha, \varphi \in [-\pi/2, \pi/2[$. Parameters (α, φ) can be found by maximizing the contrast $J_1^{(4,2)} = \varepsilon \sum_i C_{x_i}^{(4,2)}$, if

[†]The Matlab/Scilab codes for J_{KM} and J_{CM} will be downloadable from www.i3s.unice.fr/~comon in the near future.

sources have the same kurtosis sign, ε . It can be shown that this maximization problem admits an analytical solution given by the dominant eigenvector $\mathbf{w} = [\cos 2\alpha, \sin 2\alpha \cos \phi, \sin 2\alpha \sin \phi]^\top$ of a 3×3 real quadratic form: $\mathbf{w}^\top \mathbf{B} \mathbf{w}$.

More precisely, let $r = 2q$ be even. We give here a way to algebraically maximize $J_1^{(r,q)}$. It has been shown in [18, ch.9] that the maximization of contrast (15) amounts to diagonalizing simultaneously a set of third-order tensors. For this purpose, it has been shown that $|\mathcal{C}_{12,1}^z|^2 + |\mathcal{C}_{21,2}^z|^2$ is a quadratic form in variable $\mathbf{w} = [\cos 2\alpha, \sin 2\alpha \cos \phi, \sin 2\alpha \sin \phi]^\top$. Here we report that a similar result holds true for contrast $\Upsilon_{2,2}^{(1)}$:

$$J_1^{(4,2)}(\mathbf{U}) = \mathbf{w}^\top \mathbf{B} \mathbf{w}, \quad (20)$$

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

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

where $\mathcal{C}_{ij,kl}$ stands for $\mathcal{C}_{\bar{y};ij,kl}$. This result is new.

- Another interesting example is that of the algorithm proposed in [6], where the optimization criterion is an approximation of $J_2^{(3+4)}$ in (11), leading to an efficient numerical algorithm, in the same spirit.

5.3. Quasi algebraic algorithms for dynamic MIMO

Only SIMO channels can be identified with order 2 statistics only, unless some specific additional information is available. For instance, the non circularity of sources is used in [29], and the non stationarity is used in [50]. The best we can do with second order statistics, is to end up with a static MIMO problem. This is what linear prediction approaches do. But one can proceed directly, as proposed in [4] [16] [17] for instance, by a Joint Approximate Diagonalization as suggested in (17).

6. CONCLUDING REMARKS

Several extensions of the contrast concept have been examined, and are currently being deepened. In particular, it has been shown that contrast criteria can be combined together to build more efficient new contrasts. Algebraic block solutions become more and more attractive, especially in TDMA transmissions, because of the increased computational power. Therefore, quasi algebraic solutions become more attractive and should be the subject of increased efforts. The issue of space-time prewhitening should not be overlooked in operational systems, and has not been addressed in the present paper. Other interesting directions of research include the impact of channel length misadjustment, Non Linear or under-determined mixtures; MCMC approaches can be promising in this kind of context, for the same reason.

ACKNOWLEDGEMENT

The matter of this paper has been presented at the second IMA Conference in Communications, 16-18

december, 2002, at Lancaster University. The author wishes to thank S. Icart and J. Lebrun for useful discussions related to the proof presented in appendix.

REFERENCES

1. K. ABED-MERAIM, E. MOULINES, P. LOUBATON, "Prediction error method for second-order blind identification", *IEEE Trans. Sig. Proc.*, vol. 45, no. 3, pp. 694–705, Mar. 1997.
2. S. I. AMARI, J. F. CARDOSO, "Blind source separation - semiparametric statistical approach", *IEEE Trans. Sig. Proc.*, vol. 45, no. 11, pp. 2692–2700, Nov. 1997.
3. M. BABAIE-ZADEH, C. JUTTEN, K. NAYEBI, "Blind separation of post-nonlinear mixtures", in *Int. Conf. Indep. Comp. Ana. (ICA'01)*, San Diego, Dec. 2001, pp. 138–143.
4. P. D. BAXTER, J. Mc WHIRTER, "Blind signal separation of convolutive mixtures", in *Second IMA Int. Conf. on Mathematics in Communications*, Lancaster, UK, Dec. 2002.
5. A. BELOUCHRANI, K. ABED-MERAIM, J. F. CARDOSO, E. MOULINES, "A blind source separation technique using second order statistics", *IEEE Trans. Sig. Proc.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
6. T. BLASCHKE, L. WISKOTT, "An improved cumulant based method for independent component analysis.", in *Proc. ICANN'02*, José R. Dorronsoro, Ed. 2002, Lecture Notes in Computer Science, pp. 1087–1093, Springer.
7. J. F. CARDOSO, "High-order contrasts for independent component analysis", *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan. 1999.
8. 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.
9. P. CIBLAT, P. LOUBATON, E. SERPEDIN et al., "Performance analysis of blind carrier frequency offset estimators for noncircular transmissions", *IEEE Trans. Sig. Proc.*, vol. 50, no. 1, pp. 130–140, Jan. 2002.
10. 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.
11. 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.
12. P. COMON, "Contrasts for multichannel blind deconvolution", *Signal Processing Letters*, vol. 3, no. 7, pp. 209–211, July 1996.
13. P. COMON, "Blind equalization with discrete inputs in the presence of carrier residual", in *Second IEEE Int. Symp. Sig. Proc. Inf. Theory*, Marrakech, Morocco, Dec. 2002, invited session.
14. P. COMON, "Tensor decompositions", in *Mathematics in Signal Processing V*, J. G. McWhirter, I. K. Proudler, Eds., pp. 1–24. Clarendon Press, Oxford, UK, 2002.
15. P. COMON, P. CHEVALIER, "Source separation: Models, concepts, algorithms and performance", in *Unsupervised Adaptive Filtering, Vol. 1, Blind Source Separation*, S. Haykin, Ed., Series on Adaptive and learning systems for communications signal processing and control, pp. 191–236. Wiley, 2000.
16. P. COMON, E. MOREAU, "Blind MIMO equalization and joint-diagonalization criteria", in *ICASSP'01*, Salt Lake City, May 7-11, 2001, vol. 5, pp. 2749–2752.
17. P. COMON, L. ROTA, "Blind separation of independent sources from convolutive mixtures", *IEICE Trans. on Fundamentals of Elec. Com. Comput. Sciences*, vol. E86-A, no. 3, Mar. 2003, invited.
18. L. de LATHAUWER, *Signal Processing based on Multilinear Algebra*, Doctorate, Katholieke Universiteit Leuven, Sept. 1997.
19. L. de LATHAUWER, B. de MOOR, J. VANDEWALLE, "On the best rank-1 and rank-(R1,R2,...RN) approximation of high-order tensors", *SIAM Jour. Matrix Ana.*, vol. 21, no. 4, pp. 1324–1342, 2000.
20. N. DELFOSSE, P. LOUBATON, "Adaptive blind separation of independent sources: a deflation approach", *Signal Processing*, vol. 45, pp. 59–83, 1995.
21. Z. DING, "Matrix outer-product decomposition method for blind multiple channel identification", *IEEE Trans. Sig. Proc.*, vol. 45, no. 12, pp. 3053–3061, Dec. 1997.
22. Z. DING, Y. LI, *Blind Equalization and Identification*, Dekker, New York, 2001.
23. 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.
24. E. GASSIAT, F. GAMBOA, "Source separation when the input sources are discrete or have constant modulus", *IEEE Trans. Sig. Proc.*, vol. 45, no. 12, pp. 3062–3072, Dec. 1997.
25. D. GODARD, "Self recovering equalization and carrier tracking in two dimensional data communication systems", *IEEE Trans. Com.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
26. A. GOROKHOV, P. LOUBATON, "Blind identification of MIMO-FIR systems: a generalized linear prediction approach", *Signal Processing, Elsevier*, vol. 73, pp. 105–124, Feb. 1999.

27. O. GRELLIER, P. COMON, "Blind separation and equalization of a channel with MSK inputs", in *SPIE Conference*, San Diego, July 19-24 1998, pp. 26-34, invited session.
28. O. GRELLIER, P. COMON, "Analytical blind discrete source separation", in *Eusipco*, Tampere, Finland, 5-8 sept. 2000.
29. O. GRELLIER, P. COMON, B. MOURRAIN, P. TREBUCHET, "Analytical blind channel identification", *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2196-2207, Sept. 2002.
30. F. GUSTAFSSON, B. WAHLBERG, "Blind equalization by direct examination of input sequences", *IEEE Transactions on Computers*, vol. 43, no. 7, pp. 2213-2222, July 1995.
31. S. HAYKIN, *Unsupervised Adaptive Filtering*, vol. 1, Wiley, 2000, series in Adaptive and Learning Systems for Communications, Signal Processing, and Control.
32. S. HOUCKE, A. CHEVREUIL, P. CIBLAT, "Estimation of the symbol period: the frequency offset case", in *Proc. XIth European Sig. Proc. Conf. (Eusipco)*, Sept. 2002.
33. A. HYVÄRINEN, J. KARHUNEN, E. OJA, *Independent Component Analysis*, Wiley, 2001.
34. C. JUTTEN, J. HÉRAULT, "Independent component analysis versus PCA", in *Proc. EUSIPCO*, Grenoble, France, 1988, pp. 643-646.
35. J. L. LACOUME, P. O. AMBLARD, P. COMON, *Statistiques d'ordre supérieur pour le traitement du signal*, Collection Sciences de l'Ingénieur. Masson, 1997.
36. T. W. LEE, *Independent Component Analysis*, Kluwer, 1998.
37. T. H. LI, "Analysis of a non-parametric blind equalizer for discrete valued signals", *IEEE Trans. on Sig. Proc.*, vol. 47, no. 4, pp. 925-935, Apr 1999.
38. T. H. LI, K. MBAREK, "A blind equalizer for nonstationary discrete-valued signals", *IEEE Trans. Sig. Proc.*, vol. 45, no. 1, pp. 247-254, Jan. 1997, Special issue on communications.
39. P. LIAVAS, P. A. REGALIA, J-P. DELMAS, "On the robustness of the linear prediction method for blind channel identification", *IEEE Trans. Sig. Proc.*, vol. 48, pp. 1477-1481, May 2000.
40. P. LOUBATON, E. MOULINES, P. A. REGALIA, "Subspace methods for blind identification and deconvolution", in *Signal Processing Advances in Wireless and Mobile Communications*, Giannakis, Hua, Stoica, Tong, Eds., chapter 3. Prentice-Hall, 2001.
41. Y. LUO, J. CHAMBERS, "A modified cross-correlation and constant modulus algorithm with constellation matched error penalty", in *Second IMA Int. Conf. on Mathematics in Communications*, Lancaster, UK, Dec. 2002.
42. P. McCULLAGH, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.
43. E. MOREAU, "A generalization of joint-diagonalization criteria for source separation", *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 530-541, March 2001.
44. 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.
45. E. MOULINES, P. DUHAMEL, J. F. CARDOSO, S. MAYRAGUE, "Subspace methods for the blind identification of multichannel FIR filters", *IEEE Trans. Sig. Proc.*, vol. 43, no. 2, pp. 516-525, Feb. 1995.
46. C. B. PAPADIAS, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion", *IEEE Trans. Sig. Proc.*, vol. 48, pp. 3508-3519, Dec. 2000.
47. C. B. PAPADIAS, "A multi-user kurtosis algorithm for blind source separation", in *ICASSP*, Istanbul, June 5-9 2000.
48. N. PETROCHILOS, P. COMON, "Separation of ZCM signals: Application is ssr radar", in *GRETSI'03*, Paris, 8-11 Sept. 2003.
49. D. T. PHAM, "Blind separation of instantaneous mixture of sources via an independent component analysis", *IEEE Trans. Sig. Proc.*, vol. 44, no. 11, pp. 2768-2779, Nov. 1996.
50. D. T. PHAM, J-F. CARDOSO, "Blind separation of instantaneous mixtures of nonstationary sources", *IEEE Trans. Sig. Proc.*, vol. 49, no. 9, pp. 1837-1848, Sept. 2001.
51. 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.
52. J. G. PROAKIS, *Digital Communications*, McGraw-Hill, 1995, 3rd edition.
53. P. A. REGALIA, "Blind deconvolution and source separation", in *Mathematics in Signal Processing V*, J. G. McWhirter, I. K. Proudler, Eds., pp. 25-35. Clarendon Press, Oxford, UK, 2002.
54. P. A. REGALIA, M MBOUP, "Properties of some blind equalization criteria in noisy multiuser environments", *IEEE Trans. Sig. Proc.*, pp. 3112-3122, Dec. 2001.
55. Y. SATO, "A method of self recovering equalization for multilevel amplitude-modulation systems", *IEEE Trans. Com.*, vol. 23, pp. 679-682, June 1975.
56. D. T. M. SLOCK, "Blind joint equalization of multiple synchronous mobile users using oversampling and/or multiple antennas", in *Proc. 28th Asilomar Conf. Sig. Syst. Comp.*, Pacific Grove, CA, Oct. 31 - Nov. 2 1994, pp. 1154-1158.

57. A. L. SWINDLEHURST, S. DAAS, J. YANG, "Analysis of a decision directed beamformer", *IEEE Trans. Sig. Proc.*, vol. 43, no. 12, pp. 2920–2927, Dec. 1995.
58. A. TALEB, C. JUTTEN, "Source separation in post-nonlinear mixtures", *IEEE Trans. Sig. Proc.*, vol. 47, pp. 2807–2820, Oct. 1999.
59. 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.
60. J. R. TREICHLER, M. G. LARIMORE, "New processing techniques based on the constant modulus adaptive algorithm", *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 33, no. 2, pp. 420–431, Apr. 1985.
61. 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.
62. A. J. van der VEEN, S. TALWAR, A. PAULRAJ, "Blind estimation of multiple digital signals transmitted over FIR channels", *IEEE Sig. Proc. Letters*, vol. 2, no. 5, pp. 99–102, May 1995.
63. T. WIGREN, "Avoiding ill-convergence of finite dimensional blind adaptation schemes excited by discrete symbol sequences", *Signal Processing, Elsevier*, vol. 62, no. 2, pp. 121–162, Oct. 1997.
64. D. YELLIN, B. PORAT, "Blind identification of FIR systems excited by discrete-alphabet inputs", *IEEE Trans. Sig. Proc.*, vol. 41, no. 3, pp. 1331–1339, 1993.

APPENDIX

Proof of Lemma 17. The proof is in three stages. Assume that, for some $\mathbf{c} \in \mathbf{C}^P$, we have that $\mathbf{x}^\top \mathbf{c} \in \mathcal{A}$, $\forall \mathbf{x} \in \mathcal{A}^P$. We show below that \mathbf{c} must then be trivial. The idea is to prove that a non trivial vector \mathbf{c} generates symbols that may lie outside the convex hull of alphabet \mathcal{A} .

Preliminaries

Denote $\mathcal{A} = \{a_1, \dots, a_d\}$, $\mathbf{c}^\top = [c_1, \dots, c_P]$, and $\mathbf{1}$ the vector formed of P ones.

• Because $\mathbf{x}^\top \mathbf{c} \in \mathcal{A}$ for any vector \mathbf{x} containing elements of \mathcal{A} , it must be true in particular for $\mathbf{x}^\top = a_p \mathbf{1}^\top$. This implies that $\forall a_p \in \mathcal{A}$, $a_p \sum_i c_i \in \mathcal{A}$. In other words, we must always have

$$\sum_i c_i \in \mathcal{G}. \quad (21)$$

• Next, for any pair of distinct complex numbers a and b , define the $P \times P$ matrix $\mathbf{B} = (a-b)\mathbf{I} + b\mathbf{1}\mathbf{1}^\top$. This matrix has a determinant equal to $(a-b)^{P-1}(a+(P-1)b)$. For $a+(P-1)b \neq 0$, its inverse takes the form

$$\mathbf{B}^{-1} = \frac{1}{a-b} \left[\mathbf{I} - \frac{b}{a+(P-1)b} \mathbf{1}\mathbf{1}^\top \right] \quad (22)$$

• As a result, for any pair of distinct symbols of \mathcal{A} , a and b , there exists a vector $\boldsymbol{\alpha}$ containing P symbols of \mathcal{A} such that $\mathbf{B}\mathbf{c} = \boldsymbol{\alpha}$. From (22), we have in particular

$$\forall (a, b) \in \mathcal{A}^2 / a \neq b, a + (P-1)b \neq 0 : \sum_i c_i = [a + (P-1)b]^{-1} \sum_i \alpha_i \quad (23)$$

Stage 1: Case of real symbols, with $\mathcal{G} = \{1\}$

Denote $x_m = \min\{x, x \in \mathcal{A}\}$ and $x_M = \max\{x, x \in \mathcal{A}\}$. We Consider the i th row of $\mathbf{B}\mathbf{c} = \boldsymbol{\alpha}$. More precisely, from (21), $\sum_i c_i = 1$ and:

$$\forall a, b \in \mathcal{A}, \exists \alpha_i \in \mathcal{A} / \alpha_i = c_i a + (1 - c_i) b$$

So c_i must be real too. By induction, defining $\beta_0 = a$, the real number $\beta_k = c_i \beta_{k-1} + (1 - c_i) b$ is also in \mathcal{A} , $\forall k > 0$. Yet, the writing $\beta_k = b + c_i^k (a - b)$ shows that either $c_i = 0$, or $|c_i| = 1$, otherwise the series would generate infinitely many distinct symbols in \mathcal{A} . So we have proved so far that $c_i \in \{-1, 0, 1\}$, $\forall i, 1 \leq i \leq P$. If $c_i \in \{0, 1\}$, $\forall i$, then again from $\sum_i c_i = 1$, there is a single nonzero entry in \mathbf{c} , and \mathbf{c} is eventually trivial. So the last case to consider is: $\exists c_i = -1, 1 \leq i \leq P$. This case is irrelevant. In fact, $\beta = b + c_i(a - b) \in \mathcal{A}$ for any pair $(a, b) \in \mathcal{A}^2$; in particular for $c_i = -1$, $a = x_M$ and $b = x_m$, the symbol $\beta = x_m - (x_M - x_m)$ should belong to \mathcal{A} . But $x_M > x_m \Leftrightarrow \beta < x_m$, which contradicts the definition of x_m .

Stage 2: Complex case with $\mathcal{G} = \{1\}$

• If $d = 2$, the problem is equivalent to a particular case of real alphabet. This case has been consequently already addressed.

• Now assume $d > 2$, and choose a symbol b on the convex hull of \mathcal{A} . Since $d > 2$, b always has two distinct neighbors on the convex hull (unless symbols are real, but this has been already examined). So choose one of the two neighbors on the convex hull, denoted a , in order to also have $a + (P-1)b \neq 0$ (equality cannot hold for both). Result (23) then applies. Since $\sum_i c_i \in \mathcal{G}$, (23) yields

$$\frac{a + (P-1)b}{P} = \frac{1}{P} \sum_i \alpha_i \quad (24)$$

Let us prove first that α cannot be proportional to $\mathbf{1}$. Assume $\alpha = a_o \mathbf{1}$ for some $a_o \in \mathcal{A}$. Then from (24), $a_o = \frac{1}{P} a + \frac{P-1}{P} b$. Therefore symbol a_o is also on the convex hull of \mathcal{A} , and is closest to b than a was. This contradicts the fact that a was one of its two neighbors of b .

So assume now that vector α contains at least two distinct symbols. If these symbols are a and b , then we necessarily have 1 time a and $(P - 1)$ times b , and c is trivial, as already seen. If all symbols of α are real, this case is equivalent to stage 1, and has been already treated. Thus assume there is in α a third symbol x distinct from a and b , being not a real linear combination of a and b . From (23), there must be at least another symbol x' on the other side of the line spanned by $\{a, b\}$. But then one of them lies outside the convex hull of \mathcal{A} . This contradicts the fact that both x and x' are in \mathcal{A} .

Stage 3: Case $\mathcal{G} \neq \{1\}$

This is the general case. We have seen in section 3.2 that 1 indeed always belongs to \mathcal{G} , but that there may be other complex numbers γ such that $\gamma\mathcal{A} \subset \mathcal{A}$. We pointed out that these numbers must be roots of unity, which means that if $\sum_i c_i \in \mathcal{G}$, then there necessarily exists a number $\gamma \in \mathcal{G}$ such that $\gamma \sum_i c_i = 1$. Let $c'_i = \gamma c_i$. All the previous reasoning then applies with c'_i instead of c_i .

We have proved that vector c must be trivial in all cases of complex finite alphabets. \square