IMA Conf. Mathematics in Communications II, Lancaster University, UK, 16-18 Dec 2002

INDEPENDENT COMPONENT ANALYSIS, CONTRASTS, AND CONVOLUTIVE MIXTURES

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

Contents

The paper is organized as follows. Section 1 states the problem, assumptions, and notations, and includes a brief partial bibliographical survey. Section 2 is a reminder about whitening operations and cumulants. Section 3 defines a wide class of contrast functionals and gives several examples; some of them are new. Practical algorithms are presented in section 4. The effect of a carrier frequency offset is analyzed in section 5. Eventually, specific tools are proposed in section 6, when transmitted signals are discrete.

1. THE MIMO BLIND DECONVOLUTION PROBLEM

Blind equalization or identification schemes have been the subject of intense interest since the works of Sato [41] in 1975, and then Godard [19] 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.

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.

1.1. Modeling.

Limiting our discussion to linear modulations, the complex envelope of a transmitted signal $s(t)$ takes the form in baseband [39]: $s(t) = \sum_k g(t - kT)\, s[k]$. 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:

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

with $h[k] \stackrel{\text{def}}{=} h(kT_s)$. For Multiple Input Multiple Output (MIMO) systems, the transmitted signal $\boldsymbol{s}[k]$ and the received signal $\boldsymbol{y}[k]$ 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:

$$\boldsymbol{y}[n] = \sum_k \boldsymbol{H}[n - k]\, \boldsymbol{s}[k] \qquad (2)$$

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

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

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:

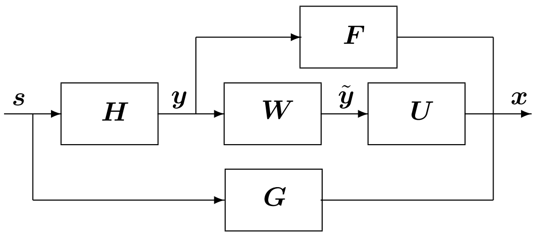
$$\boldsymbol{y}[n] = \sum_k \boldsymbol{H}(nT' - kT_s)\, \boldsymbol{s}[k]$$

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:

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

where δ is a small real number, and the dotless $\jmath$ denotes $\sqrt{-1}$. This situation will be subsequently considered as well, mainly in section 5.

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.



The transmitted sequence $\boldsymbol{s}[n]$ propagates through a channel $\boldsymbol{H}[z]$, is then whitened by a filter $\boldsymbol{W}[z]$, and is eventually deconvolved by a paraunitary equalizer $\boldsymbol{U}[z]$ to yield output $\boldsymbol{x}[n]$.

1.2. Assumptions and taxonomy.

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

Hypothesis 1 *Sources $s_j[k]$ are all i.i.d. sequences.*

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

Hypothesis 2 *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 [11]. 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 1 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 1 and 2:

Hypothesis 3 *At most one source is Gaussian.*

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 1.4). These three cases are not addressed in the present paper, and yield quite different (easier) theoretical problems. Nevertheless, the special framework of *discrete sources*, which is relevant in the context of digital communications, will be addressed in section 6.

Depending on hypotheses (independence, discrete character, SISO/MIMO, $P > K$ or not...), a whole variety of problems can be stated [11]. From now on, and unless otherwise specified, we shall concentrate only on Hypotheses 1 to 3, and on the case where $P \le K$ (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 [10].

1.3. Trivial filters.

The separating linear filter, $\boldsymbol{F}[z]$, if it exists, aims at delivering an output, $\boldsymbol{x}[n]$, which should satisfy as well as possible hypotheses 1 and 2. But it is clear that there exist some filters that do not affect them. These are called the trivial filters, and we can prove

Proposition 4 *Under hypotheses 1 to 3, trivial filters are of the form $\boldsymbol{T}[z] = \boldsymbol{P}\,\boldsymbol{D}[z]$, where $\boldsymbol{P}$ is a permutation, and $\boldsymbol{D}[z]$ a diagonal filter. In addition, because of the i.i.d. property of hypothesis 1, entries of $\boldsymbol{D}[z]$ must be of the form $D_{pp}[z] = \lambda_p\, z^{d_p}$, where d_p is an integer.*

Consequently, it is hopeless to estimate the pair $(H[z], \boldsymbol{s}[k])$. 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 $\boldsymbol{y}[n] = \boldsymbol{H}\,\boldsymbol{s}[n]$, where $\boldsymbol{y}[n]$ and $\boldsymbol{s}[n]$ are realizations of random variables. In that case, hypothesis 1 is not mandatory anymore. The estimation of the pair $(\boldsymbol{H}, \boldsymbol{s}[n])$ from the sole observations $\boldsymbol{y}[n]$ under hypotheses 2 and 3, has been introduced originally by Jutten and Comon [28] [5], and is now called *Independent Component Analysis* (ICA) [6] [37] [27] [29] [25] [10].

1.4. Bibliographical comments.

The exploitation of source color and spectral differences has been investigated in [3], among others. The case of non stationary or cyclo-stationary sources is studied for instance in

[38] [17]. MIMO blind identification/equalization via subspace fitting is described in [35] [32], and via linear prediction in [5] [15] [20]. For more references on the use of the discrete character of sources in MIMO over-determined mixtures, either static or convolutive, see [23] [48] [45] [42] [43] [47] [31] [18] [30] [24]. Underdetermined mixtures are addressed in [10] and references therein. For a general account on blind techniques, see [25] [27] [29] [16].

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

The paraunitarity means that

$$\boldsymbol{U}[z]\,\boldsymbol{U}^{\mathrm{H}}[1/z^*] = \boldsymbol{I} \tag{4}$$

where $\boldsymbol{I}$ is the identity. This property characterizes lossless filters. In the scalar case for instance, $U[z]$ is all-pass and $R[z]$ 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 [11]. From now on, we shall retain this procedure, although it is not always optimal, because is 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) [33]. 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 [33].

For a stationary process $\boldsymbol{y}[n]$, we define the ith marginal rth order cumulant as:

$$\mathcal{C}_{y_i}^{(r,p)} \stackrel{\mathrm{def}}{=} \mathrm{Cum}\{\underbrace{y_i, \ldots y_i}_{p \text{ terms}}, \underbrace{y_i^*, \ldots y_i^*}_{r-p \text{ terms}}\} \tag{5}$$

The 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.

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

Property 6 *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, $\mathtt{E}\{x^m\} = 1$, of course. But more interestingly, $\mathtt{E}\{x^p\} = 0$ 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:

Definition 7 *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. 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.

3.1. 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 proposition 4. An optimization criterion, $\Upsilon(\boldsymbol{H};\boldsymbol{x})$, is referred to as a contrast, defined on $\mathcal{H}\times\mathcal{H}\cdot\mathcal{S}$, if it satisfies the three properties below [7]:

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

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):

$$I(f_{\boldsymbol{z}}) = \int f_{\boldsymbol{z}}(\boldsymbol{u}) \log \frac{f_{\boldsymbol{z}}(\boldsymbol{u})}{\prod_p f_{z_p}(u_p)} \, d\boldsymbol{u}. \tag{6}$$

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.

However, its practical use is rather difficult, especially in large dimension (e.g. convolutive mixtures), even if some iterative algorithms have been devised [2]. Therefore, contrasts based on cumulants are often preferred.

3.2. Construction of contrast functionals

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. □

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 $\sum_i \phi(\sum_j \alpha_{ij} x_j) \leq \sum_j [\phi(x_j) - \phi(0)] + P\phi(0)$. □

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, $\boldsymbol{U}[z]$. The lemma below, already proved in [34], defines a class of contrast criteria, encompassing those proposed in [7], allowing to optimize $\boldsymbol{U}[z]$:

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

Proof. To prove this, one assumes the input of the equalizer is the source process itself, $\boldsymbol{s}[n]$. We have $\boldsymbol{x}[n] = \sum_k \boldsymbol{U}[k]\boldsymbol{s}[n-k]$. 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 triangular 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 $\boldsymbol{U}[z]$ 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 $\boldsymbol{U}[z]$ is nonzero in each row and each column, hence the result **P3**. □

Theorem 12 *Let ϕ be real positive strictly increasing convex, and ψ positive decreasing concave. Then the functional below is a contrast*

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

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 looses 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 $\Upsilon(\boldsymbol{x}) \leq \Upsilon(\boldsymbol{s})$.
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. □

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

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

$\boldsymbol{x} = \boldsymbol{T} \star \boldsymbol{s}$, for some trivial filter $\boldsymbol{T}$ of $\mathcal{H}$. This proves the theorem. □

Lastly, the theorem below is an obvious corollary of the definition of contrasts, and corollary 15 can be obtained by merely combining theorems 13 and 14:

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

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

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

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

3.3. Examples

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

$$J_{CM}(x) = \mathtt{E}\{(1-x^2)^2\} \tag{7}$$

whereas the criterion of the Maximal kurtosis (KM) is

$$J_{KM}(x) = \frac{\mathtt{E}\{|x|^4\} - \mathtt{E}\{x^2\}^2}{\mathtt{E}\{|x|^2\}} - 2 = \frac{\mathcal{C}_x^{(4,2)}}{[\mathcal{C}_x^{(2,1)}]^2} \tag{8}$$

Under mild conditions, these two criteria are equivalent, because related by a monotonous decreasing function [16, ch.4] [40].

For MIMO mixtures, the following criteria are contrasts, for any $r > 2$ [7]:

$$J_r(\boldsymbol{x}) = \sum_i |\mathcal{C}_{x_i}^{(r)}|^2 \tag{9}$$

Next, the Edgeworth expansion of the MI of a standardized real variable (i.e. after whitening) is given as follows up to an additive constant I_o, as a function of standardized cumulants (i.e. cumulants of the whitened variable) [6]:

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

This approximation is not proved to be a contrast. But from theorem 13 and (9), the criterion below is a contrast:

$$J_{3+4}(\boldsymbol{x}) = \sum_i 4\,|\mathcal{C}_{x_i}^{(3)}|^2 + |\mathcal{C}_{x_i}^{(4)}|^2$$

and has the advantage to be discriminating for either 3rd or 4th order circular variables.

4. NUMERICAL ALGORITHMS

4.1. SISO

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 is to run a gradient ascent of J_{KM}:

$$\boldsymbol{v} = \boldsymbol{f}(k) + \eta\,\boldsymbol{g}(k);\ \boldsymbol{f}(k+1) = \boldsymbol{v}/||\boldsymbol{v}|| \tag{10}$$

where $\boldsymbol{g}(k)$ denotes the gradient of $J_{KM}(\boldsymbol{f})$ calculated at $\boldsymbol{f}(k)$, 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 $J(\boldsymbol{f})$ is a rational function in the f_i's. This is the case for instance for KM and CM criteria. Then, $J(\boldsymbol{f}(k) + \eta\,\boldsymbol{g}(k))$ 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 [9]. This kind of algorithm may sometimes give the possibility to leave the attraction basin of a local minimum, if any.

4.2. Pathological cases

In the presence of carrier residual, real modulations (such as PAM or BPSK) raise problems, because they become circular (even if they are not originally circular: $\mathtt{E}\{s[n]^2\} \neq 0$). In these particular cases, one can show that the kurtosis criterion can reach a maximum for filters that contain two masses [26], instead of one as desired.

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 residual. This kind of variable is viewed as Gaussian by the algorithms based on cumulants or moments of order lower than or equal to 4.

4.3. Static MIMO

Several algorithms have been proposed in the static 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 [14]. Second, direct separating algorithms impose some structure on the filter $\boldsymbol{F}$ in order to make sure it does not become singular and delivers several times the same source; this can be implemented adaptively or not [6] [36].

As an illustration, consider the 2×2 separating filter

$$\boldsymbol{x} = \begin{pmatrix} c & s\,e^{\jmath\varphi} \\ -s\,e^{-\jmath\varphi} & c \end{pmatrix} \boldsymbol{y} \tag{11}$$

where the input $\boldsymbol{y}$ 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 $J_4 = \varepsilon \sum_i \mathcal{C}_{x_i}^{(4,2)}$, if sources have the same kurtosis sign, ε. It can be shown that this maximization problem admits an analytical solution given by the dominant eigenvector $\boldsymbol{w} = [\cos 2\alpha,\, \sin 2\alpha \cos\phi,\, \sin 2\alpha \sin\phi]^{\mathsf{T}}$ of a 3×3 real quadratic form: $\boldsymbol{w}^{\mathsf{T}}\, \boldsymbol{B}\, \boldsymbol{w}$. See [8] for the exact expression of $\boldsymbol{B}$ as a function of cumulants of $\boldsymbol{y}$.

4.4. Convolutive MIMO

Only SIMO channels can be identified with order 2 statistics only, unless some specific additional information is available. 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, as proposed already in 1990 in [5], and then later independently redeveloped [15] [1]. Approaches based on signal/noise subspaces [32] are very sensitive to channel length misadjustment, and are not considered as attractive anymore. Contrast-based approaches can be found in [44] [16] [13] [12]. Space is lacking to give here a more detailed description.

5. PRESENCE OF CARRIER RESIDUAL

In this section, we show how we can get rid easily of the carrier residual in SISO transmission schemes.

5.1. Carrier offset mitigation in SISO channels

The usual practice, for symbol sequences that are not circular at order q, is to compute the absolute maximum of some second or fourth order non-conjugated cyclic correlation [46]

$$Arg\,\underset{\alpha}{\mathrm{Max}} \left| \frac{1}{N} \sum_{n=0}^{N-1} z^q[n]\, e^{-\jmath\alpha n} \right|$$

which requires an exhaustive search. Even usual data-aided solutions require an exhaustive search [4]. On the contrary, the proposition below gives a closed-form solution [9].

Lemma 16 *Let $x[n]$ be a stationary discrete-time process with finite qth order moment, and $z[n] = x[n]\, e^{\jmath\, n\, \delta}$. Define*

$$\mu_q \overset{\mathrm{def}}{=} \mathtt{E}\{x[n]^q\}, \text{ and } m_q(T) \overset{\mathrm{def}}{=} \frac{1}{T} \sum_{n=0}^{T-1} z[n]^q. \tag{12}$$

Then, for $\delta \ll 1 \ll T$, $T\delta \leq A$, for some $A = O(1)$, we have:

$$m_q(T) \approx \mu_q\, e^{\jmath\, qT\delta/2}\, \frac{\sin qT\delta/2}{qT\delta/2} \tag{13}$$

A direct consequence of this lemma is that one can compute the carrier residual, after blind equalization, from two sample moments calculated with two different data lengths [9]:

Proposition 17 *Under the assumptions of lemma 16, δ can be estimated by*

$$\hat{\delta} = \frac{2}{qT}\, \mathrm{acos} \left(\frac{|m_q(2T)|}{|m_q(T)|} \right) \tag{14}$$

Now in the SISO case, there is a duality between source and observation. By inspection of (3), the residual can indeed be moved to the observation:

$$y[n] = e^{\jmath\, n\, \delta} \sum_k \bar{h}[n-k]\, s[k] \tag{15}$$

if we define $\bar{h}[p] = h[p]\, e^{-\jmath\, p\, \delta}$. It turns out that Proposition 17 can still be utilized to compensate the carrier mismatch *before* equalization.

Note that this quite natural role exchange does not hold however in the MIMO case, where carrier residuals have no reason to be identical for each source (think of the reception at a base-station in the up-link, for instance).

5.2. Carrier offset mitigation in MIMO channels

As just explained, it seems easier to run blind space-time channel equalization in a first stage. Then the equalizer outputs are supposed to be statistically independent. Therefore, in a second stage, one can process outputs one by one, by the procedure described for SISO channels.

6. THE CASE OF DISCRETE SOURCES

6.1. Modeling, contrast, and trivial filters

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

Assumption 18 *The sources $s_j[n]$ belong to a known finite alphabet $\mathcal{A}$ characterized by the complex roots of a polynomial $P(z) = 0$, and are sufficiently exciting.*

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

Theorem 19 *Let $\mathcal{S}$ be the set of PSK$-q$ processes, and $\mathcal{H}$ the set of $P \times P$ invertible FIR filters. Then*

$$\Upsilon(\boldsymbol{x}) \stackrel{\text{def}}{=} -\sum_i \sum_n \left| x_{i,n}^q - 1 \right|^2$$

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

Proof. This proof has been derived with the help of Jérôme Lebrun. We prove the theorem for static mixtures: $\boldsymbol{x}[n] = \boldsymbol{A}\,\boldsymbol{s}[n]$. First, the contrast is obviously null for $x \in \mathcal{S}$, and always negative. Thus $\Upsilon(\boldsymbol{x}) \leq \Upsilon(\boldsymbol{s})$. If equality holds, this means that $x_{i,n}^q = 1$ for all (i, n). We have a system of (polynomial) equations in unknowns A_{ip}. The matrix of source signals is formed of qth roots of unity. If it is full rank, this firstly means that $P \leq q$, and secondly, that there exist a submatrix $P \times P$, $\boldsymbol{S}$, related to the length$-q$ Fourier transform matrix, $\boldsymbol{F}$, by $\boldsymbol{S} = \boldsymbol{P}_1 \boldsymbol{\Delta}_1 \boldsymbol{F}$, where $\boldsymbol{P}_1$ is a permutation and $\boldsymbol{\Delta}_1$ is a $P \times q$ diagonal matrix containing qth roots of unity. Let $\boldsymbol{X} = \boldsymbol{A}\,\boldsymbol{S}$. Then $\boldsymbol{X}$ is invertible. But its entries satisfy $X_{ij}^q = 1$. Thus, there exist a permutation and a diagonal matrix of qth roots of unity such that $\boldsymbol{X} = \boldsymbol{P}_2 \boldsymbol{\Delta}_2 \boldsymbol{F}$. This implies that $\boldsymbol{A} = \boldsymbol{P}_2 \boldsymbol{\Delta}_2 \boldsymbol{\Delta}_1^- \boldsymbol{P}_1{}^{\mathsf{T}}$, which can be put in the form $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Delta}$. We have proved that $\boldsymbol{A}$ is trivial. □

In this case, trivial filters are of the form $\boldsymbol{P}\,\boldsymbol{D}[z]$, where $D_{pp}[z]$ are rotations in the complex plane of an angle multiple of $2\pi/q$ combined with a pure delay, and $\boldsymbol{P}$ are permutations.

6.2. Numerical algorithms

It is easy to run gradient ascents to find the maxima of $\Upsilon(\boldsymbol{x})$ defined in theorem 19. The same remarks already made in section 4 still hold. In the present case however, algebraic solutions can be computed, as reported in [22] [21], among others.

Other approaches, that are not based on contrast maximization, exist in the literature, including [48] [42] [45] [43] [47] [31].

7. CONCLUDING REMARKS

Several extensions of the contrast concept have been examined, and are currently being deepened. Algebraic block solutions become more and more attractive, especially in TDMA transmissions, because of the increased computational power. Other interesting directions of research include Non Linear or under-determined mixtures; MCMC approaches can be promising in this kind of context, for the same reason.

8. 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] 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.

[3] A. BELOUCHRANI, K. ABED-MERAIM, J. F. CARDOSO et al., "A blind source separation technique using second order statistics", *IEEE Trans. Sig. Proc.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[4] O. BESSON, P. STOICA, "On frequency offset estimation for flat-fading channels", *IEEE Com. Letters*, vol. 5, no. 10, pp. 402–404, Oct. 2001.

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

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

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

[8] P. COMON, "From source separation to blind equalization, contrast-based approaches", in *Int. Conf. on Image and Signal Processing (ICISP'01)*, Agadir, Morocco, May 3-5, 2001, invited plenary.

[9] 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.

[10] P. COMON, "Tensor decompositions", in *Mathematics in Signal Processing V*, J. G. McWhirter, I. K. Proudler, Eds., pp. 1–24. Oxford, UK, 2002.

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

[12] 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.

[13] P. COMON, L. ROTA, "Blind separation of independent sources from convolutive mixtures", *IEICE Trans. Fund. Elec. Com. Comp. Sc.*, Vol. E86-A, No 3, march 2003.

[14] N. DELFOSSE, P. LOUBATON, "Adaptive blind separation of independent sources: a deflation approach", *Signal Processing*, vol. 45, pp. 59–83, 1995.

[15] Z. DING, "Matrix outer-product decomposition method for blind multiple channel identification", *IEEE Trans. Sig. Proc.*, pp. 3053–3061, Dec. 1997.

[16] Z. DING, Y. LI, *Blind Equalization and Identification*, Dekker, New York, 2001.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] S. HAYKIN, *Unsupervised Adaptive Filtering*, vol. 1, Wiley, 2000, series in Adaptive and Learning Systems for Communications, Signal Processing, and Control.

[26] S. HOUCKE, A. CHEVREUIL, P. CIBLAT, "Estimation of the symbol period: the frequency offset case", in *Proc. XIth Eusipco*, Sept. 2002.

[27] A. HYVÄRINEN, J. KARHUNEN, E. OJA, *Independent Component Analysis*, Wiley, 2001.

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

[29] T. W. LEE, *Independent Component Analysis*, Kluwer, 1998.

[30] 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.

[31] 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.

[32] P. LOUBATON, E. MOULINES, P. A. REGALIA, "Subspace methods for blind identification and deconvolution", in *Sig. Proc. Advances in Wireless and Mobile Communications*, Giannakis, Hua, Stoica, Tong, Eds., chapter 3. Prentice-Hall, 2001.

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

[34] 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.

[35] E. MOULINES, P. DUHAMEL, J. F. CARDOSO et al., "Subspace methods for the blind identification of multichannel FIR filters", *IEEE Trans. Sig. Proc.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[36] 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.

[37] 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.

[38] D. T. PHAM, J-F. CARDOSO, "Blind separation of instantaneous mixtures of nonstationary sources", *IEEE Trans. Sig. Proc.*, pp. 1837–1848, Sept. 2001.

[39] J. G. PROAKIS, *Digital Communications*, McGraw-Hill, 1995, 3rd edition.

[40] 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.

[41] Y. SATO, "A method of self recovering equalization for multilevel amplitude-modulation systems", *IEEE Trans. Com.*, vol. 23, pp. 679–682, June 1975.

[42] A. L. SWINDLEHURST, S. DAAS, J. YANG, "Analysis of a decision directed beamformer", *IEEE Trans. Sig. Proc.*, vol. 43, pp. 2920–2927, Dec. 1995.

[43] 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.

[44] 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.

[45] 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.

[46] Y. WANG, E. SERPEDIN, P. CIBLAT et al., "Non data-aided feedforward cyclostationary statistics based carrier frequency offset estimators", in *Proceedings Globecom*, 2001, vol. 2, pp. 1386–1390.

[47] 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.

[48] 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.