Contrasts for Multichannel Blind Deconvolution

Pierre COMON

Abstract— **A class of optimization criteria is proposed whose maximization allows to carry out blind multichannel deconvolution in the presence of additive noise. Contrasts presented in the paper encompass those related to the sources separation and Independent Component Analysis problems. Numerical algorithms are not presented; they will be treated in a future paper.**

I. Introduction

Blind deconvolution is a subject that has raised an increasing interest during the last decade, and in particular multichannel deconvolution, based on High Order Statistics (HOS) [12] [7] [11], or second order ones [9] [6].

In fact, contrary to the scalar case [5] [10], the deviation from Gaussianity is not always necessary for the multichannel problem to be solvable.

The drawback of these approaches is that there are always restrictions concerning the noise. It is requested either to be negligible, Gaussian, or isotropic (spatial whiteness), or to have a known spatial covariance. In [2], a new class of criteria (called contrasts) was proposed, that allow the treatment of noisy multichannel observations, for the purpose of blind source separation in the presence of N sources. The advantages in using contrasts is that they are valid in presence of noise, regardless of its Gaussian character.

In addition, as will be pointed out subsequently, some contrasts are asymptotically insensitive to Gaussian noise, contrary to previously described criteria [1] [2] that require a prior standardization. In a full-length paper (in preparation), it will be also shown how a particular contrast is able also to handle the case of Gaussian sources. The corresponding numerical algorithms are not given in this paper for reasons of space, and are postponed to a full paper.

II. Observation model

Assume the observed time series, $y(t)$, is of dimension N and follows the model below:

$$y(t) = [H(z)] \cdot x(t) + v(t), \qquad (1)$$

where $H(z)$ denotes a $N \times N$ transfer matrix, $x(t)$ the so-called source process, of dimension N, and $v(t)$ an additive noise, independent of $x(t)$; the notation $[H(z)] \cdot x(t)$ stands for the discrete convolution of the series $x(t)$ by the matrix sequence $H(k)$. The case where the number of sources is smaller than N can also be addressed in a similar fashion. Nevertheless, the case where sources are more numerous than N is much more complicated, and the approaches proposed are not based on contrasts. See [4] for instance.

For convenience, assume $H(z) \in \mathcal{H}$ and $x \in \mathcal{P}$, where $\mathcal{H}$ ia a subset of the filters with finite L^2 norm, and $\mathcal{P}$ is a subset of N-dimensional stationary processes. For any process of the form (1), it will be denoted compactly that $y \in \mathcal{H} \cdot \mathcal{P}$, when $H \in \mathcal{H}$ and $x \in \mathcal{P}$.

The goal is to estimate the source process $x(t)$, assuming the basic properties:

S1. The sources $x_i(t), 1 \leq i \leq N$ are statistically independent.

S2. Sources $x_i(t)$ are each i.i.d. (independently and identically distributed) sequences with non zero variance..

Note that properties S1 and S2 can obviously be deflated to weaker properties of independence if only moments of finite order are utilized. As pointed out above, it is clear that the pair (H, x) is not uniquely defined in model (1). It is thus legitimate to assume the following additional (non restrictive) properties:

F1. The filter $H(z)$ is such that $Diag \oint H(z)^\dagger H(z)\, dz = I$. This means that the columns of $H(z)$ are normalized over the whole frequency band.

F2. $H(z)$ can be sought among the family of filters satisfying $Diag \oint H(z)\, H(z)^\dagger\, dz = I$. This means that the rows of $H(z)$ are normalized over the whole frequency band.

In fact, property F1 is acceptable if sources are not requested to have unit variance, and property F2 can be achieved by preprocessing (normalizing) each component $y_i(t)$. In other words, it is sufficient to multiply $y(t)$ and $x(t)$ by an appropriate constant diagonal matrix to meet these two requirements.

III. Contrast criteria

The optimization criteria that we are interested in are inspired by those proposed by Donoho; in his remarkable article [5], he actually gave the definition of trivial filters and contrast functionals in the scalar case, even if they were not so named. In the present paper, these definitions and properties are extended to the multichannel case, by using the tools devised in [2] in the multichannel instantaneous case.

Definition III.1: A matrix sequence $\{A(k)\}$ is referred to as trivial if and only if for every index i, there exist a single pair of indices (j, k) such that: $A_{ij}(k) \neq 0$.
A trivial filter is thus formed of a pure delay affecting each component followed by a permutation.

Definition III.2: A function Υ, mapping the probability density of a process $y(t) \in \mathcal{H} \cdot \mathcal{P}$ to a positive real number, denoted $\Upsilon(y)$, is said to be a probabilistic discriminating contrast, or **contrast** in short, over $(\mathcal{P}, \mathcal{H})$, if it meets the 3 requirements below:

C1. Υ is scale invariant; that is, $\Upsilon(\Lambda y) = \Upsilon(y)$, $\forall y \in \mathcal{H} \cdot \mathcal{P}$, and $\forall \Lambda$, regular diagonal constant matrix of $\mathcal{H}$.

[1]Manuscript received November 9, 1995; revised February 29, 1996. The associate editor coordinating the review of this letter and approving it for publication was Prof. G. B. Giannakis.

P. Comon is with CNRS – I3S, 250 av Albert Einstein, F-06560 Valbonne, pierre@thym.unice.fr, and with Thomson-Sintra, BP157, F-06903 Sophia-Antipolis Cedex, France, comon@asm.thomson.fr.

C2. If components $x_i(t)$ of a process $x(t) \in \mathcal{P}$ are statistically independent, and each temporally white, then $\Upsilon(H \cdot x) \leq \Upsilon(x), \forall H(z) \in \mathcal{H}$.

C3. Equality occurs in C2 if and only if $H(k)$ is trivial. This condition insures the **discriminating** character of the contrast.

In the scalar case ($N = 1$), this definition is compatible with [5], and it coincides with [2] for instantaneous mixtures. We refer to [2] for further justifications.

Let us give now two examples of contrast functions that can be utilized for the purpose of blind multichannel equalization. Denote $\mathcal{H}_n$ the set of invertible filters satisfying condition Fn, $n \in \{1, 2\}$, and $\mathcal{P}_r$ the set of processes stationary up to order r, having at most one null marginal cumulant of order r (at lag zero). In addition, denote $\mathcal{C}_{(r),y_j}$ the cumulant of order r of the scalar variable y_j. Then we have:

Proposition III.3: The function $\Upsilon_{1,r}(y) = \sum_j |\mathcal{C}_{(r),y_j}|$ is a contrast over $(\mathcal{P}_r, \mathcal{H}_1 \cap \mathcal{H}_2)$, if $r > 2$.

Proof: The proof begins along the same lines as in the one given by Moreau in the instantaneous case [8]. Condition C1 is insured by F2. To prove C2, one expands the contrast as:

$$\Upsilon_1(y) = \sum_i \left| \sum_{p,k} H^r_{jp}(k)\, \mathcal{C}_{(r),x_p} \right| \leq \sum_{i,p,k} |H_{ip}(k)|^r \cdot |\mathcal{C}_{(r),x_p}|.$$

But from F1, $\sum_{i,k} |H_{ip}(k)|^2 = 1, \forall p$. Thus for $r \geq 2$, one obtains indeed:

$$\Upsilon_1(y) \leq \sum_p |\mathcal{C}_{(r),x_p}| = \Upsilon_1(x).$$

To prove C3, assume that $\Upsilon_1(y) = \Upsilon_1(x)$. Then:

$$\sum_p [1 - \sum_{i,k} |H_{ip}(k)|^r] \cdot |\mathcal{C}_{(r),x_p}| = 0.$$

For any p such that $\mathcal{C}_{(r),x_p} \neq 0$, one gets: $\sum_{ik} |H_{ip}(k)|^r = 1$. The latter equality yields that $\sum_{i,p,k} |H_{ip}(k)|^r = N$. Yet, $\sum_{p,k} |H_{ip}(k)|^r \leq 1$ from F2, so that one also has:

$$\sum_{pk} |H_{ip}(k)|^r = 1, \forall i.$$

One concludes that $\{H(k)\}$ is trivial, since $r > 2$. ∎

The presence of the absolute value in contrast $\Upsilon_{1,r}$ may raise some difficulties, particularly in adaptive implementations. Therefore, it may be convenient to use the criterion below, that has not this inconvenience, though more complicated:

Proposition III.4: The function $\Upsilon_{2,r}(y) = \sum_j \mathcal{C}^2_{(r),y_j}$ is a contrast over $(\mathcal{P}_r, \mathcal{H}_1 \cap \mathcal{H}_2)$, if $r > 2$.

The proof borrows ideas from the one given by Comon in the instantaneous case [2]. In particular, it uses the following lemma:

Lemma III.5: If Q is bistochastic, that is, if Q has positive entries and if each row and column sums up to one, then it satisfies $||Q\,u|| \leq ||u||, \forall u$, for the L^2 norm.

Proof: Denote ${}^r\!H$ the matrix whose entries are ${}^r\!H_{ij} = |H_{ij}|^r$. Denote $\bar{Q}$ (resp. $\bar{u}$) the matrix (resp. the vector) whose entries are obtained by taking the modulus of the entries of Q (resp. u). Lastly, denote $\tilde{H}$ the matrix defined by $\tilde{H}_{ij} = [\sum_k \bar{H}^2_{ij}(k)]^{1/2}$. Because of properties F1 and F2, ${}^2\tilde{H}$ is bistochastic.

By triangular inequality, we have:

$$\begin{aligned}\Upsilon_{2,r}(H \cdot x) &= \sum_i |\sum_{p,k} H^r_{ip}(k)\, \mathcal{C}_{(r),x_p}|^2 \\ &\leq \sum_i |\sum_{p,k} \bar{H}^r_{ip}(k)\, \bar{\mathcal{C}}_{(r),x_p}|^2 \stackrel{\text{def}}{=} \alpha.\end{aligned}$$

Next, inequality between L^r and L^2 norms yields:

$$\alpha \leq \sum_i [\sum_p \tilde{H}^r_{ip}\, \bar{\mathcal{C}}_{(r),x_p}|^2 \leq \sum_i [\sum_p \tilde{H}^2_{ip}\, \bar{\mathcal{C}}_{(r),x_p}|^2 \stackrel{\text{def}}{=} \beta.$$

Yet, ${}^2\tilde{H}$ is bistochastic. Thus, using the lemma

$$\beta \leq \sum_i \bar{\mathcal{C}}^2_{(r),x_i} = \Upsilon_{2,r}(y),$$

which proves C2. Now to prove C3, assume that equality holds in all previous inequalities. Then

$$\sum_i [\sum_p \tilde{H}^r_{ip}\, \bar{\mathcal{C}}_{(r),x_p}]^2 = \sum_i [\sum_p \tilde{H}^2_{ip}\, \bar{\mathcal{C}}_{(r),x_p}]^2.$$

This equality holds true after removing the squares, because terms are themselves positive. This yields:

$$\sum_{i,p} (\tilde{H}^2_{ip} - \tilde{H}^r_{ip})\, \bar{\mathcal{C}}_{(r),x_p} = 0.$$

Again, every term being positive, must be null. For at least $N-1$ values of p, we have thus: $\tilde{H}^2_{ip} - \tilde{H}^r_{ip} = 0, \forall i$. In other words, since $r > 2$, $N - 1$ columns of $\tilde{H}$ contain a single non zero entry. Because $\tilde{H}$ is stochastic, it is not hard to conclude that it is a permutation. Let $\tilde{H} = P$, where P is a permutation matrix. Going back to the sequence $H(k)$, we get:

$$\sum_k H^r_{ip}(k) = P_{ip}.$$

Summing over p eventually yields $\sum_{p,k} H^r_{ip}(k) = 1$. Again, because of F2, this equality is possible only if $H(k)$ is trivial. ∎

Other contrast functions will be presented in a full-length version of this paper.

References

[1] 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.

[2] 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.

[3] P. COMON, "Contrasts for multichannel blind deconvolution", Research report no 95-44, I3S-CNRS, 8 september 1995.

[4] P. COMON, B. MOURRAIN, "Decomposition of quantics in sums of powers of linear forms", *Signal Processing*, vol. 53, no. 2, 1996. Special issue on High-Order Statistics.

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

[6] I. FIJALKOW, P. LOUBATON, "Identification of rank one rational spectral densities from noisy observations: a stochastic realization approach", *Systems and Control Letters*, , no. 24, pp. 201–205, 1995.

[7] Y. INOUYE, T. HABE, "Blind equalization of multichannel linear time-invariant systems", *The Institute of Electronics Information and Communication Engineers*, , no. 24, pp. 9–16, May 1995.

[8] E. MOREAU, *Apprentissage et adaptivité, Séparation auto-adaptative de sources indépendantes par un réseau de neurones*, Doctorat de l'Université de Paris-sud, Orsay, 1 fevrier 1995.

[9] E. MOULINES, P. DUHAMEL, J. F. CARDOSO, S. MAYRAGUE, "Subspace methods for the blind identification of multichannel FIR filters", *IEEE Trans. on Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

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

[11] A. SWAMI, G. GIANNAKIS, S. SHAMSUNDER, "Multichannel ARMA processes", *IEEE Trans. on Signal Processing*, vol. 42, no. 4, pp. 898–913, Apr. 1994.

[12] D. YELLIN, E. WEINSTEIN, "Criteria for multichannel signal separation", *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 2158–2168, Aug. 1994.