Blind Separation of Discrete Sources

Olivier GRELLIER *Student Member, IEEE,*
and Pierre COMON *Senior Member, IEEE*

Abstract— **A polynomial criterion is proposed to perform blind source separation, extending previous works to MIMO systems. The criterion is proved to be asymptotically MAP-equivalent in presence of PSK sources. An efficient minimization algorithm dedicated to polynomial criteria is then developed, improving on the fixed-step stochastic gradient previously utilized in this framework.**

I. Introduction

Even in the blind context, spatial diversity has been demonstrated to be useful in several problems, including Air Traffic Control and Mobile Communications [1] [2] [3]. Consider the following reception model:

$$\mathbf{y} = A\,\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{y}$ and $\mathbf{w}$ are random vectors of dimension K, the number of channels (which often corresponds to the number of antennas times the polarization diversity), A is an unknown $K \times P$ complex mixing matrix, and $\mathbf{x}$ is a random vector of size P, independent of $\mathbf{w}$. It is assumed in the remaining that: [**A1**] $P \leq K$, and [**A2**] every source $s = x_p$ has a known discrete distribution in the complex plane, $p_s(s)$, of support $\mathcal{C}$. The goal is to carry out a Discrete Source Estimation (DSE), solely based on successive observations of $\mathbf{y}$. This model is valid for instance for narrow-band digital communications over flat fading channels. But such static mixtures play also a key role in convolutional mixtures, sometimes implicitly; this has been noticed by several authors, the earliest reference being [5].

Blind separation techniques require neither the knowledge of the array nor learning sequences, and therefore raise increased interest, even if they are often used within a semi-blind framework [6] [12].

Blind separation algorithms can be based on the independence of the sources [1], [4], on the constance of their modulus [15], or on the discrete character of their distribution [8], [11], [14], [17]. This work is dedicated to the latter case, where sources can possibly be nonstationary or correlated.

In [11], both A and $\mathbf{x}$ are estimated in the case of BPSK sources; the optimization criterion is of ML type, and thus computationally heavy to maximize. This becomes much worse if the number of constellation symbols increases. In the present paper, the computational cost is decreased by estimating separately the rows of A^{-1} instead, as shown in section II, and also by resorting to a simpler criterion (6) of polynomial form. This criterion is similar to the one proposed in [14]. Another polynomial approach was presented in [16] to separate zero/constant modulus signals. Compared to [14], our contribution is three-fold: (i) we extend its use to the MIMO case, (ii) we prove its asymptotic MAP-equivalence for PSK sources, and (iii) a fixed-step stochastic gradient was proposed in [14] to perform blind SISO equalization, whereas our special-purpose algorithm, called AMiSRoF, performs even better than the optimal-step gradient.

Moreover, the AMiSRoF-based algorithm proposed in this paper is of deflation type [7]. Its core looks similar to the algorithm proposed in [9], except for the roots computed along each descent direction, and for the optimization criterion. Thus, AMiSRoF is applicable to MIMO convolutive mixtures as well, even if the computer experiments subsequently reported are limited to static mixtures.

Manuscript submitted March 20, 1998. Both authors are with I3S-CNRS, Sophia-Antipolis, France. P.Comon is also with Eurecom Institute, BP193, Sophia-Antipolis, France. This research was supported by CNRS Telecommunications Programme, and by Eurecom's industrial partners: Ascom, Cegetel, Hitachi, IBM France, Motorola, Swisscom, and Thomson CSF.

II. Optimization criterion

The optimal solution in the MAP sense is given by:

$$(\hat{\mathbf{x}}, \hat{A})_{MAP} = \text{Arg}\,\underset{\mathbf{x},A}{\text{Max}}\; p_{x|y,A}(\mathbf{x}, \mathbf{y}, A) \tag{2}$$

where $p_{x|y,A}(\mathbf{x}, \mathbf{y}, A) = p_x(\mathbf{x}) \cdot p_w(\mathbf{y} - A\,\mathbf{x})/p_y(\mathbf{y})$. Contrary to the so-called deterministic ML approach, the MAP approach restricts the search for $\mathbf{x}$ to the allowed constellation $\mathcal{C}$. Thus the MAP criterion is the most natural in this context. In this paper, it is aimed at minimizing this criterion. In a first stage, the search complexity is reduced by modifying the optimization problem, and appropriate approximations are then made in a second stage to allow a polynomial complexity. In presence of Gaussian noise, it is well known that the MAP solution rewrites:

$$(\hat{\mathbf{x}}, \hat{A})_{MAP} = \text{Arg}\,\underset{\hat{\mathbf{x}}\in\mathcal{C}^P,\hat{A}}{\text{Min}}\; ||\mathbf{y} - \hat{A}\,\hat{\mathbf{x}}||^2. \tag{3}$$

Complexity reduction. The first simplification consists of noticing that, if A is full column rank, one can instead minimize with respect to a $P \times K$ matrix B:

$$(\hat{\mathbf{x}}, \hat{B})_{MAP} = \text{Arg}\,\underset{\hat{\mathbf{x}}\in\mathcal{C}^P,\hat{B}}{\text{Min}}\; ||\hat{B}^*\mathbf{y} - \hat{\mathbf{x}}||^2 \tag{4}$$

This significantly reduces the computational complexity. Indeed, denote D the cardinality of constellation $\mathcal{C}$, then a single pseudo-inverse needs to be computed in (4), whereas D^{NP} of them need to be computed in (3). Furthermore, it is possible to look for the rows of B^* one by one (deflation approach), so that the optimization problem takes eventually the form below, if N independent realizations are observed, and still in the case of white Gaussian noise:

$$\hat{\mathbf{b}}_{MAP} = \text{Arg}\,\underset{\hat{\mathbf{b}}}{\text{Min}} \sum_{n=1}^{N} \underset{\hat{x}(n)\in\mathcal{C}}{\text{Min}} |\hat{\mathbf{b}}^*\mathbf{y}(n) - \hat{x}(n)|^2 \tag{5}$$

The search complexity is now impressively reduced. In fact, the search for $\hat{\mathbf{x}}$ is now performed in a set of cardinal $P\,D^N$, which is much smaller than with (4). Up to now, no approximation has been made. Unfortunately, it is not easy to find analytically a (even local) minimum of this reliable criterion.

Polynomial criterion. Now, it is shown in appendix that, under the assumption that the sources are PSK-modulated, the MAP criterion is asymptotically equivalent to

$$\Phi_N(\mathbf{b}) = \frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{D}|\mathbf{b}^*\mathbf{y}(n) - C(j)|^2. \qquad (6)$$

where $D = \text{Card}(\mathcal{C})$ and $C(j)$ are the elements of $\mathcal{C}$. This allows to replace a search over the constellation $\mathcal{C}$ by a mere product. Yet, the numerical algorithm proposed in the next section is precisely dedicated to polynomial criteria. In [13], it has been shown that the absolute minima of (6) give for large enough N the correct solutions. In section IV, the robustness of the criterion with respect to the assumptions made in appendix are investigated through computer experiments.

III. Extraction algorithm

A. Extraction of a single source

The advantage of the polynomial criterion (6) is that it can theoretically be analytically minimized. In practice, the computational burden is acceptable only in one dimension, so that we are led to construct the following algorithm.

Absolute Minimum Search by Root Finding (AMiSRoF):

1. Set $\mathbf{b}$ to the initial value $\mathbf{b}_0$, defined as the dominant eigenvector of the covariance of $\mathbf{y}$. Then, for $k = 1, 2, ..$ do the following steps
2. Compute the gradient $\mathbf{g}$ and Hessian H of $\Phi_N(\mathbf{b})$ at $\mathbf{b}_k$.
3. Compute the optimal descent direction $\mathbf{v} = H^{-1}\mathbf{g}$ and normalize it to $||\mathbf{v}|| = 1$.
4. Compute the next vector $\mathbf{b}_{k+1} = \mathbf{b}_k + \mu\mathbf{v}$, where μ minimizes $\Phi_N(\mathbf{b}_{k+1})$.

In step 4, the value of μ is obtained by rooting the first derivative of Φ_N with respect to μ, and then choosing the stationary point yielding the smallest value. Iterations are stopped when Φ_N doesn't significantly change any more. A vector quantization can follow the source extraction. This is done when Bit Error rates (BER) are evaluated (cf. section IV).

The search for the absolute minimum in the $\mathbf{v}$-space has been seen to be useful in the first 5 iterations of most experiments; it has indeed either allowed a faster convergence, or opened the way out of a local minimum. On the other hand, the remaining iterations always led to the closest minimum.

Note that the AMiSRoF algorithm could also be utilized for a MAP refinement, keeping alternatively $C(j)$ and $\mathbf{b}$ fixed.

B. Procedure for extracting successive sources

Uncorrelated Sources Extraction (UDSE) can be performed as follows.

1. For $p = 1, 2, ..$ execute the steps below
2. Run the AMiSRoF algorithm to obtain $s_p(n) = \mathbf{b}^*\mathbf{y}(n)$.
3. Quantify the signal $s_p(n)$ into $\hat{s}_p(n)$ according to the a priori on $p_s(s_p)$.
4. Remove the contribution of $\hat{s}_p(n)$ in the data by linear regression and subtraction, yielding a new set of data, $\mathbf{y}(n) \leftarrow \mathbf{y}(n) - \overline{\mathbf{y}\hat{s}_p^*}/\overline{\hat{s}_p\hat{s}_p^*}\,\hat{s}_p(n)$.

The iterations stop when no more energy is present in the data.

Note that here the number of sources does not need to be known in advance, and sources do not need to be statistically independent (but need to be uncorrelated at second order).

Now if sources were correlated, two algorithms could be envisaged to perform Correlated Discrete Source Extraction (CDSE), as explained below, but their detailed description is out of the scope of the present paper.

If the number of sources P is known in advance, it is possible to estimate all the rows of the $P \times K$ matrix B^* simultaneously. For doing this, the same type of algorithm as AMiSRoF can be run; it suffices to replace $\mathbf{b}$ by $\mathbf{vec}\{B\}$, so that there are now PK unknowns. The rest of the procedure is identical, and the last step is $B_{k+1} = B_k + \mu\text{Unvec}\{\mathbf{v}\}$.

If the number of sources is not known and sources are correlated, it is necessary to resort to other means to avoid obtaining several times the same source. The method suggested is simple and consists of including a penalty term in the new optimization criterion, as soon as one source has been extracted [10].

IV. Computer results

Simulations have been carried out using a uniformly spaced linear array with $K = 4$ sensors. The element spacing is $\lambda/2$, where λ is the wavelength of the propagating waveforms. $P = 3$ uncorrelated sources impinge on the array, and the directions of arrival are $\theta_1 = -30°$, $\theta_2 = -10°$ and $\theta_3 = 20°$. We tested our algorithm with modulations 4-QAM, $N = 2000$ and 16-QAM, $N = 1000$. Table I shows the median and the variance of the Bit Error Rate for various SNRs obtained over 200 trials; p denotes the label of the source considered.

The BER median is quite good for 4-QAM or 16-QAM until the SNR becomes lower than $6dB$, but the BER variance shows that sometimes the algorithm converges to a spurious minimum. Should this happen, the BER is so bad that a simple test allows to detect ill convergence, and the algorithm could be run again with another initial value.

The fact that our algorithm works for 16-QAM signals can seem surprising. In fact, we separate these signals by minimizing the criterion that matches 4-QAM constellations. Furthermore, this criterion turns out to be efficient for all constellations that enjoy a symmetry with respect to the origin.

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.

p	SNR (dB)	4-QAM Median	Variance	16-QAM Median	Variance
1		0	0	0	$1.3e^{-2}$
2	10	0	0	0	$1.1e^{-2}$
3		0	0	0	$2e^{-3}$
1		0	$1.5e^{-4}$	0	$2.3e^{-2}$
2	8	0	$5e^{-9}$	0	$1.1e^{-2}$
3		0	0	0	$2e^{-3}$
1		0	$1e^{-8}$	0	$5e^{-3}$
2	6	0	$2.7e^{-4}$	0	$7.7e^{-3}$
3		$1.3e^{-6}$	$3e^{-10}$	$1.8e^{-6}$	$2.8e^{-9}$
1		$1.5e^{-4}$	$6e^{-4}$	$7.5e^{-4}$	$1.1e^{-2}$
2	4	$1.2e^{-4}$	$5e^{-8}$	$5e^{-4}$	$1.5e^{-2}$
3		$6.3e^{-5}$	$2e^{-8}$	$2.5e^{-4}$	$5.7e^{-3}$
1		$1.5e^{-3}$	$3e^{-4}$	$7.7e^{-3}$	$9e^{-3}$
2	2	$1.5e^{-3}$	$2.5e^{-4}$	$5.5e^{-3}$	$2.1e^{-2}$
3		$1e^{-3}$	$2e^{-7}$	$4.5e^{-3}$	$2e^{-6}$

TABLE I

BER averaged over 200 Monte Carlo realizations.

[2] L. CASTEDO, O. MACCHI, "Maximizing the information transfer for adaptive unsupervised source separation", in *Proc. SPAWC 97 Conf.*, Paris, France, Apr. 1997, pp. 65–68.

[3] E. CHAUMETTE, P. COMON, D. MULLER, "An ICA-based technique for radiating sources estimation; application to airport surveillance", *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 395–401, Dec. 1993, Special issue on Applications of High-Order Statistics.

[4] 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.

[5] P. COMON, "Separation of Sources using High-Order Cumulants", *SPIE Conf*, San Diego, CA, pp 170-181, Aug 8-10 1989.

[6] E. de CARVALHO, D. T. M. SLOCK, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation", in *Proc. SPAWC 97 Conf.*, Paris, France, Apr. 1997.

[7] N. DELFOSSE, P. LOUBATON, "Adaptive separation of independent sources: A deflation approach", in *IEEE Int. Con. Acoust. Speech Sig. Proc. (ICASSP'94), vol.IV*, Adelaide, Australia, Apr. 1994, pp. 41–44.

[8] E. GASSIAT, F. GAMBOA, "Source separation when the input sources are discrete or have constant modulus", *IEEE Trans. Signal Processing*, vol. 45, no. 12, pp. 3062–3072, Dec. 1997.

[9] A. GOVINDARAJU, J. TUGNAIT, "Near-filed localization using inverse filter criteria-based blind separation and cumulant matching", in *IEEE SP Workshop on HOS*, Banff, Canada, July 21-23 1997, pp. 300–304.

[10] O. GRELLIER, P. COMON, "Self calibration in presence of discrete sources", in *Eusipco*, Rhodes, Greece, Sept 8–11 1998.

[11] A. KANNAN, V. U. REDDY, "Maximum likelihood estimation of constellation vectors for blind separation of co-channel BPSK signals and its performance analysis", *IEEE Trans. Sig. Proc.*, pp. 1736–1741, 1997.

[12] A. M. KUZMINSKIY et al., "Regularized semi-blind estimation of spation-temporal filter coefficients for mobile radio communications", in *XVIieme Colloque GRETSI*, Grenoble, 15-19 Sept 1997, pp. 127–130.

[13] T-H. LI, "Blind deconvolution of linear systems with nonstationary inputs", *Annals of Stat.*, vol. 23, no. 2, pp. 690–704, 1995.

[14] 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.

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

[16] A.J. VAN DER VEEN, J. TOL, "Separation of zero/constant modulus signals", *ICASSP'97 Conf.*, Munich, Germany, pp. 3445–3448, April 21-24 1997.

[17] D. YELLIN, B. PORAT, "Blind identification of FIR systems excited by discrete-alphabet inputs", *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1331–1339, 1993.

Appendix

Given a number z in the complex plane, consider the the two optimization criteria $\Psi(z) = \min_j |z - C_j|^2$ and $\Phi(z) = \prod_j |z - C_j|^2$, and denote by ϵ the distance between z and the closest element of the constellation, C_p. Then $\Phi(z)$ can be written as:

$$\begin{aligned}
\Phi(z) &= |z - C_p|^2 \prod_{j \neq p} |z - C_j|^2 \\
&= \Psi(z) \prod_{j \neq p} |C_p - C_j + \epsilon|^2 \\
&= \Psi(z) \prod_{j \neq p} |C_p - C_j|^2 \prod_{j \neq p} |1 + \frac{\epsilon}{C_p - C_j}|^2 \\
&= \Psi(z) K(p) \prod_{j \neq p} |1 + \frac{\epsilon}{C_p - C_j}|^2 \\
&= \Psi(z) K(p) \prod_{j \neq p} \left(1 + 2Re\left[\frac{\epsilon}{C_p - C_j}\right] + \frac{|\epsilon|^2}{|C_p - C_j|^2}\right) \\
&= \Psi(z) K(p) \left(1 + f_p(\epsilon)\right)
\end{aligned}$$

A constellation $\mathcal{C}$ is said to have a constant power if $\exists \gamma / \forall c \in \mathcal{C},\ c^\gamma = 1$. This property characterizes usual PSK modulations. For these constellations, $K(p)$ and $f_p(\epsilon)$ do not depend on p, and we have:

$$\Psi(z) = \frac{1}{K(1 + f(\epsilon))} \Phi(z).$$

If ϵ is small compared to the distance between constellation symbols (low noise assumption), then $|f(\epsilon)| \ll 1$, and a Taylor expansion is possible and yields:

$$\Psi(z) = \frac{1}{K} \Phi(z)(1 + g(\epsilon)) + o(\epsilon).$$

Now let $z_n = \mathbf{b}^* \mathbf{y}(n)$. Then (5) can be rewritten as $\text{Arg Min}_{\hat{\mathbf{b}}} \sum_n \Psi(z_n)$, with the notation of this appendix. Next, (5) takes the form:

$$\frac{1}{K} \text{Arg} \min_{\hat{\mathbf{b}}} \sum_n \Phi(z_n)(1 + g(\epsilon_n)) \tag{7}$$

which can be approximated by

$$\frac{1}{K} \text{Arg} \min_{\hat{\mathbf{b}}} \left[(1 + \eta) \sum_n \Phi(z_n) \right],$$

where $\eta = \left[\frac{1}{N} \sum_n \Phi(z_n) g(\epsilon_n)\right] \left[\frac{1}{N} \sum_n \Phi(z_n)\right]^{-1}$ keeps small for bounded N or uniformly small ϵ_n. This eventually shows the MAP equivalence for PSK constellations and for sufficiently low noise levels.