Optimization of the asymptotic performance of time-domain convolutive source separation algorithms

Nabil CHARKANI (1) (2), Yannick DEVILLE (1)

(1) Laboratoires d'Electronique Philips S.A.S. (LEP),
22, Av. Descartes, B.P. 15, 94453 Limeil-Brévannes, France.
(2) Laboratoire de Traitement d'Images et Reconnaissance de Formes (LTIRF - INPG),
46, Av. Félix Viallet, 38031 Grenoble Cedex, France.

E-mails: charkani@lep.research.philips.com, ydeville@lep.research.philips.com

Abstract. In this paper, we investigate the self-adaptive source separation problem for convolutively mixed signals. The proposed approach uses a recurrent structure adapted by a generic rule involving arbitrary separating functions, which is derived from a neural approach. The expression of the asymptotic error variance achieved by this rule is first determined (for strictly causal mixtures). This enables us to derive the separating functions that minimize this error variance. They are shown to be only related to the probability density functions of the sources. Simulations are performed in various conditions, ranging from artificial mixtures of synthetic sources to real mixtures of audio signals. They show that the proposed approach yields much better performance than classical rules.

1. Problem statement

Blind (or self-adaptive) source separation consists in extracting unknown independent source signals from sensor observations that are unknown linear combinations of these sources. A simple and frequently used model [5] in the case of bi-dimensional mixing systems corresponds to the source-observation relationship:

$$Y_1(z) = X_1(z) + A_{12}(z)X_2(z) \quad (1)$$

$$Y_2(z) = A_{21}(z)X_1(z) + X_2(z) \quad (2)$$

where $X_i(z)$ and $Y_i(z)$ are respectively the $\mathcal{Z}$-transforms of the source $x_i(n)$ and the observation $y_i(n)$. $A_{ij}(z)$ is the transfer function of the channel that links the source j to the sensor i. The corresponding impulse response is denoted $(a_{ij}(k))_{k\in\mathcal{Z}}$ hereafter. The mixing system is generally assumed to be causal and minimum phased so that its inverse can be implemented by a stable and

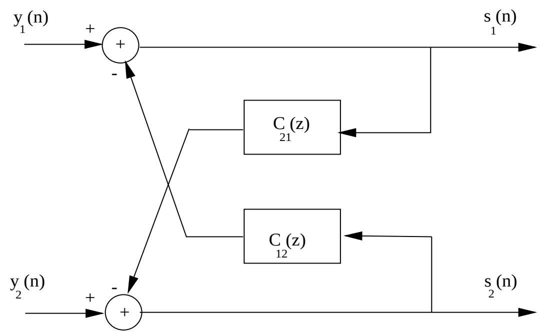

Figure 1: Recurrent structure for the separation system.

causal system. A first solution to this problem, based on a neural network, was proposed by Hérault and Jutten [4] in the case when the mixing filters become scalar coefficients ($A_{ij}(z) = a_{ij}(0)$ for $i \neq j \in \{1, 2\}$). A natural extension of this neural approach, both for the separating structure and the associated adaptation rule, was proposed by Nguyen Thi and Jutten [5]. Their approach is based on the recurrent structure of Fig. 1 where the separating filters $C_{ij}(z)$ are assumed to be M^{th}-order Moving Average filters (denoted MA(M)). Their coefficients, $(c_{ij}(k))_{0 \leq k \leq M}$, are adapted according to the stochastic rule:

$$c_{ij}(n+1, k) = c_{ij}(n, k) + \mu f(s_i(n)) g(s_j(n-k)) \quad i \neq j \in \{1, 2\}, k \in [0, M]. \tag{3}$$

The particular case $f(x) = x^3$ and $g(x) = x$ was deeply investigated from an experimental point of view but no theoretical results were provided neither on the stability of the rule (3) nor on its asymptotic behaviour. In fact, very few results are available in the literature on these two aspects. Furthermore, they are generally restricted to very specific cases (see references in [2]).
In this paper, we deal with a more general separation rule (N0) that reads:

$$c_{ij}(n+1, k) = c_{ij}(n, k) + \mu f_i(s_i(n)) g_i(s_j(n-k)) \quad i \neq j \in \{1, 2\}, k \in [0, M] \tag{4}$$

where f_i and g_i are arbitrary functions. The stability analysis of this algorithm is provided in [2],[3]. In the current paper, we investigate the asymptotic behaviour of this rule. More precisely, we derive the asymptotic error variance of the estimation of the mixing filters. The minimization of this variance (with respect to f_i and g_i) then leads to the optimum separating functions. Extended versions and applications of this rule are then presented.

2. Theoretical analysis

In this section, we consider the case of strictly causal mixing and separating filters (i.e. $a_{12}(0) = a_{21}(0) = c_{12}(0) = c_{21}(0) = 0$). The algorithm (4) can then be formulated in vector form as:

$$\theta_{n+1} = \theta_n + \mu \, H(\theta_n, \xi_{n+1}), \tag{5}$$

where μ is a small positive adaptation gain and

$$\theta_n = [c_{12}(n,1), \ldots, c_{12}(n,M), c_{21}(n,1), \ldots, c_{21}(n,M)]^T. \tag{6}$$

ξ_{n+1} and $H(\theta_n, \xi_{n+1})$ are column vectors derived from (4) and not detailed here for the sake of brevity. The separating state (that corresponds to $S_i(z) = X_i(z)$) is then

$$\theta^s = [a_{12}(1), \ldots, a_{12}(M), a_{21}(1), \ldots, a_{21}(M)]^T. \tag{7}$$

Applying the asymptotic convergence theorem established in [1], and related to the Ordinary Differential Equation (ODE) technique properties, to the algorithm (5) shows that [2], for large n and a stable state θ^s (see [2],[3]), θ_n is an asymptotically unbiased Gaussian estimator of θ^s. Its covariance matrix is μP where P is the unique symmetric and positive definite solution of the Lyapunov equation:

$$J(\theta^s)P + PJ^T(\theta^s) + R(\theta^s) = 0 \tag{8}$$

where $J(\theta^s)$ is the Jacobian matrix at the separating state [2],[3] and $R(\theta^s) = \sum_{n\in\mathcal{Z}} Cov[H(\theta^s, \xi_{n+1}), H(\theta^s, \xi_0)]$.

The asymptotic error variance of θ_n is thus $\sigma_\infty = \lim_{n\to+\infty} E[|\theta_n - \theta^s|^2] = \mu Tr(P)$. For white sources, mathematical calculations [2] show that it reads:

$$\sigma_\infty = \mu \sum_{i,j=1 \quad i\neq j}^{i,j=2} \frac{E[f_i^2(x_i)]}{E[f_i^{'}(x_i)]} \frac{E[g_i^2(x_j)]}{E[x_j g_i(x_j)]} (q_{i2} + q_{i1} a_i) \tag{9}$$

where $a_i = \frac{E^2[f_i(x_i)]}{E[f_i^2(x_i)]}$ and (q_{i1}, q_{i2}) is a couple of real constants that are only related to the mixing matrix. Differentiating σ_∞ with respect to f_i and g_i leads [2] to the extremum functions:

$$(f_{iopt}(x), g_{iopt}(x)) = \left(-\nu_{i1}\ p_{x_i}^{'}(x)/p_{x_i}(x), \nu_{i2}\ x\right) \tag{10}$$

where p_{x_i} is the p.d.f of the source x_i and (ν_{i1}, ν_{i2}) is a couple of arbitrary (apart from the stability conditions established in [2],[3]) real constants. This extremization procedure does not directly yield separating functions that minimize σ_∞, since σ_∞ still depends on the scaling factors ν_{i1} and ν_{i2}. This ambiguity can be removed by using the modified separation rule $(N1)$:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu \frac{f_i(s_i(n))}{\sqrt{E[f_i^2(s_i)]}} \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}} \quad i \neq j \in \{1,2\}, k \in [1,M] \tag{11}$$

for which all the proportional functions f_i (resp. g_i) yield exactly the same separation rule (11) and therefore the same σ_∞. It may then be shown [2] that the minimum of σ_∞ for (11) corresponds to the class of functions (10).

Extension to coloured signals: The theoretical results (9)-(10) were derived using the whiteness of the sources as a key assumption. Most often, this hypothesis does not hold for real applications. When the sources are AR processes, we propose a new separating structure that consists of: *i*) the *separating module* of Fig. 1 that implements the inverse matrix, *ii*) a *whitening module* that drives the adaptation of the separating filters. The associated adaptation rule (NW1) reads:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu \frac{f_i(v_i(n))}{\sqrt{E[f_i^2(v_i)]}} \frac{g_i(v_j(n-k))}{\sqrt{E[g_i^2(v_j)]}} \quad i \neq j \in \{1,2\}, k \in [1,M] \tag{12}$$

where $v_i(n)$ is the estimated innovation process of $s_i(n)$ provided by a q_i^{th}-order MA whitening filter, i.e:

$$v_i(n) = s_i(n) + \sum_{k=1}^{q_i} b_i(n,k) s_i(n-k), \quad i \in \{1,2\}. \tag{13}$$

The coefficients $(b_i(n,k))_{k \in [1,q_i]}$ are updated by a LMS rule with gain γ:

$$b_i(n+1,k) = b_i(n,k) - \gamma v_i(n) s_i(n-k), \quad i \in \{1,2\}. \tag{14}$$

An asymptotic behaviour analysis of $(NW1)$ [2] shows that the optimum class of separating functions is then $(f_{iopt}(x), g_{iopt}(x)) = (-\nu_{i1}\, p'_{\tilde{x}_i}(x)/p_{\tilde{x}_i}(x), \nu_{i2}\, x)$ where $p_{\tilde{x}_i}$ is the p.d.f of the innovation process associated to the source x_i.

3. Simulation results

In this section, we present some simulation results that validate the theoretical approach developed in Section 2.. The average *Signal to Noise Ratio Improvement* is used as an objective criterion for measuring the source separation performance. It is defined by $SNRI = (SNRI_1 + SNRI_2)/2$ where $SNRI_i$ denotes the Signal to Noise Ratio Improvement at the output i given by $SNRI_i = 10log_{10}\left(E[(y_i(n) - x_i(n))^2]/E[(s_i(n) - x_i(n))^2]\right)$.

3.1. Validation of the optimization analysis

The first set of simulations aims at verifying that the optimum separating functions correspond to the solution (10). Tests are simplified by using sources which have the same p.d.f and by considering uncoupled optimization i.e. (f_i varying, g_i optimum) and (f_i optimum, g_i varying). Simple forms for the optimum separating functions (10) are obtained by considering sources whose p.d.f belongs to the Generalized Gaussian Family (GGF) defined by $p_\beta(x) = K_\beta\ exp(-\frac{|x|^\beta}{\lambda_\beta})$ where K_β and λ_β are constants that depend on the paramater $\beta \geq 1$. The optimum functions f_i are then proportional to $sign(x)|x|^{\beta-1}$. This optimality can be checked by considering separating functions f_i within the family $\mathcal{F} = \{sign(x)|x|^k, k \geq 0\}$ and verifying that the minimum error variance is achieved for $k = \beta - 1$. $\mathcal{F}$ is also used to test the optimality of the separating function g_i: in this case, the optimum value is $k = 1$. In the experiments, we considered strictly causal MA(4) mixing and separating filters and white sources. Simulation results [2] show that for each considered value of β (i.e. $1, 2, 7$) the error variance associated to the separating function f_i is actually minimized for $k = \beta - 1$ (i.e. $k = 0, 1, 6$) which is exactly the expected value. They also confirm that the best separating function g_i is always the identity function (i.e. $k = 1$).

3.2. Application to real signals and comparison with classical rules

Hereafter we present some simulation results for speech signals. In [2] we show that speech can be approximated by an AR process and that its MA whitened version has a Laplace p.d.f (i.e it belongs to the GGF, with $\beta = 1$). The optimum separating function f_{iopt} then corresponds to $sign(x)$. Two types of simulations are considered hereafter to confirm this result: 1) **synthetic mixing matrix**: the mixing and separating filters are strictly causal MA(13). Three separating rules are tested using the same adaptation scheme ($NW1$). They correspond to $f_i(x) \in \{sign(x), x, x^3\}$, and $g(x) = x$. Fig. 2 represents the $SNRI$ with respect to the adaptation gain μ. It shows clearly that $f_i(x) = sign(x)$ highly outperforms the whitened version of classical rules for speech separation and confirms the theoretical results of the previous section. 2) **real mixing matrix**: the mixed signals are measured using a microphone antenna located in a room. The varying parameter here is the inter-microphone distance d that varies from 5 cm to 30 cm. Hereafter, we compare the performance of the separating rule $f_i(x) = sign(x)$ associated with the algorithm ($NW1$) to the classical adaptation approaches using: $f_i(x) \in \{x, x^3\}$ combined with the rule (N1). For all these simulations, the function g_i is set to identity. Table 1 summarizes the performance achieved by these rules. Here again, our approach yields much better performance than the classical rules. It is also more robust to bad conditioning that may occur for small inter-microphone distances.

d (cm)	$f_i(x) = x$	$f_i(x) = x^3$	$f_i(x) = sign(x)$
5	3.7	4.3	7.4
10	5.2	5.1	9.1
15	4.9	4.6	10.0
20	4.8	5.7	9.6
25	5.0	6.3	10.0
30	5.4	6.4	9.7

Table 1: $SNRI$ (dB) for real mixtures, versus inter-microphone distance d and separating function f_i.

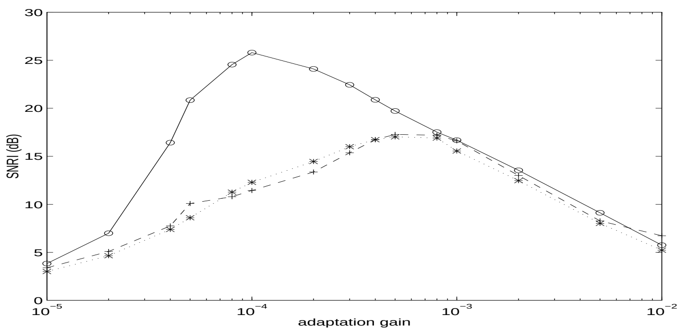

Figure 2: SNRI with respect to adaptation gain μ.
—: $f_i(x) = sign(x)$, - - -: $f_i(x) = x$, ...: $f_i(x) = x^3$

4. Conclusion

This paper presents a generic approach for the separation of two convolutively mixed signals. In the case of strictly causal mixtures, we derive the expression of the asymptotic error variance of the estimation of the mixture filters. The separating functions which minimize the error variance are then derived. They are shown to be respectively *i*) related to the probability density functions of the innovation processes of the sources and *ii*) proportional to the identity function. The validity of those results and the improvement achieved by the associated algorithms are illustrated by simulations performed in various conditions using synthetic or real signals and synthetic or real mixing matrices. These simulations especially show that the proposed approach yields a much better SNRI than classical rules when applied to real audio signals.

References

[1] A. Benveniste, M. Metivier and P. Priouret, *Adaptive algorithms and stochastic approximations,* Applications of Mathematics, vol. 22, Springer-Verlag, 1990.

[2] N. Charkani, " Séparation auto-adaptative de sources pour les mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures", PhD thesis, INP Grenoble, France, 1996.

[3] Y. Deville and N. Charkani, "Analysis of the stability of time-domain source separation algorithms for convolutively mixed signals", to appear in the Proc. ICASSP 97.

[4] C. Jutten and J. Hérault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture", Signal Processing, vol. 24, no. 1, pp. 1-10, July 1991.

[5] H.L. Nguyen Thi and C. Jutten, "Blind source separation for convolutive mixtures" Signal Processing, vol. 45, no. 2, pp. 209-229, March 1995.