A CONVOLUTIVE SOURCE SEPARATION METHOD WITH SELF-OPTIMIZING NON-LINEARITIES

Nabil Charkani

Philips Consumer Communications (PCC)
Advanced Development
Route d'Angers
72081 Le Mans Cedex 9
France.
nabil.charkani@dev-lme.pcc.philips.com

Yannick Deville

Laboratoire d'Acoustique, de Métrologie, d'Instrumentation (LAMI)
Université Paul Sabatier
38 Rue des 36 Ponts
31400 Toulouse
France.
ydeville@cict.fr

ABSTRACT

This paper deals with the separation of two convolutively mixed signals. The proposed approach uses a recurrent structure adapted by a generic rule involving arbitrary separating functions. These functions should ideally be set so as to minimize the asymptotic error variance of the structure. However, these optimal functions are often unknown in practice. The proposed alternative is based on a self-adaptive (sub-)optimization of the separating functions, performed by estimating the projection of the optimal functions on a predefined set of elementary functions. The equilibrium and stability conditions of this rule and its asymptotic error variance are studied. Simulations are performed for real mixtures of speech signals. They show that the proposed approach yields much better performance than classical rules.

1. PROBLEM STATEMENT AND CLASSICAL RESULTS

Multichannel blind (or self-adaptive) source separation is a basic topic in signal processing. It aims at extracting unknown independent signals (the so-called sources) from sensor observations that are unknown linear mixtures of these sources. A commonly used model corresponds to a two-dimensional mixing system defined by the following 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 Z-transforms of the source $x_i(n)$ and observation $y_i(n)$. $A_{ij}(z)$ is the unknown transfer function of the channel that links source j to sensor i. The corresponding impulse response is denoted $(a_{ij}(k))_{k\in Z}$ hereafter. The mixing system is assumed to be minimum-phase (i.e. to be causal and stable and to have a causal and stable inverse). Moreover, the sources $x_1(n)$ and $x_2(n)$ are assumed to be stationary, zero-mean and statistically independent.

This work was partly performed when the authors were with the Laboratoires d'Electronique Philips S.A.S (LEP), at Limeil-Brévannes, France.

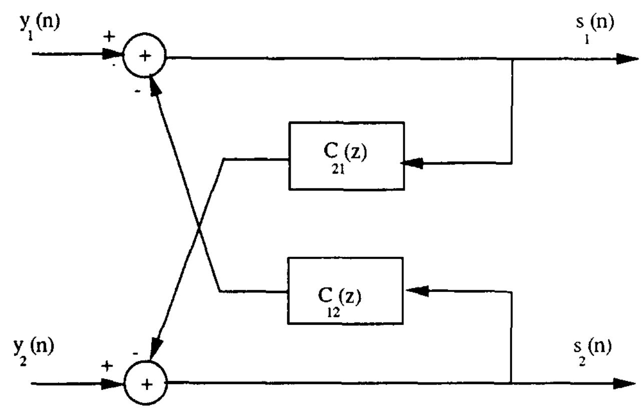

Figure 1: Recurrent convolutive source separation system.

Hérault and Jutten [5] proposed one of the very first solutions to this problem in the case of intantaneous mixtures (i.e. $A_{ij}(z) = a_{ij}(0)$ for $i \neq j \in \{1,2\}$). A natural extension of this solution for convolutive mixtures was then derived by Nguyen and Jutten [6]. Their approach is based on the recurrent structure shown in Figure 1, where the separating filters $C_{ij}(z)$ are assumed to be M^{th}-order Moving Average filters. The associated coefficients $(c_{ij}(k))_{0\leq k\leq M}$ are updated using the rule:

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

where μ is a positive adaptation gain and f and g are two possibly nonlinear (most often odd) functions. The particular case $(f(x), g(x)) = (x^3, x)$ was experimentally investigated in [6] but no theoretical results were provided neither on the stability of (3) nor on its asymptotic behaviour.

2. SUMMARY OF OUR PREVIOUS RESULTS

More recently [2]-[4], we proposed a first extension of (3) 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, except that they are requested to meet the condition

$$E[g_i(x_j)] = 0, \quad i \neq j \in \{1,2\}. \tag{5}$$

This condition is related to the state of interest of the system, i.e. to the values of the separating filters which yield separated sources at the system outputs. This state is called the separating state below and may be shown to correspond to:

$$C_{ij}(z) = A_{ij}(z), \quad i \neq j \in \{1,2\}, \tag{6}$$

which yields $S_i(z) = X_i(z)$, $i \in \{1,2\}$. The condition (5) is then set in order to ensure that this separating state is an equilibrium point of the rule (4). The stability analysis of (4) at the separating state and for independent identically distributed (i.i.d) sources is provided in [4], while its asymptotic behaviour is investigated in [3]. In the case of stricly causal mixing and separating filters (i.e. $a_{12}(0) = a_{21}(0) = c_{12}(0) = c_{21}(0) = 0$), we thus derived the expression of the asymptotic error variance of the separating system, defined as $\sigma_\infty = \lim_{n \to +\infty} E[|\theta_n - \theta^s|^2]$ where

$$\theta_n = [c_{12}(n,1), \ldots, c_{12}(n,M), c_{21}(n,1), \ldots, c_{21}(n,M)]^T \tag{7}$$

defines the system state at time n and

$$\theta^s = [a_{12}(1), \ldots, a_{12}(M), a_{21}(1), \ldots, a_{21}(M)]^T \tag{8}$$

is the separating state (as (8) is then only an explicit form of (6)).

As shown in [2],[3], the minimization of σ_∞ suffers from a multiplicative ambiguity, that can be removed by using the normalized rule that we proposed, i.e:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu \frac{f_i(s_i(n))}{\sqrt{E[f_i^2(s_i)]}} \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}} \quad i \neq j \in \{1,2\}, k \in [1,M]. \tag{9}$$

The optimal separating functions (i.e. the ones which minimize σ_∞) are then:

$$f_{i\,opt}(x) = -\nu_{i1} \frac{p'_{x_i}(x)}{p_{x_i}(x)}, \tag{10}$$

$$g_{i\,opt}(x) = \nu_{i2}\, x, \tag{11}$$

where p_{x_i} is the p.d.f (probability density function) of the source x_i, and ν_{i1} and ν_{i2} are arbitrary[1] real constants, which have no influence on the resulting rule (9).

[1] They should have the same sign to ensure the stability of the separating state.

We also extended this approach [3] to the case of possibly-coloured signals corresponding to AR processes, i.e. to sources defined in the Z-domain by:

$$X_i(z) = \frac{\tilde{X}_i(z)}{\tilde{B}_i(z)}, \quad i \in \{1,2\}, \tag{12}$$

where $\tilde{B}_i(z)$ represents the Z-transform of a causal and minimum-phase q_ith-order filter, i.e. $\tilde{B}_i(z) = \sum_{k=0}^{q_i} \tilde{b}_i(k) z^{-k}$ (with $\tilde{b}_i(0) = 1$), and where $\tilde{x}_i(n)$ are i.i.d. signals called the (normalized) innovation processes of the sources. In this case too, the rule (9) is used, except that s_i (and s_j) is replaced by its adaptively estimated innovation process v_i, computed according to:

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

where the coefficients $(b_i(n,k))_{k \in [1,q_i]}$ of the MA whitening filter are updated by a LMS rule with gain γ:

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

An asymptotic behaviour analysis [2] shows that the optimum class of separating functions is then:

$$f_{i\,opt}(x) = -\nu_{i1} \frac{p'_{\tilde{x}_i}(x)}{p_{\tilde{x}_i}(x)}, \tag{15}$$

$$g_{i\,opt}(x) = \nu_{i2}\, x. \tag{16}$$

3. DEFINITION AND THEORETICAL ANALYSIS OF A NEW APPROACH

In this section, we still consider the case of the separation of two AR processes mixed through stricly causal filters. The implementation of the above-defined optimum separating functions $f_{i\,opt}$ requires the p.d.f $p_{\tilde{x}_i}$ of the whitened versions of the sources to be known. Unfortunately, this turns out to be an unrealistic assumption in most cases. A natural solution to this problem is the estimation of these p.d.f. However, this approach is computationally expensive, hard to use in real-time applications and difficult to extend to the case of non-stationary sources. An attractive alternative is therefore proposed hereafter. It consists in having the separating system automatically determine sub-optimal estimates of these optimal separating functions. This can be achieved by estimating their optimal projections on a set of elementary functions (polynomial functions for instance). Hence, the contribution of each separating function f_i in the extension of (9) to coloured signals, i.e. $\frac{f_i(x)}{\sqrt{E[f_i^2(x)]}}$, is replaced by $h_i(x) = \sum_{k=1}^{L} \omega_{ik} \psi_{ik}(x)$, where $(\psi_{ik}(x))_{k \in [1,L]}$ is a set of continuously derivable functions that define the projection space and $(\omega_{ik})_{k \in [1,L]}$ is a set of scalar coefficients associated to $(\psi_{ik}(x))_{k \in [1,L]}$. When setting the functions g_i

to their optimal values defined in (16), the source separation rule thus obtained reads:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu h_i(v_i(n)) \frac{v_j(n-k)}{\sqrt{E[v_j^2]}} \quad i \neq j \in \{1,2\}, k \in [1,M], \tag{17}$$

still combined with (13)-(14).

The next step of this investigation consists in analyzing the theoretical properties of the new algorithm thus introduced, and especially in determining the optimum values of the coefficients $(\omega_{ik})_{k\in[1,L]}$. To this end, we first reformulate in vector form the overall set of rules defining the adaptation of all the parameters of the considered system, i.e. (14),(17). This yields:

$$\Theta_{n+1} = \Theta_n + \mu H(\Theta_n, \xi_{n+1}), \tag{18}$$

where ξ_{n+1} and $H(\Theta_n, \xi_{n+1})$ are column vectors derived from (14),(17) and where

$$\Theta_n = [\theta_{1n}, \theta_{2n}]^T, \tag{19}$$

with

$$\theta_{1n} = [c_{12}(n,1), \ldots, c_{12}(n,M), c_{21}(n,1), \ldots, c_{21}(n,M)], \tag{20}$$

$$\theta_{2n} = [b_1(n,1), \ldots, b_1(n,q_1), b_2(n,1), \ldots, b_2(n,q_2)]. \tag{21}$$

The equilibrium states of (18) are then the states Θ^* that meet:

$$E_{\Theta^*}[H(\Theta^*, \xi_{n+1})] = 0, \tag{22}$$

where $E_{\Theta^*}[]$ denotes the mathematical expectation associated to the asymptotic probability law of the vector ξ_{n+1} for a given vector Θ^*. One can easily check that the separating state defined by:

$$\Theta^s = [\theta_1^s, \theta_2^s]^T, \tag{23}$$

with:

$$\theta_1^s = [a_{12}(1), \ldots, a_{12}(M), a_{21}(1), \ldots, a_{21}(M)], \tag{24}$$

$$\theta_2^s = [\tilde{b}_1(1), \ldots, \tilde{b}_1(q_1), \tilde{b}_2(1), \ldots, \tilde{b}_2(q_2)] \tag{25}$$

is an equilibrium point of (18).

The stability analysis of (18) is based on the so-called Ordinary Differential Equation technique (ODE) [1], which approximates the discrete recurrence (18) by a continuous differential system that reads:

$$\frac{d\Theta}{dt} = \lim_{n\to+\infty} E_\Theta[H(\Theta, \xi_{n+1})]. \tag{26}$$

The differential system (26) is locally stable in the vicinity of any given equilibrium point Θ^* if and only if all the eigenvalues of the associated Jacobian matrix $J(\Theta^*)$ have negative real parts. For any state Θ, the entries of $J(\Theta)$ are defined by:

$$J_{ij}(\Theta) = \lim_{n\to+\infty} \frac{\partial(E_\Theta[H(\Theta,\xi_{n+1})]^{(i)})}{\partial\Theta^{(j)}}, \tag{27}$$

$E_\Theta[H(\Theta,\xi_{n+1})]^{(i)}$ being the i^{th} component of $E_\Theta[H(\Theta,\xi_{n+1})]$ and $\Theta^{(j)}$ the j^{th} component of vector Θ. Applying this approach to the considered adaptation rule eventually yields the following stability condition at the separating state (see details in [2]):

$$E[h_i'(\tilde{x}_i)] > 0. \tag{28}$$

The ODE asymptotic convergence theorem established in [1] is then applied to the algorithm (18). This shows that, for large n and a stable equilibrium point Θ^s, 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:

$$J(\Theta^s)P + PJ^T(\Theta^s) + R(\Theta^s) = 0 \tag{29}$$

where $R(\Theta^s) = \sum_{n\in\mathcal{Z}} Cov[H(\Theta^s,\xi_{n+1}), H(\Theta^s,\xi_0)]$.

The overall asymptotic error variance of Θ_n is thus equal to $\lim_{n\to+\infty} E[|\Theta_n - \Theta^s|^2] = \mu Tr(P)$. However, we are only interested in the eventual separation quality of the system outputs, and therefore in the asymptotic error variance related to the estimation of the separating filters, i.e. $\sigma_\infty = \lim_{n\to+\infty} E[|\theta_{1n} - \theta_1^s|^2]$. It can be shown [2] that[2] $\sigma_\infty = \mu Tr(\tilde{P})$ where $\tilde{P}$ is the unique symmetric and positive definite solution of the Lyapunov equation:

$$\tilde{J}(\theta_1^s)\tilde{P} + \tilde{P}\tilde{J}^T(\theta_1^s) + \tilde{R}(\theta_1^s) = 0 \tag{30}$$

where $\tilde{J}(\theta_1^s)$ and $\tilde{R}(\theta_1^s)$ correspond to 2M by 2M matrices obtained by keeping the first 2M raws and columns of Θ^s and $R(\Theta^s)$. Mathematical calculations then show that [2]:

$$\sigma_\infty = \mu \sum_{i=1,i\neq j}^{i,j=2} \frac{(q_{i2}E[h_i^2(\tilde{x}_i)] + q_{i1}E^2[h_i(\tilde{x}_i)])}{\sqrt{E[h_i^2(\tilde{x}_i)]}E[h_i'(\tilde{x}_i)]} \frac{1}{\sqrt{E[\tilde{x}_j^2]}}, \tag{31}$$

where q_{i1} and q_{i2} are real constants that are only related to the mixing matrix. The minimization of σ_∞ has to be undertaken under three constraints. i.e. (28), $E[h_i^2(\tilde{x}_i)] = 1$ and $E[h_i(\tilde{x}_i)] = 0$. The first one guarantees the stability of Θ^s whatever the nature of the sources (e.g. sub- or super-Gaussian). The second one results from the fact that $h_i(x)$ stands for $\frac{f_i(x)}{\sqrt{E[f_i^2(x)]}}$, as explained above. The third constraint is only used to reduce the complexity of the solution. This assumption makes sense since the optimum separating function (15) may be shown to be zero-mean, but it doesn't ensure that the sub-optimal function thus obtained is very close to the optimum. This constrained minimization of σ_∞ leads to the following set of coefficients $(\omega_{ik})_{k\geq 1}$:

$$\begin{cases} \omega_{i1} = \dfrac{sign(\sum_{k=1}^{L} \frac{d_{ik}}{d_{i1}} E[\psi_{ik}'(\tilde{x}_i)])}{\sqrt{\sum_{k,l=1}^{L} \frac{d_{ik}d_{il}}{d_{i1}^2} E[\psi_{ik}(\tilde{x}_i)\psi_{il}(\tilde{x}_i)]}} \\ \omega_{ik} = \omega_{i1}\dfrac{d_{ik}}{d_{i1}} \quad k \in [2,L] \end{cases} \tag{32}$$

[2]This result is obtained thanks to the lower-block-triangular structure of $J(\Theta^s)$.

where $d_{ik}, k \in [1, L]$, are the entries of the vector D_i defined by:

$$D_i = \Psi_i^{-1}\Gamma_i, \qquad if\ V_i = 0,$$
$$D_i = \Psi_i^{-1}\Gamma_i - (\frac{V_i^T\Psi_i^{-1}\Gamma_i}{V_i^T\Psi_i^{-1}V_i})\Psi_i^{-1}V_i, \qquad if\ V_i \neq 0, \tag{33}$$

where Ψ_i is the matrix defined as:

$$\begin{pmatrix} E[\psi_{i1}(\tilde{x}_i)\psi_{i1}(\tilde{x}_i)] & \cdots & E[\psi_{i1}(\tilde{x}_i)\psi_{iL}(\tilde{x}_i)] \\ \vdots & \vdots & \vdots \\ \vdots & E[\psi_{ik}(\tilde{x}_i)\psi_{il}(\tilde{x}_i)] & \vdots \\ \vdots & \vdots & \vdots \\ E[\psi_{iL}(\tilde{x}_i)\psi_{i1}(\tilde{x}_i)] & \cdots & E[\psi_{iL}(\tilde{x}_i)\psi_{iL}(\tilde{x}_i)] \end{pmatrix} \tag{34}$$

and where:

$$\Gamma_i^T = \left[E[\psi_{i1}^{'}(\tilde{x}_i)], \ldots, E[\psi_{iL}^{'}(\tilde{x}_i)]\right]^T, \tag{35}$$

$$V_i^T = [E[\psi_{i1}(\tilde{x}_i)], \ldots, E[\psi_{iL}(\tilde{x}_i)]]^T. \tag{36}$$

Note that d_{i1} was assumed to be non null above. If this assumption does not hold, a permutation of indices in the vector D allows to fulfill this requirement.
In practical implementations, the mathematical expectations in the above expressions are adaptively estimated using a classical first-order AR filtering algorithm, with $\tilde{x}_i$ estimated by v_i [2]. Note that the choice of the elementary functions of the projection space may take a great advantage from the a priori knowledge that we may have on some characteristics of the p.d.f of the sources (sub-Gaussianity or super-Gaussianity for example).

4. EXPERIMENTAL RESULTS

In this section, we consider the realistic situation where the sources are two speech signals limited to the telephone band $[300Hz - 3400Hz]$ and where the mixing system results from acoustic propagation inside a room. The signals are emitted by two loudspeakers and picked up by an antenna made up of eight microphones, with 40-cm inter-louspeakers and loudspeaker-antenna distances, while the distance between adjacent microphones is 5 cm.
The source separation accuracy is measured by the Average Signal to Noise Ratio Improvement defined as :

$$ASNRI = \frac{SNRI_1 + SNRI_2}{2}, \tag{37}$$

where $SNRI_i$ denotes the SNRI at output i defined as:

$$SNRI_i = 10log_{10}\left(\frac{E[(y_i(n) - x_i(n))^2]}{E[(s_i(n) - x_i(n))^2]}\right). \tag{38}$$

In these simulations, we compare the performance of two types of algorithms. The first one is the practical algorithm derived in Section 2, i.e. the adaptation rule (9) modified with the whitening algorithm (13)-(14) and operated with either of the following sets of classical separating functions:

d (cm)	$f_i(x) = x$	$f_i(x) = x^3$	$f_i(x) = h^{(3D)}(x)$
5	3.7	4.3	8.0
10	5.2	5.1	9.1
15	4.9	4.6	10.3
20	4.8	5.7	10.0
25	5.0	6.3	10.0
30	5.4	6.4	10.0

Table 1: Comparative performance for real mixtures (ASNRI (dB) vs inter-microphone distance d).

$(f_i(x), g_i(x)) \in \{(x, x), (x^3, x)\}$. The second considered approach is the new algorithm (17) introduced in this paper (again combined with (13)-(14)), with $h_i(x)$ situated in a 3-dimensional projection space (and therefore denoted $h^{(3D)}(x)$ hereafter). This projection space is defined by the following set of elementary functions: $(sign(x)\sqrt{|x|}, x, x^3)$. The results thus obtained are summarized in Table 1. This shows that the proposed projection-based approach yields much better performance than the normalized rule operated with classical separating functions. It is also more robust to bad conditioning that may occur for small inter-microphone distances.

5. CONCLUSION

In this paper, we defined and analyzed a new approach to convolutive source separation. It is based on a self-adaptive optimization of separating functions. This is performed by estimating the best projection of the optimum separating functions on a given set of functions, with no restriction on the nature of the sources. Simulations using real mixtures of speech signals show that this approach performs much better than classical rules.

6. REFERENCES

[1] A. Benvéniste, 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] N. Charkani, Y. Deville, "Optimization of the asymptotic performance of time-domain convolutive source separation algorithms", Proc. ESANN'97, pp. 273-278, Bruges, Belgium, April 16-18, 1997.

[4] Y. Deville, N. Charkani "Analysis of the stability of time-domain source separation algorithms for convolutively mixed signals", Proc. ICASSP 97, pp. 1835-1838, Munich, Germany, April 21-24, 1997.

[5] C. Jutten, 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.

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