A GEOMETRIC APPROACH FOR SEPARATING POST NON-LINEAR MIXTURES

Massoud Babaie-Zadeh[1,2] mbzadeh@yahoo.com

Christian Jutten[1] * Christian.Jutten@inpg.fr

Kambiz Nayebi[2] knayebi@sina.sharif.ac.ir

[1] Laboratoire des Images et des Signaux (LIS), Institut National Polytechnique de Grenoble (INPG), Grenoble, France.

[2] Electrical engineering department, Sharif University of Technology, Tehran, Iran.

ABSTRACT

A geometric method for separating PNL mixtures, for the case of 2 sources and 2 sensors, has been presented. The main idea is to find compensating nonlinearities to transform the scatter plot of observations to a parallelogram. It then results in a linear mixture which can be separated by any linear source separation algorithm. An indirect result of the paper is another separability proof of PNL mixtures of bounded sources for 2 sources and 2 sensors.

1 INTRODUCTION

Blind Source Separation (BSS) consists in retrieving independent source signals, say $\mathbf{s} = (s_1, \ldots, s_N)^T$, from observations consisting of a mixture of them, say $\mathbf{x} = (x_1, \ldots, x_N)^T$. There is no prior information either about the sources or the mixing system, hence the term *Blind*. This problem has first been introduced in the 80's [1] and has been studied by many researchers in the last decade (see [2] for state-of-the-art and references). There exist many algorithms, based on the estimation of a separating system such that the output signals are independent. For linear mixtures, it has been shown [3] that, when at most one of the sources is Gaussian, the independence implies the separation of the sources, up to a scale and a permutation. One then says the linear mixtures are *separable*. Conversely, general nonlinear mixtures are not separable, except if one adds structural constraints or regularization techniques. In this paper, we consider Post Non-Linear (PNL) mixtures (Fig. 1), which are special nonlinear mixtures with structural constraints, and are separable [4].

Puntonet *et al* [5] proposed a geometric approach for separating sources from two mixtures of two independent and bounded random variables s_1 and s_2. Then, the joint Probability Density Function (PDF) $p_{S_1S_2}(s_1, s_2)$ has non-zero values only within a rectangular region. Moreover, the joint PDF, $p_{X_1X_2}(x_1, x_2)$, of the mixtures has non-zero values only within a parallelogram, whose borders determine the mixing matrix $\mathbf{A}$. If the mixing matrix is:

$$\mathbf{A} = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix} \tag{1}$$

i.e. $\mathbf{x} = \mathbf{A}\mathbf{s}$, the borders of the joint PDF in (x_1, x_2) plane are straight lines with slopes equal to b and $1/a$. As a result, if the number of observed data is sufficient, we can estimate this parallelogram and consequently the parameters a and b of the mixing matrix.

In this paper, we used a geometric approach for separating two independent sources from two PNL mixture (Fig. 1). For these mixtures, the joint plot is a parallelogram in the (w_1, w_2) plane, and a 'nonlinear' region in the (e_1, e_2) plane (see Fig. 3). For separating the mixture, we first have to estimate two compensating nonlinear functions g_1 and g_2 to transform this region to a parallelogram. Then, the sensor nonlinearities are compensated, and the sources can be separated by means of any linear BSS algorithm.

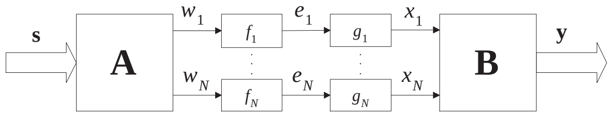

Figure 1: PNL mixing-separating model.

2 PRELIMINARY ISSUES

Suppose the joint plot of the observations is mapped, by the nonlinearities g_1 and g_2, into a parallelogram in the (x_1, x_2) plane. Does it insure that $g_1 \circ f_1$ and $g_2 \circ f_2$ are linear? The answer is positive, as proved in this section.

Theorem 1 (main theorem) *Consider the transformation:*

$$\begin{cases} x_1 = h_1(w_1) \\ x_2 = h_2(w_2) \end{cases} \tag{2}$$

where h_1 and h_2 are analytic functions. If the borders of a parallelogram in the (w_1, w_2) plane are transformed

*This work has been partially supported by European project BLISS (IST-1999-14190). Ch. Jutten is professor with ISTG, university Joseph Fourier of Grenoble.

to the borders of a parallelogram in the (x_1, x_2) plane, and the borders of these parallelograms are not parallel to the coordinate axes, then, there are constants a_1, a_2, b_1 and b_2 such that:

$$\begin{cases} h_1(x) = a_1 x + b_1 \\ h_2(x) = a_2 x + b_2 \end{cases} \tag{3}$$

Remark 1: With (2), w_2 = cte is mapped into x_2 = cte. Moreover, if h_1 and h_2 are monotonous, then the point order on this line remains unchanged (h_2 increasing) or reversed (h_2 decreasing). Therefore, the borders of any transformed region in (x_1, x_2) are the mappings of the borders of the corresponding region in (w_1, w_2).

Remark 2: The theorem shows that, as in the linear mixtures, with bounded sources, the borders of the joint plot are sufficient for separating the sources.

Remark 3: The existence of the constants b_1 and b_2, emphasizes on a 'DC' indeterminacy which always exists in separating the sources, but is generally skipped by using zero-mean sources.

Remark 4: The theorem provides a proof of the separability for two PNL mixtures of two bounded sources.

To prove the theorem, we need the following Lemma:

Lemma 1 *Let f be an analytic function on the interval D_f such that $0 \in D_f$. Suppose that for all x in a neighborhood of 0 we have:*

$$f(x) = c_1 f(c_2 x) \tag{4}$$

where $c_1 \neq 1$ and $c_2 \neq 1$.

1. *If $\exists n \in \mathbb{N}$ such that $c_1 c_2^n = 1$, then $\exists a \in \mathbb{R}$ such that $f(x) = a\, x^n, \quad \forall x \in D_f$,*
2. *if $\forall n \in \mathbb{N}, c_1 c_2^n \neq 1$, then $f(x) = 0, \quad \forall x \in D_f$.*

Proof: From (4), we deduce $f^{(n)}(x) = c_1 c_2^n f^{(n)}(c_2 x)$ where $f^{(n)}(x)$ denotes the n-th order derivative of f. By letting $x = 0$, we conclude that, $\forall n$ such that $c_1 c_2^n \neq 1$, we have $f^{(n)}(0) = 0$. Then, the final result comes from the Taylor expansion of f around $x = 0$. □

Proof of theorem 1: In the proof, we assume, without loss of generality (the general case is trivial by changing the variables), that $h_1(0) = h_2(0) = 0$ and the origin is a corner of the parallelogram.

Consider two borders of the parallelogram crossing the origin. Since the mappings of $w_2 = c_1 w_1$ and $w_2 = c_2 w_1$ (the borders in the (w_1, w_2) plane) are $x_2 = d_1 x_1$ and $x_2 = d_2 x_1$ (the borders in the (x_1, x_2) plane), we have (note that c_1, c_2, d_1 and d_2 are neither zero nor infinite, and $c_1 \neq c_2$ and $d_1 \neq d_2$):

$$\begin{cases} h_2(c_1 w) = d_1 h_1(w) \\ h_2(c_2 w) = d_2 h_1(w) \end{cases} \tag{5}$$

Combining these equations, with a little algebra, leads to:

$$h_2(w) = \frac{d_1}{d_2} h_2(\frac{c_2}{c_1} w) \tag{6}$$

The conditions of Lemma 1 are now satisfied, and we conclude that $h_2(w) = aw^n$. Then, from this result and (5), we also conclude $h_1(w) = a'w^n$. However, if $n \neq 1$, then the mappings of any borders, which are not crossing the origin, are not straight lines. Hence, $n = 1$, *i.e.* h_1 and h_2 are both linear. □

3 SEPARATING ALGORITHM

Keeping in mind the theorem 1, we separate the sources with three steps: (i) determining the borders of the joint plot, (ii) computing nonlinear transformations g_1 and g_2, which transform these curves into a parallelogram, and (iii) separating the sources from the resulting linear mixture.

3.1 Compensating for the nonlinearities

In this section, we develop an iterative method for determining the nonlinear transformations which transform the boundary curves to a parallelogram. The main idea, is to minimize the difference between the transformed boundary curves and the best parallelogram which fits on them. Hence, in each iteration, we modify g_1 and g_2 for decreasing this error, the parallelogram being assumed to be fixed, then we calculate the parallelogram which optimally fits on the transformed boundary, and so on.

Consider the vertical line $e_1 = c$, with values c for which two intersections occur with the borders of the joint plot. Let us denote $e_2 = \mathcal{L}_e(e_1)$ the set of intersection points with smaller coordinates e_2 (we call it the 'lower' boundary) and $e_2 = \mathcal{U}_e(e_1)$ the set of intersection points with larger coordinates e_2 (we call it the 'upper' boundary), as shown in Fig. 3. The corresponding curves in the (x_1, x_2) plane are denoted by $x_2 = \mathcal{L}_x(x_1)$ and $x_2 = \mathcal{U}_x(x_1)$. Moreover, for the sake of simplicity, here we assume that, in (1), a and b are both positive. This restriction insures that the functions $\mathcal{U}_e$, $\mathcal{L}_e$, $\mathcal{U}_x$ and $\mathcal{L}_x$ are all invertible. Without this restriction, the method remains applicable, but instead of working with these functions, we must consider each border piece (4 pieces), which makes the equations too complicated for stating the main points.

3.1.1 Finding compensating nonlinear functions

Suppose that we have found the 'best' parallelogram fitted on the curves $\mathcal{L}_x$ and $\mathcal{U}_x$, and denote the lower and upper borders of this parallelogram by l_l and l_u, respectively (each one is composed of two straight lines). For finding mappings g_1 and g_2, we minimize the error between this parallelogram and the actual curve in (x_1, x_2)

plane. For determining g_2, the cost function is:

$$\mathcal{E}_2 = E\left\{[\mathcal{L}_x(x_1) - l_l(x_1)]^2\right\} + E\left\{[\mathcal{U}_x(x_1) - l_u(x_1)]^2\right\} \tag{7}$$

This criterion is in fact:

$$\begin{aligned}\mathcal{E}_2 &= E\left\{\left[g_2(e_2) - l_l\big(g_1(e_1)\big)\right]^2\right\}_{e_2=\mathcal{L}_e(e_1)} \\ &\quad + E\left\{\left[g_2(e_2) - l_u\big(g_1(e_1)\big)\right]^2\right\}_{e_2=\mathcal{U}_e(e_1)} \\ &= E\left\{\left[g_2(e_2) - l_l\Big(g_1\big(\mathcal{L}_e^{-1}(e_2)\big)\Big)\right]^2\right\} \\ &\quad + E\left\{\left[g_2(e_2) - l_u\Big(g_1\big(\mathcal{U}_e^{-1}(e_2)\big)\Big)\right]^2\right\}\end{aligned} \tag{8}$$

Note that in the first equality of the above equation, the expectations are calculated on the curves $e_2 = \mathcal{L}_e(e_1)$ and $e_2 = \mathcal{U}_e(e_1)$, while on the second equality, they are taken over the whole range of e_2, since the constraints are moved into the brackets.

Now, let $\hat{g}_2 = g_2 + \epsilon_2$, where ϵ_2 is a 'small' function. Direct calculations show that, up to the first order terms, we have (note that l_l and l_u – corresponding to the parallelogram boundaries – are assumed to be fixed):

$$\begin{aligned}\frac{\hat{\mathcal{E}}_2 - \mathcal{E}_2}{2} &= E\left\{\left[g_2(e_2) - l_l\Big(g_1\big(\mathcal{L}_e^{-1}(e_2)\big)\Big)\right]\epsilon_2(e_2)\right\} \\ &\quad + E\left\{\left[g_2(e_2) - l_u\Big(g_1\big(\mathcal{U}_e^{-1}(e_2)\big)\Big)\right]\epsilon_2(e_2)\right\} \\ &= \int_{-\infty}^{+\infty} \mathcal{G}_2(e_2)\epsilon_2(e_2)p_{e_2}(e_2)de_2\end{aligned} \tag{9}$$

where $p_{e_2}(e_2)$ is the PDF of e_2 and:

$$\mathcal{G}_2(e_2) \triangleq 2g_2(e_2) - l_l\Big(g_1\big(\mathcal{L}_e^{-1}(e_2)\big)\Big) - l_u\Big(g_1\big(\mathcal{U}_e^{-1}(e_2)\big)\Big) \tag{10}$$

The above equation shows that the 'gradient' of $\mathcal{E}_2$ with respect to the function g_2 via the weighting function $p_{e_2}(e_2)$ can be defined as the function $\mathcal{G}_2$. In other words the schematic algorithm $g_2 \leftarrow g_2 - \mu\mathcal{G}_2$, where μ is a small positive constant, insures decreasing of $\mathcal{E}_2$ (provided that all the other parameters remain unchanged). In the same manner, one derives the algorithm $g_1 \leftarrow g_1 - \mu\mathcal{G}_1$, where:

$$\mathcal{G}_1(e_1) \triangleq 2g_1(e_1) - l_l^{-1}\Big(g_2\big(\mathcal{L}_e(e_1)\big)\Big) - l_u^{-1}\Big(g_2\big(\mathcal{U}_e(e_1)\big)\Big) \tag{11}$$

insures a reduction in:

$$\begin{aligned}\mathcal{E}_1 &= E\left\{\left[\mathcal{L}_x^{-1}(x_2) - l_l^{-1}(x_2)\right]^2\right\} \\ &\quad + E\left\{\left[\mathcal{U}_x^{-1}(x_2) - l_u^{-1}(x_2)\right]^2\right\}\end{aligned} \tag{12}$$

Note that, although the criteria (7) and (12) are different, their joint minimization solves our problem.

3.1.2 Fitting the parallelogram

In this subsection, we deal with the problem of fitting a parallelogram to the curves $\mathcal{L}_x$ and $\mathcal{U}_x$. As in the previous section, let l_l and l_u denote the lower and upper borders of the desired parallelogram:

$$\begin{aligned}l_l(x_1) &= \beta_{01} + \beta_1(x_1 - \alpha_1)_- + \beta_2(x_1 - \alpha_1)_+ \\ l_u(x_1) &= \beta_{02} + \beta_2(x_1 - \alpha_2)_- + \beta_1(x_1 - \alpha_2)_+\end{aligned} \tag{13}$$

where $(u)_- \triangleq \min(0, u)$ and $(u)_+ \triangleq \max(0, u)$.

Suppose that the curves $\mathcal{L}_x$ and $\mathcal{U}_x$ are known via M sample points $(x_1(k), x_{21}(k))$ and $(x_1(k), x_{22}(k))$, $k = 1, \ldots, M$, where $x_{21}(k) = \mathcal{L}_x(x_1(k))$ and $x_{22}(k) = \mathcal{U}_x(x_1(k))$. We are looking for the constants α_1, α_2, β_{01}, β_{01}, β_1 and β_2 which minimize $\mathcal{C} = \mathcal{C}_l + \mathcal{C}_u$, where:

$$\begin{aligned}\mathcal{C}_l &= \sum_k \left[x_{21}(k) - l_l(x_1(k))\right]^2 \\ \mathcal{C}_u &= \sum_k \left[x_{22}(k) - l_u(x_1(k))\right]^2\end{aligned} \tag{14}$$

Assume we know the change-points α_1 and α_2. Then we must find $\boldsymbol{\beta} = (\beta_{01}, \beta_1, \beta_2, \beta_{02})^T$ which minimizes:

$$\mathcal{C} = \|\mathbf{x}_{21} - \mathbf{L}_1\boldsymbol{\beta}\|^2 + \|\mathbf{x}_{22} - \mathbf{L}_2\boldsymbol{\beta}\|^2 \tag{15}$$

where $\mathbf{x}_{21} = (x_{21}(1), \ldots, x_{21}(M))^T$, $\mathbf{x}_{22} = (x_{22}(1), \ldots, x_{22}(M))^T$ and:

$$\mathbf{L}_1 = \begin{bmatrix} 1 & (x_1(1) - \alpha_1)_- & (x_1(1) - \alpha_1)_+ & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_1(M) - \alpha_1)_- & (x_1(M) - \alpha_1)_+ & 0 \end{bmatrix}$$

$$\mathbf{L}_2 = \begin{bmatrix} 0 & (x_1(1) - \alpha_2)_+ & (x_1(1) - \alpha_2)_- & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & (x_1(M) - \alpha_2)_+ & (x_1(M) - \alpha_2)_- & 1 \end{bmatrix}$$

Solving $\partial\mathcal{C}/\partial\boldsymbol{\beta} = 0$ leads to:

$$\boldsymbol{\beta}_{\text{opt}} = (\mathbf{L}_1^T\mathbf{L}_1 + \mathbf{L}_2^T\mathbf{L}_2)^{-1}(\mathbf{L}_1^T\mathbf{x}_{21} + \mathbf{L}_2^T\mathbf{x}_{22}) \tag{16}$$

Determining the change-points α_1 and α_2 for minimizing $\mathcal{C}$ is difficult. Instead of this, we compute the values of α_1 and α_2 which minimize $\mathcal{C}_l$ and $\mathcal{C}_u$, respectively. This is done by using the Hudson's algorithm [6].

3.1.3 Estimating borders

Equations (10) and (11) require the estimation of the borders $\mathcal{L}_e$ and $\mathcal{U}_e$. For this purpose, we first divide e_1 in K regular intervals. Let $e_1'(k)$ be the mid-point of the k-th interval. Then we denote $e_{2,L}'(k)$ the minimum value of e_2 among all the points which are in the corresponding strip of e_1, and $e_{2,U}'(k)$ their maximum. Then, we estimate $\mathcal{L}_e$ and $\mathcal{U}_e$ as the smoothing splines fitting on the points $(e_1'(k), e_{2,L}'(k))$ and $(e_1'(k), e_{2,U}'(k))$, respectively. This choice results in error smoothing, but the smoothing parameter of the splines (as defined in MATLAB spline toolbox) must be chosen close to 1, to be able to model rapid changes in the borders.

3.2 Separating Linear mixture

After having estimated the best parallelogram fitted on the joint plot in (x_1, x_2) plane, the estimated separating matrix is:

$$\mathbf{B} = \begin{bmatrix} 1 & 1/\beta_2 \\ \beta_1 & 1 \end{bmatrix}^{-1} \tag{17}$$

Finally, for limiting the degree of freedom in determining the functions g_i, we apply a smoothing procedure on these functions at each iteration (as in [7]). Moreover, the DC and energy indeterminacies are removed by filtering and normalization at each iteration.

4 EXPERIMENTAL RESULTS

Here, we present the separation result for mixtures of a sine and a triangle waveform. The sources are mixed by:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \tag{18}$$

and the sensor nonlinearities are:

$$\begin{aligned} f_1(x) &= \tanh(x) + 0.1x \\ f_2(x) &= \tanh(2x) + 0.1x \end{aligned} \tag{19}$$

The main parameters are: the sample size is 3,000, $M = 500$, $\mu = 0.05$ and $\lambda = 0.99999$ (λ is the smoothing parameter of all the smoothing splines, used for estimating $\mathcal{L}_e$ and $\mathcal{U}_e$, and for smoothing the g_i's). Fifty intervals are used in the range of variations of e_1 for estimating $\mathcal{L}_e$ and $\mathcal{U}_e$.

Figure 2 shows the output Signal to Noise Ratios (SNR's) in dB, defined by (assuming there is no permutation):

$$\mathrm{SNR}_i = 10 \log_{10} \frac{s_i^2}{(y_i - s_i)^2} \tag{20}$$

Figure 4 shows the resulting estimated parallelogram in the (x_1, x_2) plane. The fluctuations in the estimated borders comes from estimation errors due to the sparse number of points close to the borders in the joint plot in the (e_1, e_2) plane. However, the experimental results show that although these borders have been roughly estimated, the separating parameters are close to the optimal ones, due to spline smoothing estimation of g_i's.

5 CONCLUSION

In this paper, we proposed a geometric method for separating sources in PNL mixtures. The algorithm is based on the estimation of component-wise nonlinear mappings which transform the joint plot of the observations into a parallelogram, which results in compensating the sensor nonlinearities. After this compensation, we have a linear mixture which can be processed by any linear BSS algorithm. The algorithm is robust enough to errors in estimating the border of joint plot of observations, but it requires a sufficient number of samples, and its extension to mixtures of more than two sources seems tricky.

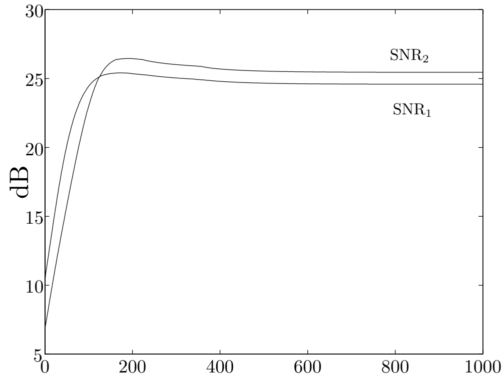

Figure 2: Output SNR's versus iteration.

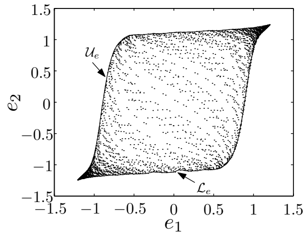

Figure 3: Scatter plot of observations and estimated curves $\mathcal{L}_e$ and $\mathcal{U}_e$.

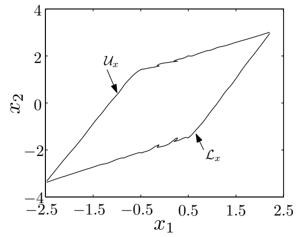

Figure 4: The curves $\mathcal{L}_x$ and $\mathcal{U}_x$.

References

[1] J. Hérault and C. Jutten, "Space or time adaptive signal processing by neural networks models", in *Intern. Conf. on Neural Networks for Computing*, Snowbird (Utah, USA), 1986, pp. 206–211.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiely & Sons, 2001.

[3] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[4] A. Taleb and C. Jutten, "Source separation in post nonlinear mixtures", *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.

[5] C. Puntonet, A. Mansour, and C. Jutten, "A geometrical algorithm for blind separation of sources", in *Actes du XVème Colloque GRETSI 95*, Juan-Les-Pins, France, Septembre 1995, pp. 273–276.

[6] D. J. Hudson, "Fitting segmented curves whose joint points have to be estimated", *Journal of the American Statistical Association*, vol. 61, no. 316, pp. 1097–1129, December 1966.

[7] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Using joint score functions in separating post non-linear mixtures", *Scientia-Iranica*, 2002, accepted.