A mutual information minimization approach for a class of nonlinear recurrent separating systems

Leonardo Tomazeli Duarte, Christian Jutten

A MUTUAL INFORMATION MINIMIZATION APPROACH FOR A CLASS OF NONLINEAR RECURRENT SEPARATING SYSTEMS

Leonardo Tomazeli Duarte and Christian Jutten

GIPSA-lab, INPG-CNRS
46 Avenue Félix Viallet, Grenoble, France
Leonardo.duarte@lis.inpg.fr, Christian.Jutten@inpg.fr

ABSTRACT

In this work, we deal with nonlinear blind source separation. Our contribution is the derivation of a learning strategy that minimizes the mutual information between the outputs of a class of nonlinear recurrent separating systems. By using the concept of the differential of the mutual information, we obtain an algorithm that does not need a precise knowledge of the source distributions, in contrast to the one obtained by a direct derivation of the minimum mutual information framework, or equally the maximum likelihood approach, for the considered model. The validity of our approach is supported by simulations.

1. INTRODUCTION

The problem of blind source separation (BSS) concerns the retrieval of an unknown set of source signals by using only samples that are mixtures of these original signals. A myriad of methods were proposed for the case in which the mixing process is of linear nature [1]. The basis of the majority of these techniques is the independent component analysis (ICA) [2]. When the mixing system is nonlinear, the BSS problem becomes more difficult given that in this situation the recovery of the independence, which is the very essence of ICA, does not guarantee, as a rule, source separation. In view of this limitation, a more reasonable approach is to consider constrained mixing systems as, for example, post-nonlinear (PNL) mixtures [3] and linear-quadratic mixtures [4].

In this work, we deal with the problem of BSS in a particular class of nonlinear systems which is related to a chemical sensing application. Our approach is based on a recursive separating system and the main contribution of this work is the derivation of a learning rule based on the ICA paradigm of mutual information minimization. The paper is organized as follows: in Section 2, we begin with a description of the mixing model as well as of the chosen separating structure. In Section 3, we discuss the application of the minimum mutual information paradigm on the considered separating structure. Firstly, in subsection 3.1, we notice through an example that a mutual information-based algorithm resulted from a direct derivation may not be able to provide source separation. Then, in subsection 3.2, a learning rule based on the concept of the differential of the mutual information is proposed. In Section 4, simulations are carried out in order to verify the viability of our proposal. Finally, in Section 5, we state our conclusions and remarks.

Leonardo Tomazeli Duarte would like to thank CNPq (Brazil) for the financial support.

2. PROBLEM STATEMENT

Blind source separation methods have been proved to be a very promising approach in chemical sensing applications. In [5, 6], for instance, BSS algorithms were used in the problem of estimating the concentrations of several ions in a solution. A remarkable feature in this kind of application is that the sensors have a nonlinear response and, thus, one must resort to nonlinear source separation techniques. A first approach in this spirit was presented in [6], in which a post-nonlinear separation algorithm was applied on a set of mixtures provided by a sensor array of ion-sensitive field-effect transistors (ISFET). Despite the good results obtained by this proposal, it is limited to situations in which all the ions in the solution have the same valence.

In the present work, we envisage the situation in which the valences are different. We take as our starting point the Nikolsky-Eisenman model [5], which gives a simple and yet adequate description of potentiometric-based ion concentration sensors, as the ISFET ones. According to this model, the response of the *i*-th ISFET sensor is given by:

$$x_i = c_{i1} + c_{i2} \log\Big(s_i + \sum_{j,j\neq i} a_{ij} s_j^{\frac{z_i}{z_j}}\Big), \qquad (1)$$

where s_i and s_j are, respectively, the concentration of the ion of interest and the concentration of the *j*-th interfering ion; and where z_i and z_j denote the valences of the ions i and j, respectively. The selective coefficients a_{ij} model the interference process; c_{i1} and c_{i2} are constants that depends on some physical parameters.

Assuming that c_{i1} and c_{i2} are previously estimated and considering a mixture of two ions, the model (1) can be simplified to the following one that will be considered in this work

$$\begin{array}{rcl} x_1 & = & s_1 + a_{12} s_2^k \\ x_2 & = & s_2 + a_{21} s_1^{\frac{1}{k}} \end{array} , \qquad (2)$$

where x_i and s_j denote the ***i***-th mixture and the ***j***-th source, respectively. In the context of this application, k is related to the ratio of the valences of the two considered ions.

2.1. Adopted separating system

For separating sources s_i from mixtures (2), we adopt the following recurrent network as separating system:

$$\begin{array}{rcl} y_1(n+1) & = & x_1 - w_{12} y_2(n)^k \\ y_2(n+1) & = & x_2 - w_{21} y_1(n)^{\frac{1}{k}} \end{array} , \qquad (3)$$

where the vector $\mathbf{w} = [w_{12}\ w_{21}]^T$ denotes the parameters to be adjusted and $\mathbf{y}(n) = [y_1(n)\ y_2(n)]^T$ corresponds to the separating system outputs at time n. For each sample of the mixtures, and for a given value $\mathbf{w}$, the system outputs are obtained after the convergence of the dynamics (3).

In order to understand how this structure works, let $\mathbf{s} = [s_1\ s_2]^T$ denote a sample of the sources. By considering (2), one can easily check that when $[w_{12}\ w_{21}]^T = [a_{12}\ a_{21}]^T$, then $\mathbf{y}(n+1) = \mathbf{y}(n) = \mathbf{s}$ correspond to an equilibrium point of (3), that is to say, it becomes possible to counterbalance the action of the mixing system without relying on its direct inversion. This sort of approach was firstly developed in [7] regarding linear BSS. Its extension to the problem of nonlinear BSS was proposed by Hosseini and Deville [4] in the context of source separation of linear-quadratic mixtures.

Given that the adopted separating system corresponds to a nonlinear dynamic, a crucial point concerns the stability of the separating solution $\mathbf{y} = \mathbf{s}$. In [8], it was shown that, for each sample of the source, a necessary stability condition is given by:

$$|a_{12} a_{21} s_1^{(\frac{1}{k}-1)} s_2^{k-1}| < 1. \qquad (4)$$

In this same work, we described a ICA learning rule for (3) based on the cancellation of the nonlinear correlations between their outputs. The main advantage of this technique relies on its simplicity, given that only the estimation of higher-order (4th) moments is required. On the other hand, there are convergence problems which are exacerbated when the sources are close to the stability boundary. Aiming to obtain a more robust learning algorithm, we discuss, in the next section, the paradigm of mutual information minimization for the considered mixing model.

3. MUTUAL INFORMATION MINIMIZATION APPROACH

Firstly, let us recall the definition of mutual information, here expressed for the outputs of (3)

$$I(\mathbf{y}) = H(y_1) + H(y_2) - H(y_1, y_2), \qquad (5)$$

where $H(\cdot)$ denotes the differential entropy [9]. A common trick to avoid the estimation of $H(y_1, y_2)$ consists in expressing it as a function of the joint entropy of the mixtures. Thus, considering that the mapping provided by (3) after the convergence is invertible in the region containing the mixed signals and applying the entropy transformation law [1], one obtains:

$$I(\mathbf{y}) = H(y_1) + H(y_2) - H(\mathbf{x}) - E\{\log(|\det \mathbf{J}|)\}, \qquad (6)$$

where $\mathbf{J}$ is the Jacobian matrix associated to the separating system mapping. Since we are interested in the minimization of (6) with respect to $\mathbf{w}$ and given that $H(\mathbf{x})$ does not depend on these parameters, then we may adopt the following expression as a separating criterion[1]

$$C(\mathbf{y}) = H(y_1) + H(y_2) - E\{\log(|\det \mathbf{J}|)\}. \qquad (7)$$

In appendix 6.1, we derive the gradient of (7) with respect to $\mathbf{w}$, which is given by:

$$\frac{\partial C(\mathbf{y})}{\partial \mathbf{w}} = E\left\{ \det \mathbf{J} \left[\begin{array}{cc} \psi(y_1)(-y_2^k) & + \\ \psi(y_1)(k a_{12} y_1^{\frac{1}{k}} y_2^{k-1}) & + \end{array} \right. \right.$$
$$\left. \left. \begin{array}{cc} \psi(y_2)(\frac{1}{k} a_{21} y_1^{\frac{1}{k}-1} y_2^k) & - \quad a_{21} y_1^{\frac{1}{k}-1} y_2^{k-1} \\ \psi(y_2)(-y_1^{\frac{1}{k}}) & - \quad a_{12} y_1^{\frac{1}{k}-1} y_2^{k-1} \end{array} \right] \right\} \qquad (8)$$

where $\psi(y_i) = -\log(p(y_i))'$ denotes the marginal score function of y_i. From this expression, one obtains the following learning rule that would minimize the mutual information:

$$\mathbf{w} \leftarrow \mathbf{w} - \mu \frac{C(\mathbf{y})}{\partial \mathbf{w}}, \qquad (9)$$

where μ is the learning rate.

Expression (8) could equally be derived through the maximal likelihood framework (see [11], for instance). Usually in this approach, the true distributions of the sources are approximated by previously defined nonlinear functions [12] or by estimations obtained from the reconstructed sources provided by the separating system. It is well known that the separation task is accomplished in the linear case even when the sources distributions are roughly approximated [12]. However, this is not true for nonlinear mixing systems. Indeed,

[1] In a nonlinear context, although equivalent in a theoretical standpoint, expressions (6) and (7) may lead to different practical algorithms due to the different statistical properties of these two estimators, as demonstrated for the case of PNL models in [10].

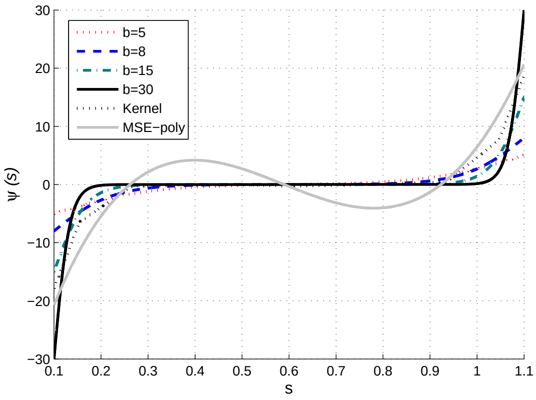


Fig. 1. Score functions.

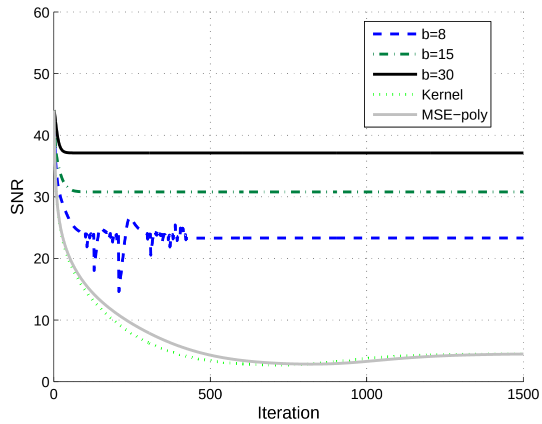


Fig. 2. Evolution of SNR for distinct score function estimation approaches.

we present, in the next section, an example that indicates that for the mixing system (2), even a satisfactory estimation of the score functions in (8) may not result in source separation.

3.1. Example of the influence of the score functions estimations on the performance of (9)

Let us consider a scenario with two sources uniformly distributed between $[0.1, 1.1]$, $k = 2$ and mixing coefficients given by $a_{12} = a_{21} = 0.5$. This set of parameters satisfies the structural stability condition given by (4). As it will be clarified in the sequel, it is useful to describe the pdf of each uniform source as

$$p(s_i) = \frac{1}{2} \lim_{b\to\infty} \tanh(b(s_i - 0.1)) - \tanh(b(s_i - 1.1)). \tag{10}$$

Therefore, it is not difficult to show that the score function of s_i is given by:

$$\psi(s_i) = \lim_{b\to\infty} \frac{b\tanh^2(b(s_i - 1.1)) - b\tanh^2(b(s_i - 0.1))}{\tanh(b(s_i - 0.1)) - \tanh(b(s_i - 1.1))}. \tag{11}$$

In fact, this limit tends to a sum of two Dirac delta functions placed at 0.1 and 1.1, respectively.

In order to verify the influence of the choice of the score function on (9), we executed this learning rule considering different values for b in (11). Also we considered the case in which the score functions are estimated from the reconstructed sources during the adaptation stage by using two distinct methods: a kernel-based approach [13] and a MSE estimation approach [3] using a fourth-order polynomial as a parametric structure. For matter of illustration, we present, in Fig. 1, the resulting functions (11) for several values of b as well as the source score function estimated with the kernel and MSE methods.

In all executions, we considered as initial conditions for $\mathbf{w}$ the theoretical solution of the problem perturbed by a small random vector. A typical evolution of a performance index (see (18) below) for each approach is depicted in Fig. 2. It is clear that there is a performance degradation as b decreases. When $b = 5$ (not depicted in the figure), for example, the adaptation goes toward a instable region with respect to the separating system. Also, it is worth special attention that, for this scenario, even the application of a satisfactory score function estimation method, such as the kernel-based one, could not be able to maintain the good initial solution.

3.2. An approach based on the notion of the mutual information differential

As discussed in the above example, a reasonable performance of the mutual information approach (or equally the maximum likelihood one) based on the minimization of (7) is depending on a very accurate knowledge of the source score functions. Evidently, this demand is a very contrasting one to the idea of blind source separation. In the sequel, aiming to overcome this requirement, we shall develop an alternative rule that directly minimizes (6). By proceeding in this way, one obtains a learning algorithm that is based on the so-called score function difference vector (see (13) below).

The cornerstone of our approach is the notion of the differential of the mutual information, proposed in [14]. In this work, it was proved that a small variation Δ of a given random vector $\mathbf{y}$ results, up to higher-order terms (expressed by $o(\Delta)$), in the following variation of the mutual information

$$I(\mathbf{y} + \Delta) - I(\mathbf{y}) = E\{\Delta^T \beta_{\mathbf{y}}(\mathbf{y})\} + o(\Delta), \tag{12}$$

where $\beta_{\mathbf{y}}(\mathbf{y})$ is the score function difference vector associated with the random variable $\mathbf{y}$. In view of this result, it can be interpreted as the gradient of the mutual information with respect to $\mathbf{y}$. In mathematical terms, the *i*-th element of $\beta_{\mathbf{y}}(\mathbf{y})$ is given by

$$\beta_{y_i}(y_i) = \left(-\frac{\partial \log p(\mathbf{y})}{\partial y_i}\right) - \left(-\frac{d \log p(y_i)}{d y_i}\right), \quad (13)$$

i.e. the difference between the *i*-th element of the joint score function of $\mathbf{y}$ and the marginal score function of y_i. It is not difficult to prove that $\mathbf{y}$ have independent components if, and only if, $\beta_{y_i}(y_i) = 0$ for every i.

The result expressed in (12) provides the guidelines for the design of a learning algorithm according to the minimum mutual information idea. The first step is to determine how a small variation of the separating system parameters, denoted by the vector $\Delta\mathbf{w} = [\Delta w_{12}\ \Delta w_{21}]^T$, affects their outputs. Then, the "gradient" of the mutual information with respect to $\mathbf{w}$ is estimated by (12), by substituting the calculated small variations of the outputs in this expression.

In order to determine the variation $\Delta\mathbf{y} = [\Delta y_1\ \Delta y_2]^T$ given a small variation $\Delta\mathbf{w}$, one may consider a linearized version of the separating system (3) with respect to $\mathbf{w}$. Considering two sources, this is expressed by

$$\begin{bmatrix} \Delta y_1 \\ \Delta y_2 \end{bmatrix} = \frac{\partial \mathbf{y}}{\partial \mathbf{w}} \Delta\mathbf{w} = \begin{bmatrix} \frac{\partial y_1}{\partial w_{12}} & \frac{\partial y_1}{\partial w_{21}} \\ \frac{\partial y_2}{\partial w_{12}} & \frac{\partial y_2}{\partial w_{21}} \end{bmatrix} \begin{bmatrix} \Delta w_{12} \\ \Delta w_{21} \end{bmatrix}. \quad (14)$$

In the appendix 6.2 the elements of $\frac{\partial \mathbf{y}}{\partial \mathbf{w}}$ are determined.

If one considers in (12) that $\Delta = \Delta\mathbf{y}$, then one readily obtains that:

$$I(\mathbf{y} + \Delta\mathbf{y}) - I(\mathbf{y}) = E\left\{\Delta\mathbf{w}^T \frac{\partial \mathbf{y}}{\partial \mathbf{w}}^T \beta_{\mathbf{y}}(\mathbf{y})\right\} + o(\Delta\mathbf{y}). \quad (15)$$

As stated above, the score function difference $\beta_{\mathbf{y}}(\mathbf{y})$ may be interpreted as the gradient of (12). Hence, considering (12) and (15), one may argue that the "gradient" of the mutual information with respect to the parameters $\mathbf{w}$ is

$$\frac{\partial I}{\partial \mathbf{w}} = E\left\{\frac{\partial \mathbf{y}}{\partial \mathbf{w}}^T \beta_{\mathbf{y}}(\mathbf{y})\right\}. \quad (16)$$

Therefore, it expected that the following learning rule minimizes the mutual information between the reconstructed sources

$$\mathbf{w} \leftarrow \mathbf{w} - \mu E\left\{\frac{\partial \mathbf{y}}{\partial \mathbf{w}}^T \beta_{\mathbf{y}}(\mathbf{y})\right\}, \quad (17)$$

where μ denotes the learning rate.

4. EXPERIMENTAL RESULTS

In order to assess the performance of the learning algorithm (17), experiments were conducted for two scenarios: $k = 2$ and $k = 3$. In both cases, the efficacy of the obtained solutions was quantified according to the following index:

$$SNR_i = 10 \log\left(\frac{E\{s_i^2\}}{E\{(s_i - y_i)^2\}}\right). \quad (18)$$

Thus, $SNR = 0.5(SNR_1 + SNR_2)$ defines a global index.

Regarding the estimation of the score function difference vector, we considered the method proposed in [13]. In short, this is a kernel-based method which differs from the classical approaches in two points: the estimation is done over a regular grid and a cardinal spline is used as kernel function. As a consequence, one obtain a much faster algorithm than the classical kernel method.

Table 1. Average SNR results over 50 experiments and standard deviation (STD).

	SNR_1	SNR_2	SNR	$STD(SNR)$
$k = 2$	44.40	40.92	42.66	7.71
$k = 3$	39.41	36.40	37.90	14.18

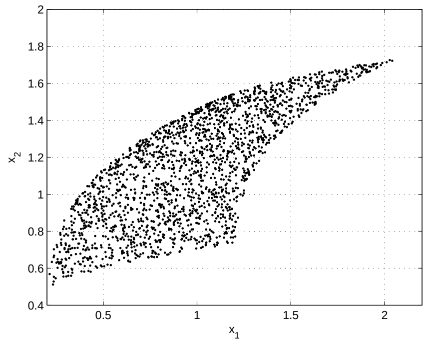
4.1. First scenario - $k = 2$

In a first scenario, we consider the separation of two sources uniformly distributed between $[0.1, 1.1]$. The mixing parameters are given by $a_{12} = 0.5$ and $a_{21} = 0.5$. Concerning the separation system, a set of 2000 samples of the mixtures was considered. The number of iterations was 2000 with a learning rate $\mu = 0.01$. The initial conditions of the dynamics (3) were chosen as $[y_1(1)\ y_2(1)]^T = [0\ 0]^T$. The results of this first case are expressed in the first row of Table 1.

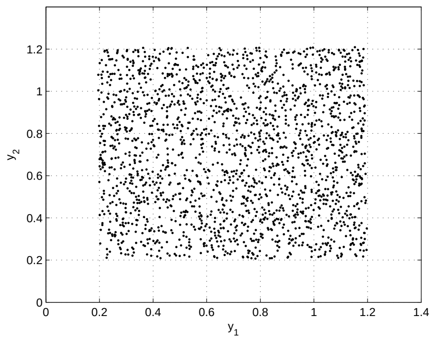
4.2. Second scenario - $k = 3$

In this situation, the separation of two sources uniformly distributed between $[0.2, 1.2]$ is conducted. The other parameters were defined in the same way as in the first scenario, except the learning step which is $\mu = 0.008$. In this situation, there are some samples of the sources that do not satisfy the stability condition (4). As a consequence, the adopted separating system will never converge to these samples which means that they cannot be retrieved. Also, these samples act as a kind of noise in the adaptation stage, in the sense that even for an ideal separating system, there are outputs that are still mixtures of the sources which may disturb the separating criteria.

Regardless the problem discussed above, our proposal does well in this scenario as can be noticed in the second row of Table 1. The high standard deviation in this case is due to the fact that the algorithm did not converge in 2 among the 50 experiments. In Fig. 3, the joint distributions of the mixtures and of the retrieved signals are depicted for a typical case. Note that the outputs of the separating system are almost uniformly distributed, which indicates that the separation task was fulfilled.



(a) Mixed signals.



(b) Retrieved signals.

Fig. 3. Second scenario - $k = 3$.

5. CONCLUSIONS

The objective of this work was to develop a mutual information framework for the adjustment of a class of recurrent separating systems. In a first moment, we showed, through an example, that the minimization of (7) leads to a learning rule that may be very dependent on an extremely accurate score function estimation stage. Then, an alternative method based on the notion of the differential of the mutual information was proposed. The obtained experimental results confirmed the efficacy of our proposal.

A first perspective of this work would be the application of the derived method on a real problem of chemical sensing. Indeed, in such an application, there are some informations, such as the positivity of the sources, that could be taken into account in the design of a separating method for the most general case of the Nikolsky-Eisenman model expressed in (1).

Also, there are two other interesting subjects for future work: $1)$ a theoretical analysis of the experimental result presented in section (3.1), and $2)$ an investigation of the cases for which our proposal did not converge. Despite the fact that just a very small number of trials did not converge, this problem may be exacerbated in the generalization of our method to scenarios with a greater number of sources.

6. APPENDIX

6.1. Derivation of (8)

Taking the derivative of (7) with respect to $\mathbf{w}$, one obtains:

$$\frac{\partial C(\mathbf{y})}{\partial \mathbf{w}} = \frac{\partial H(y_1)}{\partial \mathbf{w}} + \frac{\partial H(y_2)}{\partial \mathbf{w}} - E\left\{\frac{1}{\det \mathbf{J}} \frac{\partial \det \mathbf{J}}{\partial \mathbf{w}}\right\}. \tag{19}$$

As proved in [3]

$$\frac{\partial H(y_i)}{\partial \mathbf{w}} = E\left\{\psi(y_i)\frac{\partial y_i}{\partial \mathbf{w}}\right\}, \tag{20}$$

where $\psi(y_i)$ denotes the score function of y_1, that is to say:

$$\psi(y_i) = -\frac{d \log p(y_i)}{dy_i} = -\frac{p'(y_i)}{p(y_i)}. \tag{21}$$

By substituting (20) in (19), one obtains

$$\frac{\partial C(\mathbf{y})}{\partial \mathbf{w}} = E\left\{\psi(y_1)\frac{\partial y_1}{\partial \mathbf{w}}\right\} + E\left\{\psi(y_2)\frac{\partial y_2}{\partial \mathbf{w}}\right\} - E\left\{\frac{1}{\det(\mathbf{J})}\frac{\partial \det(\mathbf{J})}{\partial \mathbf{w}}\right\}. \tag{22}$$

In appendix 6.2, the expressions for $\frac{\partial y_i}{\partial \mathbf{w}}$ are determined

The next step is to calculate the derivative of the Jacobian determinant associated with the mapping performed by the separating system. For a given value of $\mathbf{w}$, the separating system system outputs satisfy the following equation when (3) converges

$$\begin{array}{rcl} x_1 & = & y_1 + w_{12} y_2^{k} \\ x_2 & = & y_2 + w_{21} y_1^{\frac{1}{k}} \end{array}. \tag{23}$$

In fact, this is the inverse mapping with relation to the separating system. Therefore, it is possible to calculate $\mathbf{J}$ by determining the Jacobian of this inverse mapping, which will be denoted by $\mathbf{J}'$. Obviously, there is a tacit assumption in this procedure, i.e., we assume that the mapping is invertible in the domain of the mixtures.

Considering (23), it straightforward to show that

$$\mathbf{J}' = \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{bmatrix} 1 & k w_{12} y_2^{k-1} \\ \frac{1}{k} w_{21} y_1^{\frac{1}{k}-1} & 1 \end{bmatrix}. \tag{24}$$

Given that $\mathbf{J} = \mathbf{J}'^{-1}$, then the following expression holds

$$\det(\mathbf{J}) = \frac{1}{\det(\mathbf{J}')} = \frac{1}{1 - w_{12} w_{21} y_1^{\frac{1}{k}-1} y_2^{k-1}}. \tag{25}$$

From this expression, it is straightforward to show that

$$E\left\{\frac{1}{\det(\mathbf{J})}\frac{\partial \det(\mathbf{J})}{\partial \mathbf{w}}\right\} =$$
$$E\left\{\frac{1}{1-w_{12}w_{21}y_1^{\frac{1}{k}-1}y_2^{k-1}}\left[\begin{array}{c} w_{21}y_1^{\frac{1}{k}-1}y_2^{k-1} \\ w_{12}y_1^{\frac{1}{k}-1}y_2^{k-1}\end{array}\right]\right\}. \quad (26)$$

Hence, considering (22), (26) and the results derived in the appendix 6.2, expression (8) is obtained.

6.2. Calculation of $\frac{\partial \mathbf{y}}{\partial \mathbf{w}}$

In this appendix, we are interested in the determination of

$$\frac{\partial \mathbf{y}}{\partial \mathbf{w}} = \left[\begin{array}{cc} \frac{\partial y_1}{\partial w_{12}} & \frac{\partial y_1}{\partial w_{21}} \\ \frac{\partial y_2}{\partial w_{12}} & \frac{\partial y_2}{\partial w_{21}} \end{array}\right]. \quad (27)$$

After the convergence of (3), the mapping performed by the separating system is given by (23). Therefore, the derivatives in (26) can be calculated by applying the chain rule property on (23). For instance, it is not difficult to verify from that:

$$\frac{\partial y_1}{\partial w_{12}} = -(y_2^k + w_{12}ky_2^{k-1}\frac{\partial y_2}{\partial w_{12}}). \quad (28)$$

Given that

$$\frac{\partial y_2}{\partial w_{12}} = -\frac{1}{k}w_{21}y_1^{\frac{1}{k}-1}\frac{\partial y_1}{\partial w_{12}}, \quad (29)$$

and substituting this expression in (28), one obtains:

$$\frac{\partial y_1}{\partial w_{12}} = \frac{-y_2^k}{1-w_{12}w_{21}y_1^{\frac{1}{k}-1}y_2^{k-1}}. \quad (30)$$

By conducting similar calculations, one obtains the other derivatives:

$$\frac{\partial y_2}{\partial w_{12}} = \frac{w_{21}y_1^{\frac{1}{k}-1}y_2^k}{k(1-w_{12}w_{21}y_1^{\frac{1}{k}-1}y_2^{k-1})} \quad (31)$$
$$\frac{\partial y_1}{\partial w_{21}} = \frac{kw_{12}y_1^{\frac{1}{k}}y_2^{k-1}}{1-w_{12}w_{21}y_1^{\frac{1}{k}-1}y_2^{k-1}} \quad (32)$$
$$\frac{\partial y_2}{\partial w_{21}} = \frac{-y_1^{\frac{1}{k}}}{1-w_{12}w_{21}y_1^{\frac{1}{k}-1}y_2^{k-1}} \quad (33)$$

7. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley & Sons, 2001.

[2] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

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

[4] S. Hosseini and Y. Deville, "Blind separation of linear-quadratic mixtures of real sources using a recurrent structure," in *Proceedings of the 7th International Work-conference on Artificial And Natural Neural Networks (IWANN 2003)*, 2003.

[5] S. Bermejo, C. Jutten, and J. Cabestany, "Isfet source separation: foundations and techniques," *Sensors and Actuators B*, vol. 113, pp. 222–233, 2006.

[6] G. Bedoya, C. Jutten, S. Bermejo, and J. Cabestany, "Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation," in *Sensors for Industry Conference, 2004. Proceedings the ISA/IEEE*, 2004, pp. 149–154.

[7] C. Jutten and J. Hérault, "Blind separation of sources, part i: an adapative algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[8] L. T. Duarte and C. Jutten, "Blind source separation of a class of nonlinear mixtures," Accepted for publication in ICA2007.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.

[10] S. Achard, Dinh-Tuan Pham, and C. Jutten, "Criteria based on mutual information minimization for blind source separation in post nonlinear mixtures," *Signal Processing*, vol. 85, pp. 965–974, 2005.

[11] S. Hosseini and Y. Deville, "Blind maximum likelihood separation of a linear-quadratic mixture," in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2004*, 2004.

[12] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[13] Dinh-Tuan Pham, "Fast algorithm for estimating mutual information, entropies and score functions," in *Proceedings of the Fourth International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2003*, 2003.

[14] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Differential of the mutual information," *IEEE Signal Processing Letters*, vol. 11, no. 1, pp. 48–51, January 2004.