

A NONLINEAR SOURCE SEPARATION APPROACH FOR THE NICOLSKY-EISENMAN MODEL

Leonardo Tomazeli Duarte and Christian Jutten

GIPSA-lab, INPG-CNRS, 46 Avenue Félix Viallet, Grenoble, France
leonardo.duarte@gipsa-lab.inpg.fr, christian.jutten@gipsa-lab.inpg.fr

ABSTRACT

In previous works [7, 8], we proposed source separation methods for a simplified version of the Nicolsky-Eisenman (NE) model, which is related to a chemical sensing application. In the present paper, we provide a method able to deal with the complete NE model. Basically, such a model can be seen as a composition of a non-diagonal nonlinear transformation followed by a diagonal nonlinear transformation, i.e. a set of component-wise functions. The basic idea behind the developed technique is to estimate the parameters of these two stages in a separate fashion by using a prior knowledge of the sources, namely the fact that one of the sources is constant during a certain period of time. Simulations attest the viability of the proposed technique.

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. There is a great number of methods for the situation in which the mixing process is linear. However, it seems that in some applications the linear approximation is not suitable and, as a consequence, it becomes necessary to design source separation methods that take the nonlinearity of the model into account. The main problem in the nonlinear case is that the ubiquitous tool for performing source separation, the independent component analysis (ICA), does not work in a general situation [11, 12]. In other words, there are cases in which the recovery of the statistical independence, which is the very essence of ICA, does not guarantee the separation of the sources when the mixture model is nonlinear. In view of this limitation, a more reasonable approach is to consider constrained mixing systems as, for example, post-nonlinear (PNL) mixtures [12] and linear-quadratic mixtures [10].

In recent papers [3, 8], the problem of BSS in a particular class of nonlinear systems related to a chemical sensing application was investigated. For instance, we developed in [7, 8] source separation methods for estimating the concentrations of ions in a solution when the valences of these ions are different. In our first works, we considered a simplified version of the mixing model that is pertinent only in the situations where the sensors present a Nernstian response [13]. In the present work, by relying on an additional assumption on the sources, we extend our solution to the complete model, which permits us to work in more realistic situations. Our work will be presented as follows: in Section 2 we introduce the problem. After that, in Section 3, we expose the basics of

our method. In Section 4, simulations are carried out in order to illustrate the viability of the proposal. Finally, in Section 5, we state our conclusions and perspectives.

2. PROBLEM STATEMENT

In analytical chemistry, a relevant problem is to estimate the evolution of the concentrations of several ions in a solution. An inexpensive and practical approach is based on the use of potentiometric sensors, such as glass-electrodes and ion-sensitive field-effect transistors (ISFET). Basically, the sensing mechanism in this sort of sensor is due to the dependence of the potential difference between two electrodes on the concentration of a target ion [9]. Unfortunately, there is a pronounced lack of selectivity behind this principle, i.e. the generated potential difference in the sensor can also depend on other interfering ions present in the solution.

A very common approach to deal with the interference problem is to perform a calibration of the chemical sensor by using samples from a known database [9]. However, there are two major difficulties in such an approach: 1) the acquisition of training samples may be time-demanding and 2) this calibration procedure must be performed from time to time due to the sensors drift. In this context, an approach founded on array of chemical sensors and on BSS methods emerges as an attractive alternative, since, in this case, the estimation of the evolution of the concentrations can be done by using only the sensors responses, that is, without conducting supervised learning with a known data set.

Although very general, BSS techniques usually require a parametric description of the mixing model. In our case, this information can be found on the classical model of Nicolsky-Eisenman (NE) [9], which provides a simple and yet adequate description of the potentiometric sensors. According to this model, the response of the i -th sensor is given by:

$$x_i(t) = c_i + d_i \log \left(s_i(t) + \sum_{j, j \neq i} a_{ij} s_j(t)^{\frac{z_i}{z_j}} \right), \quad (1)$$

where t corresponds to the time index; $s_i(t)$ and $s_j(t)$ denote the concentration of the target ion and of the concentration of the j -th interfering ion, respectively. z_i and z_j denote the valence of the ions i and j , respectively. The selective coefficients a_{ij} model the interference process; c_i and d_i are constants that depend on some physical parameters. Note that when the ions have the same valence, then the model (1) can be seen as a particular case of the class of PNL systems, as described in [3].

In this work, we are interested in the situation where the valences are different ($z_i \neq z_j$). In view of the toughness of the resulting model, we investigate only the case with two sources. Also, the parameters c_i may be not considered, since

Leonardo Tomazeli Duarte would like to thank the National Council for Scientific and Technological Development (CNPq-Brazil) for funding his PhD research.

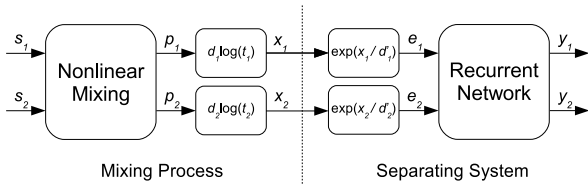


Figure 1: The Nicolsky-Eisenman model.

they only introduce an ambiguity with respect to the sources mean value. Therefore, the obtained mixing model can be described as:

$$\begin{aligned} x_1(t) &= d_1 \log(p_1(t)) = d_1 \log(s_1(t) + a_{12}s_2(t)^k) \\ x_2(t) &= d_2 \log(p_2(t)) = d_2 \log(s_2(t) + a_{21}s_1(t)^{\frac{1}{k}}) \end{aligned}, \quad (2)$$

where $k = z_1/z_2$. As illustrated in Fig. 1, this mixing model is composed of two stages, being the first one a particular nonlinear mapping that depends on the valences of the target ions and the second one a pair of nonlinear component-wise functions parametrized in d_i . As it will be clarified in the sequel, the proposed method treats these two stages in a separate way.

3. PROPOSED METHOD

A natural strategy to perform BSS on the model (2) is to adopt the two-stage separating system shown in Fig. 1. The first stage is composed of exponential component-wise functions $\exp(x_i/d_i^*)$, whereas the second one should provide a non-diagonal mapping able to counterbalance the effects introduced by the first stage of the mixing system. In [7, 8], we investigated source separation methods for this first stage and we assumed that the component-wise functions were known in advance. In fact, this situation is realistic only when the sensors have a Nernstian response, given that in such a case the parameters d_i are known in advance.

In the sequel, we will present a brief review of our previous works. Then, we will detail the main contribution of this paper which is related to the estimation of the component-wise functions, i.e., of the parameters d_i . With this new step, it becomes possible to develop a complete source separation method for the model described in (2). During our development, the following assumptions are considered:

Assumption 1 *The sources are statistically independent;*

Assumption 2 *The sources are positive and bounded, i.e., $s_i(t) \in [S_i^{\min}, S_i^{\max}]$, where $S_i^{\max} > S_i^{\min} > 0$;*

Assumption 3 *The system expressed in (2) is invertible in the region given by $[S_1^{\min}, S_1^{\max}] \times [S_2^{\min}, S_2^{\max}]$;*

Assumption 4 *k is known and takes only positive integer values;*

In the context of the ion sensing application, assumption 1 is equivalent to consider that there is no chemical reaction between the ions. Concerning the assumptions 2 and 3, our separation methods would work in a predefined range of concentrations, as is usual in commercial sensors. Furthermore, it is quite natural to consider positive sources, since

they represent concentrations. Finally, assumption 4 state a knowledge of the ions' valences. Actually, this information is available since the main interfering ions, and so their valences, are known in advance.

3.1 Review of our solution for a simplified version of (2)

When the component-wise functions in (2) are known in advance, the mixing model becomes

$$\begin{aligned} p_1(t) &= s_1(t) + a_{12}s_2(t)^k \\ p_2(t) &= s_2(t) + a_{21}s_1(t)^{\frac{1}{k}} \end{aligned}. \quad (3)$$

In order to retrieve s_i , we adopted [7] the following recurrent network as separating system:

$$\begin{aligned} y_1(n+1) &= p_1(t) - w_{12}y_2(n)^k \\ y_2(n+1) &= p_2(t) - w_{21}y_1(n)^{\frac{1}{k}} \end{aligned}, \quad (4)$$

where the vectors $\mathbf{w} = [w_{12} \ w_{21}]^T$ and $\mathbf{y}(n) = [y_1(n) \ y_2(n)]^T$ denote the parameters to be adjusted and the system outputs at time n , respectively. For a given sample of the mixtures $[p_1(t) \ p_2(t)]^T$, and for a given value \mathbf{w} , the system outputs are obtained after the convergence of the dynamics (4).

In order to adjust the parameters \mathbf{w} in (4), we proposed [8] the following learning rule that minimizes the mutual information of the vector \mathbf{y} :

$$\mathbf{w} \leftarrow \mathbf{w} - \mu E \left\{ \frac{\partial \mathbf{y}^T}{\partial \mathbf{w}} \beta_{\mathbf{y}}(\mathbf{y}) \right\}, \quad (5)$$

where μ denotes the learning rate, $\beta_{\mathbf{y}}(\mathbf{y})$ is the score function difference vector associated with the random variable \mathbf{y} and $\frac{\partial \mathbf{y}}{\partial \mathbf{w}}$ is the Jacobian of (4) with respect to \mathbf{w} .

3.2 An approach for inverting the component-wise functions

Now, let us turn our attention to the complete version of the NE model. The following parametric model can be considered in the first stage of the separating system

$$\begin{aligned} e_1(t) &= \exp\left(\frac{x_1(t)}{d_1^*}\right) \\ e_2(t) &= \exp\left(\frac{x_2(t)}{d_2^*}\right) \end{aligned}, \quad (6)$$

where d_i^* are the parameters to be adjusted.

A first natural approach to adapt the separating system would be to develop an ICA algorithm to find the parameters $\{d_1^*, d_2^*, w_{12}, w_{21}\}$. In other words, we could seek a set of parameters that optimizes a functional associated with the statistical independence, such as the mutual information. However, it would be tough to develop a gradient rule in this case given that the separating system is composed of a nonlinear dynamic system and, therefore, there would be a risk of instability during the learning phase. Besides this practical problem, there is a crucial theoretical point that should be addressed: is the considered mixing model separable, i.e., is the recovery of the statistical independence enough to assure source separation? We do not have such an answer, but recent studies [12] suggest that for nonlinear transformations like the one dealt with here (2), it is necessary to consider prior information other than independence to achieve source separation.

In view of the aforementioned problems, we believe that it is more reasonable to consider a two-step approach, in which the component-wise functions are firstly estimated and, then, the recurrent system is trained. The price to be paid is that we need the following additional assumption:

Assumption 5 *There is, at least, a period of time where one, and only one, of the sources has zero-variance.*

In view of assumption 5, if we consider, for instance, that $s_1(t)$ takes a constant value S_1 (where $S_1 \in [S_1^{\min}, S_1^{\max}]$) in a given time window, then, according to (2), we have

$$\begin{aligned} p_1(t) &= S_1 + a_{12}s_2(t)^k \\ p_2(t) &= s_2(t) + a_{21}S_1^{\frac{1}{k}}. \end{aligned} \quad (7)$$

From this expression, it is not difficult to see that

$$p_1(t) = S_1 + a_{12}(p_2(t) - a_{21}S_1^{\frac{1}{k}})^k. \quad (8)$$

Hence, under such an assumption, (8) is a polynomial of order k in the (p_1, p_2) plane. Our idea is based on this fact, in the sense that this polynomial function is lost after the application of the component-wise functions and, thus, we may invert the log functions by searching a set $\{d_1^*, d_2^*\}$ that restores a polynomial in the (e_1, e_2) plane. In the sequel, we shall detail this idea.

Firstly, let us describe the mapping between the (p_1, p_2) and (e_1, e_2) planes. By considering (2) and (7), one can rewrite (6) as:

$$\begin{aligned} e_1(t) &= \exp\left(\frac{d_1 \log(p_1(t))}{d_1^*}\right) = p_1(t)^{\frac{d_1}{d_1^*}} \\ e_2(t) &= \exp\left(\frac{d_2 \log(p_2(t))}{d_2^*}\right) = p_2(t)^{\frac{d_2}{d_2^*}}. \end{aligned} \quad (9)$$

The next step of our study is to find the relation between the data in the (e_1, e_2) plane. As stated by (8), there is a polynomial relation in the (p_1, p_2) plane, which can be expressed by

$$p_1(t) = \sum_{i=0}^k \varphi_i p_2(t)^i, \quad (10)$$

where the coefficients φ_i can be determined by the binomial expansion of (8). After a straightforward development considering (9) and (10), the following expression is obtained:

$$e_1(t) = \left[\sum_{i=0}^k \varphi_i e_2(t)^{\frac{d_2^*}{d_2} i} \right]^{\frac{d_1}{d_1^*}}. \quad (11)$$

As stated above, our initial idea to find d_1^* and d_2^* is based on the recovery of a polynomial curve of order k in the (e_1, e_2) plane. In order to verify the viability of our approach, we need to investigate the following question: for what values of d_1^* and d_2^* the function (11) is a polynomial of order k ? At a first sight, it is clear that when the optimum solution ($d_1^* = d_1, d_2^* = d_2$) is achieved, then that expression results in a polynomial curve in the (e_1, e_2) plane.

However, there is a particular situation where a polynomial curve is obtained although the mapping (9) is still nonlinear. In fact, when S_1 is null, then we can see from (8) that

all the coefficients except φ_k in (11) are null. In this situation, the following solution ($d_1^* = Dd_1, d_2^* = Dd_2$), where D is a constant, also gives a polynomial of order k although does not correspond to our desired solution. Indeed, one of the reasons behind assumption 2 is exactly to avoid this situation by considering only positive sources.

To implement our idea, we must define a way to check if a set of points in the (e_1, e_2) plane corresponds to a polynomial of order k . This can be done by defining a cost function as the mean square of the residuals resulting from the regression of the set of sample $\{(e_1(t), e_2(t))\}_{t=1}^N$ (being N the number of samples) through a polynomial of order k . In mathematical terms, this cost function can be expressed as

$$\min_{d_1^*, d_2^*} \sum_t \left(e_1(t) - \sum_{i=0}^k \alpha_i (e_2(t))^i \right)^2, \quad (12)$$

where α_i correspond to the i -th regression coefficient.

In order to gain more insight, let us substitute (11) in (12), which gives

$$\min_{d_1^*, d_2^*} \sum_t \left(\left[\sum_{i=0}^k \varphi_i (e_2(t))^{\frac{d_2^*}{d_2} i} \right]^{\frac{d_1}{d_1^*}} - \sum_{i=0}^k \alpha_i (e_2(t))^i \right)^2. \quad (13)$$

One may note that this expression is a nonlinear function with respect to the parameters $\{d_1^*, d_2^*\}$. Moreover, according to equation (11), for a given sample at time t there is an underlying relation between $e_1(t)$ and $e_2(t)$, which in turn makes the regression coefficients α_i nonlinearly dependent on the parameters to be optimized. As a consequence, it becomes difficult to obtain the derivatives of this function and, therefore, to develop a gradient-based optimization method.

An alternative approach to optimize (12) can be found on the so-called evolutionary techniques. Briefly, this class of algorithms performs a searching procedure that is based on the notion of population, i.e., it deals with a set of possible solutions (individuals). At each iteration, some perturbation (mutation and recombination, for instance) is introduced on the population and a group of individuals is selected to continue in the population of the next iteration (usually, the selected individuals are the ones with higher fitness, which is calculated through the cost function). The major benefit brought by an evolutionary method to our problem is that no information about the derivatives is needed, since the selection stage is based only on the evaluation of the cost function, which, for our problem, is a straightforward task.

3.2.1 Detection of zero-variance periods

The idea described in the last section works under the assumption that there is a time window in which one of the sources does not vary. Evidently, if a blind scenario is envisaged, then one should be able to detect this "silent period". A possible way to perform this task is to consider the problem from a geometric standpoint. Given that the mixing model is invertible and the sources are supposed bounded, the borders of the distribution in the (x_1, x_2) plane corresponds to the situation in which one of the source is constant. Therefore, at least in an ideal situation, we could detect the silent periods by estimating the borders of the distribution of the mixtures, in the same way as performed in [2]. Note that this

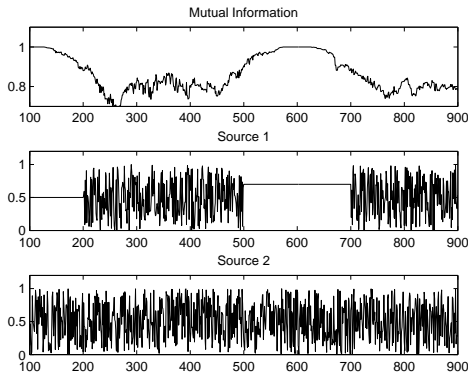


Figure 2: Mutual information between the mixtures (a) and corresponding sources (b) (c).

strategy works even when the assumption 5 is not met. Unfortunately, this procedure is difficult to implement since it demands a very accurate estimate of the borders, which may be difficult to achieve when the number of samples is small.

A second approach to search the silent periods is based on the fact that when one of the sources is constant, each sensor response corresponds to a deterministic function of the same random variable, i.e., $x_1 = g_1(s_2)$ and $x_2 = g_2(s_2)$. Therefore, we can argue that such situation is the one with maximum (nonlinear) correlation between the sensors and, given that, we may try to identify the silent periods by searching time windows for which a measure of correlation is maximized. This idea has already been developed for linear source separation [6] and, in this case, the silent periods are found by observing the second-order correlation measure. In our case, we deal with a nonlinear model and, as consequence, a measure able to detect nonlinear dependences must be employed. With this purpose in mind, we consider the mutual information between the mixing signals.

The mutual information of two continuous random variables lies on the interval $0 \leq I(x_1, x_2) < +\infty$, being zero when x_1 and x_2 are statistically independent, and tending to infinity when there is a deterministic relation between these variables. Therefore, we can find the silent periods by looking at the time windows for which the mutual information is maximized. In fact, it seems more practical to maximize a normalized version of the mutual information defined as $\zeta(x_1, x_2) = \sqrt{1 - \exp(-2I(x_1, x_2))}$, given that its maximum value is one and occurs when there is a deterministic relationship between x_1 and x_2 .

To illustrate the idea of the last paragraph, we present in Fig. 2 the evolution of the normalized mutual information between the sensors response (estimated through a time window of length 151) and the respectively sources. Note that the maximum of the mutual information occurs exactly for time windows containing a constant source.

3.3 Description of the complete algorithm

We can summarize the complete separation algorithm in Tab.1. Concerning the first step, we adopted the mutual information estimator proposed in [4]. As already discussed, the optimization of (12) in the second stage can be carried out through evolutionary methods. In this work, we chose

Table 1: Algorithm

1. Detection of silent periods
 - Estimate the mutual information between the mixtures x_1 and x_2 for a moving-time window.
 - Select the time window in which the mutual information is maximum.
2. Estimation of the component-wise functions.
 - For the selected time window, minimize expression (12) with respect to d_1^* and d_2^* .
3. Training the recurrent network
 - Determine the parameters w_{ij} of (4) through the algorithm (5). The inputs of the recurrent network are e_i .

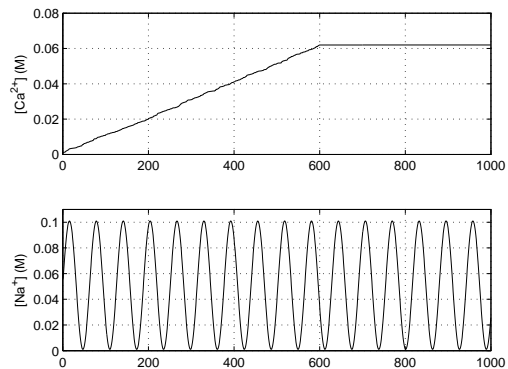


Figure 3: Sources.

the opt-aiNet algorithm. This evolutionary method has been proven to be efficient in signal processing applications (see [1], for instance). In addition to its robustness to local minima, the opt-aiNet only needs zero-order information, which, as discussed before, is a very interesting feature for our problem. The technical details of this method can be found in [5].

4. SIMULATION RESULTS

To assess the performance of the algorithm described in Tab. 1, we simulated the problem of detecting the ions Ca^{2+} and Na^+ through an array of two sensors (each one has a different ion as target). For that, we consider the parameters $a_{12} = 0.79$ and $a_{21} = 0.40$, which were taken from [13]. Also, we have assumed that both sensors have a perfect Nernstian response [9], i.e., $d_1 = 0.0129$ and $d_2 = 0.0258$. The efficacy of the obtained solution for each source was quantified according to the following index:

$$SIR_i = 10 \log \left(\frac{E\{s_i^2\}}{E\{(s_i - y_i)^2\}} \right). \quad (14)$$

Thus, $SIR = 0.5(SIR_1 + SIR_2)$ defines a global index.

Regarding the parameters of the algorithm, a set of 1000 samples was considered. The detection of the silent periods is performed by estimating the mutual information for a window of a length of 151 samples. Actually, it is difficult to achieve a reliable estimation of the mutual information with

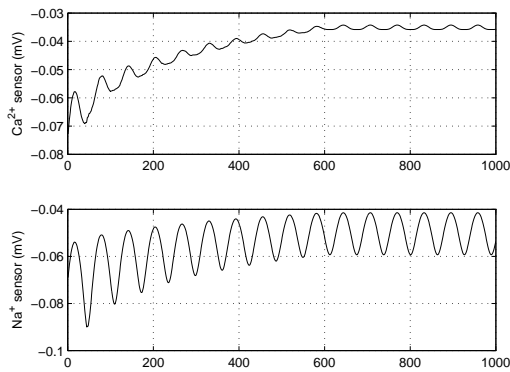


Figure 4: Mixtures.

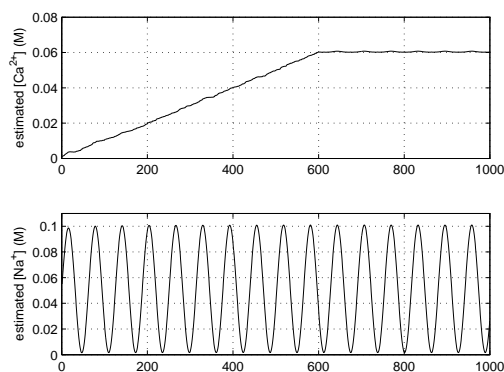


Figure 5: Retrieved sources.

a smaller number of samples. Finally, concerning the training of the recurrent network, the number of iterations was 1500 and the learning rate was $\mu = 0.02$.

Under the described scenario and with the two sources shown in Fig. 3, our method has achieved the following performance $SIR_1 = 40.52$, $SIR_2 = 38.27$ and $SIR = 39.39$. Such a result attests that our proposal does well in this case, as can be confirmed by looking at the mixing signals in Fig. 4 and at the retrieved sources in Fig. 5.

5. CONCLUSIONS AND PERSPECTIVES

In this work, we have proposed a source separation method for the NE model, which is related to a chemical sensing application. Under the assumption that one of the sources does not vary during a period of time, it became possible to estimate the component-wise functions, which correspond to the second stage of the NE model. By joining this stage to our previous solutions, we could define a complete separation framework. In order to verify the efficacy of our proposal, we conducted some experiments considering a set of parameters taken from the literature. The obtained solutions highlight that our method is a promising one to the application in mind.

There are several perspectives for this work. A first one is to investigate the extension of our method for the situation with more than two sources. Also, we intent to conduct

some real experiments to investigate the applicability of our method in a real problem. By proceeding this way, we will be able to address several questions in more details as, for instance, the study of the noisy case, starting with the problem of modeling the noise in this sort of application.

REFERENCES

- [1] R. R. F. Attux, M. Loiola, R. Suyama, L. N. de Castro, F.J. Von Zuben, and J. M. T. Romano. Blind search for optimal wiener equalizers using an artificial immune network model. *EURASIP Journal on Applied Signal Processing*, 2003:740–747, 2003.
- [2] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. A geometric approach for separating post non-linear mixtures. In *Proc. EUSIPCO 2002, Toulouse, France*, pages 11–14, 2002.
- [3] G. Bedoya, C. Jutten, S. Bermejo, and J. Cabestany. Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation. In *Proc. of IEEE SICON 2004, New Orleans, USA*, pages 149–154, 2004.
- [4] G.A. Darbellay and I. Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, 45(4):1315–1321, May 1999.
- [5] L. N. de Castro and F. J. Von Zuben. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6:239–251, 2002.
- [6] Y. Deville and M. Puigt. Temporal and time-frequency correlation-based blind source separation methods. part i: Determined and underdetermined linear instantaneous mixtures. *Signal Processing*, 87:374–407, 2007.
- [7] L. T. Duarte and C. Jutten. Blind source separation of a class of nonlinear mixtures. In *Proc. of ICA 2007, London, UK*, pages 41–48, 2007.
- [8] L. T. Duarte and C. Jutten. A mutual information minimization approach for a class of nonlinear recurrent separating systems. In *Proc. of the IEEE MLSP 2007, Thessaloniki, Greece*, pages 122 – 127, 2007.
- [9] P. Fabry and J. Fouletier, editors. *Microcapteurs chimiques et biologiques*. Lavoisier, 2003.
- [10] S. Hosseini and Y. Deville. Blind separation of linear-quadratic mixtures of real sources using a recurrent structure. In *Proc. of the IWANN 2003, Menorca, Spain*, pages 289–296, 2003.
- [11] A. Hyvärinen and P. Pajunen. Nonlinear independent component analysis: existence and uniqueness results. *Neural Networks*, 12:429–439, 1999.
- [12] C. Jutten and J. Karhunen. Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. *International Journal of Neural Systems*, 14:267–292, 2004.
- [13] Y. Umezawa, P. Bühlmann, K. Umezawa, K. Tohda, and S. Amemiya. Potentiometric selectivity coefficients of ion-selective electrodes. *Pure and Applied Chemistry*, 72:1851–2082, 2000.