Blind Compensation of Nonlinear Distortions: Application to Source Separation of Post-Nonlinear Mixtures

Leonardo Tomazeli Duarte, Ricardo Suyama, Bertrand Rivet, Romis Attux, João M. T. Romano, Christian Jutten

Blind Compensation of Nonlinear Distortions : Application to Source Separation of Post-Nonlinear Mixtures

Leonardo T. Duarte, *Member, IEEE,* Ricardo Suyama, Bertrand Rivet, Romis Attux, João M. T. Romano, *Senior Member, IEEE* and Christian Jutten, *Fellow, IEEE*

***Abstract*—In this paper, we address the problem of blind compensation of nonlinear distortions. Our approach relies on the assumption that the input signals are bandlimited. We then make use of the classical result that the output of a nonlinear memoryless system has a wider spectrum than the one of the input signal. However, different from previous works, our approach does not assume the knowledge of the input signal's bandwidth. The proposed approached is considered in the development of a two-stage method for blind source separation (BSS) in post-nonlinear (PNL) models. Indeed, once the functions present in the nonlinear stage of a PNL model are compensated, one can apply the well-established linear BSS algorithms to complete the task of separating the sources. Numerical experiments performed in different scenarios attest the viability of the proposal. Moreover, the proposed method is tested in a real situation where the data are acquired by smart chemical sensor arrays.**

***Index Terms*—onlinear memoryless systemsonlinear memoryless systemsN, blind source separation, post-nonlinear model, bandlimited signals, smart chemical sensor arrays.**

I. Introduction

In many applications, the observed signal corresponds to a nonlinearly distorted version of the desired one. For instance, this problem arises in satellite communications [2] due to the presence of amplifier stages and in chemical sensors [3] due to the nature of the transducer mechanism. Usually, the compensation of these nonlinear distortions is conducted by considering a supervised framework, in which one has access either to a set of training samples [4] or to a complete characterization of the nonlinear function [5].

When no information about the input signal and the nonlinear distortion are available, a blind (or unsupervised) framework must be considered. The main difficulty in this case is that the resulting problem is ill-posed and, thus, cannot be solved without taking into account a minimum of prior information. In this context, the seminal work of Landau and Miranker [6] on sampling of nonlinearly distorted signals eventually provided a very interesting strategy to perform blind compensation of nonlinear distortions. The prior information assumed in their approach is that the signal of interest is bandlimited. By exploiting the fact that a nonlinear function tend to spread the spectrum support of the input signal, [6] suggested that the original signal could be recovered by restoring a bandlimited signal.

The idea initially introduced in [6] has been exploited in the context of single-input single-output (SISO) models by several works [7], [8], [9]. So far, however, all these works assume that the input signal's bandwidth is known in advance. This assumption may be realistic, for instance, in applications in telecommunications and audio processing. On the other hand, in applications such as the one we are interested in this work — the development of smart chemical sensor arrays [3] — although the signals of interest can be approximately modeled as bandlimited ones, no information about these signals' bandwidth are available.

In order to obtain a more general framework of blind compensation of nonlinear functions, we propose in this work a novel strategy to restore bandlimited signals from nonlinearly distorted observations that is able to operate without the input signal's bandwidth knowledge. Although our proposal can be applied in any problem involving compensation of nonlinear functions, the main motivation of our work is to develop a novel blind source separation (BSS) [10], [11] approach for the case in which the mixing model can be modeled as a post-nonlinear (PNL) system [12]. Indeed, since this kind of model comprises a linear mixing stage followed by a set of component-wise nonlinear functions, if one is able to blindly compensate these nonlinear functions, then one ends up with a linear BSS problem, for which there are very efficient techniques [10], [13]. In other words, compensation of nonlinear functions is at the very core of source separation in PNL mixtures.

Concerning the paper's organization, we introduce in Section II the problem of blind compensation of nonlinear memoryless systems and define the notation that will be considered throughout the work. We then present, in Section III, the proposed approach to deal with nonlinear distortions in an unsupervised fashion. In Section IV, we briefly introduce the problem of blind source separation in PNL mixing models

L.T. Duarte is with the School of Applied Sciences (FCA), University of Campinas (UNICAMP), Rua Pedro Zaccaria 1300, CEP: 13484-350, Limeira, SP, Brazil (email:leonardo.duarte@fca.unicamp.br)

R. Attux and J.M.T. Romano are with the School of Electrical and Computer Engineering (FEEC), University of Campinas (UNICAMP), Av. Albert Einstein, 400, CEP: 13083-852, Campinas, SP, Brazil (email:attux@dca.fee.unicamp.br, romano@dmo.fee.unicamp.br)

R. Suyama is with the CECS, Federal University of ABC (UFABC), Av. dos Estados, 5001, CEP: 09210-170, Santo André, SP, Brazil (email:ricardo.suyama@ufabc.edu.br)

B. Rivet and C. Jutten are with the GIPSA-lab (UMR CNRS 5216), Institut Polytechnique de Grenoble, BP 46, F-38402 Grenoble Cedex, Grenoble, France (e-mail:{bertrand.rivet,christian.jutten}@gipsa-lab.grenoble-inp.fr). C. Jutten is also with the Institut Universitaire de France.

This paper was presented in part at the IEEE Workshop of Machine Learning for Signal Processing (2009) and was published [1] in its proceedings.

and show how the method proposed in Section III can be considered to provide a sound PNL source separation technique. Section V presents simulations with the aim of assessing the validity of the proposed method. Experiments with both synthetic data and actual data acquired by smart chemical sensor arrays are presented. Finally, in Section VI, conclusions and perspectives for future works are presented.

II. Blind compensation of nonlinear functions

In the problem of blind compensation of a nonlinear distortion, which is illustrated in Figure 1, the observed signal[1], $x_i(t)$, is modeled as a nonlinear function of the input signal, $z_i(t)$, as follows

$$x_i(t) = f_i(z_i(t)), \tag{1}$$

where $f_i(\cdot)$ denotes the distorting function. The goal here is thus to provide a good estimation of the input signal $z_i(t)$ by only considering the observations $x_i(t)$ and a few assumptions on the desired signal and on the distorting function. This task can be accomplished by adjusting a nonlinear compensating function $g_i(\cdot)$ so that the estimated signal, given by

$$q_i(t) = g_i(x_i(t)), \tag{2}$$

be as close as possible to the input signal $z_i(t)$. In this work, we assume that both $f_i(\cdot)$ and $g_i(\cdot)$ are monotonic functions.

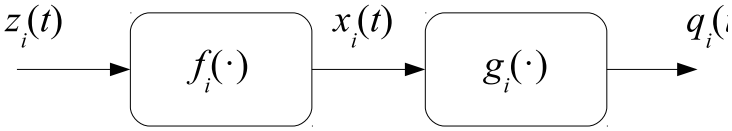


Fig. 1. Overview of the problem of blind compensation of a nonlinear function.

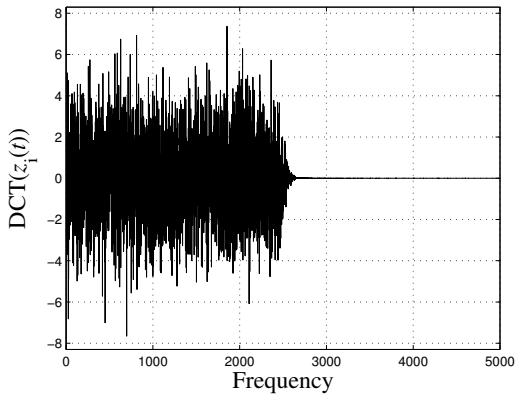
The problem of blind compensation of nonlinear distortion is ill-posed, and, thus, can be solved only when additional prior information are available. In the following, we discuss how an unsupervised framework can be built by considering the information that the input signal is bandlimited.

III. Strategies to deal with nonlinear distortions

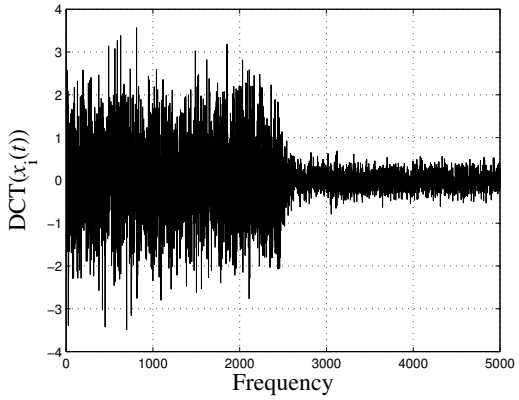
A. Spectral spreading caused by nonlinear functions

In certain applications, such as chemical sensing, the signals of interest present a slow temporal variation, thus presenting a spectral content concentrated on low-frequency bands. Motivated by this observation, we shall assume that $z_i(t)$ is a bandlimited signal with unknown maximum frequency given by B_{z_i}. Due to the action of the nonlinear functions $f_i(\cdot)$, the spectrum content of the output $x_i(t) = f(z_i(t))$ tends to be wider than the spectra of the original signals $z_i(t)$ (cf. Appendix A). It is thus expected that the maximum frequency of $X_i(\omega)$, the Fourier transform of $x_i(t)$, be larger than B_{z_i}. This phenomenon, which can also be observed is discrete signals, is illustrated in Figure 2, which shows the effects of a nonlinear memoryless system in the discrete cosine transform (DCT) representation.

[1]The index i is used here to keep a coherent notation with the one used for describing PNL systems in Section IV.



(a) DCT of $z_i(t)$.



(b) DCT of $x_i(t) = \tanh(z_i(t))$.

Fig. 2. Spectral spreading caused by a nonlinear distortion: the DCTs of a bandlimited signal $z_i(t)$ and of the distorted signal $x_i(t) = \tanh(z_i(t))$.

The spectral spreading phenomenon can be used to formulate a criterion for adapting the nonlinear function $g_i(\cdot)$. Indeed, this function can be adjusted so that it provides a signal $q_i(t) = g_i(x_i(t))$ that is bandlimited to the original bandwidth of the input signal $z_i(t)$ [7]. This condition is satisfied when $q_i(t) = g_i(f_i(z_i(t))) = \alpha z_i(t) + \beta$, where $\alpha, \beta \in \mathbb{R}$, that is, when the composition of the two functions is an affine function, which is exactly the desired solution. The practical implementation of the idea of restoring a bandlimited signal is discussed in the sequel.

B. Semi-Blind Cost Function

Henceforth, the function used for inverting $f_i(\cdot)$ will be denoted by $g_i(\cdot, \mathbf{b}_i)$, being parameterized by the vector $\mathbf{b}_i$. According to the out-of-band minimization idea mentioned before, $g_i(\cdot, \mathbf{b}_i)$ can be adjusted so that the energy of $q_i(t)$ beyond the frequency B_{z_i} be as low as possible. In mathematical terms, this can be expressed by the following optimization problem

$$\hat{\mathbf{b}}_i = \arg\min_{\mathbf{b}_i} J_1(\mathbf{b}_i), \qquad \text{with} \quad J_1(\mathbf{b}_i) = \frac{E_{q_i}^{\{f>B_{z_i}\}}}{E_{q_i}}. \tag{3}$$

In the notation adopted in this paper, E_{q_i} denotes the total energy of $q_i(t)$ and $E_{q_i}^{\{f>B_{z_i}\}}$ the energy associated with the

frequency components beyond B_{z_i}:

$$E_{q_i}^{\{f>B_{z_i}\}} = \int_{B_{z_i}}^{1} \gamma_{q_i}(f)df$$

where $\gamma_{q_i}(f)$ is the power spectral density[2] of $q_i(t)$. Note that the normalization of (3) by E_{q_i} is necessary to avoid a trivial solution in which the signal $q_i(t)$ has null energy.

The formulation expressed in (3) was considered, for instance, in [7]. However, the main drawback of (3) is that it works with the strong assumption that B_{z_i} is known, which is quite unrealistic in an unsupervised context. Yet, it is possible to define an extension of $J_1(\mathbf{b}_i)$ to a scenario with unknown B_{z_i}, by replacing B_{z_i} in (3) with a value $\hat{B}_{z_i}$ that satisfies the condition $\hat{B}_{z_i} > B_{z_i}$. To be sure that such a requirement is satisfied, one can define $\hat{B}_{z_i(t)}$ close to one. In this case, the spectral spreading in the interval $[\hat{B}_{z_i(t)}, 1]$, where $\hat{B}_{z_i(t)} > B_{z_i(t)}$ is minimized. Evidently, since this is only a necessary condition, there is no guarantee that such a procedure will lead to a proper compensation of $f_i(\cdot)$, even though this procedure usually performs well in noiseless scenarios.

When the observed signal is corrupted by noise or when the assumption of bandlimited signal is only approximated, the strategy described in the last paragraph may become rather suboptimal. For example, suppose that $\hat{B}_{z_i}$ definitively overestimates the actual bandwidth B_{z_i} of $z_i(t)$ (i.e. $\hat{B}_{z_i} >> B_{z_i}$). In this case, criterion (3) will consider only a few high-frequency components, discarding all the information available in the band $[B_{z_i}, \hat{B}_{z_i}]$. As a consequence, the resulting estimator in this case will be much less robust to noise than the estimator considering the actual value B_{z_i}. This is particularly undesirable in the present problem, given that even a low-power noise can become significant after a nonlinear distortion.

C. Blind Cost Function

The limitations associated with the blind extension of (3) can be overcome if, in addition to the vector of parameters $\mathbf{b}_i$, B_{z_i} is also seen as an unknown parameter. Having this observation in mind, the proposed approach is thus based on the following optimization problem:

$$(\hat{\mathbf{b}}_i, \widehat{\hat{B}_{z_i}}) = \arg \min_{\mathbf{b}_i, \hat{B}_{z_i}} J_2(\mathbf{b}_i, \hat{B}_{z_i}), \tag{4}$$

with

$$J_2(\mathbf{b}_i, \hat{B}_{z_i}) = \frac{E_{q_i}^{\{f>\hat{B}_{z_i}\}}}{E_{q_i}^{\{f>\hat{B}_{z_i}-\phi\}}}, \tag{5}$$

where the parameter ϕ lies in interval $]0, 1[$ and should be assigned in advance; some guidelines about this respect are discussed in Section V.

The optimization problem described in Equation (4) must be constrained with respect to the parameter $\hat{B}_{z_i}$. This can be readily verified by observing that (5) approaches zero as $\hat{B}_{z_i}$ tends to one, which may lead to a global solution that is not the desired one. Moreover, the cost function (4) cannot be evaluated for $\hat{B}_{z_i} < \phi$, since the denominator of (4) would not be defined in this case — this is actually a border effect. To overcome these problems, we assume in our approach that $\hat{B}_{z_i}$ is constrained to the interval $[\phi, 1-\phi]$. This choice is further detailed in Appendix B, which discuss this point through an illustrative example.

Concerning the parameters $\mathbf{b}_i$, they must be constrained to avoid compensating functions that are not monotonic. The type of constraint will depend on the parametric function considered to compensate $f_i(\cdot)$. For instance, if a polynomial function having only odd terms is considered, then a monotonic compensating function is obtained by imposing non-negative polynomial coefficients.

Let us now discuss the rationale behind (5). Given that (5) is the ratio between the energies of $q_i(t)$ in the bands $[\hat{B}_{z_i}, 1]$ and $[\hat{B}_{z_i}-\phi, 1]$, this cost function attains a small value whenever the energy in the band $[\hat{B}_{z_i}, 1]$ is much smaller than the energy in the band $[\hat{B}_{z_i}-\phi, 1]$. The key point here is that such a situation is expected for the desired solution to our problem, i.e. for the situation in which $(\mathbf{b} = \mathbf{b}_d, \hat{B}_{z_i} = B_{z_i})$, where $\mathbf{b}_d$ represents the parameters that provide the inversion of $f_i(\cdot)$. Indeed, consider now the following formulation, which is equivalent to (5)

$$J_2(\mathbf{b}_i, \hat{B}_{z_i}) = \frac{1}{1 + \dfrac{E_{q_i}^{\{\hat{B}_{z_i}-\phi<f<\hat{B}_{z_i}\}}}{E_{q_i}^{\{f>\hat{B}_{z_i}\}}}}. \tag{6}$$

If $g_i \circ f_i$ is approximately linear, then $q_i(t)$ is bandlimited to B_{z_i} by assumption. Therefore, in this situation, a very low energy $E_{q_i}^{\{f>B_{z_i}\}}$ is expected. On the other hand, the term $E_{q_i}^{\{B_{z_i}-\phi<f<B_{z_i}\}}$ in (6) lies within the bandwidth of $z_i(t)$. As a consequence, this latter term is expected to be much larger than $E_{q_i}^{\{f>B_{z_i}\}}$. This explains the rationale behind the minimization of (6).

An important practical point here is the role of ϕ in $J_2(\mathbf{b}_i, \hat{B}_{z_i})$. This parameter acts as a sort of frequency resolution in the sense that the difference between the terms in the ratio present in $J_2(\mathbf{b}_i, \hat{B}_{z_i})$ is the energy in a frequency interval of size ϕ. For instance, if the input signal is periodic, then the parameter ϕ should be small as the energy variations are high concentrated in the spectrum. Conversely, for aperiodic signals, the energy is less concentrated in the spectrum and, thus, a greater value for ϕ can be defined.

D. Optimization of the Cost Function (5)

It should be noted that, if the spectrum of $q_i(t)$ presents strong energy variations due, for instance, to an attenuated band, then there will be significant variations between $E_{q_i}^{\{f>\hat{B}_{z_i}\}}$ and $E_{q_i}^{\{f>\hat{B}_{z_i}-\phi\}}$. As a consequence, cost function (5) tends to present local modes around the points $\hat{B}_{z_i}$ where these variations occur, that is, $J_2(\mathbf{b}_i, \hat{B}_{z_i})$ may be multimodal. Moreover, as discussed in Appendix B, there may be a local minimum at the point $\hat{B}_{z_i} = 1-\phi$. Hence, the application of methods based on local search mechanisms,

[2]In this work, we consider discrete-time signals. So, we always refer to the normalized frequency, where $B = 1$ corresponds, in the analog domain, to $F_s/2$, where F_s is the sampling frequency.

such as pure gradient-based techniques, may lead to sub-optimal convergence.

In our proposal, in order to to circumvent the problem mentioned above, we consider an artificial immune system (AIS) conceived to deal with multimodal optimization tasks, the artificial immune network for optimization (opt-aiNet), firstly proposed in [14]). This metaheuristic, which has been applied to solve a number of signal processing tasks [15], [16], possesses the required balance between local and global search mechanisms to deal with multimodal cost functions. Besides, this method does not operate with any kind of estimation of the gradient and/or the Hessian matrix. This feature is quite interesting in our context, since the calculation of the derivatives of (5) with respect to the parameter $\hat{B}_{z_i}$ is tricky. Finally, in the opt-aiNet, the constraints on $\hat{B}_{z_i}$, and possibly on $\mathbf{b}_i$, can be handled in a straightforward fashion. This is discussed in Appendix C, in which a detailed description of the opt-aiNet is provided.

The robustness of the opt-aiNet to sub-optimal convergence comes at the expense of an increase in the computational burden in comparison to gradient-based methods. This aspect is particularly important when the search space is large and the cost function is difficult to evaluate. Fortunately, in the context of smart chemical sensor arrays, the application that has motivated the present work, the number of unknown variables is usually small. Moreover, the evaluation of the cost function (5) is a straightforward task, since it corresponds to a ratio of energies in a given band. In our work, these energies are calculated by the Euclidean norm of the discrete cosine transform (DCT) coefficients associated with the desired band. The discrete Fourier transform (DFT) could also be used, however the DCT has the advantage of being a real-valued transform.

IV. Blind Source Separation in PNL Mixtures

The strategy to deal with nonlinear distortions described in Section III can be applied in the context of Blind Source Separation (BSS). The goal in this problem is to retrieve a set of unknown signals (sources) based on observations that correspond to mixed versions of these original sources. The term *blind* is employed since only a few assumptions are made about the sources and the mixing process. The problem of BSS has been mostly tackled by methods of independent component analysis (ICA) [17], [10], which work under the hypothesis that the sources correspond to mutually statistically independent random processes. Moreover, most of the works in BSS assume a linear mixing process, which simplifies the problem to a great extent.

However, while many practical problems can be safely described by linear models, there are some applications in which the mixing process is clearly nonlinear. This is the case, for instance, in the problem that motivated the present work: the development of smart chemical sensors arrays for analyzing ionic solutions [3]. In addition to that, nonlinear BSS methods have been used in applications such as separation of scanned images [18], quantum computing [19] and particle detectors [20].

In nonlinear BSS, one should deal with problems that are absent in the linear case. For instance, it can be shown that ICA does not necessarily lead to source separation in a general nonlinear framework [21], [22]. Such limitations have been driving researchers to consider constrained classes of nonlinear models that can be useful in practice and for which ICA still allows source separation. The most studied case in this context is the class of Post-Nonlinear (PNL) models [12]. As show in Figure 3, this kind of model comprises a linear mixing stage followed by a second stage composed of component-wise nonlinear functions. The PNL model provides a good description of systems presenting amplifier stages [12] and of ISEs arrays [3].

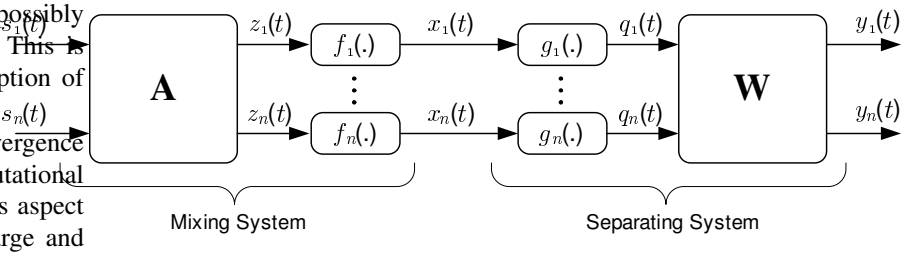


Fig. 3. The PNL problem structure.

In mathematical terms, if the n_s sources are represented by the vector $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_{n_s}(t)]^T$ and the n_m mixtures by the vector $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_{n_m}(t)]^T$, the outputs (mixtures) provided by a PNL model are given by

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{z}(t)) = \mathbf{f}(\mathbf{A}\mathbf{s}(t)), \tag{7}$$

where $\mathbf{f}(\cdot) = [f_1(\cdot), f_2(\cdot), \ldots, f_{n_s}(\cdot)]^T$ represents a set of component-wise functions that are applied to the set of linear mixtures, $\mathbf{z}(t)$, of the sources; the $n_m \times n_s$ matrix $\mathbf{A}$ models this linear mixing stage.

In order to retrieve the sources mixed by a PNL model, one can consider the separating system depicted in the right side of Figure 3. This system, which is basically a mirrored version of the mixing system, is composed of a set of nonlinear component-wise compensating functions, represented by $\mathbf{g}(\cdot) = [g_1(\cdot), g_2(\cdot), \ldots, g_{n_m}(\cdot)]^T$, followed by a separating matrix $\mathbf{W}$. Therefore, the estimated sources are given by

$$\mathbf{y}(t) = \mathbf{W}\mathbf{g}(\mathbf{x}(t)) = \mathbf{W}\mathbf{q}(t), \tag{8}$$

where $\mathbf{q}(t) = \mathbf{g}(\mathbf{x}(t))$.

Typically, the separation of PNL mixtures is accomplished by direct ICA methods, i.e. the two stages of the separating system are simultaneously adjusted. For instance, [12], [23], [24] consider a learning strategy based on the minimization of the mutual information between the elements of $\mathbf{y}(t)$. This task can be formulated as the following optimization problem

$$\min_{\mathbf{W},\mathbf{g}(\cdot)} I\left(\mathbf{y}(t)\right), \tag{9}$$

where $I\left(\mathbf{y}(t)\right)$ denotes the mutual information [25] between the elements of $\mathbf{y}(t)$. Assuming that $n_s = n_m$ and that the functions represented by $\mathbf{g}(\cdot)$ are invertible, Equation (9) can

be written as follows [12]

$$\min_{\mathbf{W},\mathbf{g}(\cdot)} \sum_{i=1}^{n_s} H(y_i(t)) - \log|\det \mathbf{W}| - E\left\{\log \prod_{i=1}^{n_s} |g_i'(x_i(t))|\right\}, \tag{10}$$

where $H(\cdot)$ denotes the Shannon's differential entropy [25]. The main drawback of direct methods is that they are dependent on a good estimation marginal entropies, which is usually a complex task. Moreover, it is difficult to formulate a multiple-input single-output (MISO) contrast based on the mutual information, rendering difficult the derivation of source extraction methods.

As an alternative to the joint or direct learning approach of (9), some works considered the so-called two-stage methods. In this approach, additional prior information on the sources is considered so the nonlinear and linear stages can be adjusted separately. Examples of two-stage methods include [26], in which one assumes that the sources are bounded, and [27], [28], which works under the hypothesis that the linear mixtures follow a Gaussian distribution.

In two-stage PNL source separation methods, the learning strategy can be formulated as the following two sequential optimization problems:

$$\begin{aligned} &\min_{\mathbf{g}(\cdot)} C_{\mathbf{g}}\left(\mathbf{y}(t)\right) \\ &\min_{\mathbf{W}} C_{\mathbf{W}}\left(\mathbf{y}(t)\right), \end{aligned} \tag{11}$$

where $C_{\mathbf{g}}(\cdot)$ and $C_{\mathbf{W}}(\cdot)$ correspond to the cost functions associated with the nonlinear and linear stages, respectively. The first optimization problem concerns the compensation of the set of nonlinear functions $\mathbf{f}(\cdot) = [f_1(\cdot), f_2(\cdot), \ldots, f_{n_s}(\cdot)]^T$, while the second one is related to a linear blind source separation problem, which can be tackled by the well-established linear BSS algorithms [10].

A. A Two-stage Approach for PNL Mixtures

By considering the proposed approach to deal with nonlinear distortions, described in Section III, it is possible to define a complete two-stage PNL source separation method. Indeed, if the sources are bandlimited signals with bandwidths given by $B_{s_1}, \cdots, B_{s_{n_s}}$, then the signal that is submitted to the nonlinear distortions, $z_i(t)$, will have a bandwidth bounded by $B_{z_i} = \max(B_{s_1}, \cdots, B_{s_{n_s}})$, thus satisfying the main assumption required by the approach introduced in Section III. As a result, the following two-stage PNL source separation method can be defined:

1) *First stage*: for each mixture $x_i(t)$, find $g_i(x_i(t), \mathbf{b}_i)$ by minimizing the cost function $J_2(\mathbf{b}_i, \hat{B}_{z_i})$, expressed in (5), through the opt-aiNet algorithm;
2) *Second stage*: the estimated sources $y_i(t)$ are obtained by applying a linear source separation or extraction method to the signals $q_i(t) = g_i(x_i(t)), \forall i = 1, \cdots, n_m$.

It is important to note in this procedure that the first step can be carried out even when the number of sources is smaller than the number of mixtures. In this case, however, one should consider in the second stage linear BSS strategies (e.g. priors like source sparsity) that are able to deal with underdetermined mixing models. Moreover, since the first step does not assume statistical independence between the sources, the estimation of the nonlinear functions can be carried out even when the sources are correlated. Finally, it is interesting to remark that the complexity of the first stage grows linearly with the number of mixtures, as the estimations of each nonlinear function is done in an independent fashion. Therefore, there is no curse of dimensionality in the first stage.

V. Experimental Results

In this section, we present a set of experiments to assess the performance of the proposed method to compensate nonlinear functions. In particular, we shall focus on the strategy for PNL source separation described in Section IV-A, giving special attention to the strategy developed for dealing with the first stage. At first, we consider the case in which the nonlinear stage of the PNL model is composed of logarithmic functions. This situation arises in the context of smart chemical sensor arrays. In this case, we test our method with synthetic data (Section V-A) but also with actual data (Section V-B). We also conduct simulations with a nonlinear stage composed of polynomial functions (Section V-C).

A. Experiments with Logarithmic Functions: the Nicolsky-Eisenman Model

1) Modeling ISE Arrays though PNL Systems: An ion-selective electrode (ISE) is a device used to estimate the ionic activity, which can be seen as a measure of effective concentration of an ion in aqueous solution [29]. Due to its simplicity, ISEs have been by far the most successful chemical sensor in commercial terms. However, this kind of sensor usually has weak selectivity, i.e. it may respond to interfering ions other than the target one.

A possible solution to deal with the interference problem in ISEs is to exploit the diversity brought by an array of ISEs. Recently, it was shown [3] that the demanding calibration stages that are typically conducted when using ISE arrays can be simplified by BSS methods. In this case, the sources represent the time series associated with the activity of each ion under analysis. Due to the interference problem, the outputs of the array correspond to mixed versions of the sources. The mixing process in this case can be modeled according to the classical formalism of the Nicolsky-Eisenman (NE) equation [29]. If the ions under analysis have the same valences, which is indeed very common in practice, then, according to the NE equation, the response of the i-th ISE within the array is given by:

$$x_i(t) = e_i + d_i \log_{10}\Big(\sum_{j=1}^{n_s} a_{ij} s_j(t)\Big), \tag{12}$$

where e_i is an unknown offset and a_{ij} denotes the selectivity coefficients. The parameter d_i, which is usually refereed to as the Nersntian slope, is approximately 0.059V for a room temperature. However, some factors such as aging and manufacturing variability may result in strong deviations from this theoretical value.

From Equation (12), one can note that the NE model is a particular case of the PNL model in which the nonlinear

mixing functions correspond to logarithms. The compensation of these functions can be achieved by means of the following parametric functions

$$q_i(t) = g_i(x_i(t), \hat{d}_i) = 10^{\frac{x_i(t)}{\hat{d}_i}} = 10^{\frac{e_i}{\hat{d}_i}} z_i(t)^{\frac{d_i}{\hat{d}_i}}. \qquad (13)$$

When $\hat{d}_i = d_i$, the composition $g_i \circ f_i$ is linear, corresponding thus to the desired solution. Note that the offset e_i of (12) cannot be estimated in a blind context, since it only introduces a scale gain in $q_i(t)$. Indeed, BSS methods are not able to retrieve the correct amplitude of the sources [10]). Therefore, in our experiments, we shall assume, without lost of generality, that $e_i = 0$.

2) Cost Function (5) in a synthetic example: To illustrate the effectiveness of our proposal for adjusting (13), we first consider a toy example with $n_s = 2$ sources and $n_m = 2$ mixtures. The synthetic sources, whose bandwidths are given by $B_{s_1} = 0.2$ and $B_{s_2} = 0.5$, were obtained from low-pass (finite impulse response) FIR filters (100 taps) driven by white Gaussian noise of zero mean and unit variance. The linear part of the PNL mixing system was given by the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.6 & 1 \end{bmatrix},$$

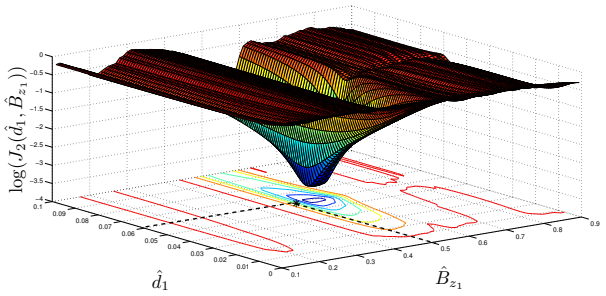
and the following Nernstian slopes were assumed: $d_1 = 0.059$ and $d_2 = 0.040$. Finally, we selected $\phi = 0.1$ in the cost function (4). This value was empirically found (after performing a series of simulations, we observed that a good rule of thumb is to select $\phi = 0.01$ for periodic signals and $\phi = 0.1$ for aperiodic signals).

Given that each separating function $g_i(x_i(t), \hat{d}_i)$ is parametrized by just a single parameter ($\hat{d}_i$), it is possible to visualize the cost function (5) in this case. For instance, considering a noiseless situation, the logarithm of the cost functions for both $g_1(\cdot)$ and $g_2(\cdot)$ are shown in Figure 4. Note that the values $\hat{d}_i$ that minimize (5) coincide with the actual values of d_i (the optimal solutions are indicated by dashed lines) in both cases. Moreover, the proposed criterion is minimized for both cases when $\hat{B}_{z_i} = 0.54$, which is close to the bandwidth of the linear mixtures ($B_{z_i} = 0.5$).

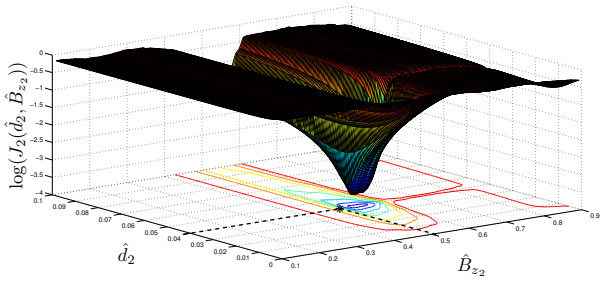
As discussed in Section III-D, local modes may appear in (5) in the presence of energy variations in the spectrum of $z_i(t)$. This is clear in Figure 4(a) where one can observe a local mode around the frequency $\hat{B}_{z_i} = 0.2$. In this case, the energy variation around this frequency takes place because $z_i(t)$ is a linear combination of two bandlimited signals, one of them having a bandwidth equal to $B_{s_1} = 0.2$.

3) Comparison between the cost functions (5) and (3): We here aim at assessing the performance of the following approaches in the task of estimating d_i: 1) the proposed cost function (5), 2) the cost function (3) assuming the knowledge of the bandwidth $B_{z_i(t)}$ (semi-blind case), which was proposed in [7], 3) the same cost function (3) but now in a complete blind situation, in which B_{z_i} is defined beforehand (we set $\hat{B}_{z_i} = 0.8$). The same scenario presented in the last experiment is considered.

In Table I, which represents the average of 100 experiments being each solution calculated through exhaustive search, one can note that, in a noiseless scenario, the three estimators give values closer to the actual ones ($d_1 = 0.059$ and $d_2 = 0.040$). However, in the presence of additive white Gaussian (AWG) noise, the blind version of (3) gives poor estimations for both d_1 and d_2 when the signal-to-noise ratio (SNR) is 20 dB, whereas the semi-blind version of (3) and our proposal (5) still give satisfactory estimation. However, it is worth mentioning that while (3) assumes the knowledge of the actual bandwidth of the input signal, our cost function (5) operates in a blind fashion.



(a) $J_2(\hat{d}_1, \hat{B}_{z_1})$.



(b) $J_2(\hat{d}_2, \hat{B}_{z_2})$.

Fig. 4. Cost functions $J_2(\hat{d}_i, \hat{B}_{z_i})$ for the NE model (synthetic sources).

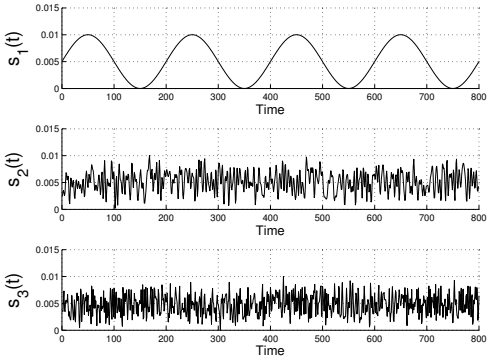
TABLE I
COMPARISON BETWEEN THE ESTIMATORS ASSOCIATED WITH THE COST FUNCTIONS (5) AND (3).

	$\hat{d}_1$ (noiseless)	$\hat{d}_2$ (noiseless)	$\hat{d}_1$ (SNR= 20 dB)	$\hat{d}_2$ (SNR= 20 dB)
$J_1(\hat{d}_i)$ (semi-blind)	0.0586	0.0395	0.0975	0.0455
$J_1(\hat{d}_i, \hat{B}_{z_i(t)})$ (blind)	0.0545	0.0398	0.9661	0.1085
$J_2(\hat{d}_i, \hat{B}_{z_i(t)})$	0.0589	0.0396	0.0885	0.0457

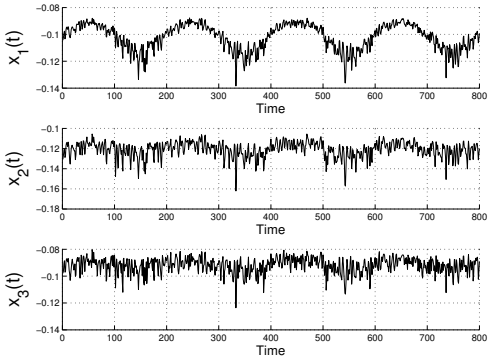
4) Example of source separation: We now present an example in which the complete procedure described in Section IV-A is applied to the NE model. We consider a scenario with $n_s = 3$ sources (a sine wave of frequency $B_{s_1} = 0.01$ and two aperiodic signals with bandwidth $B_{s_2} = 0.5$ and $B_{s_3} = 0.8$) and $n_m = 3$ mixtures. The sources are depicted in Figure 5(a) and the resulting mixtures in Figure 5(b). The PNL linear stage is given by the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.4 & 1 & 0.6 \\ 0.3 & 0.6 & 1 \end{bmatrix},$$

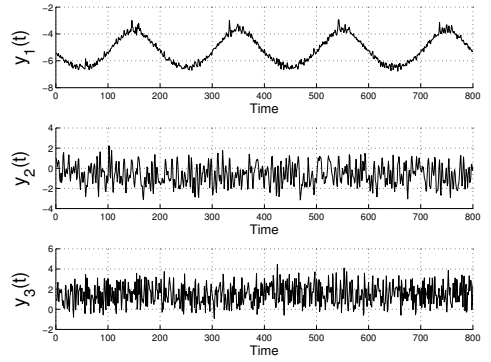
and the Nernstian slopes by $d_1 = 0.050$, $d_2 = 0.060$ and $d_3 = 0.045$. The number of available samples in this situation was 1000, and each sensor was corrupted by AWG noise of SNR $= 20\,\mathrm{dB}$.



(a) Sources.



(b) Noisy mixtures.



(c) Recovered sources.

Fig. 5. Application of the complete PNL method (the SOBI algorithm was considered in the linear stage).

According to the procedure described in Section IV-A, the first step is to apply the opt-aiNet algorithm (see Appendix C for more details) to perform the optimization of (5). After performing some preliminary tests, we considered the following set of parameters for the algorithm: $N = 9$ (initial population size), $N_c = 10$ (number of clones per individual), $\beta = 60$ and $\sigma_s = 3$ (these last two parameters are related to the mutation operator). We observed that the performance of the opt-aiNet was very robust with respect to changes in these parameters; this can be attributed to the fact that the search space in our problem is relatively small. As stopping criterion, we considered a maximum number of 1000 iterations.

The optimization of (5) via opt-aiNet algorithm led to the following estimates: $\hat{d}_1 = 0.048$, $\hat{d}_2 = 0.070$ and $\hat{d}_3 = 0.044$. Since in this toy example we have access to the outputs of PNL linear stage ($q_i(t)$) — this is unrealistic in an actual scenario — we can plot the resulting mappings between $z_i(t)$ and $q_i(t)$. As can be seen in Figure 6, these mappings are close to linear functions, indicating that the task of inverting the nonlinearities was satisfactorily accomplished. Yet, one can observed in this figure that the noise is amplified especially for high input values of $z_i(t)$. To understand why this amplification effect occurs, let us rewrite the output of the i-th sensor within the array when an AWG noise term $n_i(t)$ is present

$$x_i(t) = d_i \log_{10}(z_i(t)) + n_i(t). \tag{14}$$

One can easily obtain the mapping between $z_i(t)$ and $q_i(t)$ by applying (13) on (14), that is

$$q_i(t) = 10^{\frac{d_i \log(z_i(t)) + n_i(t)}{d_i}} = z_i(t)^{\frac{d_i}{d_i}} 10^{\frac{n_i(t)}{d_i}}. \tag{15}$$

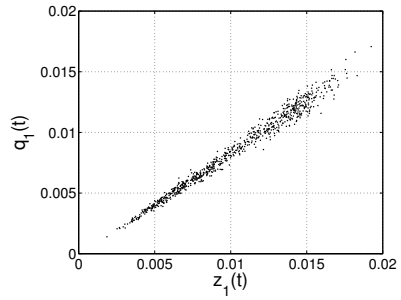
Therefore, due to the nonlinearities present in the global distorting/compensating functions, the original additive noise in $x_i(t)$ becomes a multiplicative noise with respect to the mapping between $z_i(t)$ and $q_i(t)$, which explains the noise amplification.

After estimating the separating nonlinear functions, we applied the Second Order Blind Identification (SOBI) algorithm [30] to the signals $q_i(t)$. This algorithm has been intensively applied to perform source separation in linear mixtures of colored sources. As shown in Figure 5(c), the SOBI algorithm provided good estimations of the actual sources. This is confirmed by the resulting signal-to-interference ratios[3] (SIR) for each pair actual source/estimated source: $\mathrm{SIR}_1 = 17.75\,\mathrm{dB}$ for the sine wave, and $\mathrm{SIR}_2 = 16.58\,\mathrm{dB}$ $\mathrm{SIR}_3 = 13.18\,\mathrm{dB}$ for the aperiodic signals. For matter of comparison, the SIRs obtained considering the mixtures as estimated sources are given by $\mathrm{SIR}_1 = 6.94\,\mathrm{dB}$, $\mathrm{SIR}_2 = 1.90\,\mathrm{dB}$, and $\mathrm{SIR}_2 = 3.45\,\mathrm{dB}$. We also tested a blind extraction algorithm, the second-order frequency identification (SOFI) algorithm [31], that is specially adapted to extract the smoothest signal, which, in our example, corresponds to the sine wave. As shown in Figure 7, a good estimation of the sine wave was obtained; the performance index was $\mathrm{SIR}_1 = 19.39\,\mathrm{dB}$. Note that, due to effect of noise amplification introduced by the nonlinear functions, the estimation error is more evident when the signal attains high values. The performance indices obtained when the original coefficients d_i and the original mixing matrix $\mathbf{A}$ are considered were $\mathrm{SIR}_1 = 15.43\,\mathrm{dB}$ for the sine wave, and $\mathrm{SIR}_2 = 10.01\,\mathrm{dB}$ $\mathrm{SIR}_3 = 11.69\,\mathrm{dB}$ for the aperiodic signals, revealing that, due to the noise, the solutions provided by the original parameters become suboptimal.

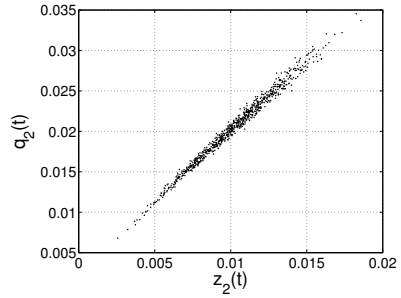
[3]The signal-to-interference ratio is given by:

$$\mathrm{SIR}_i = 10 \log \left(\frac{E\{\hat{s}_i(t)^2\}}{E\{(\hat{s}_i(t) - \hat{y}_i(t))^2\}} \right),$$

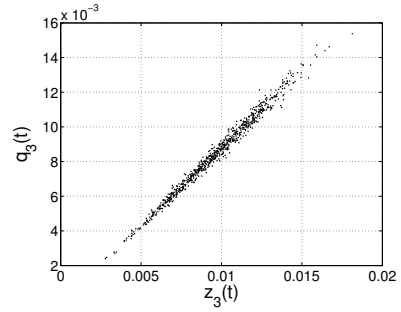
where $\hat{y}_i(t)$ and $\hat{s}_i(t)$ denote, respectively, the retrieved signal and the actual source after mean and variance normalization.



(a) $q_1(t) = g_1(z_1(t))$.



(b) $q_2(t) = g_2(z_2(t))$.



(c) $q_3(t) = g_3(z_3(t))$.

Fig. 6. Mappings between $z_i(t)$ and $q_i(t)$ for each channel.

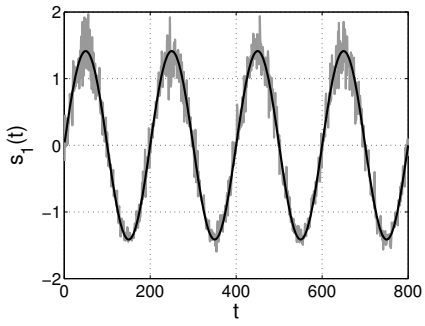
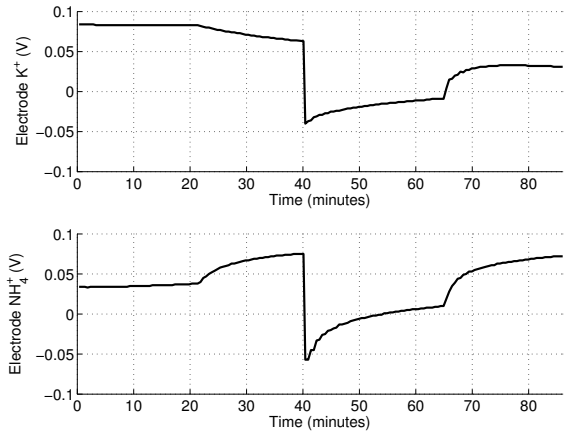


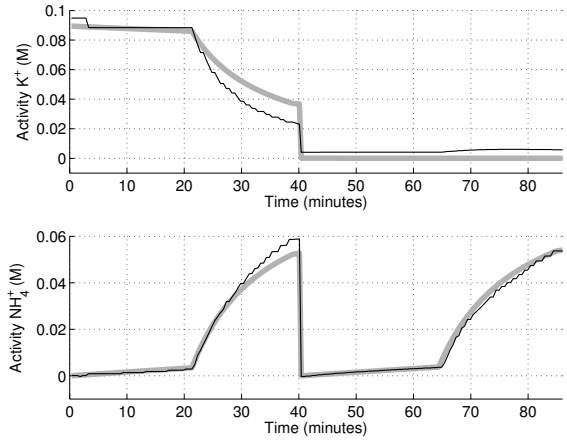
Fig. 7. Extraction of the smoothest source through the SOFI algorithm. Actual source $s_1(t)$ (black) and estimated source $y_1(t)$ (gray).

B. Experiments with Real Data

We now test our proposal in a real scenario. More precisely, we considered the experiments S1K10^{-1}NH$_4$ and S1K10^{-4}NH$_4$ of the Ion-Selective Electrode Array (ISEA) dataset [32]. This scenario corresponds to the one having the strongest interference level. In this case, there are $n_s = 2$ sources, which are related to ions ammonium (NH_4^+) and potassium (K^+), and $n_m = 2$ sensors within the array, being each one target to a different ion. These mixtures are depicted in Figure 8(a). The number of available samples in this case was 170.



(a) ISE array responses (mixtures).

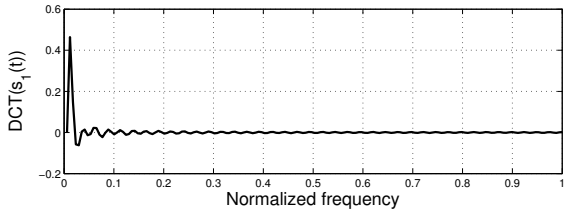


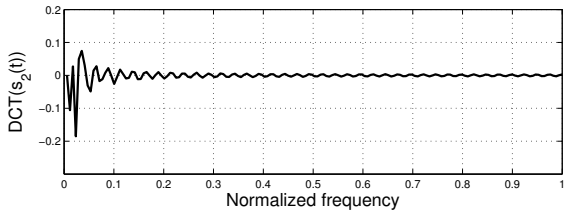
(b) Retrieved signals (black) and actual sources (gray)

Fig. 8. Experiments with actual data

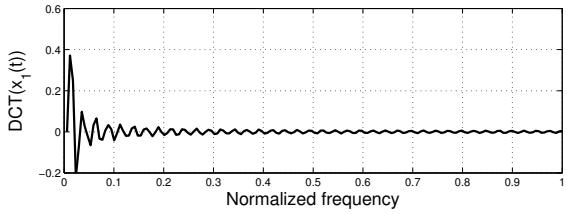
Since we also have access to the actual sources, which are shown in Figure 8(b) (gray signals), it is possible to analyze if the spectral spreading phenomenon is taking place in this situation. To that end, we show in Figure 9 the DCTs of the sources and mixtures. A first point to be made is that the sources are not bandlimited in the strict sense; there are high-frequency terms of very low energy. These terms arise because the sources, although smooth most of the time, are discontinuous around the instant $t = 40$ minutes. Yet, it is clear from Figure 9(a) that the sources' spectral content is concentrated on low-frequency bands. Another important aspect illustrated in Figure 9(b) is that the mixtures have indeed a DCT representation that is wider when compared to the sources. This is a clear indicator that the mixing process is of nonlinear nature, as predicted by the NE equation.

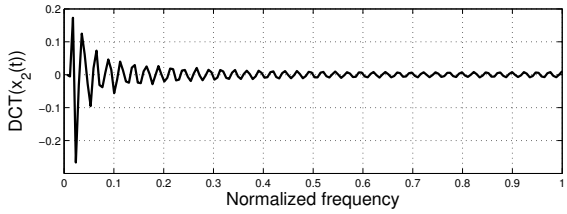
After applying the first stage of the method described in Section IV-A (the parameters of the opt-aiNet in this case were $N = 7$, $N_c = 5$, $\beta = 30$ and $\sigma_s = 3$), we obtained the following Nernstian slopes: $\hat{d}_1 = 0.031$ and $\hat{d}_2 = 0.054$. By considering the actual sources available in the dataset, we





(a) DCTs of sources.





(b) DCTs of mixtures

Fig. 9. DCTs of the sources and mixtures obtained in an experiments with actual data

compared the obtained results with the ones provided by a supervised MSE regression, which has led to the following values $\hat{d}_1 = 0.039$ and $\hat{d}_2 = 0.050$. Note that these values are close to those obtained by our approach. Finally, the bandwidths estimated by our method were $B_{s_1} = 0.020$ and $B_{s_2} = 0.032$, which are consistent with Figure 9(a), as most of the energy of the sources are found in frequency bands that are below these frequencies.

For estimating each function, the proposed method performed 5000 iterations, which took approximately 25s[4]. However, we observed that, very often, the method required much less than 5000 iterations to converge. For instance, after conducting 10 realizations of our method, we observed that, in the worst simulation, the algorithm converged after approximately 1000 iterations in the case of d_1 and after approximately 2000 iterations in the case of d_2. Therefore, good estimations were obtained after approximately 5 seconds (for d_1) and after approximately 12 seconds (for d_2).

Concerning the second stage, there is in this case an additional difficulty that limits the application of the SOBI and SOFI algorithms. Actually, the sources (gray signals in 8(b)) are highly correlated, thus violating the fundamental assumption of almost all BSS methods. To mitigate this problem, we consider the linear Bayesian source separation method proposed in [33]. Our choice is motivated by the fact that Bayesian methods may provide fair estimations even when the sources have a certain degree of correlation. Actually, differently from ICA methods, Bayesian methods do not have the independence as a central assumption.

The application of the method [33] led to the estimated sources shown in Figure 8(b) (black signals). The performance indices in this case were: $\text{SIR}_1 = 17.74\,\text{dB}$ and $\text{SIR}_2 = 22.13\,\text{dB}$. For a matter of comparison, the estimates provided by considering the SOBI algorithm in the second stage led to: $\text{SIR}_1 = 18.45\,\text{dB}$ and $\text{SIR}_2 = 9.73\,\text{dB}$.

We also applied the direct PNL ICA-based method proposed in [16]. We observed that this method achieve a poor performance in this scenario with real data. The performance indices in this case were: $\text{SIR}_1 = 7.61\,\text{dB}$ and $\text{SIR}_2 = -0.32\,\text{dB}$. This bad performance can also be attributed to the fact the sources are correlated in this situation.

C. Compensation through Polynomial Functions

Our proposal can also be applied for compensating other types of nonlinear functions. To illustrate that, we consider in this subsection the situation in which the compensating function is given by an odd polynomial, that is:

$$q_i(t) = g_i(x_i(t), \mathbf{w}_i) = \sum_{k=1}^{N_p} w_k x^{2k-1}(t), \qquad (16)$$

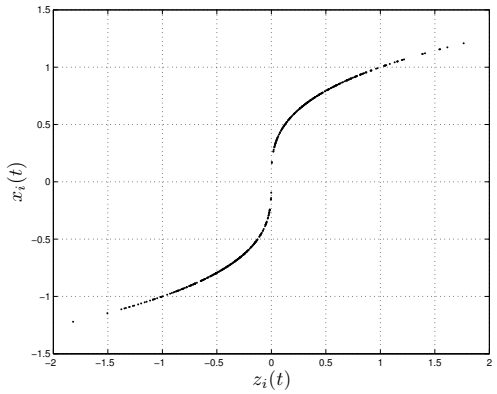
where $2N_p - 1$ is the degree of the polynomial. In the sequel, we present some results considering two situations, in which we simply tackle the problem of compensating a nonlinear function. Then, in a third scenario, we considered the problem of PNL source separation in which the compensating functions are given by polynomials.

1) Case in which perfect inversion is possible: In a first moment we consider that the nonlinear distorting function is given by $f_i(z_i(t)) = \sqrt[3]{z_i(t)}$. Therefore, in this case, (16) can perfectly invert the the distorting function. To check if our approach can be applied in this case, we considered a synthetic input signal generated by a low-pass FIR filter (100 taps) driven by white Gaussian noise of zero mean and unit variance; the number of samples in this case were 500. Finally, we defined $\phi = 0.1$ and $N_p = 3$.

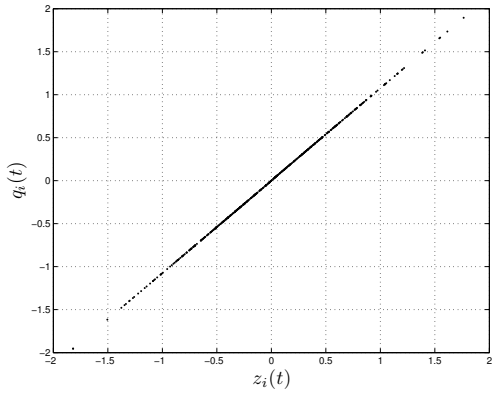
The optimization of (5) through the opt-aiNet led to a perfect inversion in all of 20 realizations executed. This is illustrated in Figure 10, which shows the joint plot between the input signal $z_i(t)$ and the observed signal $x_i(t)$ and the joint plot between the input signal $z_i(t)$ and the signal provided by the compensating function $q_i(t)$ obtained in one of the realizations. Note that this last joint plot is linear, thus indicating that a perfect compensation was achieved in this case

2) Case in which perfect inversion is not possible: The parametric form of the distorting function may not be available in some applications. As consequence, in these cases, it may not be possible to achieve a perfect inversion. In view of this

[4]We implemented our method in Matlab (Windows XP) and the simulations were performed in a Intel Core 2 duo T6400 2 GHz, 3 GB RAM



(a) Joint plot $z_i(t) \times x_i(t)$.

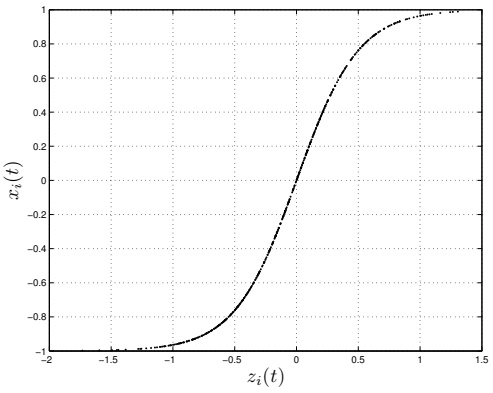


(b) Joint plot $z_i(t) \times q_i(t)$.

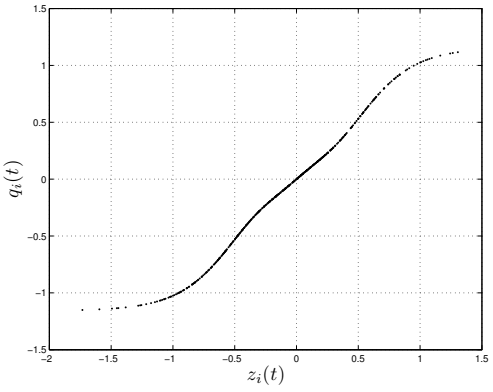
Fig. 10. Compensation of a nonlinear function when perfect inversion is possible.

limitation, it is important that a flexible enough compensating function, such as the polynomial function (16), be available. With the aim of verifying if the proposed framework is able to provide a satisfactory compensation in this case, we performed an experiment in which the nonlinear distorting function is given by $f_i(z_i(t)) = \tanh(2z_i(t))$. We set $N_p = 4$ while the other parameters were the same as the ones considered in our last experiment.

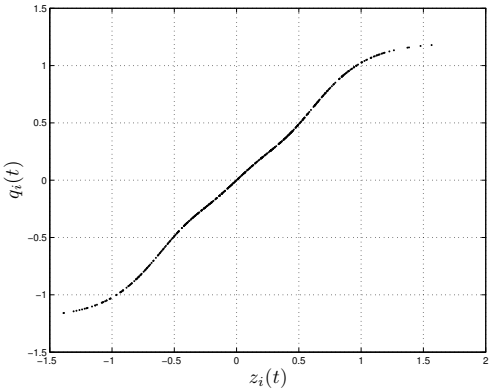
In Figure 11, we provide the joint plot between the input signal $z_i(t)$ and the observed signal $x_i(t)$ (Figure 11(a)), and the joint plot between the input signal $z_i(t)$ the signal $q_i(t)$ estimated by the proposed method ((Figure 11(b))). These results were obtained after 5000 iterations of the opti-aiNet. As expected, it is not possible to perfectly compensate the nonlinear distortion in this case, since a polynomial cannot invert a hyperbolic tangent function. Yet, the input-output relationship shown in Figure 11(b) reveals that at least the nonlinear distortion was clearly weakened. For matter of comparison, we also show in Figure 11(c) the joint plot between the input signal $z_i(t)$ and the signal $q_i(t)$ estimated by a supervised approach, in which the polynomial coefficients were adjusted to minimize the mean square error between the input signal $z_i(t)$ and the estimated one $q_i(t)$. It is interesting to note that, even in a supervised approach, there is still a remaining nonlinear distortion, which is slightly smaller than the one observed in Figure 11(b).



(a) Joint plot $z_i(t) \times x_i(t)$.



(b) Joint plot $z_i(t) \times q_i(t)$ obtained by the proposed blind method.



(c) Joint plot $z_i(t) \times q_i(t)$ obtained by a supervised approach based on the minimization of the mean square error.

Fig. 11. Compensation of a nonlinear function when perfect inversion is not possible.

3) Source Separation in PNL mixtures: Let us now consider the problem of separating $n_s = 9$ signals in the context of PNL mixtures. In these experiments, the nonlinear compensating functions were given by the polynomials (16) where $N_p = 4$. The sources were generated by low-pass FIR filters (100 taps) driven by white Gaussian noise of zero mean and unit variance. The bandwidths of theses signals were obtained from realizations of a random variable uniformly distributed in $[0.1, 0.5]$. The number of samples were 2000. The mixing matrix was obtained from realizations of a random variable

uniformly distributed in $[0.1, 0.8]$, while the elements of the main diagonal were given by 1. The distorting functions were given by $f_i(z_i(t)) = \tanh(z_i(t))$. Finally, we set $\phi = 0.05$.

In Figure 12, we show the joint plot between each linear mixture $z_i(t)$ and its corresponding nonlinear distorted signal $x_i(t)$ (observed mixtures). The application of our proposal, considering 10000 iterations of the opt-aiNet algorithm for each function (which took approximately 135 seconds), led to the joint plots $z_i(t) \times q_i(t)$ shown in Figure 13. As can be seen by comparing Figures 12 and 13, although the hyperbolic tangent functions were not perfectly compensated by the polynomial functions, the nonlinear distortions were considerably attenuated.

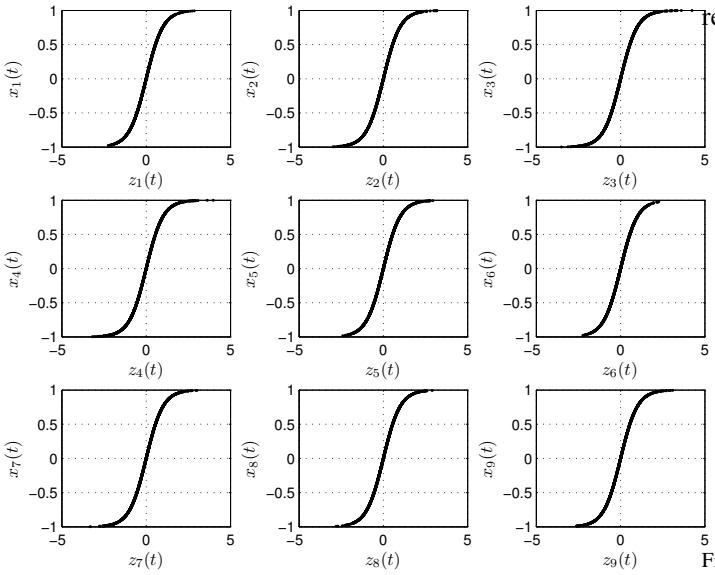


Fig. 12. Joint plots between linear mixtures $z_i(t)$ and observed signals $x_i(t)$.

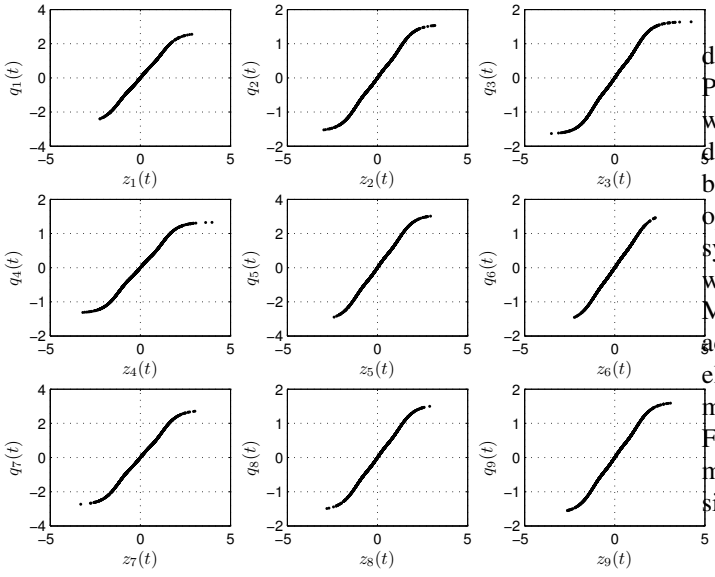


Fig. 13. Joint plots between linear mixtures $z_i(t)$ and signals $q_i(t)$, obtained after the compensation of the nonlinear function.

Having conducted the proposed method for compensating the nonlinear functions, we applied the SOBI algorithm considering as inputs the signals $q_i(t)$, which correspond to the outputs of the compensating functions. The proposed two-stage method provided good estimations of the sources, since the corresponding SIRs were given by $\text{SIR}_1 = 17.81\,\text{dB}$, $\text{SIR}_2 = 12.74\,\text{dB}$, $\text{SIR}_3 = 10.85\,\text{dB}$, $\text{SIR}_4 = 14.73\,\text{dB}$, $\text{SIR}_5 = 11.34\,\text{dB}$, $\text{SIR}_6 = 14.33\,\text{dB}$, $\text{SIR}_7 = 12.78\,\text{dB}$, $\text{SIR}_8 = 13.31\,\text{dB}$, and $\text{SIR}_9 = 14.52\,\text{dB}$. In order to illustrate this result, we plot in Figure 14 the pair source-retrieved source corresponding to the $\text{SIR}_7 = 12.78\,\text{dB}$. In this figure, the scale of the estimated source was corrected and, for the sake of visualization, only 200 samples were plotted. Note that the estimated source was indeed very close to the actual one. There is a small high frequency distortion that is mainly due to the residual nonlinear distortion.

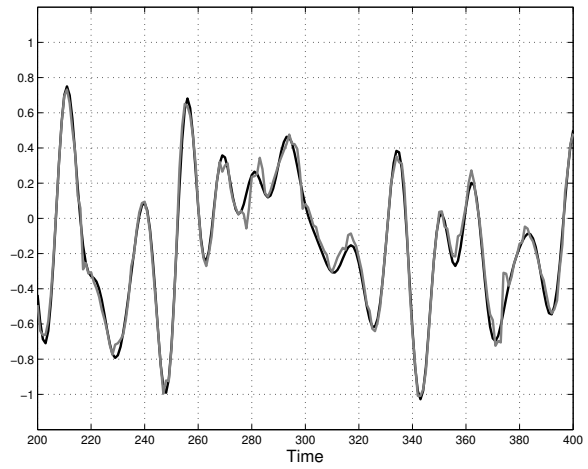


Fig. 14. Obtained signal after the application of the SOBI algorithm: actual source (black) and retrieved one (gray).

VI. Conclusions

In this work, we tackled the problem of inverting a nonlinear distortion. Our main motivation was to develop a two-stage PNL source separation method, and to accomplish that aim we assumed that the desired sources can be modeled as band-limited signals. Based on the spectral spreading introduced by the nonlinear distorting function, a framework able to operate in a blind scenario was proposed. Experiments with synthetic data pointed out that the proposal performs well even when the nonlinear distorted signals are corrupted by noise. Moreover, the proposed approach was tested considering an actual situation in which the data come from an ion-selective electrode array. Despite the reduced number of samples, our method provided good estimations of the nonlinear stage. Future works will concern the extensions of the proposed method to more general classes of signals, such as broadband signals.

Appendix A
Effect of non-linearity on signals bandwidth

Spectral spreading due to nonlinear functions is a classical result in signal processing theory [6], [7]. In order to understand why this phenomenon takes place, let us assume that

$f_i(\cdot)$ admits a power series expansion, i.e.,

$$x_i(t) = f_i\left(z_i(t)\right) = \sum_{k=1}^{\infty} f_i^{(k)}[z_i(t)]^k. \tag{17}$$

Denoting by $Z_i(\omega)$ the Fourier transform of $z_i(t)$, the Fourier transform of (17) is given by

$$\begin{aligned} X_i(\omega) = f_i^{(1)} Z_i(\omega) + f_i^{(2)} Z_i(\omega) * Z_i(\omega) + \\ f_i^{(3)} Z_i(\omega) * Z_i(\omega) * Z_i(\omega) + \cdots, \end{aligned} \tag{18}$$

where the symbol '$*$' stands for the convolution operator. A basic property of the convolution states that if $R_1(\omega)$ and $R_2(\omega)$ denote the Fourier transform of two signals bandlimited to B_1 and B_2, respectively, then $R_1(\omega) * R_2(\omega)$ is bandlimited to $B_1 + B_2$. Thus, in (18), since $Z_i(\omega)$ is bandlimited to $B_{z_i(t)}$, then $Z_i(\omega) * Z_i(\omega)$ will be bandlimited to $2B_{z_i(t)}$, $Z_i(\omega) * Z_i(\omega) * Z_i(\omega)$ to $3B_{z_i(t)}$, and so forth. As a consequence, the maximum frequency of $X_i(\omega)$ tends to become larger than B_{z_i}.

APPENDIX B
ANALYSIS OF THE COST FUNCTION (5) IN A SIMPLE CASE

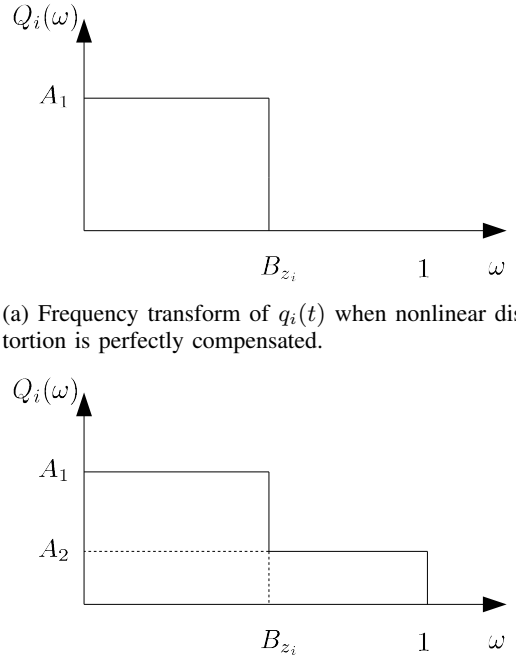
We here analyze the cost function (5) in a simple context. Our main goal is to illustrate why it is important to constrain the parameter $\hat{B}_{z_i}$ in the optimization problem expressed in (4) to the interval $[\phi, 1-\phi]$. In our analysis, we will consider a situation in which: 1) the input signal $z_i(t)$ is bandlimited with maximum frequency given by B_{z_i}; 2) the frequency transform of $z_i(t)$, denoted by $Z_i(\omega)$, takes a constant value A_1 within the interval $[0, B_{z_i}]$; 3) the nonlinear distorting function can be perfectly compensated; 4) the presence of a residual nonlinear distortion only generates, in the spectrum of the compensated signal, $q_i(t)$, a constant value A_2 within the interval $[B_{z_i}, 1]$.

In the scenario described in the last paragraph, the frequency transform of the signal $q_i(t)$, here denoted by $Q_i(\omega)$, is illustrated in two different situation: Figure 15(a) corresponds to the case in which the nonlinear distorting function is perfectly compensated — we are assuming that there is no scale ambiguity — while Figure 15(b) represents the case in which there is still nonlinear distortion.

The cost function (5) can be easily obtained in the situations depicted in Figure 15. For instance, let us consider the case where there is still a nonlinear distortion and $\hat{B}_{z_i} > B_{z_i} + \phi$. Since (5) corresponds to the ratio between the areas of $Q_i(\omega)$ in the intervals $[\hat{B}_{z_i} - \phi, 1]$ and $[\hat{B}_{z_i}, 1]$, respectively, it asserts that

$$J_2 = \frac{1 - \hat{B}_{z_i}}{1 - \hat{B}_{z_i} + \phi} \tag{19}$$

in this case. A relevant point important here is that if $\hat{B}_{z_i} = 1$, then $J_2 = 0$ in this case. In other words, since J_2 is always non-negative, there is a global minimum at the border $\hat{B}_{z_i} = 1$. To avoid that the algorithm converges to this global minimum, there are, at least, two possibilities. The first one is to select, at the end of the opt-aiNet algorithm execution, the global minimum that is not located at the border $\hat{B}_{z_i} = 1$. This is possible since the opt-aiNet algorithm provides a population of solutions instead of a single one (see Appendix C).



(a) Frequency transform of $q_i(t)$ when nonlinear distortion is perfectly compensated.

(b) Frequency transform of $q_i(t)$ when nonlinear distortion is not compensated.

Fig. 15. A simple scenario to analyze the cost function (5).

A second solution, which is the one adopted in this work, is to constrain the parameter $\hat{B}_{z_i}$ to be inferior than $1 - \phi$. This procedure, which can be easily done in the opt-aiNet algorithm (see Appendix C), aims at avoiding the existence of a global minimum other than the desired one, which takes place when $\hat{B}_{z_i} = B_{z_i}$ and there is no nonlinear distortion (situation depicted in Figure 15(a)). Indeed, if one considers, for instance, the situation described in our working example, than one can easily check that, in the presence of nonlinear distortion, (19) is always greater than zero and takes $1/2$ when $\hat{B}_{z_i} = 1 - \phi$, that is, there is no more a global minimum at the new border $\hat{B}_{z_i} = 1 - \phi$. Yet, one can readily note that, if $B_{z_i} < 1 - 2\phi$, the solution $1/2$ at $\hat{B}_{z_i} = 1 - \phi$ is a local minimum. At first sight, this local minimum would pose a problem to the implementation our proposal, especially if classical optimization methods based on gradient search were considered. However, since we are making use of an optimization method that is robust to local convergence, this minimum located at the border $\hat{B}_{z_i} = 1 - \phi$ is easily avoided.

APPENDIX C
THE OPT-AINET ALGORITHM

The opt-aiNet is based on two main theoretical pillars: the synergy between clonal selection and affinity maturation [34], [35] and the idea of immune network [36]. Each solution to the problem at hand — in our case, each possible real-valued vector $\mathbf{b}_i$ and bandwidth $\hat{B}_{z_i}$ — is assumed to correspond to an antibody (or individual). The cost function to be optimized is considered to be an index of the matching between antibodies and a given invader (antigen), being referred to as either

affinity or fitness function[5]. Finally, in contrast with classical nonlinear optimization methods, the algorithm works with a population of solutions (antibodies).

The combination between clonal selection and affinity maturation corresponds to a threefold stage in which each individual is subject to cloning, mutation and deterministic selection. In other words, each solution is replicated N_c times, being, afterwards, all clones subject to Gaussian mutation. From the group formed by the original individual and the mutated clones, only the solution with the best fitness value is kept. A relevant feature of the mutation operator is that the variance of the Gaussian perturbation is inversely proportional to the fitness of the original individual i.e. better individuals tend to be less modified, which favors a fine local search.

The immune network theory plays a key role in the opt-aiNet: that of controlling the population size by pruning redundant solutions. This process is implemented in two steps: by verifying the evolution of the average fitness of the population and, in case of stagnation, removing antibodies that are too close (in terms of Euclidean distance) to each other. The pruning process is complemented by the periodic insertion of new randomly-generated individuals.

Algorithm 1 is a summary of the opt-aiNet method. The steps $3-6$ correspond to the local search process inspired in the aforementioned duo clonal selection/affinity maturation, whereas steps $7-9$ are responsible for the network character of the method, which is decisive in terms of global search and parsimony.

The incorporation of a constraint in the optimization problem can be done in a simple manner by the opt-aiNet algorithm. For instance, as discussed in Section III-C, in our problem, $\hat{B}_{z_i}$ must lie within $[\phi, 1-\phi]$. The incorporation of this constraint in this case can be done by mapping, before each fitness evaluation, all values lower than ϕ to ϕ, and all values greater than $1-\phi$ to $1-\phi$. At the end, the same mapping must be applied in the obtained solution.

Algorithm 1 Optimization through opt-aiNet

1: *Initialization*. Randomly initialize a population of solutions with N individuals
2: **while** the stopping criterion is not met **do**
3: *Fitness evaluation*. Determine the fitness of each individual of the population.
4: *Replication*. Generate N_c copies (clones) of each individual.
5: *Mutation*. Mutate each copy $\mathbf{c}$ in accordance with the following scheme:

$$\hat{\mathbf{c}} = \mathbf{c} + \exp(-Fit)\beta^{-1}N(0,1),$$

being $\hat{\mathbf{c}}$ a mutated version of $\mathbf{c}$, $\beta^{-1}N(0,1)$ a Gaussian random vector with uncorrelated elements, null mean and standard deviation β^{-1}. Fit here corresponds to the fitness of $\mathbf{c}$.
6: *Selection*. For each group formed by a parent individual and its mutated clones, select the individual with the highest fitness and calculate the average fitness of the selected population.
7: **if** the average fitness of the population is not significantly modified, then continue; **else** return to step 3.
8: *Network interactions*. Suppress all individuals whose Euclidean distance are less than a threshold σ_s, except the individual with the highest fitness measure.
9: *Diversity introduction*. Introduce a pre-defined number of randomly-generated individuals.
10: **end while**

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their suggestions and remarks, which have greatly improved the presentation of this paper.

REFERENCES

[1] L. T. Duarte, R. Suyama, R. R. F. Attux, B. Rivet, C. Jutten, and J. M. T. Romano, "Source separation of baseband signals in post-nonlinear mixtures," in *Proc. of the IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, 2009.
[2] T. T. Ha, *Digital satellite communications*. McGraw-Hill, 1990.
[3] L. T. Duarte, C. Jutten, and S. Moussaoui, "A Bayesian nonlinear source separation method for smart ion-selective electrode arrays," *IEEE Sens. Journal*, vol. 9, no. 12, pp. 1763–1771, 2009.
[4] M. Ibnkahla, "Applications of neural networks to digital communications: a survey," *Signal Processing*, vol. 80, pp. 1185–1215, 2000.

[5]Since we are interested in minimizing (4), the fitness of a given candidate solution can be defined as

$$Fit = (J_2 + 1)^{-1}.$$

Since J_2 is always non-negative, Fit will be always between 0 and 1.

[5] J. Tsimbinos and K. Lever, "Nonlinear system compensation based on orthogonal polynomial inverses," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 48, no. 4, pp. 406–417, April 2001.
[6] H. J. Landau and W. L. Miranker, "The recovery of distorted bandlimited signals," *Journal of Mathematical Analysis and Applications*, vol. 2, pp. 97–104, 1961.
[7] K. Dogancay, "Blind compensation of nonlinear distortion for bandlimited signals," *IEEE Trans. on Circuits and Syst. I: Regular Papers*, vol. 52, no. 9, pp. 1872–1882, 2005.
[8] ——, "Adaptive pre-distortion of nonlinear systems using out-of-band energy minimization," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 2007, pp. 277–280.
[9] S. Suranthiran and S. Jayasuriya, "Signal conditioning with memoryless nonlinear sensors," *Journal of Dynamic Systems, Measurement, and Control*, vol. 126, pp. 284–293, 2004.
[10] P. Comon and C. Jutten, Eds., *Handbook of blind source separation: independent component analysis and applications*. Academic Press, 2010.
[11] J. M. T. Romano, R. R. F. Attux, C. C. Cavalcante, and R. Suyama, *Unsupervised signal processing: channel equalization and source separation*. CRC Press, 2011.
[12] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2807–2820, Oct. 1999.
[13] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2001.
[14] L. N. de Castro and J. Timmis, "An artificial immune network for multimodal function optimization," in *Proceedings of the Congress on Evolutionary Computation (CEC)*, vol. 1, 2002, pp. 699–704.
[15] R. Attux, L. de Castro, F. Von Zuben, and J. Romano, "A paradigm for blind IIR equalization using the constant modulus criterion and an artificial immune network," in *Proceedings of IEEE Workshops on Neural Networks for Signal Processing (NNSP)*, 2003.
[16] L. T. Duarte, R. Suyama, R. R. F. Attux, F. J. V. Zuben, and J. M. T. Romano, "Blind source separation of post-nonlinear mixtures using evolutionary computation and order statistics," *Lecture Notes on Computer Science*, vol. 3889, pp. 66–73, 2006.

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

[18] L. B. Almeida, "Separating a real-life nonlinear image mixture," *Journal of Machine Learning Research*, vol. 6, p. 11991229, 2005.

[19] Y. Deville and A. Deville, "Maximum likelihood blind separation of two quantum states (qubits) with cylindrical-symmetry Heisenberg spin coupling," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[20] E. F. S. Filho, J. M. Seixas, and L. P. Calôba, "Modified post-nonlinear ICA model for online neural discrimination," *Neurocomputing*, vol. 73, pp. 2820–2828, 2010.

[21] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: existence and uniqueness results," *Neural Networks*, vol. 12, pp. 429–439, 1999.

[22] C. Jutten and J. Karhunen, "Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures," *International Journal of Neural Systems*, vol. 14, pp. 267–292, 2004.

[23] S. Achard, D.-T. 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.

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

[25] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-Interscience, 1991.

[26] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "A geometric approach for separating post non-linear mixtures," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2002.

[27] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller, "Blind separation of post-nonlinear mixtures using gaussianizing transformations and temporal decorrelation," *Journal of Machine Learning Research*, vol. 4, pp. 1319–1338, 2003.

[28] J. Solé-Casals, C. Jutten, and D. T. Pham, "Fast approximation of nonlinearities for improving inversion algorithms of PNL mixtures and Wiener systems," *Signal Processing*, vol. 85, pp. 1780–1786, 2005.

[29] P. Gründler, *Chemical sensors: an introduction for scientists and engineers*. Springer, 2007.

[30] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[31] L. T. Duarte, B. Rivet, and C. Jutten, "Blind extraction of smooth signals based on a second-order frequency identification algorithm," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 79–82, 2010.

[32] L. T. Duarte, C. Jutten, P. Temple-Boyer, A. Benyahia, and J. Launay, "A dataset for the design of smart ion-selective electrode arrays for quantitative analysis," *IEEE Sens. Journal*, vol. 10, pp. 1891–1892, 2010.

[33] L. T. Duarte, C. Jutten, and S. Moussaoui, "Bayesian source separation of linear and linear-quadratic mixtures using truncated priors," *Journal of Signal Processing Systems*, pp. 1–13, 2010.

[34] F. M. Burnet, *The clonal selection theory of acquired immunity*. Cambridge University Press, 1959.

[35] L. N. de Castro and J. Timmis, *Artificial immune systems: a new computational intelligence approach*. Springer-Verlag, 2002.

[36] N. K. Jerne, "Towards a network theory of the immune system," *Ann. Immunol. (Inst. Pasteur)*, vol. 125C, pp. 373–389, 1974.