



HAL
open science

A Sparsity-Based Method for Blind Compensation of a Memoryless Nonlinear Distortion: Application to Ion-Selective Electrodes

Leonardo Tomazeli Duarte, Ricardo Suyama, Romis Attux, João Marcos Romano, Christian Jutten

► **To cite this version:**

Leonardo Tomazeli Duarte, Ricardo Suyama, Romis Attux, João Marcos Romano, Christian Jutten. A Sparsity-Based Method for Blind Compensation of a Memoryless Nonlinear Distortion: Application to Ion-Selective Electrodes. *IEEE Sensors Journal*, 2015, 15 (5), pp.2054-2061. 10.1109/JSEN.2014.2371013 . hal-01164752

HAL Id: hal-01164752

<https://hal.science/hal-01164752>

Submitted on 18 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Sparsity-based Method for Blind Compensation of a Memoryless Nonlinear Distortion: Application to Ion-Selective Electrodes

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

Abstract

In this work, we propose a method for blind compensation of a memoryless nonlinear distortion. We assume as prior information that the desired signal admits a sparse representation in a transformed domain that should be known in advance. Then, given that a nonlinear distortion tends to generate signals that are less sparse than the desired one, our proposal is to build a compensating function model that gives rise to a maximally sparse signal. The implementation of this proposal has, as central elements, a criterion built upon an approximation of the ℓ_0 -norm, the use of polynomial functions as compensating structures, and an optimization strategy based on sequential quadratic programming. We provide a theoretic analysis for an ℓ_0 -norm criterion and results considering synthetic data. We also employ the method in an actual application related to chemical analysis via ion-selective electrode arrays.

The authors thank the São Paulo Research Foundation (FAPESP, Brazil) and the National Council for Scientific and Technological Development (CNPq, Brazil) for funding their research. This work has been partly supported by the European project ERC-2012-AdG-320684-CHESS. This paper was presented in part at the 20th European Signal Processing Conference (EUSIPCO 2012) and was published in its proceedings [1]. In this new version, we extend the theoretical results and provide results considering real data.

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 (e-mail: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)

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

Index Terms

Nonlinear distortion, sparsity, ion-selective electrodes, chemical sensor arrays, blind methods.

I. INTRODUCTION

Although, historically, the vast majority of signal processing methods were conceived within the conceptual framework of linear system theory, there are many applications in which one has to deal with nonlinear distortions that arise in the processes of signal transmission and acquisition.

A potential source of nonlinear distortion is the existence of amplifier stages operating in the saturation region [2]. Such a scenario often takes place in satellite communications [3] and audio signal processing [4], [5]. Nonlinear distortions may also be a consequence of the transducer employed in the signal acquisition process. For example, in electro-optical transduction, the process of converting an optical source to a digital image may present some nonlinear effects [6]. Another example of transducers that suffer from nonlinear distortions can be found in electrochemical transduction. In particular, the conversion of chemical activity into electrical potential is often modeled by the well-known Nernst's equation, which is a nonlinear law [7].

In situations like those presented above, unless some action is taken, the acquired signal might be impaired by nonlinear distortions. A first strategy to deal with such a problem employs tools derived from system identification theory [2], [8]. In this approach, one first searches for a good model of the system, which is usually accomplished by means of a parametric model, and then develops an inversion procedure that is based on the previously estimated model. In another classical strategy, nonlinear compensation is done by directly adjusting a system devoted to counterbalance the nonlinear effects, i.e., without a system identification step in this case. This can be accomplished by placing a compensating system either after the nonlinear distortion location or before it — the latter strategy is usually known as predistortion [9], [8].

Very often, both the system identification and the inversion approaches are based on a supervised adaptation, which means that it is mandatory to have access to a set of training (calibration) points. Such requirement may be seen as a strong limitation, especially in situations in which the acquisition of training points is an expensive and/or time-consuming process.

In the case of blind (unsupervised) nonlinear compensation, training points are not available, i.e., one has only access to the output signal. Therefore, it is mandatory to consider prior information, otherwise the problem becomes ill-posed. For instance, in the seminal work of Landau and Miranker [10], which

was recently extended to blind configurations [11], [12], the main assumption is that the desired signal is bandlimited. In this case, given that the nonlinear distortion causes a spread in the spectrum of the acquired signal, a possible framework to blindly compensate the nonlinear distortion is to set up a compensating system that retrieves a bandlimited signal.

In the present work, we aim at extending of the idea introduced in [10], [11], [12]. Instead of considering a bandlimited input signal, we simply assume that the input signal admits a sparse representation in a given domain [13]. The proposed method relies on the observation that the signal obtained after the nonlinear distortion is usually less sparse than the original one. Therefore, our idea to mitigate a nonlinear distortion is to adjust a compensating function so that the estimate of the input signal be as sparse as possible. This task is performed by a polynomial function, since it can provide a flexible enough approximation of the inverse of the distortion function. With respect to the compensation criterion, our proposal is implemented by formulating an optimization problem in which an approximation of the ℓ_0 -norm¹ is used as a measure of sparsity.

Another goal of this work is to test the proposed method in an actual application related to chemical analysis through sensor arrays. In such an application, compensation of nonlinear distortions is a relevant issue, especially in the case of potentiometric sensors such as the ion-selective electrode (ISE) [14], [7].

This article is an extended version of the conference paper [1] and is organized as follows. In Section II, we describe the problem of blind compensation of nonlinear distortions and by introducing the notations. Also in Section II, we discuss the problem of nonlinear compensation that arises in the context of ISE arrays. In Section III, we introduce the proposed framework and describe a practical algorithm to implement our approach. In Section IV, a theoretical analysis of our proposal is conducted. In Section V, we provide results considering synthetic data and also actual data acquired by an ISE array. In Section VI, we present our final conclusions.

II. BLIND COMPENSATION OF A NONLINEAR DISTORTION: PROBLEM STATEMENT AND NOTATION

The problem of blind compensation of a nonlinear distortion can be summarized in Figure 1. The goal is to estimate an one-dimensional signal $s(t)$ from the observation of the one-dimensional signal $x(t) = f(s(t))$, where $f(\cdot)$ is a function that models a nonlinear memoryless distortion. To accomplish this task, one must adjust a compensating function $g(\cdot)$ so that $y(t) = g(x(t))$ be as close as possible to

¹Strictly speaking, the ℓ_0 -norm is not a mathematical norm [13]. However, we keep this nomenclature since the term “ ℓ_0 -norm” is commonly used.

$s(t)$. The problem is said to be blind (or unsupervised) because one has neither training samples from $s(t)$ nor knowledge on $f(\cdot)$. The only information that is available is the observed signal $x(t)$.

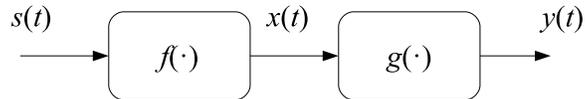


Fig. 1. The problem of blind compensation of a nonlinear distortion.

We assume that both $f(\cdot)$ and $g(\cdot)$ are monotonic functions with non-vanishing derivative. Otherwise, the problem becomes very difficult to solve, requiring strong prior information. Moreover, since we are dealing with a nonlinear blind problem, the best one can expect is to retrieve a scaled and shifted version of $s(t)$, that is, we accept estimations given by $\kappa s(t) + \nu$, where $\kappa \in \mathbb{R}^*$ and $\nu \in \mathbb{R}$.

Concerning the mathematical notation adopted in this paper, since our proposal is tailored to discrete-time one-dimensional signals, the input and observed signals will be represented by the vectors $\mathbf{s} \in \mathbb{R}^N$ and $\mathbf{x} \in \mathbb{R}^N$, respectively, where N is the sample size. Similarly, the estimated input signal will be represented by the vector $\mathbf{y} \in \mathbb{R}^N$. In our notation, the mappings between the input and observed signals and between the observed and input estimated signals will be denoted by $\mathbf{x} = f(\mathbf{s})$ and $\mathbf{y} = g(\mathbf{x})$, respectively. Therefore the i -th elements of \mathbf{x} and \mathbf{y} are given by $x_i = f(s_i)$ and $y_i = g(x_i)$, respectively. Finally, the global mapping $g \circ f$ between \mathbf{s} and \mathbf{y} is denoted by $q(\cdot)$.

In this work, we shall deal with signal transforms. These transforms will be represented through a matrix notation, i.e., the coefficients of a signal \mathbf{p} in a given transformed domain can be calculated as follows

$$\mathbf{p}_c = \mathbf{\Phi} \mathbf{p}, \quad (1)$$

where $\mathbf{\Phi}$ is the matrix associated with a given linear transform. In this work, while there are no constraints for $\mathbf{\Phi}$ with respect to the theoretical results, we consider that $\mathbf{\Phi}$ is related to the discrete cosine transform (DCT) in the examples and numerical experiments. Our choice is motivated by two reasons. The first one is that chemical signals, which is the primary focus of this work, are usually smooth and, thus, have only a few coefficients when represented in frequency-related transforms, such as DCT. As will be discussed later, this feature is a fundamental one since our proposal requires that the input signal be as sparse as possible. The second reason is that the coefficients of the DCT are real, unlike, for instance, the coefficients provided by the discrete Fourier transform (DFT). Such a feature is important since it renders easier the implementation of our approach.

A. Application to ion-selective electrodes

The problem of blind compensation of a nonlinear distortion is very important in chemical analysis. In particular, this issue is crucial when dealing with analyses performed by ion-selective electrode (ISE) arrays [15], [16]. To better understand this point, let us consider an analysis in which an array of N_m ISEs are used to estimate the activities of N_s ions. According to the Nicolsky-Eisenman (NE) equation [7], when the valences of the ions under analysis are the same, the response of the k -th ISE is given by

$$x_k(t) = e_k + d_k \log_{10}(s_k(t)), \quad (2)$$

where

$$s_k(t) = z_k(t) + \sum_{j=1, j \neq k}^{N_s} a_{kj} z_j(t) \quad (3)$$

In Equation (3), $z_k(t)$ and $z_j(t)$ correspond to the activities of the target and interfering ions, respectively. The parameters a_{kj} are known as selectivity coefficients and model the interference of j -th ion into the k -th one. The parameters e_i and d_i are unknown constants. The NE model can be seen as special case of the Wiener-Hammerstein systems. The difference is that the linear part of the NE model is not a dynamical system.

If a blind framework is considered, there are basically two approaches to deal with the nonlinear mixing process given by Equations (2) and (3). The first one is to set up a nonlinear blind source separation method that provides estimates of the signals $z_1(t), z_2(t), \dots, z_{N_s}(t)$ from the observations $x_1(t), x_2(t), \dots, x_{N_m}(t)$. This approach was considered in [15], [16].

A second approach is to consider a two-step framework in which the first stage consists in dealing with a set of N_m nonlinear distortions, each one observed in a given electrode. Once these nonlinear distortions are compensated, the mixing model given by Equation (3) can be tackled by linear blind source separation (BSS) methods [17], [18].

Note also that, even in the situation in which the effect of the mixing process is negligible, i.e. when there is only one ion within the solution under analysis, a blind estimation of the target ion activity still requires compensating the logarithmic distortion. To sum up, a fundamental step here is to estimate $s_k(t)$ from $x_k(t)$, which matches exactly the problem of compensating a nonlinear distortion.

III. A METHOD BASED ON SPARSITY RECOVERY TO BLINDLY COMPENSATE A NONLINEAR DISTORTION

A. The proposed approach

We shall introduce our idea by means of an example. Let us consider the input signal \mathbf{s} which is shown in Figure 2(a). In Figure 2(b), we plot a distorted version of the input signal given by $x_i = \sqrt[3]{s_i}$. The signal \mathbf{s} can be represented in a sparse manner after applying the DCT — this is shown in Figure 2(c).

In Figure 2(d), the DCT of the observed signal \mathbf{x} is exhibited. The key point here is that the DCT of \mathbf{x} is clearly less sparse than the one of the input signal \mathbf{s} — this feature, which is analogous to the spectral spreading phenomenon that arises when bandlimited signals are submitted to nonlinear functions [10], [11], will be further discussed in Section IV. Motivated by this observation, our proposal is to adjust the compensating function $g(\cdot)$ so that the estimated signal \mathbf{y} be as sparse as possible. We discuss in the sequel how this idea can be put into practice.

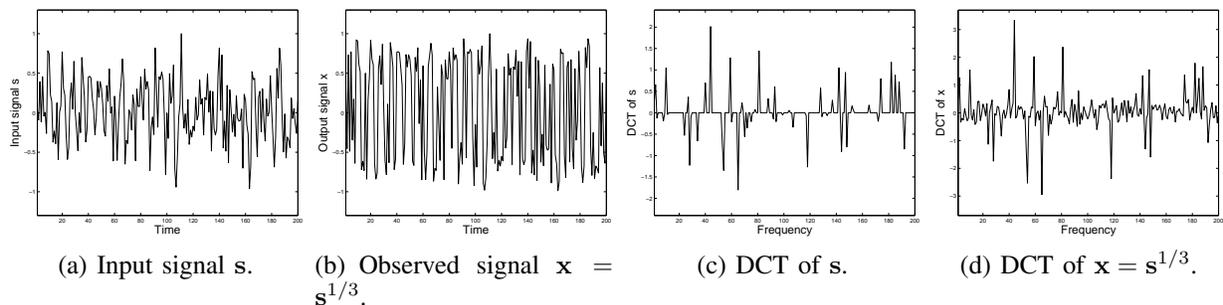


Fig. 2. Signals \mathbf{s} and \mathbf{x} and the DCT of their cubic root functions.

B. A practical method

In order to implement our proposal, we must define a parametric model for the compensating nonlinear function $g(\cdot)$. A crucial point related to this definition is the degree of flexibility of the chosen $g(\cdot)$. For instance, if one considers a compensating function based on a multilayer perceptron (MLP) neural network, then there arises the risk of overfitting [19]. In other words, due to its high degree of flexibility, the MLP model can provide a $\mathbf{y}_c = \Phi \mathbf{y}$ that is very sparse but it is not a good estimation of \mathbf{s} . Indeed, if a frequency-based transform is considered, the MLP network can be adapted so that \mathbf{y} is close to a constant signal, thus resulting in a very sparse signal in the transformed domain.

In view of the risk of overfitting, the parametric model for $g(\cdot)$ must be always monotonic. Moreover, one must consider a parametric model that is as flexible as possible, without violating the constraint of

strict monotonicity. In this work, we consider the use of a polynomial compensating function, which is given by

$$y_i = \sum_{j=1}^{N_p} w_j x_i^{2j-1}. \quad (4)$$

Monotonicity is ensured when the derivative of (4) is always positive or negative. In order to have a derivative that is always positive, only the odd terms are considered and a non-negativeness constraint must be imposed to the coefficients w_j . Equivalently, if a non-positiveness constraint is imposed to w_j , then (4) has negative derivative everywhere. In our work, we shall consider a non-negativeness constraint and, due to the scale ambiguity, there is no loss of generality in this choice.

By considering a vector notation, the estimated signal \mathbf{y} can be calculated as follows

$$\mathbf{y} = \mathbf{X}^T \mathbf{w}, \quad (5)$$

where

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ x_1^3 & x_2^3 & \cdots & x_N^3 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{2N_p-1} & x_2^{2N_p-1} & \cdots & x_N^{2N_p-1} \end{bmatrix}, \quad (6)$$

and

$$\mathbf{w} = [w_1 \ w_2 \ \dots \ w_{N_p}]^T. \quad (7)$$

Let us now turn our attention to the criterion that will be used to adjust the parameters \mathbf{w} in Equation (5). According to the idea introduced in Section III-A, the compensating function must be adjusted so as to maximize the sparsity of \mathbf{y}_c , which denotes the coefficients of the output signal \mathbf{y} in a given transformed domain generated by Φ , that is, $\mathbf{y}_c = \Phi \mathbf{y}$.

Several criteria can be applied to quantify sparsity. In this work, we shall consider the ℓ_0 -norm [13], which, for a vector $\mathbf{p} \in \mathbb{R}^N$, is given by

$$\|\mathbf{p}\|_0 = N - \sum_{i=1}^N \mathbb{1}_0(p_i), \quad (8)$$

where $\mathbb{1}_0(\cdot)$ corresponds to the indicator function on zero. Note that the ℓ_0 -norm simply counts the number of non-null elements of \mathbf{p} .

In a practical situation, sparse signals often present many coefficients that are close to zero, but are not exactly null. Such a fact restricts the direct application of the ℓ_0 -norm, thus requiring some sort

of approximation. In the present work, we make use of the smoothed version of the ℓ_0 -norm that was considered in [20]. This measure is given by:

$$S_{\ell_0}(\mathbf{p}) = N - \sum_{i=1}^N k(p_i, \sigma), \quad (9)$$

where $k(p_i, \sigma)$ is Gaussian kernel of zero mean and standard deviation σ , that is:

$$k(p_i, \sigma) = \exp\left(-\frac{p_i^2}{2\sigma^2}\right), \quad (10)$$

Note that:

$$\lim_{\sigma \rightarrow 0} S_{\ell_0}(\mathbf{p}) = \|\mathbf{p}\|_0. \quad (11)$$

Based on the elements described so far, the idea introduced in Section III-A can be implemented by formulating the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & S_{\ell_0}(\mathbf{y}_c) = S_{\ell_0}(\Phi \mathbf{X}^T \mathbf{w}) \\ \text{subject to} \quad & w_i \geq 0, \quad i = 1, \dots, N_p \\ & \|\mathbf{y}_c\|_2 = \mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} = 1. \end{aligned} \quad (12)$$

As discussed before, the inequality constraints are required to ensure the monotonicity of $g(\cdot)$. The equality constraint is necessary to avoid trivial solutions, in which all the elements of \mathbf{y}_c are null — this constraint corresponds to simply setting the energy of the estimated signal \mathbf{y} to one.

The optimization problem expressed in Equation (12) is an example of a nonlinear programming problem with equality and inequality constraints [21]. In this work, we consider an algorithm from the class of sequential quadratic programming (SQP) methods. More details on this classical optimization strategy can be found in Appendix A.

IV. THEORETICAL ASPECTS

Having defined a practical algorithm to implement our idea, let us turn our attention to an important theoretical issue: what are the conditions for which maximizing the sparsity of \mathbf{y}_c leads to a perfect compensation of the nonlinear distortion? Our analysis is based on the ℓ_0 -norm and can be summarized by the following theorem.

Theorem 1 (Perfect compensation by sparsity recovery): If the global mapping between the input sig-

nal \mathbf{s} and the estimated signal \mathbf{y} , $\mathbf{y} = q(\mathbf{s}) = g \circ f(\mathbf{s})$, can be represented by

$$\mathbf{y} = \beta\mathbf{s} + \gamma h(\mathbf{s}), \quad (13)$$

where $h(\cdot)$ is a nonlinear residual function, $\beta \in \mathbb{R}^*$ and $\gamma \in \mathbb{R}$, and if

$$\|\mathbf{s}_c\|_0 = \|\Phi\mathbf{s}\|_0 < \frac{\|\Phi h(\mathbf{s})\|_0}{2} \quad (14)$$

then,

$$\|\mathbf{y}_c\|_0 \geq \|\mathbf{s}_c\|_0,$$

being the equality achieved if and only if $\gamma = 0$ or, equivalently, if $\mathbf{y} = \beta\mathbf{s}$ (perfect compensation case).

Proof: The rationale behind the proof is to find the lower bound of $\|\mathbf{y}_c\|_0$ as a function of γ . Since the global mapping between \mathbf{s} and \mathbf{y} can be written according to Equation (13), \mathbf{y}_c is given by:

$$\mathbf{y}_c = \Phi\mathbf{y} = \Phi(\beta\mathbf{s} + \gamma h(\mathbf{s})) = \beta\mathbf{s}_c + \gamma\Phi h(\mathbf{s}). \quad (15)$$

Strictly speaking, the ℓ_0 -norm is not a mathematical norm, but it satisfies the triangle inequality, and, thus, the reverse triangle inequality [13]². Therefore, Equation (15) leads to:

$$\|\mathbf{y}_c\|_0 \geq \left| \|\mathbf{s}_c\|_0 - \|\gamma\Phi h(\mathbf{s})\|_0 \right|. \quad (16)$$

Equation (16) can be inspected in two situations: i) $\gamma \neq 0$ (situation in which there is still a nonlinear distortion) and ii) $\gamma = 0$ (case in which perfect compensation is achieved). In the first case, given that the ℓ_0 -norm is not homogeneous, which means that $\|\gamma\Phi h(\mathbf{s})\|_0 = \|\Phi h(\mathbf{s})\|_0$, Equation (16) becomes:

$$\|\mathbf{y}_c\|_0 \geq \left| \|\mathbf{s}_c\|_0 - \|\Phi h(\mathbf{s})\|_0 \right|. \quad (17)$$

If the condition (14) holds, then one can easily check from Equation (17) that

$$\|\mathbf{y}_c\|_0 > \|\mathbf{s}_c\|_0. \quad (18)$$

On the other hand, in the case in which $\gamma = 0$, it asserts, from Equation (13), that $\mathbf{y}_c = \beta\mathbf{s}_c$, and, thus, $\|\mathbf{y}_c\|_0$ attains its minimum value, which is given by $\|\mathbf{s}_c\|_0$. ■

²Given two vectors \mathbf{a} and \mathbf{b} , the reverse triangle inequality states that:

$$\|\mathbf{a} - \mathbf{b}\|_0 \geq \left| \|\mathbf{a}\|_0 - \|\mathbf{b}\|_0 \right|.$$

Theorem 1 provides two joint sufficient conditions that guarantee that the minimization of $\|\mathbf{y}_c\|_0$ leads to a perfect nonlinear compensation case. The first condition Equation (13) is not a very restrictive one, since many functions can be represented in this way by means, for instance, of a Maclaurin series [22]. A counterexample in this respect would be an even distorting function. Indeed, the Maclaurin series of an even function has only even powers and, thus, cannot be represented by (Equation (13)). Note that in this situation, $f(\cdot)$ would not be a monotonic function, which, as mentioned before, poses a tricky problem of compensation.

The second condition, expressed by Equation (14), requires that the transform of the nonlinear function $h(\mathbf{s})$ be at least half less sparse than the transform of the input signal \mathbf{s} . This result somehow extends the original work of Landau and Miranker [10] and also the more recent results provided in [11], [15]. Indeed, in these works, one assumes that the signal is bandlimited, that is, there is a sort of block sparsity representation in which the first elements of the vector \mathbf{s}_c are not null. Then, since the nonlinear distorting function spreads the spectrum of the observed signal, it is possible to perform compensation by retrieving a signal that is as bandlimited as possible. Conversely, Theorem 1, although relying on an analogous idea, does not require this block sparsity structure (bandlimited signals) that is necessary in the aforementioned works; it requires that the retrieved signal be as sparse as possible in a conjunction with the nonlinear distortion, so that Equation (14) is satisfied.

In order to illustrate the result stated in Theorem 1, let us consider an example in which the distorting function is given by

$$x_i = f(s_i) = \sqrt[3]{s_i},$$

and the compensating function by

$$y_i = g(x_i) = \beta x_i^3 + \gamma x_i.$$

Therefore,

$$y_i = q(x_i) = \beta s_i + \gamma \sqrt[3]{s_i},$$

and, thus, according to Equation (13), the residual function is given by

$$h(s_i) = \sqrt[3]{s_i}.$$

A crucial point here is that if we consider, for instance, a frequency transform such as the DCT, and if the input signal \mathbf{s} is sparse in this domain, then the signal $h(\mathbf{s})$ will tend to be much less sparse than \mathbf{s} , thus satisfying Equation (14). This result was already illustrated in the Example discussed in Section III-A

(see Figure 2).

V. RESULTS

In order to test the proposed method, we provide experiments with synthetic and actual data in Sections V-A and V-B, respectively.

The performance index adopted in our experiments is the signal-to-distortion ratio (SDR) in dB, which is defined as:

$$\text{SDR} = 10 \log \left(\frac{\mathbf{s}^T \mathbf{s}}{(\mathbf{s} - \mathbf{y}^*)^T (\mathbf{s} - \mathbf{y}^*)} \right),$$

where \mathbf{y}^* denotes the recovered signal after performing a normalization with respect to \mathbf{s} . This normalization, which is given by $\mathbf{y}^* = k\mathbf{y}$, where $k = \arg \min_{k^*} (\mathbf{s} - k^*\mathbf{y})^T (\mathbf{s} - k^*\mathbf{y})$, is necessary since the best one can expect in a blind situation is an estimation given by $\mathbf{y} = \beta\mathbf{s}$ (scale ambiguity).

In some experiments, we compare our proposal with a minimum mean-squared error (MMSE) approach, in which the coefficients of polynomial compensator (4) are adjusted by solving the following optimization problem

$$\min_{\mathbf{w}} ((\mathbf{s} - \mathbf{X}^T \mathbf{w})^T (\mathbf{s} - \mathbf{X}^T \mathbf{w})). \quad (19)$$

Note that such a solution is a supervised one since it assumes that samples of the input signal are available. Therefore, the MMSE solution can be seen as a performance bound for our proposal.

A. Experiments with synthetic data

In the experiments with synthetic data, the input signal \mathbf{s} was generated as follows:

$$\mathbf{s} = \mathbf{\Phi}^T \mathbf{s}_c = \mathbf{\Phi}^T (\mathbf{s}_{c_s} + \alpha \mathbf{r}), \quad (20)$$

where \mathbf{s}_{c_s} correspond to realizations of a Bernoulli-Gaussian process, i.e., each element of \mathbf{s}_{c_s} has a probability P of being non-zero, and, when active, it is obtained from a zero-mean Gaussian distribution of unity variance. The elements of \mathbf{r} are also obtained from a standard Gaussian distribution, and the rationale of this vector, which is weighted by the coefficient α , is to model the small residual energy typical of practical sparse signals. Therefore, when $\alpha = 0$ and P is small, \mathbf{s}_c becomes a sparse signal in the sense of the ℓ_0 -norm. Finally, in our tests, we consider that $\mathbf{\Phi}$ corresponds to the DCT matrix for the reasons exposed in Section II.

In the sequel, we shall discuss four experiments, each one carried out in a different scenario.

1) *Case in which perfect inversion is possible:* We first study the situation in which it is possible to achieve a perfect compensation of the nonlinear distortion. In particular, we address the compensation of $f(s_i) = \sqrt[3]{s_i}$ via the polynomial function (4) with $N_p = 4$. The sparse signal was generated according to (20), considering $P = 0.2$, $\alpha = 0$ (ideal sparse signal), and $N = 1000$ samples. Figure 3(a) shows the mapping between s_i and x_i .

After solving the optimization problem expressed in (12) (with $\sigma = 0.01$), we obtained a compensating function that achieves an almost perfect inversion (SDR= 57.7dB); this can be seen in Figure 3(b), which exhibits the mapping between s_i and y_i . Note, also in Figure 3(b), that there is a scale indeterminacy, since the obtained line does not have an angular coefficient equal to one.

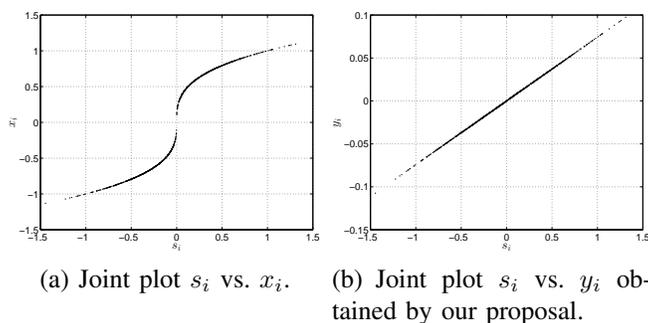


Fig. 3. Nonlinear compensation when a perfect inversion is possible.

2) *Case in which perfect inversion is not possible:* Let us now consider a more realistic experiment in which $f(s_i) = \tanh(3s_i)$; $N = 1000$, $P = 0.2$ and $\alpha = 0$. In Figure 4(a), the nonlinear distortion is illustrated by the resulting mapping between s_i and x_i .

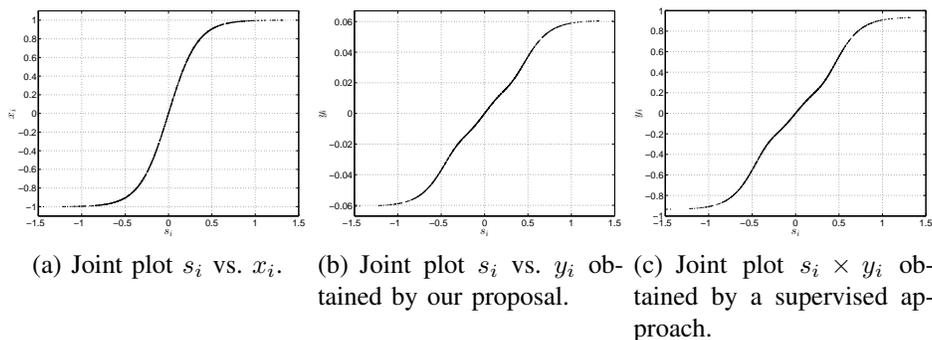


Fig. 4. Case in which a perfect inversion is not possible.

The experiment in this case was performed considering a polynomial of degree 9 ($N_p = 5$) and $\sigma = 0.1$; we noticed that very similar results are obtained for values of σ in the range $[0.001, 0.1]$. As

can be observed in Figure 4(b), our method was able to mitigate the effects of the nonlinear distortion; this is specially clear for high amplitudes. Concerning the signal-to-distortion ratio, we obtained a value of $\text{SDR} = 17.1\text{dB}$. For matter of comparison, the MMSE supervised solution (see Equation (19)) achieved a value of $\text{SDR} = 17.4$ — the resulting input signal-estimated signal mapping in this case is shown in Figure 4(c). Note that the proposed method was able to provide a solution very close to the MMSE solution.

3) *Influence of residual elements:* Let us analyze the case in which $\alpha \neq 0$ in the signal model expressed in (20). This situation simulates actual sparse signals, which usually have small coefficients that are not necessarily zero. In our analysis, we performed a set of simulations in which the parameter α was varied from 0 to 1. In these tests, we considered $N_p = 4$ and the following different bandwidths: $\sigma = 0.1$ and $\sigma = 0.01$. The nonlinear distortion is $f(s_i) = \tanh(1.5s_i)$ and $N = 1000$ samples of the input signal were generated considering $P = 0.2$ — in this case, we performed a variance normalization of the input signal (this signal was divided by its standard deviation) to obtain fair comparisons between different values of α .

As can be seen in Figure 5, which shows the average SDR obtained considering 100 trials for each α , the performance of the proposed approach becomes worse as α increases. This is expected, since the input signal tends to be less sparse as α increases, violating the basic assumption underlying our proposal. Another interesting point here is the influence of σ . A higher value of σ , in this case, leads to a better performance as α increases. Conversely, for α close to zero, one can note that defining a smaller σ leads to a better performance, very close, in fact, to that associated with the supervised solution. Finally, note that, as expected, the performance of the supervised solution does not depend on the degree of sparsity of the input signal.

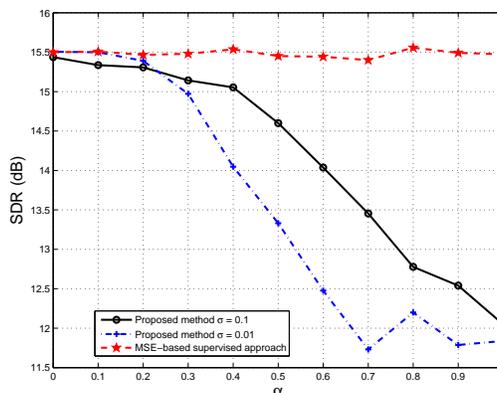


Fig. 5. Influence of α on the SDR obtained by the proposed method and the MMSE solution.

4) *Performance in the presence of noise:* Let us now consider a noisy situation in which the observed signal is given by:

$$\mathbf{x} = f(\mathbf{s}) + \mathbf{n}, \quad (21)$$

where \mathbf{n} are samples of an additive white Gaussian noise (AWGN).

We consider the same scenario of Experiment 2 ($N = 1000$, $P = 0.2$, $\alpha = 0$, $N_p = 5$ and nonlinear function $f(s_i) = \tanh(3s_i)$). A set of experiments were carried out for different signal-to-noise ratios (SNR). The evolution of the obtained SDR against the SNR is shown in Figure 6 — each point corresponds to the average obtained in 100 trials. A first point that should be stressed here is that our proposal (with both $\sigma = 0.1$ and $\sigma = 0.01$) provided SDRs very close to the ones obtained by the MSE-based supervised solution. It is also interesting to stress the effects of the noise, which in nonlinear systems can be very harmful. In this example, for instance, we note a rapid performance degradation for SNRs smaller than approximately 30dB, even for the supervised MMSE solution.

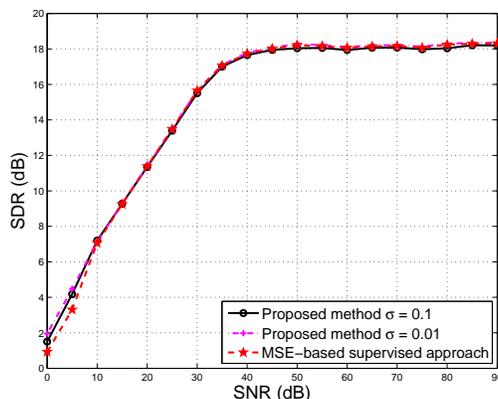


Fig. 6. Influence of noise on the SDR obtained by the proposed method and the MMSE solution.

B. Experiments with data provided by an ion-selective electrode

We here test our proposal in a real scenario. The data comes from the experiments $\text{S1K10}^{-1}\text{NH}_4$ and $\text{S1K10}^{-4}\text{NH}_4$ of the Ion-Selective Electrode Array (ISEA) dataset (see [23] for details). In this case, we are interested in the response of an ISE whose target is the ion potassium (K^+). However, in the solution under analysis, there is, besides the ion K^+ , the ion ammonium (NH_4^+), which can be seen as an interfering ion. The response of the K^+ -ISE can be modeled by Equations (2) and (3) and our goal here is to provide an estimate for $s_1(t)$ from the observation $x_1(t)$ — the index 1 means here that $x_1(t)$ corresponds to the response of the first sensor within the array.

In the ISEA database, the actual ionic activities of potassium and ammonium are known — these are, respectively, the signals $z_1(t)$ and $z_2(t)$ in Equations (2) and (3). Therefore, we can, by applying a nonlinear regression, estimate the signal $s_1(t)$, which corresponds to the desired signal in our analysis. This signal is depicted in Figure 8 (in blue), which also show the observed signal $x_1(t)$ (in gray). The number of samples in this case is $N = 169$.

In order to obtain a sparse representation of the input signal $s_1(t)$, we consider the DCT transform. In Figure 7(a), we show the DCT of the input signal, which is indeed very sparse. On the other hand, the DCT of the observed signal $x_1(t)$ is less sparse, as can be seen in Figure 7(b).

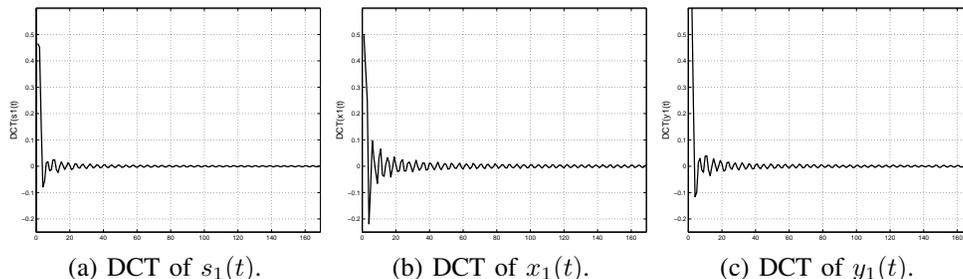


Fig. 7. DCTs of input, observed and estimated signals.

We applied our proposal with $N_p = 20$ and $\sigma = 0.05$, having obtained a good estimate of the signal $s_1(t)$, as can be seen in Figure 8 (signal in black). In Figure 7(c), the DCT of $y_1(t)$, the estimate of the input signal. Note that the algorithm provided a sparser signal with respect to the observed signal. The signal-to-distortion rate in this case was $\text{SDR} = 28.4\text{dB}$. For the sake of comparison, we calculate the SDR between the observed signal and the desired one — this is a situation in which no compensation strategy is considered. The obtained value was $\text{SDR} = 7.1\text{dB}$.

VI. CONCLUSIONS

In this work, we proposed a new approach to the problem of blind compensation of nonlinear distortions. The proposal relies on the assumption that the desired signal can be sparsely represented in a given (known) transform. We performed a set of simulations considering synthetic data in different situations, and it was observed that the new method was able to provide solutions close to those generated with an MMSE-based supervised approach.

We also tested our method in a real problem: that of nonlinear compensation in an ion-selective electrode array. The obtained results showed that our approach provided good estimates of the input signal despite the strong nonlinear distortion. In the context of ISE arrays, the use of the proposed idea

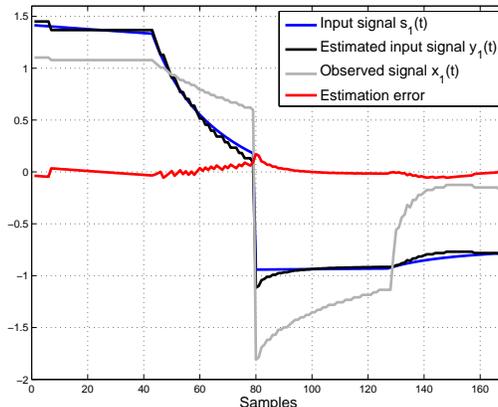


Fig. 8. Experiment with real data: input signal s_1 (blue), observed signal x_1 (gray), retrieved signal y_1 (black) and estimation error $s_1 - y_1$ (red).

can open the way for the application of linear blind source separation methods in the same way as done in [12].

There are still several points that deserve future investigation. For instance, it would be quite helpful to extend the theoretical result described in Section IV so as to consider the smoothed version of the ℓ_0 -norm [20] — this would clarify what are the limitations in using such an approximation. We also intend to study the use of more flexible structures (such as splines and monotonic neural networks) to be used in the compensating device. An extension to more general Wiener-Hammstein systems, in which the linear stage has memory, is also envisaged. Finally, we also intend to apply our method in other actual problems, e.g., to compensate nonlinear distortions introduced by amplifying stages that follow transduction.

APPENDIX A

SEQUENTIAL QUADRATIC PROGRAMMING

The optimization problem expressed in Equation (12) is a nonlinear programming problem with equality and inequality constraints. This problem can be tackled by an approach known as sequential quadratic programming (SQP). In the sequel, we provide an overview on this classical optimization strategy, and we refer the reader to [21], [24] for further details.

According to the Karush-Kuhn-Tucker (KKT) conditions [24], a set composed of a local minimizer

and Lagrange multipliers, $\{\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}\}$, for the problem (12) should satisfy:

$$\begin{aligned}
\nabla_{\mathbf{w}}L(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \mathbf{0} \\
\mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} &= 1 \\
w_i &\geq 0, \quad i = 1, \dots, N_p \\
\boldsymbol{\mu} &\geq \mathbf{0} \\
\mu_i w_i &= 0, \quad i = 1, \dots, N_p
\end{aligned} \tag{22}$$

where $L(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ corresponds to the Lagrangian function and is given by:

$$L(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = S_{\ell_0}(\boldsymbol{\Phi} \mathbf{X}^T \mathbf{w}) - \lambda_1 \mathbf{w}^T \mathbf{X} \mathbf{X}^T \mathbf{w} - \sum_{i=1}^{N_p} \mu_i w_i. \tag{23}$$

A first step in SQP is to define a initial point for $\{\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}\}$. Then, one searches for an increment $\{\boldsymbol{\delta}_{\mathbf{w}}, \boldsymbol{\delta}_{\boldsymbol{\lambda}}, \boldsymbol{\delta}_{\boldsymbol{\mu}}\}$. This search is based on a quadratic approximation of the Lagrangian function, which leads to a quadratic programming (QP) problem, for which there are very efficient solutions. After solving this QP problem, one ends up with a new increment $\{\boldsymbol{\delta}_{\mathbf{w}}, \boldsymbol{\delta}_{\boldsymbol{\lambda}}, \boldsymbol{\delta}_{\boldsymbol{\mu}}\}$ and consequently with a new candidate solution given by $\{\mathbf{w} + \boldsymbol{\delta}_{\mathbf{w}}, \boldsymbol{\lambda} + \boldsymbol{\delta}_{\boldsymbol{\lambda}}, \boldsymbol{\mu} + \boldsymbol{\delta}_{\boldsymbol{\mu}}\}$. Such a procedure can be conducted in an iterative fashion until either the increments are smaller than a pre-defined value or a maximum number of iterations is achieved.

REFERENCES

- [1] L. T. Duarte, R. Suyama R. Attux, J. M. T. Romano, and C. Jutten, "Blind compensation of nonlinear distortions via sparsity recovery," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012)*. IEEE, 2012, pp. 2362–2366.
- [2] J. Tsimbinos and K.V. 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.
- [3] T. T. Ha, *Digital satellite communications*, McGraw-Hill, 1990.
- [4] C.-T. Tan, B. C. J. Moore, and N. Zacharov, "The effect of nonlinear distortion on the perceived quality of music and speech signals," *Journal of the Audio Engineering Society*, vol. 51, no. 11, pp. 1012–1031, 2003.
- [5] W. Klippel, "Compensation for nonlinear distortion of horn loudspeakers by digital signal processing," *Journal of the Audio Engineering Society*, vol. 44, no. 11, pp. 964–972, 1996.
- [6] K. V. Asari, S. Kumar, and D. Radhakrishnan, "A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation," *IEEE Transactions on Medical Imaging*, vol. 18, no. 4, pp. 345–354, 1999.
- [7] P. Gründler, *Chemical sensors: an introduction for scientists and engineers*, Springer, 2007.
- [8] M. Ibnkahla, "Applications of neural networks to digital communications: a survey," *Signal Processing*, vol. 80, pp. 1185–1215, 2000.

- [9] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, 2006.
- [10] H. J. Landau and W. L. Miranker, "The recovery of distorted band-limited signals," *Journal of Mathematical Analysis and Applications*, vol. 2, pp. 97–104, 1961.
- [11] K. Dogancay, "Blind compensation of nonlinear distortion for bandlimited signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 9, pp. 1872–1882, 2005.
- [12] L. T. Duarte, R. Suyama, B. Rivet, R. R. F. Attux, J. M. T. Romano, and C. Jutten, "Blind compensation of nonlinear distortions: Application to source separation of post-nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5832–5844, 2012.
- [13] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer, 2010.
- [14] E. Bakker and E. Pretsch, "Modern potentiometry," *Angewandte Chemie International Edition*, vol. 46, pp. 5660–5668, 2007.
- [15] L. T. Duarte, C. Jutten, and S. Moussaoui, "A bayesian nonlinear source separation method for smart ion-selective electrode arrays," *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1763–1771, 2009.
- [16] L. T. Duarte, J. M. T. Romano, C. Jutten, K. Y. Chumbimuni-Torres, and L.T. Kubota, "Application of blind source separation methods to ion-selective electrode arrays in flow-injection analysis," *IEEE Sensors Journal*, vol. 14, no. 7, pp. 2228–2229, 2014.
- [17] P. Comon and C. Jutten, Eds., *Handbook of blind source separation: independent component analysis and applications*, Academic Press, 2010.
- [18] 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.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [20] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Transactions on Signal Processing*, vol. 57, pp. 289–301, 2009.
- [21] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, 1984.
- [22] E. Kreyszig, *Advanced Engineering Mathematics*, John Wiley & Sons, 2006.
- [23] 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 Sensors Journal*, vol. 10, no. 12, pp. 1891–1892, 2010.
- [24] A. Antoniou and W.-S. Lu, *Practical optimization*, Springer, 2007.