Blind extraction of sparse components based on L0-norm minimization
Leonardo Tomazeli Duarte, Ricardo Suyama, Romis Attux, Joao Romano, Christian Jutten

To cite this version:


HAL Id: hal-00605697
https://hal.archives-ouvertes.fr/hal-00605697
Submitted on 4 Jul 2011

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.
BLIND EXTRACTION OF SPARSE COMPONENTS BASED ON $\ell_0$-NORM MINIMIZATION

Leonardo T. Duarte$^{1*}$, Ricardo Suyama$^2$, Romis Attux$^1$, João M. T. Romano$^1$, Christian Jutten$^{3†}$

1 DSPCom Lab, University of Campinas (UNICAMP), Brazil
2 Engineering, Modeling and Applied Social Sciences Center (CECS), UFABC, Santo André, Brazil
3 GIPSA-lab, UMR CNRS 5216, Université Joseph Fourier, France
{ltduarte,romano@dmo.fee.unicamp.br, ricardo.suyama@ufabc.edu.br, attux@dca.fee.unicamp.br, christian.jutten@gipsa-lab.grenoble-inp.fr}

ABSTRACT

We investigate the application of cost functions based on the $\ell_0$-norm to the problem of blind source extraction (BSE). We show that if the sources have different levels of sparsity, then the minimization of the $\ell_0$-norm leads to the extraction of the sparsest component even when the sources are statistically dependent. We also study the conditions guaranteeing BSE when an approximation of the $\ell_0$-norm is considered. Finally, we provide a numerical example to illustrate the applicability of our proposal.

Index Terms— Blind source extraction, sparse components, $\ell_0$-norm.

1. INTRODUCTION

In blind source separation (BSS) [1], the goal is to recover a set of signals (sources) based only on the observation of mixed versions of these original sources. A closely related problem is that of blind source extraction (BSE), in which one is only interested in recovering a single source. Evidently, the BSS problem can be tackled by performing several runnings of a given BSE method so that, after a given running, the contribution of the extracted source is removed from the mixtures, e.g. via a deflation procedure [2, 3].

Most of BSE methods assume that the sources are mutually statistically independent. In this case, BSE can be performed by optimizing independent component analysis (ICA) criteria such as the ones based on cumulants [1]. More recently, some works have proposed criteria specially adapted to the case in which the sources are sparse. In [4], for instance, a nonconvex sparsity measure is adopted as cost function. Despite the good results obtained by this approach, there is no theoretical results indicating the conditions for which such an approach can be used. On the other hand, in [5], it is shown that a criterion based on the $\ell_1$-norm is indeed a contrast function [1] and, thus, can be used to perform BSE. However, [5] assumes that the sources are disjoint orthogonal, i.e., at most one source is not null at a given instant.

In the present work, we introduce a novel framework for the extraction of the sparsest source based on the minimization of the $\ell_0$-norm. We derive the conditions for which the $\ell_0$-norm is a contrast function. As will be discussed latter, these conditions are solely related to the degree of sparsity of the source. As a consequence, the proposed framework is sound even when the sources are not statistically independent or disjoint orthogonal. As a second contribution, we study an extension in which an approximation of the $\ell_0$-norm is used. Our motivation here comes from the fact that, in practical scenarios, sparse signals usually present many elements that are very close to zero although not null. In these cases, thus, one must resort to approximations of the $\ell_0$-norm.

The paper is organized as follows. In Section 2, we introduce the notation and the problem treated in this work. Section 3 presents the results concerning the use of the $\ell_0$-norm as a contrast function. In Section 4, a numerical example is conducted to illustrate the interest behind the proposed framework. Finally, in Section 5, we expose our conclusions.

2. PRELIMINARY OBSERVATIONS

Let the $i$-th source$^1$ be represented by the vector $s_i \in \mathbb{R}^{N_s \times 1}$ ($N_d$ is the number of samples) and the ensemble of $N_s$ sources by the matrix

$$
S = \begin{bmatrix}
    s_1^T \\
    \vdots \\
    s_{N_s}^T
\end{bmatrix} = \begin{bmatrix}
    s_{1}(1) & s_{1}(2) & \ldots & s_{1}(N_d) \\
    \vdots & \vdots & \ddots & \vdots \\
    s_{N_s}(1) & s_{N_s}(2) & \ldots & s_{N_s}(N_d)
\end{bmatrix}.
$$

Analogously, matrix $X$ represents the mixtures, which are assumed to be linear and instantaneous, that is, $X = AS$, where $A \in \mathbb{R}^{N_m \times N_s}$ denotes the mixing matrix. In this work, we consider that the number of mixtures is equal to the number of sources ($N_m = N_s$) and that $A$ is a full-rank matrix.

In BSE, the goal is to estimate a single source by adapting an extracting vector (represented by $w \in \mathbb{R}^{N_m \times 1}$) such that

---

$^*$L. T. Duarte thanks FAPESP for funding his post-doctoral research.

$†$C. Jutten is also with Institut Universitaire de France.

$^1$The results developed in this work are valid for signals represented in time or any other transformed domain.
the extracted source is given by
\[ y^T = [y(1) \ldots y(N_d)]^T = w^T X = w^T A S = g^T S, \quad (1) \]
where the vector \( g \in \mathbb{R}^{N_x \times 1} \) is the combined response extracting vector-mixing matrix. In most of ICA-based BSE approaches, there are scaling and permutation ambiguities [1], that is, the best one can expect is to extract a scaled version of any of the sources, without any particular order.

**Properties of the \( \ell_0 \)-norm.** The \( \ell_0 \)-norm, which corresponds to the number of non-zero elements of a vector, has been applied to quantify the sparsity of a signal [6]. It satisfies the triangle inequality and, thus, the reverse triangle inequality, i.e., given two vectors \( a \) and \( b \), then \( ||a - b||_0 \geq ||a||_0 - ||b||_0 \). On the other hand, the \( \ell_0 \)-norm is not positive homogeneous since \( ||ka||_0 = ||a||_0 \neq |k||a||_0 \) for \( k \neq 0 \). Therefore, although the term “\( \ell_0 \)-norm” is commonly used, this measure is not a true norm.

### 3. BUILDING A CONTRAST FUNCTION BASED ON SPARSITY

A BSE contrast [1] is a cost function in \( w \) whose global optima (minima in our case) are associated with a perfect source extraction. That said, we address the following question: for which conditions \( ||y||_0 \) is not a true norm? We assume, without loss of generality, that \( ||s_1||_0 \leq ||s_2||_0 \leq \ldots \leq ||s_{N_s}||_0 \).

We first address the case with two sources. In this situation, the following theorem can be derived.

**Theorem 1 (Case of binary mixtures)** Let \( w = [w_1 \ w_2]^T \) represent an extracting vector such that \( y^T = w^T X \). If at least one element of \( w \) is not null and \( ||s_1||_0 < \frac{1}{2} ||s_2||_0 \), then \( ||y||_0 \geq ||s_1||_0 \), being the equality achieved if and only if \( g^T = w^T A = [\alpha 0] \), where \( \alpha \in \mathbb{R}^* \).

**Proof** The extracted signal is given by
\[ y^T = w^T X = w^T A S = g^T S = g_1 s_1^T + g_2 s_2^T. \quad (2) \]

The theorem can be proved by analyzing the following cases: 1) \( g_1 = 0 \) and \( g_2 = 0 \). This situation is impossible to happen as, by assumption, \( w \) is not the null vector and \( A \) is a full-rank matrix. 2) \( g_1 = 0 \) and \( g_2 \neq 0 \). From the \( \ell_0 \)-norm invariance to scaling, \( ||y||_0 = ||g_2 s_2||_0 = ||s_2||_0 > ||s_1||_0 \). 3) \( g_1 \neq 0 \) and \( g_2 = 0 \). In this case, \( ||y||_0 = ||g_1 s_1||_0 = ||s_1||_0 \). 4) \( g_1 \neq 0 \) and \( g_2 \neq 0 \). From the reverse triangle inequality
\[ ||y||_0 \geq ||g_1 s_1||_0 - ||g_2 s_2||_0. \quad (3) \]

Since \( g_1 \neq 0 \) and \( g_2 \neq 0 \) in this case, and due to the \( \ell_0 \)-norm invariance to scaling, (3) can be rewritten as
\[ ||y||_0 \geq ||s_1||_0 - ||s_2||_0. \quad (4) \]

By assumption, \( ||s_1||_0 < \frac{1}{2} ||s_2||_0 \) and therefore, from (4), it asserts that \( ||y||_0 > ||s_1||_0 \).

According to Theorem 1, it is possible to extract the sparest component in a binary mixture by minimizing the \( \ell_0 \)-norm. Note that there is no permutation ambiguity as the source having the smallest \( \ell_0 \)-norm will be extracted first. It is also worth noting that Theorem 1 does not rely on joint properties such as statistical independence or correlation. In other words, when the sparsity of the sparest source is half of the other, BSE becomes possible even if the sources are dependent.

**Theorem 1 provides a sufficient but not necessary condition for the \( \ell_0 \)-norm to act like a contrast; it is in fact based on a worst-case condition. For instance, consider the class of disjoint orthogonal sources, for which the following property holds: \( ||s_1 \otimes s_2||_0 = 0 \), where \( \otimes \) stands for the element-wise multiplication. In this case, it asserts that \( ||y||_0 = ||g_1 s_1||_0 + ||g_2 s_2||_0 \). Thus, since \( g \) cannot be the null vector, \( ||y||_0 \geq ||s_1||_0 \), being the equality achieved if and only if \( g^T = w^T A = [\alpha 0] \). Therefore, in this case, \( s_1 \) can be extracted by minimizing the \( \ell_0 \)-norm no matter the values of \( ||s_1||_0 \) and \( ||s_2||_0 \). Actually, they should only be different. Considering now the case of \( N_s \) sources, we present a more general version of Theorem 1.

**Theorem 2 (General case)** Let \( w = [w_1 \ldots w_{N_s}]^T \) represent an extracting vector such \( y^T = w^T X \). If at least one element of \( w \) is non zero and
\[ ||s_1||_0 < \frac{1}{2} ||s_2||_0 \]
\[ ||s_1||_0 < \frac{1}{2} (||s_1||_0 - ||s_2||_0) \]
\[ ||s_1||_0 < \frac{1}{2} (||s_3||_0 - \cdots - ||s_{N_s-1}||_0) \]
\[ ||s_1||_0 < \frac{1}{2} \left( ||s_{N_s}||_0 - \sum_{i=2}^{N_s-1} ||s_{i}||_0 \right) \]
then \( ||y||_0 \geq ||s_1||_0 \), being the equality achieved if and only if \( g^T = w^T A = [\alpha 0 \ldots 0] \), where \( \alpha \in \mathbb{R}^* \).

**Sketch of the proof** Although the theorem is provable using the triangle inequalities for the \( \ell_0 \)-norm, we will present here a simpler line of reasoning based on the notion of mathematical induction. Starting from Theorem 1, let us deal, at first, with the case of \( N_s = 3 \) sources.

In order that the extraction of \( s_1 \) be the solution with the smallest possible \( \ell_0 \)-norm, the number of zeros generated from any linear combination including \( s_3 \) must be greater than the number of zeros generated by a solution in which \( g_2 = g_3 = 0 \) and \( g_1 \neq 0 \). If \( g_3 = 0 \), such a requirement is already covered by Theorem 1, which is exactly the first equation in (5).

If \( g_3 \neq 0 \), the least attainable value for \( ||y||_0 \) will be obtained in a situation in which all non-null elements of the other sources are used to cancel non-null elements present in \( s_3 \). This situation, which requires that, for each time instant \( n \) corresponding to a non-null element in \( s_3 \), either \( s_1(n) \) or \( s_2(n) \) is non-null, leads to a case in which the number of zeros in \( g(n) \) can be made equal to \( Z_1 = ||s_1||_0 + ||s_2||_0 + \)
$N_d - ||s_2||_0$, i.e., all non-null elements of $s_1$ and $s_2$ are used to cancel out a non-null element of $s_3$, thus increasing the number of zeros in $y$ (notice that $N_d - ||s_3||_0$ represents the number of zeros shared by the three sources). On the other hand, the number of zeros generated by the solution of interest ($g_2 = g_3 = 0$ and $g_1 \neq 0$) is simply $N_d - ||s_1||_0$.

Therefore, a valid contrast requires that:

$$||s_1||_0 + ||s_2||_0 + N_d - ||s_3||_0 < N_d - ||s_1||_0,$$

or that $||s_1||_0 < \frac{1}{2}(||s_3||_0 - ||s_2||_0)$.

The same considerations hold for cases with $N_s > 3$ as well: in order that we have a valid contrast, the sum of the $\ell_0$-norms of the $N_s - 1$ sparsest sources plus the number of zeros in common cannot exceed the number of zeros in $s_1$, as indicated in Theorem 2. This concludes the argument. \[\square\]

According to Theorem 2, the $\ell_0$-norm is a contrast when there is enough difference between the sparsities of the sources. The tighter situation, i.e. the one in which the differences between the sources’ $\ell_0$-norms are smaller, is attained when the following conditions holds: $||s_i||_0 = 2^{-2} \times (2||s_1||_0 + 1)$. For instance, if $||s_1||_0 = 100$, then such a limit is achieved for $||s_2||_0 = 201$, $||s_3||_0 = 402$, $||s_4||_0 = 804$, and so forth.

### 3.1. Extension to approximations of the $\ell_0$-norm

Despite the interesting aspects of the results presented so far, it is important to analyze them in more realistic situations. In fact, the application of the $\ell_0$-norm in a BSE context is limited, as sparse signals typically have many elements with very low energy, but not necessarily null. In view of this limitation, we here study an approximation of the $\ell_0$-norm that takes into account a small amount of modeling noise. Approximations of the $\ell_0$-norm has been already exploited in other problems such as overcomplete signal representation [7] and linear classification [8]. Our analysis addresses the case of two sources, and we consider the following approximation of the $\ell_0$-norm:

$$||y||_0^{(c)} = N_d - \sum_{i=1}^{N_d} E_c(y_i),$$

where $E_c(y_i)$ is the indicator function on the set $E_c = \{y_i | y_i \in \mathbb{R}, |y_i| \leq \epsilon\}$. Therefore, $||y||_0^{(c)}$ counts the number of elements of $y$ whose absolute value is greater than $\epsilon$

We assume the following model for the sources

$$s_i(n) = \{s | s \in \mathbb{R}, |s| \leq \delta, |s| \geq \gamma\}. \tag{7}$$

The parameter $\delta$ is an upper bound for the absolute values of the small elements that are typical of sparse signal whereas $\gamma$ models the elements of higher energy. We also assume that $\gamma > \epsilon \geq \delta$. Based on this model, the following robust version of Theorem 1 is stated.

**Theorem 3** (Case of binary mixtures · robust version)
Let $y^T = g^T S$ and consider that the sources follow the model (7). For every $g_1$ and $g_2$ satisfying

$$\begin{align*}
|g_1| \delta + |g_2| \delta &\leq \epsilon \tag{8} \\
|g_1| \gamma + |g_2| \gamma &> \epsilon \tag{9}
\end{align*}$$

if $||s_1||_0^{(c)} < \frac{1}{2}||s_2||_0^{(c)}$, then $||y||_0^{(c)} \geq ||s_1||_0^{(c)}$, being the equality achieved if and only if $g^T = [\alpha \ 0]$, where $\alpha \in \{g \ | \ g|g| \leq \epsilon, \gamma |g| > \epsilon\}$.

**Proof** Note that if the vector $g^T$ is constrained to intervals within which the measure (6) is scaling invariant ($||g_i s_i||_0^{(c)} = ||s_i||_0^{(c)} \forall g_i \neq 0$) and satisfies the triangle inequality, Theorem 3 can be proved in the same way as that of Theorem 1. Therefore, our task is to check the conditions for which these two properties are verified. Based on model (7), it is straightforward to check that (6) is scaling invariant if

$$\delta |g_i| \leq \epsilon, \gamma |g_i| > \epsilon, \ i = 1,2. \tag{10}$$

Concerning the triangle inequality, it is satisfied by measure (6) if the modulus of every linear combination of two samples smaller than $\delta$ is smaller than $\epsilon$. For this to be true, it suffices to impose (8). Note that (8) also covers $\delta |g_i| \leq \epsilon, i = 1,2$, which explain why there are only three conditions to be imposed. \[\square\]

Theorem 3 guarantees that, if $g$ is restricted to the region defined by (8) and (9), then (6) is a contrast. This feasible region is illustrated in Figure 1(a). Note that the area of this region increases as $\epsilon/\gamma$ decreases and $\epsilon/\delta$ increases. Also this illustration is helpful in pointing out the price to be paid for using an approximation of the $\ell_0$-norm. In fact, suppose that $g_1 = 1$; the ideal solution in this case is $g_2 = 0$. However, the vicinity of $g_2 = 0$ given by $|g_2| \leq \epsilon/\gamma$ is not inside the region covered by Theorem 3 and, thus, it is not assured that (6) is a contrast in this case. Given this, $\epsilon/\gamma$ can be view as a noise margin since the points $g_2$ smaller than this value may minimize $||y||_0^{(c)}$, although not being the ideal solution. Note that for the $\ell_0$-norm, this noise margin is zero.

![Fig. 1. Illustrating the regions for which (6) is a contrast.](image)
There is an interesting aspect here: Obtaining the feasible set in the $w_1 \times w_2$-plane is not a blind operation, since it requires the knowledge of $A$ to map from $g_1 \times g_2$-plane to the $w_1 \times w_2$-plane. Consequently, the shape of the feasible set in the $w_1 \times w_2$-plane (where blind algorithms will indeed operate) will depend on matrix $A$ as well.

5. CONCLUSIONS

In this work, we studied the use of the $\ell_0$-norm and a related measure as a BSE contrast. We showed that the proposed approach is valid if the sources have different degrees of sparsity. Moreover, when such a condition holds, source extraction can be conducted even when the sources are statistically dependent. This interesting feature paves the way for alternative routes to deal with signal separation problems for which the existing solutions fail.

6. REFERENCES