Blind Source Separation: Fundamentals and Perspectives on Galois Fields and Sparse Signals

Daniel G. Silva, Leonardo T. Duarte and Romis Attux

Abstract—The problem of blind source separation (BSS) has been intensively studied by the signal processing community. The first solutions to deal with BSS were proposed in the 1980's and are founded on the concept of independent component analysis (ICA). More recently, aiming at tackling some limitations of ICA-based methods, much attention has been paid to alternative BSS approaches. In this tutorial, in addition to providing a brief review of the classical BSS framework, we present two research trends in this area, namely source separation over Galois fields and sparse component analysis. For both subjects, we provide an overview of the main criteria, highlighting scenarios that can benefit from these more recent BSS paradigms.

Index Terms—BSS, ICA, Galois Field, Sparsity.

I. INTRODUCTION

BLIND Source Separation (BSS) is one of the most relevant subject in unsupervised signal processing, with a myriad of aspects worthy of investigation and analysis, such as (*i*) the separation criteria and the implied hypothesis about the sources characteristics, (*ii*) the generative model that yields the mixed signals and its association with the separation system and (*iii*) the algorithms to determine the solution parameters.

The "canonical" concept to solve BSS is the application of Independent Component Analysis (ICA) [1], [2] in the context of real- or complex-valued signals. Such approach presumes independence between the sources and, consequently, the separation strategy lies on recovering the sources from the set of dependent mixtures by searching for a recovered independent configuration. Nevertheless, there are two alternative points of view that have been consistently treated in the last years, and which deserve special attention: a) the case of linear scenarios with inherently discrete, finite-domain signals, which formally comprise the finite (or Galois) field theory and b) the use of priors based on signal sparsity (instead of independence) in the time domain or in a domain engendered by an adequate transform.

This tutorial intends to introduce and describe these two modern research trends in unsupervised signal processing: BSS over Galois fields and BSS over sparse signals. We put emphasis on the analysis of the main separation criteria and the particularities of each domain, when confronted to the canonical framework. In order to do so, the work is organized in the following sections: Section II reviews the fundamental concepts underlying BSS and the use of ICA; Section III discusses the BSS extension to the domain of Galois fields, describing the main theoretical developments and potential applications; Section IV studies the notion of sparsity in information signals and its relevance to solve BSS within different complexity instances; and Section V presents the final remarks.

II. BASIC CONCEPTS

The BSS problem can be defined, in simple terms, as that of recovering a set of information signals (sources) from mixed versions of them (mixtures). In principle, there are no limitations regarding the mixing process, which can be nonlinear, with memory, time variant etc. However, for the sake of mathematical tractability, and in view of a vast number of applications, the linear and instantaneous mixing model can be assumed as being canonical. In this model, it is considered that N sources are detected by M sensors in the form of linear combinations i.e. there is a superposition of signals with different gains, but not of delayed versions. Mathematically, if there is a source vector $\mathbf{s}(n) = [s_1(n), s_2(n), ..., s_N(n)]^T$ and a mixture vector $\mathbf{x}(n) = [x_1(n), x_2(n), ..., x_M(n)]^T$, for a given instant n, the model can be expressed as:

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n), \tag{1}$$

being $\mathbf{A}$ an $M \times N$ mixing matrix. Note that, in this explanation, the model is built without reference to measurement noise, although its presence is relevant in both theoretical and practical terms [1].

When $N > M$, there arises an *underdetermined* case, which is difficult to deal with because it maps the desired information from an original signal space onto a space of smaller dimension. On the other hand, when $M > N$, there is an *overdetermined* model, which poses, undoubtedly, fewer complications. Finally, there is the most usual case in the literature, when $M = N$, which will be the standard throughout this section.

In this last case, if the matrix $\mathbf{A}$ (which becomes square) is invertible, it is possible to formulate the problem of BSS as that of finding another square matrix $\mathbf{W}$ (called separating matrix) so that there is a vector of estimated sources

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) \tag{2}$$

giving rise to a solution as follows:

$$\mathbf{y}(n) = \mathbf{D}\mathbf{P}\mathbf{s}(n), \tag{3}$$

Daniel G. Silva is with the Department of Electrical Engineering (ENE), University of Brasília (UnB), Brasília, DF, Brazil email: danielgs@ene.unb.br.

Leonardo T. Duarte is with the School of Applied Sciences (FCA), University of Campinas (UNICAMP), Campinas, SP, Brazil email: leonardo.duarte@fca.unicamp.br .

Romis Attux is with the Department of Computer Engineering and Industrial Automation (DCA), University of Campinas (UNICAMP), Campinas, SP, Brazil email: attux@dca.fee.unicamp.br

being $\mathbf{D}$ a diagonal matrix and $\mathbf{P}$ a permutation matrix. The meaning of (3) is that, in the solution of the BSS problem as formulated (and even in a more general sense), the sources can be recovered in any order and are subject to scaling factors, which means that these information-preserving ambiguities are tolerated.

Once those remarks are made, it remains a question: how can $\mathbf{W}$ be obtained in an unsupervised (or blind) fashion? To answer it, it is necessary to make some sort of hypothesis about the sources. Although it is beyond dispute that the canonical hypothesis considers sources as mutually independent stochastic signals, there are more than one possible path to follow here, as will be seen later on. This hypothesis, which is valid in many domains [1], is a very strong one under the aegis of the defined model. In fact, as shown in [2], if the components of the vector $\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n)$ are mutually independent, the sources will have been recovered aside from the ambiguities expressed in (3). In other words, recovering the independence condition implies correct source estimation.

It is exactly because of this fact that there is a strong link between BSS and the methodology known as independent component analysis (ICA) [2], which, in contrast with the more popular technique of principal component analysis (PCA) [1], has the objective of finding projections that generate statistically independent factors (and not only uncorrelated, as is the case with PCA) underlying the focused data. By means of ICA, it is possible to build cost (or contrast) functions that allow the search for matrices $\mathbf{W}$ capable of providing efficient source recovery.

A. Criteria for Performing ICA-Based BSS

Several formulations can be used to perform ICA. Here, we will discuss three of them, based on the concepts of mutual information and non-Gaussianity (quantified in terms of kurtosis and negentropy) [3].

1) Mutual Information: A very natural criterion to quantify statistical dependence is the mutual information, which, for a general random variable vector $\mathbf{a}$, with K elements, is defined as [1]:

$$I(\mathbf{a}) = \sum_{k=1}^{K} h(a_k) - h(\mathbf{a}), \tag{4}$$

where $h(\cdot)$ is Shannon's differential entropy, given, for a vector, by [4]:

$$h(\mathbf{v}) = -\int p(\mathbf{v}) \ln\left[p(\mathbf{v})\right]\, d\mathbf{v}. \tag{5}$$

Since entropy can be seen, in simple terms, as the degree of uncertainty associated with a random variable, we may interpret (4) as being, intuitively, the difference between the total uncertainty originated by a separated observation of its components and the total uncertainty originated by a joint observation. When this difference is null, "no component carries information about the other", so to say, which in more rigorous terms implies that they are independent. If there is statistical dependence, $I(\mathbf{a}) > 0$ [4]. Hence, if the mutual information associated with $\mathbf{y}(n)$ is minimized with respect to

$\mathbf{W}$, it shall be possible to restore the independence condition and to recover the sources.

A major difficulty here lies in the question that entropy calculation requires knowledge of the involved probability densities or their estimation (something that can be quite complex in some cases). The joint entropy is not an issue, because it is possible, using the hypotheses regarding the model, to write (the time indices will be omitted for simplicity) [3]:

$$h(\mathbf{y}) = h(\mathbf{x}) - \ln\left[|\det(\mathbf{W})|\right]. \tag{6}$$

Note that $h(\mathbf{x})$ is fixed and that $\mathbf{W}$ is known, being the analyzed solution. One cannot avoid nonetheless the need for estimating the marginal entropies that form the first term of the right-hand side of (4).

This difficulty explains why the use of the mutual information in the linear and instantaneous case is not common, although it is very relevant, for instance, in the nonlinear context, which will not be dealt with here [5], and in the context of signals over finite fields, as the reader will see in Section III.

2) Non-Gaussianity: Kurtosis and Negentropy: The central limit theorem [3] can be enunciated as follows. Let a set of K continuous random variables a_i, $i = 1, ..., K$, i.i.d. (independent and identically distributed); their mean converges, in the limit $K \to \infty$, to a variable with a Gaussian density.

Well, as, in ICA, mutual independence between sources is assumed, and, moreover, in the linear case the mixtures are essentially sums, the central limit theorem implies that to mix means "to Gaussianize". In other words, it can be said that a mixture is "more Gaussian" than the sources generating it.

A way to quantify Gaussianity is to employ the kurtosis, a fourth-order statistic that, for Gaussian variables, is null. Its definition is, for a real-valued scalar variable a:

$$K(a) = E[a^4] - 3E^2[a^2]. \tag{7}$$

The use of kurtosis is more naturally explained in the domain of source extraction i.e. when the goal is to recover a single source. This leads to:

$$y_i(n) = \mathbf{w}_i^T \mathbf{x}(n), \tag{8}$$

being $\mathbf{w}_i$ interpreted as one of the lines of the separating matrix $\mathbf{W}$. It can be shown, which is intuitive in view of the central limit theorem, that to maximize the absolute value of the kurtosis with respect to $\mathbf{w}_i$, under a proper constraint, allows a source to be recovered[1]. Consequently, there arises a criterion of the form:

$$\max_{\mathbf{w}_i} |K\left[y_i(n)\right]|. \tag{9}$$

To recover all sources, it is necessary to make use of a deflation process i.e. to remove each extracted source from the remaining mixture(s) or to resort to constraints that prevent the extraction of the same signal [1].

Another form of using the idea of Gaussianization inherent to the mixing process is to bring to the scene a classical result from information theory. This result ensures that, among

---

[1] This framework is very similar to that of the Shalvi-Weinstein theorem [6] for blind equalization.

all random variables $\mathbf{a}$ with a fixed second-order statistical structure, the Gaussian density is the one with maximum differential entropy [4].

Having this in mind, let us imagine a random variable a with generic mean and covariance (we will consider the scalar case to simplify things). Let us now consider a Gaussian random variable with the same moments up to second order. It is possible to define the negentropy $\mathrm{N}(a)$ as follows:

$$\mathrm{N}(a) = h_{\mathrm{gauss}}(a) - h(a), \qquad (10)$$

where $h(a)$ is the differential entropy of the variable in question and $h_{\mathrm{gauss}}(a)$ is the entropy of a Gaussian random variable with the same moment structure up to order two. From what was discussed, it follows that $\mathrm{N}(a) \geq 0$. To obtain a proper extraction vector $\mathbf{w}_i$, $\mathrm{N}(y_i)$ must be maximized with respect to it, so that the non-Gaussianity be maximized. Again, the retrieval of all sources demands deflation or special constraints.

The use of negentropy requires, in principle, entropy estimation, which, as already mentioned, can be a complex task in some cases. Therefore, it is usual to adopt certain nonlinear functions as means to approximate it. With these functions, it is more straightforward to derive gradient or fixed-point algorithms [1].

B. Other approaches to linear-instantaneous mixing models

Even if ICA is at the origins of BSS, there is now a number of alternative approaches to deal with separation problems in linear models. For instance, another classical paradigm in BSS considers that the sources can be modeled as stochastic processes, which allows one to exploit temporal information related to the sources. Such an approach is the basis of the algorithm SOBI [7] and, more generally, of the class of algorithms that exploit the auto-correlation structure of the sources [2]. A nice aspect of such methods is that they can be applied to separate non-white Gaussian sources under the condition that these sources present different auto-correlation functions.

A different approach that has been adopted in BSS comes from the machine learning community: it is known as non-negative matrix factorization (NMF) [8], [9]. The goal in NMF is to search for an approximation for the observed non-negative matrix $\mathbf{X}$ as follows

$$\mathbf{X} \approx \mathbf{AS}, \qquad (11)$$

where $\mathbf{A}, \mathbf{S} \geq 0$. In the context of BSS, $\mathbf{A}$ and $\mathbf{S}$ are related to the mixing matrix and the sources, respectively, which are thus assumed non-negative. Such an assumption is realistic in several applications, such as those related to chemical analysis [10] and audio processing in transformed domains [11]. Although NMF lacks from separability results (see, for instance, [12]), in the sense that it is an ill-posed problem, the combination of NMF with additional prior such as sparsity or smoothness may provide sound BSS algorithms [13].

Another emblematic example of BSS paradigm is based on a Bayesian formulation of the problem [2]. This approach is suitable when there is a set of prior information that can be modeled through probability distributions. For instance, non-negative priors can be modeled in a Bayesian framework by considering distributions of non-negative support [14], whereas sparsity can be represented by, for example, a Laplacian distribution [15].

Finally, an extension to the instantaneous model can be considered i.e. the blind separation of *convolutive* mixtures. The main difference between the linear-instantaneous model and the convolutive one is that the latter becomes a superposition not only of the present values of the sources, but also of past values. In other words, the convolutive model includes the dimensions of space and time, with the multiplicity of sensors and instants engendering the mixing process. Mathematically, (1) is extended as:

$$\mathbf{x}(n) = \sum_{k=0}^{T} \mathbf{A}(k)\mathbf{s}(n - k), \qquad (12)$$

where T corresponds to the maximum delay present in any of the mixtures. Notice that (12) yields (1) when $T = 0$.

It is possible to solve this problem in the time domain (e.g. using predictive strategies [16]) or to use the property that a convolution in the time domain is a product in the frequency domain, which gives rise to a sort of instantaneous mixture [2]. In the latter case, special precautions must be taken with respect to the permutation and scale ambiguities, which may cause severe spectral distortions.

III. SEPARATION OVER FINITE FIELDS

After discussing the general aspects of BSS, now we focus on a more specific case, which necessarily deals with *digital* data. Then, BSS can be studied, for example, when signals and mixing processes are binary. This perspective was first proposed in [17] and belongs to the generic framework of source separation over *finite* or *Galois fields*.

This is the fundamental topic of this section, which is organized as follows: first, the signal representation over Galois fields is presented; then the most important criteria to separate such signals from instantaneous mixtures is discussed in Section III-B; Section III-C extends the analysis to the convolutive mixing model and, finally, Section III-D illustrates two potential applications of the techniques so far developed.

A. Signals over $GF(q)$

Fields are abstractions of familiar number systems and their essential properties [18]. A field F is defined as a set of elements associated with two operations, $+$ and $\cdot$, such that the following axioms are valid: closure, commutativity, associativity, distributivity, existence of neutral element and existence of inverse element [19].

Real and complex numbers are well-known examples of fields, both with an infinite number of elements. However, this is not a mandatory requisite: there are also finite (or Galois) fields, e.g. the set $\{0, 1\}$ with the logical operations exclusive-or (XOR) and AND as addition and product, respectively.

A finite field with q elements is named $F = GF(q)$; it is possible to show that $q = P^n$, where P is a prime and is typically called the *characteristic* of the field. If $n = 1$, F

is a *prime field* and its operations are easily defined as the product and sum *modulu* P over the elements $\{0, ..., P-1\}$. Otherwise, fields with $n > 1$ are called *extension fields* and imply a more complex definition of operations [18].

Vector spaces over finite fields can also be constructed, with a remark that such spaces are not ordered and there is no notion of orthogonality [20]. Linear mappings $\mathbf{A} : F^N \rightarrow F^M$ are represented by $M \times N$ matrices with elements in F, in accordance to the usual restrictions for having an inverse mapping — the matrix must be square and with a non-null determinant.

B. Separation over $GF(q)$ in instantaneous models

Consider the BSS formulation for the instantaneous and determined ($M = N$) case, as (1) illustrates, but with the difference that all entities and operations are defined over a field $F = GF(q)$. Hence, the problem consists of finding, in the space of all invertible N-dimension matrices — $GL(N, q)$ —, the one that recovers $\mathbf{s}(n)$, in equivalence to the definition given in (3).

The following theorem offers the possibility of achieving the solution through ICA [21]:

Theorem 1 (Identification via ICA) *Consider $F = GF(q)$ a finite field of order q. Assume that $\mathbf{s}$ is a vector of independent random variables in F, with probability distribution $p_\mathbf{s}$ such that the marginal distributions are non-uniform and non-degenerate[2]. If, for some invertible matrix $\mathbf{G}$ in F, the components of the vector $\mathbf{y} = \mathbf{G}\mathbf{s}$ are independent, then $\mathbf{G} = \mathbf{D}\mathbf{P}$ for a permutation matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{D}$.*

For instance, consider $GF(3)$ and two independent sources with marginal probability vectors given by $\mathbf{p}_{s_1} = [1/2, 3/8, 1/8]$ and $\mathbf{p}_{s_2} = [1/3, 1/6, 1/2]$. Hence, the joint distribution is

$$\mathbf{p}_\mathbf{s} = [1/6, 1/12, 1/4, 1/8, 1/16, 3/16, 1/24, 1/48, 1/16]$$

for the respective values of $\mathbf{s} = [0,0]^T, [0,1]^T, ..., [2,1]^T, [2,2]^T$. If the sources are multiplied by a matrix according to Theorem 1,

$$\mathbf{G} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix},$$

then $\mathbf{y} = \mathbf{G}\mathbf{s} = [2s_2, 2s_1]^T$ is a random vector with marginal probabilities $\mathbf{p}_{y_1} = [1/3, 1/2, 1/6]$, $\mathbf{p}_{y_2} = [1/2, 1/8, 3/8]$ and joint distribution

$$\mathbf{p}_\mathbf{y} = [1/6, 1/24, 1/8, 1/4, 1/16, 3/16, 1/12, 1/48, 1/16],$$

which yields $p_\mathbf{y}(\mathbf{k}) = p_{y_1}(k_1)p_{y_2}(k_2)$, i.e. the components of $\mathbf{y}$ are independent.

Consequently, this result indicates that one can employ ICA to perform blind separation of signals over $GF(q)$, as long as the original signals are independent and non-uniformly distributed, leading to extracted signals that differ only by

---

scale and permutation ambiguities. Since there is no definition of statistical moment for random variables over a finite field, in order to define a criterion similar to negentropy or kurtosis, it is necessary to employ the concepts that information theory offers.

An important property states that the linear combination of independent random signals results in an entropy greater than or equal to the original signals [22]. Based on this, a first separation strategy rises via source extraction [2], as already mentioned in Section II-A. The AMERICA algorithm [20] implements this technique for performing ICA over $GF(q)$, through an exhaustive search with the criterion

$$\min_{\mathbf{w}_i \in F^N} H\left[\mathbf{w}_i^T \mathbf{x}(n)\right], \tag{13}$$

to be executed N times, with the restriction that each obtained extraction vector is linearly independent from the previous ones. Note also that $H(\cdot)$ is Shannon's entropy for *discrete* random variables

$$H(v) = -\sum_{v \in F} p(v) \log p(v). \tag{14}$$

Figure 1 describes AMERICA pseudocode. Despite the adoption of an exhaustive search approach, AMERICA assures convergence to the correct inverse solution (as long as the criterion is perfectly calculated).

$\mathbf{W} \leftarrow \emptyset$;
while $\dim(\mathbf{W}) \neq N$ **do**
  $\mathbf{w}_o = \arg\min_{\mathbf{w} \in F^N \setminus span(\mathbf{W})} H\left[\mathbf{w}^T \mathbf{x}(n)\right]$;
  $\mathbf{W} \leftarrow \mathbf{W} \cup \{\mathbf{w}_o\}$;
end while
Output: $\mathbf{W}$

Figure 1. Pseudocode of AMERICA algorithm.

There are algorithms that trade convergence for a lower computational cost, through approximations of the criterion defined in (13), such as the techniques named MEXICO and CANADA [20]. The MEXICO algorithm, particularly, adopts the strategy of sequentially minimizing the entropy between *pairs* of mixtures, which does not assure global optimum convergence, but reduces the expected computational cost in comparison to AMERICA [23].

A different perspective lies on considering the same criterion with lower-cost metaheuristics that are appealing for combinatorial problems, e.g. Artificial Immune Systems (AIS) [24]. In this case, the algorithm optimizes (13), but at the end of the procedure, the N best candidate-solutions which are linearly independent represent the extraction vectors that, finally, compose the separating matrix. This *modus operandi* is possible due to the intrinsic capacity of AIS to promote *diversity* among the candidate-solutions, while the search occurs [25], which allows the algorithm to obtain the multiple solutions that are required to build the separating matrix.

Beyond the idea of exploring entropy as contrast function, a second independence criterion involves direct minimization of mutual information among the extracted signals. Mutual information is defined according to (4), remarking that, instead

of differential entropy, we consider the entropy for discrete variables. The calculation of $I(\cdot)$ among the components of the estimated sources vector (for the purpose of simplicity, we leave aside the temporal index) hence provides

$$I(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{y}). \qquad (15)$$

Fortunately, the second term on right-hand-side of (15) can be ignored, because when an invertible mapping $\mathbf{y} = \mathbf{Wx}$ of signals defined over discrete sets is considered, the following relationship holds

$$p_{Y_1,\ldots,Y_N}(\mathbf{y}) = p_{X_1,\ldots,X_N}(\mathbf{W}^{-1}\mathbf{y}), \qquad (16)$$

which, consequently, implies $H(\mathbf{y}) = H(\mathbf{x})$. Then, we obtain the final expression for the criterion:

$$\min_{\mathbf{W} \in GL(N,q)} \sum_{i=1}^{N} H(y_i),$$
$$\text{subject to } \begin{cases} \mathbf{y} = \mathbf{Wx}, \\ \mathbf{W} \text{ is invertible} \end{cases}. \qquad (17)$$

Since the search space size is proportional to q^{N^2} [26], there is a considerable increasing as compared to the space size of the first criterion, which is proportional to q^N, thus hindering the use of exhaustive search methods in this case. Then, it is possible to consider again the application of population-based metaheuristics such as AIS [27], [28], which offer signal separation with quality levels similar to exhaustive heuristics, but with a reduced computational cost. For instance, Figure 2 illustrates the successful application of the AIS-based method described in [28] for separation of black-and-white images.

C. Separation over $GF(q)$ in convolutive models

Let us consider a new situation, where there is combination of signals, defined over $GF(q)$, both in space and *time*, which yields the convolutive mixture model, mathematically described in (12).

ICA can be used once again to recover the original signals, as the authors of [29] propose. Assume that the sources are non-uniform and mutually independent (in space and time), which (again) results that the mixing process generates signals with greater entropy than the sources, in a similar fashion to AMERICA algorithm principle.

Hence, it is possible to use the extraction/deflation technique, previously mentioned in Section III-B, to revert the entropy increasing effect. A source extraction problem takes place, which consists of determining the separation filters that produce the output

$$y(n) = \sum_{k=0}^{T_e} \mathbf{w}^T(k)\mathbf{x}(n-k)$$
$$= \sum_{j=1}^{N} w_j(n) * x_j(n) \qquad (18)$$
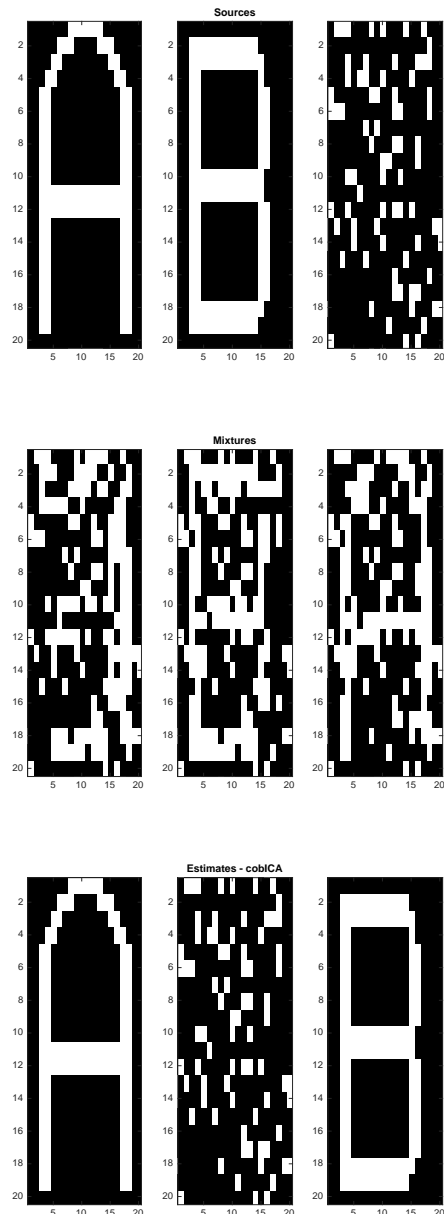$$= \sum_{j=1}^{N} u_j(n),$$

Figure 2. Application example of ICA over GF algorithm with black-and-white images.

where T_e is the maximum delay present in one of the filters $w_j(n)$ and $\mathbf{w}(n) = [w_1(n) \ldots w_N(n)]^T$. Figure 3 presents an example of convolutive mixture for $N = 2$, in association with the extraction procedure of a source signal.

Like the instantaneous case, the mixing matrix $\mathbf{A}(n)$ must be invertible, i.e. the determinant of $\mathbf{A}(n)$ must be non-null for all n. In the context of temporal filtering, this implies that if the matrix is composed of finite impulse response (FIR) filters, with input-output relationship

$$x_i(n) = \sum_{j=1}^{N} \sum_{k=0}^{T} a_{ij}(k)s_j(n-k), \quad i = 1, 2, \ldots, N, \qquad (19)$$

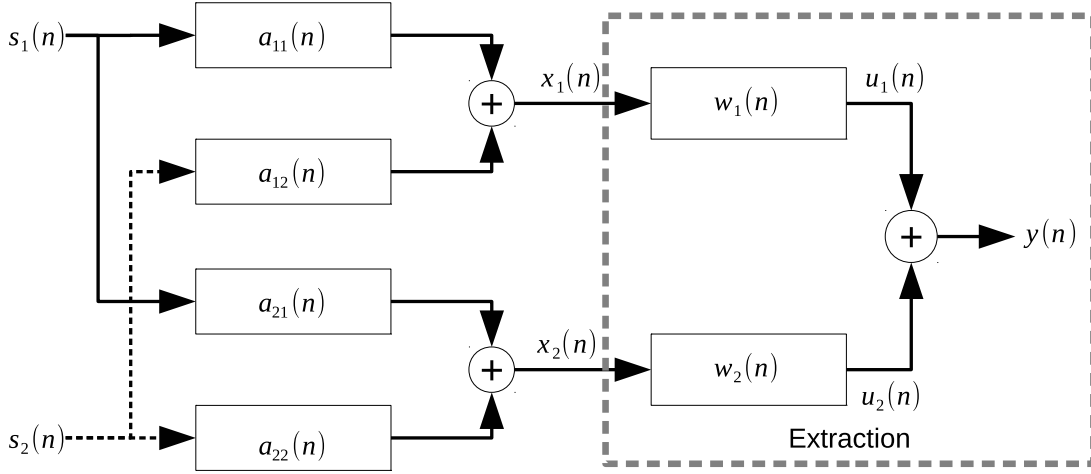the inversion is only possible if the extraction filter contains

Figure 3. Model representing the convolutive mixture problem over $GF(q)$ when $N = 2$, and the extraction system of a source.

feedback loops [29], i.e.

$$y(n) = \sum_{j=1}^{N} u_j(n)$$

$$= \sum_{j=1}^{N} \left[\sum_{k=0}^{T_b-1} b_j(k)x_j(n-k) + \qquad (20) \right.$$

$$\left. + \sum_{l=1}^{T_c} c_j(l)u_j(n-l) \right],$$

where T_b e T_c are the number of coefficients $b_j(k)$ and $c_j(l)$ of the filter, respectively. Then, the values of these parameters are estimated by minimizing the loss function of the extraction process, which is

$$\min_{b_j(\cdot),\, c_j(\cdot)} H[y(n)], \qquad (21)$$

where $y(n)$ is obtained according to (20). When the extraction succeeds, one obtains $y(n) = cs_i(n-d)$, $i \in \{1, ..., N\}$, which means that a delayed and scaled version of a source is recovered.

After extracting a source, the next step is the *deflation* process, in order to remove the recovered source from the remaining mixtures. Figure 4 details this task: assume that $y(n)$ represents the extracted signal, it must be processed by a non-causal, FIR *deflation filter*, which identifies the inter-symbolic interference signature of the source — the mixing filter $a_{ij}(n)$ — with respect to each mixture; then the signal can be properly subtracted from the mixtures.
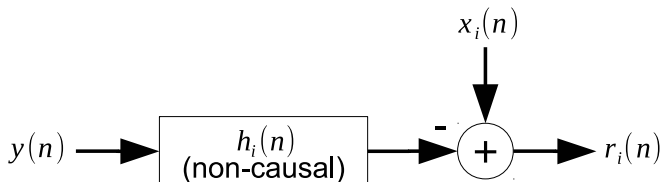
Figure 4. Representation of the deflation step, considering $y_1(n)$ the signal to be removed.

The deflation filter parameters are defined using (again) the entropy measure, via a criterion that is analogous to the employed for deflation of instantaneous mixtures [21]:

$$\min_{h_i(n)} H[x_i(n) - h_i(n) * y(n)]. \qquad (22)$$

When deflation ends, the extraction step must be repeated, in order to obtain the second source, but remember that the new mixtures are represented by the signals $r_i(n)$, $i = 1, 2, ..., N$. Therefore, both processes are alternated until all mixtures become null signals, which means that all sources were recovered.

D. Applications

Although BSS over finite fields and the associated solution strategies via ICA were initially considered only under the theoretical perspective, there are already some potential applications being developed, specially when the mixtures follow the instantaneous paradigm.

A first application lies on eavesdropping MIMO systems which employ PAM modulation and Tomlinson-Harashima pre-coding [20]. Consider a system with N transmitters and receivers, which is designed to send N binary signals to each receptor through a pure attenuation channel $\mathbf{H} \in [0, 1]^{N \times N}$. Since the transmitters known the channel characteristics, we could consider the strategy of each one sending the vector components given by $\mathbf{x}(n) = \mathbf{H}^{-1}\mathbf{s}(n)$, such that the reception would result in $\mathbf{y}(n) = \mathbf{H}\mathbf{x}(n) = \mathbf{s}(n)$.

However, if the system employs PAM modulation, hence transmitting data only in the interval $[0, 1]$, this approach would lead to transmission sequences with invalid values. In this case, the Tomlinson-Harashima spatial coding can be employed to circumvent this limitation [30]: the channel matrix is quantized into P levels (P is prime), and the inverse (over $GF(P)$) of this new matrix $\hat{\mathbf{H}}$ is applied to the transmission sequence :

$$\hat{\mathbf{x}}(n) = \frac{1}{P-1} \left[\hat{\mathbf{H}}^{-1} \left(\mathbf{s}(n)\frac{1}{2}(P-1) \right) \right]_P \qquad (23)$$

where $\hat{\mathbf{H}}^{-1}\mathbf{s}(n)$ is a "conventional" product over the real field, and $[\cdot]_P$ denotes the *modulu P* operation. This formulation results in a sequence with real values in $[0, 1]$ that can be transmitted via a PAM scheme and, in the receptor, the original values are reconstructed via the following expression [20]:

$$
\begin{aligned}
\left[(P-1)^2 \hat{\mathbf{y}}(n)\right]_P &= \left[(P-1)^2 \mathbf{H}\hat{\mathbf{x}}(n)\right]_P \\
&= \mathbf{s}(n)\frac{1}{2}(P-1).
\end{aligned}
\tag{24}
$$

In this context, this communication system can be eavesdropped, via ICA, as follows:

- A third party with another set of N antennas, intercepts the signals that are being transmitted, $\hat{\mathbf{y}}_e(n)$.
- He knows the value of P, however, he does not know the attenuation matrix between the transmitters and his antennas set, $\hat{\mathbf{H}}_e$, which is assumed to be quantized in P levels.
- When the same operations of the legitimate receivers are applied, the result is [20]

$$
\left[(P-1)^2 \hat{\mathbf{y}}_e(n)\right]_P \triangleq \hat{\mathbf{A}} \circ \left(\mathbf{s}(n)\frac{1}{2}(P-1)\right), \tag{25}
$$

where $\circ$ denotes $GF(P)$ product and $\hat{\mathbf{A}}$ is a matrix given by the composition of $\hat{\mathbf{H}}_e$ with $\hat{\mathbf{H}}^{-1}$, the latter is employed in the pre-coding step.
- Equation (25) leads to the definition of the BSS problem over finite fields, hence the application of an ICA algorithm can invert $\hat{\mathbf{A}}$ and consequently provide estimates for the transmitted sequences.

Naturally, this sort of ICA application makes use of hypotheses that restrict its viability, nevertheless, it gives us interesting insights of other potential applications that are related to coding theory. This perspective is reinforced by the second example of application, which involves ICA for improving *Network Coding* algorithms.

In simple terms, Network Coding claims that the intermediate nodes of a communication network can, instead of just forwarding data packages, process linear combinations of them, with randomly-defined coefficients over a finite field. With this idea, it is possible to show that the transmission flow over the network is maximized and the robustness against errors is increased, specially in the context of real-time applications [31].

However, in order to decode the packages at the destination nodes, the combination coefficients must be sent as a package header, which is an overhead for transmission rate, in the case of small size packages. This is the aspect to be reconsidered, then: if the coefficients are not inserted in the package, decoding still can be done by casting the problem as BSS over $GF(q)$.

This is the proposal introduced in [31], which ignores the coefficients header and substitutes it by a non-linear hashing function of each package, in order to assure that data is non-uniformly distributed — a fundamental condition to perform decoding via ICA, as seen in Theorem 1. Since it is quite usual that data traffic, in multimedia networks, presents a distribution close to uniform, e.g. compressed audio and video, the hash

mapping is necessary to increase the discriminative power of the algorithm cost function.

It is important to emphasize that the hashing function implies an overhead to each original package smaller than the conventional approach, while the failure probability on executing the separating algorithm is maintained with low values. Experimental analyses, in this context, have shown that packages with size between 1 and 1.5 kilobytes present good decoding rates by the new technique, saving about 50% of header size [31].

IV. SEPARATION OF SPARSE SIGNALS

In the present section, we shall discuss another emerging topic in BSS which has been extensively studied over the last years: the case in which the sources can be modeled as sparse signals. Besides being observed in several real applications [32], the hypothesis of sparsity allows one to develop novel methods that are able to deal with situations for which classical approaches, such as ICA, fail. The separation framework based on the sparsity hypothesis is usually referred to as sparse component analysis (SCA).

The brief overview on SCA provided in this section is organized as follows. Firstly, we discuss the notion of a sparse signal. Then, in Section IV-B, we shall discuss how the sparsity prior is exploited in the case of underdetermined mixtures. As it will be seen in Section IV-C, sparsity is also a useful information in the context of determined sources, especially when the hypothesis of independence does not hold.

A. Sparse signals

Although there is no formal definition for a sparse signal, the notion of sparsity in fields such as signal processing and machine learning is now ubiquitous and is associated with a signal that can be represented by a number of elements that is rather smaller than the signal observed dimension. In Figure 5, examples of sparse signals and images are provided. It is worth noticing the presence of a large amount of temporal samples (in the case of signals) or pixels (in the case of images) that take values that are almost null. Examples of sparse signals and images arise in different domains, including biomedical signal processing [33], geophysics [34], and audio processing [35].

Before defining separation methods based on the notion of sparsity, it is paramount to define measures of sparsity. The most natural one is the ℓ_0-pseudo-norm , which, for a discrete signal of T samples, represented by the vector $\mathbf{s}$, is defined as follows[3] [36]

$$
||\mathbf{s}||_0 = \lim_{p \to 0} ||\mathbf{s}||_0^p = \lim_{p \to 0} \sum_{n=1}^{T} |s_n|^p. \tag{26}
$$

The ℓ_0-pseudo-norm is simply the number of non-null samples of $\mathbf{s}$. Therefore, a sparse signal tends to present a small ℓ_0-pseudo-norm. Moreover, a signal can be sparse in other domains, that is, when the ℓ_0-pseudo-norm of a transformed

---

[3]The following notation will be adopted in this section: $\mathbf{s}_i$ and $\mathbf{x}_j$ represent the i-th source and j-th mixture, respectively. Each row of the matrices $\mathbf{S}$ and $\mathbf{X}$ corresponds to a given source and mixture, respectively.
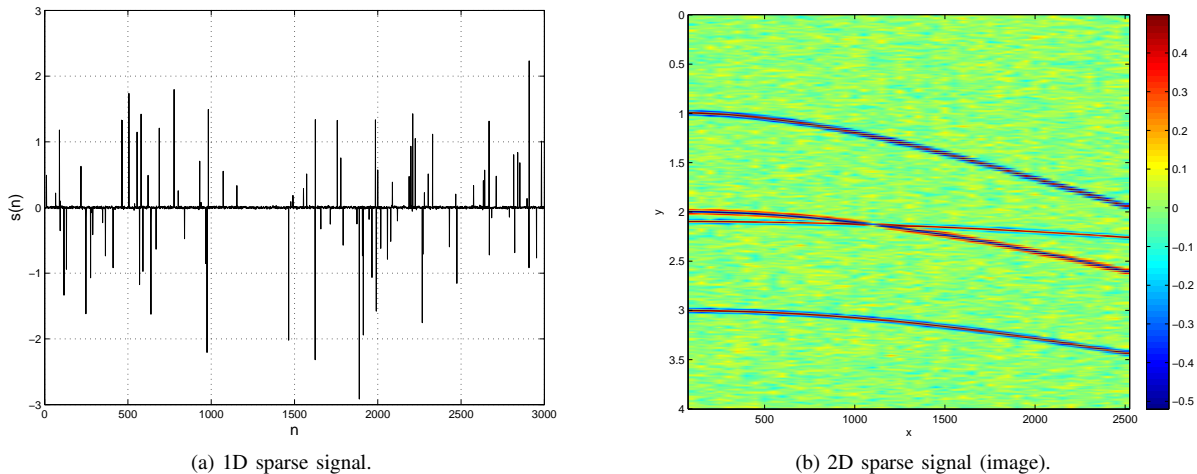
(a) 1D sparse signal.

(b) 2D sparse signal (image).

Figure 5. Examples of sparse signals.

version of **s** is low — a common example is a sine wave, which is sparse in the Fourier domain. There are other measures of sparsity. Among the most relevant ones is the ℓ_1-norm , defined as follows

$$||\mathbf{s}||_1 = \sum_{n=1}^{T} |s_n|. \tag{27}$$

Since the ℓ_1-norm engenders convex optimization problems, it is quite used in a vast number of signal processing tasks.

B. Separation of sparse signals in underdetermined models

The first works on sparse models for BSS addressed the case of underdetermined mixtures [37], [38], [39] and are based on two steps. In the first one, one searches for estimating the mixing matrix **A**. This first process is illustrated in Figure 6, which represents a BSS problem in which there are $N = 3$ sources and $M = 2$ mixtures. Given that the sources are sparse, there is a high probability that only one source is active at a given instant. For instance, let us consider the instants in which source $\mathbf{s}_1$ is much higher than $\mathbf{s}_2$. In these moments, the mixtures almost become functions of a single source, that is, $\mathbf{x}_1 = a_{11}\mathbf{s}_1$, $\mathbf{x}_2 = a_{21}\mathbf{s}_1$, and, therefore, they carry information about the first column of **A**. Analogously, when the sources $\mathbf{s}_2$ and $\mathbf{s}_3$ are exclusively active, the mixtures bring information about the second and third columns of **A**, respectively.

The fact described in the last paragraph is illustrated in Figure 6, which provides the mixtures scatter plot. One can note that the information on the columns of **A** are related to the clusters that arise when the sources are almost isolated, that is, when there is a single active source dominating the others. Therefore, a natural idea to estimate **A** is to determined the directions for which there is a relevant concentration of points — such a procedure can be carried out by clustering algorithms [38].

Having estimated the mixing matrix **A**, a second step is to solve a underdetermined linear system for estimating the sources. A first idea in that respect would be to formulate a least squares problem. However, given that the number

of unknowns is greater than the number of observatiom, the resulting problem becomes ill-posed and admits infinite solutions. As an alternative, one may consider, as prior information, the fact that the sources are sparse, which can be implemented according to the following optimization problem:

$$\min_{\mathbf{S}} \quad \sum_{i=1}^{N} ||\mathbf{s}_i||_0 \tag{28}$$
$$\text{subject to} \quad \mathbf{X} = \hat{\mathbf{A}}\mathbf{S},$$

where $\hat{\mathbf{A}}$ corresponds to the estimated version of the mixing matrix (this estimation is obtained in the first step). It is worth noticing that problems such as (28) have been extensively studied over the last years, mainly due to their applicability in compressive sensing [40], which searches for sampling signals and images by considering a rate that is lower than the Shannon-Nyquist rate.

Besides the formulation expressed in (28), there are other approaches that deal with inverse problems by making use of prior information related to sparse signals. Two notorious examples comprise a method known as the least absolute shrinkage and selection operator (LASSO) and a formulation known as basis pursuit de-noising (BPDN) [41].

Finally, it is worth mentioning that a similar approach based on a two-step strategy can also be applied in the context of sparse source separation. For instance, the algorithm DUET [42] estimates the mixing matrix by considering the disjoint orthogonality assumption, which means that only a single source can be active at a given instant. In a similar fashion, the algorithms TIFROM and TiFCorr search for regions where the sources are isolated [43] either in time or in other transformed domains.

C. Separation of sparse signals in determined models

The assumption of sparse sources is also useful as a prior in the context of determined models. A first approach in this case is similar to the one described for the case of underdetermined models (estimation of mixing matrix followed by sparse inversion). A second possibility is to set up a separation criterion
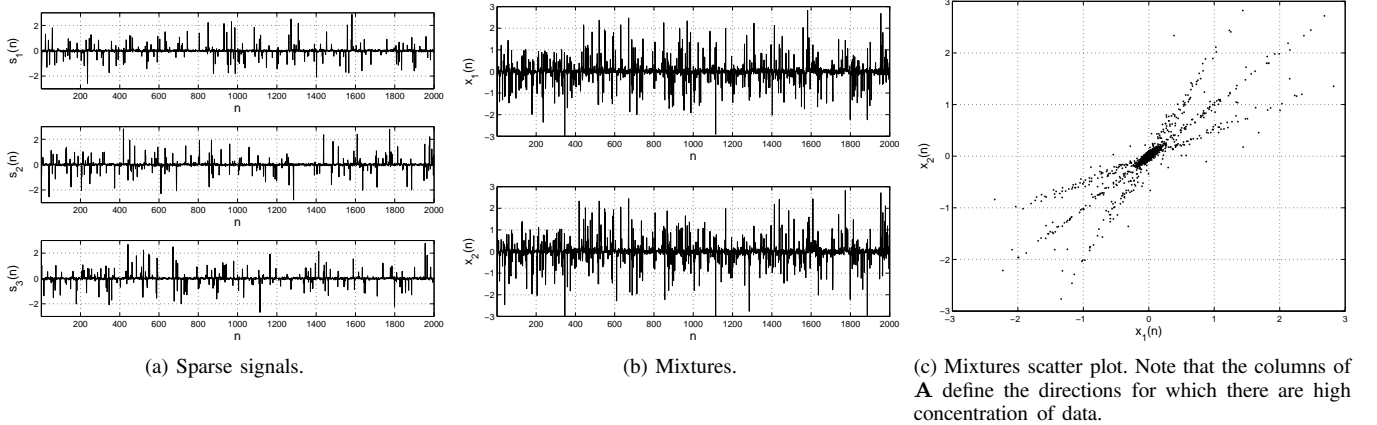
(a) Sparse signals.   (b) Mixtures.   (c) Mixtures scatter plot. Note that the columns of $\mathbf{A}$ define the directions for which there are high concentration of data.

Figure 6. Mixing matrix estimate $\mathbf{A} = [1\ 0.5\ 1; 0.5\ 1\ 1]$ by exploring the sources sparsity.

that takes into account the sparsity prior. In this case, which will be discussed in the present section, source estimation is carried out through a single stage.

Let us consider the problem of source extraction, in which the goal is to retrieve a single source from the mixtures. As discussed in Section II-A, source extraction can be conducted by estimating a vector $\mathbf{w}_i$ so that $\mathbf{y}_i = \mathbf{w}_i^T \mathbf{X}$ provides a good estimate of a given source. In the case of sparse sources, due to the action of the mixing process, the signals $\mathbf{x}_j$ are less sparse than the sources $\mathbf{s}_i$. Therefore, analogously to ICA, a natural approach to retrieve a source would be to adjust the extraction vector so that $\mathbf{y}_i$ be as sparse as possible.

In [44], extraction of sparse sources is conducted by considering a criterion based on the ℓ_1-norm, so the adjustment of $\mathbf{w}_i$ is carried out as follows:

$$\min_{\mathbf{w}_i} \quad ||\mathbf{y}_i||_1 = ||\mathbf{w}_i^T \mathbf{X}||_1 \\ \text{subject to} \quad ||\mathbf{w}_i||_2 = 1 \tag{29}$$

The restriction on the ℓ_2-norm is necessary here to avoid trivial solutions and implicitly assumes that the data is submitted to a whitening pre-processing stage. In [44], the authors have shown that (29) is indeed a contrast function when the sources are disjoint orthogonal. Besides, even when this condition is not observed, numerical experiments pointed out that the minimization of the ℓ_1-norm leads to source separation [44].

Alternatively, it is possible to retrieve sparse signals by means of separation criterion underpinned by the ℓ_0-pseudo-norm [45]. In this case, the resulting optimization problem can be expressed as follows:

$$\min_{\mathbf{w}_i} \quad ||\mathbf{y}_1||_0 = ||\mathbf{w}_i^T \mathbf{X}||_0 \\ \text{subject to} \quad \text{At least one element of } \mathbf{w}_i \text{ is not null.} \tag{30}$$

In [45], the authors proved that a sufficient condition to ensure the contrast property of (30) is given by

$$||\mathbf{s}_1||_0 < \tfrac{1}{2}||\mathbf{s}_2||_0 \\ ||\mathbf{s}_1||_0 < \tfrac{1}{2}\left(||\mathbf{s}_3||_0 - ||\mathbf{s}_2||_0\right) \\ ||\mathbf{s}_1||_0 < \tfrac{1}{2}\left(||\mathbf{s}_4||_0 - ||\mathbf{s}_3||_0 - ||\mathbf{s}_2||_0\right) \\ \vdots \\ ||\mathbf{s}_1||_0 < \tfrac{1}{2}\left(||\mathbf{s}_N||_0 - \sum_{i=2}^{N-1}||\mathbf{s}_i||_0\right), \tag{31}$$

In the particular case of $N = 2$ sources, such a condition can be simplified as [45]

$$||\mathbf{s}_1||_0 < \frac{1}{2}||\mathbf{s}_2||_0. \tag{32}$$

It is worth noticing that condition (31) allows a certain degree of overlapping between the sources, under the condition that they have different degrees of sparsity (in the sense of the ℓ_0 pseudo-norm). Another fundamental aspect here is that the obtained conditions are not expressed in a probabilistic fashion and do not require statistical independence. In other words, it is possible to separate sparse signals even in the cases in which ICA fails.

Concerning the practical implementation of methods based on (30), an important issue is related to the fact that real signals are not sparse in terms of the ℓ_0-pseudo norm. Indeed, actual signals that can be considered sparse often contain a few relevant coefficients and many coefficients that are close but not necessarily equal to zero. In order to overcome this problem, one can make use of smooth approximation for the ℓ_0-pseudo-norm as, for instance, the following one [46]

$$S_{\ell_0}(\mathbf{s}_i) = T - \sum_{n=1}^{T} \exp\left(-\frac{\mathbf{s}_i(n)^2}{2\sigma^2}\right), \tag{33}$$

where σ controls the smoothness of the approximation. If $\sigma \to 0$, then this smooth approximation approaches the ℓ_0-pseudo-norm.

V. CONCLUSION

This work is an introductory text about blind source separation and the most recent perspectives in domains beyond real or complex sets — which is the case of separation over finite fields —, and beyond the statistical independence assumption — which is the case of separation of sparse signals.

In the context of source separation over $GF(q)$, ICA-based strategies were discussed, putting emphasis on entropy-based cost functions to promote separation of signals whether the mixing model is instantaneous or convolutive. Both models implies a combinatorial optimization problem, which can be

solved via exhaustive-character search procedures or via bio-inspired strategies, e.g. the immune-inspired algorithms. Finally, two examples derived from coding theory show that BSS over Galois fields already offers preliminary contributions, in the sense of real applications.

In the case of separation of sparse signals, the two-step procedure usually employed for underdetermined models was first discussed. This approach considers sparsity both for estimating the mixing matrix and for solving the inverse problem associated with the sources estimation. In addition, the formulation of separation criteria based on sparsity for determined models was discussed. An interesting aspect, in this scenario, is that sparsity-based criteria can be applied even when sources are statistically dependent.

Naturally, the subjects that were introduced in this work are not fully explored here and, furthermore, have very interesting future perspectives, in the context of new algorithms and criteria, theoretical analyses and, ultimately, the potential association of sparsity with signals defined over a finite field.

ACKNOWLEDGMENT

REFERENCES

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, 1st ed. John Wiley & Sons, 2001.

[2] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation*, 1st ed. Academic Press, 2010.

[3] J. M. T. Romano, R. Attux, C. Cavalcante, and R. Suyama, *Unsupervised Signal Processing: Channel Equalization and Source Separation*, 1st ed. CRC Press, 2011.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[5] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, oct 1999. doi: 10.1109/78.790661

[6] O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312–321, mar 1990. doi: 10.1109/18.52478

[7] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, feb 1997. doi: 10.1109/78.554307

[8] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, aug 1999. doi: 10.1038/44565

[9] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, jun 1994. doi: 10.1002/env.3170050203

[10] L. T. Duarte, S. Moussaoui, and C. Jutten, "Source separation in chemical analysis: recent achievements and perspectives," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 135–146, may 2014. doi: 10.1109/MSP.2013.2296099

[11] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, mar 2010. doi: 10.1109/TASL.2009.2031510

[12] S. Moussaoui, D. Brie, and J. Idier, "Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, mar 2005, pp. v/289–v/292Vol.5.

[13] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, mar 2004.

[14] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4133–4145, nov 2006. doi: 10.1109/TSP.2006.880310

[15] A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–19, mar 2012. doi: 10.1186/1687-6180-2012-52

[16] R. Suyama, L. T. Duarte, R. Ferrari, L. E. P. Rangel, R. R. d. F. Attux, C. C. Cavalcante, F. J. Von Zuben, and J. M. T. Romano, "A Nonlinear Prediction Approach to the Blind Separation of Convolutive Mixtures," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 043860, sep 2007. doi: 10.1155/2007/43860

[17] A. Yeredor, "ICA in Boolean XOR Mixtures," in *Independent Component Analysis and Signal Separation*. Berlin, Heidelberg: Springer Berlin Heidelberg, sep 2007, pp. 827–835.

[18] D. Hankerson, S. Vanstone, and A. Menezes, "Finite Field Arithmetic," in *Guide to Elliptic Curve Cryptography*. Springer, 2004, no. 1, pp. 25–73.

[19] J. B. Fraleigh and V. J. Katz, *A First Course in Abstract Algebra*. Addison-Wesley, 2003.

[20] H. W. Gutch, P. Gruber, A. Yeredor, and F. J. Theis, "ICA over finite fields - Separability and algorithms," *Signal Processing*, vol. 92, no. 8, pp. 1796–1808, aug 2012. doi: 10.1016/j.sigpro.2011.10.003

[21] H. W. Gutch, P. Gruber, and F. J. Theis, *ICA over Finite Fields*. Berlin, Heidelberg: Springer Berlin Heidelberg, sep 2010, pp. 645–652.

[22] A. Yeredor, "Independent Component Analysis Over Galois Fields of Prime Order," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5342–5359, aug 2011. doi: 10.1109/TIT.2011.2145090

[23] D. G. Silva, E. Z. Nadalin, J. Montalvao, and R. Attux, "The Modified MEXICO for ICA Over Finite Fields," *Signal Processing*, vol. 93, no. 9, pp. 2525–2528, sep 2013. doi: 10.1016/j.sigpro.2013.03.021

[24] D. G. Silva, E. Z. Nadalin, G. P. Coelho, L. T. Duarte, R. Suyama, R. Attux, F. J. Von Zuben, and J. Montalvão, "A Michigan-like immune-inspired framework for performing independent component analysis over Galois fields of prime order," *Signal Processing*, vol. 96, pp. 153–163, mar 2014. doi: 10.1016/j.sigpro.2013.09.004

[25] L. N. De Castro and J. Timmis, *Artificial Immune Systems: a New Computational Intelligence Approach*. Springer, 2002.

[26] W. C. Waterhouse, "How often do determinants over finite fields vanish?" *Discrete Mathematics*, vol. 65, no. 1, pp. 103–104, may 1987. doi: 10.1016/0012-365X(87)90217-2

[27] D. G. Silva, R. Attux, E. Z. Nadalin, L. T. Duarte, and R. Suyama, "An Immune-Inspired Information-Theoretic Approach to the Problem of ICA Over a Galois Field," in *2011 IEEE Information Theory Workshop*. IEEE, oct 2011. doi: 10.1109/ITW.2011.6089571 pp. 618–622.

[28] D. G. Silva, J. Montalvao, and R. Attux, "cobICA: A concentration-based, immune-inspired algorithm for ICA over Galois fields," in *2014 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP)*. Orlando, FL: IEEE, dec 2014. doi: 10.1109/CIMSIVP.2014.7013284 pp. 1–8.

[29] D. G. Fantinato, D. G. Silva, E. Z. Nadalin, R. Attux, J. M. T. Romano, A. Neves, and J. Montalvao, "Blind Separation of Convolutive Mixtures Over Galois Fields," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, sep 2013. doi: 10.1109/MLSP.2013.6661908 pp. 1–6.

[30] F. Dietrich, *Robust Signal Processing for Wireless Communications (Foundations in Signal Processing, Communications and Networking) (No. 2)*. Berlin, Germany: Springer, 2008.

[31] I.-D. Nemoianu, C. Greco, M. Cagnazzo, and B. Pesquet-Popescu, "On a Hashing-Based Enhancement of Source Separation Algorithms Over Finite Fields With Network Coding Perspectives," *IEEE Transactions on Multimedia*, vol. 16, no. 7, pp. 2011–2024, nov 2014. doi: 10.1109/TMM.2014.2341923

[32] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge University Press, 2010.

[33] P. Georgiev, F. Theis, A. Cichocki, and H. Bakardjian, *Data Mining in Biomedicine*. Springer, 2005, ch. Sparse component analysis: a new tool for data mining, pp. 91–116.

[34] L. T. Duarte, R. R. Lopes, J. H. Faccipieri, J. M. T. Romano, and M. Tygel, "Separation of reflection from diffraction events via the crs technique and a blind source separation method based on sparsity maximization," in *Proceedings of the 13th International Congress of the Brazilian Geophysical Society (SBGf)*, aug 2013.

[35] C. Févotte and S. J. Godsill, "A bayesian approach for blind separation of sparse sources," *IEEE Transactions on Audio, Speech*

and Language Processing, vol. 14, pp. 2174–2188, nov 2006. doi: 10.1109/TSA.2005.858523

[36] M. Elad, *Sparse and redundant representations from theory to applications in signal and image processing*.   Springer, 2010.

[37] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing n sources from 2 mixtures," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 5.   IEEE, jun 2000. doi: 10.1109/ICASSP.2000.861162 pp. 2985–2988.

[38] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal processing*, vol. 81, no. 11, pp. 2353–2362, nov 2001. doi: 10.1016/S0165-1684(01)00120-7

[39] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, apr 2001. doi: 10.1162/089976601300014385

[40] E. J. Candës and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, mar 2008. doi: 10.1109/MSP.2007.914731

[41] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*.   Cambridge University Press, 2012.

[42] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, jul 2004. doi: 10.1109/TSP.2004.828896

[43] Y. Deville, "Sparse component analysis: A general framework for linear and nonlinear blind source separation and mixture identification," in *Blind Source Separation*.   Springer, may 2014, pp. 151–196.

[44] E. Z. Nadalin, A. K. Takahata, L. T. Duarte, R. Suyama, and R. Attux, *Blind Extraction of the Sparsest Component*.   Berlin, Heidelberg: Springer Berlin Heidelberg, sep 2010, pp. 394–401.

[45] L. T. Duarte, R. Suyama, R. Attux, J. M. T. Romano, and C. Jutten, "Blind extraction of sparse components based on l0-norm minimization," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, jun 2011. doi: 10.1109/SSP.2011.5967775 pp. 617–620.

[46] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed l0 norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, jan 2009. doi: 10.1109/TSP.2008.2007606

Romis Attux was born in Goiânia, Brazil, in 1978. He received the titles of Electrical Engineer (1999), Master in Electrical Engineering (2001) and Doctor in Electrical Engineering (2005) from the University of Campinas (UNICAMP), Brazil. Currently, he is an associate professor at the same institution. His main research interests are adaptive filtering, computational intelligence, dynamical systems / chaos and brain-computer interfaces.

Daniel G. Silva was born in Botucatu, Brazil, in 1983. He received the B.S. degree in computer engineering and the M.S. and Ph.D. degrees in electrical engineering, all from the University of Campinas (UNICAMP), São Paulo, Brazil, in 2006, 2009, and 2013, respectively. Currently, he is a Professor at the Department of Electrical Engineering (ENE) of the University of Brasília (UnB). His main research interests are information theoretic learning, adaptive signal processing and computational intelligence.

Leonardo T. Duarte received the B.S. and the M.Sc. degrees in electrical engineering from UNICAMP (Brazil) in 2004 and 2006, respectively, and the Ph.D. degree from the Grenoble Institute of Technology (Grenoble INP), France, in 2009. He is currently an assistant professor at the School of Applied Sciences at UNICAMP. His research interests are mainly associated with the theory of unsupervised signal processing and include signal separation, independent component analysis, Bayesian methods, and applications in chemical sensors and seismic signal processing. He is also working on unsupervised schemes for adjusting multiple-criteria decision analysis (MCDA) techniques. He is a Senior Member of the IEEE.