An Empirical Comparison of ICA Over Finite Fields Algorithms

Majid Rezaee e Daniel G. Silva

Abstract — **A series of algorithms to perform ICA over Finite Fields have been proposed in recent years. They differ from each other in terms of search strategy and independence criterion, which may yield different performances. Hence, this work intends to compare the two most relevant heuristic-based approaches, namely AMERICA and MEXICO, with the recently developed cobICA, which is based on a state-of-the-art immune inspired optimization technique. Experimental results indicate AMERICA with the best separation quality, while cobICA presents intermediate results but with a tendency of smaller computational cost in high dimension scenarios.**

Keywords — ICA, Galois Fields, Entropy, BSS

I. Introduction

Independent Component Analysis (ICA) has been shown to be an important approach to deal with the familiarly-known Blind Source Separation (BSS) problem [1], when independent, unknown source signals that were linearly mixed need to be recovered.

While the BSS/ICA problem is well-known for real- or complex-valued signals, recent works [2, 3, 4] gave the first steps to extend the BSS/ICA problem for Galois fields (GF), as well. In this context, the problem can be formulated as a combinatorial optimization problem with a cost function that estimates the independence degree between the extracted components.

A series of strategies have been proposed to deal with such an optimization task. The two pioneer algorithms AMERICA and MEXICO [3, 4] adopt heuristics which aim to minimize the entropy of linear combinations of the mixtures; on the other hand, the cobICA algorithm [5] employs a different search strategy, based on the immune-inspired algorithm cob-aiNet[C] and on a different criterion – the minimal mutual information (MMI).

Naturally, such techniques differ in terms of computational cost and separation performance, then a comparative analysis seems necessary. Hence, this work proposes to study and compare AMERICA, MEXICO and cobICA algorithms in terms of separation rate and computational burden. As the reader shall see, the simulation results show that cobICA can achieve smaller but similar separation performances to AMERICA, the best of all methods, while it presents a lower computational complexity when confronted in higher dimension scenarios.

The paper is organized as follows. Section II presents the BSS/ICA framework in its linear-instantaneous model. In Section III, the descriptive analysis of AMERICA, MEXICO and cobICA algorithms are introduced. Then, in order to compare the behavior of the three techniques altogether and to analyze the potential benefits of each one in different scenarios, a set of numerical simulations are shown and discussed in Section IV and, finally, conclusions are drawn in Section V.

Majid Rezaee and Daniel G. Silva, Department of Electrical Engineering, School of Technology, University of Brasília (UnB), Brasília-DF, Brazil, E-mails: majid.rezaee.un@gmail.com, danielgs@ene.unb.br. This work was partially supported by CAPES.

II. The BSS/ICA over Finite Fields Problem

Blind Source Separation deals with recovering an unknown set of sources from an observable set of mixed signals [1]. This framework has been applied over a wide range of engineering fields such as: array signal processing and wireless communication [6, 7], geophysical exploration [8], biomedical signal processing [9, 10], speech processing [11] and image processing [12].

In mathematical terms, the linear-instantaneous model of BSS is defined as:

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

where $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_N(n)]^{\mathrm{T}}$ is a vector of N observed random signals at the instant n, obtained from the mixing of the (unknown) source vector $\mathbf{s}(\mathrm{n}) = [s_1(n), s_2(n), \dots, s_N(n)]^{\mathrm{T}}$ by an (unknown) invertible $(N \times N)$ matrix $\mathbf{A}$.

Hence the problem involves to recover the original sources by estimating a separating matrix $\mathbf{W}$ such that

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) = \mathbf{P}\mathbf{D}\mathbf{A}^{-1}\mathbf{x}(n) = \mathbf{P}\mathbf{D}\mathbf{s}(n). \tag{2}$$

By applying an ICA method, i.e. if $\mathbf{y}(n)$ recovers the independence condition among its components, it is possible to demonstrate [1, 4, 13] that such technique will result in the recovery of the original sources up to scale (diagonal matrix $\mathbf{D}$) and permutation (permutation matrix $\mathbf{P}$) ambiguities.

Such definitions are valid for any field, like the real or complex numbers, as well for the case of a finite/Galois field $GF(q)$, $q = P^n$, where P is a prime and n is a positive integer. In the case that $n = 1$, the field is called a prime field, where the elements can be defined as the set $\{0, 1, \dots, P-1\}$ and the operations are defined as sum and product modulo P [14]. In the case of non-prime fields, we need to define the symbolic operations on polynomials, which is naturally trickier to be implemented [4].

The matrix $\mathbf{W}$ should be invertible i.e. it is part of the general linear group $GL(N, q)$, which is finite and has a number of elements given by [15]

$$|GL(N, q)| = \prod_{k=0}^{N-1} (q^N - q^k). \tag{3}$$

Note by (3) that the process of searching for the separating matrix $\mathbf{W}$ involves a finite space of solutions, whose size increases exponentially with the number of sources. Consequently, performing ICA over $GF(q)$ involves the solution of a combinatorial optimization problem where the cost function – which measures the dependence between the extracted components $\mathbf{y}(n)$ – must be minimized.

III. ALGORITHMS

After presenting the problem definition, we can discuss the three ICA over $GF(q)$ algorithms that are analyzed in this work. As the reader will see, in the following subsections, each one employs specific search heuristics and criteria in order to obtain estimates of the independent components.

A. The AMERICA algorithm

The algorithm was introduced in 2007, in the context of boolean "Exclusive Or" (XOR) mixtures [2], and a generalized version was proposed in 2010 in order to encompass fields with arbitrary order [4]. The method is supported by an important property that states, in simple terms, that a linear mixing process of independent random signals never causes a reduction of entropy [16].

By exploring the aforementioned property, the technique tries to recover the lower entropy configuration prior to the mixing. Hence, one source at a time is extracted by determining the lowest entropy linear combination of the mixtures, where each set of coefficients that extracts a source is linearly independent from the previously chosen, such that:

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \mathrm{H}(y = \mathbf{w}^{\mathrm{T}}\mathbf{x}), \tag{4}$$

where y is an estimate of one of the extracted components, $H(\cdot)$ is the entropy of a random variable, i.e. $H(X) = -\sum p_X(x)\log_2 p_X(x)$, and $\mathbf{w}_{opt}$ is the extraction vector, which yields that the lines of the separating matrix are the corresponding extraction vectors for the N sources. Note that, for the sake of simplicity, we omit the instant index n.

One can see that there are $q^N - 1$ non-trivial linear combinations which should have the corresponding entropy evaluated, for each source signal. Therefore, the AMERICA method computational cost, in terms of cost function evaluations, is $\Theta(Nq^N)$ [3].

B. The MEXICO algorithm

The MEXICO algorithm [4] follows the same basic idea of iteratively estimating the independent components by reducing their entropy values, as well as AMERICA, however, the construction scheme of the separating matrix and the search process are both different.

The main difference is that the method makes an equivalent entropy evaluation only between combined *pairs* of mixtures. In other words, given two mixtures observations at a given instant, i.e. $x_i(n) = x_i$ and $x_j(n) = x_j$ and a constant $c \in F = GF(q)$, such that

$$H(x_i + c \cdot x_j) < H(x_i), \tag{5}$$

then $c \cdot x_j$ can be removed from x_i as a step-by-step "demixing" process, where x_i is replaced by $x_i + c \cdot x_j$. This substitution can be formulated in a matrix representation, as follows:

$$\mathbf{T}_{i,j}(c) = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & c & & \ddots & \\ & & & & 1 \end{bmatrix}, \tag{6}$$

where $\mathbf{T}_{i,j}(c) \in GL(N,q)$ is an $N \times N$ identity matrix with a value c (instead of 0) in the position (i,j).

This process is repeated for all $N(N-1)$ possibilities of pairwise combinations, with $i \neq j$, and is called a "*sweep*". Then, while there is a *sweep* with some improvement (i.e., a substitution which reduces the entropy), the process is repeated, otherwise the algorithm is finished.

The computational complexity of the MEXICO algorithm, which is iterative, involves counting the number of *sweeps* and the number of cost function evaluations in each *sweep* [3]. Since the stopping condition is not deterministic, there is no fixed number of entropy evaluations as in AMERICA. Nevertheless, MEXICO has shown to be a faster method than AMERICA, but at the cost of poorer separation rates, as the simulations in [4] indicates.

C. The cobICA algorithm

The cobICA algorithm [5] is an immune-inspired algorithm to perform ICA over Galois Fields, which applies the cob-aiNet[C] algorithm with the minimal mutual information (MMI) criterion and specific full-rank-preserving operators, in order to obtain the optimal separating matrix $\mathbf{W}$.

Immune-inspired algorithms try to emulate features of the immune system in order to solve engineering tasks, e.g. pattern recognition and optimization [17]. Despite key differences, they share some ideas of population-based algorithms such as genetic algorithms and evolutionary strategies. In particular, cobICA models the problem such that each individual in a *population of cells* is a candidate to be the separating matrix and the search space is limited by adopting appropriate mutation and local search operations, which consequently allows the search only over the elements in $GL(N,q)$.

As mentioned in Section II, since the sources are mutually independent, if one finds a separating matrix that yields statistically independent signals, it is possible to recover the original sources. Then, the MMI criterion is employed to guide the search process because it is known that independent signals yield a null value of *mutual information*, an important information-theoretic measure defined as [16]

$$I(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{y}), \tag{7}$$

where $\mathbf{y}$ is the separating system output and $H(\cdot)$ is the already known entropy function.

The authors of [5] show that the search can be performed by minimizing only the first term of (7), leading to the following optimal solution:

$$\mathbf{W}_{opt} = \arg\min_{\mathbf{W}} \sum_{i=1}^{N} \mathrm{H}(y_i), \qquad \mathbf{y} = \mathbf{W}\mathbf{x}. \tag{8}$$

Since the probability distributions of the mixtures are not known in advance, an entropy estimator is employed over a set of T independent and identically distributed (iid) observations of each mixed signal, $\{x_i(1), \dots, x_i(T)\}$, in this case the maximum-likelihood estimator with the Miller-Madow bias correction [17]:

$$\hat{H}(y_i) = \frac{q-1}{2T} - \sum_{j \in GF(q)} \hat{p}_{y_i}(j) \log_q \hat{p}_{y_i}(j), \qquad (9)$$

where the probabilities are estimated according to the relative frequency of each symbol.

According to the cost function defined in (8), the population of candidate solutions is iteratively evolved through hypermutation and local search operators, and the ICA solution is the individual with the best fitness (minimal mutual information), when the algorithm finishes.

The cobICA hypermutation operator resembles the process of MEXICO algorithm. It is a routine that performs the linear combination of two randomly-chosen rows from the candidate matrix and replaces one of the original sequences, which yields a new individual defined as

$$\mathbf{B}' = \mathbf{T}_{i,j}(k)\mathbf{B}, \qquad (10)$$

where $\mathbf{T}_{i,j}(k)$ is exactly the same linear operator described in (6). Differently than MEXICO method, note that substitution occurs independently if it diminishes (or not) the entropy of the resulting signal.

On the other hand, the same operator stated in (10) is used for the local search process, but in this case, the routine updates the individual *only if* fitness improves (the cost function defined in (8) decreases). If this occurs, the procedure stops; otherwise, it continues testing all the possibilities of $i \leq N, j \leq N, i \neq j$.

This strategy also resembles MEXICO's *sweep* operations, however, the update in cobICA is performed based on a distinct cost function, the mutual information between *all* the extracted signals, while the former employs as update criterion the entropy of a *single one* extracted signal. Naturally, this may lead MEXICO to a higher risk of local convergence, despite a reduced computational cost associated with its cost function evaluation.

IV. EMPIRICAL ANALYSIS

Finally, in this section, numerical simulations results are presented for a comparative performance evaluation of cobICA, AMERICA and MEXICO techniques. The analysis is performed in different contexts, considering various numbers of samples (from $2^5 = 32$ to $2^{10} = 1024$ samples), of sources and of field orders. The performance metric is the *average success rate*: it is the mean ratio between the numbers of extracted sources and N.

The average metric is calculated for each algorithm via the mean of 20 Monte Carlo runs, where sources are generated randomly, i.e. for a single trial, the sources are generated one by one according to randomly-defined, non-uniform[1] probability distributions. The mixing matrix $\mathbf{A}$ is also randomly generated, which yields T observations from each mixture x_i to be applied as input for the algorithms.

The cob-aiNet[C] parameters used by cobICA are the same recommended in [5], i.e. maximum of 100 individuals in the population and maximum of 300 algorithm iterations.

A. Comparative simulations for variable number of samples

In the following, a comparative analysis takes place. Figures 1, 2 and 3 show the comparison between the techniques in terms of average success rate, when the number of samples increases (from 2^5=32 to 2^{10}=1024), with prime field orders $q = P = 2$, 3 and 5.

As Figure 1 shows, when the algorithms deal with less number of sources, all of them could achieve almost the same performance, which was expected because of the relatively small size of the search space. Moreover, the performances improve with higher number of samples, as expected.

On the other hand, with the addition of two more components, in the second case, a subtle discrepancy between the performances appears, with MEXICO presenting a quality level similar to cobICA, while AMERICA still maintains a perceivable margin and presents the best performance among all methods. Furthermore, this behavior of MEXICO confirms the results obtained in [4], where it also succeeded for $P = 2$ and $P = 3$.

Figure 2 shows the results for ternary fields. The performances of AMERICA and cobICA achieve full success in separation when $N = 4$, while MEXICO presents a lower performance than both. Clearly here we can see that, by increasing the search space, even for low dimension cases, where we have less number of sources, MEXICO does not achieve the same level of quality as presented in Figure 1, in contrast with the cobICA algorithm, which shows its best performance for $N = 4$ and $GF(3)$. In the second scenario ($N = 6$) the performances decrease due to the search space increasing (as expected), but AMERICA continues as the top-

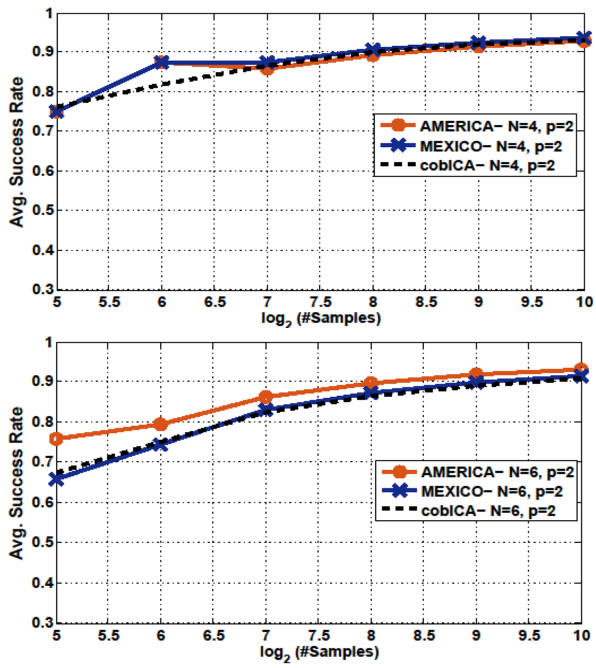

Fig. 1. Comparison among the techniques for field order $P = 2$.

[1] ICA over GF can recover independent sources only for non-uniform and non-degenerate distributions [4].

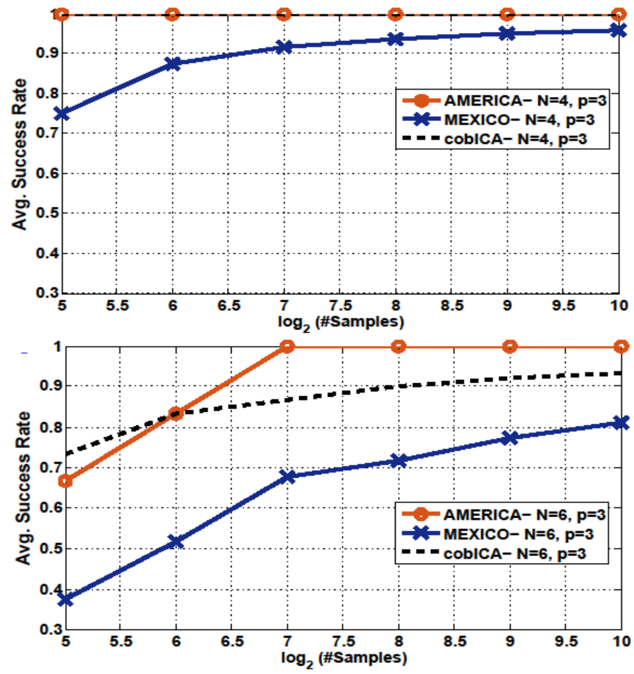

Fig. 2. Comparison among the techniques for field order $P = 3$.

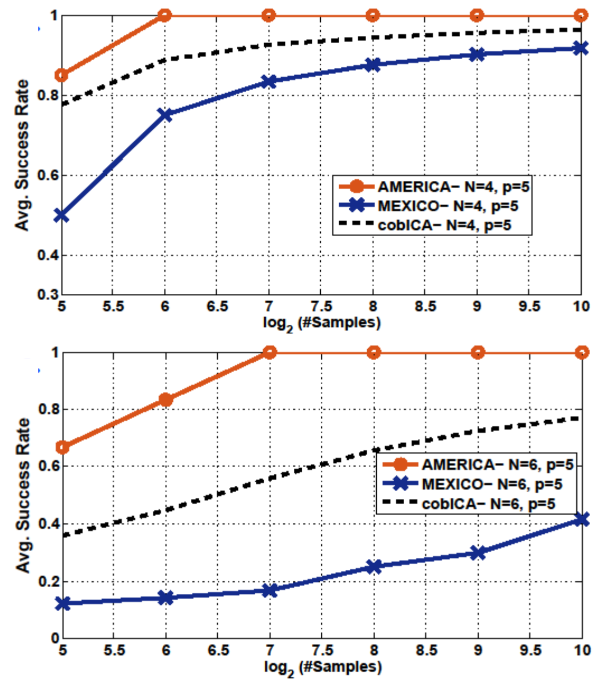

Fig. 3. Comparison among the techniques for field order $P = 5$.

ranked method, closely followed by cobICA and, as the last-ranked strategy, by MEXICO.

Finally, the first scenario of Figure 3 indicates cobICA and AMERICA algorithms with a close performance, while the line representing MEXICO shows that it performs close to the other algorithms only for an elevated number of samples, otherwise it does not present the same quality level. In the scenario $N = 6$, one can see that, although cobICA couldn't achieve a quality performance as good as AMERICA, it still remains better than MEXICO.

The overall results of this section experiments show that AMERICA and cobICA have a quite similar performance in low dimension cases, while MEXICO does not achieve the same quality of results.

B. Computational complexity comparison

The comparative results that were discussed in the previous sections point out that AMERICA yields the best separation quality, followed by the cobICA technique and, then, the MEXICO algorithm presents the lowest general performance. Notwithstanding, the three approaches present fundamental differences concerning the implementation of the search strategy and of the criterion, which may imply different computational demands when each one is executed.

In this context, this subsection experiment compares the computational costs of cobICA, AMERICA and MEXICO, by considering the number of times that the crucial function that supports each algorithm criterion, the entropy of a given component, is evaluated for each method. The field order is $P = 3$, a fixed number of samples is adopted ($T = 512$), with a varying number of sources ($N = 8, 10$ and 12).

As the authors mention in [3], for the AMERICA algorithm, an exhaustive search over all non-trivial candidate vectors is executed to extract each source, then $N.(q^N - 1)$ values of entropy are estimated, i.e. AMERICA complexity is $\Theta(Nq^N)$, as it is mentioned in Section III.A. Differently than AMERICA, MEXICO and cobICA have a non-deterministic number of entropy calculations, which is dependent on the convergence to the optimal solution, hence a numerical estimate of the average computational complexity of both methods is calculated, via the average number of entropy function evaluations over 20 independent runs.

Figure 4 shows that MEXICO offers the smallest computational complexity, however, this benefit comes with the burden of the poorest overall quality in separation, as seen in the previous subsection. Moreover, AMERICA has lower values than cobICA except for $N = 12$. It seems that, although cobICA starts with a relatively high level, it presents a smaller increasing trend than AMERICA, with respect to the number of sources. This comparison can indicate that cobICA, specifically, has an intermediate asymptotic computational cost, taking place between the most expensive AMERICA and the cheapest MEXICO.

V. Final Remarks

This work puts emphasis on three relevant algorithms to perform independent component analysis over Galois Fields, namely AMERICA, MEXICO and cobICA algorithms. A comparative analysis among the techniques is performed, in order to elucidate aspects of separation quality and computational cost.

AMERICA algorithm clearly presented the highest rates of correctly extracting independent components. On the other hand, MEXICO algorithm achieves the lowest scores and cobICA has an intermediate behavior. In terms of computational cost, we considered the number of entropy function evaluations as figure of merit and the results indicate MEXICO as the one with the lowest cost. Interestingly, cobICA presented a lower cost than AMERICA in the case of a higher number of components, which may indicate that this technique presents a compromise between scalability and separation quality, considering time constraints, specially when the number of sources is increased.

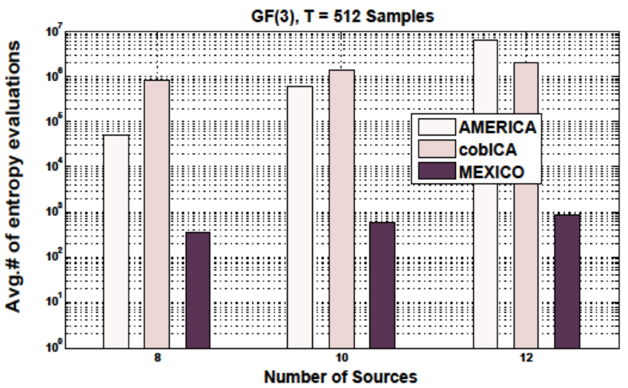

Fig. 4. Number of entropy evaluations for different values of sources when field order is $P = 3$.

ICA over Galois Fields naturally still demands theoretical developments, since it is a quite recent research subject. Despite the initial attempts in the context of coding theory problems [4, 19], seeking for improvements that expand the application possibilities of the algorithms that were discussed in this work seems a particularly promising trend.

Consequently, we consider as future work the study of potential applications for the algorithms; the association of different information-theoretic estimators with the criterion evaluation task; and a more thorough comparison between the techniques, considering larger dimensions and more emphasis on the computational cost.

REFERENCES

[1] COMON, P. and JUTTEN, C. (Ed.). (2010). "Handbook of Blind Source Separation." In: Oxford, UK: Academic Press, 2010.

[2] YEREDOR, A. (2007). "ICA in Boolean XOR mixtures." In: ICA2007—Independent Component Analysis and Signal Separation, Springer, 2007, pp.827–835.

[3] YEREDOR, A. (2011). "Independent Component Analysis Over Galois Fields of Prime Order." In: IEEE- Transactions on Information Theory, v. 57, n. 8, p. 5342–5359.

[4] GUTCH, H. W.; GRUBER, P.; YEREDOR, A. and THEIS, F. J. (2012). "ICA Over Finite Fields Separability and Algorithms." In: Signal Processing, Elsevier, v. 92, n. 8, p. 1796–1808, ago. 2012.

[5] SILVA, D. G.; MONTALVAO, J. and ATTUX, R. (2014). "cobICA: A Concentration-Based, Immune-Inspired Algorithm for ICA Over Galois Fields." In: Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2014 IEEE Symposium on (pp. 1–8). Orlando, FL: IEEE.

[6] AMARI, S.-I. and CICHOCKI, A. (1998). "Adaptive blind signal processing - neural network approaches." In: Proceedings of the IEEE, vol.86, (9), pp.2026-2048.

[7] MANSOUR, A.; BARROS, A. K. and OHNISHI, N. (2000). "Blind separation of sources: Methods, assumptions and applications", In: IEICE Trans.

[8] MAKEIG, S.; BELL, A. J.; JUNG, T.-P; and SEJNOWSKI, T.-J. (1996). "Independent component analysis of electroencephalographic data." In: Advances in Neural Information Processing Systems 8, pages 145-151. MIT Press.

[9] VIGARIO, R.; SARELL, J.; JOUSML, AKI; HAMALAINEN, M. and OJA, E. (2000). "Independent component approach to the analysis of EEG and MEG recordings." In: IEEE Transactions on Biomedical Engineering, 47(5):589{593, 2000}.

[10] IRIARTE, J; URRESTARAZU E VALENCIA, M.; ALERGE M.; MALANDA A.; VITERI C and ARTIEDA J. (2003). "Independent Component Analysis as a tool to eliminate artifacts in EEG: A quantitative study." In: J. Clin. Neurophysiol. 2003 Jul-Aug; 20(4): 249-57.

[11] LEE, T. W. (1998). "Independent Component Analysis." In: Kluwer Academic Publishers.

[12] LEE, T. W. and LEWICKI, M. S. (2000). "Unsupervised classification, segmentation and denoising of images using ICA mixture models." In: IEEE Trans. On Image Processing.

[13] HYVARINEN, A.; KARHUNEN, J. and OJA, E. (2001). "Independent Component Analysis." In: John Wiley & Sons.

[14] LIDL, R. and NIEDERREITER, H. (1997). "Finite Fields." In: vol.20, Cambridge University Press.

[15] WATERHOUSE, W. C. (1987). "How often do determinants over finite fields vanish?" In: Discrete Mathematics, v. 65, n. 1, p. 103–104.

[16] COVER, T.M. and THOMAS, J.A. (2006). "Elements of Information Theory." In: 2nd ed. Wiley-Interscience.

[17] COELHO, G.P.; DE FRANCA, F.O. and VON ZUBEN, F.J. (2011). "A concentration-based artificial immune network for combinatorial optimization." In: 2011 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2011, pp. 1242–1249.

[18] CARLTON, A.G (1969). "On the bias of information estimates." In: Psychological Bulletin 71(2)(1969)108–109.

[19] NEMOIANU, I.-D.; GRECO, C.; CAGNAZZO, M. and PESQUET-POPESCU, B. (2014). "On a Hashing-Based Enhancement of Source Separation Algorithms Over Finite Fields With Network Coding Perspectives." In: IEEE Transactions on Multimedia, v. 16, n. 7, p. 2011–2024.