



**A Study on Independent Component Analysis
Over Galois Fields**

Sayed Majid Rezaee

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Brasília, Dezembro de 2015

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**A STUDY ON INDEPENDENT COMPONENT ANALYSIS
OVER GALOIS FIELDS**

SAYED MAJID REZAEI

ORIENTADOR: DANIEL GUERREIRO E SILVA

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPGEE.DM - 610/2015

BRASÍLIA/DF: DEZEMBRO – 2015

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

A STUDY ON INDEPENDENT COMPONENT ANALYSIS OVER
GALOIS FIELDS

SAYED MAJID REZAEI

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA
ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

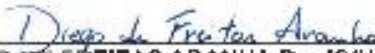
APROVADA POR:



DANIEL GUERREIRO E SILVA, Dr., ENE/UNB
(ORIENTADOR)



EDSON MINTSU HUNG, Dr., ENE/UNB
(EXAMINADOR INTERNO)



DIEGO DE FREITAS ARANHA Dr., IC/UNICAMP
(EXAMINADOR EXTERNO)

Brasília, 10 de dezembro de 2015.

FICHA CATALOGRÁFICA

REZAE, SAYED MAJID

A Study on Independent Component Analysis Over Galois Fields [Distrito Federal] 2015. xiii, 53p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2015).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

1. Teoria da informação

2. Sistemas imunológico

3. Análise de componentes independentes

4. Processamento digital de sinais

I. ENE/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

REZAE, SAYED MAJID, (2015). A Study on Independent Component Analysis Over Galois Fields. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGEE.DM – 610/2015, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 53p.

CESSÃO DE DIREITOS

AUTOR: Sayed Majid Rezaee.

TÍTULO: A Study on Independent Component Analysis Over Galois Fields.

GRAU: Mestre

ANO: 2015

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

Sayed Majid Rezaee
majid.rezaee.un@gmail.com
Brasília – DF – Brasil.

... dedicated to my dear parents,

Azar and Solieman

in love and gratitude.

ACKNOWLEDGMENTS

First, I would like to thank my professors at the department of Electronic Engineering, who helped me enormously with their constant feedback throughout my studying at University of Brasilia - UnB.

I wish to express my most sincere gratitude and appreciation to my advisor, Prof. Dr. Daniel Guerreiro e Silva, for his continuous guidance and support during the course of this work. I have the great admiration for his excellent work ethic and dedication to the science.

I am truly indebted to the Coordinator of PPGEE-UnB, Prof. Dr. Ugo Silva Dias, for his valuable advices, compassion, and help during my study at UnB.

My Special thanks goes to my thesis committee members, Prof. Dr. Diego de Freitas Aranha (IC-Unicamp) and Prof. Dr. Edson Mintsu Hung (ENE-UnB) for serving as my committee members and also for all of their guidance through this process.

I would also like to wish my deepest thanks to Prof. Dr. André Noll Barreto and Prof^a. Dr. Suélia de Siqueira Rodrigues Fleury Rosa for their scientific help during the research programs.

I am also thankful to all my friends in Brazil for the quality leisure time and memorable moments.

Finally, I would express a deep sense of gratitude to my family, whom I owe my life for their constant love, encouragement, moral support and blessings in every aspect of my life.

Last night the Shaikh went all about the city, lamp in hand, crying,

“I am weary of beast and devil, a man is my desire.”

They said, “He is not to be found, we too have searched.”

He answered, “He who is not to be found is my desire.”

(Mawlānā- Rūmī)

RESUMO

Um Estudo em Análise de Componentes Independentes em Corpos de Galois.

Nas últimas décadas, o problema de separação cega de fontes (BSS, do inglês *Blind Source Separation*) – que trata de estimar um conjunto desconhecido de fontes de sinais a partir de versões misturadas destes – tornou-se relevante em vários campos da engenharia, incluindo o processamento matricial, comunicações sem fio, processamento de sinais médicos, processamento de voz e engenharia biomédica.

A fim de resolver o problema de BSS no contexto de modelos lineares, considerando-se várias técnicas possíveis, a Análise de Componentes Independentes (ICA, do inglês *Independent Component Analysis*) – que utiliza a independência estatística das fontes como uma premissa – demonstrou ser uma das mais importantes estratégias de solução. Além disso, embora o modelo de BSS/ICA para sinais reais ou complexos esteja bem estabelecido, a recente perspectiva de uma formulação do problema com sinais e modelos definidos em corpos de Galois oferece várias possibilidades de análise e contribuições.

Esta dissertação de mestrado realiza um estudo da Análise de Componentes Independentes em corpos de Galois, considerando os conceitos teóricos e abordagens para o problema, assim como dos algoritmos estado-da-arte até agora propostos, em termos de suas capacidades de separação e custo computacional. Especificamente, as técnicas dos algoritmos AMERICA e MEXICO são estudadas juntamente com o algoritmo cobICA.

Como as simulações experimentais indicam, devido à sua complexidade computacional menor e uma qualidade de desempenho satisfatório, o algoritmo cobICA apresenta-se como uma solução de compromisso entre os algoritmos AMERICA e MEXICO para executar BSS/ICA em corpos de Galois.

ABSTRACT

A Study on Independent Component Analysis Over Galois Fields.

Over the past decades, the Blind Source Separation (BSS) problem – which deals with estimating an unknown set of source signals from their measured mixtures –has become prevalent in several engineering fields, including array processing, wireless communications, medical signal processing, speech processing and biomedical engineering.

In order to solve the BSS problem in the context of linear models, considering several possible techniques, Independent Component Analysis (ICA) – which uses statistical independence of the source signals as a premise – has been shown to be one of the most important approaches. Furthermore, although the BSS/ICA framework for real- or complex-valued signals is firmly established, the recent perspective of a BSS/ICA formulation where the signals and models are defined over Galois fields gives several possibilities of analyzes and contributions.

This Master’s thesis performs a study on Independent Component Analysis over Galois fields, considering the theoretical concepts and aspects of the problem and the investigation, in terms of capability and efficiency, of the state-of-the-art algorithms so far introduced.

In this context, AMERICA and MEXICO techniques are studied, along with cobICA algorithm – a bioinspired framework based on cob-aiNet[C] immune-inspired algorithm –, mainly focusing on comparing the quality of separation and on discussing the computational burden of each technique.

As the experimental simulations indicate, due to its lower computational complexity and a satisfactory performance quality, cobICA takes place as a compromise solution between AMERICA and MEXICO algorithms, to perform BSS/ICA over Galois fields.

CONTENTS

| | |
|---|-----------|
| 1 - INTRODUCTION..... | 1 |
| 2 - INDEPENDENT COMPONENT ANALYSIS OVER GALOIS FIELDS | 4 |
| 2.1 - BLIND SOURCE SEPARATION AND INDEPENDENT COMPONENT ANALYSIS..... | 4 |
| 2.2 - BSS/ICA OVER GALOIS FIELDS..... | 9 |
| 2.3 - INFORMATION-THEORETIC MEASURES FOR DISCRETE SOURCES..... | 12 |
| 2.4 - CANONICAL ALGORITHMS FOR ICA OVER GF..... | 14 |
| 2.4.1 - The AMERICA algorithm..... | 14 |
| 2.4.2 - The MEXICO algorithm..... | 16 |
| 2.5 - CONCLUDING REMARKS..... | 18 |
| 3 - BIOINSPIRED SEARCH AND INFORMATION THEORETIC LEARNING TO PERFORM ICA OVER GALOIS FIELDS: THE cobICA ALGORITHM..... | 19 |
| 3.1 - ARTIFICIAL IMMUNE SYSTEMS..... | 19 |
| 3.2 - THE cob-aiNet [C] ALGORITHM..... | 21 |
| 3.2.1 - Representation and affinity metrics..... | 23 |
| 3.2.2 - Concentration model and suppression..... | 24 |
| 3.2.3 - Cloning and Mutation..... | 24 |
| 3.2.4 - Fitness-based selection and insertion..... | 25 |
| 3.3 - THE cobICA ALGORITHM..... | 25 |
| 3.3.1 - General aspects..... | 25 |
| 3.3.2 - Criterion and Representation..... | 26 |
| 3.3.3 - The mutation operator..... | 28 |
| 3.3.4 - The local search process..... | 29 |
| 3.4 – CONCLUDING REMARKS..... | 30 |
| 4 - SIMULATION RESULTS..... | 32 |
| 4.1 - SIMULATIONS of cobICA..... | 33 |
| 4.2 - COMPARATIVE SIMULATIONS FOR VARIABLE NUMBER OF SAMPLES..... | 37 |

| | |
|--|-----------|
| 4.3 - COMPARATIVE SIMULATIONS FOR A VARIABLE NUMBER OF SOURCES..... | 41 |
| 4.4 - COMPUTATIONAL COMPLEXITY COMPARISON..... | 43 |
| 5 – CONCLUDING REMARKS..... | 46 |
| 5.1 - FUTURE PERSPECTIVES..... | 47 |
| REFERENCES..... | 48 |

LIST OF FIGURES

| | |
|--|----|
| Figure 2.1 - The BSS problem in Secondary Surveillance Radar (SSR). | 5 |
| Figure 2.2 - The BSS approach in the cocktail-party problem. | 6 |
| Figure 2.3 - The basic block diagram of BSS problem. | 7 |
| Figure 3.1 - The cobICA algorithm key components. | 26 |
| Figure 4.1.1 - Performance of cobICA for different values of sources (N), $P = 2, 3$ and 5 | 34 |
| Figure 4.1.2 - Performance of cobICA for different field orders, $N = 4$ | 36 |
| Figure 4.1.3 - Performance of cobICA when the non-uniform character of distributions is changed, (a) $N = 4, P = 3$, (b) $N = 6, P = 5$ | 37 |
| Figure 4.2.1 - Comparison among the techniques for field orders $P = 2$, (a) $N = 4$, (b) $N = 5$, (c) $N = 6$ | 38 |
| Figure 4.2.2 - Comparison among the technique for field order $P = 3$, (a) $N = 4$, (b) $N = 5$, (c) $N = 6$ | 39 |
| Figure 4.2.3 - Comparison among the techniques for field order $P = 5$, (a) $N = 4$, (b) $N = 5$ and (c) $N = 6$ | 40 |
| Figure 4.3.1 - Comparison among the technique for different values of sources when field order $P = 3$, (a) 32 samples, (b) 64 samples and (c) 512 samples. | 42 |
| Figure 4.3.2 - Comparison among the techniques for different values of sources when field order $P = 5$, (a) 32 samples, (b) 64 samples. | 43 |
| Figure 4.4.1 - Number of entropy evaluation for different values of sources when field order $P = 3$ | 44 |
| Figure 4.4.2 - Number of entropy evaluation for different numbers of samples when field order $P = 3$ and $N = 8$ | 45 |

LIST OF TABLES

| | |
|---|----|
| Table 4.1 - Simulation parameter's of cobICA algorithm..... | 33 |
|---|----|

LIST OF SYMBOLS

| | |
|-----|---|
| BSS | Blind Source Separation |
| ICA | Independent Component Analysis |
| iid | independent and identically distributed |
| ITL | Information Theoretic Learning |
| GF | Galois Fields |
| PMF | Probability Mass Function |
| MI | Mutual Information |
| RV | Random Variable |

CHAPTER 1

1 – Introduction

During the last decades several techniques based on Machine Learning, Computational Intelligence and Adaptive Signal Processing have been developed to deal with the increasing availability and complexity of data (HAYKIN, 1994; ADALI; HAYKIN, 2010; ROMANO *et al.*, 2011), for which conventional concepts such as linearity, continuity, gaussianity and second-order statistical information may not be sufficient.

In this context, Information Theoretic Learning (ITL) (PRINCIPE, 2010) has shown to be an important approach to solve engineering problems which deal with information content measuring of events and sources. In order to train the parameters of a learning machine, ITL seems an intuitive way because the main aspect of learning is to transfer the information contained in the available data onto the adaptive system that is intended to model this data and the underlying problem. The learning capability of adaptive responsive systems is essential in the implementation of intelligent algorithms (PRINCIPE *et al.*, 2000).

Independent Component Analysis (ICA) can be considered an example of ITL application, with a general framework applicable to several modern challenges in signal processing, such as in speech recognition systems, telecommunications and medical signal processing (CRISTESCU *et al.*, 2000; JUNG *et al.*, 2001; VALKAMA *et al.*, 2001; ZHANG; KASSAM, 2001).

When linear models fit well, ICA has been shown to be one of the most important approaches in separating independent source signals that were linearly mixed, with no prior information on the sources contents or on the mixtures parameters. This is the familiarly-known Blind Source Separation (BSS) problem (COMON; JUTTEN, 2010). Hence, the goal of an ICA solution for BSS is to recover independent sources, given only sensor observations that are linear mixtures of the unobserved independent source signals.

While the BSS/ICA problem is a well-known approach for real- or complex-valued signals, recently Arie Yeredor (YEREDOR, 2007) gave the first steps to extend the BSS/ICA problem for Galois fields (GF), as well. Hence, the problem can be formulated as a

combinatorial optimization problem with a cost function that estimates the independence degree between the extracted components $\mathbf{y}(n)$, which defined in Equation 2.3.

This Master's thesis proposes to study the problem of blind source separation and its associated solution via independent component analysis, when the signals and models are defined over Galois fields. Moreover, a study and comparative analysis is performed for two heuristic methods denoted as AMERICA and MEXICO (YEREDOR, 2011a; GUTCH *et al.*, 2012), along with cobICA (SILVA *et al.*, 2014), a Bioinspired framework based on cob-aiNet[C] immune-inspired algorithm (COELHO; VON ZUBEN, 2010).

The two pioneer algorithms AMERICA and MEXICO adopt a criterion based on the lowest entropy linear combination of the mixtures, while cobICA algorithm (SILVA *et al.*, 2014) employs a different search strategy, based on cob-aiNet[C] mechanism, a different criterion, the minimal mutual information (MMI), and specific full-rank-preserving operators.

According to the simulation results, the cobICA technique can achieve good performances for small number of sources. Additionally, the comparison between the computational complexities of the methods, when the number of sources increases, indicates that the cobICA technique presents a lower computational complexity than AMERICA, which highlights the advantage of the cobICA technique over other methods in high-dimension scenarios.

This dissertation is divided into 5 chapters, including this introduction. Chapter 2 presents the BSS/ICA framework in its linear-instantaneous model, first in the canonical formulation, for real- or complex-valued signals, and then in the context of signals defined over Galois fields. In the sequence, some fundamental concepts of information theory are presented, such as Shannon's entropy and mutual information for discrete random variables. Those definitions are essential to formulate ICA over GF algorithms, including AMERICA and MEXICO, which are the final subject of this chapter.

In Chapter 3, basic concepts of immune-inspired algorithms and, specifically the key aspects of cob-aiNet[C] algorithm are discussed, then, a descriptive analysis of cobICA algorithm is presented: an immune-inspired implementation of ICA over GF, supported by a mutual information-based criterion.

In order to compare the behavior of the three techniques altogether and to analyze the potential benefits of each one in different scenarios, an extensive set of numerical simulations are shown and discussed in Chapter 4 and, finally, conclusions are drawn in Chapter 5.

CHAPTER 2

2 - Independent Component Analysis over Galois fields

In order to perform blind source separation, different techniques have been discovered based on statistical features of the source signals and on structural aspects of the mixing process. Independent component analysis is one of the most important solutions, in the case of separating sources that are mutually independent.

The idea of performing ICA specifically over Galois or finite fields was first proposed by Arie Yeredor in 2007, in the context of boolean “Exclusive Or” (XOR) mixtures (YEREDOR, 2007), then, the idea was generalized by GUTCH *et al.*, (2010) in order to encompass arbitrary discrete and finite sets of numbers, yielding the ICA framework over $GF(q)$.

In this chapter the BSS problem, the ICA model, Galois fields and information-theoretic concepts are discussed. Then, the pioneering algorithms AMERICA and MEXICO are presented.

2.1 - Blind Source Separation and Independent Component Analysis

Blind Source Separation, which deals with recovering an unknown set of sources from an observable set of mixed signals (COMON; JUTTEN, 2010), has been applied over a wide range of engineering fields such as: array signal processing and wireless communication (AMAR; CICHOCKI, 1998; MANSOUR *et al.*, 2000), geophysical exploration (MAKEIG *et al.*, 1996), biomedical signal processing (VIGARIO *et al.*, 2000; IRIARTE *et al.*, 2003), speech processing (LEE, 1998) and image processing (LEE; LEWICKI, 2000).

As an example, Figure 2.1 shows the BSS model applied in Secondary Surveillance Radar (SSR) (PETROCHILOS, 2002; ICAO, 2004), which is a radar system employed in air traffic control centers. The SSR system uses an interrogation signal to detect and measure the position of aircrafts while it also requests additional information from the aircraft, such as its identity and altitude, and the aircrafts reply to each request by transmitting a response containing encoded data.

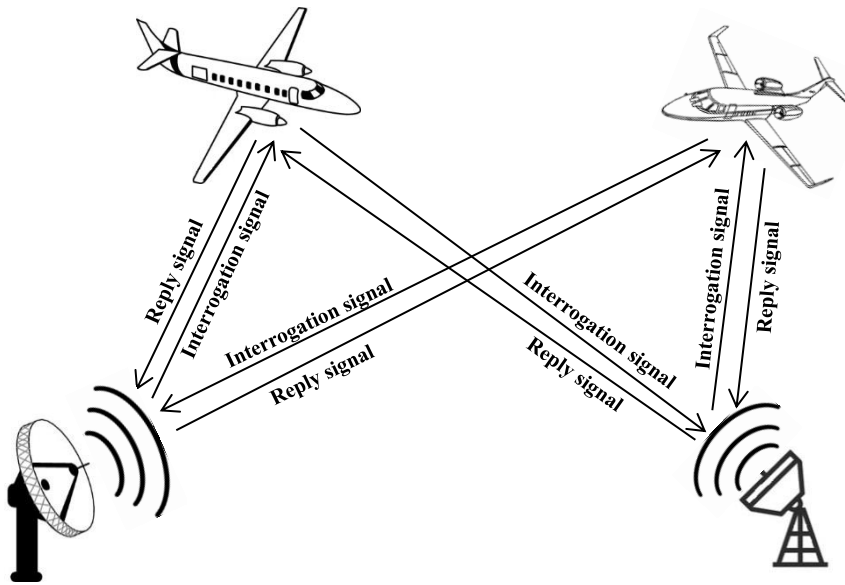


Figure 2.1 - The BSS problem in Secondary Surveillance Radar (SSR), adapted from (ICAO, 2004).

Another example for the BSS approach is in the cocktail-party problem (HYVARINEN *et al.*, 2001), where the combination of voices from different people, in the same room, are captured by a set of microphones, and we need to estimate the original speech signals from this set of mixtures.

As an example, consider the case where two individuals are having a conversation in a room with two microphones, which results in two recorded signals. To illustrate, consider the waveforms in Figure 2.2, where these recorded signals are mixed in two mixture signals and, hence, the goal is to recover the two original speech signals (HYVARINEN *et al.*, 2001; COMON; JUTTEN, 2010).

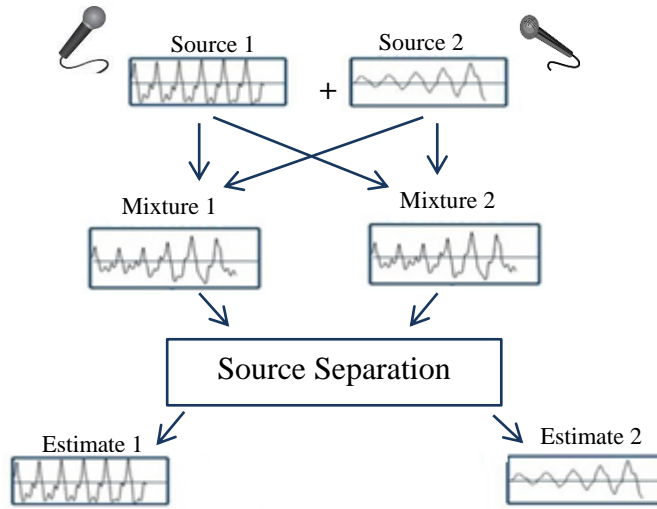


Figure 2.2 - The BSS approach for the cocktail-party problem.

In mathematical terms, the linear-instantaneous model of BSS for real- and complex-valued signals is defined as follows:

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n), \quad (2.1)$$

where $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_N(n)]^T$ is a vector of N observed random signals at the instant n , obtained from the mixing of $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$, a vector of the source components at the same instant and \mathbf{A} is an invertible ($N \times N$) matrix, since we are considering the determined model, i.e. the number of sources is equal to the number of mixtures.

The basic BSS block diagram is shown in Figure 2.3, observe that the source data $\mathbf{s}(n)$ and the mixing matrix \mathbf{A} are both unknown.

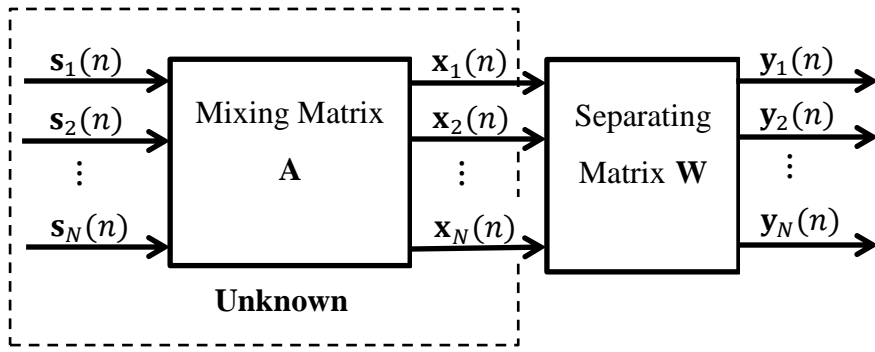


Figure 2.3 - The basic block diagram of BSS problem.

As shown in (COMON, 1994), if the following assumptions are valid, the BSS problem for real and complex signals can be solved via ICA:

1. The sources are statistically mutually independent, where the independence condition is defined as follow (KAGAN *et al.*, 1973):

$$P(\mathbf{s}(n)) = \prod_{i=1}^N P[s_i(n)]. \quad (2.2)$$

This means that the joint probability density of the source vector – $P(\mathbf{s}(n))$ – is equal to the product of the marginal probability densities – $P[s_i(n)]$ – of the individual signals.

2. At most one of the sources has Gaussian distribution.
3. The mixing matrix \mathbf{A} is invertible.

In this context, the ICA strategy consists of seeking for a separating matrix that produces an output vector whose components are maximally independent. It is possible to demonstrate (HYVARINEN *et al.*, 2001; COMON; JUTTEN, 2010) that such technique will result in the recovery of the original sources, more specifically, the mathematical form involves to recover the original sources by estimating the unmixing (or separating) matrix \mathbf{W} , such that

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) = \mathbf{P}\mathbf{A}^{-1}\mathbf{x}(n) = \mathbf{P}\mathbf{D}\mathbf{s}(n), \quad (2.3)$$

where $\mathbf{y}(n)$ recovers the independence condition among its components up to scale and permutation ambiguities, which are respectively represented by a diagonal matrix \mathbf{D} and a permutation matrix \mathbf{P} .

Hence, when the mentioned assumptions are valid for the BSS problem, its solution is possible via ICA, which basically attempts to find an inverse mapping that maximizes independence between the output components $\mathbf{y}(n)$, as described by Equation 2.3.

There are several criteria to measure independence between the extracted components and, consequently, to estimate \mathbf{W} , such as Higher Order Statistics (HOS) like kurtosis, cumulants and others that attempt to explore the connection between non-Gaussianity and independence, e.g. negentropy (CARDOSO, 1992; NIKIAS; MENDEL, 1993; HYVARINEN *et al.*, 2001).

Notwithstanding, if there are Gaussian-distributed sources, but with certain temporal structure (temporally dependent), then criteria that are based on Second Order Statistics (SOS) (CARDOSO, 1989) between time samples are capable of obtaining the BSS solution, instead of ICA.

In the context of real- or complex-valued signals, ICA has been successfully applied as a tool to solve BSS problems in many different fields. As an example, in wireless communication, ICA is employed to extract the original signal which was transmitted through an unknown channel (CASTEDO *et al.*, 1997; FENG; KAMMAYAR, 1999; CRISTESCU *et al.*, 2000; VALKAMA *et al.*, 2001; ZHANG; KASSAM, 2001).

Furthermore, one can find ICA applications in biomedical signal processing and analysis of seizures (artifacts removing) (VIGARIO *et al.*, 2000; JUNG *et al.*, 2001); in image processing and feature extraction (BELL; SEJNOWSKI, 1997; HYVARINEN; HOYER 2000); and in audio processing, where applications are found in multiple speakers separation (IKEDA; MURATA, 1999; TORKKOLA, 1999).

Based on the concepts of BSS/ICA for real- or complex-valued signals that this work discussed so far, it is possible to move towards the extension to the case of finite/Galois fields. But, before specifically discussing the BSS/ICA problem, we need to analyze some fundamental theoretical aspects of Galois fields, which are presented in the following section.

2.2 - BSS/ICA over Galois fields

In simple words, a field is the mathematical formalization of sum and product operations for arbitrary sets. The real numbers and the complex numbers are two familiar examples of fields.

A field F is a set with two operations \cdot and $+$, such that the following axioms hold (LIDL; NIEDERREITER, 1997):

1. Closure: For every $a, b \in F \rightarrow a \cdot b, a + b \in F$.
2. Commutativity: For every $a, b \in F \rightarrow a + b = b + a$ and $a \cdot b = b \cdot a$.
3. Associativity: For every $a, b, c \in F \rightarrow a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
4. Distributivity: For every $a, b, c \in F \rightarrow a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
5. Identity elements: There are elements $0 \in F$ and $1 \in F, 0 \neq 1$, such that $a + 0 = a$ and $a \cdot 1 = a, \forall a \in F$.
6. Inverse elements: For every $a \in F, \exists b \in F$ such that $a + b = 0$, and if $a \neq 0; \exists c \in F$ such that $a \cdot c = 1$.

If the number of elements of a field is finite, it is called a Finite or Galois field of size q , denoted $GF(q)$, where q has to be some prime power, i.e. $q = P^n$, where P is a prime and n 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 the sum and product in modular arithmetic (LIDL; NIEDERREITER, 1997), i.e. sum modulo P and product modulo P . In the case of non-prime fields, i.e. $n > 1$, we need to define the symbolic operations on polynomials, which is naturally more tricky to be implemented than the prime field case (GUTCH *et al.*, 2012).

It is also possible to construct vector spaces over finite fields, with some important considerations (GUTCH *et al.*, 2012; SILVA, 2013):

1. Considering a vector space V with dimension N in a finite field $F = GF(q)$, it can be written as $V = F^N$, such that a member of V is a N -dimension column vector with components that belong to F .

2. A linear mapping $A: F^N \rightarrow F^M$ can be represented by a matrix $(M \times N)$ with values in F , and the composition of matrices is denoted as "matrix product", by noticing that element-wise operations are performed in F .
3. The determinant of a square matrix \mathbf{A} is calculated as in the "traditional" manner, for example, by Cramer's rule or by the Laplace formula, with the operations between the elements naturally defined in F ; moreover, \mathbf{A} is invertible if its determinant is different than the field's null element.
4. The set of $(N \times N)$ invertible matrices with elements defined in $F = GF(q)$ is denoted by $GL(N, q)$.
5. Unlike the reals, the elements of a finite field cannot be ordered. Therefore, the product of two vectors in V , defined by

$$\langle \mathbf{a}, \mathbf{b} \rangle := \sum_{i=1}^N a_i b_i \quad (2.4)$$

is not positive definite and therefore cannot be a scalar product.

6. There is a nonzero vector $\mathbf{c} \in V$ such that $\langle \mathbf{c}, \mathbf{c} \rangle = 0$ – for example, let $F = GF(2)$ and $\mathbf{c} = (1, 1)^T$ – which means that there is no concept of orthogonality for vector spaces in finite fields.

With the basic concepts of Galois fields, we are ready to analyze the BSS/ICA problem in this context, which is mathematically formulated in the same manner as Equations 2.1 and 2.3 states, but with the crucial difference that the values of all elements ($\mathbf{x}(n)$, $\mathbf{s}(n)$, $\mathbf{y}(n)$, \mathbf{A} , \mathbf{W} , \mathbf{P} , \mathbf{D}) and the \cdot , $+$ operations are defined over $GF(q)$.

The ICA solution provided by Equation 2.3, specifically, remains valid for finite fields, but, instead of the pre-requisite that restricts the presence of Gaussian distributions, the following theorem shows that non-uniformity of the sources is fundamental to perform ICA over $GF(q)$ (YEREDOR, 2011a):

Theorem 2.1 Assume that \mathbf{s} be an N -dimensional independent random vector over a finite field $F = GF(q)$ with joint probability distribution $p_s(\cdot)$, such that the marginal distributions are not uniform and also don't have null probability values.

If any $\mathbf{G} \in GL(N, q)$ could be found such that $\mathbf{y} = \mathbf{G}\mathbf{s}$ be composed of independent components again, then $\mathbf{G} = \mathbf{P}\mathbf{D}$, for some permutation matrix \mathbf{P} and some diagonal matrix \mathbf{D} (GUTCH *et al.*, 2010; SILVA *et al.*, 2014a).

As Theorem 2.1 shows – its proof can be seen in (GUTCH *et al.*, 2010) – there is a restriction to sources whose distribution is uniform. This behavior refers to the statement that a combination of two independent signals, with one uniformly-distributed, results in a new signal that maintains the independence condition (YEREDOR, 2011a), which makes ICA useless, in this case.

Note that the definition of a non-singular, hence invertible, separating matrix in terms of a non-null determinant also remains valid in this new scenario. Furthermore, as shown in (WATERHOUSE, 1987), the subset of invertible matrices $GL(N, q)$ is, naturally, finite and has a number of elements given by

$$|GL(N, q)| = \prod_{k=0}^{N-1} (q^N - q^k), \quad (2.5)$$

while the cardinality of the set of $(N \times N)$ square matrices is q^{N^2} . Therefore, 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.

Hence, we conclude that ICA over $GF(q)$ can be formulated as a combinatorial optimization problem with a cost function that measures the dependence between the extracted components $\mathbf{y}(n)$, such that the solution is obtained by the signals that minimize this criterion (or maximize an independence measure, equivalently). It is fundamental, then, to define an appropriate form of measuring the degree of dependence between random signals, which is possible by considering some key concepts derived from information theory, to be presented in the following section.

2.3 - Information-theoretic measures for discrete sources

In order to find a solution in signal processing problems based on measuring the degree of dependence between random signals, Information Theory (SHANNON, 1948) lies as an important framework to deal with information sources of both nature, discrete and continuous.

Given that ICA over $GF(q)$ involves discrete entities, we consider for this specific case the domain of discrete random variables (RV), e.g. X , for which the entropy $H(\cdot)$ – the average degree of uncertainty of a RV – is defined as (SHANNON, 1948):

$$H(X) = - \sum_x p_X(x) \log p_X(x), \quad (2.6)$$

where $p_X(x)$ is the probability mass function (PMF) associated with X . Recall that the definition considers that $0 \log 0 = 0$.

Furthermore, its possible to define the joint entropy of two random variables X and Y , as follows:

$$H(X; Y) = - \sum_x \sum_y p_{XY}(x, y) \log p_{XY}(x, y), \quad (2.7)$$

where $p_{XY}(x, y)$ is the joint PMF concerning X and Y .

Based on the definition in Equation 2.7, we can write

$$H(X; Y) \leq H(X) + H(Y), \quad (2.8)$$

where equality is only possible if X and Y are statistically independent, i.e. when

$$p_{XY}(x, y) = p_X(x) p_Y(y). \quad (2.9)$$

The Mutual Information (MI) measures how much (on average) the realization of random variable Y tells us about the realization of X , (COVER; THOMAS, 2006). Furthermore, mutual information is symmetric, which means that X tells us exactly as much about Y as Y tells us about X (CHAOLI; SHEN, 2011).

The Mutual Information (MI) between X and Y is defined as

$$I(X; Y) = H(X) - H(X|Y), \quad (2.10)$$

and

$$I(X; Y) = H(X) + H(Y) - H(X; Y), \quad (2.11)$$

where the conditional entropy $H(X|Y)$ tells how much information about X is still unknown after observing Y . It is defined as

$$\begin{aligned} H(X|Y) &= \sum_y p(y) H(X|Y=y) \\ &= - \sum_y p_Y(y) \sum_x p_{X|Y=y}(y) \log p_{X|Y=y}(y), \end{aligned} \quad (2.12)$$

where $H(X|Y=y)$ is the entropy of the variable X conditional on the variable Y taking a certain value y .

Another important property for entropy is the chain rule (COVER; THOMAS, 2006), which states that the entropy of a collection of random variables is the sum of the conditional entropies, as follows:

$$H(X_1, X_2, \dots, X_N) = \sum_{i=1}^N H(X_i | X_{i-1}, \dots, X_1), \quad (2.13)$$

where (X_1, X_2, \dots, X_N) are drawn according to $p_{X_1}(\cdot), p_{X_2}(\cdot), \dots, p_{X_N}(\cdot)$, respectively.

Note that the MI of a collection of random variables can also be defined (COVER; THOMAS, 2006):

$$I(X_1; X_2; \dots; X_N) = \sum_{i=1}^N H(X_i) - H(X_1, X_2, \dots, X_N). \quad (2.14)$$

Also based on Equation 2.10 we can write MI in terms of probabilities

$$I(X; Y) = \sum_x \sum_y p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x) p_Y(y)}, \quad (2.15)$$

and, thus, relate it to the relative entropy or Kullback-Leibler divergence (COVER; THOMAS, 2006), as follows:

$$D_{KL}(p; q) = \sum_u p(u) \log \frac{p(u)}{q(u)}, \quad (2.16)$$

where p and q are two PMFs with equivalent support and consider that $0 \log \frac{0}{0} = 0$.

From Equation 2.15 and 2.16 we can show that mutual information can be formulated as the relative entropy between the joint distribution and the distribution given by the product of marginal distributions (COVER; THOMAS, 2006):

$$I(X; Y) = D_{KL}(p_{XY}(x, y); p_X(x) p_Y(y)). \quad (2.17)$$

2.4 – Canonical algorithms for ICA over GF

After presenting the problem definition, the fundamental concepts regarding finite fields and information-theoretic measures, we can discuss two pioneering propositions of algorithms to perform ICA over $GF(q)$. As the reader will see, in the following subsections, both employ specific search heuristics that optimize the entropy value of the outputs of the separating system, in order to obtain estimates of the original, independent components.

2.4.1 - The AMERICA algorithm

AMERICA was introduced by Arie Yeredor in 2007, in the context of boolean “Exclusive Or” (XOR) mixtures (YEREDOR, 2007), and subsequently was generalized by H.W. Gutch *et al.*, in 2010 in order to encompass fields with arbitrary order, yielding the ICA framework over $GF(q)$. Yeredor’s work contains the proof of a separation theorem for signals in $GF(2)$ and also presents an important property:

Lemma 2.1 Let U and V be two independent RVs defined over $GF(2)$, where $W = U + V$, then $H(W) \geq H(V)$ and $H(W) \geq H(U)$.

Recall that $H(\cdot)$ is the entropy of a random variable, represented in Equation 2.7. Lemma 2.1 can be generalized to any linear combination of discrete, independent RVs (COVER; THOMAS, 2006; YEREDOR, 2011a) and in simple terms, ensures that a linear mixing process of independent random signals never causes a reduction of entropy.

By exploring the aforementioned property, in order to solve the problem of BSS/ICA over Galois fields, the AMERICA technique (Ascending Minimization of EntRopies) tries to recover the lower entropy configuration previous to the mixing in order to extract the signals sources (COMON; JUTTEN, 2010), in which this extraction should be performed N times, once for each signal source (GUTCH *et al.*, 2012).

Based on this, the lowest entropy linear combination of the mixtures is determined by AMERICA algorithm search process, where each set of coefficients that extracts a source is linearly independent from the previously chosen (otherwise the same source would be repeatedly extracted), such that:

$$\mathbf{w}_{\text{opt}} \leftarrow \arg \min_{\mathbf{w} \in \mathbb{F}^N} H(y = \mathbf{w}^T \mathbf{x}), \quad (2.18)$$

where y is an estimate of one of the independent sources and \mathbf{w}_{opt} is the vector that indicates the optimal linear combination of the mixtures.

AMERICA first proposal (YEREDOR, 2007) employed, after each source extraction, a deflationary procedure (DELFOSSÉ; LOUBATON, 1995) to remove it from the mixtures, running the search process again to extract the remaining sources. The definitive version (YEREDOR, 2011a; GUTCH *et al.*, 2012), however, adopts the aforementioned strategy of obtaining each extraction vector as being linearly independent from the previous ones, which results in the lines of the separating matrix being the corresponding extraction vectors of each source.

Note that there are, in the worst case, $q^N - 1$ non-trivial linear combinations which should have the corresponding entropy evaluated, for each source signal. Therefore, the computational cost, in terms of cost function evaluations, of the search increases exponentially with the number of components, in other words the computational complexity of AMERICA is approximately $O(Nq^N)$ (YEREDOR, 2011a).

The Pseudo-code 2.1 presents the steps of the AMERICA algorithm, where the technique corresponds to an exhaustive search to determine the linear combination of the mixtures, which retrieves each of the sources through the condition minimum entropy, until it could find all N sources. Consider that a set of T samples of each mixture is available, resulting in a $N \times T$ mixed sample matrix \mathbf{X} .

Pseudo-code 2.1 Algorithm AMERICA

Input : $(N \times T)$ mixed sample matrix \mathbf{X}

Output: separating matrix \mathbf{W}

$\mathbf{W} \leftarrow \emptyset$;

do

$\mathbf{w}_{\text{opt}} \leftarrow \arg \min_{\mathbf{w} \in \mathbb{F}^N} H(\mathbf{w}^T \mathbf{X})$

s. t. \mathbf{w} is L.I.

$\mathbf{W} \leftarrow \mathbf{W} \cup \{\mathbf{w}_{\text{opt}}\}$;

while $\text{Size}(\mathbf{W}) = N$;

$\mathbf{W} \leftarrow$ matrix built from the row vectors in \mathbf{W} ;

2.4.2 - The MEXICO algorithm

In order to achieve the lowest entropy linear combination of the mixtures in ICA model, another useful technique is the MEXICO algorithm (Minimizing Entropies by eXchanging In Couples), which was first proposed by Gutch *et al.*, (2010), with the name “Entropy Based Demixing”, then, in Gutch *et al.*, (2012), the authors renamed it as MEXICO.

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

The main difference is that MEXICO makes an equivalent entropy evaluation only between combined pairs of mixtures, while AMERICA evaluates the entropy of an extracted signal which is generated by combining N mixtures.

In other words, if we have two mixtures observations at a given instant, i.e. $x_i(n) = x_i$ and $x_j(n) = x_j$ and a constant $c \in \mathbb{F} = GF(q)$, such that

$$H(x_i + c x_j) < H(x_i), \quad (2.19)$$

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

$$\mathbf{T}_{i,j}(c) = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & c & & \ddots & \\ & & & & 1 \end{bmatrix}, \quad (2.20)$$

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

More specifically, the Pseudo-code 2.2 presents the steps of the MEXICO algorithm. Consider again that x_i and x_j are two components of the mixtures vector \mathbf{x} ; the process starts by determining $c \in GF(q)$ in $\bar{x}_i = x_i + cx_j$ such that $H(\bar{x}_i)$ has the minimum value, if $H(x_i)$ is greater than the entropy of the combination, i.e. $H(x_i) > H(\bar{x}_i)$, then x_i is replaced by \bar{x}_i .

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

Pseudo-code 2.2 Algorithm MEXICO

Input : $(N \times T)$ mixed sample matrix \mathbf{X}

Output: separation matrix \mathbf{W}

$\mathbf{W} \leftarrow \mathbf{I}_N \in GL(N, q);$

while $\exists(i, j, c) \mid H(x_i + cx_j) < H(x_i)$ **do**

$\mathbf{W}_0 \leftarrow \mathbf{I}_N \in GL(N, q);$

$(\mathbf{W}_0)_{ij} \leftarrow c;$

$x_i \leftarrow x_i + cx_j$

$\mathbf{W} \leftarrow \mathbf{W}_0 \mathbf{W}$

end while

The computational complexity of MEXICO algorithm, which is an iterative algorithm, involves counting the number of sweeps and the number of cost function evaluations in each sweep (YEREDOR, 2011a). Since the stopping condition is not deterministic, there is no fixed number of entropy evaluations as AMERICA.

Whatsoever, MEXICO have shown to be a faster method than AMERICA, in terms of time processing, as the simulations in GUTCH *et al.*, (2012) indicates. On the other hand, AMERICA presented a level of quality superior than MEXICO, which couldn't achieve similar marks.

2.5 – Concluding Remarks

In this chapter we discussed about the BSS/ICA problem for real- or complex-valued signals and its extension to the domain of finite or Galois fields. In this scenario, BSS/ICA is formulated as a combinatorial optimization problem whose cost function to be minimized measures the dependence degree between the extracted components $\mathbf{y}(n)$, which are obtained as defined in Equation 2.3.

In the sequence, a review on AMERICA and MEXICO algorithms, which can be considered state-of-the-art heuristics for ICA over GF, was developed. In the next chapter, we discuss the possibility of joining Bioinspired algorithms with Information-theoretic criteria in order to obtain a different approach for this same problem.

CHAPTER 3

3 - Bioinspired Search and Information Theoretic Learning to perform ICA over Galois fields: the cobICA algorithm

Recently, D. G. Silva, *et al.*, (2014) proposed a novel Bioinspired algorithm for ICA over finite fields, which is based on the combinatorial version of the Concentration-based Artificial Immune Network - cob-aiNet[C] (COELHO *et al.*, 2011). This mechanism is used as the optimization algorithm for the framework, which is called cobICA.

In this chapter, we present a brief overview on objectives of Bioinspired algorithms, specifically Artificial Immune Systems, then, the mechanism of cob-aiNet[C] is introduced and, finally, the cobICA algorithm (SILVA *et al.*, 2014) is detailed.

3.1 – Artificial Immune Systems

Bioinspired algorithms are engineering / computational tools that rely on biological models to solve diverse tasks, such as optimization (DE CASTRO, 2006). In this case, the goal is to find the best parameter settings for maximizing (or minimizing) a given function (DE CASTRO; TIMMIS, 2002a). Bioinspired methods are general purpose tools that have shown efficient solutions to optimization problems that are hard to be tackled by classical methods, e.g high-dimension combinatorial optimization problems or optimization of real-valued functions that are non-continuous or non-differentiable.

An important class of Bioinspired algorithms are Artificial Immune Systems (AIS) – computational systems inspired by functions, principles and models of the vertebrates' immune system (DE CASTRO; TIMMIS, 2002a). In comparison with other strategies such as genetic algorithms, immune-inspired algorithms are preferable because they present higher diversity maintenance capabilities (DE FRANCA *et al.*, 2010), which means that they possess an increased possibility of obtaining multiple optima points, at the end of the search process.

The natural immune system provides defense mechanisms against infectious agents and, in order to do so, it has many useful properties from a computational point of view, such as self-organization, recognition, adaptability and robustness. Those properties begin to

emerge when the system extracts features of the invader and adapts its internal structure to the environment, without any external supervision (JERNE, 1974).

Imagine that an attack occurs by an unknown antigen (molecular signatures of the foreign agent), then, in order to block the damaging action of the invader, two sub-systems of the immune system act: the innate sub-system tries to act quickly by blocking the invader via “conventional” defense mechanisms, while the adaptive system recognizes the specific agent and produces B-cells with “customized” molecules (antibodies) that neutralize the agent. The recognition capacity of a given antigen by a B-cell is called affinity. Thus, the adaptive immune response is continuously improved to recognize antigens (DE CASTRO, 2001).

When those ideas are used as inspiration for Artificial Immune Systems, in the optimization context, a population of candidate solutions represent the B-cells and their respective affinities with the antigen are represented by the objective function to be optimized. Furthermore, in cob-aiNet[C] algorithm, the Clonal Selection theory and the Immune Network theory are the two fundamental models that inspire the technique.

The Clonal Selection Principle is one of the key theories that support a series of immune-inspired algorithms, there are several clonal selection-based algorithms discussed in the literature (DE CASTRO; VON ZUBEN, 2002). They seek to improve candidate solutions for a given problem by means of the tasks of cloning, mutation and selection.

The theory of clonal selection was first proposed by Frank Macfarlane Burnet in 1959 (BURNET, 1959), it explains how the immune system is able to recognize and produce antibodies only against the antigens to which the body was exposed.

This means that, when antigens of a foreign pathogen invading the organism are identified by the immune cells, these suffer a cloning process which is followed by a controlled mutation and by a selective pressure, in order to improve the recognition capability of the cells and, consequently, the organism can properly respond against the infection.

Clonal selection algorithms have been applied to optimization problems e.g. CLONALG (DE CASTRO; VON ZUBEN, 2002), opt-IA (NICOSIA, 2004), the B-Cell algorithm (KELSEY; TIMMIS, 2003) and to multi-objective optimization (CRUZ CORTES; COELLO COELLO, 2003).

The other fundamental theory, the Immune Network Theory (JERNE, 1974), discusses that the immune cells are also capable of recognizing each other, via a positive response or a negative response. A positive response is characterized by the activation of the recognizing cell and its cloning and mutation; a negative response is characterized by tolerance and possibly the suppression (elimination) of the recognized cell, since there may be another sufficiently similar cell with equivalent or better affinity with the antigen.

3.2 – The cob-aiNet[C] algorithm

The Concentration-based Artificial Immune Network (cob-aiNet) is an AIS (DE CASTRO, 2006) that implements the fundamental ideas of the Clonal Selection Principle (JERNE, 1974) and the Immune Network Theory (DE CASTRO, 2002) for optimization tasks. In order to solve, specifically, combinatorial optimization problems, the cob-aiNet[C] version (COELHO *et al.*, 2011) can be used.

In the following subsections, the main steps of cob-aiNet[C] are discussed. The full details of the technique are presented in its original paper, if the reader is interested (COELHO; VON ZUBEN, 2010; COELHO *et al.*, 2011).

The Pseudo-code 3.1 presents the main steps of the cob-aiNet[C] algorithm.

Pseudo-code 3.1 Algorithm cob-aiNet[C]

Parameters:

- nAB : initial number of cells;
- $maxAB$: maximum number of cells;
- nCl^{max} : maximum number of clones per cell;
- nCl^{min} : minimum number of clones per cell;
- C_0 : initial concentration;
- σ_s : suppression threshold;
- β^i : initial mutation parameter;
- β^f : final mutation parameter;
- LS_{it} : number of local search iterations;
- LS_{freq} : number of iterations between consecutive local search steps;
- $maxIT$: maximum number of iterations;

1. Randomly create the initial population of size nAB ;
 2. Evaluate the fitness of the cells in the initial population;
 3. Evaluate the affinity among cells in the initial population;
 - while** (iteration $\leq maxIT$) **do**
 4. Define the number of clones nCl_i that must be generated for cell i ;
 5. Generate nCl_i clones for each cell i in the population;
 6. Apply the mutation operator to each of the generated clones;
 7. Evaluate the fitness of the new cells;
 8. Select those cells that must be kept in the next generation (with insertion);
 - if** ((iteration **mod** LS_{freq}) == 0) **then**
 9. Apply local search to all cells and update their fitness;
 - end if**
 10. Evaluate the affinity among cells in the population;
 11. Update the concentration of all cells;
 12. Remove from the population those cells with null concentration;
 - end while**
 13. Apply local search to all cells and update their fitness;
 14. Evaluate the affinity among cells in the population;
 15. Update the concentration of all cells;
 16. Remove from the population those cells with null concentration;
-

3.2.1 - Representation and affinity metrics

In order to quantify the interactions between the elements of the system, affinity measures should be chosen. The concept of affinity maturation guarantees that the immune system becomes increasingly better at the task of optimization.

The cob-aiNet[C] mechanisms are based on two affinity metrics such that both are evaluated, for each cell \mathbf{u} in the population, at each iteration t . Considering a minimization task, the first measure is the affinity with antigens or fitness, called $f^{Ag}(\mathbf{u}_t)$, which is defined by:

$$f^{Ag}(\mathbf{u}_t) = 1 - \overline{f(\mathbf{u}_t)}, \quad (3.1)$$

where $f^{Ag}(\mathbf{u}_t) \in [0, 1]$ is the affinity with antigens of cell \mathbf{u} at iteration t and $\overline{f(\mathbf{u}_t)} \in [0, 1]$ is the normalized value of the cost function for cell \mathbf{u} at iteration t .

The second affinity measure is the affinity with antibodies or affinity with other cells, called $f^{Ab}(\mathbf{u}_t)$, defined by:

$$f^{Ab}(\mathbf{u}_t) = \begin{cases} \frac{\sum_{\mathbf{v}_t \in \mathbf{J}} C(\mathbf{v}_t) \cdot [\sigma_s - d(\mathbf{u}_t, \mathbf{v}_t)]}{\sum_{\mathbf{v}_t \in \mathbf{J}} C(\mathbf{v}_t)} & \text{if } \mathbf{J} \neq \emptyset \\ 0 & \text{otherwise} \end{cases}, \quad (3.2)$$

where $f^{Ab}(\mathbf{u}_t)$ is the total affinity between cell \mathbf{u}_t and all the other cells in the population, at iteration t , \mathbf{J} is the set of cells that are not worse than \mathbf{u}_t ($f^{Ag}(\mathbf{v}_t) \geq f^{Ag}(\mathbf{u}_t)$) and that are reasonably similar to it, i.e. they are within a radius σ_s (parameter of the algorithm) from \mathbf{u}_t , according to a distance (dissimilarity) metric $d(\mathbf{u}_t, \mathbf{v}_t)$, and $C(\mathbf{v}_t)$ is the concentration of cell \mathbf{v}_t at iteration t . In simple words, the affinity with antibodies value is higher, for a given cell, when there are many high-concentrated cells that are too similar and with better fitness than this particularly one.

3.2.2 - Concentration model and suppression

Inspired by the Immune Network Theory, the dynamic behavior of the algorithm is implemented by a concentration model that is directly ruled by its past values and by the affinity with other cells, such that at each iteration t , the concentration $C(\mathbf{u}_t) \in [0, 1]$ of a given cell \mathbf{u}_t in the population (initially defined with value C_0) is updated as follows:

$$C(\mathbf{u}_{t+1}) = \max \left[0, \min \left[\left(aC(\mathbf{u}_t) - f^{Ab}(\mathbf{u}_t) \right), 1 \right] \right], \quad (3.3)$$

where a is a regulatory factor expressed as:

$$a = \begin{cases} 1 + 0.1 \cdot f^{Ag}(\mathbf{u}_t) & \text{if } f^{Ab}(\mathbf{u}_t) = 0 \\ 0.7 & \text{otherwise} \end{cases}. \quad (3.4)$$

Note that non-null values of affinity ($f^{Ab}(\mathbf{u}_t)$) help decreasing the concentration, otherwise the concentration is increased in proportion with the cell's fitness ($f^{Ag}(\mathbf{u}_t)$). Besides, if a given cell obtains a null concentration, it is eliminated from the population. This situation is typical when the cell represents, for some iterations, a poor quality solution and/or is too similar to other cells.

3.2.3 - Cloning and Mutation

In the cloning step of cob-aiNet[C] algorithm, a given number of clones $nCl(\mathbf{u}_t)$ is created, at each iteration t , for each cell \mathbf{u}_t in the population such that:

$$nCl(\mathbf{u}_t) = C(\mathbf{u}_t) \cdot (nCl^{max} - nCl^{min}) + nCl^{min}, \quad (3.5)$$

where nCl^{max} and nCl^{min} are the limiting values defined by user.

In the sequence, the process of mutation is applied for the new cells, where each cell \mathbf{u} at iteration t has its parameters modified by a mutation operator $n^{mut}(\mathbf{u}_t)$ times, which is given by

$$n^{mut}(\mathbf{u}_t) = \max \left[\left\lfloor \beta(t) \cdot e^{-f^{Ag}(\mathbf{u}_t) \cdot C(\mathbf{u}_t)} \right\rfloor, 1 \right], \quad (3.6)$$

where $\beta(t)$ is the value of parameter β at iteration t , which indirectly influences the range of modifications made to the clones and is changed with respect to t according to a

mirrored-sigmoid function (COELHO; VON ZUBEN, 2010). Note that higher concentration levels imply a higher number of cell's clones and higher fitness values imply that each clone should be submitted to fewer rounds of mutations, or, in the opposite way, bad solutions are mutated in a more intense manner.

3.2.4 - Fitness-based selection and insertion

The selection operator in cob-aiNet[C] also presents the ability for the insertion of new cells in the population. First, a fitness-based selection mechanism is applied to each subset composed by the parent cell and its mutated clones. If the best clone cell is distant enough from its parent, it is kept together with the parent, increasing the population; otherwise, just the best one between the clone and its parent is kept for the next iteration. Therefore, during runtime, cob-aiNet[C] adjusts the population size dynamically by the insertion operator combined with the suppression of cells.

3.3 – The cobICA algorithm

ICA over Galois Fields poses, as seen in Chapter 2, a combinatorial optimization task where the objective function is some measure of (in)dependence between the extracted signals. Notwithstanding, cob-aiNet[C] is a state-of-the-art immune-inspired optimization method that can be adopted in the context of high-dimension domains.

In this section, we are able to join those concepts by studying the cobICA algorithm (SILVA *et al.*, 2014): an immune-inspired algorithm to perform ICA over Galois Fields, which applies 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} . In the following, the algorithm proposal is detailed.

3.3.1 - General aspects

In a big picture, cobICA overall organization is shown in Figure 3.1. The algorithm models the problem of ICA over GF such that each individual of the population (cell) is a candidate to be the separating matrix and the search space is limited by adopting appropriate mutation and local search operations.

As mentioned in Section 2.1, the BSS/ICA problem assumes that the mixing and the separating matrix are invertible. Therefore, the search space should be the set of invertible matrices over $GF(q)$ – i.e. $GL(N, q)$, for the N-dimension case – and, in this sense, the algorithm does not violate the full rank restriction by using customized mutation and local search operators, which consequently avoids the search for inappropriate solutions.

In order to control the diversity degree among the individuals of the population (as stated in cob-aiNet[C] algorithm specification, recall Section 3.2.1), cobICA chooses Hamming distance – the number of different symbols that exist between two sequences – as the dissimilarity metric.

The remaining fundamental parts of cobICA are discussed in the following subsections.

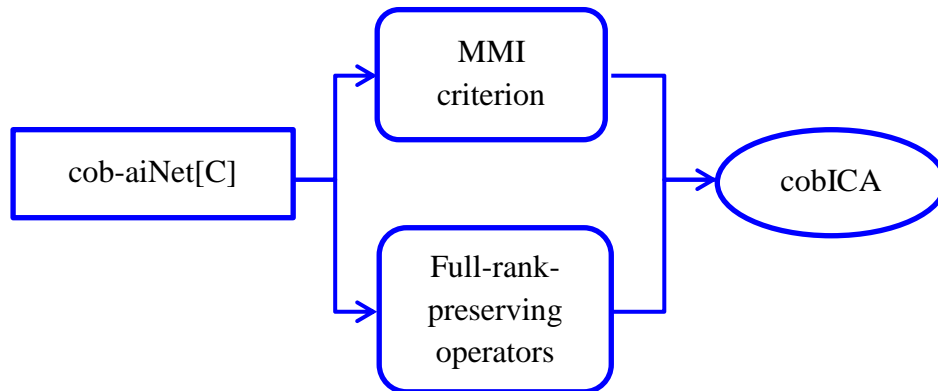


Figure 3.1 - The cobICA algorithm key components. Adapted from (SILVA *et al.*, 2014).

3.3.2 - Criterion and Representation

As mentioned in Section 2.2, in order to solve the BSS over GF problem, since the sources are mutually independent and non-uniform, if one finds a separating matrix that yields statically independent signals, it is possible to recover the original sources.

Based on this principle and the fact that independent signals yield a null value of mutual information, the minimal mutual information criterion is employed to guide the search process.

Recall that the ICA model considers the separating system output as $\mathbf{y} = \mathbf{W}\mathbf{x}$ (see Equation 2.3), where the instant indexes are omitted for simplicity purposes. Since \mathbf{W} is an invertible linear mapping over $GF(q)$, we have (GUTCH *et al.*, 2012):

$$p_Y(\mathbf{y}) = p_X(\mathbf{W}^{-1}\mathbf{y}). \quad (3.7)$$

This shows that bijective linear transformations over discrete spaces yield a joint PMF for random vector \mathbf{y} that is just a rearrangement of elements of $p_X(\mathbf{x})$.

As a consequence, if the mutual information definition for random vectors, previously defined in Equation 2.14, is applied to the output \mathbf{y} , the joint entropy term is not changed after the transformation, hence we can write:

$$I(\mathbf{y}) = \sum_{i=1}^N H(y_i) - H(\mathbf{x}), \quad (3.8)$$

where the term $H(\mathbf{x})$ is invariant. Thus, the search can be performed by minimizing only the first term of Equation 3.8, leading to the cost function of the algorithm, as follows:

$$J(\mathbf{W}) = \sum_{i=1}^N \hat{H}(y_i),$$

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \quad (3.9)$$

Since it is not known, in advance, the probabilities distributions of the mixtures, an entropy estimator is employed over a set of T independent and identically distributed (iid) observations of each mixed signal, $\{x_i(1), \dots, x_i(T)\}$, therefore $\hat{H}(y_i)$ is the maximum-likelihood estimator with the Miller-Madow bias correction (CARLTON, 1969):

$$\hat{H}(y_i) = \frac{q-1}{2T} - \sum_{j \in GF(q)} \hat{p}_{y_i}(j) \log_q \hat{p}_{y_i}(j), \quad (3.10)$$

$$\hat{p}_{y_i}(j) = \frac{1}{T} \sum_{n=1}^T 1_j(y_i(n)), \quad (3.11)$$

where $1_j(\cdot)$ is the indicator function, defined as:

$$1_a(x) = \begin{cases} 1, & \text{if } x = a, \\ 0, & \text{otherwise.} \end{cases} \quad (3.12)$$

According to the cost function defined in Equation 3.9, each individual cell of the population represent a candidate separating matrix and the whole population is evolved in order to find the ICA solution via the individual with the best fitness, when the algorithm finishes. In simple words, it tries to find the minimal value of the objective function defined in Equation 3.9.

3.3.3 - The mutation operator

The cobICA algorithm proposes a mutation operator as a routine that performs the linear combination of two rows, randomly chosen from the candidate matrix, and then replaces one of the original sequences.

This operation between rows can be represented as a left product by an elementary matrix and preserves the full rank property (CARLTON, 1969), consequently, there is no risk of a candidate matrix to become an unfeasible solution after a mutation. Assume that matrix $\mathbf{B} \in GL(N, q)$ represents the individual to be mutated, the corresponding elementary matrix that adds row \mathbf{B}_j multiplied by a scalar k to row \mathbf{B}_i is defined as the same in Equation 2.20, as follows

$$\mathbf{T}_{i,j}(k) = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & k & & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1 \end{bmatrix}, \quad (3.13)$$

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

Hence, the new cell (or individual) after the mutation is represented by

$$\mathbf{B}' = \mathbf{T}_{i,j}(k)\mathbf{B}. \quad (3.14)$$

The idea practiced in this technique resembles the process of MEXICO algorithm, which is discussed in Section 2.5, but there is a subtle difference. While MEXICO algorithm performs the row substitution exclusively based on the entropy reduction of a single extracted signal, cobICA mutation operator performs the modification independently if it may promote an improvement in the solution.

The Pseudo-code 3.2, adapted from (SILVA *et al.*, 2014) presents the main steps of the mutation process.

Pseudo-code 3.2 The main steps of the mutation process.

Input parameters:

B : individual / matrix to be mutated;

N : number of mixtures / matrix dimension;

q : order of the field;

1- Randomly choose $i \leq N, j \leq N, i \neq j$;

2- Randomly choose $k \in GF(q)$;

3- $\mathbf{B} \leftarrow T_{i,j}(k)\mathbf{B}$;

Return B

3.3.4 - The local search process

The authors of cob-aiNet[C] recommend the inclusion of a local search operator, in order to refine the optimal solution (COELHO *et al.*, 2011). In the context of cobICA, hence, the same operator stated in Equation 3.14 is used for the local search process, with the difference that, this time, the routine updates the individual *only if* the fitness is improved (the cost function value, defined in Equation 3.9, 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.

The Pseudo-code 3.3, adapted from (SILVA *et al.*, 2014) presents the main steps of the local search process, note also that the local search executes only one *sweep* process and that it does not lead to null determinant matrices, as well.

Pseudo-code 3.3 The main steps of the local search process.

Input parameters:

B : individual / matrix to be mutated;

N : number of mixtures / matrix dimension;

q : order of the field;

```
for all  $i \leq N$  do
  for all  $j \leq N, i \neq j$  do
    1- Randomly choose  $k \in \text{GF}(q)$ ;
    2- Calculate  $\mathbf{B}' \leftarrow T_{i,j}(k)\mathbf{B}$ ;
    If cost function ( $\mathbf{B}'$ ) < cost function ( $\mathbf{B}$ ) then
      3-  $\mathbf{B} \leftarrow \mathbf{B}'$ 
      return  $\mathbf{B}$ 
    end if
  end for
end for
return  $\mathbf{B}$ 
```

3.4 – Concluding Remarks

In this chapter we took a brief overview on an important class of Bioinspired algorithms, Artificial Immune Systems (AIS), which are computational systems inspired by functions, principles and models of the vertebrates' immune system. Specifically, we studied the combinatorial version of the Concentration-based Artificial Immune Network – cob-aiNet[C].

Based on cob-aiNet[C], the cobICA technique was presented in the sequence, which is an immune-inspired search algorithm to perform ICA over Galois Fields. A particularity of cobICA, with respect to previous approaches (YEREDOR, 2011a; SILVA *et al.*, 2014), is the employment of the MMI criterion and problem-tailored full-rank-preserving operators.

In order to analyze, for the first time, the comparative performance of all the algorithms presented in this work, in the next chapter simulation results are presented, regarding the cobICA, AMERICA and MEXICO algorithms.

CHAPTER 4

4 - Simulation Results

In this chapter, 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 MATLAB code of cobICA is available in <http://danielgs.weebly.com> and AMERICA and MEXICO code are available at <http://www.eng.tau.ac.il/~arie/ICA4GFP.rar>.

The performance metric to analyze the quality of each proposal in the separation task is the average success rate: it is the mean ratio between the numbers of extracted sources and N , where each algorithm estimates a separating matrix \mathbf{W} and the quality of \mathbf{WA} is evaluated by counting the number of rows with strictly one non-null entry, indicating an extracted source.

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 PMFs. The mixing matrix \mathbf{A} is also randomly generated, which yields T i.i.d. observations from each mixture x_i to be applied as input for the algorithms.

Recall that ICA only works for non-uniform and non-degenerate distributions (see Section 2.2), so we consider for the simulations probability vectors whose Kullback–Leibler divergences, provided by Equation 2.16, to the uniform distribution are greater than or equal to 0.2 and with all symbol probabilities within the interval $(0, 0.98]$.

The cob-aiNet[C] parameters used by cobICA are defined as the same that were adopted after a cross-validation procedure performed in (SILVA *et al.*, 2014), displayed in Table 4.1, and are fixed for all experiments that follow.

The simulation results are organized according to the different scenarios in the following four sub-sections, in order to facilitate the reader’s understanding. In Sub-section 4.1, we study the behavior of cobICA by itself, with fields order $q = 2$, $q = 3$, $q = 4$ and $q = 5$.

Subsequently, in Sub-section 4.2, the comparison between the techniques in terms of average success rate analysis, when the number of samples increases (from $2^5=32$ to $2^{10}=1024$), is presented. Due to the lack of implementations of AMERICA and MEXICO that operate with non-prime fields, the comparative analysis consider only prime fields, i.e. $q = P$, specifically $P = 2, 3$ and 5 .

Similarly, in Sub-section 4.3 the comparison between the techniques continues, in terms of average success rate analysis, but, now the number of sources (N) increases (from 2 to 8), with field orders $P = 3$ and $P = 5$, and with a fixed number of samples.

Finally, an empirical computational complexity comparison between the techniques is drawn in Sub-section 4.4.

Table 4.1 - Simulation parameter's of cobICA algorithm. Adapted from (SILVA *et al.*, 2014).

| | |
|----------------|-------------|
| n_{AB} | 2 |
| max_{AB} | 100 |
| n_{Cl}^{min} | 2 |
| n_{Cl}^{max} | 10 |
| β^i | $0.8 N^2$ |
| β^f | $0.008 N^2$ |
| LS_{it} | 1 |
| LS_{freq} | 1 |
| $maxIT$ | 300 |

4.1 – Simulations of cobICA

In Figures 4.1.1 (a, b and c), the behavior of cobICA with fields order $q = 2$, $q = 3$ and $q = 5$ indicates that, in order to get a better performance for high dimension cases, more samples are needed to estimate the fitness / objective function.

Although Figure 4.1.1.a shows that a less number of sources with higher number of samples performs better, since the study is implemented in $GF(2)$, which presents a small

search space size, the gain between the different number of sources performance is close to each other.

On the other hand, in Figures 4.1.1.b and 4.1.1.c, due to the search space dimension increasing, the differences between the performances increased as well. Also, the result in Figure 4.1.1.b shows that the cobICA algorithm has the best performance, in this numerical simulation, for $N = 4$ and $GF(3)$.

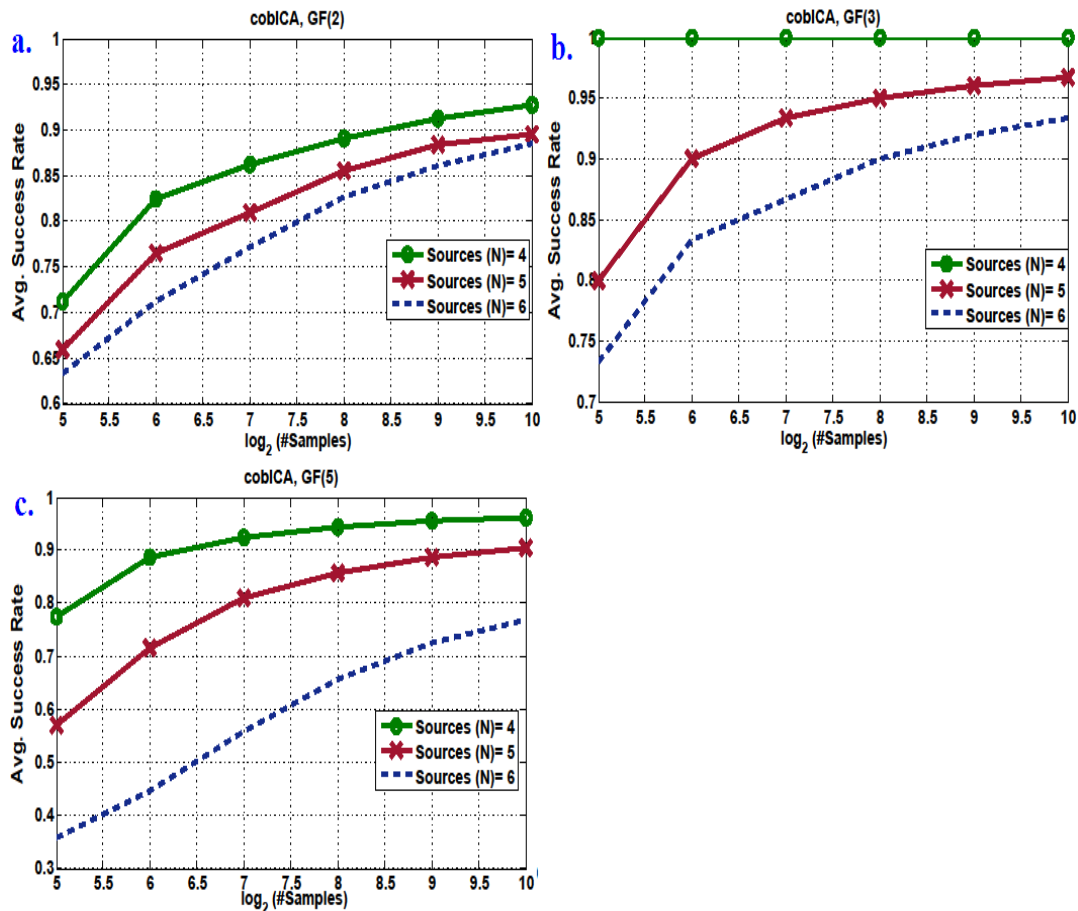


Figure 4.1.1 - Performance of cobICA for different values of sources (N), $P = 2, 3$ and 5 .

Figure 4.1.2 exhibits the analysis for different field orders, when the number of sources is fixed with $N = 4$. Since $GF(4)$ implies more number of symbols than $GF(2)$ and $GF(3)$, the associated search space became larger, and so, it presents a lower quality performance than the others, which was expected as a natural behavior of cobICA algorithm. On the other hand, $GF(2)$ implies the smallest number of symbols and, consequently, a smaller search

space, it was expected to provide a better result than the $GF(3)$ case, however the empirical results show an opposite behavior, where the $GF(2)$ case presents a worse performance than $GF(3)$.

In order to find an hypothesis for this behavior we refer to the results obtained by a Michigan-like¹ implementation of cob-aiNet[C] for ICA over GF in (SILVA *et al.*, 2014a), where a similar behavior between the $GF(2)$ and $GF(3)$ scenarios was noticed. In that work, this fact is interpreted due a possible damage caused by the “*imperfection*” of the dissimilarity metric, which is also the Hamming distance. This means, in other words, that some good candidate solutions may be mistakenly suppressed during the algorithm execution.

Although there are clear differences between cobICA algorithm and the Michigan-like approach, where, especially in cobICA algorithm, the technique works with whole separating matrices as being candidate solutions and with the MMI criterion, instead of extraction vectors (lines of the separating matrix) and with minimum-entropy criterion in the Michigan-based approach, Hamming distance is the dissimilarity metric among the solutions for cobICA algorithm, as well.

Hence, we consider still valid the hypothetical explanation presented by SILVA *et al.*, (2014a), where the authors referred to WATERHOUSE (1987), who has shown that, based on Equation 2.5, the probability that a matrix ($N \times N$) in $GF(q)$ has null determinant is

$$Prob[N, q] = 1 - (1 - q^{-1}) (1 - q^{-2}) \dots (1 - q^{-N}). \quad (4.1)$$

This equation indicates that $Prob[4; 2] \approx 0.6923$, while $Prob[4; 3] \approx 0.4365$ and $Prob[4; 5] \approx 0.2392$. These values can induce that a mistaken suppression of a solution candidate (an invertible matrix) when $q = 2$ reduces the already relatively small set of invertible matrices (in comparison to the set of non-invertible ones) that comprise the feasible solutions set, as stated in Equation 4.1.

¹ The Michigan method for population-based algorithms states that, at the end of the execution, the problem solution is composed by the entire population, while in the Pittsburgh approach, only the individual with best fitness composes the final solution (BACK *et al.*, 2000).

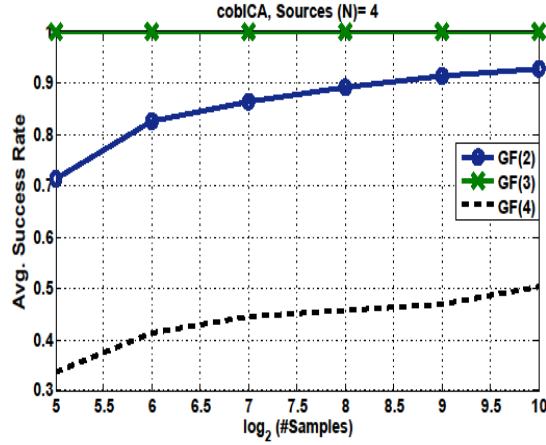


Figure 4.1.2 - Performance of cobICA for different field orders, $N = 4$.

In order to analyze the sensitivity to the non-uniform restriction (recall Section 2.2), Figures 4.1.3 (a and b) show the behavior of cobICA with different thresholds for the Kullback-Leibler divergence between the PMF of each source, defined in Equation 2.16, and the uniform distribution.

The results reinforce that ICA over GF only works for non-uniform distributions, as already mentioned in Section 2.2, such that, the higher is the threshold, the more “non-uniform” the sources are generated.

As Figure 4.1.3.a shows, when $q = 3$ a similar performance can be achieved with higher number of samples ($T = 2^{10}$), which compensates the characteristics of the distribution. On the other hand, as Figure 4.1.3.b shows, with 6 sources and $q = 5$, the discrepancy is increased with higher number of samples. This indicates that as the search space increases in $GF(5)$, the algorithm becomes more sensitive to the threshold value, in contrast with $GF(3)$ in Figure 4.1.3.a, where a similar performance can be achieved for increased number of samples.

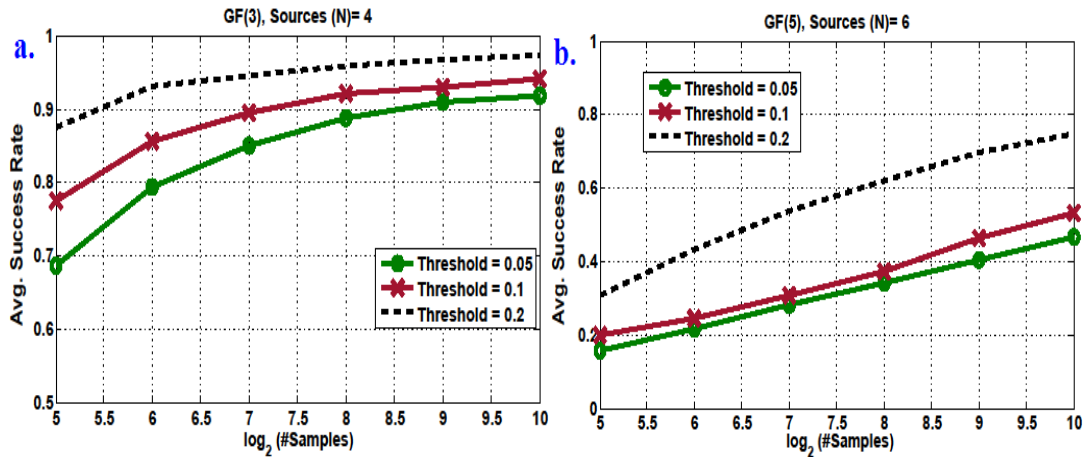


Figure 4.1.3 - Performance of cobICA when the non-uniform character of distributions is changed, (a) $N = 4$, $P = 3$, (b) $N = 6$, $P = 5$.

4.2 – Comparative simulations for variable number of samples

In the following, a comparative analysis takes place. Figures 4.2.1, 4.2.2 and 4.2.3 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 $P = 2, 3$ and 5 .

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

As Figure 4.2.1.a shows, when the algorithms imply less number of sources, all the methods could achieve almost the same performance, which was expected because of the small size of the search space, and also the performances increased with higher number of samples, due to the natural improvement on estimating the cost functions associated to each method.

On the other hand, with the addition of one more source in the case depicted by Figure 4.2.1.b, a discrepancy between the performances appears, such that AMERICA and MEXICO present better results than cobICA, although their results become closer with higher number of samples.

A similar situation can be seen in Figure 4.2.1.c too, with the difference that MEXICO presents a quality level closer to cobICA, while AMERICA still maintains a perceivable margin and presents the best performance among all methods.

Note that in Figure 4.2.1, since it is employed a field order $P = 2$, with a relatively small size of the search space for all scenarios, generally MEXICO shows better results than cobICA. Furthermore, this behavior of MEXICO is quite similar to GUTCH *et al.*, (2012) simulation's results, where for the fields order $P = 2$ and $P = 3$, MEXICO performances are close to AMERICA, mainly for the less number of sources scenarios.

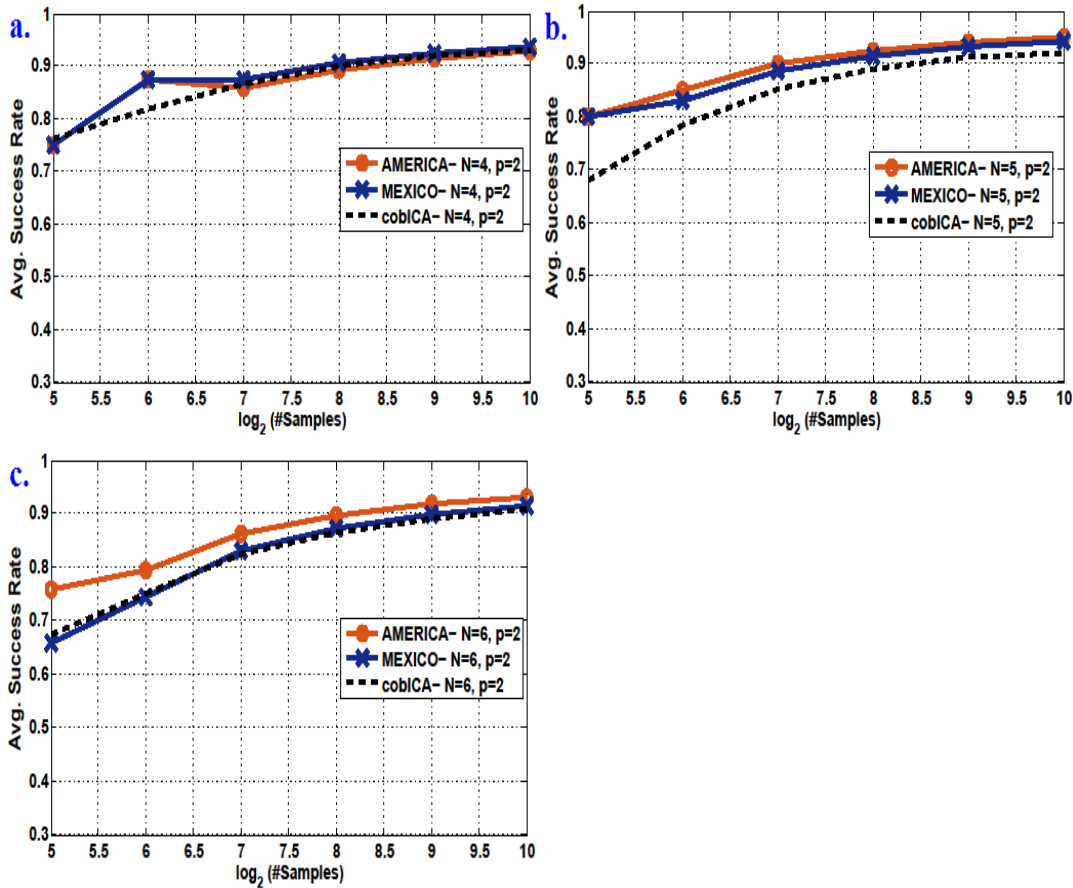


Figure 4.2.1 - Comparison among the techniques for field orders $P = 2$, (a) $N = 4$, (b) $N = 5$, (c) $N = 6$.

In Figure 4.2.2 (a, b and c) the comparison between the techniques is again analyzed in terms of average success rate, but, this time with field order $P = 3$.

As Figure 4.2.2.a shows, the performances of AMERICA and cobICA achieve full success in separation, while MEXICO presents a lower performance than the both ones. 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 4.2.1, in contrast with the cobICA algorithm, which shows its best performance for $N = 4$ and $GF(3)$.

Furthermore, in both Figures 4.2.2.b and 4.2.2.c, as expected the performances get decreased due to the search space increasing, with AMERICA being the top-ranked method, closely followed by cobICA and, as the last-ranked strategy, MEXICO.

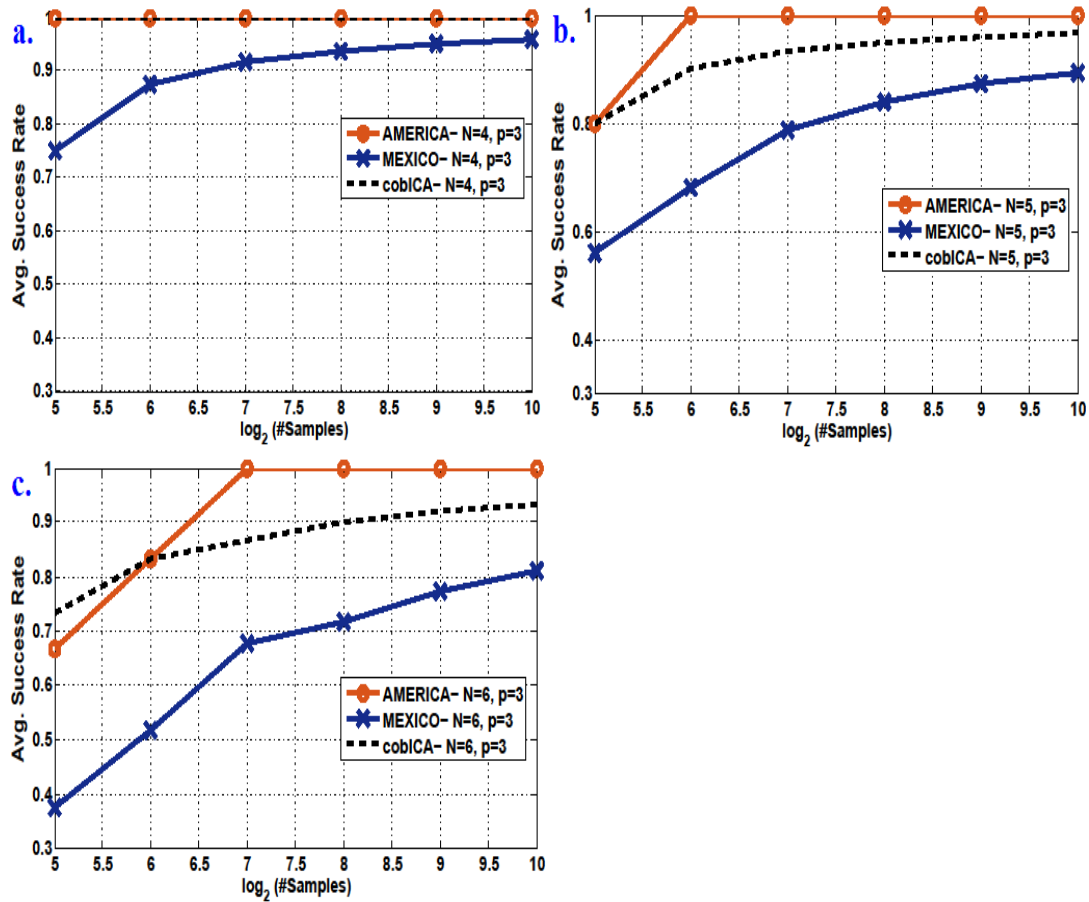


Figure 4.2.2 - Comparison among the techniques for field order $P = 3$, (a) $N = 4$, (b) $N = 5$, (c) $N = 6$.

Figure 4.2.3 (a, b and c) shows the same type of comparative analysis but, now, with a field order $P = 5$.

Figure 4.2.3.a indicates cobICA and AMERICA algorithms with a very close performance, even for low number of samples, while the line representing MEXICO shows that it performs close to the others algorithms only when the number of samples is increased, otherwise it doesn't present the same quality level.

Also, in Figures 4.2.3.b and 4.2.3.c, both graphics show that, although cobICA couldn't achieve a quality performance as good as AMERICA, it is still better than MEXICO, especially in Figure 4.2.3.c.

Note that in Figure 4.2.3.c, where the search space is large enough, the gain of the MEXICO increased more, compared with others, even for high number of samples.

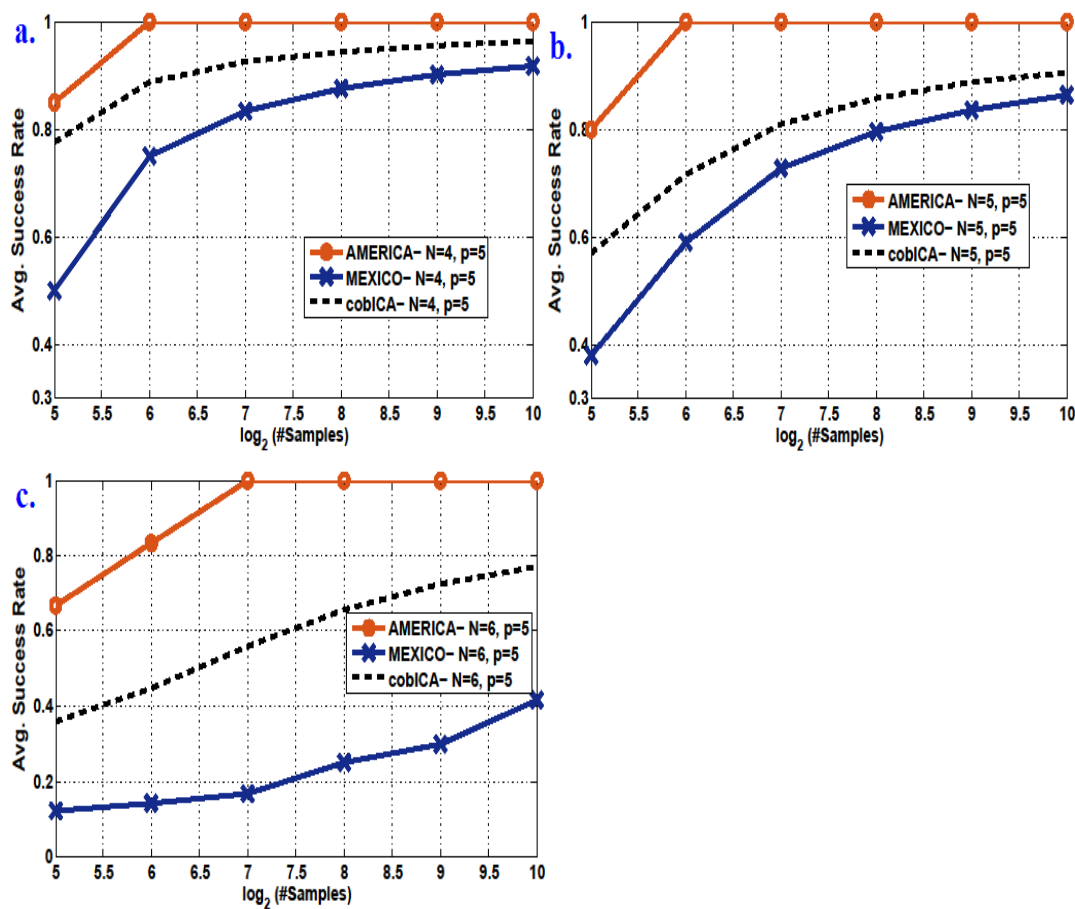


Figure 4.2.3 - Comparison among the techniques for field order $P = 5$, (a) $N = 4$, (b) $N = 5$ and (c) $N = 6$.

4.3 – Comparative simulations for a variable number of sources

Now considering a field order $P = 3$, a different analysis is displayed in Figure 4.3.1 (a and b), which shows the comparison between the techniques when N is increased, for a fixed number of samples: $T = 32$ (a) and $T = 64$ (b). A relatively small number of samples were initially chosen in order to emphasize the performances variability between the methods, possibly avoiding that a certain technique obtains full separation in all trials. The results show that the performance of AMERICA and cobICA are similar, with a superiority for the former, while MEXICO presents a lower quality degree; naturally, when N increases all algorithms decrease performance, but, interestingly, for $T = 32$ AMERICA degrades quicker than cobICA.

Furthermore, Figure 4.3.1.c shows the same type of comparison between the techniques, but with higher numbers of sources and a higher number of samples ($T = 512$). It is noticeable the considerable reduction of MEXICO algorithm results, in comparison to the other two approaches, which are reasonably capable of maintaining good performances until $N = 12$, with, again, an advantage for AMERICA technique.

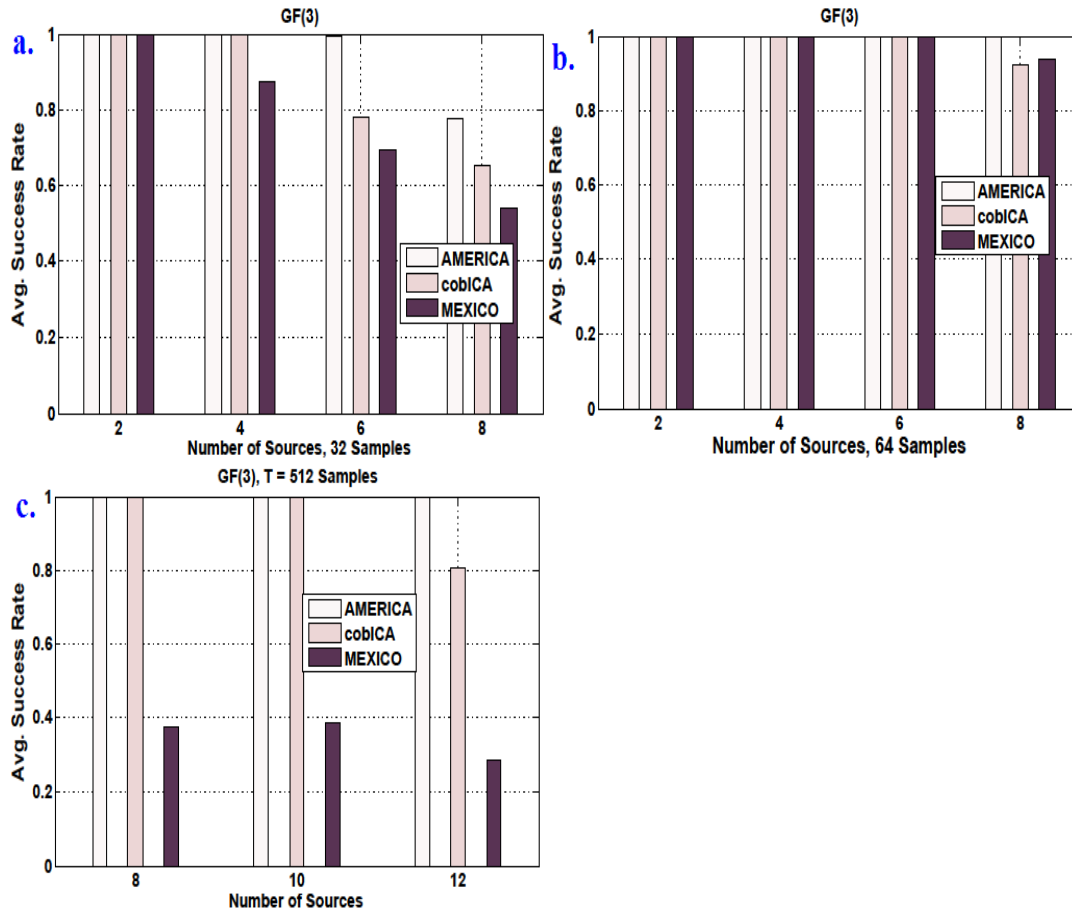


Figure 4.3.1 - Comparison among the techniques for different values of sources when field order $P = 3$, (a) 32 samples, (b) 64 samples and (c) 512 samples.

We repeat the analysis of quality versus number of sources previously seen in Figure 4.3.1, but for a field order $P = 5$ in Figure 4.3.2. The results show that the performance of AMERICA and cobICA are close in the lowest dimension cases, while MEXICO is not successful enough.

Although for $T = 32$, which is shown in Figure 4.3.2.a, all the algorithms degrade their performance levels fast, cobICA results seem to degrade quicker than the others. Moreover, Figure 4.3.2.b reminds that, as expected, when N increases all algorithms performances decrease as well.

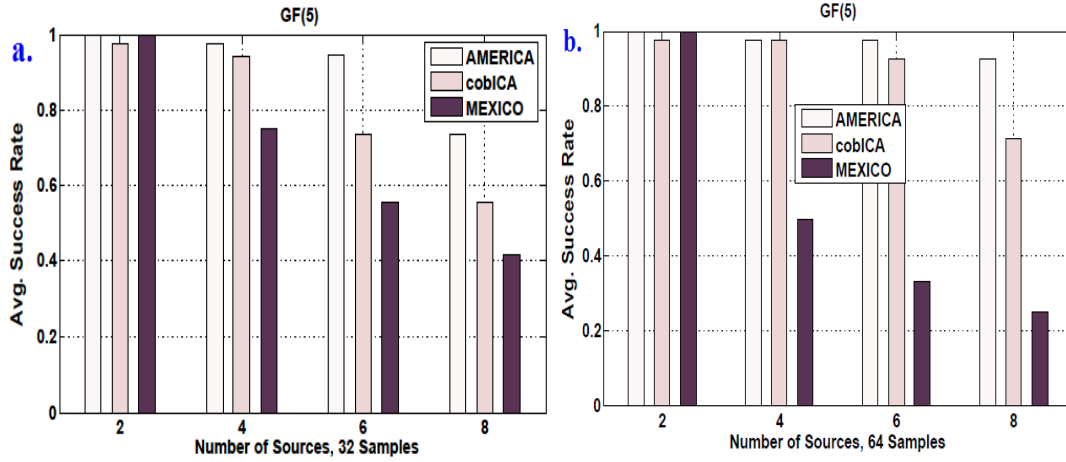


Figure 4.3.2 - Comparison among the techniques for different values of sources when field order $P = 5$, (a) 32 samples, (b) 64 samples.

4.4 – Computational complexity comparison

The comparative results that were discussed in the previous sections point out that AMERICA yields the best separation quality, followed by 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 sub-section experiment compares the computational costs of cobICA, AMERICA and MEXICO, by considering the number of times that the crucial function that is the basis of each algorithm criterion, the entropy of a given component, is evaluated for each method, since this operation is, besides the explicit importance, the most costly operation for all techniques. 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 author mentions in (YEREDOR, 2011a), for AMERICA algorithm, an exhaustive search over all non-trivial candidate vectors is executed to extract each source, then $N \cdot (P^N - 1)$ values of entropy are estimated or, in big-O notation, AMERICA is $O(NP^N)$, as it is mentioned in Section 2.4. Differently than AMERICA, MEXICO and cobICA have a non-deterministic number of entropy calculations, 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 algorithm runs.

As Figure 4.4.1 shows, although MEXICO offers a smaller computational complexity than the other two methods, as already shown in the previous experiments, especially in Figure 4.3.1.c, this benefit comes with the burden of the poorest overall quality in separation. Moreover, AMERICA has lower values than cobICA until $N = 12$, when in this case the exponential computational cost points the highest cost level. In contrast, although cobICA starts with a relatively high level, the algorithm seems to present a lower increasing pattern, with respect to the number of sources. Despite the actual number of entropy evaluations of cobICA and MEXICO may vary due to their non-deterministic character, this comparison indicates that cobICA, specifically, has an intermediate computational cost, taking place between the most expensive AMERICA and the cheapest MEXICO, for higher numbers of sources.

Therefore we can infer that cobICA technique presents a compromise between scalability and separation quality, considering time constraints, specially when the number of sources is increased.

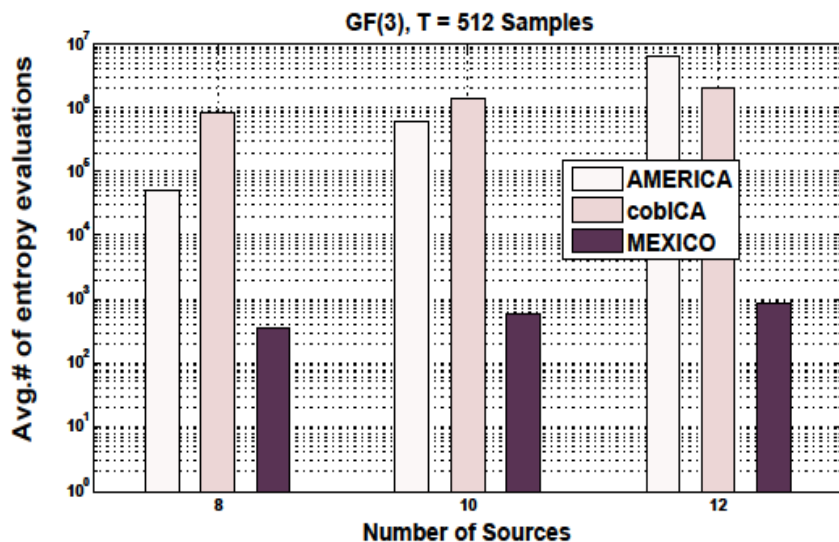


Figure 4.4.1 - Number of entropy evaluations for different values of sources when field order $P = 3$.

Figure 4.4.2 presents the same type of comparison between the techniques, but when the number of samples increases (from $2^5 = 32$ to $2^{10} = 1024$) and a fixed number of sources is adopted ($N = 8$).

The result shows the same performance already achieved in Figure 4.4.1, where in $N = 8$, MEXICO has the smallest computational complexity and cobICA offers the biggest one, while MEXICO has the worst quality, recall Figure 4.3.1.c.

But differently than Figure 4.4.1, where the computational complexity of all algorithms has considerably increased by increasing the number of sources, in Figure 4.4.2 the performances of all the algorithms remain almost fixed when the number of samples increases. This means that increasing the number of the samples may not change the algorithms number of iterations, and, consequently, it does not cause significant changes in the number of entropy evaluations.

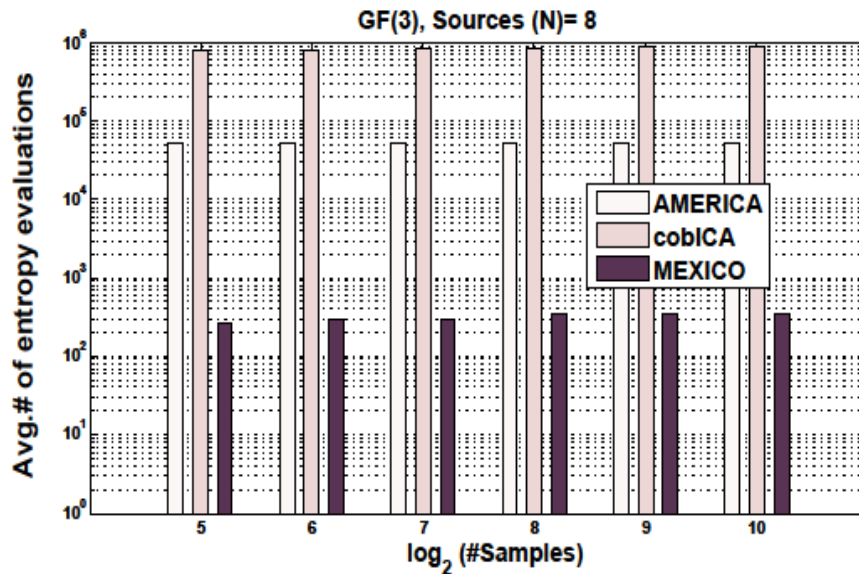


Figure 4.4.2 - Number of entropy evaluations for different numbers of samples when field order $P = 3$ and $N = 8$.

CHAPTER 5

5 - Concluding Remarks

This dissertation studies Blind Source Separation and its associated solution via Independent Component Analysis, in the context of signals and models defined over Galois fields. Three modern algorithms for ICA over GF, namely AMERICA, MEXICO (GUTCH *et al.*, 2012) and cobICA (SILVA *et al.*, 2014), are studied and analyzed in a pioneering comparative context.

The goal of ICA over GF is to recover independent components from a set of mixtures that are originated by an (unknown) linear mixing process of (also unknown) independent source signals, as shown in Equation 2.3. Due to the discrete nature of all entities that are involved, the problem can be formulated as a combinatorial optimization task. In order to solve it, a cost function that measures the dependence degree between the extracted components $\mathbf{y}(n)$ should be defined.

In this context, the AMERICA and MEXICO algorithms, which try to find the lowest entropy linear combination of mixtures to yield independent signals, are effective heuristics that have shown results of good reliability. While AMERICA and MEXICO follow the same base idea, iteratively estimating the independent components by reducing their entropy values, their main difference is in the search process: AMERICA evaluates the entropy of an extracted signal which is generated by combining N mixtures, but MEXICO makes an equivalent entropy evaluation only between combined pairs of mixtures.

Under a different perspective, the cobICA algorithm implements an immune-inspired search strategy, the cob-aiNet[C] algorithm, to optimize a mutual information-based cost function. Nevertheless, the operators of mutation and local search still resemble heuristic strategies, such as the full-rank operations adopted in MEXICO technique. According to the experimental results, the cobICA algorithm has shown to be a competitive method for small number of sources and, also, presented good success rates for more complex instances.

By shifting the context to the computational burden of each technique, a comparison between the methods shows that cobICA costs less than AMERICA when the number of sources increases, with respect to the number of entropy evaluations. Notwithstanding, the less costly algorithm according to the experimental evidences is MEXICO, but this efficiency is provided due to a search procedure that is prone to local convergence and, consequently, less quality separation results (which were confirmed by the experimental analysis).

Therefore, it is possible to infer that the main advantage of cobICA technique lies on offering good separation solutions with an asymptotic smaller computational complexity, for which the experimental results yield to a lower increasing behavior, with respect to AMERICA, as shown in Figure 4.4.1. Those evidences promote cobICA technique to be a possible solution in allowing the application of ICA to BSS and factor analysis in the context of large datasets of discrete nature. On the other hand, for small numbers of components, AMERICA algorithm is confirmed, by the studies developed in this dissertation, as the best approach for ICA over GF.

5.1 - Future perspectives

According to the achieved results, there are several possible future works to be performed, while several potential applications are already being explored, such as: ICA for eavesdropping a Tomlinson-Harashima pre-coded MIMO channel (YEREDOR 2011a; GUTCH *et al.*, 2012); a binary independent component analysis approach (NGUYEN, ZHENG, 2011; NGUYEN, ZHENG, 2013) and the application of ICA in the context of Network Coding (NEMOIANU *et al.*, 2013).

Moreover, it would be interesting to extend the comparative analysis of cobICA, AMERICA and MEXICO for non-prime fields, orders higher than $P = 5$ and to study cobICA with different dissimilarity metrics and different cost function estimators.

In terms of practical applications, there are open possibilities of implementing ICA over GF in real time applications for digital telecommunication systems, image processing, as well as in radar, sonar and multidimensional digital data in a general perspective.

References

- ADALI, T. and HAYKIN, S. (Ed.) (2010). “Adaptive Signal Processing: Next Generation Solutions.” In: *Hoboken, USA: Wiley-IEEE Press*, 2010.
- 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.
- BACK, J. FOGEL, D. B. and MICHALEWICZ, Z. EDS. (2000). “Evolutionary Computation1: basic algorithms and Operations.” In: *Taylor & Francis*. Bristol, UK.
- BELL, A. and SEJNOWSKI, T. (1997). “The independent components of natural scenes are edgeless.” In: *Vision Research*, 37(23):3327{3338, 1997}.
- BURNET, F. M. (1959). “The Clonal Selection Theory of Acquired Immunity.” In: *Cambridge University Press*, 1959.
- CARDOSO, J.-F. (1992). “Iterative techniques for blind source separation using only fourth-order cumulants.” In: *Proc. EUSIPCO*, pages 739-742, Brussels, Belgium.
- CARDOSO, J.-F. (1989). “Source separation using higher order moments”, In: *Proc. ICASSP'89*, pages 2109-2112.
- CARLTON, A.G (1969). “On the bias of information estimates.” In: *Psychological Bulletin* 71(2)(1969)108–109.
- CASTEDO, L.; ESCUDERO, C. and DAPENA, A. (1997). “A blind signal separation method for multiuser communications.” In: *IEEE Transactions on Signal Processing*, 45(5):1343{1348, 1997}.
- 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.
- COELHO, G.P. and VON ZUBEN, F.J. (2010). “A concentration-based artificial immune network for continuous optimization.” In: *2010 IEEE Congresson Evolutionary omputation (CEC), IEEE*, 2010, pp.1–8.
- COMON, P. (1994). “Independent component analysis, a new concept.” In: *Signal Processing*, 36(3):287{314, 1994}.
- COMON, P. and JUTTEN, C. (Ed.). (2010). “Handbook of Blind Source Separation.” In: *Oxford, UK: Academic Press*, 2010. Chapter 12 pages, 21, 23, 24, 29, 30, 32, 38, 40, 58, 65 and 87.
- COVER, T.M. and THOMAS, J.A. (2006). “Elements of Information Theory.” In: *2nd ed. Wiley-Interscience*.

- CHAOLI WANG and SHEN HAN-WEI. (2011). “Information Theory in Scientific Visualization”. In: *Entropy* 2011, 13, 254-273; doi:10.3390/e13010254. Available at: ([http:// www.mdpi.com/journal/entropy](http://www.mdpi.com/journal/entropy)).
- CRISTESCU, R.; RISTANIEMI, T.; JOUSENSALO, J. and KARHUNEN, J. (2000). “Delay estimation in CDMA communications using a Fast ICA algorithm.” In: *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, ICA 2000*, pages 585{590, 2000}.
- CRUZ CORTES, N. and COELLO COELLO, C. (2003). “Multi objective optimization using ideas from the clonal selection principle.” In: *E. Cantu-Paz (Ed.), Genetic and Evolutionary Computation (GECCO)*, vol. 1, 2003, pp. 158–170.
- DE CASTRO, L. N. (2001). “Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais”. In: *Ph.D. Thesis*, — Universidade Estadual de Campinas, 2001.
- DE CASTRO, L.N.. (2006). “Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications.” In: *Chapman & Hall/CRC*.
- DE CASTRO, L. N. and J. TIMMIS. (2002a). “An Artificial Immune Network for Multimodal Function Optimization”. In: *2002 IEEE Congress on Evolutionary Computation (CEC)*. 2002. p. 699–704.
- DE CASTRO, L.N. and J. TIMMIS. (2002). “Artificial Immune Systems: A New Computational Intelligence Approach.” In: *Springer*.
- DE CASTRO, L.N. and VON ZUBEN, F.J. (2002). “Learning and optimization using the clonal selection principle.” In: *IEEE Transactions on Evolutionary Computation* 6(3) (2002) 239–251.
- DE FRANCA, F.O.; COELHO, G.P. and VON ZUBEN, F.J. (2010). “On the diversity mechanism of opt-ainet; a comparative study with fitness sharing.” In: *2010 IEEE Congress on Evolutionary Computation(CEC), IEEE, 2010*, pp. 1-8.
- DELFOSSÉ, N. and LOUBATON, P. (1995). “Adaptive blind separation of independent sources: a deflation approach.” In: *Signal Processing* 45(1)(1995) 59–83.
- FENG, M. and KAMMAYAR, K.-D. (1999). “Application of source separation algorithms for mobile communications environment.” In: *Proc. of the First International Workshop on Independent Component Analysis and Signal Separation, ICA'99*, pages 431{436, 1999}.

- GUTCH, H. W.; GRUBER, P. and THEIS, F. J. (2010). "ICA over finite fields." In: *ICA2010—Latent Variable Analysis and Signal Separation*, Springer, 2010, pp. 645–652.
- 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. ISSN 01651684. Chapter 6, pages 25, 31, 57, 68, 76 and 80.
- HAYKIN, S. (1994). "Communication Systems. 3. ed." [S.l.] In: *John Wiley & Sons*.
- HYVARINEN, A. and HOYER, P. (2000). "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces." In: *Neural Computation*, 12(7):1705{1720, 2000}.
- HYVARINEN, A.; KARHUNEN, J. and OJA, E. (2001). "Independent Component Analysis." In: *John Wiley & Sons*.
- ICAO. (2004). "Manual on the Secondary Surveillance Radar (SSR) Systems". In: *International Civil Aviation Organization*. Approved by the Secretary General and published under his authority. Third Edition – 2004.
- IKEDA, S. and MURATA, N. (1999). "A method of ICA in time-frequency domain." In: *Proc. of the First International Workshop on Independent Component Analysis and Signal Separation*, ICA'99, pages 365{370, 1999}.
- 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.
- JUNG, T.; MAKEIG, S.; MCKEOWN, M.; BELL, A.; LEE, T. and SEJNOWSKI, T. (2001). "Imaging brain dynamics using independent component analysis." In: *Proceedings of the IEEE*, 9(7):1107{1122, 2001}.
- JERNE, N.K. (1974). "Towards a network the oryof the immune system." In: *Annales d'immunologie* 125(1–2) (1974)373–389.
- KAGAN, A. M.; LINNIK, Y. V. and RAO, C. R. (1973). "Characterization Problems in Mathematical Statistics." In: *Wiley*, New York.
- KELSEY, J. and TIMMIS, J. (2003). "Immune inspired somatic contiguous hyper mutation for function optimisation." In: *Genetic and Evolutionary Computation Conference*, in: *Lecture Notes in Computer Science*, vol. 2723, Springer, 2003, pp. 207–218.
- LEE, T. W. (1998). "Independent Component Analysis." In: *Kluwer Academic Publishers*.

- 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*.
- LIDL, R. and NIEDERREITER, H. (1997). "Finite Fields." In: *vol.20, Cambridge University Press*.
- 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.
- MANSOUR, A.; BARROS, A. K. and OHNISHI, N. (2000). "Blind separation of sources: Methods, assumptions and applications", In: *IEICE Trans. On Fundamental of Electronics, Communications and Computer Sciences*, vol.E83-A, pp. 1498-1512, 2000.
- NEMOIANU, I.; GRECO, C.; CASTELLA M.; PESQUET-POPESCU B. and CAGNAZZO M. (2013). "On a practical approach to source separation over finite fields for network coding applications," IN: *Proc., IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, pp. 1335 – 1339, 2013.
- NGUYEN HUY and ZHENG RONG, (2011). "Binary independent component analysis with or mixtures," In: *Signal Processing, IEEE Transactions on*, vol. 59, no. 7, pp. 3168–3181.
- NGUYEN HUY and ZHENG RONG. (2013). "A binary independent component analysis approach to tree topology inference," In: *Signal Processing, IEEE Transactions on*, vol. 61, no. 12, pp. 3071– 3080.
- NICOSIA, G. (2004). "Immune algorithms for optimisation and protein structure prediction." *Ph.D. Thesis*, Department of Mathematics and Computer Science, University of Catania, Italy.
- NIKIAS and MENDEL J. (1993). "Signal processing with higher-order spectra." In: *IEEE Signal Processing Magazine*, pages 10-37, July 1993.
- PETROCHILOS, N. (2002). "Algorithms for Separation of Secondary Surveillance Radar Replies." *PhD Thesis*, Delft University Press.
- PRINCIPE, J. C. (2010). "Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives." In: *New York, USA: Springer, 2010*. ISBN 9781441915696. Chapter 8, pages 12, 14, 16, 18, 19, 20, 63 and 100.
- PRINCIPE, J. C.; XU DONGXIN; ZHAO QUN and FISHER III JOHN W. (2000). "Learning from Examples with Information Theoretic Criteria" In: *Journal of VLSI*

- signal processing systems for signal, image and video technology*, August 2000, Volume 26, Issue 1, pp 61-77.
- ROMANO, J. M. T.; ATTUX, R.; CAVLCANTE, C. and SUYAMA, R. (2011). “Unsupervised Signal Processing: Channel Equalization and Source Separation.” In: *CRC Press*, 2011.
- SILVA, D. G.; (2013). “Aprendizado de Máquina Baseado na Teoria da Informação: Contribuições à Separação de Sinais em Corpos Finitos e Inversão de Sistemas de Wiener.” *Ph.D. Thesis*, Department of Electrical and Computer Engineering, University of CAMPINAS, SP, Brazil.
- SILVA, D. G.; NADALIN, E. Z.; COELHO, G. P.; DUARTE, L. T.; SUYAMA, R.; ATTUX, R.; VON ZUBEN, F. J. and MONTALVAO, J. (2014a). “A Michigan-like immune-inspired framework for performing independent component analysis over Galois fields of prime order.” In: *Signal Processing 96(2014)153–163*.
- 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.
- SHANNON, C. E. (1948). “A Mathematical Theory of Communication.” In: *The Bell system Technical Journal*, v. 27, p. 379–423, 623–656, 1948. Chapter 4, pages 7, 8, 10 and 11.
- TORKKOLA, K. (1999). “Blind separation of audio signals.” In: *Proc. of the First International Workshop on Independent Component Analysis and Signal Separation, ICA'99*, pages 239{244, 1999}.
- VALKAMA, M.; RENFORS, M. and KOIVUNEN, V. (2001). “Advanced methods for i/q imbalance compensation in communication receivers.” In: *IEEE Transactions on Signal Processing*, 49:2335{2344, 2001}.
- 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}.
- WATERHOUSE, W.C. (1987). “How often do determinants over finite fields vanish ?.” In: *Discrete Mathematics 65(1)(1987)103–104*.
- YEREDOR, A. (2007). “ICA in Boolean XOR mixtures.” In: *ICA2007—Independent Component Analysis and Signal Separation, Springer*, 2007, pp.827–835.

- YEREDOR, A. (2011a). "Independent Component Analysis Over Galois Fields of Prime Order." In: *IEEE- Transactions on Information Theory*, v. 57, n. 8, p. 5342–5359.
- YEREDOR, A. (2011). "MATLAB Code for ICA Over GF(P), AMERICA and MEXICO." Available at: <http://www.eng.tau.ac.il/~arie/ICA4GFP.rar>.
- ZHANG, Y. and KASSAM, S. (2001). "Blind separation and equalization using fractional sampling of communication signals." In: *Signal Processing*, 81(12):2591{2608, 2001}.