



Daniel Guerreiro e Silva

# **Aprendizado de Máquina Baseado na Teoria da Informação: Contribuições à Separação de Sinais em Corpos Finitos e Inversão de Sistemas de Wiener**

Campinas

2013





UNIVERSIDADE ESTADUAL DE CAMPINAS  
Faculdade de Engenharia Elétrica e de Computação

Daniel Guerreiro e Silva

**Aprendizado de Máquina Baseado na Teoria da Informação:  
Contribuições à Separação de Sinais em Corpos Finitos e  
Inversão de Sistemas de Wiener**

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica, na Área de Engenharia de Computação.

Orientador: Prof. Dr. Romis Ribeiro de Faissol Attux

Co-orientador Prof. Dr. Jugurta Rosa Montalvão Filho

Este exemplar corresponde à versão final da tese defendida pelo aluno Daniel Guerreiro e Silva, e orientada pelo Prof. Dr. Romis Ribeiro de Faissol Attux

---

Campinas

2013

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

Si38a Silva, Daniel Guerreiro e, 1983-  
Aprendizado de máquina baseado na teoria da informação : contribuições à separação de sinais em corpos finitos e inversão de sistemas de Wiener / Daniel Guerreiro e Silva. – Campinas, SP : [s.n.], 2013.

Orientador: Romis Ribeiro de Faissol Attux.  
Coorientador: Jugurta Rosa Montalvão Filho.  
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Aprendizado de máquina. 2. Teoria da informação. 3. Sistemas imunológico. 4. Análise de componentes independentes. 5. Processamento digital de sinais. I. Ribeiro de Faissol Attux, Romis, 1978-. II. Montalvão Filho, Jugurta Rosa. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Information theoretic learning : contributions to signal separation over finite fields and inversion of Wiener systems

**Palavras-chave em inglês:**

Machine learning  
Information theory  
Immune systems  
Independent component analysis  
Digital signal processing

**Área de concentração:** Engenharia de Computação

**Titulação:** Doutor em Engenharia Elétrica

**Banca examinadora:**

Romis Ribeiro de Faissol Attux [Orientador]  
Luis Cláudius Coradine  
Charles Casimiro Cavalcante  
Reginaldo Palazzo Júnior  
Leonardo Tomazeli Duarte

**Data de defesa:** 31-10-2013

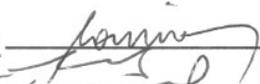
**Programa de Pós-Graduação:** Engenharia Elétrica

## COMISSÃO JULGADORA - TESE DE DOUTORADO

**Candidato:** Daniel Guerreiro e Silva

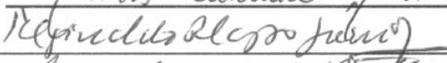
**Data da Defesa:** 31 de outubro de 2013

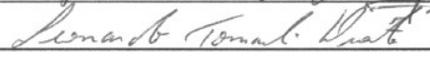
**Título da Tese:** "Aprendizado de Máquina Baseado na Teoria da Informação: Contribuições à Separação de Sinais em Corpos Finitos e Inversão de Sistemas de Wiener"

Prof. Dr. Romis Ribeiro de Faissol Attux (Presidente): 

Prof. Dr. Luis Cláudius Coradine: 

Prof. Dr. Charles Casimiro Cavalcante: 

Prof. Dr. Reginaldo Palazzo Júnior: 

Prof. Dr. Leonardo Tomazeli Duarte: 



# Resumo

Esta tese de doutorado possui como tema geral o desenvolvimento de algoritmos de Aprendizado de Máquina Baseado na Teoria da Informação (ITL — *Information Theoretic Learning*). O paradigma de ITL propõe o uso de critérios de treinamento baseados em medidas como entropia e informação mútua, em substituição aos tradicionais critérios baseados em estatísticas de segunda ordem.

Os problemas de inversão cega de sistemas de Wiener e separação cega de sinais em corpos de Galois são os objetos de estudo e desenvolvimento dessas ferramentas. Estes problemas apresentam características marcantes quanto à necessidade de descritores estatísticos de ordem superior, por isso, apresenta-se uma série de contribuições que se baseiam em critérios de ITL e empregam algoritmos imuno-inspirados (ou heurísticas de busca) para adaptar os parâmetros dos modelos envolvidos.

As propostas desenvolvidas abrem a perspectiva de futuras aplicações em áreas como genômica, codificação e geofísica. Espera-se também que os resultados apresentados fortaleçam um entendimento mais amplo de ITL, a fim de abranger estratégias mais genéricas de busca, de estimação de informação e de modelagem de dados.

**Palavras-chaves:** Aprendizado de Máquina; Teoria da Informação; Sistemas Imunológicos Artificiais; Análise de Componentes Independentes; Corpos Finitos; Inversão Não-Linear; Sistemas de Wiener; Sistemas de Hammerstein.

# Abstract

The main theme of this thesis is the development of *Information Theoretic Learning* (ITL) algorithms. The ITL paradigm proposes the adoption of training criteria based on information measures such as entropy and mutual information, instead of traditional criteria based on second order statistics.

The problems of blind inversion of Wiener systems and blind separation of signals over Galois fields are the tasks over which these tools are applied. Such problems present key aspects that establish a demand for higher order statistics, hence we present several contributions that are based on ITL criteria and employ immune-inspired algorithms (or heuristic-based methods) to perform the adaptation of the parameters of each related model.

The developed proposals open the perspective of future applications in genomic, coding theory and geophysics. Furthermore, we expect that the presented results support a wider understanding of ITL, in order to comprise more general strategies of search, information estimation and data modeling.

**Keywords:** Information Theoretic Learning; Artificial Immune Systems; Independent Component Analysis; Finite Fields; Nonlinear Inversion; Wiener Systems; Hammerstein Systems.

# Sumário

<b>Introdução</b> . . . . .	<b>1</b>
<b>I Fundamentos</b>	<b>5</b>
<b>1 Aprendizado de Máquina Baseado na Teoria da Informação: Origem e Fundamentos</b> . . . . .	<b>7</b>
1.1 Teoria da Informação . . . . .	7
1.1.1 Fontes Discretas . . . . .	8
1.1.2 Fontes Contínuas . . . . .	11
1.1.3 A Entropia de Rényi . . . . .	12
1.2 Aprendizado de Máquina Baseado na Teoria da Informação . . . . .	14
1.2.1 O Estimador de Entropia Quadrática de Rényi . . . . .	17
1.2.2 Aplicações . . . . .	19
1.3 Uma Visão Ampliada de ITL . . . . .	20
1.4 Conclusões . . . . .	22
<b>2 Separação de Sinais em Corpos Finitos</b> . . . . .	<b>23</b>
2.1 Análise de Componentes Independentes . . . . .	23
2.2 Corpos Finitos . . . . .	24
2.2.1 Probabilidade e Estatística em Corpos Finitos . . . . .	26
2.3 ICA em Corpos Finitos . . . . .	27
2.3.1 Trabalhos Desenvolvidos . . . . .	28
2.4 Conclusões . . . . .	31
<b>3 Inversão Cega de Sistemas de Wiener</b> . . . . .	<b>33</b>
3.1 Inversão de Sistemas . . . . .	33
3.1.1 Inversão Cega de Sistemas de Wiener por Sistemas de Hammerstein . . . . .	34
3.2 Trabalhos Desenvolvidos . . . . .	35
3.2.1 A Metodologia de Taleb, Sole-Casals e Jutten . . . . .	37
3.2.2 A Metodologia de Rojas, Sole-Casals e Puntonet . . . . .	39
3.3 Conclusões . . . . .	40
<b>4 Computação Imuno-Inspirada</b> . . . . .	<b>43</b>
4.1 Princípios Básicos do Sistema Imunológico Natural . . . . .	43
4.1.1 Seleção Clonal . . . . .	45
4.1.2 Rede Imunológica . . . . .	45
4.2 Sistemas Imunológicos Artificiais . . . . .	46

4.2.1	O Algoritmo CLONALG . . . . .	47
4.2.2	O Algoritmo cob-aiNet . . . . .	49
4.2.2.1	Representação e Métricas de Afinidade . . . . .	50
4.2.2.2	O Modelo de Concentração e Supressão . . . . .	51
4.2.2.3	Clonagem e Mutação . . . . .	52
4.2.2.4	Seleção e Inserção de Novas Células . . . . .	53
4.3	Conclusões . . . . .	54
<b>II</b>	<b>Contribuições</b>	<b>55</b>
<b>5</b>	<b>Contribuições à Separação de Sinais em Corpos Finitos . . . . .</b>	<b>57</b>
5.1	Otimização Baseada no CLONALG e no Critério de Mínima Informação Mútua	57
5.1.1	Simulações . . . . .	59
5.1.2	Considerações Sobre a Proposta . . . . .	61
5.2	Otimização pela cob-aiNet[C] e o Critério de Mínima Entropia . . . . .	61
5.2.1	Simulações . . . . .	66
5.2.1.1	Cenário 1: $P$ é Variável . . . . .	66
5.2.1.2	Cenário 2: Comparativo com os Algoritmos AMERICA e MEXICO . . . . .	68
5.2.1.3	Cenário 3: Sensibilidade à Distância de Hamming . . . . .	72
5.2.1.4	Cenário 4: Análise Empírica de Estabilidade . . . . .	73
5.2.2	Considerações sobre a Proposta . . . . .	75
5.3	O Algoritmo MEXICO Modificado . . . . .	75
5.3.1	O Algoritmo MEXICO . . . . .	76
5.3.2	Modificações do Algoritmo MEXICO . . . . .	77
5.3.3	Simulações . . . . .	79
5.3.3.1	Cenário 1: Desempenho Geral da Nova Proposta . . . . .	80
5.3.3.2	Cenário 2: Comparativo . . . . .	80
5.3.4	Considerações sobre a Proposta . . . . .	85
5.4	Misturas Convolutivas em Corpos Finitos . . . . .	85
5.4.1	Proposta de Modelo . . . . .	85
5.4.2	Algoritmo de Extração / Deflação Baseado na Entropia . . . . .	87
5.4.3	Condições <i>Zero-Forcing</i> para Separação . . . . .	89
5.4.4	Simulações . . . . .	92
5.4.5	Considerações sobre a Proposta . . . . .	93
5.5	Conclusões . . . . .	96
<b>6</b>	<b>Contribuições à Inversão Cega de Sistemas de Wiener . . . . .</b>	<b>97</b>

6.1	Um <i>Framework</i> Bio-Inspirado Baseado no Critério da Informação Mútua . . .	97
6.1.1	Função Objetivo: Definição e Estimadores . . . . .	98
6.1.1.1	Estimador de IM Baseado em Histograma . . . . .	99
6.1.1.2	Estimador de Informação Mútua Quadrática de Cauchy-Schwarz	100
6.1.2	Método de Busca e Componentes do Modelo . . . . .	101
6.2	Um Experimento Preliminar: Inversão Linear Via Estimação da Informação Mútua . . . . .	103
6.2.1	Canal de Fase Mínima . . . . .	104
6.2.2	Canal de Fase Não-Mínima . . . . .	108
6.3	Análise Experimental do <i>Framework</i> . . . . .	109
6.3.1	Experimento com uma Segunda Não-Linearidade . . . . .	117
6.4	Conclusões . . . . .	120
	<b>Conclusão . . . . .</b>	<b>121</b>
	<b>Referências . . . . .</b>	<b>125</b>



*Dedico esta tese à minha querida Rafinha,  
aos meus queridos pais, Alberto e Marta,  
e ao meu querido irmão, Rodrigo.*



# Agradecimentos

Como outros empreendimentos ao longo da vida, o esforço pessoal despendido neste trabalho de doutorado é diretamente ligado ao apoio que tive de várias pessoas especiais, as quais não posso deixar de mencionar.

Primeiramente, agradeço ao Romis, por ter sido um extraordinário orientador e parceiro. Sua amizade e gentileza comigo, além da incondicional disponibilidade e paixão pelo ofício que exerce, aspectos estes que pude testemunhar ao longo de todo o doutorado, compõem o melhor exemplo de professor que tenho para a minha carreira.

Agradeço ao meu co-orientador, Jugurta, pela grata experiência de ter recebido seus empenhados e minuciosos conselhos, não obstante a grande distância física que havia entre nós. De uma certa maneira, foi como se estivéssemos convivendo no mesmo lugar.

Ao meu orientador do mestrado, professor Mario Jino, pela nítida contribuição ao início da minha carreira como pesquisador e pelo gentil e atencioso auxílio na minha transição para o doutorado, com um novo projeto e um novo orientador.

Aos membros do DSPCom, eu agradeço enormemente a calorosa acolhida e parceria em todos estes anos, que resultaram não só em valiosas colaborações científicas, mas também em agradáveis momentos de lazer, interessantes debates e amizades.

Aos membros do LBiC — especialmente, os professores Guilherme Coelho e Fernando Von Zuben — e aos colegas do Capítulo de Inteligência Computacional do IEEE - Unicamp, agradeço pela parceria nos trabalhos resultantes deste doutorado e pelas profícuas discussões e troca de conhecimentos.

Aos meus amigos, pelo carinho, paciência, alegria e paz que me trazem, principalmente nos momentos mais aflitivos e difíceis da vida. Aos amigos da *Unicamp Swimming Society Reloaded*, mais uma vez obrigado pela companhia e alegria, dentro ou fora da piscina.

Agradeço à minha família, em especial aos meus pais e ao meu irmão, por todo o amor, apoio e carinho que tive durante esta importante fase. Cada passo dado em minha vida inicia-se no amor de vocês.

À minha querida Rafinha, obrigado pelo privilégio de sua tão amorosa companhia. Obrigado pela paciência nos momentos ruins, pela cumplicidade, carinho e alegria que me traz, a cada dia que vivemos juntos.

E obrigado a Deus, pela vida plena e feliz que tenho tido a possibilidade de realizar.



*“Combati o bom combate, completei a corrida, guardei a fé.”*  
*(2ª carta de São Paulo a Timóteo, 4:7)*



# Lista de ilustrações

Figura 1 – Esquema geral do problema de filtragem. . . . .	15
Figura 2 – Modelo de mistura linear e instantânea do problema de BSS. . . . .	24
Figura 3 – Modelagem inversa de sistemas. . . . .	33
Figura 4 – O sistema de Wiener. . . . .	35
Figura 5 – O sistema de Hammerstein. . . . .	35
Figura 6 – Ilustração da variação sigmoideal espelhada de $\beta$ conforme as iterações da cob-aiNet. Este gráfico foi extraído de (COELHO; VON ZUBEN, 2010) com os parâmetros $\beta^i = 3.0$ , $\beta^f = 0.001$ e $maxIT = 1000$ . . . . .	53
Figura 7 – Desempenho médio dos algoritmos de separação para o cenário com duas fontes. . . . .	60
Figura 8 – Desempenho médio dos algoritmos de separação para o cenário com quatro fontes. . . . .	61
Figura 9 – Desempenho médio dos algoritmos de separação para o cenário com oito fontes. . . . .	62
Figura 10 – Desempenho do algoritmo cob-aiNet[C] com $N = 8$ . . . . .	67
Figura 11 – Comparação da cob-aiNet[C] com outras técnicas: acima, caso com $N = 6$ , $P = 5$ , abaixo, o caso com $N = 12$ , $P = 3$ . . . . .	69
Figura 12 – Desempenho do algoritmo cob-aiNet[C] com $N = 14$ , $P = 3$ . . . . .	70
Figura 13 – Número de avaliações da entropia para quantidades distintas de fontes, entre a cob-aiNet[C] e AMERICA, com $P = 3$ . . . . .	72
Figura 14 – Desempenho da cob-aiNet[C] em casos particulares de misturas: acima, caso com $N = 5$ , $P = 2$ , $p(0) = 0.2$ , $p(1) = 0.8$ ; abaixo, caso com $N = 5$ , $P = 2$ e perfil de probabilidade aleatório. . . . .	74
Figura 15 – Desempenho do algoritmo MEXICO-m: acima, sinais em $GF(3)$ ; abaixo, sinais em $GF(5)$ . . . . .	81
Figura 16 – Taxa de separação total para sinais em $GF(2)$ (acima) e $GF(3)$ (abaixo). . . . .	84
Figura 17 – O modelo de mistura convolutiva em corpos finitos e da estrutura de extração, no caso de um sistema MIMO $2 \times 2$ . . . . .	86
Figura 18 – Estrutura para realizar a deflação no problema de misturas convolutivas. . . . .	89
Figura 19 – Taxa de separação do algoritmo de Extração/Deflação em BSS convolutivo. Canais com 2 coeficientes. . . . .	94
Figura 20 – Taxa de separação do algoritmo de Extração/Deflação em BSS convolutivo. Canais com 3 coeficientes. . . . .	94

Figura 21 – Taxa de falha ao extrair a primeira fonte, no espaço das PMFs de $s_1(n)$ e $s_2(n)$ . . . . .	95
Figura 22 – Proposta de <i>framework</i> para inversão cega de sistemas de Wiener. . . . .	98
Figura 23 – Esquema de codificação dos indivíduos para o algoritmo de inversão dos sistemas Wiener-Hammerstein, cada coeficiente é um valor real. . . . .	103
Figura 24 – IM com relação aos parâmetros filtro FIR - $H(z) = 1 + 0.5z^{-1}$ . . . . .	105
Figura 25 – MSE entre $y(n)$ e $s(n - d)$ para o melhor $d$ , com relação aos parâmetros do filtro FIR - $H(z) = 1 + 0.5z^{-1}$ . . . . .	105
Figura 26 – Histograma do MSE para o filtro FIR MMI - $H(z) = 1 + 0.5z^{-1}$ . . . . .	106
Figura 27 – Histograma da IM para o filtro FIR MMI - $H(z) = 1 + 0.5z^{-1}$ . . . . .	106
Figura 28 – IM com relação aos parâmetros filtro IIR - $H(z) = 1 + 0.5z^{-1}$ . . . . .	107
Figura 29 – MSE entre $y(n)$ e $s(n - d)$ para o melhor $d$ , com relação aos parâmetros do filtro IIR - $H(z) = 1 + 0.5z^{-1}$ . . . . .	107
Figura 30 – Histograma do MSE para o filtro IIR MMI - $H(z) = 1 + 0.5z^{-1}$ . . . . .	108
Figura 31 – Histograma da IM para o filtro IIR MMI - $H(z) = 1 + 0.5z^{-1}$ . . . . .	108
Figura 32 – IM com relação aos parâmetros filtro FIR - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	110
Figura 33 – MSE entre $y(n)$ e $s(n - d)$ para o melhor $d$ , com relação aos parâmetros do filtro FIR - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	110
Figura 34 – IM com relação aos parâmetros filtro IIR - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	111
Figura 35 – MSE entre $y(n)$ e $s(n - d)$ para o melhor $d$ , com relação aos parâmetros do filtro IIR - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	111
Figura 36 – Histograma do MSE para o filtro FIR MMI - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	112
Figura 37 – Histograma da IM para o filtro FIR MMI - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	112
Figura 38 – Histograma do MSE para o filtro IIR MMI - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	112
Figura 39 – Histograma da IM para o filtro IIR MMI - $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	112
Figura 40 – Experimento com o sinal original uniforme processado pelo canal $H(z) = 1 + 0.5z^{-1}$ e $f(v) = \text{sign}(v)\sqrt[3]{ v }$ , estimador QMI como função custo. A partir do topo à esquerda, em sentido horário: $g \circ f(\cdot)$ ; correção não-linear $g(\cdot)$ ; histograma de $y(n)$ ; $s(n)$ , $y(n)$ e sinal de erro. . . . .	115
Figura 41 – Histogramas de $\mathcal{P}$ para o experimento com o canal de fase não-mínima, $s(n)$ uniforme e filtro FIR. . . . .	116
Figura 42 – Desvio padrão da estimativa de IM entre duas VAs independentes. . . . .	117
Figura 43 – Experimento com o sinal original uniforme processado pelo canal $H(z) = 1 + 0.5z^{-1}$ e $f(v) = \tanh(3v)$ , estimador QMI como função custo. A partir do topo à esquerda, em sentido horário: $g \circ f(\cdot)$ ; correção não-linear $g(\cdot)$ ; histograma de $y(n)$ ; $s(n)$ , $y(n)$ e sinal de erro. . . . .	119

# Lista de tabelas

Tabela 1	– Parâmetros gerais para todos os experimentos com a cob-aiNet[C]. . . . .	66
Tabela 2	– Parâmetros específicos para as simulações com $N = 8$ , $N = 6$ , $GF(5)$ e $N = 12$ , $GF(3)$ . . . . .	67
Tabela 3	– Parâmetros específicos para a simulação com $N = 14$ , $GF(3)$ . . . . .	71
Tabela 4	– Média amostral e desvio padrão da medida NAT entre as versões do algoritmo MEXICO. . . . .	82
Tabela 5	– Média amostral e desvio padrão do número de fontes recuperadas entre as versões do algoritmo MEXICO, $GF(2)$ . . . . .	83
Tabela 6	– Média amostral e desvio padrão do número de fontes recuperadas entre as versões do algoritmo MEXICO, $GF(3)$ . . . . .	83
Tabela 7	– Filtros FIR e IIR obtidos pelo critério MMI para inverter o canal $H(z) = 1 + 0.5z^{-1}$ . . . . .	104
Tabela 8	– Filtros FIR e IIR obtidos pelo critério MMI para inverter o canal $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . . . . .	108
Tabela 9	– Resultados do algoritmo imuno-inspirado no problema de Wiener - Hammerstein — fase mínima. Nas linhas superiores, média e desvio padrão de $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado. . . . .	113
Tabela 10	– Resultados do algoritmo imuno-inspirado no problema de Wiener - Hammerstein — fase não-mínima. Nas linhas superiores, média e desvio padrão de $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado. . . . .	114
Tabela 11	– Análise comparativa dos métodos — fase mínima, $f(v) = \tanh(3v)$ . Nas linhas superiores, média e desvio padrão de $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado. . . . .	118
Tabela 12	– Comparativo das estratégias de busca com o estimador QMI — fase mínima, $f(v) = \tanh(3v)$ . Nas linhas superiores, média e desvio padrão de $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado. . . . .	120



# Lista de Acrônimos e Abreviações

- BSS Do inglês *Blind Source Separation*
- CIP Do inglês *Cross-Information Potential*
- EBD Do inglês *Entropy Based Demixing*
- EBED Do inglês *Entropy Based Extraction/Deflation*
- EPBED Do inglês *Extreme Probability Based Extraction/Deflation*
- EQM Erro Quadrático Médio
- ESO Estatísticas de Segunda Ordem
- FIR Do inglês *Finite Impulse Response*
- GA Do inglês *Genetic Algorithm*
- GFFT Do inglês *Galois Field Fourier Transform*
- ICA Do inglês *Independent Component Analysis*
- iid Independente e identicamente distribuído
- IIR Do inglês *Infinite Impulse Response*
- IM Informação Mútua
- IP Do inglês *Information Potential*
- ITL Do inglês *Information Theoretic Learning*
- LTI Do inglês *Linear and Time-Invariant*
- MEE Minimização da Entropia do Erro
- MISO Do inglês *Multiple Input Single Output*
- ML Do inglês *Maximum Likelihood*
- MLP Do inglês *Multilayer Perceptron*
- MMI Do inglês *Minimal Mutual Information*

MMSE Do inglês *Minimum Mean Square Error*

NAT Número de Avaliações de Troca

PCA Do inglês *Principal Component Analysis*

PDF Do inglês *Probability Density Function*

PMF Do inglês *Probability Mass Function*

PNL Do inglês *Post-Nonlinear*

QMI Do inglês *Quadratic Mutual Information*

SIA Sistema Imunológico Artificial

TI Teoria da Informação

VA Variável Aleatória

ZF Do inglês *Zero-Forcing*

# Introdução

O desenvolvimento das disciplinas de Aprendizado de Máquina, Inteligência Computacional e Processamento Adaptativo de Sinais vem ocorrendo, ao longo dos últimos anos, de forma alinhada aos desafios impostos pelo considerável aumento da disponibilidade e complexidade de dados associados às mais diversas esferas do conhecimento.

O surgimento de novos problemas nesse contexto, além da reformulação de problemas já existentes em um âmbito mais elaborado, tem causado a necessidade de estudo de modelos e estratégias de solução que possivelmente abandonem premissas tradicionalmente utilizadas, como a de que o modelo seja um sistema cuja relação entre a entrada e a saída é *linear* e a de que os sinais envolvidos estejam sujeitos a um ruído de distribuição *gaussiana*. Por isso, medidas estatísticas que estão presentes em abordagens clássicas passam a ser questionadas quanto à sua completude.

Neste cenário, emerge a disciplina de Aprendizado de Máquina Baseado na Teoria da Informação (em inglês, *Information Theoretic Learning*, ITL), que é o tema central desta tese de doutorado. Nela, abordaremos o desenvolvimento e aplicação de algoritmos de ITL para a solução de dois problemas não-supervisionados — separação de sinais em corpos finitos e inversão de sistemas de Wiener. O paradigma de ITL propõe o uso de critérios de otimização baseados em medidas como entropia e informação mútua, em substituição aos tradicionais critérios baseados em estatísticas de segunda ordem.

Dadas as características dos problemas analisados, propõe-se uma série de contribuições que, além do objetivo direto de melhoria de soluções em termos de qualidade e custo computacional, visam dar subsídios para se ampliar o escopo de ITL. Esse novo domínio abrangeria *(i)* estratégias de busca não necessariamente baseadas em gradiente, como os algoritmos imunológicos; *(ii)* funções objetivo não restritas às definições de Rényi; e *(iii)* dados de natureza discreta ou simbólica.

## Objetivos e Organização da Tese

Esta tese se organiza em duas partes: a primeira consiste na fundamentação teórica do paradigma de ITL, dos dois problemas que são abordados e da área de computação imuno-inspirada; a segunda parte apresenta as contribuições realizadas junto aos problemas, calcadas nos conceitos e definições apresentadas na parte anterior. Detalhemos a seguir o que cada capítulo compreende.

O Capítulo 1 introduz os conceitos do Aprendizado de Máquina Baseado na Teoria da Informação, partindo de aspectos primordiais da Teoria da Informação, para daí chegar à sua conexão com o processamento adaptativo de sinais e o aprendizado de máquina.

O Capítulo 2 apresenta o problema de separação cega de sinais em corpos finitos, associado à estratégia de solução pela análise de componentes independentes. São discutidos os conceitos referentes a álgebra e probabilidade para sinais definidos em corpos de Galois, além dos principais algoritmos propostos para o problema.

O Capítulo 3 analisa o segundo problema tratado na tese, o da inversão cega de sistemas de Wiener. O modelo não-linear do problema é descrito, assim como a estratégia de emprego do sistema de Hammerstein para a sua inversão e os principais algoritmos elaborados para determinar os parâmetros desse sistema inversor.

O Capítulo 4 encerra a parte de fundamentos da tese apresentando conceitos básicos sobre Sistemas Imunológicos Artificiais, o que primeiramente engloba a explicação dos mecanismos primordiais de operação do sistema imunológico e que são a base das duas propostas de algoritmos imuno-inspirados empregados na tese, descritas em seguida: o CLONALG e a cob-aiNet.

A parte de contribuições da tese inicia-se com o Capítulo 5, que apresenta as propostas para a separação de sinais em corpos de Galois. Elas compreendem dois novos algoritmos que empregam computação imuno-inspirada em critérios baseados na teoria da informação; uma versão alternativa de um algoritmo já estabelecido na área e a extensão do problema para o modelo de misturas convolutivas.

O Capítulo 6 traz as contribuições relativas ao problema de inversão de sistemas de Wiener, que consistem em um novo *framework* para solução do problema com o uso de um critério de mínima informação mútua, cujos parâmetros do modelo são otimizados pelo algoritmo CLONALG. Adicionalmente aos experimentos numéricos de validação da nova técnica, o capítulo apresenta, antes, uma análise empírica do critério não-supervisionado frente ao tradicional critério do mínimo erro quadrático médio.

Apresentadas as contribuições, a tese se encerra com o capítulo de conclusão, contendo as considerações finais sobre o trabalho como um todo e as perspectivas futuras.

## Trabalhos Publicados Durante o Doutorado

A seguir, apresentamos uma relação contendo as publicações feitas ao longo do doutorado.

- 
- Artigos completos publicados em ou submetidos a periódicos:
    1. Montalvão, J., Attux, R., & **Silva, D. G.** (2012). Simple Entropy Estimator for Small Datasets. *Electronics Letters*, 48(17), 1059–1061.
    2. **Silva, D. G.**, Nadalin, E. Z., Montalvão, J., & Attux, R. (2013). The Modified MEXICO for ICA Over Finite Fields. *Signal Processing*, 93(9), 2525–2528.
    3. Montalvão, J., Attux, R., & **Silva, D. G.** (2013). A pragmatic entropy and differential entropy estimator for small datasets. **Submetido ao** *Journal of Communication and Information Systems*.
    4. **Silva, D. G.**, Nadalin, E. Z., Coelho, G. P., Duarte, L. T., Suyama, R., Attux, R., Von Zuben, F. J., & Montalvão, J. (2014). A Michigan-like Immune-Inspired Framework for Performing Independent Component Analysis over Galois Fields of Prime Order. *Signal Processing*, 96(0), 153-163.
  
  - Trabalhos completos publicados ou aceitos para publicação em anais de congressos:
    1. **Silva, D. G.**, Attux, R., Nadalin, E. Z., Duarte, L. T., & Suyama, R. (2011). An Immune-Inspired Information-Theoretic Approach to the Problem of ICA Over a Galois Field. Em 2011 IEEE Information Theory Workshop (pp. 618–622). IEEE.
    2. **Silva, D. G.**, Nadalin, E. Z., Attux, R., & Montalvão, J. (2012). A Modified Version of the MEXICO Algorithm for Performing ICA Over Galois Fields. Em 2012 IEEE International Workshop on Machine Learning for Signal Processing (pp. 1–6). IEEE.
    3. Boccato, L., **Silva, D. G.**, Fantinato, D., Nose-Filho, K., Ferrari, R., Attux, R., Neves, A. O., Montalvão, J., & Romano, J. M. T. (2013). Error Entropy Criterion in Echo State Network Training. Em European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (pp. 35–40).
    4. Fantinato, D. G., **Silva, D. G.**, Nadalin, E. Z., Attux, R., Romano, J. M. T., Neves, A. O., & Montalvão, J. (2013). Blind Separation of Convolutional Mixtures Over Galois Fields. Em 2013 IEEE International Workshop on Machine Learning for Signal Processing (pp. 1-6). IEEE.
    5. Montalvão, J., **Silva, D. G.**, & Attux, R. (2013). Symbolic Dynamic Analysis Through Coincidence Detection - An Oversimplified Method. Em XI Simpósio Brasileiro de Automação Inteligente - SBAI-DINCON 2013.
  
  - Resumos expandidos publicados em anais de congressos:

1. **Silva, D. G.**, & Attux, R. (2010). Aprendizado Baseado em Teoria da Informação: Fundamentos e Perspectivas. Em Primeiro Simpósio de Processamento de Sinais da Unicamp, 2010, Campinas.
2. **Silva, D. G.**, & Attux, R. (2011). Aprendizado Baseado na Teoria da Informação: Fundamentos e Perspectivas. Em Quarto Encontro dos Alunos e Docentes do DCA - FEEC, 2011, Campinas.
3. Soriano, D. C., Olivatto, V. B., **Silva, D. G.**, & Attux, R. (2011). Algumas Reflexões sobre o Poder Computacional de Paradigmas Analógicos - Parte 1: Fundamentos. Em Quarto Encontro dos Alunos e Docentes do DCA - FEEC, 2011, Campinas.
4. Soriano, D. C., Olivatto, V. B., **Silva, D. G.**, & Attux, R. (2011). Algumas Reflexões sobre o Poder Computacional de Paradigmas Analógicos - Parte 2: Aspectos Formais e Especulações. Em Quarto Encontro dos Alunos e Docentes do DCA - FEEC, 2011, Campinas.
5. **Silva, D. G.**, Attux, R., Nadalin, E. Z., Coelho, G. P., Von Zuben, F. J., Duarte, L. T., & Suyama, R. (2011). ICA Over Finite Fields: an Immune-Inspired and Information-Theoretic Approach. Em Segundo Simpósio de Processamento de Sinais da Unicamp, 2011, Campinas.
6. Fantinato, D. G., **Silva, D. G.**, Attux, R., Ferrari, R., Duarte, L. T., Suyama, R., Neves, A. O., Montalvão, J., & Romano, J. M. T. (2012). Optimal Time-Series Prediction over Galois Fields. Em Terceiro Simpósio de Processamento de Sinais da Unicamp, 2012, Campinas.

Parte I

Fundamentos



# 1 Aprendizado de Máquina Baseado na Teoria da Informação: Origem e Fundamentos

Este capítulo apresenta o tema de fundo desta tese, que é o Aprendizado de Máquina Baseado na Teoria da Informação. Antes de discuti-lo, porém, faz-se uma apresentação dos aspectos fundamentais e históricos da Teoria da Informação e a definição de suas principais medidas para dados discretos e contínuos. Então, finalmente chega-se à sua conexão com o processamento adaptativo de sinais e o aprendizado de máquina, sendo apresentada a principal grandeza para construção de critérios de filtragem, a entropia quadrática de Rényi, além de exemplos de aplicação e argumentos que visam ampliar o escopo desta recente área de pesquisa.

## 1.1 Teoria da Informação

A humanidade testemunhou, ao longo dos últimos dois séculos, um desenvolvimento gradual e verdadeiramente revolucionário das tecnologias de transmissão, armazenamento e processamento de dados. Este processo gerou uma série de disciplinas que podem ser vistas, de forma harmoniosa, como sendo suportadas por um conceito único: o de *informação* (GLEICK, 2013).

Apesar das relevantes contribuições dadas por Nyquist (1924) e Hartley (1928), é consenso que a área de pesquisa denominada Teoria da Informação (TI) toma sua forma concreta com o esplendoroso trabalho de uma das personalidades científicas mais marcantes do século XX: Claude Elwood Shannon. O próprio título do trabalho já indica a grandeza de seu escopo — “*A Mathematical Theory of Communication*” (SHANNON, 1948) — e, no entanto, após a última página, o leitor só pode ter a impressão de que o texto superou as mais altas expectativas.

Embora a aplicação da Teoria da Informação ao Aprendizado de Máquina esteja concentrada nos seus conceitos fundamentais, que, em sua maioria, surgem neste trabalho de 1948, achamos relevante descrever a seguir, além deles, alguns aspectos adicionais que são essenciais do ponto de vista histórico e para a compreensão desta área do conhecimento como um todo.

### 1.1.1 Fontes Discretas

O artigo de Shannon lida com fontes de informação de naturezas discreta e contínua, e ele também estabelece uma divisão entre processos de transmissão com ou sem ruído. Na primeira parte do trabalho, que é dedicado à análise do caso discreto e sem ruído, Shannon apresenta uma medida fundamental  $H(\cdot)$ :

$$H(X) = -K \sum_{x \in \mathcal{X}} p_X(x) \log p_X(x), \quad (1.1)$$

em que  $p_X(x)$  é a função massa de probabilidade (PMF, do inglês *Probability Mass Function*) da variável aleatória (VA)  $X$ . Considera-se que  $0 \log 0 = 0$ .

Após indicar na Equação 1.1 que, sem perda de generalidade, pode-se fazer  $K = 1$ , Shannon diz que os valores desta medida são relevantes como “medidas de informação, escolha e incerteza” (SHANNON, 1948). Devidamente estabelece-se, assim, uma conexão com a Mecânica Estatística, e a medida  $H(\cdot)$  é batizada de entropia, cuja unidade é o bit se for considerado o logaritmo de base 2. Feynman (1998) relata que a escolha deste nome por Shannon foi uma sugestão de John von Neumann, que supostamente argumentou que o uso do termo “entropia” lhe daria “...uma grande vantagem em debates, porque ninguém sabe o que é entropia, de qualquer maneira.”

A expressão proposta para  $H(\cdot)$  é justificada em termos de uma série de propriedades que uma definição de entropia deveria possuir, como continuidade sobre os valores de probabilidade e crescimento monotônico com respeito ao número de eventos no caso uniforme (SHANNON, 1948). A seguir, apresentamos algumas propriedades importantes de  $H(X)$ , que serão úteis para considerações futuras:

1.  $H(X) = 0$  se e somente se existe um único evento com probabilidade não nula (i.e., unitária). Isto significa que a entropia é nula quando não há incerteza a respeito do resultado de um experimento aleatório.
2. Dado um número  $n$  de possíveis resultados, a entropia é máxima se todas as probabilidades equivalem a  $1/n$ . Isto significa que o caso de maior entropia ocorre com uma distribuição uniforme, o que é satisfatório do ponto de vista intuitivo, dado que as probabilidades não fornecem nenhuma “dica” para alguém que esteja interessado em prever o resultado — a incerteza é máxima.
3. A entropia conjunta de duas variáveis aleatórias,  $X$  e  $Y$ , é definida por

$$H(X; Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log p_{XY}(x, y), \quad (1.2)$$

onde  $p_{XY}(x, y)$  é a PMF conjunta referente a  $X$  e  $Y$ . Esta definição origina, por sua vez, uma importante desigualdade:

$$H(X; Y) \leq H(X) + H(Y), \quad (1.3)$$

cujas situação de igualdade é possível somente se  $X$  e  $Y$  são estatisticamente independentes, i.e., quando

$$p_{XY}(x, y) = p_X(x)p_Y(y). \quad (1.4)$$

Note que a desigualdade dada pela Equação 1.3 é bastante intuitiva: sempre que existe algum grau de dependência entre variáveis, a incerteza associada ao conhecimento conjunto delas será menor que a soma das incertezas dadas por uma análise separada. Quando as variáveis são independentes, temos que qualquer uma delas é inútil na tarefa de reduzir a quantidade de incerteza associada à outra: portanto, a igualdade ocorre.

4. A entropia de uma distribuição condicional, chamada de entropia condicional, é dada por:

$$H(Y|X) = - \sum_{x \in \mathcal{X}} p_X(x) \sum_{y \in \mathcal{Y}} p_{Y|X=x}(y) \log p_{Y|X=x}(y). \quad (1.5)$$

A partir desta definição, pode-se demonstrar que

$$H(X; Y) = H(X) + H(Y|X), \quad (1.6)$$

o que significa que a incerteza conjunta de  $X$  e  $Y$  pode ser vista como a quantidade de incerteza vinculada a  $X$  mais a incerteza associada a  $Y$  quando  $X$  é conhecido (repare que a dependência tem um papel fundamental). A Equação 1.5 também pode ser reescrita em termos de  $H(Y)$  e  $H(X|Y)$ .

5. A partir da Equação 1.3 e da Equação 1.6, é possível obter a expressão

$$H(Y) \geq H(Y|X), \quad (1.7)$$

que revela que a incerteza de uma VA nunca é aumentada pelo conhecimento de uma outra variável. O caso extremo — de independência — leva à igualdade.

Após definir o conceito fundamental de sequência típica, Shannon prova um teorema que estabelece, inexoravelmente, que a entropia da fonte de informação é o limite alcançável de eficiência para qualquer processo de codificação (note que, no caso sem ruído, o objetivo é de que haja compressão sem perdas, i.e., remoção de redundância).

Na sequência, ele analisa o caso de canais discretos com a presença de ruído. O princípio básico deste cenário é avaliar a possibilidade de transmitir dados mesmo sabendo que há

a chance de equívocos no processo de reconstrução, no receptor. Naturalmente, descarta-se a transmissão dos dados no formato “cru”, e a estratégia reside em utilizar algum método de codificação, mas determinar qual é o código apropriado é uma tarefa longe de ser trivial (COVER; THOMAS, 2006).

Shannon, com o seu estilo lúcido e leve, argumenta que, para um canal com ruído, a “taxa de transmissão real” (SHANNON, 1948) é dada por  $H(X) - H(X|Y)$  — sendo  $X$  relacionado ao transmissor e  $Y$  ao sinal recebido. Esta quantidade é o que denominamos de informação mútua (IM), definida então como

$$I(X; Y) = H(X) - H(X|Y). \quad (1.8)$$

A sua linha de raciocínio é direta:  $H(X)$  representa a entropia da fonte e  $H(X|Y)$  é uma medida de equívoco, i.e., a ambiguidade média do sinal recebido. Note que a Equação 1.8 pode ser reescrita de duas formas:

$$I(X; Y) = H(Y) - H(Y|X), \quad (1.9)$$

e

$$I(X; Y) = H(X) + H(Y) - H(X; Y). \quad (1.10)$$

Shannon fornece em seu artigo interpretações interessantes de ambas as expressões. A Equação 1.9 indica a quantidade de informação recebida menos aquilo que é devido ao ruído, e a Equação 1.10 é “a soma das duas quantidades (informação transmitida e informação recebida) menos a entropia conjunta e, portanto, é o número de bits por segundo em comum” (SHANNON, 1948).

Podemos escrever  $I(X; Y)$  em termos de probabilidades

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \quad (1.11)$$

e, assim, relacioná-la com a divergência de Kullback-Leibler

$$D_{KL}(p; q) = \sum_u p(u) \log \frac{p(u)}{q(u)}, \quad (1.12)$$

em que  $p$  e  $q$  são duas PMFs e se considera que  $0 \log \frac{0}{0} = 0$ . Com estas duas expressões, temos a relação:

$$I(X; Y) = D_{KL}(p_{XY}(x, y); p_X(x)p_Y(y)). \quad (1.13)$$

A divergência de Kullback-Leibler, também chamada de entropia relativa, é uma medida de distância entre duas distribuições e, no caso da propriedade apresentada pela Equação 1.13, mostra que a informação mútua também pode ser interpretada como a entropia

relativa entre a distribuição conjunta e a distribuição dada pelo produto das distribuições marginais (COVER; THOMAS, 2006). Veja que a informação mútua se anula no caso em que as variáveis em análise são independentes, o que a credencia como uma medida direta dessa condição estatística.

Retornando ao artigo de Shannon, ele prossegue provando um resultado impressionante: se um valor  $C$  chamado de capacidade do canal não é excedido pela taxa de informação produzida na fonte, existe necessariamente um sistema de codificação capaz de resultar em uma transmissão com uma taxa de erro arbitrariamente pequena. À primeira vista, isto soa quase como inacreditável — um canal com ruído pode ser utilizado para enviar informação com um erro de reconstrução arbitrariamente pequeno sem que a taxa tenda a zero. Não pertence ao escopo desta tese discutir em detalhes o método elegante que foi utilizado para provar este resultado, mas é importante registrar que a capacidade de um dado canal é definida em termos da maximização da informação mútua entre as variáveis vinculadas ao transmissor e ao receptor:

$$C = \max_{p_X(x)} I(X, Y). \quad (1.14)$$

Esta maximização é feita com respeito à estrutura de probabilidades da fonte de informação.

### 1.1.2 Fontes Contínuas

Na sequência, o caso de fontes contínuas torna-se o foco do trabalho de Shannon. Um primeiro aspecto de suma importância, nesse sentido, é a extensão da definição de entropia para o caso de uma variável aleatória contínua com função densidade de probabilidade (PDF, do inglês *Probability Density Function*)  $f_X(x)$ :

$$h(X) = - \int f_X(x) \log f_X(x) dx. \quad (1.15)$$

Esta é uma escolha intuitiva que preserva muitas das propriedades válidas para o caso discreto, previamente estudado, mas existem também diferenças importantes. O aspecto mais relevante apontado por Shannon é que, em contraste com o caso discreto, para o qual a entropia corresponde a uma medida absoluta de incerteza, no caso contínuo, a definição leva, em geral, a resultados diferentes para sistemas de coordenadas diferentes. No entanto, a diferença entre entropias não é afetada por esta modificação potencial, o que significa que medidas como a capacidade de canal são imunes a este fato (SHANNON, 1948). Conforme expressa pela Equação 1.15, a entropia de uma variável contínua é também chamada de entropia diferencial (COVER; THOMAS, 2006), e seu valor é comumente expresso em nats, sendo usado o logaritmo natural.

As propriedades dadas pela Equação 1.6 e Equação 1.7 são diretamente extensíveis, assim como as definições dadas pela Equação 1.9 e Equação 1.10. Se considerarmos a definição da divergência de Kullback-Leibler para o caso contínuo, dada por

$$D_{KL}(f; g) = \int f(u) \log \frac{f(u)}{g(u)} du, \quad (1.16)$$

onde  $f$  e  $g$  são duas PDFs, também é possível preservar a propriedade dada pela Equação 1.13. Ademais, são fornecidas outras três propriedades importantes no artigo de Shannon:

1. A entropia diferencial não é limitada por baixo ou, em outras palavras, ela pode assumir valores negativos.
2. Sob a restrição de que uma VA contínua é limitada a um volume finito do espaço, a PDF com entropia máxima é a uniforme.
3. Se a matriz de covariância de uma VA contínua é fixada *a priori*, a PDF com entropia máxima é a gaussiana.

O restante do trabalho de Shannon corresponde a um excepcional esforço que estabeleceu uma série de resultados-chave para fontes e canais contínuos. Graças a esse artigo, a Teoria da Informação se tornou uma área própria da ciência, e uma quantidade enorme de estudos foram desenvolvidos a fim de compreender, consolidar e ampliar as contribuições iniciais dadas por Shannon. Como estes resultados estão além do escopo desta tese, não seguiremos mais o curso desse belo trabalho. Em vez disso, enfocaremos uma medida generalizada de informação que estendeu a definição original da Equação 1.1 e, vários anos depois, acabou viabilizando o surgimento das principais ferramentas de aprendizado de máquina baseadas na TI.

### 1.1.3 A Entropia de Rényi

Após Shannon, a Teoria da Informação tornou-se um novo campo científico e extrapolou o campo das comunicações, sendo utilizada, por exemplo, na física, na biologia, na estatística e em outras áreas da engenharia como processamento de sinais e aprendizado de máquina (PRINCIPE, 2010). Esse relativo sucesso passa pelo fato de que as medidas de entropia e informação mútua são descritores escalares que indicam, de uma forma não ambígua, a carga informacional da estrutura probabilística envolvida nos dados observados.

O emprego de TI nestes mais diversos contextos ocorreu em paralelo ao desenvolvimento de novas medidas que estendem as formas canônicas propostas por Shannon. Nesse sentido, Alfred Rényi, em meados dos anos 50, desenvolveu uma generalização matemática

da entropia, usualmente chamada de entropia de ordem  $\alpha$  de Rényi. Ela é expressa, no caso unidimensional e discreto como (RÉNYI, 1961)

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} p_X^\alpha(x), \quad (1.17)$$

e para variáveis contínuas

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \int f_X^\alpha(x) dx. \quad (1.18)$$

A intenção de Rényi era desenvolver uma medida generalizada de informação que possuísse a propriedade aditiva de sistemas estatisticamente independentes e que respeitasse os axiomas de probabilidade de Kolmogorov (LEON-GARCIA, 1994). Esta definição permaneceu praticamente ignorada na área de comunicações, onde a proposta pioneira de Shannon foi adotada com extremo sucesso, provavelmente porque possui propriedades matemáticas que levam a uma interpretação mais compreensível do fluxo de informação nestes sistemas. Porém, existem aplicações da entropia de Rényi em outras áreas como, por exemplo, teoria de códigos (CAMPBELL, 1965), mecânica quântica (BENGTSSON; ZYCZKOWSKI, 2008), sistemas dinâmicos caóticos (GRASSBERGER; PROCACCIA, 1983) e economia (HART, 1975).

O parâmetro  $\alpha$  da entropia de Rényi permite diversas medidas de incerteza para a mesma distribuição. Considerando o caso contínuo, dois cenários são importantes de se mencionar, nesse contexto: (i) quando  $\lim_{\alpha \rightarrow 1} h_\alpha(X) = h(X)$ , o que implica que a entropia de Shannon é um caso especial da definição de Rényi e (ii) quando  $\alpha = 2$ , em que temos a chamada entropia quadrática

$$h_2(X) = -\log \int f_X^2(x) dx = -\log E[f_X(x)], \quad (1.19)$$

sendo  $E[\cdot]$  o operador de *esperança* ou *valor médio* (LEON-GARCIA, 1994).

A entropia quadrática tem um papel importante no aprendizado de máquina baseado na TI porque ela origina uma família de estimadores que tem características interessantes (da perspectiva de aprendizado de máquina) como serem não-paramétricos, contínuos e computacionalmente simples de calcular. Embora seja possível obter estimadores análogos para qualquer valor de  $\alpha$ , o caso quadrático associado à estimação de densidade de probabilidade utilizando *kernels* gaussianos (PARZEN, 1962) permite a aplicação de um “truque” interessante para criar o estimador de entropia, que será demonstrado em mais detalhes na seção 1.2.

Rényi (1961) também propôs nos seus estudos uma medida generalizada de divergência em espaços de probabilidades, a divergência- $\alpha$  de Rényi:

$$D_\alpha(f; g) = \frac{1}{\alpha-1} \log \int f(x) \left( \frac{f(x)}{g(x)} \right)^{\alpha-1} dx, \quad (1.20)$$

onde  $f$  e  $g$  são duas PDFs. Analogamente à entropia, a divergência- $\alpha$  converge para a divergência de Kullback-Leibler quando  $\alpha \rightarrow 1$ . Além disso, de maneira direta, pode-se definir a informação mútua de ordem  $\alpha$

$$I_\alpha(X; Y) = \log \int \int f_{XY}(x, y) \left( \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} \right)^{\alpha-1} dx dy, \quad (1.21)$$

A divergência- $\alpha$ , assim como a divergência de Kullback-Leibler, é uma medida assimétrica. Entre várias propostas que contornam essa questão, uma delas é a divergência de Cauchy-Schwarz:

$$D_{CS}(f; g) = -\frac{1}{2} \log \frac{(\int f(x)g(x)dx)^2}{\int f^2(x)dx \int g^2(x)dx}. \quad (1.22)$$

Baseados nesta divergência, podemos obter uma medida alternativa de independência, a informação mútua quadrática de Cauchy-Schwarz (QMI, do inglês *Quadratic Mutual Information*) (PRINCIPE, 2010). A medida QMI pode ser aplicada no contexto de diversos problemas e será mais detalhada nos capítulos adiante.

## 1.2 Aprendizado de Máquina Baseado na Teoria da Informação

Os trabalhos pioneiros de Wiener (1958) e Kolmogorov (1962), pilares da disciplina de processamento estatístico de sinais, foram cruciais para que se estabelecesse a possibilidade de projeto de estruturas adaptativas de filtragem sob uma perspectiva probabilística. Duas premissas foram fundamentais para a consolidação dessa linha de pesquisa: o emprego de estruturas de adaptação lineares e de um critério para ajuste de parâmetros baseado em estatísticas de segunda ordem (ESO) (HAYKIN, 2001), como a esperança do erro quadrático, a variância e a correlação.

A Figura 1 ilustra o problema de filtragem adaptativa ótima, onde  $x(n)$  é o sinal original,  $y(n)$  é o sinal após o processo de filtragem,  $r(n)$  é um sinal que indica a resposta desejada do filtro (opcional) e  $e(n)$  é o sinal de erro (ou somente de saída) que é provido como entrada para o critério de ajuste dos parâmetros do filtro. Considerando as duas premissas já mencionadas, um exemplo tradicional de critério *supervisionado* para o problema é a minimização do Erro Quadrático Médio (EQM) entre  $r(n)$  e  $y(n)$ , com respeito ao conjunto de parâmetros  $\mathbf{w}$  do filtro:

$$J(\mathbf{w}) = E[(r(n) - y(n))^2]. \quad (1.23)$$

A razão para que seja comum o uso de um filtro cujo mapeamento  $y(n) = \mathcal{F}(\mathbf{w}, x(n))$  que o descreve é *linear* está no fato de que essas estruturas representam um compromisso

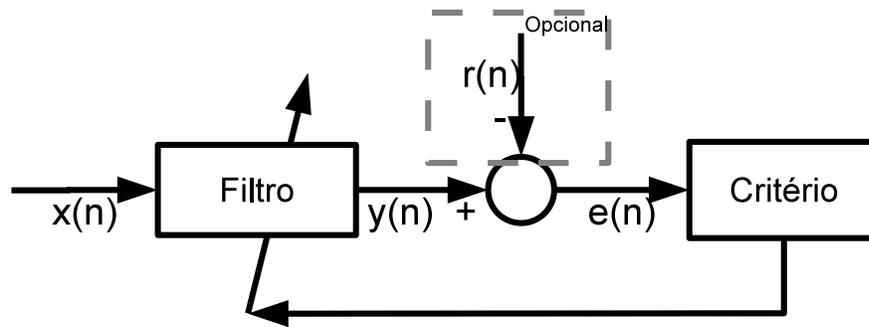


Figura 1 – Esquema geral do problema de filtragem.

adequado entre eficácia e tratabilidade matemática. Quando então se associa esse filtro a um critério baseado em ESO, se considerarmos que os dados disponíveis seguem uma distribuição gaussiana — uma hipótese que ocorre amiúde em situações nas quais é aplicável o teorema do limite central (LEON-GARCIA, 1994) —, o casamento entre filtro e critério de ajuste será ótima do ponto de vista, por exemplo, de uma solução de máxima verossimilhança (DUDA *et al.*, 2001).

Não suficiente, há outros benefícios no emprego de critérios baseados em ESO (ERDOGMUS; PRINCIPE, 2006; WANG; BOVIK, 2009): eles levam, em geral, (i) a funções custo relativamente simples no que se refere à manipulação matemática; as quais (ii) tipicamente engendram propriedades interessantes no contexto de otimização, como diferenciabilidade e simetria; e que (iii) se associam a uma enorme variedade de algoritmos adaptativos eficientes.

Em paralelo aos desenvolvimentos ocorridos no âmbito do tratamento mais extensivo da tarefa de modelagem de sinais aleatórios, emerge de maneira definitiva, a partir da década de 1950, a disciplina de aprendizado de máquina (ou, em inglês, *machine learning*), em cujo seio foram desenvolvidos os primeiros modelos de redes neurais (MCCULLOCH; PITTS, 1943), sistemas nebulosos (ZADEH, 1965) e algoritmos evolutivos (HOLLAND, 1962; FOGEL *et al.*, 1966; KOZA, 1992). A posterior conexão desses desenvolvimentos com a área de processamento de sinais levou à popularização do emprego de filtros não-lineares e também ampliou as possibilidades de projeto de dispositivos de tratamento da informação graças à adoção de meta-heurísticas bio-inspiradas capazes de lidar efetivamente com a questão de otimalidade local e com funções de difícil manipulação (ROMANO *et al.*, 2011). Pode-se dizer que, atualmente, ferramentas desse tipo vêm se tornando parte do que poderíamos denominar “estado da arte” em certos domínios desse campo de pesquisa (ADALI; HAYKIN, 2010).

Naturalmente, o progresso das técnicas de *machine learning* e processamento de si-

nais andou lado-a-lado com a crescente complexidade dos problemas a serem tratados, que frequentemente não se adequam às premissas clássicas de linearidade e/ou gaussianidade. Não é errado afirmar que houve um notável desenvolvimento dos recursos computacionais ao longo das últimas décadas e que ocorre atualmente uma “explosão” de informação e dados obtidos nos mais diferentes contextos práticos, o que sustenta o aparecimento de problemas mais complexos de tratamento e processamento de tais dados (HEY *et al.*, 2009). Tornam-se plausíveis, no contexto atual, dados que não obedecem à distribuição normal, que possuem *outliers* ou que são não-estacionários (PRINCIPE, 2010).

Portanto, um critério baseado em ESO pode não ser o mais adequado para ajustar os parâmetros do modelo durante o processo de aprendizado. Uma alternativa que vem se consolidando, nesse sentido, é o emprego de descritores baseados na teoria da informação. Como visto na seção 1.1, as medidas baseadas em TI como a entropia, informação mútua e divergência são intrinsecamente definidas de modo a contemplar toda a estrutura estatística dos dados (ao invés de um ou mais momentos) e podem ser utilizadas para se criar critérios de otimização que se ajustem melhor ao problema, obtendo soluções mais acuradas e robustas.

Embora sejam reconhecidas as iniciativas pioneiras, neste sentido, dos trabalhos de Wiggins (1978), Donoho (1981), Linsker (1988) e Bell e Sejnowski (1995), esta nova metodologia de fato ganhou impulso com os esforços do grupo de pesquisa liderado pelo Prof. José C. Principe (Universidade da Flórida - Gainesville), que muito contribuíram para o estabelecimento da área de pesquisa denominada *Information Theoretic Learning* (ITL) (PRINCIPE, 2010) ou Aprendizado de Máquina Baseado na Teoria da Informação. O cerne da proposta de ITL é a implementação do conceito de Filtragem Adaptativa de Informação, que consiste em substituir a função custo da Equação 1.23 (ou qualquer outra baseada em ESO) por um descritor baseado em TI.

Vale ressaltar que este novo paradigma não necessita obrigatoriamente da presença de um sinal da resposta desejada, em outras palavras, estratégias não-supervisionadas também podem ser desenvolvidas no contexto de ITL. Por exemplo, entre os diversos métodos para Análise de Componentes Independentes (ICA, do inglês *Independent Component Analysis*) — ferramenta amplamente utilizada em separação cega de fontes (COMON; JUTTEN, 2010) e no domínio de equalização temporal (ROMANO *et al.*, 2011) — há algoritmos que se baseiam nas medidas de TI como a entropia e/ou informação mútua (HYVARINEN *et al.*, 2001).

Por fim, há uma importante ressalva a se fazer: é comum que não se conheça a distribuição de probabilidade dos dados, estando disponível somente um conjunto de observações. Por isso, é necessário lançar mão de técnicas de estimação para as probabilidades e para as medidas de TI, a fim de implementar os algoritmos adaptativos de ITL. A proposta desenvol-

vida pelo grupo do Prof. Principe tem como premissas básicas as de que os dados fornecidos ao sistema adaptativo são contínuos e o critério de otimização deve utilizar um estimador de densidade de probabilidade não-paramétrico, usualmente a técnica de estimação por *kernels*, também conhecida como método de janela de Parzen (1962). A metodologia se associa de modo marcante às definições generalizadas propostas por Rényi (vide subseção 1.1.3) a fim de obter estimadores diferenciáveis, que por sua vez viabilizam o uso de algoritmos de ajuste baseados em busca por gradiente.

### 1.2.1 O Estimador de Entropia Quadrática de Rényi

O estimador pelo método de janela de Parzen é uma técnica que se baseia no uso de funções *kernel* para aproximar a PDF  $f(\mathbf{x})$ <sup>1</sup> de um vetor de variáveis aleatórias contínuas  $X$ . O problema pode ser descrito da seguinte forma: seja  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$  um conjunto de  $T$  observações  $m$ -dimensionais da PDF desconhecida  $f(\mathbf{x})$ ; assume-se que existe uma aproximação adequada  $\hat{f}(\mathbf{x})$  dada por

$$\hat{f}(\mathbf{x}) = \frac{1}{T} \sum_{i=1}^T \phi(\mathbf{x} - \mathbf{x}_i, \sigma), \quad (1.24)$$

onde  $\phi(\cdot)$  é chamada de função janela e  $\sigma$  é o parâmetro de largura da janela. Parzen demonstra que  $\hat{f}(\mathbf{x})$  converge para a densidade real se  $\phi(\cdot)$  e  $\sigma$  são escolhidas de forma a atender às seguintes propriedades (PARZEN, 1962):

1. A função janela deve ser uma função de densidade não-negativa e com valores finitos, de forma que  $\int \phi(\mathbf{u}, \sigma) d\mathbf{u} = 1$ .
2. O parâmetro de largura da janela deve ser uma função de  $T$  tal que  $\lim_{T \rightarrow \infty} \sigma(T) = 0$ , e  $\lim_{T \rightarrow \infty} T\sigma^m(T) = \infty$ .

As funções *kernel* são uma classe especial de funções que atende às propriedades requisitadas para o papel de função janela e, por isso, são as mais utilizadas no método. Mais especificamente, a função *kernel* gaussiana circular é uma das mais adotadas, tendo como expressão

$$\kappa(\mathbf{u}, \sigma) = \frac{1}{(2\pi)^{m/2} \sigma^m} \exp\left(-\frac{\mathbf{u}^T \mathbf{u}}{2\sigma^2}\right). \quad (1.25)$$

Com esta escolha de função janela, a expressão final do estimador de PDF fica:

$$\hat{f}(\mathbf{x}) = \frac{1}{T} \sum_{i=1}^T \kappa(\mathbf{x} - \mathbf{x}_i, \sigma). \quad (1.26)$$

<sup>1</sup> Por razões de simplicidade, a partir deste ponto descartamos o subscrito referente à variável aleatória ao referenciar a sua PDF ou PMF.

Estabelecido o método para estimar a função de probabilidade que descreve os dados, pode-se aplicar a sua expressão às definições de entropia diferencial de Shannon ou de Rényi, o que resulta em estimadores conhecidos como métodos *plug-in* (BEIRLANT *et al.*, 1997). Porém, a definição de Shannon ainda requer que uma operação relativamente complexa de integração sobre uma função logarítmica seja avaliada, enquanto, na versão de Rényi, Principe (2010) mostra que, empregando a Equação 1.26, a integral resultante tem uma solução consideravelmente mais simples.

Considere o caso unidimensional ( $m = 1$ ), se substituirmos na Equação 1.19 o valor exato da PDF pelo estimador de Parzen:

$$\begin{aligned}
 \hat{h}_2(X) &= -\log \int \hat{f}^2(x) dx \\
 &= -\log \int \left[ \frac{1}{T} \sum_{k=1}^T \kappa(x - x_k, \sigma) \right]^2 dx \\
 &= -\log \frac{1}{T^2} \int \left[ \sum_{j=1}^T \sum_{k=1}^T \kappa(x - x_k, \sigma) \kappa(x - x_j, \sigma) \right] dx \\
 &= -\log \frac{1}{T^2} \sum_{j=1}^T \sum_{k=1}^T \int \kappa(x - x_k, \sigma) \kappa(x - x_j, \sigma) dx \\
 &= -\log \frac{1}{T^2} \sum_{j=1}^T \sum_{k=1}^T \kappa(x_k - x_j, \sqrt{2}\sigma). \tag{1.27}
 \end{aligned}$$

Repare que o “truque” está no fato de que a integral do produto de duas gaussianas é exatamente uma gaussiana avaliada na diferença entre os argumentos e com uma variância final igual à soma das variâncias originais. Além disso, este estimador é bastante atraente, do ponto de vista de aprendizado adaptativo, porque ele é não-paramétrico, contínuo e diferenciável. Tais propriedades permitem a concepção de algoritmos adaptativos baseados em técnicas por gradiente.

O argumento da função log na Equação 1.19,  $E[f_X(x)]$ , é chamado de potencial de informação (IP, do inglês *Information Potential*) —  $V(X)$  — e, durante o processo de adaptação de um algoritmo de aprendizado, podemos considerar somente a otimização de  $V(X)$ , uma vez que  $h_2(X)$  é uma função monotônica. Para chegar ao seu estimador, basta também considerarmos somente o argumento na Equação 1.27, ou seja:

$$\hat{V}(X) = \frac{1}{T^2} \sum_{j=1}^T \sum_{k=1}^T \kappa(x_k - x_j, \sqrt{2}\sigma). \tag{1.28}$$

Este resultado é um dos mais importantes em ITL, pois demonstra que a medida IP, escalar, pode ser estimada *diretamente* a partir das amostras avaliadas em pares, sem

dependem do formato da PDF. Assemelha-se, assim, aos conhecidos estimadores de média e variância que operam sem a dependência da distribuição, com a exceção de neste caso haver também a variância da gaussiana (também chamada de tamanho do *kernel*): um parâmetro livre que deve ser definido com relativo cuidado, pois pode modificar completamente a natureza da superfície de otimização (PRINCIPE, 2010).

Vale também ressaltar que o estimador de potencial de informação desempenha um papel chave em uma série de outros estimadores baseados nas medidas de Rényi. Por exemplo, ao adotar a mesma estratégia de uso do método de Parzen com função *kernel* gaussiana nas medidas de informação mútua e divergência vistas na subseção 1.1.3, chega-se a expressões descritas em termos da quantidade IP.

## 1.2.2 Aplicações

A área de ITL permite elaborar uma série de associações entre estruturas de filtragem e critérios de adaptação, supervisionados ou não-supervisionados, que, conseqüentemente, levam à possibilidade de aplicação nos mais diversos problemas. A título de ilustração, vale mencionar:

### 1. Aprendizado Supervisionado

- a) Identificação e Predição: Erdogmus e Principe (2002) empregam o critério de minimização da entropia do erro (MEE) em um novo algoritmo de treinamento para redes neurais do tipo *multilayer perceptron* (MLP), que desempenha a tarefa de prever e identificar sistemas dinâmicos não-lineares.
- b) Equalização: nesse problema, encontramos algoritmos apoiados pelo critério MEE para equalização linear e não-linear (SANTAMARIA *et al.*, 2002), e também algoritmos que se baseiam na correntropia (SANTAMARIA *et al.*, 2006) — uma medida generalizada de correlação que é fundada na medida IP —, de modo a treinar equalizadores lineares pelo critério de maximização da correntropia entre o sinal de saída e o desejado (SINGH; PRINCIPE, 2009).
- c) Classificação: a aplicação de redes do tipo MLP treinadas com o critério MEE também é encontrada nos problemas de classificação de padrões, como atestam os trabalhos de Silva *et al.* (2006), Silva *et al.* (2010). Entretanto, as análises desenvolvidas mostram que este critério possui uma diversidade de comportamentos que depende das características estatísticas dos dados de entrada, de forma que ainda não há conclusões sólidas afirmando que o novo critério pode superar o critério de EQM mínimo (PRINCIPE, 2010).

## 2. Aprendizado Não-Supervisionado

- a) Desconvolução / Equalização cega: este problema é um dos que mais comportam abordagens por ITL. Há, por exemplo, os algoritmos baseados (i) na minimização da entropia da saída (WIGGINS, 1978; DONOHO, 1981; ERDOGMUS *et al.*, 2004), (ii) no casamento de correntropias (SANTAMARIA *et al.*, 2006) ou de PDFs (CAVALCANTE *et al.*, 2002; CAVALCANTE; ROMANO, 2009) e (iii) na minimização da taxa de informação mútua para canais não-lineares (TALEB *et al.*, 2001).
- b) Separação Cega de Fontes: ICA tem sido a principal ferramenta para solução deste problema, e existem diversas propostas de algoritmos para ICA que consideram, de maneira comum, o objetivo de minimizar a informação mútua entre os componentes da saída do modelo. Entre elas, destacam-se os algoritmos com o critério de negentropia (HYVARINEN *et al.*, 2001), o critério Infomax (BELL; SEJNOWSKI, 1995) e, mais recentemente, as propostas baseadas na minimização da informação mútua de Rényi (HILD II; ERDOGMUS, 2001; HILD II *et al.*, 2006).
- c) Clusterização: uma aplicação interessante de ITL encontra-se no uso da informação mútua para determinar consenso em comitês de algoritmos de clusterização, como mostram os trabalhos de Strehl e Ghosh (2002), Topchy *et al.* (2005) e Montalvao e Canuto (2010).

Como dito, os exemplos de aplicações previamente listados foram escolhidos para prover uma visão geral da utilização das ferramentas de ITL, e, de forma nenhuma, esgotam as possibilidades de emprego do arcabouço teórico aqui apresentado em domínios diferentes. Por isso, recomenda-se fortemente ao leitor que consulte (PRINCIPE, 2010) caso deseje obter mais exemplos.

### 1.3 Uma Visão Ampliada de ITL

Pode-se dizer que a proposta inicial de ITL delineada pelo grupo do Prof. Principe inspira-se em um estimador contínuo e diferenciável — o de potencial de informação — a fim de desenvolver algoritmos de ajuste baseados no cálculo do gradiente para os mais diversos critérios, como a própria entropia quadrática, a informação mútua de Rényi, a divergência quadrática ou outras medidas que daí são derivadas.

Sem dúvida, o histórico de propostas e resultados até agora apresentados mostram que a abordagem tem obtido considerável sucesso na solução de problemas das áreas de processamento de sinais e aprendizado de máquina, como ilustram alguns dos exemplos mencionados

na subseção 1.2.2 (ERDOGMUS; PRINCIPE, 2002; SILVA *et al.*, 2006; WIGGINS, 1978; TOPCHY *et al.*, 2005). Entretanto, há três ressalvas importantes que devem ser feitas ao paradigma “canônico” de ITL. A primeira é que, além das questões práticas de implementação, não há garantias teóricas que sustentem de forma geral a escolha da entropia de Rényi — em especial a quadrática — em detrimento da entropia de Shannon, o que significa que pode haver domínios em que tal medida não funcione tão bem quanto a abordagem clássica. Por exemplo, Hild II *et al.* (2006) indicam que o algoritmo de mínima informação mútua de Rényi para separação cega de fontes pode funcionar como função contraste (COMON; JUTTEN, 2010) para a família de distribuições da exponencial generalizada e, em sequência, Vrins *et al.* (2007) demonstram que, mesmo neste caso, a escolha depende do parâmetro  $\alpha$  e da distribuição das fontes, o que leva à conclusão de que o critério pode falhar, dependendo dos dados disponíveis.

A segunda ressalva diz respeito ao fato de que o uso de algoritmos de busca baseados no cálculo do gradiente da função custo traz o risco de prover soluções localmente ótimas. Uma vez que os resultados experimentais têm demonstrado que o contexto mais vantajoso para se aplicar algoritmos de ITL frequentemente apresenta as características de não-linearidade e não-gaussianidade (ERDOGMUS; PRINCIPE, 2006), é plausível conceber que a superfície de otimização seja multimodal com relação aos parâmetros do modelo. Portanto, algoritmos de busca por gradiente possuem uma razoável chance de proverem soluções sub-ótimas em um cenário como este, dependendo das condições de inicialização (LUENBERGER, 2004).

A terceira ressalva é que, nesta configuração, ITL ignora os problemas cujos dados envolvidos são discretos ou simbólicos. Em um mundo no qual o crescimento da presença dos dados digitais é vertiginoso (HEY *et al.*, 2009; GLEICK, 2013), surgem inúmeras possibilidades de aplicação do ferramental baseado na teoria da informação para extrair conhecimento. Temos, por exemplo, os problemas no domínio da genômica (HU; PAN, 2007), da mineração de dados e da recuperação/extração de informação na Web.

A fim de superar estes obstáculos, pode-se considerar uma extensão do escopo de ITL, que (i) permita algoritmos de busca mais gerais, o que é possível caso se amplie o leque de estratégias de otimização utilizadas — por exemplo, seguindo a já comentada linha de meta-heurísticas bio-inspiradas como algoritmos genéticos (HOLLAND, 1962) e sistemas imunológicos artificiais (DE CASTRO; TIMMIS, 2002b). Desta forma, passa a ser (ii) permitido trabalhar com estimadores não necessariamente diferenciáveis, o que por sua vez significa (iii) que a escolha de um critério baseado em TI torna-se mais flexível, sendo possível utilizar as definições de Shannon ou Rényi.

Ademais, nessa nova perspectiva, dados discretos poderiam ser igualmente contem-

plados em ITL, e, nesse sentido, novas soluções para problemas não-contínuos complexos podem ser obtidas. A adoção de algoritmos de busca bio-inspirados é natural também à luz de sua reconhecida competência em determinar soluções interessantes para problemas combinatórios (HOLLAND, 1962). Conforme será apresentado nos próximos capítulos, esta visão ampliada de ITL é o que permeia praticamente todas as contribuições desta tese.

## 1.4 Conclusões

Este capítulo apresentou os principais fundamentos e as origens da área denominada de Aprendizado de Máquina Baseado na Teoria da Informação ou *Information Theoretic Learning*, iniciando pela introdução à Teoria da Informação, um campo do conhecimento científico cujo principal pioneiro é Claude E. Shannon, e que revolucionou o desenvolvimento das tecnologias de comunicação e computação, e que segue trazendo impactos nas mais distintas áreas do conhecimento, como a biologia, física e economia. Em seguida, introduziu-se a motivação e a origem para a área de ITL, iniciada no contexto da extensão de critérios de adaptação baseados em ESO.

Assim, uma série de algoritmos e ferramentas vêm sendo propostas para os mais diversos problemas, por exemplo no contexto de classificação, identificação, predição, equalização, entre outros. Inicialmente calcados no estimador de entropia quadrática de Rényi ou no conceito subjacente de potencial de informação, as técnicas possuem em comum a estratégia de efetuar a busca através do cálculo do gradiente da função custo. Finalmente, o capítulo se encerra propondo ampliar este escopo de modo a permitir estratégias de busca alternativas, associadas a descritores de TI não necessariamente diferenciáveis e que, desta forma, também podem contemplar problemas de natureza discreta.

O próximo capítulo apresenta um dos dois problemas gerais tratados nesta tese, à luz do paradigma de ITL: a separação de sinais em corpos finitos. Além de elucidar os conceitos algébricos deste “novo mundo” e descrever matematicamente o problema, ele traz uma discussão das principais abordagens desenvolvidas até então, incluindo as perspectivas de melhoria.

## 2 Separação de Sinais em Corpos Finitos

Este é o capítulo que apresenta o primeiro problema abordado nesta tese de doutorado, o da separação de sinais em corpos finitos. Ele se inicia com uma introdução à técnica de ICA e o seu papel como solução para o problema de separação cega de fontes. Em seguida, discutimos os aspectos peculiares de corpos finitos, para, daí, chegar ao recente modelo de ICA em corpos de Galois. São apresentados os principais desenvolvimentos neste novo tema de pesquisa, incluindo os algoritmos de separação mais importantes e fundamentos teóricos. Ao final, são levantados diversos aspectos que o assunto suscita, e que motivaram contribuições da tese, as quais serão apresentadas mais adiante.

### 2.1 Análise de Componentes Independentes

A análise de componentes independentes (ICA) foi concebida, a princípio, como uma ferramenta para solução de problemas de processamento de sinais não-supervisionados, em particular, do problema de separação cega de fontes (BSS, do inglês *Blind Source Separation*) (COMON; JUTTEN, 2010). Adicionalmente, ICA pode ser vista como uma extensão natural da análise de componentes principais (PCA, do inglês *Principal Component Analysis*) (HOTELLING, 1933), que é uma ferramenta clássica baseada em estatísticas de segunda ordem para realizar análise de fatores e redução de dimensionalidade.

O problema de BSS corresponde à tarefa de recuperar um conjunto de sinais (fontes), com um certo conteúdo de interesse, a partir de um conjunto de sinais (misturas) observados. Em termos matemáticos, seja  $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_N(n)]^T$  um vetor composto de  $N$  observações em um dado instante  $n$ , que foi originado pela combinação linear de  $N$  sinais independentes. Isto resulta em

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

onde  $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$  é o vetor de fontes, que são considerados também independentes e identicamente distribuídos (iid) no tempo, e  $\mathbf{A}$  é denominada matriz de mistura, que tem dimensão  $N \times N$  e deve ser inversível. Esta formulação corresponde ao modelo *determinado* de mistura *linear, instantânea* e sem ruído, a qual estabelece o problema de encontrar uma matriz de separação  $\mathbf{W}$  tal que

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) = \mathbf{PDA}^{-1}\mathbf{x}(n) = \mathbf{PD}\mathbf{s}(n). \quad (2.2)$$

Nesta condição ideal, os elementos de  $\mathbf{y}(n)$  são iguais às fontes exceto por ambiguidades de escala e permutação, representadas, respectivamente, por uma matriz diagonal  $\mathbf{D}$  e uma matriz de permutação  $\mathbf{P}$ . A Figura 2 ilustra o modelo do problema de BSS: note que, como é inerente a um problema não-supervisionado, a parte esquerda da figura (fontes e matriz de mistura) é desconhecida, e, portanto, seus valores não podem ser utilizados para se descobrir a solução.

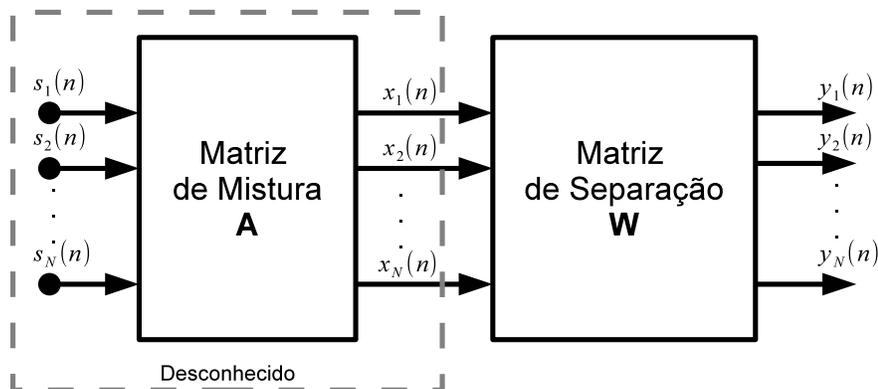


Figura 2 – Modelo de mistura linear e instantânea do problema de BSS.

A conexão entre BSS e ICA se estabelece através da constatação de que, se (i) há no máximo uma fonte gaussiana, (ii) as fontes são mutuamente independentes e (iii) a matriz de mistura  $\mathbf{A}$  é inversível, é possível obter uma solução na forma descrita pela Equação 2.2 por meio da busca de uma matriz de separação que maximize a independência entre os componentes de  $\mathbf{y}(n)$  (COMON, 1994; COMON; JUTTEN, 2010).

A partir desta relação, foram elaborados algoritmos baseados nos mais variados critérios, todos partindo da premissa de tentar restaurar a condição original de independência, como: (i) negentropia (HYVARINEN *et al.*, 2001), que é um critério utilizado para permitir a maximização de não-gaussianidade; (ii) o princípio Infomax (BELL; SEJNOWSKI, 1995), que se baseia na idéia de maximizar o fluxo de informação entre as misturas e as saídas do separador; (iii) cumulantes (COMON, 1994); (iv) máxima verossimilhança (CARDOSO, 1998) e (v) curtose (HYVARINEN *et al.*, 2001).

## 2.2 Corpos Finitos

Corpos formalizam os conceitos matemáticos e conexões subjacentes à adição, à subtração, à multiplicação e à divisão de números. A definição de um corpo é (LIDL; NIEDERREITER, 1997):

**Definição 2.1** *Um corpo  $F$  é um conjunto com duas operações,  $\cdot$  e  $+$ , tal que os seguintes axiomas são válidos:*

- Fechamento:  $\forall a, b \in F \rightarrow a \cdot b, a + b \in F$
- Comutatividade:  $\forall a, b \in F \rightarrow a + b = b + a$  e  $a \cdot b = b \cdot a$
- Associatividade:  $\forall a, b, c \in F \rightarrow a + (b + c) = (a + b) + c$  e  $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
- Distributividade:  $\forall a, b, c \in F \rightarrow a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
- Elemento neutro: *Existem os elementos  $0 \in F$  e  $1 \in F$ ,  $0 \neq 1$ , tais que  $a + 0 = a$  e  $a \cdot 1 = a$ ,  $\forall a \in F$*
- Elemento inverso:  $\forall a \in F, \exists b \in F$  tal que  $a + b = 0$ , e se  $a \neq 0, \exists c \in F$  tal que  $a \cdot c = 1$

A notação usual da aritmética sobre números reais também se aplica aqui, o que significa que podemos denotar  $-a$  o inverso com respeito à adição de um elemento  $a \in F$  e  $a^{-1}$  o inverso com respeito à multiplicação (caso exista), além de dar às respectivas operações os nomes de subtração e divisão; pode-se dar maior precedência à multiplicação e divisão que à soma e subtração e omitir o símbolo “ $\cdot$ ”.

Os conjuntos dos números reais e o conjunto dos complexos fornecem dois dos exemplos mais conhecidos de corpos, ambos de tamanho infinito. Porém, a cardinalidade infinita não é obrigatória segundo a Definição 2.1: um corpo finito, também denominado corpo de Galois em homenagem ao seu criador, o brilhante matemático francês Évariste Galois, é um corpo que possui somente um número  $q$  de elementos. Tome, como exemplo, um conjunto binário  $F = \{0, 1\}$  e as operações descritas pelo produto “usual” e pela operação de OU-exclusivo —  $0 + 1 = 1, 0 + 0 = 1 + 1 = 0, 0 \cdot 0 = 0 \cdot 1 = 0, 1 \cdot 1 = 1$  —: é diretamente verificável, neste caso, que todos os axiomas que caracterizam um corpo são respeitados.

Demonstra-se que, para um corpo de Galois de tamanho (ou ordem)  $q$ , denominado  $GF(q)$ ,  $q$  deve ser a potência de algum primo, i.e.,  $q = P^n$ ,  $P$  sendo um primo e  $n$  um inteiro positivo. Se  $n = 1$ , o corpo é chamado de corpo primo e seus valores são definidos pelo conjunto  $\{0, 1, \dots, P - 1\}$  (ou por alguma versão isomórfica) com as operações sendo simplesmente o produto e a soma “usuais” *módulo*  $P$  (LIDL; NIEDERREITER, 1997). Por outro lado, corpos finitos com um número não-primo de elementos implica uma definição mais complexa das operações.

É também possível construir espaços vetoriais em corpos finitos, com as seguintes observações (GUTCH *et al.*, 2012):

1. Considerando um espaço vetorial  $V$  de dimensão  $N$  em um corpo finito  $F = GF(q)$ , podemos denotá-lo  $V = F^N$  e seus elementos são vetores dispostos em coluna, em que cada posição é um valor pertencente a  $F$ .

2. Um mapeamento linear  $\mathbf{A} : F^N \rightarrow F^M$  pode ser representado por uma matriz  $M \times N$  com valores em  $F$ , e a composição de matrizes de tamanho compatível é denominada, como já conhecemos usualmente, de “produto matricial”, tomando o devido cuidado de que as operações elemento a elemento sejam feitas em  $F$ .
3. O cálculo do determinante de uma matriz quadrada  $\mathbf{A}$  é calculado da maneira convencional, por exemplo, pela regra de Cramer ou pela fórmula de Laplace associada às operações elemento a elemento definidas em  $F$ ; ademais,  $\mathbf{A}$  é inversível se o seu determinante for diferente do elemento nulo.
4. O conjunto de matrizes  $N \times N$  inversíveis em  $F = GF(q)$  é denotado por  $GL(N, q)$  ou  $GL(N, F)$ .
5. Diferentemente do conjunto dos reais, elementos de um corpo finito não podem ser ordenados. Em outras palavras, os elementos são usualmente considerados “símbolos”, em vez de valores numéricos. Por exemplo, em  $GF(3)$  com um conjunto de elementos  $\{0, 1, 2\}$ , não podemos afirmar que o elemento “2” é maior que “1”, pois existe um conjunto isomórfico  $\{0, 1, -1\}$  em que o símbolo “-1” equivale ao “2” original. Isto implica que o produto entre dois vetores em  $V$  definido por

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

não é definido positivo e, por isso, não pode ser considerado um produto interno. Além disso, existe um vetor não-nulo  $\mathbf{c} \in V$  tal que  $\langle \mathbf{c}, \mathbf{c} \rangle = 0$  — por exemplo, tome  $F = GF(2)$  e  $\mathbf{c} = [1, 1]^T$  —, o que significa que não existe o conceito de ortogonalidade para espaços vetoriais em corpos finitos.

### 2.2.1 Probabilidade e Estatística em Corpos Finitos

De maneira geral, os conceitos de probabilidade e estatística empregados para variáveis aleatórias discretas são válidos para variáveis definidas em um corpo finito. Seja  $X$  uma VA cujos valores observáveis pertencem a  $F$ : ela pode ser totalmente descrita pela sua PMF  $p_X : F \rightarrow \mathbb{R}$

$$p_X(x) = Prob[X = x] \in [0, 1], \quad (2.3)$$

tal que  $\sum_{x \in F} p_X(x) = 1$ . De maneira análoga, temos o caso da PMF conjunta  $p_Z : F^N \rightarrow \mathbb{R}$  de um vetor aleatório  $Z$  com possíveis valores definidos em  $F^N$

$$p_Z(\mathbf{z}) = Prob[Z = \mathbf{z}] \in [0, 1], \quad (2.4)$$

tal que  $\sum_{\mathbf{z} \in F^N} p_Z(\mathbf{z}) = 1$ .

Se temos uma transformação linear inversível  $\mathbf{A} \in GL(N, q)$  e queremos saber a PMF do vetor aleatório  $U = \mathbf{A}Z$ , uma sutil diferença aparece em relação ao caso de variáveis reais:

$$p_U(\mathbf{u}) = p_Z(\mathbf{A}^{-1}\mathbf{u}). \quad (2.5)$$

Note que, diferentemente do caso em que transformações sobre variáveis contínuas requerem que haja a divisão da PDF original por  $\det(\mathbf{A})$  (LEON-GARCIA, 1994), aqui não há esta necessidade, porque transformações bijetoras em uma variável discreta somente rearranjam os seus elementos.

Finalmente, uma diferença importante em relação a VAs reais ou complexas é que não existe o conceito de momento para VAs em corpos finitos. Isto ocorre porque, se analisarmos, por exemplo, a definição “ingênua” de esperança de uma variável  $E[X] = \sum_{u \in F} u p_X(u)$ , surge na expressão um produto entre reais e elementos pertencentes a  $F$ , o que não faz sentido. Isto valoriza o emprego dos descritores baseados em TI neste contexto, pois eles são funcionais dependentes exclusivamente das probabilidades e, por isso, são formas consistentes de sumarizar através de medidas escalares a estrutura probabilística das variáveis.

## 2.3 ICA em Corpos Finitos

Até recentemente, os algoritmos de ICA lidavam exclusivamente com sinais compostos por valores reais ou complexos (HYVARINEN *et al.*, 2001). Entretanto, o panorama muda quando Arie Yeredor (2007) estende a aplicabilidade de ICA para sinais binários para os quais a adição tem o significado da operação booleana de “Ou Exclusivo”, que é o caso de  $GF(2)$ , exemplificado na seção 2.2. Seu esforço pioneiro foi elegantemente estendido para compreender conjuntos genéricos de números discretos e finitos — os corpos de Galois —, em 2010, pelo trabalho de Gutch *et al.* (2010).

O interesse em sinais deste tipo pode ser justificado em termos teóricos — tanto do ponto de vista de construir um entendimento mais completo do problema de BSS como das potenciais relações com a teoria de codificação — e também em termos práticos — sob a visão, por exemplo, dos desafiadores problemas de mineração de dados impostos pela genética moderna e dos problemas relacionados à transmissão e armazenamento na computação tradicional ou na computação bio-inspirada (ADLEMAN, 1994).

O modelo de ICA em GF é formulado matematicamente da mesma maneira que está expresso na Equação 2.1 e na Equação 2.2, com a diferença de que os valores de todas as entidades ( $\mathbf{s}(n), \mathbf{x}(n), \mathbf{A}, \mathbf{W}, \mathbf{P}, \mathbf{D}$ ) e as operações “ $\cdot$ ”, “ $+$ ” são definidas em  $GF(q)$ . Consequentemente, todos os aspectos discutidos na seção 2.2 devem ser considerados neste

problema.

Note que a cardinalidade do conjunto de matrizes quadradas  $N \times N$  é  $q^{N^2}$ , consequentemente, o subconjunto  $GL(N, q)$  também é finito, embora possua um menor número de elementos devido à restrição de não-singularidade, o que resulta em (WATERHOUSE, 1987):

$$|GL(N, q)| = \prod_{k=0}^{N-1} (q^N - q^k), \quad (2.6)$$

onde  $|\cdot|$ , neste caso, representa o operador de cardinalidade de um conjunto.

Por exemplo, se estamos realizando ICA com 4 componentes e  $q = 2$ , existem 65536 possíveis matrizes quadradas, mas somente 20160 são inversíveis. De toda maneira, a expansão da Equação 2.6 mostra que  $|GL(N, q)|$  ainda possui um crescimento assintótico proporcional a  $q^{N^2}$ .

Estabelece-se, assim, que o processo de encontrar uma solução de separação  $\mathbf{W}$  acontece em um espaço de busca cujo tamanho cresce exponencialmente com o quadrado do número de fontes. Tal busca pode ser formulada como um problema de otimização combinatoria com uma função objetivo que avalie o grau de independência entre os componentes extraídos ( $\mathbf{y}(n)$ ). A solução ótima é aquela que maximiza esta função custo, i.e., que minimiza a dependência entre as componentes.

Claramente, adotar a busca exaustiva como algoritmo para determinar a solução é inviável para valores arbitrários de componentes e número de símbolos. Por isso, estratégias mais sofisticadas devem ser elaboradas, algumas das quais são apresentadas a seguir.

### 2.3.1 Trabalhos Desenvolvidos

Como já citado, Arie Yeredor (YEREDOR, 2007) foi o primeiro a propor a idéia de realizar ICA em corpos de Galois, mais especificamente, no contexto de misturas booleanas de OU-Exclusivo. Seu trabalho contém a prova de um teorema de separação para sinais em  $GF(2)$  e também apresenta uma propriedade importante:

**Lema 2.1** *Sejam  $U$  e  $V$  duas VAs independentes, definidas em  $GF(2)$ , e seja  $W = U + V$ . Então,  $H(W) \geq H(V)$  e  $H(W) \geq H(U)$ .*

Recorde-se que  $H(\cdot)$  é a entropia de uma VA, apresentada na Equação 1.1 do Capítulo 1. Este lema pode ser generalizado para qualquer combinação de VAs em um corpo primo (YEREDOR, 2011a) e, em termos simples, garante que o processo de mistura linear de sinais aleatórios independentes nunca causa a redução da entropia.

Baseado nisso, Yeredor propôs o algoritmo AMERICA, um método que trata a separação dos sinais pela ótica de um problema de extração cega de fontes (COMON; JUTTEN, 2010), em que esta extração deve ser executada  $N$  vezes, uma para cada sinal. Essencialmente, a técnica corresponde a uma busca exaustiva para determinar a combinação linear (não-trivial) de misturas (componentes de  $\mathbf{x}^1$ ) que recupera uma das fontes, através da condição de entropia mínima. Matematicamente, este procedimento equivale ao problema de otimização

$$\mathbf{w}_{\text{opt}} = \arg \min_{\mathbf{w} \in F^N} H(y = \mathbf{w}^T \mathbf{x}), \quad (2.7)$$

em que  $y$  é a estimativa de uma das fontes independentes e  $\mathbf{w}_{\text{opt}}$  é o vetor que indica a combinação linear ótima das misturas.

Após a extração desta fonte, um procedimento de deflação (DELFOSSÉ; LOUBATON, 1995) é aplicado para removê-la das misturas e o processo de busca é executado novamente. Ao final, as linhas da matriz de separação são os vetores correspondentes a essas combinações lineares. Note que o custo computacional da busca cresce exponencialmente com o número de componentes, uma vez que há, no pior caso,  $2^N - 1$  combinações lineares não-triviais para serem checadas para cada sinal a ser extraído.

Em 2010, Gutch *et al.* (2010), como já mencionado, estenderam os resultados de Yeredor (2007) para corpos finitos de ordem arbitrária. Eles apresentaram a prova de um teorema generalizado de separação, que estabelece as condições suficientes para realizar ICA em GF.

**Teorema 2.1** *Seja  $q$  a potência de um primo e  $F = GF(q)$  o corpo finito com  $q$  elementos. Assuma que  $\mathbf{s}$  é um vetor de  $N$  VAs independentes em  $F$  com distribuição de probabilidade  $p_{\mathbf{s}}$  tal que as distribuições marginais não são uniformes e não são degeneradas<sup>2</sup>. Então, se, para alguma matriz de posto completo  $\mathbf{G}$  em  $F$ , a distribuição de  $\mathbf{y} = \mathbf{G}\mathbf{s}$  é independente, então  $\mathbf{G} = \mathbf{P}\mathbf{D}$  para alguma matriz de permutação  $\mathbf{P}$  e alguma matriz diagonal  $\mathbf{D}$ .*

A prova do teorema pode ser vista em (GUTCH *et al.*, 2010). Note que há uma restrição para fontes cuja distribuição seja uniforme, o que é consequência de uma propriedade que afirma que a combinação de dois sinais independentes, um uniforme e outro de qualquer distribuição, resulta em um novo sinal que mantém a condição de independência (YEREDOR, 2011a) e que se torna, portanto, indistinguível (em termos de ICA) em relação à configuração original. Este comportamento “problemático” é similar ao da densidade gaussiana em ICA para sinais reais, o que indica a possibilidade de haver paralelos com noções como mínima

<sup>1</sup> Por razões de simplicidade, o índice referente ao instante será omitido daqui em diante.

<sup>2</sup> Uma distribuição de probabilidade é *degenerada* quando há pelo menos um evento do espaço amostral com probabilidade nula de ocorrência.

entropia e negentropia (COMON; JUTTEN, 2010). No entanto, repare que a condição de separação neste novo domínio é mais restritiva, pois não permite a presença de *nenhuma* fonte uniforme.

Com base nestes fundamentos teóricos, Gutch *et al.* (2010) desenvolveram três diferentes algoritmos de separação:

1. *Entropy Based Extraction/Deflation (EBED)*: este algoritmo é a generalização do algoritmo AMERICA proposto por Yeredor, no sentido de lidar com corpos finitos de ordem arbitrária. Note que o custo computacional exponencial permanece nesta técnica.
2. *Entropy Based Demixing (EBD)*: este método utiliza um princípio similar ao anterior, porém emprega uma comparação de entropias entre pares de fontes, a fim de remover uma fonte que esteja misturada com uma segunda, uma de cada vez. O algoritmo é consideravelmente mais rápido que a primeira proposta, mas é suscetível a soluções localmente ótimas.
3. *Extreme Probability Based Extraction/Deflation (EPBED)*: o algoritmo consiste em ordenar as probabilidades das observações a fim de determinar a direção de cada fonte (uma coluna da matriz de mistura) a partir das diferenças entre observações consecutivas na lista ordenada. Este algoritmo é consideravelmente mais rápido que as duas técnicas anteriores, porém é sujeito a menor acurácia e ao risco de convergência local.

Já em 2011, Yeredor publica um novo trabalho (YEREDOR, 2011a) que revisita todos os desenvolvimentos teóricos realizados até aquele momento e estende as contribuições originais de (YEREDOR, 2007) para além do caso de sinais binários. No artigo, ele também demonstra que a independência entre todas as combinações par a par de misturas implica na independência total, para os casos de  $GF(2)$  e  $GF(3)$ , o que caracteriza uma condição de separação.

Dois algoritmos iterativos são propostos para lidar especificamente com corpos primos: uma versão aprimorada do AMERICA de 2007, que é baseada na identificação sequencial da combinação linear de misturas de menor entropia, com a restrição de que cada vetor de extração de uma fonte seja linearmente independente dos vetores previamente selecionados; e o algoritmo MEXICO, que realiza a minimização da entropia de combinações de pares de misturas, em um esquema equivalente à abordagem do método EBD apresentado por Gutch *et al.* (2010).

Yeredor também apresenta uma análise preliminar de desempenho dos dois algoritmos, com resultados de simulação que demonstram uma boa qualidade de soluções, principalmente para o algoritmo AMERICA, embora o algoritmo MEXICO tenha exibido um tempo

de execução menor. Em termos de complexidade computacional, o método AMERICA é  $O(N^2q^N)$ , enquanto o método MEXICO é  $O(q^2q^N)$ . Por fim, o trabalho avança com uma primeira aplicação potencial, que seria a interceptação de um sistema de comunicação digital MIMO (do inglês *Multiple Input, Multiple Output*) que emprega o método de codificação Tomlinson-Harashima (DIETRICH, 2008).

Em 2012 foi publicado um trabalho feito em colaboração por Gutch, Yeredor et al. (GUTCH *et al.*, 2012). O artigo novamente reforça os desenvolvimentos teóricos existentes e sintetiza todas as propostas anteriores de algoritmos em três abordagens, agora válidas para qualquer ordem de corpo: os já conhecidos algoritmos AMERICA e MEXICO, e o algoritmo EPBED proposto por Gutch em 2010, agora rebatizado como CANADA. Simulações foram realizadas para comparar os métodos e levaram à conclusão de que, enquanto o algoritmo AMERICA apresenta o maior número médio de fontes extraídas, ele tem um crescimento exponencial no tempo de execução com respeito ao número de fontes; por outro lado, o algoritmo CANADA é o mais rápido, mas possui os resultados de pior qualidade; e o algoritmo MEXICO apresentou-se como uma proposta intermediária em termos de qualidade de soluções e velocidade de execução. Os experimentos numéricos também mostraram a viabilidade, ainda que limitada, do algoritmo AMERICA para interceptar transmissões MIMO que utilizam a codificação Tomlinson-Harashima, o que havia sido visto somente como uma possibilidade futura no trabalho de Yeredor (2011a).

A restrição de não-uniformidade apresentada pelo Teorema 2.1 pode ser um fator limitante para aplicação das técnicas de ICA em problemas práticos, por exemplo, na recuperação de pacotes em redes de computadores, pois a distribuição dos símbolos neste contexto tende a ser uniforme, então Nemoianu *et al.* (2013) propõem, em 2013, uma melhoria no algoritmo AMERICA de forma que este passe a incorporar uma etapa prévia de aplicação de códigos de paridade às amostras das misturas. A codificação por bits de paridade é não-linear e, por isso, tende a alterar a distribuição dos símbolos e aumentar o poder discriminativo do algoritmo de separação, conforme as simulações desenvolvidas pelos autores mostraram.

## 2.4 Conclusões

Este capítulo apresentou o problema de separação de sinais em corpos finitos e a sua abordagem através da análise de componentes independentes. Partindo da formulação “canônica” de BSS para sinais reais / complexos, descreve-se como pode ser feita a sua extensão para um novo domínio em que os sinais e estrutura misturadora pertencem a um corpo finito. Esta estrutura algébrica apresenta certas peculiaridades no que diz respeito ao tratamento estatístico e probabilístico, as quais também são discutidas, para em seguida

analisarmos as principais contribuições realizadas pelos pesquisadores pioneiros no tema.

Embora a motivação inicial tenha sido puramente teórica, já se vislumbram potenciais aplicações no contexto de interceptação de transmissões MIMO e decodificação de pacotes em redes de computadores, além de perspectivas em mineração de dados digitais e genômica. Entretanto, podemos afirmar que o problema em estudo é muito recente e, por isso, o amadurecimento das propostas já publicadas ainda não foi suficiente para testemunharmos a aplicação efetiva e perene em cenários reais. Dado este panorama atual da pesquisa, verificamos alguns aspectos que permitem o desenvolvimento de novas contribuições:

- Os algoritmos AMERICA e MEXICO, os quais representam a classe de técnicas com melhor qualidade de separação, empregam estratégias de busca exaustiva para obter a solução, o que naturalmente se reflete na alta complexidade computacional que apresentam (ambos exponenciais com relação ao número de fontes). Além disso, uma estrutura tensorial é proposta para estimar os valores de entropia das misturas, o que implica numa estrutura de dados de tamanho  $q^N$  e um custo de memória também exponencial com o número de fontes.
- É reconhecido o emprego de algoritmos bio-inspirados para a solução dos mais diversos problemas combinatórios (HOLLAND, 1962; DE CASTRO; TIMMIS, 2002b; DE CASTRO, 2006). Por isso, há um grande potencial para desenvolver novas modelagens do problema dentro deste contexto, com a finalidade de diminuir o custo computacional e aumentar a qualidade das soluções.
- O modelo de ICA em GF apresentado está limitado, por enquanto, a misturas *instantâneas* de fontes, o que deixa em aberto a possibilidade de estender o problema para o caso em que os sinais sofrem um processo de mistura tanto no espaço como no tempo. Este é o caso do problema de separação de sinais em misturas convolutivas (COMMON; JUTTEN, 2010), que já possui diversas abordagens para sinais reais, mas ainda não foi explorado a fundo no domínio de corpos finitos.

Consequentemente, as propostas desenvolvidas nesta tese se pautam por estes pontos, e serão detalhadas no Capítulo 5. Antes disso, porém, o próximo capítulo apresenta o segundo problema abordado nesta tese, o da inversão cega de sistemas de Wiener.

## 3 Inversão Cega de Sistemas de Wiener

Após a análise do domínio de corpos de Galois, ainda consideravelmente novo no âmbito da área de processamento de sinais, passamos para a apresentação, neste capítulo, de um segundo problema que nos traz de volta ao “confortável” domínio do processamento de sinais de valores reais. Aqui tratamos, especificamente, do problema de inversão de sistemas de Wiener, discutindo aspectos introdutórios para a construção de modelos inversos e descrevendo as demandas técnicas subjacentes à tarefa de realizar a inversão do sistema não-linear de Wiener por meio de métodos não-supervisionados. Apresentamos também as duas principais metodologias pertencentes a esse paradigma, acrescidas de uma análise crítica sobre os principais pontos passíveis de melhorias.

### 3.1 Inversão de Sistemas

Modelagem inversa adaptativa é uma das aplicações tradicionais de processamento de sinais (HAYKIN, 2001). Neste problema, que é ilustrado pela Figura 3, um filtro adaptativo é utilizado para realizar, dinamicamente, a tarefa de contrabalancear os efeitos de um sistema desconhecido. Repare que, para a maioria das aplicações, é aceitável que o processo de inversão suscite um atraso e até um fator de escala, uma vez que ambas as operações ainda preservam a informação presente (HAYKIN, 2001).

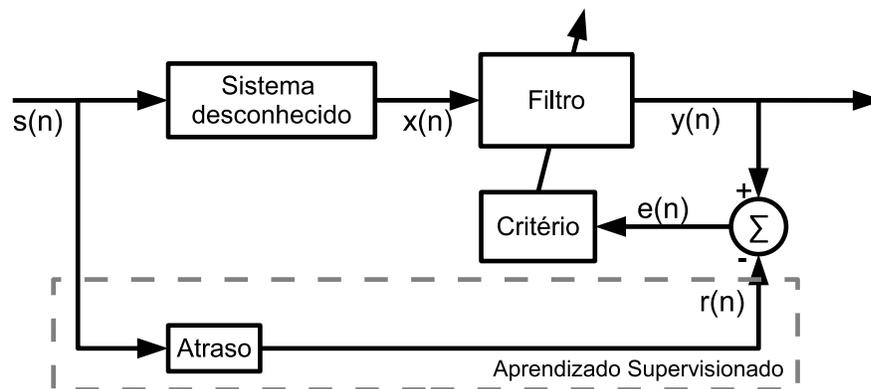


Figura 3 – Modelagem inversa de sistemas.

No contexto da teoria de comunicações, a tarefa de inversão é comumente denominada de *equalização*, associação essa que vem do fato do sinal a ser recuperado ser tipicamente um processo estocástico independente, logo, com densidade espectral de potência uniforme.

De uma perspectiva mais geral em processamento de sinais, pode-se dar a esse problema o nome de *desconvolução*. Quando o sinal original é “branco”, equalização e desconvolução são tarefas equivalentes, mas a desconvolução, que é apenas uma inversão de canal linear, pode até recuperar sinais não equalizados.

A teoria de desconvolução / equalização foi inicialmente desenvolvida sob a hipótese de linearidade do sistema (também denominado de canal) e do filtro inverso. Adicionalmente, assumiu-se que há um sinal de referência (também chamado de sinal desejado) disponível para auxiliar na escolha dos parâmetros da estrutura de filtragem. A presença deste tipo de sinal caracteriza a abordagem como sendo *supervisionada*; por outro lado, as dificuldades encontradas para se ter acesso a este sinal de informação em certos casos encorajou o desenvolvimento de estratégias *não-supervisionadas* ou *cegas* (ROMANO *et al.*, 2011).

A inversão de sistemas não-lineares é uma tarefa que demanda maiores cuidados. Isto ocorre devido aos seguintes fatores:

1. O tratamento matemático de um problema não-linear exige, até certo ponto, um desvio do elegante e conceitualmente útil arcabouço teórico de sistemas lineares.
2. Efeitos não-lineares são capazes de “destruir” a informação contida em  $s(n)$ , o sinal a ser recuperado, o que pode comprometer as perspectivas estruturais de inversão.
3. A ambiguidade introduzida por um canal não-linear genérico é um fator de degradação de desempenho para esquemas supervisionados. Se o contexto, então, for restrito a métodos não-supervisionados, ela pode ser ainda mais impactante, eventualmente criando situações impeditivas para uma solução (JUTTEN; KARHUNEN, 2004). Tal cenário explica por que, neste último caso, é usual limitar-se a formulações contendo não-linearidades perfeitamente inversíveis, i.e., não-linearidades que podem ser descritas em termos de um mapeamento um-a-um.

Essas dificuldades não são, porém, impossibilidades *a priori*. O uso consistente de métodos não-supervisionados no contexto de sistemas não-lineares tem lugar em uma série de trabalhos importantes, como (TALEB; JUTTEN, 1999; TALEB *et al.*, 2001; HOSSEINI; DEVILLE, 2003; DUARTE *et al.*, 2009; DUARTE *et al.*, 2012).

### 3.1.1 Inversão Cega de Sistemas de Wiener por Sistemas de Hammerstein

Nesta tese, consideramos um modelo que pode ser compreendido como um híbrido entre o efeito convolutivo e uma distorção não-linear inversível — o modelo de Wiener (TALEB *et al.*, 2001), que é ilustrado pela Figura 4. O modelo consiste de um sub-sistema linear

e invariante no tempo (LTI, do inglês *linear and time-invariant*) com resposta ao impulso  $h(n)$ , ao qual se segue uma transformação não-linear  $f(\cdot)$ .

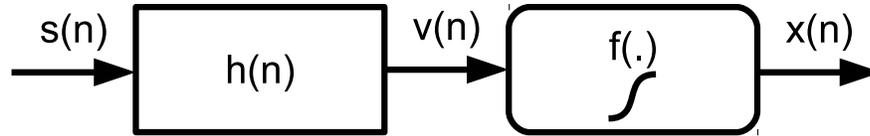


Figura 4 – O sistema de Wiener.

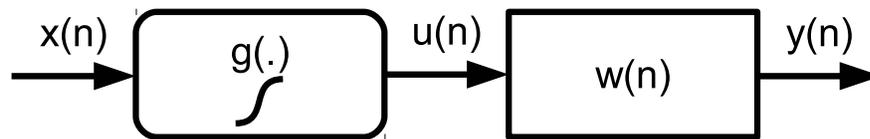


Figura 5 – O sistema de Hammerstein.

O modelo de Wiener é atraente porque ele, em um certo sentido, insinua a possibilidade de lidar separadamente com os estágios linear e não-linear, conseqüentemente preservando uma série de dispositivos teóricos interessantes. O modelo também é representativo de diversos cenários práticos, como: a dinâmica das fibras musculares em biologia (HUNTER; KORENBERG, 1986), controle de plantas de destilação na indústria (BARS *et al.*, 1990), rastreamento de pessoas em visão computacional (MA *et al.*, 2006) e cancelamento de ecos acústicos (COSTA *et al.*, 2003).

Uma opção natural para desempenhar o papel de modelo inverso de um sistema de Wiener é o denominado sistema de Hammerstein (HUNTER; KORENBERG, 1986), o qual é composto de uma transformação não-linear sem memória (função  $g(\cdot)$ ) seguida por um sub-sistema LTI (um filtro com resposta ao impulso  $w(n)$ ), conforme ilustra a Figura 5. Assim, o problema de inversão cega consiste em determinar  $w(n)$  e  $g(\cdot)$  tal que a saída do sistema de Hammerstein seja a mais fiel possível ao sinal original  $s(n - d)$ , onde  $d$  é um atraso, quando sua entrada é  $x(n)$ . A seguir descrevemos matematicamente como esse processo de inversão cega se viabiliza.

## 3.2 Trabalhos Desenvolvidos

Considere que  $\mathcal{S} = \{s(n)\}$  é um processo estocástico de tempo discreto composto de amostras iid / não-gaussianas, e que tanto o sub-sistema LTI com resposta ao impulso  $h(n)$  como o mapeamento não-linear  $f(\cdot)$  são inversíveis. Podemos representar a saída do sistema

de Wiener como:

$$x(n) = f[v(n)] = f[s(n) * h(n)] = f \left[ \sum_k h(k)s(n-k) \right], \quad (3.1)$$

em que  $s(n) * h(n)$  é a convolução entre  $s(n)$  e  $h(n)$ , cujo resultado é submetido à distorção não-linear  $f(\cdot)$ .

Como previamente citado, o sistema de Hammerstein que a Figura 5 apresenta é a estrutura empregada para neutralizar o efeito do sistema de Wiener, o qual acaba sendo, de uma certa forma, uma “imagem espelhada” do sistema original. Analogamente à idéia adjacente à Equação 3.1, a saída do modelo é descrita por:

$$y(n) = w(n) * u(n) = \sum_k w(k)u(n-k), \quad (3.2)$$

na qual  $u(n) = g[x(n)]$ . O modelo do filtro linear com resposta ao impulso  $w(n)$  é normalmente uma estrutura *feedforward* (TALEB *et al.*, 2001; SOLE-CASALS *et al.*, 2002; ROJAS *et al.*, 2007) cuja relação entrada-saída é definida por

$$y(n) = \sum_{k=0}^M a_k u(n-k), \quad (3.3)$$

sendo que há  $M + 1$  coeficientes ajustáveis  $a_k$  do modelo. Note que, neste caso específico,  $w(k) = a_k, k = 0, \dots, M$ , e, por isso, essa estrutura é amplamente conhecida como filtro de resposta ao impulso finita (FIR, do inglês *finite impulse response*) (OPPENHEIM; SCHAFER, 2009). Duas características são importantes para justificar o amplo uso dessa estrutura: (i) sua tratabilidade do ponto de vista de obtenção das derivadas de uma função custo de filtragem e (ii) sua estabilidade intrínseca. Por outro lado, possui maiores generalidade e capacidade de representação uma estrutura linear com resposta ao impulso infinita (IIR, do inglês *infinite impulse response*), cuja relação entrada-saída é definida por

$$y(n) = \sum_{k=0}^M a_k u(n-k) + \sum_{k=1}^Q b_k y(n-k). \quad (3.4)$$

Ademais, há outros aspectos chave que precisam ser considerados para chegarmos a uma solução completa para o problema: (i) é preciso formular um modelo para a não-linearidade do sistema inverso, (ii) deve-se definir um critério para ajustar os parâmetros livres dos estágios linear e não-linear do sistema de Hammerstein e (iii) é preciso definir um método para encontrar os valores ótimos de tais parâmetros. As propostas existentes são caracterizadas por opções distintas para lidar com estes três pontos levantados — por isso, discutiremos separadamente as abordagens canônicas a seguir.

### 3.2.1 A Metodologia de Taleb, Sole-Casals e Jutten

Taleb *et al.* (2001) apresentaram o primeiro método cego para inverter sistemas de Wiener. Eles desenvolveram uma análise teórica do problema que levou à proposta de um algoritmo baseado em gradiente que busca minimizar a taxa de informação mútua do sinal extraído, o que, idealmente, recupera o sinal original. Conforme mostram os autores, sob as hipóteses levantadas no artigo, recuperar a independência entre as amostras significa que o sistema de Hammerstein será o inverso do sistema de Wiener.

É possível traduzir a notação das Equações 3.1 e 3.2 para a forma matricial, considerando que qualquer sinal pode ser representado por um vetor de dimensão infinita e que um filtro é representado por uma matriz de Toeplitz de tamanho também infinito. Esta abordagem pretende transformar, assim, um problema intrinsecamente temporal em um problema “pseudo-espacial”, estabelecendo uma conexão com o já apresentado problema de separação cega de fontes e, mais especificamente, com o modelo *Post-Nonlinear* (PNL) (TALEB; JUTTEN, 1999). Então, Taleb *et al.* conjecturam que, em configurações Wiener-Hammerstein, analogamente à teoria de sistemas PNL, a inversão é possível desde que  $\{s(n)\}$  seja composto de amostras independentes e  $h(n)$  não seja um filtro de atraso puro (TALEB *et al.*, 2001).

Como consequência destes elementos, chegamos em um primeiro critério: buscar o conjunto de parâmetros do sistema de Hammerstein que resulta em um valor mínimo (idealmente nulo) de IM para as variáveis aleatórias correspondentes à saída do sistema. Em outras palavras, o sinal original será recuperado, com exceção de ambiguidades de atraso e escala, se  $\mathcal{Y} = \{y(n)\}$  for um processo iid.

Já apresentamos no Capítulo 1 o conceito de informação mútua, uma entidade que intuitivamente descreve a quantidade de informação “compartilhada” por duas ou mais variáveis. No caso de um vetor aleatório  $\mathbf{z} \in \mathbb{R}^m$ , ela é expressa por

$$I(\mathbf{z}) = I(z_1; \dots; z_m) = \int_{\mathbb{R}^m} f_{\mathbf{z}}(\mathbf{u}) \log \left( \frac{f_{\mathbf{z}}(\mathbf{u})}{\prod_{i=1}^m f_{z_i}(u_i)} \right) d\mathbf{u} \quad (3.5)$$

ou, em termos das entropias diferenciais

$$I(\mathbf{z}) = \sum_{i=1}^m h(z_i) - h(\mathbf{z}), \quad (3.6)$$

em que  $h(z_i)$  e  $h(\mathbf{z})$  são a entropia diferencial de  $z_i$  (vide Equação 1.15) e a entropia diferencial conjunta de  $\mathbf{z}$ , respectivamente<sup>1</sup>. Embora não tenha sido apresentada a expressão da entropia diferencial conjunta no Capítulo 1, ela é uma extensão direta do caso unidimensional:

$$h(\mathbf{z}) = h(z_1; \dots; z_m) = - \int_{\mathbb{R}^m} f_{\mathbf{z}}(\mathbf{u}) \log f_{\mathbf{z}}(\mathbf{u}) d\mathbf{u}. \quad (3.7)$$

<sup>1</sup> Observe que estamos adotando uma notação com as variáveis aleatórias em minúsculo, a fim de não causar ambiguidades com a notação para matrizes.

Uma vez que estamos lidando com processos estocásticos, e.g.  $\mathcal{Z} = \{z(n)\}$ , é necessário acrescentar a noção de taxa de entropia diferencial, definida por

$$h(\mathcal{Z}) = \lim_{T \rightarrow \infty} \frac{h(z(-T); \dots; z(T))}{2T + 1} \quad (3.8)$$

quando o limite existe. Cover e Thomas (2006) demonstram que isto ocorre quando  $\mathcal{Z}$  é estacionário. Portanto, a partir da Equação 3.6, temos condições de apresentar a taxa de informação mútua (TALEB *et al.*, 2001) do processo estocástico estacionário  $\mathcal{Z}$ :

$$\begin{aligned} I(\mathcal{Z}) &= \lim_{T \rightarrow \infty} \frac{1}{2T + 1} \left[ \sum_{n=-T}^T h(z(n)) - h(z(-T); \dots; z(T)) \right] \\ &= h(z(\tau)) - h(\mathcal{Z}), \end{aligned} \quad (3.9)$$

em que  $\tau \in \{\dots, -2, -1, 0, 1, 2, \dots\}$  é arbitrário devido à premissa de estacionariedade. Se associarmos tais definições ao contexto da estrutura Wiener-Hammerstein apresentada na Figura 4 e na Figura 5, e dado que  $\mathcal{S}$  é estacionário e  $h(n), w(n)$  são sistemas LTI, então temos que

$$I(\mathcal{Y}) = h(y(\tau)) - h(\mathcal{Y}) \quad (3.10)$$

é a função objetivo a ser minimizada, porque ela resulta em um valor nulo quando as amostras de  $\mathcal{Y}$  são independentes.

Para modelar  $g(\cdot)$ , Taleb *et al.* originalmente propuseram um método *quasi* não-paramétrico (com respeito ao estágio não-linear), porém uma alternativa com um modelo paramétrico foi introduzida em um trabalho a seguir, de Sole-Casals *et al.* (2002), adotando redes neurais artificiais ou funções polinomiais. Independentemente da abordagem, o gradiente da função custo apresentada na Equação 3.10 é calculado com respeito aos parâmetros do filtro — uma estrutura FIR anti-causal — e com respeito à não-linearidade. As expressões finais levam ao cálculo das conhecidas funções *score* (COMON; JUTTEN, 2010), o que é possível com a ajuda do método de Parzen (vide subseção 1.2.1) para estimação das densidades de probabilidade.

Tanto o trabalho de Taleb *et al.* (2001) como o de Sole-Casals *et al.* (2002) demonstram, por fim, exemplos de aplicabilidade do algoritmo no (i) problema de recuperação de sinais sísmicos, cujo modelo de refração após o experimento claramente comprova-se não-linear; e no (ii) problema de inversão para sinais correlacionados, obtendo como solução, neste caso, o processo de inovação que deve ser em seguida aplicado a um modelo auto-regressivo, conhecido *a priori*, para se chegar à solução final.

Após a publicação destes dois trabalhos pioneiros, algumas extensões importantes foram propostas:

- Babaie-Zadeh *et al.* (2003) propõem um novo algoritmo dividido em uma etapa de minimização pelo gradiente com respeito ao próprio sinal  $y(n)$ , seguido por uma etapa de atualização dos parâmetros da estrutura de Hammerstein e consequente projeção de um novo valor de saída do sistema.
- Zhang e Chan (2004) e Sole-Casals *et al.* (2005) sugerem o emprego de técnicas de gaussianização do sinal de entrada  $x(n)$  a fim de reduzir os efeitos da não-linearidade.
- Sole-Casals e Caiafa (2012) propõem o uso de interpolação por funções *spline* para diminuir o custo de estimação das funções *score* e, conseqüentemente, o tempo de execução do algoritmo.

Apesar da diversidade de contribuições e desenvolvimentos, permanece a necessidade de utilizar alguma variedade de busca baseada em gradiente, o que significa que as soluções obtidas podem ser sub-ótimas. Além disso, provavelmente também devido ao uso de métodos de busca por gradiente, o filtro linear que compõe o sistema de Hammerstein é sempre modelado como uma estrutura FIR. Uma escolha desse tipo é restritiva do ponto de vista estrutural, pois a presença de laços de realimentação pode reduzir significativamente o número de parâmetros livres em casos práticos.

### 3.2.2 A Metodologia de Rojas, Sole-Casals e Puntonet

A fim de se apresentar como uma alternativa aos algoritmos baseados em gradiente, Rojas *et al.* (2007) apresentam um novo trabalho — sequência de uma primeira abordagem para o problema de inversão linear (ROJAS *et al.*, 2006) — que desta vez emprega um algoritmo genético (GA, do inglês *genetic algorithm*) (BACK *et al.*, 2000) para executar a tarefa de busca pelos valores ótimos dos parâmetros em um problema de inversão Wiener-Hammerstein. A parte não-linear do sistema inverso é composta de um polinômio com monômios exclusivamente de potência ímpar e, em vista do fato de que a condição teórica de separabilidade (TALEB; JUTTEN, 1999) exige que  $g \circ f(\cdot)$  seja uma função monotônica, todos os seus coeficientes devem ser positivos. Semelhantemente à metodologia proposta por Taleb *et al.*, a parte linear é um filtro FIR anti-causal.

O critério escolhido para realizar o ajuste do modelo é a maximização da *curtose* da saída do sistema de Hammerstein. O Teorema do Limite Central atesta que a soma normalizada de VAs contínuas e independentes gera, quando o número de termos tende ao infinito, uma VA de distribuição normal (LEON-GARCIA, 1994). Portanto, o efeito convolutivo de um sistema linear sobre o processo original, que é iid e não-gaussiano, faz com que o formato de cada densidade de probabilidade do processo de saída se torne mais similar a uma gaus-

siana, em comparação às distribuições das amostras de entrada. De forma análoga a uma família de critérios desenvolvida para ICA (COMON; JUTTEN, 2010), é possível empregar este tipo de medida de não-gaussianidade para lidar com sistemas Wiener-Hammerstein.

A curtose é uma medida clássica de não-gaussianidade, definida por (HYVARINEN *et al.*, 2001)

$$\text{Kurt}[z] = E[z^4] - 3E[z^2]^2. \quad (3.11)$$

Alternativamente, a curtose normalizada

$$\overline{\text{Kurt}}[z] = \frac{E[z^4]}{E[z^2]^2} - 3 \quad (3.12)$$

pode ser empregada. Quanto “mais não-gaussiana” for uma variável aleatória  $z$ , maior será o valor absoluto de  $\text{Kurt}[z]$ . Em vista da potencial existência de dependência entre amostras relacionadas por atrasos distintos de tempo, Rojas *et al.* estabeleceram uma função objetivo que consiste em maximizar o valor absoluto da curtose normalizada para vários “sub-sinais” (ROJAS *et al.*, 2007):

$$J_k(g; w) = \sum_{i=1}^B |\overline{\text{Kurt}}[y_i]|, \quad (3.13)$$

em que  $B$  é o número de partições do sinal  $y(n)$  original.

Apesar da simplicidade desta função custo, ela é inegavelmente um critério de inversão menos rigoroso que a informação mútua, principalmente por causa de três pontos:

1. A definição apresentada na Equação 3.12 mostra que somente informação associada a estatísticas de até quarta ordem está sendo considerada; portanto, a presença de dependências de ordem maior não é capturada por esse descritor estatístico.
2. A curtose não é uma medida simétrica quando se comparam distribuições leptocúrticas com distribuições platicúrticas, o que pode resultar em comportamentos distintos durante um processo de adaptação para cada caso.
3. O efeito convolutivo do sub-sistema linear  $h(n)$  tem, de fato, um impacto tangível em termos de gaussianização, mas o efeito não-linear de  $f(\cdot)$  pode ter uma influência significativa em distorcer ou até mesmo anular esse comportamento (JUTTEN; KARHUNEN, 2004).

### 3.3 Conclusões

O capítulo que se encerra apresentou o problema de inversão de sistemas, que, em termos simples, consiste em reverter os efeitos distorcivos produzidos por um sistema dinâ-

mico desconhecido, a fim de restaurar o sinal de informação que existia previamente. Isto posto, acrescentamos alguns fatores complicantes ao cenário: realizar a inversão de um sistema não-linear — o sistema de Wiener — possuindo de conhecimento *a priori* sobre o sinal fonte somente o fato de que este é um processo estocástico iid e não-gaussiano.

Partindo da analogia com o problema de separação de fontes em modelos PNL, mostramos a conjectura de que a solução do problema é obtida se for recuperada a condição de independência no sinal de saída e, com base nessa teoria, discutimos as principais metodologias propostas até então, que se dividem entre a abordagem direta de estimação da taxa de informação mútua como critério para um processo de busca por gradiente e a abordagem de busca utilizando um algoritmo genético com a curtose como critério indireto de independência.

Ao longo da descrição das propostas, foram levantados alguns pontos limitantes que sintetizamos aqui:

- O algoritmo proposto por Taleb et al. baseia-se na técnica de busca por gradiente e, conseqüentemente, é vulnerável ao risco de convergência prematura para um mínimo local.
- Já o algoritmo proposto por Rojas, embora seja mais robusto em termos de risco de convergência local, é atrelado à medida de curtose como função custo, o que explora de uma maneira limitada a estrutura estatística dos sinais envolvidos.
- Ambos os métodos restringem o estágio linear do sistema de Hammerstein ao uso somente de estruturas FIR.

É viável mitigar essas limitações se elaborarmos uma nova metodologia que (i) evite o caráter de busca exclusivamente local de algoritmos por gradiente, que (ii) considere uma função objetivo apropriada e que (iii) implemente estruturas de desconvolução mais flexíveis que os modelos FIR. Esta nova abordagem é que será fruto de análise mais adiante, no Capítulo 6. Já o próximo capítulo apresenta a área de computação imuno-inspirada e as propostas de sistemas imunológicos artificiais que compõem, no papel de ferramentas de otimização, a maioria das contribuições desta tese.



## 4 Computação Imuno-Inspirada

Como já aventamos nos capítulos anteriores, flexibilizar a estratégia de busca dos parâmetros de um sistema adaptativo pode ser um caminho interessante para os algoritmos de ITL. Isto se vincula diretamente ao que este capítulo apresenta: os algoritmos baseados no sistema imunológico. Aqui vamos discutir, primeiro, os aspectos básicos de funcionamento do sistema imune de animais vertebrados, para, em seguida, relacioná-los a duas propostas de algoritmos computacionais que visam atuar como ferramentas para a solução de diversos problemas: os algoritmos CLONALG (DE CASTRO; VON ZUBEN, 2002) e cob-aiNet (COELHO; VON ZUBEN, 2010). Os esquemas de operação de ambos os métodos serão apresentados, assim como as semelhanças e diferenças entre eles.

### 4.1 Princípios Básicos do Sistema Imunológico Natural

Durante o curso da vida de qualquer ser, é natural que ele esteja exposto a ameaças ao bom funcionamento do seu corpo, como, por exemplo, a invasão de microorganismos ou substâncias causadoras de doenças. Neste contexto é que se posiciona, de maneira fundamental para a sua sobrevivência, o *sistema imunológico*, o qual nada mais é que o sistema que defende o animal contra o ataque constante de elementos invasores (TIZARD, 1995).

A importância crucial do sistema imunológico para a sobrevivência do animal requer que ele opere de maneira eficiente; ele é um sistema complexo construído, de forma redundante, com uma série de componentes e mecanismos distintos, os quais são ajustados para combater um único invasor ou para proteção contra uma série de agentes infecciosos. Destacam-se dois aspectos fundamentais para a operação do sistema: os mecanismos de *memória* e *aprendizado*. O sistema é capaz de capturar informações a respeito dos agentes infecciosos e armazená-las para serem utilizadas posteriormente, caso o mesmo agente ou um similar invada o organismo.

A gênese dos estudos sobre imunologia se deve a Edward Jenner, criador em 1796 da vacina contra a varíola, embora nada soubesse acerca dos agentes infecciosos que causam doenças (JANEWAY *et al.*, 2000). Tal enigma foi solucionado um pouco mais tarde, no século XIX, quando Robert Koch demonstrou que as doenças infecciosas são causadas por microorganismos *patogênicos* (também chamados de *patógenos*), que, atualmente, são divididos em quatro grandes classes: as dos vírus, bactérias, fungos e parasitas. Já ao final do mesmo século, sabia-se da existência das *células brancas* e do fato de que elas eram respon-

sáveis pela destruição dos organismos invasores, produzindo, para isso, *anticorpos*. Ao final do século XX, a imunologia já se encontrava sustentada por teorias relativamente sólidas, capazes de descrever com boa precisão os dispositivos de reconhecimento e eliminação de corpos estranhos (DE CASTRO, 2001).

Um *antígeno* é qualquer molécula que o sistema imune consegue reconhecer. Dada a sua presença no organismo, hoje sabemos que o sistema imunológico desempenha sua primordial tarefa por meio da ação coordenada de dois sub-sistemas: o sistema imune *inato* e o sistema imune *adaptativo*. O primeiro é composto de células de defesa prontas, os *granulócitos* e os *macrófagos*, que proveem uma resposta imediata e cuja forma de atuação é idêntica em todo indivíduo, independentemente do tempo de vida. O segundo é o sistema composto de células denominadas de *linfócitos*, dos tipos B e T, que suscitam uma resposta específica para cada patógeno cujo tempo de ocorrência depende de se já houve uma exposição prévia a este agente.

As células B são as responsáveis pela produção de anticorpos — moléculas de forma específica que se ligam a um antígeno para sinalizar que este deve ser ingerido e destruído por células especializadas (e.g. macrófagos). Além disso, as células B podem se converter em células de *memória*, que são capazes de reconhecer o mesmo antígeno em um evento futuro e, portanto, podem outra vez produzir anticorpos para evitar que a doença se reestabeleça. Dessa forma, a resposta imune adaptativa aperfeiçoa-se a cada encontro com um antígeno (DE CASTRO, 2001).

A existência das duas formas de resposta imune é interessante, já que elas adquirem um aspecto de complementariedade. Veja, por exemplo, o caso em que ocorre a invasão por um antígeno desconhecido: o sistema inato pode atuar rapidamente tentando bloquear (ou pelo menos retardar) a ação danosa do invasor, enquanto o sistema adaptativo reconhece a especificidade do agente e produz anticorpos preparados para eliminá-lo efetivamente. Além disso, o sistema inato é responsável por uma série de mecanismos de sinalização essenciais para o funcionamento do sistema adaptativo. Entretanto, interessa-nos aqui o estudo da resposta adaptativa porque ela apresenta a capacidade de aprendizado, o que é essencial na concepção de algoritmos inteligentes.

Existem duas teorias fundamentais que descrevem este processo de aprendizado: a teoria da *Seleção Clonal* e a teoria da *Rede Imunológica*, que descrevemos em mais detalhes a seguir.

### 4.1.1 Seleção Clonal

O princípio ou teoria da seleção clonal foi proposto por Frank Macfarlane Burnet em 1959 (BURNET, 1959), e sua notória contribuição lhe rendeu o prêmio Nobel de medicina do ano seguinte, sendo até hoje aceita como a teoria que explica como o sistema imune é capaz de reconhecer e produzir anticorpos somente contra os antígenos aos quais o organismo foi exposto.

Seu *modus operandi* pode ser explicado, em termos simples, como se segue. Pela nossa corrente sanguínea estão circulando constantemente inúmeros linfócitos com características diferentes — diversidade essa que decorre do processo de produção das células na medula e no timo —, ou seja, com receptores diferentes. Quando um indivíduo é exposto a um antígeno, o *reconhecimento* deste por um linfócito se dá quando os receptores que ele possui em sua membrana externa são compatíveis com a estrutura molecular antigênica. O linfócito ativado então passa a sofrer um processo de clonagem, gerando cópias que se diferenciam em (i) células especializadas (denominadas de *plasmócitos*) capazes de secretar altas taxas de anticorpos para aquele antígeno, ou em (ii) células de memória, que possuem longos períodos de vida e permanecem inativas circulando pelo organismo até que haja uma re-exposição ao estímulo antigênico, o que dispara novamente — e de forma mais rápida do que na primeira ocorrência — o processo de diferenciação em plasmócitos e secreção dos anticorpos.

Ademais, esse rápido processo de replicação é acompanhado de um mecanismo de *hipermutação* que visa introduzir modificações aleatórias tão mais intensas quanto menos efetiva a célula for no reconhecimento antigênico, ou seja, quanto menor for a sua *afinidade* com o corpo estranho. Então, as melhores células produzidas são selecionadas, de tal forma que o repertório de linfócitos como um todo se desenvolva e seja mais preciso ao reconhecer os antígenos.

### 4.1.2 Rede Imunológica

Em 1974, Niels Kaj Jerne propôs a teoria da Rede Imunológica (JERNE, 1974), uma contribuição fundamental para o entendimento do sistema imunológico como uma *rede regulada*, ou seja, um sistema em que as células são capazes de reconhecer, além de antígenos, *umas às outras*. Enquanto a teoria da Seleção Clonal se ocupou de descrever exclusivamente o comportamento frente à presença de patógenos, nesta proposta, Jerne postula que as células brancas permanecem em interação mesmo quando não estão dedicadas ao reconhecimento de um antígeno.

Desse fato decorrem duas respostas distintas que o sistema imune pode apresentar: uma resposta *positiva*, caracterizada pela ativação do linfócito que reconhece um antígeno,

desencadeando os mecanismos da Seleção Clonal previamente descritos; e uma resposta *negativa*, caracterizada pela tolerância ou supressão de uma outra célula branca reconhecida. Portanto, a teoria estabelece que os principais aspectos do sistema imune são propriedades emergentes que resultam de mecanismos regulatórios intencionados a manter a rede em um estado de equilíbrio dinâmico (BERSINI, 2003).

## 4.2 Sistemas Imunológicos Artificiais

Como visto, o sistema imune dos animais vertebrados apresenta uma notável capacidade de defender o organismo de agentes patogênicos, o que é fruto de diversos mecanismos que garantem a emergência das propriedades de *reconhecimento de padrões*, *memória* e *aprendizado*. Desta forma, o sistema não só reconhece e elimina organismos prejudiciais, como também desenvolve e aprimora seus dispositivos de defesa para que tenha um desempenho ainda mais efetivo contra ataques futuros. Tais propriedades são características de sistemas que são reconhecidos como *adaptativos* e, por isso, é natural surgir o interesse em replicar os princípios básicos do sistema imunológico no desenvolvimento de sistemas computacionais inteligentes.

A partir desta constatação, desenvolveu-se, desde a década de 90 do século passado, o conceito de um Sistema Imunológico Artificial (SIA) e a área de pesquisa associada (DE CASTRO; TIMMIS, 2002b). Sua premissa fundamental é o uso de idéias da imunologia para a construção de sistemas adaptativos capazes de realizar diversas tarefas computacionais, como reconhecimento de padrões, agrupamento de dados e otimização.

Em um contexto mais amplo, os alternativamente intitulados *algoritmos imunológicos* ou *imuno-inspirados* pertencem à classe dos algoritmos *bio-inspirados* (DE CASTRO, 2006), a qual agrega uma série de paradigmas baseados em modelos biológicos distintos. Porém, é possível afirmar que, no contexto da aplicação destes em problemas de otimização, todos são capazes de explorar o espaço de busca (em graus diferentes) a fim de encontrar a melhor solução possível, dados os recursos computacionais disponíveis.

A maioria dos algoritmos bio-inspirados são técnicas de propósito geral. Isto é particularmente relevante quando o desenvolvimento de um algoritmo específico para o problema é inviável, deficiente ou se limita a instâncias menores, o que é o caso dos problemas tratados nesta tese. Entretanto, esses métodos possuem a desvantagem de que, uma vez que o conhecimento sobre o problema está implícito no algoritmo, provas formais de tempo de execução e /ou qualidade de aproximação são difíceis de se obter<sup>1</sup>. Por isso, é comum en-

<sup>1</sup> É importante mencionar, no entanto, a existência de esforços como a clássica teoria dos esquemas (HOLLAND, 1962).

contrarmos somente análises estatísticas sobre esses quesitos para a maioria das abordagens bio-inspiradas (NEUMANN; WITT, 2010), como, por exemplo, a análise markoviana do impacto de operadores evolutivos na possibilidade de convergência global (RUDOLPH, 1994; RUDOLPH, 1996).

Voltando a tratar especificamente de algoritmos imunológicos, eles possuem a distintiva vantagem sobre outras propostas bio-inspiradas de apresentar uma maior capacidade de manutenção de *diversidade* de soluções, o que os torna mais interessantes para a prevenção de convergência prematura e para a localização de múltiplos pontos ótimos (DE CASTRO; TIMMIS, 2002b; DE FRANÇA *et al.*, 2010b). Esta é a principal razão para que dois algoritmos imuno-inspirados fossem utilizados nas contribuições aos problemas abordados nesta tese, sendo seus detalhes esmiuçados a seguir.

### 4.2.1 O Algoritmo CLONALG

Na comunidade internacional de computação imuno-inspirada, um dos algoritmos mais empregados e divulgados é fruto do trabalho conjunto de dois notáveis pesquisadores brasileiros da área, os professores Leandro Nunes de Castro e Fernando José Von Zuben. Com mais de 1800 citações pelo *Google Scholar*<sup>2</sup>, o artigo de 2002 que apresenta o algoritmo CLONALG — *Learning and Optimization Using the Clonal Selection Principle* (DE CASTRO; VON ZUBEN, 2002) — pode ser considerado um texto fundamental para quem se propõe a trabalhar com SIAs. Tal sucesso se deve, ao menos em parte, à simplicidade de implementação e à flexibilidade para uso tanto em reconhecimento de padrões como em otimização, levando a um reconhecido desempenho em diversas aplicações ao longo dos anos.

Como o próprio nome já sugere, o algoritmo se baseia no princípio da Seleção Clonal para operar e, para compreender sua proposta, é preciso fazer algumas correspondências entre as entidades do sistema imune e aquelas relacionadas a um problema de otimização<sup>3</sup>. Temos uma população de elementos ocupando o espaço de busca, sendo que cada um deles (i.e. cada solução candidata) corresponde à estrutura de uma célula B associada a um anticorpo específico ( $Ab$ ), o qual podemos denominar somente célula. A função objetivo para este elemento em  $Ab$  representa a sua afinidade com um certo antígeno ( $Ag$ ). Passemos à apresentação, então, dos seus passos principais no Pseudocódigo 4.1.

O algoritmo inicia com uma população  $Ab$  de anticorpos ou células<sup>4</sup>, definida de maneira aleatória, sendo que cada  $Ab_i$  representa um elemento do espaço de parâmetros que

<sup>2</sup> Conforme aferição realizada em 5/9/2013.

<sup>3</sup> Embora o algoritmo tenha sido proposto também para problemas de reconhecimento de padrões, consideramos para os problemas da tese somente a abordagem para otimização.

<sup>4</sup> Assumimos, no contexto dos algoritmos imuno-inspirados, que os dois nomes são equivalentes.

---

**Pseudocódigo 4.1** Passos principais do algoritmo CLONALG para otimização.
 

---

**Parâmetros:**

- $nAB$ : número inicial de células;
- $nCl$ : número de clones por célula;
- $maxIT$ : número máximo de iterações;
- $b$ : percentual de novas células inseridas;

1- Crie uma população inicial de  $nAB$  elementos;

**while** (iteração  $\leq maxIT$ ) **do**

2- Avalie a afinidade (*fitness*) das células em  $Ab$ ;

3- Gere  $nCl$  clones para cada célula  $i$  em  $Ab$ ;

4- Aplique o operador de mutação em cada clone gerado;

5- Avalie a afinidade dos clones mutados;

6- Selecione entre o pai e os seus clones aquele que permanece na nova população;

7- Substitua as  $b$  piores células por novas geradas aleatoriamente;

**end while**

---

se pretende otimizar. No laço principal, a avaliação de afinidade implica em determinar o valor da função objetivo para cada  $Ab_i$ , o qual é subsequentemente clonado ( $nCl$  cópias por indivíduo). Então, os clones sofrem um processo de mutação que é inversamente proporcional à afinidade do seu pai, o que é geralmente implementado através da definição da probabilidade de mutação conforme a expressão

$$p_m = e^{-\rho f^{Ag}(Ab_i)}, \quad (4.1)$$

onde  $f^{Ag}(Ab_i)$  é o *fitness* da célula, escalonado no intervalo  $[0, 1]$ , e  $\rho$  é um parâmetro de ajuste do decaimento da curva exponencial. Realizada esta etapa, temos um novo conjunto  $C^*$  que corresponde aos clones que sofreram o processo de mutação, e que agora devem ter a afinidade avaliada a fim de, então, ocorrer o processo de seleção do melhor candidato entre a célula original e o grupo de clones que esta gerou. O laço termina com a criação de um percentual  $b$  de novos anticorpos para substituir os indivíduos de  $Ab$  com as piores afinidades. O processo assim se repete até o número  $maxIT$  de iterações ser atingido, quando então o elemento  $Ab_i$  que possui a maior afinidade é considerado a solução do problema de otimização em questão.

É possível reparar uma notável semelhança do algoritmo com os métodos evolutivos, como os algoritmos genéticos (BACK *et al.*, 2000): ambos são estratégias populacionais e compartilham do uso de operadores de seleção, avaliação de *fitness* e mutação. Entretanto, além da diferença quanto à inspiração, o algoritmo CLONALG se diferencia principalmente devido ao esquema de clonagem, à estratégia de mutação inversamente proporcional à afinidade e à inserção regular de novos candidatos na população. O resultado é que esta proposta de SIA possui, ao contrário de um GA clássico, a capacidade intrínseca de manter soluções

diversificadas durante a sua execução (DE CASTRO; TIMMIS, 2002b), o que pode ser muito relevante para incrementar a probabilidade de que o ótimo global ou um bom ótimo local seja encontrado.

### 4.2.2 O Algoritmo cob-aiNet

Como visto, o algoritmo CLONALG inspira-se exclusivamente no princípio da Seleção Clonal. Subsequentemente, outras propostas surgiram com o intuito de incluir a teoria da Rede Imunológica, pois ela permitiria acrescentar às técnicas a capacidade de ajuste dinâmico do número de indivíduos na população. O pioneirismo nesse sentido pertence ao algoritmo aiNet — *Artificial Immune Network* (DE CASTRO; TIMMIS, 2002a) —, que inaugurou uma família de algoritmos desenvolvidos, com o decorrer dos anos, sob a mesma inspiração (DE FRANÇA *et al.*, 2010a). O “espécime” mais recente desta família, apresentado em 2010, é o algoritmo cob-aiNet (COELHO; VON ZUBEN, 2010). O método apresenta uma nova formulação baseada no modelo atualizado de rede imunológica proposto por Bersini (2003). Além da versão original projetada para otimização contínua, também se propôs, em seguida, a cob-aiNet[C] (COELHO *et al.*, 2011), a fim de lidar especificamente com problemas combinatórios.

Trataremos de maneira geral do algoritmo cob-aiNet nesta seção, descrevendo os mecanismos principais que existem tanto na abordagem contínua como na combinatória. De maneira simplificada, primeiramente, o método é composto pelos mesmos mecanismos inspirados na Seleção Clonal: ele promove a evolução de uma população de células, que correspondem às soluções candidatas do problema, através de uma sequência de operações de clonagem, mutação e seleção, direcionadas por meio da qualidade de cada solução individual (afinidade com os antígenos). Então são incorporadas novas operações inspiradas na teoria da Rede Imunológica, que, em suma, consistem na comparação das células umas com as outras e, sempre que uma célula é semelhante a uma outra, de melhor *fitness*, acima de um pré-determinado limiar, ela tem sua *concentração* reduzida (o que corresponde à resposta negativa descrita na teoria da Rede Imunológica). Caso haja a contínua redução de concentração de uma célula, ao longo das iterações do algoritmo, isto pode levar à sua eliminação (supressão) quando este valor chegar a zero. A concentração de cada célula também pode aumentar (resposta positiva) conforme o seu *fitness*, o que revela a importância desta propriedade para o algoritmo: ela determina o número de clones a serem gerados, a intensidade do processo de mutação sofrido por cada clone e quando uma dada célula deve ser eliminada.

A estrutura geral da cob-aiNet é apresentada pelo Pseudocódigo 4.2. O leitor pode observar que o algoritmo é mais sofisticado que o seu “ancestral”, o CLONALG, por isso vamos discutir os seus passos principais com mais detalhes a seguir.

---

**Pseudocódigo 4.2** Passos principais do algoritmo cob-aiNet.
 

---

**Parâmetros:**

- $nAB$ : tamanho inicial da população;
- $maxAB$ : tamanho máximo da população;
- $nCl^{max}$ : número máximo de clones por célula;
- $nCl^{min}$ : número mínimo de clones por célula;
- $C_0$ : concentração inicial de cada célula;
- $\sigma_s$ : limiar de supressão;
- $\beta^i$ : parâmetro inicial do operador de mutação;
- $\beta^f$ : parâmetro final do operador de mutação;
- $LS_{it}$ : número de iterações de busca local;
- $LS_{freq}$ : frequência da busca local;
- $maxIT$ : número máximo de iterações;

- 1- Crie uma população inicial de  $nAB$  elementos;
  - 2- Avalie o *fitness* das células da população inicial;
  - 3- Avalie a afinidade entre as células da população inicial;
  - while** (iteração  $\leq maxIT$ ) **do**
    - 4- Defina o número de clones  $nCl_i$  que deve ser gerado para cada célula  $i$ ;
    - 5- Gere  $nCl_i$  clones para cada célula  $i$  da população;
    - 6- Aplique o operador de mutação a cada um dos clones gerados;
    - 7- Avalie o *fitness* das novas células geradas;
    - 8- Selecione as células que deverão ser mantidas na próxima geração (com inserção);
    - if** ((iteração **mod**  $LS_{freq}$ ) == 0) **then**
      - 9- Aplique o operador de busca local a todos os indivíduos e atualize os respectivos *fitness*;
    - end if**
    - 10- Avalie a afinidade entre as células da população;
    - 11- Atualize a concentração de todas as células;
    - 12- Elimine da população as células com concentração nula;
  - end while**
  - 13- Aplique o operador de busca local a todos os indivíduos e atualize os respectivos *fitness*;
  - 14- Avalie a afinidade entre as células da população;
  - 15- Atualize a concentração de todas as células;
  - 16- Elimine da população as células com concentração nula;
- 

#### 4.2.2.1 Representação e Métricas de Afinidade

Inicialmente, precisamos definir uma *representação* das soluções candidatas e a função objetivo a ser otimizada. Para propósitos gerais, vamos assumir a partir daqui que os candidatos a solução são vetores de valores reais  $\mathbf{u}$  e que o objetivo do algoritmo passa a ser a minimização de uma função objetivo  $f(\mathbf{u})$ .

Os mecanismos da cob-aiNet se baseiam em duas métricas de *afinidade*: *afinidade com*

*antígenos* (ou *fitness* –  $f^{Ag}(\mathbf{u}_t)$ ), definição esta que já conhecemos do algoritmo CLONALG, e a nova *afinidade com anticorpos* (ou *afinidade com outras células* –  $f^{Ab}(\mathbf{u}_t)$ )<sup>5</sup>. Estas medidas são definidas pelas Equações 4.2 e 4.3, e são avaliadas para cada célula representada por  $\mathbf{u}$  em cada iteração  $t$ . Matematicamente, temos:

$$f^{Ag}(\mathbf{u}_t) = 1 - \overline{f(\mathbf{u}_t)}, \quad (4.2)$$

$$f^{Ab}(\mathbf{u}_t) = \begin{cases} \frac{\sum_{\mathbf{v}_t \in J} C(\mathbf{v}_t)[\sigma_s - d(\mathbf{u}_t, \mathbf{v}_t)]}{\sum_{\mathbf{v}_t \in J} C(\mathbf{v}_t)}, & \text{se } J \neq \emptyset, \\ 0, & \text{caso contrário,} \end{cases} \quad (4.3)$$

em que  $f^{Ag}(\mathbf{u}_t) \in [0, 1]$  é o *fitness* da célula  $\mathbf{u}$  na iteração  $t$  e  $\overline{f(\mathbf{u}_t)} \in [0, 1]$  é o valor escalonado da função objetivo para a célula  $\mathbf{u}$  na iteração  $t$ . Na Equação 4.3,  $f^{Ab}(\mathbf{u}_t)$  é a afinidade total entre a célula  $\mathbf{u}_t$  e todas as outras células da população na iteração  $t$ ,  $J$  é o conjunto de células que não são piores que  $\mathbf{u}_t$  ( $f^{Ag}(\mathbf{v}_t) \geq f^{Ag}(\mathbf{u}_t)$ ) e que se localizam dentro de um raio  $\sigma_s$  (parâmetro do algoritmo) conforme uma medida de dissimilaridade  $d(\mathbf{u}_t, \mathbf{v}_t)$ , e, finalmente,  $C(\mathbf{v}_t)$  é a concentração da célula  $\mathbf{v}_t$  na iteração  $t$ .

Observe pela Equação 4.3 que somente células próximas de  $\mathbf{u}_t$  a partir de um dado limiar  $e$  melhores ou equivalentes a  $\mathbf{u}_t$  (aquelas pertencentes a  $J$ ) são consideradas no cálculo de  $f^{Ab}(\mathbf{u}_t)$ . Portanto, para determinar quem pertence a  $J$ , é necessário realizar comparações par-a-par entre *todas* as células da população. A métrica de distância  $d(\mathbf{u}_t, \mathbf{v}_t)$  deve ser escolhida de acordo com o problema, por isso a discutiremos mais detidamente no Capítulo 5.

#### 4.2.2.2 O Modelo de Concentração e Supressão

O modelo de concentração empregado na cob-aiNet é responsável pelo comportamento dinâmico do algoritmo, pois a concentração de cada célula na população afeta o número de clones que será gerado, a “intensidade” da mutação que cada clone sofrerá e também quando uma célula deve ser eliminada (sua concentração é, no caso, nula). A concentração  $C(\mathbf{u}_t) \in [0, 1]$  de uma célula  $\mathbf{u}_t$  na iteração  $t$  é definida inicialmente por  $C_0$  e é atualizada ao final de cada iteração de acordo com:

$$C(\mathbf{u}_{t+1}) = \max[0, \min[(\alpha C(\mathbf{u}_t) - f^{Ab}(\mathbf{u}_t)), 1]], \quad (4.4)$$

onde  $f^{Ab}(\mathbf{u}_t)$  é a afinidade total dada pela Equação 4.3 e  $\alpha$  é um fator regulatório definido por

$$\alpha = \begin{cases} 1 + 0.1 f^{Ag}(\mathbf{u}_t), & \text{se } f^{Ab}(\mathbf{u}_t) = 0, \\ 0.7, & \text{caso contrário.} \end{cases} \quad (4.5)$$

<sup>5</sup> Com o fim de evitar ambiguidade dos conceitos, passaremos a denominar a primeira métrica de *fitness* e a segunda somente de *afinidade*.

Note pela Equação 4.4 e Equação 4.5 que as melhores células nas suas respectivas vizinhanças (aquelas que possuem afinidade nula com as outras —  $f^{Ab}(\mathbf{u}_t) = 0$ ) sofrem um acréscimo de até 10% na concentração (*resposta positiva*), enquanto células que estão na vizinhança de outras que representam melhores soluções acabam sofrendo um decréscimo na concentração de pelo menos 30% (*resposta negativa*).

#### 4.2.2.3 Clonagem e Mutaç o

Embora as etapas de clonagem e muta o da cob-aiNet baseiem-se na Sele o Clonal assim como o algoritmo CLONALG, h  uma s rie de diferen as sutis nesta nova proposta. Inicialmente, o n mero de clones deixa de ser fixo para cada c lula e passa a depender da sua concentra o: a cada itera o  $t$  um dado n mero de clones  $nCl(\mathbf{u}_t)$    criado para cada c lula  $\mathbf{u}$  da popula o, conforme a express o

$$nCl(\mathbf{u}_t) = C(\mathbf{u}_t)(nCl^{max} - nCl^{min}) + nCl^{min}, \quad (4.6)$$

em que  $nCl^{max}$  e  $nCl^{min}$  s o os valores limitantes, definidos *a priori* pelo usu rio.

Ap s a clonagem, as novas c lulas sofrem um processo de muta o com variabilidade gen tica inversamente proporcional ao *fitness* e   concentra o da c lula “pai”, dessa forma os descendentes vindos de indiv duos melhores e mais isolados s o menos modificados do que aqueles originados de c lulas piores e/ou mais agrupadas. A proposta sugere que sejam implementados dois ou mais operadores de muta o — sendo um escolhido aleatoriamente para cada c lula  $\mathbf{u}$  a cada itera o  $t$  — e o n mero de vezes  $n^{mut}(\mathbf{u}_t)$  que se aplica   definido por:

$$n^{mut}(\mathbf{u}_t) = \max[\lfloor \beta(t)e^{-f^{Ag}(\mathbf{u}_t) \cdot C(\mathbf{u}_t)} \rfloor, 1], \quad (4.7)$$

na qual  $\beta(t)$    o valor do par metro  $\beta$  na itera o  $t$ . O controle da intensidade das muta es passa a ser exercido, ent o, pela quantidade de aplica es do operador sobre o indiv duo, diferentemente do esquema utilizado pelo CLONALG e que est  indicado na Equa o 4.1.

O par metro  $\beta(t)$  influencia indiretamente a amplitude de modifica es realizadas nos clones e varia em fun o de  $t$  conforme uma fun o sigmoideal espelhada, ilustrada na Figura 6. Este comportamento representa um dos aspectos principais do algoritmo, que   a mudan a de um est gio altamente explorat rio no in cio para uma fase de refinamento (*exploitation*) no final da execu o.

Adicionalmente, o m todo tamb m adota um operador de busca local com a finalidade de refinar o conjunto de solu es da popula o, e que deve ser escolhido conforme o problema em quest o. Por exemplo, para o caso de um problema de otimiza o real, Coelho e Von Zuben (2010) sugerem o uso do operador de busca local *LocalSearch1*, proposto no algoritmo MTS

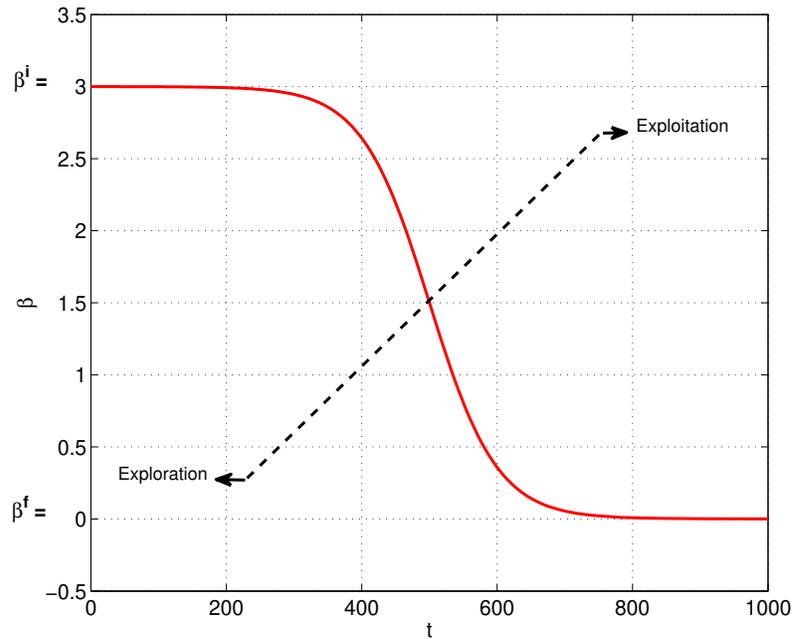


Figura 6 – Ilustração da variação sigmoideal espelhada de  $\beta$  conforme as iterações da cob-aiNet. Este gráfico foi extraído de (COELHO; VON ZUBEN, 2010) com os parâmetros  $\beta^i = 3.0$ ,  $\beta^f = 0.001$  e  $maxIT = 1000$ .

de otimização global de larga escala (TSENG; CHEN, 2008). Esta técnica opera em cada dimensão da solução candidata de maneira sequencial e independente, através do acréscimo ou diminuição dos seus valores conforme um *intervalo de busca*, que é dinamicamente alterado quando a modificação não leva a uma melhoria da solução.

#### 4.2.2.4 Seleção e Inserção de Novas Células

Após a clonagem e a mutação, o mecanismo de *seleção* entra em cena, sendo aplicado a cada subconjunto de células formado pelo “pai” e seus clones modificados. O aspecto distintivo desta operação no caso da cob-aiNet é que ela também é responsável pela inserção de novas células na população. Sempre que um clone modificado pela mutação está posicionado a uma distância (segundo a métrica de dissimilaridade escolhida) maior do que o valor do limiar de supressão  $\sigma_s$ , em relação à célula progenitora e às outras da população, ele é inserido junto com o “pai” desde que o tamanho da população permaneça menor ou igual a  $maxAB$ .

Este operador de inserção associado ao esquema de supressão de células previamente discutido permite que a cob-aiNet ajuste dinamicamente o tamanho da sua população durante a execução, o que é um caráter peculiar dos SIAs inspirados na teoria da Rede Imunológica.

### 4.3 Conclusões

Este capítulo apresentou uma breve introdução à computação imuno-inspirada, que se vincula ao conceito de Sistema Imunológico Artificial. O sistema imunológico dos vertebrados é a origem do modelo geracional de uma série de algoritmos e ferramentas computacionais que podem ser aplicados aos mais diversos problemas como aprendizado de máquina, clusteração, reconhecimento de padrões, otimização, entre outros (DE CASTRO; VON ZUBEN, 2002).

Apresentamos os elementos constituintes e mecanismos básicos do sistema imunológico, que formam uma entidade extremamente complexa e poderosa no sentido de apresentar as faculdades de memorizar, aprender com, reconhecer e agir contra os agentes invasores do organismo ao qual pertence. Assim, o sistema opera adaptativamente através da contínua defesa contra patógenos e melhoria dos seus próprios mecanismos. As duas teorias principais que explicam essa forma de operar foram discutidas também neste capítulo, iniciando com a teoria da Seleção Clonal e encerrando com a teoria da Rede Imunológica.

A metade final do capítulo foi dedicada aos paradigmas artificiais, tendo sido introduzidos os dois algoritmos empregados nas contribuições desta tese, o CLONALG e a cob-aiNet. Enquanto a primeira técnica já pode ser considerada uma abordagem canônica na área de SIAs, a segunda é bastante recente e incorpora vários mecanismos sofisticados de auto-regulação da população e do caráter de busca do algoritmo ao longo da sua execução.

À primeira vista, pode soar inadequado o uso do CLONALG em detrimento da cob-aiNet nas aplicações atuais, mas essa afirmação é incorreta, pois a nova abordagem é mais sofisticada ao preço de um custo computacional maior, advindo da incorporação dos novos operadores inspirados no modelo de rede imunológica. Dessa forma, o emprego de cada técnica deve ser decidido conforme a especificidade do problema em questão, avaliando-se os aspectos de tempo para determinação das soluções e complexidade do problema, que aí pode levar à escolha de uma técnica mais enxuta ou de uma abordagem mais poderosa em termos de capacidade de gerar soluções diversas.

A exposição do arcabouço teórico subjacente aos problemas a serem abordados e ao ferramental para se construir soluções calcadas no paradigma de ITL se encerra com este capítulo. No próximo capítulo, passaremos a tratar das contribuições realizadas nesta tese, iniciando por aquelas relacionadas ao problema de separação de sinais em corpos finitos.

Parte II

Contribuições



## 5 Contribuições à Separação de Sinais em Corpos Finitos

Neste capítulo, apresentaremos as contribuições desenvolvidas nesta tese para abordar o problema de separação cega de sinais em corpos finitos. Tais propostas abrangem: (i) o desenvolvimento de novos algoritmos baseados em otimização via computação imuno-inspirada e critérios baseados na TI, (ii) a melhoria de um algoritmo já existente e (iii) a generalização do problema original para o contexto de misturas convolutivas.

### 5.1 Otimização Baseada no CLONALG e no Critério de Mínima Informação Mútua

Como vimos no Capítulo 2, quando o problema de ICA é formulado para sinais em corpos de Galois, o processo de encontrar uma matriz de separação  $\mathbf{W}$  para recuperar as fontes é, necessariamente, um problema de otimização combinatória. Uma consequência disto é que os algoritmos baseados na busca com algum caráter exaustivo — como os métodos AMERICA e MEXICO (GUTCH *et al.*, 2012) — tendem a se tornar proibitivos à medida que a dimensão do problema aumenta.

Tendo isto em vista, desenvolvemos uma técnica de ICA em corpos finitos que se baseia em duas novas características (SILVA *et al.*, 2011): (i) o uso da informação mútua no lugar da entropia como a medida de TI para quantificar independência — o que permite uma recuperação direta de todas as fontes, em vez da extração uma a uma e (ii) o uso de um algoritmo imuno-inspirado, o CLONALG (DE CASTRO; VON ZUBEN, 2002), para evitar a necessidade de recurso a um método de otimização exaustivo. Esta proposta será testada no contexto de  $GF(2)$  para diversos conjuntos de fontes, sendo seu desempenho comparado com o do algoritmo AMERICA.

Em termos matemáticos, retomemos os vetores<sup>1</sup>  $\mathbf{x}$  e  $\mathbf{y}$  do problema de ICA apresentado na seção 2.3 do Capítulo 2. Temos que  $\mathbf{y} = f(\mathbf{x}) = \mathbf{W}\mathbf{x}$  e  $\mathbf{W}$  é uma matriz inversível  $N \times N$ , definidos em  $GF(2)$ . Utilizando a propriedade apresentada na Equação 2.5, a PMF de  $\mathbf{y}$  é:

$$p_Y(\mathbf{y}) = p_X(\mathbf{W}^{-1}\mathbf{y}) \quad (5.1)$$

<sup>1</sup> Neste capítulo, para simplificar a notação e dado que estamos tratando do problema de ICA/BSS instantâneo, ignoraremos o termo referente ao tempo nas variáveis.

Graças ao teorema de identificação em ICA (Teorema 2.1), é possível realizar a separação dos sinais através da minimização da informação mútua entre os componentes de  $\mathbf{y}$

$$I(y_1; \dots; y_N) = I(\mathbf{y}) = \sum_{i=1}^N H(y_i) - H(\mathbf{y}). \quad (5.2)$$

Repare que a Equação 5.1 implica que a PMF de  $\mathbf{y}$  equivale à PMF de  $\mathbf{x}$ , tendo sido somente o conjunto domínio modificado por  $\mathbf{W}$ . Portanto, tem-se que  $H(\mathbf{y}) = H(\mathbf{x})$  e, consequentemente

$$I(\mathbf{y}) = \sum_{i=1}^N H(y_i) - H(\mathbf{x}). \quad (5.3)$$

A entropia conjunta  $H(\mathbf{x})$  é invariante com relação aos parâmetros do problema (os elementos de  $\mathbf{W}$ ): portanto, é possível realizar a busca minimizando somente o primeiro termo da Equação 5.3, o que nos leva ao problema de otimização

$$\min_{\mathbf{W}} \left[ \sum_{i=1}^N H(y_i) \right], \quad (5.4)$$

$\mathbf{y} = \mathbf{W}\mathbf{x}$ ,  $\mathbf{W}$  é não-singular.

Esta formulação é bastante familiar na pesquisa em ICA para valores reais (HYVARINEN *et al.*, 2001; COMON; JUTTEN, 2010), mas a sua adoção para sinais finitos e discretos é inédita. Já sabemos, através da Equação 2.6 no Capítulo 2, que a cardinalidade do conjunto de soluções candidatas cresce proporcionalmente a  $2^{N^2}$  para  $GF(2)$ , o que faz com que a aplicação de algoritmos baseados em busca exaustiva se torne proibitiva a partir de uma certa dimensão.

Por isso, é promissora a estratégia de lidar com um problema desta natureza por meio de algoritmos bio-inspirados, como os sistemas imunológicos artificiais, introduzidos no Capítulo 4. Eles são métodos de propósito geral que combinam mecanismos de busca global e local com a finalidade de prover soluções boas, em um tempo de execução viável, a problemas de otimização com domínios de caracterização árdua ou de grande dimensionalidade. Neste contexto, mostra-se adequado o emprego do algoritmo CLONALG, que combina as características previamente mencionadas e apresenta uma atrativa simplicidade de implementação.

Como não é conhecida, *a priori*, a PMF das misturas, o cálculo exato dos valores de entropia expressos na Equação 5.4 é inviável. Em seu lugar, podemos empregar uma técnica de estimação a partir das amostras disponíveis. Supondo haver um conjunto de  $T$  observações iid para cada mistura,  $\{x_i(1), \dots, x_i(T)\}$ , escolhemos o estimador de máxima verossimilhança (ML, do inglês *Maximum Likelihood*) (PANINSKI, 2003) para calcular a entropia de cada

componente:

$$\begin{aligned}\hat{H}(y_i) &= - \sum_{j=0}^{P-1} \hat{p}_{y_i}(j) \log_P \hat{p}_{y_i}(j), \\ \hat{p}_{y_i}(j) &= \frac{1}{T} \sum_{n=1}^T \mathbf{1}_j(y_i(n)),\end{aligned}\tag{5.5}$$

na qual  $P$  é a ordem do corpo finito e  $\mathbf{1}_j(\cdot)$  é a função indicador

$$\mathbf{1}_\alpha(x) = \begin{cases} 1, & \text{se } x = \alpha, \\ 0, & \text{caso contrário.} \end{cases}\tag{5.6}$$

Adotamos o logaritmo de base  $P$  a fim de ter a entropia sempre com valor máximo 1, independentemente da cardinalidade.

Para o CLONALG resolver o problema posto pela Equação 5.4, a função objetivo (*fitness*) é definida por

$$\begin{aligned}g_{MI}(\mathbf{y}) &= 1 - \frac{1}{N} \sum_{i=1}^N \hat{H}(y_i), \\ \mathbf{y} &= \mathbf{W}\mathbf{x},\end{aligned}\tag{5.7}$$

onde cada indivíduo da população representa um possível conjunto de valores para  $\mathbf{W}$  e  $\hat{H}(y_i)$  é a entropia estimada de um componente de  $\mathbf{y}$ . Repare que a maximização da função de fitness dada pela Equação 5.7 é equivalente ao problema de minimização original descrito na Equação 5.4.

Para finalizar a descrição da proposta, são empregadas algumas modificações na implementação, com base nas análises de resultados obtidos em simulações preliminares: (i) uma função heurística foi incluída — para ser executada a cada 100 iterações — para operar sobre os possíveis valores de cada linha da matriz que um indivíduo representa, buscando a primeira modificação que melhora o *fitness* e mantém a restrição de não-singularidade; (ii) a substituição dos piores indivíduos por células novas geradas aleatoriamente ocorre somente a cada 100 iterações; e (iii) a fase de mutação é implementada com o operador de *bit random resetting* (BACK *et al.*, 2000), sendo incluída a restrição de só serem aceitas mutações que resultem em uma matriz não-singular.

### 5.1.1 Simulações

Uma série de simulações em MATLAB foi realizada para comparar o desempenho, em  $GF(2)$ , da nova proposta (chamada nos gráficos de CLONALG) com o do algoritmo AMERICA (YEREDOR, 2011a) — a técnica canônica que apresenta os resultados mais consistentes. Os parâmetros do algoritmo imunológico, definidos após um conjunto de testes

preliminares, são:  $nAB = 5N$ ;  $nCl = 0.1(nAB)$ ;  $maxIT = 2000$ ;  $b = 0.2(nAB)$ . O cenário da simulação compreende 200 execuções com fontes binárias de distribuição aleatoriamente definida, i.e. cada execução baseia-se em um valor aleatório para as probabilidades  $p(0)$  e  $p(1)$ , 4 valores distintos ( $T$ ) de observações para estimação do *fitness* (100, 1000, 5000 e 10000 amostras) e 3 quantidades distintas de fontes ( $N$ ): 2, 4 e 8.

A Figura 7 exhibe o número médio de fontes recuperadas<sup>2</sup> para o caso com  $N = 2$ . Note que, quando  $T = 100$ , o algoritmo AMERICA supera o CLONALG, mas a partir de  $T = 1000$ , ambas as propostas atingem desempenhos similares. Ademais, para  $T = 10000$ , o caso no qual as melhores estimativas são empregadas, a nova proposta é levemente superior. De toda forma, para uma quantidade tão pequena de componentes, ambas as abordagens podem ser consideradas plenamente eficazes.

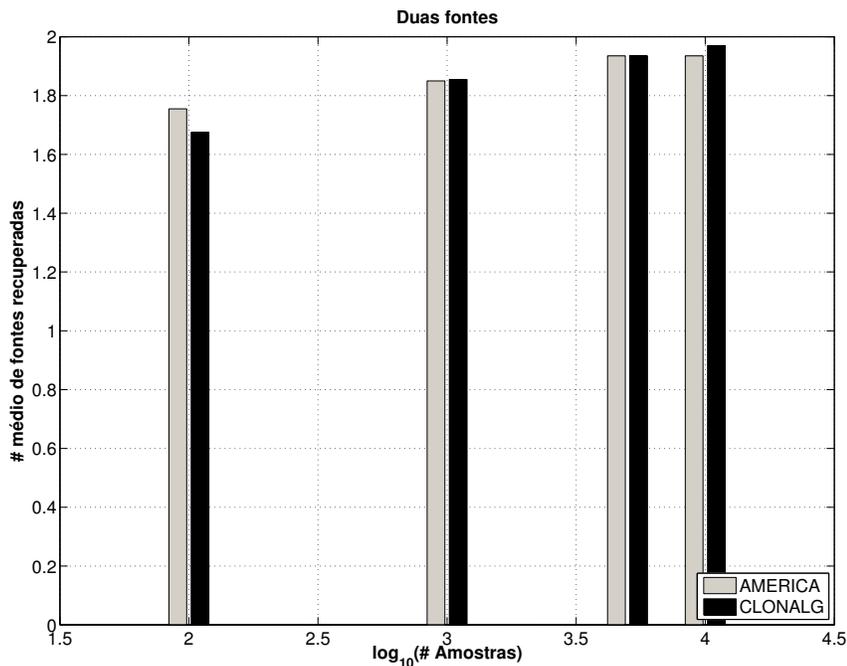


Figura 7 – Desempenho médio dos algoritmos de separação para o cenário com duas fontes.

Na Figura 8, comparamos os desempenhos para  $N = 4$  fontes; note que o comportamento é similar ao caso anterior: ambas as propostas apresentam resultados próximos até que uma leve vantagem para o CLONALG é observada com 10000 amostras. Já para o caso em que  $N = 8$ , apresentado na Figura 9, podemos notar que, para as três primeiras quantidades

<sup>2</sup> Para determinar quantas fontes foram recuperadas pelo CLONALG e por todos os métodos mencionados no restante do capítulo, adotamos o seguinte procedimento: o algoritmo sugere uma matriz de separação  $\mathbf{W}$  e a qualidade de  $\mathbf{W}\mathbf{A}$  é mensurada através da contagem do número de linhas com estritamente uma entrada não-nula, o que indica o número de fontes extraídas.

de amostras, o método AMERICA possui vantagem, enquanto, com a quantidade máxima, os dois métodos são equivalentes na prática.

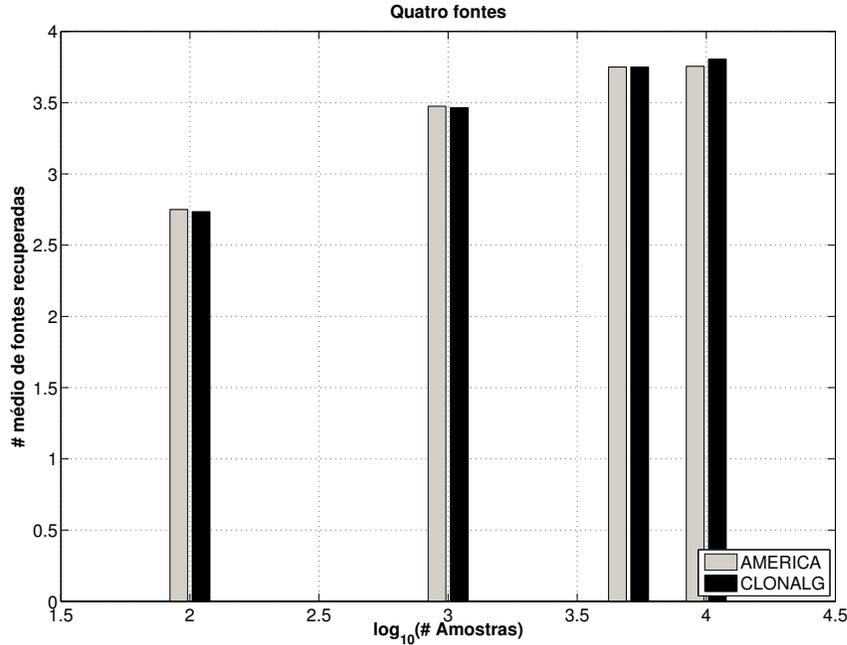


Figura 8 – Desempenho médio dos algoritmos de separação para o cenário com quatro fontes.

### 5.1.2 Considerações Sobre a Proposta

De uma perspectiva ampla, os resultados apontam que a proposta do CLONALG para ICA em  $GF(2)$  pode funcionar em um nível similar ao atingível pelo algoritmo AMERICA, a técnica estado-da-arte para este problema. Dado que a abordagem possui um caráter não-exaustivo, fica a perspectiva de aplicação desta em cenários de maior número de fontes, em que o custo exponencial do método AMERICA pode inviabilizar sua adoção.

## 5.2 Otimização pela cob-aiNet[C] e o Critério de Mínima Entropia

Após a formalização da proposta do algoritmo CLONALG para ICA em  $GF(2)$ , identificamos algumas possibilidades de desenvolvimento que poderiam levar a uma nova técnica de separação de sinais em corpos finitos, que é apresentada nesta seção (SILVA *et al.*, 2014). A essência desta segunda contribuição é utilizar uma estratégia de busca que mantém soluções diversas ao longo do processo, em conjunto com uma nova formulação do problema que é um meio termo entre (i) buscar a matriz inteira de separação e (ii) buscar iterativamente cada

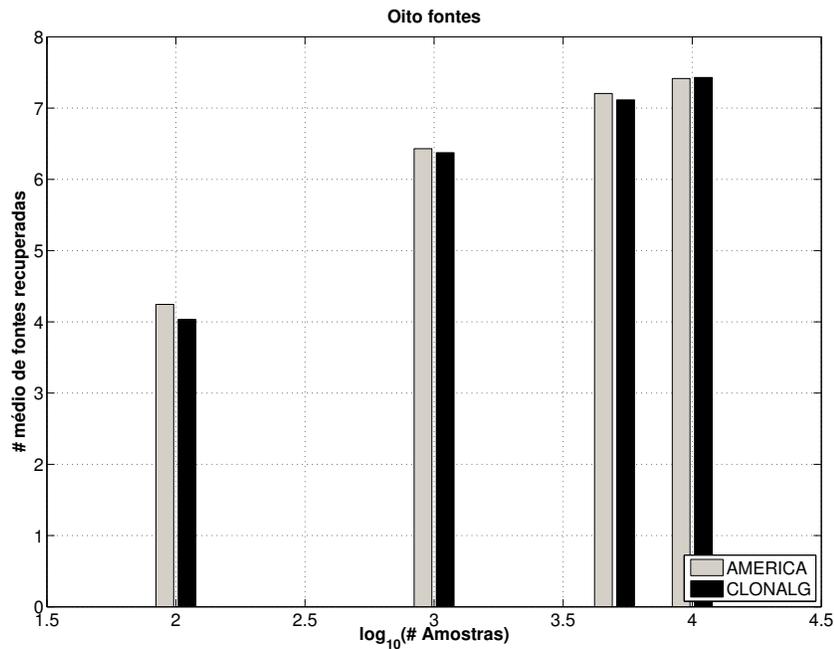


Figura 9 – Desempenho médio dos algoritmos de separação para o cenário com oito fontes.

linha (vetor de extração de um sinal) que a compõem. Adicionalmente, o problema passa a ser analisado para sinais em corpos primos, de ordem  $P$ .

Recordemos o Lema 2.1, que é válido para corpos primos e que atesta que a combinação linear de VAs independentes aumenta ou mantém igual a entropia da variável resultante. Este lema é uma propriedade chave, que viabiliza a construção de algoritmos que extraem iterativamente cada componente independente através da busca por um vetor de separação que produza um sinal de mínima entropia. Tal estratégia é análoga à abordagem por *deflação* em ICA para sinais e sistemas reais (DELFOSSE; LOUBATON, 1995).

A nova proposta considera esta propriedade, mas no contexto de um algoritmo de busca populacional cujo propósito geral é determinar, em uma população de soluções candidatas,  $N$  vetores *distintos* de separação que minimizam a entropia e que, ao final da busca, são utilizados para ser composta a matriz de separação.

Desta maneira, os indivíduos da população são candidatos a serem linhas da matriz e, com a progressiva melhora do conjunto todo, espera-se obter, após várias iterações, a solução final como sendo um subconjunto de  $N$  linhas candidatas que comporão  $\mathbf{W}$ . Esta idéia aproxima-se conceitualmente da abordagem *Michigan* para algoritmos genéticos, na qual a população inteira é considerada a solução do problema em questão (BACK *et al.*, 2000).

Matematicamente, o problema de otimização que esta proposta impõe é:

$$\begin{aligned} \min_{\mathbf{w}_i} \quad & H(y), \\ & y = \mathbf{w}_i \mathbf{x}, \end{aligned} \tag{5.8}$$

onde  $\mathbf{w}_i$  representa uma linha de  $\mathbf{W}$  e  $H(y)$  é a entropia do respectivo sinal extraído  $y$ . A fim de que se encontre uma matriz de inversão que seja composta por vetores que extraiam componentes *diferentes* das misturas, é necessário associar o objetivo descrito pela Equação 5.8 a um algoritmo de busca que mantenha diversidade na população. Caso contrário, o subconjunto final de soluções conterá diversas versões de uma mesma fonte. Para atender a esta propriedade fundamental, escolhamos o algoritmo cob-aiNet[C] (COELHO *et al.*, 2011) — a implementação da cob-aiNet para problemas combinatórios —, o qual é intrinsecamente capaz de manter diversidade na população, como apresentado no Capítulo 4.

Como visto anteriormente, não temos acesso às distribuições de probabilidade de  $\mathbf{x}$  e, conseqüentemente, de  $\mathbf{y}$ : por isso, novamente lançamos mão de técnicas de estimação, partindo de um conjunto de  $T$  observações iid das misturas. Propõe-se aqui substituir a entropia teórica na Equação 5.8 pelo já conhecido estimador de entropia ML, agora acrescido da correção de Miller-Madow (CARLTON, 1969):

$$\begin{aligned} \hat{H}(y) &= \frac{P-1}{2T} - \sum_{j=0}^{P-1} \hat{p}_y(j) \log_P \hat{p}_y(j), \\ \hat{p}_y(j) &= \frac{1}{T} \sum_{n=1}^T \mathbf{1}_j(y(n)). \end{aligned} \tag{5.9}$$

$P$  é a ordem do corpo primo e  $\mathbf{1}_j(\cdot)$  é a já conhecida função indicador (vide Equação 5.6).

É sabido que este estimador, mesmo com a correção, ainda possui *bias* e, por isso, muitas propostas para reduzir essa distorção podem ser encontradas na literatura (SCHÜR-MANN, 2004). Entretanto, isto não representa uma questão tão crucial em problemas de otimização, pois neles o interesse jaz somente nos pontos onde a função objetivo assume seus valores extremos (máximos ou mínimos). Portanto, somente avaliações *relativas* são necessárias, e os valores reais de entropia não são mandatórios (PRINCIPE, 2010).

Como já mencionamos, os indivíduos ou células da população representam candidatos a serem linhas da matriz  $N \times N$  de separação, i.e., são vetores  $1 \times N$  definidos em  $GF(P)$ . Seus valores de *fitness* são calculados através da Equação 5.9, submetida a uma transformação para operar como uma função objetivo de maximização, escalonada entre 0 e 1:

$$\begin{aligned} g_E(y) &= 1 - \frac{\hat{H}(y) - \hat{H}_{min}}{\hat{H}_{max} - \hat{H}_{min}}, \\ & y = \mathbf{w} \mathbf{x}, \end{aligned} \tag{5.10}$$

onde  $\hat{H}_{min}$  e  $\hat{H}_{max}$  são, respectivamente, os valores mínimo e máximo de entropia calculados entre os indivíduos da população, na iteração em questão. Ademais, os vetores escolhidos ao final do processo precisam formar uma matriz de separação não-singular, devido à especificação do problema de BSS. Para que isto ocorra, a `cob-aiNet[C]` é executada e, ao final da execução, um procedimento de busca gulosa é empregado para selecionar os  $N$  melhores indivíduos, em uma ordem decrescente de *fitness*, que sejam *linearmente independentes*. Repare que esta rotina é similar à adotada pelo algoritmo AMERICA — que seleciona os  $N$  vetores de mínima entropia e linearmente independentes em uma ordem não-decrescente (de entropia) —, com a diferença crucial de que o espaço de busca, em nossa proposta, restringe-se aos membros da população final, enquanto, no método AMERICA, todas as possibilidades de vetores são checadas.

É importante reforçarmos, novamente, que a proposta é viável graças à reconhecida capacidade do algoritmo `cob-aiNet[C]` de manter um conjunto diversificado de indivíduos na sua população. Mais ainda, essa propriedade é crucialmente afetada pela escolha de uma métrica de distância (ou dissimilaridade) adequada que compare os candidatos de uma maneira representativa da natureza do problema.

Após uma série de testes preliminares com diferentes abordagens, propomos a adoção da distância de Hamming entre as soluções candidatas, a qual consiste em contar as posições onde os dois vetores diferem. Isto implica na escolha de uma métrica que opera no espaço *genotípico* (dos vetores de extração).

Se pensarmos puramente em termos de recuperação dos sinais, seria preferível uma métrica que operasse no espaço *fenotípico* (dos sinais extraídos), e que discriminasse indivíduos que resultam em sinais independentes. Em outras palavras, se dois vetores distintos extraem a mesma fonte / componente ou fontes que são consideravelmente dependentes, do ponto de vista estatístico, a métrica de dissimilaridade deveria indicar esta situação com um valor pequeno. No entanto, esta opção é por ora rejeitada porque ela afeta severamente a velocidade do algoritmo:

- Quando empregamos a distância de Hamming, uma métrica com custo  $O(N)$ , a complexidade computacional do processo de avaliação da afinidade (vide subseção 4.2.2) entre todos os pares de células é  $O((nAB)^2 N)$  por iteração.
- Por outro lado, a complexidade computacional da alternativa fenotípica consiste na avaliação da informação mútua<sup>3</sup> entre os sinais extraídos por aquele par, a qual é  $O(NT + T^2 + P^2)$ , resultando em  $O((nAB)^2 (NT + T^2 + P^2))$  por iteração. Não obstante, é usual que  $T \gg N$ , o que reforça a opção pela distância de Hamming.

<sup>3</sup> Aqui consideramos o custo da informação mútua estimada pela abordagem por histograma.

Porém, a escolha da métrica genotípica traz um contratempo na definição do valor do limiar de supressão ( $\sigma_s$ ). Assuma o cenário onde as fontes são apresentadas diretamente ao algoritmo, i.e., em que não houve mistura. A solução para este caso é uma matriz de permutação onde qualquer linha tem distância de Hamming 2 em relação a todas as outras. Se a cob-aiNet[C] é configurada com  $\sigma_s > 2$ , a concentração de algumas células que compõem a solução correta decairá até que sejam eliminadas, dado que o algoritmo preferirá erroneamente matrizes com distâncias par-a-par maiores. Para evitar esta situação, recomendamos que  $\sigma_s$  seja ajustado com valores menores que 2.

Como já detalhamos na subseção 4.2.2 do Capítulo 4, a cob-aiNet[C] requer a escolha de operadores de mutação e busca local que se adequem ao domínio do problema. Neste caso, definimos uma heurística de busca local que percorre os  $N$  elementos de cada vetor candidato, testando se o acréscimo ou decréscimo de 1 (a unidade multiplicativa do corpo) ao seu valor melhora o *fitness*, quando isto ocorre, o elemento do vetor é atualizado e o procedimento passa a ser realizado com o próximo indivíduo da população.

Quanto aos operadores de mutação, os escolhidos são (i) a versão discreta da mutação *creep*, a qual escolhe aleatoriamente um elemento do vetor e adiciona (ou subtrai, o que também se define aleatoriamente) 1 ao seu valor; e (ii) a mutação de reinicialização aleatória (*random resetting*) (EIBEN; SMITH, 2007), a qual simplesmente sorteia um novo valor em  $GF(P)$  para um elemento do vetor, também escolhido aleatoriamente.

Encerrada a descrição do novo método, cujos passos principais são sumarizados no Pseudocódigo 5.1, podemos identificar algumas melhorias em comparação com os algoritmos previamente apresentados:

- Além de permitir a separação de sinais em qualquer corpo primo, a nova abordagem é mais eficiente que o primeiro método proposto nesta tese, visto na seção 5.1, porque (i) o tamanho dos indivíduos é proporcional a  $N$ , ao invés de  $N^2$ ; (ii) a função de *fitness* é a entropia do sinal extraído por um vetor candidato, o que é mais barato computacionalmente que a soma das entropias de todos os sinais; e (iii) o algoritmo cob-aiNet[C] possui mecanismos de controle automático do tamanho da população e um potencial de manutenção de diversidade superior ao do algoritmo CLONALG, que é a base da proposta prévia.
- A proposta não emprega técnicas de otimização exaustiva para obter as soluções, diferentemente dos algoritmos canônicos que foram propostos por Yeredor (2011a) e Gutch *et al.* (2010), o que efetivamente permite que ICA seja aplicado a cenários de BSS caracterizados por um número relativamente elevado de fontes, como na análise de fatores (COMON; JUTTEN, 2010) em grandes bases de dados.

---

**Pseudocódigo 5.1** Passos principais do algoritmo `cob-aiNet[C]` para ICA em  $GF(P)$ .

---

**Entrada:** conjunto de misturas  $\{\mathbf{x}(1), \dots, \mathbf{x}(T)\}$ , ordem do corpo  $P$ .

**Métricas:** função de *fitness*  $\hat{H}()$ , distância de Hamming como métrica de dissimilaridade.

- 1- Execute a `cob-aiNet[C]`, obtendo o conjunto  $AB$  de soluções (vetores de extração) com os seus respectivos valores de *fitness*.
- 2- Execute o algoritmo guloso para selecionar em  $AB$  os  $N$  melhores indivíduos (conforme o *fitness*) que sejam linearmente independentes, obtendo a matriz de separação  $\mathbf{W}$  e  $\mathbf{y}$ .

**Saída:**  $\mathbf{W}$  e  $\mathbf{y}$ .

---

### 5.2.1 Simulações

Para avaliar o desempenho da proposta, implementamos o algoritmo em MATLAB para realizar uma série de simulações. Foi definido um conjunto de parâmetros da `cob-aiNet[C]` com valores fixos, que são exibidos na Tabela 1, para todos os experimentos que seguem. A definição destes valores e daqueles mencionados nas subseções a seguir foi obtida através de testes preliminares do algoritmo, para cada cenário em questão.

Tabela 1 – Parâmetros gerais para todos os experimentos com a `cob-aiNet[C]`.

$maxAB$	2000
$nCl^{min}$	2
$nCl^{max}$	6
$\beta^i$	$0.8N$
$\beta^f$	$0.008N$
$LS_{it}$	1
$LS_{freq}$	45
$maxIT$	300

#### 5.2.1.1 Cenário 1: $P$ é Variável

Nesta primeira simulação, realizamos 100 execuções do algoritmo para quantidades distintas de amostras iid das misturas ( $\mathbf{x}$ ) — de  $T = 32$  a  $T = 4096$  amostras. Em cada execução, a distribuição de probabilidade de cada fonte é gerada aleatoriamente — sorteiam-se os valores das probabilidades  $p(0), \dots, p(P - 1)$  e, então, ela é amostrada —, assim como a matriz de mistura  $\mathbf{A}$ . Devido à restrição de que a técnica de ICA somente opera com distribuições não-degeneradas e não-uniformes, consideramos somente aquelas cuja divergência de Kullback-Leibler (vide Equação 1.12) com respeito à distribuição uniforme fosse maior ou igual a 0.2 e nas quais todas as probabilidades dos símbolos estivessem no intervalo  $(0, 0.98]$ .

Posto isto, analisamos um cenário com  $N = 8$ , fixo, e com números distintos de símbolos (ordem do corpo -  $P$ ): 2, 3, e 5. A primeira coluna da Tabela 2 descreve os parâmetros específicos do algoritmo para este primeiro experimento. Na Figura 10, podemos observar os resultados, dados pela *razão média de sucesso* (média amostral da razão entre o número de fontes extraídas e  $N$ ) para cada caso.

Tabela 2 – Parâmetros específicos para as simulações com  $N = 8$ ,  $N = 6, GF(5)$  e  $N = 12, GF(3)$ .

	$N = 8$	$N = 6, GF(5)$	$N = 12, GF(3)$
$n_{AB}$	200	200	500
$C_0$	0.5	0.9	0.9
$\sigma_s$	1.2	0.9	1.8

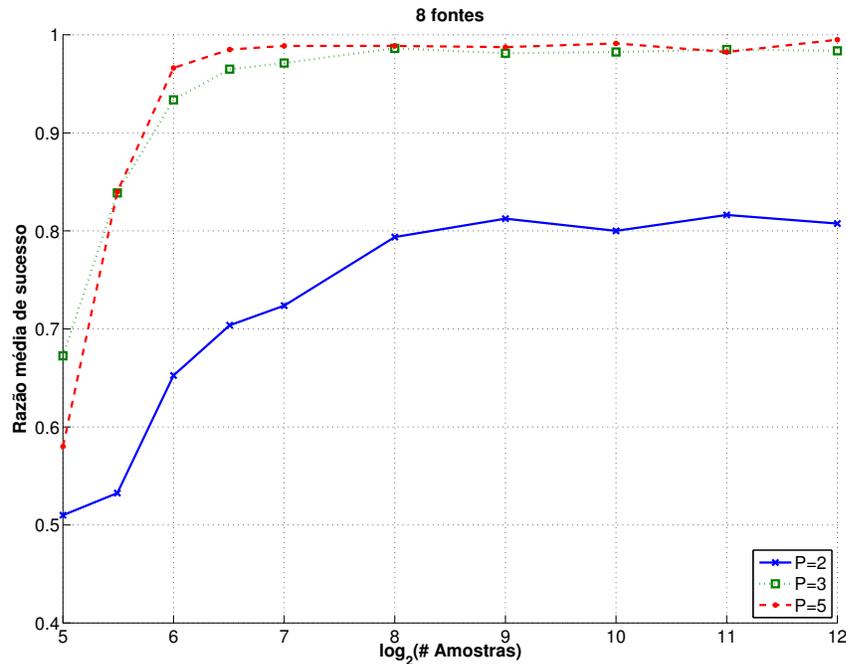


Figura 10 – Desempenho do algoritmo cob-aiNet[C] com  $N = 8$ .

Curiosamente, note que o caso mais simples,  $P = 2$ , leva a um desempenho inferior aos outros dois. Isto pode ser explicado por meio do fato de que o algoritmo não armazena uma base do espaço vetorial  $F^N$ ,  $F = GF(P)$ , e, conseqüentemente, emprega a distância de Hamming como medida de dissimilaridade entre as soluções: uma vez que essa métrica não é perfeitamente apropriada para indicar quais vetores de extração levarão a sinais diferentes (e independentes), podemos esperar que algumas soluções candidatas sejam erroneamente suprimidas durante a execução do algoritmo. Baseada na Equação 2.6, Waterhouse (1987)

prova que a probabilidade de que uma matriz  $N \times N$  em  $GF(P)$  seja singular (determinante nulo) é

$$Prob[N, P] = 1 - (1 - P^{-1})(1 - P^{-2}) \dots (1 - P^{-N}), \quad (5.11)$$

o que implica  $Prob[8, 2] \approx 0.7101$ , enquanto  $Prob[8, 3] \approx 0.4398$  e  $Prob[8, 5] \approx 0.2397$ . Isto significa que a supressão por engano de soluções candidatas quando  $P = 2$  reduz o já relativamente pequeno conjunto de matrizes de *rank* completo que poderia ser formado no final do procedimento. Em poucas palavras, podemos conjecturar que, quanto menor for a ordem do corpo, maior é o dano causado pela “imperfeição” da métrica de dissimilaridade.

Também podemos observar que, para todas as escolhas de  $P$ , há um desempenho crescente quando o número de amostras aumenta, até surgir um comportamento aproximadamente estável com  $T = 2^8 = 256$ .

#### 5.2.1.2 Cenário 2: Comparativo com os Algoritmos AMERICA e MEXICO

O segundo experimento visa comparar a abordagem imuno-inspirada com os algoritmos AMERICA e MEXICO (código disponível em (YEREDOR, 2011b)), que são os métodos com melhores taxas de sucesso entre as propostas apresentadas por Gutch *et al.* (2012). Realizamos novamente 100 execuções com 6 fontes geradas aleatoriamente em  $GF(5)$  e 12 fontes em  $GF(3)$ , seguindo os mesmos padrões de experimento da subseção 5.2.1.1. A segunda e terceira colunas da Tabela 2 apresentam os parâmetros específicos desta simulação.

A Figura 11 apresenta a razão média de sucesso nos dois casos mencionados. Pode-se ver claramente que o método AMERICA e a *cob-aiNet[C]* obtiveram um desempenho equivalente no cenário menor, separando todas as fontes em todas as rodadas com 256 ou mais amostras. No cenário maior, o algoritmo AMERICA exibe o melhor desempenho, enquanto a *cob-aiNet[C]* tem um resultado inferior, mas ainda assim satisfatório. Em ambos os cenários, o algoritmo MEXICO não conseguiu atingir a mesma qualidade nos resultados que os seus concorrentes, o que pode ser parcialmente explicado pelo fato de que a recuperação correta por esta técnica requer que  $P \leq 3$ .

Em suma, podemos afirmar que, em cenários caracterizados por um número reduzido de fontes e símbolos, o algoritmo AMERICA deve apresentar um desempenho melhor do que a *cob-aiNet[C]*, embora, como é visto nos resultados anteriores, esta também é capaz de obter uma qualidade bastante razoável. Entretanto, em cenários de maior dimensão, os métodos propostos por Gutch *et al.* (2012) compartilham um aspecto que restringe seu uso: com base nas observações do sinal, é estimada uma estrutura tensorial das probabilidades conjuntas a fim de acelerar a avaliação do critério empregado por cada técnica; o número total de elementos desta matriz é  $P^N$ , portanto a demanda por memória para executar os

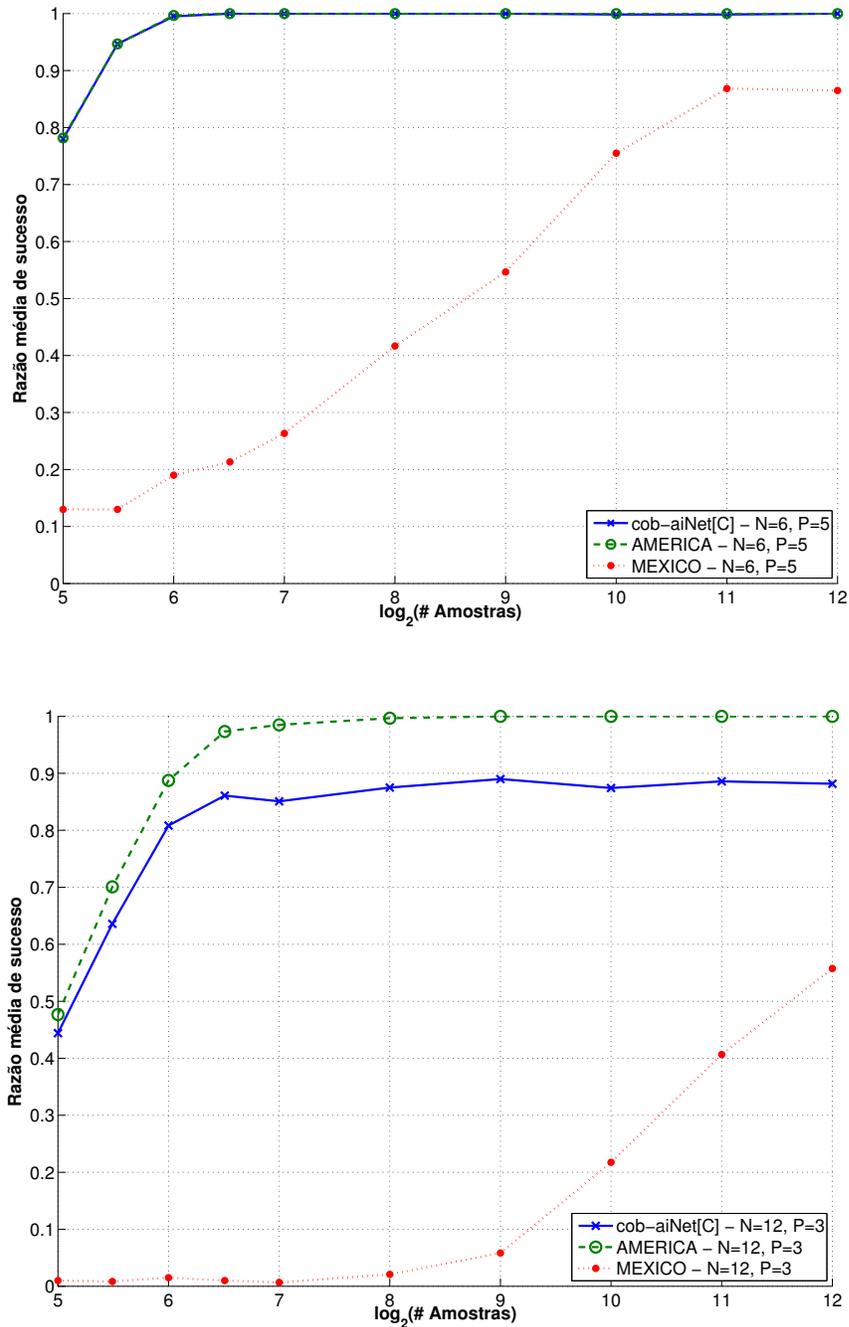


Figura 11 – Comparação da cob-aiNet[C] com outras técnicas: acima, caso com  $N = 6, P = 5$ , abaixo, o caso com  $N = 12, P = 3$ .

algoritmos cresce *exponencialmente* com o número de fontes.

Isto nos revela que há uma clara relação de compromisso para aplicar as ferramentas de ICA em GF, a qual envolve o tempo de execução *versus* os requisitos de memória. Se considerarmos a possibilidade de aplicar ICA em situações com um número maior de fontes ou símbolos, a abordagem imuno-inspirada é aquela que se apresenta como a mais escalável, dado o nosso conhecimento. Ademais, não se deve esquecer que, além do crescimento da demanda por memória, os algoritmos AMERICA e MEXICO também estão sujeitos a dificuldades vindas da explosão dimensional do espaço de busca, dado o caráter exaustivo que possuem. A proposta imunológica, embora não ofereça garantias no que diz respeito a convergência global, possui o potencial de prover ao menos soluções boas, até para instâncias mais complexas do problema.

Porém, mesmo possuindo uma maior escalabilidade, isto não implica que esta nova proposta esteja livre de melhorias futuras, como podemos observar nos resultados que a Figura 12 exhibe. Estes dados representam a razão média de sucesso da cob-aiNet[C] em um experimento de 50 rodadas com 14 fontes geradas aleatoriamente, em  $GF(3)$ , conforme os mesmos critérios definidos na subseção 5.2.1.1. Os parâmetros específicos desta simulação se encontram na Tabela 3.

Repare que, no melhor caso, obteve-se uma média de 7.34 fontes por execução (razão

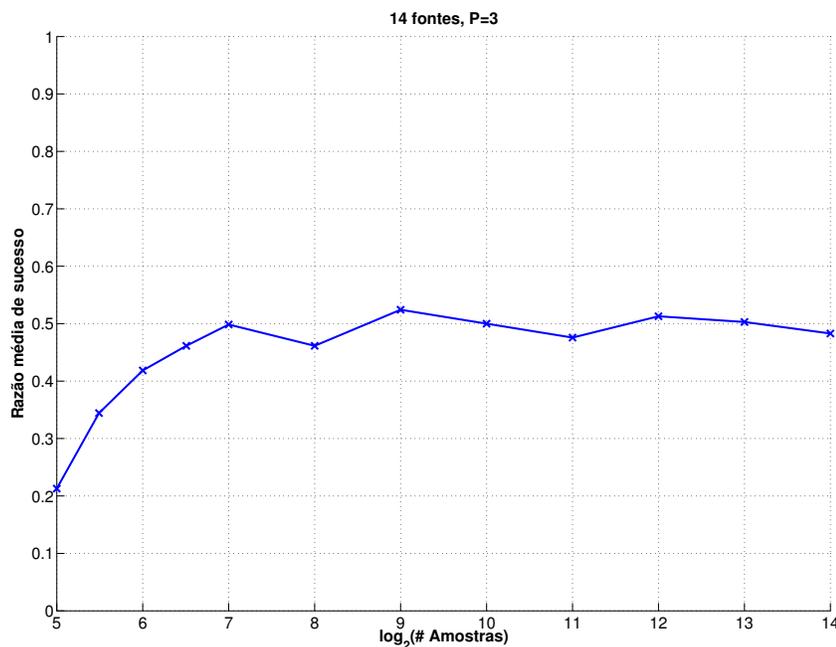


Figura 12 – Desempenho do algoritmo cob-aiNet[C] com  $N = 14$ ,  $P = 3$ .

Tabela 3 – Parâmetros específicos para a simulação com  $N = 14$ ,  $GF(3)$ .

$nAB$	1000
$maxAB$	3000
$C_0$	1.0
$\sigma_s$	0.14
$LS_{it}$	2

de 52.43%). Os resultados mostram que o aumento no tamanho da população (vide Tabela 3), em comparação ao primeiro experimento desta subseção, não permitiu alcançar um desempenho similar. É nítido que a complexidade do problema, que pode ser inferida pelo tamanho do espaço de busca, cresceu consideravelmente: considerando um cenário inicial com 8 fontes, saltamos de um espaço com  $3^8 = 6561$  possíveis vetores candidatos para um espaço com  $3^{14} = 4782969$  soluções. Isto é, ainda, sem considerar as possibilidades de combinações entre os vetores para formar a matriz de separação, o que resultaria numa ordem de grandeza ainda maior.

Para encerrar este segundo experimento, comparamos o custo computacional da proposta com o do algoritmo AMERICA. A operação mais custosa das duas técnicas é a estimação da entropia do sinal gerado pelo vetor de extração candidato, então analisaremos o número de ocorrências desta operação em ambos os métodos. O algoritmo MEXICO não é considerado porque, embora ele tenha uma complexidade computacional menor que o AMERICA, a qualidade dos resultados é consideravelmente inferior.

Na implementação do AMERICA considerada, ocorre uma busca exaustiva sobre todas as possibilidades (exceto a solução trivial) de vetor, portanto  $P^N - 1$  valores de entropia precisam ser estimados (YEREDOR, 2011a) ou, em notação *big-O*, o algoritmo AMERICA é  $O(P^N)$ . Por outro lado, o número de avaliações de *fitness* (entropia) que ocorrem durante a execução da proposta imuno-inspirada é  $O((nCI^{max} + N)(maxAB)(maxIT))$  no pior caso. Esta comparação indica, portanto, que a cob-aiNet[C] é assintoticamente menos custosa que o AMERICA no sentido de avaliações de entropia. Porém, os reais valores de custo do algoritmo imunológico podem variar devido a seu comportamento dinâmico. Por isso, para fazermos uma comparação prática, o valor analítico de custo do AMERICA é comparado com a média amostral sobre as rodadas de execução da cob-aiNet[C] em 5 cenários com  $P = 3$  e  $T = 512$  amostras, que o leitor pode analisar na Figura 13.

O algoritmo AMERICA possui os menores valores até  $N = 14$ , mas, como esperado, o custo da proposta de Yeredor cresce exponencialmente com o número de fontes, diferentemente da cob-aiNet[C], que cresce linearmente. Podemos, então, reafirmar que, com o crescimento de  $N$ , a técnica aqui proposta apresenta a escalabilidade mais adequada, considerando

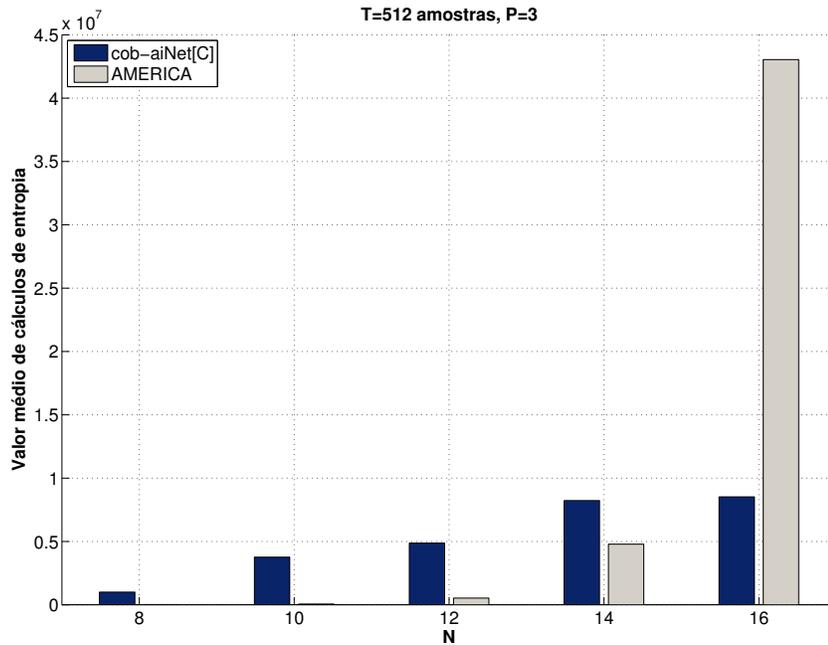


Figura 13 – Número de avaliações da entropia para quantidades distintas de fontes, entre a cob-aiNet[C] e AMERICA, com  $P = 3$ .

tanto as restrições de memória quanto as de tempo.

### 5.2.1.3 Cenário 3: Sensibilidade à Distância de Hamming

O terceiro experimento estuda a influência da distância de Hamming no comportamento do algoritmo quando há diferentes casos de misturas. Seja  $N = 5$ , com  $GF(2)$ , e considere, em um primeiro caso, que todas as fontes tem as suas amostras geradas a partir das probabilidades  $p(0) = 0.2, p(1) = 0.8$ . Considere também que haja três maneiras de se escolher uma matriz de mistura para aplicar às fontes:

1. uma matriz cuja inversa  $\mathbf{W}$  tem uma distância de Hamming (denominada de  $D_H$ ) mínima de 2 entre qualquer par de linhas;
2. uma matriz não-singular sem qualquer restrição adicional;
3. uma matriz cuja inversa possui ao menos  $m$  pares de linhas com  $D_H = 1$ .

Foram realizadas 100 rodadas em que o algoritmo era aplicado às misturas geradas por cada possibilidade de matriz, com diferentes valores de amostras. Os valores de parâmetros do experimento com  $N = 8, P = 2$  foram reaproveitados aqui, e podem ser revistos na Tabela 2.

O gráfico superior da Figura 14 exhibe os resultados deste primeiro conjunto de simulações. Nota-se que o algoritmo tem melhor desempenho com matrizes restritas a  $D_H \geq 2$  do que com matrizes não-singulares arbitrárias. Ademais, os conjuntos com um par e dois pares de linhas possuindo  $D_H = 1$  possuem uma crescente degradação da qualidade de separação, nesta ordem, o que é esperado pois  $\sigma_s = 1.2$ , o que neste caso leva o algoritmo a descartar linhas idênticas e aquelas que possuam uma posição diferente. Consequentemente, soluções que são compostas de pares de linhas tais que  $D_H = 1$  não podem ser totalmente formadas pelo algoritmo, neste cenário analisado.

Adicionalmente, o gráfico inferior da Figura 14 apresenta os resultados de um segundo conjunto, onde uma *única* matriz é escolhida para cada uma das possibilidades descritas previamente (exceto o caso sem restrição) e esta é então aplicada 100 vezes sobre 5 fontes que possuem a mesma distribuição de probabilidade, que agora é aleatoriamente definida a cada rodada. Esta segunda configuração exhibe um comportamento parecido com aquele que se observa no primeiro cenário, apesar da aleatoriedade das probabilidades.

A despeito de os parâmetros escolhidos não terem sido ajustados especificamente para este experimento, a análise relativa dos resultados ajuda a esclarecer ainda mais a mudança de desempenho que a técnica proposta pode apresentar quando é submetida a diferentes condições de mistura. Uma hipótese para minimizar tais efeitos é utilizar  $\sigma_s < 1$ , o que, consequentemente, reduz a diversidade de soluções, uma vez que o algoritmo só eliminaria vetores idênticos.

#### 5.2.1.4 Cenário 4: Análise Empírica de Estabilidade

Como outras metaheurísticas populacionais, os algoritmos imuno-inspirados são não-determinísticos, i.e., não há garantia de que, para a mesma entrada, a solução permanecerá constante. Dessa forma, o quarto e último experimento é uma análise empírica da capacidade do algoritmo de determinar a solução de ICA. Repetindo os respectivos parâmetros indicados na Tabela 2, 100 execuções da cob-aiNet[C] foram realizadas sobre um *único* conjunto de observações,  $T = 128$ ,  $N = 6$ ,  $GF(5)$ , gerados por uma matriz de mistura e fontes que obedecem às restrições expressas na subseção 5.2.1.2.

Em termos de número de fontes extraídas, os resultados mostram que a ferramenta apresenta um comportamento bastante estável: em 98 vezes de 100, o algoritmo resultou na separação total dos sinais. A média amostral é de 5.98, enquanto o desvio padrão é 0.1407. Este resultado indica que o método é capaz de prover soluções confiáveis de maneira consistente.

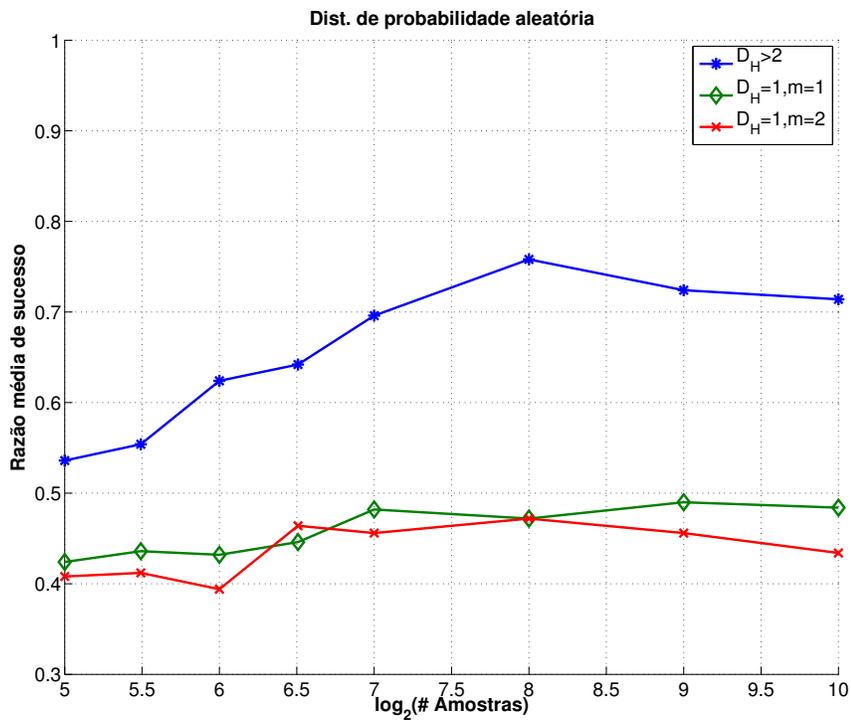
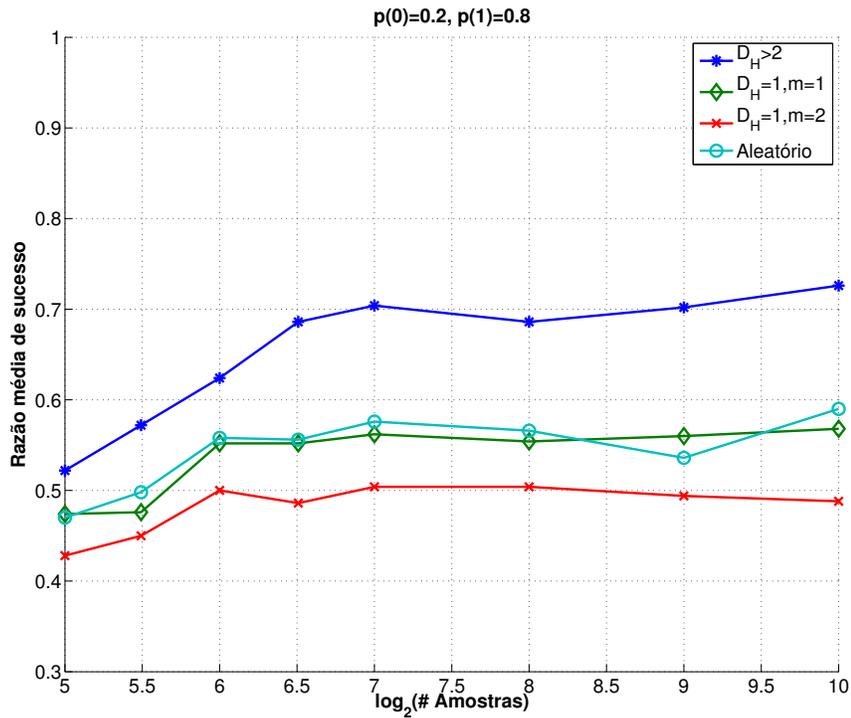


Figura 14 – Desempenho da cob-aiNet[C] em casos particulares de misturas: acima, caso com  $N = 5, P = 2, p(0) = 0.2, p(1) = 0.8$ ; abaixo, caso com  $N = 5, P = 2$  e perfil de probabilidade aleatório.

### 5.2.2 Considerações sobre a Proposta

Esta segunda proposta de metodologia imuno-inspirada para realizar ICA em corpos finitos foi uma evolução natural do primeiro trabalho realizado nesta tese, que se baseou no algoritmo CLONALG. Além de a nova implementação ser capaz de lidar também com corpos primos de ordem superior a 2, há novidades na modelagem do problema que a diferencia da técnica antecessora: utilizando-se a capacidade do algoritmo cob-aiNet[C] de manter diversidade entre as células durante a busca, a população pode ser evoluída a fim de manter soluções candidatas que extraíam componentes independentes *distintos*. Isto foi viabilizado através do emprego da entropia do sinal de saída como função custo e, encerrada a execução da cob-aiNet[C], a matriz completa de separação é montada a partir do subconjunto com os  $N$  melhores indivíduos linearmente independentes.

Os resultados dos experimentos mostraram que a nova técnica apresenta um desempenho muito bom — ainda que sub-ótimo — para quantidades pequenas de fontes e também fornece uma taxa de sucesso razoável para instâncias maiores. Este fato — associado à questão da explosão dimensional do espaço de busca — revelam que a proposta pode ser essencial em permitir a aplicação de ICA a BSS e análise de fatores no contexto de grandes bases de dados.

Este estudo, naturalmente, ainda deixa diversas perspectivas de melhoria para o futuro, como: (*i*) estender o algoritmo para corpos de ordem arbitrária; (*ii*) testar a ferramenta com novas métricas de dissimilaridade, em especial aquelas que quantifiquem diretamente o grau de dependência estatística entre os sinais extraídos e que não penalizem tanto o custo computacional; e (*iii*) analisar o desempenho do método com abordagens alternativas para estimação da entropia.

## 5.3 O Algoritmo MEXICO Modificado

As duas contribuições apresentadas nas seções anteriores possuem o paradigma comum de empregar a computação imuno-inspirada como método de busca global, a fim de encontrar propostas acuradas e, ao mesmo tempo, mais baratas computacionalmente. Tal estratégia, porém, não é exclusiva para se atingir esses objetivos: pode-se lançar mão de aprimoramentos dos algoritmos baseados em heurísticas de busca, como os pioneiros AMERICA, MEXICO e CANADA.

É neste contexto que desenvolvemos esta terceira contribuição, que diz respeito à modificação do algoritmo MEXICO, um método que pode ser visto como uma solução de compromisso entre o custoso, mas preciso, algoritmo AMERICA e o mais rápido, porém de

menor acurácia, algoritmo CANADA (GUTCH *et al.*, 2012). O algoritmo MEXICO original é baseado na extração iterativa de componentes por meio de comparações das entropias de pares de sinais, e a nova abordagem, que será denominada algoritmo MEXICO Modificado (MEXICO-m), apresenta-se capaz de, por meio de procedimentos diferentes de busca e de um novo critério de parada, obter uma melhora de desempenho sob as óticas de velocidade de convergência e taxa de separação das fontes (SILVA *et al.*, 2012; SILVA *et al.*, 2013).

Antes de apresentar a nova proposta, comecemos discutindo com mais detalhes o funcionamento da técnica original, a seguir.

### 5.3.1 O Algoritmo MEXICO

O algoritmo MEXICO, assim como o AMERICA, segue o princípio de extrair iterativamente os sinais por meio da redução dos respectivos valores de entropia. Porém, a construção da matriz de separação e o processo de busca ocorrem por meios bem distintos, pois, enquanto o AMERICA já avalia a entropia de cada sinal gerado pela combinação de até  $N$  misturas, o MEXICO faz a avaliação equivalente somente entre combinações de *pares* de misturas.

Mais especificamente, sejam  $x_i$  e  $x_k$  duas componentes (misturas) do vetor de observações  $\mathbf{x}$ . Primeiro, o algoritmo determina  $c \in GF(q)$  em  $\bar{x}_i = x_i + cx_k$  tal que  $H(\bar{x}_i)$  tenha valor mínimo, e, se  $H(x_i)$  é maior que a entropia da combinação,  $x_i$  é substituído por  $\bar{x}_i$ . Este processo é repetido para todas as  $N(N - 1)$  combinações, com  $i \neq k$ , o que caracteriza uma iteração ou “*sweep*”. Então, as iterações são repetidas até que ocorra um *sweep* sem que haja uma atualização (i.e., sem uma substituição que reduza a entropia), o que encerra o algoritmo.

Repare que a operação de substituição pode ser representada em termos de uma matriz com valores iguais a 1 na diagonal, o valor  $c$  na posição  $(i, k)$  e valor 0 nas posições restantes, o que é claramente uma transformação linear inversível. Além disso, essa é uma operação que reduz o valor da soma das entropias das componentes de  $\mathbf{x}$  e faz com que as duas componentes envolvidas na substituição se tornem “menos estatisticamente dependentes”, uma vez que a informação mútua  $I(\bar{x}_i; x_k) = H(\bar{x}_i) + H(x_k) - H(\bar{x}_i, x_k)$  se torna menor porque a entropia conjunta  $H(\bar{x}_i, x_k)$  permanece igual a  $H(x_i, x_k)$  (a transformação é inversível) (YEREDOR, 2011a).

De maneira prática, o algoritmo começa com uma estimativa inicial de matriz  $\mathbf{B} = \mathbf{I}_N$ , onde  $\mathbf{I}_N$  é uma matriz identidade  $N \times N$ , e as sucessivas operações de troca são registradas pelo produto à esquerda com  $\mathbf{B}$ . Quando o algoritmo termina,  $\mathbf{B}$  é a estimativa da matriz de separação  $\mathbf{W}$ . O Pseudocódigo 5.2 apresenta o passos principais do método. Relembre que

todas as entidades e operações presentes estão definidas em  $GF(q)$ .

---

**Pseudocódigo 5.2** Passos principais do algoritmo MEXICO.

---

```

B  $\leftarrow$  IN;
while parar=Falso do
  parar  $\leftarrow$  Verdadeiro;
  for all  $i \in \{1, \dots, N\}$  do
    for all  $j \neq i, j \in \{1, \dots, N\}$  do
       $c_0 = \arg \min_c H(x_i + cx_j)$ 
      HJ  $\leftarrow H(x_i + c_0x_j)$ ;
      if HJ <  $H(x_i)$  then
        B0  $\leftarrow$  IN; B0( $i, j$ )  $\leftarrow c_0$ ; B  $\leftarrow$  B · B0;
         $x_i \leftarrow x_i + c_0x_j$ ; parar  $\leftarrow$  Falso;
      end if
    end for
  end for
end while

```

---

### 5.3.2 Modificações do Algoritmo MEXICO

Nesta seção, propomos duas modificações para o algoritmo MEXICO que são, conforme as simulações realizadas em uma série de cenários representativos, capazes de trazer melhorias na velocidade de convergência e na taxa de sucesso, em relação ao algoritmo tradicional. Para justificar o *modus operandi* da nossa proposta, primeiro vamos revisar a questão da convergência do MEXICO original para mínimos locais e o seu critério de parada, à vista do que Yeredor (2011a) discute. O autor argumenta que há um risco de convergência local do seu método, com base em um exemplo em  $GF(2)$  com a seguinte matriz de mistura

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad (5.12)$$

e as fontes apresentam PMFs idênticas, de tal forma que não há combinação dois-a-dois de misturas que reduz a entropia. Neste caso, a entropia de cada mistura está diretamente vinculada ao número de valores 1 na correspondente linha da matriz. Se tentamos combinar quaisquer duas linhas, esta ação nunca resultará no decréscimo de valor da entropia: o número de 1s na nova linha será igual ou maior que da configuração prévia. Portanto, o algoritmo MEXICO se encerrará de maneira prematura neste cenário.

Interessantemente, esta situação pode ser contornada se aceitarmos aleatoriamente combinações lineares que resultem no mesmo valor de entropia. A implementação em MA-

TLAB do MEXICO, fornecida por Yeredor (2011b), de fato implementa essa idéia, mas a especificação do algoritmo no artigo original não a considera. No exemplo em questão, se aceitamos a troca da linha 2 pela sua combinação com a linha 3

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad (5.13)$$

seguido pela substituição da linha 1 pela sua combinação com a linha 2, temos

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad (5.14)$$

o que nos mostra que, por meio desta sequência de operações lineares, uma primeira fonte foi extraída e pode-se continuar com esta estratégia para extrair as componentes restantes.

Implicitamente, isto nos leva a um novo critério de parada para ser adotado no novo algoritmo, diferente do MEXICO padrão: o método pára somente quando todas as possíveis combinações dois-a-dois de misturas resultam no aumento de entropia ou, nos casos de empate, não se tenha aceitado aleatoriamente a troca. Observe também que a característica estocástica da regra para aceite de uma troca assemelha-se ao conceito chave da técnica de *Simulated Annealing* (KIRKPATRICK *et al.*, 1983) e, conseqüentemente, permite ao método escapar de soluções locais — embora não seja possível assegurar a separação perfeita em todos os casos devido a duas razões principais: (i) a proposta minimiza a independência entre pares de componentes, o que não implica na independência total, exceto para os casos de corpos em  $GF(2)$  e  $GF(3)$  (YEREDOR, 2011a); e (ii) há a limitação do número de amostras para se estimar a entropia, portanto flutuações do estimador podem levar a comparações erradas entre as misturas e, conseqüentemente, a soluções sub-ótimas.

Há uma segunda proposta de melhoria no método, que consiste em modificar a operação de troca: em vez de substituir  $x_i$  pela combinação  $x_i + cx_k$ , podemos trocar a “pior mistura”, i.e., aquela com o maior valor de entropia. Mais especificamente, o novo procedimento é

- Determine a componente, entre  $x_i$  e  $x_k$ , que possui maior entropia.
- Se  $x_i$  é a “pior” mistura, então determine  $c$  em  $\bar{x}_i = x_i + cx_k$  tal que  $H(\bar{x}_i)$  é mínima e substitua  $x_i$  se  $H(x_i)$  for maior.

- Se  $x_k$  é a “pior” mistura, então determine  $c$  em  $\bar{x}_k = cx_i + x_k$  tal que  $H(\bar{x}_k)$  é mínima e substitua  $x_k$  se  $H(x_k)$  for maior.

Esta mudança pretende acelerar a convergência do método através do aumento da possibilidade de substituição vinda da comparação dois-a-dois (há duas misturas candidatas a substituição, em vez de uma). Como consequência, espera-se que a convergência do algoritmo aconteça com um número menor de avaliações de entropia — o que é interessante, pois esta é a operação mais custosa da técnica. Se ambas as misturas apresentam o mesmo valor de entropia, em outras palavras, não há um candidato “pior”, também empregamos o princípio de escolher aleatoriamente uma das duas misturas para ser comparada. O Pseudocódigo 5.3 apresenta os passos da nova versão do algoritmo, com as duas propostas de mudança.

---

**Pseudocódigo 5.3** Passos principais do algoritmo MEXICO-m.

---

```

B ← IN;
while parar=Falso do
  parar ← Verdadeiro;
  for all  $i \in \{1, \dots, N\}$  do
    for all  $j \neq i, j \in \{1, \dots, N\}$  do
      if  $H(x_i) > H(x_j)$  then
         $k \leftarrow i; l \leftarrow j; HW \leftarrow H(x_i);$ 
      else if  $H(x_j) > H(x_i)$  then
         $k \leftarrow j; l \leftarrow i; HW \leftarrow H(x_j);$ 
      else
         $k, l \leftarrow \text{escolha-aleatória}(i, j); HW \leftarrow H(x_k);$ 
      end if
       $c_0 = \arg \min_c H(x_k + cx_l)$ 
       $HJ \leftarrow H(x_k + c_0x_l);$ 
      if  $HJ = HW$  then
        moeda ← escolha-aleatória(0, 1);
      end if
      if  $HJ < HW$  OR moeda=1 then
        B0 ← IN; B0( $k, l$ ) ←  $c_0$ ; B ← B · B0;
         $x_k \leftarrow x_k + c_0x_l$ ; parar ← Falso;
      end if
    end for
  end for
end while

```

---

### 5.3.3 Simulações

Realizamos uma série de simulações em MATLAB para avaliar o comportamento da nova proposta e se as modificações introduzidas promovem os efeitos esperados, em com-

paração ao algoritmo MEXICO padrão. Para simplificar a implementação computacional, continuamos a considerar experimentos realizados sempre com sinais em corpos primos. Para obter o valor da entropia, aplicamos novamente o estimador ML com correção de bias (vide Equação 5.9) sobre um conjunto de  $T$  observações iid das misturas.

### 5.3.3.1 Cenário 1: Desempenho Geral da Nova Proposta

O primeiro experimento tem o objetivo de avaliar exclusivamente o novo algoritmo MEXICO-m e se o seu funcionamento é adequado. O desenho do experimento assemelha-se àquele descrito para os testes com a cob-aiNet[C] na subseção 5.2.1.1: o algoritmo é executado 2000 vezes com um conjunto de  $N$  fontes de distribuição aleatória misturadas por uma matriz também definida aleatoriamente. As PMFs geradas atendem às mesmas restrições de não-uniformidade e não-degeneração observadas nos experimentos anteriores deste capítulo.

A Figura 15 exhibe o comportamento do algoritmo MEXICO-m para diferentes valores de  $N$  e  $q$ . O eixo  $y$  representa a razão média de sucesso e o eixo  $x$  representa o número de observações disponíveis, na escala logarítmica. Observe que, para o cenário com 3 símbolos, o algoritmo já possui, para  $N < 10$ , um bom desempenho com 1024 amostras, e há uma melhora significativa, em todos os casos, com o aumento dessa quantidade. Apesar do aparente comportamento assintótico das curvas, a separação perfeita em todas as rodadas nunca foi atingida, o que parece consequência das flutuações estatísticas vindas da estimação de entropia, que podem influenciar em um cenário formado por um espaço de busca com tantas possibilidades e possivelmente configurações entrópicas similares.

Para o caso de  $GF(5)$ , a situação torna-se mais complexa devido ao crescimento da dimensão do espaço de soluções — o que cria um cenário mais árduo do ponto de vista de estimação — e da inexistência de uma base teórica que assegure à técnica obter a separação total das fontes (vide subseção 5.3.2). É visível no gráfico que se requer mais de 4096 observações para obter taxas de sucesso superiores a 50% quando  $N = 8$ , enquanto que a diferença de desempenho entre cada curva é maior que aquela verificada no caso de  $GF(3)$ .

### 5.3.3.2 Cenário 2: Comparativo

Neste segundo experimento, passamos à análise comparativa entre o MEXICO-m, o algoritmo MEXICO original — implementado estritamente de acordo com (YEREDOR, 2011a; GUTCH *et al.*, 2012) — e a implementação feita em MATLAB fornecida por Yeredor (2011b), que considera a questão da troca de misturas com entropia igual, a qual denominamos de MEXICO2. Os resultados agora são obtidos sobre 500 execuções de cada proposta, para cada conjunto de amostras.

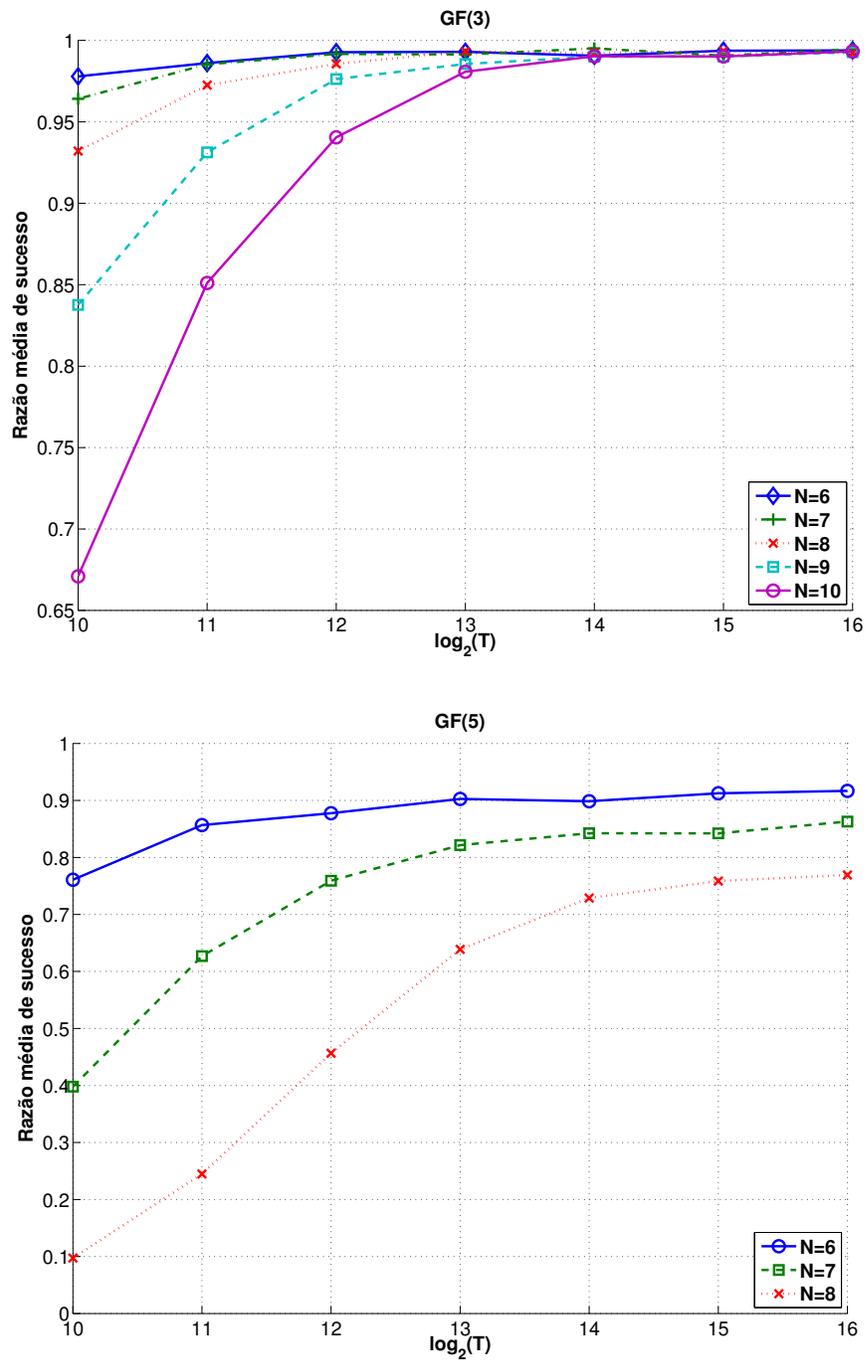


Figura 15 – Desempenho do algoritmo MEXICO-m: acima, sinais em  $GF(3)$ ; abaixo, sinais em  $GF(5)$ .

Começamos pela análise do número de vezes que cada método executa o cálculo das entropias e verifica se faz a substituição de uma mistura, com diferentes quantidades de fontes. A métrica em questão é aqui denominada de “Número de Avaliações de Troca” (NAT). Para evitar que ocorram resultados com convergência prematura devido a flutuações da estimação de entropia, o número de amostras é fixo em  $T = 65536$ . A Tabela 4 exibe a média amostral da métrica NAT para cada método e a razão entre os respectivos valores. Observe que o algoritmo MEXICO-m requer a menor quantidade de operações entre todos e que seu desempenho comparativo melhora com o crescimento da dimensão do problema, nas duas ordens de corpo. É notável, também, a considerável semelhança de resultados entre os algoritmos MEXICO e MEXICO2, o que indica que a presença (ou não) do sorteio para decidir a troca de misturas com entropia igual não influenciou na convergência.

Tabela 4 – Média amostral e desvio padrão da medida NAT entre as versões do algoritmo MEXICO.

$P$	$N$	MEXICO-m (a)	MEXICO (b)	MEXICO2 (c)	$a/b$	$a/c$
2	4	26.64 ( $\pm$ )4.98	35.03 ( $\pm$ )8.17	35.03 ( $\pm$ )8.17	0.76	0.76
	6	80.40 ( $\pm$ )15.12	127.31 ( $\pm$ )21.53	127.31 ( $\pm$ )21.53	0.63	0.63
	8	163.63 ( $\pm$ )23.51	308.97 ( $\pm$ )43.44	308.97 ( $\pm$ )43.44	0.53	0.53
	10	288.90 ( $\pm$ )39.67	619.61 ( $\pm$ )74.10	619.65 ( $\pm$ )74.10	0.47	0.47
	12	455.93 ( $\pm$ )68.39	1101.60 ( $\pm$ )110.96	1101.67 ( $\pm$ )110.86	0.41	0.41
	14	688.32 ( $\pm$ )94.75	1790.58 ( $\pm$ )163.93	1791.06 ( $\pm$ )164.38	0.38	0.38
3	16	984.00 ( $\pm$ )120.12	2728.94 ( $\pm$ )217.78	2730.34 ( $\pm$ )217.82	0.36	0.36
	4	29.30 ( $\pm$ )6.11	43.12 ( $\pm$ )8.05	43.12 ( $\pm$ )8.05	0.68	0.68
	6	89.52 ( $\pm$ )13.29	160.65 ( $\pm$ )23.14	160.65 ( $\pm$ )23.14	0.56	0.56
	8	197.01 ( $\pm$ )32.77	414.56 ( $\pm$ )49.38	414.56 ( $\pm$ )49.38	0.48	0.48
	10	373.14 ( $\pm$ )65.01	848.29 ( $\pm$ )88.83	848.29 ( $\pm$ )88.83	0.44	0.44
	12	644.16 ( $\pm$ )110.04	1522.49 ( $\pm$ )150.73	1522.49 ( $\pm$ )150.73	0.42	0.42

Analisemos a seguir a qualidade das soluções obtidas, para as mesmas 500 execuções independentes em um intervalo maior de amostras disponíveis para estimação da entropia. A Tabela 5 e a Tabela 6 sintetizam os resultados para diferentes quantidades de fontes ( $N$ ) e amostras ( $T$ ), em  $GF(2)$  e  $GF(3)$ , respectivamente.

Observa-se que, tanto em  $GF(2)$  como em  $GF(3)$ , todos os métodos possuem desempenho similar com 8 fontes, entretanto, com 12 e 16 fontes, o algoritmo MEXICO-m é vantajoso. Por isso, podemos afirmar, a partir dos dados, que a discrepância entre as propostas aumenta com o tamanho do problema, o que confirma a validade das modificações que foram propostas.

É necessário fazer uma menção aos altos valores de desvio padrão que encontramos

Tabela 5 – Média amostral e desvio padrão do número de fontes recuperadas entre as versões do algoritmo MEXICO,  $GF(2)$ .

$N$	$\log_2(T)$	6	8	10
8	<b>MEXICO</b>	5.64 ( $\pm$ )1.61	7.76 ( $\pm$ )0.76	7.99 ( $\pm$ )0.10
	<b>MEXICO2</b>	5.78 ( $\pm$ )1.55	7.80 ( $\pm$ )0.61	7.99 ( $\pm$ )0.11
	<b>MEXICO-m</b>	5.87 ( $\pm$ )1.55	7.81 ( $\pm$ )0.54	7.98 ( $\pm$ )0.25
12	<b>MEXICO</b>	4.82 ( $\pm$ )2.66	10.51 ( $\pm$ )2.54	11.85 ( $\pm$ )0.95
	<b>MEXICO2</b>	5.56 ( $\pm$ )2.72	10.79 ( $\pm$ )2.34	11.82 ( $\pm$ )1.03
	<b>MEXICO-m</b>	6.03 ( $\pm$ )2.67	11.13 ( $\pm$ )1.84	11.99 ( $\pm$ )0.19
16	<b>MEXICO</b>	1.78 ( $\pm$ )2.27	9.75 ( $\pm$ )5.45	14.82 ( $\pm$ )3.12
	<b>MEXICO2</b>	2.61 ( $\pm$ )2.54	10.23 ( $\pm$ )5.34	14.74 ( $\pm$ )3.15
	<b>MEXICO-m</b>	3.34 ( $\pm$ )3.10	11.24 ( $\pm$ )5.11	15.50 ( $\pm$ )2.02

Tabela 6 – Média amostral e desvio padrão do número de fontes recuperadas entre as versões do algoritmo MEXICO,  $GF(3)$ .

$N$	$\log_2(T)$	10	12	14
8	<b>MEXICO</b>	7.20 ( $\pm$ )2.04	7.87 ( $\pm$ )0.90	7.96 ( $\pm$ )0.54
	<b>MEXICO2</b>	7.20 ( $\pm$ )2.04	7.87 ( $\pm$ )0.90	7.96 ( $\pm$ )0.54
	<b>MEXICO-m</b>	7.45 ( $\pm$ )1.77	7.91 ( $\pm$ )0.82	7.98 ( $\pm$ )0.38
10	<b>MEXICO</b>	6.23 ( $\pm$ )4.16	8.82 ( $\pm$ )2.77	9.67 ( $\pm$ )1.54
	<b>MEXICO2</b>	6.26 ( $\pm$ )4.15	8.82 ( $\pm$ )2.77	9.67 ( $\pm$ )1.54
	<b>MEXICO-m</b>	6.70 ( $\pm$ )4.16	9.49 ( $\pm$ )1.92	9.90 ( $\pm$ )0.95
12	<b>MEXICO</b>	2.61 ( $\pm$ )3.91	7.33 ( $\pm$ )4.82	10.18 ( $\pm$ )3.56
	<b>MEXICO2</b>	2.61 ( $\pm$ )3.94	7.32 ( $\pm$ )4.81	10.18 ( $\pm$ )3.56
	<b>MEXICO-m</b>	2.69 ( $\pm$ )4.29	8.62 ( $\pm$ )4.97	11.18 ( $\pm$ )2.70

nos resultados. Observamos durante as simulações que as 3 propostas apresentam um comportamento similar: se o método não separa todos os sinais, é provável que o resultado seja ruim. Consequentemente, aumentar o número de execuções não implica na redução da variância. Porém, isto pode ser conseguido através do aumento do número de amostras, o que, no limite, levará à separação total dos sinais em todas as rodadas.

Neste contexto, consideramos que acrescentaria conteúdo a este experimento analisar o número de vezes que, nas 500 execuções, cada algoritmo consegue recuperar *todas* as fontes. A Figura 16 detalha os resultados com esta métrica, de  $2^6$  a  $2^{16}$  amostras em  $GF(2)$  e  $GF(3)$ . Observe que o MEXICO-m consistentemente supera as outras duas versões em todas as 4 possíveis combinações de  $P$  e  $N$ . Podemos notar também que a técnica apresentada requer menos amostras para alcançar os resultados equivalentes dos algoritmos originais.

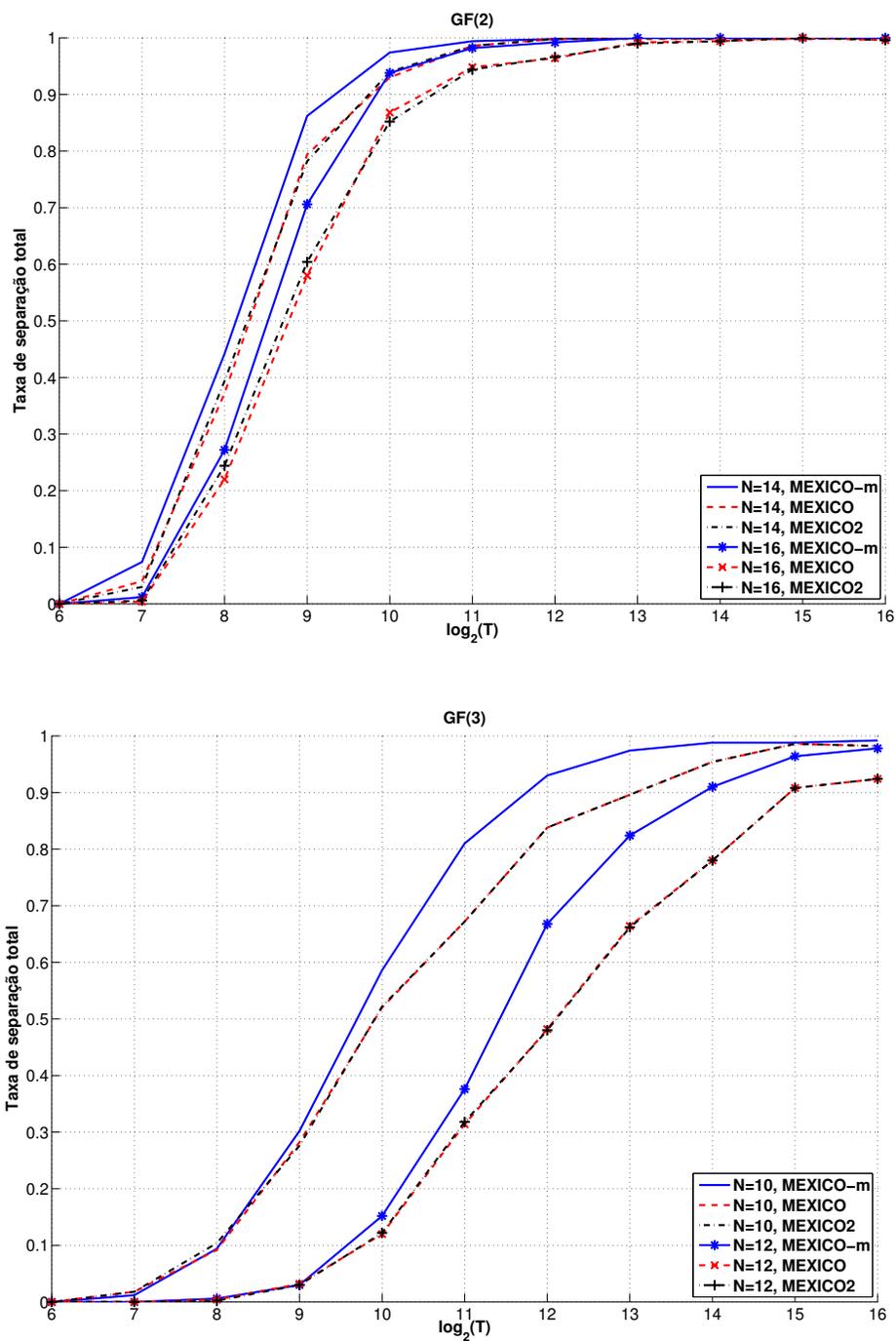


Figura 16 – Taxa de separação total para sinais em  $GF(2)$  (acima) e  $GF(3)$  (abaixo).

### 5.3.4 Considerações sobre a Proposta

Podemos concluir, com base na análise das simulações realizadas em um conjunto representativo de cenários, que as modificações propostas ao algoritmo MEXICO, que compõem sua nova versão, denominada MEXICO-m, aceleram a convergência do método sem implicar em alguma redução de qualidade das soluções.

Uma perspectiva de trabalho futuro é o estudo mais aprofundado do processo de estimação de entropia do método, bem como de maneiras para melhorar a operação tanto em termos de custo computacional como de qualidade de estimação, visto que as propostas atuais demandam uma quantidade considerável de amostras para ter estimativas confiáveis da PMF conjunta dos dados.

## 5.4 Misturas Convolutivas em Corpos Finitos

Até agora, a teoria de ICA / BSS em corpos finitos em geral e as contribuições propostas nesta tese se inserem no âmbito de hipóteses de linearidade e ausência de memória do sistema de mistura. Pode-se então indagar se tais hipóteses podem ser suprimidas ou relaxadas.

A fim de responder a pelo menos uma parte desta pergunta, desenvolvemos uma quarta e última contribuição ao problema de separação de sinais em corpos finitos, que consiste em estender o modelo de mistura instantânea para o cenário de misturas convolutivas (FANTI-NATO *et al.*, 2013).

Os modelos convolutivos, no caso de BSS para sinais reais ou complexos, são ferramentas essenciais para modelar a superposição de sinais tanto no espaço como no tempo. Exemplos práticos disto são encontrados nas comunicações sem fio, no processamento de áudio / fala e na geofísica (HAYKIN, 1994). Além de lidar pela primeira vez com esse modelo ampliado em corpos finitos, apresentaremos uma técnica baseada em deflação para lidar com o problema resultante, a qual será validada em simulações com sinais binários.

### 5.4.1 Proposta de Modelo

Considere um novo problema de mistura de sinais em que temos  $N_s$  sinais  $s_1(n), \dots, s_{N_s}(n)$  em  $GF(q)$  que são independentes (espacialmente e temporalmente), não-uniformes e não-degenerados. Este problema pode ser interpretado como a transmissão de sinais por meio de

um canal que provoca distorções descritas pela seguinte transformação linear:

$$\begin{aligned} x_i(n) &= \sum_{j=1}^{N_s} \sum_{k=0}^{L_h-1} h_{ij}(k) s_j(n-k) \\ &= \sum_{j=1}^{N_s} h_{ij}(n) * s_j(n), \end{aligned} \quad (5.15)$$

com  $i = 1, \dots, N_m$ , em que  $N_m$  é o número de misturas,  $h_{ij}(n)$  representa a resposta ao impulso dos canais individuais associados à  $i$ -ésima mistura e à  $j$ -ésima fonte,  $L_h$  é o comprimento de  $h_{ij}(n)$  e  $x_i(n)$  é a  $i$ -ésima mistura observada no instante  $n$ . Deve-se ter em mente que todos os sinais, somatórios e operações são definidos em um corpo de Galois, como no restante deste capítulo. Utilizando uma notação matricial, podemos reescrever a Equação 5.15 como:

$$\mathbf{x}(n) = \sum_{k=0}^{L_h-1} \mathbf{H}(k) \mathbf{s}(n-k) = \mathbf{H}(n) * \mathbf{s}(n), \quad (5.16)$$

onde  $\mathbf{x}(n) = [x_1(n), \dots, x_{N_m}(n)]^T$ ,  $\mathbf{s}(n) = [s_1(n), \dots, s_{N_s}(n)]^T$  e a matriz  $\mathbf{H}(n)$  é composta dos elementos  $h_{ij}(n)$  de cada canal individual (ROMANO *et al.*, 2011).

A Figura 17 ilustra graficamente o modelo apresentado, considerando o caso de duas fontes e duas misturas, também denominado sistema MIMO  $2 \times 2$ .

Do ponto de vista do receptor, os canais e as fontes são desconhecidos. Partindo do princípio de tentar recuperar uma fonte de cada vez, a partir dos sinais recebidos, define-se um problema de extração de sinais que consiste em determinar os coeficientes  $w_i(n)$  dos filtros de separação que produzem como saída

$$y(n) = \sum_{i=1}^{N_m} y_i(n) = \sum_{i=1}^{N_m} w_i(n) * x_i(n), \quad (5.17)$$

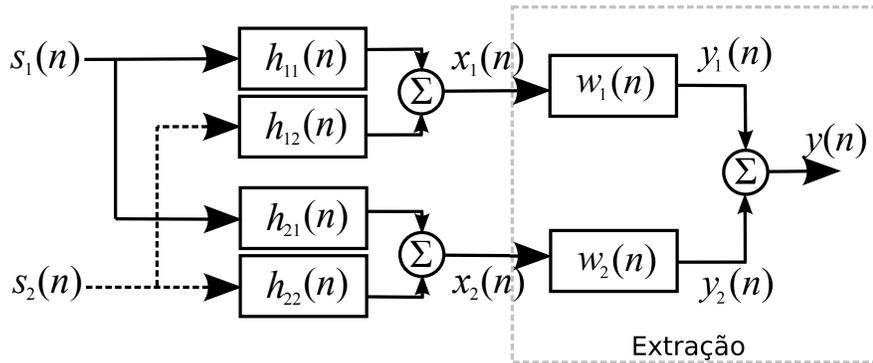


Figura 17 – O modelo de mistura convolutiva em corpos finitos e da estrutura de extração, no caso de um sistema MIMO  $2 \times 2$ .

de tal forma que este sinal corresponda a um dos  $N_s$  sinais originais, i.e.,  $y(n) = cs_j(n - d)$ , onde  $c \in GF(q)$  é um “fator de ganho” não-nulo e  $d$  é um atraso de tempo. Novamente fazendo uso da notação vetorial, podemos expressar a Equação 5.17 como

$$y(n) = \sum_{k=0}^{L_w-1} \mathbf{w}(k)\mathbf{x}(n - k) = \mathbf{w}(n) * \mathbf{x}(n) \quad (5.18)$$

onde  $L_w$  é o comprimento dos filtros  $w_i(n)$  e  $\mathbf{w}(n) = [w_1(n), \dots, w_{N_m}(n)]$ . Apesar de existirem ambiguidades de escala, atraso e permutação dos sinais, note que não há indeterminação de filtragem, dado que se assume que as amostras das fontes  $s_j(k)$  e  $s_j(l)$  são iid, para  $k \neq l$ , e que as fontes são independentes, o que configura uma situação semelhante à do problema de desconvolução cega (COMON; JUTTEN, 2010). Este panorama permite-nos propor um método para recuperar as fontes, quanto isto é viável em termos estruturais, que apresentamos a seguir.

#### 5.4.2 Algoritmo de Extração / Deflação Baseado na Entropia

Nesta seção, propomos um método para solucionar o problema de BSS em misturas convolutivas sobre corpos de Galois que se baseia no mesmo princípio de extração / deflação adotado por Yeredor (2007) e Gutch *et al.* (2010) no algoritmo EBED (vide subseção 2.3.1) para misturas instantâneas. Empregamos, nesta abordagem, filtros  $w_i(n)$  do tipo IIR para compor o sistema de separação, uma vez que, em corpos finitos, esta classe de filtros é potencialmente capaz de remover as dependências temporais de uma mistura sem deixar erros residuais, o que, por outro lado, pode ocorrer com filtros FIR. Ao aplicarmos esta estrutura de filtragem à formulação da Equação 5.17, temos

$$\begin{aligned} y(n) &= \sum_{i=1}^{N_m} \left( \sum_{j=0}^{L_a-1} a_i(j) * x_i(n - j) + \sum_{k=1}^{L_b} b_i(k) * y_i(n - k) \right) \\ &= \sum_{i=1}^{N_m} (a_i(n) * x_i(n) + b_i(n) * y_i(n)), \end{aligned} \quad (5.19)$$

em que  $L_a$  e  $L_b$  são as quantidades de coeficientes  $a_i(j)$  e  $b_i(k)$ , respectivamente, dos filtros IIR de extração. Estes coeficientes pertencem a  $GF(q)$  e são determinados de acordo com uma metodologia calcada em duas premissas:

1. Como é assumido que as fontes e suas observações são independentes, é possível empregar o procedimento de minimização da entropia do sinal de saída para extrair uma versão atrasada de uma única fonte, à luz da propriedade vista no Lema 2.1.
2. Tendo em mãos a versão de uma das fontes recuperadas de acordo com o passo 1, é possível novamente recorrer ao princípio de extração pela mínima entropia para remover

a influência desta fonte das misturas ainda presentes. Como esta fonte está presente na mistura devido a um processo convolutivo, a sua remoção é feita através da otimização dos parâmetros de um filtro que, de uma certa maneira, identifica a assinatura de interferência intersimbólica da fonte com respeito a cada mistura.

Sob condições ideais de estimação da entropia e otimização dos parâmetros, este procedimento deve extrair uma das fontes e removê-la de todas as misturas. Se então for repetido, ele deve permitir a recuperação de todas as fontes, contanto que não haja restrições estruturais, assunto este que discutiremos mais adiante.

Em termos matemáticos, o primeiro estágio da proposta corresponde a solucionar (mais uma vez) um problema de otimização — minimizar, com respeito aos coeficientes dos filtros de separação, a entropia do sinal de saída  $y(n)$  (vide Figura 17):

$$\begin{aligned} \min_{\mathbf{w}(n)} H(y), \\ y(n) = \mathbf{w}(n) * \mathbf{x}(n). \end{aligned} \tag{5.20}$$

Quando a saída  $y(n)$  de menor entropia é encontrada — que, neste caso, será denominada  $y_e(n)$  — surge a segunda tarefa: remover sua contribuição das misturas de tal forma que não mais interfira nos passos de separação seguintes. Como mencionado anteriormente, idealmente  $y_e(n)$  será idêntico a uma das fontes com exceção de um fator de “escala” e de um atraso de tempo indeterminado, i.e.,  $y_e(n) = cs_e(n - d)$ . Dadas a natureza convolutiva do problema e a indeterminação de atraso, o processo de deflação deve ser realizado aplicando um filtro FIR não-causal,  $g(n)$ , à fonte extraída  $y_e(n)$ . O resultado desta filtragem é então subtraído (ou adicionado, tanto faz em corpos de Galois) de cada mistura, levando a novos sinais  $r_i(n)$ ,  $i = 1, \dots, N_m$ . Encerrada esta etapa, o processo de extração é repetido para recuperar outra fonte, mas agora considerando  $r_i(n)$  como as novas misturas no lugar de  $x_i(n)$ . Analogamente à determinação do fator  $c$  que minimiza  $H(x_i + cs_1)$  no caso exclusivamente espacial (GUTCH *et al.*, 2010), os coeficientes do filtro de deflação  $g(n)$  são escolhidos de forma a minimizar  $H(r_i) = H(x_i + g * y_e)$ , como ilustra a Figura 18. Enfim, o processo todo se repete até que as misturas se tornem sinais nulos, o que implica que todas as fontes foram extraídas.

Os passos principais do algoritmo podem ser vistos no Pseudocódigo 5.4, que, por razões de simplicidade, considera o caso de um sistema MIMO  $2 \times 2$ . Todavia, o método pode ser diretamente estendido para abordar casos maiores, uma vez que o algoritmo, na verdade, decompõe um problema MIMO em uma série de  $N_s$  problemas MISO (do inglês *Multiple Input Single Output*).

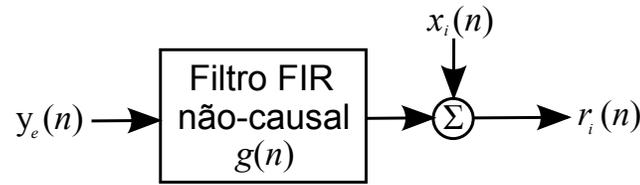


Figura 18 – Estrutura para realizar a deflação no problema de misturas convolutivas.

Como já sabemos, corpos finitos são estruturas inerentemente discretas e finitas, o que nos permite, no contexto do problema, obter  $w_i(n)$  e  $g(n)$  através de uma busca exaustiva. O espaço de busca dependerá da ordem dos filtros e da ordem do corpo. Para os filtros de separação  $w_i(n)$ , o número de possibilidades é  $q^{L_a+L_b}$ , e, para o caso do filtro de deflação  $g(n)$ , temos  $q^{L_g}$ , onde  $L_g$  é o número de coeficientes do filtro. Estes valores mostram claramente que, embora a busca exaustiva seja viável em dimensões relativamente menores, deve-se prestar atenção ao seu custo computacional para tarefas de maior escala.

---

**Pseudocódigo 5.4** Passos principais do algoritmo de Extração/Deflação para misturas convolutivas, no caso de um sistema MIMO  $2 \times 2$ .

---

**1. Extração:**

- 1.1 Determine, por busca exaustiva,  $w_1(n)$  e  $w_2(n)$  tal que  $H(y)$  é mínima.
- 1.2 Extrair sinal  $y_e(n)$ .

**2. Deflação:**

- 2.1 Para  $i = 1, 2$ , determine, por busca exaustiva,  $g(n)$  tal que  $H(r_i) = H(x_i + g * y_e)$  é mínima. Saída desta etapa é  $r_i(n)$ .
  - 2.2 Substitua as misturas por  $r_1(n)$  e  $r_2(n)$ .
3. Repita as etapas 1 e 2 para extrair as fontes restantes.
- 

### 5.4.3 Condições *Zero-Forcing* para Separação

As discussões sobre o modelo e o algoritmo proposto para extrair os sinais, até então, supõem que a estrutura descrita pela Figura 17 é inversível. No entanto, esta suposição pode não ser verdadeira; por isso, é necessário fazer uma análise de quais condições estão implícitas a sua aceitação. Nesta seção, apresentamos então as condições de *Zero-Forcing* (ZF) para esta estrutura.

A análise consiste em definir as condições necessárias para se obter uma inversão perfeita de um modelo de misturas convolutivas. Novamente, consideraremos o caso MIMO  $2 \times 2$ , embora a linha de desenvolvimento possa ser diretamente estendida para casos de maior

ordem no futuro. Neste contexto, portanto, as misturas  $\mathbf{x}(n)$  podem ser expressas como

$$\begin{aligned}\mathbf{x}(n) &= \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ h_{21}(n) & h_{22}(n) \end{bmatrix} * \begin{bmatrix} s_1(n) \\ s_2(n) \end{bmatrix} \\ &= \mathbf{H}(n) * \mathbf{s}(n).\end{aligned}\quad (5.21)$$

Embora a estratégia de extração / deflação que apresentamos na seção anterior seja uma abordagem tradicional no contexto de separação cega de sinais, é importante enfatizarmos que o problema de misturas convolutivas também é comumente tratado por meio de uma estrutura “espelho” do problema, que seria capaz de recuperar todas as fontes de uma vez (ROMANO *et al.*, 2011). Neste caso, idealmente, é necessário que a estrutura de separação seja um sistema MIMO e que  $\mathbf{H}(n)$  seja inversível, o que em outros termos significa que o determinante de  $\mathbf{H}(n)$  não seja nulo. A fim de avaliar esta situação, empregamos a transformada de Fourier para corpos de Galois (GFFT, do inglês *Galois Field Fourier Transform*), que é definida pelas equações

$$V(m) = \sum_{n=0}^{T-1} \alpha^{nm} v(n), \quad m = 0, \dots, T-1, \quad (5.22)$$

e

$$v(n) = T^{-1} \sum_{m=0}^{T-1} \alpha^{-nm} V(m), \quad n = 0, \dots, T-1; \quad (5.23)$$

onde  $v(n) \in GF(P)$  é um sinal contendo  $T$  amostras,  $P$  é um primo e  $\alpha \in GF(P^d)$  é um elemento de ordem  $T$  (SAKK; SMALL, 2003).

O teorema da convolução é válido com esta transformação, o que nos leva à aplicação da Equação 5.22 à matriz do canal  $\mathbf{H}(n)$ , resultando em

$$\det(\mathcal{F}\{\mathbf{H}(n)\}) = H_{11}(m)H_{22}(m) - H_{12}(m)H_{21}(m) \neq 0, \quad (5.24)$$

em que recordemos que a subtração é definida também para um corpo finito e 0 representa o respectivo elemento nulo,  $H_{11}(m)H_{22}(m)$  e  $H_{12}(m)H_{21}(m)$  são a GFFT das convoluções  $h_{11}(n) * h_{22}(n)$  e  $h_{12} * h_{21}(n)$ , e  $\mathcal{F}\{\mathbf{H}(n)\}$  é a GFFT de  $\mathbf{H}(n)$ . Podemos interpretar a Equação 5.24 como sendo necessária para obtermos uma condição ZF, já que ela se associa à possibilidade de recuperação perfeita.

Mesmo que tenhamos considerado, no processo de extração por entropia mínima, que somente uma fonte é separada por vez, as condições ZF apresentadas na Equação 5.24 ainda são necessárias. Pois se desejamos, por exemplo, extrair inicialmente o sinal  $s_1(n)$ , a saída combinada  $y(n)$  dos filtros de separação pode ser descrita vetorialmente como

$$\begin{aligned}y(n) &= \begin{bmatrix} w_1(n) & w_2(n) \end{bmatrix} * \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} \\ &= \mathbf{w}(n) * \mathbf{x}(n) = s_1(n).\end{aligned}\quad (5.25)$$

Aplicando a Equação 5.21 na Equação 5.25, temos

$$y(n) = \begin{bmatrix} w_1(n) & w_2(n) \end{bmatrix} * \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ h_{21}(n) & h_{22}(n) \end{bmatrix} * \begin{bmatrix} s_1(n) \\ s_2(n) \end{bmatrix}. \quad (5.26)$$

Expandindo os seus termos e aplicando em seguida a GFFT, chega-se às seguintes igualdades derivadas da Equação 5.26:

$$\begin{aligned} W_1(m)H_{11}(m) + W_2(m)H_{21}(m) &= 1, \\ W_1(m)H_{12}(m) + W_2(m)H_{22}(m) &= 0. \end{aligned} \quad (5.27)$$

Neste caso, a extração perfeita pode ser realizada se

$$\begin{aligned} W_1(m) &= \frac{H_{22}(m)}{H_{11}(m)H_{22}(m) - H_{12}(m)H_{21}(m)}, \\ W_2(m) &= \frac{-H_{12}(m)}{H_{11}(m)H_{22}(m) - H_{12}(m)H_{21}(m)}. \end{aligned} \quad (5.28)$$

Observe que esta solução é válida somente se  $H_{11}(m)H_{22}(m) - H_{12}(m)H_{21}(m) \neq 0$ , i.e., se nenhum dos dois denominadores na Equação 5.28 tornam-se nulos, condição esta idêntica à colocada pela Equação 5.24. Se, ao invés de  $s_1(n)$ , executamos o mesmo procedimento para extrair primeiramente  $s_2(n)$ , chegamos a uma situação complementar à descrita pela Equação 5.27:

$$\begin{aligned} W_1(m)H_{11}(m) + W_2(m)H_{21}(m) &= 0, \\ W_1(m)H_{12}(m) + W_2(m)H_{22}(m) &= 1, \end{aligned} \quad (5.29)$$

que nos leva à solução

$$\begin{aligned} W_1(m) &= \frac{-H_{21}(m)}{H_{11}(m)H_{22}(m) - H_{12}(m)H_{21}(m)}, \\ W_2(m) &= \frac{H_{11}(m)}{H_{11}(m)H_{22}(m) - H_{12}(m)H_{21}(m)}, \end{aligned} \quad (5.30)$$

cujos denominadores são iguais aos observados na Equação 5.28. Por isso, a mesma condição ZF deve permanecer válida para se extrair o sinal com sucesso e, conseqüentemente, a desigualdade da Equação 5.24 pode ser considerada como a principal restrição estrutural para a existência de soluções de extração do problema de BSS convolutivo em corpos finitos.

Entretanto, a condição ZF para a extração dos sinais *subsequentes* pode diferir daquela apresentada na Equação 5.24, uma vez que os efeitos do processo de deflação devem ser acrescentados ao cenário. Tal análise, infelizmente, não pôde ser realizada a tempo durante o doutorado.

Em suma, o que podemos afirmar é que se a condição apresentada nesta seção não for válida, o algoritmo não terá condições estruturais de obter as soluções. Caso a condição

seja respeitada, podemos esperar que o método proposto na subseção 5.4.2 consiga recuperar todos ou pelo menos um subconjunto dos sinais originais, dadas as ambiguidades previamente mencionadas e com a ressalva de que a análise completa da condição estrutural de separação dos sinais subsequentes ainda permanece em aberto.

#### 5.4.4 Simulações

Para analisar o algoritmo e as questões teóricas levantadas, implementamos o método proposto em MATLAB para realizar uma série de simulações, considerando o caso de corpos binários —  $GF(2)$  — e um sistema convolutivo MIMO  $2 \times 2$ . A opção por esse cenário relativamente simples foi feita para permitir que o desempenho do método pudesse ser avaliado com o mínimo de peculiaridades teóricas que, potencialmente, fariam a análise ser menos informativa. Verificada a validade da proposta, naturalmente se coloca como tarefa futura a ampliação dos cenários de estudo.

Como feito nos casos anteriores, o método para estimar a entropia dos sinais é o estimador ML, apresentando na Equação 5.9. Dadas as considerações teóricas feitas previamente, se as condições ZF forem atendidas, esperamos que a principal fonte de limitação de desempenho do algoritmo seja a quantidade  $T$  de amostras disponíveis para o estimador, dado que cada passo de minimização da entropia será penalizado por flutuações causadas por uma estimação imperfeita.

Realizamos o tradicional esquema de simulações *Monte Carlo* compreendendo 50 execuções do algoritmo para diferentes valores de  $T$  e diferentes ordens de filtros:  $L_h = L_a = L_g = 2$ ,  $L_b = 3$  e  $L_h = L_a = L_g = 3$ ,  $L_b = 5$ . Cada execução tem as probabilidades dos símbolos 0/1 definidas aleatoriamente no intervalo  $[0.05, 0.3] \cup [0.7, 0.95]$ , e os coeficientes dos filtros  $h_{ij}(n)$  do sistema de mistura também são definidos aleatoriamente, desde que atendam à condição ZF imposta na subseção 5.4.3.

A Figura 19 exhibe os resultados, em termos de frequência de separação, para o cenário com canais de 2 coeficientes. O gráfico ilustra que o algoritmo conseguiu extrair pelo menos uma fonte em todas as execuções quando o número de amostras foi maior ou igual a 128. A frequência de separação total (das duas fontes) é próxima de 50% no ponto inicial com 64 amostras e, como esperado, cresce com o seu respectivo aumento até atingir o patamar de 100% de sucesso com 2048 amostras.

Já a Figura 20 apresenta os resultados do cenário com 3 coeficientes nos canais do sistema misturador. Claramente, a primeira conclusão possível de se formular é que o aumento da dimensão do problema impacta diretamente na capacidade de separação dos sinais. No entanto, o método ainda apresenta resultados bem satisfatórios, mais uma vez pelo menos

uma fonte foi extraída em todas as rodadas quando  $T \geq 128$ . Ademais, a frequência de separação completa teve valor inicial de 30% e cresceu continuamente até atingir 100% de sucesso com 8192 amostras.

Notamos, ao analisar os dois cenários, que o sucesso na separação da segunda fonte requer um número maior de amostras. Isto é explicado pelo fato de que a extração da segunda fonte requer a deflação da fonte previamente extraída, seguido por uma nova rodada de otimização da estrutura de extração (vide Figura 17), o que significa duas etapas adicionais baseadas na estimação da entropia — com as suas respectivas imprecisões — e que potencialmente influenciam a qualidade da solução obtida. Não obstante, pudemos observar experimentalmente que, com o atendimento das condições ZF para extrair a primeira fonte, a segunda extração também é bem-sucedida se há um número suficientemente grande de amostras para estimação.

Portanto, concluímos que a técnica aqui proposta, embora efetiva, depende de uma estimação precisa da entropia para prover resultados adequados, o que se torna cada vez mais complexo quando cresce o número de fontes e a dimensão das estruturas envolvidas (sejam de mistura ou de separação).

Para concluir este estudo experimental, foi feita uma análise do efeito da restrição de não-uniformidade sobre os resultados do algoritmo. Considerando os parâmetros  $T = 256$ ,  $L_h = L_a = L_g = 2$  e  $L_b = 3$ , todos fixos, desta vez realizamos 50 execuções do algoritmo em cada ponto de um espaço de possíveis PMFs para  $s_1(n)$  e  $s_2(n)$ . Então, contabilizamos o número de vezes em que o algoritmo falhava na *primeira* tentativa de extração de qualquer fonte. O resultado do experimento pode ser observado na Figura 21.

O gráfico ilustra de maneira bastante didática como a eficácia do princípio de extração por entropia mínima reduz à medida que as duas distribuições se tornam “mais uniformes”. Na região central da superfície, quando os dois sinais possuem PMFs aproximadamente uniformes, suas respectivas entropias já possuem valor praticamente máximo e, portanto, o processo de mistura não gera incremento numericamente significativo de forma que a estrutura de separação consiga reduzir tal efeito *a posteriori*. Nas regiões onde somente uma das distribuições é uniforme ou próxima disso, a taxa de falha permanece nula porque ainda há uma fonte não-uniforme, com entropia não-máxima, e que portanto o método é capaz de extrair.

#### 5.4.5 Considerações sobre a Proposta

Após termos explorado novas possibilidades de algoritmos para o problema de separação cega de sinais em corpos finitos, no contexto de misturas instantâneas, a quarta e última

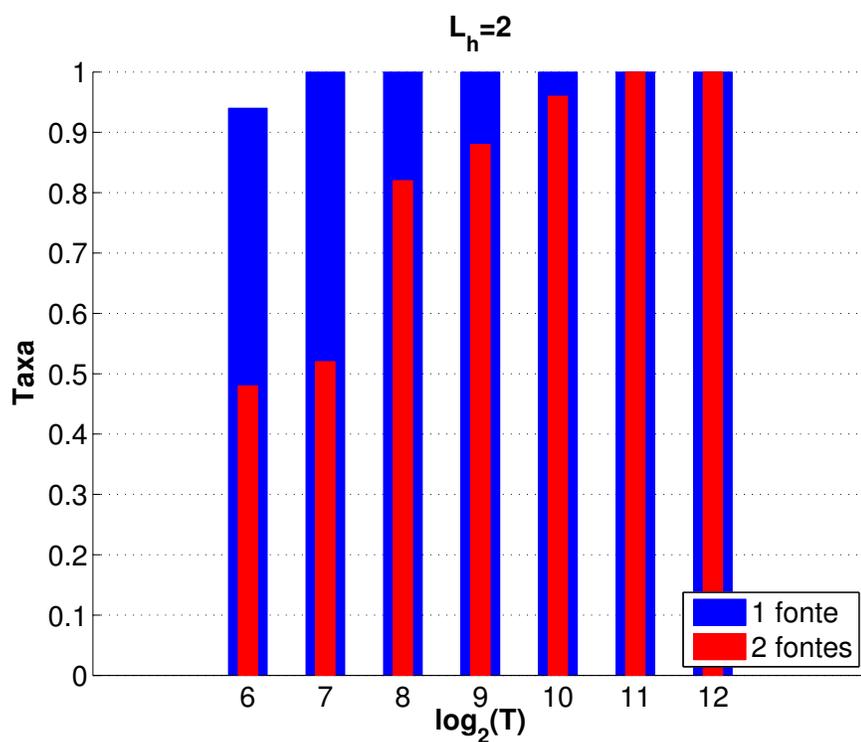


Figura 19 – Taxa de separação do algoritmo de Extração/Deflação em BSS convolutivo. Canais com 2 coeficientes.

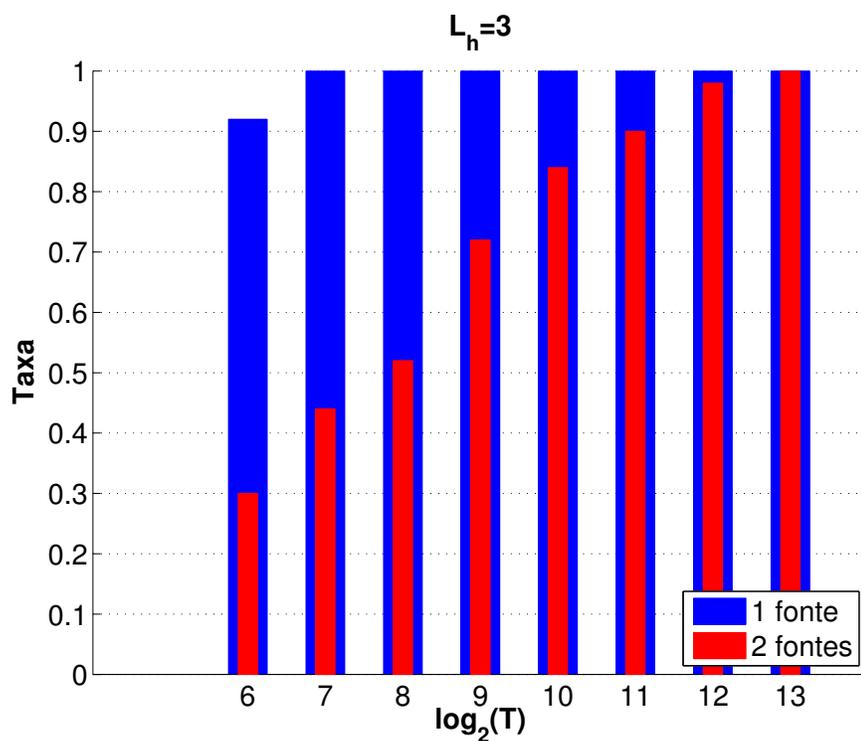


Figura 20 – Taxa de separação do algoritmo de Extração/Deflação em BSS convolutivo. Canais com 3 coeficientes.

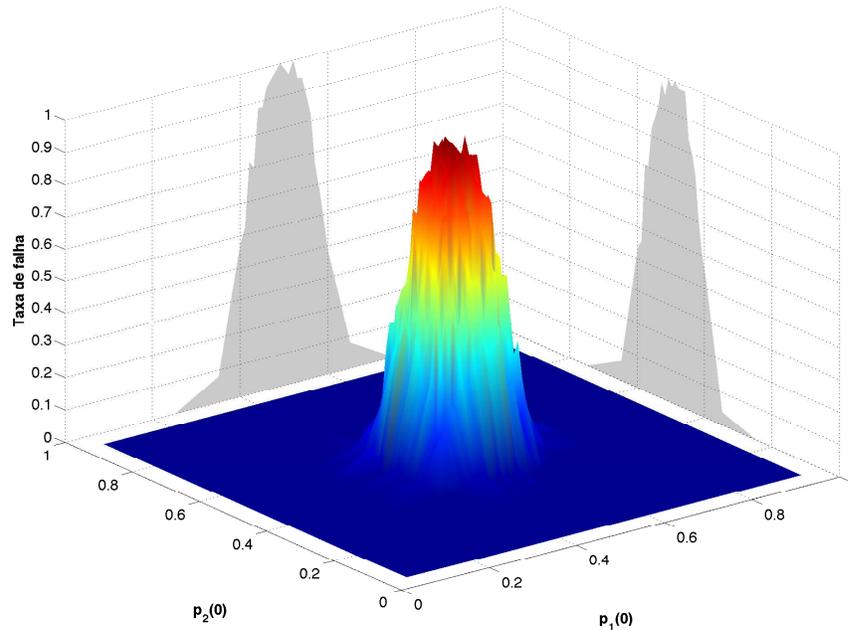


Figura 21 – Taxa de falha ao extrair a primeira fonte, no espaço das PMFs de  $s_1(n)$  e  $s_2(n)$ .

contribuição apresentada neste capítulo trata principalmente da formulação de um novo problema, o de BSS em corpos finitos no contexto de misturas convolutivas. Além de apresentar este novo modelo, desenvolvemos um algoritmo baseado em deflação para sua solução e um estudo inicial das condições estruturais para que a inversão ocorra.

Ao debruçar-nos sobre os resultados das simulações, podemos concluir que o método pode operar eficientemente na separação dos sinais desde que haja um número adequado de amostras. Em outras palavras, o desempenho do algoritmo é, essencialmente, limitado pela qualidade das estimativas subjacentes de entropia. Como esperado, os resultados indicaram que sistemas mais complexos (em termos espaciais ou temporais) tendem a necessitar de mais amostras para que a entropia seja apropriadamente estimada e para que as tarefas tocantes à otimização sejam eficazes.

Como fica implícito ao longo da proposta, o trabalho desenvolvido é germinal e, assim, provê diversas perspectivas de estudos futuros, como o detalhamento dos experimentos para cenários com mais fontes e ordens maiores de corpos, além do uso de procedimentos de busca não-exaustivos e da análise das condições ZF para as extrações dos sinais subsequentes.

## 5.5 Conclusões

O capítulo que se encerra apresentou as contribuições desenvolvidas neste doutorado para o problema de separação cega de sinais em corpos de Galois. O algoritmo de ICA em  $GF(2)$  composto pelo CLONALG e pelo critério de minimização da informação mútua, mais especificamente a soma das entropias marginais, é o ponto de partida das propostas elaboradas. Os resultados das simulações indicam que a abordagem por metaheurísticas populacionais é viável para o problema em questão.

A estratégia permanece em uso na proposta seguinte, uma nova técnica composta pelo algoritmo cob-aiNet[C] associado ao critério da mínima entropia. Além da já conhecida tática de utilizar uma técnica bio-inspirada a problemas de engenharia, o algoritmo é inovador ao formular o problema sustentando-se numa abordagem *Michigan* de construção de algoritmos evolutivos, para que a matriz de separação seja composta ao final da execução por um subconjunto de indivíduos bem adaptados da população. Os experimentos demonstraram o bom funcionamento do método, similar ao de técnicas estado-da-arte, e sua potencial aplicação a cenários com números relativamente grandes de fontes.

Aparte do uso de algoritmos imuno-inspirados, a terceira contribuição consistiu na melhoria da heurística de busca do algoritmo MEXICO, resultando em uma nova implementação que traz melhorias em termos de convergência e potencial de busca global. E por fim a generalização do modelo de BSS linear instantâneo para o cenário onde ocorrem misturas também no domínio do tempo foi o tema da quarta contribuição desenvolvida, acompanhada de um algoritmo por deflação para solução do problema e da análise das condições de inversão.

De maneira ampla, todas as proposições mostraram-se viáveis e com resultados competitivos ou até superiores às ferramentas já estabelecidas. Além disso, percebe-se que os trabalhos que se desenvolveram compartilham do objetivo comum de trazer maior robustez e escalabilidade à ICA em corpos finitos, visto que as melhorias se posicionaram sempre no sentido de melhoria de custo computacional, qualidade de soluções ou escopo do modelo atendido.

Os trabalhos propostos permitiram explorar praticamente todos os aspectos que permeiam as soluções baseadas em ITL, que são: o critério de otimização, o método de ajuste / busca dos parâmetros envolvidos e o modelo implícito do problema a ser tratado. No próximo capítulo, retornamos aos sinais com valores reais e tratamos das contribuições desenvolvidas para o problema de inversão cega de sistemas de Wiener.

## 6 Contribuições à Inversão Cega de Sistemas de Wiener

Este capítulo encerra a parte referente às contribuições desta tese, abordando aquelas desenvolvidas no contexto do problema de inversão cega de sistemas de Wiener por sistemas de Hammerstein. Em consonância com o paradigma de construção de algoritmos de ITL que vem sendo empregado na tese, apresentaremos uma proposta que emprega o algoritmo CLONALG para otimização dos parâmetros do sistema inversor de acordo com o critério de minimização da informação mútua do sinal de saída. A formulação deste novo *framework* é acompanhada de um estudo empírico sobre as condições para solução do sub-problema linear e, por fim, de simulações contemplando o problema por inteiro.

### 6.1 Um *Framework* Bio-Inspirado Baseado no Critério da Informação Mútua

O Capítulo 3 introduziu o problema de inversão cega de sistemas de Wiener por sistemas de Hammerstein, e nele foi exposto, dentre várias questões, que o algoritmo proposto por Taleb *et al.* (2001) baseia-se na busca descendente por gradiente e, em consequência disso, é inerentemente vulnerável ao risco de convergência prematura para um mínimo local. Por outro lado, verificamos também que a técnica elaborada por Rojas *et al.* (2007), embora seja mais robusta em termos do potencial de convergência local, é implementada com a curtose como função custo, o que a faz explorar de maneira limitada a estrutura estatística dos sinais envolvidos. Ademais, ambos os métodos limitam a estrutura linear do sistema de Hammerstein a um modelo FIR.

Com o objetivo de *(i)* evitar o caráter exclusivo de busca local dos algoritmos de gradiente, *(ii)* considerar um função objetivo que seja a mais rigorosa possível e *(iii)* implementar estruturas de desconvolução mais genéricas que aquelas compostas somente por modelos FIR, propomos uma nova técnica para realizar a inversão cega de sistemas de Wiener que:

1. Realiza a otimização dos parâmetros através de um algoritmo bio-inspirado — mais especificamente um SIA, que, como discutido no Capítulo 4, possui um interessante equilíbrio entre as estratégias de busca local e global.
2. Emprega a informação mútua como referência para elaboração do critério de filtragem.

3. Considera o sistema de Hammerstein como um modelo paramétrico que aceita tanto filtros FIR como IIR para a estrutura de inversão linear (de resposta ao impulso  $w(n)$ ), o que teoricamente aumenta a capacidade de encontrar boas soluções dentro de um número limitado de coeficientes.

Observe que esta proposta, similarmente a algumas das contribuições desenvolvidas no capítulo anterior, alinha-se ao princípio de propor algoritmos de ITL que empregam estratégias de busca além dos tradicionais métodos de gradiente, por meio das metaheurísticas populacionais. Outra questão relevante é que o emprego destas técnicas de busca permite considerar um espectro amplo de funções objetivo, dado que não se exige, por exemplo, que a função seja diferenciável; ademais, a adaptação dos parâmetros de estruturas recorrentes é relativamente direta, pois não há a necessidade de lidar com técnicas de retropropagação (*backpropagation*) através do tempo e, conseqüentemente, o risco de instabilidade é mitigado graças à existência de uma população de filtros candidatos a solução.

A Figura 22 fornece a visão geral da proposta, com cada um dos seus aspectos distintivos destacados pelas caixas em azul. A seguir, detalhamos cada um destes aspectos: a estimação da função custo, as questões estruturais do sistema de inversão e a otimização dos seus parâmetros.

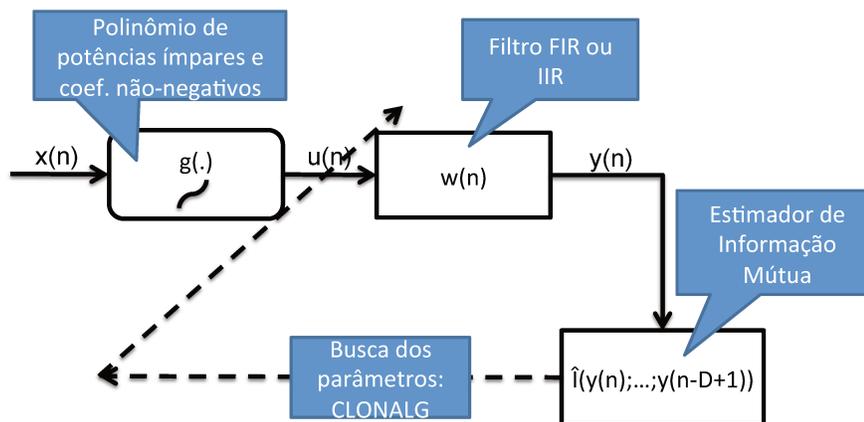


Figura 22 – Proposta de *framework* para inversão cega de sistemas de Wiener.

### 6.1.1 Função Objetivo: Definição e Estimadores

Embora a função objetivo apresentada na Equação 3.10 seja extremamente rica do ponto de vista teórico, ela é difícil de ser avaliada na prática porque o segundo termo do lado direito da fórmula não pode ser obtido sem que haja o *inteiro* conhecimento da probabilidade conjunta, com *quaisquer* (e infinitos) instantes de tempo, do processo da saída  $\mathcal{Y}$ .

Em termos mais objetivos, se temos disponível um conjunto de  $T$  amostras  $x(1), \dots, x(T)$  do sinal distorcido, é necessário que consideremos para a saída do sistema de Hammerstein o conjunto das  $T$  amostras  $y(1), \dots, y(T)$  (se desprezarmos o efeito de espalhamento no tempo do filtro de inversão).

Como uma aproximação a essa formulação ideal, usamos uma função custo baseada em estimadores de informação mútua para variáveis aleatórias (recorde-se da definição da Equação 3.6), truncando o número de atrasos considerados. É possível adotar as mais diversas estratégias para estimação de  $IM^1$ , entretanto, nesta proposta, escolhemos dois estimadores que consideramos representativos do compromisso entre simplicidade e acurácia. Descrevamos os dois, portanto.

### 6.1.1.1 Estimador de IM Baseado em Histograma

Quando pensamos em estimar probabilidades como sendo a frequência com que os eventos ocorrem, é bastante natural considerarmos o uso de histogramas, à vista da objetividade conceitual e simplicidade que apresentam. Ao empregar histogramas, é possível derivar um estimador *plug-in* de IM relativamente simples, cujo desempenho pode ser devidamente avaliado em termos de *bias* e variância (MODDEMEIJER, 1989).

O estimador é válido para qualquer número de VAs, mas por razões de simplicidade, vamos apresentar a sua derivação para o caso bidimensional, referindo-se às variáveis  $X_1$  e  $X_2$ . Primeiramente, dividimos o espaço de observações  $X_1 X_2$  em  $(k_{X_1} \times k_{X_2})$  segmentos de tamanho idêntico  $(\Delta_{X_1} \times \Delta_{X_2})$ . Este particionamento tem sua origem e o tamanho dos segmentos definidos de forma que se cubra um intervalo de  $6\sigma_{X_1}$  e  $6\sigma_{X_2}$  ao redor da média amostral, onde  $\sigma_{X_1}$  e  $\sigma_{X_2}$  são os desvios padrão amostrais das VAs.

Se contamos o número  $c_{ij}$  de observações que ocorrem dentro de cada célula  $(i, j)$  do espaço particionado e assumimos que a densidade de probabilidade é constante dentro de cada região,  $c_{ij}$  equivale à PMF conjunta. Posto isto, o estimador de informação mútua pode ser definido através da definição canônica para VAs discretas (COVER; THOMAS, 2006):

$$\hat{I}(X_1; X_2) = \sum_{j=1}^{k_{X_2}} \sum_{i=1}^{k_{X_1}} \frac{c_{ij}}{T} \log \frac{c_{ij} T}{c_i c_j}, \quad (6.1)$$

no qual  $c_i$  e  $c_j$  são os valores marginais de  $c_{ij}$ . Moddemeijer (1989) derivou fórmulas aproximadas para o *bias* e a variância deste estimador, chegando a uma versão do estimador com *bias* corrigido, a qual é considerada nesta proposta. As expressões calculadas são

$$\text{Bias}\{\hat{I}(X_1; X_2)\} = \frac{(k_{X_1} - 1)(k_{X_2} - 1)}{2T} \quad (6.2)$$

<sup>1</sup> Caso o leitor se interesse, recomendamos consultar o artigo de revisão de Walters-Williams e Li (2009).

e

$$\text{Var}\{\hat{I}(X_1; X_2)\} = \frac{1}{T} \left( \sum_{i,j} \frac{c_{ij}}{T} \log^2 \frac{c_{ij}T}{c_i c_j} - \left( \sum_{i,j} \frac{c_{ij}}{T} \log \frac{c_{ij}T}{c_i c_j} \right)^2 \right). \quad (6.3)$$

Percebemos pela análise da Equação 6.2 que o *bias* cresce com o aumento de partições, enquanto diminui com o aumento do número de amostras. Interessantemente, podemos ver nisto uma relação do excesso de particionamento com um sobre-ajuste do modelo aos dados, enquanto é bastante intuitivo observar no aumento das amostras um maior conhecimento sobre a real distribuição probabilística.

### 6.1.1.2 Estimador de Informação Mútua Quadrática de Cauchy-Schwarz

Principe e seus colaboradores (PRINCIPE, 2010) propuseram várias medidas de independência que podem ser empregadas em algoritmos de ITL. Uma delas, a ser apresentada nesta seção, é baseada na divergência de Cauchy-Schwarz entre duas PDFs  $f_{X_1}(\cdot)$  e  $f_{X_2}(\cdot)$ :

$$D_{CS}(f_{X_1}; f_{X_2}) = D_{CS}(X_1; X_2) = -\log \frac{\left( \int f_{X_1}(x) f_{X_2}(x) dx \right)^2}{\int f_{X_1}^2(x) dx \int f_{X_2}^2(x) dx}. \quad (6.4)$$

Note que  $D_{CS}(X_1; X_2) \geq 0$ , e a igualdade ocorre se, e somente se, as duas PDFs forem idênticas. É possível derivar um estimador para  $D_{CS}$  através do emprego do já conhecido método de janela de Parzen (PARZEN, 1962) — com uma função *kernel* gaussiana —, cuja definição está na Equação 1.26. Aplica-se então este estimador de PDF à Equação 6.4 e, novamente, deparamo-nos com o “truque” da função *kernel* gaussiana, que resulta na solução analítica das integrais envolvidas. A expressão final é descrita, então, em termos das medidas de potencial de informação (recorde-se da Equação 1.28):

$$\hat{D}_{CS}(X_1; X_2) = \log \frac{\hat{V}_{X_1} \hat{V}_{X_2}}{(\hat{V}_{X_1 X_2})^2}, \quad (6.5)$$

onde  $\hat{V}_{X_1}$  e  $\hat{V}_{X_2}$  são a medida IP com respeito às PDFs  $f_{X_1}$  e  $f_{X_2}$ , respectivamente, e  $\hat{V}_{X_1 X_2}$  é o potencial de informação cruzada (CIP, do inglês *Cross-Information Potential*), dado por

$$\hat{V}_{X_1 X_2} = \frac{1}{T^2} \sum_{j=1}^T \sum_{k=1}^T \kappa(x_{1_k} - x_{2_j}, \sqrt{2}\sigma), \quad (6.6)$$

em que  $x_{1_k}$  e  $x_{2_j}$  são a  $k$ -ésima e  $j$ -ésima observações associadas às distribuições  $f_{X_1}$  e  $f_{X_2}$ , respectivamente, e assumimos por simplicidade que ambos os conjuntos de dados possuem o mesmo número de amostras  $T$ .

De maneira análoga à relação que mostra que a divergência de Kullback-Leibler entre a PDF conjunta e o produto das PDFs marginais equivale à informação mútua (vide Equação 1.13), podemos partir da divergência recém apresentada na Equação 6.4 para se definir uma nova medida, a Informação Mútua Quadrática de Cauchy-Schwarz (QMI), que no caso de duas VAs  $X_1$  e  $X_2$  é expressa por

$$I_{CS}(X_1; X_2) = D_{CS}(f_X(x_1, x_2); f_{X_1}(x_1)f_{X_2}(x_2)). \quad (6.7)$$

Pode-se constatar que  $I_{CS}(X_1; X_2)$  é uma medida apropriada de independência porque seu valor anula-se somente se  $X_1$  e  $X_2$  são independentes. A definição pode ser diretamente estendida para vetores aleatórios  $m$ -dimensionais:

$$I_{CS}(X_1; \dots; X_m) = D_{CS}(f_X(x_1, \dots, x_m); \prod_{i=1}^m f_{X_i}(x_i)). \quad (6.8)$$

Estas relações estabelecidas entre  $I_{CS}$  e  $D_{CS}$  permitem de maneira direta definir o estimador da medida de independência a partir do estimador da divergência, o que resulta em

$$\hat{I}_{CS}(X_1; X_2) = \log \frac{\left( \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \hat{V}_{X_1}(i, j) \hat{V}_{X_2}(i, j) \right) (\hat{V}_{X_1} \hat{V}_{X_2})}{\left( \frac{1}{T} \sum_{i=1}^T \hat{V}_{X_1}(i) \hat{V}_{X_2}(i) \right)^2}, \quad (6.9)$$

onde as expressões do estimador IP são dadas por:

$$\begin{aligned} \hat{V}_{X_k}(i, j) &= \kappa(x_{k_i} - x_{k_j}, \sqrt{2}\sigma), \\ \hat{V}_{X_k}(i) &= \frac{1}{T} \sum_{j=1}^T \hat{V}_{X_k}(i, j), \quad \hat{V}_{X_k} = \frac{1}{T} \sum_{i=1}^T \hat{V}_{X_k}(i), \quad k = 1, 2. \end{aligned} \quad (6.10)$$

É possível derivar a partir da Equação 6.9 uma expressão similar para o caso multivariado, mas consideraremos na proposta somente a derivação bidimensional. O estimador neste caso apresenta uma complexidade computacional de  $O(T^3)$ , porém uma implementação rápida que usa a técnica de decomposição de Cholesky (SETH; PRINCIPE, 2009) pode reduzir este custo para  $O(TC^2)$ , onde  $C$  é a dimensão da matriz de decomposição,  $C \ll T$ .

### 6.1.2 Método de Busca e Componentes do Modelo

Conforme mencionado, propomos que, à luz do trabalho anterior de Rojas *et al.* (2007), seja mantida a estratégia de uso da computação bio-inspirada para a otimização dos parâmetros do sistema de Hammerstein, segundo o critério dado por um dos dois estimadores

estudados na subseção 6.1.1. Porém, é realizada uma mudança que consiste em trocar o algoritmo genético pelo algoritmo imuno-inspirado CLONALG, apresentado na subseção 4.2.1. Entre as razões para a troca, está o fato de que este método e outras abordagens de SIAs vêm sendo regularmente aplicadas com sucesso a diversos problemas em processamento de sinais — vide, por exemplo, os resultados de Dias *et al.* (2009), Wada *et al.* (2009) e Romano *et al.* (2011) — e de que o método possui a (já mencionada) capacidade intrínseca de manter diversidade de soluções ao longo da execução, o que pode ser muito relevante para aumentar a probabilidade de encontrar o ótimo global ou um bom ótimo local. Mesmo assim, do ponto de vista computacional, há muitos pontos de contato entre o *modus operandi* do CLONALG e dos métodos evolutivos, como pôde ser observado no seu detalhamento, no Capítulo 4.

Quanto aos parâmetros a serem otimizados, os componentes estruturais do sistema de Hammerstein, que são:

1. Uma função não-linear  $g(\cdot)$  dada por um polinômio de ordem  $k$  (composto somente por monômios de grau ímpar) com coeficientes estritamente positivos. Estas restrições fazem com que o polinômio seja monotonicamente crescente (e portanto inversível), dado que sua primeira derivada é positiva para qualquer valor. Conseqüentemente, é um modelo factível para aproximar a inversa da distorção original  $f(\cdot)$ , exceto por uma ambigüidade de sinal — o que não é relevante, pois este ponto pode ser compensado pelo filtro linear subsequente.
2. Um sub-sistema LTI  $w(n)$  com um número finito de coeficientes ( $M + 1$  na parte FIR e  $Q$  na parte IIR) que descrevem a sua função de transferência — vide Equação 3.4.

O algoritmo CLONALG, na sua implementação para otimização real, é responsável por buscar os parâmetros de  $g(\cdot)$  e  $w(n)$  que minimizam a função custo  $J$ , avaliada através de um dos dois estimadores descritos na subseção 6.1.1

$$J(g; w) = \hat{I}(y(n); \dots; y(n - D + 1)) \approx I(\mathcal{Y}), \quad (6.11)$$

em que  $D$  é número de atrasos referente ao sinal de saída  $y(n)$ . Os indivíduos da população representam os parâmetros do sistema de Hammerstein, que são organizados conforme o diagrama apresentado na Figura 23. O algoritmo adota o operador de mutação gaussiana, o qual soma um número aleatório gerado por uma distribuição normal a cada componente do vetor candidato (BACK *et al.*, 2000).

De uma perspectiva ampla, esta proposta de *framework* aborda o problema de inversão cega com um algoritmo de busca capaz de escapar de soluções localmente ótimas e, ao mesmo tempo, permite empregar funções custo mais complexas (não obrigatoriamente diferenciáveis) que possivelmente levam a soluções melhores.

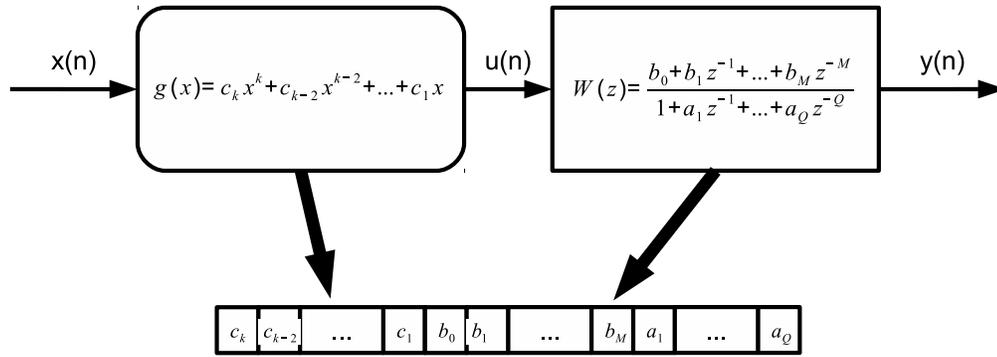


Figura 23 – Esquema de codificação dos indivíduos para o algoritmo de inversão dos sistemas Wiener-Hammerstein, cada coeficiente é um valor real.

## 6.2 Um Experimento Preliminar: Inversão Linear Via Estimação da Informação Mútua

Após as apresentações dos conceitos acerca do problema, das estratégias de solução existentes e da proposta que introduzimos nesta tese, pode ser interessante verificar, por razões de precaução, se a informação mútua da saída do sistema de fato funciona como um contraste. Em outras palavras, tentaremos responder à seguinte questão, antes de prosseguir para os testes do *framework*: supondo que a distorção não-linear já tenha sido neutralizada, o valor mínimo do estimador de IM, calculado entre  $y(n)$  e  $y(n - 1)$ , corresponderá aos parâmetros de um filtro  $w(n)$  que recupera (ou aproxima o máximo possível)  $s(n - d)$ ?

Ignoremos por ora a parte não-linear e analisemos a comportamento da função custo (que por simplicidade é somente o estimador por histograma, definido na subseção 6.1.1.1) em um cenário equivalente ao problema de desconvolução linear cega (ROMANO *et al.*, 2011). A implementação em MATLAB do estimador por histograma é fornecida por Moddemeijer (1989). O procedimento consiste então em realizar uma *busca exaustiva* no espaço dos parâmetros do filtro de desconvolução a fim de construir o mapeamento da função custo, levando em conta diferentes tipos de canais processados por (i) um filtro FIR de três coeficientes, especificados de acordo com um sistema de coordenadas esféricas

$$W_{\theta,\phi}(z) = \cos(\theta) + \sin(\theta) \sin(\phi)z^{-1} + \sin(\theta) \cos(\phi)z^{-2},$$

e (ii) um filtro IIR com dois pólos e dois zeros que são especificados conforme o sistema de coordenadas polares

$$W_{\theta,\phi}(z) = \frac{\cos(\phi) + \sin(\phi)z^{-1}}{\cos(\theta) + \sin(\theta)z^{-1}}.$$

Este experimento é importante porque, se a inversão não se mostrar viável com o problema linear, não há razão prática para estender a análise para o modelo não-linear. Para

que haja sucesso, então, o que se espera de resultado é que o sinal de saída do filtro  $y(n)$  seja o mais similar possível a  $s(n - d)$ . O critério para medir tal solução é comparar o filtro de IM mínima (ou MMI, do inglês *Minimal Mutual Information*) com a solução que seria obtida pelo critério *supervisionado* de mínimo erro quadrático médio (MMSE, do inglês *Minimum Mean Square Error*).

### 6.2.1 Canal de Fase Mínima

O experimento preliminar é iniciado com  $T = 20000$  amostras de um sinal iid, uniforme, de média zero e variância unitária, que sofre a distorção de um canal fase mínima<sup>2</sup>  $H(z) = 1 + 0.5z^{-1}$ . A Tabela 7 apresenta a solução MMI obtida para cada modelo de filtro e a solução MMSE equivalente.

Tabela 7 – Filtros FIR e IIR obtidos pelo critério MMI para inverter o canal  $H(z) = 1 + 0.5z^{-1}$ .

		$W(z)$	MSE	Atraso Recuperado
FIR	MMSE	$-0.89 + 0.42z^{-1} - 0.17z^{-2}$	0.0118	0
	MMI	$-0.86 + 0.45z^{-1} - 0.26z^{-2}$	0.0248	0
IIR	MMSE	$1.00/(1.00 + 0.50z^{-1})$	0.0000	0
	MMI	$(1.00 + 0.08z^{-1})/(1.00 + 0.57z^{-1})$	0.0020	0

A Figura 24 exibe o gráfico da IM entre dois atrasos ( $D = 2$ ) de  $y(n)$ , com relação aos parâmetros do filtro FIR; o ponto vermelho indica o seu valor mínimo, o que equivale ao filtro apresentado na segunda linha da Tabela 7, enquanto o ponto verde indica a solução MMSE. Observe que as duas soluções estão relativamente próximas, ademais, a solução MMI aproxima-se do sinal original com atraso zero, o que é o melhor atraso a ser considerado segundo o critério MMSE.

Este fato pode ser melhor compreendido se analisarmos o gráfico do menor valor de MSE de cada solução candidata com respeito ao atraso de sinal, que é exibido na Figura 25. Ambas as soluções são próximas com relação ao mesmo ponto mínimo e nota-se que há outros mínimos associados a diferentes atrasos. O MSE obtido pelo filtro MMI é 0.0248, o que é naturalmente maior, mas ainda assim próximo da solução MMSE.

Considerando a natureza não-supervisionada do critério e que a estimação da informação mútua é somente entre dois atrasos, o resultado observado pode ser definido como bom. Embora seja conhecido que existem técnicas cegas mais eficientes e poderosas para

<sup>2</sup> Um canal de fase mínima é aquele cujos zeros da função de transferência estão localizados dentro do círculo de raio unitário, no plano  $z$  (HAYKIN, 2001).

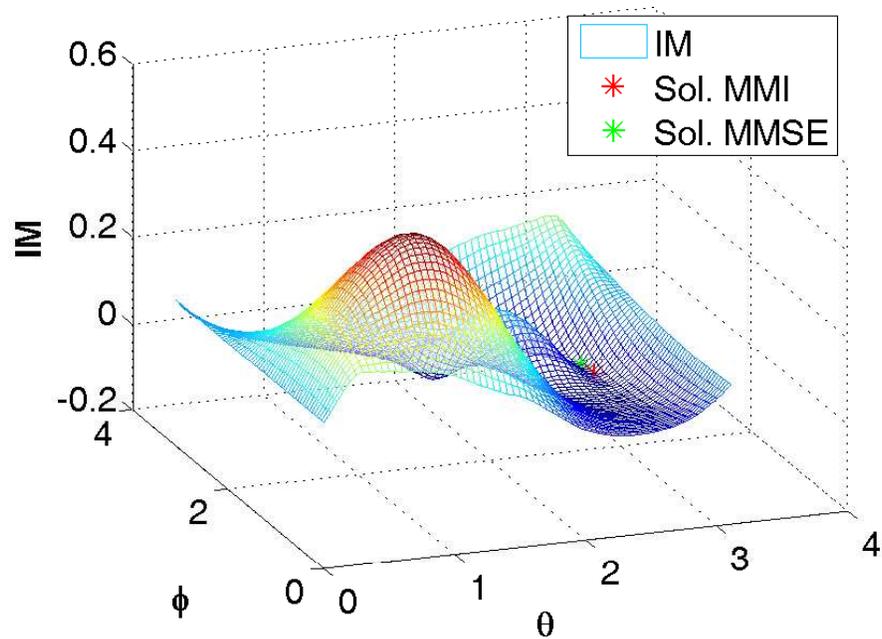


Figura 24 – IM com relação aos parâmetros filtro FIR -  $H(z) = 1 + 0.5z^{-1}$ .

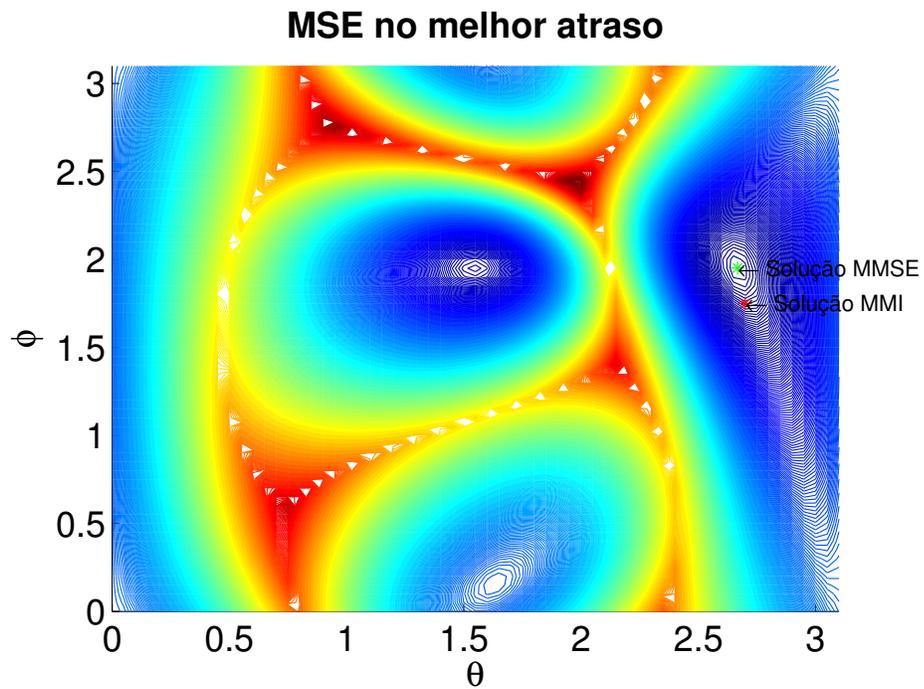


Figura 25 – MSE entre  $y(n)$  e  $s(n-d)$  para o melhor  $d$ , com relação aos parâmetros do filtro FIR -  $H(z) = 1 + 0.5z^{-1}$ .

equalizar / inverter canais lineares de fase mínima (HAYKIN, 2001), o critério que está em análise nesta proposta é interessante porque permitirá, pelas análises a virem na seção 6.3, a inversão simultânea das partes linear e não-linear do sistema de Wiener.

Repetimos o procedimento mais 50 vezes a fim de avaliar se a solução MMI que encontramos é estável. Como a Figura 26 exhibe, em mais de 90% das ocorrências o valor de MSE das soluções propostas se encontra abaixo de 0.2, enquanto que a Figura 27 indica que o valor MMI estimado a cada rodada permanece bem concentrado ao redor da solução ideal, que é zero.

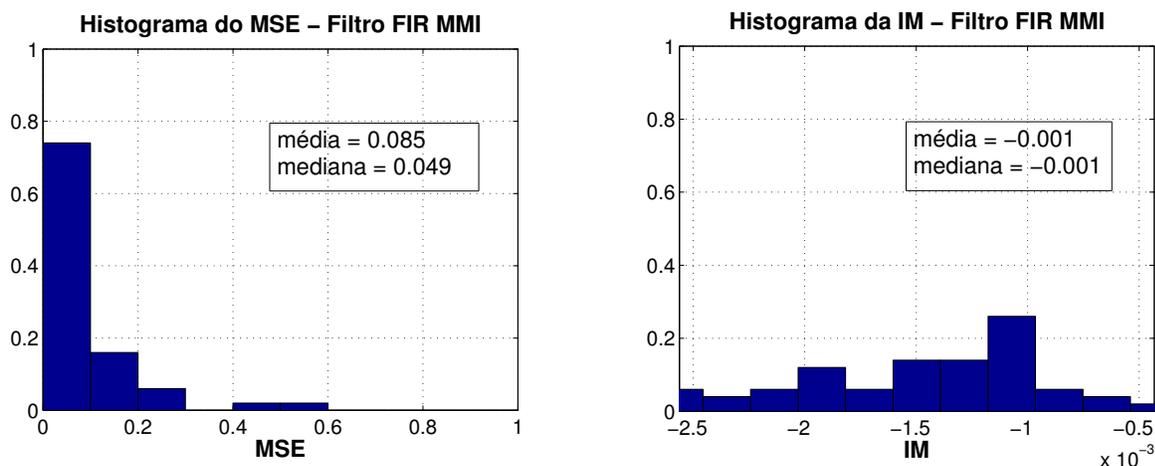


Figura 26 – Histograma do MSE para o filtro FIR MMI -  $H(z) = 1 + 0.5z^{-1}$ .      Figura 27 – Histograma da IM para o filtro FIR MMI -  $H(z) = 1 + 0.5z^{-1}$ .

Analisemos, a seguir, a solução fornecida por um filtro IIR. As linhas 3 e 4 da Tabela 7 demonstram que o filtro IIR MMI determinado pela busca exaustiva é semelhante à solução MMSE, o que resulta na inversão perfeita do canal. A Figura 28 ilustra a superfície da IM com esta estrutura e a Figura 29 apresenta o valor de MSE equivalente: a região vazia dos gráficos corresponde às soluções que divergem, e que por isso foram ignoradas. Observe que a Figura 28 é bem diferente do que foi visto para o cenário com o filtro FIR, as regiões de pontos ótimos são mais fáceis de se identificar devido ao formato de “onda” do gráfico, ao contrário do formato de “vale” que vimos no caso prévio. Ademais, a Figura 29 confirma que o critério MMI obteve uma solução próxima à equivalente pelo critério MMSE.

Analogamente ao caso FIR, repetimos a análise 50 vezes e o filtro IIR acabou apresentando um desempenho mais estável que a configuração FIR, conforme indicam os histogramas da Figura 30, com valor médio de MSE igual a 0.036 e mediana igual a 0.030, enquanto que os valores de IM exibidos na Figura 31 estão concentrados ao redor de zero. Adicionalmente, em 47 ocorrências o melhor atraso (0) foi recuperado. É previsível que o filtro IIR forneça so-

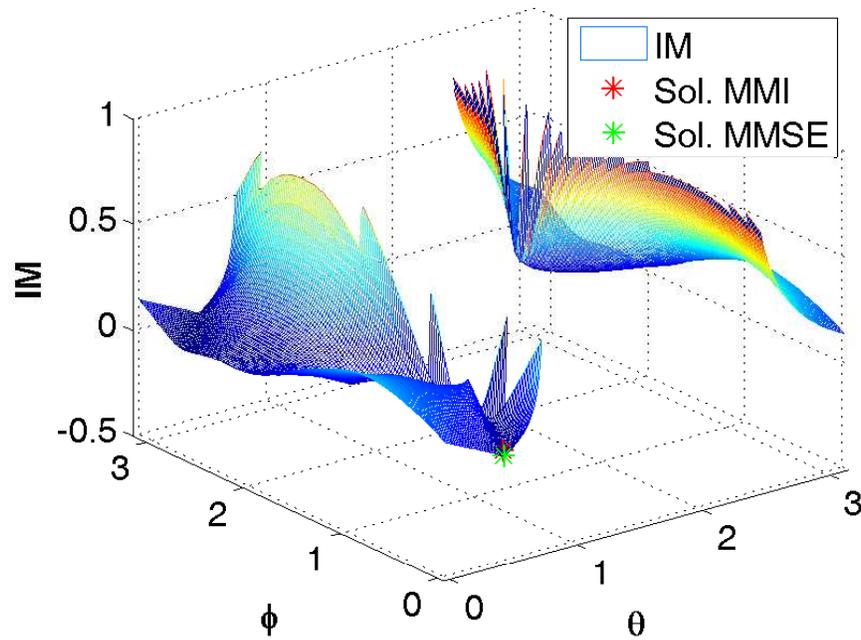


Figura 28 – IM com relação aos parâmetros filtro IIR -  $H(z) = 1 + 0.5z^{-1}$ .

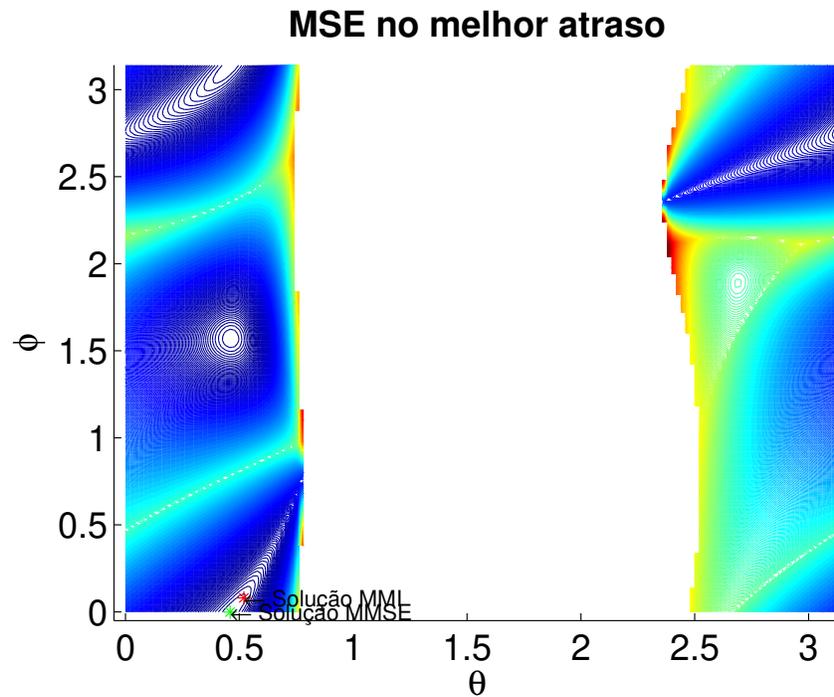


Figura 29 – MSE entre  $y(n)$  e  $s(n-d)$  para o melhor  $d$ , com relação aos parâmetros do filtro IIR -  $H(z) = 1 + 0.5z^{-1}$ .

luções melhores para canais de fase mínima, uma vez que estes são perfeitamente inversíveis. Portanto, é válido analisar, a seguir, o comportamento de ambos os filtros com um canal que não permita a inversão perfeita, a fim de determinar se o critério MMI preserva a capacidade de captura do melhor atraso e consegue obter um filtro semelhante ao derivado pelo critério MMSE.

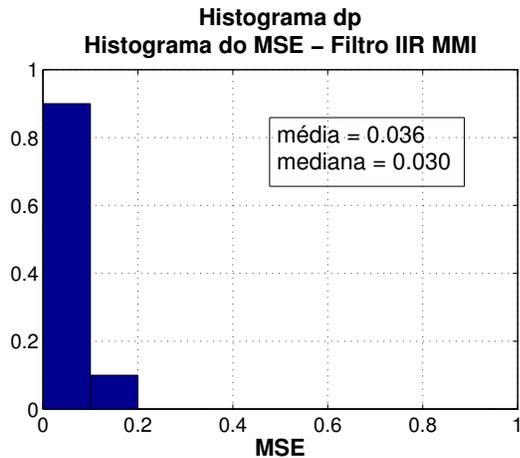


Figura 30 – Histograma do MSE para o filtro IIR MMI -  $H(z) = 1 + 0.5z^{-1}$ .

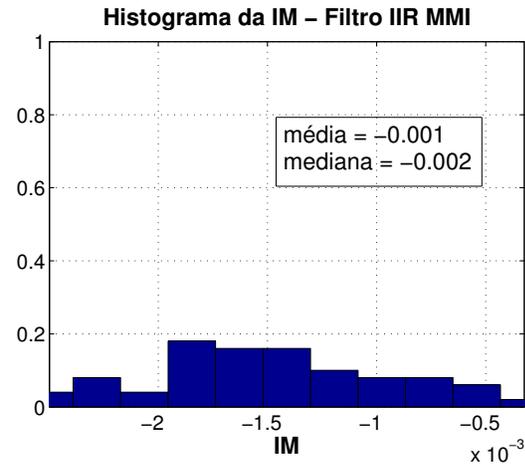


Figura 31 – Histograma da IM para o filtro IIR MMI -  $H(z) = 1 + 0.5z^{-1}$ .

## 6.2.2 Canal de Fase Não-Mínima

Repetimos o teste com a mesma quantidade  $T = 20000$  amostras de um sinal iid, uniforme, com média zero e variância unitária, porém agora distorcido por um canal de fase mista  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ . A Tabela 8 exibe os resultados da busca exaustiva pela solução MMI utilizando as duas estruturas de filtragem possíveis, com os resultados dos filtros MMSE equivalentes.

Tabela 8 – Filtros FIR e IIR obtidos pelo critério MMI para inverter o canal  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

		$W(z)$	MSE	Atraso Recuperado
FIR	MMSE	$-0.33 + 0.91z^{-1} - 0.27z^{-2}$	0.1328	2
	MMI	$-0.42 + 0.84z^{-1} - 0.34z^{-2}$	0.2026	2
IIR	MMSE	$(-0.42 + 0.91z^{-1}) / (1.00 + 0.42z^{-1})$	0.1259	2
	MMI	$(-0.5474 + 0.8369z^{-1}) / (1.00 + 0.55z^{-1})$	0.1763	2

No caso do filtro FIR, as soluções MMI e MMSE são similares, o mesmo atraso foi recuperado e o MSE obtido pela solução MMI é razoavelmente próximo da abordagem su-

pervisionada. Entretanto, ao observar a Figura 32 podemos notar que a região de valores mínimos é um vale, conseqüentemente, a variabilidade na estimação da IM pode levar a uma solução MMI com um valor de MSE consideravelmente alto e possivelmente com a recuperação de um atraso inadequado, dado que pequenas mudanças no espaço dos parâmetros pode aumentar consideravelmente o sinal de erro, conforme a Figura 33 ilustra.

O filtro IIR MMI é levemente superior à abordagem FIR equivalente, com um valor de MSE menor e qualitativamente mais próximo à solução MMSE. A Figura 34 e a Figura 35 exibem o gráfico da função objetivo e o valor de MSE equivalente para cada solução, respectivamente, e novamente podemos identificar que a superfície da primeira não apresenta pontos mínimos proeminentes, o que dificulta a identificação destes. No entanto, ao analisar o gráfico da Figura 35 percebe-se que um certo grau de variabilidade da solução MMI não implicará em aumento tão significativo de MSE como no caso FIR, previamente comentado.

Esta constatação pode ser mais solidamente embasada após a análise estatística de 50 execuções da busca. Por meio dos histogramas do MSE para ambos os modelos de filtro (Figura 36 e Figura 38), percebemos que o filtro IIR MMI produz uma concentração de valores de MSE menores que os do método FIR. Já ao analisar o histograma da IM (Figura 37 e Figura 39), o filtro IIR MMI apresenta um valor absoluto médio um pouco menor que o filtro FIR, mas estatisticamente não podemos concluir que tal diferença é significativa. Todas as execuções de ambos os modelos levaram à correta recuperação do atraso de sinal.

Com estas análises preliminares, apesar do cenário restrito de canais e de tamanho de filtros, o desempenho geral do critério não-supervisionado de mínima informação mútua, para o caso linear, pode ser considerado bom, com soluções capazes de se aproximar do método MMSE, seja com a estrutura FIR ou IIR, especialmente no caso de fase mínima. Adicionalmente, o truncamento considerável do cálculo da IM para somente dois atrasos ainda foi suficiente para obter soluções que capturam o melhor atraso de  $s(n)$ , no sentido MMSE.

## 6.3 Análise Experimental do *Framework*

Com o experimento preliminar realizado podemos considerar que o critério MMI possui potencial para ser agora testado na condição completa do problema de inversão cega de sistemas de Wiener, no contexto do *framework* proposto na seção 6.1. Recorde-se que, além do estimador de IM baseado em histograma, aqui consideraremos também o estimador QMI, cuja implementação foi fornecida por Seth e Principe (2009).

A função objetivo é definida conforme a Equação 6.11 com  $D = 2$ , i.e., a IM é esti-

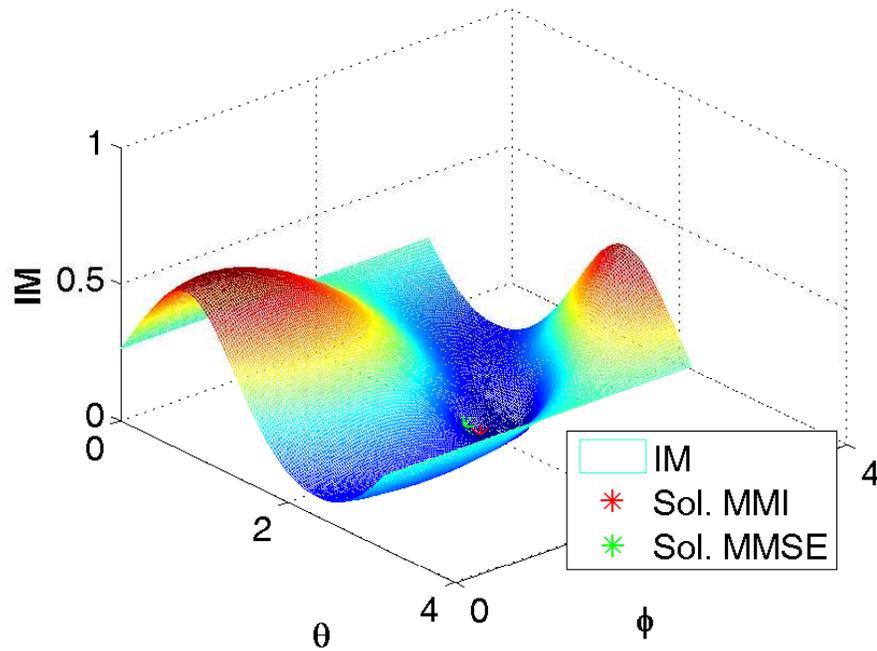


Figura 32 – IM com relação aos parâmetros filtro FIR -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

### MSE no melhor atraso

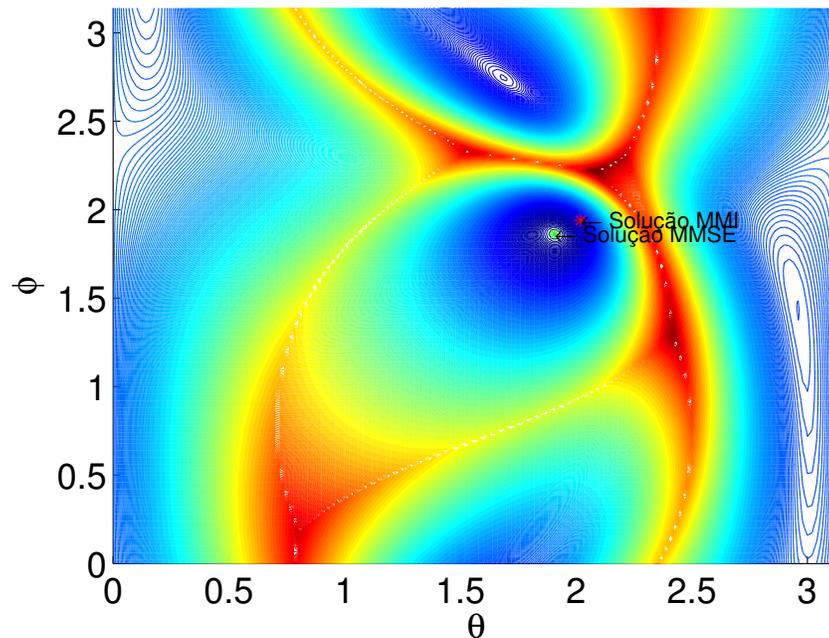


Figura 33 – MSE entre  $y(n)$  e  $s(n-d)$  para o melhor  $d$ , com relação aos parâmetros do filtro FIR -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

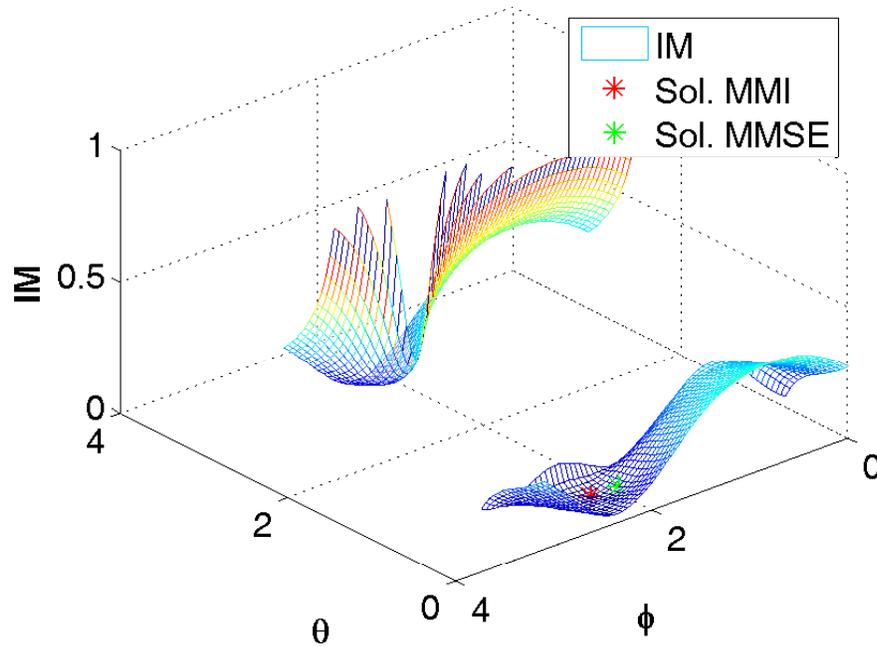


Figura 34 –  $IM$  com relação aos parâmetros filtro IIR -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

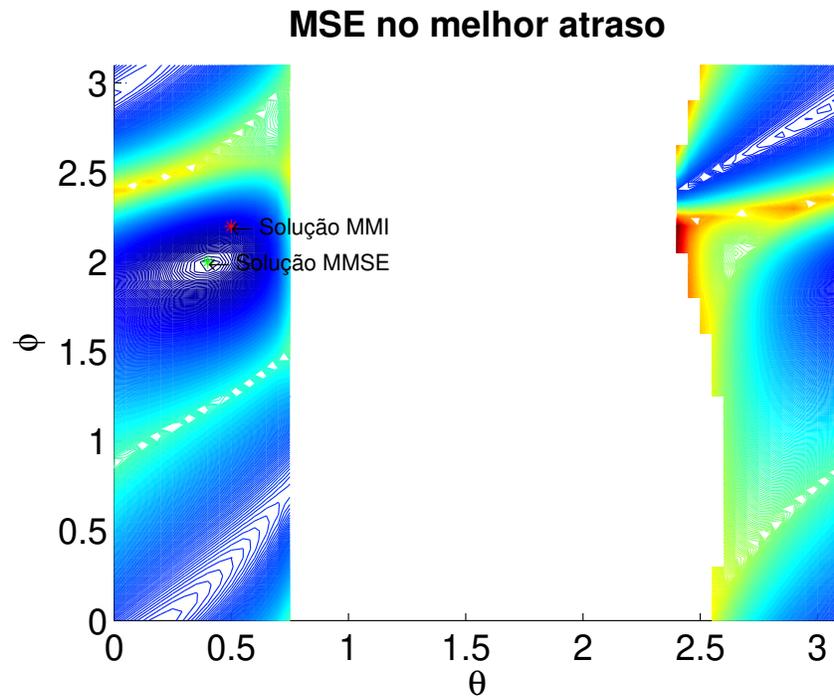


Figura 35 – MSE entre  $y(n)$  e  $s(n-d)$  para o melhor  $d$ , com relação aos parâmetros do filtro IIR -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

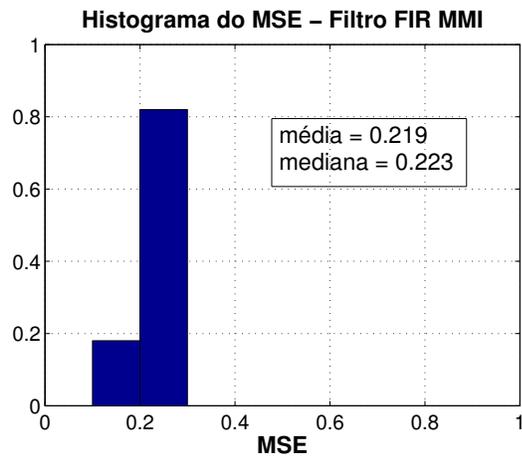


Figura 36 – Histograma do MSE para o filtro FIR MMI -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

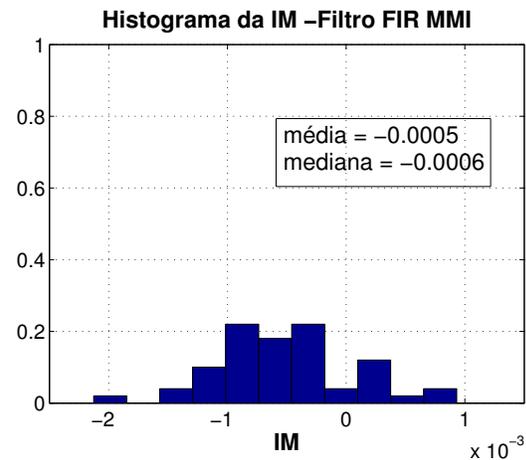


Figura 37 – Histograma da IM para o filtro FIR MMI -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

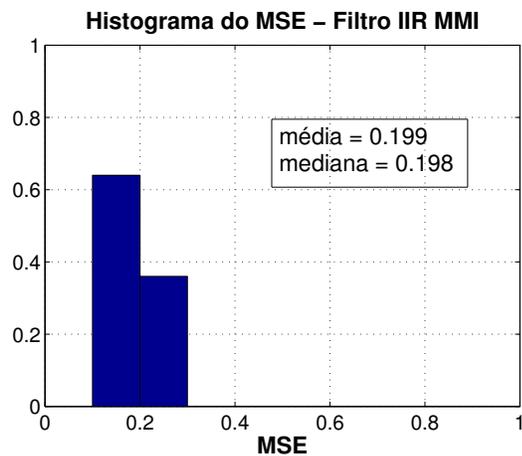


Figura 38 – Histograma do MSE para o filtro IIR MMI -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

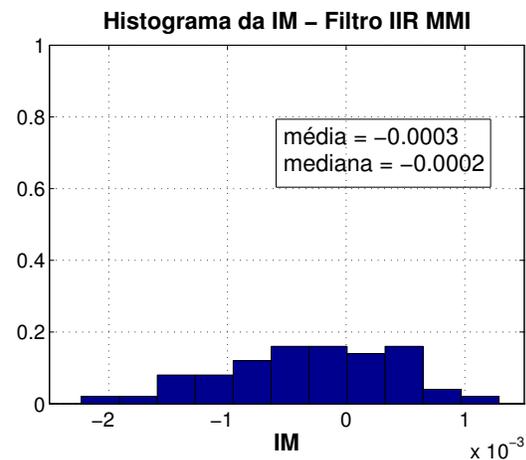


Figura 39 – Histograma da IM para o filtro IIR MMI -  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$ .

mada com respeito a  $y(n)$  e  $y(n-1)$ . A motivação por trás desta considerável simplificação é que a complexidade computacional dos estimadores cresce com o número de sinais atrasados incluídos no cálculo e, conseqüentemente, afeta gravemente o tempo de execução do método. Por isso, esta simplificação é considerada e sua adequação será posta em teste pelos experimentos.

O algoritmo CLONALG foi implementado em MATLAB. Seus parâmetros (vide Pseudocódigo 4.1) foram determinados após um conjunto de testes preliminares, e os valores são fixos para todos os experimentos que seguem:  $n_{AB} = 50$ ,  $maxIT = 200$ ,  $n_{Cl} = 10\%$ ,

$b = 10\%$ ,  $\rho = 4$ . O tamanho da janela da função *kernel* do estimador QMI é  $\sigma = 1$ . Embora possam ser empregados métodos mais sofisticados para determinação deste parâmetro, notamos experimentalmente que o valor definido é apropriado para o algoritmo.

No primeiro cenário, o canal  $h(n)$  é um sistema de fase mínima com resposta ao impulso  $H(z) = 1 + 0.5z^{-1}$ , a distorção não-linear é  $f(v) = \text{sign}(v)\sqrt[3]{|v|}$ , e a sequência iid de entrada  $s(n)$  é gerada conforme duas distribuições diferentes (a primeira sub-gaussiana e a segunda super-gaussiana): uniforme e laplaciana, ambas com média zero e variância unitária.  $T = 2000$  amostras de  $x(n)$  são fornecidas ao algoritmo, que é executado 50 vezes para cada possível configuração de estimador, estrutura de filtragem e sinal original.

De acordo com a notação definida na Figura 23, cada indivíduo possui a parte do polinômio com 3 coeficientes, i.e.,  $k = 5$ , e o comprimento do filtro  $w(n)$  é  $M = 1, Q = 0$  no caso FIR e  $M = 0, Q = 1$  no caso IIR. A medida de desempenho escolhida para o estudo é o MSE medido em decibéis, entre a saída do sistema e o sinal original (para o melhor atraso):

$$\mathcal{P} = 10 \log_{10} \left\{ \frac{1}{T} \sum_{n=1}^T [y(n) - s(n)]^2 \right\}, \quad (6.12)$$

considerando que  $y(n)$  foi apropriadamente atrasado e reescalado. A Tabela 9 exhibe a média e desvio padrão de  $\mathcal{P}$  nas 50 execuções e a taxa de sucesso em recuperar o melhor atraso de sinal, de acordo com o critério de Wiener/MMSE.

Em todas as situações podemos notar que o algoritmo conseguiu 100% de recuperação do atraso correto, o qual é zero para este canal. Ademais, os resultados como um todo indicam que o sinal desejado é recuperado com um grau significativo de precisão. É notável que, em todos os casos, a estrutura IIR foi mais eficiente do ponto de vista de inversão do sistema, o que ilustra as vantagens potenciais de se utilizar um filtro recorrente no sistema de Hammerstein. Este caso foi um teste interessante para este aspecto porque a parte linear do sistema de Wiener pode ser invertida perfeitamente por uma estrutura somente com pólos na função de transferência.

Tabela 9 – Resultados do algoritmo imuno-inspirado no problema de Wiener - Hammerstein — fase mínima. Nas linhas superiores, média e desvio padrão de  $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado.

	Uniforme		Laplaciana	
	FIR	IIR	FIR	IIR
<b>QMI</b>	-12.21 ( $\pm 0.39$ ) 100	-31.10 ( $\pm 5.38$ ) 100	-11.84 ( $\pm 0.60$ ) 100	-23.52 ( $\pm 5.43$ ) 100
<b>Histograma</b>	-11.45 ( $\pm 1.20$ ) 100	-28.14 ( $\pm 6.07$ ) 100	-11.36 ( $\pm 1.06$ ) 100	-17.69 ( $\pm 5.45$ ) 100

Ressalta-se também que o estimador QMI, comparado ao estimador baseado em histograma, levou o método a resultados melhores e com menor variabilidade, e que os melhores desempenhos foram obtidos para o sinal distribuído uniformemente, o que pode ser uma consequência da “cauda” relativamente mais longa da distribuição laplaciana, a qual pode causar dificuldades no processo de estimação devido à potencial existência de *outliers*.

A Figura 40 ilustra um exemplo dos resultados do algoritmo para o sinal original de distribuição uniforme, com o modelo IIR e o estimador QMI.

O próximo cenário expressa uma situação mais complexa, com o sistema de Wiener formado por um canal de fase não-mínima  $H(z) = 1 + 2z^{-1} + 0.8z^{-2}$  e a mesma distorção não-linear utilizada no primeiro cenário. Cada solução candidata tem um tamanho definido por  $k = 5, M = 2, Q = 0$  no caso FIR e  $k = 5, M = 1, Q = 1$  no caso IIR. A Tabela 10 apresenta os resultados: é nítido que o algoritmo, para este caso, não consegue obter o mesmo desempenho do caso com o canal de fase mínima. Os resultados associados ao estimador por histograma são bem distintos quanto à distribuição ser uniforme ou laplaciana, e podemos notar que, apesar de o estimador QMI ser levemente pior que a abordagem por histograma no caso laplaciano, o seu desempenho geral é mais homogêneo.

Tabela 10 – Resultados do algoritmo imuno-inspirado no problema de Wiener - Hammerstein — fase não-mínima. Nas linhas superiores, média e desvio padrão de  $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado.

	Uniforme		Laplaciana	
	FIR	IIR	FIR	IIR
<b>QMI</b>	-5.56 ( $\pm 2.36$ ) 92	-5.28 ( $\pm 2.33$ ) 78	-5.36 ( $\pm 2.52$ ) 90	-5.17 ( $\pm 2.43$ ) 92
<b>Histograma</b>	-2.01 ( $\pm 2.00$ ) 46	-2.78 ( $\pm 1.86$ ) 54	-6.09 ( $\pm 1.62$ ) 94	-6.88 ( $\pm 0.95$ ) 100

Não obstante, percebe-se que a taxa de sucesso na recuperação do atraso é baixa para a combinação sinal uniforme / estimador por histograma e, diferentemente do cenário anterior, a adoção do filtro IIR não levou necessariamente a um melhor desempenho em comparação com a estrutura FIR — uma consequência do caráter de fase não-mínima do estágio linear do sistema de Wiener.

Possivelmente, as *variâncias* dos estimadores de IM empregados são um fator importante no processo de análise do desempenho. Observe a Figura 41, que exhibe os histogramas de  $\mathcal{P}$  para o algoritmo aplicado ao cenário Uniforme / FIR e empregando em cada quadro uma das duas técnicas de estimação da IM, que os resultados apresentam uma diferença significativa. Os gráficos mostram que o *framework* com o estimador baseado em histograma

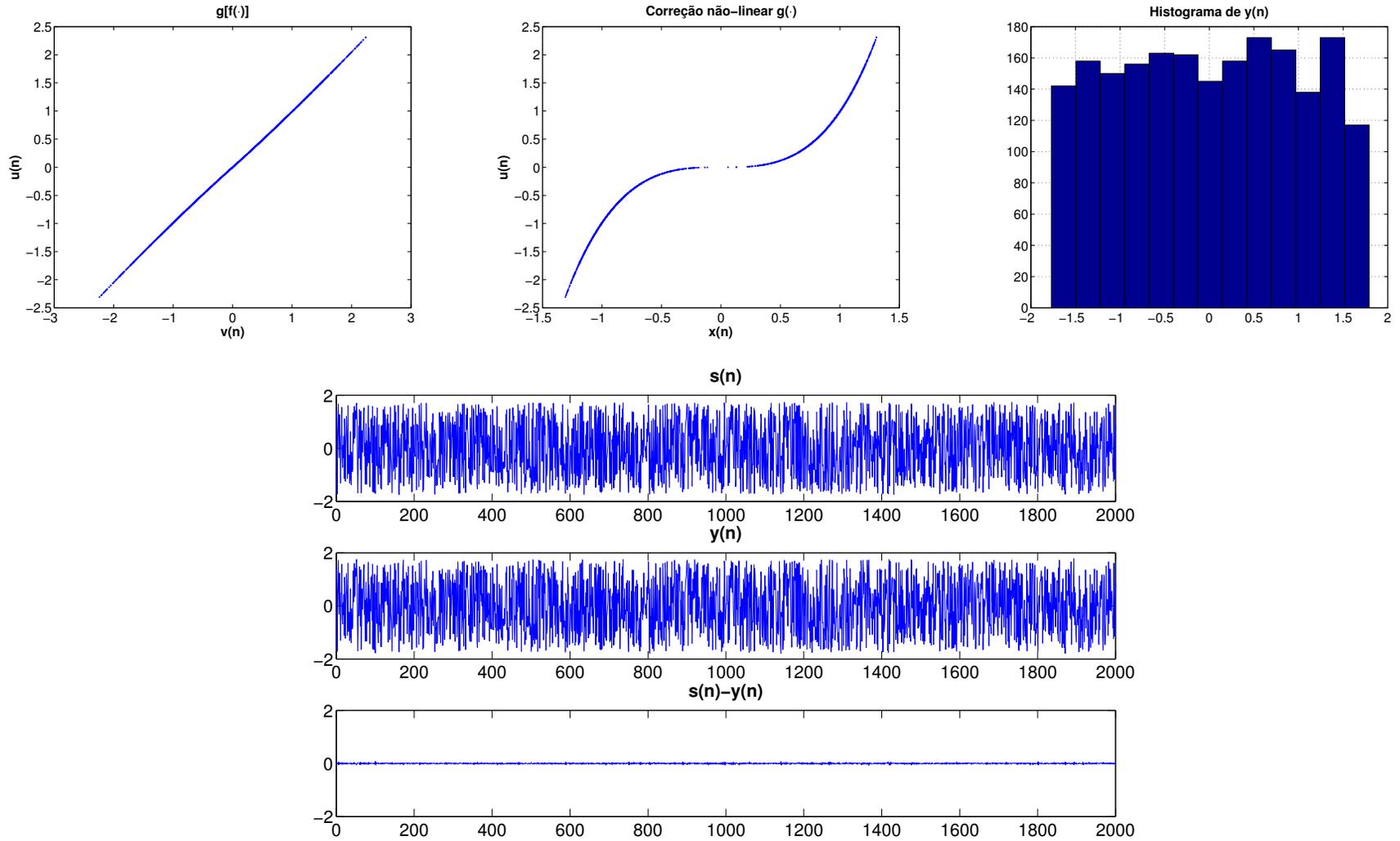


Figura 40 – Experimento com o sinal original uniforme processado pelo canal  $H(z) = 1 + 0.5z^{-1}$  e  $f(v) = \text{sign}(v)\sqrt[3]{|v|}$ , estimador QMI como função custo. A partir do topo à esquerda, em sentido horário:  $g \circ f(\cdot)$ ; correção não-linear  $g(\cdot)$ ; histograma de  $y(n)$ ;  $s(n)$ ,  $y(n)$  e sinal de erro.

possui um comportamento insatisfatório para este cenário, com valores de desempenho ruins e altamente dispersos, enquanto o algoritmo empregando o estimador QMI apresenta um comportamento mais estável, com aproximadamente 80% dos valores concentrados em até -6dB. No entanto, há ainda um certo grau de variabilidade nos resultados, como podemos ver nos valores acima de -3dB presentes.

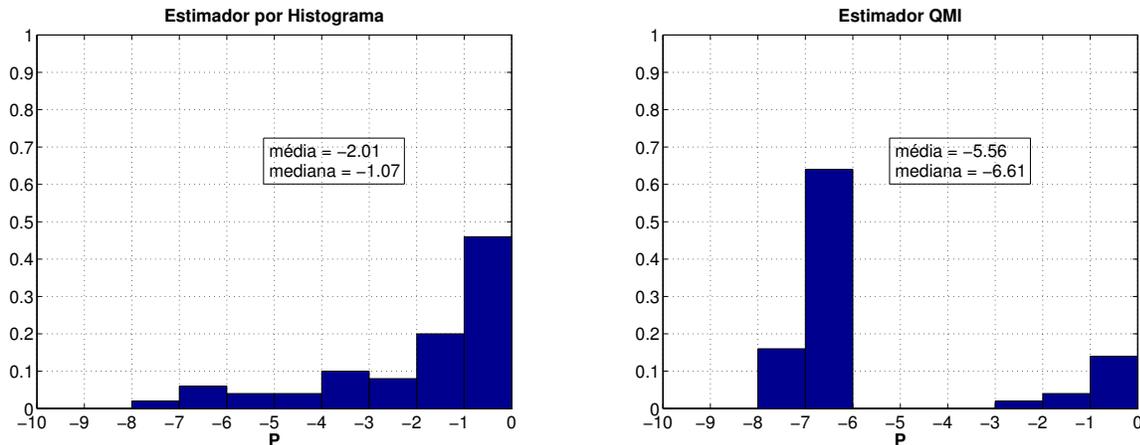


Figura 41 – Histogramas de  $\mathcal{P}$  para o experimento com o canal de fase não-mínima,  $s(n)$  uniforme e filtro FIR.

Para terminarmos de “ligar os pontos”, ou seja, associar a variabilidade da estimação de IM com a variabilidade do sinal  $y(n)$  extraído pelo algoritmo (em outras palavras, a variabilidade de  $\mathcal{P}$ ), ainda é preciso analisar a variância dos estimadores de IM. A Figura 42 ilustra uma simulação Monte Carlo que consiste em calcular a IM entre duas VAs uniformes e entre duas laplacianas, independentes, com ambas as técnicas. Cada ponto calculado representa o desvio padrão sobre 100 realizações independentes com  $T$  amostras para cada variável. Claramente, uma conclusão análoga à derivada para a Figura 41 pode ser feita neste caso, no qual o estimador QMI é mais estável que a abordagem baseada em histograma, com ambas as distribuições.

Com este resultado exclusivamente focado na estimação, podemos então estabelecer a ligação direta com as evidências vistas na Figura 41, conjecturando que uma maior variância na estimação da IM causa uma incoerência entre o ótimo da função objetivo e a solução ideal (no sentido MMSE), e conseqüentemente prejudica a qualidade do *framework*. Tal efeito indesejável pode ser potencialmente mitigado com um aumento no número de amostras disponíveis.

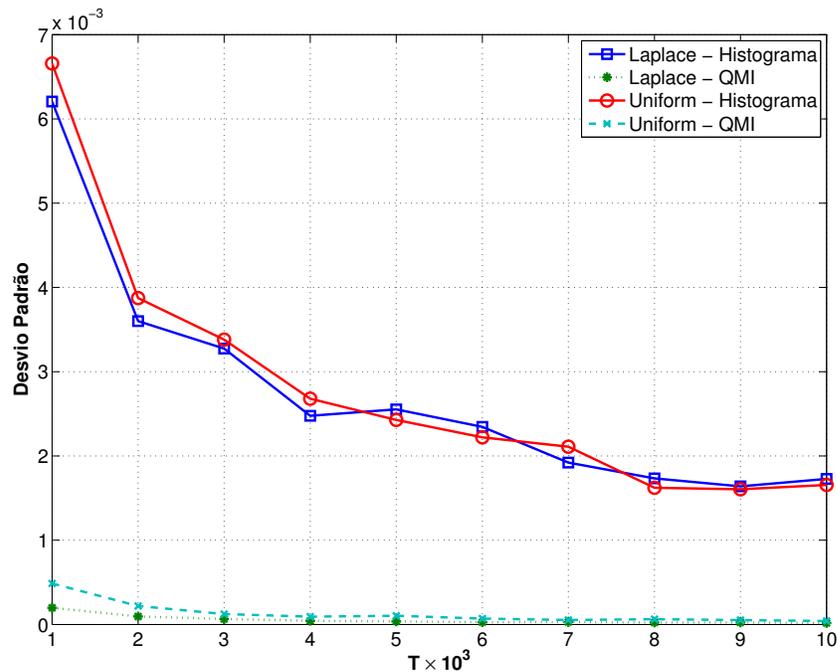


Figura 42 – Desvio padrão da estimativa de IM entre duas VAs independentes.

### 6.3.1 Experimento com uma Segunda Não-Linearidade

Devido aos resultados superiores apresentados pelo algoritmo empregando o estimador QMI, vamos tomar esta configuração como base para comparação com as propostas de Taleb *et al.* (2001) e Rojas *et al.* (2007) em um segundo experimento composto de um sinal iid uniforme e um sistema de Wiener com o estágio linear  $H(z) = 1 + 0.5z^{-1}$  e uma não-linearidade mais forte,  $f(v) = \tanh(3v)$ . Os parâmetros do CLONALG são iguais aos do experimento anterior. Cada indivíduo tem a sua estrutura de dados definida por  $k = 11$ ,  $M = 1$ ,  $Q = 0$  no caso FIR e  $k = 11$ ,  $M = 0$ ,  $Q = 1$  no caso IIR. O algoritmo proposto por Taleb *et al.* é configurado para ter um número máximo de 200 iterações e o filtro FIR com 3 coeficientes. Esta escolha foi feita especificamente para fornecer condições estruturais semelhantes para os três métodos.

A Tabela 11 exhibe os resultados, em 50 execuções de cada técnica. Observe que o método proposto neste capítulo consegue resultados melhores em termos da medida de desempenho e também no aspecto da taxa de recuperação do atraso, com ambas as estruturas de filtragem (FIR e IIR). Por outro lado, o algoritmo CLONALG com a curtose como função custo possui o pior desempenho com os dois tipos de filtro (sendo o filtro IIR não considerado anteriormente por Rojas *et al.*), enquanto o método proposto por Taleb *et al.* ocupa uma posição intermediária.

Tabela 11 – Análise comparativa dos métodos — fase mínima,  $f(v) = \tanh(3v)$ . Nas linhas superiores, média e desvio padrão de  $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado.

	QMI	Curtose	Alg. de Taleb <i>et al.</i>
<b>FIR</b>	-8.00 ( $\pm 1.13$ ) 98	-5.62 ( $\pm 0.15$ ) 90	-6.57 ( $\pm 0.36$ ) 100
<b>IIR</b>	-9.80 ( $\pm 0.34$ ) 100	-2.06 ( $\pm 4.28$ ) 84	N/A

Os resultados indicam que, como esperado, o estimador QMI resulta em uma função custo mais robusta que a curtose e, conseqüentemente, é uma referência de maior respaldo conceitual para o projeto do sistema de Hammerstein. Ademais, para o cenário estudado, o *framework* proposto foi capaz de alcançar soluções melhores que a técnica proposta por Taleb *et al.* — a qual pode ser considerada como a abordagem canônica para o problema em questão —, considerando um número igual de iterações e dimensões similares de estruturas de filtragem. A Figura 43 mostra um exemplo dos resultados do algoritmo, com a estrutura IIR e o estimador QMI: observe que a parametrização polinomial não é capaz de inverter perfeitamente a tangente hiperbólica.

Mantendo a mesma parametrização para o sistema de Wiener, o sinal uniforme para  $s(n)$  e a dimensão dos indivíduos, encerramos a análise experimental comparando o desempenho do *framework* com uma versão sua alterada para empregar um algoritmo genético (HOLLAND, 1962) no lugar do algoritmo CLONALG. O objetivo desta simulação é verificar se o algoritmo imunológico de fato pode trazer benefícios na determinação das soluções do problema de Wiener-Hammerstein. O estimador QMI é mantido como função objetivo e a implementação do algoritmo genético é fornecida pela *toolbox* de computação evolutiva disponível no MATLAB, com os operadores de mutação gaussiana e *crossover* de ponto único. Os dois métodos de busca são executados 50 vezes, para os dois modelos de filtros, com o mesmo número máximo de iterações (200) e de indivíduos (50).

A Tabela 12 exhibe os resultados, que indicam que o emprego da proposta imuno-inspirada determinou melhores soluções. Embora a comparação não cubra um conjunto amplo de cenários e possibilidades de parâmetros para o GA, a diferença nos valores de desempenho alcançados para as duas possibilidades de filtros fornecem um razoável argumento inicial de que a adoção do CLONALG pode ser benéfica para a determinação de soluções melhores.

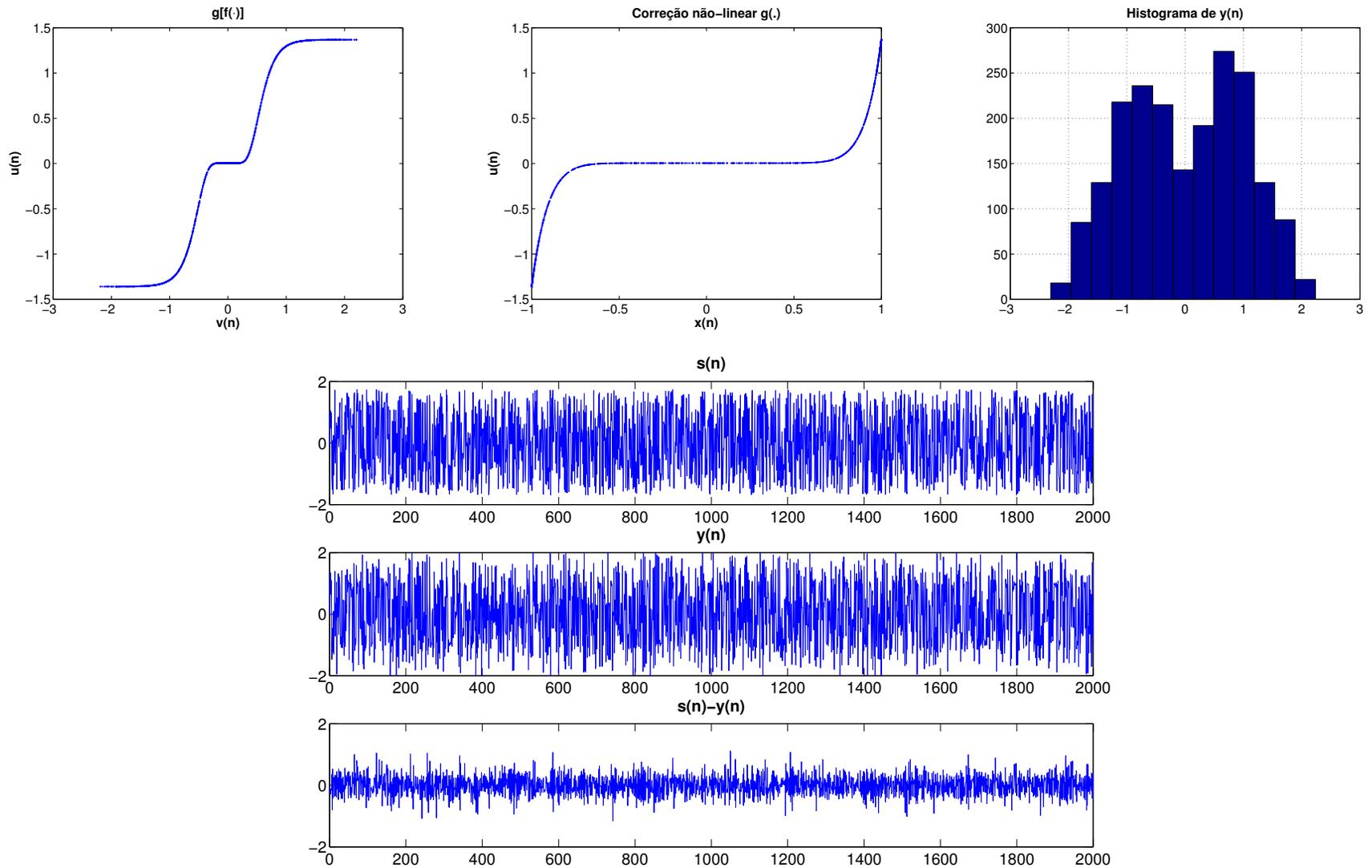


Figura 43 – Experimento com o sinal original uniforme processado pelo canal  $H(z) = 1 + 0.5z^{-1}$  e  $f(v) = \tanh(3v)$ , estimador QMI como função custo. A partir do topo à esquerda, em sentido horário:  $g \circ f(\cdot)$ ; correção não-linear  $g(\cdot)$ ; histograma de  $y(n)$ ;  $s(n)$ ,  $y(n)$  e sinal de erro.

Tabela 12 – Comparativo das estratégias de busca com o estimador QMI — fase mínima,  $f(v) = \tanh(3v)$ . Nas linhas superiores, média e desvio padrão de  $\mathcal{P}$ , em dB; nas linhas inferiores, percentual de acerto na recuperação do atraso adequado.

	CLONALG	GA
<b>FIR</b>	-8.46 ( $\pm 0.57$ ) 100	-6.40 ( $\pm 1.38$ ) 92
<b>IIR</b>	-9.83 ( $\pm 0.35$ ) 100	-6.94 ( $\pm 0.54$ ) 100

## 6.4 Conclusões

Este capítulo apresentou as contribuições da tese ao problema de inversão cega de sistemas de Wiener, que consistem na proposição de um novo algoritmo que se difere, de maneira resumida, em três aspectos das soluções existentes: (i) o método baseia-se no algoritmo imunológico CLONALG para a otimização dos parâmetros, em contraste com os métodos de gradiente ou algoritmos genéticos; (ii) a função custo é definida com estimadores de IM, em oposição ao uso da curtose; e (iii) o estágio linear do sistema de Hammerstein é modelado através de um filtro FIR ou IIR.

Ademais, realizamos um estudo de caráter qualitativo da relação entre as soluções não-supervisionadas obtidas pelo critério MMI e as soluções supervisionadas obtidas pelo critério MMSE, no contexto do problema de desconvolução linear, o qual permitiu averiguar de maneira empírica se critérios baseados na informação mútua possuem os requisitos mínimos para a recuperação de sinais iid.

O relativo sucesso nesta análise preliminar permitiu então a realização dos experimentos de análise do novo algoritmo proposto para o problema de Wiener-Hammerstein, e os resultados indicaram que o *framework* é viável e que é mandatório que o estimador de IM seja escolhido cuidadosamente, a fim de assegurar um desempenho satisfatório do método. Neste contexto, o estimador QMI apresentou a menor variância e, conseqüentemente, produziu os resultados de melhor qualidade em comparação com o emprego do estimador por histograma. Ao final, a análise comparativa mostrou que o novo algoritmo operou de maneira superior às duas metodologias existentes, o que reafirma que o emprego do critério MMI e da técnica imuno-inspirada de busca pode levar a configurações apropriadas do sistema inverso.

O próximo capítulo encerra esta tese, apresentando as considerações finais a respeito de todo o trabalho desenvolvido e as perspectivas futuras de pesquisa e estudo que se abrem.

# Conclusão

Este trabalho de doutorado é o resultado do estudo e proposição de algoritmos de ITL para abordar os problemas não-supervisionados de separação de sinais em corpos finitos e inversão de sistemas de Wiener. Dadas as características dos dois problemas mencionados, desenvolvemos contribuições que se baseiam não somente no paradigma tradicional de ITL — algoritmos de ajuste por gradiente e medidas de informação de Rényi —, mas também incorporam as estratégias imuno-inspiradas de busca, a melhoria de heurísticas de busca específicas, e o emprego de estimadores de TI baseados nas definições de Shannon.

Para levar a termo esta proposta de trabalho, a tese acaba tocando, ao longo de seu conteúdo, em aspectos de diversas disciplinas como Álgebra, Estatística, Teoria da Informação, Processamento de Sinais, Filtragem Adaptativa, Inteligência Computacional, e Complexidade de Algoritmos.

Recapitulando o conteúdo da tese, no Capítulo 1 foram apresentados os aspectos históricos e definições essenciais da Teoria da Informação, e o vínculo posterior dessa disciplina primordial com o desenvolvimento da nova área de Aprendizado de Máquina Baseado na Teoria da Informação, iniciada no contexto da extensão de critérios de algoritmos adaptativos baseados em ESO. Por fim, o capítulo se encerra propondo a ampliação do escopo desta nova área de pesquisa.

Os dois capítulos subsequentes, 2 e 3, são dedicados à apresentação formal dos problemas que são abordados na tese, a separação cega de sinais em corpos finitos e a inversão cega de sistemas de Wiener, respectivamente. Além de descrevê-los matematicamente, apresentou-se uma revisão bibliográfica acerca das principais abordagens para cada problema, acompanhada do levantamento de fatores que geram as oportunidades de contribuições pautadas na tese.

O Capítulo 4 encerrou a parte referente aos Fundamentos apresentando uma introdução à computação imuno-inspirada, em especial aos Sistemas Imunológicos Artificiais e a seus representantes CLONALG e cob-aiNet, algoritmos que são empregados em algumas das contribuições analisadas na parte seguinte.

O Capítulo 5, primeiro da parte de Contribuições, apresentou os trabalhos desenvolvidos para ICA e BSS em corpos finitos, que trazem soluções baseadas no *(i)* algoritmo CLONALG com o critério de mínima IM, no *(ii)* algoritmo cob-aiNet[C] com o critério de mínima entropia, e no *(iii)* novo algoritmo MEXICO-m. As simulações para os três algoritmos

indicaram que os métodos possuem desempenho competitivo frente às técnicas existentes, tornando mais concreta a possibilidade de aplicação em cenários com números maiores de fontes. Ademais, foi proposta a extensão do problema de BSS original para o modelo de misturas convolutivas, incluindo aí a análise da questão estrutural de inversibilidade e de um algoritmo iterativo para extração dos sinais. Portanto, podemos reafirmar que os trabalhos que se desenvolveram possuem o objetivo comum de trazer maior robustez e escalabilidade a ICA em corpos finitos.

Por fim, o Capítulo 6 apresenta a proposta de uma nova metodologia para a inversão cega de sistemas de Wiener, baseada no algoritmo CLONALG e estimadores de IM como função objetivo do problema de otimização dos parâmetros. Além da estruturação deste *framework*, que possui flexibilidade quanto à modelagem do sistema de Hammerstein no que diz respeito à não-linearidade e modelo de filtro (FIR ou IIR), realizamos um estudo qualitativo da relação entre o critério MMI e o critério MMSE para a inversão do canal linear, que forneceu indícios da viabilidade do algoritmo para o cenário não-linear. Uma série de simulações foi realizada, incluindo a comparação com as técnicas estado-da-arte, as quais apontaram que o novo método apresenta desempenho equivalente ou até superior nos cenários avaliados.

## Perspectivas Futuras

Dados os resultados, há uma série de trabalhos futuros possíveis de se realizar, e listamos a seguir alguns deles.

Com respeito ao problema de separação de sinais em corpos de Galois:

- É importante analisar em detalhes as conexões práticas e conceituais entre o processamento de sinais em corpos finitos e a teoria de codificação.
- Uma perspectiva de aplicação real consiste em empregar ICA em corpos finitos como uma ferramenta para realizar análise de fatores de dados discretos. Esta possibilidade pode ter impacto significativo na área de mineração de dados em pelo menos dois domínios de grande impacto: grandes bases de dados armazenadas em formato digital e bases de dados genômicos. No primeiro caso, uma escolha natural seria  $GF(2)$ , mas, no segundo caso, a idéia consistiria em determinar uma matriz capaz de explicar uma massa de dados como uma combinação de fatores independentes em  $GF(3)$  ou  $GF(2)$ , sendo a ordem do corpo definida em concordância com a cardinalidade do conjunto factível de pares de bases nitrogenadas do DNA. Esta possibilidade parece também ser

---

relevante à vista dos recentes esforços (FARIA *et al.*, 2010) para se associar a Teoria da Informação à Genética.

- No caso específico da proposta de ICA com a *cob-aiNet[C]*, é interessante futuramente estender o algoritmo para sinais em corpos não necessariamente primos, aprofundar o estudo da metodologia com diferentes métricas de dissimilaridade, em especial aquelas que operam no espaço fenotípico, e simplificar o processo de determinação dos parâmetros de operação do algoritmo.
- É importante que se analise de maneira mais completa a relação entre o número de amostras disponíveis e o desempenho atingível por métodos operando sobre corpos finitos. Isso também motiva a busca por estimadores mais eficientes e a aplicação sistemática de abordagens já propostas, inclusive o método de estimação de entropia desenvolvido, em caráter de cooperação, durante o trabalho de doutorado (MONTALVAO *et al.*, 2012).
- Em consonância com a estratégia de explorar novas formulações de misturas de sinais em corpos de Galois, como realizado com o caso de misturas convolutivas, é interessante avaliar a possibilidade de lidar com modelos de mistura não-lineares e inversíveis. Uma perspectiva interessante nesse sentido parece ser buscar diretamente permutações das possíveis entradas do sistema misturador com a ajuda de algoritmos de otimização combinatória.
- Finalmente, há a perspectiva de estudar os problemas de separação de sinais e ICA dentro de formulações com outras estruturas algébricas, como as Álgebras de Lie (SAN MARTIN, 2010) e as extensões de corpos (FRALEIGH, 2002).

Com respeito ao problema de inversão de sistemas de Wiener:

- Estender a análise comparativa entre as metodologias existentes, para que contemple um maior número de cenários de canais, sinais de origem e não-linearidades.
- Estudar a aplicação de estimadores de IM de custo computacional menor no *framework* proposto nesta tese, a fim de aumentar o número de sinais atrasados considerados no cálculo da função objetivo. Neste sentido, o trabalho desenvolvido em (MONTALVAO *et al.*, 2013) — que propõe um estimador de entropia *diferencial* baseado em contagem de coincidências — apresenta-se como uma possibilidade promissora, pois é mais simples e menos custoso que métodos de estimação baseados em Parzen, mesmo em dimensões altas, podendo fornecer uma qualidade de estimação comparável a estimadores mais sofisticados.

- É interessante realizar uma análise abrangente do risco de a busca por gradiente levar a soluções localmente ótimas para o problema, análise esta que seria importante aliar a uma avaliação do papel de meta-heurísticas frente a tal limitação.
- É possível estudar novas medidas de independência no papel de função objetivo do problema, como, por exemplo, a correntropia (SANTAMARIA *et al.*, 2006). Para o caso de sinais advindos de sistemas determinísticos, também seria relevante considerar métricas desenvolvidas no estudo de dinâmica não-linear como mapas de retorno (MARWAN *et al.*, 2007).
- A conexão entre o problema de Wiener-Hammerstein e o problema de BSS para misturas PNL precisa ser desenvolvida no sentido de estabelecer algoritmos comuns aos dois métodos e para que se demonstre, de fato, a capacidade de inversão cega do sistema de Wiener por meio da recuperação de independência, o que hoje é apenas conjecturado.

# Referências

- ADALI, T.; HAYKIN, S. (Ed.). *Adaptive Signal Processing: Next Generation Solutions*. Hoboken, USA: Wiley-IEEE Press, 2010. Citado na página 15.
- ADLEMAN, L. M. Molecular Computation of Solutions to Combinatorial Problems. *Science*, v. 266, n. 5187, p. 1021–1024, 1994. Citado na página 27.
- BABAIE-ZADEH, M.; SOLE-CASALS, J.; JUTTEN, C. Blind inversion of Wiener system using a minimization-projection (MP) approach. In: *ICA 2003*. [S.l.: s.n.], 2003. p. 681–686. Citado na página 39.
- BACK, T.; FOGEL, D. B.; MICHALEWICZ, Z. (Ed.). *Evolutionary Computation 1: Basic Algorithms and Operators*. Bristol, UK: Taylor & Francis, 2000. Citado 5 vezes nas páginas 39, 48, 59, 62 e 102.
- BARS, R.; BÈZI, I.; PILIPAR, B.; OJHELYI, B. Nonlinear and Long Range Control of a Distillation Pilot Plant. In: *Proc. Ident. Syst. Parameter Estim. 9th IFAC/IFORS Symp.* [S.l.: s.n.], 1990. p. 848–853. Citado na página 35.
- BEIRLANT, J.; DUDEWICZ, E. J.; GYÖRFI, L.; VAN DER MEULEN, E. C. Nonparametric Entropy Estimation: An Overview. *International Journal of Mathematical and Statistical Sciences*, v. 6, p. 17–40, 1997. Citado na página 18.
- BELL, A. J.; SEJNOWSKI, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, v. 7, n. 6, p. 1129–59, nov. 1995. ISSN 0899-7667. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/7584893>>. Citado 3 vezes nas páginas 16, 20 e 24.
- BENGTSSON, I.; ZYCZKOWSKI, K. *Geometry of Quantum States*. [S.l.: s.n.], 2008. ISBN 052189140X. Citado na página 13.
- BERSINI, H. Revisiting Idiotypic Immune Networks. In: *Proc. of the 7th European Conference on Advances in Artificial Life (ECAL)*. [S.l.: s.n.], 2003. p. 164–174. Citado 2 vezes nas páginas 46 e 49.
- BURNET, F. M. *The Clonal Selection Theory of Acquired Immunity*. [S.l.]: Cambridge University Press, 1959. Citado na página 45.
- CAMPBELL, L. A Coding Theorem and Rényi's Entropy. *Information and Control*, v. 8, n. 4, p. 423–429, ago. 1965. ISSN 00199958. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0019995865903323>>. Citado na página 13.
- CARDOSO, J.-F. Blind Signal Separation: Statistical Principles. *Proceedings of the IEEE*, v. 86, n. 10, p. 2009–2025, 1998. ISSN 0018-9219. Citado na página 24.
- CARLTON, A. G. On the Bias of Information Estimates. *Psychological Bulletin*, v. 71, n. 2, p. 108–109, 1969. Citado na página 63.

- CAVALCANTE, C. C.; CAVALCANTI, F. R. P.; MOTA, J. C. M. A PDF Estimation-Based Blind Criterion for Adaptive Equalization. In: *Proceedings of International Telecommunications Symposium (ITS 2002)*. [S.l.: s.n.], 2002. Citado na página 20.
- CAVALCANTE, C. C.; ROMANO, J. M. T. On the Relationships between MMSE and Information-Theoretic-Based Blind Criterion for Minimum BER Filtering. In: ADALI, T.; JUTTEN, C.; ROMANO, J. M. T.; BARROS, A. K. (Ed.). *Independent Component Analysis and Signal Separation SE - 3*. Springer Berlin Heidelberg, 2009, (Lecture Notes in Computer Science, v. 5441). p. 17–24. ISBN 978-3-642-00598-5. Disponível em: <[http://dx.doi.org/10.1007/978-3-642-00599-2\\_3](http://dx.doi.org/10.1007/978-3-642-00599-2_3)>. Citado na página 20.
- COELHO, G. P.; DE FRANÇA, F. O.; VON ZUBEN, F. J. A Concentration-Based Artificial Immune Network for Combinatorial Optimization. In: *2011 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2011. p. 1242–1249. ISBN 978-1-4244-7834-7. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5949758>>. Citado 2 vezes nas páginas 49 e 63.
- COELHO, G. P.; VON ZUBEN, F. J. A Concentration-Based Artificial Immune Network for Continuous Optimization. In: *2010 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.: s.n.], 2010. p. 1–8. Citado 5 vezes nas páginas xix, 43, 49, 52 e 53.
- COMON, P. Independent Component Analysis, A New Concept? *Signal Processing*, v. 36, n. 3, p. 287–314, abr. 1994. ISSN 01651684. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/0165168494900299>>. Citado na página 24.
- COMON, P.; JUTTEN, C. (Ed.). *Handbook of Blind Source Separation*. Oxford, UK: Academic Press, 2010. Citado 12 vezes nas páginas 16, 21, 23, 24, 29, 30, 32, 38, 40, 58, 65 e 87.
- COSTA, J. P.; LAGRANGE, A.; ARLIAUD, A. Acoustic Echo Cancellation Using Nonlinear Cascade Filters. In: *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*. [S.l.: s.n.], 2003. p. 389–392. Citado na página 35.
- COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. 2. ed. Hoboken, USA: Wiley-Interscience, 2006. (Wiley Series in Telecommunications and Signal Processing, 3). Citado 4 vezes nas páginas 10, 11, 38 e 99.
- DE CASTRO, L. N. *Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais*. 302 p. Tese (Doutorado) — Universidade Estadual de Campinas, 2001. Citado na página 44.
- DE CASTRO, L. N. *Fundamentals of Natural Computing: Basic Concepts, Algorithms and Applications*. Boca Raton, USA: Chapman & Hall/CRC, 2006. Citado 2 vezes nas páginas 32 e 46.
- DE CASTRO, L. N.; TIMMIS, J. An Artificial Immune Network for Multimodal Function Optimization. In: *2002 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.: s.n.], 2002. p. 699–704. Citado na página 49.

- DE CASTRO, L. N.; TIMMIS, J. *Artificial Immune Systems: a New Computational Intelligence Approach*. London, UK: Springer Verlag, 2002. Citado 5 vezes nas páginas 21, 32, 46, 47 e 49.
- DE CASTRO, L. N.; VON ZUBEN, F. J. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, v. 6, n. 3, p. 239–251, jun. 2002. ISSN 1089-778X. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1011539>>. Citado 4 vezes nas páginas 43, 47, 54 e 57.
- DE FRANÇA, F. O.; COELHO, G. P.; CASTRO, P. A. D.; VON ZUBEN, F. J. Conceptual and Practical Aspects of the aiNet Family of Algorithms. *International Journal of Natural Computing Research*, v. 1, n. 1, p. 1–35, jan. 2010. ISSN 1947-928X. Disponível em: <<http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jncr.2010010101>>. Citado na página 49.
- DE FRANÇA, F. O.; COELHO, G. P.; VON ZUBEN, F. J. On the Diversity Mechanisms of opt-aiNet: A Comparative Study With Fitness Sharing. In: *2010 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2010. p. 1–8. ISBN 978-1-4244-6909-3. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5585964>>. Citado na página 47.
- DELFOSSÉ, N.; LOUBATON, P. Adaptive Blind Separation of Independent Sources: A Deflation Approach. *Signal Processing*, v. 45, n. 1, p. 59–83, jul. 1995. ISSN 01651684. Disponível em: <[http://dx.doi.org/10.1016/0165-1684\(95\)00042-C](http://dx.doi.org/10.1016/0165-1684(95)00042-C)>. Citado 2 vezes nas páginas 29 e 62.
- DIAS, T. M.; ATTUX, R.; ROMANO, J. M. T.; SUYAMA, R. Blind Source Separation of Post-Nonlinear Mixtures Using Evolutionary Computation and Gaussianization. *ICA 2009, LNCS 5441*, v. 5441, p. 235–242, 2009. Citado na página 102.
- DIETRICH, F. *Robust Signal Processing for Wireless Communications (Foundations in Signal Processing, Communications and Networking) (No. 2)*. Berlin, Germany: Springer, 2008. Citado na página 31.
- DONOHO, D. On Minimum Entropy Deconvolution. *Applied Time Series Analysis II*, p. 565–608, 1981. Citado 2 vezes nas páginas 16 e 20.
- DUARTE, L.; JUTTEN, C.; MOUSSAOUI, S. A Bayesian Nonlinear Source Separation Method for Smart Ion-Selective Electrode Arrays. *Sensors Journal, IEEE*, v. 9, n. 12, p. 1763–1771, 2009. Citado na página 34.
- DUARTE, L.; SUYAMA, R.; RIVET, B.; ATTUX, R.; ROMANO, J. M. T.; JUTTEN, C. Blind Compensation of Nonlinear Distortions: Application to Source Separation of Post-Nonlinear Mixtures. *IEEE Transactions on Signal Processing*, v. 60, n. 11, p. 5832–5844, 2012. ISSN 1053-587X. Citado na página 34.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. New York, USA: Wiley-Interscience, 2001. Citado na página 15.

- EIBEN, A. E.; SMITH, J. E. *Introduction to Evolutionary Computing*. Berlin, Germany: Springer, 2007. Citado na página 65.
- ERDOGMUS, D.; HILD II, K. E.; PRINCIPE, J. C.; LAZARO, M.; SANTAMARIA, I. Adaptive Blind Deconvolution of Linear Channels Using Renyi's Entropy with Parzen Window Estimation. *IEEE Transactions on Signal Processing*, v. 52, n. 6, p. 1489–1498, 2004. Citado na página 20.
- ERDOGMUS, D.; PRINCIPE, J. C. An Error-Entropy Minimization Algorithm for Supervised Training of Nonlinear Adaptive Systems. *IEEE Transactions on Signal Processing*, v. 50, n. 7, p. 1780–1786, 2002. Citado 2 vezes nas páginas 19 e 21.
- ERDOGMUS, D.; PRINCIPE, J. C. From Linear Adaptive Filtering to Nonlinear Information Processing. *IEEE Signal Processing Magazine*, n. November, p. 14–33, 2006. Citado 2 vezes nas páginas 15 e 21.
- FANTINATO, D. G.; SILVA, D. G.; ATTUX, R.; NEVES, A. O.; MONTALVAO, J. Blind Separation of Convolutional Mixtures Over Galois Fields. In: *2013 IEEE International Workshop on Machine Learning for Signal Processing*. [S.l.]: IEEE, 2013. p. 1–6. Citado na página 85.
- FARIA, L. C. B.; ROCHA, A. S. L.; KLEINSCHMIDT, J. H.; PALAZZO, R.; SILVA-FILHO, M. C. DNA Sequences Generated by BCH Codes Over GF(4). *Electronics Letters*, v. 46, n. 4, p. 203–204, 2010. Citado na página 123.
- FEYNMAN, R. P. *Feynman Lectures on Computation*. Boston, USA: Addison-Wesley Longman Publishing Co., Inc., 1998. Citado na página 8.
- FOGEL, L. J.; OWENS, A. J.; WALSH, M. J. *Artificial Intelligence Through Simulated Evolution*. [S.l.]: John Wiley & Sons, 1966. Citado na página 15.
- FRALEIGH, J. B. *A First Course in Abstract Algebra*. 7. ed. [S.l.]: Pearson, 2002. Citado na página 123.
- GLEICK, J. *A Informação: Uma História, Uma Teoria, Uma Enxurrada*. [S.l.]: Companhia das Letras, 2013. Citado 2 vezes nas páginas 7 e 21.
- GRASSBERGER, P.; PROCACCIA, I. Characterization of Strange Attractors. *Physical Review Letters*, v. 50, n. 5, p. 346–349, 1983. Citado na página 13.
- GUTCH, H. W.; GRUBER, P.; THEIS, F. J. ICA over Finite Fields. In: *ICA 2010, LNCS 6365*. [S.l.: s.n.], 2010. p. 645–652. Citado 6 vezes nas páginas 27, 29, 30, 65, 87 e 88.
- GUTCH, H. W.; GRUBER, P.; YEREDOR, A.; THEIS, F. J. ICA Over Finite Fields - Separability and Algorithms. *Signal Processing*, Elsevier, v. 92, n. 8, p. 1796–1808, ago. 2012. ISSN 01651684. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0165168411003513>>. Citado 6 vezes nas páginas 25, 31, 57, 68, 76 e 80.
- HART, P. E. Moment Distributions in Economics: An Exposition. *Journal of the Royal Statistical Society. Series A (General)*, v. 138, n. 3, p. 423–434, 1975. Citado na página 13.

- HARTLEY, R. V. L. Transmission of Information. *Bell System Technical Journal*, p. 535, 1928. Citado na página 7.
- HAYKIN, S. *Communication Systems*. 3. ed. [S.l.]: John Wiley & Sons, 1994. Citado na página 85.
- HAYKIN, S. *Adaptive Filter Theory*. 4. ed. [S.l.]: Prentice Hall, 2001. Citado 4 vezes nas páginas 14, 33, 104 e 106.
- HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). *The Fourth Paradigm*. Redmond: Microsoft Research, 2009. ISBN 9780982544204. Citado 2 vezes nas páginas 16 e 21.
- HILD II, K. E.; ERDOGMUS, D. Blind Source Separation Using Renyi's Mutual Information. *IEEE Signal Processing Letters*, v. 8, n. 6, p. 174–176, 2001. Citado na página 20.
- HILD II, K. E.; ERDOGMUS, D.; PRINCIPE, J. C. An Analysis of Entropy Estimators for Blind Source Separation. *Signal Processing*, v. 86, n. 1, p. 182–194, jan. 2006. ISSN 01651684. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0165168405001994>>. Citado 2 vezes nas páginas 20 e 21.
- HOLLAND, J. H. Outline for a Logical Theory of Adaptive Systems. *Journal of the ACM*, v. 9, n. 3, p. 297–314, 1962. Citado 6 vezes nas páginas 15, 21, 22, 32, 46 e 118.
- HOSSEINI, S.; DEVILLE, Y. Blind Separation of Linear-Quadratic Mixtures of Real Sources Using a Recurrent Structure. *Artificial Neural Nets Problem Solving Methods*, v. 2687, p. 241–248, 2003. Citado na página 34.
- HOTELLING, H. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, v. 24, n. 6, p. 417–441, 1933. Citado na página 23.
- HU, X.; PAN, Y. *Knowledge Discovery in Bioinformatics: Techniques, Methods, and Applications*. [S.l.]: Wiley-Interscience, 2007. Citado na página 21.
- HUNTER, I. W.; KORENBERG, M. J. The Identification of Nonlinear Biological Systems: Wiener and Hammerstein Cascade Models. *Biological Cybernetics*, v. 55, n. 2-3, p. 135–144, 1986. Citado na página 35.
- HYVARINEN, A.; KARHUNEN, J.; OJA, E. *Independent Component Analysis*. [S.l.]: John Wiley & Sons, 2001. ISBN 0471221317. Citado 6 vezes nas páginas 16, 20, 24, 27, 40 e 58.
- JANEWAY, C. A.; TRAVERS, P.; WALPORT, M.; CAPRA, J. D. *Imunobiologia: O Sistema Imunológico na Saúde e na Doença*. 4. ed. [S.l.]: Artes Médicas, 2000. Citado na página 43.
- JERNE, N. K. Towards a Network Theory of the Immune System. *Annales d'Immunologie*, v. 125, n. 1-2, p. 373–389, 1974. Citado na página 45.
- JUTTEN, C.; KARHUNEN, J. Advances in Blind Source Separation (BSS) and Independent Component Analysis (ICA) for Nonlinear Mixtures. *International Journal of Neural Systems*, v. 14, n. 5, p. 267–92, out. 2004. ISSN 0129-0657. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15593377>>. Citado 2 vezes nas páginas 34 e 40.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by Simulated Annealing. *Science*, v. 220, n. 4598, p. 671–80, maio 1983. ISSN 0036-8075. Disponível em: <<http://www.sciencemag.org/content/220/4598/671.abstract><http://www.ncbi.nlm.nih.gov/pubmed/17813860>>. Citado na página 78.

KOLMOGOROV, A. *Interpolation and extrapolation of stationary random processes*. Santa Monica: Rand Co. (tradução do Russo), 1962. Citado na página 14.

KOZA, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge: MIT Press, 1992. Citado na página 15.

LEON-GARCIA, A. *Probability and Random Processes for Electrical Engineering*. [S.l.]: Addison Wesley, 1994. Citado 4 vezes nas páginas 13, 15, 27 e 39.

LIDL, R.; NIEDERREITER, H. *Finite Fields*. 2. ed. Cambridge: Cambridge University Press, 1997. Citado 2 vezes nas páginas 24 e 25.

LINSKER, R. Self-Organization in a Perceptual Network. *Computer*, v. 21, n. 3, p. 105–117, 1988. Citado na página 16.

LUENBERGER, D. G. *Linear an Nonlinear Programming*. [S.l.]: Kluwer Academic Publishers, 2004. Citado na página 21.

MA, W.; LIM, H.; SZNAIER, M.; CAMPS, O. Risk Adjusted Identification of Wiener Systems. In: *Decision and Control, 2006 45th IEEE Conference on*. [S.l.: s.n.], 2006. p. 2512–2517. Citado na página 35.

MARWAN, N.; CARMEN ROMANO, M.; THIEL, M.; KURTHS, J. Recurrence Plots for the Analysis of Complex Systems. *Physics Reports*, v. 438, n. 5-6, p. 237–329, jan. 2007. ISSN 03701573. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0370157306004066>>. Citado na página 124.

MCCULLOCH, W. S.; PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, v. 5, n. 4, p. 115–133, 1943. Citado na página 15.

MODDEMEIJER, R. On Estimation of Entropy and Mutual Information of Continuous Distributions. *Signal Processing*, v. 16, n. 3, p. 233–248, mar. 1989. ISSN 01651684. Disponível em: <[http://dx.doi.org/10.1016/0165-1684\(89\)90132-1](http://dx.doi.org/10.1016/0165-1684(89)90132-1)>. Citado 2 vezes nas páginas 99 e 103.

MONTALVAO, J.; ATTUX, R.; SILVA, D. G. Simple Entropy Estimator for Small Datasets. *Electronics Letters*, IET Digital Library, v. 48, n. 17, p. 1059–1061, ago. 2012. ISSN 0013-5194. Disponível em: <<http://digital-library.theiet.org/content/journals/10.1049/el.2012.2002>>. Citado na página 123.

MONTALVAO, J.; ATTUX, R.; SILVA, D. G. A Pragmatic Entropy and Differential Entropy Estimator for Small Datasets. *Submetido ao Journal of Communication and Information Systems*, 2013. Citado na página 123.

- MONTALVAO, J.; CANUTO, J. Clustering Ensembles and Space Discretization – A New Regard Toward Diversity and Consensus. *Pattern Recognition Letters*, v. 31, n. 15, p. 2415–2424, nov. 2010. ISSN 01678655. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0167865510002424>>. Citado na página 20.
- NEMOIANU, I.; GRECO, C.; CASTELLA, M.; PESQUET-POPESCU, B.; CAGNAZZO, M. On a Practical Approach to Source Separation Over Finite Fields for Network Coding Applications. In: *IEEE Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2013. p. 1335–1339. ISBN 9781479903566. Citado na página 31.
- NEUMANN, F.; WITT, C. *Bioinspired Computation in Combinatorial Optimization: Algorithms and Their Computational Complexity*. [S.l.]: Springer, 2010. Citado na página 47.
- NYQUIST, H. Certain factors affecting telegraph speed. *Transactions of the American Institute of Electrical Engineers*, v. 43, p. 412–422, 1924. Citado na página 7.
- OPPENHEIM, A. V.; SCHAFER, R. W. *Discrete-Time Signal Processing*. 3. ed. Upper Saddle River, USA: Prentice Hall, 2009. Citado na página 36.
- PANINSKI, L. Estimation of Entropy and Mutual Information. *Neural Computation*, v. 15, n. 6, p. 1191–1253, jun. 2003. ISSN 0899-7667. Disponível em: <<http://www.mitpressjournals.org/doi/abs/10.1162/089976603321780272>>. Citado na página 58.
- PARZEN, E. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, v. 33, n. 3, p. 1065–1076, 1962. Citado 3 vezes nas páginas 13, 17 e 100.
- PRINCIPE, J. C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York, USA: Springer, 2010. ISBN 9781441915696. Citado 8 vezes nas páginas 12, 14, 16, 18, 19, 20, 63 e 100.
- RÉNYI, A. On Measures of Entropy and Information. In: *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.: s.n.], 1961. p. 547–561. Citado na página 13.
- ROJAS, F.; SOLE-CASALS, J.; MONTE-MORENO, E.; PUNTONET, C. G.; PRIETO, A. A Canonical Genetic Algorithm for Blind Inversion of Linear Channels. In: *ICA 2006, LNCS 3889*. [S.l.: s.n.], 2006. p. 238–245. Citado na página 39.
- ROJAS, F.; SOLE-CASALS, J.; PUNTONET, C. G. An Evolutionary Approach for Blind Inversion of Wiener Systems. In: *ICA 2007, LNCS 4666*. [S.l.: s.n.], 2007. p. 260–267. Citado 6 vezes nas páginas 36, 39, 40, 97, 101 e 117.
- ROMANO, J. M. T.; ATTUX, R.; CAVALCANTE, C. C.; SUYAMA, R. *Unsupervised Signal Processing: Channel Equalization and Source Separation*. [S.l.]: CRC Press, 2011. Citado 7 vezes nas páginas 15, 16, 34, 86, 90, 102 e 103.
- RUDOLPH, G. Convergence Analysis of Canonical Genetic Algorithms. *IEEE Transactions on Neural Networks*, v. 5, n. 1, p. 96–101, 1994. Citado na página 47.

- RUDOLPH, G. Convergence of evolutionary algorithms in general search spaces. In: *Proceedings of IEEE International Conference on Evolutionary Computation*. IEEE, 1996. p. 50–54. ISBN 0-7803-2902-3. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=542332http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=542332>>. Citado na página 47.
- SAKK, E.; SMALL, S. The Fourier Convolution Theorem over Finite Fields: Extensions of Its Application to Error Control Coding. In: *Fourier Transform Applications*. [S.l.]: InTech, 2003. cap. 9, p. 300. ISBN 978-953-51-0518-3. Citado na página 90.
- SAN MARTIN, L. A. B. *Álgebras de Lie*. 2. ed. [S.l.]: Editora da Unicamp, 2010. Citado na página 123.
- SANTAMARIA, I.; ERDOGMUS, D.; PRINCIPE, J. C. Entropy Minimization for Supervised Digital Communications Channel Equalization. *IEEE Transactions on Signal Processing*, v. 50, n. 5, p. 1184–1192, 2002. Citado na página 19.
- SANTAMARIA, I.; POKHAREL, P.; PRINCIPE, J. Generalized Correlation Function: Definition, Properties, and Application to Blind Equalization. *IEEE Transactions on Signal Processing*, v. 54, n. 6, p. 2187–2197, jun. 2006. ISSN 1053-587X. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1634815>>. Citado 3 vezes nas páginas 19, 20 e 124.
- SCHÜRMAN, T. Bias analysis in entropy estimation. *Journal of Physics A: Mathematical and General*, v. 37, n. 27, p. L295–L301, jul. 2004. ISSN 0305-4470. Disponível em: <<http://stacks.iop.org/0305-4470/37/i=27/a=L02?key=crossref.270d8e62aad1ebe6fd9bb6089cd9d71e>>. Citado na página 63.
- SETH, S.; PRINCIPE, J. C. On speeding up computation in information theoretic learning. *2009 International Joint Conference on Neural Networks*, Ieee, p. 2883–2887, jun. 2009. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5178933>>. Citado 2 vezes nas páginas 101 e 109.
- SHANNON, C. E. A Mathematical Theory of Communication. *The Bell System Technical Journal*, v. 27, p. 379–423, 623–656, 1948. Citado 4 vezes nas páginas 7, 8, 10 e 11.
- SILVA, D. G.; ATTUX, R.; NADALIN, E. Z.; DUARTE, L. T.; SUYAMA, R. An Immune-Inspired Information-Theoretic Approach to the Problem of ICA Over a Galois Field. In: *2011 IEEE Information Theory Workshop*. IEEE, 2011. p. 618–622. ISBN 978-1-4577-0437-6. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6089571>>. Citado na página 57.
- SILVA, D. G.; NADALIN, E. Z.; ATTUX, R.; MONTALVAO, J. A Modified Version of the MEXICO Algorithm for Performing ICA Over Galois Fields. In: *2012 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE, 2012. p. 1–6. ISBN 978-1-4673-1026-0. ISSN 1551-2541. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6349741http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6349741>>. Citado na página 76.

SILVA, D. G.; NADALIN, E. Z.; COELHO, G. P.; DUARTE, L. T.; SUYAMA, R.; ATTUX, R.; VON ZUBEN, F. J.; MONTALVAO, J. A Michigan-like Immune-Inspired Framework for Performing Independent Component Analysis over Galois Fields of Prime Order. *Signal Processing*, v. 96, p. 153–163, 2014. Citado na página 61.

SILVA, D. G.; NADALIN, E. Z.; MONTALVAO, J.; ATTUX, R. The Modified MEXICO for ICA Over Finite Fields. *Signal Processing*, v. 93, n. 9, p. 2525–2528, set. 2013. ISSN 01651684. Disponível em: <<http://dx.doi.org/10.1016/j.sigpro.2013.03.021>>. Citado na página 76.

SILVA, L. M.; FELGUEIRAS, C. S.; ALEXANDRE, L. A.; SÁ, J. M. de. Error Entropy in Classification Problems: a Univariate Data Analysis. *Neural computation*, v. 18, n. 9, p. 2036–61, set. 2006. ISSN 0899-7667. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16846386>>. Citado 2 vezes nas páginas 19 e 21.

SILVA, L. M.; SÁ, J. M. de; ALEXANDRE, L. A. The MEE Principle in Data Classification: a Perceptron-Based Analysis. *Neural computation*, v. 22, n. 10, p. 2698–728, out. 2010. ISSN 1530-888X. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20569178>>. Citado na página 19.

SINGH, A.; PRINCIPE, J. C. Using Correntropy as a cost function in linear adaptive filters. *2009 International Joint Conference on Neural Networks, Ieee*, n. 1, p. 2950–2955, jun. 2009. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5178823>>. Citado na página 19.

SOLE-CASALS, J.; CAIAFA, C. F. A Fast Gradient Approximation for Nonlinear Blind Signal Processing. *Cognitive Computation*, n. 1063, out. 2012. ISSN 1866-9956. Disponível em: <<http://link.springer.com/10.1007/s12559-012-9192-x>>. Citado na página 39.

SOLE-CASALS, J.; JUTTEN, C.; PHAM, D. Fast Approximation of Nonlinearities for Improving Inversion Algorithms of PNL Mixtures and Wiener Systems. *Signal Processing*, v. 85, n. 9, p. 1780–1786, set. 2005. ISSN 01651684. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0165168405001088>>. Citado na página 39.

SOLE-CASALS, J.; JUTTEN, C.; TALEB, A. Parametric approach to blind deconvolution of nonlinear channels. *Neurocomputing*, v. 48, n. 1-4, p. 339–355, 2002. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231201006518>>. Citado 2 vezes nas páginas 36 e 38.

STREHL, A.; GHOSH, J. Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, v. 3, p. 583–617, 2002. Citado na página 20.

TALEB, A.; JUTTEN, C. Source Separation in Post-Nonlinear Mixtures. *Signal Processing, IEEE Transactions on*, v. 47, n. 10, p. 2807–2820, 1999. Citado 3 vezes nas páginas 34, 37 e 39.

TALEB, A.; SOLE-CASALS, J.; JUTTEN, C. Quasi-Nonparametric Blind Inversion of Wiener Systems. *IEEE Transactions on Signal Processing*, v. 49, n. 5, p. 917–924, 2001. Citado 7 vezes nas páginas 20, 34, 36, 37, 38, 97 e 117.

- TIZARD, I. R. *Immunology An Introduction*. 4. ed. [S.l.]: Saunders College Publishing, 1995. Citado na página 43.
- TOPCHY, A.; JAIN, A. K.; PUNCH, W. Clustering Ensembles: Models of Consensus and Weak Partitions. *IEEE transactions on pattern analysis and machine intelligence*, v. 27, n. 12, p. 1866–81, dez. 2005. ISSN 0162-8828. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16355656>>. Citado 2 vezes nas páginas 20 e 21.
- TSENG, L.-Y.; CHEN, C. Multiple Trajectory Search for Large Scale Global Optimization. In: *2008 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.: s.n.], 2008. p. 3052–3059. Citado na página 53.
- VRINS, F.; PHAM, D.-T.; VERLEYSSEN, M. Is the General Form of Renyi's Entropy a Contrast for Source Separation? In: *ICA 2007, LNCS 4666*. [S.l.: s.n.], 2007. v. 1, p. 129–136. Citado na página 21.
- WADA, C.; CONSOLARO, D. M.; FERRARI, R.; SUYAMA, R.; ATTUX, R.; VON ZUBEN, F. J. Nonlinear Blind Source Deconvolution Using Recurrent Prediction-Error Filters and an Artificial Immune System. *ICA 2009, LNCS 5441*, v. 5441, p. 371–378, 2009. Citado na página 102.
- WALTERS-WILLIAMS, J.; LI, Y. Estimation of mutual information: a survey. *Rough Sets and Knowledge Technology*, v. 5589, p. 389–396, 2009. Disponível em: <<http://www.springerlink.com/index/p747r77451013704.pdf>>. Citado na página 99.
- WANG, Z.; BOVIK, A. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, v. 26, n. 1, p. 98–117, jan. 2009. ISSN 1053-5888. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4775883>>. Citado na página 15.
- WATERHOUSE, W. C. How often do determinants over finite fields vanish? *Discrete Mathematics*, v. 65, n. 1, p. 103–104, maio 1987. ISSN 0012365X. Disponível em: <[http://dx.doi.org/10.1016/0012-365X\(87\)90217-2](http://dx.doi.org/10.1016/0012-365X(87)90217-2)>. Citado 2 vezes nas páginas 28 e 67.
- WIENER, N. *Nonlinear Problems in Random Theory*. Boston: MIT, 1958. Citado na página 14.
- WIGGINS, R. Minimum entropy deconvolution. *Geoexploration*, v. 16, n. 1-2, p. 21–35, 1978. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0016714278900054>>. Citado 3 vezes nas páginas 16, 20 e 21.
- YEREDOR, A. ICA in Boolean XOR Mixtures. In: *ICA 2007, LNCS 4666*. [s.n.], 2007. p. 827–835. Disponível em: <<http://www.springerlink.com/index/DG4UKW86P1510K16.pdf>>. Citado 5 vezes nas páginas 27, 28, 29, 30 e 87.
- YEREDOR, A. Independent Component Analysis Over Galois Fields of Prime Order. *IEEE Transactions on Information Theory*, v. 57, n. 8, p. 5342–5359, 2011. Citado 11 vezes nas páginas 28, 29, 30, 31, 59, 65, 71, 76, 77, 78 e 80.

YEREDOR, A. *MATLAB Code for ICA Over GF(P), AMERICA and MEXICO*. 2011. Citado 3 vezes nas páginas 68, 78 e 80.

ZADEH, L. A. Fuzzy Sets. *Inf. Control*, v. 8, p. 338–353, 1965. Citado na página 15.

ZHANG, K.; CHAN, L.-W. Practical Method for Blind Inversion of Wiener Systems. In: *2004 IEEE International Joint Conference on Neural Networks IJCNN-04*. IEEE, 2004. p. 2163–2168 vol.3. ISBN 0-7803-8359-1. Disponível em: <[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1380954http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1380954](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1380954http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1380954)>. Citado na página 39.